

Data – 622

Homework # 1

Due date Feb 19, 2021- 11:59 EST

Let's use the Penguin dataset for our assignment. To learn more about the dataset, please visit:

<https://allisonhorst.github.io/palmerpenguins/articles/intro.html>

For this assignment, let us use 'species' as our outcome or the dependent variable.

1. Logistic Regression with a binary outcome. (40)
 - a. The penguin dataset has 'species' column. Please check how many categories you have in the species column. Conduct whatever data manipulation you need to do to be able to build a logistic regression with binary outcome. Please explain your reasoning behind your decision as you manipulate the outcome/dependent variable (species).
 - b. Please make sure you are evaluating the independent variables appropriately in deciding which ones should be in the model.
 - c. Provide variable interpretations in your model.
2. For your model from #1, please provide: AUC, Accuracy, TPR, FPR, TNR, FNR (20)
3. Multinomial Logistic Regression. (40)
 - a. Please fit it a multinomial logistic regression where your outcome variable is 'species'.
 - b. Please be sure to evaluate the independent variables appropriately to fit your best parsimonious model.
 - c. Please be sure to interpret your variables in the model.
4. **Extra credit:** what would be some of the fit statistics you would want to evaluate for your model in question #3? Feel free to share whatever you can provide. (10)