Stat 6021: Guided Question Set 3

We will look at a data set concerning adult penguins near Palmer station, Antarctica. The data set, `penguins` comes from the `palmerpenguins` package. Be sure to install and load the `palmerpenguins` package. I recommend reading the documentation of this data set by typing `?penguins`

We will explore the relationship between the response variable body mass (in grams), `body_mass_g`, and the predictor length of the flippers (in mm), `flipper_length_mm`.

1. Produce a scatterplot of the two variables. How would you describe the relationship between the two variables? Be sure to label the axes and give an appropriate title. Based on the appearance of the plot, does a simple linear regression appear reasonable for the data?

2. Produce a similar scatterplot, but with different colored plots for each species. How does this scatterplot influence your answer to the previous part?

3. Regardless of your answer to the previous part, produce a scatterplot of body mass and flipper length for Gentoo penguins. Based on the appearance of the plot, does a simple linear regression appear reasonable for the data?

4. What is the correlation between body mass and flipper length for Gentoo penguins. Interpret this correlation contextually. How reliable is this interpretation?

   For the rest of the questions, assume the assumptions to perform linear regression on Gentoo penguins are met.

5. Use the `lm()` function to fit a linear regression for body mass and flipper length for Gentoo penguins. Write out the estimated linear regression equation.

6. Interpret the estimated slope contextually.

7. Does the estimated intercept make sense contextually?

8. Report the value of R^2 from this linear regression, and interpret its value contextually.

9. What is the estimated value for the standard deviation of the error terms for this regression model, $\hat{\sigma}$?

10. For a Gentoo penguin which has a flipper length of 220mm, what is its predicted body mass in grams?

11. Produce the ANOVA table for this linear regression. Using only this table, calculate the value of R^2.

12. What are the null and alternative hypotheses for the ANOVA F test?

13. Explain how the F statistic of 118.01 is found.

14. Write an appropriate conclusion for the ANOVA F test for this simple linear regression model.