

Additive and Interaction Models with Categorical Explanatory Variables

Visualizing an Additive Model with Categorical Variables

Our fitted multiple regression equation relating the average bill depth of Antarctic penguins (depth_i , measured in mm) to their bill length (length_i , measured in mm) and species indicators (Adelie_i , for Adelie penguins, and Gentoo_i , for Gentoo penguins) is

$$\widehat{\text{depth}}_i = 8.66 + 0.20 (\text{length}_i) + 1.93 (\text{Adelie}_i) - 3.17 (\text{Gentoo}_i)$$

1. What does this equation suggest is the fitted regression line relating bill length and bill depth in Adelie penguins? Find this line by evaluating the species indicators for an Adelie penguin (i.e., plugging in 1s and 0s, as appropriate) and then simplifying.
2. What is the fitted regression line relating bill length and bill depth in Chinstrap penguins?
3. What is the fitted regression line relating bill length and bill depth in Gentoo penguins?
4. Sketch a graph of the fitted multiple regression equation above. Your graph should have one line for each of the three penguin species. Clearly label which line corresponds to which species. Then, label your graph to show where the slope for bill length ($\hat{\beta}_1 = 0.20$), the coefficient for the Adelie indicator ($\hat{\beta}_2 = 1.93$), and the coefficient for the Gentoo indicator ($\hat{\beta}_3 = -3.17$) all appear in the plot.

Interaction Models with Categorical Variables

If we allow for a possible interaction between penguin species and bill length, then our population regression equation for the mean bill depth becomes

$$\begin{aligned}\mathbb{E}[\text{depth}_i | \text{length}_i, \text{species}_i] = & \beta_0 + \beta_1 (\text{length}_i) + \beta_2 (\text{Adelie}_i) + \beta_3 (\text{Gentoo}_i) \\ & + \beta_4 (\text{length}_i) (\text{Adelie}_i) + \beta_5 (\text{length}_i) (\text{Gentoo}_i)\end{aligned}$$

Just as was the case for the additive “parallel slopes” model on the previous page, this one regression equation can be used to produce three different regression lines: one for each of the different penguin species. Evaluate the species indicators, collect like terms, and simplify in order to:

1. Find the population regression line relating bill length to mean bill depth in Adelie penguins. What term(s) in the original interaction model make up the new intercept? What term(s) make up the new slope?
2. Find the population regression line relating bill length to mean bill depth in Chinstrap penguins. What term(s) in the original interaction model make up the new intercept? What term(s) make up the new slope?
3. Find the population regression line relating bill length to mean bill depth in Gentoo penguins. What term(s) in the original interaction model make up the new intercept? What term(s) make up the new slope?