How can I compute indirect effects with imputed data using Method 2?

Authored by

stats writer

July 1, 2024

Method 2 for computing indirect effects with imputed data is a statistical approach that involves imputing missing data values and then using a regression-based method to estimate the indirect effects. This method allows researchers to estimate the effects of one variable on another variable through a third variable, even in cases where there are missing data points. By imputing the missing data, the indirect effects can be accurately estimated and potential biases from missing data can be reduced. This formal description outlines the process of using Method 2 to compute indirect effects with imputed data, allowing for more comprehensive and accurate analysis in statistical research.

How can I compute indirect effects with imputed data? (Method 2) | Stata FAQ

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

Mediation analysis with multiply imputed data takes a few more step than for
a conventional non-imputed model. We looked at one approach on our page
How can I compute indirect effects with imputed data? (Method 1).
The approach shown on this page is a bit easier to implement and less convoluted.

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.

Let's go ahead and start our example analysis by performing 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. We also need the
equivalent of the nlcom command.
We can do this by adding effects directly to the mi analyze command.
As shown below, we have added indirect and total effects in parentheses. Each of the effects
is labeled, ind_eff for the indirect effect and tot_eff for the total effect..

mi estimate (ind_eff: _b*_b) ///
(tot_eff: _b*_b + _b), cmdok: ///
sureg (read math)(science read math)

Multiple-imputation estimates Imputations = 20

Number of obs = 200

Average RVI = 0.1138

Largest FMI = 0.1688

DF adjustment: Large sample DF: min = 686.04

avg = 1495.76

max = 2377.41

--

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

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

read |

math | .7038385 .0628724 11.19 0.000 .5805186 .8271584

_cons | 15.10742 3.321185 4.55 0.000 8.594701 21.62014

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

science |

read | .3721145 .0692032 5.38 0.000 .2363619 .5078671

math | .4097015 .0780576 5.25 0.000 .256441 .562962

_cons | 11.00745 3.273573 3.36 0.001 4.585789 17.42911

--

Transformations Average RVI = 0.1315

Largest FMI = 0.1098

DF adjustment: Large sample DF: min = 1607.17

avg = 1652.73

max = 1698.29

ind_eff: _b*_b
tot_eff: _b*_b + _b

| Coef. Std. Err. t P>|t|
-------------+--
ind_eff | .2618675 .0538953 4.86 0.000 .1561551 .36758
tot_eff | .671569 .0623058 10.78 0.000 .5493647 .7937733

As you can see, the information for the indirect and total effects is added on below the
results
for the sureg command. If we divide the indirect effect by the total effect we
can see the proportion of the total effect that is mediated.

display .26186753/.67156901

.38993391

In this example approximately 39% of the total effect is

mediated.

This method of computing indirect effects is superior than
Method 1
because it computes the indirect effects for each imputed dataset than then
combines them using Rubin's rules rather than computing the indirect effects
once on the final imputed sureg.