Logistic and Probit Regression in Stata

Yingying Ji

2019/04/10

Introduction

This report introduces the command codes about different implementations of logistic and probit regressions in Stata. And also, demonstrate logit and probit methods in Classification leture by translating the "Ketchup" example in class into Stata language. Additionally, compare these two methods in this case.

Required commands in Stata

- For logistic regression
  - ***logit***: for binary selection model.
  - ***mlogit***: for multinomial selection model, deal with individual-specific explanatory variables, required wide type data, dependent variable can be categorical data.
  - ***clogit***: for multinomial selection model, deal with alternative-specific explanatory variables, required long type data, dependent variable need to be binary.
  - ***asclogit***: for multinomial selection model, deal with both individual-specific and alternative-specific explanatory variables, required long type data, dependent variable need to be binary.
- For probit regression
  - ***probit***: for binary selection model.
  - ***mprobit***: for multinomial selection model, deal with individual-specific explanatory variables,required wide type data, dependent variable can be categorical data.
  - ***asmlogit***: for multinomial selection model, deal with both individual-specific and alternative-specific explanatory variables, required long type data, dependent variable need to be binary.

Ketchup example in Stata

- Import, read the data and start a log file to restore the process and data results at first

```
clear //clear memory and remove data
set more off
capture log close //If there is any log file open, close it.
cd "C:\Users\15068\Desktop\HW3" //Change the working directory to a specific folder
log using hw3_YingyingJi_15220162202134,replace //Start a log file. If the log file a

import excel Ketchup, firstrow //load the data
des //read the data
sum //read the data
```

- Multinomial selection model with individual-specific explanatory variables in two regressions

```
* mlogit & mprobit
* Since both require wide type data
gen id=_n
gen choice1= .
replace choice1=1 if choice == "stb"
replace choice1=2 if choice == "heinz"
replace choice1=3 if choice == "hunts"
replace choice1=4 if choice == "delmonte"
drop choice
rename choice1 choice

label define choice 1 "stb" 2 "heinz" 3 "hunts" 4 "delmonte"
label values choice choice // Assign value label to variables

mlogit choice income, base(1) // based on store brand refenrence level
predict cl1 cl2 cl3 cl4
mprobit choice income, base(1)
predict cp1 cp2 cp3 cp4
```

- Multinomial selection model with alternative-specific explanatory variables in logistic regression

```
* clogit
* Since clogit requires long type data and binary dependent variable
reshape long price, i(id) j(brand) string
gen brand1=.
replace brand1=1 if brand == "stb"
replace brand1=2 if brand == "heinz"
replace brand1=3 if brand == "hunts"
replace brand1=4 if brand == "delmonte"
drop brand
rename brand1 brand
label values brand choice

gen choice2=0
replace choice2=1 if choice ==brand
drop choice
rename choice2 choice

clogit choice i.brand price, group(id)
predict ccl
```

- Multinomial selection model with both individual-specific and alternative-specific explanatory variables

```
* asclogit & asmprobit
* Since asclogit & asmprobit both require long type data and binary dependent variable
asclogit choice price, case(id) alternatives(brand) casevar(income) base(1)
predict ascl

asmprobit choice price, case(id) alternatives(brand) casevar(income) base(1)
predict asmp

log close //close the log file.
```

Conclusion

Since the choices in our case exist correlation,which does not satisfy the **IIA** condition, the regression results [1] in probit regression seems more reliable.

[1]The results can be seen by running the do file or seeing in the log file, which is also attached.