MIS771 Descriptive Analytics and Visualisation	DEAKIN UNIVERSITY
Topic 4 Tutorial – Simple Linear Regression	

Introduction

In this tutorial you will cover simple linear regression.

Specifically, the aims of this tutorial are to:

- To generate a correlation analysis to help identify a linear relationship.
- To generate a scatter plot to help confirm a linear relationship.
- To generate a simple linear regression model to illustrate the key features of simple regression modelling.
- Construct and interpret residual plots.

Scenario

We continue with the analysis of the BLITZ employee survey data set (Note that this is a new dataset).

Management is concerned about the wide variation in productivity amongst employees, and suspects the amount of unpaid overtime hours worked may play a part in this problem. To further investigate this issue, the company has conducted a survey of 48 of their staff, collecting data on several factors that may be relevant to employee productivity.

Management has asked you to investigate the data and examine if their suspicion about employee productivity is realised. Specifically, you are asked to develop a model to help predict productivity based on the number of unpaid overtime hours.

Step 1. Open the data file and install the Data <u>Analysis Tool Pak</u>

a) Download the file **BLITS_Dataset_Tut4.xl**s from

"Content → Learning Resources → Topic 4 Folder" in Cloud Deakin. **Save it** to your hard drive.

b) Open the file in Excel.

c) Install the data <u>Analysis Tool Pak</u>.

Instruction:

From the top of *Excel* (*Microsoft Office Ribbon*), click on **File** tab (Figure 1a), select **Options** (Figure 1b), choose **Add-ins** (Figure 1c), and then press **Go…** button to *manage excel add-ins* (Figure 1c). Finally, select **Analysis Tool Pack** and press **OK** (Figure 1d).

Step 2. Scatter diagrams and Correlation analysis

(a) Construct a **scatter diagram** between <u>Productivity</u> and <u>Unpaid Overtime</u> along with a **correlation analysis.**

See Tutorial 1 instructions for how to create a scatter diagram.
To calculate correlation coefficient, go to working worksheet, in a cell type the following:

> *=CORREL (UnpaidOverTime, productivity)*
> Note: instead of typing variable labels, you can select the range of values for each variable.
> *=CORREL ('Survey Data'!D1:D49,'Survey Data'!K1:K49)*

(b) Using your diagram in (a) and the fact that r = -0.808, how would you describe the relationship between the variables.

Step 3. Regression Analysis

(a) Build a **regression model** between <u>Productivity</u> and <u>Unpaid Overtime</u>.

Instruction:

Make sure you are on the *Working worksheet*. From the *Data* tab, select *Data Analysis* and then choose <u>Regression</u>. The pop-up window in Figure 1 will appear. Enter the **Y-Range**, **X-Range**, select the **Labels checkbox** and specify the **Output Range** as shown in Figure 1. Finally, press *OK*.
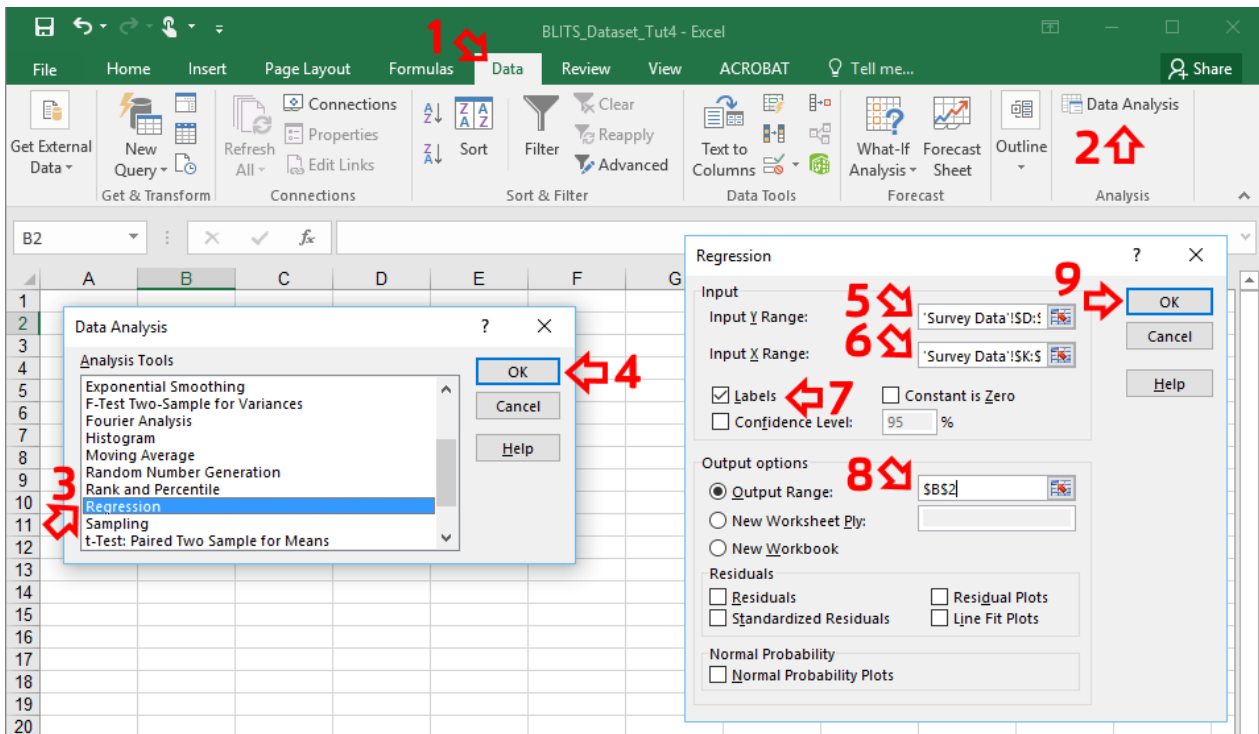

Figure 1.

You should obtain the following equation:

$$\text{Prdtvty} = 107.961 - 1.283 \times \text{UOvTime}$$

(b)   Explain, in practical terms, the values of b_0 and b_1 in the above equation.

(c)   From your regression output, write down the practical interpretations of for R^2 and s_{yx}.

Step 4. Residual Analysis

    (a)   Check the model you created in question 3 does not violate any of the underlying assumptions of regression by doing a **residual plot(s)**.

Instruction:

Make sure you are on the *Working worksheet*. Repeat **Step 3**. This time set the **Output Range** to *Working!B54* and select the **Residuals**, **Standardized Residuals** and **Residual Plots** checkboxes (see Figure 2).
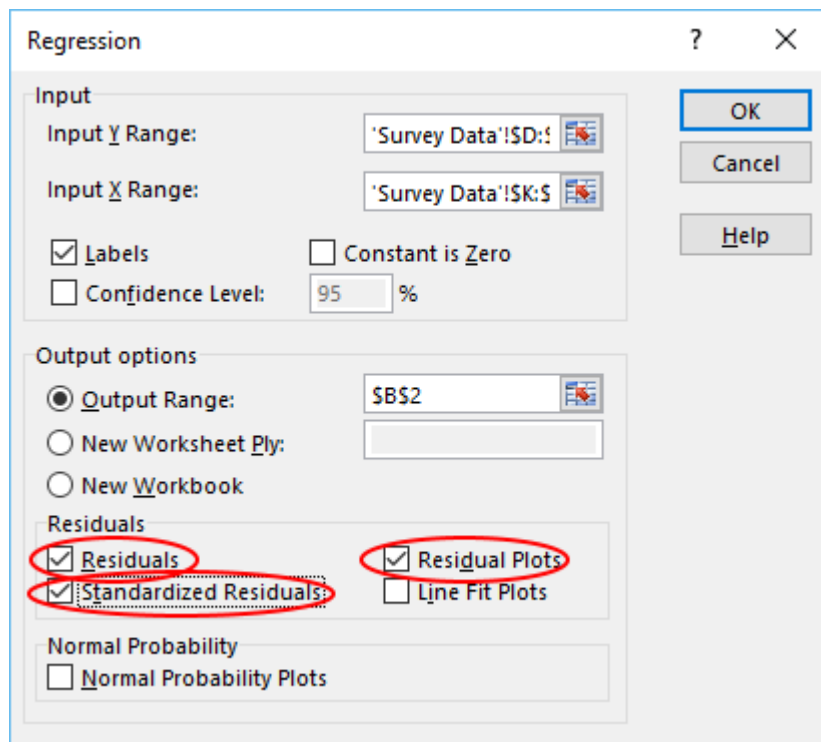

Figure 1.

    (b)   Interpret the residual plot.

    (c)   Are there any outliers (influential values)?

Step 5. Use the Regression model

    (a)   Use a calculator to predict the productivity on an employee who works 10 hours of unpaid overtime. (That is, determine a point estimate).

(b) Interpret the following interval estimates for an employee who works 10 hours of unpaid overtime.

	95% Lower	95% Upper
Confidence Interval	94.026	96.245
Prediction Interval	89.743	100.528

(c) Comment on the reliability of the result if we were to use the model to predict the productivity of an employee working 20 hours of unpaid overtime.