

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 $\beta = 20\%$.

Решение: для решения задачи используем калькулятор <https://www.statulator.com/SampleSize/ss2M.html>. Ищем Expected Means. Значение Mean of the Reference Group задаем 25.

Mean of the Test Group равен 27,5

Показателю Standard Deviation присваиваем значение 156.

Щелкнув по кнопке Options устанавливаем:

Desired Power: $1 - \beta = 0,8$

Level of Significance: $\alpha = 0,05$

Alternate Hypothesis: Two Sided

Group Sizes = 20000

Результат расчета Calculate выдает 77858. С учетом двух вариантов теста имеем формулку 😊 $77858 \times 2 = 155716$.

Сколько в день наш сайт посещает посетителей (aDAU)? ~ 20 000 пользователей. Делим $\frac{155716}{20000} \approx 7,8$, т.е. около 8-ми дней должен длиться тест, если мы его реализуем на всех пользователях.

Sample Size Calculation Results

Results and Live Interpretation

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 10 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.

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 $\beta = 20\%$.

Решение: на сайте <https://outgrow.co/blog/how-to-lower-bounce-rates> сказано: “By now we understand that bounce rate and conversion rates very much go hand in hand. Hence, aim to reduce your bounce rate and you should see your conversion rates rise as a result, as the customer will be more willing to engage”.

В калькуляторе <https://www.evanmiller.org/ab-testing/sample-size.html> определяем:

Baseline conversion rate: 40%

Minimum Detectable Effect: 20%

Относительные изменения легче применять к различным метрикам, поэтому устанавливаем переключатель Relative, абсолютные метрики менее удобны.

Statistical power $1-\beta$ устанавливаем на 80% и α (false positive) на 5%.

Нужно по 592 юзера. В лекции <https://gb.ru/lessons/328106> на 16:25 сказано, что это ещё нужно на два умножить, когда считаем итоговую выборку, т.е. 1184 пользователя.

Сколько в **день** наш сайт посещает посетителей (aDAU)? ~ 4000 пользователей.

Делим $\frac{1184}{4000} \approx 0,3$, т.е. примерно один рабочий день (8 часов) нужно дней держать эксперимент.