

## POS 3713: Assignment 7

Assigned: Wednesday, 4/12/2000

Due date: In class, Wednesday, 4/19/2000

Tutorial Session: Thursday, 4/13/2000, 9am-10:45am; and Friday, 4/14/2000, 9am-10:45am

Professors Mitchell and Lubell

The purpose of this assignment is to introduce bivariate and multiple regression. We will be using both the 1996 National Election Study and the 1995 World Survey as the basis for this assignment. Make sure you have copies of both datasets before you begin. When discussing the results of the analyses, make sure you refer directly to the numbers in the SPSS output.

### Part A. Bivariate and Multivariate Explanations for Presidential Approval

Throughout this class, we have been looking at examples of the factors US citizens use to evaluate the President of the United States. A simple model of Presidential approval might take into account three factors: the citizen's party identification, evaluation of Presidential job performance, and evaluation of Presidential personality. In mathematical form, Presidential Evaluation =  $f$  (party identification, job performance, personality). This exercise is designed to find out which of these factors has the strongest influence on Presidential approval. We will be examining 4 variables in this exercise: V960272 (*ClinTher*), V960298 (*Economy*), V960420 (*PartyID2*), and V960426 (*Leader*).

The first thing you must do is rename each of these variables to the names indicated in the parentheses. You may have already renamed some of these (e.g., *ClinTher*) from previous exercises. Second, you need to prepare each of these variables for analysis in regression. Check the frequencies for all these variables to make sure they are coded as consecutive integer scales. For example, *Leader* is already coded as a [1,4] scale where 1 = Extremely Well and 4 = Not well at all. Fortunately, only two variables need attention in this case. For *PartyID2*, recode the single respondent from a minor party into system-missing; you may have done this already in a previous exercise. For *Economy*, recode the value 4 = 3, and the value 5 = 4. Then label value 4 = Disapprove Strongly and 3 = Disapprove Not Strongly. This recode procedure essentially transforms the *Economy* variable into a 4-point scale ranging from [1,4] where 1 = Approve Strongly and 4 = Disapprove Strongly. You have already learned how to execute these data preparation procedures in previous assignments, so we will not repeat detailed instructions here. Remember, to label variable values just double click on the variable heading and click the "Labels" command button on the pop-up text box. Once the data preparation is complete, you should have 4 scales. *ClinTher* is a 100-point scale ranging from [1,100]; *Economy* is a 4-point scale ranging from [1,4]; *PartyID2* is a 7-point scale ranging from [1,7]; and *Leader* is a 4-point scale ranging from [1,4]. Check the frequencies of each variable a second time to make sure the values are correct, and also to think about what concept each variable is measuring.

### Exercise 1: Bivariate Regression

- a) Throughout these exercises, we will be using *ClinTher* as the dependent variable measuring Presidential approval, while *economy*, *partyID2*, and *leader* are independent variables. What component of the presidential performance model does each independent variable measure? In words, describe the causal relationship you expect between each independent variable and Presidential approval.
- b) Many times researchers begin a data analysis task by examining the bivariate relationships between the dependent variable and all possible independent variables, just to get a feel for the nature of the relationships. In other words, estimate the coefficients for the following three bivariate regression models:

- I.  $ClinTher = a + B_1(Economy)$
- II.  $ClinTher = a + B_1(PartyID2)$
- III.  $ClinTher = a + B_1(Leader)$

- Select the “Analyze” menu, select “Regression”, and select “Linear”. This will bring up the regression analysis interface.
  - Use the arrow key to move *ClinTher* into the “Dependent” text box, and *Economy* into the “Independent(s)” text box. Click “OK”. This estimates the first bivariate regression using *Economy* as the independent variable.
  - Repeat the above two steps using *PartyID2* and *Leader*. Your output should now have the results of 3 bivariate regression models.
- c) In symbolic form, state the null and research hypotheses for the slope coefficients in each of the bivariate regression models. Evaluate the goodness of fit of each regression model by interpreting  $R^2$  and the significance of the F-test. Based on the t-statistics, are the slope coefficients in each model significantly different from zero? In words, interpret the meaning of each slope coefficient. Based on these results, which of these variables influence Presidential approval?

### Exercise 2: Multiple Regression

- a) Now that you have gained some intuition about the relationships between the independent variables and Presidential approval, you can use multiple regression to see how all the variables work together. Use SPSS to estimate the following multiple regression model:

$$ClinTher = a + B_1 (Economy) + B_2 (PartyID2) + B_3 (Leader)$$

- Select the “Analyze” menu, select “Regression”, and select “Linear”. This will bring up the regression analysis interface.
  - Use the arrow key to move *ClinTher* into the “Dependent” text box, and *Economy*, *PartyID2*, and *Leader* into the “Independent(s)” text box all at the same time. Click “OK”. This estimates the multiple regression equation in one step.
- b) The null and research hypotheses for each slope coefficient are the same as above. However, the slope coefficients are now different, reflecting the influence of each independent variable while controlling for the influence of the others. Evaluate the goodness of fit of the multiple regression using the F-test and  $R^2$ . Based on the t-statistics, are the partial slope coefficients significantly different from zero?
- c) In words, interpret each of the slope coefficients. Do you notice any changes in the magnitude of the slope coefficients going from the bivariate regression models to the multiple regression models? Looking at the standardized partial slope coefficients, which variable appears to have the largest influence on Presidential approval? What are some possible theoretical explanations for these results?
- d) Regression equations can be used to predict levels of presidential approval for different types of citizens. Using the unstandardized partial slope coefficients estimated in the three variable model above, make two predictions about presidential approval. First, predict the level of approval for a Strong Democrat when *Economy* and *Leader* are at their mean values in the sample. Second, predict the level of approval for Strong Republican when *Economy* and *Leader* are at their mean values in the sample. To find the mean levels for *Economy* and *Leader*, you need to run descriptive statistics for these variables. For an example of how to use regression to make predictions, read section 18.6 (starting p.456) in Healey. Discuss whether or not the predictions corroborate your theoretical expectations.

## Part B: Constructing a Multiple Regression Model for Gross Domestic Product

Researchers in comparative politics spend a lot of time trying to figure out the factors that affect the economic performance of different countries. In this exercise, you will use the 1995 World Survey to construct and test your own model to explain economic performance. The main dependent variable for your model will be per capita Gross Domestic Product (GDP; called *gdp\_cap* in the dataset), which is measured in dollars. Your task is to pick three independent variables that you think provide the best explanation for the variance in Gross Domestic Product.

### Exercise 1: Construct Your Model

- a) Which three independent variables do you think will have the largest influence on per capita GDP? State the hypothesized direction of each causal relationship and explain the theoretical reasoning behind each of your selections.
- b) Write the regression equation for your proposed model in mathematical form in the same manner as the equation for Presidential approval from Part A, Exercise 2 (a) above.

### Exercise 2: Estimate Your Model

- a) Use SPSS to estimate your regression model:
  - Select the “Analyze” menu, select “Regression”, and select “Linear”. This will bring up the regression analysis interface.
  - Use the arrow key to move *gdp\_cap* into the “Dependent” text box, and move all the independent variables in your model into the “Independent(s)” text box all at the same time. Click “OK”. This estimates the multiple regression equation for economic performance in one step.
- b) Evaluate the goodness of fit of your model using the F-test and  $R^2$ . Are the partial slope coefficients significantly different from zero? Explain in words the substantive interpretation of each partial slope coefficient. Looking at the standardized partial slope coefficients, which independent variable has the largest influence on economic performance? Discuss how the results of your findings corroborate or contradict your theoretical expectations.