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.

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

(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

(i) a Spanish mark for the student who scored 55 marks in their French test,

(ii) 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)

DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA

Question 4 continued

Question 4 continued

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Question 4 continued

(Total for Question 4 is 12 marks)

DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA