1   Introduction to Using Excel Spreadsheets

This section of the guide is based on the file (a faux grade sheet created for messing with)
- http://cda.morris.umn.edu/∼mcquarrb/SurveyofMath/excel/ExcelIntro.xls

Important: Download the file to the desktop before you begin to work with it. If you open it while it is still "online", some of the Excel functions I refer to later will be disabled. This is just meant to get us comfortable with Excel, we will get more practice in class and on assignments throughout the semester.

Changing Font

You can change the font on an individual cell if you like. Let's change the font for the entire spreadsheet. You can do that by first selecting all the cells in the spreadsheet, which is accomplished by clicking on the corner rectangle, which is the rectangle to the left of A and above 1. Once you have selected all the cells, change the font face and size using the *Formatting Toolbar* (there are other ways of doing this, but using the *Formatting Toolbar* is the easiest).

Changing Formatting Options for Row 1

The first row can be selected by left clicking on the 1. Once you have selected the first row, you can adjust the format of it. Right click anywhere in the selected region, choose FORMAT CELL, and see what can be changed.
You can change the number format, text alignment, font format, patterns (background colour), among other things.

Undo

If you don't like a change you made, you can undo it by selecting EDIT > UNDO from the *Menu Bar* at the top of the document.

Changing Width of Columns & Rows

If you changed the text earlier, you may have a cell that is too small to contain the text. You can adjust the width of columns or rows by right clicking the column and then selecting either ROW HEIGHT or COLUMN WIDTH.

Working with Sheets

Excel files can have more than one Sheet. This file has one Sheet, titled **Grades**. You can change the name of the sheet by right clicking on the *Worksheet Tab* Grades and selecting RENAME. You can also copy and paste sheets at this point. Try it!

Relative Cell Reference: Sum, Copy, Paste

OK, let's do something useful with the data. To sum the potential grades a student can earn, we should enter a formula in cell K2. To enter a formula or data into a cell you first left click on the cell. You can also type the cell reference you wish to work with in the *Name Box*, and the active cell will move to that cell.
The formula we want should tell Excel to sum the entries from cell B2 to J2. This is done by typing

=sum(B2:J2)

This should be evaluated and give 525 in cell K2. Notice the formula appears in the *Formula Bar*, but it is the value of the evaluated formula that appears in the active cell in the spreadsheet.
We want this sum for all the students, but we don't want to type this for each student. Instead, we can copy and paste the formula from cell K2, and Excel will adjust the formula to have the correct form for each row! How cool is that? The answer is very cool, of course. Once you start cutting and pasting formulas in Excel it becomes an extremely useful program. Here's how we do it:

Right click on cell K2, and select COPY.
Use the mouse to select cells K3 to K16 (left click, hold, and drag).
Right click anywhere in the group of cells you just selected, and select PASTE.

You should have the totals for all students now. What has happened here is Excel has used <u>relative cell references</u>, and adjusted each formula to be the correct formula depending on what row the formula is in. If you click on cell K9, you will see the formula in the *Formula Bar* is =SUM(B9:J9).

> Another way to copy is to highlight the cell you want to copy, and hold left click the small black square at the lower right of the cell. You can now drag the cell and it will be copied. This is a useful technique if you want to copy a cell down a column, or across a row.

Absolute Cell Reference

If we want a final percentage grade for each student, we want to divide each student's total grade by 525 (the total possible).

To get started, we should work on the total for Chris. We want 357/525, but we want to enter this in a way that if we change the grade Chris received on an assignment we wouldn't have to adjust the formula here as well. You might think entering in cell L3 something like K3/K2 would work, but you will find it won't. Anytime we enter a formula we need to tell Excel that is going to calculate something, and we do that with an equal sign. So we should enter

   =K3/K2

This will work for Chris, but remember the power of Excel is when we can copy and paste formulas and have them work elsewhere. Copy and paste cell L3 into L4, and see what happens. Yikes! Deb scored 1.241 (124%!), which is clearly wrong.

The problem is that we need an <u>absolute cell reference</u> for the cell K2. When we copy and paste the formula we create in cell L3 to cell L4, we want the K3 value to change to L4 but the K2 to remain as K2.

We can modify our formula in cell L3 so this will happen.

   =K3/\$K\$2

If we copy and past this formula to a different row, Excel will adjust the K3 to the correct cell for that row, but will always divide by the value in cell K2. The dollar signs make this an absolute cell reference.

Formatting Cells to Represent Money (Currency) or Percentages

Cells can be formatted to automatically include a dollar sign if the cell value represents money. Select the cell or cells that you want to represent as currency. From the *Menu Bar*, choose FORMAT > CELLS and then from the *Number* tab select CURRENCY. If the cell represents a percentage, select the cell or cells and then from the *Menu Bar*, choose FORMAT > CELLS and then from the *Number* tab select PERCENTAGE. Change the percentage grades you just calculated to show as a percentage.

Delete, Insert

You can delete or insert rows or columns. Let's delete Geri from the class.

Right click on the 6 beside Geri to select the entire row associated with her. Now a right click in the selected cells will allow you to copy this row, delete this row, or insert a row here. Select DELETE and the row will be removed, the rows below shifting up. If you decide you deleted Geri by mistake, use the *Menu Bar* to select EDIT > UNDO DELETE.

Paste

Paste can be tricky in Excel, since sometimes you want to paste a formula, and sometimes you want to paste the number that formula gives you. By default, Excel will try to paste the formula. If you want to paste the number, you need to choose PASTE SPECIAL.

Select cell L8, which on the spreadsheet is a number but in the *Formula Bar* contains a formula. Try to copy and paste into cell C19 and see what happens.

What happens is Excel tried to paste the formula, and adjust the formula based on the new cell the formula is in. The *Formula Bar* shows the formula in C19 is =B19/\$K\$2.

To get the number instead of a formula, again copy cell L8. Click on cell C19, but this time right click and choose PASTE SPECIAL. Change from ALL to VALUES and you will get the number.

Showing the Formulas

Sometimes it can be useful to have Excel show all the formulas in the spreadsheet, rather than what they evaluate to (this can be useful if you have an error somewhere and don't know where it is). To have Excel show you the formulas, hold down `Ctrl` ` (control, accent) all at once. To switch back to showing the evaluation of the formulas, hold down `Ctrl` ` again. Try it!

Sort

You can sort the data by the value in a particular column if you like. Select the rows 3 through 16 (all the students). Then, from the *Menu Bar* choose DATA > SORT.
If you choose to sort by Column A, the data with be alphabetized by student. If you choose to sort by Column L (Column L might show as a 1 in the list), the data will be sorted by student grade.

Freezing Panes

From the *Menu Bar*, choose WINDOW > FREEZE PANES and see what that does. This is very useful if you have large data sets to manipulate, and you want to always see the row and column labels.

Comment Boxes in Excel

The DRAWING TOOLBAR (usually shown at the bottom of the spreadsheet) is used to add a box to a spreadsheet which you can write comments in. To make sure the DRAWING TOOLBAR is displayed, verify that the box next to DRAWING is checked by choosing from the *Menu Bar* VIEW > TOOLBARS > DRAWING.
From the DRAWING TOOLBAR, choose AUTOSHAPES > BASIC SHAPES > (CHOOSE SHAPE), then left click on the spreadsheet to place a box, which can then be resized and to which you can add text. This provides a convenient way of adding explanation to a spreadsheet.

Dates in Excel

The date January 1, 2005 can be entered in Excel using the formula `=DATE(2005,1,1)`.
This is a great benefit in personal finance problems, because you can then add 1 month to this date using the formula `=DATE(2005,1+1,1)` which will yield February 1, 2005.
You can add 4 months using `=DATE(2005,1+4,1)`, which yields May 1, 2005.
To add 1 day to the date, and get an output of January 2, 2005, you can enter the formula =DATE(2005,1,1+1)-. In a similar fashion you can get any date you want from the initial date.
If you add 12 months to a date, the year will increment by one, so `=DATE(2005,1+12,1)` will yield January 1, 2006.
Here is one way (there are others) that you can conveniently enter dates in Excel, when the dates are changing in the month. This would be useful in a problem where interest was being compounded monthly. This way is convenient because you can copy/paste to get more dates!

	A	B
1	Compounding Period	Date
2	0	=DATE(2005,1,1)
3	1	=DATE(2005,1+A3,1)
4	2	=DATE(2005,1+A4,1)

Once you have the formulas in Excel, you can format how the date is presented. After selecting the cells that represent dates, from the *Menu Bar* choose FORMAT > CELLS and from the *Number* tab then select DATE.

2 Analyzing Data Using Excel

Importing Data Into Excel

Getting data into Excel can be a bit tricky, but it is important if you hope to be able to analyse data effectively using Excel. Here is an outline of how you can do it.

I have some data at http://cda.morris.umn.edu/~mcquarrb/SurveyofMath/excel/data1.txt that you can download into Excel. First try a simple copy/paste and see what goes wrong.

> An easy way to paste delimited data into Excel is the following. Select and copy the data you wish to save. Click on the cell where you want to start pasting the data, then left click and choose PASTE SPECIAL. Choose TEXT. Now, click on the clipboard image Excel has placed beside the newly pasted text, and choose USE TEXT IMPORT WIZARD. Follow the steps to get the data aligned correctly in the cells.

If the above doesn't work, you can always do the following:

- Store the Data in a text file such as data.txt as comma delimited text, or tab delimited text.
  You can use Notepad or some other editor to do this, and can copy and paste the data from a webpage.
- In Excel, choose from the drop down menu DATA > IMPORT EXTERNAL DATA > IMPORT DATA. You will have some choices to make, so make the appropriate choices based on how you stored the data in the text file. You can view how the data will be imported into Excel at this stage.

XY Plots

To do linear regression (and nonlinear regression as well) and determine the correlation r between two variables, we can first create a scatterplot of the data, and then get Excel to figure out the equation of the regression line and the correlation from the scatterplot. Let's treat the data you pasted into Excel as (x, y) pairs, and create a scatterplot for the data using the following process.

Creating a Scatterplot

Here is one way you can create a scatterplot in Excel.

- First, highlight the data you want in your scatterplot. You can remove or add data later, so feel free to highlight more than you need. If your data is not in two columns that are beside each other, don't highlight anything, and move on to the next step.
- From the *Menu Bar* select INSERT > CHART, and you will get a popup window that offers you some choices. Choose XY (SCATTER) to get the scatter plot. You can choose to have only points, points connected by lines, etc. Once the chart is created, you can modify it, so the choice you make here isn't very important. Choose only the points, since we will be putting a regression line on the plot later. Click NEXT.
- Click the tab SERIES. The series represent the X and Y values that go into your plot, so a series will be two columns, one the X values, one the Y values. You can plot multiple series, which means multiple X,Y pairs, on a single plot. You will now have a view of the plot you will create. If you need to modify the choose of data somehow, here are some things you can do:
  - If you did not select any data to plot earlier, your plot will appear empty.

    To select the X values to plot, click on the small image of a spreadsheet to the right of the box beside X VALUES. The window will collapse to a line, and you can now select the X values you want to plot (you can even select them from another sheet in the Excel file!). When you have selected the X values you want to plot, click on the small image of a spreadsheet, and the cell references will be placed into the box. You could type the cell reference by hand, but this process is usually easier.

    To get Y values, you would do the same process as above. Once you have selected the new values you want to plot, click OK and your plot will be updated.

– If you selected more than two columns before you began your chart earlier, you can remove some of the data if you like. Removing a series is like removing a column of data from the plot.

Once you are happy with your data selection, click NEXT.

- Now you can start to format the plot you are creating, by labeling the axes, giving the plot a title, etc. You can change other things, but generally the defaults are fine and you just need to add titles at this step. Once you have added appropriate titles, click NEXT.
- Finally, you have a choice to place the scatterplot in the current sheet, or in its own sheet. Once you've decided where you want the scatterplot, click FINISH and the plot will be created.

You can move the scatterplot around in the sheet once it has been created. If you click on the chart to make it active, you will get a tab in the MENU BAR called CHART. From this drop down menu you can change your chart, and add things to the chart. If you spent time earlier thinking about what you want in your chart, the only thing you will want to add at this time is a regression line and the correlation, but other changes are possible. You can also double click on the chart itself to make changes (sometimes you need to change the format of the data points, which you can do by double clicking on the data points). You can even change the source data that is being plotted if you like. You can also drag a corner of the chart to change its size.

Adding Regression Line and Correlation to Scatterplot

Once you have a plot of the data, click on it to make it active, then from the *Menu bar* choose CHART > ADD TRENDLINE. Choose LINEAR, then click the tab OPTIONS and check the boxes next to DISPLAY EQUATION and DISPLAY R-SQUARED VALUE. Click OK.

You can move the equation to another part of the chart if you like.

You should find that $y = 0.501x + 1.6686$ and $R^2 = 0.4911$. The R^2 is the square of the correlation r, and with the equation of the linear regression line you are able to make predictions of future behaviour of the variables.

Formulas

Here are some of the Excel formulas that can be used for statistical analysis.

I've assumed the data x_1, x_2, x_3, x_4 is stored in cells A1, A2, A3, A4. **Vary the Excel formula for the data you are working with.**

Statistical Quantity	Mathematical Formula	Excel Formula
mean	$\bar{x} = \frac{1}{4} \sum_{i=1}^{4} x_i$	=AVERAGE(A1:A4)
standard deviation	$s = \sqrt{\frac{1}{3} \sum_{i=1}^{4} (x_i - \bar{x})^2}$	=STDEV(A1:A4)
variance	s^2	=VAR(A1:A4)
median		=MEDIAN(A1:A4)
first quartile		=QUARTILE(A1:A4,1)
third quartile		=QUARTILE(A1:A4,3)
maximum		=MAX(A1:A4)
minimum		=MIN(A1:A4)

Once the five number summary quantities have been calculated for the data, a box plot can be drawn. These can be drawn by hand, or using Excel with the processes described below.

Analysis Toolpack

There is an add-in called the Analysis Toolpack for Excel that can really help with statistical analysis, but it is not always installed. Everything we have done so far has not used the Analysis Toolpack. The toolpack gives you access to even more statistical tools than these.

You can see if the version of Excel you are working with has the toolpack installed by going to the *Menu Bar* and choosing TOOLS > ADD INS > ANALYSIS TOOLPACK. This will make available from TOOLS > DATA ANALYSIS additional statistical packages. If it is available, you can use Excel to create descriptive statistics, histograms, random numbers, and lots of other things.

Histograms with Analysis Toolpack

Let's construct histograms of the following two datasets:
http://cda.morris.umn.edu/~mcquarrb/SurveyofMath/excel/data2.txt. We will do this using the tools from the Analysis toolpack. First, Download the data sets into Excel.

We need to decide what our bounding boxes will be for the histogram we will create. Say we want intervals from 0-20, 20-40, 40-60, 60-80, 80-100. We need to make a list of the upper numbers in the bins in a column. In cell D1, type the heading INTERVAL, and then put in D2 the number 20, D3 40, D4 60, D5 80, D6 100.

To generate the histogram, from the *Menu Bar* go to TOOLS > DATA ANALYSIS and select *Histogram*. The Input Range should be the data you want the histogram for. The Bin range is the list of values you just entered in D1:D5. Output Range allows you to choose where the histogram should be placed, and make sure you check Chart Output to get the chart!

If you don't like the gap betweens the bars in your histogram, you can set the gap width to zero. Click on the histogram bars and you will see a rectangle in each bar–double click on any one of them and a dialogue box will appear. Under OPTIONS you can set the gap width to zero.

You can get rid of the More data by clicking on the chart and then dragging the *Bin/Frequency* data in the spreadsheet (notice it is highlighted when you click on the chart) so it does not include the More category. You can also remove the Legend, since it is not very useful most of the time. Just click on the legend and choose CLEAR.

Boxplots

These are a bit tricky to create in Excel, although getting the five-number summary that is used to create them is easy! Suppose we treat the data you downloaded previously (data2.txt) as two separate data sets. Calculate the QUARTILE(,1), MIN, MEDIAN, MAX and QUARTILE(,3) in that order for each data set. Arrange the results on an Excel worksheet in exactly this order:

	A	B	C
1		Data Set 1	Data Set 2
2	Q1	6.98	37.665
3	min	0.000922	8.694
4	median	22.94	66.93
5	max	96.44	99.76
6	Q3	57.7	79.13

Note that you can do this for Data Set 1, and then copy all the cells to get the numbers for Data Set 2.
The following steps will create the boxplots:

1. Select the whole table above, including labels, then from *Menu Bar* choose INSERT > CHART.

2. Select a Line Chart.

3. Click NEXT, choose SERIES IN ROWS, (the default is Columns), then Finish.

4. Select each data series in turn and use FORMAT DATA SERIES to remove the connecting lines in the PATTERNS tab.

5. Select any of the data series and then FORMAT DATA SERIES, select the OPTIONS tab and check High-Low lines and Up-Down bars.

Now you have a pretty good approximation to a box plot for your data! And you've got the boxplots on the same set of axes, so it is easy to compare the two distributions. If you have more than two distributions to compare, you simply add more columns to your five number summary table.

Formatting Charts

In Excel, rarely are the default options for a chart either pleasing to the eye or the best way to convey information. You should always include axes labels, include a title, remove unnecessary legends, and change grid lines and other parts of the chart to make it easy to understand and (frankly) enjoyable to look at.