How can I get margins for a multiply imputed survey logit model?

Authored by
stats writer

July 1, 2024

The process of obtaining margins for a multiply imputed survey logit model involves using statistical software to calculate the predicted values for the outcome variable based on the imputed datasets. These predicted values are then averaged across the multiple imputed datasets to provide a single set of margins. This allows for a more accurate estimation of the model's effects and can help in interpreting the results of the survey logit model.

How can I get margins for a multiply imputed survey logit model? | Stata FAQ

The margins command introduced in Stata 11
is a very popular post-estimation command. However, it can be tricky to use in
conjunction with multiple imputation and survey data.

Let's begin by looking at the data.

use https://stats.idre.ucla.edu/stat/data/hsbmar, clear

sum honors female prog read math science socst, sep(0)

Variable | Obs Mean Std. Dev. Min Max
-------------+---
honors | 200 .265 .4424407 0 1
female | 185 .5459459 .4992356 0 1
prog | 200 2.025 .6904772 1 3
read | 185 51.61622 10.19104 28 76

math | 190 52.17895 9.246168 33 75
science | 193 51.57513 9.86396 26 74
socst | 188 51.59043 10.44862 26 71

As you can see from the table above, all of the variables except for honors and prog have missing values.

honors is the binary response variable while female (two level categorical) and prog (three level categorical) are the research variables of interest with read, math, science and socst serving as control variables. Our primary interest is in the female-by-prog interaction. We will want to compute the predicted probabilities for each of the six cells of the 2-by-3 interaction.

So, what's the big deal?

Why not just impute the data and then run the margins command.
Well, we can impute the data, but we need a way to run both svy logit and margins on each

imputed dataset and then combine the margins results into a single output. The issue is that margins does not work with mi estimate.

We can accomplish this by writing a wrapper program called mimargins and saving it in a file called mimargins.ado. It contains both the svy logit and margins commands. By setting the option properties to mi, mimargins can be used with mi estimate. We also need to declare mimargins to be an eclass program.

Here is what the mimargins program looks like.

```
program mimargins, eclass properties(mi)
version 12
svy: logit honors i.female##i.prog read math science socst
margins female#prog, atmeans asbalanced post
end
```

Here is how you use mimargins in the calling program.

mi estimate, cmdok: mimargins 1

The cmdok is needed because Stata does not recognize mimargins as an mi estimable program.

Next, we need to note that our data are not truly survey data. We are going to fake this by declaring that the values of write are the pweights and that ses is the stratification variable. Since this is part of a multiple imputation we need to run the survey set command as mi svyset. Here is the code for performing the multiple imputation using chained equations creating 10 imputed datasets. Note, the value 10 for the number of imputed datasets was selected for demonstration purposes and does not represent a recommendation.

set seed 1234543

mi set mlong

mi register imputed female math read science socst

mi svyset , strata(ses)

mi impute chain (logit) female (regress) math read science socst = ///

write awards, add(10)

Conditional models:
science: regress science math socst i.female read write awards
math: regress math science socst i.female read write awards
socst: regress socst science math i.female read write awards
female: logit female science math socst read write awards
read: regress read science math socst i.female write awards

Performing chained iterations ...

Multivariate imputation Imputations = 10
Chained equations added = 10

Imputed: m=1 through m=10 updated = 0

Initialization: monotone Iterations = 100
burn-in = 10

female: logistic regression
math: linear regression
read: linear regression
science: linear regression
socst: linear regression

--

| Observations per m

|--

Variable | Complete Incomplete Imputed | Total

-------------------+------------------------------------+----------

female | 185 15 15 | 200

math | 190 10 10 | 200

read | 185 15 15 | 200

science | 193 7 7 | 200

socst | 188 12 12 | 200

--

(complete + incomplete = total; imputed is the minimum across m

of the number of filled-in observations.)

Next, we can run our survey logit model and check the interaction. Please note the order of the commands: The mi estimate: comes first, followed by the svy:, which in turn, is followed by the logit command itself.

mi estimate: svy: logit honors i.female##i.prog read math science socst

Multiple-imputation estimates Imputations = 10
Survey: Logistic regression Number of obs = 190

Number of strata = 3 Population size = 9,998
Number of PSUs = 190
Average RVI = 0.0660
Largest FMI = 0.2469
Complete DF = 187
DF adjustment: Small sample DF: min = 75.62
avg = 156.92
max = 181.78
Model F test: Equal FMI F(9, 182.6) = 5.06
Within VCE type: Linearized Prob > F = 0.0000

--

--

honors | Coef. Std. Err. t P>|t|

-----------------+--

---

female |

female | 1.669564 1.06815 1.56 0.120 -.438678 3.777806

|

prog |

academic | .706834 1.040896 0.68 0.498 -1.347074 2.760742

vocation | -.6572194 1.126282 -0.58 0.560 -2.879486 1.565048

|

female#prog |

female#academic | -.5020129 1.200932 -0.42 0.676 -2.87197 1.867944

female#vocation | 1.264679 1.36103 0.93 0.354 -1.421087 3.950444

|

read | .0579493 .0365918 1.58 0.117 -.0149354 .1308341

math | .1131006 .0383768 2.95 0.004 .0372635 .1889377

science | .0709565 .0405595 1.75 0.082 -.0092108 .1511239

socst | -.0009834 .0323599 -0.03 0.976 -.0649752

.0630084
_cons | -15.40424 2.485827 -6.20 0.000 -20.31064
-10.49784

--

mi test 1.female#2.prog 1.female#3.prog
note: assuming equal fractions of missing information

(1) 1.female#2.prog = 0
(2) 1.female#3.prog = 0

F(2, 183.4) = 1.26
Prob > F = 0.2850

Unfortunately our interaction was not statistically significant. However, we will push
ahead and compute the predicted cell probabilities for the 2×3 interaction just to show how it can be done.

mi estimate, cmdok: mimargins 1

Multiple-imputation estimates Imputations = 10
Adjusted predictions Number of obs = 190

Number of strata = 3

Average RVI = 0.0279

Largest FMI = 0.0586

Complete DF = 187

DF adjustment: Small sample DF: min = 164.05

avg = 176.22

Within VCE type: Delta-method max = 183.42

--

| Coef. Std. Err. t P>|t|

----------------+---

female#prog |

male#general | .0716598 .0630814 1.14 0.257 -.0528264 .196146

male#academic | .1348606 .0586423 2.30 0.023 .0190696 .2506515

male#vocation | .0384081 .0288355 1.33 0.185 -.0184993 .0953155

female#general | .2891648 .0954564 3.03 0.003 .1007761 .4775536

female#academic | .3328262 .0879882 3.78 0.000 .1592084 .5064441

female#vocation | .427272 .1585705 2.69 0.008 .1144153

.7401288

--

And that is how you can compute adjusted predictions for multiply imputed survey data.
This approach will generalize to other estimation commands as well as other margins commands.