

## PS206 Homework #5 (Due Tuesday, May 11th)

Listed under Homework #5 on our class page is a tab-delimited text dataset called “ps206hw5.txt”. This data is from a survey conducted just before the 1992 U.S. presidential election, and contains 1650 observations on five variables – opinions on changes in personal finances over the last year (-1 for got worse, 0 for no change, 1 for got better), age (in years), a 7-point scale for education (with higher numbers indicating more education), a dummy variable for gender (0 for men, 1 for women), and a 101 point “thermometer” rating for George Bush (Senior). Download the data and do the following four problems:

(1) Estimate an ordered probit model for examining the effects of gender, age, education, and thermometer rating for George Bush on an individual’s opinion about the state of their personal finances.

The command for an ordered probit in **Stata** is `oprobit`. In **R**, you can estimate an ordered probit using the command `polr`. First, though, you need to load a library called **MASS** (it should have come installed with your version of **R**): just type `library("MASS")`. To estimate an ordered probit with dependent variable “y” and independent variables “x” and “z” from dataset “mydata”, type: `polr(y ~ x+z, data = mydata, method="probit")`. You can feed this to the summary command as usual. To instead estimate an ordered logit model, replace “probit” with “logistic” in the prior command. Note: before you estimate the model, make sure that **R** knows which variables are categorical using the command `factor`.

Interpret the results of this ordered probit model as far as you can.

(2) Now we will look at some predicted probabilities. As in Homework #3, we will first calculate these predicted probabilities the long way. Come up with a hypothetical individual, and calculate the probability of this hypothetical individual being in each category of the dependent variable. Recall that the **Stata** command to calculate the CDF of a normal and save it as a variable called “p1” is `generate p1 = normprob(X)`, where “X” is the argument of the function (in this case, the equation with the estimated coefficients and hypothetical values of the independent variables). In **R**, you calculate the standard normal CDF using `pnorm`, and you can save the results of the calculation using `p1 <- pnorm(X)`. Remember, you have three different probabilities to calculate for this hypothetical individual. Describe what you find.

(3) Now examine how changes in age affect the probability of being in each response category for this hypothetical individual. You might want to look at the probabilities that occur at the minimum and maximum values for this variable, and compare them to the probabilities we see for the mean value. Describe what you find.

(4) Now we will calculate predicted probabilities with standard errors. For **Stata** you should use *Clarify*, while for **R** you will want to modify the code I provided for Homework #3 (substituting the `plogis` command for the `pnorm` command so that you use the CDF of the logistic rather than the normal distribution).

Using the same dependent and independent variables as in problem 1, estimate an ordered logit model. Then calculate the changes in the probability of each opinion on changes in personal finances as your hypothetical individual moves from the minimum to the maximum on the thermometer scale for George Bush, holding all else constant. Describe what you find.