Problem Set 7: Linear Regression

Joe Ornstein

Due November 11, 2020

Create an R project, an R script, load tidyverse, and import the modified Freedom In The World dataset you created for the midterm.

Line of Best Fit

1. Estimate a line of best fit with `liberty_score` as the outcome and `democracy_score` as the explanatory variable.
2. Visualize the line of best fit with `geom_smooth(method = 'lm')`.
3. What is the slope of the relationship? What is the 95% confidence interval around that slope?
4. Are there any outliers with higher or lower liberty scores than you would expect given their democracy score? Which countries are they? (Hint: there is a vector called `residuals` in the `lm` object you just created. Add it to your FIW dataframe).

Data Wrangling and OLS

I'm including another dataset that I pulled from the World Bank with GDP per capita figures for each country since 1960. Man, it's a mess. We're going to have to tidy it up.

5. Read the dataset into `R` and pivot the data so each row represents a country-year.
6. Keep the most recent year of gdp per capita
7. Create a new variable, `log_gdp_per_capita`, equal to the logarithm of GDP per capita.
8. Save your cleaned up World Bank dataset to the `data/` folder.
9. Merge your cleaned up World Bank dataset with the FIW dataset.
10. Estimate a multivariable linear model with `liberty_score` as the outcome variable and `democracy_score` and `log_gdp_per_capita` as the explanatory variables. What are the slope coefficients and 95% confidence intervals?