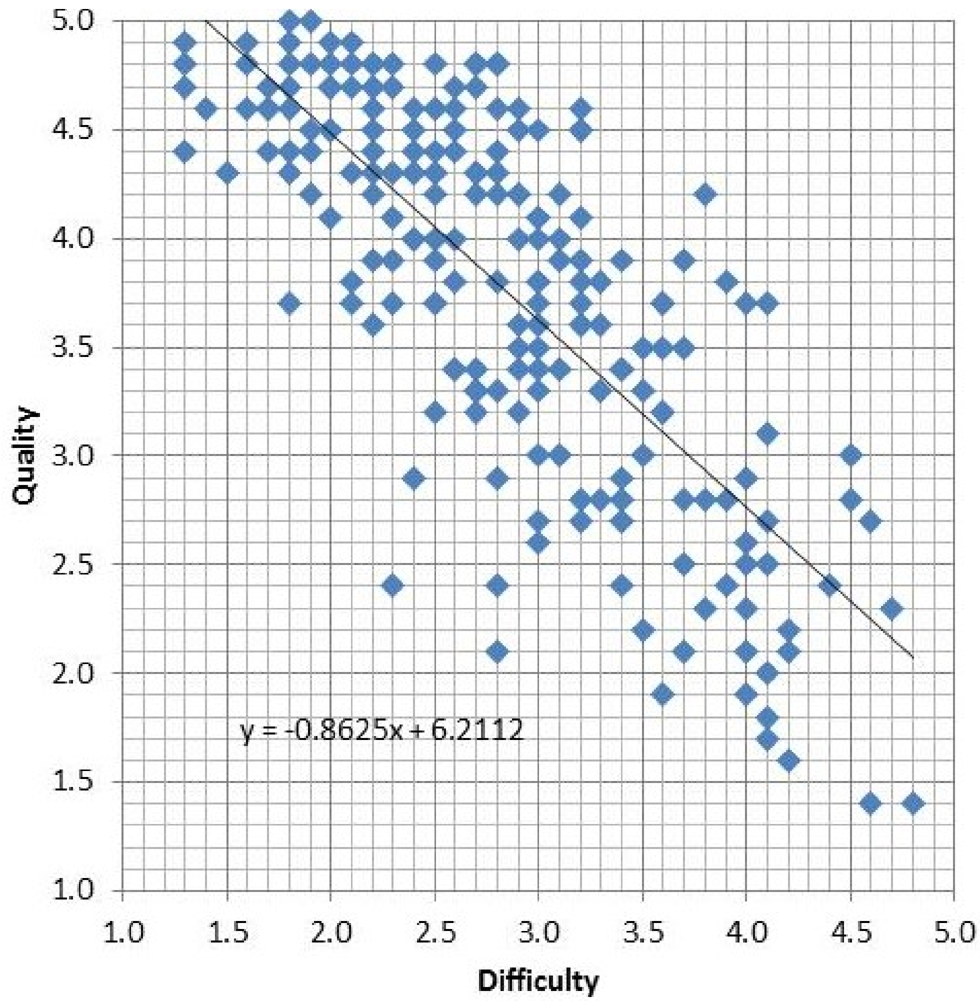
David Youngberg
BSAD 210—Montgomery College

LECTURE 22: SIMPLE LINEAR REGRESSIONS II

I. On Causation and Evaluation.

a. Let's revisit our regression from last class.

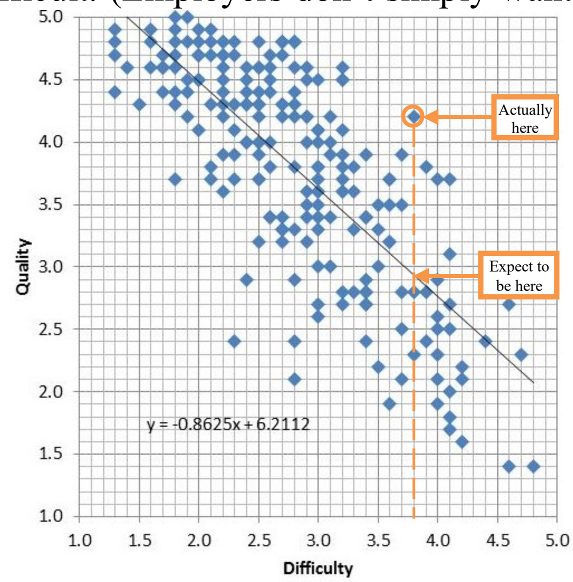
b. Here is the graph with DIFFICULTY causing QUALITY:



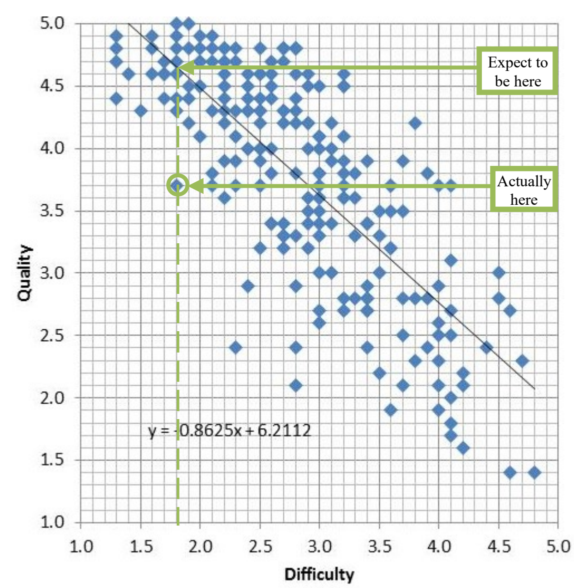
i. I made this with the Add Trendline… option found after you right-click the data on a scatter plot. By opting to Display Equation, it will show you the line's equation.

ii. Note (a) it's in a slightly different format (y=mx+b) and (b) it doesn't give you statistical significance. It's better to use Data Analysis to run the regression. But this option is useful for visual purposes.

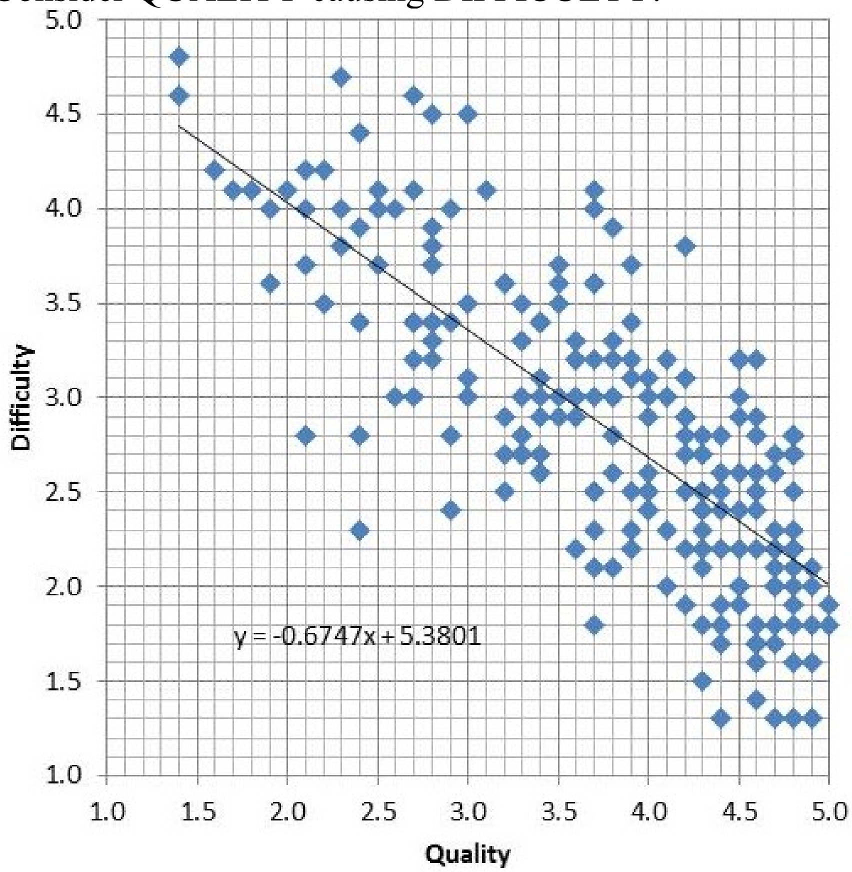
c. Interestingly, we could use this to evaluate professors. A good professor gets high ratings while being difficult. (Employers don't simply want "A" students. They want "A" students who had to work really hard for the grade.)



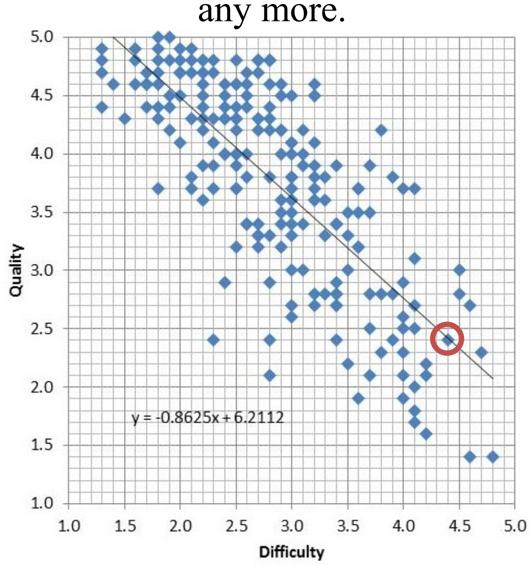
   i. Professors above the line having a higher quality than you'd expect given their difficulty rating.
   ii. Professors below the line have a lower quality than you'd expect given their difficulty rating.
   iii. The professor highlighted with the orange circle has a good quality rating (4.2) but when you consider how hard s/he is (difficulty is 3.8), it's much more impressive. You'd expect the quality rating to be only about 2.9 with a course that difficult. Despite being hard, the students like the professor. That difference is the error term, ε, mentioned earlier.
   iv. The professor highlighted with the green circle seems to be pretty good (quality of 3.7) but with a difficulty of 1.8, you'd expect a rating of about 4.7. The quality rating is quite low for how difficulty the professor is. Again, that difference is the error term, ε, mentioned earlier.

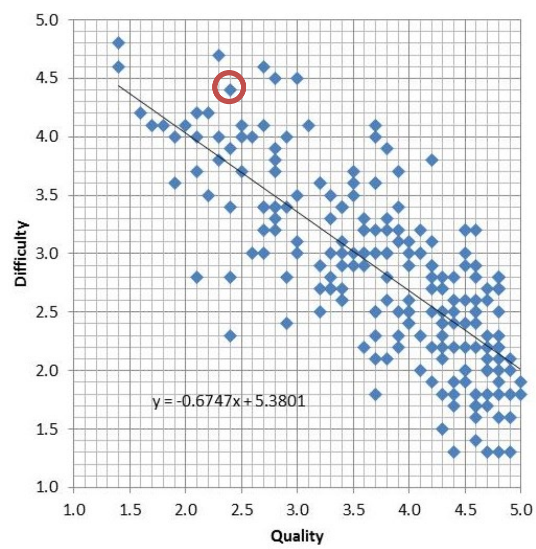


d. Consider QUALITY causing DIFFICULTY:



i. The professor with a QUALITY of 2.4 and DIFFICULTY of 4.4 is right on the predicted line in the first graph. But reversing the causation moves that professor above the line. S/he is not average any more.





e. Why does it change? Linear regressions minimize the summed and squared *vertical* distance. What's set as the y variable and what's set as the x variable determines the line. Swapping the two variables results in a fundamentally different line.