Chapter 9

# Try It Yourself: Exploring R-Squared in Multiple Linear Regression

## Excel Instructions

Fit Regression Model:

1. Open the penguins.csv dataset
2. Create indicator variables for sex and species first
   1. In cell I1, type “sex\_male”
   2. In cell I2, type:

=IF(G2="male",1,0)

* 1. Double-click the marker at the bottom right corner of the cell to apply to all cells.
  2. In cell J1, type “species\_adelie”
  3. In cell J2, type:

=IF(A2="Adelie",1,0)

* 1. Double-click the marker at the bottom right corner of the cell to apply to all cells.
  2. In cell K1, type “species\_chinstrap”
  3. In cell K2, type:

=IF(A2="Chinstrap",1,0)

* 1. Double-click the marker at the bottom right corner of the cell to apply to all cells.

1. Click “Data” tab
2. Click “Data Analysis”
3. Click “Regression”
4. Click “OK”
5. Input Y Range → select all data in the bill\_depth\_mm column D1:D334
6. Input X Range → select all data in the columns for the variables you want to include in your model(e.g., including body\_mass\_g would mean selecting F1:F334)
   1. Be sure that your columns do not contain any missing values!
   2. Be sure not to include bill\_depth\_mm in your X Range!
   3. If your model includes sex, then include the sex\_male variable
   4. If your model includes species, then include both the species\_adelie and the species\_chinstrap variables
   5. Be sure that all variables you want to include are in columns next to each other. This may require copying existing columns to new empty columns in column L or farther to the right.
7. Check the box for “Labels”
8. Click “OK”
9. Record R-squared value (cell B5) in a new column (P) in the original worksheet, with the name “R-squared” at the top
10. Record the number of variables in your model in a new column (O), with the name “Number of Variables” at the top
11. Take note of the slope coefficient (cell B18) for the model based on body\_mass\_g only.
12. Repeat steps 3-12 for each model.

Create Summary Plot of Models:

1. Select the data.
   1. Click on the column headers for both R-squared (column P) and number of variables (Column O) to select all the data in those columns.
2. Insert a scatter plot.
   1. With Columns O and P highlighted, go to the Insert tab on the Excel ribbon.
   2. Click on the Scatter (X,Y) icon in the Charts group.
   3. Choose the first scatter chart option, a simple scatter chart with no lines.