

Problem Set 6 — Political Science 599

Due Tuesday, 6 December 2005

Instructions: Follow the same general guidelines as for Problem Set 5.

In this problem set you will perform a regression analysis relating a measure of a substantively interesting dependent variable to measures of substantively interesting explanatory variables. You can use the NES to find individual, survey-based measures of the political attitudes, behaviors, and other characteristics of individuals in the U.S., or you can use a dataset of your choosing. Complete parts (a - c) as follows.

(a) **Motivation and theory:** Develop and describe a hypothesis about a phenomenon or condition you are interested in. Your hypothesis should be such that multiple linear regression analysis will allow you to assess the evidence—either consistent or disconfirming—that your dependent variable is simultaneously and additively predicted by at least 3 causal factors.¹ In particular, consider these aspects:

- *Motivation.* Explain why you (or someone) might be interested in your dependent variable and its purported causes? Your answer might address a normative component as well as a contribution to scientific progress, i.e. the refinement of an existing area of research.
- *Theory.* Explain briefly your hypothesis about the causal explanation of your dependent variable afforded by your explanatory factors, and (as best you can) any theory it derives from. In other words, how do your variables explain why people might report having certain opinions or characteristics, or engage in specific behaviors?²

Use no more than 1 page, single-spaced for part (a).

(b) **Methods:**

- *Measures.* You'll need measures to study the phenomenon or condition of interest. Develop a measurement strategy (using the NES or a dataset of your choice) to evaluate your theory. In addition to using variables explicitly coded in the dataset you can create measures that you think somehow indicate/capture the variables of interest.³ Explain and defend the measurement strategy. What are the weaknesses of the strategy? Don't overstate the case.
- *Inference.* Explain why regression analysis might be appropriate for your investigation. Can you think of alternative methods that would allow you to study the phenomenon or condition of interest? How does regression analysis size up in terms of the goals of scientific inquiry, potential pitfalls, and practicality/convenience in terms of time and money?
 - (a) Linear model and least squares estimates. Write down your model, the relationship between the measure of your dependent variables and the measures of your explanatory variables (using clearly labeled variables or matrices). Then give the formulas for the least squares estimates of your coefficients for the coefficients and their standard errors. You'll need to find the residual standard deviation.
 - (b) Likelihood and MLEs. Write down a likelihood (using clearly labeled variables or matrices) that could be maximized to find MLEs for the regression coefficients and the residual standard deviation. What is the general form for the asymptotic standard error? Why would/wouldn't you need "optim", i.e. a program that finds extrema, to find the MLEs? Without actually computing the MLEs, compare them to the least-squares estimates.

Use no more than 3 pages, single-spaced for part (b).

¹Be wary that you are selecting variables for which measures exist in the data such that regression analysis is a reasonable way to seek evidence of a relationship. See (b) for consideration of measurement.

²You might also want to give competing theories, or other explanations that you think might be incomplete.

³As in Problem Set 5, you'll have to make sure your measures are appropriately coded, and deal with missing data.

(c) **Results:**

- *Descriptives.* Present descriptives of your measures, and bivariate plots to get an initial sense of the relationship between your explanatory variables and the dependent variable.
- *Analysis.* Carry out the proposed regression analysis using a pre-packaged regression program.⁴ Provide the appropriate documentation of your fitted linear regression model in a table, just as you would in a paper/article.
- *Diagnostics.* Give the reader confidence in the validity of your analysis. Consider and explain each of the following aspects:
 - Linearity. Show that the effects are linear (or, if they aren't, revise your model). What problem would non-linearity pose?
 - Heteroskedasticity. Determine whether the errors are heteroskedastic (we'll learn ways to deal with this in the future...).
 - Influence. Show that the effects are not being driven by one or two influential observations (or, if they are, explain why these few observations are running the show if you can). What is the problem with influential observations?
 - Specification. Fit two additional regression models, one in which you add a predictor to the model that you find to be correlated with an explanatory measure, and another in which you drop all predictors but the measure whose coefficient originally had the smallest absolute magnitude.⁵ Help the reader think about the outcome—both real and anticipated—in terms of proxy variable effects and the multicollinearity the world generates.
 - OLS assumptions. How (if at all) do the above diagnostics address OLS assumptions? Are there any that you haven't examined? What are the consequences of the failure of these assumptions for your estimates?
- *Interpretation.* Interpret your results for the reader in a way that the reader will understand. Relate the linear regression of your measures to your original hypothesis about the variables and phenomenon or condition of interest. Give examples of substantively meaningful effects. What do you conclude? Don't claim more for your results than they deserve.

Use no more than 6 pages, single-spaced, not including graphics or tables, for part (c).

⁴Hint: Don't forget the constant. Most software includes that automatically.

⁵Don't drop the constant though.