STATS 3001 / STATS 4104 / STATS 7054
Statistical Modelling III
Workshop 1 - beyond linear models

John Maclean

Week 1

The goal of these workshops is to get to mixed effect models in twelve weeks. These are relatively simple extensions of basic linear modelling that will massively increase the power and flexibility of your analyses.

This workshop takes the first step. Let's look at a dataset and work out what linear models *can't* do.

Data

Load the data in `Clams.txt`. Hint: if using tidyverse, then `read_delim` will work.

Note: the data holds measurements on clams found on a particular beach in one year. Variables include the month when the clams were measured, the length of the clam, and the weight (AFD). There are two other variables - see if you can guess what they are.

EDA

Have a look at the data. We want to predict weight. Look at the relationship of weight with month and with length.

Modelling

You are told to fit a linear model with clam weight as response variable and both month and length as predictors.

Go ahead and follow the standard modelling approach (e.g. from your prac). Pay careful attention to assumptions and use the EDA above to help.

We will consider and discuss:

- issues including month as a predictor
- whether length is a problematic predictor
- which assumptions of linear models are violated (and whether that is a problem)
- how to detect the violation of assumptions (possible discussion: how to detect in big datasets)
- Key discussion: *what simple thing is missing from the linear model, that would give us what we need?*