

Project 3: Library Card Data

McCourt School of Public Policy, Georgetown University

Overview

This project will focus on binary dependent variables. In other words, the y variable only takes value of 0 or 1.

In quant class, you'll cover several approaches to regression with binary dependent variables. LPM, Probit, Logit

Week 1:

Key Ideas:

- commands for estimating LPM, probit, and logit
- generating predicted values after these commands

Key Commands / Concepts:

- regress, probit, logit
- twoway line, scheme(s2mono)
- predict, pr
- margins, post

Questions

3.1 Download and open data

- The data set this week is named card.dta and is available through Blackboard.
- Download and save this data set.

3.2 LPM

- Consider the simple bivariate relationship between a person's education and the probability that they have a library card.
- Estimate a linear probability model with `library card` as the dependent variable and `education` as the independent variable.
- Note: There is no special command for a linear probability model. Just use basic linear regression.
- Use the predict command to generate a new variable with predicted probabilities.
- Label this new variable, "LPM".
- Create a twoway line graph of the predicted variable against education.

- Note: twoway line connects the points of your data in the order that they appear in your data set.
- If your graph looks like a big scribble, try sorting your data: `sort educ.`

3.3 Probit

- Repeat the steps above using a probit regression.
- Remember to give your predicted probability variable a different name, and label it "Probit"

3.4 Logit

- Repeat the steps above using a logit regression.
- Remember to give your predicted probability variable a new name, and label it "Logit"

3.5 Visual Comparison

- Create a twoway line graph plotting each of the three predicted probabilities against education.
- All three line plots should be on the same graph.

3.6 Graph options

- Add the options `yline(0)` and `yline(1)` to add horizontal lines to your graph at $y=0$ and $y=1$.
- Notice the logit and probit predicted values will never cross the 0 or 1 line, unlike predictions from the Linear Probability Model.
- Suppose you wanted to print this graph out in black and white to include on a problem set.
- Without color, you will not be able to identify the lines for logit vs. probit.
- You can create a graph in black and white using the `scheme` option.
- Add the option `scheme(slmono)` to redisplay your graph in black and white.
- See `help scheme` for other schemes you can apply to your graphs.

3.7 Graph export

- Save your graph as an image, as if you are including it in a problem set.
- You can also do this from your do-file using the command `graph export.`

3.8 Average predicted value

- Find the mean of the predicted probabilities of `libcrd14` that were generated from the logit regression results.
- Another way to calculate this number is using the `margins` command.
- Re-run (or replay) your logit regression to make sure it is the most recently-run command.
- run the command `margins` to calculate the mean of the predicted probabilities from the previous regression.
- Is the result the same as the value you found previously?

- Can you explain this result? Hint: Pay attention to the number of observations used in each calculation.
- Can you run a summarize command that produces a result matching that from the margins command?

3.9 Margins, post

- The current output from margins is not very interesting.
- You might not be interested in putting these results into a nice table.
- But, we will be generating more interesting and useful results using the margins command.
- At some point, you will want to use outreg2 on the results from margins.
- Try running outreg2 after margins.
- Normally, margins does not overwrite the most recent estimation commands, so you just get the logit results.
- You can change this behavior by adding the `post` option to the `margins` command.
- Try adding the `post` option and rerunning your margins command and `outreg2` command.

3.10 Margins, at

- Run a final margins command, but use the option `atmeans`.
- How does this calculation differ from the previous one?
- This is a single predicted value, calculated at the mean value of education.
- The previous calculation was the mean of the predicted values at observations value of education.
- Challenge: Can you produce the predicted value of `libcrd14` at the mean value of education without using the margins command?