

Chapter 823

Tests of Mediation Effect in Logistic Regression

Introduction

This procedure computes power and sample size for a test of the mediation effect in a logistic regression with a binary dependent (output) variable Y and an independent (input) variable X . Interest focuses on the interrelationship between Y , X , and a third independent variable called the mediator M . The sample size calculations are based on the work of Vittinghoff, Sen, and McCulloch (2009). Note that their work has been extended in Vittinghoff and Neilands (2015). We are looking into adding those extensions in a later procedure.

Mediation Model

Vittinghoff, Sen, and McCulloch (2009) derived a sample size formula for testing the mediation effect based on testing the significance of β_M in the logistic regression model

$$\log\left(\frac{P}{1-P}\right) = \beta_0 + \beta_X X + \beta_M M$$

They showed that testing $\beta_M = 0$ is equivalent to testing for a significant mediation effect. In addition to the notation above, they use ρ_{XM} as the correlation between the independent variables X and M .

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: $z_\beta = \sqrt{N\sigma_M^2\beta_M^2(1 - \rho_{XM}^2)P(1 - P)} - z_{1-\alpha}$.
3. Calculate: 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.

Procedure Options

This section describes the options that are specific to this procedure. These are located on the Design tab. For more information about the options of other tabs, go to the Procedure Window chapter.

Design Tab

The Design tab contains most of the parameters and options that you will be concerned with.

Solve For

Solve For

This option specifies the parameter to be solved for from the other parameters. Under most situations, you will select either *Power* or *N (Sample Size)*.

Select *Sample Size* when you want to calculate the sample size needed to achieve a given power and alpha level.

Select *Power* when you want to calculate the power of an experiment.

Test Direction

Alternative Hypothesis

Specify whether the hypothesis test is one-sided or two-sided. When a two-sided test is selected, the value of alpha is automatically divided by two. A two-sided test requires alpha to be less than 0.50.

Power and Alpha

Power

This option specifies one or more values for power. Power is the probability of rejecting a false null hypothesis and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected.

Values must be between zero and one. Historically, the value of 0.80 (Beta = 0.20) was used for power. Now, 0.90 (Beta = 0.10) is also commonly used.

A single value may be entered here or a range of values such as *0.8 to 0.95 by 0.05* may be entered.

Alpha

This option specifies one or more values for the probability of a type-I error (alpha). A type-I error occurs when you reject the null hypothesis when in fact it is true.

Values of alpha must be between zero and one. Historically, the value of 0.05 has been used for alpha. This means that about one test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You may enter a range of values such as *0.01 0.05 0.10* or *0.01 to 0.10 by 0.01*.

Sample Size

N (Sample Size)

This option specifies the value(s) for *N*, the sample size. Note that $3 < N$.

Tests of Mediation Effect in Logistic Regression

Effect Size

β_M (Reg Coef of M)

Enter one or more values for β_M , which is the coefficient of M in the logistic regression model

$$\log(P/(1-P)) = \beta_0 + \beta_X(X) + \beta_M(M)$$

where $P/(1-P)$ is the odds of P, X is the primary predictor variable, and M is the mediator variable.

In this model, β_M can be interpreted as the log of the odds or logit. Hence, if you want to test for an odds of 1.5, enter $\log(1.5) = 0.4055$ here.

Other common values are

Odds $\ln(\text{Odds})$

0.25	-1.3863
0.33	-1.1087
0.5	-0.6931
0.8	-0.2231
0.9	-0.1054
1.0	0.0 (Not Allowed)
1.1	0.0953
1.2	0.1823
1.3	0.2624
1.4	0.3365
1.5	0.4055
2.0	0.6931
3.0	1.0986
4.0	1.3863

Range

β_M can be any value other than zero.

ρ_{XM} (Correlation of X and M)

Enter one or more values for the correlation of X and M.

Range

$0 < \rho_{XM} < 1$.

Type of Mediator, M

Indicate whether the mediator variable, M, is continuous or binary.

σ_M (Standard Deviation of M)

Enter one or more values for the standard deviation of M, the mediator variable.

Range

$0 < \sigma_M$.

Tests of Mediation Effect in Logistic Regression**Probability M = 1**

Enter one or more values of the probability in the population that $M = 1$ when M is a binary variable that takes on the values 0 or 1. This value is used to compute the variance of M using the formula: $\sigma_M^2 = P(M=1) \times (1 - P(M=1))$.

Range

Since this is a probability, it can be any value between 0 and 1.

P (Marginal Prevalence of Y)

Enter one or more values of the marginal prevalence of the binary outcome variable Y . This is the probability in the population that $Y = 1$ when Y is a binary dependent variable that takes on the values 0 or 1.

Range

Since this is a probability, it can be any value between 0 and 1. That is, $0 < P < 1$.

Example 1 – Finding Sample Size

Researchers are studying the relationship between a binary 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 significance test of the mediator term in a logistic regression. Using prior analyses, they decide to use $\beta_M = 0.6, 0.8, 1.0$, $\rho_{XM} = 0.3$, $\sigma_M = 0.4$, and $P = 0.45$. They set the power at 0.9 and the two-sided significance level at 0.05.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests of Mediation Effect in Logistic Regression** procedure. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	N (Sample Size)
Alternative Hypothesis	Two-Sided
Power	0.9
Alpha	0.05
β_M (Reg Coef of M)	0.6 0.8 1.0
ρ_{XM} (Correlation of X and M)	0.3
Type of Mediator, M	Continuous
σ_M (Standard Deviation of M)	0.4
P (Marginal Prevalence of Y)	0.45

Annotated Output

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

Numeric Results

Numeric Results
Two-sided alternative hypothesis

	Sample Size	Reg Coef of M	Corr of X and M	Std Dev of M	Marg Prev of Y	Alpha
Power	N	β_M	ρ_{XM}	σ_M	P(Y=1)	
0.9000	810	0.600	0.300	0.400	0.450	0.050
0.9003	456	0.800	0.300	0.400	0.450	0.050
0.9004	292	1.000	0.300	0.400	0.450	0.050

References

Vittinghoff, E., Sen, S., and McCulloch, C.E. 2009. 'Sample size calculations for evaluating mediation.'
Statistics in Medicine, Vol. 28, Pages 541-557.

Tests of Mediation Effect in Logistic Regression

Report Definitions

Hypotheses: $H_0: \beta_M = 0$ versus $H_1: \beta_M \neq 0$. (Two-Sided).

X is the primary predictor. It is an independent variable.

M is the mediator. It is a continuous variable.

Model: $\log(P/(1-P)) = \beta_0 + \beta_X(X) + \beta_M(M)$.

Power is the probability of rejecting a false null hypothesis.

N is the number of observations on which the multiple regression is computed.

β_M is the regression coefficient of the mediator in the model.

ρ_{XM} is the correlation between X and M.

σ_M is the standard deviation of M.

$P(Y=1)$ is the marginal prevalence of the binary outcome, Y.

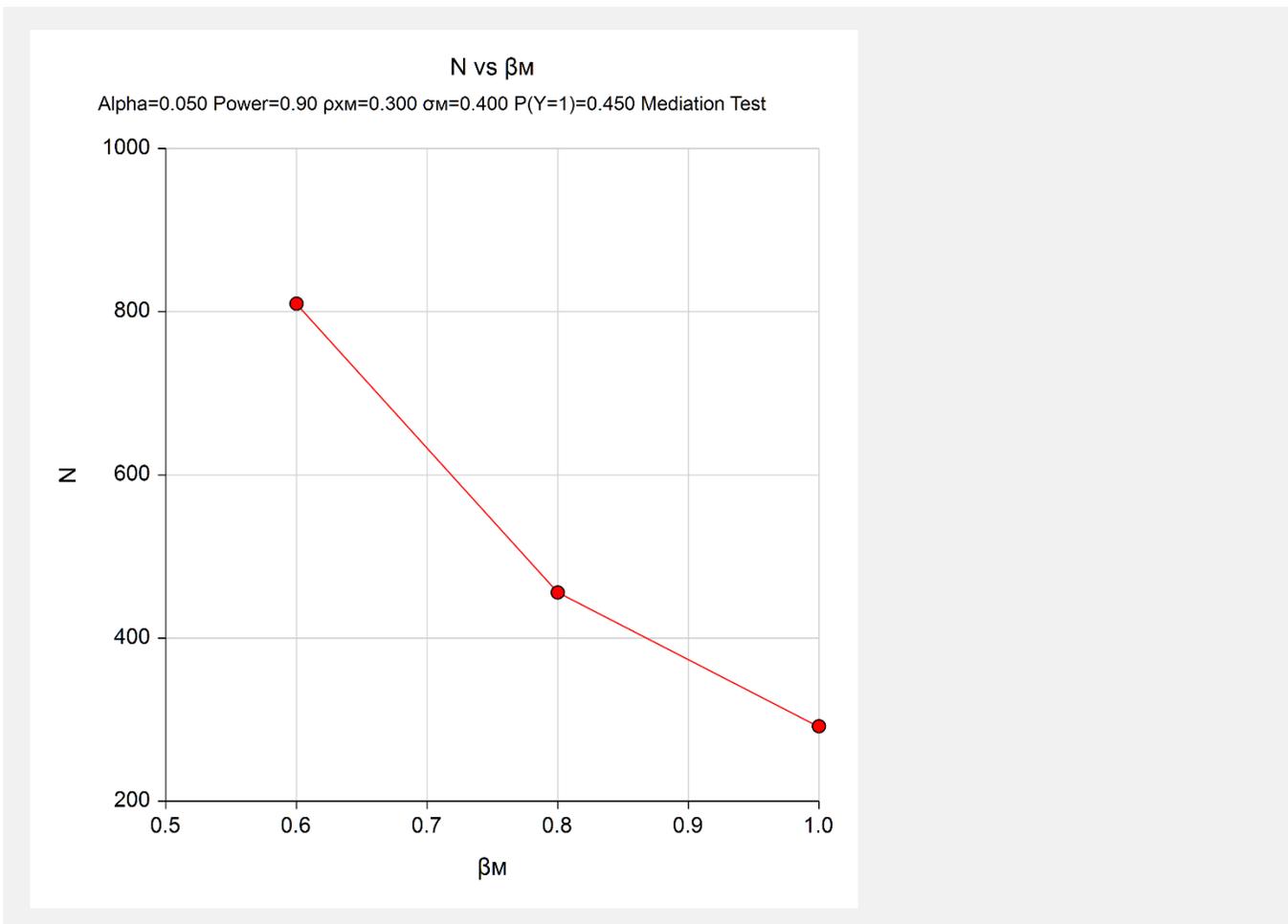
Alpha is the probability of rejecting a true null hypothesis. It should be small.

Summary Statements

A sample size of 810 achieves 90% power to detect a mediation effect (as measured by the regression coefficient of M, β_M) of at least 0.600 when the two-sided significance level (alpha) is 0.050. The correlation between X (the primary predictor) and M (mediator) is 0.300. The mediator, M, has a standard deviation of 0.400. The marginal prevalence of the outcome variable is 0.450.

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

Plots Section



This plot shows the relationship between sample size and effect size.

Tests of Mediation Effect in Logistic Regression

Example 2 – Validation using Vittinghoff (2009)

Vittinghoff et al. (2009) present an example on page 545 in which $\beta_M = \log 1.5$ (0.4055), $\rho_{XM} = 0.5$, $\sigma_M = 1$, and $P = 0.5$. They set the power at 0.8 and the one-sided significance level at 0.025. The computed sample size is 255.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests of Mediation Effect in Logistic Regression** procedure. You may then make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	N (Sample Size)
Alternative Hypothesis	One-Sided
Power	0.8
Alpha	0.025
β_M (Reg Coef of M)	0.4055
ρ_{XM} (Correlation of X and M)	0.5
Type of Mediator, M	Continuous
σ_M (Standard Deviation of M)	1
P (Marginal Prevalence of Y)	0.5

Output

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

Numeric Results

Numeric Results						
One-sided alternative hypothesis						
	Sample Size	Reg Coef of M	Corr of X and M	Std Dev of M	Marg Prev of Y	Alpha
Power	N	β_M	ρ_{XM}	σ_M	P(Y=1)	
0.8006	255	0.406	0.500	1.000	0.500	0.025

PASS matches the calculation of $N = 255$.