AN INTRODUCTION TO EXCEL 2010
EXCEL LAB #1

Overview

This lab is written for Excel 2010, which is available to students in the library. The notation => can be read as "go to" or "click on." This notation will most often be used when navigating the menu or toolbars in Excel. To indicate a command or icon that you might click on or search for in Excel, **bold** will be used. Likewise, anything that you are to type into Excel will be **bolded** in the instructions. Do not enter such text as bolded text unless the instructions ask you to do so.

Tutorial

1. Start 2010 Excel by clicking on the icon on the desktop or by running the program through the Start Menu under Microsoft Office.

2. *Importing .txt Data* and *Saving Files*.

    1. In Excel, **File** => **Open** => and browse for the file wherever you saved it. Because **Highschools.txt** is a text file (and in particular, not a .xlsx file), you should change the extensions of the files Excel searches by changing **Files of Type** to **All Files**. When you find the file, click on Highschools.txt => **Open**.

    2. Because **Highschools.txt** is not an Excel file, Excel will run its Text Import Wizard when opening this file. When the Text Import Wizard box opens, look at the text in the box and see how your data is listed. If it would be in columns, then you would click on **Fixed Width**. However, this isn't the case here. With this file, items are separated by a common character, specifically a comma, so click: **Delimited box** => **Next** => **Comma** => **Next** => **Finish**.

    3. In order to make sure you do not need to repeat this process, save your file as an Excel file called **Highschools.xlsx**. You will need to change the **Save Type** to allow for a .xlsx extension which is the Excel Workbook choice at the top of the **Save as Type** choices. Once you have created an Excel file, you should save your work frequently by either clicking on the **disk icon** in the upper-left or **File** => **Save** or by pressing **CTRL-s**. You also must know where you save files.

3. *Renaming Worksheets*.

   1. Excel keeps data in worksheets. The bottom toolbar shows that Excel has named the worksheet **Highschools**. Right click on the worksheet name => **Rename** => then rename the worksheet **Original Data**.

Check that your new Excel data matches the original data by comparing it to the codebook (which is found at the end of the lab). Specifically, your file should contain 14 variables and 588 observations.

- Each one of the 14 variables is contained in a separate row, labeled A through N (N being the 14th letter of the alphabet).

- Each observation in the data is listed in a unique row, labeled by numbers along the left-hand side of the worksheet. Currently there are 588 observations in the data as rows 1 through 588 contain data.

4. *Inserting a blank row.*

   Insert a blank row at the top of your spreadsheet: **right-click on Row 1** => **Insert**.

5. Type the variable names in the cells in row 1 according to the codebook (which can be found at the end of the lab instructions). For example, enter **schlname** in cell A1; enter **schltype** in cell B1; and so on, finishing with **pnotkr** in cell N1. Variable names should now be in cells A1 through N1, while the data are in cells A2 through N589. Notice that the last of the 588 observations is now in row 589. This will obviously happen whenever one puts variable names in the first row.

6. *Setting Font Type and Size*.

   1. Change the font to Times New Roman 10 point for all 14 columns of data: left click on the A column and drag right to highlight columns A – N => **Home** => **Font** box => **Times New Roman** & **10**.

   2. Bold all variable names in row 1: **left click on row 1** => **Home** => **Font** box => **B** (for bold). Similar steps can be taken to underlined or italics text.

7. *Changing Column Width.*

   1. Change the width of column A to 35: **right click on column A** => **Column Width** => **35** => **OK**.

   2. Repeat for Column C: **right click on column C** => **Column Width** => **35** => **OK**.

   3. Change column B's width to 8: **right click on column B** => **Column Width** => **8** => **OK**.

   4. Change the width of columns D through N to 8: **left click on column D and drag through column N** => **right click** anywhere in the shaded area => **Column Width** => **8** => **OK**.

8. *Changing Alignment.*

One changes the alignment of cells (left justified, center justified, right justified) by highlighting the designated cells and then clicking on the appropriate icon in the Alignment box under the Home tab on the toolbar.

Right align all cells in columns E through N: **left click on column E and drag through column N** => **Home** tab => **Alignment** box => **Right Justified** icon (third from the left on the second line of the Alignment box).

9. *Freeze Panes.*

Freeze panes allows one to always keep certain columns or rows in view.

1. Employ Freeze Panes so that the school name (column A) and the variable names (row 1) are always in view: **left click on cell B2** => **View** tab => **Window** box => **Freeze Panes Freeze Panes** (you could freeze just the column or the row too if you wanted by selecting appropriately). You will see a dark horizontal line between rows 1 and 2 and a dark vertical line between columns A and B. Freeze panes will keep the data outside of these lines always in view. Thus, regardless of where you scroll, you will always see the variable names on top and the district names on the left.

2. Don't do this now, but to turn off freeze panes, repeat the process: **View** tab => **Window** box => **Freeze Panes** => **Unfreeze Panes**.

10. *Changing Number Formats.*

1. The variable **white** reports the percent of students in each school who are white, measured from 0 to 100. For example, the first observation is A-C Central High School, for which 97.7% of the students are white. The data for this variable, however, are not all reported to the same number of decimal points. Change the numeric formal so that all data for white is reported to 1 decimal point: r**ight click on column E** (the **white** variable) => **Format Cells** => **Number** (tab) => **Number** (category) => **1** (decimal places) => **OK**.

2. Change the number format so that both **psae_s** and **psae_d** also report only 1 digit after the decimal place: **left click on column L and drag through column M** (the **psae_s** and **psae_d** variables) => **Format Cells** => **Number** (tab) => **Number** (category) => **1** (decimal places) => **OK**.

3. Change the number format so that **fundppd**, **eavppd**, **texpppd**, and **pnotkr**, each report 3 digits after the decimal: **left click on column G and drag through column I** (the **fundppd**, **eavppd**, and **texpppd** variables) => **right click anywhere in the shaded area** => **Format Cells** => **Number** (tab) => **Number** (category) => **3** (decimal places) => **OK**. Right click on column N and repeat (for the **pnotkr** variable).

The **Format Cells** option (under the **category** tab) allows you to manipulate the cell values in many ways – using currency labels, scientific notation, percentage, allowing for different sorts of dates, and so on. Whenever you want to manipulate how numbers look, this is where you would go.

11. *Sorting Data*.

One of the most powerful tools in Excel is the ability to sort data. For example, consider the following question. How many high schools are in Peoria School District #150? You should be able to find the answer fairly easily by scrolling down. (The answer is four.) A question about a particular high school, however, would be more difficult to answer because the high schools are not in alphabetical order. For example, what school district is Lincolnwood High School in? As the data is not sorted by high schools, and you are not lucky enough to have Lincolnwood High School be in the Lincolnwood School District, you would potentially have to read through all 588 observations to find Lincolnwood High School. The easier thing to do is to sort the data by high school.

Note that whenever you are sorting your data, it is extremely important that you highlight all of the data. You do not want to simply sort the list of high schools; you want to sort all of the data associated with each high school. This is why one must highlight all of the data, columns A through N in what we do below.

1. Sort all of the data by high school name: **left click on column A and drag through column N** => **Data** tab => **Sort & Filter** box) => **click on the A to Z↓** box. To check, you should have that A B Shepard High School (Campus) is now in the first row.

Now by scrolling down to Lincolnwood High School, we see that Lincolnwood High School is in Panhandle Community Unit School District #2.

Many times we want to sort by a variable other than the one listed on the far-left of the data. For example, consider the following question. What high school has the greatest percentage of students passing standards on the PSAE? To answer this, we need to sort the data by **psae_s**.

2. Sort the data by **psae_s**: **left click on column A and drag through column N** => **Data** tab => **Sort** tab => change **Sort By** to **psae_s** => change **Order** to **Z to A** in order to sort from highest to lowest => **OK**. The data are now sorted by **psae_s**. Northside College Preparatory High School having the highest pass rate of 96.4 percent.

Consider the following question. What ***unit district*** has the greatest percentage of students meeting or exceeding standards on the PSAE? One way to answer this question is to sort the data by **psae_d**, and then read down the data until one gets to the first unit district (rather than a high school district). And in this simple problem, that would be fine. For other problems, however, that might not be all that useful. The way to formally handle this problem is to sort by two variables – have Excel sort by district type first (**disttype**), and then within district type sort by district PSAE pass rate (**psae_d**).

3. Sort the data by **disttype** and then **psae_d**: **left click on column A and drag to column N** => **Data** tab => **Sort** => change **Sort By** to **disttype** => change **Order** to **A to Z** if need be => **Add Level** => change **Then By** to **psae_d** => change **Order** to **Largest to Smallest**.

By scrolling down, we now see that the first Unit district type is now located in row 146. To get there, look at column D, **disttype**, and scroll down until you get to the Unit district type. From here we see that the Naperville Community Unit District #203 has the greatest district-wide PSAE pass rate at 84.1 percent.

4. At this point, to keep everyone on the same page, sort the data in the **Original Data** worksheet one final time. This time, sort by **schlname** (A to Z) as the first variable and then by **distname** (A to Z) as the second variable: **left click on column A and drag through column N** => **Data** tab => **Sort** => set **Sort By** to **schlname** and set **Order** to **A to Z** => **Add Level** => set **Then By** to **distname** and set **Order** to **A to Z**.

12. *Inserting and Renaming a Worksheet.*

Although we now want to leave the original data alone as sorted, we still want to manipulate the data to answer a few more questions. The way to do this is to copy the data into a new worksheet, and manipulate the data there. To do this:

1. Create a new blank worksheet: **Home** tab => **Cells** box => **Insert** => **Insert Sheet**. This can also be done by left clicking on the "create sheet" icon that is located immediately to the left of the current worksheet name Original Data on the bottom toolbar. A blank worksheet can also be created by typing Shift=F11.

2. A blank Excel worksheet will appear. Excel has named this Sheet 1. Rename this worksheet **School Data**: **right click on Sheet 1** => **Rename** => **School Data**.

3. Order the worksheets so that **Original Data** is on the left and **School Data** is on the right: left click and drag the **School Data** worksheet to the right of the **Original Data** worksheet.

13. *Copying and Pasting Columns of Data.*

The above step created a new worksheet, but it didn't actually copy the data. We will now copy and paste (not cut and paste) the data from **Original Data** to **School Data**.

1. Left click on the **Original Data** worksheet tab in the bottom left of the screen.

2. Left click on column A => drag to highlight through column N.

3. Right click anywhere in the shaded area => **Copy**. Alternatively one can copy using **Home** tab => **Clipboard** box => **Copy** icon. Alternatively one can use CTRL-C.

4. Left click on the **School Data** worksheet tab on the bottom toolbar.

5. Right click on cell A1 (or on column A) => **right click in the highlighted area** => **Paste** (first icon in list). Alternatively one can paste using **Home** tab => **Clipboard** box => **Paste**. Alternatively one can use CTRL–V.

6. In the **School Data** worksheet, you only want to keep four variables: **white**, **enroll**, **psae_s**, and **psae_d**: **left click on column A and drag through column D** => **right click in the shaded area** => **Delete** => **left click on column C and drag through column G** (variables **fundpp** through **tsttkr**) => **right click in the shaded area** => **Delete** => **right click on column E** (variable **pnotkr**) => **right click in the shaded area** => **Delete**.

One can also make a new worksheet that has all the data from another worksheet by right-clicking on the worksheet (**Original Data** in this case) => **Move or Copy** => select **Create a Copy** => **OK**. After the copy has been made, one can rename it and move it according to step 12 above.

In the **School Data** worksheet, you should now have four variables: **white** in column A, **enroll** in column B, **psae_s** in column C, and **psae_d** in column D.

14. In cells G1 - J1, type **white**, **enroll**, **psae_s**, and **psae_d** respectively. In cells F2 through F9 type **Sum**, **Number of Obs**, **Mean**, **Median**, **Minimum**, **Maximum**, **Sample Variance**, and **Sample Stnd. Dev**.

15. Change the font to Times Roman 10 point for columns E through J: **left click on column E and drag through column J** => **Home** tab => **Font** box => **change the font to Times Roman and change the font size to 10 point**.

16. Change column F's width to 15: **right click on column F** => **Column Width** => **15** => **OK**.

17. Make columns A through N have one number after the decimal point: **left click on column A and drag through column N** => **right click in the shaded area** => **Format Cells** => **Number** (tab) => **Number** (category) => **1** (decimal places) =>**OK**.

18. As student enrollment (**enroll**) is never reported with decimal points, go back and change these values to be integers: **right click column B** => **Format Cells** => **Number** (tab) => **Number** (category) => **0** (decimal places) => **OK**.

Excel lets you manipulate data and calculate descriptive statistics by entering formulas and by using functions (such as **AVERAGE**) that Excel has defined. To indicate that you are doing this, Excel requires that all formulas and functions begin with an **=** sign. The lab directions will include all formulas in capital letters, but it is not necessary to enter them this way.

19. We want the sum of all of the **white** values to appear in cell G2: in cell G2 enter **=SUM(A2:A589)**. (You don't type in the bold of course.) You should get a value of 46,318.6.

    The function **SUM** for Excel means to add up all of the values in the specified cells. Moreover, one tells Excel which cells to apply the function to by placing the cells in parentheses. In this case, **=SUM(A2:A589)** tells Excel to add up all of the values in cells A2 through A589. Notice that the colon, **:**, means "through."

Next we want to put the number of non-missing observations for **white** in cell G3. We suspect that this number will be 588, but maybe there is a missing value or two in the data set.

20. Put the number of non-missing observations for **white** in cell G3: in cell G3 enter **=COUNT(A2:A589)**. The value in cell G3 should be 588 now, indicating that there were no missing values.

    Notice that the **COUNT** formula in Excel does not count up the data (**SUM** does that), but rather **COUNT** adds up the number of observations which report any data value.

In G2 is the sum of the observations on **white** and in G3 is the number of **white** observations. Thus, we can calculate the mean value of **white** by dividing the value in G2 by the value in G3. We could do this by entering **=G2/G3** in cell G4. Rather than entering the formula by hand, however, we can also let Excel do it.

21. Put the average value of **white** in cell G4: in cell G4 enter **=AVERAGE(A2:A589)**. The value in cell G4 should now be 78.8.

22. Put the sample median for **white** in cell G5: in cell G5 enter **=MEDIAN(A2:A589)**. The median value of **white** is 95.2.

23. Put the sample minimum and sample maximum for **white** in cells G6 and G7: in cell G6 enter **=MIN(A2:A589)** and in cell G7 enter **=MAX(A2:A589)**. The respective values should be 0.0 and 100.0.

24. Put the sample variance for **white** in cell G8: in cell G8 enter **=VAR(A2:A589)**. The sample variance for **white** is 976.9.

25. Put the sample standard deviation for **white** in cell G9: in cell G9 enter **=STDEV(A2:A589)**. The sample standard deviation for **white** is 31.3. Notice that, as we know the sample standard deviation of a variable is the square root of the variance for the variable, we could have entered in cell G9 **=sqrt(G8)** and we would have gotten the same answer.

26. ***Copying and Pasting Functions.***

    Next we want to enter the same eight descriptive statistics for each of the three remaining variables (**enroll**, **psae_s**, and **psae_d**) in cells H2 through J9. Of course we could retype all of the functions again, three times each. Instead of doing this, though, we will take advantage of Excels ability to "roll" functions when copying and pasting functions. Do the following: **left click on cell G2 and drag to highlight through G9** => **right click in the shaded area** => **Copy** => **left click on cell H2 and drag through J9** => **right click in the shaded area** => **Paste** (first icon on the left).

    When you entered the paste, values automatically appeared in cells H2 through H9. A value of 535,639 should be in cell H2. Left click on cell H2. If you look in the textbox above the worksheet, instead of seeing the value 535,639, which is the sum of all enrollments, you will see the function command **=SUM(B2:B589)**. The point is that when you copied cell G2 to cell H2, which is copying one cell and moving it one to the right, Excel automatically rolls the cell range one to the right as well. Specifically in this case, Excel changed the **(A2:A589)** which is located in cell G2 to be **(B2:B589)** in cell H2. Moreover this happened for all of the functions now in cells H2 through H9 as well. For example, cell H5, which reports a value of 567.0 is actually arrived at by the function as listed in the textbox as **=MEDIAN(B2:B589)**.

    To demonstrate that Excel does the rolling automatically, and for as many columns or rows as indicated by the copying-and –pasting, look at the textbox for the statistics in columns I and J. All of the functions in column I apply to cells **(C2:C589)** and all of the functions in column J apply to cells **(D2:D589)**. Note: when you copy and paste a function, if you don't want Excel to automatically roll the column and row identifiers for the cells, you need to place a dollar sign, $, in front of the column or row identifier that you do not want to have rolled. We will see an example of this shortly.

27. ***Descriptive Statistics (Data Analysis Toolpak).***

   Instead of specifying each descriptive statistic separately, we could have used Excel's built-in Descriptive Statistics feature of its Data Analysis Toolpak to produce all of these statistics (and many more) in just a few clicks.

   1. Launch the Data Analysis Toolpak: **File** (tab) => **Options** (toward the bottom of the list of options on the left hand side) => **Add Ins** (toward the bottom of the list of options on the left hand side) => **left click on Data Analysis ToolPak - VBA** => **Go** (bottom middle of the box) => **check the Data Analysis ToolPak box** => **OK**.

   To check that you have installed the toolpak, click on the **Data** tab. One of the boxes on the toolbar should now be **Data Analysis**. Usually it appears on the far right: **Connections** then **Sort and Filter** then **Data Tools** then **Outline** and then **Data Analysis**.

   2. Use Excel's Data Analysis ToolPak to calculate all of the descriptive statistics plus more (to be reported in a separate worksheet) that we previously calculated using Excel's built-in function: **Data** tab => **Data Analysis** => **Descriptive Statistics** => **OK** =>
      1. Input Range: **A1:D589**. (Note: enter A1 and not A2.)
      2. Select Columns.
      3. Select Labels in First Row (which is why you entered A1 rather than A2).
      4. Select New Worksheet Ply
      5. Select Summary Statistics.
      6. Click **OK**.

   Upon clicking OK, a new worksheet is created called **Sheet**. Rename the worksheet **Descriptive Statistics** and move it to the right of **School Data**. Compare the results from Excel's Descriptive Statistics tool to the results in G2 - J9. They should match perfectly.

28. In the School Data worksheet, type **WHITE** in cell F11. Make the word **WHITE** bold and underlined using the **Font** box in the **Home** tab on the toolbar. In H12 through J12 enter **enroll**, **psae_s**, and **psae_d** respectively (not bolded or underlined). In cell G13 enter **Covariance**. In cell G14 enter **Correlation** (neither bolded). If necessary, widen column G so that both words are completely visible.

29. In cell H13, enter the covariance of **white** and **enroll**: **=COVAR(A2:A589,B2:B589)**. The value of the covariance is -10,927.1.

   The covariance function requires two variables, A2:A589 are the values for **white** and B2:B589 are the values for **enroll**. Because the covariance between **white** and **enroll** is negative, there is a negative relationship between the percent of students who are white and school size (enrollment). That is, white students tend to be at smaller schools.

30. In cell H14, enter the correlation coefficient between **white** and **enroll** using Excel's correl function: **=CORREL(A2:A589,B2:B589)**. If necessary, change the formatting so that 3 decimal points appear on the correlation. The correlation coefficient is -0.418.

   Recall that the correlation coefficient must be between -1 and 1, which it is; and it must be the same sign as the covariance, which it is.

Now we want to fill in the covariance and correlation between **white** and the two test score variables. We cannot do this simply by copying and pasting, however, because Excel will automatically roll the column (and row) identifiers. If we did this, for example, the numbers in cells I13 and I14 would contain the covariance and correlation of **enroll** and **psae_s**, not the covariance and correlation of **white** and **psae_s**.

31. To demonstrate: **right click on cell H13** => **Copy** => **right click on cell I13** => **Paste**.

    The value 1,096.4 appears in cell I13, but this is not the covariance between **white** and **psae_s**. Click on cell I13, and look at the textbox. It shows **=COVAR(B2:B589, C2:C589)**. Although the C column is correctly identifying the **psae_s** column, the problem is that Excel automatically rolled the A column to the B column as well.

32. To fix this, return to cell H13 and enter: **=COVAR($A2:$A589,B2:B589)**. The dollar signs in front of the A's tell Excel to not roll the A column. The fact that there is no dollar sign in front of the B tells Excel to roll the B column. Also go to cell H14 and enter: **=CORREL($A2:$A589,B2:B589)**. Again the dollar signs tell Excel to not roll the A column. Now copy and paste cells H13 & H14 into cells I13 through J14: **left click on cell H13 and drag through cell H14** => **right click in the shaded area** => **Copy** => **left click on cell I13 and drag through cell J14** => **right click in the shaded area** => **Paste**.

    Notice that the formula in the textbox for I13 is **=COVAR($A2:$A589,C2:C589)**. The A column has not rolled, which is what we want as we want the covariance between **white** and **psae_s**. The covariance between **white** and **psae_s** is 361.6, and the correlation coefficient between **white** and **psae_d** is 0.632.

To emphasize the importance of rolling and not rolling formulas in Excel when copying and pasting (and for a bit of practice), we will use Excel's formulas to create a multiplication table.

33. Create a new worksheet called **Multiplication** Table placed to the right of the **Descriptive** Statistics worksheet: **Home** tab => **Cells** box => **Insert** => **Insert Sheet** => **right click on the new worksheet's tab** => **Rename** (rename it **Multiplication Table**) => drag the newly titled **Multiplication Table** worksheet to the right of **Descriptive Statistics** (**left click on the Multiplication Table worksheet tab** => **drag it to the right of the Descriptive Statistics worksheet tab**.

34. In cell A1 of this new worksheet, enter **MULTIPLICATION TABLE (0 through 10)**. Use the **Font** box under the **Home** tab to make the contents of cell A1 bold with Arial font size 14. Again using the **Font** box, make the formatting of rows 2 through 13 to be not bold with Arial font size 9: **left click on row 2 and drag through row 13** => **Home** tab => **Font** box => **select Arial and font size 9**.

35. In cell A3, enter: 0. In cell A4 enter: **=A3+1**. The value in cell A4 will be 1, while the textbox shows **=A3+1**. Copy cell A4, and paste it into cells A5 through A13: **right click on cell A4** => **Copy** => **left click in cell A5 and drag through cell A13** => **right click anywhere in the shaded area** => **Paste**.

36. In cell B2, enter: 0. In cell C2, enter: **=B2+1**. The value in cell C2 will be 1, while the textbox shows **=B2+1**. Copy cell C2, and paste it into cells D2 through L2: **right click on cell C2** => **Copy** => **left click in cell D2 and drag through cell L2** => **right click anywhere in the shaded area** => **Paste**.

37. Change the column width for all 12 columns (A through L) to 6: **left click on column A and drag through column L** => **right click in the shaded area** => **Column Width** => **6** => **OK**.

38. Change the alignment for all 12 of these columns to center-justified: **left click on column A and drag through column L** => **Home** tab => **Alignment** box => **click on the center justified icon**.

39. ***Merging Cells.***

    In order to center the title of the table above all 12 rows, merge the cells in the first row (cell A1 through cell L1): **left click on cell A1 and drag through L1** => **right click in the shaded area** => **Format Cells** => **Alignment** (tab) => **Merge Cells** => **OK**.

40. Increase the height of row 1 to 24: **right click on row 1** => **Row Height** => **24**.

41. Increase the height of rows 2 through 13 to 16: **left click on row 2 and drag through row 13** => **right click in the shaded area** => **Row Height** => **16** => **OK**.

42. ***Centering Cells Vertically.***

    1. Center the title vertically: **click on cell A1** (or anywhere from A1 through L1 as the cells are now merged) => **Home** tab => **Font** box => **Center Vertically** icon (second from left on first row).

    2. Alternatively, once can format vertically by doing the following: **right click on cell A1** => **Format Cells** => **Alignment** (tab) => **Center** (vertical) => **OK**.

43. Our first attempt at filling in the table will be wrong, but it is what springs to mind.

    1. In cell B3, enter: **=A3*B2**. A value of 0 appears in cell B3, as it should be as 0 x 0 = 0.

    2. Copy cell B3, and paste it into cells B3 through L13: **right click on cell B3** => **Copy** => **left click on cell B3 and drag through cell L13** => **right click anywhere in the shaded area** => **Paste**.

    All values are 0, and this is of course wrong. But why? The problem is Excel's automatic rolling.

44. To correctly copy and paste the formula, the original formula needs to be changesd to prevent the automatic rolling.

   1. In cell B3, enter **=$A3*B$2**.

      A value of 0 appears in cell B3. But what is important is the dollar signs. The $A tells Excel to not roll off of the A column for the first number, but it can roll down the A column as there is no dollar sign in front of the 3. Similarly, the $2 tells Excel to not roll off of the 2 for the second number, but it can roll off the B column as there is no dollar sign in front of the B.

   2. Copy and paste B3 into cells B3 through L13: **right click on cell B3** => **Copy** => **left click on cell B3 and drag through cell L13** => **right click anywhere in the shaded area** => **Paste**. You will notice that all values are correct.

45. *Cell Borders.*

   Before filling in the multiplication table, let's make the table look nicer with some borders: **left click on cell A1 and drag through cell L13** => **right click in the shaded area** => **Format Cells** => **Border** (tab) => **Outline** and **Inline**=> **OK**.

46. Save your file, which includes four worksheets – **Original Data**, **School Data**, **Descriptive Data**, and **Multiplication Table** – to a file called YourName_Lab1_Tutorial.xlsx. For example, if Dave Smith were saving the file, he would create a file called DaveSmith_Lab1_Tutorial.xlsx.

Exercises

1. Import the file **Orphans.txt** into Excel. The codebook for this data is at the end of the lab. This is a text file with six variables on 92 children in a school in Uganda during the summer of 2004. Loveena Dookhony, a former Lake Forest College student collected the data for her Senior Thesis while volunteering at the school over the summer. Rename the worksheet **Orphan Data**.

2. Have all cells in the table be Arial 9 point font.

3. Insert a top row, and in that row label the variables **age**, **female**, **weight**, **income**, **orphan**, and **test**.

4. Use freeze panes so that the first row (with variable names) will always be in sight. Center justify (vertically and horizontally) all six columns of data.

5. In cells H4, H5, and H6 enter **Mean**, **Median**, and **Variance**. In cells I3 through N3 enter **Age**, **Female**, **Weight**, **Income**, **Orphan**, and **Test**. Right justify all cells in columns H through N.

6. In cells I4 through I6, enter formulas to calculate the mean, median, and variance of **age**.

7. Copy and paste the formulas in cells I4 through I6 into cells J4 through N6.

8. For cells I4 through N6, have all statistics associated with **age**, **weight**, and **test** report only one decimal place. Have the statistics associated with **Income** report zero decimal places. And have the statistics associated with **female** and **orphan** report three decimal places.

9. In cell I9 enter **<u>Female</u>** (both bold and underline). In cell J11 enter **Covariance** and in J12 enter **Correlation**. In cells K10 through N10 center **Weight**, **Income**, **Orphan**, and **Test** respectively.

10. In cells K11 and K12 enter a formula that calculates the covariance and correlation coefficient between **female** and **weight** respectively. These formulas must be entered in such a way so that they can be copied and pasted into cells L11 through N12 and have the correct values appear there as well. (Do this by copying and pasting.)

11. On the answer sheet (located at the end of the lab, which you will turn in), answer the following questions.
    1. What percent of the observations are female students?
    2. What percent of the observations are orphaned students?
    3. Are test scores positively or negatively related to being female?
    4. Is the correlation between weight and female expected? Explain.

12. Using the data analysis toolpak, generate all of the descriptive statistics (called the summary statistics in Excel) for the six variables in the data set. Have the descriptive statistics appear in their own worksheet. Name this worksheet **Descriptive Statistics**, and place it to the right of **Orphan Data**.

13. Condense the descriptive statistics by moving the variable name over one column and deleting all but the first column of statistics names. The resulting table should look like the following.

	age	female	weight	income	orphan	test
Mean	11.97826087	0.652173913	82.2173913	29236.95652	0.347826087	55.7826087
Standard Error	0.176888662	0.049927761	2.031631497	3529.316751	0.049927761	2.172990726
Median	12	1	78	20000	0	57
Mode	13	1	68	20000	0	27
Standard Deviation	1.696656446	0.47889026	19.48672475	33852.01706	0.47889026	20.84259485
Sample Variance	2.878643096	0.229335882	379.7324415	1145959059	0.229335882	434.4137602
Kurtosis	-0.160622659	-1.613517478	-0.246981717	6.845822136	-1.613517478	-0.984956664
Skewness	0.076094663	-0.649650112	0.624622908	2.274877289	0.649650112	-0.044047398
Range	9	1	92	199700	1	87
Minimum	8	0	42	300	0	13
Maximum	17	1	134	200000	1	100
Sum	1102	60	7564	2689800	32	5132
Count	92	92	92	92	92	92

14. Open a new worksheet, and place it to the right of **Descriptive Statistics**. Name the new worksheet **Revenue Table**. Have all cells be Arial 10 point font. Set the column width of column A at 25. Set the column width of columns B through K at 8.

15. In A2 through A7, enter Week, Units Sold, Price, Weekly Revenue, Total Revenue, and Average Price Per Unit Sold respectively.

16. In cell A1 enter the title (in bold) to the table: **Weekly Revenue Numbers for the First 10 Weeks**.

17. Let the title span cells A1 through K1, and have the title be horizontally and vertically center justified.

18. Make row 1 have a height of 24. Make rows 2 through 7 have a row height of 16.

19. Have column A be left justified and columns B through K be center justified.

20. In row 2, enter weeks 1 through 10 in cells B2 through K2. Do this by entering 1 in cell B2 and then entering a formula in cell C2 that can be copied and pasted into cells D2 through K2. (Do this by copying and pasting.) To check your work, cell K2 should be 10.

21. In cells B3 through K3, enter the following numbers for the Units Sold: 4, 2, 8, 10, 3, 6, 2, 6, 4, 3.

22. In cell B4 enter a price of $1.25. In cells C4 through K4, let price increase by $0.25 each week. Do this by entering a formula in cell C4 that can be copied and pasted into cells D4 through K4. To check your work, cell K4 should be $3.50.

23. Enter a formula in cell B5 that calculates the weekly revenue for week 1. The formula should return a value of $5.00 as 4 units were sold at a price of $1.25 each (and 4 x $1.25 = $5.00). Copy and paste this formula into cells C5 through K5. To check your work, cell K5 should be $10.50 as 3 units were sold in the 10th week for $3.50 each (and 3 x $3.50 = $10.50).

24. In cell B6, enter a formula that calculates total revenue for the first week. (Hint: this isn't any different from cell B5.)

25. In cell C6, enter a formula that can be copied and pasted into cells D6 through K6 that calculates total revenue for all weeks up to the second week (then the third, then the fourth, and so on). Do the copying and pasting. To check your week, cell C6 should be $8.00 (the $5.00 of revenue from week 1 plus the $3.00 of revenue from week 2), and cell K6 should be $110.75.

26. In formulas in cells B7 through K7 that enters the average price of each unit so far sold. To check your work, the average price per unit sold in week 10 is $2.31.

    Extra credit: enter a formula in cell B7 that can be copied and pasted into cells C7 though K7 that enters the average price of each unit so far sold. Warning: the calculation is obvious for week 1 (total revenue of $5.00 and 4 total units sold means the average price per unit is $1.25), but if you enter **=B5/B3** in cell B7 this will not correct for the subsequent cells as this would simply calculate each week's price.

27. Lastly, go back and make all values in cells B4 through K7 have a currency format (meaning with a $) and with 2 decimal places. Your table should look like the following (where any numbers not given to you above have been omitted and font size has been changed to fit the page).

Weekly Revenue Numbers for the First 10 Weeks										
Week	1	2	3	4	5	6	7	8	9	10
Units Sold	4	2	8	10	3	6	2	6	4	3
Price	$1.25									$3.50
Weekly Revenue	$5.00									$10.50
Total Revenue	$5.00									$110.75
Average Price Per Unit Sold	$1.25									$2.31

28. On the answer sheet (at the end of the lab, which you will turn in), answer the following questions.
    5. According to your revenue table, after sales have been completed for the fourth week, what is total revenue for weeks 1 through 4?
    6. According to your revenue table, after the six units in week eight have been sold, what is the average price at which all units have been sold?

29. Save this file as YourName_Lab1_Exercises.xlsx.

Turning in your work

Email both YourName_Lab1_Tutorial.xlsx and YourName_Lab1_Exercises.xlsx to your professor as file attachments to a single email with the subject heading Excel Lab 1: Your Name. Turn in your filled-in answer sheet during class.

Codebook for Highschools.txt.

There are 588 observations in the data set, one for (almost) every public high school in Illinois. The variable names and descriptions follow.

schlname	School Name
schltype	School Type Code (all are high schools)
distname	District Name
disttype	District Type Code (high school or unit)
white	Percent students in school who are white
enroll	School enrollment
fundpp	Total district funding per pupil in $1000s.
eavppd	District's per pupil equalized assessed property value in $1000s
texpppd	Per pupil expenditures on instruction & operations in $1000s
tstenr	Number of students in the school who should take the PSAE
tsttkr	Number of students in the school who took the PSAE
psae_s	Percent of students in school who passed the PSAE standards
psae_d	Percent of students in district who passed the PSAE standards
pnotkr	Percent of students in school who did not take the PSAE

Codebook for Orphans.txt.

There are 92 observations in the data set, one for each student at a particular Ugandan school in the summer of 2004. The variable names and descriptions follow.

age	Age of student
female	=1 if the student is female; =0 if the student is male
weight	Student's weight in kilograms
income	Student's family's income (in Ugandan schillings
orphan	=1 if the student is an orphan; =0 if the student is not
test	Score (out of 100 percent) on a general skills test

Answer Sheet for Lab #1: An Introduction to Excel

Name: _____

1. What percent of the observations are female students?

2. What percent of the observations are orphaned students?

3. Are test scores positively or negatively related to being female?

4. Is the correlation between weight and female expected? Explain.

    _____

    _____

    _____

    _____

5. According to your revenue table, after sales have been completed for the fourth week, what is total revenue for weeks 1 through 4?

6. According to your revenue table, after the six units in week eight have been sold, what is the average price at which all units have been sold?