Dummy Variables and Analysis of Covariance

So far we have mostly seen quantitative variables in regression models, but many variables of interest are qualitative (sex, status, etc). To add such information to a model, we can set up a **indicator/dummy variable**.

For example, we could set up a variable X_{i2} representing sex as 0 for type A and 1 for other for the i th individual. The resulting model would look something like:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$$

Where X_{i2} is 0 when the individual is type A. So for type A:

$$E[Y_i] = \beta_0 + \beta_1 X_{i1} + \beta_2 * 0 = \beta_0 + \beta_1 X_{i1}$$

But for any other type:

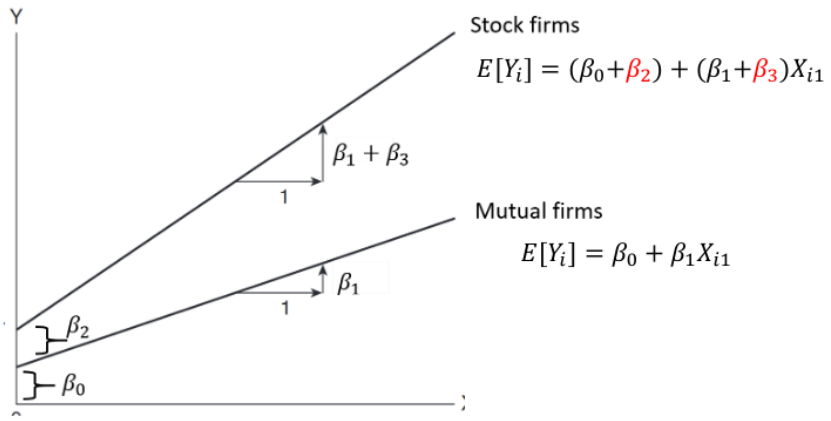
$$E[Y_i] = \beta_0 + \beta_1 X_{i1} + \beta_2 * 1 = (\beta_0 + \beta_2) + \beta_1 X_{i1}$$

Models with Interaction

If we have a first-order regression model with an interaction we can represent it with an interaction term:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2} + \varepsilon_i$$

We can illustrate interaction as follows:



Beta2 indicates how much greater (smaller) the Y intercept for the class coded 1 than that of the class coded 0
Beta3 indicates how much greater (smaller) the slope for the class coded 1 then that of the class coded 0

Raw Mean

A raw mean is simply an average of the observations without considering other covariates. **Least square means** (sometimes called adjusted mean) are adjusted for other covariates, since it is estimated from a linear regression.

Qualitative Variable with 2+ Classes

If there are more than 2 classes to a qualitative variable, we require additional indicator variables in the regression model.

$$\text{☞} X_2 = \begin{cases} 1 & if\ tool\ model\ 1 \\ 0 & Otherwise \end{cases}$$

$$\text{☞} X_3 = \begin{cases} 1 & if tool\ model\ 2 \\ 0 & Otherwise \end{cases}$$

$$\text{☞} X_4 = \begin{cases} 1 & if\ tool\ model\ 3 \\ 0 & Otherwise \end{cases}$$

Resulting in a model like:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \varepsilon_i$$

And the X matrix would look like:

Tool Model	X_0	X_1	X_2	X_3	X_4
M1	1	X_{i1}	1	0	0
M2	1	X_{i1}	0	1	0
M3	1	X_{i1}	0	0	1
M4	1	X_{i1}	0	0	0

Model 4: $E[Y] = \beta_0 + \beta_1 X_1$

Model 1: $E[Y] = \beta_0 + \beta_1 X_1 + \beta_2$

Model 2: $E[Y] = \beta_0 + \beta_1 X_1 + \beta_3$

Model 3: $E[Y] = \beta_0 + \beta_1 X_1 + \beta_4$

Revision #5
Created 29 September 2022 22:02:51 by Elkip
Updated 5 October 2022 17:05:29 by Elkip