

Lecture 27

INTERACTIONS IN REGRESSION

In this lecture, we will discuss interactions in regression. An interaction is modeled as the product of two predictor variables. Often, one of the predictors will be an indicator variables. For instance, in our allometric model of bird heart rate based on egg mass, there were 3 categories of birds: altricial, semi-altricial, and semi-precocial and precocial birds lumped together. We might ask if the same relationship holds for each species by examining the interaction of species group and egg mass. The data can be found at:

<http://mypage.iu.edu/~ehouswor/Fall2008/BioZ620/AllometricData2.txt>

Interactions

Suppose you regress Y on some predictors X_1, X_2, \dots, X_k . An interaction between two variables is a new variable which is the product of the two. For instance $X_1 \times X_2$. A full interaction model would contain all products of pairs of predictor variables. Generally, if there is a group of predictor variables that are indicator functions for the same thing, all interactions with them should be included in the model if any one interaction is included.

Example 1 For the allometric data predicting heart rates of fetal birds from their egg mass, there were 3 categories of birds. Precocial and semi-precocial birds were lumped together in one categories. These birds are mobile immediately upon birth, are hatched with their eyes open, and differ not in their ability to move so much as in whether they choose to move soon after birth. Semi-precocial birds choose to stay at the nest and be fed by their parents. Semi-altricial birds are incapable of leaving the nest and are fed by their parents. Some are hatched with their eyes open but most have their eyes closed at birth. They are covered with down at birth. Altricial birds are born with their eyes closed, without much down if any, and are dependent on their parents for food.

We creates indicator variables for the category of bird and use two of the three indicator variables in our analysis. A full interaction model for predicting heart rate on egg mass would then be

$$\begin{aligned}\text{log(Heart Rate)} = & \beta_0 + \beta_1 \text{log(Egg Mass)} + \beta_2 \text{Precocial} + \beta_3 \text{Semi-altricial} \\ & + \beta_4 \text{log(Egg Mass)} \times \text{Precocial} + \beta_5 \text{log(Egg Mass)} \times \text{Semi-altricial}\end{aligned}$$

In this example, altricial birds are the reference group. The coefficient β_1 is the slope for altricial birds. The coefficients β_4 and β_5 measure how much precocial and semi-altricial birds differ in slope from the altricial birds. Similarly, the intercept β_0 is the intercept for altricial birds. The coefficients β_2 and β_3 measure how much the intercept differs from altricial birds for precocial and semi-altricial birds, respectively.

In Minitab, you have to create columns containing the variables $\log(\text{Egg Mass}) \times \text{Precocial}$ and $\log(\text{Egg Mass}) \times \text{Semi-altricial}$ by actually multiplying the appropriate columns together.

Nested F-tests can tell us whether we should include different slopes, different intercepts, or both. ■

We will carry this out with the data in class. For this example, you can get a picture of what is going on by using the Graphs > Scatterplot... menu with multiple regressions option.

Exercises for Lecture 27

1. –

2. –