

Linear Regression - Categorical Predictor

Reading

Ch. 8.1-8.2

Activity Description

We will continue using the SAT GPA data set from Lab 3. This data set contains one categorical predictor - sex. In this activity, we will explore the relationship between sex and first year GPA. The variable sex is dichotomized; it is coded as a 1 for males, and as 2 for females.

Instructions

1. Open the Markdown file from Lab 3 that we were working on in class on Wednesday. You will complete the following at the end of the file, under “Model the relationship between first year college GPA and sex”. You should answer all of these questions in your Markdown file, adding R chunks where necessary.
2. Run the R chunk (already in your Markdown file) that turns sex into a factor. Why do you think this is a necessary step? (Hint: what values are used to code for males versus females?)
3. Make a side-by-side boxplot using the ggplot2 library to show the relationship between FYGPA and sex. See Lab 2 (last part) if you do not remember how to do this.
4. Write down the equation for the linear model like we have in the notes. Be sure to define what x and y are.
5. What values can x take on here?
6. Fit a linear model that uses sex to predict FYGPA; assign this linear model to `lm_FYGPA_sex` (as we did with other linear models - see the first part of this lab if you need a refresher).
7. Use the summary function on your linear model to print out the model summary details.
8. What is the estimate for $\hat{\beta}_0$? What does this mean? Interpret. (Hint: this will be similar to what we have been doing for numerical predictors.)
9. What is the estimate for $\hat{\beta}_1$? What do you think it means? Try to interpret - we will talk about this as a class, but I want you to think about it first. (Hint: think about the values that x can have.)
10. What is the R^2 ? Interpret this value in the context of the problem.
11. Here we have a categorical predictor with two levels (male or female). Is the linearity assumption satisfied? Why or why not? (No R required.)