Mgmt 469

Interpreting Dummy Variables When You Run Interactions

Funny things can happen when you estimate interactions using dummy variables. If you compare a model that has dummies with a model that has dummies plus interactions, you may notice that the coefficients on the dummy variables can change by a lot. If you work through the algebra, however, you will find that nothing is wrong.

For example, suppose you are examining yogurt sales for two stores, 5 and 14. To make it interesting, let's suppose your LHS variable is lsales = log(sales). You first run a simple dummy variable regression:

lsales = - 3.15 – 3.5(lprice) + .3(dummy14)

This tells us several things.

1) If the price is .10, then for store 5, lsales = -3.15 –3.5(log(.10)) = 4.91. This implies sales = 136 units.

2) If the price is .10, then for store 14, lsales = 5.21, implying sales = 183 units.

3) Thus, the sales at store 14 are higher by 35 percent. This is confirmed by directly examining the coefficient on the store 14 dummy. In particular: exp(0.3) = 1.35

Now suppose you run a store*price interaction. Your new model is:

lsales = -2.0 – 3(lprice) –2(dummy14) –1(dummy14*lprice)

If you examine the coefficient on dummy14, it appears that sales are much lower in store 14 than in store 5. This seems to contradict both the earlier findings and the raw data. But let's use algebra to see what is really going on. Remember, when we use an interaction, the level of sales depends on the price.

1) If the price is .10, then for store 5, lsales = -2 –3(log(.10)) = 4.9, implying sales = 134.

2) If the price is .10, then for store 14, lsales = -2 –3(log(.10)) –2 –1(log(.10)) = 5.2, implying sales = 181.

Thus, the results are identical to those in the non-interacted model, except for a bit of rounding error.