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, Z, is the mediator. It "mediates" the relationship between a predictor, X, and an outcome. Graphically, mediation can be depicted in the following way:

$$X \xrightarrow{f} Z \xrightarrow{g} Y$$

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

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 *h* is also a direct effect).

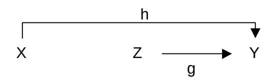
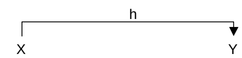
$$\begin{tikzcd} X \arrow[r, "f"'] \arrow[rr, bend left, "h"] & Z \arrow[r, "g"'] & Y \end{tikzcd}$$

	Analysis	*Visual Depiction*
Step 1	Conduct a simple regression analysis with X predicting Y to test for path *h* alone, $Y = B_0 + B_1 X + e$	$X \xrightarrow{h} Y$
Step 2	Conduct a simple regression analysis with X predicting Z to test for path *f*, $Z = B_0 + B_1 X + e$.	$X \xrightarrow{f} Z$
Step 3	Conduct a simple regression analysis with Z predicting Y to test the significance of path *g* alone, $Y = B_0 + B_1 Z + e$.	$Z \xrightarrow{g} Y$
Step 4	Conduct a multiple regression analysis with X and Z predicting Y, $Y = B_0 + B_1 X + B_2 Z + e$	$\begin{tikzcd} X \arrow[rr, bend left, "h"] & Z \arrow[r, "g"'] & Y \end{tikzcd}$

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. 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 Z (path *g*) remains significant after controlling for X. If X is no longer significant when Z is controlled, the finding supports *full mediation*. If X is still significant (i.e., both X and Z significantly predict Y), the finding supports *partial mediation*.

Calculating the indirect effect

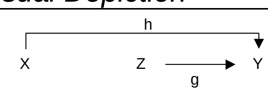
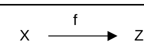
The above four-step approach is the general approach many researchers use. There are potential problems with this approach, however. One problem is that we do not ever really test the significance of the indirect pathway—that X affects Y through the compound pathway of *f* and *g*. A second problem occurs if there is a suppressed relationship in Steps 1-3. The 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 Z. 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.

Judd & Kenny Difference of Coefficients Approach		
	Analysis	*Visual Depiction*
Model 1	$Y = B_0 + B_1X + B_2Z + e$	
Model 2	$Y = B_0 + BX + e$	

The approach involves subtracting the simple regression coefficient obtained from Model 2, *B*, from the partial regression coefficient obtained in Model 1, B_1. 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.

Sobel Product of Coefficients Approach		
	Analysis	*Visual Depiction*
Model 1	$Y = B_0 + B_1X + B_2Z + e$	
Model 2	$Z = B_0 + BX + e$	

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 Z. A product is formed by multiplying two coefficients together, the partial regression effect for Z predicting Y, B_2, and the simple coefficient for X predicting Z, *B*:

$$B_{indirect} = \left(B_2\right)\left(B\right)$$

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, be sure to use *unstandardized* coefficients.

Statistical tests of the indirect effect

Once the regression coefficient for the indirect effect is calculated, it needs to be tested for significance. There has been considerable controversy about the best way to estimate the standard error used in the significance test, however. There are quite a few approaches to calculation of standard errors and a recent paper by MacKinnon, Lockwood, Hoffman, West, and Sheets (2002) gives a thorough review and comparison of the approaches. This paper reports the results from a

Monte Carlo study of a variety of methods for testing the significance of indirect effects and examined the Type I and Type II error rates of each. Although most of the approaches controlled Type I errors well, they did differ on statistical power. Two approaches developed by MacKinnon, using tailor-made statistics, *P* and *z'*, appear to have the highest power. Significance tables for these two approaches, which need to be conducted by hand, are available through MacKinnon's website (see link below). An alternative, developed by Goodman (1960), has lower power but might be a next best alternative because it can be conveniently calculated automatically online (compliments of Preacher & Leonardelli): http://www.unc.edu/~preacher/ (click on "On-line Sobel Test Calculator, but use Goodman II). An alternative approach recently proposed by Shrout and Bolger (2002) uses bootstrapping for standard errors and may have greater power in small samples, but there is a fairly complicated process that requires special structural modeling software.

Online resources

Online calculation of Goodman and Sobel tests: http://www.unc.edu/~preacher/

Dave McKinnon's website on mediation analysis: http://www.public.asu.edu/~davidpm/ripl/mediate.htm

David Kenny also has a webpage on mediation: http://nw3.nai.net/~dakenny/mediate.htm

Further reading on mediation

Judd, C.M. & Kenny, D.A. (1981). Process Analysis: Estimating mediation in treatment evaluations. Evaluation Review, 5(5), 602-619.

Goodman, L. A. (1960). On the exact variance of products. Journal of the American Statistical Association, 55, 708-713.

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. & Dwyer, J.H. (1993). Estimating mediated effects in prevention studies. Evaluation Review, 17(2), 144-158.

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, D. P., Warsi, G., & Dwyer, J. H. (1995). A simulation study of mediated effect measures. Multivariate Behavioral Research, 30(1), 41-62.

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.