

Experiment

0

Introduction to Data Analysis Using an Excel Spreadsheet

I. Purpose

The purpose of this introductory lab is to teach you a few basic things about how to use an EXCEL 2010 spreadsheet to do simple data analysis in the labs.

II. References

Read this write-up and consult the Help button in Microsoft Office Excel 2010

III. Equipment

You will need a computer with Microsoft Office Excel 2010 installed. Note: the layout of the 2010 version of Excel is similar to the 2007 version but differs substantially from all earlier versions, such as Excel 2003. You need the 2010 version to do this lab.

IV. Pre-Lab Questions

There are no Pre-lab questions for Lab 0. For later labs, the Pre-lab questions are due before you start each lab.

Now would be a good time to read over the course syllabus to learn how much the Pre-lab questions count in your grade, when they are due and how you are supposed to turn them in. Also, check your course syllabus to see when this lab needs to be completed and when your first lab section meets. Be sure not to miss your first lab!

V. Introduction

In Physics 261 you will use computer spreadsheets to record and analyze your data. This lab is a tutorial on spreadsheets in general and specifically introduces you to EXCEL. You can use your own computer if you have EXCEL 2010 installed. Otherwise, stop by the Physics 261 lab if it is open or use a computer in one of the campus computer rooms.

Section VI of this write up discusses what a spreadsheet is and how it works. If you are already comfortable with spreadsheets, you can skip this section. Section VII specifically describes EXCEL 2010. Once again, if you are comfortable with this version of EXCEL, you can skip this section.

Finally, Section VIII is a tutorial exercise that takes you through a few spreadsheet operations that are used in the labs. You must complete this exercise and turn in a copy of both your spreadsheet and graph to get credit for doing this lab. It should take you less than an hour.

VI. Introduction to Data Analysis with Spreadsheets

Spreadsheets - What they are and what they are good for.

A spreadsheet is a computer program that turns your computer screen into a smart piece of paper. It removes much of the grunt work associated with repetitive calculations and lets you easily see the results of your work. We have chosen to use spreadsheets in the physics labs, because they reduce the amount of time needed to look at and understand the data taken in the labs. Traditionally accountants have used spreadsheets to do bookkeeping and budgets, but they make outstanding tools for scientists as well. With a spreadsheet, we can enter raw data, manipulate it and plot it all with a few simple commands. They are especially useful because of their built in ability to plot data.

One of key aspect that makes a spreadsheet so powerful is that whenever you change a number or formula in your spreadsheet, everything else in the spreadsheet that depends on that number or formula gets automatically recalculated, including plots. So if you make a mistake and have to re-measure a quantity or change a formula, all numbers connected to it are automatically updated. Another key aspect that makes a spreadsheet so powerful is that everything you do in a spreadsheet is saved and displayed in an intuitive graphical interface (the spreadsheet). With everything displayed in front of you, it is easier to understand your data and spot potential mistakes.

Yet another reason spreadsheets are extremely useful for data analysis is because of their ability to plot data. Rather than having to draw a graph by hand, you can just select the numbers you want to plot and the spreadsheet will do the work. Even better, if you change the numbers or formulas, the graph changes automatically. Still another advantage of a spreadsheet is that it can easily handle the **statistical analysis** of data sets with hundreds, thousands, or even tens of thousands of points, something that you would never want to try doing on a calculator. Several examples of functions that are particularly useful in analyzing real data are the **average**, the **standard deviation** and a least square fits of a straight line (known as a linear regression) to find the **slope** and the **intercept**.

Think about these advantages for a moment and then please take your calculator and put it away for the rest of the semester - you should not use a calculator for any of the labs because it is very poorly suited to the job of analyzing real data. Needless to say, spreadsheets such as Excel do have limits and will not be appropriate for every data set you encounter in the future. In particular, data sets with more than about 10,000 points, or that require symbolic manipulation, extensive signal analysis or image processing, are best handled using more sophisticated general purpose software (such as MatLab, Maple or Mathematica) or special purpose software.

The Basics

A spreadsheet consists of a collection of **cells** arranged in a big table. The cells are labeled by their column and row location (see Table 1 below). For example cell A4 is in the first column, the fourth cell down. A cell can contain a label (text), a number, or a formula. If you click on a cell, type in some text and hot enter, the spreadsheet will display the text in the cell. See cell A1 or cell C1 below in Table 1. We can also enter numbers like 15 (see A2). We can also put in formulas (see A2 and A3). Cell B1 contains a very simple **formula** that is, of course, equal to 6. In Excel, all formulas begin with an equals sign. If you type this formula into cell B1, you will see that the spreadsheet shows the number 6 instead of the formula $=3*2$ (see Table 2).

Table 1- This shows what to put in various cells in the spreadsheet

	A	B	C	D	E	F	G
1	text	=3*2	mass				
2	15	=2*A2					
3	=A2+1	=2*A3					
4							

Table 2- This shows what the spreadsheet displays

	A	B	C	D	E	F	G
1	text	6	mass				
2	15	30					
3	16	32					
4							
5							

The formula in cell B2 is different. It says =2*A2. What this means is twice the value of the cell A2. Since A2 is currently 15, B2 displays the value 30 (see Table 2). Cell A3 has a similar formula =A2+1, so it shows 16. One of the spreadsheet's greatest advantages can be seen when the number 15 is changed to 20. Automatically, B2 gets changed to 40 and A3 gets changed to 21. Also B3 would become 42.

Another really convenient feature of spreadsheets is the ability to replicate formulas. As an example, suppose that you wanted to extend the above spreadsheet so that the numbers go from 15 and 16 all the way to 25. Of course you could type 17, 18, ... into cells A4, A5 *etc.*, but this is unnecessary. Instead, you can use the **copy** and **paste** tools to replicate the formula in cell A3 to the cells A4 to A12. Cell A3 has the formula =A2+1 in it. Now what you would like in cell A4 is not exactly the same formula as A3, but you would like it to say A3+1 (not A2+1). This way it will become 17. When you tell the spreadsheet to copy and paste a formula, the formula is automatically changed in just this way. If you copy the formula A3 and paste it into cells A4 through A12, the formulas will become =A3+1, =A4+1, =A5+1 *etc.* all the way to A11+1. If you were to copy the formula in B2 and paste it into C2, it would change in a similar manner. It would change from =2*A2 to =2* B2, and in your example the value would become 2*(2*15) or 60. The key thing to realize is that you rarely have to type a formula more than once, even if it is used frequently. Also if you have a row of formulas and you want to change it, you can make the change once and copy it to all the other cells. Similarly if you change the number 15 in A1, all the numbers below it change.

Sometimes you will want to copy a formula, or move it from one place to another, without having it change. There are a several ways to do this in Excel. The simplest way is to use

cut and paste instead of copy and paste. For another way, see the discussion of the \$ symbol below.

VII. Basics of Using EXCEL 2010

The best way to learn how to use a spreadsheet is to just go ahead and start using one.

You can start Excel by clicking the Microsoft EXCEL icon in Windows. Once EXCEL is started you should see a screen that should look similar to the one shown below. The appearance of the window will change if you stretch it or click on some of the buttons, tabs or sliders.

Moving around the spreadsheet

You can move from cell to cell by using your mouse, the cursor keys, or by using the PgUp and PgDn keys. You can also go directly to a specific cell by pressing the F5 key and entering the cell address. The cell addresses in EXCEL are in the form sheet1!:C23. This is cell C23 on sheet1. You can also move around on the spreadsheet without changing cells by using the vertical and horizontal scroll bars (see Figure 1).

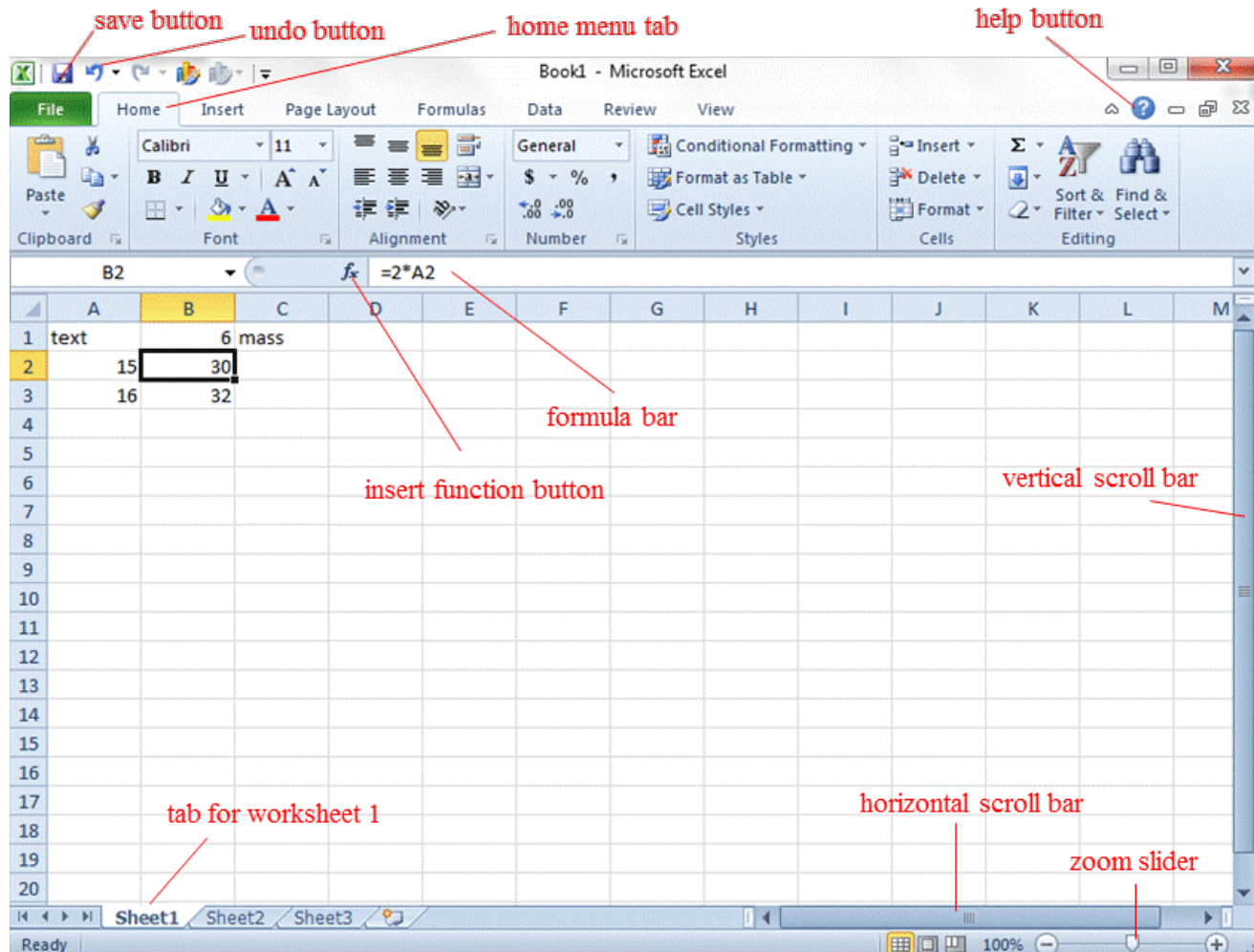


Figure 1. Layout of an Excel 2010 spreadsheet, highlighting some of the most important buttons, bars and menus. The “Home” menu is clicked and some text and formulas have been entered into the worksheet. Cell B2 has been selected and the contents are displayed in the formula bar.

EXCEL lets you have many worksheets in the same “notebook” which can all be connected to one another. Small tabs on the bottom of the page let you switch from sheet to sheet (see Figure 1). It is also possible to have more than one notebook open at the same time and the window menu at the top of the sheet lets you toggle between notebooks.

Entering Information

To enter information (text, numbers, formulas) into a cell, move to that cell and type the information. When you start to type in a cell, whatever you type appears in the cell and also on the top left of the page in the formula bar or display line (see Figure 1). If you press enter after typing something, it will enter it on the sheet; or if you press the escape key, it will abort the entry and leave the cell unchanged. When you are entering an expression in a cell, you can edit that expression by moving the mouse to the spot you want to change (on display line). Pressing enter or moving the cursor to another cell enters the expression. If you want to change a previously entered expression, you can move to the cell and retype the whole thing or move the mouse to the display line, click the left mouse button, and then revise the expression.

Entering Text

You will need to put labels on your data to make your spreadsheet understandable. It is really simple to set up nice looking labels. Just type words into a cell and they will show up in the cell. For example if you type “text” in cell A1, it will show up. If the text is longer than the cell width (which is adjustable) the whole text will show up if there are empty cells to the right of the cell with the text in it. If the string is long and the cell to the right occupied, the string will show up cut off - but it is all still there.

Entering Numbers

Numbers are simply entered in a cell. A number like 15 can just be typed and its value is entered in the cell. If the number is negative just start it with a minus sign.

Entering Formulas

Formulas always start with an equal sign. A formula can be simple numeric expression such as =2*3, or it can include more complicated expressions involving other cells and statistical and mathematical functions. A formula like =2*B5^3 follows the rules of programming where this means 2 times the contents of cell B5 raised to the third power. Use parentheses in a natural way - for example, you can write:

$$=1/((1/B3)^2+(1/B4)^2)^(0.5))$$

The spreadsheet has numerous mathematical functions such as sine (=SIN), cosine (=COS), and square Root (=SQRT), which are almost the same as in a normal mathematical expression. =SIN(C3) takes the sine of the contents of cell C3 (the angle must be in radians). Many other functions exist. The list of available functions can be viewed by clicking on the function button f_x (see Figure 1) and choosing “all”.

Functions, such as the sine, take normal numerical arguments. Some statistical functions take lists of cells or a range of cells. An example of this type of function is the average. The average function will compute the average of a list of numbers or the average of the numbers in a list of cells. So if you enter =AVERAGE(5,6,7,8) the value of the cell will be 6.5. This type of function is most useful when you give it a list of cells. If you wish to find the average of a

column of numbers that are in say cells A1 to A20, we just enter =AVERAGE(A1:A20). Most spreadsheet functions are “smart” so that if your range includes empty cells, they are not counted in the average or other statistical function. (**Note: there is a distinction between empty and zero.**)

Pointing at Cells

When you first start using a spreadsheet, you may often find yourself typing cell numbers into equations and expressions. All this typing is unnecessary and can lead to errors. The easiest way to put a cell number into an expression is to click on the cell, or “point” at it. For example, suppose that you have a number in Cell A3 and are writing an expression in which you want the cosine of the number in Cell A3. Simply type “=COS(“ then move the mouse or cursor to cell A3 using the page up/down/left right buttons or by clicking on the cell. Then the expression you are typing will look like =COS(A3. At this point do not press enter because you have not completed the expression. Just type the closing parenthesis and then hit enter and you will have =COS(A3) .

Selecting Ranges

You can also point at groups of cells. This is useful for functions such as the average or standard deviation that expect ranges as arguments. It is also useful when you want to mark a region that needs to be moved or copied. Anytime you need a range of numbers, use the mouse to move the cursor to the first cell in the range and hold down the left mouse button. Now “drag” the cursor to the last cell in the range and release the button. You should now have the range highlighted and displayed in the display line.

Copying a formula

Suppose you have a column of numbers in cells D2,...D20, and you wish to evaluate the sine of each number (see Figure 2). To begin, enter the numbers into D2... to D20. Next in cell E2 enter the formula =SIN(D2). Now copy the formula to the entire range E2,...E20. To do this put the cursor on E2 and press the copy button (on the keyboard type either Ctrl-C, or move to the menu, click on edit and select copy). Now it expects a cell or range to be the object to be copied. Use the techniques described above to select the output range E2,...E20. (It is OK to copy the formula onto itself). Hit enter and your formula will be replicated over the entire range.

Note that in cell E2 the formula is =sin(D2), but in cell E3 it has been updated to =sin(D3) rather than Sin(D2). This smart copying by Excel is very useful and leaves you with column E with the Sine of the values in column D, in just the way you would want.

There is another type of copying where the cell location does not change. To fix a cell location so that it does not change in copying, you can put a \$ in front of the cell name, e.g. \$B\$1. Note that there are two dollar signs in this expression, because it is possible to fix the column and row or both so that they do not change when you copy. An example where you might want to use this is: You have a column (A1,...A20) of measured accelerations, and you want a column of forces and all objects have the same mass. If the mass is in cell D1 your formula in B1 might be =A1*\$D\$1. This way when you copy the formula to B2 it will be =A2*\$D\$1. If

you did not use \$ signs the formula would have be \$A2*D2 which would be wrong if the mass was in D1.

Plotting Data

One very useful feature of spreadsheets is the capability to plot graphs. Using EXCEL, it is pretty simple to plot one column of numbers versus the numbers in a second column.

For example, suppose you want to plot $\sin(x)$ vs x . To do this, you will need to have some x -values that we can take the sine of (if you did not already do this when working through the previous part). Start by entering the label “X” into cell D1. Next enter 0 into cell D2. Then in D3 enter $=D2+0.3$. Next copy this formula from D2, D3, ... D20. Next enter the label “Sin(X)” into cell E1. Now in E2, enter $=\text{SIN}(D2)$. Copy this formula to cells E2,...E20 (see Figure 2).

To plot this data, go to the top of the spreadsheet and click on the **Insert** tab (see Figure 2). Now use your mouse to point at and select cells D1 to D20 and cells E1 to E20 (just click on cell D1, hold the left mouse button down as you drag the mouse to cell E20, and then release the mouse button). Then look for the section of the menu at the top of the spreadsheet that says **Charts** and click on **Scatter** (see the note below about why you should never use anything other than Scatter for plotting). You will be presented with a sub-menu of different types of scatter plots (just points, just lines, lines and points, etc.), choose any one of these by clicking on it. You should now see a plot and notice that the menu at the top of the spreadsheet has changed. Now click on the **Quick Layout** menu and try selecting Layout #10, which has a pretty nice format. All of your plots should have labels on the x and y -axes. In this case, Layout #10 just has default labels that say “Axis Title”. To change the axes labels, just click on them and type in X for the x -axis label and $\sin(x)$ for the y -axis label.

Your Spreadsheet should now look like Figure 2. If your plot is on top of the x and $\sin(x)$ columns, just click on the cahrt and you can move it with the mouse. Notice that when you click on the chart, Excel will highlight which columns are being plotted (see Figure 2). This is very useful because one of the most common mistakes students make in the labs is to plot the wrong set of numbers.

In some of the experiments, you will need to plot different curves on the same plot. The easiest way to do this is to start by plotting the first curve. After you have this plotted, then right click on the plot and click on **Select Data** in the pop-up menu that appears. A new pop-up menu will open and you should click on **Add**. Once you do this, yet another pop-up menu will appear which will allow you to enter a series name and select the x and y columns for the second data series. Once you fill these in, just click OK and you should see your new plot with both curves.

Sometimes you will make a plot and forget to add axes titles. To go back to a plot and add axes titles to a graph that does not have any, click on the graph and notice that a new menu appears at the top of the spreadsheet called **Chart Tools**. Select the **Layout** tab and then select **Axes Titles** from this menu and figure it out from there. Another way to add axes labels is to click on the chart and select the **Design** tab in the **Chart Tools** menu. The **Design** menu will appear and you can then pick from one of the **Chart Layouts** that has axes labels already in them (you will need to edit the default titles but that just involves clicking on the label and typing). In the examples shown below in Figures 2 and 3, we selected Layout 10.

You should be aware that there are some potential issues with plotting graphs in Excel. One problem is that, although there are many different types of plots in Excel, only a **Scatter** plot will plot an x -value and a y -value as a point at location (x,y) in the Cartesian plane. Some of the

other types of plots can fool you into thinking that they are plotting (x,y) points, but they are not. For example, a closer look at a Line plot reveals that it actually plots the y-values versus the order in which they are given. To be clear, you should only use **Scatter** plot for all of your plotting - never use a Line plot or a Bar chart or any of the other types of charts. Another problem with plotting in Excel is that it is not so easy to find what you want in all of the menus and sub-menus that you can access for changing a plot. The best thing to do is play around with Excel and after a while you will start to get a feel for where everything is and what the various buttons do.

Fitting Data

The spreadsheet allows you to do least square fits to data. To see how this works, make some data that follows a linear trend by copying the same numbers you have in column D into column F and label them with T for time. Next enter $=2 * F2 + 1$ into cell G2 and copy this down into cells G3 to G20. Excel has built-in functions that use a least square fit to determine the **slope** and **intercept** that best fit a set of data. To determine the **slope**, go to H2, and enter

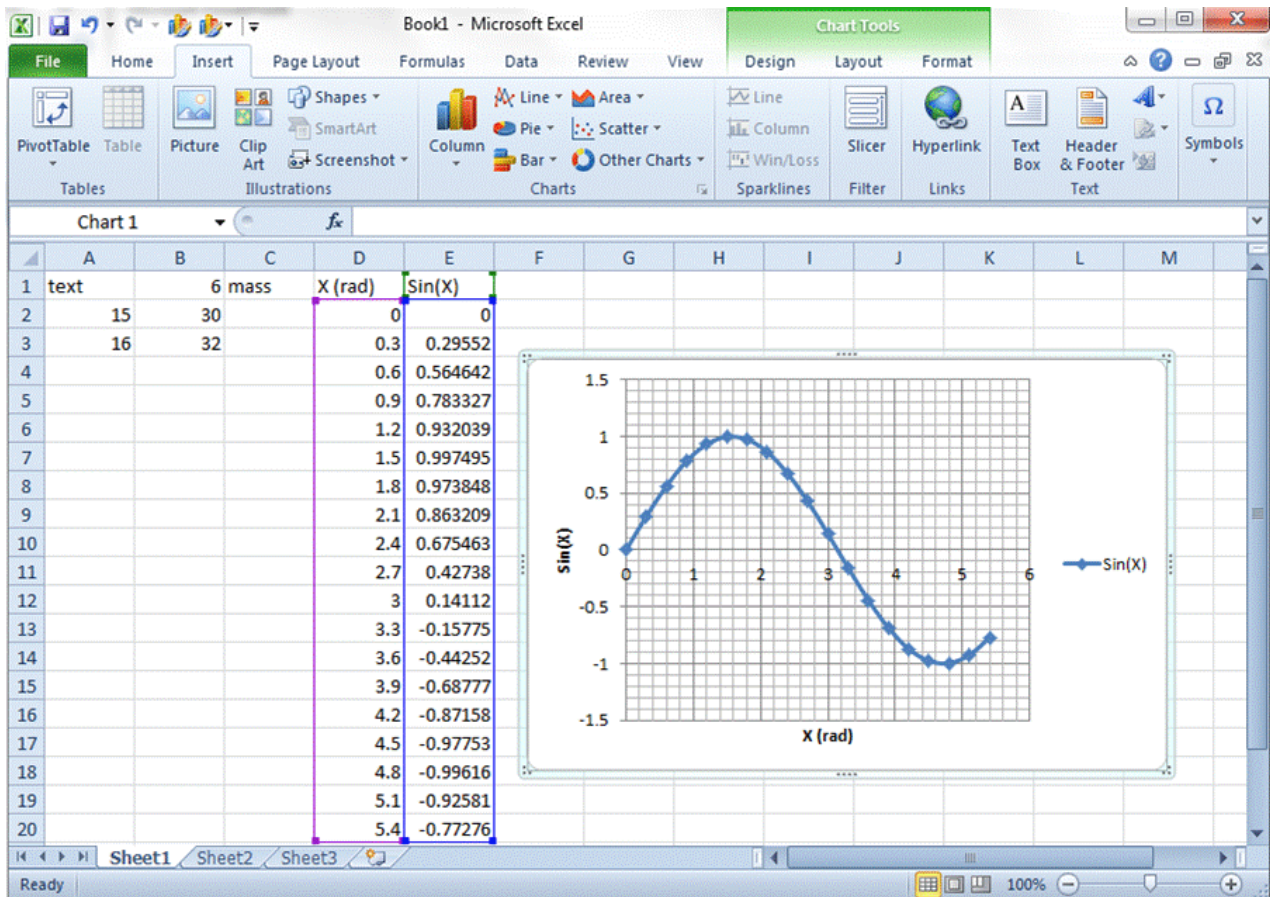


Figure 2. Scatter plot of Sin(x) versus x. Note that the chart was clicked on and this caused Excel to highlight the x and y columns being plotted and display the Chart Tools menu.

$=SLOPE(G2:G20,F2:F20)$. To determine the best estimate for the **intercept**, enter into cell H3 the formula $=INTERCEPT(G2:G20,F2:F20)$. Using these slope and intercept values, you can easily calculate the best estimate for the y value for any given x value.

Another way to fit a straight line to a set of data is to perform a least square fit. To do

this, you need to first click on the **Data** menu tab from on the top of the spreadsheet, then click on **Data Analysis** (it's on the far right side), select Regression and fill in the pop-up menu.

Finally, it is worth noting that in some of the labs, you will fit data to a straight line or some other function using yet another way, called χ^2 minimization. You need to use χ^2 in a situation where different data points have different uncertainties. If you tried to use a least-square fit technique (such as used in SLOPE, INTERCEPT or the linear regression routine) you would get a result, but it would not actually be the best fit to the data.

VIII. An Exercise

Suppose that you have performed an experiment to measure the speed of a cart rolling down an inclined plane. The cart has mass 0.25 kg and you made speed measurements at 1 second intervals from 0 to 10 seconds. We expect there should be a linear relationship between speed and time. Here is the data:

Time (s)	Speed (m/s)
0	0
1	0.8
2	1.5
3	1.6
4	2.5
5	2.7
6	3.2
7	3.9
8	4.5
9	5.0
10	5.5

Step 1 - Turn on the computer and start Excel

Step 2 - Enter your name and section starting in cell A1

Step 3 - Enter the label "Mass(Kg)" in cell A2 and put the value 0.25 in cell B2.

Step 4 - Label cell A4 "Time (s)" and cell B4 "Speed (m/s)"

Step 5 - Enter the above data show into the spreadsheet below the Time and Speed labels.

Step 6 - Now create a new column starting with the label K.E. in cell C4.

Step 7 - In cell C5, enter the formula for the kinetic energy of the object with the speed shown in cell A5. (Remember that kinetic energy is $1/2 mv^2$).

Step 8 - Check that you have fixed the cell location of the mass. For your formula in cell C5 you should have $0.5*\$B\$2*B5^2$

Step 9 - Copy the formula in C5 to the range C5...C15

Step 10 - Check that your screen looks pretty much like the one shown below:

Step 11 - Now calculate the slope and intercept of the data in column B with respect to column A. Put the slope in D5 and the intercept D7 and add some appropriate labels above both these cells so you don't forget what they are.

Step 12 - Next set up a column that contains the best fit to the data. This column should start with the label "Best Fit Speed" in E4. The best fit for the speed (which you should think of as the y-values) as a function of the time (which you should think of as the X-values) are found by remembering that a line is $Y = mX + b$. Also remember that the slope (D5)

and intercept (D7) should be “fixed” with \$ signs so that when we copy the formula they don’t change location.

Step 13 - Plot both the data and the best fit together on the same graph.

Step 14 - Label the axis on your graph.

Step 15 - Figure out how to make it so that the data is plotted as blue points and the best fit is plotted as a red solid line without points (see Figure 3). Hint: try right-clicking on a point in your plot and then select Format Data Series and follow where it leads.

Step 16 - Check that your spreadsheet and graph look like the one shown below.

Step 17 - If you have not already done so, save your spreadsheet.

Step 18 - Submit a copy of your spreadsheet to the ELMS Blackboard Physics 261 site (if your instructor requires this) and/or print out a copy (if your instructor requires a printed copy). If you print a copy, you must first select the region of the spreadsheet to print. **Use a white background in your plot to save toner, and to obtain a better looking graph.**

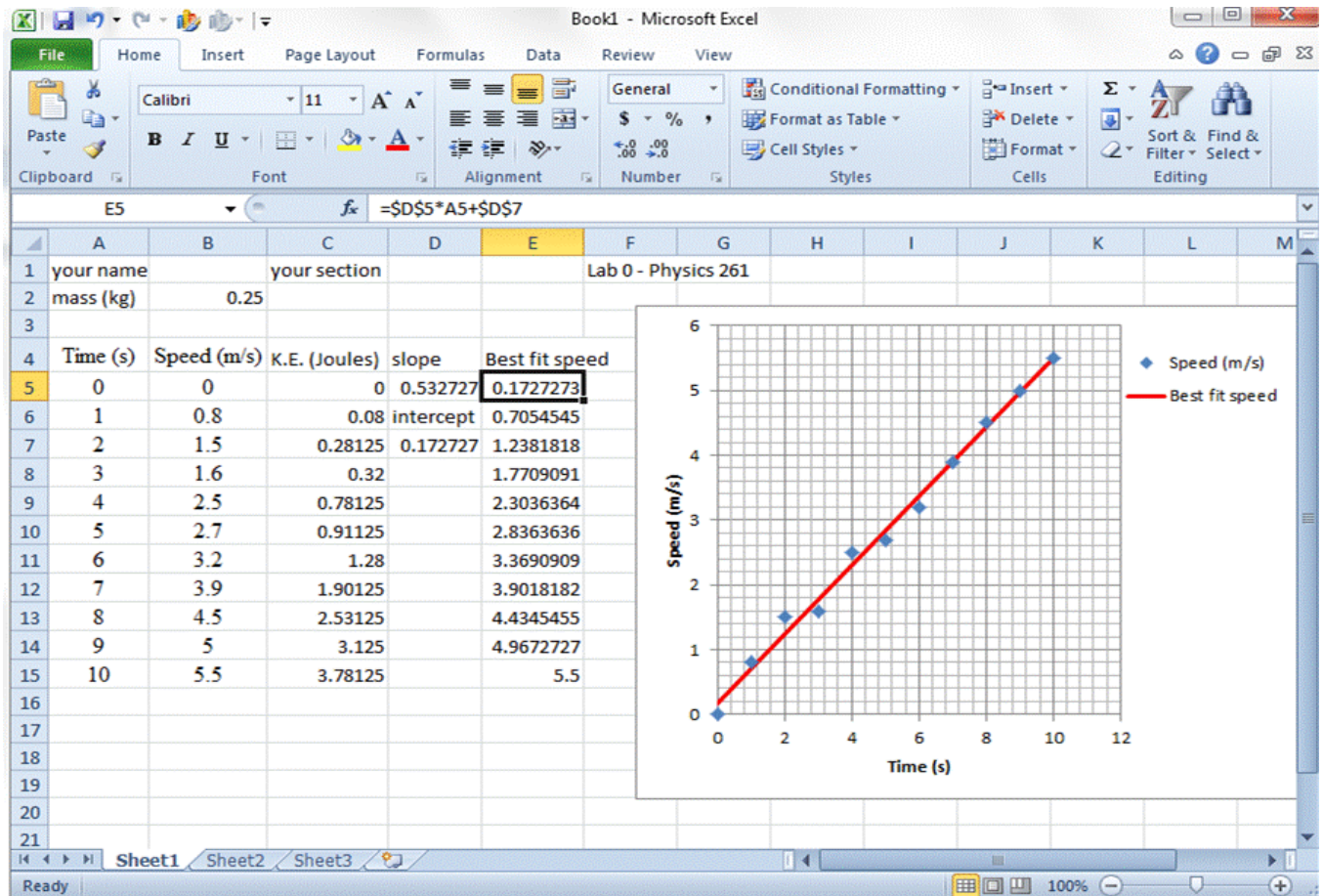


Figure 3. Spreadsheet for the exercise.