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

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

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

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

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

$$\sigma^2_{e_Y} = \sigma^2_Y - \sigma^2_X \beta_{YX}(2\beta_{OLS} - \beta_{YX})$$

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

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 β_1 is 1.05 to 1.3 by 0.05, β_0 is 1.4, $\rho^2(XG)$ is 0.01 or 0.02, $\sigma(X)$ is 1, and $\sigma(Y)$ is 10.

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 ....**Sample Size**
Alternative Hypothesis ....**Two-Sided**
Power..**0.80**
Alpha...**0.05**
β_1 (2SLS of Y on X|G) ....**1.05 to 1.30 by 0.05**
β_0 (OLS of Y on X) **1.40**
$\rho^2(XG)$...**0.01 0.02**
$\sigma(X)$..**1**
$\sigma(Y)$...**10**

Output

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

Numeric Reports

Numeric Results

Solve For: Sample Size
Alternative Hypothesis: Two-Sided

Power	Sample Size N	Regression Coefficient of Y on X β1(2SLS)	β0(OLS)	R² of the Regression of X on G ρ²(XG)	σ(X)	σ(Y)	Standard Error of the Estimate of β1 σ(b1)	Non-Centrality Parameter NCP	Alpha
0.8	69817	1.05	1.4	0.01	1	10	0.3750	7.8489	0.05
0.8	34909	1.05	1.4	0.02	1	10	0.3750	7.8490	0.05
0.8	63600	1.10	1.4	0.01	1	10	0.3928	7.8490	0.05
0.8	31800	1.10	1.4	0.02	1	10	0.3928	7.8490	0.05
0.8	58180	1.15	1.4	0.01	1	10	0.4106	7.8490	0.05
0.8	29090	1.15	1.4	0.02	1	10	0.4106	7.8490	0.05
0.8	53427	1.20	1.4	0.01	1	10	0.4285	7.8490	0.05
0.8	26714	1.20	1.4	0.02	1	10	0.4285	7.8491	0.05
0.8	49236	1.25	1.4	0.01	1	10	0.4463	7.8489	0.05
0.8	24618	1.25	1.4	0.02	1	10	0.4463	7.8489	0.05
0.8	45522	1.30	1.4	0.01	1	10	0.4641	7.8489	0.05
0.8	22761	1.30	1.4	0.02	1	10	0.4641	7.8489	0.05

Y   The continuous outcome variable.
X   The exposure variable.
G   The genetic variant or instrumental variable used to divide the subjects up into groups.
Power   The probability of rejecting a false null hypothesis when the alternative hypothesis is true.
N   The sample size, the number of subjects in the study.
β1(2SLS)   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)   The regression parameter (slope) of the regression of Y on X which is estimated by ordinary least squares.
ρ²(XG)   The proportion of the variance of X explained by the regression of X on G.
σ(X)   The standard deviation of the X variable.
σ(Y)   The standard deviation of the Y variable.
σ(b1)   The standard error of the estimate of β1.
NCP   The non-centrality parameter for testing the significance of β1.
Alpha   The probability of rejecting a true null hypothesis.

Summary Statements

A Mendelian randomization study design with a continuous outcome variable (Y) will be used to test whether there is a causal relationship between Y and an exposure variable (X), with a genetic variable (G) as an instrumental variable that divides the subjects into groups. The comparison will be made using a two-sided two-stage test involving a regression of X on G (first stage) and a regression of Y on X (second stage). The Type I error rate (α) will be 0.05. The proportion of the variance of X explained by the regression of X on G is assumed to be 0.01. The standard deviations of X and Y are assumed to be 1 and 10, respectively. To detect a causal effect of X on Y (β1) of 1.05 when the ordinary least squares regression slope of Y on X (β0) is 1.4, with 80% power, the number of needed subjects will be 69817.

Dropout-Inflated Sample Size

Dropout Rate	Sample Size N	Dropout-Inflated Enrollment Sample Size N'	Expected Number of Dropouts D
20%	69817	87272	17455
20%	34909	43637	8728
20%	63600	79500	15900
20%	31800	39750	7950
20%	58180	72725	14545
20%	29090	36363	7273
20%	53427	66784	13357
20%	26714	33393	6679
20%	49236	61545	12309
20%	24618	30773	6155
20%	45522	56903	11381
20%	22761	28452	5691

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, 87272 subjects should be enrolled to obtain a final sample size of 69817 subjects.

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.

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

Plots Section

Plots
_____
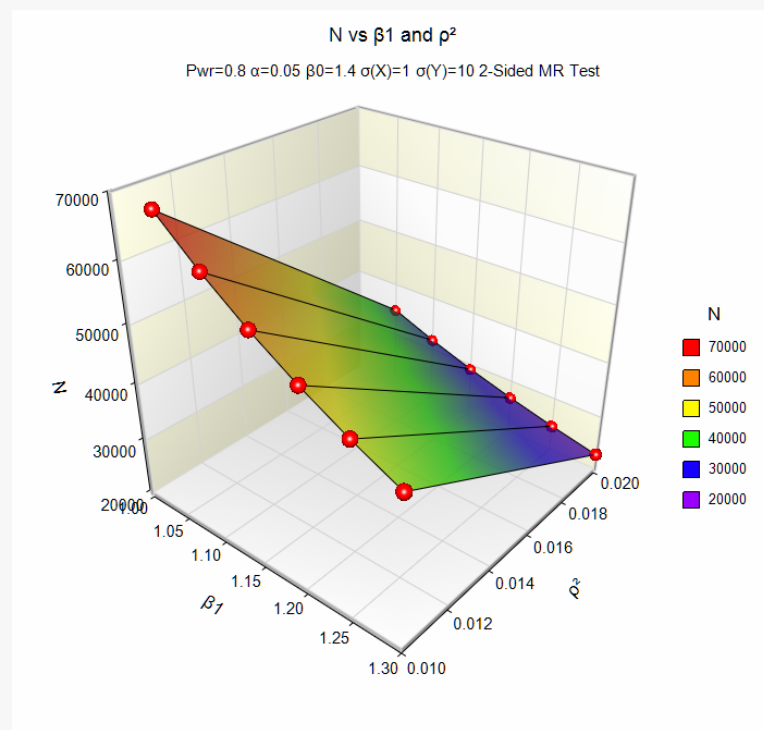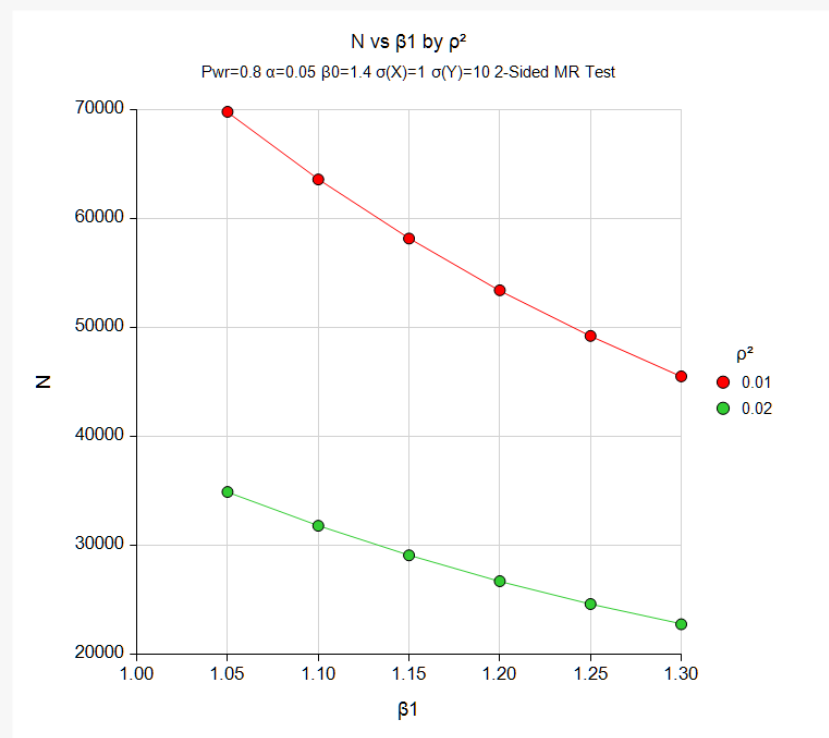
These plots show the relationship among the varying parameters.

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, β_1 is 1.3, β_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

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 ....**Sample Size**
Alternative Hypothesis ....**Two-Sided**
Power...**0.80**
Alpha...**0.05**
β_1 (2SLS of Y on X|G) **1.3**
β_0 (OLS of Y on X) **1.41**
$\rho^2(XG)$..**0.01**
$\sigma(X)$...**1**
$\sigma(Y)$...**10.79815**

Output

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

Numeric Results

Solve For: Sample Size
Alternative Hypothesis: Two-Sided

Power	Sample Size N	Regression Coefficient of Y on X		R² of the Regression of X on G	Standard Deviation		Standard Error of the Estimate of β1	Non-Centrality Parameter	Alpha
		β1(2SLS)	β0(OLS)	ρ²(XG)	σ(X)	σ(Y)	σ(b1)	NCP	
0.8	53218	1.3	1.41	0.01	1	10.7982	0.4641	7.8489	0.05

PASS also calculates N as 53218.