

## Trends in Two-Variable Data Sets

### Terminology

**Line of Best Fit** – A line that passes through, or as near as possible to, the points on a scatter plot.

**Regression Equation** – The equation of a line of best fit.

**Correlation Coefficient** – A value between -1 and +1 which represents how close the points in a data set are, on average, from the line of best fit. The closer the value of r is to -1 or +1, the closer the fit. The correlation coefficient does not indicate whether there is a causal relationship between the variables.

**Interpolation** – To estimate or calculate a value between two other values in a data set.

**Extrapolation** – To estimate or calculate a value beyond a given data set.

**Outlier** – An extreme value in a set of data or a value separated from the main body of data, which does not follow the trend of the other values in the data set.

**Influential Point** – A data point with extreme values that greatly affects the slope of the regression line. Does not decrease the coefficient of determination, r<sup>2</sup>, rather it increases it.

**Cause and Effect Relationship** – A relation where a change in the independent variable has a predictable effect on the dependent variable.