

# Political Methodology Comprehensive Examination

Department of Political Science

George Washington University

September 2007

*Instructions:* Answer 5 of the 6 questions in Part I and the data analysis question in Part II. You have five hours to complete the exam.

## Part I

1. The majority party in the Freedonian legislature (the Freedom Fighters of Freedonia, more commonly known as the FFF) holds 70% of the seats.

- (a) If members are randomly assigned to the 30 member foreign affairs committee, what is the probability that the majority party has less than a majority of the seats (i.e., less than 16 seats) on the committee?
- (b) If the legislature has 15 committees with 30 members each, and all committees are assigned by random, what is the probability that the majority party has a majority on all 15 committees?

[*Note:* For both parts, you may assume sampling with replacement. When a member is selected more than once for the same committee, she is able to name a substitute from her party.]

2. Say you estimated a linear regression model and obtained the following results:

$$\hat{y} = 10 + 5 \cdot \text{income} + 4 \cdot \text{male} + 3 \cdot \text{north} + 2 \cdot \text{male} \cdot \text{north},$$

where *income* is a continuous variable, *male* is a dummy variable equalling 1 for males, 0 otherwise, *north* is a dummy variable that is 1 for people in the north, 0 otherwise, and *male* · *north* is an interaction of the two dummy variables. What parameter estimates for  $\beta_0$ ,  $\beta_1$ ,  $\delta_1$ ,  $\delta_2$ , and  $\delta_3$  would you obtain if you estimated the following equation instead?

$$y = \beta_0 + \beta_1 \cdot \text{income} + \delta_1 \cdot NM + \delta_2 \cdot NF + \delta_3 \cdot SF + u,$$

where *NM*, *NF*, and *SF* are dummies for northern males, northern females, and non-northern males, respectively?

3. You have estimated a linear regression model, and in the course of analyzing the residuals, you conclude that the residuals are not normally distributed.

- (a) Could the OLS estimates be BLUE? Explain.
- (b) What are the implications of non-normal residuals? If they are a problem, what can/should you do to “fix” the problem?

4. 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 write out the following multilevel models: [YOUR NOTATION NEEDS TO BE PRECISE.]

- (a) A varying intercept model.
- (b) A varying intercept and varying slopes model.

(c) A varying intercept and varying slopes model estimating the between group correlation parameter  $\rho$ .

5. The latent variable  $y_i$  is assumed to be linearly related to the observed  $x$ 's through the model:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2)$$

The latent variable  $y_i$  is linked to the observed binary variable  $y^*$  by the equation:

$$y_i^* = \begin{cases} 1 & \text{if } y_i > \tau \\ 0 & \text{if } y_i \leq \tau \end{cases}$$

where  $\tau$  is the *threshold* or *cutpoint*. Formally write out the following:

(a) The probability that  $y_i = 0$  given  $x_i$

(b) The log-likelihood Function

(c) Do not assume that the variance equals 1, but instead you want to model the variance with a predictor,  $z_i$ . Write the log-likelihood function.

6. In recent years, political scientists in a variety of subfields have used game-theoretic signaling models to provide insight into decision making in the presence of asymmetric information. Discuss the basic features of signaling models, as well as several substantive applications.

## Part II

7. You are given a Stata data set, `anes2004.dta`, which contains data from the American National Election Study in 2004. You are interested in studying the determinants of the respondents' vote choice between George W. Bush and John Kerry. Your dependent variable is therefore "bush" in the data set, with "party," "ideology," "income," "education," "financial," and "religion" as possible independent variables (which you can recode if you would like). A codebook for the variables can be obtained by typing "codebook" in the Stata command window.

Analyze the data. More specifically:

(a) Estimate an appropriate model using these variables or variables derived from them.

(b) Interpret the coefficients after conducting relevant hypothesis tests. Which independent variables have significant effects on the dependent variable, and in what direction? Are the results expected?

(c) Do the independent variables explain vote choice well? By what criteria?

(d) What assumptions, if any, might this model violate? Please provide relevant diagnostic tests and any appropriate corrective action.

(e) Standardize all the independent variables by subtracting the mean and dividing by the standard deviation. Answer (a)-(c) with these new variables.

## Part III

Submit an empirical research paper along with the exam.