Chapter 822

Tests of Mediation Effect using the Sobel Test

Introduction

This procedure computes power and sample size for a mediation analysis of a continuous dependent (output) variable Y and an independent (input) variable X. Interest focuses on the interrelationship between Y, X, and a third variable called the mediator M. The sample size calculations are based on the work of Sobel (1982).

Mediation Model

An in-depth discussion of mediation can be found in Hayes (2018). A popular method for testing for mediation is that of Baron and Kenny (1986). In this method, three regression models are fit where $M \sim \mathrm{N}(\mu_M, \sigma_M^2)$ and $X \sim \mathrm{N}(\mu_x, \sigma_X^2)$.

$$(1)\quad M = \theta_0 + \theta_X X + e_M,\ \ e_M \sim \mathrm{N}\left(0, \sigma_{e_M}^2\right)$$

$$(2)\quad Y = \beta_0^* + \beta_X^* X + e_{Y^*}$$

$$(3)\quad Y = \beta_0 + \beta_X X + \beta_M M + e_Y,\ \ e_Y \sim \mathrm{N}\left(0, \sigma_{e_Y}^2\right)$$

These coefficients have the following interpretations: β_X is called the *direct effect* of X on Y, $\theta_X \beta_M$ is called the *indirect effect* of X on Y, and β_X^* is called the *total effect* of X on Y.

The indirect effect comes from substituting the equation for M in (1) into equation (3) and then rearranging terms.

$$\begin{aligned} Y &= \beta_0 + \beta_X X + \beta_M (M) + e_Y \\ &= \beta_0 + \beta_X X + \beta_M (\theta_0 + \theta_X X + e_M) + e_Y \\ &= \beta_0 + \beta_X X + \beta_M \theta_0 + \beta_M \theta_X X + \beta_M e_M + e_Y \\ &= (\beta_0 + \beta_M \theta_0) + (\beta_X + \boldsymbol{\beta_M \theta_X}) X + (\beta_M e_M + e_Y) \end{aligned}$$

Comparing coefficients, we can see that the *total effect* is exactly equal to the *direct effect* plus the *indirect effect*.

Mediation is likely if all four of the following tests are significant:

1. Test of θ_X.
2. Test of β_X^*.
3. Test of β_M.
4. Sobel's (1982) Test of whether β_X is significantly smaller than β_X^* using

$$z = \theta_X \beta_M / \sqrt{\theta_X^2 \mathrm{V}(\beta_M) + \beta_M^2 \mathrm{V}(\theta_X)}.$$

where $\mathrm{V}(\theta_X) = \frac{\sigma_{e_M}^2}{N\sigma_X^2}, \mathrm{V}(\beta_M) = \frac{\sigma_{e_Y}^2}{N\sigma_{e_M}^2},\ \rho_{XM} = \frac{\theta_X \sigma_X}{\sigma_M}$, and $\sigma_{e_M}^2 = \sigma_M^2(1 - \rho_{XM}^2)$

This procedure provides a power analysis and sample size calculation of the z test in step 4.

Calculating the Power

Power calculations are based on standard normal distribution. They proceed as follows:

1. Determine the critical value $z_{1-\alpha}$ from the standard normal distribution where α is the probability of a type-I error.
2. Calculate: $z_\beta = \frac{\theta_X \beta_M}{\sqrt{\theta_X^2 \mathrm{V}(\beta_M) + \beta_M^2 \mathrm{V}(\theta_X)}} - z_{1-\alpha}$.
3. Calculate: $\text{Power} = \Phi(z_\beta)$.

Notes

1. Use $\frac{\alpha}{2}$ instead of α for two-sided test.
2. $\sigma_M^2 = \Pr(M = 1)\Pr(M = 0)$ if M is binary.
3. $\sigma_X^2 = \Pr(X = 1)\Pr(X = 0)$ if X is binary.

Example 1 – Finding Sample Size

Researchers are studying the relationship between a dependent variable (Y) and an independent variable (X). They want to understand the impact of a third variable (M) on the relationship between X and Y, so they decide to carry out a mediation analysis. They decide to use Sobel's test for their sample calculation. Using prior analyses, they decide to use θ_X = 0.2, 0.3, 0.5; β_M = 0.2, 0.4, 0.6, σ_X = 0.6, σ_M = 0.5, and σ_e = 0.2. They set the power at 0.9 and the two-sided significance level at 0.05.

Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 1** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab

Solve For	**N (Sample Size)**
Alternative Hypothesis	**Two-Sided**
Power	**0.90**
Alpha	**0.05**
θ_X (Reg Coef of X)	**0.2 0.3 0.5**
β_M (Reg Coef of M)	**0.2 0.4 0.6**
Type of Primary Predictor, X	**Continuous**
σ_X (Standard Deviation of X)	**0.6**
Type of Mediator, M	**Continuous**
σ_M (Standard Deviation of M)	**0.5**
σ_e (Standard Deviation of e_Y)	**0.2**

Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Reports

Numeric Results

Solve For: N (Sample Size)
Alternative Hypothesis: Two-Sided
Hypotheses: H0: $\theta_X(\beta_M) = 0$ versus H1: $\theta_X(\beta_M) \neq 0$

		Regression Coefficient			Standard Deviation				
Power	Sample Size N	X θ_X	M β_M	Product $\theta_X(\beta_M)$	X σ_X	M σ_M	e_Y σ_e	Product $\sigma(\theta_X(\beta_M))$	Alpha
0.9006	217	0.2	0.2	0.04	0.6	0.5	0.2	0.012	0.05
0.9014	184	0.2	0.4	0.08	0.6	0.5	0.2	0.025	0.05
0.9002	177	0.2	0.6	0.12	0.6	0.5	0.2	0.037	0.05
0.9003	119	0.3	0.2	0.06	0.6	0.5	0.2	0.018	0.05
0.9012	83	0.3	0.4	0.12	0.6	0.5	0.2	0.037	0.05
0.9002	76	0.3	0.6	0.18	0.6	0.5	0.2	0.056	0.05
0.9022	85	0.5	0.2	0.10	0.6	0.5	0.2	0.031	0.05
0.9071	36	0.5	0.4	0.20	0.6	0.5	0.2	0.061	0.05
0.9003	26	0.5	0.6	0.30	0.6	0.5	0.2	0.093	0.05

Model 1 $M = \theta_0 + \theta_X(X) + e_M$
Model 2 $Y = \beta_0 + \beta_X(X) + \beta_M(M) + e_Y$. The e_Y's are normally distributed.
X The primary predictor. It is a continuous, independent variable.
M The mediator. It is a continuous variable.

Power The probability of rejecting a false null hypothesis when the alternative hypothesis is true.
N The number of observations on which the multiple regression is computed.
θ_X The regression coefficient of the primary predictor in Model 1. It is sometimes referred to as a path coefficient.
β_M The regression coefficient of the mediator in Model 2. It is sometimes referred to as a path coefficient.
$\theta_X(\beta_M)$ The product of the coefficients θ_X and β_M.
σ_X The standard deviation of X.
σ_M The standard deviation of M.
σ_e The standard deviation of ei in Model 2.
$\sigma(\theta_X(\beta_M))$ The standard deviation of the coefficient product $\theta_X(\beta_M)$.
Alpha The probability of rejecting a true null hypothesis.

Summary Statements

A mediation effect (single group, Y versus X with mediator M) design will be used to test whether the indirect effect ($\theta_X(\beta_M)$) is different from 0 (H0: $\theta_X(\beta_M) = 0$ versus H1: $\theta_X(\beta_M) \neq 0$). The comparison will be made using a two-sided Sobel (1982) test of the product coefficient ($\theta_X(\beta_M)$), with a Type I error rate (α) of 0.05. The continuous primary predictor, X, is assumed to have a standard deviation of 0.6. The continuous mediator, M, is assumed to have a standard deviation of 0.5. The standard deviation of the residuals from the two-variable regression model is assumed to be 0.2. To detect a primary predictor (X) regression coefficient of 0.2 and a mediator (M) regression coefficient of 0.2 (a mediation effect of 0.04), with 90% power, the number of needed subjects will be 217.

Dropout-Inflated Sample Size

Dropout Rate	Sample Size N	Dropout-Inflated Enrollment Sample Size N'	Expected Number of Dropouts D
20%	217	272	55
20%	184	230	46
20%	177	222	45
20%	119	149	30
20%	83	104	21
20%	76	95	19
20%	85	107	22
20%	36	45	9
20%	26	33	7

Dropout Rate The percentage of subjects (or items) that are expected to be lost at random during the course of the study and for whom no response data will be collected (i.e., will be treated as "missing"). Abbreviated as DR.
N The evaluable sample size at which power is computed. If N subjects are evaluated out of the N' subjects that are enrolled in the study, the design will achieve the stated power.
N' The total number of subjects that should be enrolled in the study in order to obtain N evaluable subjects, based on the assumed dropout rate. After solving for N, N' is calculated by inflating N using the formula N' = N / (1 - DR), with N' always rounded up. (See Julious, S.A. (2010) pages 52-53, or Chow, S.C., Shao, J., Wang, H., and Lokhnygina, Y. (2018) pages 32-33.)
D The expected number of dropouts. D = N' - N.

Dropout Summary Statements

Anticipating a 20% dropout rate, 272 subjects should be enrolled to obtain a final sample size of 217 subjects.

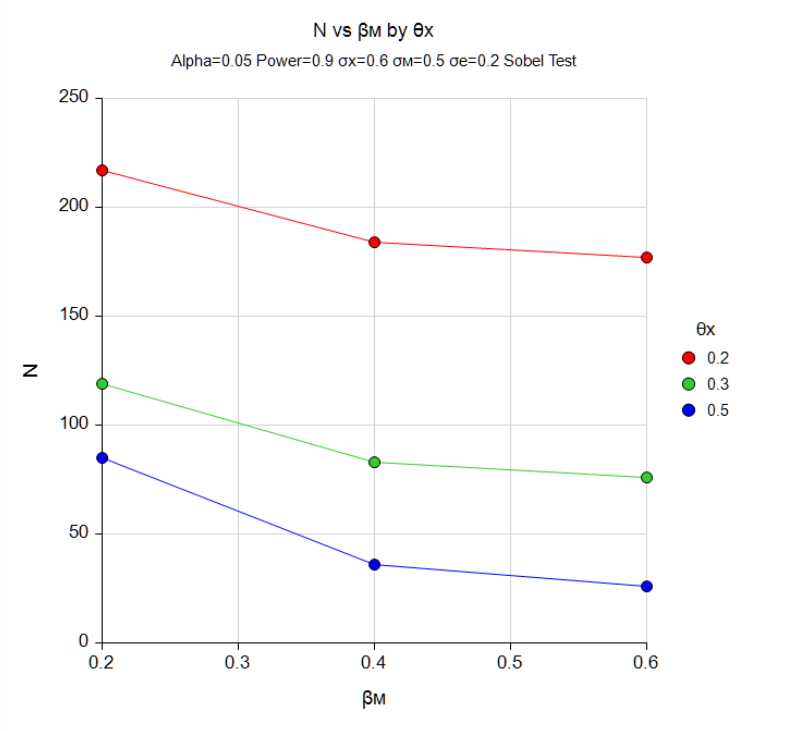
References

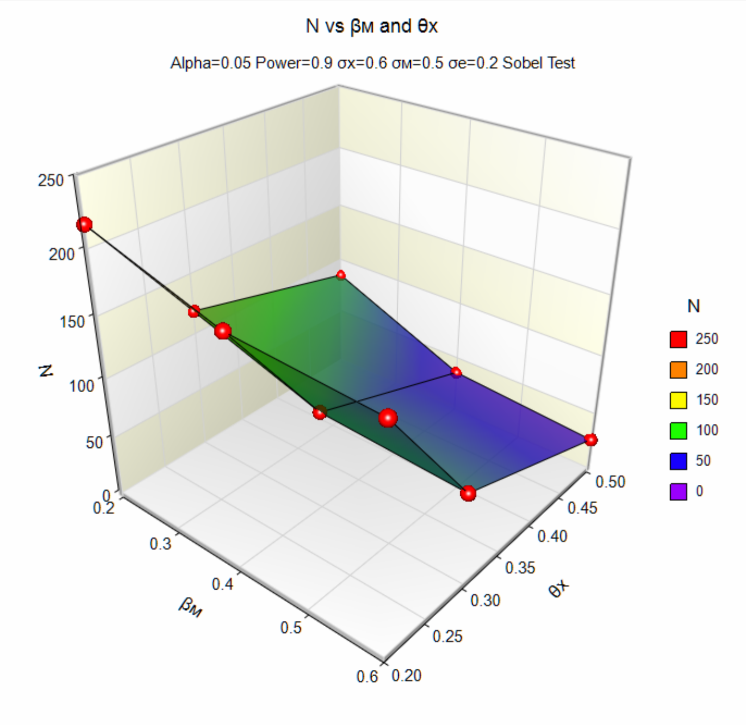
Fritz, M.S. and MacKinnon, D.P. 2007. 'Required Sample Size to Detect the Mediated Effect.' Psychological Science, Vol. 18, No. 3, Pages 233-239.
Hayes, A.F. 2018. Introduction to Mediation, Moderation, and Conditional Process Analysis, Second Edition. CRC The Guilford Press. New York.
Sobel, M. E. 1982. 'Asymptotic confidence intervals for indirect effects in structural equation models.' Sociological Methodology. Vol. 13, Pages 290-312.
Vittinghoff, E., Sen, S., and McCulloch, C.E. 2009. 'Sample size calculations for evaluating mediation.' Statistics in Medicine, Vol. 28, Pages 541-557.

This report shows the necessary sample sizes. The definitions of each of the columns is given in the Report Definitions section.

Plots Section

Plots





These plots show the relationship between sample size and effect size.

Example 2 – Validation using Hand Calculations

We were unable to find a validation example in the literature, so we will present an example calculated by hand. We will use N = 100, θ_X = 0.3; β_M = 0.6, σ_X = 0.7, σ_M = 0.8, and σ_e = 0.2. The two-sided significance level is set at 0.05. The calculation could proceed as follows:

$$\rho_{XM} = \frac{\theta_X \sigma_X}{\sigma_M} = \frac{(0.3)(0.7)}{0.8} = 0.2625$$

$$\sigma_{e_M}^2 = \sigma_M^2(1 - \rho_{XM}^2) = 0.8^2(1 - 0.2625^2) = 0.5959$$

$$\mathrm{V}(\theta_X) = \frac{\sigma_{e_M}^2}{N\sigma_X^2} = \frac{0.5959}{100(0.7^2)} = 0.01216122$$

$$\mathrm{V}(\beta_M) = \frac{\sigma_{e_Y}^2}{N\sigma_{e_M}^2} = \frac{0.2^2}{100(0.5959)} = 0.00067125357$$

$$\begin{aligned} z_\beta &= \frac{\theta_X \beta_M}{\sqrt{\theta_X^2 \mathrm{V}(\beta_M) + \beta_M^2 \mathrm{V}(\theta_X)}} - z_{1-\frac{\alpha}{2}} \\ &= \frac{(0.3)(0.6)}{\sqrt{0.3^2(0.00067125357) + 0.6^2(0.01216122)}} - 1.959964 \\ &= \frac{0.18}{\sqrt{0.00006041282 + 0.0043780392}} - 1.959964 \\ &= 0.741858 \end{aligned}$$

$$\text{Power} = \Phi(0.741858) = 0.7709$$

Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 2** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab

Solve For .... **Power**
Alternative Hypothesis .... **Two-Sided**
Alpha.. **0.05**
N (Sample Size)... **100**
θ_X (Reg Coef of X) **0.3**
β_M (Reg Coef of M)....................................... **0.6**
Type of Primary Predictor, X.......................... **Continuous**
σ_X (Standard Deviation of X)......................... **0.7**
Type of Mediator, M .... **Continuous**
σ_M (Standard Deviation of M) **0.8**
σ_e (Standard Deviation of e_Y)........................ **0.2**

Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results

Solve For: Power
Alternative Hypothesis: Two-Sided
Hypotheses: H0: $\theta_X(\beta_M) = 0$ versus H1: $\theta_X(\beta_M) \neq 0$

		Regression Coefficient			Standard Deviation				
Power	**Sample Size N**	**X θ_X**	**M β_M**	**Product $\theta_X(\beta_M)$**	**X σ_X**	**M σ_M**	**e_Y σ_e**	**Product $\sigma(\theta_X(\beta_M))$**	**Alpha**
0.7709	100	0.3	0.6	0.18	0.7	0.8	0.2	0.067	0.05

PASS matches the calculation by hand of power = 0.7709.