How can I compute indirect effects with imputed data using Method 1?

Authored by

stats writer

July 1, 2024

Method 1 is a statistical approach used to compute indirect effects with imputed data. This method involves imputing missing data values and then using a regression model to estimate the indirect effects of a particular variable. The process starts by creating multiple imputed datasets, which are used to estimate the missing values. These datasets are then combined to create a single dataset, and the indirect effect is computed by running a regression analysis. This method is useful when dealing with incomplete data as it allows for a more accurate estimation of indirect effects and can provide valuable insights into the relationship between variables.

How can I compute indirect effects with imputed data? (Method 1) | Stata FAQ

NOTE: Code for this page was tested in Stata 12.

Computing indirect effects with multiply imputed data takes a few more step than for a conventional non-imputed model.

Let's begin by looking at the data.

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

sum science read math female

Variable | Obs Mean Std. Dev. Min Max

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

science | 193 51.57513 9.86396 26 74

read | 185 51.61622 10.19104 28 76

math | 190 52.17895 9.246168 33 75

female | 185 .5459459 .4992356 0 1

As you can see from the table above, all of
the variables have a different number of observations.
For our example science
is the dependent variable, read is the mediator, math
is the independent variable and female is a covariate.

The method we will use to compute an indirect effect involves the sureg
and nlcom commands to get the product of coefficients.

So, what's the problem?

Why not just impute the data and run the analyses. Well, sureg does not
work with imputed data unless we add the cmdok option to our mi estimate.
But then nlcom can't find the results in the correct location. nlcom
looks for e(b) and e(V) but mi estimate
saves the results in e(b_mi) and e(V_mi).

We can move the results to the correct location by writing an eclass

program which we are calling called myeret.ado.
It makes use of the ereturn repost command.
Here is what the program looks like.

```
prog myeret, eclass
ereturn repost b=b V=V
end
```

It has to be declared to be an eclass command. The repost
places matrix b into e(b) anf V into
e(V).

We saved the program in our ado/personal directory.

Now, let's start our example analysis by running the multiple imputation.

```
mi set mlong

mi register imputed read math science female

set seed 485769

mi impute mvn read math science female = write, add(20)
```

Performing EM optimization:
observed log likelihood = -1349.5408 at iteration 7

Performing MCMC data augmentation ...

Multivariate imputation Imputations = 20
Multivariate normal regression added = 20
Imputed: m=1 through m=20 updated = 0

Prior: uniform Iterations = 2000
burn-in = 100
between = 100

| Observations per m
|----------
Variable | Complete Incomplete Imputed | Total
----------+----------+----------
read | 185 15 15 | 200
math | 190 10 10 | 200
science | 193 7 7 | 200
female | 185 15 15 | 200

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

of the number of filled-in observations.)

If you try to run mi estimate: sureg (read math female)(science read math female)
you will get an error message that sureg is not officially supported.
However if you add the cmdok option it will run just fine.

mi estimate, cmdok: sureg (read math female)(science read math female)

Multiple-imputation estimates Imputations = 20
Number of obs = 200
Average RVI = 0.1236
Largest FMI = 0.1816
DF adjustment: Large sample DF: min = 594.26
avg = 1602.26
max = 3257.71

--
| Coef. Std. Err. t P>|t|
-------------+--
read |
math | .7030755 .0627357 11.21 0.000 .5800272 .8261238

female | -.8018491 1.156265 -0.69 0.488 -3.068929 1.465231
_cons | 15.58747 3.385455 4.60 0.000 8.949015 22.22592
-------------+--
science |
read | .3661471 .0686624 5.33 0.000 .23146 .5008342
math | .411848 .077228 5.33 0.000 .2602325 .5634634
female | -1.772933 1.122191 -1.58 0.115 -3.976875 .4310101
_cons | 12.17901 3.384061 3.60 0.000 5.53725 18.82077

Next, we need to copy the e(b_mi) and e(V_mi) matrices.

```
matrix b = e(b_mi)
matrix V = e(V_mi)
```

To get all of the pieces into the right place we quietly run sureg on the
imputation zero data followed by our myeret command that was shown above.

```
quietly sureg (read math female)(science read math female) if _mi_m == 0
```

myeret

We are finally able to use the nlcom command to get the product of the coefficients.

nlcom _b*_b

_nl_1: _b*_b

| Coef. Std. Err. z P>|z|
-------------+--
_nl_1 | .257429 .0531467 4.84 0.000 .1532635 .3615946

To get the total effect of math just add the direct effect to the nlcom
command above.

nlcom _b*_b + _b

_nl_1: _b*_b + _b

| Coef. Std. Err. z P>|z|

```
-------------+----------------------------------------------------------------
_nl_1 | .669277 .0617048 10.85 0.000 .5483379 .7902161
------------------------------------------------------------------------------
```

And that is how you can compute an indirect effect using multiply imputed data.