STATS 3001 / STATS 4104 / STATS 7054
Statistical Modelling III
Workshop 6 - clams II

John Maclean

In this workshop let's use previous methods to solve problems introduced in the first workshop.

Load the data in `Clams.txt`. Recall that our best model from week 1 tried to predict log(weight) using month as a factor and log(length) as a numerical predictor (with an interaction term).

1. Perform a quick EDA: look at relationship between log(length) and log(weight) and make boxplots for the distribution of log(weight) in each month.
2. Fit the linear model log(weight) ~ log(length) * month
3. Consider which of GAM (fit nonlinear response) or GLS (model heteroscedasticity) is more likely to be useful.
4. Fit your chosen model, selecting the input parameters appropriately.
5. Are the assumptions of your model satisfied?
6. Interpret your model output. Choose the preferred model out of all considered, and explain why.