

Microsoft Excel 2016 Tutorial

Microsoft Excel spreadsheets are a powerful and easy to use tool to record, plot and analyze experimental data. Excel is commonly used by engineers to tackle sophisticated computations and produce detailed optimization studies of real data. Excel is used extensively in every engineering discipline and in a wide variety of courses. This introductory tutorial will enable you to complete useful and meaningful data analysis and visualization using Excel, and will act as a foundation in further building these skills throughout your academic career. The video tutorials cover Excel fundamentals, including how to get started using Excel, using formulas or functions, creating visual representations of data and conducting data analysis.

This written tutorial is an additional resource to the video series: “FEAS Microsoft Excel Tutorials.” In any split sections, the left column indicates an action and the right column describes the steps required to complete the action. In instances where the process for Mac is significantly different than the process for PC, a separate section will include how to complete the process for Mac. Any instances where a shortcut is present for PC (ie. <Ctrl> + C) can be adapted to Mac by changing <Ctrl> to <⌘> (ie. <⌘> + C).

Table of Contents

1. Excel Basics.....	1
1.1 Opening/Closing the file	1
1.2 Saving the file	1
1.3 Installing the “Analysis ToolPak” for PC	1
1.4 Installing the “Analysis ToolPak” for Mac	2
2. Entering data into Excel	2
2.1 Simple cell formatting.....	2
2.2 Using formulas and functions	5
3. Data visualization	7
3.1 Inserting a scatter plot	8
3.2 Selecting data series	8
3.3 Chart formatting.....	8
3.4 Extracting tables and graphs	12
4. Data analysis.....	13
4.1 Linear regression	13
4.2 Descriptive Statistics	16

1. Excel Basics

This section explains how to get started with Excel with the basic operations of opening, closing and saving an Excel file, known as a workbook. A workbook is comprised of multiple spreadsheets, known as worksheets, which can be used and manipulated separately.

1.1 Opening/Closing the file

The first step to using Excel is launching the program and knowing how to close it when you're finished.

To open a new Excel workbook:

Simply double click the Excel icon and select the **Spreadsheet** option

In order to close the workbook:

Click the "**X**" in the upper right hand corner of the window

1.2 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.

In order to save your Excel file:

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. Give a descriptive file name and save it as an Excel Workbook, meaning it will have ".xlsx" as an extension, indicating that it is an Excel 2013 document.

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

1.3 Installing the "Analysis ToolPak" for PC

Plugins in Excel can be installed to increase the functionality of the program and allow the user to complete operations that the original program doesn't have automatically installed. Many engineering courses require the "**Analysis ToolPak**" plugin to be installed in Excel.

To install the "Analysis ToolPak":

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**", and then click **OK**.

1.4 Installing the “Analysis ToolPak” for Mac

Plugins in Excel can be installed to increase the functionality of the program and allow the user to complete operations that the original program doesn't have automatically installed. Many engineering courses require the “**Analysis ToolPak**” plugin to be installed in Excel.

Please note that in Microsoft Excel 2016 the “Analysis ToolPak” is now available for Mac.

It is expected that all students are working in the 2016 version of Microsoft Excel.

Microsoft Office 2016 can be downloaded for free by all Queen's University students at the following link:

<http://www.queensu.ca/its/microsoft-office-365/office-365-proplus/tutorials/download-install-mac>

To install the “Analysis ToolPak”:

Click the **Tools** tab and select **Add-Ins**. Navigate to the **Add-Ins available** box, select the “**Analysis ToolPak**” check box, and click **OK**. If the “**Analysis ToolPak**” is not listed, click **Browse** to search for it. If the “**Analysis ToolPak**” does not appear installed on your computer, click **Yes** to install it. You should quit and restart Microsoft Excel. **Data Analysis** should now be available on the **Data** tab.

2. Entering data into Excel

Using a spreadsheet to effectively visualize and analyze data requires proper formatting. The boxes that make up the spreadsheet are called ‘cells’, and each cell can be characterized by its row (numbered) and column (lettered). An example of a cell referenced using its row and column is H7, which is cell in the 7th row of column H. This is especially useful to know when working with formulas.

2.1 Simple cell formatting

This section describes how to enter and format numbers into a spreadsheet.

2.1.1 Entering data into cells

The first thing to know when using worksheets is how to enter data into a cell.

To do this, simply type in all the raw data by **left-clicking** the cell under the correct column and typing in the numbers only, NO UNITS. In Excel, cells with numbers are automatically right aligned and cells with any lettered elements are left aligned. This can be changed using the **Alignment** options in the **Home** tab.

2.1.2 Spreadsheet Organization

It can be useful to organize worksheets using an identification section, especially when you have more than one Excel document in progress at one time. An identification section includes a title, the author's name, the date of creation, and the file name.

To add an identification section:

Click on cell A1 of Sheet 1 to make it the active cell. Type the title and press **<Enter>**. The cursor should then move to Cell A2. Type in your name and the name of the file in cells A2 and A3 respectively. The name of the file may or may not be the same as the title. Select cell C2 and type today's date in the format year/month/date and press **<Enter>**.

It's also useful to name worksheets, especially if you're using more than one worksheet in a workbook.

To rename a worksheet:

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

2.1.3 Subscripts and superscripts

In some notation schemes, the use of subscripted and superscripted numbers and text is essential in effectively labeling data.

In order to subscript or superscript a character or string of characters:

Highlight the text you wish to change. 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** or **Superscript** box depending on which you desire. Then press **OK** or press **<Enter>**.

2.1.4 Widening/condensing column widths

An important visual characteristic of your Excel tables is the column width. Sometimes it is desirable to widen or condense columns to improve the readability of your spreadsheet.

To change column width:

Simply 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.

2.1.5 Merge and center

To improve clarity and make tables aesthetically pleasing, it may be of use to have multiple cells merge into one.

To merge cells together:

Select all the cells you wish to merge, then click the **Merge and center** button in the **Alignment** group of the **Home** tab.

2.1.6 Wrapping Text

If the content of a cell is significantly longer than the column width, the contents can be wrapped. This means that the cell will lengthen automatically such that the content fits within the cell.

To wrap text:

Select the cell or column that you wish to format then in the **Alignment** Tab in the **Home** menu select **Wrap Text**.

2.1.7 Cell type

Sometimes it's useful to define the type of number contained within a cell. For example, if you're dealing with monetary values then formatting the cell to contain a currency automatically places a currency sign before the value. There are a number of cell types to select from with different formatting schemes.

To define cell type:

Use the drop down menu in the **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** is sufficient).

2.1.8 Decimal numbers

To adjust the number of decimal places shown in a cell:

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.

2.1.9 Autofill with patterns

When entering data, If there is a pattern to the raw data (for example if you want to calculate the areas of circles with radii of 1m, 2m, 3m, 4m etc.) then use can use an Excel trick to populate desired cells without typing all the values. This is called auto-filling cells.

In order to autofill cells:

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. 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.

This method also works with days of the week, months and written patterns (such as Week 1, Week 2, Week 3 etc.).

2.1.10 Sorting data

It is often necessary to sort data in a spreadsheet according to one of the columns, for example from smallest to largest or vice versa.

To sort data according to a specific scheme:

Select all data columns and navigate to **Sort** in the **Sort and Filter** group of the **Data** tab. 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. 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 like it to be rearranged. 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. Once you're satisfied with the sort criteria, click **OK**.

2.1.11 Transposing data

It may be necessary to copy and paste a column of data into a row (or row to column) 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.

To transpose data:

Copy the data then right-click on the first cell in which you wish the transposed data to be located. In the resultant **Quick Menu**, select **Transpose** under the **Paste Options**.

2.2 Using formulas and functions

Next we're going to discuss formulas and functions. Formulas and functions in Excel are essential for the analysis of data sets, and a firm grasp of this functionality will be invaluable in upper year laboratory courses for any discipline.

2.2.1 Custom formulas

We'll start with custom formulas. Excel can be used to perform custom mathematical operations on data sets.

To perform a custom mathematical operation:

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>**.

For example, if you wish for the contents of cell B3 to be twice the value of cell A3, cell B3 would contain the following formula: **"=A3*2"**. When you press **<Enter>**, cell B3 will evaluate and the formula will disappear.

To edit a formula:

Select the cell and press **F2** or click in the formula bar just above the column headers.

2.2.2 Built-in Functions

Excel also has built in functions for a number of mathematical operations; a few illustrative functions are included below in Table 1. Functions are usually common mathematical actions that are challenging to write formulas for (such as finding the average or mean of a dataset or calculating the sine of an angle).

Entering a function is similar to a formula:

Type '=' then follow with the desired function. **<Enter>** will evaluate the cell.

The Excel help menu and online forums give detailed information on all available built in functions. Using functions can greatly speed up completing calculations and prevent human error in entering formulas.

Table 1: A few examples of built-in functions in Excel

Name	Syntax	Action
Sum	SUM(element1, element2...)	Sums all the elements in parenthesis
Average	AVERAGE(element1, element2...)	Finds the average of the input elements
Sine	SIN(angle)	Finds the sine of a given angle

2.2.3 Copying formulas/functions

It is often useful to apply the same formula to a column of data. This can be achieved using two methods: the click and drag trick or copy and paste.

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.

To Copy and paste formulas and functions:

Select the cell to copy and press **<Ctrl> and C** and then paste it in the cells intended for evaluation using **<Ctrl> and V**.

2.2.4 Absolute and relative referencing

When copying a formula to multiple rows, Excel will keep shifting the cells used in the formula down by one row. This is called *relative referencing* and is Excel's default for copying formulas and functions. If you wish to always evaluate a formula or function with reference to the value in one specific cell, this is called *absolute referencing*.

To use absolute referencing in a formula:

Take the cell reference that you intend to make absolute (for example B4) and add '\$' symbols

before the value (row and/or column) that you want to fix. Thus B4 becomes \$B\$4 if you want to fix both the row and column, while B4 becomes B\$4 if you only want to fix the row (4).

3. Data visualization

In this section we are discussing data visualization. This section covers how to create a scatter plot and how to properly format graphs. To describe these concepts, data from an example lab is used. The concepts and data discussed in this section are entirely for illustrative purposes so don't worry about understanding the physics, just focus on the Excel skills.

In this example lab, you are asked to compare the *Elasticity* of a ping-pong ball and a rubber ball. Elasticity is defined as the ratio of bounce height, h_b , to initial drop height, h_i . The initial drop height and bounce heights along with the associated measured error of the rubber ball and the ping-pong ball can be seen in Table 2.

Table 2: Experimental data for the bounce heights of the rubber ball and ping pong ball.

Trial #	Rubber Ball				Ping Pong Ball			
	h_{ri} [cm]	Error	h_{rb} [cm]	Error	h_{pi} [cm]	Error	h_{pb} [cm]	Error
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. We will determine whether our collected data fits this linear model using graphs for visualization.

3.1 Inserting a scatter plot

The Scatter Plot will be used for many future engineering applications, and plots each data point separately on the chart area. Do not confuse it with the Line Chart, which is similar but uses lines to connect data points. It should be noted that in the example lab 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. The concepts of dependent and independent variables will be explored further in other courses.

To insert a scatter plot:

Go to the **Charts** group in the **Insert** tab and select **Scatter**.

Note that an inserted plot will be blank until you add one or more data series to the plot area.

3.2 Selecting data series

The next step is to add a data series to the blank plot:

To do this, right-click on the blank plot and click on **Select Data...** in the Quick Menu. In the **Select Data Source** window, click on **Add**. Give a descriptive name in the **Series Name** dialogue box (in the example lab we'll plot the rubber ball data first and thus will type **Rubber Ball** into the 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 then click **OK** again.

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.

3.3 Chart formatting

An essential aspect of data visualization is proper formatting. An improperly formatted chart will likely not describe data in an adequate manner.

3.3.1 Title, legend, labels

Always 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 improve readability. 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>**.

To add a title, legend and labels:

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.

3.3.2 Tick marks

It's useful to tidy up the axis by displaying major and minor tick marks. This will clarify data ranges and improve the readability of the plot.

To alter tick marks:

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.

3.3.3 Axis ranges

Data should always be spread evenly across a plot area, as opposed to being bunched in one half of a chart. Bunching can be fixed by altering the axis ranges.

To modify the horizontal and vertical axis ranges and specify the types of numbers on the graph:

Right-click on the vertical axis, and then select **Format Axis**. In the **Axis Options** section, change the **Bounds**. This sometimes takes some playing around to get a nice spread of data. In the sample lab, having a **Minimum of 0** and a **Maximum of 120** produces an aesthetically pleasing plot. Note that you can always reset the axes to the initial auto-generated range by clicking the **Reset** button next to **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.

3.3.4 Gridlines and border styles

Another formatting option is to alter gridlines and border styles. There is no standard formatting for this, so feel free to play around. However, always make sure that formatting makes your chart easier to read and understand.

To 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**.

3.3.5 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 specified uncertainty. In the example lab, error for each point is specified in Table 2. Error bars for each series will be added separately.

The process for adding error bars to data is different in Excel for Mac, therefore a section is included highlighting difference in this process when using Excel for Mac.

The first step is to open the **Error Bar Options** toolbar:

To do this, 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.

Next, you need to select the correct type of error:

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.

To add custom error bars:

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 $\pm 2\text{cm}$, 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.

The last step is to format the data markers to make error bars easier to read:

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 (4 is usually a good size). 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.

A properly formatted graph of the ball data, created by following the above steps, is shown below in Figure 1.

3.3.5.1 Error Bars for Mac

The process for adding error bars in Excel for Mac is very similar, with the sole exception being opening the **Error Bar Options** toolbar.

To open the **Error Bar Options** toolbar in Excel for Mac:

Click on any of the rubber ball data points to highlight all the points in the series. Under the **Chart Layout** tab, select the **Error Bars** button in

the **Analysis**. Under the drop down menu, select **Error Bars Options...** to bring up the window. Select the Error Bars tab.

After this step, follow the same process described above to select the appropriate type of error and add the bars for both the x and y axis error. Regardless of your version of Excel, a properly formatted graph should look like the one included below in Figure 1.

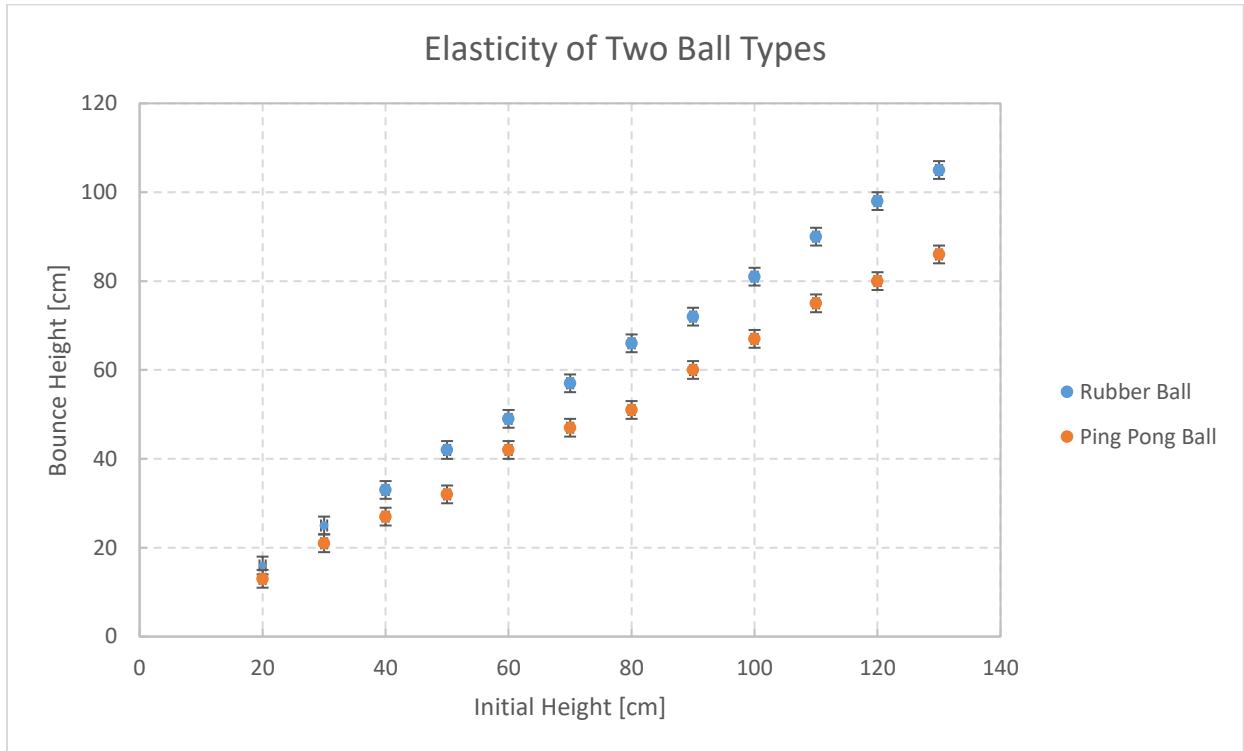


Figure 1: The final plot of the ball lab, showing both rubber and ping pong ball data

3.4 Extracting tables and graphs

To display plots and tables in your report, you need to move them from Excel in to Microsoft Word. 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. Because of this link 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.

To copy a table or graph:

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.

4. Data analysis

4.1 Linear regression

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, such as the relationship between drop and bounce heights explored in the example lab. When dealing with linear relationships, the best method to “draw” the perfect line is to use **linear regression**. The theory behind linear regression will be explored in upper year statistics courses, at this point it is sufficient to know the two methods of completing linear regression: trendlines and regression analysis using the Analysis ToolPak on PC/Mac (**only available in Microsoft Excel 2016**).

4.1.1 Trendlines

We’ll start with trendlines. It is useful to have the equations of your lines of best fit on your graph, along with the line itself. After plotting the data onto the chart area and completing relevant formatting, the following steps allow for trendline addition.

The first step is adding the trendline to the plot:

To do this, right-click on any of the data points 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.

The trendline equation that was added to your plot is likely overlapping your data points and/or your trendline. To move the equation:

Simply click and drag the equation such that it is located beside the trendline.

You can also change the variables in the equation:

Excel produces a default equation in the form of $y = mx + b$. In the example lab 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 other data series on the plot.

Ensure that both equations show an appropriate number of decimal places for the slope and intercept. These can be adjusted in the same manner as changing the variables. An example plot from the sample lab with two trendlines and equations is shown in Figure 2.

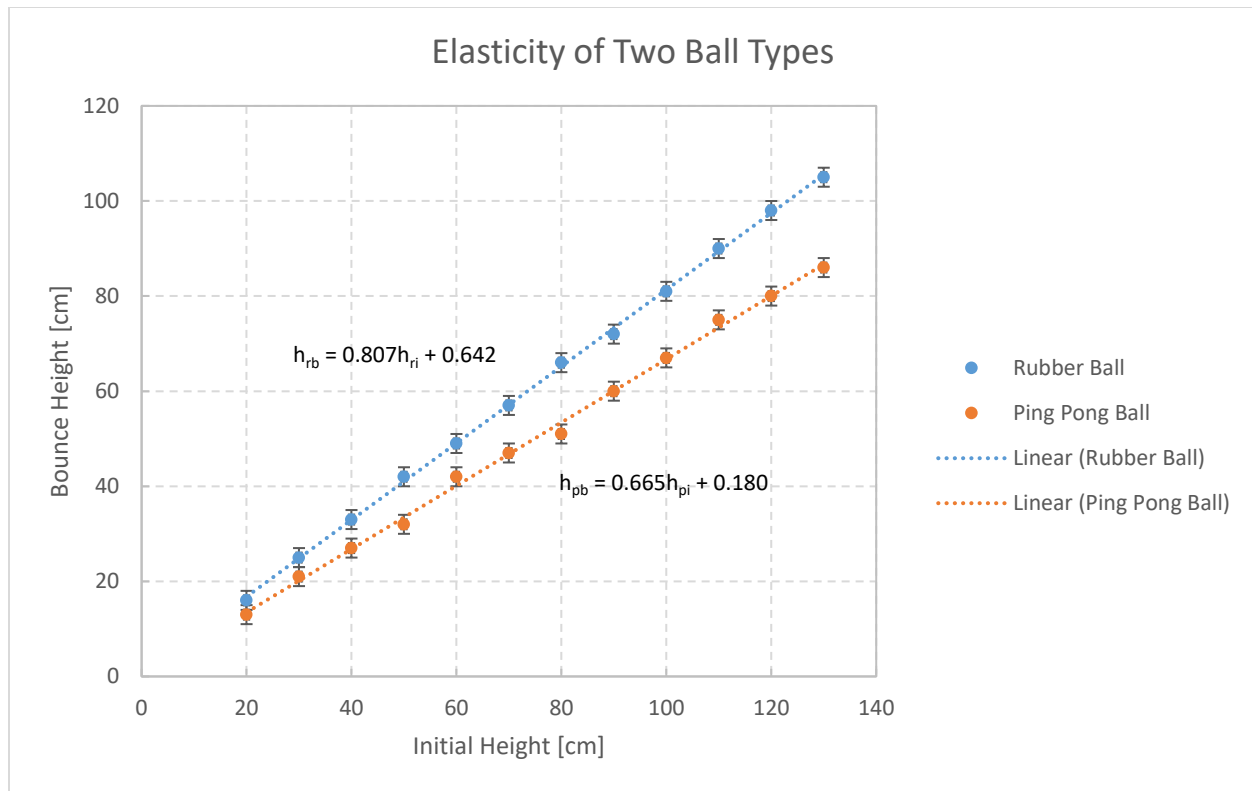


Figure 2: Linear regression analysis of rubber and ping pong ball drop data

4.1.2 Regression Tool in the Analysis ToolPak for PC/Mac

Another method of completing regression analysis is using the Regression Tool in the Analysis ToolPak within Excel. This tool completes rigorous statistical regression analysis on your data and generates a number of useful items such as residuals, standard error, and the coefficient of determination. These concepts will be explored in upper year statistics courses, but learning the tool can be done early.

Instructions for downloading the “Analysis ToolPak” for PC/Mac have been included in Section 1.3 and Section 1.4 above.

The first step is to launch the Regression tool.

To do this, 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** and then click **OK**.

Next, 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 **y-axis data** (in the example lab this is the bounce height, h_b) followed by **<Enter>**. Then click the **Collapse** button beside the **Input X Range** box and select the x-axis data (the initial height, h_i , in the example lab) followed by **<Enter>**.

Now you need to specify whether you included labels on your data and whether your relationship passes through the origin.

To do this, 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, changing the level to 68%.

Finally, 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 a descriptive title in the adjacent dialogue box. Check off the **Residuals**, **Residuals Plot** and **Line Fit Plot** boxes. Click **OK**.

A new worksheet with your title 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.

To format the Line Fit Plot:

Right-click on any of the Predicted 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 data points along the line of best fit.

4.1.4 Explanation of outputs

This section briefly explains some of the regression results from the Regression Tool. Again, this information is not essential to know but this section can be used a reference in upper years.

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 wish 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.:

The highlighted area in Table 3 displays the values which make up the linearly regressed trendline equation. The column headers are described in further detail below.

Table 3: Sample Regression output with critical values highlighted.

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	-1.220588235	1.374910683	-0.88776	0.38967	-4.169478366	1.728301895
hri [cm]	20.32941176	0.156179359	130.167	5.5E-23	19.99444036	20.66438317

Coefficients – this column provides the y-intercept and slope of the line of best fit. The Intercept row indicates the y-intercept, and the h_{ri} row [cm] indicates the slope. The regression should not be forced through the origin unless the data point (0,0) was recorded experimentally. 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 experimentally by the following:

$$h_b = E h_i + \text{intercept}$$

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. 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.

Residuals:

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.

4.2 Descriptive Statistics

Other common analyses of data include finding the mean, median and standard error. This can be easily completed using Descriptive Statistics in the “Analysis ToolPak” for PC/Mac.

4.2.1 Using the Analysis ToolPak for PC/Mac

First, launch Descriptive Statistics:

By selecting the **Data** tab, then **Data Analysis** in the **Analysis** group, then **Descriptive Statistics**.

Next, we want to select the data series.

Click on the input range and select the data for the sample from your worksheet.

Finally, we need to set the analysis parameters and output location.

To do this, check the box for **Summary Statistics**, and use a 68% confidence level for the mean.

Select an open area of the worksheet or a new worksheet for the output as desired. Click **OK**.