

Intermediate Social Statistics: Hilary 2010

Homework 1

Raymond Duch

Nuffield College Oxford University

www.raymond Duch.com

1 Homework Questions

Question 1

I have estimated a logistic model to explore how judicial selection procedures impact court rulings for or against the governor when he or she is a party in the case (governors are named in many cases). My main independent variable is whether the judge is elected or appointed. I also know whether the judge is the same party as the governor named in the case.

Data: 1000 rulings on cases in which governors are named as a party in the case. Drawn from 42 states over 10 years.

Dependent Variable: coded 1 for a ruling for the governor and 0 for a ruling against.

Independent variables: Elected = 1 if judge was elected, 0 if appointed Party = 1 if judge is same party as governor 0 if otherwise

Table 1: Results

Result	Coefficient
Elected	.07
Party	1 .61
Party*Elected	.21
Constant	.01

1. What is the stochastic component of this model?
2. What is the systematic component of this model?
3. What is the difference in the expected probability of ruling in the governors' favor between two judges who are both from the governor's party but one of whom is elected and the other appointed?
4. What is the difference in the expected probability of ruling in the governor's favor between two judges, both of whom are elected, but one who is from the governor's party and one who is not?
5. Comment on the strengths and weaknesses of this model in this situation

Question 2

A five member committee votes 3-2 in favor of a proposal. Assume voting is independent. Let p be the probability that a committee member votes for the proposal.

1. We have no information with which to distinguish committee members (in the language of Bayesian statistics, we'd say that the committee members are exchangeable, but I digress). What is the maximum likelihood estimate (MLE) of p , the probability that any particular committee member votes for the proposal?
2. What is the log-likelihood of $p = .5$? Compare this value of the log-likelihood function with that attained at the MLE with a likelihood ratio test. What does this say about the plausibility of $H_0 : p = .5$?
3. How would your conclusion about the plausibility of $H_0 : p = .5$ change if we observed
 - (a) i. a 10 person committee splitting 6-4 in favor of the proposal?
 - (b) ii. a 50 person assembly splitting 30-20 in favor of the proposal? i.e., what is happening to the likelihood function and/or the log-likelihood function in these cases relative to the case of a five person committee? In particular, what is happening to the 2nd derivative of the log-likelihood function in the neighborhood of the MLE?

Question 3

Download the file `nagler.asc.dta` from my web site (www.raymond Duch.com). This file contains 98,857 cases (welcome to large n research!) from the 1984 Current Population Survey, analyzed by Jonathan Nagler in two articles: *The Effects of Registration Laws and Education on Voter Turnout* *American Political Science Review*, 1991, 85:1393–1405; *Scobit: an alternative estimator to logit and probit* *American Journal of Political Science*, 1994, 38:230–255. The data in the file comprise the following variables (in column order): `turnout` 1 if the respondent reports turning out to vote in the 1984 presidential election, 0 otherwise. `educ` 1 for 0-4 yrs education; 2 for 5-7 yrs; 3 for 8 yrs; 4 for 9-11 yrs; 5 for 12 yrs; 6 for 1-3 yrs college; 7 for 4 yrs college; 8 for 5+ yrs college `age` age of respondent, in years `south` 1 if respondent lives in the South, 0 otherwise. `govelec` 1 if a gubernatorial election coincided with the presidential election `days` number of days before election day that voter registration closes in the respondents state The following questions ask to you to estimate a series of logistic regression models. Construct a publication-quality table with the parameter estimates and standard errors for each the models, along with some summary information (e.g., goodness-of-fit, deviance, etc).

1. Estimate a logit model predicting turnout with the predictors `educ` and `age` and the square of each of these predictors. Provide a brief write-up of the parameter estimates (i.e., assess statistical significance and substantive implications) and the goodness-of-fit of the logistic regression model.
2. How many unique predicted probabilities are produced by this model? Explain how you derived your answer.

3. Compare the predicted probabilities from the logit model with the corresponding predicted probabilities from a probit model. How and why do they differ, if at all? Is there any statistical basis for preferring logit over probit or vice-versa?
4. Augment your logit model from the first part of this question with the following additional contextual predictors: south, govelec, and closing, and interactions between the two education variables (educ and educ2) and the closing date variable (i.e., make the effects of closing date quadratically conditional on the categorical education measure). Discuss the estimates and goodness-of-fit of this model in contrast with those obtained from the model for the previous question. Report a likelihood ratio test of the joint significance of the new predictors.
5. Using the estimates from the second model, plot the implied coefficient for closing as a function of education, given the interaction effects estimated above. Overlay 95 percent confidence intervals around the point estimates. Offer a substantive interpretation of what this plot reveals.
6. Using the estimates from the second model, consider a hypothetical nonsoutherner, in a state without a gubernatorial election, who has 12 years of education and has the median age of a non-southerner with 12 years of education. Plot the predicted probability of turnout for this person, as the closing date requirement varies over the range of closing date requirements observed in non-southern states. Overlay 95 percent confidence intervals around the point estimates.
7. Using the estimates from the second model, consider a hypothetical nonsoutherner, in a state without a gubernatorial election, who has 5+ years of college and has the median age of a non-southerner with 5+ years of college. Plot the predicted probability of turnout for this person, as the closing date requirement varies over the range of closing date requirements observed in non-southern states. Overlay 95 percent confidence intervals around the point estimates. Briefly compare the answers from this question with those from the previous question.

Due in class Tuesday, Week 5.