Name:	Math 130H Lab 5
	Excel: Multiple Linear Regression

In the previous 2 labs the goal was to be able to predict the value of y if we were given the value of one variable x. Sometimes however, to make a good prediction for y you will need to know the value of many variables $x_1, x_2, ..., x_n$. This is called multiple regression. So if you think y depends on 2 variables, then you model the data with the equation $y = ax_1 + bx_2 + c$ where a, b, and c are numbers. If you think y depends on 3 variables, then you model the data with the equation $y = ax_1 + bx_2 + cx_3 + d$ where a, b, c, and d are numbers, and so on. In the 2 independent variable case, the sample data can be graphed, but it would be graphed in 3 dimensional space. For more than 2 independent variables, the data can't even be graphed, but nevertheless the ideas are the same as for linear regression back in lab 3. That is, we are looking for the best values for a, b, c, and so on that make your equation as close as possible to all of the data points. Then once you have the equation, you can use it to make predictions.

Here's how you have excel find the best values of *a*, *b*, *c*, and so on...

1) Type the data exactly the way it appears below in excel

у	x1	x2
15	3	-1
16.5	5	2
25.9	8	1
38.2	14	4

- 2) Since there are 2 independent variables x_1 and x_2 , the equation that we want to fit the data to
- is $y = ax_1 + bx_2 + c$
- 3) In excel, click on the "Data" tab on the top
- 4) Then click on "Data Analysis" near the top right corner (if "Data Analysis" is not yet installed on your computer's version of excel, go to my website <u>www.gregsriohondomathpage.com</u> and under the link for this lab is a link for how to add it in to excel)
- 5) From the scroll down window, select "Regression"

6) Click on the text box next to "Input Y Range" then select the *y* column of your data (including the label *y*)

7) Click on the text box next to "Input X Range" then select the x_1 and x_2 column of your data (including the x_1 and x_2 labels)

8) Check the labels box, then click ok

9) Once you click ok, you will see lots of numbers in your excel file. The numbers you want are on the bottom left. The number next to x_1 is the coefficient of x_1 , the number next to x_2 is the coefficient of x_2 and the number next to "Intercept" is the constant term. Locate these numbers in your excel file. The equation of the form $y = ax_1 + bx_2 + c$ that best fits this data is $y = 2.7x_1 - 1.3x_2 + 5.6$ and now you can use this equation to make predictions for y

<u>Your Turn (Section 14.3 Hw #28)</u>: **Income** An economist was interested in modeling the relation among annual income, level of education, and work experience. The level of education is the number of years of education beyond eighth grade, so 1 represents completing 1 year of high school, 8 means completing 4 years of college, and so on. Work experience is the number of years employed in the current profession. From a random sample of 12 individuals, he obtained the following data:

Annual Income (\$ thousands)	Work Experience (years)	Level of Education
34.7	21	6
17.9	14	3
22.7	4	8
63.1	16	8
33	12	4
41.4	20	4
20.7	25	1
14.6	8	3
97.3	24	12
72.1	28	9
49.1	4	11
52	15	4

a) Find the least-squares multiple linear regression equation $y = ax_1 + bx_2 + c$ where x_1 is work experience, x_2 is level of education, and y is the response variable, annual income.

b) Predict the income of an individual whose experience is 12 years and level of education is 4.