Lecture 26

Interactions in Regression

In this lecture, we will discuss interactions in regression. An interaction is the product of two predictor variables in a regression model. Often, one of the predictors will be an indicator variable. The data we will use is the allometric data of bird heart rate based on egg mass. In that data set, we combined categories of birds: altricial, semi-altricial, semi-precocial and precocial birds. We will now separate these birds out into three categories using two indicator variables. Altricial and semi-altricial birds are born without the ability to leave the nest. The parents find food for them and bring it back to them at the nest. These will form one category. Semi-precocial birds can leave the nest within two days of birth, but choose to remain at the nest for longer. They will be a second category. Precocial birds leave the nest within 2 days of birth and follow their parents around to obtain food, usually on their own but sometimes with parental assistance. They will be the third category.

Including interaction terms in the model will allow us to determine if the same allometric relationship holds for each type of bird, or if separate relationships hold.

Interactions

Suppose you regress Y on some predictors X_1, X_2, ..., 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 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. Semi-altricial and altricial birds will be considered one category for this analysis.

To test whether the regression is the same for all three categories of birds, we form indicator variables for Precocial and Semi-precocial birds, leaving altricial and semi-altricial birds as the reference category. A full interaction model for predicting heart rate on egg mass would then be

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

In this example, altricial/semi-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(Egg Mass) × Precocial and log(Egg Mass) × Semi-precocial by actually multiplying the appropriate columns together.

The Minitab output is as follows:

```
The regression equation is
Log Heart Rate = 5.92 - 0.121 Log Egg Mass + 0.156 group_precocial
                 + 0.063 group_semi-precocial - 0.0018 precocial*log_egg
                 + 0.0233 semi-precocial*log_egg


Predictor                   Coef  SE Coef       T      P     VIF
Constant                 5.91725  0.03220  183.78  0.000
Log Egg Mass            -0.12066  0.01607   -7.51  0.000   4.649
group_precocial          0.15608  0.08748    1.78  0.086   6.908
group_semi-precocial      0.0630   0.2647    0.24  0.814  24.754
precocial*log_egg       -0.00178  0.02267   -0.08  0.938  12.605
semi-precocial*log_egg   0.02327  0.06014    0.39  0.702  26.360


S = 0.0878642   R-Sq = 86.3%   R-Sq(adj) = 83.8%


Analysis of Variance

Source          DF       SS       MS      F      P
Regression       5  1.31236  0.26247  34.00  0.000
Residual Error  27  0.20844  0.00772
Total           32  1.52080
```

From this output, it does not look like the groups differ. However, the p-values tell only how much a variable contributed *given* that all of the other variables are in the model. The appropriate way to tell whether this full model works better than the model with only one intercept and slope is to do the Nested Models F-Test.

So we run the reduced model and get its ANOVA table:

```
The regression equation is
Log Heart Rate = 5.90 - 0.0940 Log Egg Mass
```

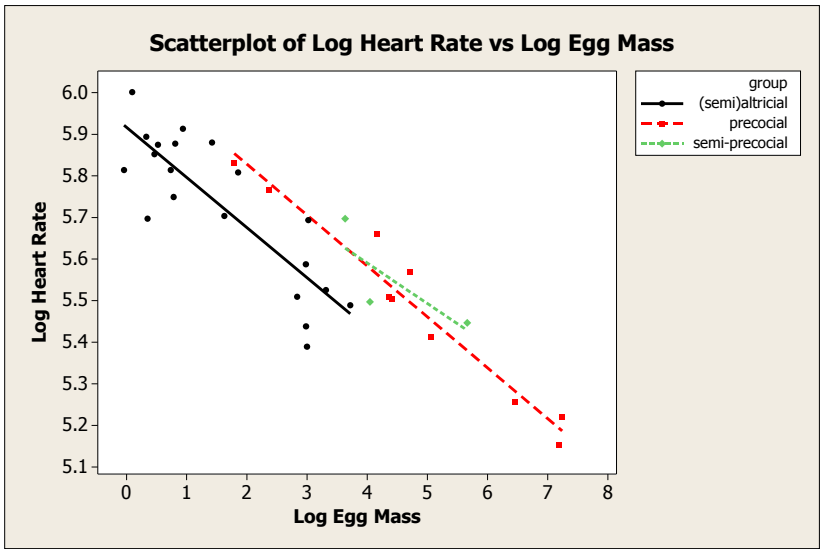
```
Predictor          Coef    SE Coef       T      P    VIF
Constant        5.90194    0.02874  205.38  0.000
Log Egg Mass  -0.093957   0.008249  -11.39  0.000  1.000


S = 0.0972689   R-Sq = 80.7%   R-Sq(adj) = 80.1%


Analysis of Variance

Source           DF      SS      MS       F      P
Regression        1  1.2275  1.2275  129.74  0.000
Residual Error   31  0.2933  0.0095
Total            32  1.5208
```

The F test is

$$F(4, 27) = \frac{(0.2933 - 0.20844)/4}{(0.20844/27)} = 2.748$$

Using the Calc > Probability Distributions > F... menu, we can calculate the p-value as one minus the cumulative probability. The p-value is 0.0488. With an absolute 5% cut-off, this is significant meaning that you would not consider the reduced model as good at explaining the heart rate as the full model is at the 5% significance level. ■

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. That is because only one predictor variable is continuous and the others are dichotomous.



It looks like the separate regressions are parallel (equal slopes) but that the precocial and semi-precocial birds are shifted (different intercept) from the altricial/semi-altricial birds. We will look at this model in class.

References and Readings

[1] Hiroshi Tazawa, James T. Pearson, Takashi Komoro, and Amos Ar. Allometric relationships between embryonic heart rate and fresh egg mass in birds. *The Journal of Experimental Biology*, 204:165–174, 2001.

Exercises for Lecture 26

1. –

2. –