# Assignment 4: Simple and Multiple Regression

Due Friday, October 24, 2025

Insert Name Here

**Instructions**

Answer the questions within this document. Keep the questions included below. Make sure to answer the questions in sufficient detail. Report all regression results using *texreg*. A texreg template is included in the folder.

- Submit your rendered .pdf in Blackboard.

# 1. Simple Regression

Consider the following hypothesis: *In comparing U.S. states, as the percentage of college graduates increases, the percentage of people turning out to vote in elections will increase as well.* Use the "states.dta" dataset to answer the following questions. In that dataset, percent college grads = "BA_or_more." Voter turnout = "vep16_turnout." Note that this is percent turnout of the *voter eligible population.*

A. What are the units of analysis? What is N?

B. What is the independent variable and what is the dependent variable?

C. Generate a scatterplot between the dependent and independent variable. Include the line of best fit. Interpret the results.

D. Estimate a simple regression. Show your output. Interpret the intercept and slope. Be specific.

E. How would a 10% increase in the college graduation rate influence turnout?

F. Use the regression results to draw a conclusion about whether college graduation rate exhibits a statistically significant effect on voter turnout. Be specific about which information you're using and the decision rule for drawing your conclusion.

G. Interpret $R^2$.

# 2. Multiple Regression

A critic suggests that your regression results in number 1 are highly suspect because you haven't controlled for a key variable: the percentage of whites in each state. This critic contends that because African Americans had their suffrage rights suppressed for so many years, they have shorter histories of voting participation, are still subject to voter suppression, and therefore, have lower "base rates" of voting than whites. As a result, states' turnout rates are not influenced by college graduation rates, per se, but instead by the percentage of whites in a state.

A. For percent whites in each state, use "whitepct_2016." Generate a scatterplot between

turnout and percent white (include the line of best fit). Intrepret your results.

B. Now estimate a multiple regression model (with both college grad rate and %white) to evaluate the critic's claim. *Interpret your results in sufficient detail (direction, size, and statistical significance).* Is the critic right or wrong? Estimate this multiple regression alongside your simple regression from 1D.

C. Now evaluate the critic's claim about African Americans in particular. Instead of using percent white, use percentage of African Americans ("blackpct_2016") and Hispanics ("hispanicpct_2016"). First, generate two scatterplots: (1) between voter turnout and Black percent and (2) between voter turnout and Hispanic percent. Include lines of best fit in each graph. Also, include state labels using the two-letter state abbreviations ("StateID"). You may need to use ggreprel. Interpret your results. Any trends you see?

D. Next, evaluate whether the results in part B necessarily mean that it's percentage of African Americans per se that is responsible for lower voter turnout in states. Estimate a multiple regression model including %Black and %Hispanic instead of %White (and continue to include college graduate rate). Interpret the results in sufficient detail (direction, size, and statistical significance). What do your results have to say about this question?

E. Interpret and compare your adjusted $R^2$ values between your models estimated in parts A and B. What conclusions can you draw?

## 3.  OPTIONAL Maps Exercise

To what extent are your results about race generalizable? Use Tigris to generate two geographic (national) choropleth maps—one for %Black and one for %Hispanic. Do the results seem generalizable or are they particular to certain parts of the country? Connect these results to your scatterplots you generated in Number 2, Part C. Draw some conclusions about this issue.