Chapter 419

# Two-Sample T-Tests using Effect Size

## Introduction

This procedure provides sample size and power calculations for one- or two-sided two-sample t-tests when the effect size is specified rather than the means and variance(s). The details of this procedure are given in Cohen (1988). The design corresponding to this test procedure is sometimes referred to as a *parallel-groups* design. In this design, two groups from independent, normally distributed populations are compared by considering the difference in their means scaled be their common standard deviation.

This procedure is specific to the two-sample t-test assuming equal variance. If the variances are known to be significantly different, this procedure can still be used if the group sample sizes are equal and the average of the variances is used.

## Test Assumptions

When running a two-sample equal-variance t-test, the basic assumptions are that the distributions of the two populations are approximately normal, and that the variances of the two distributions are the same. If the variances are different, this procedure can still be used if the two group sample sizes are nearly equal.

## Test Procedure

If we assume that $\mu_1$ and $\mu_2$ represent the means of the two populations of interest and their common (unknown) standard deviation is $\sigma$, the effect size is represented by $d$, where

$$d = \frac{\mu_1 - \mu_2}{\sigma}$$

The null hypothesis is $H_0$: d = 0 and the alternative hypothesis depends on the number of "sides" of the test:

| | | | |
|---|---|---|---|
| Two-Sided: | $H_1: d \neq 0$ | or | $H_1: \mu_1 - \mu_2 \neq 0$ |
| Upper One-Sided: | $H_1: d > 0$ | or | $H_1: \mu_1 - \mu_2 > 0$ |
| Lower One-Sided: | $H_1: d < 0$ | or | $H_1: \mu_1 - \mu_2 < 0$ |

A suitable Type I error probability ($\alpha$) is chosen for the test, the data is collected, and a *t*-statistic is generated using the formula:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

This *t*-statistic follows a *t* distribution with $n_1 + n_2 - 2$ degrees of freedom. The null hypothesis is rejected in favor of the alternative if,

for $H_1: d \neq 0$ or $H_1: \mu_1 - \mu_2 \neq 0$

$$t < t_{\alpha/2} \text{ or } t > t_{1-\alpha/2},$$

for $H_1: d > 0$ or $H_1: \mu_1 - \mu_2 > 0$

$$t > t_{1-\alpha},$$

Or, for $H_1: d < 0$ or $H_1: \mu_1 - \mu_2 < 0$

$$t < t_{\alpha}.$$

Comparing the *t*-statistic to the cut-off *t*-value (as shown here) is equivalent to comparing the *p*-value to $\alpha$.

# Power Calculation

The power is calculated using the same formulation as in the *Two-Sample T-Tests Assuming Equal Variances* procedure with the modification that the σ used in that procedure is set equal to one.

If the variances cannot be assumed to be equal, the modification suggested by Cohen (1988) is used. This modification is to substitute an average value of the two variances and then proceeding as if the variances were equal. The average value is computed using

$$\sigma = \sqrt{\frac{\sigma_1^2 + \sigma_2^2}{2}}$$

Cohen remarks that this method is only accurate if the two sample sizes are (nearly) equal.

# The Effect Size

If we assume that $\mu_1$ and $\mu_2$ represent the means of the two populations of interest and their common (unknown) standard deviation is $\sigma$, the effect size is represented by *d*, where

$$d = \frac{\mu_1 - \mu_2}{\sigma}$$

Cohen (1988) proposed the following interpretation of the *d* values. A *d* near 0.2 is a *small* effect, a *d* near 0.5 is a *medium* effect, and a *d* near 0.8 is a *large* effect. These values for small, medium, and large effects are popular in the social sciences. However, this convention is not as popular among the medical sciences since the scale of the effect is left unstated which makes interpretation difficult.

# Example 1 – Finding the Sample Size

Researchers wish to compare two types of local anesthesia using a balanced, parallel-group design. Subjects in pain will be randomized to one of two treatment groups, the treatment will be administered, and the subject's evaluation of pain intensity will be measured on a seven-point scale.

The researchers would like to determine the sample sizes required to detect a small, medium, and large effect size with a two-sided t-test when the power is 80% or 90% and the significance level is 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 ...................................................... **Sample Size**
Alternative Hypothesis .................................. **Two-Sided**
Power........................................................... **0.8 0.9**
Alpha............................................................ **0.05**
Group Allocation ........................................... **Equal (N1 = N2)**
d.................................................................. **0.2 0.5 0.8**

## Output

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

### Numeric Reports

**Numeric Results**

Solve For: Sample Size
Test Type: Two-Sample Equal-Variance T-Test
Alternative Hypothesis: H1: d ≠ 0

| Power | | Sample Size | | | Effect Size | |
|---|---|---|---|---|---|---|
| Target | Actual | N1 | N2 | N | d | Alpha |
| 0.8 | 0.8004 | 393 | 393 | 786 | 0.2 | 0.05 |
| 0.9 | 0.9002 | 526 | 526 | 1052 | 0.2 | 0.05 |
| 0.8 | 0.8015 | 64 | 64 | 128 | 0.5 | 0.05 |
| 0.9 | 0.9032 | 86 | 86 | 172 | 0.5 | 0.05 |
| 0.8 | 0.8075 | 26 | 26 | 52 | 0.8 | 0.05 |
| 0.9 | 0.9015 | 34 | 34 | 68 | 0.8 | 0.05 |

Target Power The desired power. It may not be achieved because of integer N1 and N2.
Actual Power The achieved power. Because N1 and N2 are integers, this value is often (slightly) larger than the target power.
N1 and N2 The number of items sampled from each population.
N The total sample size. N = N1 + N2.
d Effect Size. Cohen recommended Low = 0.2, Medium = 0.5, and High = 0.8. d = (μ1 - μ2) / σ.
Alpha The probability of rejecting a true null hypothesis.

**Summary Statements**

A parallel two-group design will be used to test the difference between two means. The comparison will be made using a two-sided, two-sample equal-variance t-test, with a Type I error rate (α) of 0.05. The effect size is defined as d = (μ1 - μ2) / σ, where σ is the common standard deviation for both groups. To detect a population effect size of 0.2 with 80% power, the number of needed subjects will be 393 in Group 1 and 393 in Group 2.

**Dropout-Inflated Sample Size**

| | Sample Size | | | Dropout-Inflated Enrollment Sample Size | | | Expected Number of Dropouts | | |
|---|---|---|---|---|---|---|---|---|---|
| Dropout Rate | N1 | N2 | N | N1' | N2' | N' | D1 | D2 | D |
| 20% | 393 | 393 | 786 | 492 | 492 | 984 | 99 | 99 | 198 |
| 20% | 526 | 526 | 1052 | 658 | 658 | 1316 | 132 | 132 | 264 |
| 20% | 64 | 64 | 128 | 80 | 80 | 160 | 16 | 16 | 32 |
| 20% | 86 | 86 | 172 | 108 | 108 | 216 | 22 | 22 | 44 |
| 20% | 26 | 26 | 52 | 33 | 33 | 66 | 7 | 7 | 14 |
| 20% | 34 | 34 | 68 | 43 | 43 | 86 | 9 | 9 | 18 |

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.
N1, N2, and N The evaluable sample sizes at which power is computed. If N1 and N2 subjects are evaluated out of the N1' and N2' subjects that are enrolled in the study, the design will achieve the stated power.
N1', N2', and N' The number of subjects that should be enrolled in the study in order to obtain N1, N2, and N evaluable subjects, based on the assumed dropout rate. After solving for N1 and N2, N1' and N2' are calculated by inflating N1 and N2 using the formulas N1' = N1 / (1 - DR) and N2' = N2 / (1 - DR), with N1' and N2' 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.)
D1, D2, and D The expected number of dropouts. D1 = N1' - N1, D2 = N2' - N2, and D = D1 + D2.

**Dropout Summary Statements**

Anticipating a 20% dropout rate, 492 subjects should be enrolled in Group 1, and 492 in Group 2, to obtain final group sample sizes of 393 and 393, respectively.

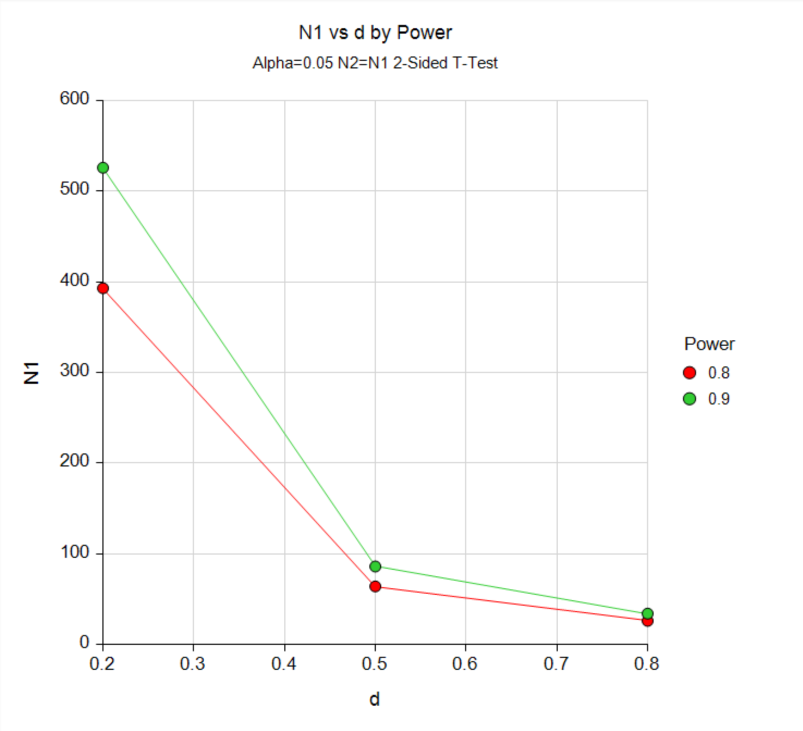
**References**

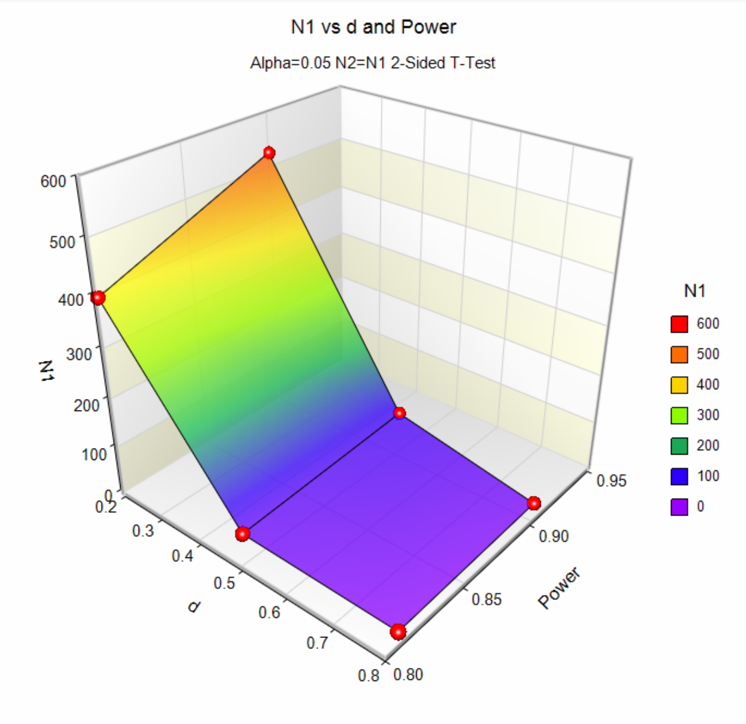
Cohen, Jacob. 1988. Statistical Power Analysis for the Behavioral Sciences. Lawrence Erlbaum Associates. Hillsdale, New Jersey
Julious, S. A. 2010. Sample Sizes for Clinical Trials. Chapman & Hall/CRC. Boca Raton, FL.
Machin, D., Campbell, M., Tan, B. T., Tan, S. H. 2009. Sample Size Tables for Clinical Studies, 3rd Edition. Wiley-Blackwell.
Ryan, Thomas P. 2013. Sample Size Determination and Power. John Wiley & Sons. New Jersey.

These reports show the values of each of the parameters, one scenario per row.

## Plots Section

### Plots




These plots show the relationship between effect size, power, and sample size.

# Example 2 – Finding the Power

Researchers wish to compare two types of local anesthesia using a balanced, parallel-group design. Subjects in pain will be randomized to one of two treatment groups, the treatment will be administered, and the subject's evaluation of pain intensity will be measured on a seven-point scale.

The researchers would like to determine the power to detect a small, medium, and large effect size with a two-sided t-test for group sample sizes of 25, 50, 100, 200, 400 and a significance level of 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 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** |
| Alpha | **0.05** |
| Group Allocation | **Equal (N1 = N2)** |
| Sample Size Per Group | **25 50 100 200 400** |
| d | **0.2 0.5 0.8** |

# Output

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

**Numeric Results**

Solve For: Power
Test Type: Two-Sample Equal-Variance T-Test
Alternative Hypothesis: H1: d ≠ 0

| | Sample Size | | | Effect Size | |
|---|---|---|---|---|---|
| Power | N1 | N2 | N | d | Alpha |
| 0.1066 | 25 | 25 | 50 | 0.2 | 0.05 |
| 0.1677 | 50 | 50 | 100 | 0.2 | 0.05 |
| 0.2906 | 100 | 100 | 200 | 0.2 | 0.05 |
| 0.5141 | 200 | 200 | 400 | 0.2 | 0.05 |
| 0.8073 | 400 | 400 | 800 | 0.2 | 0.05 |
| 0.4101 | 25 | 25 | 50 | 0.5 | 0.05 |
| 0.6969 | 50 | 50 | 100 | 0.5 | 0.05 |
| 0.9404 | 100 | 100 | 200 | 0.5 | 0.05 |
| 0.9988 | 200 | 200 | 400 | 0.5 | 0.05 |
| 1.0000 | 400 | 400 | 800 | 0.5 | 0.05 |
| 0.7915 | 25 | 25 | 50 | 0.8 | 0.05 |
| 0.9773 | 50 | 50 | 100 | 0.8 | 0.05 |
| 0.9999 | 100 | 100 | 200 | 0.8 | 0.05 |
| 1.0000 | 200 | 200 | 400 | 0.8 | 0.05 |
| 1.0000 | 400 | 400 | 800 | 0.8 | 0.05 |

# Example 3 – Validation using Another Procedure

This procedure should give identical results to the **Two-Sample T-Tests Assuming Equal Variance** procedure when the value of σ there is set to one. We will use this fact to provide a validation problem for this procedure.

If we run that procedure with power = 0.90, alpha = 0.05, $\mu_1$ = 1, $\mu_2$ = 0, σ = 1, and solve for sample size with N1 = N2, we obtain N1 = N2 = 23.

## 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 3** 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 | **0.90** |
| Alpha | **0.05** |
| Group Allocation | **Equal (N1 = N2)** |
| d | **1** |

## Output

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

**Numeric Results**

Solve For: Sample Size
Test Type: Two-Sample Equal-Variance T-Test
Alternative Hypothesis: H1: d ≠ 0

| Power | | Sample Size | | | Effect Size | |
|---|---|---|---|---|---|---|
| Target | Actual | N1 | N2 | N | d | Alpha |
| 0.9 | 0.9125 | 23 | 23 | 46 | 1 | 0.05 |

This procedure also calculated N1 = N2 = 23, thus the procedure is validated.