Project 3: Library Card Data

McCourt School of Public Policy, Georgetown University

Overview

If you've covered Probit and Logit in class, you've learned that you can't interpret coefficients the same way as OLS coefficients. Today, we will work with tools that people use to interpret results after Logit and Probit. These include predicted values, odds ratios, and marginal effects. The margins command is central to producing these results. We will work with the margins command and see how it interacts with factor variable notation.

Week 2:

Key Ideas:

- predicted probabilities after logit/probit
- using margins after logit/probit
- producing odds ratios after logit
- use of factor variables with margins

Key Commands / Concepts:

- probit,
- normprob()
- logit
- invlogit()
- margins , at()
- marginsplot
- margins , dydx()
- margins , post
- logistic
- logit, or
- outreg2 , eform
- help fvvarlist

Questions

3.11 Predicted Probability at specific values

- Run a probit regression of libcrd14 on educ and IQ.
- You can calculate predicted values after probit using the `normprob()` function (or `invlogit()` for logit)
- For example, what is the predicted probability of having a library card for someone with average IQ and 12 years of education?
- There are several ways you could do this. Three ways are demonstrated below.
- Choose one of these three methods to calculate the predicted probability for educ=10 and IQ at the mean value.

```
display normprob(-2.074503 + .1095198*12 + .0127187*102.4498)
probit, coeflegend
display normprob(_b[_cons] + _b[educ]*12 + _b[IQ]*102.4498)
margins , atmeans at(educ=12)
```

3.12 margins, at()

- The `margins` command makes it very easy to calculate predicted probabilities for different x-values.
- Example1: `margins , atmeans at(educ=(8 10 12 14 16))`
- Example2: `margins , atmeans at(educ=(8(2)16))`
- To see how you can specify different values of X-variables using at(), go to the help page for `margins`.
- The section you are looking for begins with the example: `at(age=20) fixes covariate age...`
- Hint: Use `ctrl+f` (or `apple+f`) to search within the help page.
- Calculate predicted probabilities at values of IQ from 50 to 150 in steps of 10. Hold educ at 12 years.

3.13 marginsplot

- You can use `marginsplot` after `margins` to graph your predicted probabilities with confidence intervals.
- Add the command `marginsplot` after each of your `margins` commands from question 3.12.
- Re-run each command together with `marginsplot`

3.14 margins, asobserved

- What happens if you don't specify the at() option?

- The default behavior is `asobserved`.
- As we saw last week, margins calculates predicted values at the observed data points.
- It is possible to hold some variables constant and use observed data for others.
- For example, calculate predicted probabilities for educ=12 and IQ at observed values.

```
gen preduc12 = normprob(_b[_cons] + _b[educ]*12 + _b[IQ]*IQ)
sum preduc12 if e(sample)
margins , at(educ=12)
```

- Note: adding `if e(sample)` to the summarize command ensures that only the observations from the previous regression are used.
- Try running the summarize command without `if e(sample)`. What if you add `if libcrd14!=.` instead?
- Calculate predicted probabilities for education = (8 10 12 14 16) and IQ at observed values.
- Use `marginsplot` to graph predicted probabilities.

3.15 Marginal Effects

- One of the most common uses of `margins` is to calculate marginal effects.
- If you haven't covered marginal effects in class yet, don't worry.
- We can make a small change to the `margins` command to change the output to marginal effects.
- The rest of the command options work exactly the same.
- The key option is: `dydx(varlist)`, replacing varlist with one or more x-variable names.
- Alternatively, to calculate marginal effects for all x-variables, use `dydx(*)`
- Write a `margins` command to calculate the average marginal effects of `educ` and `IQ` over all observed data points.
- Write a `margins` command to calculate the marginal effects at the average values of `educ` and `IQ`.

3.16 Odds Ratios

- After `logit` regression, you may want to report marginal effects or odds ratios.
- Run a `logit` regression using the same specification we've been using for `probit`.
- Look at the help page for `logit` to find the option to report odds ratios.
- Hint1: Search in the page for `odds` if you can't find the option right away.
- Hint2: There are two common ways of getting Stata to report odds ratios from a logit regression.

3.17 Reporting Results

- When making tables of regression results, you may want to report coefficient estimates, marginal effects, and/or odds ratios (logit only).
- You can do all of these with `outreg2`, but there are some tricks.
- For odds ratios, you must include the option `eform` on your outreg2 command.
- For marginal effects, you must include the option `post` on the margins command.
- The option `post` will make `margins` overwrite your previous estimation results in Stata's memory.
- Consider the following example:

```
logit libcrd14 educ IQ lwage exper
outreg2 using mylogitregs , excel word ///
 replace ctitle(Coefficients)
outreg2 using mylogitregs , excel word ///
 append eform ctitle(Odds Ratios)
margins , dydx(*) post
outreg2 using mylogitregs , excel word ///
 append ctitle(Marginal FX)
```

3.18 Factor Variables - Categorical

- `margins` is very good at handling categorical variables specified using factor variable notation: `help fvvarlist`
- Let's try it with education categories.
- First, create a new variable named `educcat` with the following educational categories: 1-11, 12, 13-15, 16-18
- Run the following logit regression: `logit libcrd14 i.educcat IQ black exper expersq`
- Get predicted probabilities for each education category: `margins i.educcat`
- Get predicted probabilities for each education category, at IQ values from 50 to 150 in steps of 10.
- Use `marginsplot` to graph the predicted probabilities for each educcat group.

3.19 Factor Variables - Binary/Dummies

- In addition to being convenient, sometimes estimates are "better" if you specify variables using factor variable notation.
- Consider the binary variable `black` from the previous regression.
- Calculate the average marginal effect of the variable `black`
- Rerun the regression using `i.black` instead of `black`.
- Are the coefficients the same?
- Recalculate the average marginal effect of the variable `black`
- Is the average marginal effect the same?

- Using `i.black` tells Stata that this variable can only take values of 0 and 1.
- Stata can use that information in the calculation of average marginal effects, resulting in a better estimate.

3.20 Factor Variables - Interactions

- Next, consider the following OLS regression with a quadratic term: `reg lwage exper expersq`
- Is the quadratic term significant? Is that reflected in predicted values?
- Use `marginsplot` to plot the following predicted probabilities: `margins, at(exper=(0(5)25))`
- Stata does not know that exper and expersq are based on the same underlying characteristic, and should move together.
- You can let Stata know this by specifying the quadratic term using factor variable notation.
- From the help page: `help fvvarlist`

```
c.age                same as age
c.age#c.age          age squared
```

- Rerun the initial regression, specifying `expersq` using factor variable notation.
- Rerun the `margins` and `marginsplot` commands exactly as before.
- Do the plots look different?
- `margins` works like this for any interaction terms, not just for continuous variables.
- Any time you are using `margins`, always use factor variable notation to specify interaction or quadratic terms.