Section 3: Bivariate data

Notes and Examples

These notes contain subsections on

- **Scatter diagrams and correlation coefficients**

Scatter diagrams and correlation coefficients

You probably remember meeting scatter diagrams at GCSE level. You probably used them to gain a visual impression of whether there was any correlation in a set of bivariate data. You probably also drew lines of best fit by eye and may have used them to predict values.

In your work using large data sets, you will be able to use a spreadsheet to produce a scatter diagram and also to calculate a correlation coefficient. This is a value between -1 and 1 which gives a measure of how good the correlation is.

- A correlation coefficient of zero means that there is no correlation
- A correlation coefficient of 1 means that there is perfect positive correlation and so the data lie on a straight line with positive gradient
- A correlation coefficient of -1 means that there is perfect negative correlation and so the data lie on a straight line with negative gradient.

Although the calculation of a correlation coefficient will give you more reliable information than the “by eye” methods you used at GCSE, scatter diagrams are still an important part of the process.

- The scatter diagram allows you to spot any obvious outliers which may affect the results of your calculation.

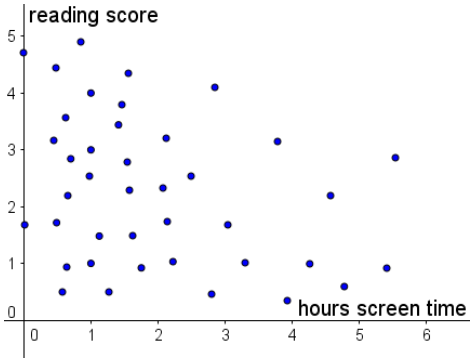
- You can get some idea of the degree of correlation, and whether it is positive or negative, from the scatter diagram, which might allow you to spot if you have made an error in your spreadsheet calculation such as selecting the wrong set of data.

- You can see any other features of the data. For example the points might form two distinct groups.

Example 1

A newspaper publishes an article about children’s use of screens. It includes the scatter diagram below which shows the number of hours of screen time per day for a sample of 40 seven-year olds and their reading scores.

Edexcel S1 Correlation 1 Notes and Examples



The correlation coefficient for the sample is -0.36.
The newspaper article states that the data show that screen time has a negative effect on children's reading scores.
Comment on this claim.

Solution
The scatter diagram and correlation coefficient show that there is weak negative correlation between the number of hours of screen time and the reading scores.
However this does not mean that screen time has a negative effect on a child's reading ability.

In the example above, the value of the correlation coefficient suggests that there may be a **connection** between the two variables. It does **not** mean that screen time has a negative effect on a child's reading ability. For example, it might be the case that children who read well spend less time on screens because they enjoy reading books, or perhaps parents who spend time helping their children with reading may also be those who are more likely to restrict their children's screen time. Issues such as these are very complex, and all that the correlation coefficient shows is that there is a slight connection between the two variables.

This illustrates a very important general point, that **correlation does not imply causation**.