STAT 387 (Introduction to Statistical Learning, Winter 2023)

Project 2

Instructions:

- Due date: **March 21**
- Submit a typed report.
- Do a good job.

Preamble:

Consider the snail data available in the `Snail2.csv`. Take `Length` as the response variable and we would like to understand how it is related with other variables in the data.

Problem statements:

(a) Perform an exploratory analysis of data.

(b) Is `Length` appropriate as a response variable or a transformation is necessary? In case a transformation of response is necessary, try the natural log transformation or some other simple transformation and use it for the rest of this problem.

(c) Do part (a) of Exercise 15 in Chapter 3 for these data.

(d) Do part (b) of Exercise 15 in Chapter 3 for these data.

(e) Build a reasonably "good" multiple regression model for these data. Be sure to explore interactions of `ShellType` with other predictors. Carefully justify all the choices you make in building the model and verify the model assumptions.

(f) Write the final model in equation form, being careful to handle qualitative predictors and interactions (if any) properly.

(g) Use the final model to predict the `Length` of a `Type1` snail with other predictors set equal to their sample means. Also provide a 95% prediction interval for the response and a 95% confidence interval for the mean response. Repeat for a `Type2` snail, and compare the answers.