Stat 311 Homework 8

This assignment uses the file PenguinsHW8.csv. This is a popular data set; the data were collected and made available by <u>Dr. Kristen Gorman</u> and the <u>Palmer Station, Antarctica LTER</u>, a member of the <u>Long</u> <u>Term Ecological Research Network</u>. (Gorman, Williams, and Fraser 2014). This version of the data set was downloaded from <u>here</u> (on GitHub Gist).

The goal is to predict penguin bill length from penguin bill depth and perhaps species. Bill length and bill depth are shown in the image below (Image copied from <u>here</u>.)



Regression Model

- 1. In the HW8 template, we provide code that creates a scatterplot of bill length on bill depth. Describe the joint relationship between bill length and depth.
- 2. In the HW8 template, we provide code that creates a second scatterplot that adds color coded plotting symbols by species. Describe the joint relationship(s) taking species into account.
- 3. Run a linear regression of bill length on bill depth. Show the regression summary (no interpretations needed).
- 4. Write out the regression equation using information from the regression summary.
- 5. Interpret the estimated slope of the regression equation in the context of the problem.
- 6. Report and interpret R^2 .
- 7. In the HW8 template, we provide ggplot code to create model diagnostic plots (residual plot, histogram of residuals, normal QQ plot of residuals). Do you think that the assumptions for inference are met? Explain addressing specific assumptions.

Inference for Regression (assume all assumptions for inference are met)

- 8. Find a 95% confidence interval for the slope parameter using "by-hand" calculations in a code chunk. Interpret the interval. [Hint: pull the numbers you need from the regression summary]
- 9. Perform a hypothesis test to determine if the slope parameter is different than zero. State the hypotheses using symbols, report the test statistic and degrees of freedom, and the p-value. Include your decision and an interpretation in the context of the problem. Use a 5% significance level. [Hint: no code needed]
- 10. Observation 194 is a Gentoo penguin with a bill length of 49.6 mm and a bill depth of 16.0 mm. Using the model from problem 4, what is the residual for this penguin? Does the model over or underestimate the bill length for this penguin?
- 11. Find the 99% confidence interval for the mean bill length when the bill depth is 16.0 mm. Interpret this interval.
- 12. Find the 99% prediction interval for a penguin that has a bill depth of 16.0 mm. Interpret this interval.

Consider Species

13. In the HW8 template, we include code to run a model that includes bill depth and species. We start by allowing the slopes to vary by species (using an asterisk between bill depth and species). Since there are

Stat 311 Homework 8

three species, the first species will be included in the intercept term (Adeline) and the other two species will have their own coefficients. You should see that both additional interaction terms are statistically significant. What does this mean?

- 14. Write out the equation for the model for Gentoo penguins, rounding all coefficients to two decimal places.
- 15. Calculate the residual for observation 194 (see Problem 10) based on the new model. Based on the residual, does this new model do a better job of predicting the bill length for the Gentoo penguin in observation 194?