WFC 198 “Sampling Animal Populations” – Homework Week 3

Due April 25th, 9am

Please complete the following exercises in R, “knit” your script into a html report (as you learned in lab) and submit it on Canvas under Assignments.

Please preface every piece of code with a comment that states to which question the code belongs, for example: “#Task 3a” before the code used to tackle Task 3a. If we cannot follow your code, we cannot grade it. If you are unsure of how to format your script, see the short example homework script on Canvas (Files – Homework – Examplescript.R).

Work through the questions in order, as they build on each other.

Task 1: Simulating logistic regression data (6 Pts)

Simulate data for a logistic regression with one predictor variable

1. The intercept of the linear predictor is -0.2 (that is, minus 0.2); the slope (describing the effect of covariate X) is 1.3. The sample size (number of data points) is 30. (1 Pt)
2. The predictor variable, X, comes from a uniform distribution with lower limit -1 (that is, minus 1) and upper limit 1. (1 Pt)
3. Calculate the expected values both on the logit and on the response (i.e., probability) scale (2 Pts)
4. Use the rbinom() function to generate the response variable Y. Note that you need to specify three arguments in rbinom(): how many observations to generate (n), the number of Bernoulli trials for each of these observations (which is 1, because each data point in a logistic regression is Binary), and the probability (ie, expected value on the natural scale). Use ?rbinom for more help. (2 Pts).

Task 2: Analyzing data from Task 1 (8 Pts)

1. As we have done in lab, analyze the data you generated under Q1 using logistic regression. Include the covariate in your model. Make sure your “knitted” report shows the summary output from your model\* (2 Pts)
2. Calculate the differences between the parameter estimates and the true input parameters, and save these differences in a numerical vector. Make sure your “knitted” report shows the values in the numerical vector\*. TIP: you can use the coefficients() function to extract parameter estimates from a model object (2 Pts)
3. Generate expected values of Y ***on the link scale*** (ie., the logit scale) based on the model fit under a) for a sequence of values of X ranging from -1 to 1 at intervals of 0.2. (2 Pts)
4. Plot these expected values against the sequence of values of X generated in c). Appropriately label your axes. (2 Pts)

\*This means that your script needs to contain code to look at the model summary (or at other objects for which I ask you to make sure its value appear in the report). For example, if you have a model object called “*mod*”, for the summary output to appear in the “knitted” report, you would have to include the following line of code:

summary(mod)

Or, if I asked you for the content of a vector, *vec*, you would need a line of code just calling up that vector:

vec

Question 3: Analyzing Poisson regression data (7 Pts)

The data file “HW3data.csv” contains two columns, “Species”, showing the number of bird species detected at a given site (response variable, count), and “Quality”, and index of habitat quality at that site (predictor variable), where higher numbers indicate higher quality.

1. Read the data file into R and perform a Poisson regression (with the covariate) on the data (2 Pts)
2. Generate expected values of Y ***on the response scale*** based on the model fit under a) for a sequence of habitat quality values, ranging from the minimum observed quality to the maximum observed quality, at intervals of 0.1 (2 Pts)
3. Plot these expected values against the range of habitat quality values generated in b). Adjust the y axis to range from 0 to 7 using the *ylim* argument\*\*. Label your axes appropriately. (2 Pts).
4. Add the observations to the plot generated in c). (1 Pt)

\*\*If the lab didn’t cover this because it ran out of time, look at the last couple of lab slides, and the lab code for help

Question 5 (not an R question): (4 Pts)

Interpret the results from the Poisson regression in Q3. In 2-3 sentences, quantitatively\*\*\* describe the effect of the predictor variable on the response variable (both direction and magnitude; use units); what the null hypothesis is in this context, which decision you make about the null hypothesis, based on what part of the results.

Type your answer into your R script; remember to comment it out, since it is not code! Please use line breaks, rather than writing everything into one very long line.

\*\*By that I mean, it is not enough to say the effect is positive or negative, strong or weak, you need to use the parameter estimate in your description, and explain what that number means in the context of this example.

TOTAL POINTS: 25