Appendix 5: Microsoft Excel 2013 Tutorial

Spreadsheets are widely used to record, plot and analyze experimental data due to their power and ease of use. They are very common technical computer software tools used by engineers to tackle sophisticated computations and produce detailed optimization studies of real data. This introductory document will provide the means to start producing proper engineering reports that can be reviewed and understood by other engineers. In this instance, the computer software to be used to create and manipulate spreadsheets is Microsoft Excel 2013.

The majority of the material covered in this tutorial will be used in the APSC-100 Module 2 course; however, many formatting and computational techniques will be used throughout your entire engineering career. Particular attention should be given to the graph and table labelling sections as these are of paramount importance in professional documentation. The marking of assignments and reports will reflect this importance. The first section (Section 5.1) is intended only for those who have never used Excel. Other students may begin the tutorial at Section 5.2. There is also an important summaries in Section 5.5 and Section 5.6 that can be used as quick refreshers for upper year students.

5.1 Entering Data into Excel

This section is a brief introduction to making tables and performing calculations in Excel and is intended for those who have never used Excel.

- Launch Microsoft Excel 2013.
- Select (**left-click**) the cell where you wish to place the worksheet title. This title usually exists in cell A1.
- Type in the title and press **<Enter>** (pressing *Enter* indicates you are finished editing the contents of a cell).
- Type in column titles in the same manner, while ensuring units are included. The arrow keys can be used to move between cells.
    - An important visual characteristic of your Excel tables is the column width. For cells with longer entries, put the cursor on the line separating the letters at the top of the Excel window, changing the cursor to a vertical line with arrows pointing in opposite directions. Click and drag the cursor to the right to adjust the column width manually, or double-click to auto-adjust the column width to the longest entry.
- Type in all raw data by selecting the cell under the correct column and typing in the numbers only, NO UNITS.
    - If there is a pattern to the raw data you are entering (i.e.: you want to calculate the areas of circles with radii of 1m, 2m, 3m, 4m etc.), type in the first few entries into the column. Then, select all the column entries thus far by clicking the middle of the topmost cell and dragging until all the column entries are highlighted and surrounded by the green border, as shown in Figure 1. Release the click, then move your cursor to the bottom right of your selected cells where there is a small green square, changing the cursor to a black plus sign. Click and drag down to populate as many lower cells as you intend.

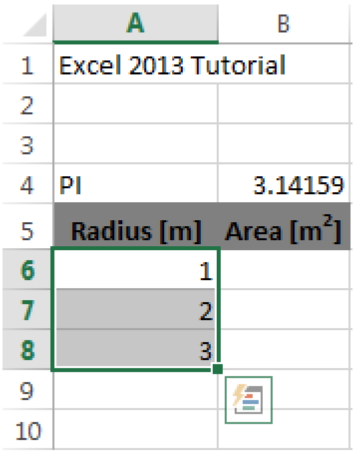

Figure 1: Selecting a group of cells.

- To enter a formula (calculation) into a cell, select the cell and begin by typing "=". Use parentheses, operations (+, -, *, /, ^) and click on cells you wish to reference in the formula, and evaluate the cell by pressing **<Enter>**. An example formula can be seen in Figure 2.

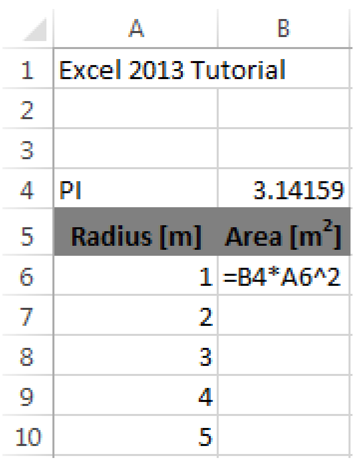

Figure 2: Typing in a formula into a cell.

  - When you press enter, your cell will evaluate and the formula will disappear. To edit the formula at a later time, select the cell in which the formula was used and press **F2** or click in the formula bar just above the column headers.
- To copy a formula to a group of rows, select the cell containing the original formula, then move the cursor to the bottom right corner of the cell where the small green square is located, changing the cursor to a black plus sign. Click and drag the cursor down to the last row in which you want the formula, or double-click to have the formula populate automatically to the lowest row with data. One can also copy the cell (select the cell and press **<Ctrl> and C**) and the paste it in the cells intended for evaluation (select all the cells you wish the formula to populate, then press **<Ctrl> and V**)
  - When copying a formula to multiple rows, Excel will keep shifting the cells used in the formula down by one row. This method is called relative referencing. If you wish to always

evaluate the formula referencing one particular cell (absolute referencing), edit the formula such that instead of calculating using cell ***B4***, the formula will read ***B4*** (try this method so that the cell containing π will always be used in each row!)

- To adjust the number of decimal places a cell shows, select the cells you wish to format and use the **One Less Decimal** or **One More Decimal** buttons located in the **Number** group of the **Home** tab, as shown in Figure 3.

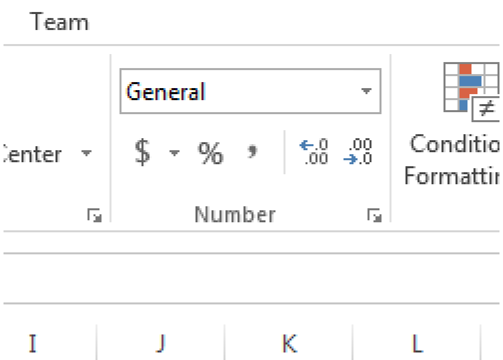

Figure 3: The Number category of the Home tab at the top of the Excel window.

  - You can also define what type of number the cell contains using the drop down menu in the same **Number** group in the **Home** tab. You can select from **General, Number, Scientific, Percentage**, and others. Use this formatting to show your data in the most appropriate method (in most cases, General will suffice).
  - You can also format cells further by selecting the group of cells you wish to format, **right-clicking** to bring up the Quick Menu, and selecting **Format Cells…**

5.2 Displaying Experimental Data

In this section, some of the fundamentals of Excel will be explained, including the following operations:

- Setting up a spreadsheet with a proper identification section
- Creating, editing and saving a file
- Creating properly structured/formatted graphs
- Printing a spreadsheet and a graph
- Plotting theoretical curves and experimental data on the same graph

It is recommended that you create a separate folder for the exercises performed in this tutorial for future reference. A useful title for this folder would be "*APSC100 Module 2 Excel Tutorial*".

It is most useful to become familiarized with Excel through the use of real situations, such as the collection of data during a first year physics experiment called The Super Ball Lab.

5.2.1 Basic Spreadsheet Operations

You are asked to compare the **Elasticity** of a ping pong ball and a rubber ball from the data collected during a lab. Elasticity is defined as the ratio of bounce height, h_b, to initial drop height, h_i. The initial and bounce heights, along with the associated measured error, of the rubber ball and the ping pong ball can be seen in Table 1.

Table 1: Experimental data for the bounce heights of the rubber ball and ping pong ball.

	Rubber Ball				Ping Pong Ball			
Trial #	**h_{ri} [cm]**	**h_{ri} Error [cm]**	**h_{rb} [cm]**	**h_{rb} Error [cm]**	**h_{pi} [cm]**	**h_{pi} Error [cm]**	**h_{pb} [cm]**	**h_{pb} Error [cm]**
1	20	0.5	16	2	20	0.5	13	2
2	30	0.5	25	2	30	0.5	21	2
3	40	0.5	33	2	40	0.5	27	2
4	50	0.5	42	2	50	0.5	32	2
5	60	0.5	49	2	60	0.5	42	2
6	70	0.5	57	2	70	0.5	47	2
7	80	0.5	66	2	80	0.5	51	2
8	90	0.5	72	2	90	0.5	60	2
9	100	0.5	81	2	100	0.5	67	2
10	110	0.5	90	2	110	0.5	75	2
11	120	0.5	98	2	120	0.5	80	2
12	130	0.5	105	2	130	0.5	86	2

In this scenario, it is expected that the initial and bounce heights relate according to the following equation:

$$h_b = E h_i$$

where h_b is the bounce height (in m), *E* is the elasticity of the ball (dimensionless), and h_i is the initial drop height (in m). It should be noted that this relationship resembles the equation of a straight line intersecting the origin.

Exercise: The first step in analyzing data is to create a spreadsheet. Launch Microsoft Excel 2013 and select the "New Blank Workbook" option, producing a blank Excel file with three empty worksheets. The following sections of the tutorial will utilize two columns: the left column will describe the intended action and the right column will proceed through the steps required to accomplish the intended action.

Always start a spreadsheet with an **identification section**. This section includes a **title**, the **author's name**, the **date** of creation, and the **file name**.

Click on cell A1 of **Sheet 1** to make it the active cell. Type **Super Ball Experiment <Enter>**. The cursor should then move to Cell A2. Type in your name and "Exercise 1-1.xlsx" in cells A2 and A3 respectively. Select cell C2 and type today's date in the format **2014/05/15 <Enter>**.

Note that Excel has multiple ways of expressing the date. Right-click on the cell containing the date, then select "**Format Cells…**" and select the **Number** tab on the resulting menu. Click on the **Date** category and select the **Type** in which you wish the date to be expressed, such as **May 15, 2014**. Click OK. The date should now appear as **May 15, 2014**.

The Worksheet name appears at the bottom left of the Excel Window – you will see "Sheet 1, Sheet 2, Sheet 3". It is helpful to name your worksheets, particularly in larger files with many worksheets.

At the bottom of the page, double-click on the title **Sheet 1** (or right-click and select **Rename**). Type in the name **Exercise 1-1**, then press **<Enter>.**

5.2.2 Saving the File

It is of extreme importance that you save your Excel workbook as you progress through your analysis to ensure your work is not lost for any reason.

Save the file under a desired and informative name and into the desired folder that you previously created.

Select the **File** tab at the top left to go to the **Backstage**. Select **Save As** and navigate to the correct directory using **Browse** to locate your folder (either **OneDrive** for cloud storage or **Computer** for local storage). Type **Exercise 1-1** for the File Name and save it as an Excel Workbook type to give in a **.xlsx** extension, indicating that it is an *Excel 2013* document.

The saving process can also be expedited by using the keyboard shortcut "**Ctrl**" **and** "**S**", which automatically saves the document to the directory specified at the first save. Use this shortcut to quickly save while progressing through your analysis.

5.2.3 Setting up the Spreadsheet

Entering and plotting data are the first steps in data analysis using Excel. Open the spreadsheet named Exercise 1-1 and prepare a table similar to that shown in Table 1. Make sure to include column headings with proper units and use correct formatting.

Subscripts and Superscripts: The column headings need to be formatted to show the appropriate subscripts. Double-click on the cell containing the heading "**hri**". Highlight the letters "**ri**" either by using the mouse to click and drag or by holding **Shift** and using the **arrow keys**. In the **Home** tab under the **Font** group, bring up the **Font** menu by clicking the arrow in the bottom right corner of the group. In the **Font** tab on the resulting menu, tick the **Subscript box**. Then press **OK** or press **<Enter>**.

Merge and Center: It may be of use to have multiple cells merge into one, as was done in the "Rubber Ball" and "Ping Pong Ball" data headers in Table 1. To accomplish this manipulation of cells, select all the cells you wish to merge, then click the **Merge and Center** button in the **Alignment** group of the **Home** tab.

5.2.4 Tables and Graphs Formatting

Tables and graphs are two common ways of presenting data in engineering. You may want to turn on gridlines and print them when printing raw data. Your data also appears more professional when borders are used around tables and graphs.

- To turn **gridlines** on to print, select the **Page Layout** tab. Under the **Sheet Options** group, check off the **Print** box under the **Gridlines** section. Note that this button will also place dotted lines on the spreadsheet indicating page boundaries. To eliminate these boundaries while you are working, select the **File** tab in the top left to bring up the **Backstage**, then click **Options** in the left hand menu to bring up the Excel Options menu. Select the **Advanced** tab in the left menu, then scroll down to deselect the "**Show page breaks**" box under **Display options for this worksheet**.
- To create borders around tables in your spreadsheet, use the **Borders** button (shaped like a 2x2 table) under the **Font** group in the **Home** tab. First, highlight the cells on which you would like to place the borders. Then select as many border options as necessary.

Let us move on to plotting the Super Ball data. For this example, a scatter plot will best serve the purpose of data analysis.

The **Scatter Plot** will be used for many future engineering applications. Do not confuse it with the **Line Chart**. It should be noted that the initial height, h_i, is the independent variable and should be placed on the horizontal axis, whereas the bounce height, h_b, is the dependent variable and thus should be placed on the vertical axis.

On the **Charts** group in the **Insert** tab, select **Scatter with only Markers**. Right-click on the resultant blank plot and click on **Select Data…** in the Quick Menu. In the **Select Data Source** window, click on **Add** to add a series of data to the plot. The first series will be of the rubber ball, and hence type in "**Rubber Ball**" in the **Series Name** dialogue box. In the **Series X values** dialogue box, click the **Collapse** button beside the box. Now, select all the numbers (the column under the column header) representing the h_{ri} values for the rubber ball, then press **<Enter>** to return to the Edit Series window. In the same manner, select all the h_{rb} values for the rubber ball for the **Series Y values**. Click **OK** to return to the Select Data Source window.

If the resultant plot is sitting on top of your data table, simply click and drag the plot window away from the data to another location in your spreadsheet. It is now time to add the second set of data points to the same graph. Use the same procedure as that of the rubber ball to add the series for the ping pong ball. The resultant graph should possess two sets of data; however, it needs to be formatted to look professional.

Add a **title, axis labels**, and a **legend** to your scatter plot. It should be noted that in professional reports, titles are NOT included on graphs; however, it is useful to include titles in your Excel worksheet to distinguish graphs during large data analysis applications with multiple graphs. To delete the title when you copy the Excel plot into Microsoft word during your report write up, simply select the title and press **<Delete>**.

Left-click anywhere on the chart and note that a green plus sign appears at the top left corner. Click on the green plus sign and check off the **Axis Titles, Chart Title** and **Legend** boxes. The small arrow to the right of the boxes allows for specifications of each addition. This addition of elements can also be accomplished by selecting the **Add Chart Element** in the **Chart Layouts Group** of the **Chart Tools** tab that appears at the top of the window when a graph is selected. Type in your own chart and axis titles by clicking on the titles in the graph. Make sure to include **units** in the axis titles.

Tidy up the axis by displaying major and minor tick marks on the axes.

Right-click on the vertical axis line, then select **Format Axis**. A toolbar will appear on the right side of the window. Under the **Tick Marks** section, use the drop down menu to change **Major type** to **Inside**. Repeat this procedure for the horizontal axis tick marks.

Modify the horizontal and vertical axis ranges and specify the types of numbers on the graph.

Right-click on the vertical axis, then select **Format Axis**. In the **Axis Options** section, change the **Bounds** to have a **Minimum** of 0 and a **Maximum** of 120, if they are not already. Note that you can always reset the axes to the initial auto-generated range by clicking the **Reset** button next to the Minimum and Maximum. Under the **Number** section, you can specify what **Type** of number the axis represents, and change the displayed number of decimals on the major tick marks. Repeat this procedure for the horizontal axis.

Format the gridlines and the border styles.	Right-click on the horizontal gridlines (inside the graph axes) and then select **Format Gridlines**. Under **Dash Type,** select the **Dash** option. You may have to add vertical gridlines by right-clicking on the horizontal axis, then selecting **Add Major Gridlines…** before formatting them to the dashed line. Then, right-click on the white space within the plot area and select **Format Plot Area** from the Quick Menu. Under the **Border** section, select **Solid line**.

5.2.5 Error Bars

The last, and extremely important, step in presenting data is the addition of error bars on each of the points in accordance with the uncertainty specified in Table 1. The error bars will be added to the two series separately.

Click on any of the rubber ball data points to highlight all the points in the series. Under the **Chart Tools - Design** tab, select the **Add Chart Element** button in the **Chart Layouts** group in the top left corner. Under **Error Bars**, select **More Error Bars Options…** to bring up the toolbar on the right side of the window. In bold green, Excel will specify whether you are formatting the **Vertical Error Bar** or the **Horizontal Error Bar**. To switch between the two, click the bolded **Error Bar Options** drop down at the top of the toolbar and select between **Series "Rubber Ball" X Error Bars** or **Series "Rubber Ball" Y Error Bars**. We will specify the vertical error bars for the rubber ball data first.

If the error amount is the same for all data points, the easiest method is to type in the uncertainty in the **Fixed Value** box under **Error Amount**. The **Percentage** box may also be used if the error is always some fixed percentage of the value. In most cases, the error amounts will differ for each data point and may even differ between the positive and negative errors. Thus, the formatting of error bars requires use of the **Custom** button.

Select the **Custom** button and then click the **Specify Value** button. This action will bring up dialogue boxes for the positive and negative errors that work in the same manner as selecting data. The order in which the error data is selected corresponds to the points on which the error bars will be placed in your graph (first selected cell corresponds to error on the first data point in the series, second cell corresponds to error on second, etc.). Click the **Collapse** button beside the positive error box and then select all the errors for the bounce heights of the rubber ball, then press **<Enter>**. In this case, the error for the bounce height is $\pm$2cm, and thus the positive and negative errors for the bounce height are the same. As a result, the negative errors can be specified by the same cells as the positive errors. Click **OK** after specifying both errors, but do **NOT** close the toolbar on the right side of the window. Now click the bold **Error Bar Options** drop down as before to bring up the **Series "Rubber Ball" X Error Bars** and repeat the same procedure.

To reduce the size of the data point markers, thus making the error bars more visible, right-click on any of the markers then select **Format Data Series…** to bring up the right side toolbar. Underneath the bolded **Series Options** title, click the **Fill & Line** (paint can) button, then click **Marker**. Under **Marker Options**, change the marker type to **Built-in** and specify an appropriate marker size (in this case, use 4). Repeat the same error bars procedure for both series markers. It should be noted that you can also alter the format of the error bars in a similar manner by right-clicking on an error bar.

The resultant graph should now resemble the plot shown in Figure 4.

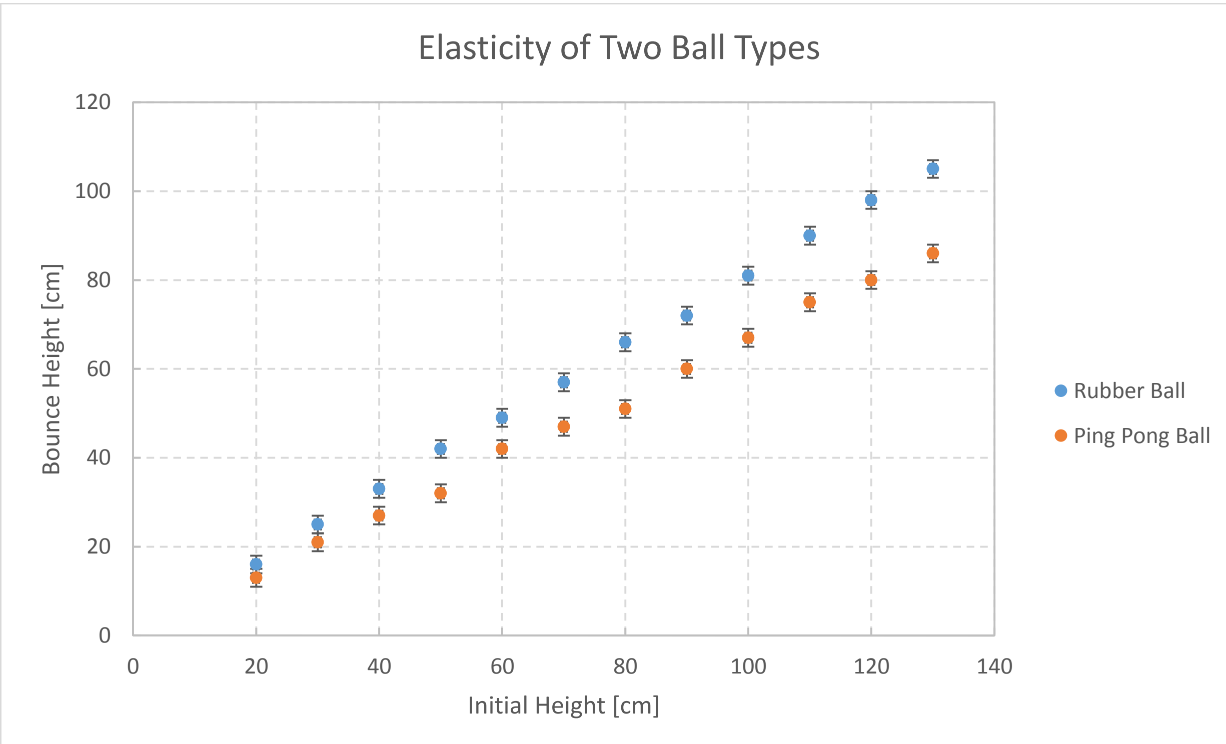

Figure 4: The final plot of Exercise 1-1, showing both the rubber ball and ping pong ball data series.

5.2.6 Evaluation Checklist

Your graphs and tables should ALWAYS display the professional format indicated in the previous section.

Spreadsheets in General:

- Have you chosen appropriate settings for typeface, font size and special formats (such as dates or percentages)? Have you used subscripts and superscripts where necessary?
- Have you forgotten any special characters or symbols? Use the **Symbol** button in the **Symbols** group of the **Insert** tab.
- Does your workbook have a meaningful title? Your name? Creation date? Do individual sheets within the workbook have distinguishing names?
- Is there enough information to understand the problem you are analyzing? Have you including input data (parameters), relationships (equations), as needed?

Tables:

- Do you tables all have column headings? With units?
- Are the column headings meaningful?
- Are the numbers in proper formats? Are you displaying the appropriate number of decimal places?
- Are the columns aligned properly with the text?
- Do your tables have neat borders?

Graphs:

- Does each graph have a real title (more than "Y vs X")?
- When you have pasted your graph into your Word file for your final report, have you eliminated the graph title?
- Are your graph axes properly labelled with appropriate units?
- If there are multiple symbols (data series) on a graph, do you have a legend to say what each symbol represents (with meaningful names for each series)?
- Have you used a second axis if necessary?
- Do the background gridlines interfere with printing or viewing the information on the graph?
- Are the axis scales appropriate for the range of data values? Do the curves take up most of the chart area (good) or are there large blank areas (bad)?
- Have you used professional formatting?

5.3 Fitting Linear Models to Data: Regression

One of the most common tasks associated with experimental data is fitting the best curve through the data in order to determine a desired quantity (for example, the slope). In this section, the tutorial will progress through the required steps to use one of Excel's built-in analysis tool packs called Regression, which is frequently used in engineering applications. Specifically, this section will cover the following points:

- Set up the Regression Data Analysis Tool for fitting a straight line to data
- Assess the validity of the Regression results
- Express the equation of the straight line derived from the Regression analysis
- Understand and report the errors in the slope and intercept of the produced line in a standard format

The previous problem and data on the elasticity of the rubber ball and ping pong ball will continue to be used.

5.3.1 Mathematical Models

Many engineering labs and projects involve fitting a mathematical model or equation to experimental data. This analysis may be done to determine the effect of one variable on another. An engineering application could be in the manufacture of plastics, in which the strength of the plastic film in an important quality measure, yet difficult to obtain. However, mathematical models can be developed easily to relate measured parameters, such as manufacturing line speed and temperature, to strength. These types of models allow engineers to estimate the strength of plastics by simply measuring the line speed or temperature.

A model should always be an **adequate fit** for the data, otherwise it is inaccurate. There are always two parts to modelling experimental data: first, drawing the line and second, evaluating relationship between the points and the line. The best method to "draw" the perfect line is to use **linear regression**.

Linear regression involves obtaining the equation of a straight line of best fit, and then using the equation to calculate an estimate $\hat{y}$ value for each x coordinate. The differences between the experimental y-values and the estimated $\hat{y}$-values are called the **residuals** of the data. The line of best fit is the line that minimizes the sum of the squares of the residuals for each point, as given by the following expression:

$$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

where y_i is the original data point and $\hat{y_i}$ is the estimated data point using the equation for line of best fit. This equation will be explained further in Statistics courses; however, it is useful to examine the method that Excel uses to perform regression.

5.3.2 Estimating the Elasticity (E) and its Uncertainty

Let us carry out a linear regression of the bounce height and initial height data of the rubber ball and ping pong ball from Table 1. If you have never used the regression tool in Excel prior to this tutorial, it will be required to be loaded as an add-on.

Click the **File** tab to bring up the **Backstage**, and then select **Options** in order to bring up the Excel Options menu. Navigate to the **Add-Ins** tab on the left menu. At the bottom, click the **Go** button beside the drop down menu, ensuring that you are managing **Excel Add-ins**. A pop-up window will appear with several unchecked boxes. Check the box that corresponds to **Analysis ToolPak**, then click **OK**.

Re-open your Exercise 1-1 workbook, then click the **File** tab and **Save-As**. Save the file under a new name "**Exercise 2-1**". Using Save As in this way will not overwrite the original file, but it will save the same workbook as the new file name you specified.

Launch the Regression tool.

Click the **Data** tab, then click **Data Analysis** in the Analysis group on the right. A window with a list of analysis tools will appear. Select **Regression**, then click **OK**.

Specify the data on which you wish to perform regression analysis.

Click the **Collapse** button beside the **Input Y Range** dialogue box and select the bounce height, h_b, data for the rubber ball (including the column heading), followed by **<Enter>.** Then click the **Collapse** button beside the **Input X Range** box and select the initial height, h_i, data for the rubber ball (including the column heading), followed by **<Enter>**.

Specify whether you included labels on your data and whether your relationship passes through the origin.

Check the **Labels** and **Constant is Zero** boxes. Note that Constant is Zero will force the regression line through zero, and you will not want this specified for all cases. Also check the **Confidence Level** box, leaving the level at **95%**.

Specify where you wish the regression results to appear. In this case, we wish to have the results in a new sheet in our workbook, and include the residuals plot along with a graph of the original data and fitted line.

Make sure you have selected **New Worksheet Ply**, then type the title "**Reg 2-1A Rubber**" in the adjacent dialogue box. Check off the **Residuals**, **Residuals Plot** and **Line Fit Plot** boxes. Click **OK**.

A new worksheet titled Reg 2-1A Rubber will then be added to your workbook containing several tables, a residual plot and a line fit plot. The residual plot indicates the adequacy of the line of best fit. If the line is of adequate fit, the residuals should be **randomly distributed** above and below the zero line (a perfect fit requires that all residuals are zero). If there is a systematic pattern in the residuals, this correlation suggests that a **higher-order fit is necessary**. The residual plot from this data should possess initial height, h_i, on its horizontal axis and "residuals" on its vertical axis.

The Line fit plot should show the line of best fit and the experimental points on the same graph. The graph should display h_b vs. h_i.

Right click on any of the **Predicted $\boldsymbol{h_b}$** points and select **Format Data Series…** to bring up a toolbar on the right side of the window. Under the **Fill & Line** section (paint can), change the line **Type** to **Solid line**. Then click the **Marker** section and drop down the **Marker Options** selections to change the marker **Type** to **None**. Close the toolbar by pressing the **X** at the top right.

The Line Fit Plot should now display the experimental rubber ball points along with a regression line of best fit.

Summary Output Table: The following points will explain some key information regarding the output tables of the regression tool.

Underneath the table labelled Regression Statistics:

Observations – these are the number of data values that were used in the calculations (number of trials in the experiment). It is good practice to ensure that this number matches the number of data points you wished to include in the regression AND ensure that you have checked the Labels box before performing the regression.

R-Square – this value provides a measure of the "goodness of fit" of the regression line. A value of 1 indicates that the data was perfectly linear and thus the line passed directly through all data points. In practice, this degree of perfection will never be the case. An R-Square value greater than 0.95 indicates a good fit.

From the 3rd table with column headings including Coefficients, Standard Error, etc.:

Coefficients – this column provides the y-intercept and slope of the line of best fit. The **Intercept row** indicates the y-intercept, and the **hri row [cm]** indicates the slope. The intercept in this case should be exactly zero as we forced the regression through the origin in the regression tool. In this example, the slope of the line is the same as the elasticity (E) of the rubber ball. Therefore, the equation of this line of best fit can be represented by the following:

$$h_b = Eh_i$$

Standard Error – this error is the value most frequently used in APSC-100 Module 2. It directly corresponds to the value of the "propagated error" that will be calculated in your error analysis. It is also comparable to the standard error that one can obtain by using a statistical analysis (for example, in a titration lab). Standard error corresponds to a 68% confidence interval, and twice the standard error corresponds to a 95% confidence interval. As shown in the table, the regression tool calculates a standard error on the intercept as well as the slope.

Lower 95% and Upper 95% -- these values are the lower and upper values bounding the confidence intervals for the slope and y-intercept. In other words, there is a 95% probability that the slope and intercept lie within the upper and lower limits. For example, if the upper and lower limits on the slope are 0.820 and 0.808, then the slope should be reported as $E = 0.814 \pm (0.820 - 0.808) = 0.814 \pm 0.012$. Thus, ± 0.012 is the **uncertainty** or **error** associated with the estimate of $E = 0.814$ with a 95% confidence level.

Now, the regression will be repeated without specifying that the y-intercept is zero to see if a more satisfactory result is obtained. Use the same procedures specified in the regression instructions above; however, ensure that the box adjacent to **Constant is Zero** is not checked off. In the **New Worksheet Ply** dialogue box, type the title "Reg 2-1B Rubber". Click **OK** to perform the regression.

After comparing “Reg 2-1A Rubber” and “Reg 2-1B Rubber”, it should be apparent that the residuals are more randomly distributed about zero in the latter regression, thus indicating a slightly better fit.

Important Note: When you report your results in a lab report, make sure that you include the equation of the line of best fit as well as the error on the slope and intercept. When you report the error, you must state which type of error you are using (for example, standard error or 95% CI). In this APSC-100 course, the standard error will be used. Your reported values possess no meaning without their uncertainties.

5.3.3 Plotting the Regression Results

It is useful to have the equations of your lines of best fit on your graph. Return to the initial worksheet containing the graph of the rubber ball and ping pong ball data sets.

Right click on any of the data points for the rubber ball, then select **Add Trendline…** to bring up the toolbar on the right hand side of the window and place a dotted trendline through your data. Under **Trendline Options**, ensure that the trendline is **Linear**. Excel will produce a default name for the trendline; however, you may specify a specific name in the **Custom** dialogue box under **Trendline Name**. Check off the box beside **Display Equation on chart**, then close the toolbar. This trendline will match the line of best fit that was plotted on the Line Fit Plot from the regression output. Verify this method by confirming that the slope and intercept is the same on your graph and in the results of the regression.

The trendline equation that was added to your plot is likely overlapping your data points and/or your trendline. **Click and drag** the equation such that it is located beside the trendline for the rubber ball data. Additionally, Excel produces a default equation in the form of $y = mx + b$. We have plotted h_{rb} vs. h_{ri} for the rubber ball. Change the variables in the equation by clicking within the equation and replacing “y” and “x” with the appropriate variables. Use the method of producing subscripts discussed earlier. Repeat this same procedure for adding a trendline to the ping pong ball data. Ensure that both equations show an appropriate number of decimal places for the slope and intercept. Your resultant graph should resemble that of the plot shown in Figure 5.

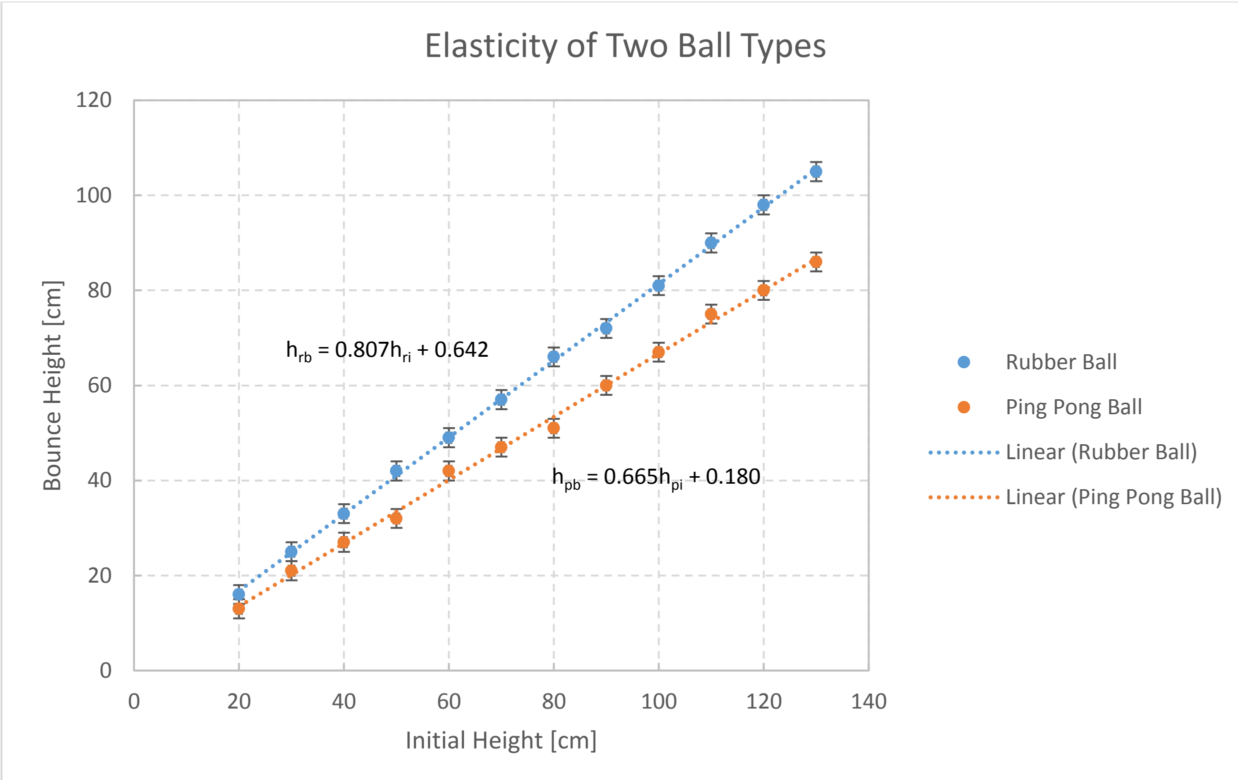

Figure 5: The resultant plot of the rubber ball and ping pong ball data sets.

5.4 Frequency Distributions and Importing Data

In this section, the following points will be discussed:

- Calculating Frequency Distributions and plotting them as Histograms
- Using the Excel average, median and standard deviation functions
- Sorting and filtering data
- Using Edit Fill
- Importing data from an ASCII file

As an exercise, consider one of the most common examples of frequency distributions – the analysis of student final marks in a class.

5.4.1 The Problem

A department is trying to determine which method of teaching is best for the students in a particular class. The class gets divided randomly into two sections. Section A gets taught the traditional way (with lectures and notes) and Section B received material and interacted with the professor and others solely through the internet. The final grades of the students in both sections can be seen in Table 2.

Table 2: The grades for the students in a particular class for sections A and B.

Section	Final Grade
A	85.2
A	81.7
A	84.5
A	82.4
A	84.8
A	83.1
A	85.1
A	84.5
A	83.7
A	86.8
B	87.1
B	86.1
B	85.2
B	88.5
B	84.5
B	86.1
B	87.7
B	86.3
B	85.8
B	86.3

One way to determine which method of learning is best is to compare the distribution of marks in the two sections. This analysis can be done by using the following steps:

- Set up a spreadsheet to obtain the frequency distributions for both sections, and plot the distributions in the same chart
- Determine the average grade, median grade and the standard deviation of grades for each section
- Report the results in a summary section at the bottom of the spreadsheet

Open a new workbook in Excel and save it under the name "Exercise 4-1.xlsx". Begin the spreadsheet with the proper identification section described earlier and using the title "Exercise 4-1: Analysis of Instructional Methods" in cell A1. In cell A4, begin adding the data from Table 2 such that the first mark of 85.2 will appear in cell B5.

The main idea behind a frequency distribution is to set up a series of bins, then count the number of items in each bin. In Excel, you specify the bins through their **upper edge**. An ideal bin size for this scenario may be bins that are "1 mark wide".

Create bins that are one mark wide.

In cell **B27**, label a column with the name "**Bin**". Below this cell, type the marks corresponding to the upper edge of each bin (i.e.: 79.9, 80.9, 81.9, etc.) until you reach 89.9.

Create a column for each of the sections and adjust the cell height.

Label cell C27 "**Section A Frequency**" and cell D27 "**Section B Frequency**". As these are large headings, it is prudent to make the cells two rows high. Select the two headers, the click the **Wrap Text** button in the **Alignment** group of the **Home** tab.

To compute the frequency distributions for each section, a histogram will be constructed. Click the **Data** tab and then the **Data Analysis** button in the **Analysis** group (same button to get to the regression tool). Select **Histogram** and then fill out the dialogue box as follows.

Specify the data you wish to analyze and where the output will appear.

For the **Input Range**, click the **Collapse** button and select cells B5:B14 to include all the marks for Section A, followed by **<Enter>**. Click the **Collapse** button for the **Bin Range**, then select cells B28:B38, corresponding to the upper bin limits. Hit **<Enter>**. Select **New Worksheet Ply** and specify the title "**Sec A Histogram**". Select **Chart Output**, and then click **OK.**

The histogram for the current data will be presented in a new worksheet called "Sec A Histogram". The sheet will also possess a table in which the first column is headed *Bin* and the second column is headed *Frequency*.

Return to Sheet 1 and repeat the procedure using marks from Section B, assigning the analysis to a sheet titled "Sec B Histogram". It is now possible to create a combined chart displaying both results.

5.4.2 Column Charts of Frequency Distributions

In Sheet 1, there should be a column with bin edges and two empty columns for the frequencies of each section. Copy and paste the frequency numbers from Sec A Histogram and Sec B Histogram into these columns. Create horizontal axis labels in column A such that your table appears as shown in Table 3.

Table 3: A completed frequency distribution chart for both sections.

Final Grade	Bin	Section A Frequency	Section B Frequency
<80	79.9	0	0
80-80.9	80.9	0	0
81-81.9	81.9	1	0
82-82.9	82.9	1	0
83-83.9	83.9	2	0
84-84.9	84.9	3	1
85-85.9	85.9	2	2
86-86.9	86.9	1	4
87-87.9	87.9	0	2
88-88.9	88.9	0	1
89-89.9	89.9	0	0
>89.9	∞	0	0

It is now required to create a column plot of both data sets.

Select the data to be plotted and use the chart tools to create a column chart.

Select **A28:A40** (grade ranges for the x-axis), then hold down the **<Ctrl>** key and use your mouse to select **C28:D40** (frequencies for each Section). In the **Insert** tab under the **Charts** group, select the **Column** chart. Properly label the graph with an informative title and axes titles.

Finally, it is often useful to be able to conduct a statistical analysis of a set of data. For the purpose of the labs in APSC-100 Module 2, you will need to calculate the **Mean** (average) and the **Standard Error** (the standard error determined statistically is directly comparable to the Standard Error found through regression analysis on a linear set of data points such as was described earlier). In order to obtain these statistics, the **Descriptive Statistics** data analysis tool must be used. Click the **Data** tab and then select **Data Analysis** in the **Analysis** group (same button as for histogram and regression tools). In this instance, use the **Descriptive Statistics** tool. Click on the **Collapse** button for the **Input Range** and select the data for the students grades from Section A (NOT the frequencies), then press **<Enter>.** Next, check the box adjacent to **Summary Statistics**. Finally, specify the name of the new worksheet in which the summary statistics will be produced. Click **OK**. The Mean and Standard Error are the first two values in the table. Repeat this procedure for Section B.

5.4.3 Sorting Data

It is often necessary to sort data in a spreadsheet according to one of the variables. This sorting is done by selecting **Sort** in the **Sort and Filter** group of the **Data** tab. Select both columns for section and grade, including the headers, before pressing the **Sort** button. Column headings can be included in the block as long as the **My data has headers** box is ticked when performing the sort. Click the **Sort** button to bring up the Sort window. Ensure that the **My data has headers box** is ticked in the top right of the window. Next, use the drop down menus to specify the criterion by which the data is to be sorted and in which order to you would it to be rearranged. In this instance, sort the data by **Final Grade** under the **Sort by** menu and select **Largest to Smallest** under the **Order** menu. The **Options** button can be used if you wish the sort to be case sensitive or to have a special sort order such as days of the week or months of the year. Click **OK**.

5.4.4 Importing Data from an ASCII File

There are some experiments in which there are hundreds of data points and typing them all into Excel manually would be extremely inefficient. Fortunately, such lengthy data can be entered into Excel as an ASCII text file. ASCII stands for American Standard Code for Information Interchange. ASCII is an internationally agreed upon standard which assigns standard numbers to each key on a keyboard.

The data is saved in the file as one set of readings per line; however, these data values must be separated somehow, or it will be impossible to distinguish where one stops and the next begins. This separation can be done using a **Delimiter** or by using a **Fixed Width**. The comma is an example of a typical delimiter, in which case the type of file is called a comma separated values (.CSV) file.

The data can consist of text and/or numerical values. Text is sometimes identified by being put in double quotes. When importing data into a spreadsheet, it is required that the user specifies to the program how the data is delimited or whether a fixed width is being used.

A common method of importing data is to use the **Text Import Wizard**.

Use a plain text editor such as **Notepad** to create a comma-delimited ASCII file of the data in Exercise 4-1 and name the file “Ex4-2.txt”. This action is accomplished by typing in each row of data from Table 2, separating the section from the grade using a comma and giving each data pair its own row. Next, use the **From Text** button located in the **Get External Data** group in the **Data** tab of Excel. This button will bring up the Text Import Wizard window, which guides the user through 3 steps to import the file. First, specify the file type as either **Delimited** or **Fixed Width** (in this case, select Delimited), then click **Next**. Then, check off the boxes that represent the delimiter used in the file (in this case, check off the box adjacent to **Comma**) before clicking **Next**. The third step specifies what type of values the file contains, with the default as **General**. This step can be left unchanged. Click **Finish**, then check **New Worksheet** on the resultant pop-up window to complete the import. Check that the imported data is consistent with the data on the original sheet.

A more simplified copy and paste operation can also be used to bring the data into an Excel worksheet from the text file. When Notepad is open to the text file, press **<Ctrl> and A** at the same time to select all the contents (this shortcut works for an entire Excel worksheet as well). Then use **Ctrl and C** to copy the text to your clipboard. Go to your Excel Worksheet and select the cell in which you would like the data to start, then press **Ctrl and V** to paste the text. You will notice that you need to place the data values in separate columns in order to analyze them in Excel as the text file has placed the section and grade in the same cell. After pasting the data into the worksheet, click the arrow underneath the **Paste** icon in the **Clipboard** group of the **Home** tab, the select the **Text Import Wizard**.

5.5 Quick Excel Summary

5.5.1 Summary of Statistical Analysis

The mean and standard error (also known as standard deviation of the mean) can be determined using Excel using multiple methods. In order to use the Descriptive Statistics tool, perform the following actions:

- Select: **Data** tab, then **Data Analysis** in the **Analysis** group, then **Descriptive Statistics**.
- Click on the input range and select the data for the sample from your worksheet.
- Check the box for **Summary Statistics**.
- Click **OK**.

The mean and standard error are given in the first two rows of the resulting table. These represent the best estimate of the value being measured and the uncertainty of this estimate. The standard deviation is also given in the output table; however, this error is not a value that you will need in the experiments in APSC-100 Module 2.

5.5.2 Graphical Analysis

The following instructions will explain how to obtain a plot of data with error bars, produce a model of the data (regression), and to obtain the uncertainty associated with the slope and intercept of the line of best fit.

- Enter the data points are their uncertainties into Excel in an organized table.
- Select the data and then click the **Scatter with only markers** button from the **Charts** group in the **Insert** tab.
- Use the **Select Data** button in the **Data** group of the **Design** tab (appears while the graph is selected) or right-click on the plot, then select **Select Data** from the Quick Menu. Name the data series appropriately.
- Ensure you have added all the necessary Chart Elements, such as axis titles, a legend, error bars and trendlines. Professional formatting is also required.
    - To avoid having to do the formatting specified in Section 5.2.4 for each and every graph you produce for the foreseeable future, you can save your formatting as a chart template for scatter plots. To create a template, format a plot according to the specified steps, then right-click on your plot to bring up the Quick Menu. Select **Save as Template…** to bring up a window already specified to the correct location to save templates in your local hard drive. **Name** your template, then save it. To use the template in later plots, select the data you wish to plot and choose the correct type of graph (i.e. scatter, column, etc.) to create an unformatted plot of your data. Under the **Design** tab (appears when you have selected your plot), click **Change Chart Type** under the **Type** group. A window will appear with an option to select **Templates** from the left menu, which should allow you to select from your saved templates.

- To add error bars, select your graph and then click the **Add Chart Element** button in the **Design** tab under the **Chart Layouts** group. Go to **Error Bars**, and then to **More Error Bars options…** in order to bring up the toolbar on the right side of the window. Make sure to have **Both** selected for **Direction** and include a **Cap**. Then specify whether you wish to have a **Fixed Value**, **Percentage**, or **Custom**.
- To add a trendline and its corresponding equation, right-click on the data series for which you wish to produce a trendline, then select **Add Trendline…** from the Quick Menu to bring up the toolbar on the right side of the window. You can then specify the line **Type**, **Width**, and **Dash type** in the **Fill & Line** group (paint can button). In the **Trendline Options** group (bar chart button), specify the **Type** of trendline, **Trendline Name**, and check the box to **Display Equation on chart**. Remember to change the default variables "x" and "y" in your plot to correspond to your actual data.

5.5.3 Regression Analysis

- To perform a regression analysis on a set of data, click the **Data Analysis** button under the **Analysis** group in the **Data** tab. Select **Regression** from the list of possible analysis tools.
- Click the collapse buttons on the **Input Range** and **Output Range** dialogue boxes to specific the data you wish to analyze.
- Select **Residual Plots** to see whether the data is adequately explained by a linear relationship, and specify that you wish the regression output to appear as a **New Worksheet Ply** with an appropriate name.
- The main table of interest in the regression output is the third table containing the row headers **Coefficient** and **Intercept**. Coefficient represents the **slope** of the line and Intercept represents the **intercept** of the line. The **Standard Errors** for these parameters are also reported and should be included when you report the slope or intercept in a professional report
    - Note that the **Coefficient** for the **X variable** should be the same as the slope of the trendline in your plot.

5.6 Important Tips and Tricks for your Reports and Analysis

There are several tricks that one acquires over the course of one's academic career, especially with respect to Excel and Word. Some of these tricks are explained below and can be used to greatly expedite the process of analysis and formatting using Excel and Word.

5.6.1 Copying Figures or Tables as Pictures

When you bring your tables and graphs from Excel to Word for the creation of your reports, the default is for a link between the Word table or graph and the Excel table or graph. This default means that when the table or graph is updated in Excel, it will also be updated in Word. There is the potential for Word to run slowly, particularly for large documents with many linked figures and tables. To avoid this delay, there is the ability to copy a table or graph into Word as a picture. Instead of simply using **<Ctrl> and V** to paste a graph into your Word document, right-click where you wish to paste the table or graph, then select **Paste as Picture** from the **Paste Options**. You may need to format the picture to fit within the margins of your page.

5.6.2 Table Borders in Word

Although Excel possesses gridlines that allow you to create bolder borders in some cells and leave gridlines between others, Word does not. This lack of gridlines means that when tables are pasted into Word, the lines separating some cells may be missing. To fix this problem, paste the table into Word and then click inside one of the cells of the table. A symbol should appear at the top left corner of the table. Left-click on this symbol to select all the cells in the table. Next, click the **Design** tab at the top of the page and select the A**rrow** underneath the **Borders** button in the **Borders** group, then select **All Borders**.

5.6.3 Calling Variables in Excel and Managing Them

This section may become more important in upper year courses; however, it is worth noting as it allows user to implement equations in a much more symbolic and intuitive method in Excel. When one types in a formula using Excel and calls for cells to be multiplied together, the resultant formula looks very busy and it is difficult to discern what is occurring. For a simple example, we shall calculate power in terms of voltage and current measurements utilizing the following expression:

$$P = VI$$

where P is the power (in Watts), V is the voltage (in volts) and I is the current (in amperes). An Excel data table was created such that voltage, current and power measurements possessed their own columns. The usual way of calculating power would be to multiply cells in one row, and then copy the formula down through all the rows, as shown in Figure 6.

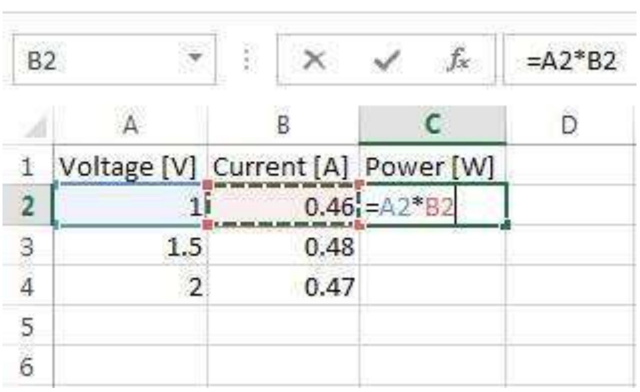

Figure 6: The normal method of using formulae in Excel.

Excel has the ability to name a cell, group of cells or entire column/row such that the user can type the actual variable name into the formula. To name a group of cells, simply highlight the cells you wish to name, as shown in Figure 7.

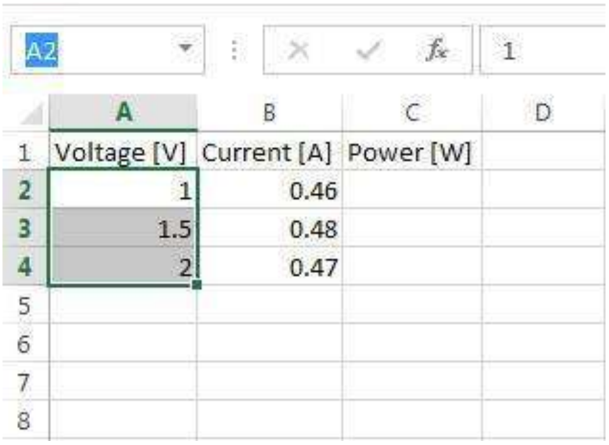

Figure 7: Highlighting a group of cells to be renamed.

It should be noted that the "**A2**" in the top left corner of Figure 7 is the **Name** of that particular group of cells, in this case the top cell by default. Use your mouse to highlight the default name of the cells and change it to the variable the data represents. An example of this renaming is shown in Figure 8 when the cells containing the current measurements were renamed to "**Current**".

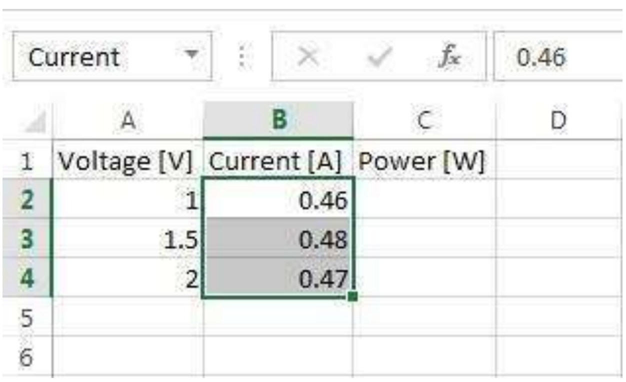

Figure 8: Renaming a group of cells in Excel.

In order to perform the Power calculation, the formula may contain the names "Voltage" and "Current" instead of "A2" and "B2", as shown in Figure 9.

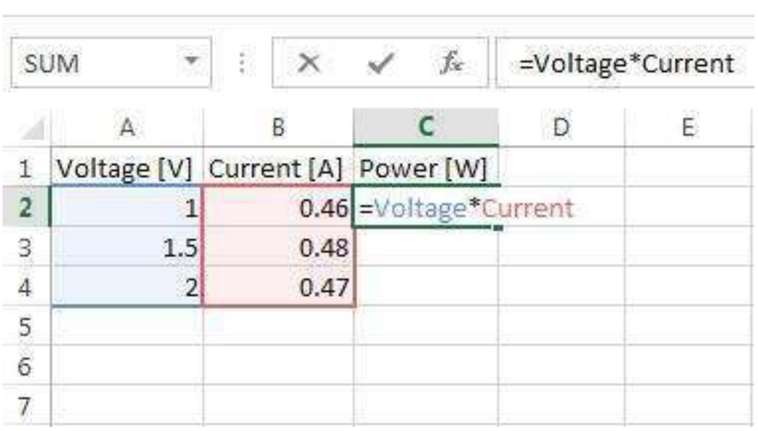

Figure 9: The final formula using variables instead of cell references.

When the formula is copied down in rows 3 and 4, “Voltage” and “Current” refer to the cells in those rows. The final numerical result is left unchanged; however, it becomes much easier to debug an Excel formula using variables, particularly when formulae become very large and reference many cells. In order to edit which cells, groups of cells or columns/rows are named, go to the **Formulas** tab and select the **Name Manager** in the **Defined Names** group. This manager also allows for the deletion of variable names.

5.6.4 Selecting All Data in a Column

You will frequently need to copy and paste, edit, or simply select an entire column of data. Rather than using the mouse to scroll to the end of the column where the last data point is located, Excel possesses a keyboard shortcut to select all data in a row. First, select the top cell in the column where your data is located. Then, press the keys **<Ctrl>** and **<Shift>** and the **Down Arrow** at the same time. You have now selected all cells in that column that possess contents. This shortcut works with other arrow keys to select a row as well.

5.6.5 Transposing Column Data into a Row

It may be necessary to copy and paste a column of data into a row depending on the arrangement of your spreadsheet. A common instance of this transposition is the need to place experimentally obtained data from a data acquisition device into a worksheet that was set up prior to performing the experiment. First, copy the column of data using the normal method (select all the required cells and then press **<Ctrl> and C**). Then, right-click on the first cell in which you wish the ROW of data to be located. In the resultant Quick Menu, select **Transpose** under the **Paste Options**.

5.6.6 Converting to Different Units

Measurement devices and data acquisition systems often do not record data in units that are appropriate for further analysis. As engineers, you will be expected to be able to work with multiple unit systems, including imperial and metric. Fortuitously, Excel possesses an internal function named **“Convert”** that can convert data to a desired unit system. This function is used in a very similar manner to the cosine or sine function. Start typing a formula into an empty cell by left-click the cell, then pressing **<Enter>**. Type “**convert**”, and Excel should show a small menu beneath the cell, confirming the existence of the Convert function. Use a **left parenthesis** after the word “convert” to begin using the function, as shown in Figure 10.

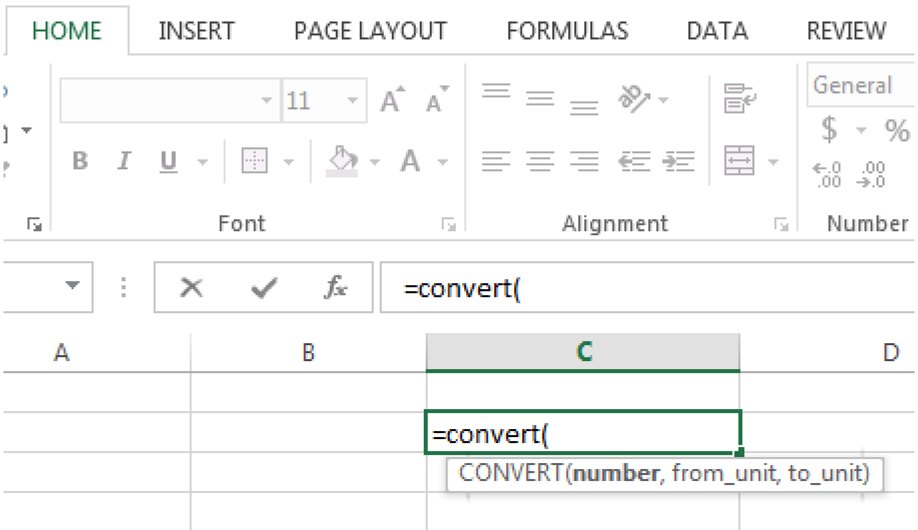

Figure 10: Using the "Convert" function in an Excel formula.

Next, follow the syntax specified under the cell. The first input, "**number**", represents the number or cell that you wish to convert to different units. Enter the number or click the cell you wish to convert, then type a **comma**. You are then required to specify the original units of the cell as the "**from_unit**" input. Excel will automatically bring up a drop down menu from which you can choose the original units with a **double click**. Type another **comma** to proceed to the "**to_unit**" input, specifying which unit you would like the resultant cell to contain. Your formula should appear similar to that shown in Figure 11.

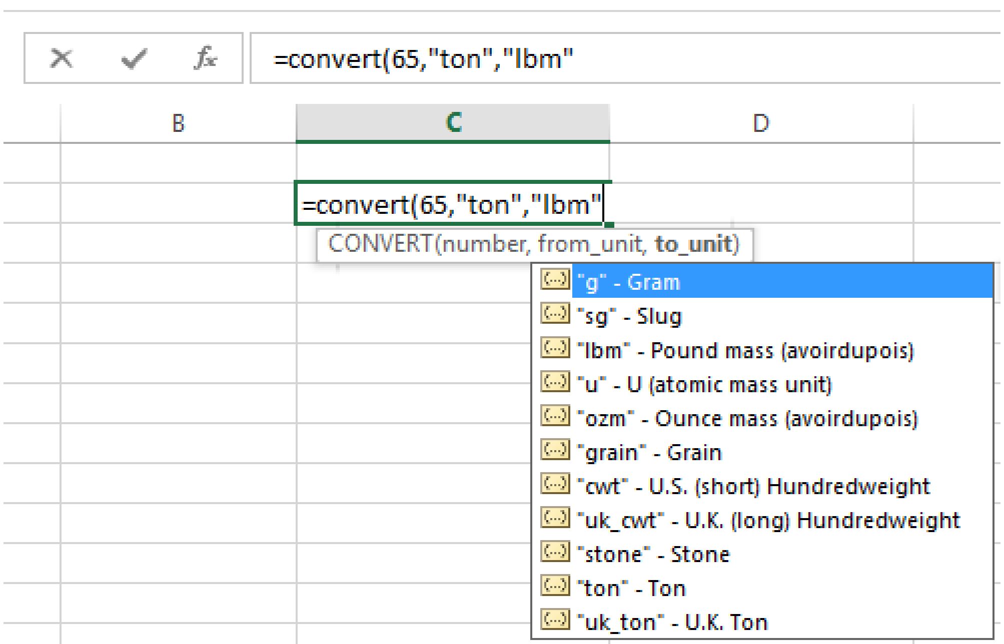

Figure 11: The final appearance and syntax of the "Convert" function in Excel.

Finalize the formula by closing it with a **right parenthesis**. Press **<Enter>** to evaluate the cell.