

Political Methodology Minor Exam
Fall 2008

If you believe a question provides insufficient information to answer, explain why you think so, state some minimal additional assumptions necessary to answer, make those assumptions, and answer accordingly. There should be enough information, but this might earn partial credit. For questions that require calculation, please show your work.

The time for the exam is four hours. This means that you have from 10 am to 2 pm on September 22. Handwritten parts of the exam must be turned into Retta no later than 3:30 on September 22. If handwritten parts are not turned in by that time, those parts will assumed to be missing from your answer.

This exam is open note and open book. You are not allowed to collaborate, however.

PART I Take approximately 2.5 hours for Part I.

PART I-A Answer all questions. The answer to this question may be handwritten.

1. Below is probability density function (pdf) for a binomial random variable. Write down the likelihood, log-likelihood, and derive the MLE for the sample proportion. Prove that the ML estimate of the sample proportion is unbiased.

$$\binom{n}{y} \times p^y \times (1-p)^{n-y}$$

PART I-B Answer Question 1 and then answer either Question 2 OR Question 3

1. A measure of the perceived prestige of 102 occupations was analyzed via ordinary least squares multiple regression; the prestige measure ranges from 3 to 97 with a mean of 47.7. Three predictors were used: (1) percentage of males who were high school graduates, (mean 53, minimum 7, maximum 100); (2) a variable for type of occupation: blue collar, professional, and white collar; (3) income (mean 42, minimum 7, maximum 81) the percentage earning \$3500 or more. The occupation variable is entered as dummy variables for professional and white collar. The researcher estimates three different models which are in Table 1.
 - (a) Interpret the coefficients for the Professional and White Collar dummy variable, and the coefficient for log Income in Model 2.
 - (b) Write down the marginal effect for the interaction between education and professionals with education held at the mean. Provide a substantive interpretation of the interaction and calculate the test statistic for the marginal effect for professionals. The variance-covariance matrix is at the end of the question should you need it.

Table 1: Regression Results For Occupational Prestige Data

	Model		
	1	2	3
Intercept	0.28 (5.9)	-41 (8.7)	-2.85 (9.87)
Education	0.9 (0.08)	0.36 (0.12)	1.0 (0.37)
Professional		16 (7.2)	43 (19.5)
White Collar		-16 (6.3)	16 (23.0)
Log Income		18 (3.01)	
Education X Professional			-0.51 (0.42)
Education X White Collar			-0.63 (0.49)

- (c) Assume that Model 1 has an omitted variable. Write down the proof for how this will affect the coefficient on education. Under what conditions is it possible that an omitted variable will not affect the coefficient for education. Elaborate all possibilities for the direction of the bias for the coefficient on education. [The proof may be handwritten].

Variance-Covariance Matrix for Occupational Prestige Model 3

	(Intercept)	education	typeprof	typewc	educ:prof	educ:white
(Intercept)	97.548496	-3.4806036	-97.548496	-97.548496	3.4806036	3.4806036
education	-3.480604	0.1373922	3.480604	3.480604	-0.1373922	-0.1373922
typeprof	-97.548496	3.4806036	379.303487	97.548496	-6.8103526	-3.4806036
typewc	-97.548496	3.4806036	97.548496	548.254153	-3.4806036	-10.2757165
educ:prof	3.480604	-0.1373922	-6.810353	-3.480604	0.1783318	0.1373922
educ:whitecol	3.480604	-0.1373922	-3.480604	-10.275716	0.1373922	0.2478819

2. Say you have a bivariate regression

$$Y = \beta_0 + \beta_1 X + \epsilon$$

Assume X is measured with normal random measurement error. In other words, $X = X^* + U_i$, where $U_i \sim N(0, 1)$. Sketch proof for the effect of this measurement error on the estimate of β_1 . Does this proof generalize to the multivariate case? [The proof may be handwritten].

3. In this model, the outcome is attitudes toward inequality in 24 emerging democracies or autocratic countries. This outcome is modeled as a function of GDP and the GINI coefficient a measure of income inequality. Figures 1 and 2 contains a number of diagnostics plots. For Figure 1 ignore the Residuals vs Leverage plots. For each plot (except the Residuals vs Leverage plot) in Figure 1, explain the plot and how it is used in regression diagnostics. Report what you conclude from each plot about this specific model. Do the same for Figure 2.

PART I-C Answer Question 1 and then answer either Question 2 OR Question 3

1. Table 2 contains a logistic regression model of voter turnout as a function of education in years (1-15, mean 12), age (18-85, mean 47), weekly earnings (0-2884, mean 158), a dummy for females, and a dummy for blacks. Calculate the change in predicted probabilities for a ten year increase in age from the mean for a white male with average education level. Why is it possible to calculate statistical significance for a coefficient in a logistic regression in the exact same way as for a linear regression model? Describe one method for calculating a confidence interval for a predicted probability? Why is such a method necessary?

Table 2: Logit Model of Voter Turnout

Education	0.27 (0.05)
Age	0.035 (0.007)
Earnings	0.019 (0.0002)
Female	0.02 (0.22)
Black	0.15 (0.43)
Constant	-12.03 (2.12)

2. What is the Independence of Irrelevant Alternatives assumption? Use conditional probabilities to sketch how it might be violated. What does IIA imply for models of unordered choice? What models might one use to model data when IIA is violated?
3. Models of civil war have become common in international relations. Table 3 contains the results from a Cox model with the outcome being the length of the civil war in years. This is modeled as a function of a series of dummy variables. The first is whether the war started from a popular revolution, whether the country is in Eastern Europe, whether the war is between an ethnic minority and majority, and finally whether

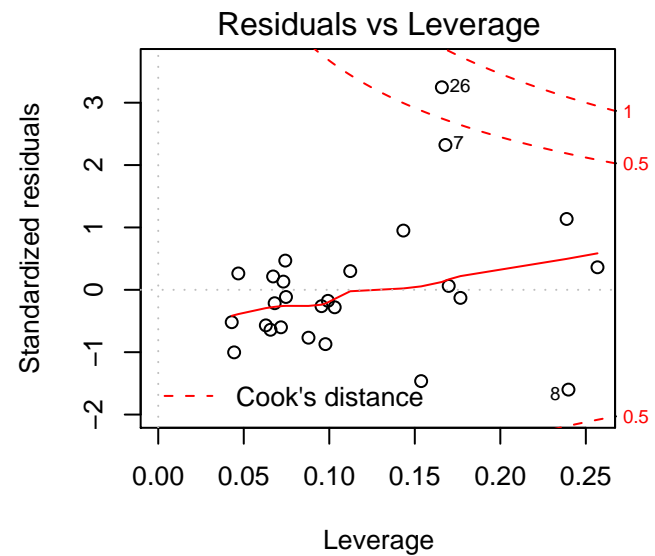
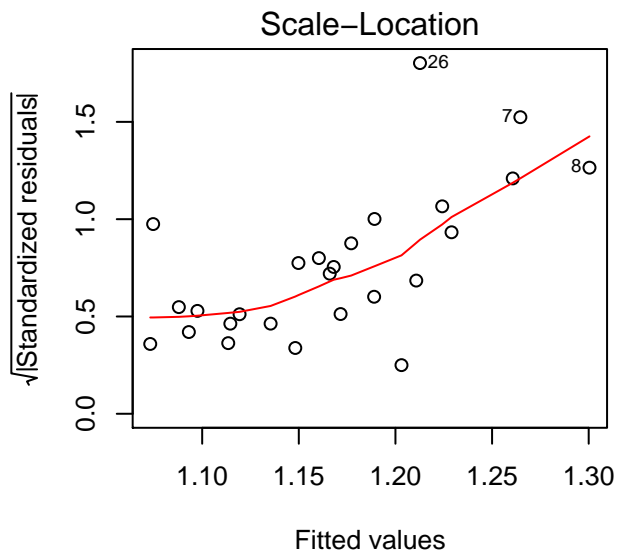
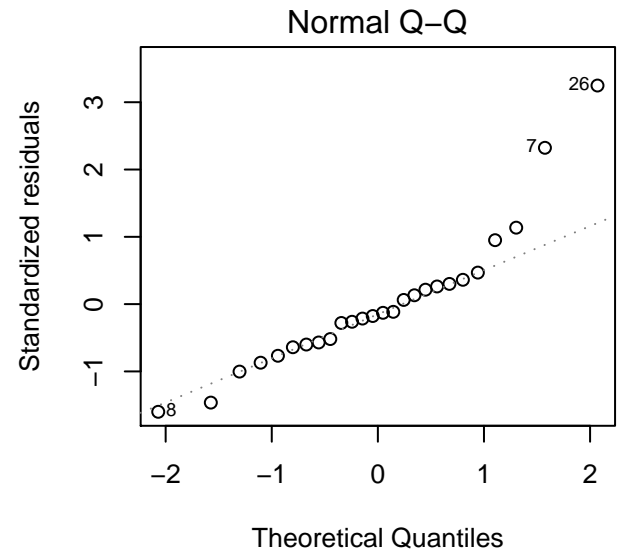
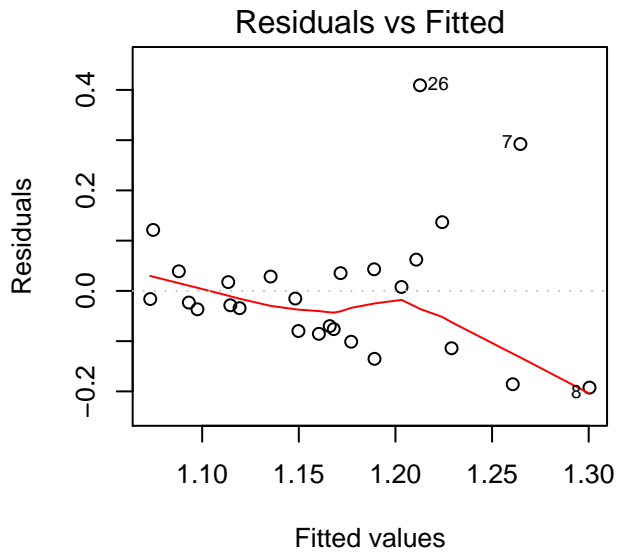


Figure 1: Diagnostic Plots For Attitudes Toward Inequality

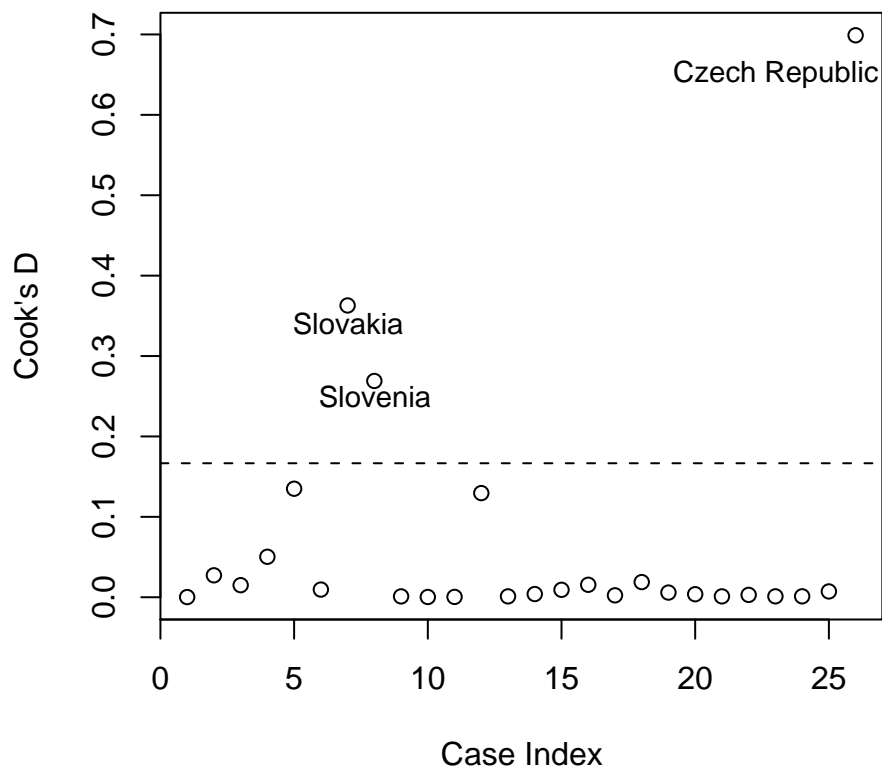


Figure 2: Plot of Cook's D Against Case Index

the rebels receive financing from the drug trade. Use hazard ratios to interpret the variable for popular revolutions and the drugs variable. What is the key diagnostic for Cox models?

Table 3: Cox Model of Civil War Duration

Popular Rev.	1.38 (0.27)
Eastern Europe	1.22 (0.34)
Ethnic War	-1.33 (0.40)
Drugs	-1.06 (0.43)

PART II. Answer the question below. Take no more than 30 minutes for this section.

1. Within positivist social science a debate has emerged between scientific realists who think the surest road to useful knowledge about politics is through the identification of causal mechanisms, which may or may not travel to other cases; and empiricists who think it is through the identification of lawlike generalizations, which are necessarily much broader in scope. Perhaps ideally we would want both kinds of information, but since different methods and logics of inquiry follow from the two approaches individual social scientists have to choose between them in their research projects. Compare and contrast the arguments made by each side in this debate about why their approach should be favored. On balance, what is your view, and why?

PART III. Answer two of the questions below. Take no more than 60 minutes for this section.

1. Matching has developed as an alternative to regression models for the estimating of causal effects. What assumptions underlie a matching analysis? How do those assumptions compare to the usual regression model assumptions? What role does sensitivity analysis play in an analysis based on matching? Assume you have conducted a sensitivity analysis with $\Gamma = 2$ and the p -value is now .26. What does this imply and how would you interpret this?
2. What is balance and what role does it play in a matching analysis? Why have some authors argued that it is not enough to compare the means across treated and control when assessing balance? Besides comparing means what other methods of comparison should an analyst use to assess balance? What role do p -values play in assessing balance and how does that differ from the p -value from classical hypothesis tests? What strategies can an analyst use to achieve better balance?

3. For this question use the following estimated error correction model:

$$\Delta y_t = -.75y_{t-1} + .25\Delta x_t + .10x_{t-1} \quad (1)$$

Calculate the total long run multiplier. Calculate the short term effect. Now rewrite the above model as an auto-distributed lag (ADL) model. How would you calculate these same quantities using the ADL?

4. A key component of time series analysis is diagnosing the properties of the dependent variable. Outline a strategy for testing whether variables are stationary, integrated, or fractionally integrated. Discuss the tests for each type of property. What models are appropriate depending on what you find?
5. Explain Krosnick's application of the satisficing notion to the evaluation of survey questions, and also explain Tourangeau's four-component theory of survey response. Would you consider these two approaches to be opposing one another or complementing one another, and why? Using these theories, which types of survey questions are least likely to obtain valid answers, and what are some of the ways to handle those problems?
6. The standard advice with surveys is to maximize response rates, but several recent authors argue that it is not important to do so. Explain the evidence that high response rates are not necessary, as well as the argument that a high response rate is essential. In light of this debate, to what extent would you suggest that researchers should invest in maximizing their response rate when they contract for a survey? What might the budget for a survey spend money on to reduce survey error, if maximizing response rates is not essential?
7. Figure 3 contains a simultaneous move game in normal form. Do the following for this game:
- Draw an extensive form representation of this game.
 - What are the Nash equilibria if $x = 0$? (Consider both pure and mixed strategies).
 - What are the pure-strategy Nash equilibria if $x = 2$?
 - What are the Nash equilibria if $x = -2$? (Consider both pure and mixed strategies.)
 - Characterize the pure-strategy Nash equilibria for all possible values of x between negative infinity and infinity.

		Player 2	
		c	d
Player 1	C	$(x, -x)$	$(0, 0)$
	D	$(1, 1)$	$(x, -x)$