STOR 455 Homework #1

25 points - Due Friday 9/3 at 5:00pm

Directions: This first assignment is meant to be a brief introduction to working with R in RStudio. You may (and should) collaborate with other students. If you do so, you must identify them on the work that you turn in. You should complete the assignment in an R Notebook, including all calculations, plots, and explanations. Make use of the white space outside of the R chunks for your explanations rather than using comments inside of the chunks. For your submission, you should knit the notebook to PDF and submit the file to Gradescope.

Eastern Box Turtles: The Box Turtle Connection is a long-term study anticipating at least 100 years of data collection on box turtles. Their purpose is to learn more about the status and trends in box turtle populations, identify threats, and develop strategies for long-term conservation of the species. Eastern Box Turtle populations are in decline in North Carolina and while they are recognized as a threatened species by the International Union for Conservation of Nature, the turtles have no protection in North Carolina. There are currently more than 30 active research study sites across the state of North Carolina. Turtles are weighed, measured, photographed, and permanently marked. These data, along with voucher photos (photos that document sightings), are then entered into centralized database managed by the NC Wildlife Resources Commission. The *Turtles* dataset (found at the link below) contains data collected at The Piedmont Wildlife Center in Durham.

https://raw.githubusercontent.com/JA-McLean/STOR455/master/data/Turtles.csv

1) The *Annuli* rings on a turtle represent growth on the scutes of the carapace and plastron. In the past, it was thought that annuli corresponded to age, but recent findings suggest that this is not the case. However, the annuli are still counted since it may yield important life history information. Construct a least squares regression line that predicts turtles' *Annuli* by their *Mass*.

2) Produce a scatterplot of this relationship (and include the least squares line on the plot).

3) The turtle in the tenth row of the *Turtles* dataset has a mass of 325 grams. What does your model predict for this turtle's number of *Annuli*? What is the residual for this case?

4) Which turtle (by row number in the dataset) has the largest positive residual? What is the value of that residual?

5) Which turtle (by row number in the dataset) has the most negative residual? What is the value of that residual?

6) Comment how the conditions for a simple linear model are met this model. Include at least two plots (in addition to the plot in question 2) - with commentary on what each plot tells you specifically about the appropriateness of conditions.

7) Experiment with at least two transformations to determine if models constructed with these transformations appear to do a better job of satisfying the simple linear model conditions. Include the summary outputs for fitting these model and scatterplots of the transformed variable(s) with the least square lines.

8) For your model using the best transformation from question 7, plot the raw data (not transformed) with the model (likely a curve) on the same axes.

9) Again, the turtle in the tenth row of the *Turtles* dataset has a mass of 325 grams. For your model using the best transformation from question 7, what does this model predict for this turtle's number of *Annuli*? In terms of *Annuli*, how different is this prediction from the observed value?

10) For your model using the best transformation from question 7, could the relationship between *Mass* and *Annuli* be different depending on the *Sex* of the turtle? Construct two new dataframes, one with only male turtles, and one with only female turtles. Using your best transformation from question 7, construct two new models to predict *Annuli* with *Mass* for male and female turtles separately. Plot the raw data for *Anulli* and *Mass* as well as each of these new models on the same plot. You should use different colors for each model (which are likely curves). What does this plot tell you about the relationship between *Mass* and *Annuli* depending on the *Sex* of the turtle?