

Political Science 104 Homework #4 Due Wednesday, May 27th, at the end of lecture

To complete this homework you'll need to use SPSS. You can either work in the computer lab, or download the data from my website and work at home if you have access to SPSS at home.

This data is the same data you saw in section and for Homework #3. It consists of 11 variables observed in all 50 US states and the District of Columbia just before the 2004 Presidential election. The variables are (1) the name of the state, (2) the number of soldiers from that state killed in the Iraq War up to that point, (3) the state population, (4) the state's median family income, (5) the percentage of people in the state with a college education, (6) the unemployment rate in the state, (7) the per capita amount of federal aid distributed to the state, (8) the percentage of the state population that is white, (9) the percentage of people in the state that own their own homes, (10) the per capita energy consumption rate (in British Thermal Units), and (11) the percentage of voters in the state who voted for George W. Bush.

Use a 5% significance level and two-tailed tests when testing the following hypotheses.

Question 1: Many pundits have argued that the US in 2004 was a "50-50" nation, with support almost evenly split between the Democrats and Republicans.

- Use SPSS to test the hypothesis that the mean vote share across states for Bush was 50. Report the mean vote share for Bush, the t score of the test, and the significance level of this test.
- Would you reject or accept (fail to reject) the null hypothesis that the mean vote share for Bush in 2004 was 50? Why?

Question 2: Suppose you were interested in testing the hypothesis "States that consume more energy per capita supported George W Bush at higher rates."

- Run a regression to test your hypothesis. What is the intercept of the regression line, and what does it tell you in this case? What is the slope of the regression line, and what does it tell you in this case?
- What is the R squared of this regression line? What does the R-squared of this regression line tell you?
- Based on these regression results, would you accept or reject this hypothesis? Explain why you made this decision.

- Suppose someone criticizes your hypothesis test, and argues that it is actually income that is driving these results. This person thinks that higher income states can afford to consume more energy, and are also more likely to support George W Bush. Run another regression to test the hypothesis above, but this time include median family income as a control variable. Report the intercept, slope coefficients, and R-squared of this regression line.
- Did your original decision to accept or reject this hypothesis change once you included the control variable? Why or why not?

Question 3: Now generate your own original hypothesis about a relationship between two variables in this data.

- Run a regression to test your hypothesis. What is the intercept of the regression line, and what does it tell you in this case? What is the slope of the regression line, and what does it tell you in this case? What is the R squared of this regression line, and what does this tell you in this case?
- Based on these regression results, would you accept or reject this hypothesis? Explain why you made this decision.
- Think of a plausible control variable that you might want to include in your regression. Run another regression to test the hypothesis above, but this time including this control variable. Report the intercept, slope coefficients, and R-squared of this regression line.
- Did your conclusion about your hypothesis change once you included the control variable? Why or why not?