# STATISTICS 3.9

## Relating to page 54 of *Level 3 Statistics Learning Workbook*

### *Drawing a scatter plot on iNZight*

The statistical tool iNZight will draw a scatter plot, fit a trend line and give you the equation and the correlation coefficient (*r*) very easily (see Chapter 1: Time series for guidance on how to format data files to import into iNZight).

**Example**

**Import the data file:**

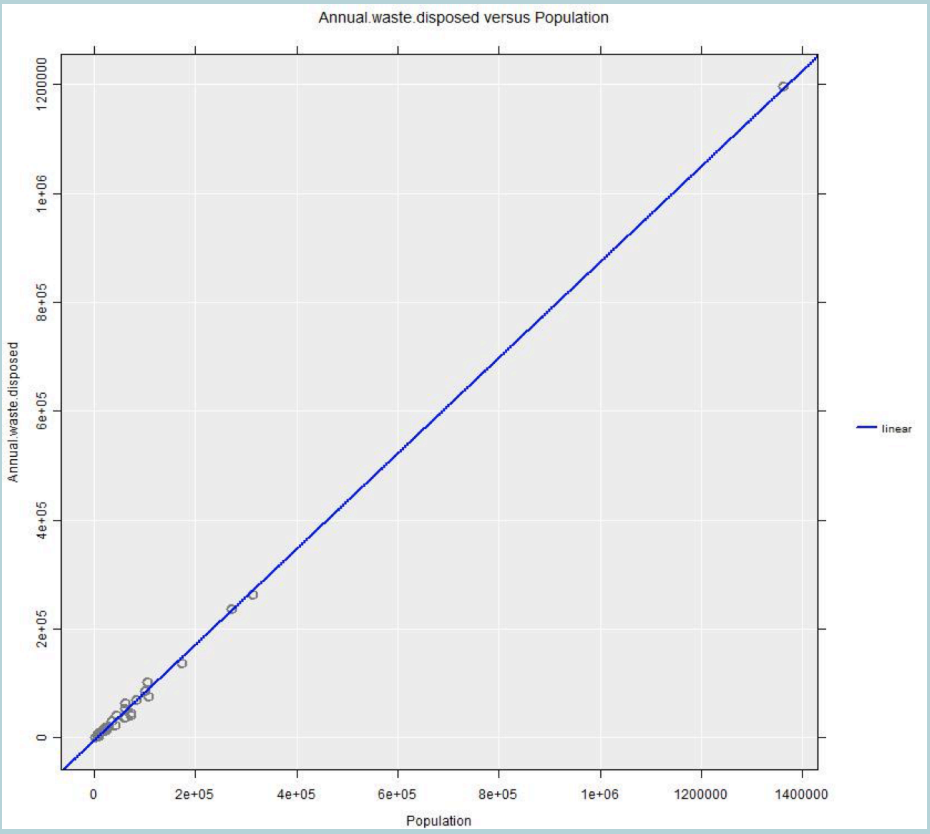
Oregon waste disposal.csv RESOURCES

**Creating the scatter graph**

- Click on **View Variables**
- Click on **Population** and drag to **select/drag-drop** for Variable 2 (this means that 'Population' will be on the horizontal axis in the scatter plot). Alternatively select Population from the drop-down menu.
- Click on **Annual waste disposed** and drag to **Select/Drag-drop** for Variable 1 (this means that 'Annual waste disposed' will be on the vertical axis in the scatter plot) Alternatively, select Annual.waste.disposed from the drop-down menu.

**Adding a line of best fit**

- Click on  symbol (**Add to plot**) at the bottom right of the screen
- Choose **Trend Lines and Curves** in the drop-down menu beside **Add to plot**
- Tick box beside **Linear**



**Finding the equation of the line and *r***

- Click on **Home** at the bottom of the screen
- Click on **Get Summary**

Your screen should now show:

```
Linear trend:

    Disposed = -4764 + 0.8784 * Population
    Linear correlation: 1
```

This means that the annual waste disposed in Oregon in 1998 can be estimated by multiplying the population by 0.8784 and subtracting 4764.

The correlation coefficient *r* is 1, meaning there is a perfect correlation between the two variables.

## *Drawing a scatter plot on NZGrapher*

*The following instructions are used to produce a scatter graph for Annual solid waste disposal in settlements in Oregon, 1998.*

First upload your file of data. In the area above the spreadsheet

select **Choose File** and

upload **Oregon waste disposal.csv**

In the area below the table, select your graph type and variables:

In **Graph Type** box select **Scatter graph** from the drop-down menu

In **Variable 1** box select **Population** from the drop-down menu

In **Variable 2** box Select **Disposed** from the drop-down menu

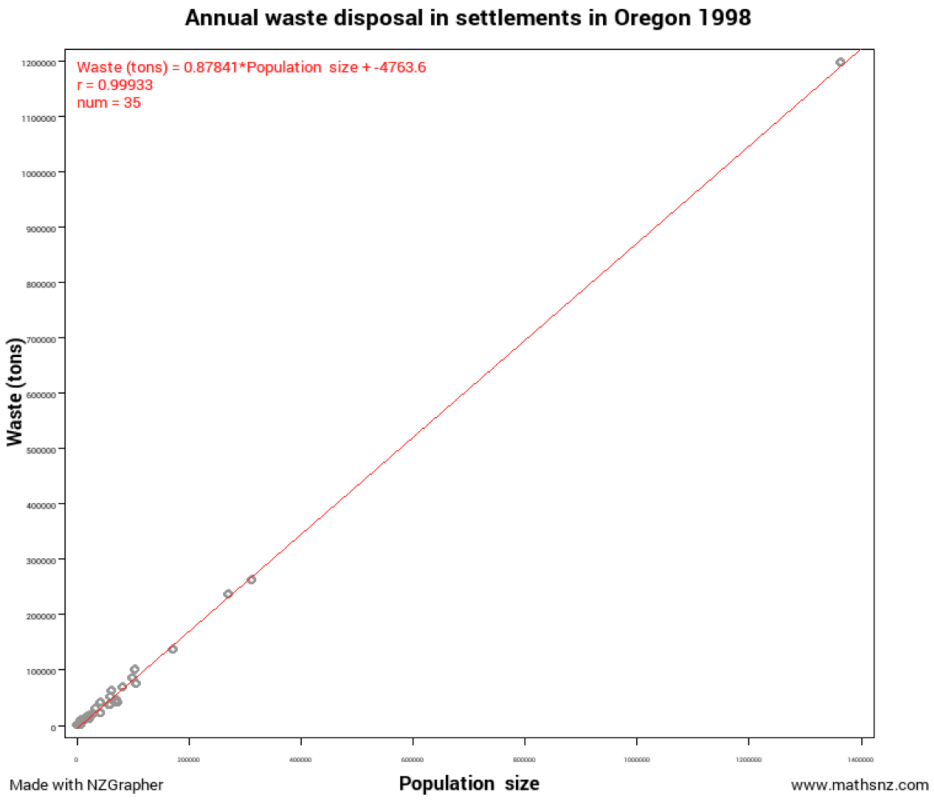
This produces a scatter graph.

Next add a trend line and equation. In the area below the graph

select **Regression line**

adjust labels for **title** and **axes** as desired and press **Update Graph**

**Right click** on graph to copy.



## *Drawing a scatter plot on Excel 2010*

The following example illustrates the process of using the spreadsheet Excel to create a labelled scatter graph, fit a trend line and calculate an $R^2$-value.

Open the file **Oregon waste disposal**.
(The data for this file is available on the ESA website RESOURCES.)

### Creating the scatter graph

- Using the mouse, highlight the cells A3, B3–A37, B37
- Click **Insert**, **Scatter** and then select the scatter plot with unconnected points (top left option).

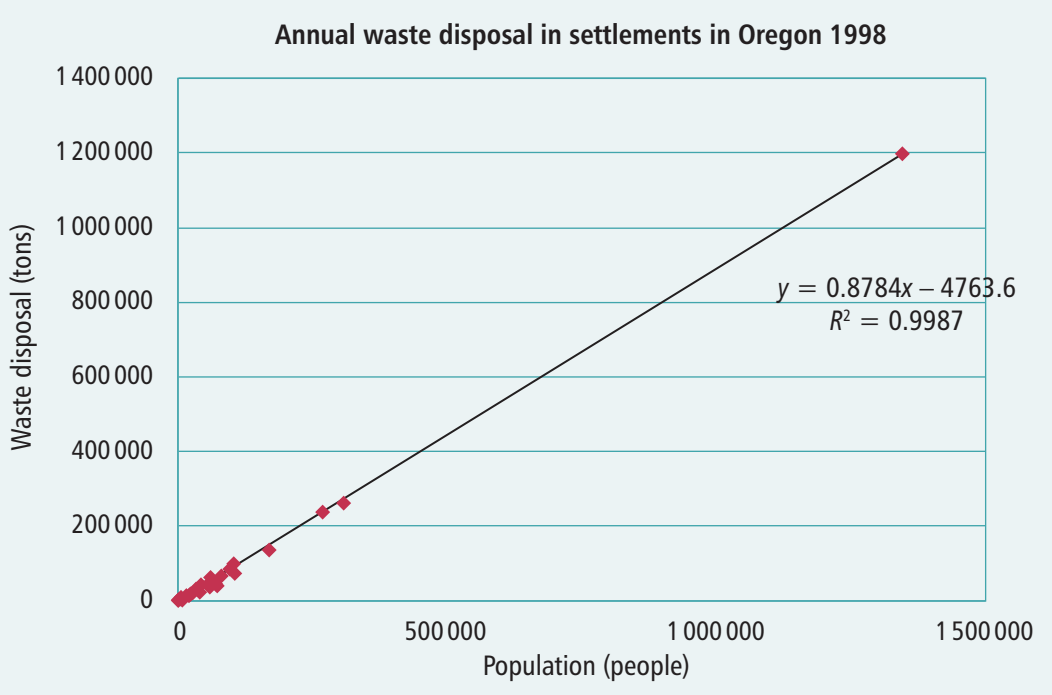
### Adding labels for the axes and a title

- Select scatter graph (click anywhere on graph), then click on Layout.
- Click on **Chart Title** and select placement; then type 'Annual waste disposed in settlements in Oregon 1998' and press enter.
- Click on **Axis Titles**, select vertical axis title and placement, then type 'Annual waste disposed (tons)'.
- Click on **Axis Titles**, select horizontal axis title and placement, then type 'Population (people)'.
- Remove 'Series1' from graph by right-clicking on the word and selecting delete.

### Adding a line of best fit

- Select graph (or right-click on any point), then click on **Layout** then Trendline then **More Trendline Options**.
- Select **Linear**, **Display equation on chart** and **Display R-squared value on chart**, then **Close**.

Note: You can move the equation and $R^2$ value by clicking on them and dragging them to the side of the graph.




On your scatter graph you should have a line passing through the points, its equation ($y = 0.8784x - 4763.6$) and the coefficient of determination value $R^2 = 0.9987$.

The correlation coefficient $r = +0.9993$ (line slopes up so take the *positive* square root of $R^2$-value).