**STATA ASSIGNMENT**

Go to the Problem Set 3 Folder in the Assignments section of Blackboard. There, you’ll find data, a codebook for the data and the journal article the data are from. Download these files and answer the questions below. Include all Stata outputs with your answers. Be sure Stata outputs are readable, clear and professional (such as a screen shot, regression table or other easily readable format).

At the end of your problem set, copy and paste your annotated do file.

**Data Exploration**

1. Read the article, esp pp. 13-39 of the article for background information.

2. Import problemset3summers2.csv file into Stata.

3. Familiarize yourself with the dataset.

4. Note that variables that end in “adv” refer to conquerors, while variables that do not refer to defenders. Not all variables are defined in the codebook, so learn about what variables mean from the article, the codebook and this problem set. Use only what you can understand, and email me with questions.

5. What is the unit of analysis? How many observations are there? How many variables are there?

Is this dataset a cross-section, time series, or time series-cross section? How do you know?

6. The variable cinc is a measure of military capacity of defenders. Among observations in the bottom decile of the cinc variable, which regions occur most frequently? Among observations in the top decile of the cinc, which regions occur most frequently? Is there anything you observe among observations in the lowest decile of of the cinc variable and in the highest decile?

7. Generate a variable avg cinc and populate all observations with missing values. Replace these values with a ‘1’ if the observation had at or above average (mean) of military capacity (cinc). Replace these values with a ‘0’ if the observation had lower than average (mean) military capacity. Do places with above average wealth have a higher polity score (polity2, meaning they are more democratic)? Do places with below average military capacity face challengers with higher rates of military expenditures (milexadv)? Report your results and be sure to discuss the confidence with which you report them. What might explain these results?

8. The author believes that conquest during war (cowwar) could be due to the population of the conquered state (popul). The author also believes that conquest during war (cowwar) could be related to shared ethnic ties between conquered and conqueror (ethnic). Specify the null and alternative hypothesis with respect to these questions. Investigate the data to identify the test best suited to answer this question. Report and interpret your results and be sure to discuss the confidence with which you report them.

9. How correlated are the variables polity2, polity2adv, milex and cinc? Are any of these relation- ships statistically significant? Substantively interpret the relationship you find most interesting.

10. Produce a scatter plot of the relationship between the polity score of the adversary (polity2adv) and the polity score of the defender country (polity2) . Include the best fit line. Interpret this figure.

11. Produce a scatter plot of the relationship between cowwar and the natural resources available in conquered places (tcresource). Include the best fit line. Interpret this figure. Is a line a good way to represent this relationship? Why or why not?

**Data Analysis**

In this section we will work with the authors’ data to develop our own models. Review the article again, especially pp. 13-39.

12. You are interested in understanding if one country’s level of democracy (polity2) affects the level of a democracy in an adversarial country (polity2adv). Investigate the way these two variables are constructed and determine which regression estimator will you use (logit or linear). Once you estimate this regression, substantively interpret your results by calculating how many standard deviations polity2adv changes as a result of a two-standard deviation increase in polity2. Ad- ditionally, explain how much polity2adv changes when polity2 moves from the first quartile to the third quartile.

13. You are now interested in conducting a multivariate analysis using the variable polity2adv as your dependent variable again. Estimate a theoretically informed model by selecting from the remaining variables in the dataset to use as your independent variables. What estimator (linear regression, logistic, t-test, cross-tabs, correlation) will you use? Why did you include these variable(s)? Are there any additional variables **not** in the dataset that you would like to include but cannot? Report the results of your model below.

14. Interpret the results of your model. Which variables are significant? Do these variables have a positive or negative effect on the dependent variable you selected? Interpret the substantive effects of each of the variables you included such that a smart person who has never taken statistics could understand your results.

15. Like the author, you are also interested in what causes conquest to occur within the context of war (cowwar) and want to use the appropriate regression estimator (logit or linear) to examine relationships between them. First, how is this variable constructed (categorical, binary, continuous, ordinal)? Which estimator (logit or linear) will you use? Estimate a theoretically informed bivariate model. Substantively interpret the results of this bivariate analysis to include size and significance.

16. You believe other variables might also be related to conquest during war. Estimate a theoretically informed model by selecting from the remaining variables in the dataset to use as your independent variables. What estimator (linear regression, logistic, t-test, cross-tabs, correlation) will you use? Why did you include these variable(s)? Are there any additional variables **not** in the dataset that you would like to include but cannot? Report the results of your model below.

17. Interpret the results of your model. Which variables are significant? Do these variables have a positive or negative effect on the dependent variable you selected? Interpret the substantive effects of each of the variables you included such that a smart person who has never taken statistics could understand your results.

18. The author uses the same dependent variable you did in 15 and 16 in their work. In Table 1, Model 6 of the main text of the paper (p. 43), he includes the following control variables: popul, ethnic, island, colony, logmilexshare, rspolity2, rspolity2adv, rspolity2int. Exe- cute the author’s model, then compare your model from Question 15 to the author’s model. Did the author include any variables that you think you should have? Is the author missing any variables? Does your model or the author’s model have a better fit? How do you know?