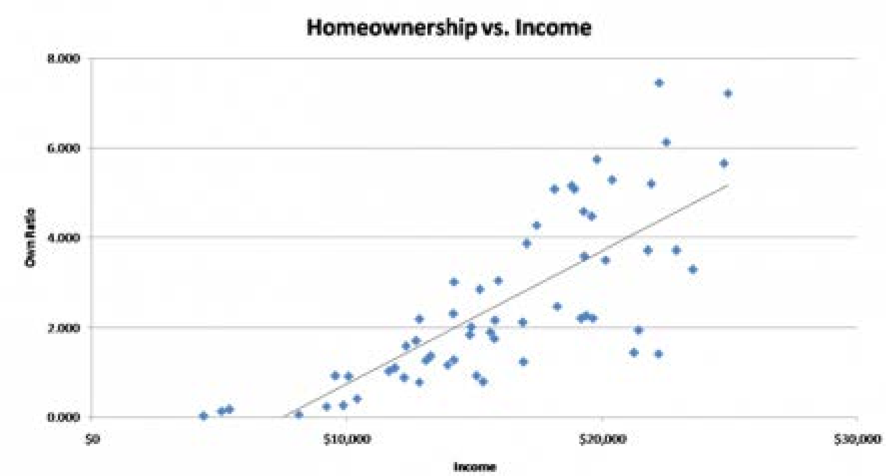
David Youngberg
BSAD 210—Montgomery College

LECTURE 29: SIMPLE LINEAR REGRESSION II

I. Assumptions:
   a. That the regression is linear, as the name implies.
   b. That the residuals follow a normal distribution (from the CLT).
   c. That the residuals don't exhibit a pattern across the independent variable. There should be statistical independence of errors.
      i. In other words, you can't predict the residual of the one observation with the residual of the previous observation.
      ii. <u>Example</u>: If as AGE increases, the residuals decreased.
   d. That there is homoscedasticity; this requires some explanation.

II. Homoscedasticity
   a. *Homoscedasticity* is that the variance (or the deviation) from the regression line is the same, regardless the value of the independent variable(s).
   b. When we lack homoscedasticity we have heteroscedasticity, or the variance is not the same for all values of our independent variable.
      i. Heteroscedasticity can show up in different ways. Here we see how variance increases as income increases. But if variance decreased, or increased and then decreased, or decreased and then increased, etc. we'd still have a problem.



   c. Why should we care? Heteroscedasticity biases our error which means our t-statistic is higher (or lower) than it should be.
      i. In practice, it is not much higher so if you're significant to the 1% level, you're probably fine.

ii. But if your values are barely significant (close to 5%), then you have a problem. If you'd adjust for heteroscedasticity, your significant result might cease being so!

d. The simplest way to detect heteroscedasticity is to make a scatter plot and add a regression line. This visualization test is intuitive (but not precise).

III. Natural log

a. Suppose you don't think the relationship between variables is linear. You can perform a nonlinear transformation on the data itself. This doesn't violate our assumptions; it's actually a solution. Applying a linear regression to a pattern that is, say, U-shaped, is a huge no-no.

i. For example, income should matter a lot for life expectancy when you are poor but its effect should taper off as you become wealthier.

ii. Natural log is a good way to approach this problem and a great standard non-linear transformation.

b. What is natural log? It is a logarithm with a base of e (the number).

i. We use it when the raw data have a strong positive skew.

ii. We sometimes use it to help combat heteroscedasticity.

iii. We use it when we want the independent variable to have a percent change influence on the dependent variable (see below).

iv. Natural logs change how we interpret the regression. Now we think in terms of percent increases rather than raw numbers.

c. If just X is logarithmic…

i. Add one to the percent increase in X (in decimal form) and then multiply the result by β.

$$\Delta Y = \beta \ln(1 + \%\Delta X)$$

ii. Suppose β=0.2 for a logarithmic variable, X. If X increases by 1%, then: (0.2)ln(1.01) = 0.00199; Y increases by 0.00199.

iii. This is typically simplified as a one percentage point change in X means Y changes by β/100 units.

d. If just Y is logarithmic…

i. Take *e* (the number) to the power of β and then subtract one (in decimal form) to find the percent change of Y for every one unit increase in X. If you want X to change by more than 1, multiply that change in X by β first before adjusting with e.

$$\%\Delta Y = e^{\beta(\Delta X)} - 1$$

ii. Suppose β=0.2 and the dependent variable, Y, is logarithmic. If X increased by 1 then: $e^{0.2(1)} - 1 = 1.221 - 1 = 0.221$; Y increases by 22.1% for each 1 unit increase in X.

iii. This is typically simplified as a one-unit change in X results in a (β*100) percentage points change in Y.

e. If both X and Y are logarithmic…

i. Suppose X increases by 1%; how much will Y increase?

ii. Add one to the decimal form of 1%, or 1.01. Take this to the power of β and then subtract 1 from the result. This is the percent change (in decimal form) of Y.

$$\%\Delta Y = (1 + \%\Delta X)^{\beta} - 1$$

iii. Suppose β=0.2 for a logarithmic variable, X. If X increases by 1%, then: $(1.01)^{0.2} - 1 = 1.00199 - 1 = 0.00199$; Y increases by 0.199%.

iv. This is typically simplified as a 1 percentage point change in X results in a β percentage points change in Y.