glm (multiple regression) exercise - badgers

NES8010 Data Analaysis and Modelling

February 2021

# Introduction

This exercise is to allow you to develop skills in data exploration and model building. Please use the BirdNest example and the protocol below to help you get started with model building

## Badger Characteristics

# Dataset

The dataset are of badger characteristics, measured at postmotem. The badgers were collected from across the country over a number of years. We are interested to determine if we can describe the weight of badgers in relation to the other covariates provided.

The variables include Area and Sett (Codes to determine where the animal came from), Year (Year of Post mortem), Age (adult/juvenile) , Weight (g), Length (cm) Sex (M/F) of the individual a variable called tooth to be used as a proxy for Age

You should consider the relationship between weight and length and determine if there are differences between sex and between animals sampled from different locations and at different times (perhaps use adult badgers only)

Your task is to determine the most parsimonious model and describe the diagnostics that you have used.

Remember to use a new script and to save all of your code including a line to set the working directory to where the data an dcode are saved.

**Please keep your code as this will be useful for the next workshop**.

wants <- c("ggfortify", "here", "nlme", "lattice","ggplot2")  
has <- wants %in% rownames(installed.packages())  
if(any(!has)) install.packages(wants[!has])  
  
  
library(here)

## here() starts at C:/Users/nacm4/OneDrive - Newcastle University/Documents/MSc/NES8010 Quantitative Skills/NES8010 Quantitative Skills

badger<- read.csv(here("Data", "bad4.csv"))  
  
library(nlme)  
library(ggplot2)  
library(lattice)

# Modelling protcol

1. DEFINE YOUR PERCEIVED RELATIONSHIP (A HYPOTHESIS) e.g. badger weight is a function of length
2. PLOT THE RESPONSE VARIABLE AGAINST THE PREDICTOR e.g. plot(badger$Length,badger$Weight)
3. WRITE THE MODEL INTO R test.glm<- glm(formula=response~predictor, family=errorfamily, dataframe name) e.g. test.glm <- glm(Weight~Length, gaussian, badger)
4. SUMMARY MODEL OUTPUT eg summary(test.glm) assess estimatse and coefficients (are they significant ?) assess amount of deviance (variation explained) (NULL-RESIDUAL)/NULL
5. ASSESS ERROR MODEL SUITABILITY e.g. hist(residuals(test.glm)) ~ bell-shaped curved centred on zero formally: qqnorm(residuals(test.glm) ~ should be straight line