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.