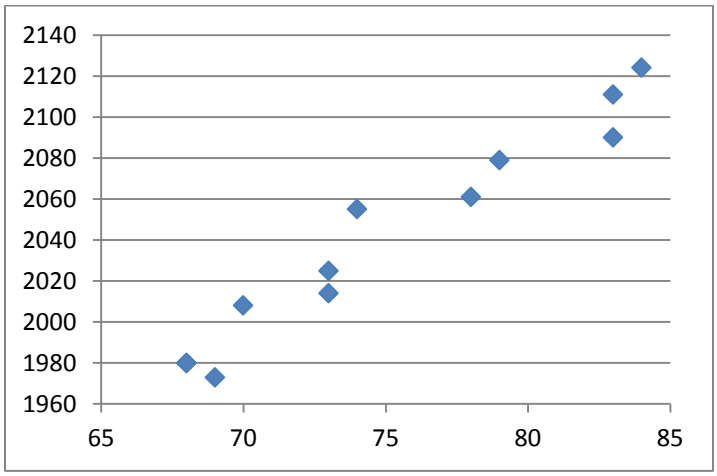
Name: Math 130H Lab 3

Excel: Linear Curve Fitting

One topic in statistics is regression analysis, which is the analysis of paired data. The idea is that you want to see if one quantity (y) depends on another quantity (x) and if so, you can make predictions for y by knowing the value of x.

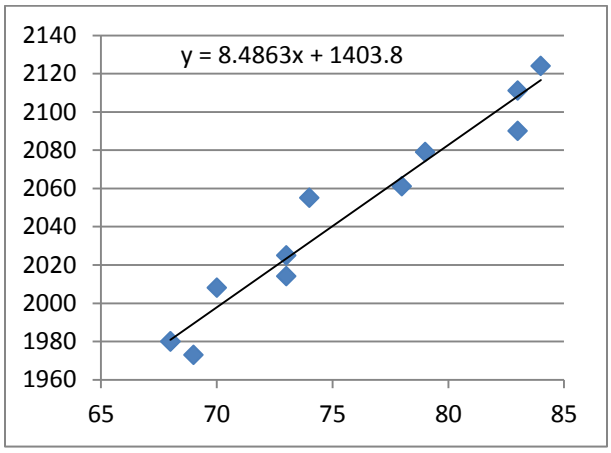
Example: In Physics you learn that when you heat most objects, they expand. Likewise, when you cool most objects, they contract. Let's say that you want to study this effect by trying to investigate how the height of a door in your house depends on the outside temperature. You decide to take data by measuring 2 quantities: (1) The height of the door and (2) The outside temperature. You do this once a month for a year and get the data in the table below. This is paired data, meaning that every time you collected data, you collected 2 quantities. Think of the paired data as a point that you can plot. That means that one of the quantities represents x (as in x-coordinate) and the other quantity represents y (as in y- coordinate). Which one is x and which one is y? Usually the end product of such a problem is to predict the value of one of the letters when you are given the value of the other letter. So in this problem, let's assume that we will want to predict the height of the door when the outside temperature is given, so this means that the height of the door is y (and the outside temperature is x). The data is graphed in below.

Outside Temperature x (F)	68	69	70	73	74	78	83	84	83	79	73	68
Door Height y (mm)	1980	1973	2008	2025	2055	2061	2090	2124	2111	2079	2014	1980



Physics says that if you graph the equation that relates the height of the door and the outside temperature, you should get a straight line. The graph above is not exactly a straight line. Why? The answer is because many different factors can mess up the data. The instruments you used to measure the outside temperature or the height of the door can be inaccurate or old, or you may have made mistakes when making the measurements, or the door can be expanding for other reasons besides just the outside temperature (like if there is a full moon out, or you have termites).

There is no single line that that will pass through all of the points, so we want a line that is "as close as possible" to all of the points. This is called the least squares regression line. The equation of this line can be found using Calculus, or using the computer program Excel, and once we have the equation of the line, we can use it to make predictions. Excel gives the following equation as the best fitting line.



So if you were asked to estimate the height of the door when the outside temperature is 89° F, you plug in 89 for x and solve for y. After doing this, you get $y = 2159$. This means that you predict that the door height will be 2159 mm.

Your Turn

Problem 1:

1) Find a computer that has Excel on it and type the data below (make sure to type the x numbers in the 1st row or 1st column and type the y numbers in the 2nd row or 2nd column)

x	3	5	2	9	12
y	23	24	29	18	14

2) Graph the data on Excel.
   a) Do this by first clicking on the cell with the number 3 on it, hold the mouse down and drag the mouse to the number 14 to highlight all of the numbers
   b) Click the insert menu at the top, then click scatter under that charts area. Select the icon that has no lines drawn in it
   c) Get the regression line equation by first clicking (just once) on the points in your scatter plot, then right click, and select "Add Trendline." Then in the format trendline box, make sure "Linear" is selected, and the "Display equation on chart" box is selected then finally click ok.
3) Print this graph and turn it in with this lab
4) Predict the value of y when $x = 38$

Problem 2: The following data represent the price of a random sample of used Chevy Camaros, by age.

Age x (years)	2	5	2	5	4	5	1	2	6	1	1	4
Price y ($)	15900	10988	16980	9995	11995	10995	20365	16463	10824	19995	18650	10488

a) Use Excel to draw a scatter diagram with the regression line and regression line's equation displayed on the graph.

b) Assuming that there is a linear relationship between the data, even though it might not look like it in your scatter plot, predict the price of a randomly selected Chevy Camaro that is 8 years old.