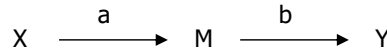


Testing Mediation with Regression Analysis

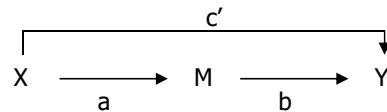
Mediation is a hypothesized causal chain in which one variable affects a second variable that, in turn, affects a third variable. The intervening variable, M, is the mediator. It “mediates” the relationship between a predictor, X, and an outcome. Graphically, mediation can be depicted in the following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

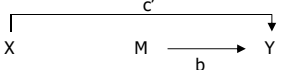
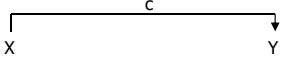


	Analysis	Visual Depiction
Step 1	Conduct a simple regression analysis with X predicting Y to test for path <i>c</i> alone, $Y = B_0 + B_1X + e$	
Step 2	Conduct a simple regression analysis with X predicting M to test for path <i>a</i> , $M = B_0 + B_1X + e$.	
Step 3	Conduct a simple regression analysis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

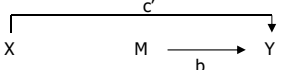

The above four-step approach is the general approach that most researchers followed traditionally. There are potential problems with this approach, however. One problem is that we do not ever really test the significance of the indirect pathway—that X affects Y through the compound pathway of *a* and *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 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 and the best way to construction the confidence interval, however. 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 currently—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. 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, using the products of standardized coefficients from Model 1 and Model 2 above, $(\beta_2)(\beta)$, though they may not be reported by the software program, and other methods such as the ratio of indirect to total effect can also be computed (see Preacher & Kelley, 2011, for a review).

Statistical Software

There are several possible computer methods of estimating and testing indirect effects, and I will focus on two (the PROCESS macro for SPSS and the `mediation` package for R) that use the percentile bootstrap method in the subsequent handout "Testing Mediation with Regression Analysis Example." Both of these methods estimate the indirect tests and confidence limits from the model and data in a single step.¹ The `RMediation` package (Tofighi & MacKinnon, 2011) will estimate confidence limits with the Monte Carlo method.

Another approach is to use a structural equation modeling (SEM, also called covariance structure analysis) software program (e.g., Mplus, lavaan package in R, AMOS, LISREL). SEM is designed, in part, to test these more complicated models in a single analysis instead of testing separate regression analyses, but simple mediation models (or "path" models) can also be tested. Most SEM software packages now offer indirect effect tests using one of the above approaches for determining significance (for an example, see https://web.pdx.edu/~newsomj/semclass/ho_pathex.pdf). In addition, the SEM analysis approach provides model fit information that provides information about consistency of the hypothesized mediational model to the data. Measurement error is a potential concern in mediation testing because of attenuation of relationships (Baron & Kenny, 1986; Fritz, Kenny, and MacKinnon, 2016; VanderWeele, Valeri, & Ogburn, 2012) and the SEM approach can address this problem by removing measurement error from the estimation of the relationships among the variables. I will save more detail on this topic for another course, however.

Online resources

David Kenny also has a webpage on mediation: <http://davidakenny.net/cm/mediate.htm>
Andrew Hayes' Process macro site: <http://processmacro.org/index.html>
Mediation package for R: <https://cran.r-project.org/web/packages/mediation/mediation.pdf>
Kristopher Preacher's site with software approaches for the Monte Carlo method among other related material: <http://www.quantpsy.org/medmc/medmc.htm>
RMediation package: <https://cran.r-project.org/web/packages/RMediation/RMediation.pdf>

References and Further Reading

- Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51, 1173-1182.
- Biesanz, J. C., Falk, C. F., & Savalei, V. (2010). Assessing mediational models: Testing and interval estimation for indirect effects. *Multivariate Behavioral Research*, 45(4), 661-701.
- Goodman, L. A. (1960). On the exact variance of products. *Journal of the American Statistical Association*, 55, 708-713.
- Fritz, M. S., Kenny, D. A., & MacKinnon, D. P. (2016). The combined effects of measurement error and omitting confounders in the single-mediator model. *Multivariate Behavioral Research*, 51, 681-697.
- Fritz, M. S., Taylor, A. B., & MacKinnon, D. P. (2012). Explanation of two anomalous results in statistical mediation analysis. *Multivariate Behavioral Research*, 47(1), 61-87.

¹ Of course, the indirect coefficient can be computed by hand (as described above) from the product of B_2 (from Model 1) and B (from Model 2). The standard error can also be computed by hand and a significance test or confidence limits can be obtained in the usual manner. The standard errors can be computed more conveniently using Preacher and Leonardelli's online calculator, <http://quantpsy.org/sobel/sobel.htm>, if the appropriate (direct effect) unstandardized coefficients and their standard errors are entered. The online calculator returns Sobel, Goodman, and Aroian versions of the indirect standard error. Although the Sobel and Aroian tests will likely give fairly accurate statistical tests and confidence limits for large sample sizes, the bootstrap or Monte Carlo tests are preferred for best accuracy in the widest range of circumstances.

- Judd, C.M. & Kenny, D.A. (1981). Process Analysis: Estimating mediation in treatment evaluations. *Evaluation Review*, 5(5), 602-619.
- Hayes, A. F. (2020). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach, second edition*. Guilford Press.
- Hayes, A. F., & Scharkow, M. (2013). The relative trustworthiness of inferential tests of the indirect effect in statistical mediation analysis does method really matter?. *Psychological Science*, 24(10), 1918-1927.
- 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.
- MackKinnon, D.P. (2008). *Introduction to statistical mediation analysis*. Mahwah, NJ: Erlbaum.
- MackKinnon, D.P. & Dwyer, J.H. (1993). Estimating mediated effects in prevention studies. *Evaluation Review*, 17(2), 144-158.
- MackKinnon, D.P., Fairchild, A.J., & Fritz, M.S. (2007). Mediation analysis. *Annual Review of Psychology*, 58, 593-614.
- MackKinnon, 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.
- 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.
- Preacher, K. J., & Kelley, K. (2011). Effect size measures for mediation models: quantitative strategies for communicating indirect effects. *Psychological methods*, 16(2), 93.
- 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.
- Tofighi, D., & MackKinnon, D. P. (2011). RMediation: An R package for mediation analysis confidence intervals. *Behavior Research Methods*, 43, 692-700.
- Tofighi, D., & MackKinnon, D. P. (2016). Monte Carlo Confidence Intervals for Complex Functions of Indirect Effects. *Structural Equation Modeling: A Multidisciplinary Journal*, 23, 194-205.
- VanderWeele, T. J., Valeri, L., & Ogburn, E. L. (2012). The role of measurement error and misclassification in mediation analysis. *Epidemiology*, 23, 561-564.