Computer lab 1: Excel Warmup
MATH 5003/7003 - Swinarski
Wednesday, August 25, 2010

http://www.math.uga.edu/~davids/lab/lab1.pdf

1. Goals

- Learn basic tricks for working with formulas in Excel:
  - dragging formulas
  - protecting an expression with dollar signs
  - entering a range of values using :
- Paste special
- Format a column or row (insert, change width, hide)
- Discuss spreadsheet design
- Learn to create plots (bar graphs, pie) and histograms
- Learn to import data
- Sort data alphabetically or numerically
- Discuss Excel functions:
  - AVERAGE, STDEV, SUM, MAX, MIN, COUNTIF

2. Section 14.2, Question 7

Let's revisit one of your homework questions, #7 from Section 14.2 from Beckmann's textbook. Here was the question:

Table 14.6 shows data about 4th graders' mathematics performance on the 2005 National Assessment of Educational Progress (NAEP).

a. *Use some of the data in Table 14.6 to make a graphical display useful for comparing the percentage of 4th graders performing at or above the proficient level in mathematics in the states in the table.*
b. *Use some of the data in Table 14.6 to make a graphical display useful for comparing the percentage of 4th graders performing at or below the basic level in mathematics in the states in the table.*
c. *Use some of the data in Table 14.6 to make a graphical display useful for seeing how the percentage of 4th graders performing below the basic level in mathematics is related to the percentage of 4th graders performing at the proficient level.*

Table 14.6

Selected States	Below Basic	At Basic	At Proficient	At Advanced
Alabama	34	46	19	2
California	29	43	24	4
Georgia	24	47	26	4
Massachusetts	9	42	41	8
South Carolina	19	46	31	5
Texas	13	47	35	5

3. Bar or column graphs for Question 7

(1) Download and open the .xls file http://www.math.uga.edu/~davids/lab/lab1.xls. I typed Table 14.6 into Excel for you already.
(2) *Page view or normal view?* If you want Excel to break up the screen according to how many cells fit on a page, select Page Layout under the View menu. If you want to cut out the margins while you are working onscreen, go to View, and select Normal.
(3) *Sheets and books.* Each "screen" in Excel is one spreadsheet; each .xls file is a workbook. You can keep more than one sheet inside a workbook. In the workbook lab1.xls, there are

3 sheets. The first one is labeled 2005NAEP (look for the tab at the bottom of the screen). You can change this by right-clicking on the tab and selecting Rename.

(4) *Single quotation marks.* If you want Excel to display exactly what you type in a cell, it's a good idea to type a single quotation mark ' before the letters.

For example, let's say in Cell A1 you want to type 5-10, and in Cell B1 you plan to type the number of children in your class whose age is between 5 and 10. If you just type 5-10, Excel will probably change it to May 10. If you type '5-10, it will display the way you want. (You could also type `5--10` and it will display correctly.)

When I typed the table, I put a single quotation mark ' in front of all the cells that contain text.

(5) *Entering and dragging formulas.* Let's do Part a. of Question 7 from Section 14.2 in Excel.

We want to create a column telling us the percentage of children at or above proficient in each state. That is, in Column G, we want to see the sum of Columns D and E.

Make sure the formula bar is displayed (put a check next to "formula bar" under the View menu if you don't see it).

Click on Cell G2. You could type `= D2+E2` if you want. Alternatively, just type =, and then click on D2 and E2; Excel will interpret this to mean you want to add these in Cell G2.

Now, you could proceed to do this for every row, but that would take a while if there were lots of rows. Here's a shortcut: Click on the lower right hand corner of Cell G2, hold the mouse button down, and start dragging down over Column G. The cursor will change into a diamond shape. Stop when you reach Cell G7. Let go of the mouse button, and Excel should automatically fill in Cells G3 through G7. You can check that Excel entered the correct formulas; for instance, if you select Cell D6, in the formula bar, you'll see `=D6+E6`, just like we wanted.

(6) Now let's make a chart. Go to Insert > Chart. Let's select a Bar Graph.

Excel will insert an empty chart in the spreadsheet. Next, we should link it to the data. Right-click on the empty chart and select Select Data. In the next box, you may have something crazy in the top line, "Chart data range." Clear this out. Click on the upper left of our data (Cell A1), then press and hold the shift key while you click on the lower right of our data (Cell G7). Excel should enter `='2005NAEP'!$A$1:$G$7` in "Chart data range." If you get anything else (my most common error somehow produces plus signs), try again.

Now, Excel is already guessing that our columns might contain sets of data. In the lower left, you see "Series." At first all of these may say `<blank series>`, but if you click here, and your Chart data range is correct, this list should change. Over to the right, we see that the X-values come from A2 through A7, and the Y-values come from B2 through B7. In other words, Excel guessed (correctly) that Cells A1 and B1 were labels for data sets, and the data themselves were in the lower rows.

You may notice there's an ugly gap in the chart. This is because I left Column E blank. Under "Series" we see `<blank series>`. Remove it; then we get a very nice-looking graph!

Now, this master graph has ALL the data... there's almost too much there for us to focus on one thing at a time. For Part a of Question 7, we just want to graph Column G vs. Column A. This is easy to do; just remove all the other series, and we'll be left with the one we want for Part a.

(7) Now, try to do the same thing for Part b.

When I did this, I didn't like the color that resulted. Just right-click on the bars, and select "Format data series." Change Color > Automatic to the color you want.

(8) One way to do Part c. is with a double bar graph. How could you create this in Excel?

(9) In Excel, Bar Graphs have the bars going horizontally. Excel calls a bar graph with the bars going vertically a "Column Graph." If this is what you want instead, just change the type of the graph after you make it.

4. Pie charts for Question 7

(1) A second way to do Part c. is by drawing a pie chart for each state.
(2) Let's insert a pie chart for Alabama. Again, Insert > Chart, and this time select a pie chart. Again, right-click, choose Select Data, and select the range A1 to I7.

This time, though, we want all the data associated to Alabama — that is, we want to make a graph out of the numbers in a row, instead of the numbers in a column. To do this, just hit the "Switch rows/columns" button. Now, under Series, you should see the states' names. Delete all the others except Alabama and we'll have what we want...

...almost. Notice there are some ugly spots coming from my empty columns F and H. I don't know how to get rid of these easily, but here's one way to do it: Delete the chart. Then delete these empty columns and start over. Voila!

5. A scatter plot

Say we want to see a scatter plot of US states population by area. (Remember the graph from the textbook where we saw population of world countries versus area?)

(1) First, let's find the data. One source I found is `http://www.demographia.com/db-statedens.htm`
(2) Next, we want to import the data here into Excel. First, save this webpage to the Desktop. Then, in Excel, select File > Import. Choose the name of the file on the Desktop.

We don't want to save all the formatting of the webpage. Select all the data, and find or create an empty sheet in our original workbook. Select Cell A1. Under Edit, choose Paste Special, then Values. This will paste all the data (text, numbers, etc.) without copying the formatting (colors, spacing, etc.) that we don't want.

It's a good practice to keep track of where we got this data from. Let's insert a row at the top of the sheet (Insert > Rows) and copy the URL into a cell in the top row so we have it recorded.

(3) *Hiding rows and columns.*

Notice that in Column D, the website already computed the density for us. Suppose you didn't want this column to show. You could click on the "D" for Column D, and then go to Format > Column > Hide. If you change your mind later, to bring it back, select columns C and E, and do Format > Column > Unhide.

(4) *Highlighting.* Say we want to highlight the data related to a specific state, like Georgia. Select those cells (or that row, or a column) and click on the bucket icon in the toolbar to highlight.
(5) Now create a scatter plot from this data. Insert > Chart, select X-Y Scatter, then link to the data, ...
(6) MAX, MIN, SUM, AVERAGE. Let's use some of Excel's function to analyze the data we downloaded.

Move to a row below all the data. Notice the last row is totals for the whole US. Let's replicate those calculations ourselves:

In Cell B56, type `SUM(B3:B53)`. The colon : means that we are entering a whole range of cells (B3, B53, and everything in column B between these two cells). This should agree with the website's data for the total area of the U.S.

(7) MAX, MIN, Find, and Sort.

Next, we could ask, "What is the largest state?" In Cell B57, type `MAX(B3:B53)`. You'll see the value 570374. But which state has this area? There are a few ways to find out.

- Go to Edit > Find and search for 570374. Excel will find all the cells on the sheet where this number appears, notably, cell B4 next to Alaska.

- Alternatively, we could sort the whole list in increasing order of area. Select the whole range A3 to D53. Go to Data > Sort >. Where it says "Sort by" use the menu to change State to Square Miles and hit OK. We'll get a list of states from smallest to largest. Alaska is at the bottom. (Or you could ask Excel to give a list from largest to smallest by choosing "descending" instead of "ascending.")

Use MAX, MIN, Find or Sort to answer the following questions:

(a) Which state has the smallest population, and what is its population?
(b) Which state has the highest population density, and what is its value? Which state has the lowest population density, and what is its value?

(8) AVERAGE. Suppose we wanted to know the population density of the U.S. Here are two approaches:

Approach 1. We can find the total population of the US and divide by the total area of the US to get the population density of the US.

The creators of the website have done this already for us, but we could easily recreate their work by SUMming columns B and C and performing the division.

Approach 2. We can find the population density of each state, and then average these numbers.

To carry out this approach: in Column D, the makers of the website computed the population densities for us. In Cell D55, type `AVERAGE(D3:D53)`. Does this match the number we get from Approach 1? What is going on here?

6. A dot plot or histogram

The third sheet in the workbook contains data from a test I have last fall (though I changed the names to protect my students' privacy.)

(1) We see each student's score on each question. First, compute each student's total score (use SUM across row 3, then drag down).
(2) Next, let's compute the average and standard deviation for the test.
(3) Now, let's create a dot plot or histogram showing how many students scored in each range 0-49, 50-59, ..., 80-89, 90-100.

First, let's copy the total scores, and paste special into a new column off to the side. (What happens if you paste instead of paste special?)

We could sort this new list if we wanted to (though this is not necessary for what follows).

Excel can count how many entries in an array satisfy a certain property. For instance, we can count the number of entries in L1:L32 which are less than 50 by using `COUNTIF(L1:L32,''<50'')`.

Try dragging this formula. Notice that the range L1:L32 starts to change, but we don't want that! To *stop* Excel from changing something when you drag, you can put dollar signs around the letter, like so: `COUNTIF($L$1:$L$32,''<50'')`.

We can drag this down five cells or so, and manually change the 50 to 60, 70, etc.

This is *almost* what we want. We had Excel count the number of scores below 50 and the number of scores below 60. To get the number of scores between 50 and 60, we can just subtract.

Thus, we can produce a column of data which counts the number of scores in each range.

(4) Now we can produce a bar graph (or column graph) of this data.
(5) Which question on this test was the easiest? Which question was the most difficult? Which question on this exam had the largest variation in student's scores?