STT 380

In-Class Assignment 24

For the penguins dataset,

1. Create a single-variable model, using bill length to predict body mass.

    a. summary(lm(data = penguin, body_mass_g ~ bill_length_mm))

2. Next, use bill length and species to predict body mass. What are the level shifts between the species?

    a. summary(lm(data = penguin, body_mass_g ~ bill_length_mm + species))

    b. A categorical variable as an input functions as a "level shift" - so the level shifts would be both bill length and species.

3. Create separate models for the 3 species. How do the models compare, in terms of errors and coefficients? Which of the models are similar to each other (if at all)?

    a. mods <- list()

    b.

    c. spec <- c("Adelie", "Chinstrap", "Gentoo")

    d. for(i in 1:3){

    e.   sset <- subset(penguin, subset = species == spec[i])

    f.   mods[[i]] <- lm(data = sset, body_mass_g ~ bill_length_mm)

    g. }

    h. summary(mods[[1]])

    i. summary(mods[[2]])

    j. summary(mods[[3]])

4. Which model is the "best?" Why?

    a. Adelie model has Adjusted R-squared:0.2966 and p-value: 2.955e-13.

    b. Chinstrap model has Adjusted R-squared:0.2527 and p-value: 7.48e-06

    c. Gentoo model has Adjusted R-squared:0.4432 and p-value: < 2.2e-16. by these results we would be able to say that model 1 is the "best."