Programming Assignment 8
STAT 311

Please complete the following problems and submit a file named `STAT311-HW8.R` to Gradescope. You should start from the provided `STAT311-HW8.R` file on Canvas.

You may need to install the `palmerpenguins` package.

Code is provided to plot the data that will be used for analysis, but ensure that the plotting code is commented out for submission to gradescope or the autograder will fail.

Overview

Address each of the following questions. As always, you should start from the provided skeleton code.

All of these questions look at modeling the relationship between bill length and bill depth of penguins, specifically trying to predict length from depth.

Each question should utilize the model associated with the question group.

Note that answer need to be formatted correctly; gradescope will not accept 'named numbers', which can be converted to numeric variables by using double square braces: `[[1]]`.

Output by functions such as `lm` will often by named numbers; be careful when extracting values from these functions, and use double square braces when appropriate: ie `lm(...)$coefficients[[2]]`.

Be sure to format responses as specified, especially for intervals which should be in the format `c(lowerbound, upperbound)`.

Question 1 - (6 points)

Fit a model to predict bill length from bill depth for the palmerpenguins data, just called `penguins`.

a) What is the estimated intercept term of the model? Save your answer in the variable `q1.a`

b) What is the estimated slope term of the model? Save your answer in the variable `q1.b`

c) Predict the expected bill length for a penguin with a bill depth of 14mm. Save your answer in the variable `q1.c`

d) Construct a 90% prediction interval for the bill length of a penguin with a bill depth of 19mm. Save your answer in the variable `q1.d`

e) Construct a 99% confidence interval for the average bill length of a penguin with a bill depth of 19mm. Save your answer in the variable `q1.e`

f) What is the p-value for determining if there is evidence that the slope of this model is different from 0? Save your answer in the variable `q1.f`

Question 2 - (6 points)

Fit a model to predict bill length from bill depth with different intercepts (but not slopes) based on sex for the palmerpenguins data.

a) What is the estimated intercept term for female penguins? Save your answer in the variable `q2.a`

b) What is the estimated intercept term for male penguins? Save your answer in the variable `q2.b`

c) What is the expected bill length for a male penguin with a bill depth of 14mm? Save your answer in the variable `q2.c`

d) Construct a 95% prediction interval for the bill length of a female penguin with a bill depth of 19mm. Save your answer in the variable `q2.d`

e) Construct a 95% confidence interval for the slope of the regression equation. Hint: The degrees of freedom depends on the number of parameters in the model, but can be found from the output of the summary(lm(...)) command. Save your answer in the variable `q2.e`

f) TRUE or FALSE: Using the p-values to determine parameter significance, we should keep the sex differentiated intercept in this model. Save your answer in the variable `q2.f`

Question 3 - (6 points)

Fit a model to predict bill length from bill depth with different intercepts and slopes based on sex for the palmerpenguins data.

a) What is the estimated slope term for female penguins? Save your answer in the variable `q3.a`

b) What is the estimated slope term for male penguins? Save your answer in the variable `q3.b`

c) Construct a 90% prediction interval for the bill length of a male penguin with a bill depth of 15mm. Save your answer in the variable `q3.c`

d) Construct a 99% confidence interval for the average bill length of a female penguin with a bill depth of 18mm. Save your answer in the variable `q3.d`

e) What is the p-value for determining the signficance of the difference in slope for male penguins? Save your answer in the variable `q3.e`

f) TRUE or FALSE: Using the p-values to determine parameter significance, we should keep the sex differentiated intercept and slope in this model. Save your answer in the variable `q3.f`

Question 4 - (7 points)

Fit a model to predict bill length from bill depth with different intercepts and slopes based on species for the palmerpenguins data

a) What is the estimated intercept term for Adelie penguins? Save your answer in the variable `q4.a`

b) What is the estimated intercept term for Gentoo penguins? Save your answer in the variable `q4.b`

c) What is the estimated slope term for Adelie penguins? Save your answer in the variable `q4.c`

d) What is the estimated slope term for Chinstrap penguins? Save your answer in the variable `q4.d`

e) What is the expected bill length for a Gentoo penguin with a bill depth of 15mm? Save your answer in the variable `q4.e`

f) Construct a 80% prediction interval for the bill length of a Chinstrap penguin with a bill depth of 20mm. Save your answer in the variable `q4.f`

g) Construct a 90% confidence interval for the average bill length of an Adelie penguin with a bill depth of 17mm. Save your answer in the variable `q4.g`

Question 5 - (1 points)

Which model do you think best describes the data? Your answer should be one of the following
`q5.answers=c("Model 1", "Model 2", "Model 3", "Model 4")`.
You can ensure correct format by selecting your answer using `q5.answers[#]`.

a) Save your answer in the variable `q5.a`