1. Typically candidates successfully used the correct formula in order to calculate the product moment correlation coefficient in part (a). However, a number of candidates lost the accuracy mark by only giving a rounded answer to two decimal places. Providing an interpretation of their value of the correlation coefficient was less straightforward. Most frequently candidates made general remarks and described the correlation as positive without relating this to the context of the question. Of those who did attempt to provide an interpretation, many failed to appreciate that it was the attendance at the matches being compared to the total number of goals scored and not the number of home matches that were played.

Part (c) was answered well overall and correct answers were often justified by accompanying statements which indicated that linear coding does not affect the product moment correlation coefficient. Some candidates, however, seemed unaware of this fact and a common mistake was to divide their original product moment correlation coefficient by 100. In addition many candidates failed to recognise the significance of them being asked to write down their answer and chose to perform a full calculation in order to obtain the product moment correlation coefficient, which sometimes led to processing errors.

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. The vast majority scored full marks in part (a). The most common reason for losing marks for the correlation coefficient was for rounding to less than 3 significant figures without having stated the more accurate answer first. A large proportion of candidates still believe that stating 'it's a high level of correlation' will be enough to gain the mark for interpretation. A fully contextual comment is required here, using the named variables of pressure and temperature and not just the letters p and t.

4. This question was usually answered well. In part (b) some did not realise that they needed to check the lower limit as well in order to be sure that 110 was the only outlier. Part (c) was answered very well although some lost the last mark because there was no gap between the end of their whisker and the outlier. Part (d) was answered very well and most gave the correct values for $\sum y$ and $\sum y^2$ in the appropriate formula. A few tried to use the $\sum (y - \bar{y})^2$ approach but this requires all 10 terms to be seen for a complete "show that" and this was rare.

Part (e) was answered well although some gave the answer as –5.7 having forgotten the 10^{-3}, or failed to interpret their calculator correctly. Many candidates gave comments about the correlation being small or negative in part (f) but they did not give a clear reason for rejecting the parent's belief. Once again the interpretation of a calculated statistic caused difficulties.

5. Most candidates knew how to carry out the required calculations in parts (b) and (c) and these were usually completed accurately and with suitable working shown. Although the majority gave an answer of £17 in part (a) £60 and £–3 were sometimes seen. The coding on the variable *m* also caused some confusion with candidates using a value of 261 for $\bar{m}$ and then trying to combine this with the sums of squares given in the question.

In part (d) most knew that the correlation coefficient remain unchanged but some thought the value should be increased by 20 and a few candidates found new values of $\sum m, \sum m^2$ and $\sum tm$ and then seemed surprised when their correlation coefficient was unchanged. In part (e), the commonest response was to simply state that 0.914 represented strong positive correlation whilst 0.178 was weak correlation rather than attempting to interpret the values in terms of time spent shopping and amount of money spent as required. There were a number of sensible practical suggestions offered in response to part (f).

6. Parts (a) and (b) were extremely well answered by candidates; the value of 664 for S_{yy} was occasionally miscopied as 646 from part (a) to part (b). Candidates found it surprisingly difficult to obtain both marks in part (c), with a contextual relationship frequently being omitted. In part (d) the calculation of the mean was straightforward for nearly all candidates. Those candidates who were able to provide a correct formula also accurately found the standard deviation; however, too many candidates at this level were quoting an incorrect formula. Part (e) proved a good discriminator, with relatively few concise solutions; some candidates managed to obtain the correct value of *a* after a page or so of working. Only a handful of candidates were able to see that the number of press-ups is a discrete variable, whereas normal distributions are continuous.

7. 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.

8. Parts (a) and (b) were generally well answered with many candidates gaining full marks. This being said, it was not unusual to see ridiculous values for the correlation coefficient and for candidates to follow this through into part (c). Many candidates realised that the value of the correlation coefficient would be the same in (c)(i) and those that attempted (c)(ii) often did so without reference to the context of the question.