Political Methodology Comprehensive Examination
September 2012
Department of Political Science, George Washington University

Part I: Answer all 5 questions

1. In a particular state, a constitutional referendum needs 60% of the public to vote in favor of the referendum in order to pass. A survey based on a random sample of 800 registered voters indicates 63% support (i.e., a sample proportion of .63). Construct a 95% confidence interval for the population proportion. If you had to bet based on the confidence interval, would you bet that the referendum will pass? Why or why not?

2. A friend of yours comes to you with survey data. She wants to examine the impact of race on evaluations of President Obama. She believes African-Americans are most favorable toward Obama, followed by Hispanics, followed by Whites, then Asian-Americans. Ultimately, she wants to examine differences in Obama evaluations between each racial category. In her data set, she has a "race" variable, where1=White, 2=African-American, 3=Hispanic, 4=Asian-American. She has additional independent variables that are already coded and ready to be entered into a regression model. Her question to you is: How should she operationalize, specify, and interpret the race variable in her regression model in order to examine these differences between racial groups?

3. In a series of articles, Alan Gerber, Neil Malhotra, and some co-authors have demonstrated a striking effect in the political science literature. An unexpectedly large proportion of reported t-statistics and z-statistics in published journal articles have values slightly larger than 1.65 and 1.96. Offer two possible explanations for why this might be the case. Do you think this a problem for the discipline? If not, why not? If so, what solutions could be implemented to limit the problem?

4. After presenting a paper with the main model specification being a linear regression model, your discussant objects to your interpretation of the main independent variable of interest. He argues that the variable might be endogenous, which would lead to biased and inconsistent estimates. He suggests you use matching methods to deal with the issue of endogenity. Does this make sense as a strategy? More generally, to deal with the issue of endogeneity, what other statistical analyses and tests could you conduct? What are the advantages and disadvantages of the techniques you could apply?

5. When estimating models for event counts, there are a large number of models to choose among, ranging across Poisson regression, negative binomial models, zero inflated models, hurdle models, and so forth. Given this surplus of models, how can one choose among the models? Does it make sense to just estimate them all on your data and see which one fits best? Why or why not?

Part II: Answer ONE of these two questions

6a. Discuss the key issues that make event history modeling superior to simple OLS or logit/probit when one has time-to-event data. Put another way, what aspects of duration data do event history models directly incorporate/handle that simple OLS or logit/probit models do not? Within the event history family of models, discuss and evaluate some of the issues one needs to consider when employing the parametric versus semi-parametric approach.

6b. Some would argue that the three most important considerations in the modeling of longitudinal data are:

unobserved heterogeneity, pooling, and dynamics. First, discuss what each concept means. Second, discuss why each is an important consideration when modeling longitudinal data structures. In other words, what are the consequences of NOT accounting for each? Third, discuss methods that explicitly account for each consideration. What statistical modeling specifications are available, and which specifications are most appropriate under different conditions?