Chapter 852

Mendelian Randomization with a Continuous Outcome

Introduction

This module computes the sample size and power of the causal effect in Mendelian randomization studies with a continuous outcome. This analysis is used in observational studies where clinical trials are not possible.

Analogous to randomized clinical trials (RCT), Mendelian randomization (MR) divides subjects into two or more groups. However, MR uses a genetic variable, such as the state of a certain gene, to form the groups. The state of the gene is assumed to be random. Using two-stage least squares and making several assumptions, the causal impact of an exposure variable on the outcome variable can be measured.

For further reading, we recommend the book by Burgess and Thompson (2015) which is completely devoted to this topic. We have used the paper by Brion, Shakhbazov, and Visscher (2013) for sample size formulas. We also recommend the papers by Burgess (2014) and Freeman, Cowling, and Schooling (2013).

Technical Details

Causal Relationship Test

The following details follow closely the results in Brion, Shakhbazov, and Visscher (2013). Assume that we are interested in assessing the causal relationship between an *outcome* variable Y and an *exposure* variable X. A genetic variable G is available to use as an *instrumental* variable. A sample of *n* will be selected. The basic models are

$$Y = \beta_{YX} X + e_Y$$

$$X = \beta_{XG} G + e_X$$

An estimate of β_{YX} will be obtained using two-stage least squares and called b_{2SLS}. The mean and variance of b_{2SLS} is given by

$$E(b_{2SLS}) = \beta_{YX} + \mathrm{cov}(e_Y, e_X)/(n\rho_{XG}^2)$$

$$\sigma_{b_{2SLS}}^2 = \sigma_{e_Y}^2/(n\rho_{XG}^2\sigma_X^2)$$

We are using ρ_{XG}^2 to represent the proportion of the variation in X explained by G in the population.

Power Calculation

The power is given by

$$\text{Power} = 1 - P\left(\chi^{2\prime}_{df,NCP} > \chi^2_{df,1-\alpha}\right)$$

where $\chi^{2\prime}_{df,NCP}$ is a non-central chi-square with df = 1 and non-centrality parameter NCP. Also, $\chi^2_{df,1-\alpha}$ is the quantile of a central chi-square with df = 1 and probability α. Note that this is a two-sided test. The result for a one-sided test is obtained by the usual adjustment.

The value of NCP is given by

$$NCP = \frac{E(b_{2SLS})}{\sigma^2_{b_{2SLS}}}$$

This can be evaluated if we note that

$$\text{cov}(e_Y, e_X) = (\beta_{OLS} - \beta_{YX})\sigma_X^2$$

$$\sigma_{e_Y}^2 = \sigma_Y^2 - \sigma_X^2\beta_{YX}(2\beta_{OLS} - \beta_{YX})$$

where β_{OLS} is the expected value of the ordinary least squares (OLS) regression coefficient of the regression of Y on X.

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. This chapter covers four procedures, each of which has different effect size options. However, many of the options are common to all four procedures. These common options will be displayed first, followed by the various effect size options.

Solve For

Solve For

This option specifies the parameter to be solved for from the other parameters. The parameters that may be selected are *Power* and *Sample Size*.

Test

Alternative Hypothesis

Specify whether the statistical test is two-sided or one-sided.

- **Two-Sided**
  This option gives the two-sided test.

- **One-Sided**
  This option gives the approximate result for a one-sided test by making an appropriate substitution for α.

Power and Alpha

Power

This option specifies one or more values for power. Power is the probability of rejecting a false null hypothesis. Values must be between zero and one. Historically, the value of 0.80 was used for power. Now, 0.90 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. A type-I error occurs when you reject the null hypothesis of equal survival curves when in fact the curves are equal.

Values of alpha must be between zero and one. Historically, the value of 0.05 has been used for a two-sided test and 0.025 has been used for a one-sided test. 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)

Enter a value for the sample size, N. This is the number of subjects in the study. You can enter one or more positive integers greater than or equal to 3. Note that in these type of studies, very large sample sizes (50000+) are often needed.

You may also enter a range such as “10000 to 50000 by 10000” or a list of values separated by commas or blanks such as “2000 4000 6000 8000.”

Effect Size

β1 (2SLS of Y on X|G)

Enter one or more values for β1, the parameter of interest. This is the value of β_{YX}. If all assumptions are met, β1 is a measure of the causal effect of X on Y. It is estimated using two-stage least squares (2SLS), where G is the genetic, or instrumental, variable.

You should keep this value in the same scale with your entries for σ(X) and σ(Y). This value can be any value greater than zero. The results for negative values will be identical to those of |β1|.

β0 (OLS of Y on X)

Enter one or more values for β0, regression parameter from the OLS regression of Y on X. This is the value of β_{OLS}. You should keep this value in the same scale with your entries for σ(X) and σ(Y). This value can be any value greater than zero. The results for negative values will be identical to those of |β0|.

ρ^2(XG)

Enter one or more value for ρ^2(XG), the proportion of the variance of X explained by the regression of X on G. In practice, it is estimated by the R^2 of this regression. It is routinely very small. Values of 0.01 and 0.02 are common. This value varies between zero and one. For mathematical reasons, it cannot be exactly 0.

σ(X)

Enter one or more value for σ(X), the standard deviation of the X values. This value can be any number greater than 0.

σ(Y)

Enter one or more value for σ(Y), the standard deviation of the Y values. This value can be any number greater than 0.

Example 1 – Finding the Sample Size

Researchers are planning an observation study to determine the causal effect of an exposure X on a continuous outcome variable Y using an instrumental variable G. They want to determine the sample sizes necessary to have 80% power and 5% significance using a two-sided test when β1 is 1.05 to 1.3 by 0.05, β0 is 1.4, ρ^2(XG) is 0.01 or 0.02, σ(X) is 1, and σ(Y) is 10.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Mendelian Randomization with a Continuous Outcome** procedure window by expanding **Regression** and then clicking on **Mendelian Randomization with a Continuous Outcome**. 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**.

Option	Value
Design Tab	
Solve For	**Sample Size**
Alternative Hypothesis	**Two-Sided**
Power	**0.80**
Alpha	**0.05**
β1 (2SLS of Y on X|G)	**1.05 to 1.30 by 0.05**
β0 (OLS of Y on X)	**1.40**
ρ^2(XG)	**0.01 0.02**
σ(X)	**1**
σ(Y)	**10**

Annotated Output

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

Numeric Results

Numeric Results for the Two-Sided, Mendelian Randomization Test of a Continuous Outcome

Power	N	β1(2SLS)	β0(OLS)	ρ^2(XG)	σ(X)	σ(Y)	σ(b1)	NCP	Alpha
0.8000	69817	1.0500	1.4000	0.0100	1.0000	10.0000	0.3750	7.8490	0.0500
0.8000	34908	1.0500	1.4000	0.0200	1.0000	10.0000	0.3750	7.8489	0.0500
0.8000	63600	1.1000	1.4000	0.0100	1.0000	10.0000	0.3928	7.8489	0.0500
0.8000	31800	1.1000	1.4000	0.0200	1.0000	10.0000	0.3928	7.8491	0.0500
0.8000	58180	1.1500	1.4000	0.0100	1.0000	10.0000	0.4106	7.8490	0.0500
0.8000	29090	1.1500	1.4000	0.0200	1.0000	10.0000	0.4106	7.8490	0.0500
0.8000	53427	1.2000	1.4000	0.0100	1.0000	10.0000	0.4285	7.8489	0.0500
0.8000	26713	1.2000	1.4000	0.0200	1.0000	10.0000	0.4285	7.8490	0.0500
0.8000	49236	1.2500	1.4000	0.0100	1.0000	10.0000	0.4463	7.8490	0.0500
0.8000	24618	1.2500	1.4000	0.0200	1.0000	10.0000	0.4463	7.8490	0.0500
0.8000	45523	1.3000	1.4000	0.0100	1.0000	10.0000	0.4641	7.8490	0.0500
0.8000	22761	1.3000	1.4000	0.0200	1.0000	10.0000	0.4641	7.8490	0.0500

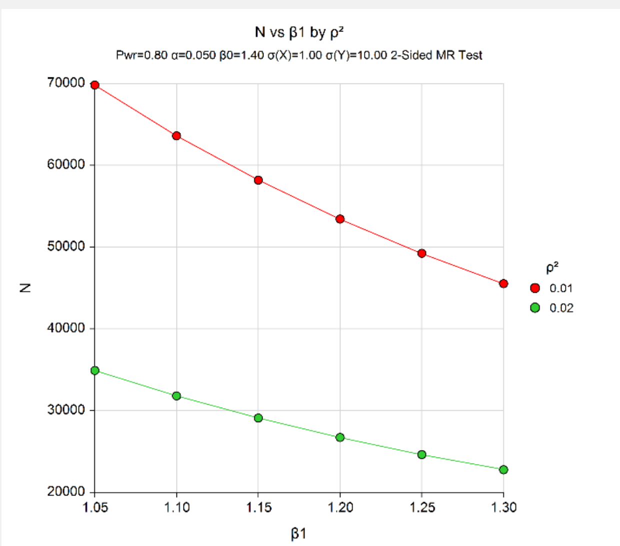
References
Burgess, S. and Thompson, S.G. 2015. Mendelian Randomization Methods for Using Genetic Variants in Causal Estimation. Chapman & Hall/CRC Press. New York.
Burgess, Stephen. 2014. 'Sample size and power calculations in Mendelian randomization with a single instrumental variable and a binary outcome.' International Journal of Epidemiology, 43, pages 922-929.
Freeman, G., Cowling, B.J., Schooling, C.M. 2013. 'Power and sample size calculations for Mendelian randomization studies using one genetic instrument.' International Journal of Epidemiology, 42, pages 1157-1163.
Brion, M.J.A., Shakhbazov, K., Visscher, P.M. 2013. 'Calculating statistical power in Mendelian randomization studies.' International Journal of Epidemiology, 42, pages 1497-1501.

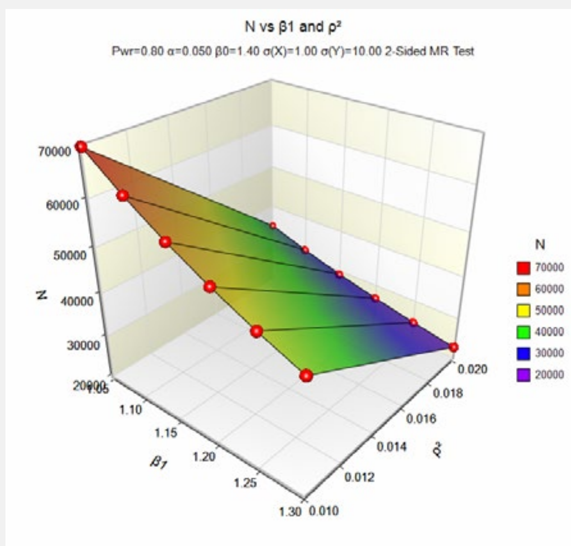
Report Definitions
Y is the continuous outcome variable.
X is the exposure variable.
G is the genetic variant or instrumental variable used to divide the subjects up into groups.
Power is the probability of rejecting a false null hypothesis.
N is the sample size, the number of subjects in the study.
β1(2SLS) is the parameter of interest. It is a measure of the causal effect of X on Y. It can be estimated using two-stage least squares (2SLS) where G is the genetic, or instrumental, variable.
β0(OLS) is the regression parameter (slope) of the regression of Y on X which is estimated by ordinary least squares.
ρ^2(XG) is the proportion of the variance of X explained by the regression of X on G.
σ(X) is the standard deviation of the X variable.
σ(Y) is the standard deviation of the Y variable.
σ(b1) is the standard error of the estimate of β1.
NCP is the non-centrality parameter for testing the significance of β1.
Alpha is the probability of rejecting a true null hypothesis.

Summary Statements
A two-sided test in a Mendelian randomization study for a continuous outcome variable that was based on a sample of 69817 subjects achieves 80.0% power at a 0.0500 significance level to detect a causal effect (β1) of 1.0500 when the ordinary least squares estimate of the regression slope (β0) is 1.4000. The standard deviations of X and Y are 1.0000 and 10.0000, respectively. The proportion of the variance of X explained by its regression on G, the instrumental variable, is 0.0100.

This report presents the calculated sample sizes for each scenario as well as the values of the other parameters.

Plots Section





These plots show the relationship among the varying parameters.

Example 2 – Validation using Brion, et al. (2013)

Brion, et al. (2013) give an example in the online tool (cnsgenomics.com/shiny/mRnd/) associated with their paper in which the power is 0.80, alpha = 0.05, $\beta 1$ is 1.3, $\beta 0$ is 1.41, $\rho^2(XG)$ is 0.01, $\sigma(X)$ is 1, and $\sigma(Y)$ is 10.79815. They obtain N = 53218.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Mendelian Randomization with a Continuous Outcome** procedure window by expanding **Regression** and then clicking on **Mendelian Randomization with a Continuous Outcome**. 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**.

Option	Value
Design Tab	
Solve For	**Sample Size**
Alternative Hypothesis	**Two-Sided**
Power	**0.80**
Alpha	**0.05**
β1 (2SLS of Y on X|G)	**1.3**
β0 (OLS of Y on X)	**1.41**
ρ²(XG)	**0.01**
σ(X)	**1**
σ(Y)	**10.79815**

Output

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

Numeric Results

Numeric Results for the Two-Sided, Mendelian Randomization Test of a Continuous Outcome

Power	N	β1(2SLS)	β0(OLS)	ρ²(XG)	σ(X)	σ(Y)	σ(b1)	NCP	Alpha
0.8000	53219	1.3000	1.4100	0.0100	1.0000	10.7982	0.4641	7.8490	0.0500

PASS calculated N as 53219 because 53218 gives a power slightly less than the goal of 0.80.