Preacher and Hayes 2008: Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models

* Hypothesis using mediation is very common in behavioral science/psychology
* Mediation is when a predictor affects a dependent variable INDIRECTLY through at least one intervening variable (the mediator).
  + Assessing multiple simultaneous mediators is difficult, and has not been studied, but is clearly needed.
* Overview of simple and multiple mediation, as well as several approaches to investigate the process.
  + Additionally how to contrast 2+ mediators within a model.
* Correlations b/w variables is IMPORTANT, as correlation is a necessary but not sufficient condition to prove relationship.
  + HOW or WHY a causal is also interesting, but generally involves *mediation* analysis, how some variables affect others through intervening (mediating) variables.

Simple Mediation

Diagram

Description automatically generated

* How X affects Y through M.
  + NOTE: It is important to establish the causal order of X, M, and Y on theoretical/procedural ground
  + Total effect of X on Y is the sum of direct and indirect effects, *c = c’+ ab*, thus *c’* is the difference b/w the total effect of X on Y and the indirect effect of X on Y through M.
    - *c’ = c – ab* : These identities hold in regression and SEM where M and Y are continuous, but not where one or more of the dependent variables are binary, we then need to use logistic or probit regression, in which case the identity does not hold (MacKinnon & Dwyer, 1993).

General overview of how to test mediation hypothesis have been proposed (see MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002, for an overview).

* Commonly used path is *causal steps strategy* (Baron and Kenny 1986)
  + Investigator estimates paths of model in figure 1 using OLS regression or SEM, and assess the extent to which some criteria are met.
  + Variable M is a mediator if X significantly accounts for variability in M, X significantly accounts for variability in Y, and M significantly STILL accounts for variability in Y after controlling for X, and the effect of X on Y decreases substantially when M is entered simultaneously as a predictor of Y.
  + The last criterion is satisfied when the first and third criteria are satisfied, and when the signs of the effects are consistent w/ the proposed mediation process.
  + Using the diagram above, criteria requires paths *a, b,*  and *c*, to be significant, *c’* to be smaller than *c* by a nontrivial amount.
  + Note some authors (Collins, Graham, & Flaherty, 1998; Judd & Kenny, 1981; Kenny

et al., 1998; MacKinnon, 1994, 2000; MacKinnon, Krull, & Lockwood, 2000; Shrout & Bolger, 2002) argue that a significant total effect of X on Y (quantified as *c* in the diagram) is NOT necessary for mediation to occur

Other approaches are not based on individual paths in the mediation model, but instead on the product term *ab*, because this product is equal to the difference b/w total and direct effect!

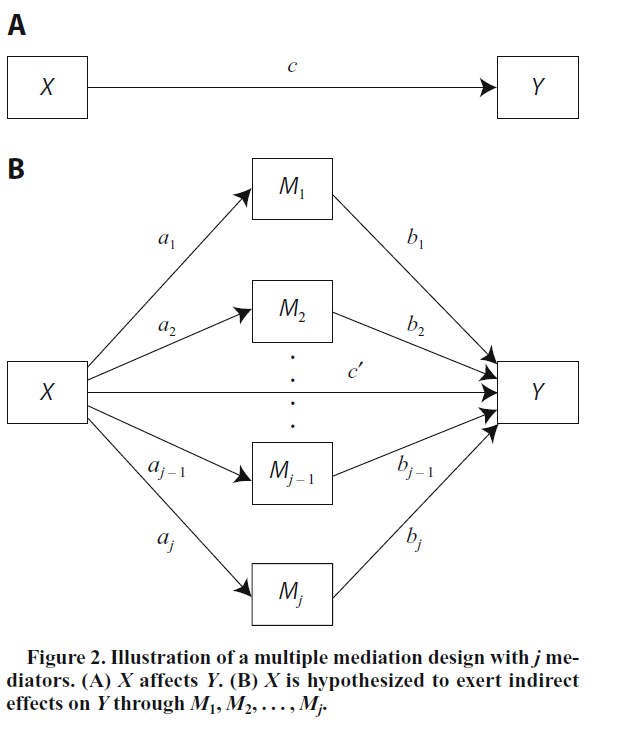
* The Sobel Test (Sobel, 1982, 1986) aka *product of coefficients* approach requires directly computing the ratio of ab to it’s estimated standard error (SE). Several formulas exist to estimate them, but the differences are negligible.
  + A *p* value is computed in reference to standard normal distribution, and significance itself supports the mediation hypothesis.
* However, using a standard normal for deriving *p*  for the indirect effect could be a problem b/ the sampling distribution of *ab* is only normal in LARGE samples.
  + *Distribution of the product approach* is a solution, based on the inference of the mathematical derivation of the distribution of the product of two normally distributed variables. Acknowledging how the distribution of products will be skewed and not requiring the assumption of normality. R code available (MacKinnon, Fritz, Williams, & Lockwood, 2007).

*Bootstrapping,* the nonparametric resampling procedure, is ANOTHER method for testing mediation that DOES NOT require normality on the sampling distribution.

* Computationally intensive, requires repeatedly sampling from the data set and estimating the indirect effect in each resampled data set.
  + By repeating thousands of times, an empirical APPROXIMATION of the sampling distribution for *ab* is built and used to construct confidence intervals for the indirect effect.
  + Details can be found with Bollen and Stine (1990), Lockwood and MacKinnon (1998), MacKinnon et al. (2004), Shrout and Bolger (2002), and Preacher and Hayes (2004, 2008).

Overall, these methods have been examine with simulations to asses Type I error rates and power, and the distribution of product approach or bootstrapping are seen as better than Sobel test or causal steps approach

* Because the first two have higher power while maintaining reasonable control over Type I error rate.
* Especially the causal steps strategy CANNOT be recommended except in large samples, see Preacher and Hayes (2004) and MacKinnon et al. (2002)

Multiple Mediation

* Simultaneous mediation by multiple variables w/ several
* Can have several X’s purportedly affecting a single Y
* The analytic methods for multiple mediation are MORE complex than for simple mediation
* Illustrated through figure 2, there are many paths, the indirect effects of X on Y via the j number of mediators.
  + Specific indirect effect of X on Y via mediator *i* however is the product of the two unstandardized paths linking x to y via that mediator
  + The effect of X on Y through M1 is a1b1, the TOTAL indirect effect of x on y is the sum of the specific indirect effects, the TOTAL effect of X on Y is the sum of the direct effect and all j of the specific indirect effects.
    - C = c’ + sum of indirect effects, the total indirect effect is c – c’
* Testing multiple mediation model instead of several separate simple mediation models is good!
  + Testing total effect of X on Y is like regression w/ several predictors
  + If an effect is found, can conclude that these j variables mediate the effect of X on Y.
  + And, what extent specific M variables mediate the X into Y effect, conditional on the presence of other mediators in the model.
  + Likelihood of parameter bias due to omitted variables is reduced!
    - Otherwise, can have biased parameter estimates, and has the issue of examining multiple comparisons instead of less comparisons (inflates type I error rate)
* It’s difficult to tease apart individual mediating effects that can often overlap in content.
  + The specific effect of M3 for example, isn’t just the effect of M3 alone, but conditional on all the other Mi­ in the model.
  + Thus, multicollinearity is an issue (same as in multiple regression!)

Investigating multiple mediation should involve 2 parts:

1. Investigating the total indirect effect (deciding which set of mediators translates the effect of X on Y
2. Testing hypothesis regarding individual mediators in the context of a multiple mediator model (the specific indirect effect associated w/ each mediator)
   1. A significant total indirect effect is NOT needed to investigate specific indirect effects.
   2. It is possible to find specific indirect effects to be significant in the presence of a nonsignificant total indirect effect!

Several approaches exist, primarily similar in ways to the original methods for testing simple mediation.

Causal Steps Approach

Generally used to find out whether or not c – c’ is a mediation effect. Here, the investigator asks whether the paths defining a specific indirect effect (ai and bi) are significant. If either of the paths through variable Mi is NOT different from 0, then Mi is not a mediator for effect of X on Y.

* Note: It’s possible for one variable to act as mediator M1, and another to act as a suppressor M2, in effect M1 and M2 cancel each other out (see MacKinnon et al., 2000)
* This is an EASY to understand extension, but it has the same problems as the single mediator version, thus it’s not used particularly commonly. It relies on a set of tests for individual *a* and *b* paths rather than testing the specific indirect effects, AND yields no point estimate or SE of the mediation effect.

Product-of-Coefficients Approach

* This works just fine for multiple mediators! Uses multivariate delta method to derive the SE of the total indirect effect (aka c – c’).
* The specific indirect effects can be investigated later through individual mediators.
* Total indirect effect for a model including the three mediators is the sub of the specific indirect effects!
  + ***F*** = a1b1 + a2b2 + a3b3
  + A more complex formula, using methods from Bollen (1987, 1989) generates the variance of the effect, **F**.
    - The square root of var(F) is the first order SE of the total indirect effect in a 3-mediator model. Assuming normality for the total indirect effect
    - A second-order ver. Of the multivariate delta method can be used, but the accuracy is only slightly improved.
  + Can either add or remove terms as necessary for larger or smaller than 3 mediator models from formula 1 in the paper.
* If using path analysis or SEM to fit, the residuals should be allowed to covary.
  + Obviously, because they all mutually depend on X! Thus covariance is somewhat expected.
  + If covariances are constrained to zero and actually correlated, the model will be misspecified, and the SE’s will be very biased, but the point estimates of a or b coefficients will still be correct.

Distribution of the Product Strategy

Can test specific indirect effects, however the distribution of sums or differences of products (needed to test hypothesis about total indirect effects/pairwise contrasts is NOT currently solved. Does not exist!

Bootstrapping

Super practical method! One assumption needed for use of SE’s derived via delta method (or a limitation of this multivariate extension for the product-of-coefficients strategy) is the need for multivariate normality!

* Not just the paths, but the sampling distributions of the total and specific indirect effects must be assumed to be normal for p-o-e strategy!
* We can use bootstrapping to solve multiple mediation, because we don’t HAVE to assume normality!

To bootstrap the sampling distribution of the specific and total indirect effects take a sample of size n cases with replacement from the original sample.

* A given case can be selected as part of a bootstrap sample 0, 1, 2 or even MORE times!
* Using this resample of size N, reestimate all j values of ai and bi­ and then calculate the product (ai \* bi) and the sum, from our resampled dataset.
  + Repeat this process k times, where k is at least 1000, yielding k estimates of the total and specific indirect effects of X on Y.
* Distributions of these K estimates serve as empirical, nonparametric approximations of the sampling distributions of the indirect effects of interest.
* Our bootstrap confidence interval for the population specific indirect effect through M1 is derived by sorting the k values of ai \* bi from low to high.
  + The lower and upper 100(alpha/2)% of the distribution are then found and taken as the lower and upper limits of the 100(1-alpha)% CI for the population indirect effect, where alpha is our desired nominal Type I error rate.
  + For example, with alpha of .05, we have 95% CI, and if K is 1000, we use the 25th and 976th values of ai \* bi in our sorted distribution. This is a PERCENTILE bootstrap CI.
    - Note, these can be asymmetrical b/c it’s based on empirical estimation rather than the assumption that the distribution is normal
* Bootstrapping is generally superior to product-of-coefficients in small/moderate samples

Contrasting Indirect Effects in Multiple Mediator Models

Sometimes need to test hypothesis that two indirect effects are equal in size.

* For example, which theory has greater impact in multiple mediation of the same model.

MacKinnon (2000) provides the only statistical treatment of contrast hypotheses such as these.

Example of Multiple Mediation

Testing hypothesis about early employee socialization (preentry knowledge, helpfulness of socialization agents) and how it affects socialization outcomes (job satisfaction, commitment, and cliarity).

* Proposed mediators are different types of socialization content expertise: Organizational goals and values, people, history, job performance proficiency, and politics (5 mediators)
* Theory was that early socialization experience is related to socialization outcomes, through how they affect the various elements of organizational socialization.
* Example is sub-set of hypothesis linking helpfulness of socialization agents to future job satisfaction.
  + Mediated by job proficiency, good work relationships, and understanding of workplace politics.
  + Only found indirect effect for the ‘people’ dimension.
* The total indirect effect of X on Y is f = a1b1 + a2b2 + a3b3 = .1074
  + Solved using equation 1, with var(f) = 0.0009719
  + Then Z = f/sqrt(var(f)) = 3.445
    - Reject H0 that the indirect effect is zero, p = 0.0006
* However, when directly comparing whether or not the effect of politics was significantly different than the effect of people, we can contrast the sampling variance, and determine that for both of them, there is 0 contained w/in the interval, thus even though people is significant, there is no significant difference b/w people and politics on how helpfulness affects job satisfaction.

Note, one big advantage of using SEM w/ latent variables is that unlike regression, you explicitly model measurement error. This allows you to test hypothesis using latent constructs rather than imperfect measured indicators!

MacKinnon 2002: A Comparison of Methods to Test Mediation and Other Intervening Variable Effects

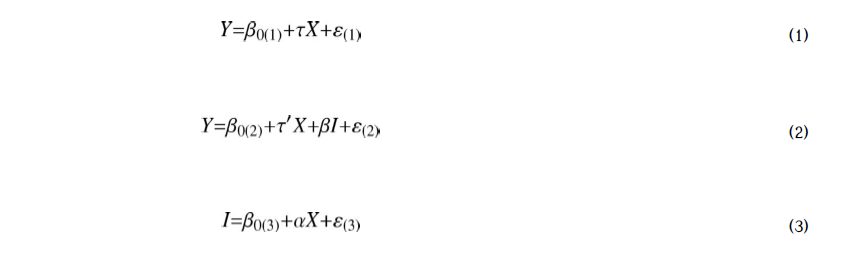
* Overview of a Monte Carlo study comparing various methods to test statistical significance of intervening variable effect (aka a mediator). The mediator transmits the effect of an independent variable to a dependent variable.
  + 2 methods based on distribution of product and 2 methods based on difference-in-coefficient methods have the most accurate type I error rates and greatest statistical power.
* X > I > Y , the effect of independent variable X on intervening variable I to affect dependent variable Y
  + Not used as commonly as should be, b/c either people don’t know the methods, there are too many methods to choose, or some ppl feel like the methods have too low statistical power.

There are three general schools of thought for methods when analyzing mediation.

1. Causal Steps Approach: Specifies a series of tests of links in a ‘causal chain’. Traditionally based on the work of Judd and Kenny (1981), and Baron and Kenny (1986) and is very commonly used.
2. Difference-in-Coefficients: Methods such like those that compare the difference b/w a regression coefficient before and after adjusting for the mediator (Freedman & Schatzkin, 1992;McGuigan & Langholtz, 1988;Olkin& Finn, 1995). Some of these methods test hypothesis about intervening variables that diverge from what psychologists are ‘used to’.
3. Product-of-Coefficients: Uses the product of coefficients involving paths in a path model (aka the indirect effect; Alwin & Hauser, 1975;Bollen, 1987;Fox, 1980;Sobel, 1982,1988).
   1. We use term intervening variable to refer to all non causal steps approaches for analyzing mediation.

Main goal is to simulate and thus compare/contrast Type I error rates and Statistical Power for all of these different mediation methods

* If power is too low, will not detect real effects in the population
* With Type I error too high, you can risk finding nonexistent effects.



Basic intervening Variable Model

* X is independent variable, Y is dependent, and I is intervening, our three B0 are the population regression intercepts for equation 1,2, and 3. t represents the relationship b/w independent and dependent variables after adjustment for intervening variables in equation 2, alpha represents the relationship b/w independent and intervening variables in equation 3, and beta represents tine relation b/w intervening and dependent variables adjusted for the effect of independent variable in equation 2.