Problem Set 4: Co-Ethnic Candidates and Voter Turnout

Due 4/1

For this problem set, we will analyze data from the following article:

Fraga, Bernard. (2015) "Candidates or Districts? Reevaluating the Role of Race in Voter Turnout," *American Journal of Political Science*, Vol. 60, No. 1, pp. 97–122.

Fraga assesses the theory that minority voters are more likely to vote in elections featuring co-ethnic candidates. He analyzes data from elections to the US House of Representatives. A description of the variables is listed below:

Name	Description
`year`	Year the election was held
`state`	State in which the election was held
`district`	District in which the election was held (unique within state but not across states)
`turnout`	The proportion of the black voting-age population in a district that votes in the general election
`CVAP`	The proportion of a district's voting-age population that is black
`candidate`	Binary variable coded "1" when the election includes a black candidate (there is a co-ethnic candidate); "0" when the election does not include a black candidate (there is no co-ethnic candidate)

Question 1

Fraga analyzes turnout data for four different racial and ethnic groups, but for this analysis we will focus on the data for black voters. Load `blackturnout.csv`. Which years are included in the dataset? How many different states are included in the dataset?

Question 2

Create a boxplot that compares black turnout in elections with and without a black candidate. Be sure to use informative labels. Interpret the resulting graph.

Question 3

Run a linear regression with black turnout as your dependent variable and whether there are black candidates as your independent variable. Report the coefficient of your independent variable and the intercept. Interpret these coefficients. Do not merely comment on the direction of the association (i.e., whether the slope is positive or negative). Explain what the value of the coefficients mean in terms of the units in which each variable is measured. In addition, comment on whether we can reject the null hypothesis or not. Based on your analysis, what would you conclude about black voter turnout and co-ethnic candidates?

Question 4

You decide to investigate the results of the previous question a bit more carefully because the elections with co-ethnic candidates may differ from the elections without co-ethnic candidates in other ways. Create

a scatter plot where the x-axis is the proportion of black voting-age population and the y-axis is black voter turnout. Color your points according to candidate co-ethnicity. That is, make the points for elections featuring co-ethnic/black candidates one color, and make the points for elections featuring no co-ethnic candidates a different color. Interpret the graph.

Question 5

Run a linear regression with black turnout as your dependent variable and with candidate co-ethnicity and co-ethnic voting-age population as your independent variables. Report the coefficients and interpret them. Explain what each coefficient represents in terms of the units of the relevant variables. Comment on the statistical significance of the coefficients.

Question 6

Based on the regression model with one predictor (Question 3), what do you conclude about the relationship between co-ethnic candidates and black voter turnout? Based on the regression model with two predictors (Question 5), what do you conclude about the relationship between co-ethnic candidates and black voter turnout?

Question 6

Based on the regression model from Question 5, plot the predicted turnout (y-axis) for the range of values of a district's voting-age population that is black (x-axis). Assume that a black candidate is running for office. Be sure to also plot the 95 percent confidence interval. Interpret.