# Homework (HW08) Linear Modeling

### General Instructions

For this homework you will upload 1 R file into blackboard.

Reminder:

* All HW must start with an Identification Block like this sample…

################################################

# IST 387/687, Standard Homework Heading

#

# Student name:

# Homework number:

# Date due:

#

# Attribution statement: (choose the statements that are true)

# 1. I did this work by myself, with help from the book and the professor

# 2. I did this work with help from the book and the professor and these Internet sources: <provide the urls>

# 3. I did this work with coaching from <Name of another student> but did not cut and paste any code

# Run these three functions to get a clean test of homework code

dev.off() # Clear the graph window

cat('\014') # Clear the console

rm(list=ls()) # Clear all user objects from the environment!!!

# Set working directory

# Change to the folder containing your homework data files

setwd("~/MyDesktop/ISTX87/Homework")

* This homework builds on our efforts from the Prep Exercise and depends on a careful read of Chapter 13 of *An Introduction to Data Science*. As usual we’ll use the dataframe we created in our Prep Ex. Reminder: our dataframe contains data about median incomes in zipcodes. We also used the merge function to add state names. In this homework we continue to use ggplot, together with a companion package called ggmap. This companion package enhances the capabilities of ggplot by adding the capability to draw geographic outlines (polygons), shading, labeling, and other map markings.

This homework builds on our efforts from the Prep Exercise and depends on a careful read of Chapter 16 of *An Introduction to Data Science*. As usual we’ll use the Prep Ex to clean up our data. We’ll revisit the spring fawn count dataframe that you created in the PE. This homework uses one of the most common models used by data scientists, linear modeling. Using linear modeling data scientists are able to measure the effects of a variable x against a variable y, with x referring to the independent variable and y the dependent.

### HW08

**Step 1: Visualizing a Linear Model.**

1. Using ggplot, create a bivariate plot of the number of baby fawns versus adult antelope population. Make sure the Y-axis and X-axis are labeled. Keeping in mind that the number of fawns is the outcome (dependent) variable, which axis should it occupy in your plots? Comment your observations.
2. Using similar code, create another bivariate plot of the number of baby fawns versus precipitation that year and comment your observations.
3. Create a final bivariate plot that examines the number of baby fawns versus severity of the winter and comment your observations.

**Step 2: Creating a regression model.**

1. Using the above independent variables (adult antelope pop., precipitation that year, and severity of winter) create a **SINGLE** regression model that predicts the number of fawns.

(**HINT**: If you are stuck, refer to page 202 in the textbook for syntax and explanations of lm())

1. In a comment, record the R-Squared value of the model and explain its significance to the model.
2. In a comment, note the predictor that is most statistically significant. How were you able to decide which predictors were significant and which were not?

**Step 3: Interpreting the model.**

1. Write a block comment that explains in a paragraph your overall interpretation of the model.
2. In a comment, write out the full multiple regression equation and then explain how to interpret the equation

***You must submit all Homework to blackboard prior to the deadline specified for each assignment.***

Late HW assignments will not be accepted for credit.

*Materials provided by the instructors of this course are copyrighted 2018 by Jeffrey Saltz and Jeffrey Stanton, or by the respective instructor that produced them. These materials are provided to enrolled students for their exclusive use during the semester of instruction. Uploading instructor-produced materials to Internet sites (e.g., Coursehero) without permission is considered a breach of professional ethics and will be reported as a possible academic integrity violation.*