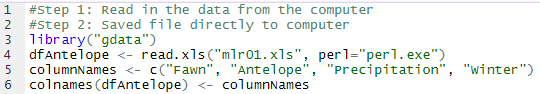
Jeffrey Chao

IST 687 – Homework 8

Date Due: 03/10/2020

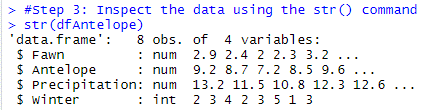
Date Submitted: 03/07/2020

**Step 1+2: Read the data from a file saved directly to the computer**



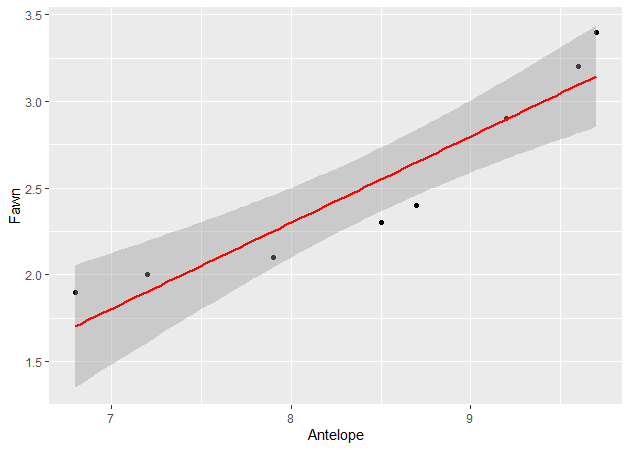
**Step 3: Inspect the data to verify all cases and variables have been read**



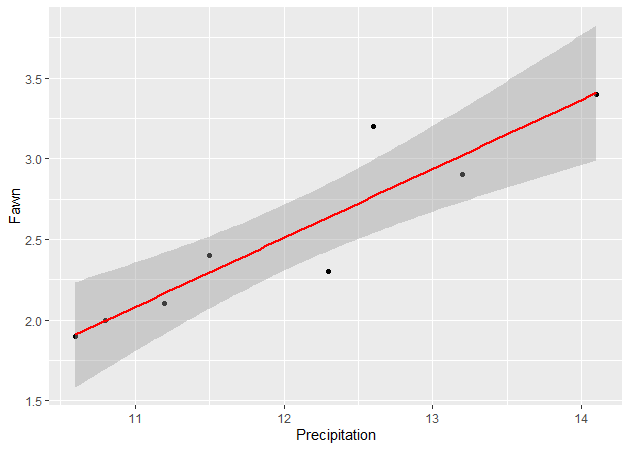


**Step 4: Create bivariate plots**

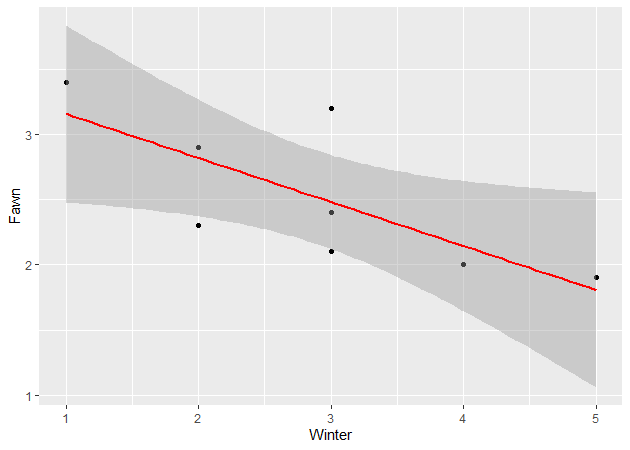








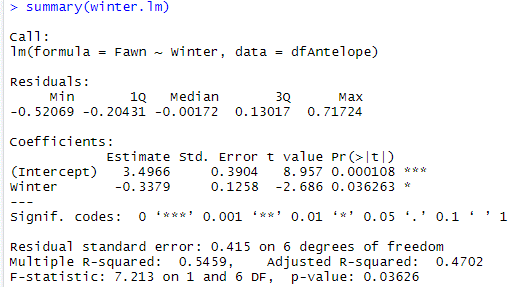




**Step 5: Create regression models of increasing complexity**

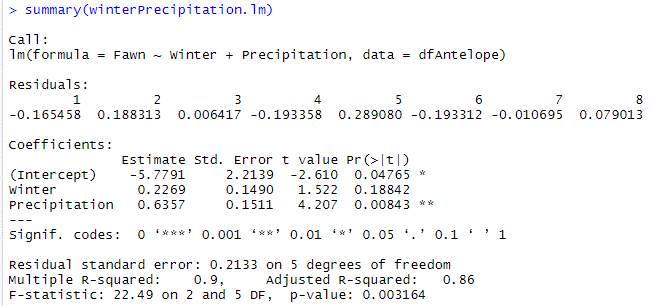
***First model: predict the number of fawns from the severity of the winter***





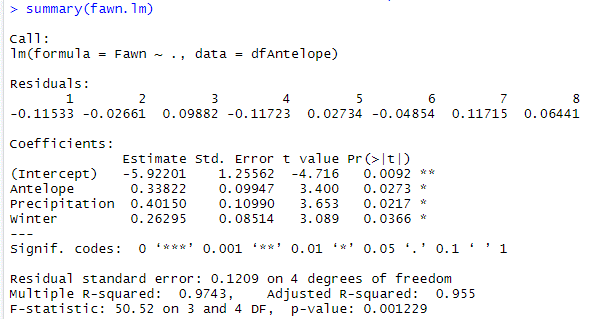
***Second model: predict the number of fawns from two variables (one should be winter)***





***Third model: predict number of fawns from the three other variables***





***Which model works best?***

The third model that contains all three independent variables works best at predicting the number of fawns, with an adjusted R-squared value of **0.955**

***Which of the predictors are statistically significant in each model?***

Using the standard value of less than 0.05 p-value to indicate statistical significance…  
First model: Winter is statistically significant with a p-value of 0.036  
Second model: Precipitation is statistically significant with a p-value of 0.008  
 Winter is not statistically significant with a p-value of 0.188  
Third model: All variables have a statistical significance (p-value < 0.05)

***If you wanted to create the most parsimonious model, what would it contain?***

In my opinion, the most parsimonious model would only use the Antelope variable. This would lead to a linear model with an adjusted R-squared value of 0.862, the highest value among the three independent variables.

Even adding Precipitation (the next highest adjusted R-squared value) to the model would increase the adjusted R-squared value only very slightly to 0.878 while doubling the amount of “observations” that need to be recorded.

