Lesson 17. Comparing Two Regression Lines – Part 1

1 Overview

- So far, we have worked with multiple linear regression models when all predictors are quantitative

- This lesson: multiple linear regression models with a categorical predictor

- By including a categorical predictor in our model, we can make comparisons between groups and make better predictions

2 Using one model to fit two lines (with different intercepts)

- Suppose we have a categorical variable that divides our observations into two groups, A and B

- We include the categorical predictor as an **indicator variable**:

- The **full model** is:

- For observations in group A, the model reduces to:

- For observations in group B, the model reduces to:

- Coefficients:

  ○ β_0:

  ○ β_1:

  ○ β_2:

3 Why fit a separate model for each group?

- Pros:

  - We already know how to do it
  - We would learn about the patterns for each group

- Cons:

  - Difficult to <u>formally</u> compare the two groups
  - Difficult to use a common parameter for each group (if desired)
  - We have to estimate "extra" parameters

- Good news! The flexibility of multiple linear regression allows us to fit one model that helps us compare relationships for different groups

Switch to Part 2 for an example...

4 Allowing different intercepts AND different slopes

- Suppose again that we have a categorical variable that divides our observations into two groups, A and B

- Now, we still include the categorical predictor as an indicator variable:

$$GroupB = \begin{cases} 1 & \text{if in Group B} \\ 0 & \text{otherwise} \end{cases}$$

- We also include an **interaction term** as a predictor:

  - This term <u>multiplies</u> two predictors together
  - This allows the slopes to be different for each level of the binary categorical variable

- The full model is:

- For observations in group A, this reduces to:

- For observations in group B, this reduces to:

- Coefficients:

  - β_0:

  - β_1:

  - β_2:

  - β_3:

Switch to Part 2 for an example...