

Chapter 631

Meta-Analysis of Tests for Two Means using a Random-Effects Model

Introduction

This procedure calculates power and number of studies for random-effects meta-analyses of parallel group studies that are summarized by the difference in two group means. The sample size formulas used by this procedure were published in Hedges and Pigott (2001) and Harrer *et al.* (2022).

Meta-Analysis

Meta-analysis refers to the systematic assessment of a group of two or more individual studies (experiments or clinical trials). The studies must be similar so that their results may be numerically compared. For example, in this procedure all studies must have as their outcome two group means whose difference is used to compare their values.

By combining the results of several studies, their individual results may be summarized and assessed. The power of the combined test will be increased since it represents a larger sample.

Fixed-Effects versus Random-Effects Models

The *fixed-effects model*, which is assumed by another procedure, assumes that all studies have the same true effect size. That is, it assumes that all factors that might influence the effect size are the same so that the true effect size for all studies is the same. Therefore, only the *within study variability* needs to be considered since the effect size variation occurs because of random error.

The *random-effects model*, which is assumed by this procedure, assumes that the studies in the analysis are similar, but not to the extent that they all have the same true effect size. Instead, it is assumed that each study has its own true effect size and that the particular outcome of the study is a sample from a distribution with that mean. The random-effects model must not only account for the *within study variability*, but it must also account for the *between study variability*.

Technical Details

The Meta-Analysis of the Difference of Two Means

The sample size formulas used by this procedure were published in Hedges and Pigott (2001).

A series of K studies are assumed to each have an effect size δ which is the standardized difference between two group means. The results of these individual studies may be combined into one overall statistical test that draws a more precise conclusion.

Test Statistic

Suppose the i^{th} study yields the two group sample sizes $N1$ and $N2$ as well as a *standardized difference* D which is calculated as

$$D_i = \frac{\bar{X}_{1i} - \bar{X}_{2i}}{S_{D_i}}$$

Using a random-effect model, these results may be summarized and tested using the statistic

$$Z = \frac{\bar{D} - \delta_0}{S_{\bar{D}}}$$

where \bar{D} is a weighted average of the individual D_i and $S_{\bar{D}}$ is a weighted estimate of its standard deviation that includes an estimate of the between-study variation.

We assume that Z follows the standard normal distribution and so it can be used to test the null hypothesis $H_0: \delta = \delta_0$ at significance level of α .

Hypotheses

Three statistical hypotheses commonly used are

$$H_0: \delta \leq \delta_0 \quad \text{vs.} \quad H_1: \delta > \delta_0$$

$$H_0: \delta \geq \delta_0 \quad \text{vs.} \quad H_1: \delta < \delta_0$$

$$H_0: \delta = \delta_0 \quad \text{vs.} \quad H_1: \delta \neq \delta_0$$

Power Calculation

The power of a one-tailed test that $\delta = \delta_0$ is given by

$$Power = 1 - \Phi(c_\alpha - \lambda)$$

where δ is set by the alternative hypothesis, c_α is the $1 - \alpha$ quantile of the standard normal distribution, and

$$\lambda = \frac{\delta - \delta_0}{\sqrt{V_{\bar{D}}}}$$

$$V_F = \left(\frac{\bar{N}_1 + \bar{N}_2}{\bar{N}_1 \bar{N}_2} + \frac{\delta^2}{2(\bar{N}_1 + \bar{N}_2)} \right)$$

$$V_{\bar{D}} = \frac{1}{K} (V_F + R V_F)$$

Here V_F represents the within-study variance and R is the heterogeneity ratio. This ratio allows us to specify the between-study variance, V_B , as a multiple of V_F . Hedges and Pigott (2001) recommend using 0.333 for a *small* value, 0.667 for a *medium* value, and 1.0 for a *large* value. Setting R to 0 results in the fixed-effect result.

The power of a two-tailed test that $\delta = \delta_0$ is given by

$$Power = 1 - \Phi(c_{\alpha/2} - \lambda) + \Phi(-c_{\alpha/2} - \lambda)$$

These formulas can be used in search algorithms to determine the necessary number of studies required to achieve a desired level of power.

I^2 - Heterogeneity Index

Some authors state the amount of heterogeneity in the study using the heterogeneity index I^2 (also called the *degree of heterogeneity* or *inconsistency*) instead of R . This value is related to R using the relationship

$$R = \frac{I^2}{1 - I^2}$$

Example 1 – Calculating Sample Size

Researchers wish to determine a reasonable number of studies for a random-effects meta-analysis with the following characteristics: two-group, equal-allocation trials, two-sided significance level of 0.05 and a power of 0.9, average group size of all studies is 25, $\delta_1 = 0.3$, and $\delta_0 = 0.0$. They want to compare the results of setting $R = 0, 0.333, 0.667$, or 1.0.

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	Number of Studies
Alternative Hypothesis	Two-Sided (H1: $\delta \neq \delta_0$)
Power.....	0.90
Alpha.....	0.05
N1 (Average Sample Size of Group 1).....	25
N2 (Average Sample Size of Group 2).....	N1
δ_0 (Standardized Mean Difference H0).....	0.0
δ_1 (Standardized Mean Difference H1).....	0.3
Calculate Between-Study Variance From	R (Between-Study Var / Within-Study Var)
R (Between-Study Var / Within-Study Var)	0 0.333 0.667 1

Output

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

Numeric Reports

Numeric Results

Solve For: [Number of Studies](#)
 Hypotheses: $H_0: \delta = \delta_0$ vs. $H_1: \delta \neq \delta_0$
 Between-Study Variance Input: R (Between-Study / Within-Study Variances)

Power		Average Sample Size Per Study			Number of		Standardized Mean Difference		Between-Study Variance Calculation		Alpha
Target	Actual	Group 1 N1	Group 2 N2	Total N	Studies K	Subjects KN	δH_0 δ_0	δH_1 δ_1	R	I ²	
0.9	0.91550	25	25	50	10	500	0	0.3	0.000	0.00000	0.05
0.9	0.90888	25	25	50	13	650	0	0.3	0.333	0.24981	0.05
0.9	0.90451	25	25	50	16	800	0	0.3	0.667	0.40012	0.05
0.9	0.90164	25	25	50	19	950	0	0.3	1.000	0.50000	0.05

Target Power	The power desired. Power is the probability of rejecting a false null hypothesis.
Actual Power	The power actually achieved. Power is the probability of rejecting a false null hypothesis about the standardized mean difference (SMD). This is the power calculated at the actual SMD.
N1, N2, N	The average sample sizes of groups 1 and 2 along with the total of both groups calculated on all studies included in the meta-analysis.
K	The number of studies included in the meta-analysis.
KN	The number of subjects included in the meta-analysis. $KN = K \times N$.
δ_0	The value of the standardized mean difference (SMD) assumed by the null hypothesis, H_0 . Note that $\delta = [(\mu_1 - \mu_2) / \sigma(\delta)]$ where $\sigma(\delta)$ is the standard deviation of the estimate of δ .
δ_1	The value of the SMD assumed by the alternative hypothesis, H_1 . This is the value of δ at which the power is calculated.
R	The value of the heterogeneity ratio used to specify the between study variance. The formula is (Between Study Variance) = R x (Within Study Variance). The following values are often used: Small = 0.333, Medium = 0.667, Large = 1.000.
I ²	The value of the heterogeneity index used to specify the Between Study variance. The formula is (Between Study Variance) = I ² / (1 - I ²) x (Within Study Variance). The following values are often used: Small = 0.25, Medium = 0.5, Large = 0.75.
Alpha	The probability of rejecting a true null hypothesis.

Summary Statements

A collection of two-group studies will be used to test whether the Group 1 mean is different from the Group 2 mean ($H_0: \delta = 0$ versus $H_1: \delta \neq 0$, $\delta = [(\mu_1 - \mu_2) / \sigma(\delta)]$). The meta-analysis comparison will be made using a two-sided random-effects-model Z-test, with a Type I error rate (α) of 0.05. The heterogeneity ratio used to create the between-study variance is assumed to be 0. With average study sample sizes of 25 for Group 1 and 25 for Group 2, to detect a standardized mean difference (δ) of 0.3 with 90% power, the number of needed studies will be 10.

References

Borenstein, M., Hedges, L.V., Higgins, J.P.T., and Rothstein, H.R. 2021. Introduction to Meta-Analysis, Second Edition. John Wiley and Sons.

Cooper, J., Hedges, L.V., Valentine, J.C. 2019. The Handbook of Research Synthesis and Meta-Analysis, Third Edition. Russell Sage Foundation.

Griffin, J.W. 2021. 'Calculating statistical power for meta-analysis using metapower', The Quantitative Methods for Psychology, Vol. 17, No. 1, pages 24-39.

Harrer, M., Cuijpers, P., Furukawa, T.A., and Ebert, D.D. 2022. Doing Meta-Analysis with R: A Hands-On Guide. CRC Press.

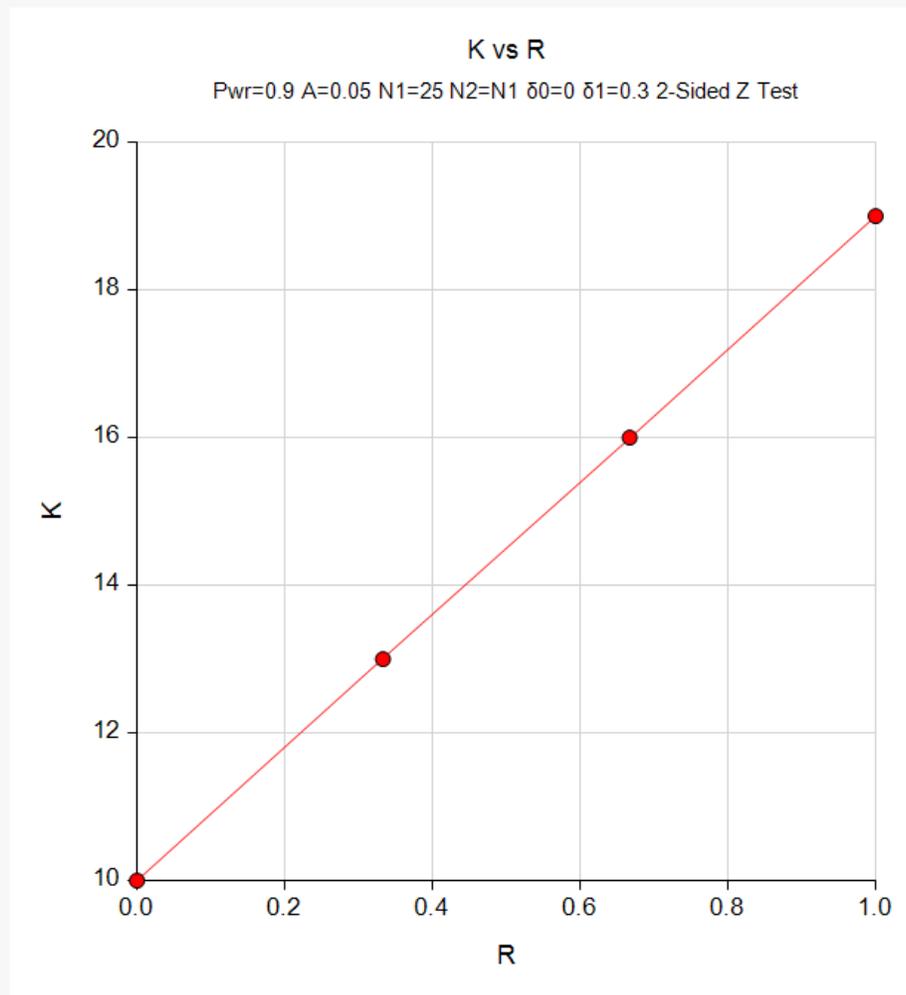
Hedges, L.V. and Pigott, T.D. 2001. 'The Power of Statistical Tests in Meta-Analysis', Psychological Methods, Volume 6, No. 3, pages 203-217.

Valentine, J.C., Pigott, T.D., and Rothstein, H.R. 2010. 'How Many Studies Do You Need? A Primer on Statistical Power for Meta-Analysis', Journal of Education and Behavioral Statistics, Volume 35, No. 2, pages 215-247.

This report shows the sample sizes for the indicated scenarios.

Plots Section

Plots



This plot represents the required number of studies for various values of the heterogeneity index.

Example 2 – Validation using Hedges and Pigott (2001)

Hedges and Pigott (2001) page 213 present an example of solving for power with a two-group, two-sided significance level of 0.05, average group sizes of 12 and 36 in groups 1 and 2, $\delta_1 = 0.2$, and $\delta_0 = 0.0$. The value of R is set to 0.333. The computed power is 0.60.

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 (H1: $\delta \neq \delta_0$)**
 Alpha..... **0.05**
 K (Number of Studies) **18**
 N1 (Average Sample Size of Group 1)..... **12**
 N2 (Average Sample Size of Group 2)..... **36**
 δ_0 (Standardized Mean Difference|H0)..... **0.0**
 δ_1 (Standardized Mean Difference|H1)..... **0.2**
 Calculate Between-Study Variance From **R (Between-Study Var / Within-Study Var)**
 R (Between-Study Var / Within-Study Var) **0.333**

Output

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

Numeric Results

Solve For: [Power](#)
 Hypotheses: H0: $\delta = \delta_0$ vs. H1: $\delta \neq \delta_0$
 Between-Study Variance Input: R (Between-Study / Within-Study Variances)

Power	Average Sample Size Per Study			Number of		Standardized Mean Difference		Between-Study Variance Calculation		Alpha
	Group 1 N1	Group 2 N2	Total N	Studies K	Subjects KN	$\delta H_0 \delta_0$	$\delta H_1 \delta_1$	R	I ²	
0.59513	12	36	48	18	864	0	0.2	0.333	0.24981	0.05

PASS also calculates a power that rounds to 0.60 which validates the procedure.