

Testing Mediation with Regression Analysis

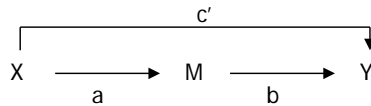
Mediation is a hypothesized causal chain in which one variable affects a second variable that, in turn, affects a third variable. The intervening variable, M, is the mediator. It “mediates” the relationship between a predictor, X, and an outcome. Graphically, mediation can be depicted in the following way:



Paths *a* and *b* are called direct effects. The mediational effect, in which X leads to Y through M, is called the *indirect effect*. The indirect effect represents the portion of the relationship between X and Y that is mediated by M.

Testing for mediation

Baron and Kenny (1986) proposed a four step approach in which several regression analyses are conducted and significance of the coefficients is examined at each step. Take a look at the diagram below to follow the description (note that *c'* could also be called a direct effect).



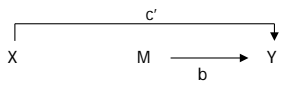
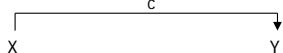
	<i>Analysis</i>	<i>Visual Depiction</i>
<i>Step 1</i>	Conduct a simple regression analysis with X predicting Y to test for path <i>c</i> alone, $Y = B_0 + B_1X + e$	
<i>Step 2</i>	Conduct a simple regression analysis with X predicting M to test for path <i>a</i> , $M = B_0 + B_1X + e$.	
<i>Step 3</i>	Conduct a simple regression analysis with M predicting Y to test the significance of path <i>b</i> alone, $Y = B_0 + B_1M + e$.	
<i>Step 4</i>	Conduct a multiple regression analysis with X and M predicting Y, $Y = B_0 + B_1X + B_2M + e$	

The purpose of Steps 1-3 is to establish that zero-order relationships among the variables exist. If one or more of these relationships are nonsignificant, researchers usually conclude that mediation is not possible or likely (although this is not always true; see MacKinnon, Fairchild, & Fritz, 2007).

Assuming there are significant relationships from Steps 1 through 3, one proceeds to Step 4. In the Step 4 model, some form of mediation is supported if the effect of M (path *b*) remains significant after controlling for X. If X is no longer significant when M is controlled, the finding supports *full mediation*. If X is still significant (i.e., both X and M both significantly predict Y), the finding supports *partial mediation*.

Calculating the indirect effect

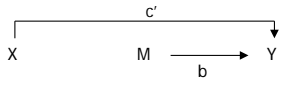

The above four-step approach is the general approach many researchers use. There are potential problems with this approach, however. One problem is that we do not ever really test the significance of the indirect pathway—that X affects Y through the compound pathway of *a* and *b*. A second problem is that the Barron and Kenny approach tends to miss some true mediation effects (Type II errors; MacKinnon et al., 2007). An alternative, and preferable approach, is to calculate the indirect effect and test it for significance. The regression coefficient for the indirect effect represents the change in Y for every unit change in X that is mediated by M. There are two ways to estimate the indirect coefficient. Judd and Kenny (1981) suggested computing the difference between two regression coefficients. To do this, two regressions are required.

<i>Judd & Kenny Difference of Coefficients Approach</i>		
	<i>Analysis</i>	<i>Visual Depiction</i>
<i>Model 1</i>	$Y = B_0 + B_1X + B_2M + e$	
<i>Model 2</i>	$Y = B_0 + BX + e$	

The approach involves subtracting the partial regression coefficient obtained in Model 1, B_1 from the simple regression coefficient obtained from Model 2, B . Note that both represent the effect of X on Y but that B is the zero-order coefficient from the simple regression and B_1 is the partial regression coefficient from a multiple regression. The indirect effect is the difference between these two coefficients:

$$B_{indirect} = B - B_1.$$

An equivalent approach calculates the indirect effect by multiplying two regression coefficients (Sobel, 1982). The two coefficients are obtained from two regression models.

<i>Sobel Product of Coefficients Approach</i>		
	<i>Analysis</i>	<i>Visual Depiction</i>
<i>Model 1</i>	$Y = B_0 + B_1X + B_2M + e$	
<i>Model 2</i>	$M = B_0 + BX + e$	

Notice that Model 2 is a different model from the one used in the difference approach. In the Sobel approach, Model 2 involves the relationship between X and M. A product is formed by multiplying two coefficients together, the partial regression effect for M predicting Y, B_2 , and the simple coefficient for X predicting M, B :

$$B_{indirect} = (B_2)(B)$$

As it turns out, the Kenny and Judd difference of coefficients approach and the Sobel product of coefficients approach yield identical values for the indirect effect (MacKinnon, Warsi, & Dwyer, 1995).

Note: regardless of the approach you use (i.e., difference or product) be sure to use *unstandardized* coefficients if you do the computations yourself.

Statistical tests of the indirect effect

Once the regression coefficient for the indirect effect is calculated, it needs to be tested for significance or a confidence interval needs to be constructed. There has been considerable controversy about the best way to estimate the standard error used in the significance test or confidence interval, however, and there are quite a few proposed approaches to calculation of standard errors. One of the problems is that the sampling distribution of the indirect effect may not be normal, and this has led to more emphasis on confidence intervals, which can be constructed to be asymmetric.

There are two general approaches to testing significance of the indirect effect that appear to perform better than the alternatives in simulation studies—bootstrap methods (sometimes called "nonparametric resampling") and the Monte Carlo method (sometimes called "parametric resampling"). For the bootstrap method, software for testing indirect effects generally offers two

options. One, referred to as "percentile" bootstrap, involves confidence intervals using usual sampling distribution cutoffs without explicit bias corrections. The accelerated bias-corrected bootstrap estimates correct for a bias in the average estimate and the standard deviation across potential values of the indirect coefficient ("accelerated" refers to the graded correction in the standard deviation). The Monte Carlo approach involves computation of the indirect effect and the standard error estimates for the separate coefficients for the full sample. Resampling is then used to estimate the standard errors for the indirect effects using these values. The bias-corrected bootstrap method may result in Type I error rates that are slightly higher than the percentile bootstrap method (Biesanz, Falk, & Savalei, 2010; Fritz, Taylor, & MacKinnon, 2012). Tofighi and MacKinnon (2016) find that both the percentile bootstrap confidence intervals and the Monte Carlo method provide good tests with good Type I error rates and statistical power but that the Monte Carlo approach had somewhat better power in one circumstance. Standardized coefficients can be computed, but they are often not computed with macros or packages that test indirect effects. Computation is simple by hand using the products of standardized coefficients from Model 1 and Model 2 above, $(\beta_2)(\beta)$. Or one can pre-standardize the variables and run the analysis, although you should ignore the significance tests if using this approach. Other methods such as the ratio of indirect to total effect have been suggested for gauging the magnitude of effect (see Preacher & Kelley, 2011, for a review), although MacKinnon and colleagues found this measure to be unstable with smaller sample sizes .

Mediation models (and the indirect effects) can be tested with regression analysis. Macros or preprogramed procedures, such as Andrew Hayes's PROCESS macro, the mediation package in R, or the RMediation package in R, run the separate regression models described above, calculate the indirect effect coefficient (or coefficients), and then use a method such as bootstrap or Monte Carlo to test the indirect effect for significance.

Many structural equation modeling packages, such as Mplus or R lavaan, can conduct the same types of tests of the indirect effects. There are a couple of differences from regression macros/packages. SEM tests the paths specified in the model, and, upon request, can conduct the same types of tests of the indirect effects. For measured variables and continuous variables, this approach is equivalent to the regression approach. But SEM makes it possible to test more complicated models, with multiple mediators or multiple links in the chain, all tested as part of the usual model testing process rather than use of regressions conducted in separate steps. In addition, the SEM analysis approach provides model fit information that provides information about consistency of the hypothesized mediational model to the data (more on this issue later). Measurement error is a potential concern in mediation testing because of attenuation of relationships, and the SEM approach can address this problem by removing measurement error from the estimation of the relationships among the variables when latent variables are incorporated

Online resources

David Kenny also has a webpage on mediation: <http://davidakenny.net/cm/mediate.htm>
Preacher's Sobel test calculator: <http://quantpsy.org/sobel/sobel.htm>
Hayes's PROCESS macro: <https://processmacro.org/index.html>
Mediation package in R: <https://cran.r-project.org/web/packages/mediation/mediation.pdf>
RMediation package in R: <https://cran.r-project.org/web/packages/RMediation/RMediation.pdf>

References and Further Reading

Judd, C.M. & Kenny, D.A. (1981). Process Analysis: Estimating mediation in treatment evaluations. *Evaluation Review*, 5(5), 602-619.
Fritz, M. S., Taylor, A. B., & MacKinnon, D. P. (2012). Explanation of two anomalous results in statistical mediation analysis. *Multivariate Behavioral Research*, 47, 61-87.
Goodman, L. A. (1960). On the exact variance of products. *Journal of the American Statistical Association*, 55, 708-713.
Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis*. New York: The Guilford Press.
Hoyle, R. H., & Kenny, D. A. (1999). Statistical power and tests of mediation. In R. H. Hoyle (Ed.), *Statistical strategies for small sample research*. Newbury Park: Sage.
MacKinnon, D.P. (2008). *Introduction to statistical mediation analysis*. Mahwah, NJ: Erlbaum.
MacKinnon, D.P. & Dwyer, J.H. (1993). Estimating mediated effects in prevention studies. *Evaluation Review*, 17(2), 144-158.
MacKinnon, D.P., Fairchild, A.J., & Fritz, M.S. (2007). Mediation analysis. *Annual Review of Psychology*, 58, 593-614.

- MacKinnon, D.P., Lockwood, C.M., Hoffman, J.M., West, S.G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods, 7*, 83-104.
- MacKinnon DP, Warsi G, Dwyer JH. 1995. A simulation study of mediated effect measures. *Multivariate Behavioral Research, 30*:41–62
- Preacher, K.J., & Hayes, A.F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers, 36*, 717-731.
- Shrout, P.E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods, 7*, 422-445.
- Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. In S. Leinhardt (Ed.), *Sociological Methodology 1982* (pp. 290-312). Washington DC: American Sociological Association.
- Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). *Mediation: R package for causal mediation analysis*. Retrieved from <ftp://cran.r-project.org/pub/R/web/packages/mediation/vignettes/mediation.pdf>