

Political Science 150C/350C
Spring 2007
Problem Set 2

1. Consider the data frame `nes96` in the R package `faraway`. This data frame consists of 944 responses to the 1996 National Election Study. A seven point party identification measure is in the data frame as variable `PID` (an ordered factor). Model the ordinal responses as a function of income and levels of education. Fit a series of models to the data: a null “intercept-only” model, a model with income as a predictor, and a model with income and levels of education as a predictor. Use likelihood ratio tests and other goodness of fit criteria to compare the models. For your preferred model, generate a graph showing how the predicted probabilities of specific party identification categories change as a function of income, for one or two different values of the education variable. Provide a paragraph or two of supporting text detailing what strikes you as interesting/compelling in the data, as revealed by the model fitting and/or the predicted probability graphs etc.

Notes:

- The `income` variable is given as an ordered factor. Before using the variable in analysis, convert this to a continuous variable by assigning the mid-point of each income category (in thousands of dollars) to the corresponding level of the ordered factor. You can do this with the following R commands:

```
inca <- c(1.5,4,6,8,9.5,10.5,11.5,12.5,13.5,14.5,  
         16,18.5,21,23.5,  
         27.5,32.5,37.5,42.5,47.5,55,67.5,82.5,97.5,115)  
nes96$income <- inca[unclass(nes96$income)]
```

- The `educ` variable is also an ordered factor. Convert this to a regular factor in R with the command:

```
nes96$educ <- factor(nes96$educ,ordered=FALSE)
```

- I will supply R code that will generate (1) `hitmiss` tables and (2) McFadden’s pseudo- r^2 for objects of class `polr` (and classes `glm` and `multinom`, for that matter). Watch my web site and/or your e-mail.

2. **Download** the file `nomocc2.dta` from my web site. This file is in `stata` format, and you’ll need the `foreign` library in R in order to be able to read it. The file contains 337 observations from the General Social Survey. The outcome of interest is occupational type, with (unordered) categories, “menial”, “blue collar”, “craft”, “white collar” and “professional”. The predictors are

- `white`, 1 if white, 0 otherwise

- `ed`, years of education
- `exper`, an estimate of the possible number of years that the respondent could have been in the workforce (age minus years of education minus 5).

Analyze these occupational attainment data via multinomial logit analysis, so as to address the following questions in a short write-up (say, two to three pages).

- Leaving aside other the predictors, does the distribution of occupational attainment vary by race?
- After controlling for education and experience, does race plays a role in determining occupational type?
- What is the role of education in determining occupational type; that is, how does the predicted probability of different levels of occupational attainment vary by education, net of the effects of other predictors?
- Are the effects of education are conditional on race? (i.e., does the same marginal change in levels of education translates into different levels of occupational attainment, depending on whether the subject is white or non-white?)
- Based on these data and your modeling, does it seem that education and experience help equalize racial differences in occupational attainment?
- How well do your models fit the data?

Include a table of coefficients and standard errors for your preferred model, along with any supporting tables demonstrating model comparison tests and/or goodness of fits. Your write up should also include some graphs demonstrating marginal effects, where appropriate.

Due in class, Wednesday, May 2, 2007.