Problem Set 1 — Political Science 599

Due Tuesday, 20 September 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 (i.e. graphics) 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, ask &/or look at a recent copy of the *APSR*. In responding to the questions use your own words rather than quoting from books or articles. If you quote from someone else, 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. (If you used a spreadsheet, write step-by-step instructions manually. If you recorded an interactive session, clean the log file so that it becomes a naturally executable and well-commented program file.) This appendix should have separate sections for each question. Someone should be able to "run" a given segment of your appendix to produce the results for a given question.

1. Draw two data distributions for which the standard deviation would be relatively uninformative or potentially misleading. Give real-world examples that more-or-less represent each of the 2 distributions you just drew. For what sort of distribution would the standard deviation offer more informative summary? What kind of information does the standard deviation offer for such a distribution? What are two real-world examples where this might make sense? What's one instance in which people use this sort of distributional assumption incorrectly?

2. Use Tables X and XII from Merriam and Gosnell's *Non-Voting*. What kind of probability distribution might be thought of as generating the indifference variable? Even though you don't have the actual dataset in your hands, you can still make a guess at how that variable looked in their original dataset. Provide a small example of that dataset (with no more than 5 rows), filling in some made up values that could plausibly have existed in it. Now, back to the table: Calculate the mean, mode, median and standard deviation for four variables: men's age, women's age, men's responses of general indifference, and women's responses of general indifference from Table X by hand. Show your work. What does each of these numbers have to say about what's going on among the respondents? What inferences do they allow you to draw? Does Table XII help you interpret the results? If Gosnell were evaluating his study today, what would he have to say? What would he say about the nature of his own measurements? About the connection between concepts and measures? What would he say about the sample?

3. Use the Cumulative NES File. Choose a variable that shows up in at least 4 of the years of the data set. First, justify the coding of the variable. Recode the variable if necessary. Present the frequencies of the variable. Graph the distribution of the variable for each of 4 years. Calculate the mean, mode, median and standard deviation of for that variable, for each of the 4 years. What inferences do you draw from these data about the way the variable changed over time?

4. Use the University of Virginia's web site access for the Historical Census of the United States (http://fisher.lib.virginia.edu/collections/stats/histcensus/). Draw two random samples of 3

states. Describe how you did this. Choose a variable that exists in 3 of the censuses — in one early, one middle, and one late in the series.Use these two samples (each of 3 states), plus another sample that you create by pooling the first two (a sample of 6 states), to answer the following questions: Have the states become more or less homogeneous on that variable over time? Does the answer to this question depend on which sample you use? Provide evidence. What are the mean, median, and standard deviation of the variable across the states in your three samples for each of the 3 years. On the basis of your work, what advice about sampling would you give a researcher interested in finding out whether the states have become more homogeneous over time?[1]

5. You've just been asked to help a newly democratizing country design an election study parallel to the NES. What advice would you give them about sampling? Be specific. Discuss the population of interest, the inferences one would like to draw, and the strategies the local researchers should use to draw a sample. What about measurement? Again be specific. They're starting at the beginning and need a lot of pretty detailed advice. They've never done this sort of thing before. And they don't have a lot of money. So help them spend wisely. Cite the readings to build your case.

6. From Wonnacott and Wonnacott, do problems: 1-1 (p 8), 1-18 (p 23), 2-2 (p 30), and 2-6 (p 31). Show your work.

---

[1]For this problem you don't have to show code for downloading the data file but you do have to show code for how you drew the random sample, how you read the data into your data analysis environment, and how you did the analysis.