Chapter 4: Simulations

- Mediation equations.
- Other tests of mediation.
- Comparison of mediation tests.
- Statistical Simulation Studies

1

Mediation Regression Equations

- Tests of mediation use information from some or all of the three equations.
- The coefficients in the equations may be obtained using methods such as ordinary least squares regression, covariance structure analysis, or logistic regression.

2

Three Major Types of Single Sample Tests for the Mediation Effect

- (1) Causal Steps (Baron & Kenny, 1986; Judd & Kenny, 1981).
- (2) Difference in Coefficients: $\hat{c}-\hat{c}'$ estimator (e.g., Clogg et al., 1992)
- (3) Product of Coefficients: $\hat{a}\hat{b}$ estimator (e.g., Sobel, 1982)
- See MacKinnon et al. (2002), Psychological Methods article for a review and comparison of single sample tests

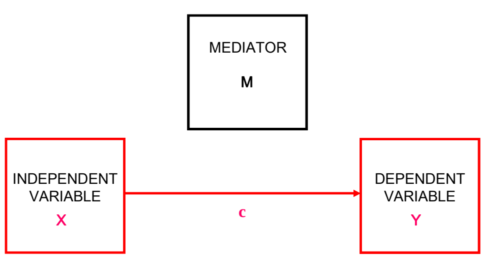
3

Causal Steps Tests of Mediation

- Causal Step 4 from Judd & Kenny (1981): test that $\hat{c}' = 0$ is nonsignificant (i.e., complete mediation required).
- Causal Step 4 from Baron & Kenny (1986): drop in magnitude of sample estimates from $\hat{c}$ to $\hat{c}'$
- Test of joint significance: test whether the $\hat{a}$ and $\hat{b}$ paths are statistically significant (MacKinnon et al., 2002).

4

Equation 1

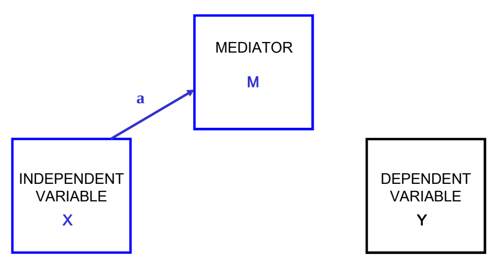


1. The independent variable is related to the dependent variable:

$$Y = i_1 + \hat{c}X + e_1$$

5

Equation 2

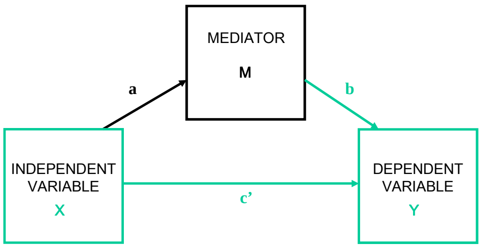


2. The independent variable is related to the potential mediator:

$$M = i_2 + \hat{a}X + e_2$$

6

Equation 3



3. The mediator is related to the dependent variable controlling for exposure to the independent variable:

$$Y = i_3 + \hat{c}' X + \hat{b} M + e_3$$

7

Mediated Effect Measures

Indirect Effect = Mediated effect = $\hat{a}\hat{b} = \hat{c} - \hat{c}'$

Direct effect= $\hat{c}'$

Total effect= $\hat{a}\hat{b} + \hat{c}' = \hat{c}$

8

Product of Coefficients

Corresponding standard errors of *ab*:

$$s_{First} = \sqrt{\hat{a}^2 s_{\hat{b}}^2 + \hat{b}^2 s_{\hat{a}}^2}$$

$$s_{Second} = \sqrt{\hat{a}^2 s_{\hat{b}}^2 + \hat{b}^2 s_{\hat{a}}^2 + s_{\hat{a}}^2 s_{\hat{b}}^2}$$

$$s_{Unbiased} = \sqrt{\hat{a}^2 s_{\hat{b}}^2 + \hat{b}^2 s_{\hat{a}}^2 - s_{\hat{a}}^2 s_{\hat{b}}^2}$$

9

Difference in Coefficients

General standard error formula:

$$s_{\hat{c}-\hat{c}'} = \sqrt{s_{\hat{c}}^2 + s_{\hat{c}'}^2 - 2 r_{\hat{c}\hat{c}'} s_{\hat{c}} s_{\hat{c}'}}$$

Clogg, Petkova, and Shihadeh (1992) variance:

$$s_{\hat{c}-\hat{c}'} = \sqrt{s_{\hat{c}'}^2 |r_{XM}|}$$

Covariance between $\hat{c}$ and $\hat{c}'$, McGuigan & Langholz (1988) generalized to more cases by MacKinnon et al. (2002)

$$s_{\hat{c}\hat{c}'} = r_{\hat{c}\hat{c}'} s_{\hat{c}} s_{\hat{c}'} = MSE / ((s^2{}_X)(N))$$

10

Tests evaluated

(1) Baron & Kenny Causal Steps (Baron & Kenny, 1986)
(2) Joint Significance test (MacKinnon et al., 2002)
(3) Delta Method is the first order standard error of *ab* (Sobel, 1982).
(4) Distribution of the Product (MacKinnon et al., 2002) uses the distribution of the product to form a confidence intervals and assesses significance by evaluating whether 0 is in the confidence interval.
(5) Lots of other tests evaluated in the simulation study. Resampling tests will be described later, e.g., the bootstrap. See the cited articles for more on these tests.

11

Steps in a Statistical Simulation

(1) Generate sample data under a known population model.
(2) Estimate model coefficients and standard errors in the sample.
(3) Save the estimates, standard errors, and results of statistical tests in the sample.
(4) Repeat Steps 1 to 3 a large number of times. The number of times that steps 1-3 are repeated are the replications.
(5) Compare results across all replications to the population values. Which tests led to the most accurate decisions about the population value?

12

Simulation Design: MacKinnon et al., 2002

- All possible combinations of *a, b*, and *c*ʹ effect sizes for zero, small (2% variance explained), medium (13%), and large (26%) effects.
- 5 Sample sizes, N= 50, 100, 200, 500, and 1000
- 500 Replications of each of the 4 X 4 X 4 X 5 generated data sets.
- Type I error and Power
- 14 Tests
- Causal step, difference in coefficients, and product of coefficients tests.

Simulation Results: MacKinnon et al., (2002) Conclusions

Taking all situations, both paths zero, one path zero and the other path nonzero for Type I error rates, and power for nonzero mediation relations.

Tests differ widely in statistical performance.

Best tests are: (1) the joint significance test of the *a* and *b* paths
(2) a test based on forming confidence limits using the distribution of the product (test significance by whether 0 is in the confidence interval).

Reasons for Differences Among Methods

- Requirement for significant total effect, $\hat{c}$, and requirement that $\hat{c}'$ is nonsignificant reduces statistical power of BK and JK causal steps methods.
- Assumption that the mediated effect divided by its standard error has a normal distribution is incorrect in some situations.
- Mediation is fundamentally a test of two paths corresponding to the *a* and *b* paths.

Fritz & MacKinnon, 2007

- Purpose of the study is to obtain required sample size to have .8 power to detect the mediated effect given population values of *a, b*, and *c*ʹ effect sizes for small (S), medium (M), halfway between small and medium (H), and large (L) effects.
- Table 3 presents these values for Baron & Kenny, joint significance, Delta (first order), PRODCLIN, percentile bootstrap, and bias-corrected bootstrap methods.
- Required sample size determined empirically using a iterative procedure.

Empirical Sample size estimates for .8 power to detect the mediated effect

Test	S-S	S-M	S-L	M-S	M-M	M-L	L-S	L-M	L-L
Baron/Kenny ($\tau' = 0$)	20886	3039	1561	2682	397	204	1184	175	92
a & b Joint	530	403	403	405	74	58	405	59	36
Delta	667	422	412	421	90	66	410	67	42
PRODCLIN	539	401	402	404	74	57	404	58	35

Note: Table entries are based on empirical simulation so they are not exact. Fritz & MacKinnon (2007).

Results: Fritz & MacKinnon, 2007

- Sample size requirements are large for .8 power to detect a mediated effect—around 400 if one of the effects is not small.
- Excessive sample size requirements for the Baron & Kenny method because of the requirement for a significant total effect $\hat{c}$. This occurs because when the direct effect is zero the value of $\hat{c}$ is the product of the two paths in the mediated effect. So if both paths are small then the total effect is the product of two small effects.
- Excessive sample size for .8 power to detect $\hat{c}$ for the product of two small mediation paths is correct (p. 238).
- Best tests: joint significance, distribution of the product or bias-corrected bootstrap (there is some evidence that the bias-corrected bootstrap has increased Type I error rates in some, albeit rare, situations).

New Methods for Power for Complex Mediation Models

- Thoemmes, MacKinnon, & Reiser (2010) describe a general procedure to calculate power for any mediation model. The paper uses Mplus to conduct the power calculations.
- Some of the models covered in that paper are multiple mediator models, latent variable models, moderator and mediator models, and longitudinal mediation models.
- This does require that you can come up with educated guesses of the parameter values and variability for many different parameters.

Thoemmes, F., MacKinnon, D. P., & Reiser, M. R. (2010). Power Analysis for Complex Mediational Designs Using Monte Carlo Methods. *Structural Equation Modeling*, *17*, 510-534.

19

Mediation as a Way of Increasing Power

- O'Rourke and MacKinnon (2013) discuss situations in which including a mediator will increase power to detect effects over a bivariate relation between X and Y
- When *ab* is equal to *c* (*c'* is zero), the test of mediation will always have more power than the test of the total effect
- This occurs when the standard error of *c* is larger than the standard error of ***ab.***
- These results also apply to the two mediator and sequential mediation models.

O'Rourke, H. P., & MacKinnon, D.P. (2015). When the test of mediation has more power that the test of the total effect. *Behavior Research Methods, 47*, 424-442.
Ways to increase statistical power: Fritz, M. S., *Cox, M. G., & MacKinnon, D. P. (2015). Increasing Statistical Power in Mediation Models without Increasing Sample Size. *Evaluation and the Health Professions*, *38*(3), 343-366.

20

Confidence Limits (MacKinnon, Lockwood, & Williams, 2004)

- Many single sample tests have low power
- Earlier studies (MacKinnon et al., 1995) found that confidence limits for the mediated effect are imbalanced especially for small sample sizes and small effect sizes
- Some problems with testing for mediation because the distribution of the product is normal only in special cases.
- Resampling methods may solve the problem.

21

Options to make Confidence Limits

- Normal theory yields symmetric confidence limits.
- Distribution of the Product for asymmetric confidence limits.
- Resampling methods for asymmetric confidence limits—many different types of resampling methods including the bootstrap and jackknife.
- Which confidence limits are the most accurate?

22

Resampling Steps: Confidence Limits

1. Estimate mediated effect in the original sample
2. Generate new data based on rearranging or sampling original data
3. Calculate effect in the generated data
4. Repeat steps 2 and 3 a large number of times
5. Create empirical distribution of the effect from generated and original data
6. Compute UCL and LCL in the empirical distribution

23

Resampling Simulation Design

- 10 combinations of effect size for the a and b paths: z,z; z,s; z,m; z,l; s,s; s,m; s,l; m,m; m,l; l,l
- 4 Sample sizes, N= 25, 50, 100, and 200
- 1000 Replications so there are 4 X 10 X 1000 = 40,000 generated data sets in Study 1. But there are also 1000 resamples in Study 2 so that there are actually, 40,000,000 data sets in that study.
- Study 1 compared normal and distribution of the product confidence limits. Study 2 evaluated many resampling tests
- Type I error, Power, Confidence limit coverage

24

Results (MacKinnon, Lockwood, & Williams, 2004) #1

- Study 1 demonstrated the superiority of the distribution of the product confidence limits over the normal theory confidence limits.
- Study 2 demonstrated that resampling methods work as well as the distribution of the product and both are better than normal theory based confidence limits

25

Results (MacKinnon, Lockwood, & Williams, 2004) #2

Bias-corrected bootstrap most accurate overall but can be cumbersome and there are situations where the Type I error rate is over .05 (see Fritz et al., 2012). Percentile method works well.

Bootstrap is available in Amos (Arbuckle & Wothke, 1999) EQS (Bentler, 1997), LISREL (Joreskog & Sorbom, 2001), Mplus (Muthen & Muthen) and a SAS program (Lockwood & MacKinnon, 1998), SAS and SPSS (Preacher & Hayes, 2008)

Single sample Distribution of the Product CL is the best single sample method and does not have cases where the Type I error rate is as high as the bias-corrected bootstrap.

26

Other Mediation Simulation Studies

Inconsistent Mediation (MacKinnon, Krull, & Lockwood, 2000, *Prevention Science*).

Logistic and probit regression (MacKinnon et al. 2007, *Clinical Trials*).

Path Analysis models (Williams & MacKinnon, *Structural Equation Modeling*, 2008)

Multilevel models. (Krull & MacKinnon, 2001)

Pituch, Whittaker, & Stapleton (2005) replicated superior results of the distribution of the product methods (*Multivariate Behavioral Research*)

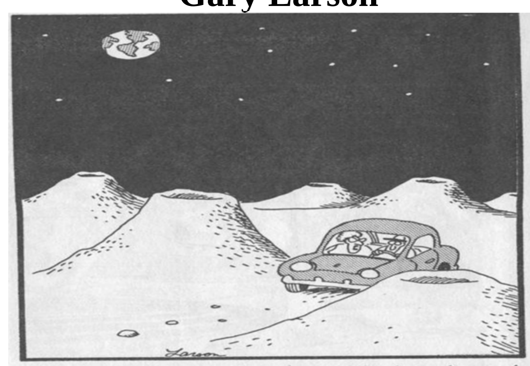
Bayesian Mediation Analysis (Yuan & MacKinnon, 2009, *Psychological Methods*.

Median Regression Mediation Analysis (Yuan & MacKinnon 2014, *Psychological Methods*).

.

27

Gary Larson



"For heaven's sake, Elroy! . . . NOW look where the earth is! . . . Move over and let me drive!"