Problem Set 5 — Political Science 599

Due Thursday, November 17, 2005

Instructions: Please type this problem set single spaced. When presenting tables, do not cut and paste from statistical programs directly into your word processor. Make real tables with sensible significant digits. Embed any figures or tables into the text. Make equations, when necessary, that look nice, and that are well explained in terms of the symbols and variables used. Use numbers for equations, figures, and tables. If you have questions about how this should work, look at a recent copy of the APSR. In responding to the questions use your own words rather than quoting from books. If you quote from books, explain the quotes in your own words. Include an appendix at the end of the problem set containing the code you used to read, manipulate, and analyze the datasets. The appendix should have separate sections for each question. Someone should be able to "run" a given section of your appendix to produce the results for a given question. You may hand-write (neatly) the solutions to the math problems or use an equation editor or LaTeX. And of course, show your work.

1. **Lipset article**

   What concepts does Lipset want to measure? How does he measure them? Are these good measures? What would you do to be more certain about how good these measures are? (Lipset, Seymour Martin. 1959. "Some Social Requisites of Democracy." *APSR.* 53:1, p 69-105.)

   (Use a maximum of 1 page single spaced.)

2. **MLE / Italian gov'ts revisited**

   In the last problem set you calculated the mean rate of the fall of Italian governments using maximum likelihood and calculated the confidence interval based on the Hessian matrix. First, explain briefly why we use the Hessian matrix in the calculation of confidence intervals for maximum likelihood estimators (this will involve telling me what the Hessian matrix is). Generate samples from the distribution of the mean using the bootstrapping procedure we provide (in R). Now build a basic bootstrap confidence interval for this mean. How does the bootstrap confidence interval compare to the one you just calculated in the previous problem set? Finally compare briefly the assumptions that the intervals require, and state which interval you prefer for whatever reason.

   (Use a maximum of 2 pages single spaced excluding tables and plots.)

   Extra credit: Can you think of yet a third way to do a confidence interval, and explain this? Produce the interval (it's not hard) and compare it to the other two.

3. **ANOVA**

   - Choose two dependent variables that you'd like to understand or explain from an NES data set (say which NES data set you're using). Both variables must be ordered in some sense (so not categorical), and at least one must fall on an interval scale with at least a handful of different possible responses. Please do not use gender or respondent characteristics that are rarely explained (caused) by other social variables.
     - Describe your dependent variables are measured. How are the measures related to the concepts you'd like them to measure? Please pay careful attention to making the coding of your variables sensible (so you *need* to worry about the coding of missing values or other special codes in the NES). Recode the dependent variables if necessary to make them suitable for analysis.
     - Present descriptive statistics for your variables (but for each, *only those that are appropriate for its scale*). Note outliers. Include qq-plot as appropriate.

     (Use a maximum of 1/2 page single spaced excluding tables and plots.)
   - Choose two discrete explanatory variables from that same NES data set – variables that you think explain, cause, or otherwise influence the response, or outcome, of one or both of the variables you chose in the previous question. (You can recode/discretize variables if it makes sense to do so.) What concepts do you hope these variables measure?
     - Describe how your explanatory variables are measured. How are the measures related to the concepts you'd like them to measure? Please pay careful attention to making the coding of your variables sensible. Why would some coding schemes be unsuitable for analysis? Given an example of an unsuitable coding scheme. Recode the explanatory variables if necessary to make them suitable for analysis.

- Present descriptive statistics for your variables (but for each, *only those that are appropriate for its scale*).
- Do your dependent variables act like samples from normal distributions? Make sure you keep this in mind as you do the rest of the problem set.

(Use a maximum of 1/2 page single spaced excluding tables and plots.)

- Carry out two ANOVAs of the dependent variables you selected above, using the explanatory variables. Present and interpret these 2 tables. As you do so explain ANOVA. Demystify the method, and explain any complications that arise. Keep the concepts you're really interested in – and the relationships you are really interested in – in mind and in front of the reader. *Don't let the reader ever wonder why you're carrying out a particular analysis.* Explain what you hope to learn from the work you do. And explain what you did learn from the work you did. Keep in mind that the reader will want to know exactly what all of these numbers she's seeing on these tables really mean. So tell her.

  (Use a maximum of 2 pages single spaced excluding tables and plots.)

- Draw a random sample of 5 cases from the data set you're using. Present the frequencies on 2 of the variables you analyzed earlier, for these 5 cases. Carry out ANOVA using the independent variable as you did before (only this time, using you new and smaller data set of 5 cases). Do this work by hand and report all of the math (you can handwrite this neatly if you like). What do you conclude? What is the relationship between these results and the results you got with the whole dataset? Why? Finally, using what we have learned in class, relate ANOVA, briefly, to regression.

  (Use a maximum of 2 pages single spaced.)