



GuideMyGrades

— YOUR GOALS. OUR EXPERTISE. —

EDEXCEL INTERNATIONAL A LEVEL MATHS: QUESTIONS BY TOPIC

Correlation & Regression (S1.3)

International A Level Mathematics · S1: Statistics 1 (WST01)

17 questions · 219 marks · Mark schemes included

Past papers October 2020 - June 2025 · 1 sub-topic

 guidemygrades.com

 @Guidemygrades

 @Guidemygrades

 @Guidemygrades

 info@guidemygrades.com

 Piccadilly Business Centre, Unit C, Aldow Enterprise Park
Blackett Street, Manchester, M12 6AE

HOW TO USE THIS BOOKLET

Questions are grouped by sub-topic and run from the lowest to the highest mark tariff, then oldest paper first.

Click any row below to jump straight to that sub-topic or question.

On every question page, click ANSWER in the top-right corner to jump to its mark scheme; < QUESTION brings you back.

A gold bar in the left margin marks exactly where the question sits on the original page.

Diagrams, graphs, grids and tables are printed exactly as they appear in the original papers.

Each question is filed under its main topic; longer questions may also test skills from other topics.

QUESTION	SOURCE	MARKS	PAGE
S1.3a CORRELATION & REGRESSION		17 Qs · 219 marks	2
1. Q2	IAL S1 (WST01) January 2022	6	3
2. Q2	IAL S1 (WST01) October 2024	11	7
3. Q2	IAL S1 (WST01) October 2021	12	11
4. Q6	IAL S1 (WST01) October 2023	12	15
5. Q2	IAL S1 (WST01) January 2024	12	18
6. Q4	IAL S1 (WST01) January 2024	12	22
7. Q6	IAL S1 (WST01) January 2025	12	26
8. Q6	IAL S1 (WST01) January 2022	13	30
9. Q2	IAL S1 (WST01) October 2022	13	34
10. Q2	IAL S1 (WST01) June 2023	13	38
11. Q4	IAL S1 (WST01) June 2024	13	42
12. Q2	IAL S1 (WST01) June 2022	14	46
13. Q6	IAL S1 (WST01) January 2023	14	50
14. Q3	IAL S1 (WST01) June 2025	14	55
15. Q5	IAL S1 (WST01) October 2020	15	59
16. Q6	IAL S1 (WST01) June 2021	16	63
17. Q5	IAL S1 (WST01) January 2021	17	66



QUESTIONS

17 questions · 219 marks · 1 sub-topic

SUB-TOPIC S1.3a

CORRELATION & REGRESSION

Sub-topic S1.3a · 17 questions · 219 marks · lowest tariff first



Question 2 continued

Leave
blank

Handwriting practice area with horizontal lines.



Question 2 continued

Leave
blank

Lined area for writing the answer to Question 2.

Question 2 continued

Leave
blank

Lined area for writing the answer to Question 2.

(Total 6 marks)

Q2







(Total for Question 2 is 11 marks)



Question 2 continued

Leave
blank

Lined area for writing the answer to Question 2.



Question 2 continued

Leave
blank

Lined area for writing the answer to Question 2 continued.



Question 2 continued

Leave
blank

Lined area for writing the answer to Question 2.

(Total 12 marks)

Q2



- | | Regression equation |
|------------|---------------------|
| y on x | $y = 1.4x + 1.5$ |
| x on y | $x = 1.2 + 0.2y$ |

- (4)

$$\sum xy = \frac{6961}{60}$$

- (4)

- (4)



Question 6 continued

Lined area for writing the answer to Question 6.

(Total for Question 6 is 12 marks)

TOTAL FOR PAPER IS 75 MARKS



2. The average minimum monthly temperature, x degrees Fahrenheit ($^{\circ}\text{F}$), and the average maximum monthly temperature, y degrees Fahrenheit ($^{\circ}\text{F}$), in Kolkata were recorded for 12 months.

Some of the summary statistics are given below.

$$\sum x = 862 \quad \sum x^2 = 62\,802 \quad S_{yy} = 413.67 \quad S_{xy} = 512.67 \quad n = 12$$

- (a) (i) Calculate the mean of the 12 values of the average **minimum** monthly temperature.
- (ii) Show that the standard deviation of the 12 values of the average **minimum** monthly temperature is 8.57°F to 3 significant figures.

(3)

- (b) Calculate the product moment correlation coefficient between x and y

(3)

For comparative purposes with a UK city, it was necessary to convert the temperatures from degrees Fahrenheit ($^{\circ}\text{F}$) to degrees Celsius ($^{\circ}\text{C}$).

The formula used was

$$c = \frac{5}{9}(f - 32)$$

where f is the temperature in $^{\circ}\text{F}$ and c is the temperature in $^{\circ}\text{C}$

- (c) Use this formula and the values from part (a) to calculate, in $^{\circ}\text{C}$, the mean and the standard deviation of the 12 values of the average **minimum** monthly temperature in Kolkata.

Give your answers to 3 significant figures.

(4)

Given that

- u is the equivalent temperature in $^{\circ}\text{C}$ of x
- v is the equivalent temperature in $^{\circ}\text{C}$ of y

- (d) state, giving a reason, the product moment correlation coefficient between u and v

(2)



Question 2 continued

Lined area for writing the answer to Question 2.



(Total for Question 2 is 12 marks)



4. A French test and a Spanish test were sat by 11 students.

The table below shows their marks.

Student	A	B	C	D	E	F	G	H	I	J	K
French mark (f)	24	30	32	32	36	36	40	44	50	60	68
Spanish mark (s)	16	90	24	28	32	36	38	44	48	48	68

Greg says that if these points were plotted on a scatter diagram, then the point (30, 90) would be an outlier because 90 is an outlier for the Spanish marks.

An outlier is defined as a value that is

$$\text{greater than } Q_3 + 1.5 \times (Q_3 - Q_1) \text{ or smaller than } Q_1 - 1.5 \times (Q_3 - Q_1)$$

- (a) Show that 90 is an outlier for the Spanish marks.

(3)

Ignoring the point (30, 90), Greg calculated the following summary statistics.

$$\sum f = 422 \quad \sum s = 382 \quad S_{ff} = 1667.6 \quad S_{fs} = 1735.6$$

- (b) Use these summary statistics to show that the equation of the least squares regression line of s on f for the remaining 10 students is

$$s = -5.72 + 1.04f$$

where the values of the intercept and gradient are given to 3 significant figures. You must show your working.

(3)

- (c) Give an interpretation of the gradient of the regression line.

(1)

Two further students sat the French test but missed the Spanish test.

- (d) Using the equation given in part (b), estimate

- a Spanish mark for the student who scored 55 marks in their French test,
- a Spanish mark for the student who scored 18 marks in their French test.

(3)

- (e) State, giving a reason, which of the two estimates found in part (d) would be the more reliable estimate.

(2)





Question 4 continued

Lined area for writing the answer to Question 4.

(Total for Question 4 is 12 marks)



6. As a part of a selection process, applicants for a television game show have to take a test, and then complete a task as quickly as possible. The test score, w , and the time taken to complete the task, t minutes, are recorded for each applicant.

The summary statistics below represent the data for a random sample of 30 applicants.

$$S_{wt} = -1648.83 \quad S_{ww} = 2396.97 \quad \sum w = 839 \quad \sum t = 635 \quad \sum t^2 = 14837$$

- (a) Show that the product moment correlation coefficient for these data is -0.901 to 3 significant figures.

(2)

A scatter diagram of t against w is plotted for these data.

- (b) State **two** features of the graph that you would expect to see, given the correlation coefficient in part (a)

(2)

- (c) Calculate the equation of the regression line of t on w in the form

$$t = a + bw$$

Give the values of the constants a and b to 3 significant figures.

(4)

- (d) Give an interpretation of the gradient of this regression line.

(1)

The test score, w , was a score out of 50

The manager of the selection process now decides to double all the values of w to make them into percentages.

The manager then recalculates the product moment correlation coefficient and the equation of the regression line.

- (e) State, for **each** of the following, whether the value would increase **or** decrease **or** stay the same as a result of applying this change.

- (i) The product moment correlation coefficient
- (ii) The magnitude of the gradient of the regression line
- (iii) The t intercept of the regression line

(3)

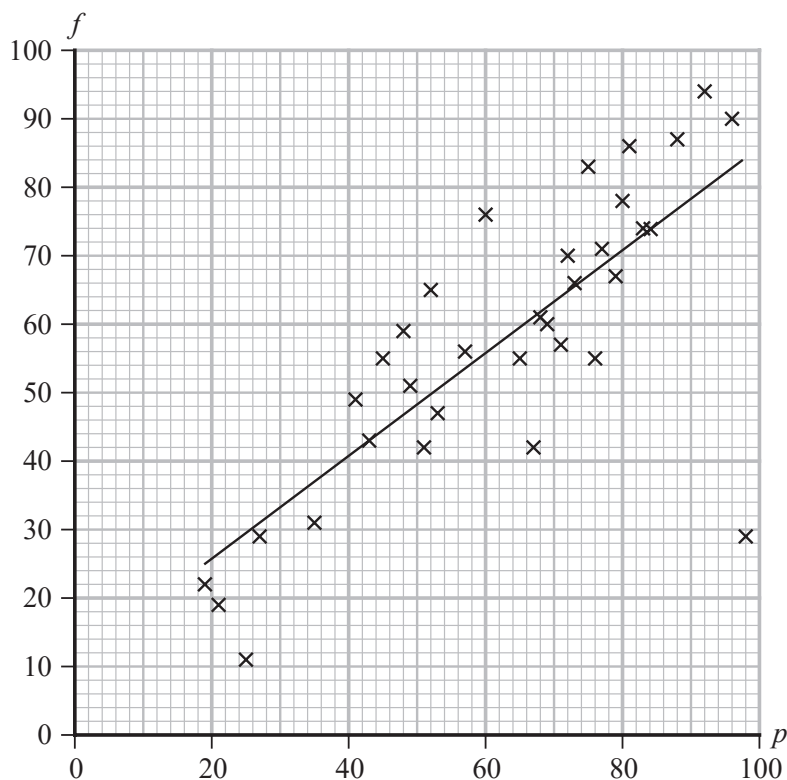


(Total for Question 6 is 12 marks)



Leave
blank

6. Students on a psychology course were given a pre-test at the start of the course and a final exam at the end of the course. The teacher recorded the number of marks achieved on the pre-test, p , and the number of marks achieved on the final exam, f , for 34 students and displayed them on the scatter diagram.



The equation of the least squares regression line for these data is found to be

$$f = 10.8 + 0.748p$$

For these students, the mean number of marks on the pre-test is 62.4

- (a) Use the regression model to find the mean number of marks on the final exam.

(2)

- (b) Give an interpretation of the gradient of the regression line.

(1)

Considering the equation of the regression line, Priya says that she would expect someone who scored 0 marks on the pre-test to score 10.8 marks on the final exam.

- (c) Comment on the reliability of Priya's statement.

(1)

- (d) Write down the number of marks achieved on the final exam for the student who exceeded the expectation of the regression model by the largest number of marks.

(1)

Question 6 continues on page 24.





Leave
blank

Question 6 continued

- (e) Find the range of values of p for which this regression model, $f = 10.8 + 0.748p$, predicts a greater number of marks on the final exam than on the pre-test.

(3)

Later the teacher discovers an error in the recorded data. The student who achieved a score of 98 on the pre-test, scored 92 not 29 on the final exam.

The summary statistics used for the model $f = 10.8 + 0.748p$ are corrected to include this information and a new least squares regression line is found.

Given the **original** summary statistics were,

$$n = 34 \quad \sum p = 2120 \quad \sum pf = 133486 \quad S_{pp} = 15573.76 \quad S_{pf} = 11648.35$$

- (f) calculate the gradient of the new regression line. Show your working clearly.

(5)



Leave
blank

Q6





(Total for Question 2 is 13 marks)



Question 2 continued

Lined area for writing the answer to Question 2.



Question 2 continued

Lined area for writing the answer to Question 2.



Question 2 continued

Lined area for writing the answer to Question 2.

(Total for Question 2 is 13 marks)



Question 4 continued

Lined area for writing the answer to Question 4 continued.



Question 4 continued

Lined area for writing the answer to Question 4 continued.



(Total for Question 4 is 13 marks)



Question 2 continued

Lined area for writing the answer to Question 2 continued.



Question 2 continued

Lined area for writing the answer to Question 2.

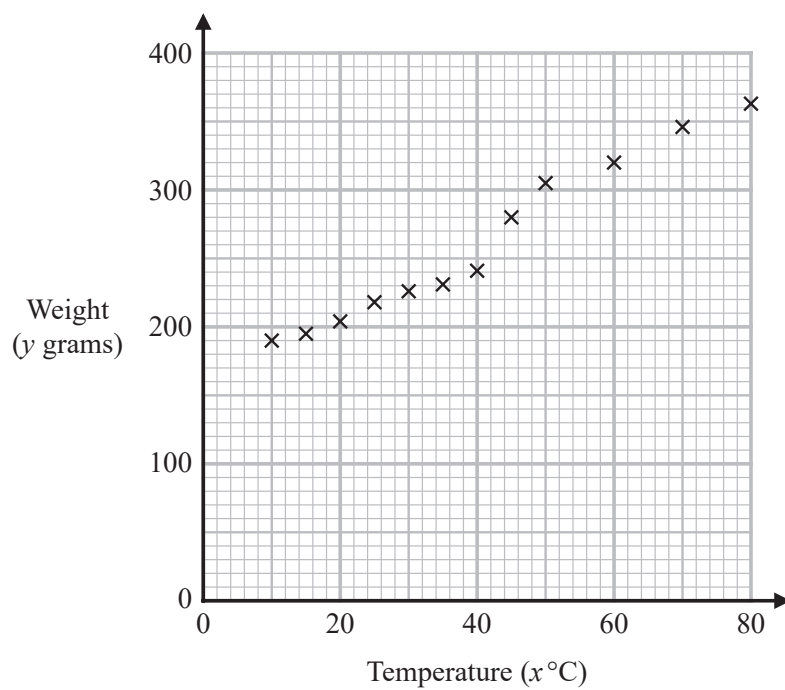


(Total for Question 2 is 14 marks)









TOTAL FOR PAPER IS 75 MARKS

3. Statistical models are a cheap and quick way to make predictions about real-world situations.

(a) Give one other reason why statistical models are used.

(1)

Madison wants to develop a model to describe the relationship between the average daily temperature, t °C, and a household's daily gas consumption, d m³, in winter.

Madison takes a random sample of 12 days in winter and codes the daily gas consumption so that $w = 11.5d$

These data are summarised as follows

$$S_{tt} = 26.43 \quad S_{tw} = -91.55 \quad \sum w = 339.25 \quad \sum t = 9.1 \quad \sum w^2 = 10\,036.45$$

(b) Show that $S_{ww} = 445.57$ to 2 decimal places.

(1)

(c) Find the value of S_{td} and the value of S_{dd}

(3)

(d) Find the product moment correlation coefficient between d and t

(2)

(e) Give an interpretation, in context, of your product moment correlation coefficient.

(1)

(f) Show that the equation of the regression line of w on t is

$$w = -3.46t + 30.9$$

where the values of the intercept and the gradient are given to 3 significant figures.

(3)

(g) Write down an equation of the regression line of d on t

(1)

(h) Using your equation in part (g)

(i) estimate the daily gas consumption in winter when the temperature is 2 °C

(ii) interpret the effect an increase of 1 °C in average daily temperature is expected to have on the daily gas consumption in winter.

(2)



Question 3 continued

Lined area for writing the answer to Question 3 continued.



Question 3 continued

Lined area for writing the answer to Question 3.

(Total for Question 3 is 14 marks)



Leave
blank



Leave
blank

Question 5 continued

Leave
blank

Lined area for writing the answer to Question 5.

(Total 15 marks)

Q5	
----	--



Leave
blank

6. Two economics students, Andi and Behrouz, are studying some data relating to unemployment, $x\%$, and increase in wages, $y\%$, for a European country. The least squares regression line of y on x has equation

$$y = 3.684 - 0.3242x$$

and $\sum y = 23.7$ $\sum y^2 = 42.63$ $\sum x^2 = 756.81$ $n = 16$

- (a) Show that $S_{yy} = 7.524375$ (1)

- (b) Find S_{xx} (4)

- (c) Find the product moment correlation coefficient between x and y . (3)

Behrouz claims that, assuming the model is valid, the data show that when unemployment is 2% wages increase at over 3%

- (d) Explain how Behrouz could have come to this conclusion. (1)

Andi uses the formula

$$\text{range} = \text{mean} \pm 3 \times \text{standard deviation}$$

to estimate the range of values for x .

- (e) Find estimates of the minimum value and the maximum value of x in these data using Andi's formula. (3)

- (f) Comment, giving a reason, on the reliability of Behrouz's claim. (2)

Andi suggests using the regression line with equation $y = 3.684 - 0.3242x$ to estimate unemployment when wages are increasing at 2%

- (g) Comment, giving a reason, on Andi's suggestion. (2)

Leave
blank

Question 6 continued

Leave
blank

Handwriting practice area with horizontal lines.

(Total 16 marks)

Q6

Grading box for Q6.

TOTAL FOR PAPER: 75 MARKS

END



Leave
blank

5. A company director wants to introduce a performance-related pay structure for her managers. A random sample of 15 managers is taken and the annual salary, y in £1000, was recorded for each manager. The director then calculated a performance score, x , for each of these managers.

The results are shown on the scatter diagram in Figure 1 on the next page.

- (a) Describe the correlation between performance score and annual salary. (1)

The results are also summarised in the following statistics.

$$\sum x = 465 \quad \sum y = 562 \quad S_{xx} = 2492 \quad \sum y^2 = 23\,140 \quad \sum xy = 19\,428$$

- (b) (i) Show that $S_{xy} = 2006$ (1)

- (ii) Find S_{yy} (2)

- (c) Find the product moment correlation coefficient between performance score and annual salary. (2)

The director believes that there is a linear relationship between performance score and annual salary.

- (d) State, giving a reason, whether or not these data are consistent with the director's belief. (1)

- (e) Calculate the equation of the regression line of y on x , in the form $y = a + bx$. Give the value of a and the value of b to 3 significant figures. (4)

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

- (g) Plot your regression line on the scatter diagram in Figure 1 (2)

The director hears that one of the managers in the sample seems to be underperforming.

- (h) On the scatter diagram, circle the point that best identifies this manager. (1)

The director decides to use this regression line for the new performance related pay structure.

- (i) Estimate, to 3 significant figures, the new salary of a manager with a performance score of 30 (2)

Leave
blank



MARK SCHEMES

official mark schemes, in the same order as the questions



2. (a)	$S_{dp} = 5240.8 - \frac{1029 \times 50.8}{10} [= 13.48]$	M1	
	$r = \frac{'13.48'}{\sqrt{344.9 \times 0.576}}$	M1	
	$= 0.9563834526 \dots$	A1	
	awrt 0.956		(3)
(b)(i)	$w = 50 - p$	B1	
(ii)	-1	B1	(2)
(c)	-0.956	B1ft	(1)
		[6]	
Notes			
(a)	1 st M1 correct expression for S_{dp} 2 nd M1 valid attempt at r with their S_{dp} not equal to 5240.8 and the correct denominator A1 awrt 0.956		
(b)(i)	B1 allow equivalent rearrangements		
(ii)	B1 - 1 cao		
(c)	B1ft follow through -1 × their(a) providing $-1 < \text{their (a)} < 1$		

Question Number	Scheme	Marks
-----------------	--------	-------

Question Number	Scheme		Marks
2 (a) (i)	$S_{yy} = 81938.5 - \frac{2015^2}{50} [= 734] *$		B1*
(ii)	$r = \frac{219.55}{\sqrt{734 \times 72.25}} = 0.95338...$	awrt 0.953	M1 A1
			(3)
(b)	e.g. [In general] the longer the rabbit the greater the weight		B1ft
			(1)
(c)	Consistent/Yes as r /PMCC is close to 1		B1ft
			(1)
(d)	$b = \frac{219.55}{734} = 0.2991...$		M1 A1
	$a = \left(\frac{125}{50}\right) - 'b' \left(\frac{2015}{50}\right) [= -9.554...]$		M1
	$w = -9.55 + 0.299y$		A1
			(4)
(e)	$'-9.55' + '0.299' \times 45 = 3.905$	awrt 3.91	M1 A1ft
			(2)
Notes			Total 11
(a)(i)	B1*	Answer is given so a correct numerical expression and no incorrect working seen	
(ii)	M1	For use of $\frac{S_{yw}}{\sqrt{S_{yy} \times S_{ww}}}$ May be implied by awrt 0.953	
	A1	awrt 0.953	
(b)	B1ft	A correct interpretation ft their r value (provided that $ r < 1$) e.g as length/ y increases the weight/ w increases Ignore any figures quoted. Do not accept comments about correlation on their own	
(c)	B1ft	A correct statement with a correct reason ft their r value (provided that $ r < 1$) Allow $0.953 \approx 1$ Allow 'my value' to imply r	
(d)	M1	A correct method to find the gradient (May be implied by awrt 0.299 or $\frac{4391}{14680}$)	
	A1	awrt 0.299 (Condone awrt 0.3 if M1 scored) May be implied by a correct gradient in the regression line.	
	M1	A correct method to find the intercept ft their b (May be implied by awrt -9.55)	
	A1	For $w = (\text{awrt}) - 9.55 + (\text{awrt})0.299y$ Must be seen in part (d)	
(e)	M1	For substitution of 45 into their regression equation	
	A1ft	For awrt 3.91 or ft their regression line, provided their final answer is > 0 . If regression line is not correct, then you will need to check their answer	

Question Number	Scheme		Marks
3 (a)			B1 B1 B1
			(3)
(b) (i)	$\frac{14}{200}$		B1ft (1)
(ii)	$\frac{33+11+42+32}{200} = \frac{118}{200}$ or $\frac{200-29-22-14-17}{200} = \frac{118}{200}$		M1 A1
			(2)
(c)	$\frac{14+11}{14+11+22+33} = \frac{25}{80}$		M1 A1
			(2)
Notes			Total 8
(a)	B3	For a fully correct Venn diagram (B2 for at least 6 numbers in the correct place on the Venn diagram) (B1 for at least 3 numbers in the correct place on the Venn diagram) Treat blanks on the diagram as zero Condone correct probabilities instead of frequencies e.g. $\frac{22}{200}$ oe or 0.11 as 22	
(b) (i)	B1ft	For $\frac{14}{200}$ oe ft their Venn diagram	
(ii)	M1	$\frac{33+11+42+32}{200}$ or $\frac{200-29-22-14-17}{200}$ ft their Venn diagram. May be implied by $\frac{118}{200}$ oe	
	A1	For $\frac{118}{200}$ oe	
(c)	M1	$\frac{n}{14+11+22+33}$ provided the answer gives a probability and $0 < n < 80$ ft the Venn diagram for the denominator or $\frac{m}{0.4}$ where $0 < m < 0.4$ ft the Venn diagram for the denominator	$\frac{n}{80}$ provided $0 < n < 80$ ft the Venn diagram for numerator or $\frac{m}{0.4}$ where $0 < m < 0.4$ ft the Venn diagram for numerator
	A1	For $\frac{25}{80}$ oe Allow 0.313	

Question Number	Scheme	Marks
2 (a)	$[S_{xp}] = 2347 - \frac{93 \times 273}{12}$ or $2347 - \frac{25\,389}{12} [= 231.25] (*)$	B1cso (1)
(b)	$[S_{pp}] = 6602.72 - \frac{273^2}{12} = [391.97]$	M1
	$[r] = \frac{231.25}{\sqrt{148.25 \times 391.97}}$	M1
	$= 0.959307... \text{ awrt } \underline{0.959}$	A1 (3)
(c)	$\left[b = \frac{S_{xp}}{S_{xx}} = \frac{231.25}{148.25} [= 1.559865...] \right]$	M1
	$a = \frac{273}{12} - "1.56" \times \frac{93}{12}$ or $22.75 - "1.56" \times 7.75 [= 10.66...]$	M1
	$b = \text{awrt } 1.6 \text{ or } a = \text{awrt } 11$	A1
	$\underline{p = 10.7 + 1.56x}$	A1 (4)
(d)	e.g. each extra employee costs the company (on average)["\$"]156" a year in paper	B1 (1)
(e)	$[\text{New } p =] 0.8 \times "10.66..." + \frac{"1.559..." \times 93}{2} [= 14.573...]$	M1
	$[\text{compared with } \bar{p} = 22.75] \text{ so percentage saving is } \frac{22.75 - 14.573...}{22.75} [\times 100]$	M1
	$= 35.94... \text{ awrt } \underline{36\%}]$	A1 (3)
	Notes	[12]
(a)	B1 for either correct expression [don't need $= 231.25$]	
(b)	1 st M1 for attempt at correct expression for S_{pp} Allow one transcription error e.g. 6620.72 May be seen in part (a) If no correct expression seen allow $S_{pp} = \text{awrt } 392$ or correctly placed in formula for r 2 nd M1 for a correct expression for r , ft their S_{pp} A1 for awrt 0.959	
(c)	1 st M1 for a correct expression for b (may be implied by a correct value of awrt 1.56) 2 nd M1 for a correct expression for a ft their value for b . May be implied by awrt 10.7 1 st A1 for $b = \text{awrt } 1.6$ or $a = \text{awrt } 11$ 2 nd A1 for correct equation in p and x with $b = \text{awrt } 1.56$ and $a = \text{awrt } 10.7$	
(d)	B1 for a suitable contextual comment that mentions their value of b Allow multiples eg every extra 100 employees costs the company "\$15600". Condone missing \$ sign or use of £. Do not allow " \$1.56" for 1 person unless indicates in 100's	
(e)	1 st M1 for a correct expression for average value of p using new model [ft their a and b] 2 nd M1 for correct percentage saving calculation using 22.75 (e.g. $\frac{14.573...}{22.75} [\times 100]$) Allow use of $"10.7" + 1.56" \times \frac{93}{12} [\approx 22.79]$ for 22.75 May be implied by correct answer. A1 for awrt 36 SC use of 93 throughout part (e) rather than 7.75 leading to awrt 48 or 0.48 (they will need to use the regression line from part(c) to calculate the original value) gains M0M1A0 SC use of 93 in part(c) Answer of 36% gains M1M1A1, 64% or 0.64 gains M1M1A0.	

Question Number	Scheme			Marks
6 (a)(i)	$x = 1.2 + 0.2(1.4x + 1.5)$ o.e or $y = 1.4(1.2 + 0.2y) + 1.5$ o.e			M1
	$x = \frac{25}{12}$ $y = \frac{53}{12}$			A1A1
	(ii) $[\sum x =] \frac{25}{12} \times 12 [= 25]$			A1*
				(4)
(b)	$[\sum y =] \left(\frac{53}{12}\right)' \times 12 = 53$			M1A1ft
	$S_{xy} = \frac{6961}{60} - \frac{(25 \times '53')}{12} = 5.6$			M1 A1
				(4)
(c)	$\frac{'5.6'}{S_{xx}} = 1.4$ and $\frac{'5.6'}{S_{yy}} = 0.2$	$\frac{'5.6'}{\sqrt{\frac{'5.6'}{1.4} \times \frac{'5.6'}{0.2}}}$	$\frac{S_{xy}}{S_{xx}} = 1.4$ and $\frac{S_{xy}}{S_{yy}} = 0.2$	M1
	$S_{xx} = 4$ and $S_{yy} = 28$	$\frac{5.6}{\sqrt{1.4 \times 0.2}}$	$r^2 = 1.4 \times 0.2$	A1
	$r = \frac{'5.6'}{\sqrt{'4' \times '28'}} = 0.5291...$	$\sqrt{1.4 \times 0.2} = 0.5291$		M1 dA1
				awrt 0.529
				(4)
	Notes			Total 12
(a)(i)	M1	For either of the two equations o.e or an attempt to solve the two equations simultaneously. May be implied by $x = \frac{25}{12} / 2.08$ or better or $y = \frac{53}{12} / 4.42$ or better		
	A1	For either $x = \frac{25}{12} / 2.08$ or better or $y = \frac{53}{12} / 4.42$ or better		
	A1	For both $x = \frac{25}{12} / 2.08$ or better and $y = \frac{53}{12} / 4.42$ or better (May be written as a coordinate) NB This is M1 on EPEN		
(ii)	A1*	For $\frac{25}{12} \times 12$ Allow use of $\sum x$ rather than $\bar{x}$ e.g. $\sum x = 14.4 + 0.2(1.4 \sum x + 18)$ oe As the answer is given no incorrect working must be seen. NB Working must be shown		
(b)	M1	For $\left(\frac{53}{12}\right)' \times 12$ ft their y coordinate. Allow use of $\sum y$ rather than $\bar{y}$ e.g. $\sum y = 1.4(14.4 + 0.2 \sum y) + 18$ oe		
	A1ft	For $\sum y = 53$ or ft their y coordinate $\times 12$ (An answer of exactly 5.6 implies M1A1)		
	M1	Use of $S_{xy} = \frac{6961}{60} - \frac{25 \times \sum y'}{12}$ ft their $\sum y$ If $\sum y$ is not stated then M0 is awarded		
	A1	5.6 (Allow awrt 5.6)		

(c)	M1	For use of the gradient to find S_{xx} and S_{yy} ft their S_{xy} or use of $\frac{S_{xy}}{\sqrt{\frac{S_{xy}}{S_{xx}} \times \frac{S_{xy}}{S_{yy}}}}$ or setting both $\frac{S_{xy}}{S_{xx}}$ and $\frac{S_{xy}}{S_{yy}}$ equal to their respective gradients
	A1	$S_{xx} = 4$ and $S_{yy} = 28$ or $\frac{S_{xy}}{\sqrt{1.4 \times 0.2}}$ or $\frac{(S_{xy})^2}{S_{xx} \times S_{yy}} = 1.4 \times 0.2$
	M1	For a correct expression for r ft their S_{xy} , S_{xx} and S_{yy} or $\sqrt{1.4 \times 0.2}$ If answer is incorrect then you must see their stated values substituted into a correct expression for r . An answer of $\frac{\sqrt{7}}{5}$ implies M1A1M1 only
	dA1	Dependant on all previous marks being awarded. awrt 0.529

Question Number	Scheme		Marks
1 (a)	$2 \times 36 = 72 \quad 8 \times 4 = 32$		M1 A1
			(2)
(b)	$[13] + \frac{(204 - 184)}{120} \times 2$	$[13] + \frac{(204.5 - 184)}{120} \times 2$	M1
	$= \frac{40}{3} = \text{awrt } 13.3$		A1
			(2)
(c)	Symmetrically distributed/No skew as the mean $\approx$ median		B1
			(1)
(d)	$\frac{32}{4} + 152 + \frac{120}{2} [= 220]$		M1
	$\frac{'220'}{408} \times \frac{'219'}{407}$		M1
	$\frac{365}{1258}$ or 0.2901... awrt 0.29		A1
			(3)
	Notes		Total 8
(a)	M1	For any equivalent method to find either frequency Maybe implied by either correct frequency Also maybe implied by two frequencies which add to 104 Also maybe implied by a correct scale on the fd axis , at least 3 labels	
	A1	For 72 and 32	
(b)	M1	For any equivalent method to find the median e.g. $\frac{Q_2 - 13}{15 - 13} = \frac{204 - 184}{304 - 184}$ or $\frac{15 - Q_2}{Q_2 - 13} = \frac{304 - 204}{204 - 184}$ allow working downwards $[15] - \frac{(304 - 204)}{120} \times 2$	
	A1	awrt 13.3	
(c)	B1	For a correct identification of skew [which must either be symmetric/no skew or (slight) negative skew] with a correct supporting reason. Condone mean < median so negative skew Allow use of 'their median' in the comparison provided 'their median' ... 13.2 Allow $Q_1 = \text{awrt } 10.8$ or awrt 10.9 and $Q_3 = \text{awrt } 15.1$ and $Q_2 - Q_1 > Q_3 - Q_2$ so negative skew. Comments referring only to the diagram (being symmetrical therefore no skew) send to review	
(d)	M1	For a correct method to find the number of plants between 8cm and 14cm (may be implied by sight of 220)	
	M1	For $\frac{n}{408} \times \frac{n-1}{407}$ or $\left(\frac{n}{408}\right)^2$ with 210 ,, n ,, 230	
	A1	awrt 0.29 may see $\frac{3025}{10404}$ from $\left(\frac{220}{408}\right)^2$	

Question Number	Scheme		Marks
2 (a)(i)	Mean = 71.83...	awrt 71.8	B1
(ii)	Standard deviation = $\sqrt{\frac{62802}{12} - \left(\frac{862}{12}\right)^2}$ or variance = $\frac{62802}{12} - \left(\frac{862}{12}\right)^2$		M1
	$\sqrt{73.47...} = 8.571...$	8.57 * (to 3s.f.)	A1*
			(3)
(b)	$S_{xx} = 62802 - \frac{862^2}{12} [= \frac{2645}{3} = 881.66...]$		M1
	$r = \frac{512.67}{\sqrt{413.67 \times 881.66...}}$		M1
	= 0.8489...	awrt 0.849	A1
			(3)
(c)	Mean = $\frac{5}{9} \times ('71.8' - 32)$		M1
	= 22.11...	awrt 22.1	A1ft
	Standard deviation = $\frac{5}{9} \times 8.57...$		M1
	= 4.76...	awrt 4.76	A1
			(4)
(d)	$r = '0.8489...' / \text{same (as for x and y)}$		M1
	r not affected by (linear) coding oe		A1
			(2)
Notes			Total 12
(a)(i)	B1	awrt 71.8 Allow $\frac{431}{6}$ oe	
(ii)	M1	A correct method to find the standard deviation or the variance ft their mean for M1 only Also allow s.d. = $\sqrt{\frac{S_{xx}}{n}}$	
	A1*	Must see at least one simplification of working and the given answer 8.57. e.g. $\sqrt{73.47...}$ or 8.572 or 8.571... or $\frac{23\sqrt{5}}{6}$ or $\sqrt{\frac{2645}{36}}$ therefore s.d. = 8.57* (to 3s.f.) $\sqrt{\frac{62802}{12} - 71.8^2}$ scores M1A0 (use of 71.8 or 71.83 always scores M1A0). To get required accuracy must see at least 71.833 used i.e. $\sqrt{\frac{62802}{12} - 71.833^2}$	
(b)	M1	A correct method to find S_{xx} (implied by awrt 882)	
	M1	A correct method to find PMCC using their value of S_{xx}	
	A1	awrt 0.849	
(c)	M1	A correct method to find the mean ft their mean in part (a)	
	A1ft	awrt 22.1 ft their mean in part (a)	
	M1	A correct method to find the standard deviation (do not isw if any further calculation is done after multiplying by $\frac{5}{9}$)	
	A1	awrt 4.76	
(d)	M1	$r =$ their part (b) provided $-1 \leq r \leq 1$ allow 2 s.f. on the ft	
	A1	Any correct reasoning but M1 must be scored. Allow e.g. 'addition/subtraction and multiplication/division does not affect r ' ,	

Question Number	Scheme		Marks
4 (a)	LQ = 28 or UQ = 48		B1
	'48'+1.5('48'- '28') [= 78]		M1
	90 > 78 so, 90 is an outlier*		A1*
			(3)
(b)	$b = \frac{1735.6}{1667.6} [= 1.04...]$		M1
	$a = 38.2 - 'b'(42.2) [= -5.72....]$		M1
	$s = -5.72 + 1.04 f$ *		A1*
			(3)
(c)	For every extra mark (oe) in French/f , Spanish/s goes up (oe) by [on average] 1.04 marks		B1
			(1)
(d) (i) (ii)	$s = -5.72 + 1.04 \times 55 = 51.48$ awrt 51.5		M1 A1
	$s = -5.72 + 1.04 \times 18 = 13$		A1
			(3)
(e)	- The first estimate is an interpolation/The second estimate is an extrapolation 55 is within the range of data/18 is not within the range of data 55 is closer to the mean/18 is further away from the mean so 51.5 is the more reliable estimate		M1
			A1
			(2)
Notes			Total 12
(a)	B1	For either LQ or UQ correct (may be seen in calculation for M1)	
	M1	Correct use of $Q_3 + 1.5 \times (Q_3 - Q_1)$ ft their LQ and their UQ provided their UQ > their LQ	
	A1*	For both LQ and UQ correct and identifying 90>78 or 90 is an outlier Answer is given so no incorrect working can be seen	
(b)	M1	For a correct method to find the gradient	
	M1	For a correct method to find the intercept (division by 11 is M0)	
	A1*	Cao (dep on both M marks) must see printed answer $s = -5.72 + 1.04 f$	
(c)	B1	For a correct numerical interpretation of the gradient in context which must include <u>marks</u> at least once	
(d) (i)	M1	For a correct substitution into the regression equation. May be seen in (i) or (ii) or implied by one correct answer	
	A1	awrt 51.5 Allow 51 or 52	
(ii)	A1	13 or awrt 13.0	
(e)	M1	For any equivalent correct reason Ignore extraneous non-contradictory comments For the second bullet point must be clear that they are referring to French marks (24,, f ,, 68). Do not allow comments that refer to the range of Spanish marks e.g. ‘51.5’ is within the range of data/‘13’ is not within the range of data’ Do not allow ‘55 is closer to the median (than 18)’	
	A1	For clearly identifying the estimate from part (d)(i): 51.5 or 55 or (i) or ‘the first estimate’ , etc.	

Question Number	Scheme		Marks
6 (a)	$S_{tt} = 14837 - \frac{635^2}{30} \left(= \frac{8377}{6} = 1396.1666... \right)$		M1
	$r = \frac{-1648.83}{\sqrt{2396.97 \times 1396.166...}} = -0.9013136...^*$		A1*
			(2)
(b)	Possible linear relationship between t and w / points lie close to a (straight) line a negative gradient/slope / as w increases t decreases		B1 B1 (2)
(c)			
	$b = \frac{S_{wt}}{S_{ww}} = \frac{-1648.83}{2396.97}$		M1
	$= -0.68788... \qquad \qquad \qquad \text{awrt } \underline{-0.69}$		A1
	$a = \frac{635}{30} - (-0.68788...) \times \frac{839}{30} = 40.404...$ $t = 40.4 - 0.688w$		M1 A1
			(4)
(d)	On average as score increases by 1, time decreases by '0.688' minutes		B1ft (1)
(e)(i)	(pmcc would) stay the same		B1
(ii)	(Magnitude of gradient would) decrease		B1
(iii)	(Intercept would) stay the same		B1
			(3)
	Notes		Total 12
(a)	M1	Use of correct formula to find S_{tt} implied by 1396 or better	
	A1*	Correct calculation shown to find r and answer awrt -0.901	
(b)	B1	One correct feature Mention of linear relationship/close to a straight line oe Just mentioning the word line is not enough – must imply the points/graph form a (straight) line or negative gradient/slope oe allow eg downward gradient for negative gradient Do not allow negative correlation/negative trend/negative relationship. A single comment eg 'The gradient of the line is negative' scores B1B0 (since the mention of line here does not imply that the points form a line)	
	B1	Two correct features Mention of linear relationship/close to a straight line oe Just mentioning the word line is not enough – must imply the points/graph form a (straight) line and negative gradient/slope oe allow eg downward gradient for negative gradient Do not allow negative correlation/negative trend/negative relationship. A single comment eg 'The gradient of the line is negative' scores B1B0 (since the mention of line here does not imply that the points form a line)	
(c)	M1	Correct method for b (implied by awrt -0.69)	
	A1	awrt -0.69 (may be seen in final equation)	
	M1	Correct method to find a ft their b (implied by awrt 40.4)	
	A1	Fully correct equation must be in terms of t and w with awrt 40.4 and awrt -0.688 . No fractions.	
(d)	B1ft	For a numerical interpretation which must mention score (oe) and time/minutes (oe) or w and t , ft their ' -0.688 '. Condone eg as score increases by 1, time increases by ' -0.688 ' ft interpretation must be compatible with the sign of their gradient.	
(e)(i)	B1	oe allow -0.901	
(ii)	B1	oe condone 'halves'	
(iii)	B1	oe allow '40.4'	



Question Number	Scheme	Marks
6. (a)	$\bar{f} = 10.8 + 0.748\bar{p} = 10.8 + 0.748(62.4)$ awrt 57.5	M1 A1 (2)
(b)	For each additional mark scored on the pre-test , the average mark on the final exam increases by 0.748	B1 (1)
(c)	The statement is not reliable as there is no data below 19 (extrapolation).	B1 (1)
(d)	76	B1 (1)
(e)	$p < 10.8 + 0.748p$ $0.252p < 10.8$ $p < \text{awrt } 42.9$	M1 M1 A1 (3)
(f)	[No change to] $S_{pp} = 15\,573.76$	
	$\sum pf = 133486 - 2842 + 9016$ [= 139660]	$\sum pf$ increases by $98(92 - 29)$ [= 6174] M1
	$\sum f = "57.47" \times 34 + (92 - 29)$ or $\frac{133486 - 11648.35}{2120} \times 34 + (92 - 29)$ [= 1954 + 92 - 29 $\approx$ 2017]	$\frac{\sum p \sum f}{n}$ increases by $\frac{2120(92 - 29)}{34}$ [= 3928.235...] M1
	$S_{pf} = "139660" - \frac{2120 \times "2017"}{34}$ [= 13894...]	S_{pf} increases by '6174' - '3928.235' [= 2245.764...] dM1
	$b = \frac{"13894..."}{15573.76}$ [= 0.89...]	$b = \frac{11648.35 + "2245.764"}{15573.76}$ M1
	awrt 0.9	A1 (5)
	Notes	[13]
(a)	M1 for substituting 62.4 into the regression equation. Allow answer between 57 and 58 A1 awrt 57.5	
(b)	B1 must include context and reference to 0.748 Needs to refer to each mark being 0.748 or a multiple of eg 10 marks is 7.48 Allow equivalent words eg score/ point for mark, pre or test for pre-test, exam or final for final exam	
(c)	B1 Not reliable with correct supporting reason eg it (10.8) is an outlier, outside the range	
(d)	B1 76 cao	
(e)	1 st M1 for setting up inequality in p only or for drawing the line $f = p$ on the graph. May be implied by $p < n$ (ignore any lower limit) where $40 \leq n < 46$ (allow incorrect inequality sign or =) Allow trial and improvement. 2 nd M1 rearranging to the form $ap < b$ with correct inequality sign. Allow $(1 - 0.748)p < 10.8$ May be implied by $p < n$ (ignore any lower limit) where $42 < n < 44$ A1 $p < \text{awrt } 42.9$ (ignore any lower limit) ISW	
(f)	1 st M1 Correct method to find new $\sum pf$ or change in $\sum pf$ 2 nd M1 Correct method to find new $\sum f$ or change in $\frac{\sum p \sum f}{n}$ Allow 2018 or 2017 3 rd dM1 dep on both previous method marks being awarded. Correct method to find new S_{pf} with their changed $\sum pf$ and $\sum f$ or change in S_{pf} 4 th M1 expression for $b = \frac{S_{pf}}{15573.76}$ with their changed S_{pf} and unchanged S_{pp} A1 awrt 0.9 (from correct working)	

Question Number	Scheme	Marks
1. (a)	[Area = $k \times \text{frequency} \rightarrow 16.5 = k \times 12 \rightarrow$] Area = $\frac{16.5}{12} \times 18$ oe $= \underline{24.75} \text{ (cm}^2\text{)}$	M1 A1 (2)
(b)	fd method $\frac{24}{58-55} [= 8]$ and $\frac{35}{55-50} [= 7]$ or Area method $\frac{16.5}{12} \times 24 [= 33]$ and $\frac{16.5}{12} \times 35 [= 48.125]$ or $\frac{16.5}{12} \times \frac{24}{3} [= 11]$ and $\frac{16.5}{12} \times \frac{35}{5} [= 9.625]$ Let h = height of the 2 nd tallest bar [$h = \frac{10}{8} \times 7$] or [$h = \frac{48.125 \times 10 \times 3}{5 \times 33}$] or [$h = 9.625 \times \frac{10}{11}$] $= \underline{8.75} \text{ (cm)}$	M1 dM1 A1 (3)
(c)(i)	[$Q_2 = 50 + \frac{7}{35} \times 5$] or [$Q_2 = 55 - \frac{28}{35} \times 5$] $= \underline{51} \text{ (cm)}$	M1 A1
(ii)	[$Q_3 = 55 + \frac{2}{24} \times 3$] or [$Q_3 = 58 - \frac{22}{24} \times 3$] and "55.25" – 45 $= \underline{10.25} \text{ (cm)}$	M1 A1 (4)
(d)	"55.25" – 2("51") + 45 "55.25" – 45 [$= -0.17073... < 0$] negative [skew].	M1 A1ft (2)
	Notes	[11]
(a)	M1 allow equivalent eg $16.5 \times \frac{3}{2}$ A1 for 24.75 allow 24.8	
(b)	M1 correct method for finding the frequency density or area for the highest and 2nd highest bars Allow if 8 and 7 seen or 33 and 48.125 seen or 9.625 rather than 48.125 and/or 11 seen rather than 33 dM1 dep on previous M mark awarded. A fully correct expression for h or a fully correct equation to enable h to be found eg $\frac{33}{10 \times 3} = \frac{48.125}{5h}$ A1 8.75 oe NB answer of 8.75 seen as final answer 3/3	
(c)(i)	M1 for $50 + \frac{7}{35} \times k$ or $55 - \frac{28}{35} \times k$ or $\frac{Q_2 - 50}{k} = \frac{60 - 53}{88 - 53}$ or $\frac{55 - Q_2}{k} = \frac{88 - 60}{88 - 53}$ where 4,, k ,, 6 oe (condone use of $n + 1$ ie 7.5 rather than 7, 27.5 rather than 28 or 60.5 rather than 60) A1 51 (condone for use of $n + 1$ awrt 51.1)	
(ii)	M1 $55 + \frac{2}{24} \times t$ or $58 - \frac{22}{24} \times t$ or $\frac{Q_3 - 55}{t} = \frac{90 - 88}{112 - 88}$ or $\frac{58 - Q_3}{t} = \frac{112 - 90}{112 - 88}$ oe where 2,, t ,, 4 and using "their Q_3 " – 45 (condone use of $n + 1$ ie 2.5 or 2.75 rather than 2, 21.5 or 21.25 rather than 22 or 90.5 or 90.75 rather than 90) A1 10.25 oe eg 41/4 allow 10.3 from correct working	

(d)	M1 substitution of their values from (c) seen or awrt -0.17 or $-7/41$ A1ft dependent on M1 being scored. Correct description of skewness consistent with their values from part (c) ignore the final answer if working shown. Only allow no skew or symmetrical if their value should be 0 Ignore correlation.
-----	---

Question Number	Scheme	Marks
2. (a)	$[S_{tt}] = 82\,873 - \frac{1361^2}{40}$ $[S_{ct}] = 83\,634 - \frac{1634 \times 1361}{40}$ $[S_{tt}] = 36\,564.975$ $[S_{ct}] = 28\,037.15$	M1 A1 A1 (3)
(b)	$[r] = \frac{28\,037.15}{\sqrt{28\,732.1 \times 36\,564.975}} = 0.865\dots$ awrt 0.865	M1 A1 (2)
(c)	In general, films with higher cost have higher ticket sales .	B1ft (1)
(d)	$[b] = \frac{28\,037.15}{28\,732.1} [= 0.9758\dots]$ $[a] = \frac{1361}{40} - 'b' \times \frac{1634}{40}$ or $34.025 - 'b' \times 40.85$ $t = -5.8369\dots + 0.9758\dots c$ $t = \text{awrt } -5.84 + \text{awrt } 0.976c^*$	M1 M1 A1cso* (3)
(e)	$t = -5.84 + 0.976 \times 90$ $t = \text{£}82 \text{ million}$ awrt £82 million	M1 A1 (2)
(f)	$-5.84 + 0.976c < 0.8c \rightarrow 0.176c < 5.84$ $c < 33.1818$ $c < \text{awrt } (\text{£}) \text{ 33.2}$	M1 A1 (2)
	Notes	Total 13
(a)	Mark part (a) and (b) together M1 either correct expression A1 36 564.975 or exact equivalent A1 28037.15 or exact equivalent SC M1A1 A0 for awrt 36565 and awrt 28037	
(b)	M1 valid attempt at r with their $S_{cc} \neq 28732.1$ and their $S_{ct} \neq 83634$ A1 awrt 0.865	
(c)	B1ft only ft if $ r < 1$ For a correct comment in context. Must include words underlined (Allow use of c for cost and t for sales) and be compatible with their value in (b). Must have that as one increases/decreases the other increases/decreases. Allow other words eg goes up Do not accept t and c are similar.	
(d)	M1 correct numerical expression for b ft their $S_{ct} \neq 83634$ Implied by awrt 0.9758 or better M1 attempt at a with their value of b substituted. Implied by awrt 5.837 or better A1*cso answer given so both method marks must be awarded and no incorrect working seen. Either $b = 0.9758$ (or better) or $a = 5.837$ (or better) must be seen somewhere along with the correct equation in t and c (do not allow fractions).	
(e)	M1 for substituting $c = 90$ into $t = \text{awrt } -5.84 + \text{awrt } 0.976c$ Implied by awrt 82 A1 £82 million (must include units. Allow 82 million pounds). Allow awrt £82 million	
(f)	M1 forming inequality (allow $>$ or $<$ or $=$ or $\geq$ or $\leq$ or $\dots$) with $0.8c$ A1 correct inequality in c (allow any letter) with awrt 33.2 (units not required). Do not allow as a fraction. Ignore any lower limit. Condone awrt 33200000 or awrt 33.2 million	

Qu	Scheme	Marks
1(a)	eg 60 people = 1.5 large squares/6 medium squares/150 small squares or eg 1 person = 0.025 large squares/0.6 medium squares/2.5 small squares or eg [1 small square =] 0.4 people/[1 medium square =] 10 people/[1 large square =] 40 people eg a correct f.d. eg 60/(20 – 10) [= 6] eg a correct frequency, 100, 70, 20, 24 associated with the appropriate bar	B1
	eg $\frac{8}{10} \times 20$ or $\frac{15}{30} \times 24$ or 8×2 or 0.8×15 or $\frac{40}{2.5}$ or $\frac{30}{2.5}$ or $8 \times 5 \times 0.4$ $2 \times 15 \times 0.4$ or 16 or 12 or 70×0.4	M1
	28 people	A1
		(3)
(b)	Median = $[5] + \frac{5}{70} \times 37$ or $[10] - \frac{5}{70} \times 33$	M1
	= 7.642...	A1
	awrt 7.64	A1
		(2)
(c)	$\sum \text{midpoint} \times \text{freq} = 2.5 \times 100 + 7.5 \times 70 + 15 \times 60 + 25 \times 20 + 45 \times 24 [= 3255]$	M1
	Mean = $\frac{"3255"}{274}$	dM1
	= 11.879...	A1
	awrt 11.9	A1
		(3)
	Notes	Total 8
(a)	B1: for finding a ratio between people and area. Allow just the numbers for 1 person or for 1 square ie 0.025, 0.6, 2.5, 0.4, 1.66 or 40 or calculating f.d. for any bar correctly, fd = 20, 14, 6, 2 or 0.8 Information may be seen on diagram. Must be clear it is a fd either by correct use, seen on axes associated with correct bar or stated as an fd. May be implied by M1	
	M1: for a correct method to find the number of people between 22 and 30 km or 30 and 45 km or between 22 and 45 km	
	A1: 28	
	NB An answer of 28 gains 3/3 unless from obvious incorrect working	
(b)	M1: Allow equivalent for $n + \frac{5}{70} \times 37$ or $\frac{Q_2 - n}{5} = \frac{137 - 100}{170 - 100}$ or $n - \frac{33}{70} \times 5$ or $\frac{n - Q_2}{5} = \frac{170 - 137}{170 - 100}$ oe Allow alternative methods eg $\frac{Q_2 - 5}{10 - Q_2} = \frac{137 - 170}{170 - 137}$ Allow 37.5 for 37, 137.5 for 137, 32.5 for 33	
	A1: awrt 7.64 or $\frac{107}{14}$ or allow awrt 7.68 or $\frac{215}{28}$ if using $n + 1$ Allow awrt 7640 m or 7680 m but must have units.	
(c)	M1: Attempt at correct expression for $\sum \text{midpoint} \times \text{freq}$ - at least 3 products with correct midpoints added with at least 1 of these products fully correct . Allow for 3255	
	M1: dep on M1 being awarded for dividing “their sum” by 274	
	A1: awrt 11.9 or $\frac{3255}{274}$ Allow awrt 11900 m but must have units	

Qu	Scheme	Marks
2(a)	$S_{tw} = 2304.53 - \frac{297.8 \times 114.8}{15}$ or $S_{ww} = 6089.12 - \frac{297.8^2}{15}$	M1
	$S_{tw} = 25.367...$ awrt 25.4	A1
	$S_{ww} = 176.797$ awrt 177	A1
		(3)
(b)	$r = \frac{"25.367"}{\sqrt{5.3173 \times "176.797..."}}$	M1
	$= 0.82735...$ awrt 0.827 or 0.828	A1
		(2)
(c)	$b = \frac{"25.367..."}{5.3173} [= 4.77065...]$	M1
	$a = \frac{297.8}{15} - \frac{"25.367"}{5.3173} \times \frac{114.8}{15} [= -16.658...]$	M1
	$b = 4.771$ or better or $a = -16.66$ or better seen and $w = -16.7 + 4.77t^*$	A1*cso
		(3)
(d)	[On average,] for each cm/1 cm of tail length/t the weight/w increases by 4.77 g/grams	B1
		(1)
(e)	$w = -16.7 + 4.77 \times 2 [= -7.16]$ or $4.77 \times 2 [= 9.54]$ or $[t =] \frac{16.7}{4.77} [= 3.5]$ or sd = awrt 0.6	M1
	$[w =] -7.16$ or $9.54 < 16.7$ or $2 < 3.5$ which is negative/weight cannot be negative or for sd extrapolation since a 2 cm tail is (approx 9 sd)/(more than 3 sd) from the mean	A1
		(2)
(f)	0.827	B1ft
		(1)
(g)	$2y + 10 = -16.7 + 4.77(x + 6)$ oe	B1ft
		(1)
Notes		Total 13
(a)	M1 for a correct expression for S_{tw} or S_{ww}	
	A1 awrt 25.4	
	A1 awrt 177	
(b)	M1 for a valid attempt at r with their S_{tw} not equal to 2304.53 and S_{ww} not equal to 6089.12	
	A1 (M2 on open) awrt 0.827 or awrt 0.828	
(c)	1 st M1 for a correct method to find the value of b	
	2 nd M1 ft their b . For a correct method to find a . Minimum shown $a = \text{awrt } 19.9 - \text{"their } b" \times \text{awrt } 7.65 [= -16.658]$	
	A1* Both method marks must be awarded, equation stated (no fractions) and sight of (4.771 or better) or (-16.66 or better)	
(d)	B1 For a suitable contextual comment that implies that as length increases by 1 cm weight increases by 4.77g. Allow multiples eg each 10 cm increase in tail length weight increases by 47.7g Allow in terms of t and w	
(e)	M1 for a correct method to calculate the value of w (condone if written as a fraction) or $4.77 \times 2 [= 9.54]$ or correct method to find tail length when $w = 0$ or sd = awrt 0.6	
	A1 Method mark must be awarded. For -7.16 or $9.54 < 16.7$ or $2 < 3.5$ with a relevant explanation stating that weight is negative. If sd = awrt 0.6 is given allow extrapolation since a 2 cm tail is (approx 9 sd)/(more than 3 sd) from the mean.	
(f)	B1ft follow through their answer to (b)	
(g)	B1 ISW no need to be simplified. Allow equivalent eg $y = \frac{-16.7 + 4.77(x + 6)}{2} - 5$ The correct simplified equation is $y = 2.385x + 0.96$ allow awrt 2.39 and 0.96 – 0.98	

Question	Scheme		Marks
4 (a)	$S_{dg} = 141978.84 - \frac{1456.8 \times 713.2}{8}$ or $S_{gg} = 72675.98 - \frac{713.2^2}{8}$		M1
	$S_{dg} = 12105.12$		A1
	$S_{gg} = 9094.2$		A1
			(3)
(b)	$r = \frac{"12105.12"}{\sqrt{16769.78 \times "9094.2"}}$		M1
	$= 0.9802....$ awrt 0.98		A1
			(2)
(c)	$b = \frac{"12105.12"}{16769.78} [= 0.7218...]$		M1
	$a = \frac{713.2}{8} - "0.7218..." \times \frac{1456.8}{8} [= -42.297...]$		M1
	$g = -42.3 + 0.722d *$		A1*cso
			(3)
(d)	for each 1 [cm] increase in length/d the girth/g increases by “0.722...”		B1
			(1)
(e)(i)	138.2	awrt 138	B1
(ii)	[unreliable] as get a negative girth		B1
			(2)
(f)	$0.722x = 17.3$		M1
	$x = 23.96...$ awrt 24		A1
			(2)
	Notes		Total 13
(a)	M1	for a correct expression for S_{dg} or S_{gg}	
	A1	for 12105.12 Allow $\frac{302628}{25}$	
	A1	for 9094.2 Allow $\frac{45471}{5}$	
		If exact answers are not seen then SC award M1A0A1 for both awrt 12100 and awrt 9090 if correct methods are seen If exact answers are not seen then SC award M1A0A0 for both awrt 12100 and awrt 9090 if no working seen	
(b)	M1	for a valid attempt at r with their S_{dg} not equal to 141978.84	
	A1	awrt 0.98	
(c)	M1	for a correct method to find the value of b May be implied by 0.7218 or better	
	M1	for a correct method to find a ft their b May be implied by -42.29 or better	
	A1*	both method marks must be awarded with sight of 0.7218 or better or -42.29 or better	
(d)	B1	for a suitable contextual comment that mentions 0.722 (or better) If units are stated they must be correct.	
(e)(i)	B1	awrt 138 Allow 1.38m	
(ii)	B1	For a correct reason eg sd = 45.8 cm so girth is nearly 3sd below mean so likely outlier Allow substitution of 50 leading to $g = -6.2$ and suitable reason e.g. this is not possible/this is negative Do not allow substitution of 0.5 to imply the girth is negative	
(f)	M1	for a correct equation. implied by awrt 24	
	A1	awrt 24	

Question Number	Scheme	Marks
2(a)	$S_{gg} = 3624.41 - \frac{144.84^2}{9} [= 1293.4516]$	M1
	$r = \frac{40.25}{\sqrt{1293.4516 \times 1.29}}$	M1
	$= 0.985....$ awrt 0.985	A1
		(3)
(b)	As the population/ t increases, GDP/ g increases oe	B1
		(1)
(c)	$b = \frac{40.25}{1.29} [= 31.20155...]$	M1
	$a = \frac{144.84}{9} - "31.20155..." \times \frac{7.87}{9} [= -11.19068....]$	M1
	$g = ...31.20155...t$	A1
	$g = -11.2 + 31.2t$	A1
		(4)
(d)	The GDP/g increases by (an average of) " 31.2 " billion [dollars] when the population/t increases by one million .	B1
		(1)
(e)(i)	$"-11.2" + "31.2" \times 7$	M1
	$= 207.2...$ awrt 207	A1
		(2)
(ii)	Unreliable as 7 000 000 is much greater than the mean population/ $\bar{t}$ for the 9 years.	B1
		(1)
(f)	$0.1 = "31.2" x$	M1
	$x = 0.003205... \text{million people}$ awrt 0.0032	A1
		(2)
Notes		Total 14
(a)	M1 Correct method for finding S_{gg} (implied by awrt 1290 to 3sf)	
	M1 Correct method for finding r using their S_{gg} If $S_{gg} = 3624.41$ is used here, then M0.	
	A1 awrt 0.985 (correct answer only scores M1M1A1)	
(b)	B1 A correct interpreted contextual statement including population (or t) and GDP (or g). 'Strong positive correlation between population and GDP' on its own is B0.	
(c)	M1 Correct method for finding b	
	M1 Correct method for finding a using their b $a = 16.0... - "31.20155..." \times 0.874... [= -11.19068....]$	
	A1 Only dep on 1 st M1 awrt 31.2 in a regression equation (allow any variables in the equation).	
	A1 Correct equation $g = \text{awrt } -11.2 + \text{awrt } 31.2t$ (must be g and t) Do not allow fractions.	
(d)	B1 Idea that the GDP increases by "Their b " billion dollars for every 1 million increase in population	
(e)(i)	M1 Correct method. Allow substitution of 7 000 000 instead of 7	
	A1 awrt 207 (billion) (isw after an answer of 207 is seen)	
(ii)	B1 Unreliable with a correct supporting comment which must reference t or $\bar{t}$ [=0.874] or population 'Extrapolation so unreliable' on its own is B0. Reference to g out of range is also B0.	
(f)	M1 Equating 0.1 with "their b " $\times x$ Or substituting two values for g with a difference of 0.1 in their equation leading to a value of x	
	A1 awrt 0.0032 (million) Allow awrt 3200 (to 2sf) Do not allow fractions.	

Question	Scheme		Marks
6 (a)	An increase/change of 1°C will allow an extra 2.72 grams [of sugar] to dissolve		B1
			(1)
(b)	151.2 + 2.72 × 90 = 396		M1 A1
			(2)
(c)	The temperature/90[°C] is outside of the range ; so (may be) unreliable		B1 ; dB1
			(2)
(d)	Use of $\bar{y} = 151.2 + 2.72\bar{x}$ So $\sum x = \left(\frac{\frac{3119}{12} - 151.2}{2.72} \right) \times 12 = 479.63235...$		M1 A1
	$S_{yy} = 851093 - \frac{3119^2}{12} [= 40412.9166...]$		M1
	$S_{xx} = 24500 - \frac{479.63235...^2}{12} [= 5329.4005...]$		M1
	$S_{xy} = 2.72 \times 5329.4005... [= 14495.9693...]$		M1
	$r = \frac{14495.9693...}{\sqrt{5329.4005... \times 40412.9166...}}$ or $r = 2.72 \times \sqrt{\frac{5329.4005...}{40412.9166}}$		M1
	= 0.988 *		A1*
			(7)
(e)	e.g. the points lie reasonably close to a straight line/positive correlation and the PMCC is close to 1 therefore supports a linear model		B1 B1
			(2)
	Notes		Total 14
(a)	B1	for a correct interpretation of the gradient in context including grams and degrees	
(b)	M1	for substitution of 90 into the regression line	
	A1	cao 396 on its own scores 2 out 2	
(c)	B1	for a comment that implies the temperature/90[°C] is outside of the range. Allow extrapolation if not linked to 396. (Do not allow comments that imply that 396 is out of range or the use of “it”)	
	dB1	dependent on 1 st B1 for a correct conclusion	
(d)	M1	for clear use of the regression line to find $\sum x$ or $\bar{x}$ (may be implied by 3 rd M1)	
	A1	$\sum x$ = awrt 480 or $\bar{x}$ = awrt 40 (may be implied by 3 rd M1)	
	M1	for a correct expression for S_{yy} May be implied by awrt 40400	
	M1	for a correct expression for S_{xx} ft their $\sum x$ or $\bar{x}$ May be implied by awrt 5330	
	M1	for use of the gradient to find S_{xy} ft their S_{xx} May be implied by awrt 14500 or use of $r = b \sqrt{\frac{S_{xx}}{S_{yy}}}$	
	M1	for a correct expression for r ft their S_{xy} , S_{xx} and S_{yy} or 2.72, ' S_{xx} ' and ' S_{yy} ' If these are not correct then they must be labelled before an expression for r is given for this mark to be awarded	
	A1*	Answer is given so a fully correct solution must be seen	
(e)	B1	for either the points lie reasonably close to a straight line/points or data are linear/positive correlation or the PMCC is close to 1 (Ignore any reference to strength)	
	B1	for both the points lie reasonably close to a straight line/points or data are linear/positive correlation and the PMCC is close to 1 (Ignore any reference to strength) with a correct conclusion	

Question	Scheme	
3 (a)	eg They represent / analyse / solve (real world problems) They improve understanding / simplify (the real world) Can be repeated/ adapted / refined To show the relationship between variables/show trends	B1
		(1)
(b)	$[S_{ww}] = 10036.45 - \frac{339.25^2}{12} = 445.57$ (to 2 decimal places)*	B1*
		(1)
(c)	$S_{td} = \frac{-91.55}{11.5}$ or $S_{td} = 14.4 - \frac{29.5 \times 9.1}{12}$ or $S_{dd} = \frac{445.57}{11.5^2}$ or $S_{dd} = 75.9 - \frac{29.5^2}{12}$	M1
	$S_{td} = -7.9608...$ awrt -7.96 $S_{dd} = 3.3691...$ awrt 3.37	A1A1
		(3)
(d)	$[r] = \frac{-91.55}{\sqrt{445.57 \times 26.43}}$ or $[r] = \frac{"-7.9608..."}{\sqrt{"3.3691..." \times 26.43}}$ $= -0.8436...$ awrt -0.844	M1 A1
		(2)
(e)	As the temperature increases, gas/consumption decreases	B1ft
		(1)
(f)	$b = \frac{-91.55}{26.43} [= -3.4638...]$ $a = \frac{339.25}{12} - "-3.4638" \times \frac{9.1}{12} [= 30.897...]$ or $a = \text{awrt } 28.3 - "-3.4638" \times \text{awrt } 0.76$ $w = -3.46t + 30.9 *$	M1 M1 A1*
		(3)
(g)	$11.5d = -3.46t + 30.9$ oe eg $d = -0.301t + 2.69$	B1
		(1)
(h)(i)	$2.0852... (m^3)$ awrt 2.09	B1
(ii)	the gas consumption will decrease by 0.3 m³	B1ft
		(2)
		Total 14
(a)	B1 for one valid reason why models are used B0 for comments relating to time , cost , and predictions only. "Easy"/ "Easy to use" on its own is B0 Ignore extraneous non-contradictory comments once B1 has been scored.	
(b)	B1* Correct numerical method to find S_{ww} shown with an answer awrt 445.57	
(c)	M1 Method to find S_{dd} or S_{td} (implied by awrt -7.96 or awrt -7.97 or awrt 3.37 or awrt 3.38) A1 awrt -7.96 or awrt 3.37 condone awrt -7.97 or awrt 3.38 for this mark A1 awrt -7.96 and awrt 3.37	
(d)	M1 for a correct method with $S_{ww} \neq 10036.45$ A1 awrt -0.844 correct answer scores 2 out of 2 an answer of awrt -0.843 scores M1A0	
(e)	B1ft for correct contextual interpretation with bold words o.e., ft "their (d)" Allow as gas increases, temp(eration) decreases o.e. B0 for "their (d)" outside the range [-1, 1] Interpreting numerically as a gradient is B0.	
(f)	M1 for a correct numerical expression to find the value of b M1 for a correct method to find a allow "their b " or just b in the expression $w = \text{awrt } 28.3 - "-3.46..."(t - \text{awrt } 0.76)$	
	A1* (dep on M1M1) for correct given equation. Allow $w = \text{awrt } -3.46t + \text{awrt } 30.9$ No fractions.	
(g)	B1 Any correct equation seen (ISW) (allow fractions). allow $d = -0.3t + 2.69$, $d = \frac{-3.46t + 30.9}{11.5}$	
(h)(i)	B1 awrt 2.09	
(ii)	B1ft for decrease awrt 0.3m ³ must include units ft their changed gradient B0 for decrease by -0.3m ³	



Question Number	Scheme	Marks
5. (a)	$y = 6.066 + 0.136 \times 80$ $= 16.946$ (so annual rent is) \$ 16 946	M1 A1 (2)
(b)	$S_{yy} = 3434 - \frac{183^2}{10}$ or $S_{xx} = 84\,818 - \frac{900^2}{10}$ $S_{yy} = \underline{85.1}$ $S_{xx} = \underline{3818}$	M1 A1 A1 (3)
(c)	Need S_{xy} so use b so $S_{xy} = b \times S_{xx} = 0.136 \times 3818$ or 519.248 $[r =] \frac{0.136 \times "3818"}{\sqrt{"3818" \times "85.1"}}$ $= 0.9109448...$ awrt 0.911	M1; A1 M1 A1 (4)
(d)	Since (new $x = 90$ and [original or] new $\bar{x} = 90$) the term $(x - \bar{x})$ will be 0 Therefore (the 11 th shop makes no change) S_{xy} stays the same	M1 A1 (2)
(e)	S_{xx} will be the same so b will be the same New $\bar{y} = \frac{183+15}{11} = 18$ (or a is reduced by 0.3) Equation is $y = 5.766 + 0.136x$	M1 M1 A1 (3)
(f)	$x = 300$ is outside the range $300 \gg 90$ [$300 \gg 90 + 3\sigma = 90 + 3 \times 18.63 \approx 146$] So not suitable (since involves extrapolation) (o.e.)	B1 (1)
[15]		
Notes		
(a)	M1 for substituting $x = 80$ into the given equation A1 for awrt \$ 16 900 (or better)(allow "16.9 thousand dollars"). Must have some units. Condone $y =$	
(b)	M1 for a correct expression for either (can be implied by sight of either correct answer) 1 st A1 for 85.1 2 nd A1 for 3818 or accept 3820	
(c)	1 st M1 for an attempt to use gradient of regression line to find S_{xy} 1 st A1 for awrt 519 2 nd M1 for a correct expression using their values (M0 if $S_{xy} = 900 \times 183 = 164700$) 2 nd A1 for awrt 0.911	
(d)	M1 for stating or showing [old or] new $\bar{x} = 90$ (new $x = 90$ implied) or stating that $(x - \bar{x})$ term = 0 A1 for a fully correct argument mentioning new $x = \bar{x} = 90$ <u>and</u> that extra $(x - \bar{x})$ term = 0 Condone using $\bar{y} = 18.3$ instead of 18	
(e)	1 st M1 for a correct statement about S_{xx} or b (may be implied by 0.136 used correctly) 2 nd M1 for a correct value for new $\bar{y}$ (calculation may be seen in (d) scores here when 18 is used) A1 for $y = 5.766$ (or awrt 5.77 or awrt 5.76) + 0.136x (correct equation scores 3/3)	
(f)	B1 for suitable comparison (must see 300 vs 90 or 3000 vs 900) that says or implies that 300 will be <u>outside</u> the range and therefore <u>not</u> suitable. Not sufficient to just say "larger"	

Question Number	Scheme	Marks
6. (a)	$\{S_{yy} = \} 42.63 - \frac{23.7^2}{16} = [7.524375]$	B1 (1)
(b)	Use of $\bar{y} = 3.684 - 0.3242\bar{x}$; so $\sum x = 16 \times \left(\frac{3.684 - \frac{23.7}{16}}{0.3242} \right) = 108.71067..$ $\{S_{xx} = \} 756.81 - \frac{("108.71..")^2}{16}$; = 18.18435... awrt 18.2	M1; A1 M1; A1 (4)
(c)	$b = \frac{S_{xy}}{S_{xx}} \Rightarrow S_{xy} = "18.1843.." \times (-0.3242) [= -5.8953..]; r = \frac{"-5.89536.."}{\sqrt{"18.184.." \times 7.524375}}$ $= -0.50399... = \underline{\underline{-0.49 \sim -0.51}}$	M1; M1 A1 (3)
(d)	Sub $x = 2$ in the regression line gives $y = 3.0356$	B1 (1)
(e)	St.dev = $\sqrt{\frac{S_{xx}}{n}} = \sqrt{\frac{"18.184..." }{16}} = 1.066...$ So limits are: $\frac{"108.71..." }{16} \pm 3 \times "1.066..." = 3.5965... \sim 9.9929... = \text{awrt } \underline{\underline{3.6 \sim 10}}$	M1 M1, A1 (3)
(f)	The probability of $\underline{x=2}$ being in the range is very small; so Behrouz's estimate is <u>unreliable</u>	B1ft; dB1ft (2)
(g)	Should use regression of x on y to estimate unemployment or equivalent So Andi's suggestion is not suitable <u>or</u> not to be recommended	B1 dB1 (2)
[16 marks]		
Notes		
(a)	B1 Value given so must see sight of a correct expression – allow 561.69 for 23.7^2	
(b)	1 st M1 for clear use of regression line with $\bar{y}$ <u>or</u> $\sum y$ 1 st A1 for $\sum x = \text{awrt } 109$ 2 nd M1 for a correct expression for S_{xx} ft their $\sum x$ 2 nd A1 for awrt 18.2	
(c)	1 st M1 for use of gradient to find S_{xy} 2 nd M1 for a correct expression for r ft their S_{xy} and S_{xx} A1 for an answer in the range $-0.49 \sim -0.51$	
(d)	B1 for sight of $y = 3.03...$ or better. Allow 3.04	
(e)	1 st M1 for a correct attempt at st. dev. ft their S_{xx} or $\sqrt{\frac{756.81}{16} - \left(\frac{"108.71..." }{16} \right)^2}$ ft their $\sum x$ 2 nd M1 for one correct calc...ft their values A1 for a range awrt 3.6~10	
(f)	1 st B1ft for a correct reason ft their range in part (e) eg $x = 2$ is <u>outside</u> the range. Allow extrapolation	
(g)	2 nd dB1ft dep on 1 st B1 for stating a correct conclusion for their range 1 st B1 for a suitable reason based on reg line, eg regression line (y on x) can only be used to estimate wages. Allow x instead of unemployment and y instead of wages 2 nd dB1 dep on 1 st B1 for suggesting not suitable (or equivalent)	

Question Number	Scheme	Marks
5 (a)	Positive (correlation) <u>or</u> e.g. “salary (y) increases as performance (x) increases” [NB “Positive skew” is B0]	B1 (1)
(b)(i)	$19428 - \frac{465 \times 562}{15}$ <u>or</u> $19428 - \frac{261330}{15} = 2006$ (*)	B1cso (1)
(ii)	$[S_{yy} =] \quad 23140 - \frac{562^2}{15}$ $= 2083.7333... \quad \text{awrt } \underline{2080}$	M1 A1 (2)
(c)	$[r =] \frac{2006}{\sqrt{2492 \times 2083.73...}}$; $= 0.8803104... \quad \text{awrt } \underline{0.880}$	M1;A1 (2)
(d)	Is consistent and the points on the scatter diagram lie close to a straight line <u>or</u> r is close to 1 <u>or</u> strong/high (positive) correlation (o.e.)	B1 (1)
(e)	$b = \frac{2006}{2492} ; = 0.80[497...] \quad ; \quad a = 37.46... - "b" \times 31 [= 12.512...]$ $\underline{y = 12.5 + 0.805x}$	M1;A1;M1 A1 (4)
(f)	An increase of <u>1 (performance) point</u> gives an extra <u>£800</u> (1 sf) in salary (o.e.)	B1 (1)
(g)	Line must cross $x = 9$ and $x = 50$ to score either of these marks Line for 9~50 Intercept (extend line if necessary) at “12.5” (accept 11.5~13.5) Line for 9~50 At $x = 50$ $y = 52.8$ (accept 52~54)	B1ft B1 (2)
(h)	For the point (25, 48) circled. (If more than one of the given points circled B0)	B1 (1)
(i)	"12.5" + 30×0.805 [= 36 ~37] <u>or</u> allow 2sf from their diagram Salary of awrt (£) <u>36 700</u> (or 36.7 thousands)	M1 A1 (2)
[17 marks]		
Notes		
(b)(i)	B1 for correct expression, all correct values must be seen (either of the printed expressions) Correct answers to parts (b)(ii), (c), (e) & (i) with no incorrect working score full marks	
(ii)	M1 for a correct expression A1 for awrt 2080 (expect to see 2084 but allow $\frac{31256}{15}$)	
(c)	M1 for a correct expression but ft their $S_{yy} \neq 23140$ <u>or</u> answer only of 0.88 A1 for awrt 0.880 (accept 0.88 from a correct expression with $S_{yy} = [2083 \sim 2084]$)	
(d)	B1 [no ft] for “yes” (o.e.) and a suitable reason based on scatter diagram <u>or</u> value of r	
(e)	1 st M1 for a correct expression for b 1 st A1 for $b = 0.80$ or better (allow $\frac{1003}{1246}$ but not $\frac{2006}{2492}$) 2 nd M1 for a correct expression for a (allow $\frac{562}{15}$ for 37.46... and $\frac{465}{15}$ for 31) 2 nd A1 for correct equation in y and x with $b =$ awrt 0.805 and $a =$ awrt 12.5(no fractions)	
(f)	B1 for a comment mentioning their value in £ of $b \times 1000$ (awrt 1 sf) per performance point Condone use of \$ rather than £	
(g)	1 st B1ft for correct intercept for their line (± 1) 2 nd B1 for $y = 52 \sim 54$ when $x = 50$	
(i)	M1 for using $x = 30$ in their equation ft their a and b to any accuracy A1 for awrt 36 700 (Answer only of awrt 37 000 can score M1A0)	

