Activity description

Students use an online statistics module to learn about regression and correlation

Suitability and Time

Level 3 (Advanced), 2–4 hours

Resources and equipment

Student sheets, computers with Excel and internet access, slideshow (optional)

Key mathematical language

Scatter graph, dependent and independent variables, regression line, line of best fit, least squares method, standard error, correlation (positive, negative, linear, strong, weak), Pearson's product moment correlation coefficient, interpolation, extrapolation, causation.

Notes on the activity

The slideshow can introduce the activity. It includes a weblink to the DISCUSS module students will use (http://www.icse.xyz/discuss/regression/). You may not need students to study some of the more advanced sections of this module which are time-consuming.

To save time, ask students to omit the following sections if you wish:
Lines webpage: omit the 'Scaling' and 'Other Criteria' sections.
Goodness of Fit webpage: omit use of the spreadsheet in the 'Coefficient of Determination –r^2' section.
More correlation webpage: omit use of the spreadsheets in the 'Correlation and Significance' section.
Assumptions webpage: omit use of the spreadsheets in the 'Checking Assumptions' section.
Predictions webpage: omit all except the first 4 lines of the 'Margins of Error for Predictions' section.
Questions on the last slide of the slideshow help students reflect on their work.

During the activity

Students could work in pairs.

Points for discussion

Discuss the use of a scatter graph to find the relationship between two variables. Stress that the relationship could be linear or non-linear, so the use of a straight line of best fit and Pearson's correlation coefficient may not be appropriate in some cases.

Students should appreciate that strong correlation does not always mean that a change in one of the variables causes a change in the other.

Discuss the use of interpolation to estimate values, but point out the dangers of extrapolation.