**POS 3713: Assignment 5**
**Assigned: Wednesday, 6/7/2000**
**Due date: In class, Friday, 6/16/2000**
**Tutorial Session: Friday, 6/9/2000, 9am–10:45am and 2:00-3:15pm in Room 66 Bellamy**

The purpose of this assignment is to introduce you to correlation and regression. We will be using the 1996 National Election Study. When discussing the results of the analyses, make sure you refer directly to the numbers in the SPSS output.

**Part A: Measures of Association for Interval Variables (Pearson's *r* correlation)**
First, we want to examine the relationship between respondents' feelings for Bill and Hillary Clinton. To what extent is someone more/less likely to evaluate Hillary Clinton favorably if they evaluate Bill Clinton favorably? In this case, we have two interval measures (the feeling thermometers. Remember, we can at least *assume* these are measured at the interval level), so we can examine the correlation (Pearson's *r*) between them.

➤ Click on "Analyze", "Correlate", "Bivariate". Move the variables "Clinther" and "Hillther" into the Variables box and then click OK. You should have already renamed and recoded these variables in previous assignments.
➤ The output will contain a correlation matrix listing both variables in the rows and columns of the table. Thus the correlation between Clinther and Hillther will be listed twice in the table, once below and once above the diagonal of the matrix, which is where the same-variable correlations equal 1.

**Exercise A:** Use the information calculated above to answer the following questions:
1. What is the correlation between the feeling thermometer scores for Bill and Hillary Clinton? Is this relationship positive or negative? Explain how to interpret the sign of this correlation in terms of these variables. Is the size of the correlation between these feeling thermometers small or large?
2. Is the correlation statistically significant? Conduct a hypothesis test by comparing the p-value (Approx. Sig.) to the value of $\alpha = .05$.

**Part B. 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. Produce a frequency distribution for each variable to see how they are coded.

Recoding V960298 (*Economy*): You will need to recode values 0, 8, and 9 as system missing and recode value 4=3 and 5=4.

➢ Click on "Transform", "Recode", "Into Same Variable"; move the variable *Economy* into the Numeric Variables box and click on the "Old and New Values" box at the bottom. Enter 0 in the Value box on the Old Value side, and then click on "System-missing" on the New Value side. Click on Add (the box will contain the statement "0 → SYSMIS" if you have done this correctly. Now enter the value of 0 in the old value box with 8, click Add and do the same for 9. Next, enter the value 4 in the Old Value box, and the value 3 in the New Value box, click Add (it should read 4 → 3). Now enter the value 5 in the Old Value box, and the value 4 in the New Value box (5 → 4). When you are finished, click on Continue and OK.

➢ Next find the variable *Economy* in the data window and double click on the column. Click on "labels"; you will need to change the value label for #4 to #3 and the label for #5 to #4. Click on Continue and OK.

➢ Click on "Analyze", "Descriptive Statistics", "Frequencies" and move the variable *Economy* into the Variables box and then click OK. If you have done everything correctly, you will have the following table. Notice that higher values on the scale imply greater *dissatisfaction* with Clinton's handling of the economy.

**CLINTON ECONOMY STRENGTH**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 1. Approve strongly | 539 | 31.4 | 32.4 | 32.4 |
| | 2. Approve not strongly | 562 | 32.8 | 33.8 | 66.2 |
| | 3. Disapprove not strongly | 251 | 14.6 | 15.1 | 81.3 |
| | 4. Disapprove strongly | 311 | 18.1 | 18.7 | 100.0 |
| | Total | 1663 | 97.0 | 100.0 | |
| Missing | System | 51 | 3.0 | | |
| Total | | 1714 | 100.0 | | |

Now, you need to recode the values 8 and 9 as system missing for the variable *Leader*. Follow the instructions above; when you click on Transform, Recode, Into Same Variables, click on the Reset button before you begin recoding this variable. If you have recoded this variable correctly, it will produce the following frequency table. Notice again that higher values imply lower evaluations of Clinton's leadership abilities.

**CLINTON STRONG LEADER**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 1. Extremely well | 249 | 14.5 | 14.7 | 14.7 |
| | 2. Quite well | 770 | 44.9 | 45.4 | 60.0 |
| | 3. Not too well | 486 | 28.4 | 28.6 | 88.7 |
| | 4. Not well at all | 192 | 11.2 | 11.3 | 100.0 |
| | Total | 1697 | 99.0 | 100.0 | |
| Missing | System | 17 | 1.0 | | |
| Total | | 1714 | 100.0 | | |

Finally, recode the values 7, 8, & 9 for *PartyID2* as system missing using the instructions above. If you saved your data set from previous assignments, you should have done this already. The correct frequency table is printed below.

**SUMMARY R PARTY ID**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0. Strong Democrat (1,1,0 in K1, K1a/b, | 329 | 19.2 | 19.4 | 19.4 |
| | 1. Weak Democrat (1,5/8/9,0 in K1, K1a/b | 334 | 19.5 | 19.7 | 39.1 |
| | 2. Independent-Democrat (3/4/5,0,5 in K1 | 233 | 13.6 | 13.7 | 52.9 |
| | 3. Independent-Independent (3,0,3/8/9 in | 145 | 8.5 | 8.6 | 61.4 |
| | 4. Independent-Republican (3/4/5,0,1 in | 183 | 10.7 | 10.8 | 72.2 |
| | 5. Weak Republican (2,5/8/9,0 in K1, K1a | 257 | 15.0 | 15.2 | 87.4 |
| | 6. Strong Republican (2,1,0 in K1, K1a/b | 214 | 12.5 | 12.6 | 100.0 |
| | Total | 1695 | 98.9 | 100.0 | |
| Missing | System | 19 | 1.1 | | |
| Total | | 1714 | 100.0 | | |

**Exercise B1: 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 (remember that higher numbers for some of these variables indicate a lower evaluation of the president). 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 B2: Multiple Regression**
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.

a) 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 $R^2$. Does the multiple regression model explain the variance in presidential approval better than the bivariate regression models (compare the $R^2$ of the four estimated models)? Based on the t-statistics, are the partial slope coefficients significantly different from zero?

b) Interpret each of the slope coefficients substantively. Do you notice any changes in the magnitude of the slope coefficients going from the bivariate regression models to the multiple regression models?

c) 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 <u>two</u> predictions about presidential approval. First, predict the level of approval for a <u>Strong Democrat</u> when *Economy* and *Leader* are at their mean values in the sample. Second, predict the level of approval for <u>Strong Republican</u> 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 you theoretical expectations.