STATS 201 Data Analysis
Assignment 4, NEFU, 2021

Instructions concerning this assignment:

We are providing you an R Markdown document called **`STATS201_2021_NEFU_A4.Rmd`** which will have some answers already filled in. You will need to fill in and complete the rest of the document. The data files you will be using for the assignment are described in the questions and are available online. Make sure you put these datasets in the same place you put the R markdown document because it is going to look for them there. The first change you need to make to the markdown document is put your name and ID number at the top.

Question 1. [17 Marks]

A marine scientist was interested in modelling the population of geoducks from the west coast of the North Island taking into account their age and an extreme storm event 10 years before. A random sample of geoducks (a big clam, the name is actually pronounced "gooey-duck") was taken off the west coast of a North Island surf beach in 2013. Each geoduck was aged by counting rings in the shell, and the count of geoduck at each age was recorded. The population theory of geoducks suggests using the following model for the expected count at each age:

$$log(E[Count_i]) = \beta_0 + \beta_1 \times Age_i$$

Here, we'll be using a modified version of this model since there was as extreme storm event that all geoduck of age 10 or older experienced. The initial model to be fitted includes the storm effect and its interaction with age in the following way:

$$log(E[Count_i]) = \beta_0 + \beta_1 \times Age_i + \beta_2 \times Storm_i + \beta_3 \times Age_i \times Storm_i$$

where $Count_i$ is the geoduck count at age i and $Storm_i$ is an indicator variable that takes the value 0 if age is less than 10, or 1 if age is at least 10.

The data is stored is `Geoduck.txt` which contains the following variables:

`Count`	Number of geoducks counted at given value of age
`Age`	Years of age of geoduck
`Storm`	Indicator variable (0 if age<10, or 1 if age ≥ 10)

The questions we want answered are: how does the population of geoducks change with age, did the storm event have any impact on the population, and did the relationship between age and the population count change after the storm?

- Comment on the plot of the data.
- Fit an appropriate Poisson GLM to model the number of geoducks. Determine whether the model can be simplified and determine an appropriate final model. Generate confidence interval output for this model.
- Write a Method and Assumption Checks section.
- Write an Executive Summary.

Question 2. [17 Marks]

Now for some gruesome research from the 1950s (a very different time from today – we hope). A group of researchers investigated the utility of Streptomycin as a means of treating radiated mice. A total of 522 mice were irradiated by a burst of fast neutrons (each randomly assigned to get either a high or low dose of radiation). The mice were then randomly assigned to a treatment group, either Control or Streptomycin. The outcome of the experiment was then whether a mouse survived for 10 days or not.

The data is stored is `mice.csv` which contains the following variables:

`n`	The number of mice in the group
`Survived`	The number of mice in the group that survived for 10 days
`Died`	The number of mice in the group that died within 10 days
`Propn`	The proportion of mice in the group that survived for 10 days
`Treatment`	The treatment the group received (`Strept` or `Control`)
`Dose`	The dose level of radiation the group was given (`high` or `low`)

We wish to determine whether or not Streptomycin was effective as a radiation treatment and whether or not the effectiveness depended on the level of radiation dose.

- Comment on the plots of the data.
- Fit an appropriate logistic regression model to investigate the effect of the predictors on the proportion of mice that survived. Carry out model checks and amend the model as needed. Determine whether the model can be simplified and determine an appropriate final model. Generate confidence interval output for this model.
- Write a Method and Assumption Checks section.
- Write an Executive Summary.