

IPSA Multi-Methods: Lab 2

Due on Second Day

Jason Seawright

Problem 1

Bivariate Regression in R

We will be analyzing a data set on inequality for a collection of 118 countries during the mid-1990s.

Begin by loading the data set into R, using the following command:

```
inequality <- read.csv("https://raw.githubusercontent.com/jnseawright/practice-of-multimethod/main/inequality.csv")
```

Look at the names of the variables in the data set:

```
names(inequality)
```

Does democracy cause a relative increase in economic equality? Carry out a bivariate regression, using the Gini coefficient (a measure of inequality) as the dependent variable and the Polity variable (a measure of democracy) as the independent variable. Report all the relevant results from this analysis.

Problem 2

Multivariate Regression in R

You should now expand the model by incorporating logged GDP as a second independent variable, to control for level of development as a possible confounder. The R commands generalize straightforwardly:

```
ineqlm2 <- lm(Gini ~ Polity + log(GDP), data=inequality)
summary(ineqlm2)
```

(Note that we are able to incorporate functions like log directly into the regression command.)
Do these results have causal meaning?

Problem 3

Identifying Regression Assumptions

This problem works well as a group activity.

Look up Spruyt et al.'s 2016 *Journal of Political Research Quarterly* article, “Who Supports Populism and What Attracts People to It?.” Consider in particular the set of regression analyses reported in Table 2. Taking into account both the results and the information in the text about the research design, what assumptions do you think must be made for these causal inferences to succeed? Do you find those assumptions plausible? Why or why not?

What, if anything, does the article say to justify the necessary assumptions?

Problem 4

Control Variables in Civil War Regressions

This problem works well as a group activity.

Find three quantitative articles on the causes of civil wars. Make a master list of all the control variables used in those articles, as well as which articles use each variable. For each variable, determine whether it might be a good control variable that helps eliminate confounding — but also whether it might be a bad control variable that is either an instrument or post-treatment. How close do your articles come to including the good control variables and excluding the bad ones?

Problem 5

Regression in your Research Area

Find an example of research using regression in a way that you regard as successful from your area of research. What makes this application of regression successful? What is regression used for, what assumptions are needed, and to what extent do you regard the results as credible? Prepare to discuss this analysis in class tomorrow morning.