Chapter 827

# Joint Tests of Mediation in Linear Regression with Continuous Variables

## Introduction

This procedure computes power and sample size for evaluating mediation analysis of a continuous dependent variable Y and continuous independent variables X and M. It is assumed that joint testing will be made of both links (regression models) of an indirect pathway assuming that the tests are asymptotically independent. The sample size calculations are based on the work of Vittinghoff and Neilands (2015).

## 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 $X \sim \mathrm{N}(\mu_x, \sigma_X^2)$.

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

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

$$(3)\ \ Y = \beta_0^* + \beta_X^* X + e_{Y^*}$$

Vittinghoff, Sen, and McCulloch (2009) derived sample size formulas based on testing the significance of $\beta_M$ in model 2. Vittinghoff and Neilands (2015) extended these results to give results for testing $\theta_X = 0$ in model 1 and $\beta_M = 0$ in model 2 simultaneously. They added adjustment factors for confounders in model 1 and model 2. They also added an adjustment for cluster randomized designs, if they are used.

### 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 $\alpha$ is the probability of a type-I error.
2. Calculate: $z_{\beta 1} = \sqrt{\frac{N\sigma_X^2\theta_X^2(1-R_X^2)}{\delta\sigma_M^2(1-\rho_{XM}^2)}} - z_{1-\alpha}$.
3. Calculate: $z_{\beta 2} = \sqrt{\frac{N\sigma_M^2\beta_M^2(1-\rho_{XM}^2)(1-R_M^2)}{\delta\sigma_{e_Y}^2}} - z_{1-\alpha}$.
4. Calculate: $\text{Power} = \Phi(z_{\beta 1})\Phi(z_{\beta 2})$.

Notes

1. Use $\frac{\alpha}{2}$ instead of $\alpha$ for two-sided test.
2. $\rho_{XM} = \theta_X \sigma_X / \sigma_M$ must be between -1 and 1.
3. $R_X^2$ is squared multiple correlation of X on any confounding variables.
4. $R_M^2$ is squared multiple correlation of M on any confounding variables.
5. $\delta$ is the design effect which is defined as the ratio of the actual variance of Y with clustering to its variance under simple random sampling.

# 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 determine the sample size based on the joint significance tests. Using prior analyses, they decide to use $\theta_X$ = 0.5, 0.7; $\beta_M$ = 0.5, 0.6, 0.7, $\sigma_X = \sigma_M = \sigma_e$ = DE = 1; $R^2_X$ = 0.1; and $R^2_M$ = 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** |
| $\theta_X$ (Reg Coef of X in Model 1) | **0.5 0.7** |
| $\beta_M$ (Reg Coef of M in Model 2) | **0.5 0.6 0.7** |
| $\sigma_X$ (Standard Deviation of X) | **1** |
| $\sigma_M$ (Standard Deviation of M) | **1** |
| $\sigma_e$ (Standard Deviation of $e_Y$) | **1** |
| DE (Design Effect) | **1** |
| $R^2_X$ ($R^2$ of Confounders in Model 1) | **0.1** |
| $R^2_M$ ($R^2$ of Confounders in Model 2) | **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: $\beta_M = 0$ versus H1: $\beta_M \neq 0$

| | Individual Power | | | Regression Coefficient | | | Standard Deviation | | | R² if X is Regressed on Confounders in Model 1 | R² if M is Regressed on Confounders in Model 2 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Joint Power | Model 1 Pw1 | Model 2 Pw2 | Sample Size N | X θx | M βM | Correlation of X and M ρXM | X σX | M σM | eY σe | R²X | R²M | Design Effect DE | Alpha |
| 0.9002 | 0.996 | 0.904 | 71 | 0.5 | 0.5 | 0.5 | 1 | 1 | 1 | 0.1 | 0.2 | 1 | 0.05 |
| 0.9030 | 0.979 | 0.923 | 53 | 0.5 | 0.6 | 0.5 | 1 | 1 | 1 | 0.1 | 0.2 | 1 | 0.05 |
| 0.9044 | 0.953 | 0.949 | 44 | 0.5 | 0.7 | 0.5 | 1 | 1 | 1 | 0.1 | 0.2 | 1 | 0.05 |
| 0.9027 | 1.000 | 0.903 | 104 | 0.7 | 0.5 | 0.7 | 1 | 1 | 1 | 0.1 | 0.2 | 1 | 0.05 |
| 0.9018 | 1.000 | 0.902 | 72 | 0.7 | 0.6 | 0.7 | 1 | 1 | 1 | 0.1 | 0.2 | 1 | 0.05 |
| 0.9024 | 1.000 | 0.902 | 53 | 0.7 | 0.7 | 0.7 | 1 | 1 | 1 | 0.1 | 0.2 | 1 | 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 an independent variable.
M The mediator. It is the dependent variable in Model 1 and an independent variable in Model 2.

Power The probability of rejecting a false pair of null hypotheses that both $\theta_X$ and $\beta_M$ are zero.
Pw1 The probability of rejecting a false null hypotheses that $\theta_X$ is zero in model 1.
Pw2 The probability of rejecting a false null hypotheses that $\beta_X$ is zero in model 2.
N The number of observations on which the multiple regressions are computed.
$\theta_X$ The regression coefficient of X in Model 2.
$\beta_M$ The regression coefficient of the mediator in Model 2.
$\rho_{XM}$ The correlation between X and M.
$\sigma_X$ The standard deviation of X.
$\sigma_M$ The standard deviation of M.
$\sigma_e$ The standard deviation of $e_Y$ in the model.
$R^2_X$ The $R^2$ achieved when X is regressed on any confounder variables in Model 1.
$R^2_M$ The $R^2$ achieved when M is regressed on any confounder variables in Model 2.
DE The design effect (the ratio of actual variance of the outcome, Y, with clustering to its variance assuming simple random sampling).
Alpha The probability of rejecting a true null hypothesis.

**Summary Statements**

A two-model mediation effect (single group, Y versus X with mediator M) design will be used to test whether the regression coefficient of X ($\theta_X$) in model 1 is different from 0 (H0: $\theta_X = 0$ versus H1: $\theta_X \neq 0$), and whether the mediation effect ($\beta_M$) in model 2 is different from 0 (H0: $\beta_M = 0$ versus H1: $\beta_M \neq 0$). Models 1 and 2 are assumed to be independent. The comparisons will be made using two two-sided linear regression tests of the two coefficients ($\theta_X$ and $\beta_M$), with each test having a Type I error rate ($\alpha$) of 0.05. The standard deviation of the primary predictor variable, X, is assumed to be 1. The standard deviation of the mediator variable, M, is assumed to be 1. The standard deviation of the residuals in model 2 is assumed to be 1. The correlation between X (primary predictor) and M (mediator) is assumed to be 0.5. The $R^2$ achieved when X is regressed on any confounder variables in model 1 is assumed to be 0.1. The $R^2$ achieved when M is regressed on any confounder variables in model 2 is assumed to be 0.2. The design effect (the ratio of the variance of the outcome (Y) with clustering to its variance assuming simple random sampling) is assumed to be 1. To detect a regression coefficient, $\theta_X$, in model 1 of 0.5, and a mediation effect (mediator regression coefficient, $\beta_M$) in model 2 of 0.5, with 90% (joint) power, the number of needed subjects will be 71 (the individual power associated with model 1 is 0.996 and the individual power associated with model 2 is 0.904).

**Dropout-Inflated Sample Size**

| Dropout Rate | Sample Size N | Dropout-Inflated Enrollment Sample Size N' | Expected Number of Dropouts D |
|---|---|---|---|
| 20% | 71 | 89 | 18 |
| 20% | 53 | 67 | 14 |
| 20% | 44 | 55 | 11 |
| 20% | 104 | 130 | 26 |
| 20% | 72 | 90 | 18 |
| 20% | 53 | 67 | 14 |

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, 89 subjects should be enrolled to obtain a final sample size of 71 subjects.

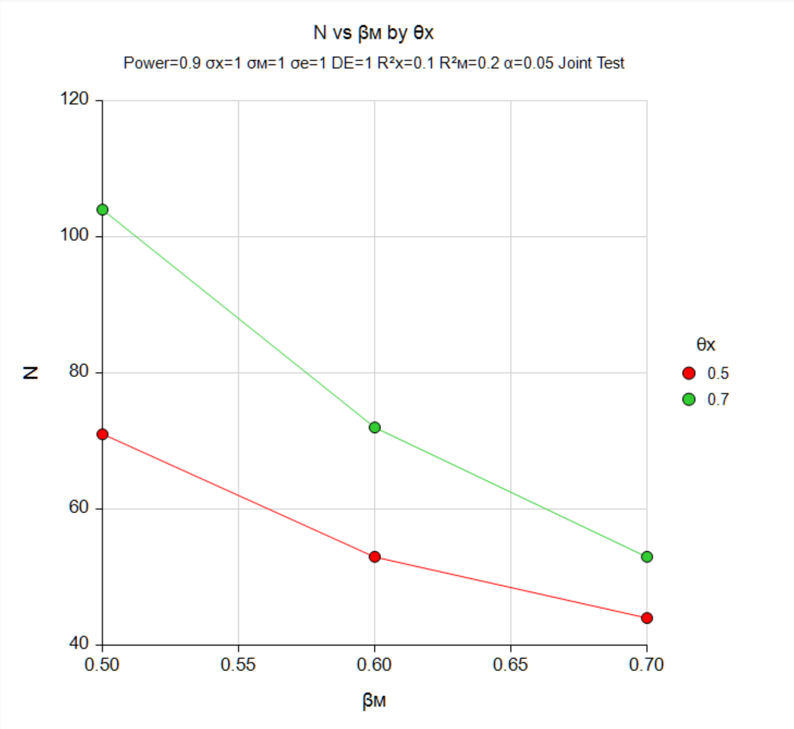
**References**

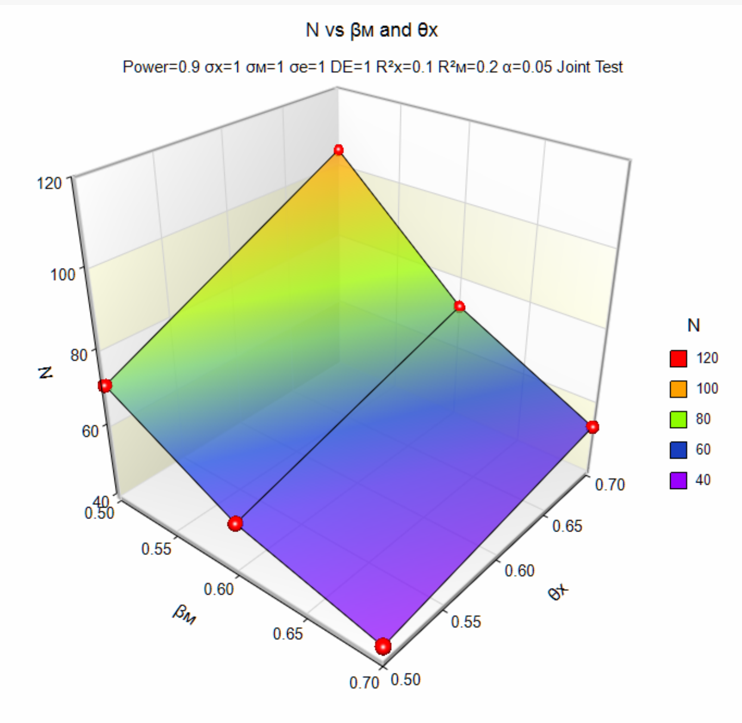
Vittinghoff, E. and Neilands, T.B. 2015. 'Sample size for Joint Testing of Indirect Effects.' Prevention Science, Vol. 16(8), Pages 1128-1135.
Vittinghoff, E., Sen, S., and McCulloch, C.E. 2009. 'Sample size calculations for evaluating mediation.' Statistics in Medicine, Vol. 28, Pages 541-557.

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 Vittinghoff (2015)

Vittinghoff et al. (2015) present an example which $\theta_X = 0.25$; $\beta_M = 0.2$, $\sigma_X = \sigma_M = \sigma_e = DE = 1$; $R^2_X = 0.0$; and $R^2_M = 0.09$. They set the power at 0.8 and the two-sided significance level at 0.05. The computed N is 240.

## 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 ...................................................... **N (Sample Size)**
Alternative Hypothesis .................................. **Two-Sided**
Power............................................................ **0.80**
Alpha............................................................ **0.05**
$\theta_X$ (Reg Coef of X in Model 1)....................... **0.25**
$\beta_M$ (Reg Coef of M in Model 2) ...................... **0.2**
$\sigma_X$ (Standard Deviation of X).......................... **1**
$\sigma_M$ (Standard Deviation of M) ........................ **1**
$\sigma_e$ (Standard Deviation of $e_Y$)........................ **1**
DE (Design Effect) ........................................ **1**
$R^2_X$ ($R^2$ of Confounders in Model 1) ............... **0.0**
$R^2_M$ ($R^2$ of Confounders in Model 2) ............. **0.09**

## Output

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

**Numeric Results**

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

| | Individual Power | | | Regression Coefficient | | | Standard Deviation | | | R² if X is Regressed on Confounders in Model 1 | R² if M is Regressed on Confounders in Model 2 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Joint Power | Model 1 Pw1 | Model 2 Pw2 | Sample Size N | X $\theta_X$ | M $\beta_M$ | Correlation of X and M $\rho_{XM}$ | X $\sigma_X$ | M $\sigma_M$ | $e_Y$ $\sigma_e$ | $R^2_X$ | $R^2_M$ | Design Effect DE | Alpha |
| 0.8014 | 0.98 | 0.818 | 241 | 0.25 | 0.2 | 0.25 | 1 | 1 | 1 | 0 | 0.09 | 1 | 0.05 |

**PASS** obtained 241 which is one higher than the 240 they calculated. It turns out that 240 has achieves a power that is slightly less than the 0.8 that was desired. Thus, 241 is the correct answer.