How can I analyze multiple mediators in Stata?

Authored by
stats writer

July 1, 2024

Analyzing multiple mediators in Stata involves using statistical methods to examine the relationship between multiple independent variables and a single dependent variable. This can be accomplished through a process called mediation analysis, which identifies and measures the indirect effects of the independent variables on the dependent variable through one or more mediators. Stata provides various tools and commands for conducting mediation analysis, such as the mediation and medeff commands, which can help researchers understand the complex relationships between variables and potentially identify important mediators in their data. By utilizing these tools, researchers can gain a deeper understanding of the underlying mechanisms driving their results and make more accurate interpretations of their data.

How can I analyze multiple mediators in Stata? | Stata FAQ

Preacher and Hayes (2008) show how to analyze models with multiple mediators in SPSS and SAS, how can I analyze multiple mediators in Stata?

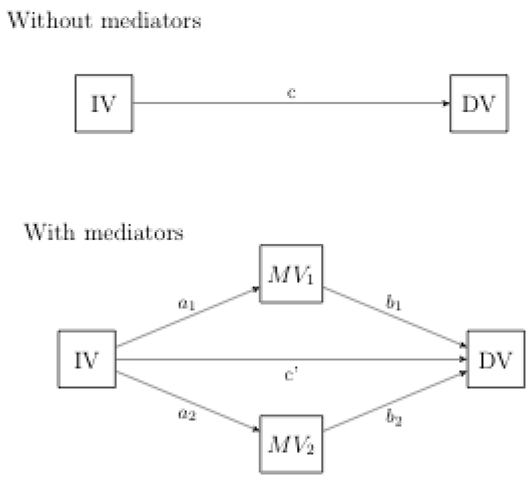
Here is the full citation:

Preacher, K.J. and Hayes, A.F. 2008. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. Behavioral Research Methods, 40, 879-891.

NOTE: If running the code on this page, please copy it all into a do-file and run all of it.

Mediator variables are variables that sit between

independent variable and dependent variable
and mediate the effect of the IV on the DV. A model with two mediators is shown in the figure below.



In the figure above a1 represents the regression coefficient for the IV when the MV is regressed on the IV while b is the coefficient for the MV when the DV is regressed on MV and IV. The symbol c' represents the direct effect of the IV on the DV. Generally, researchers want to determine the indirect effect of the IV on the DV through the MV. One common way to compute the indirect effect is by using the product of the coefficients method. This method determines the indirect effect by multiplying the regression

coefficients, for example, a1*b1 = a1b1. In addition to computing the indirect effect we also want to obtain the standard error of a1b1. Further, we want to be able to do this for each of the mediator variables in the model.

Thus, we need the a and b coefficients for each of the mediator variable in the model. We will obtain all of the necessary coefficients using the sureg (seemingly unrelated regression) command as suggested by Maarten Buis on the Statalist. The general form of the sureg command will look something like this:

sureg (mv1 iv)(mv2 iv)(dv mv1 mv2 iv)

Example 1

The hsb2 dataset with science as the dv, math as the iv and read and write as the two mediator variables.

We will need the coefficients for read on math and write on math as well as the coefficients for science on read and write from

the equation that also includes math.

```
use https://stats.idre.ucla.edu/stat/data/hsb2, clear

sureg (read math)(write math)(science read write math)

Seemingly unrelated regression
----------------------------------------------------------------------
Equation Obs Parms RMSE "R-sq" chi2 P
----------------------------------------------------------------------
read 200 1 7.662848 0.4386 156.26 0.0000
write 200 1 7.437294 0.3812 123.23 0.0000
science 200 3 6.983853 0.4999 199.96 0.0000
----------------------------------------------------------------------

------------------------------------------------------------------------------
| Coef. Std. Err. z P>|z|
-------------+----------------------------------------------------------------
read |
math | .724807 .0579824 12.50 0.000 .6111636 .8384504
_cons | 14.07254 3.100201 4.54 0.000 7.996255 20.14882
-------------+----------------------------------------------------------------
write |
math | .6247082 .0562757 11.10 0.000 .5144099 .7350065
_cons | 19.88724 3.008947 6.61 0.000 13.98981 25.78467
```

-------------+--
science |
read | .3015317 .0679912 4.43 0.000 .1682715 .434792
write | .2065257 .0700532 2.95 0.003 .0692239 .3438274
math | .3190094 .0759047 4.20 0.000 .170239 .4677798
_cons | 8.407353 3.160709 2.66 0.008 2.212476 14.60223

Now we have all the coefficients we need to compute the indirect effect coefficients and their standard errors. We can do this using the nlcom (nonlinear combination) command.
We will run nlcom three times: Once for each of the two specific indirect effects for read and write and once for the total indirect effect.

To compute an indirect direct we specify a product of coefficients. For example, the coefficient for read on math is _b and the coefficient for science on read is _b. Thus, the product is _b*_b. To get the total indirect effect we just add the two product terms together in the nlcom command.

```
/* indirect via read */

nlcom _b*_b

_nl_1: _b*_b

------------------------------------------------------------------------------
| Coef. Std. Err. z P>|z|
-------------+----------------------------------------------------------------
_nl_1 | .2185523 .05229 4.18 0.000 .1160659 .3210388
------------------------------------------------------------------------------

/* indirect via write */

nlcom _b*_b

_nl_1: _b*_b

------------------------------------------------------------------------------
| Coef. Std. Err. z P>|z|
-------------+----------------------------------------------------------------
_nl_1 | .1290183 .0452798 2.85 0.004 .0402715 .2177651
------------------------------------------------------------------------------

/* total indirect */

nlcom _b*_b+_b*_b
```

_nl_1: _b*_b+_b*_b

```
------------------------------------------------------------------------------
| Coef. Std. Err. z P>|z|
-------------+----------------------------------------------------------------
_nl_1 | .3475706 .0594916 5.84 0.000 .2309693 .4641719
------------------------------------------------------------------------------
```

The results above suggest that each of the separate indirect effects as well as the total indirect effect are significant. From the above results it is also possible to compute the ratio of indirect to direct effect and the proportion due to the indirect effect. These computations require an estimate of the direct effect, which can be found in the sureg output. In this example the direct effect is given by the coefficient for math in the last equation (.3190094). Here are the manual computations for the ratio of indirect to direct and the proportion of total effect that is mediated.

```
/* ratio of indirect to direct */

display .3475706/.3190094
```

1.0895309

/* proportion of total effect that is mediated */

display .3475706/(.3475706+.3190094)

.52142369

nlcom computes the standard errors using the delta method which assumes that the estimates of the indirect effect are normally distributed. For many situations this is acceptable but it does not work well for the indirect effects which are usually positively skewed and kurtotic. Thus the z-test and p-values for these indirect effects generally cannot be trusted. Therefore, it is recommended that bootstrap standard errors and confidence intervals be used.

Below is a short ado-program that is called by the bootstrap command. It computes the indirect effect coefficients as the product of sureg coefficients (as before) but does not use the nlcom command since the standard

errors will be computed using the bootstrap.

bootmm is an rclass program that produces three return values which we have called "indread", "indwrite" and "indtotal." These are the local names for each of the indirect effect coefficients and for the total indirect effect.

We run bootmm with the bootstrap command. We give the bootstrap command the names of the three return values and select options for the number of replications and to omit printing dots after each replication.

Since we selected 5,000 replications you may need to be a bit patient depending upon the speed of your computer.

```
capture program drop bootmm
program bootmm, rclass
syntax
sureg (read math)(write math)(science read write math)
`if' `in'
return scalar indread = _b*_b
```

```
return scalar indwrite = _b*_b
return scalar indtotal = _b*_b+_b*_b
end

bootstrap r(indread) r(indwrite) r(indtotal), bca reps(5000): bootmm

Bootstrap results Number of obs = 200
Replications = 5000

command: bootmm
_bs_1: r(indread)
_bs_2: r(indwrite)
_bs_3: r(indtotal)

------------------------------------------------------------------------------
| Observed Bootstrap Normal-based
| Coef. Std. Err. z P>|z|
-------------+----------------------------------------------------------------
_bs_1 | .2185523 .0544617 4.01 0.000 .1118094 .3252953
_bs_2 | .1290183 .0498037 2.59 0.010 .0314048 .2266318
_bs_3 | .3475706 .0653076 5.32 0.000 .2195701 .4755711
------------------------------------------------------------------------------
```

We could use the bootstrap standard errors to see if the

indirect effects are significant but it is usually recommended that bias-corrected or percentile confidence intervals be used instead. These confidence intervals are nonsymmetric reflecting the skewness of the sampling distribution of the product coefficients. If the confidence interval does not contain zero than the indirect effect is considered to be statistically significant.

estat boot, percentile bc bca

Bootstrap results Number of obs = 200
Replications = 5000

command: bootmm
_bs_1: r(indread)
_bs_2: r(indwrite)
_bs_3: r(indtotal)

--

| Observed Bootstrap
| Coef. Bias Std. Err.

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

_bs_1 | .21855231 -.0009252 .05446169 .1116576

.3263005 (P)

| .1140179 .3286456 (BC)

| .1140179 .3286456 (BCa)

_bs_2 | .12901828 .0009822 .04980373 .0375536 .2286579 (P)

| .0375377 .22842 (BC)

| .0333511 .2264691 (BCa)

_bs_3 | .34757059 .000057 .0653076 .2181866 .4773324 (P)

| .2209776 .4805473 (BC)

| .2158857 .4752103 (BCa)

--

(P) percentile confidence interval

(BC) bias-corrected confidence interval

(BCa) bias-corrected and accelerated confidence interval

In this example, the total indirect effect of math through read and write is significant as are the individual indirect effects.

Example 2

What do you do if you also have control variables? You just add them to each of the equations in the

sureg model. Let's say that socst is a covariate. Here is how the bootstrap
process would work.

```
capture program drop bootmm
program bootmm, rclass
syntax
sureg (read math socst)(write math socst)(science read write math socst) `if' `in'
return scalar indread = _b*_b
return scalar indwrite = _b*_b
return scalar indtotal = _b*_b + ///
_b*_b
end

bootstrap r(indread) r(indwrite) r(indtotal), bca reps(5000) nodots: bootmm

Bootstrap results Number of obs = 200
Replications = 5000

command: bootmm
_bs_1: r(indread)
_bs_2: r(indwrite)
_bs_3: r(indtotal)
```

| Observed Bootstrap Normal-based

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

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

_bs_1 | .1561855 .040306 3.87 0.000 .0771872 .2351837

_bs_2 | .0890589 .0352121 2.53 0.011 .0200444 .1580733

_bs_3 | .2452443 .0477817 5.13 0.000 .1515939 .3388947

----

estat boot, percentile bc bca

Bootstrap results Number of obs = 200

Replications = 5000

command: bootmm

_bs_1: r(indread)

_bs_2: r(indwrite)

_bs_3: r(indtotal)

| Observed Bootstrap

| Coef. Bias Std. Err.

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

_bs_1 | .15618546 -.0016606 .04030598 .0784972 .2359464 (P)

| .0836141 .2407494 (BC)

| .0838816 .2413034 (BCa)

_bs_2 | .08905886 .0000963 .0352121 .0241053 .163005 (P)

| .0274379 .1664222 (BC)

| .0260387 .164438 (BCa)

_bs_3 | .24524432 -.0015643 .0477817 .1536668 .341307 (P)

| .1581453 .3477974 (BC)

| .1581453 .3477974 (BCa)

(P) percentile confidence interval

(BC) bias-corrected confidence interval

(BCa) bias-corrected and accelerated confidence interval

Although the total and individual indirect are much smaller in the model with the covariate, they are still statistically significant using the 95% confidence intervals.