David Youngberg
BSAD 210—Montgomery College

LECTURE 27: UNDERSTANDING REGRESSIONS I

I. Predicting & Interpreting

a. There are two kinds of prediction you can make with a regression line.

   i. What the dependent variable would be if you put in various values for the explanatory variables.

   ii. How much you can expect the dependent variable to change if you change one explanatory variable.

b. Consider the regression line we did earlier:

$$PRICE = 79 + 237(SQFT) - 23{,}792(MILES)$$

c. Suppose we wanted a house that was in bad shape so we could buy it, renovate it, and sell it for a profit (called "flipping"). Such a house would be going for less than what we would expect given its location and square footage. If we have both variables, we can predict what the value of the house would be if it was in typical shape.

   i. Imagine a house had 2,000 square feet and was 3 miles from the city center. We can expect that house to go for:

$$PRICE = 79 + 237(2{,}000) - 23{,}792(3)$$
$$PRICE = 79 + 474{,}000 - 71{,}376$$
$$PRICE = 402{,}703$$

   ii. So if we see a price for $200,000 or $300,000, we can infer either (a) the seller isn't asking enough for that house or (b) the house is in really terrible shape. Either way, it's a candidate to invest in.

d. Suppose you wanted to know how much cheaper a house could be if it was farther from the city. Since we're changing an explanatory variable (in this case, increasing the MILES variable), we only need to look at the coefficient for MILES to answer the question; the rest of the equation doesn't matter.

   i. Here's the proof. Suppose $MILES_N$ is the miles from city center we are considering now and $MILES_B$ is the miles from city center we were considering. $PRICE_N$ and $PRICE_B$ are the price now and the price before, respectively. Thus we have two equations:

$$PRICE_N = 79 + 237(SQFT) - 23{,}792(MILES_N)$$
$$PRICE_B = 79 + 237(SQFT) - 23{,}792(MILES_B)$$

ii. We are curious how much the price changed. We want to know what this is:

$$PRICE_N - PRICE_B = ?$$

iii. Let's put in the equations from before and do some algebra.

$$[79 + 237(SQFT) - 23{,}792(MILES_N)] - [79 + 237(SQFT) - 23{,}792(MILES_B)]$$
$$79 + 237(SQFT) - 23{,}792(MILES_N) - 79 - 237(SQFT) + 23{,}792(MILES_B)$$
$$79 - 79 + 237(SQFT) - 237(SQFT) - 23{,}792(MILES_N) + 23{,}792(MILES_B)$$
$$-23{,}792(MILES_N) + 23{,}792(MILES_B)$$
$$-23{,}792(MILES_N - MILES_B)$$
$$-23{,}792(\Delta MILES)$$

iv. The Δ symbol is delta and stands for change. Getting one mile farther from the city center drops the price by $23,792.

v. Going an additional 1.5 miles farther out drops the price by $35,688 ($23,792 times 1.5).

vi. Going an additional 2.7 miles farther out drops the price by $64,238.4 ($23,792 times 2.7).

vii. Note all of this keeps the size of the house the same. All other coefficients don't matter because all other variables are held constant.

II. Percentage Points

a. Sometimes the explanatory or dependent variable is a value from zero to one. That could be a dummy variable or simply a percent (such as the percent of a population that's below the poverty line).

b. We shouldn't think of increasing or decreasing in terms of percent. We should think in terms of increasing or decreasing in terms of percentage points.

i. Imagine unemployment is 6%. If it increases to 9%, how much did it increase?

ii. If you said "3%", you'd be wrong because a 3% increase would be 6.18% (3% of 6 is 0.18). This is a 50% increase.

iii. Instead, we can say it increased by three percentage points.

c. Consider the following hypothetical regression (all variables are statistically significant):

$$SALES\ (K) = 50 - 9.4(RIVALS) + 300(\%HOMEOWNERS)$$

i. This regression predicts weekly sales (in thousands) of a home improvement store location based on the number of rivals in a 20 mile radius and the percent of people in that radius who own their own home.
ii. If the percent of people who own their home increases from 35% to 36% (0.35 to 0.36), then it increased by one percentage point, ***not 1%***.
iii. A one percentage point increase in homeownership increases weekly sales by $3 thousand (300 times 0.01 results in 3, or $3,000).

d. Consider the following hypothetical regression (all variables are statistically significant):

$$EMPLOYED? = 0.6 + 0.08(GPA) - 0.12(FEMALE?)$$

i. This regression uses a dummy variable as the dependent variable. Thus, all coefficients involve moving the dependent variable closer to either zero or one.
ii. For any predicted value, you'll get a number between zero and one; this should be interpreted as a percent chance. For example, the predicted value of a male with a 3.0 GPA is 0.84. Such a person has an 84% chance of being employed.
iii. Changing the explanatory variable changes the percentage points. If that same 3.0 student was female, the chance of being employed would fall by 12 percentage points to 72%.
iv. If you increase a person's GPA by 1 point, the chance that person is employed rises by 8 percentage points. It *does not* rise by 8%.

e. Consider the following hypothetical regression (all variables are statistically significant):

$$\%HAPPY = 1.3 - 0.07(PRICE) - 0.05(TIME) + 0.16(CHEESE)$$

i. This regression predicts customer satisfaction for pizza delivery. Lower prices, faster deliveries, and more cheese make for happier customers.

ii. Like the previous example, it's bounded between zero and one (or 0% and 100%). But notice that the constant is greater than one! How is that possible?