

Political Methodology Comprehensive Examination
January 2010
Department of Political Science, George Washington University

Part I

1. In a large class, the correlation between midterm scores and final scores is found to be nearly .5, every term. The scores are normally distributed. Predict the percentile rank on the final for a student whose percentile rank on the midterm is:

- (a) 15%
- (b) 95%
- (c) 50%

What fundamental statistical property is at work here in the answers to (a)-(c)?

2. Much of econometrics consists of examining the consequences of and remedies for violations of regression assumptions. Consider the various tests for functional form, heteroskedasticity, and serial correlation. In words, describe the commonalities across the tests. Is there a common logic that applies to the tests? Are there any problems with this common logic?

3. Given a regression equation $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i$, what are the consequences of the following:

- (a) y_i is measured with error.
- (b) x_{i1} is measured with error.
- (c) Both x_{i1} and x_{i2} are measured with error.

If there are negative consequences of (a) - (c), what remedies exist to mitigate the effects?

4. Political scientists often confront the following situation. The outcome of interest to be explained by a statistical model is ordinal, with a small number of values (3, 4, or 5, say). A standard model to estimate in such situations is either the ordered probit or ordered logit model. Why do analysts argue that such models are preferable to OLS? Are there any situations where OLS would be preferable? What is the proper way to interpret the coefficient estimates in an ordered probit or ordered logit model?

5. Another common data structure in International Relations takes the following form. The unit of analysis is a dyad—a pair of countries. The dependent variable is dichotomous (in or out of war; in or out of some trade agreement, etc.), and the analyst has data arrayed over time. For each dyad, then the dependent variable is a string of 0s and 1s. What are some statistical problems that such data structure present? What solutions are available that account for the data structure to solve the problems?

6. Answer one of the two following questions.

(a) A colleague of yours comes to you with some questions about Time Series Cross Sectional (TSCS) data. She has minimal prior experience working with this data. She has the following questions: (1) What are the most important issues (both substantively and statistically) that I should consider when examining TSCS data? (2) What types of statistical models are available and appropriate to use to analyze TSCS data? (3) How do I know

which model is best for my data? Answer your colleague's questions.

(b) Consider the goal of estimating the probability of voting for a Republican presidential candidate (there are only two candidates running for office) where n individuals are nested within j states. At the individual level we have a single predictor, income, and at the state-level we also have a single predictor, state-level income. Formally and precisely write out the following multilevel models:

1. A varying intercept model.
2. A varying intercept and varying slopes model.
3. A varying intercept and varying slopes model estimating the between group correlation parameter ρ .

Part II

Either submit an empirical research paper along with the exam or schedule an oral exam after the written exam.