

StatNews #39

Interpreting Regression Coefficients

June 2000

Linear regression is one of the most popular statistical techniques used by researchers. Despite its popularity, interpretation of the regression coefficients of any but the simplest models is sometimes difficult. This newsletter explains how to interpret the coefficients of continuous and categorical variables. A subsequent newsletter will discuss interpreting the coefficients of interaction terms. Although the example used here is a linear regression model with two predictor variables, the same approach can be applied when interpreting coefficients from any regression model, including logistic and proportional hazards models.

A linear regression model with two predictor variables can be expressed with the following equation:

$$Y = B_0 + B_1 * X_1 + B_2 * X_2 + E.$$

The variables in the model are Y, the response variable; X1, the first predictor variable; X2, the second predictor variable; and E, the residual error, which is an unmeasured variable. The parameters in the model are B0, the Y-intercept; B1, the first regression coefficient; and B2, the second regression coefficient.

One example would be a model of the height of a shrub (Y) based on the amount of bacteria in the soil (X1) and whether the plant is located in partial or full sun (X2). Height is measured in cm, bacteria is measured in thousand per ml of soil, and type of sun = 0 if the plant is in partial sun and type of sun = 1 if the plant is in full sun. Let us say it turned out that the regression equation was estimated as follows:

$$Y = 42 + 2.3 * X_1 + 11 * X_2$$

Interpreting the Intercept

B0, the Y-intercept, can be interpreted as the value you would predict for Y if both X1 = 0 and X2 = 0. We would expect an average height of 42 cm for shrubs in partial sun with no bacteria in the soil. However, this is only a meaningful interpretation if it is reasonable that both X1 and X2 can be 0, and if the dataset actually included values for X1 and X2 that were near 0. If neither of these conditions are true, then B0 really has no meaningful interpretation. It just anchors the regression line in the right place. In our case, it is easy to see that X2 sometimes is 0, but if X1, our bacteria level, never comes close to 0, then our intercept has no real interpretation.

Interpreting Coefficients of Continuous Predictor Variables

Since X1 is a continuous variable, B1 represents the difference in the predicted value of Y for each one-unit difference in X1, if X2 remains constant. This means that if X1 differed by one unit, and X2 did not differ, Y will differ by B1 units, on average. In our example, shrubs with a 5000 bacteria count would, on average, be 2.3 cm taller than those with a 4000/ml bacteria count, which likewise would be about 2.3 cm taller than those with 3000/ml bacteria, as long as they were in the same type of sun. Note that since the bacteria count was measured in 1000 per ml of soil, 1000 bacteria represent one unit of X1.

Interpreting Coefficients of Categorical Predictor Variables

Similarly, B_2 is interpreted as the difference in the predicted value in Y for each one-unit difference in X_2 , if X_1 remains constant. However, since X_2 is a categorical variable coded as 0 or 1, a one unit difference represents switching from one category to the other. B_2 is then the average difference in Y between the category for which $X_2 = 0$ (the reference group) and the category for which $X_2 = 1$ (the comparison group). So compared to shrubs that were in partial sun, we would expect shrubs in full sun to be 11 cm taller, on average, at the same level of soil bacteria.

Interpreting Coefficients of Associated Predictor Variables

It is important to keep in mind that each coefficient is influenced by the other variables in a regression model. Because predictor variables are nearly always associated, two or more variables may explain the same variation in Y . Therefore, each coefficient does not explain the total effect on Y of its corresponding variable, as it would if it were the only variable in the model. Rather, each coefficient represents the *additional* effect of adding that variable to the model, if the effects of all other variables in the model are already accounted for. Therefore, each coefficient will change when other variables are added to or deleted from the model.

If you have any questions about how to interpret your regression coefficients, please contact any of the consultants in the Office of Statistical Consulting.

Author: Karen Grace-Martin

[Back to StatNews Table of Contents](#)

(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>).