evanmiller.org/ab-testing/sample-size.html

Evan's Awesome A/B Tools (home):

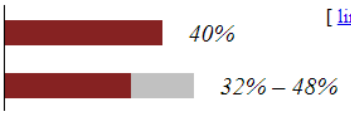
Sample Size Calculator | Chi-Squared Test | Sequential Sampling | 2 Sample T-Test | Survival Times | Count Data

Question: How many subjects are needed for an A/B test?

Baseline conversion rate: 40 %

Minimum Detectable Effect: 20 %

[link]



The Minimum Detectable Effect is the smallest effect that will be detected (1-β)% of the time.

○ Absolute
◉ Relative

Conversion rates in the gray area will not be distinguishable from the baseline.

Sample size:

592

per variation

Statistical power 1−β: 80% *Percent of the time the minimum effect size will be detected, assuming it exists*

Significance level α: 5% *Percent of the time a difference will be detected, assuming one does NOT exist*

See also: How Not To Run an A/B Test

*Need A/B sample sizes on your iPhone or iPad? Download **A/B Buddy** today.*

Sample Size Calculator for Comparing Two Independent Means

✔ Provides live interpretations.
✔ Evaluates the influence of changing input values.
✔ Adjusts sample sizes for continuity and clustering.

Equality | Non-inferiority | Superiority | Equivalence

Calculate | Visualise | Tabulate

Input Values

Select one of the two options to specify input values. Hover over the sign to obtain help.

- Expected Means
  - Mean of the Reference Group: 25
  - Mean of the Test Group: 27,5
  - Standard Deviation: 156
- Expected Difference between Means

Click the Options button to change the default options for Power, Significance, Alternate Hypothesis and Group Sizes. Use the Adjust button to adjust sample sizes for t-distribution (option applied by default), and clustering.

▶ Calculate | Options | Adjust | Reset

Results and Live Interpretation

Download

Assuming a pooled standard deviation of 156 units, the study would require a sample size of:

77858

for each group (i.e. a total sample size of 155716, assuming equal group sizes), to achieve a power of 80% and a level of significance of 5% (two sided), for detecting a true difference in means between the test and the reference group of 2.5 (i.e. 27.5 - 25) units.

In other words, if you select a random sample of 77858 from each population, and determine that the means of the test and the reference groups are 27.5 and 25 units, respectively, and the standard deviation is 156 units, you would have 80% power to declare that the two groups have significantly different means, i.e. a two sided p-value of less than 0.05.

Reference: Dhand, N. K., & Khatkar, M. S. (2014). Statulator: An online statistical calculator. Sample Size Calculator for Comparing Two Independent Means. Accessed 4 April 2024 at http://statulator.com/SampleSize/ss2M.html

Note: Statulator used the input values of a power of 80%, a two sided level of significance of 5% and equal group sizes for sample size calculation and adjusted the sample size for t-distribution. You may change the options by clicking here or the 'Options' button and the adjustments by clicking here or the 'Adjust' button.