Stat 6021: Guided Question Set 8

We will revisit the data set `penguins` from the `palmerpenguins` package. The data set contains size measurements for adult foraging penguins near Palmer Station, Antarctica. In this set of questions, we focus on exploring the relationship between body mass (y) and bill depth (x_1) of three species of penguins.

1. Create a scatterplot of the body mass against the bill depth of the penguins. How would you describe the relationship between these two variables?

2. Create the same scatterplot but now with different colored plots for each species. Also be sure to overlay separate regression lines for each species. How would you now describe the relationship between the variables?

3. Create a regression with interaction between bill depth and species, i.e.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 I_1 + \beta_3 I_2 + \beta_4 x_1 I_1 + \beta_5 x_1 I_2 + \epsilon,$$

   where I_1 and I_2 are indicator variables where $I_1 = 1$ for Chinstrap penguins and 0 otherwise, and $I_2 = 1$ for Gentoo penguins and 0 otherwise. Write down the estimated regression equation for this model.

4. Carry out the relevant hypothesis test to see if the interaction terms can be dropped. What is the conclusion?

5. Based on your answer in part 4, write out the estimated regression equations relating body mass and bill depth, for each species of the penguins.

6. Assess if the regression assumptions are met, for the model you will recommend to use (based on part 4). Also, be sure to carry out Levene's test of equality of variances since we have a categorical predictor.

7. Briefly explain if we can conduct pairwise comparisons for the difference in mean body mass among all pairs of species for given values bill depth, i.e.,

   (a) Adelie and Chinstrap,
   (b) Adelie and Gentoo,

(c) Chinstrap and Gentoo.

If we are able to, conduct Tukey's multiple comparisons and contextually interpret the results of these hypothesis tests.