Math1070: Excel Project

Fall, 2011
Instructor: Kelly MacArthur
(This project was actually created by Greg Rice.)

Introduction: Microsoft Excel is a commercial spreadsheet application written and distributed by Microsoft for Microsoft Windows and Mac OS X operating systems. It features calculation, graphing tools, pivot tables and a macro programming language called Visual Basic for Applications. Its ease of use has made it an industry standard in many areas, including medicine and business, which is why it is imperative that you all have at least a basic understanding of how it works. Like most computer programs/languages, they take a little bit of time and work to understand and to become familiar with. The nice thing about Excel though is that there are an abundance of helpful tutorials and walkthroughs available online. You are encouraged to use those as much as possible.

I'm not attached to using Excel, by the way. Feel free to do this project in Open Office, Google Spreadsheets or any other spreadsheet application. Excel is simply the most widely used and as such, you may want to be familiar with it.

Remark on using Excel: It is common practice in the "real world" to have to learn something on your own. Learning how to use computer software is one of the most typical examples of this. Therefore, I am giving all of you some incentive to learn how to use computer software the way everyone else does, by searching the internet for tutorials.

Due Date: This project is due December 9th, 2011, in class. I will not accept a late project under **ANY** circumstance.

Structure: This project will consist of four separate exercises, each one being worth 30 points. Each exercise will focus on a different statistical feature of Excel. The exercises are outlined in detail below.

Grading: Each exercise will be graded individually. The points will be allotted as follows.

10 points- Accuracy: Are numerical values in the spread sheet accurately computed? Are the formulas used accurate? etc.

10 points- Appearance/Organization : Are all components of the spreadsheet well labeled? Is your spread sheet easy to read? Are the graphs and tables properly labeled.

10 points- Completion: Did you finish all components of the exercise?

How your project should be handed in: Each exercise will involve some sort of calculation or graph done on a spreadsheet. A paper copy of this spreadsheet is to be turned in for every exercise. Also, some exercises ask for explanations, these should be attached on a separate page.

Brief intro to the workings of Excel:

- Short Description: Excel is a form of spreadsheet software. When you open Excel, you will see lots and lots of "cells" (ergo Ex'cel'). Each cell has a name corresponding to which row and column it is in. For example, the cell in the upper left corner is called A1 (since it is in the first row ant A'th column). One of the simplest things you can do is write stuff in the cells. If you click on cell A1 and then write "Hello" and then press enter, Excel will save the word Hello in cell A1. Although this is useful for labeling your spreadsheet (which you are required to do), typically what we want to do is save numbers in the cells, this takes a bit more work. If you just type "3" into a cell, the program will not recognize it as a number. Instead what you have to do is type "=3". The equal sign tells Excel that you want to write a number in the cell. This has to be done in order to do arithmetic in-between the cells.

-Arithmetic Between Cells and Using Canned Commands: Once numbers are saved into the cells (using the = sign) , one can perform arithmetic operations between them. For example, suppose we write "=3" in cell A1 and "=2" in cell B1 (i.e. we save the number 3 in A1 and 2 in B1). Then if in cell C1 we write "=A1+B1 (remember the =since these are numbers), a 5 will show up in this cell. Why is this useful? Now if we go back and write "=5" in A1, C1 will automatically be updated and show a 7 (A1+B1). Therefore, what we can do is set up formulas which relate the cells in the spreadsheet, and if we feel like tweaking those values, we just change the cell we want to tweak and then the calculations are updated automatically, which is pretty nice.

Popular formulas are often times saved as pre canned commands in Excel. For example Excel can compute the average, median, standard deviation, correlation, and many, many more things with very little effort. To compute the average of some numbers (say we have 6 of them). You can write them into the cells A1 thru A6. Then if in B1 you write "=AVERAGE(A1:A6)" (note the syntax here, the first number of the column, a colon, and then the last number in the column), B1 will then display the Average between the numbers A1 thru A6; it's that simple. Here is a list of all of the pre canned commands you will need for this project:

AVERAGE
MEDIAN
MAX
MIN
QUARTILE
NORMDIST
NORMINV
CORREL

To see how these commands work you can search for them in the Excel help bar, or online.

Graphs and Pictures: Part of the project will be to make graphs and pictures which illustrate data. This task is fairly simple using Excel. To make a pie chart or a bar graph, you can store your frequency data in one column, and your labels in another. Then simply highlight both columns and click "Insert" and then "Chart". Follow the direction in the chart wizard to make a pie chart or a bar graph. The Chart wizard also has options about how to label your chart, this should always be done. The scatter plot is a bit more difficult. To do a scatter plot, record the ordered pairs of data in corresponding rows (i.e save the first data point in cells A1 and B1, and then continue to A2, B2 and so on). Once you have recorded all of your data, highlight it and then click the chart button as before. This time select the XY scatter chart and follow the instructions as before. Once the chart is made, you can insert a least squares regression line quite easily. Just click on the scatter plot, and then click on the chart menu and choose add trend line, it's as easy as that. Of course, be sure to label your scatter plot.

The Random Number Generator: Excel has a fairly good random number generator. If you enter "=Rand()" into a cell, it will input a random number uniformly taken between 0 and 1 into the cell. To get a number between 0 and 10, use "=Rand()*10".

The Exercises:

Below are the descriptions of the exercises.

Exercise 1: The mean, standard deviation, and the five number summary.

In your excel spreadsheet, use the Rand function to create 10 random numbers between 0 and 10 in the cells A1 through A10. Then in some other cells compute the following (be sure to label which is which): the mean of A1-A10, the median, the min and max, the range, the standard deviation, Q1, Q3, and the inter-quartile range.

Exercise 2: Pie Charts and Bar Graphs:

You are doing a study where you wait outside of a local bar and survey the college students who come in as to what year they are in. The responses are Freshman, Sophomore, Junior, Senior, and Graduate Student. Suppose you see 200 college students, and that the tallies of the evening amount to:

Freshman	Sophomore	Junior	Senior	Graduate
20	25	47	64	44

Display this data in an array on a spread sheet and then create a Pie-Chart and a Bar Graph, using Excel to illustrate this data.

Exercise 3: The Normal Distribution.

Excel has a fairly nice canned command for computing normal probabilities. The command NORMDIST(x, mean=m, standard deviation=s, TRUE) computes the probability that a normal observation with a fixed mean and standard deviation is less than x. There is also a nice command for computing the inverse operation to the one described above. The command NORMINV(Probability=P, mean=m, standard deviation=s) computes a value z such that the probability that a N(m,s) is less than z is equal to P (this is the sort of backwards calculation which has come up several times now). Use these commands to answer the following problems

1) Compute the probability that an observation from a N(3, 5) population is less than 7.
2) Find a number Z so that P{ N(4, 4) < Z } = .75 (the probability that a normal 4, 4 is less than Z equals .75)
3) Compute the probability that a N(14,3) observations is larger than 12.
4) If the heights of 17-your-old girls are known to be normally distributed with mean 64 inches and standard deviation 3 inches, how tall must Taylor be when she is 17 so that she will be taller than 73% of other girls her age?

Exercise 4: Correlation and Least Squares Regression

1) As a first introduction to correlation, create two sets of 15 random numbers uniformly picked between zero and 10 (Label one column of numbers the X-coordinates, and the other column the Y-coordinates). Use the command CORREL to compute the correlation coefficient between the two columns. Now create a Scatter Plot showing the relationship between the two sets of numbers, then add to it the least squares regression line for the data. Give a brief explanation about what you expect the correlation coefficient to be and why you observe any deviations from what you expect. This should be done on a separate sheet of paper.

2) Now suppose Cletus the Creamer is driving an Ice-Cream Van around the University, and that he keeps track of how much ice cream cones he sells each day as well as the temperature on that day. The data he recorded is as follows.

Temp	# of Cones Sold
70	70
64	30
48	12
88	102
92	115
67	30
103	145
36	3
76	68
86	87
89	90
68	40
82	87
98	104
67	35

Do a complete analysis of this data. Using Excel, compute the mean temperature, the mean number of ice cream cones sold, the standard deviation of the temperature data, the standard deviation of the number of ice cream cones data, and the correlation coefficient between the two data sets. Also create a scatter plot with a least squares regression line for the data. Then use the least squares regression line to predict how much Ice-Cream Cletus should stock in his van if the forecasted temperature is 95 degrees.