PSC 103B

Homework 7

Winter 2024

**Instructions**

Please use R/RStudio to complete the following questions. You will submit your filled-out version of this document **as a PDF** on Canvas. Make sure your PDF looks as expected before submitting. Unless otherwise specified, please **always** **include the code you used to generate your answer for each question, or the steps you used to calculate the answer (when relevant), as well as the final answer and/or relevant output** (output is what comes out in the console when you run an R command, e.g., the results of the model when you run summary(fit1)).It’s a good idea to organize your R code in the R script and save it, so if you need to modify or recalculate one of the questions, that’s easy to do (see the R scripts provided in previous weeks for tips on how to organize your code). If you copy and paste code or output into this document (screenshots are also acceptable), **please format the code and output using a fixed-width font** (e.g., Courier) so it’s easier to read.

You may consult with your classmates while working on the assignment, but **you must do all the work yourself – everything you turn in must be your own code and words**. Academic dishonesty will not be tolerated.

Please submit **a pdf version of this document** with your answers on Canvas by **1:59pm on Friday, March 19.**

**Logistic Regression**

For these logistic regression equations, we will be using the *wine* dataset that we used in lab, and that is available on the Canvas page for this homework assignment. We will be exploring how different predictors relate to whether a red wine is rated as good or bad.

**Question 1**

Write a logistic regression model predicting the wine quality from the predictor of residual sugar. Be sure to create a binary version of the quality variable like I did in lab! Show your code and summary output. (1 point)

**Question 2**

Write out the regression model predicting the log-odds of whether a wine was rated as good or not. Interpret the intercept and slope in terms of the log-odds. (3 points)

**Question 3**

Write out the regression model predicting the *odds* of whether or not a wine was rated as good or not. Interpret the intercept and slope in terms of the odds. (3 points)

**Question 4**

Now center your predictor variable of residual sugar, and re-run the logistic regression model. Show your code and summary output (1 point)

**Question 5**

What is the new interpretation of the intercept (in terms of both the log-odds and the odds). (2 points)

**Question 6**

Assume we had a wine that had a residual sugar content of 13. Calculate and interpret: the log-odds, the odds, and the probability of a wine with this residual sugar content being rated as good. (3 points).

**Question 7**

Fit a logistic regression model with both residual sugar and citric acid (neither of them centered) as predictors. Show your code and summary output (1 point).

**Question 8**

Interpret the intercept and both slopes in terms of the log-odds (3 points).

**Question 9**

Interpret the intercept and slopes in terms of the odds. (3 points)

**Dummy Coding**

Use the penguins dataset from the palmerpenguins package in R for this question.

**Question 10**

Fit a dummy-coded linear regression model using bill length as the outcome and Island as the predictor. Show your code and model output (1 point).

**Question 11**

Interpret the intercept and slopes of the regression output. Which groups are significantly different from the reference group? (3 points).