Lesson 18. 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 want to predict the quantitative variable Y based on a quantitative variable X
- We also have a categorical variable that divides our observations into two groups, A and B
- We code the categorical variable as an **indicator variable**:

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

- The model is:

$$Y = \beta_0 + \beta_1 X + \beta_2 GroupB + \varepsilon \qquad \varepsilon \sim \text{iid } N(0, \sigma_\varepsilon^2)$$

- 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:

⇒ β_2 represents the difference in the magnitude of Y due to membership in group B versus group A

Switch to Part 2 for an example...

3 Using one model to fit two lines with different intercepts AND different slopes

- Same setup as before:
  - We want to predict the quantitative variable Y based on a quantitative variable X
  - We also have a categorical variable that divides our observations into two groups, A and B
- Now, we still code 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:

$$X \times GroupB$$

  - This term multiplies two predictors together
  - This allows the slopes to be different for each level of the binary categorical variable
- The full model is:

$$Y = \beta_0 + \beta_1 X + \beta_2 GroupB + \beta_3 (X \times GroupB) + \varepsilon \qquad \varepsilon \sim \text{iid } N(0, \sigma_\varepsilon^2)$$

- 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:

  $\Rightarrow$ β_3 represents the difference in the rate of change in Y vs. X due to membership in group B versus group A

Switch to Part 2 for an example...