Homework 2

Econ 322: Econometrics

Spring 2019

Prof. Amanda Agan

Upload to Sakai by **Friday May 3 Uploaded to Sakai by 11:55pm**

Please upload one .pdf document with your responses**. Name that document FirstnameLastname\_Econ322HW2.pdf (4 points).** Also upload your .R command file separately (6 points)**. Failure to follow the naming convention or upload a .pdf will result in the loss of points**

You may work in groups of 2-3 students, but you must do the write-up on your own. ***Put names of all group members on Homework*.** Groups may not turn in identical write-ups.

**Part I: Presidential Voting (40 points)**

This data is based on a survey of voting behavior from the 2000 presidential election – votingdata.csv. The codebook is important and also available for download with the assignment on Sakai.

1. Generate a new variable
   1. Generate a new variable high\_income =1 if the person’s family income is above the 68th percentile (income=4 or income=5), high\_income=0 otherwise
2. Run a regression of voting Republican on your new high income variable – use probit. In your table report the marginal effect at the mean of the independent variables! (see Class 20 slides and the programming resources below for how to do this)
   1. Why are we using probit? Why do we report marginal effects and not the direct coefficients?
   2. How do you interpret this marginal effect?
3. Re-run this regression but add age, female, and race (use otherrace as the reference group) to your probit model. Comment on the effect of each of these variables on the probability of voting Republican and whether they are significant. Be sure that after your estimation you estimate the marginal effects at the means of the independent variable to report in the table.
4. Use this data to tell me something else interesting about the determinants of voting republican in the 2000 presidential election (make this a 3rd regression), and interpret your new marginal effect of interest for me.

Create a table that shows these results. Each regression should be a different column – so you should have 3 columns of results. **YOU SHOULD REPORT MARGINAL EFFECTS, NOT THE COEFFICIENTS.** Put standard errors in parenthesis under the marginal effects. Put appropriate stars on the coefficients to indicate statistical significance (\*\*\* for p<0.01, \*\* for p<0.05, \* for p<0.1). Include the number of observations. You need to add an appropriate title and table notes. Table should have actual word NOT variable names. [Note – for this problem, do not worry about heteroskedasticty-robust standard errors. We didn’t really get into this in class, here is an interesting blog post about it:

<https://davegiles.blogspot.com/2013/05/robust-standard-errors-for-nonlinear.html>. You also do not need to include the R^2. Again, it’s a bit more complicated in the probit model]

**Programming Resources**:

* On creating new categorical variables: <http://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/R/R-Manual/R-Manual4.html>
* To get marginal effects in R at the means, install package mfx [install.packages(“mfx”)], then library(mfx). Then you can use probitmfx command, probitmfx(depvar ~ indvar, data=mydata)

**Part II: Determinants of the State Suicide Rate (50 points)**

The following questions are based on a state level panel data on the suicide rates from 1990-2000 and was originally used in the paper “Are Mental Health Insurance Mandates Effective? Evidence from Suicides” *Health Economics* (2006) by Jonathan Klick and Sara Markowitz. The codebook is important and available to download with this assignment from Sakai.

1. Run a regression of the adult suicide rate on percent unemployed, per capita state income, percent living in rural areas, percent of population with a college degree, and percent of uninsured
   1. How do you interpret the coefficient on the unemployment rate?
   2. Are you surprised by the sign of the coefficient?
2. Re-run the regression but include both state and year fixed effects. (see programming resources below and lecture notes from Class 25)
   1. Which of your coefficients changed sign and significance from the regression in (1)
   2. What are potential explanations for these changes?
3. Re-run your regression but now also include the available religion variables
   1. How do you interpret the coefficient on percent southern Baptist
   2. Are the religion variables jointly significant? How do you know? (what is the value of the statistic you used? And the p-value?)
4. Use this dataset to tell me something else interesting about the suicide rate, be sure to include state and year fixed effects. Make this a 4th regression and interpret 1 or more coefficients for me. Do you think that this effect you estimated is causal? Do we have more confidence in the causal interpretation than if we ran a cross-sectional regression that used data in a single year only? Why or why not?
5. You used heteroscedasticity-robust standard errors, do you have reason to believe these standard errors may be incorrect? Why or why not? **Extra Credit:** rerun your regression in (3) with the appropriate standard errors and show me, did anything change?

Create a table in that shows these results. Each regression should be a different column – so you should have 4 columns of results. Put heteroskedasticity-robust standard errors in parenthesis under the coefficients. **Do NOT show the coefficients on the fixed effects**. Instead, include a row called “State Fixed Effects” and indicate either No or Yes in each column for whether those are included, same for year fixed effects. Put appropriate stars on the coefficients to indicate statistical significance (\*\*\* for p<0.01, \*\* for p<0.05, \* for p<0.1). Include the adjusted R^2 and the number of observations. You need to add an appropriate title and table notes. Table should have actual word NOT variable names.

**Programming Resources**:

* See the R file I went over in class on joint hypothesis testing
* Fixed effects in R (p. 6-11): <https://www.princeton.edu/~otorres/Panel101$.pdf>
* <https://rpubs.com/rslbliss/fixed_effects>
* See R command plm: install.packages(“plm”), library(plm)

**Additional Notes:**

* Your code should be saved in a .R file as good practice. This code should have comments at the top that include your name, and that this is code for Econ 322 Homework 2. Please turn in this .R file and upload as a SEPARATE file onto Sakai. This .R file should have all the commands you actually used to run your homework. I.e. if I opened the .R file and ran it (after of course changing the directory) then I would be able to see all the variables you created, regressions you ran, and tests you did.  **A CORRECT R FILE WILL BE PART OF THE GRADE THIS TIME. IT MUST INCLUDE THE read.csv COMMAND AT THE BEGINNING and have appropriate comments**
* You may work in groups of up to 3 students on the code**; however each student must turn in their *own* write-up**. The names of your collaborators should be included in your document.
* Your tables should be “professional” and include actual words as variable names, not “R names”, an appropriate title, and helpful table notes (i.e. what do the stars mean? What is in parenthesis? What is the omitted category for any set of mutually exclusive dummy variables?
* In your table, round each number to have only 3 digits after the decimal

**Sample Table Format (where the last two rows are only included for the table in Part II. And In Part I you should report marginal effects, not coefficients). Total Num of Columns are determined by the Part as well, this is just a template**

TITLE

|  |  |  |  |
| --- | --- | --- | --- |
|  | (1) | (2) | (3) |
| Variable 1 | Coefficient1  (std error1) | Coefficient1  (std error1) |  |
| Variable 2 | Coefficient2  (std error2) | Coefficient2  (std error2) |  |
|  |  | Etc… |  |
|  |  |  |  |
| Constant | Constant1 | Constant2 |  |
|  |  |  |  |
| (leave blank line above # of obs) |  |  |  |
| Number of Observations | N1 | N2 |  |
| Adj R2 | [R21] | [R22] |  |
| State Fixed Effects | No | Yes | Yes |
| Year Fixed Effects | No | Yes | Yes |

**Notes: [**Any important Notes to help the reader understand the table]