

Урок 6. Расчёт длительности A/B теста

- 1) Продакт на главной mail.ru решил протестировать в рекомендательной ленте контента вместо карточек со статьями видеоплеер с короткими видео. Нынешний таймспент на юзера в день в среднем равен 25 минут, а стандартная ошибка (SD) равна 156. Мы предполагаем, что в новой версии таймспент на юзера в день изменится на 10%. Средний трафик 20000 человек в день. Посчитайте сколько дней необходимо держать эксперимент при $\alpha = 5\%$ и $\beta = 20\%$.

Statulator Sample Size Descriptive Analysis Statistical Tests About Blog Contact us

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:

61124

for each group (i.e. a total sample size of 122248, 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 61124 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 9 June 2023 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.

© Statulator 2023 All Rights Reserved

Statulator Sample Size Descriptive Analysis Statistical Tests About Blog Contact us

Calculate Visualise Tabulate

Input Values

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

☐ Expected Means ⓘ

☒ Expected Difference between Means ⓘ

Difference between Two Means: ⓘ 2.5

Expected Standard Deviation: ⓘ 156

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:

61124

for each group (i.e. a total sample size of 122248, 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 units.

In other words, if you select a random sample of 61124 from each population, and determine that the difference in the two means is 2.5 units, and the pooled 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 9 June 2023 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.

© Statulator 2023 All Rights Reserved

About Blog Contact us Privacy Terms


Для того, чтобы получить статистически значимый результат при заданном уровне конверсии на обеих выборках, нам нужно направить на каждый вариант теста (A и B) по 61124 юзера, а в общем 122248.


$122248 / 20000 = 6,1124$ – то есть для проведения теста необходимо чуть больше 6 или 7 полных дней.

- 2) Наша продуктовая команда в ecommerce магазине планирует запустить тест, направленный на ускорение загрузки сайта. Одна из основных метрик bounce rate в GA = 40%. Мы предполагаем, что при оптимизации сайта она изменится минимум на 20%. Средний трафик 4000 человек в день. Посчитайте сколько нам нужно дней держать эксперимент при alpha = 5% и beta = 20%

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% [\[link\]](#)

Minimum Detectable Effect: %  32% – 48%

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: % Percent of the time the minimum effect size will be detected, assuming it exists

Significance level α: % 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.

Для того, чтобы получить статистически значимый результат при заданном уровне конверсии на обеих выборках, нам нужно направить на каждый вариант теста (А и В) по 592 юзера, а в общем 1184.

$1184 / 4000 = 0,296$ - то есть для проведения теста необходимо треть дня или один полный день.