“How can I perform post estimation tests with multiply imputed datasets?”

Authored by

stats writer

July 1, 2024

Performing post estimation tests with multiply imputed datasets involves utilizing statistical techniques to analyze and draw conclusions from data that has been imputed multiple times. This process allows for a more accurate and comprehensive analysis, as it takes into account the uncertainty and variability introduced by the imputation process. Various methods such as Rubin's rules and pooled analysis can be used to combine the results from each imputed dataset and conduct post estimation tests. These tests can then provide valuable insights and assess the validity of the imputed data and the overall analysis.

How can I perform post estimation tests with multiply imputed datasets? | Stata FAQ

Below we show how to perform post estimation hypothesis tests on models based on multiply imputed data with mi estimate, mi test and mi testtransform.

The example for this faq uses data on high school students. The variables read, write, and math give the student's scores in reading, writing, and math respectively. The variable female is equal to one if the student is female and zero otherwise. Finally, prog contains information on the type of program the student is in either general, academic, and vocational. The multiply imputed datasets are

created using mi impute and are saved into in a single file which contains all 10 imputations
as well as the original data. The variable _mi_m gives the imputation number, _mi_m = 0
contains the original data.

use https://stats.idre.ucla.edu/stat/data/hsbmar, clearmi set mlongmi register imputed female math read science socstmi impute chain (logit) female (regress) math science socst read = ///ses write awards, add(10) force

Below we use mi estimate:regress to fit a linear regression model. The mi estimate:
prefix informs Stata that we want to analyze multiply imputed
datasets, without it, the command would be performed on the dataset as though it
were a single dataset, rather than a series of multiply imputed
datasets.

mi estimate: regress read write i.female math i.prog

Multiple-imputation estimates Imputations = 10
Linear regression Number of obs = 199

Average RVI = 0.1481
Largest FMI = 0.2715
Complete DF = 193
DF adjustment: Small sample DF: min = 69.10
avg = 127.30
max = 174.50
Model F test: Equal FMI F(5, 172.3) = 33.18
Within VCE type: OLS Prob > F = 0.0000

--
read | Coef. Std. Err. t P>|t|
-------------+--
write | .4059465 .0840575 4.83 0.000 .2395667 .5723263
|
female |
female | -3.18419 1.296704 -2.46 0.016 -5.763144 -
.6052363
math | .3980997 .0898392 4.43 0.000 .21888 .5773194
|
prog |
academic | 1.467474 1.444345 1.02 0.311 -1.385213
4.320161
vocation | -.7782421 1.586677 -0.49 0.624 -3.90979
2.353306

|

_cons | 11.04381 3.939979 2.80 0.006 3.260903 18.82671

--

Once the model is estimated the mi test command with the prefix

can be used to perform multiple degree of freedom tests. One common use for this is to

test for an overall effect ofa nominal variable represented by a series of dummy variables.

Below we use mi test: to test for an overall effect of type of program (prog).

mi test 2.prog 3.prog

(1) 2.prog = 0

(2) 3.prog = 0

F(2, 181.6) = 1.23

Prob > F = 0.2952

The mi test command can also be used to test nested models, where the null

hypothesis is that the coefficients on two or more variables are simultaneously equal to zero.

mi test math write

(1) math = 0
(2) write = 0

F(2, 132.6) = 52.12
Prob > F = 0.0000

It is also possible to test linear combinations of variables. Below we test a model
with an interaction between math and female. The variable female
is dummy coded (0=male, 1=female). First we create the interaction as we
normally would, then we use the regress command with the mi estimate: prefix to
fit a regression model.

Then the mi estimate: and mitesttransform command
can be used to test the null hypothesis that the effect of math on read is zero when
female=1. The coeflegend option specifies the legend of coefficients and
how to specify them in an expression. We will need these coefficient names in order to estimate

the effect of math for female=1.

mi estimate (math_slope_female:_b + _b), coeflegend:
///
regress read write i.female##c.math i.prog

Transformations Average RVI = 0.1536
Largest FMI = 0.1379
Complete DF = 192
DF adjustment: Small sample DF: min = 124.38
avg = 124.38
Within VCE type: OLS max = 124.38

math_slope~e: _b + _b

--
read | Coef. Legend
---------------+--
math_slope_female | .4542803 _b
--

mi testtransform math_slope_female

math_slope~e: _b + _b

(1) math_slope_female = 0

F(1, 124.4) = 20.67
Prob > F = 0.0000