Lab Assignment 2 – Generalized linear models (GLMs)

WFSC 570 Wildlife Habitat Analysis

For this assignment, you will practice simulating data under a binomial GLM, fitting a binomial GLM to those simulated data, and plotting your simulated data against the expected value of the response variable from your fitted model.

This lab will have two parts both designed to show you how changing our sample size changes our ability to precisely estimate the true values of our model parameters. This is something we can only examine using simulated data because 1) we know our true parameter values when we simulate data and 2) we can easily generate simulated data sets of varying sample sizes. In lab, we used large simulated data sets so as to not create confusion about why our coefficient estimates may differ more-or-less from our true parameter values. But here you get to see how large those differences might be.

I have provided R script for both parts which will simulate the data. There are a few lines of code that you must change to complete this exercise but you should not need to change other parts of the code nor do you necessarily need to understand what is happening in these parts of the code. This exercise does use 95% confidence intervals (CI) which are a measure of the precision of a parameter estimate. The narrower the 95% CI, the more precise our parameter estimate.

*Note: Be sure to install the ggplot2 package on your computer and load it at the beginning of your script using library(ggplot2).*

**Part I:** Effects on parameter estimate precision

Refer to Part I of the lab assignment R script. On line 5 you will see the following vector:

sample\_sizes <- c(A, B, C, D)

This vector will provide the sample sizes to simulate data according to a binomial GLM through a for-loop on lines 14-36. This for-loop will simulate data, fit a binomial GLM, and then store the estimate and 95% confidence interval (CI) for the intercept and slope. You will need to change each of the four letters into numbers and then run the for-loop. You can highlight the entire for-loop and click “Run” to run it all at once. The for-loop will store the coefficient estimates and their 95% CI in the data frame *results.* I have then provided some code to plot these estimates and their 95% CI using the *ggplot* package.

Your four sample sizes should include a range of realistic sample sizes based on your experiences as wildlife biologists but also include an unrealistically large sample size (e.g., 1000, 5000) for comparison. Run the for-loop, plot the data, and complete the following tasks.

* Verbally describe how the differences in the width of the 95% CI across your four sample sizes.
* Save the two plots you created and include them in your report.
* Researchers often use the 95% CI to determine which coefficient estimates should be used for making statistical inferences. Specifically, many researchers only make inferences from slope estimates (β1) whose 95% **exclude** zero. Using this criterion, how might sample size affect any inferences made from the slope estimate?

***Note:*** *If you use sample sizes smaller than 15-20 you will probably see some error messages about algorithms not converging or fitted probabilities of 0 or 1 occurring. This means the Binomial GLM is having trouble estimating the model parameters, probably because of data insufficiency and/or small sample sizes. Increasing the sample sizes will fix this. The red message you see saying “Waiting for profiling to be done…” is OK. This is just a normal output of the function I used to calculate the 95% CI.*

**Part II:** Plotting expected values from a binomial GLM

This part is designed to help you see how the expected (i.e., predicted) values from a binomial GLM change as we change the intercept and the effect size (i.e., slope). I have provided a custom function, *plot.glm.R*, which will allow you to easily change sample size (n), intercept (b0), and slope (b1) and see a plot of the expected values and their 95% CI. Complete the following tasks:

* Pick a single, realistic sample size and change b0 and b1 in *plot.glm.R*.
* Run *plot.glm.R* for two different effect sizes (b1) keeping sample size (n) and intercept (b0) the same. Describe how the expected values change between the two values of b1, save the two plots, and include them in your report.

**Part III:**

This part is designed to help you see how the expected values from a binomial GLM change as we increase our sample size whole holding our model parameters (the intercept and slope) constant. Using the *plot.glm.R* function to complete the following tasks:

* Pick a single, reasonable value for b0 (intercept) and b1 (slope)
* Pick four different values of sample size (n) and see how the expected values and their 95% CI change. Save these plots and include them in your report. Trying going as low as a sample size of 15 and then work up from there to an unrealistically large sample size (e.g., 1000, 5000).

Your assignment is to work in pairs, complete the following tasks and answer the following questions as a pair, and submit a single written report for each pair describing how you completed the tasks and your answers to the questions. Reports should be written using complete sentences and paragraph structure. Also include your R script (either as a separate file and copied-and-pasted into the end of your report).

**This assignment will be due on D2L by the beginning of lab (2:00) on Tuesday September 24, 2024.**