

Sample Size Workbook: User's Guide

Table of Contents

Overview.....	2
Getting Started.....	2
Using the Sample Size Workbook	4
AIPE: Single Mean.....	4
AIPE: Two Means.....	4
AIPE: Single Proportion.....	4
AIPE: Two Proportions.....	4
AIPE: Correlation.....	5
SPA: Mean Difference	5
SPA: Difference in Proportions.....	5

Overview

The **Sample Size** Workbook (SSW) is an Excel workbook that allows you to compute sample size requirements for various types of simple quantitative evaluation designs.

The SSW recognizes two competing methods for computing required sample sizes for quantitative studies. The accuracy in parameter estimation (AIPE) method seeks to determine the sample size necessary to achieve a predetermined confidence interval.¹ The second method, *statistical power analysis* (SPA), seeks to determine the sample size necessary to ensure a certain level of statistical power for a hypothesis test, given an assumption about the population.

The SSW relies on the assumption that SPA is most appropriate for studies that compare two groups and less appropriate for studies of a single group. Therefore, the workbook assists with sample size planning based on SPA for evaluation designs that compare two groups. On the other hand, the AIPE method is appropriate for a wide range of situations, so functionality for planning based on AIPE is provided for both designs with one group and designs that compare two groups.

The SSW is intended to be used when the data collection process will result in independent observations. Independence means that the value of data collected for one individual in the sample provides no information about the likely value of data collected for another individual in the sample. Independence excludes evaluation designs that involve clustering, for instance, if data for students will be obtained by first sampling schools and then collecting measurements for students in those schools. If you need to compute sample size requirements for designs of these types you could consult, for example, the Optimal Design software (<http://hlmsoft.net/od/>) or the PowerUp! software (<https://www.causalevaluation.org/power-analysis.html>).

Getting Started

To use the SSW, it is necessary to determine (1) the type of variable that will serve as the outcome variable for the main evaluation question (continuous or binary), (2) the analytic technique that will be used to answer the main evaluation question, and (2) the sample size planning method that will be used (AIPE or SPA).

If you use the AIPE method, the SSW may compute the necessary sample size when (1) a single mean will be estimated (continuous outcome), (2) a single proportion will be estimated (binary outcome), (3) means in two groups will be compared (continuous outcome), (4) proportions in two groups will be compared (binary outcome), or (5) a correlation will be computed (continuous outcome). Here, it is assumed that a 95 percent confidence interval is desired. This functionality is provided in the first tab of the Sample Size Workbook, *Confidence Interval*.

¹ Maxwell, S. E., Kelley, K., & Rausch, J. R. (2008). Sample size planning for statistical power and accuracy in parameter estimation. *Annual Review of Psychology*, 59, 537–563.
<https://doi.org/10.1146/annurev.psych.59.103006.093735>

If you plan to use SPA, the SSW may compute the necessary sample size when (1) means in two groups will be compared (continuous outcome) or (2) proportions in two groups will be compared (binary outcome). Here, it is assumed that you will use a hypothesis test with a level of significance at 0.05 and a statistical power of 0.80. The continuous outcome analysis may be conducted within the *Power-Mean Difference* tab, and the binary outcome analysis may be conducted within the *Power-Proportion Difference* tab.

Using the **Sample Size** Workbook

AIPE: Single Mean

In the *Confidence Interval* tab, find the heading *Confidence Interval for a Single Mean* (cell B4). In cell C5, enter the desired width of the confidence interval. As explained in Module 4, Chapter 3 of the Program Evaluation Toolkit, this length should be in standard deviation units. For instance, if you desire to be 95 percent confident that the population mean is within a 2-unit interval, and the standard deviation of the main outcome variable is believed to be 4 units, then you should enter $2/4=0.5$ in cell C5. Once a value has been entered in cell C5, the required sample size will appear in cell C6.

AIPE: Two Means

In the *Confidence Interval* tab, find the heading *Confidence Interval for Comparing Two Means* (cell B8). In cell C9, enter the desired width of the confidence interval. As explained in Module 4, Chapter 3 of the Program Evaluation Toolkit, this length should be in standard deviation units. For instance, if you desire to be 95 percent confident that the difference between the two population means is within a 2-unit interval, and the standard deviation of the main outcome variable is believed to be 4 units in each group, you should enter $2/4=0.5$ in cell C9. In cell C10, you should enter the proportion of the sample that is expected to come from the first group. For instance, if a sample size of 100 is expected to contain 25 individuals from Group 1 and 75 individuals from Group 2, you should enter 0.25.

AIPE: Single Proportion

In the *Confidence Interval* tab, find the heading *Confidence Interval for a Single Proportion* (cell B13). Two values must be entered to compute the required sample. In cell C14, you should specify the desired interval width in percentage points. In cell C15, you should estimate the value of the true population proportion of “successes.” For instance, if you believe that the true proportion of successes will be 0.20, and it is desired for the confidence interval to be 10 percentage points wide, then 10 should be entered in cell C14, and 0.20 should be entered in cell C15.

AIPE: Two Proportions

In the *Confidence Interval* tab, find the heading *Confidence Interval for Comparing Two Proportions* (cell B18). You must enter four values to compute the required sample size. In cell C19, you should specify the desired interval width for the difference between the proportions in percentage points. In cell C20, you should estimate the true population proportion of “successes” in the first group. In cell C21, you should estimate the true population proportion of “successes” in the second group. Finally, in cell 22, you should enter the proportion of the sample that is expected to come from the first group.

AIPE: Correlation

In the *Confidence Interval* tab, find the heading *Confidence Interval for Correlation* (cell B25). You should enter the desired width of the confidence interval in cell C26. In cell C27, you should enter the estimated value of the correlation in the population. Note that the correlation must range between -1 and 1 and the confidence interval must be contained within these bounds. Therefore, certain entries in cells C26 and C27 are not allowable. For instance, it would not be allowable to enter 0.4 for the desired interval width along with an expected correlation of 0.9 because the expected upper limit of the confidence interval would then be 1.1, which is larger than 1.

SPA: Mean Difference

In the *Power-Mean Difference* tab, find the heading *Power Analysis for Comparing Two Means* (cell B4). The calculations assume that the difference between the means in the two groups will be estimated by multiple linear regression, with an indicator variable for group membership as well as other covariates added to the equation as additional independent variables.

Three values are needed to compute the necessary sample size. In cell C5, you should enter the expected difference in average outcomes between the two groups, expressed in standard deviation units. For instance, if you expect that the mean outcome in the first group will exceed the mean difference in the second group by 2 and that the standard deviation of the outcomes is 10, then you should enter $2/10=0.20$ in cell C5. In cell C6, you should enter the expected proportion of variance in the outcome variable that is explained by its relationship with variables in the multiple regression model *other than* the treatment indicator variable (this is a number between 0 and 1). Finally, in cell C7, you should enter the proportion of the sample that is expected to come from the first group. Once the three values have been entered, the required sample size will appear in cell C8.

SPA: Difference in Proportions

In the *Power-Proportion Difference* tab, find the heading *Power Analysis for Comparing Two Proportions* (cell B4). Three values are needed to compute the necessary sample size. In cell C5, you should specify the estimated value of the true population proportion of “successes” in the first group. In cell C6, you should estimate the value of the true population proportion of “successes” in the second group. Finally, in cell 7, you should enter the proportion of the sample that is expected to come from the first group. Once the three values have been entered, the required sample size will appear in cell C8.