Residuals

In a scattergraph of bivariate data, each point is a certain vertical distance from the trend line. The difference between the actual *recorded value* of the response variable, and the *value predicted by the trend line* is called a **residual**.

Residual = observed value − trend (predicted) value

If a point lies above the trend line, the value of its residual is positive; if a point lies below the trend line, the value of its residual is negative.

Calculating and graphing residuals by spreadsheet

The following explanation uses the spreadsheet for Oregon settlement population versus annual waste (1998), with figures rounded to the nearest whole number. This data set is available at ▶ESA Online .

Population (people)	Disposed (tons)	Population (people)	Disposed (tons)
16 700	12 376	313 000	261 958
71 860	45 551	43 200	41 127
34 700	30 716	107 090	75 807
42 300	23 004	29 200	20 052
61 400	37 434	271 750	237 166
16 650	14 232	1 363 100	1 196 486
22 000	12 264	6 500	5 586
104 900	101 313	9 400	5 893
100 300	86 369	58 940	36 790
2 100	1 320	1 900	1 092
8 000	3 174	24 000	15 063
7 600	1 794	60 600	52 484
19 500	14 931	24 400	15 610
172 800	136 337	7 200	4 526
8 709	8 709	22 600	17 997
73 000	40 518	1 600	359
62 000	62 603	82 460	68 901
7 400	6 361	Source: QELP	

Annual waste disposal in settlements in Oregon 1998

$y = 0.8784x - 4763.6$
$R^2 = 0.9987$
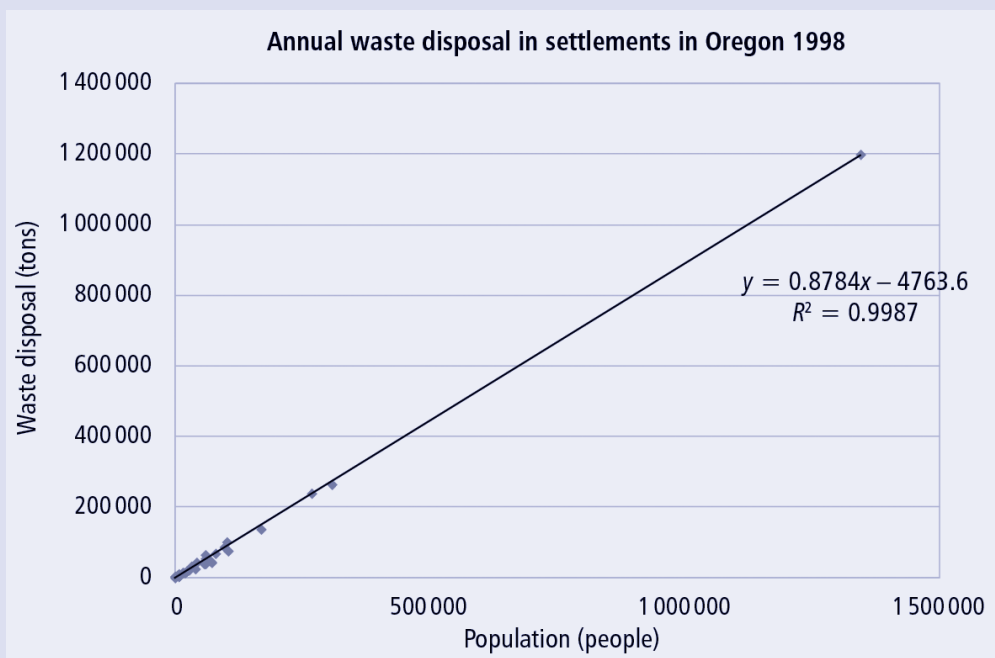
Finding trend (predicted) values on Excel

In cell C1 type: Trend values

In cell C2 type: tons

In cell C3 type: $=0.8784*A3$ (or click on the cell A3) $- 4763.6$

Press **Enter** (the number 9 906 should appear). This is the annual number of tons of waste predicted for a settlement of 16 700 people, found by using the regression equation).

Click on cell C3 and click on the small black square at the bottom right-hand corner (a cross should appear).

Left-click on the cross, and holding the mouse key down, drag the cross down to cell C37. When you release the mouse key, cells C3–C37 should be filled with predictions from the regression equation.

Calculating residuals on Excel

In cell D1 type Residuals

In cell D2 type tons

In cell D3 type $=B3-C3$

Press **Enter** (the number 2470 should appear)

Click on cell D3, hold and drag down (as done for column C) to D37

Cells D3–D37 should be filled with the values of the residuals.
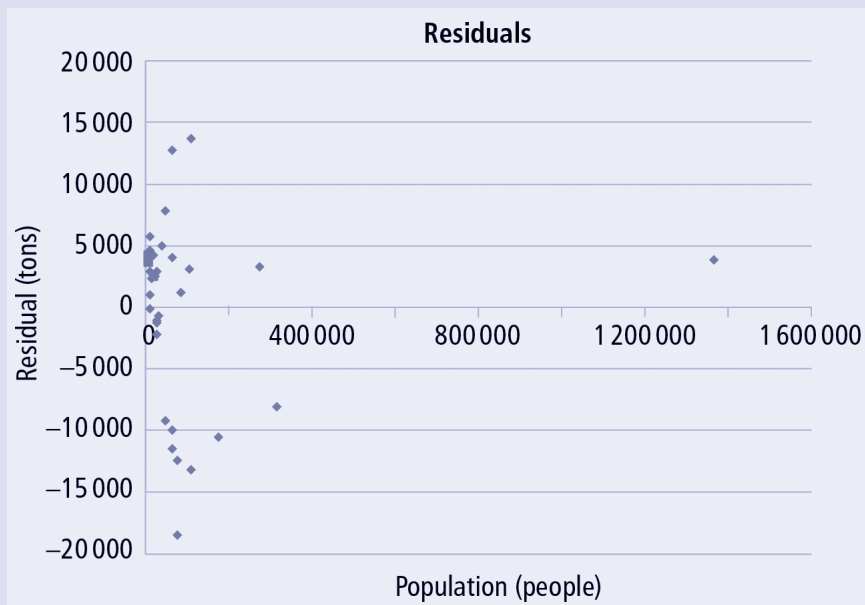
Graphing the residuals

Click on the **A** at the top of column A (column A should now be highlighted)

Press **Ctrl** on the keyboard and click on the **D** at the top of column D

Press **Insert** then **Scatter** selecting the scatter plot with unconnected points (top left).

As before, insert a title for the graph, and label the axes.

Your graph should look like this:

Interpreting the residuals

The graph of the residuals can be used to decide whether a linear trend line is an appropriate model for the relationship between the variables.

- If the residuals appear *randomly scattered* above and below the *x*-axis, then a linear relationship between the variables you are investigating would be appropriate.
- If the residuals form a *pattern* when graphed, then a linear relationship between the variables you are investigating would not be appropriate.

Example

In the graph of residuals shown below, the residuals form a curved pattern.

Residuals for a power function given a linear trend line
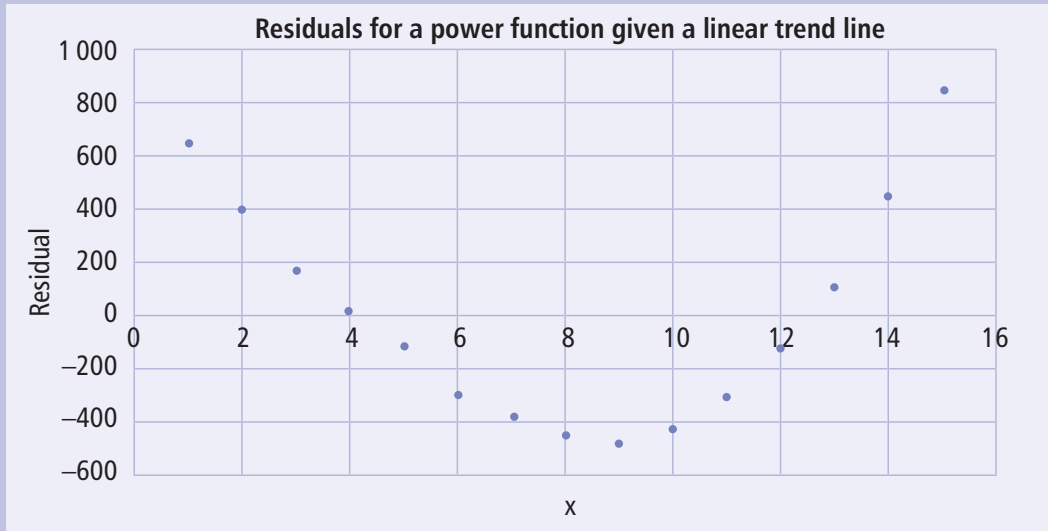
This means that a linear trend line is not appropriate for the data.

Exercise: Residuals

Graph the residuals for the following data sets (you have already found the equations of the trend lines in Activity 5B pages 77–80). Interpret the residuals.

1. Columbia River

_____

_____

_____

_____

2. Wrist/neck measurement

_____

_____

_____

_____

3. Sacramento Valley drainage

_____

_____

_____

_____

4. California power plants

_____

_____

_____

_____

5. Zinc v lead concentrations

_____

_____

_____

_____

6. Rainbow Glacier retreat

_____

_____

_____

_____