Cornell University
Office of Statistical Consulting
Colleges of Human Ecology and Agriculture and Life Sciences
Division of Nutritional Sciences

StatNews #42: Interpreting the Importance of Interactions

October 27, 2000

It is often challenging to gain an understanding of the substantive meaning of statistical interactions in a regression model. The purpose of this newsletter is to demonstrate a method for doing this. The demonstration will be with an interaction between two continuous variables, but the same method could be used with categorical variables as well. Previously, StatNews #40 discussed a method to assess interaction that examines how the effect of one predictor variable depends upon the value of another predictor variable. The method discussed here involves examining how the response variable is predicted to differ according to various combinations of values of the predictor variables.

As an example (from the M.S. thesis of Holly Gump), we consider the regression of women's body mass index (a measure of relative weight or body fatness) on number of persons in the household, age in years, years of education, and the natural logarithm of income in thousands of dollars. A regression equation with estimates found to fit the data is:
$\text{BMI} = -3.998 + 7.113*\text{persons} + 0.245*\text{age} + 1.624*\text{education} - 0.947*\text{log\_income} - 0.545*\text{persons}*\text{education}$

To help interpret the interaction of persons and education, set the variables not involved in the interaction to some central value. Set age=27, near its mean. Set log_income=0, corresponding to income=1000, near its mean.

Then, the equation with these values set is:

$$\text{BMI} = 2.617 + 7.113*\text{persons} + 1.624*\text{education} - 0.545*\text{persons}*\text{education}$$

Consider two values of persons, 3 and 6, and two values of education, 10 and 16. These sets of values were chosen to represent the width of the distributions of persons and education, but not to be at the extreme range.

We can then put the four combinations of these values into the equation and form a table with the resulting predicted values for BMI:

		Education	
		10	16
Persons	3	23.85	23.78
	6	28.84	18.96

From this table, we can see that, for households with fewer persons (i.e., 3), education makes no difference to BMI and women have normal relative weight. For households with more persons (i.e., 6), lower education is related to higher BMI; women with lower education are overweight and those with higher education are underweight. Also, for those with lower education (i.e., 10), larger households have women with higher BMI. For those with higher education, larger households have women with lower BMI.

Whereas the method described in StatNews #40 focuses directly on the differential effects of the predictor variables, the method described here focuses on seeing how the effects are manifested in the prediction of the response variable. The choice of method for a given study will depend primarily upon which is best at helping to understand and communicate the results.

Author: Edward A. Frongillo, Jr.

This newsletter was distributed to faculty and graduate students in the Division of Nutritional Sciences and the College of Human Ecology, and faculty in the College of Agriculture and Life Sciences, by the Office of Statistical Consulting, Cornell University. Please forward it to any interested colleagues and research staff. Anyone not receiving this newsletter who would like to be added to the mailing list for future newsletters should e-mail statcons@cornell.edu. Information about the Office of Statistical Consulting and copies of previous newsletters can be obtained at World Wide Web address http://www.human.cornell.edu/Admin/StatCons/.