

PS206 Homework #4 (Due Tuesday, May 11th)

Listed under Homework #4 on our class page is a tab-delimited text dataset called “ps206hw3.txt”. This data is the same as the data used for Homework #3, and is from a survey conducted just before the 1992 U.S. presidential election. It contains 678 observations on five variables – vote (0 for Bush, 1 for Clinton), education (a 7 point scale with higher numbers meaning more education), age (in years), gender (0 for men, 1 for women), opinions on limiting imports (0 for opposes limits, 1 for favors limits), and party identification (0 for Democrat, 1 for independent, 2 for Republican). Download the data and do the following three problems:

(1) This first problem shows you how to estimate a likelihood-ratio test. Suppose we wanted to estimate a model examining the effect of gender and education on vote choice, but someone else thinks age and opinions on import limits should also be included. To perform a likelihood-ratio test comparing these two models do the following.

In Stata:

- (1) Estimate the unconstrained model. (Which one is the unconstrained model?)
- (2) Type `estimates store A`. This will save the log-likelihood of the unconstrained model.
- (3) Estimate the constrained model.
- (4) Type `estimates store B`.
- (5) Type `lrtest A B`.

In R:

- (1) Estimate the unconstrained model and save the results as usual. (Which one is the unconstrained model?)
- (2) Estimate the constrained model and save the results as usual.
- (3) Type the following, assuming that you saved the constrained model as “mod1” and the unconstrained model as “mod2”: `anova(mod1, mod2, test="Chisq")`. The Analysis of Deviance table will print to the screen; the fifth column (“Deviance”) is the likelihood ratio test statistic and the sixth column is the p-value.

Interpret the results of this likelihood-ratio test. Don’t forget these models should have the same number of observations. In Stata, you might need to use the “if” command to estimate the constrained model omitting the data missing from the unconstrained model – for instance, you would type `logit y x1 x2 x3 x4 if x3!=. & x4!=.` to estimate a logit only in cases where x3 and x4 were not missing.

In R, throw out cases with missing data using the `na.omit` command, saving the listwise deleted set as a new data set in R, e.g. `mydata.nona <- na.omit(mydata)`, where “mydata” is the name

of your data set.

(2) Do the same calculations as above, except comparing a model that looks at the effects of gender and education on vote choice with a model that looks at the effects of gender, education, and age on vote choice. Interpret the results of this likelihood-ratio test.

(3) Now estimate either a logit or a probit using vote choice as the dependent variable, and all other variables as independent variables. Then do a Wald test to see if the coefficients on limiting imports and party identification are equal. In **Stata** you can simply type `test limit_imports = partyID`.

Wald tests in **R** are a bit more involved. After running your logit or probit, download a package that contains some nice code for Wald tests by typing `install.packages("aod")`, and then typing `library("aod")`. To run a Wald test in **R** you'll need to specify a restriction matrix that indicates that you're setting two coefficients to be equal. In this case, you would type (assuming "partyID" and "limit_imports" were the 4th and 5th independent variables in the model) `Q <- matrix(c(0,0,0,1,-1),1,5,byrow=T)`. Multiplying this matrix by the coefficients will give you the coefficient on "partyID" minus the coefficient on "limit_imports." We then set this sum equal to something – in our case since we're testing to see if these coefficients are equal, we'll set the sum to zero. Do this by typing `R <- 0`. Then, assuming we've saved the covariance matrix of our coefficients as "covmat" and the coefficients as "coeffs", we can then run our Wald test by typing:

```
wald.test(Sigma=covmat, b=coeffs, L=Q, H0=R)
```

Interpret the results of this Wald test.