



StatNews #40

Interpreting Interactions in Regression September 2000

Adding interaction terms to a regression model can greatly expand understanding of the relationships among the variables in the model and allows more hypotheses to be tested. A previous newsletter, StatNews #39, discussed how to interpret coefficients in regression models. This newsletter will extend those ideas to explain how to interpret the coefficients of interaction terms.

The example from StatNews #39 was a model of the height of a shrub (Height) based on the amount of bacteria in the soil (Bacteria) and whether the shrub is located in partial or full sun (Sun). Height is measured in cm, Bacteria is measured in thousand per ml of soil, and Sun = 0 if the plant is in partial sun and Sun = 1 if the plant is in full sun. The regression equation was estimated as follows:

$$\text{Height} = 42 + 2.3 * \text{Bacteria} + 11 * \text{Sun}$$

It would be useful to add an interaction term to the model if we wanted to test the hypothesis that the relationship between the amount of bacteria in the soil on the height of the shrub was different in full sun than in partial sun. One possibility is that in full sun, plants with more bacteria in the soil tend to be taller, whereas in partial sun, plants with more bacteria in the soil are shorter. Another possibility is that plants with more bacteria in the soil tend to be taller in both full and partial sun, but that the relationship is much more dramatic in full than in partial sun.

The presence of a significant interaction indicates that the effect of one predictor variable on the response variable is different at different values of the other predictor variable. It is tested by adding a term to the model in which the two predictor variables are multiplied. The regression equation will look like this: $\text{Height} = B_0 + B_1 * \text{Bacteria} + B_2 * \text{Sun} + B_3 * \text{Bacteria} * \text{Sun}$

Adding an interaction term to a model drastically changes the interpretation of all of the coefficients. If there were no interaction term, B1 would be interpreted as the unique effect of Bacteria on Height. Since the interaction indicates that the effect of Bacteria on Height is different for different values of Sun, the unique effect of Bacteria on Height is not limited to B1, but also depends on the values of B3 and Sun. The unique effect of Bacteria is represented by everything that is multiplied by Bacteria in the model: $B_1 + B_3 * \text{Sun}$. B1 can now be interpreted as the unique effect of Bacteria on Height only when Sun = 0.

In our example, once we add the interaction term, our model looks like the following:

$$\text{Height} = 35 + 4.2 * \text{Bacteria} + 9 * \text{Sun} + 3.2 * \text{Bacteria} * \text{Sun}$$

Notice that adding the interaction term changed the values of B1 and B2. The effect of Bacteria on Height is now $4.2 + 3.2 * \text{Sun}$. For plants in partial sun, Sun = 0, so the effect of Bacteria is $4.2 + 3.2 * 0 = 4.2$. So for two plants in partial sun, a plant with 1000 more bacteria/ml in the soil would be expected to be 4.2 cm taller than a plant with less bacteria. For plants in full sun, however, the effect of Bacteria is $4.2 + 3.2 * 1 = 7.4$. So for two plants in full sun, a plant with 1000 more bacteria/ml in the soil would be expected to be 7.4 cm taller than a plant with less bacteria.

Because of the interaction, the effect of having more bacteria in the soil is different if a plant is in full or partial sun. Another way of saying this is that the slopes of the regression lines between height and bacteria count are different for the different categories of sun. B3 indicates how different those slopes are.

Interpreting B2 is more difficult. B2 is the effect of Sun when Bacteria = 0. Since Bacteria is a continuous variable, it is unlikely that it equals 0 often, if ever, so B2 can be virtually meaningless by itself. Instead, it is more useful to understand the effect of Sun, but again, this can be difficult. The effect of Sun is $B2 + B3 * \text{Bacteria}$, which is different at every one of the infinite values of Bacteria. For that reason, often the only way to get an intuitive understanding of the effect of Sun is to plug a few values of Bacteria into the equation to see how Height, the response variable, changes. A subsequent newsletter will illustrate this approach.

If you have any questions about how to use or interpret interactions, please contact any of the consultants in the CSCU Office.

Author: Karen Grace-Martin

(This newsletter was distributed by the Cornell Statistical Consulting Unit. Please forward it to any interested colleagues, students, and research staff. Anyone not receiving this newsletter who would like to be added to the mailing list for future newsletters should contact us at cscu@cornell.edu. Information about the Cornell Statistical Consulting Unit and copies of previous newsletters can be obtained at <http://www.cscu.cornell.edu>).