

# Assignment 02

## Simple Linear Regression: Description

This goal of this assignment is to give you experience fitting and interpreting regression models. Submit your responses to each of the questions below in a printed document. All graphics should be resized so that they do not take up more room than necessary and also should have an appropriate caption. This assignment is worth 10 points. (Each question is worth 1 point unless otherwise noted.)

Should more money be spent on public schools or should that money be spent elsewhere? Both sides of this ongoing public debate have been argued passionately, using a multitude of anecdotal evidence. Although we will not settle this debate, we will examine data akin to the types of data that policy makers use to make funding decisions. Specifically, we will examine whether teacher salaries are related to SAT scores at the state level. For this assignment, you will use the file *state-education.csv* (see the [data codebook](#)).

### Preparation

Before carrying out any analyses, create a predictor called `salary_thousand` that indicates the average state salary in thousands of dollars (e.g., `salary = 52143`; `salary_thousand = 52.143`). This variable (not `salary`) should be used in all analyses in Part I.

### Questions

1. Create a plot of the marginal distribution of SAT scores. Make sure your plot has a caption.
2. Examine the structure and formatting of Table 1 at <http://zieff0002.github.io/epsy-8251/misc/creating-tables/creating-tables.html>. Mimic the format and structure of this table to create a table to present the numerical summary information for the marginal distributions of SAT scores and salaries. Provide the same measures for these variables as is given in Table 1 in the article. Re-create the formatting of Table 1 as closely as you can. Finally, make sure the table you create also has an appropriate caption.
3. Create a scatterplot of the distribution of SAT scores ( $Y$ ) *conditioned on* teacher salaries ( $X$ ). Make sure your plot has a caption.
4. Describe the relationship between SAT scores and teacher salaries. Be sure to comment on the structural form, direction and strength of the relationship. Also comment on any potential observations that deviate from following this relationship (unusual observations or clusters of observations).
5. Regress SAT scores on teacher salaries. Write the *fitted equation* using Equation Editor (or some other program that correctly types mathematical expressions).
6. Interpret the value of the intercept from the regression equation using the context of the data.
7. Interpret the value of the slope from the regression equation using the context of the data.
8. Compute, report, and interpret the value for  $R^2$  based on values from the ANOVA decomposition.
9. Compute and report the predicted mean SAT score for Minnesota students based on the average teacher salary in the state using the fitted regression equation.

10. Compute and report the residual for Minnesota. Explain what the sign and magnitude of the residual value tells you about how Minnesota's mean SAT score compares to the mean SAT score for other states having the same average teacher salary?