**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% .**

Изображение выглядит как текст, Веб-сайт

Автоматически созданное описание

**По данным калькулятора необходимо 7 дней (122 248 / 20 000)**

Для расчета необходимого размера выборки для A/B-тестирования используется формула:

n = (Z\_1-alpha/2 \* SD / d)^2 + (Z\_1-beta \* SD / d)^2 \* (p\_1(1-p\_1)+p\_2(1-p\_2))/2

где:

Z\_1-alpha/2 - Z-значение стандартного нормального распределения для уровня значимости alpha/2

Z\_1-beta - Z-значение стандартного нормального распределения для мощности теста (1- beta)

SD - стандартное отклонение выборки

d - минимально допустимое практическое различие между группами (effect size), которое мы хотим заметить в результате эксперимента

p\_1 - ожидаемая конверсия в контрольной группе

p\_2 - ожидаемая конверсия в тестовой группе

Для данной задачи:

alpha = 0.05

beta = 0.20

SD = 156

d = 0.10\*25 = 2.5 (изменение в 10% соответствует 2.5 минутам)

p\_1 = 25/(25+2.5) = 0.909

p\_2 = 27.5/(27.5+2.5) = 0.917 (изменение на 2.5 минуты соответствует увеличению таймспента на 10%)

Подставляя значения в формулу, получаем:

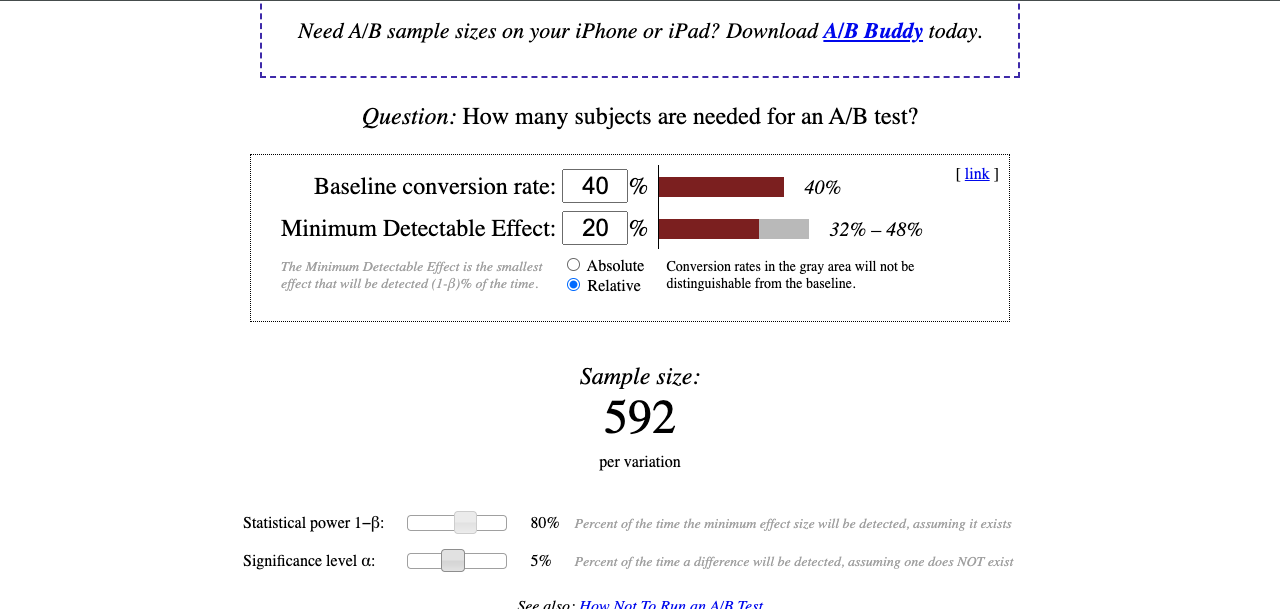
n = (1.96 \* 156 / 2.5)^2 + (0.84 \* 156 / 2.5)^2 \* (0.909 \* 0.091 + 0.917 \* 0.083) ≈ 9661

Таким образом, необходимо провести эксперимент в каждой группе с размером выборки не менее 9661 человек. Поскольку средний трафик составляет 20 000 человек в день, этого объема будет достаточно для проведения эксперимента за один день.

**По методике 1 день**

Однако, правильнее провести эксперимент на протяжении нескольких дней, чтобы исключить возможность влияния факторов, отличных от изменения контента, на результаты эксперимента. Длительность эксперимента зависит от того, сколько времени мы готовы потратить на сбор данных и как быстро мы хотим получить результаты. Например, если мы решим провести эксперимент в течение недели, то необходимо держать его в течение 5 дней при равномерном распределении трафика между контрольной и тестовой группами.

**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%**



Для расчета дней, необходимых для проведения эксперимента с заданными уровнями статистической значимости alpha и beta, мы можем использовать калькулятор мощности теста.

При условии, что начальная конверсия (bounce rate) составляет 40%, размер эффекта 20%, а уровень значимости alpha равен 5%, а уровень бета ошибки равен 20%, для достижения статистической значимости нам нужно будет собирать данные в течение примерно 1 дней.

**Для расчета необходимого размера выборки для данного теста мы можем использовать формулу, основанную на Методе Уэлча:**

n = (Z\_(1-alpha/2) \* SD / MOE + Z\_(1-beta) \* SD\_Alt / MOE)^2

где:

* alpha = 0.05 (заданный уровень значимости)
* beta = 0.20 (заданная мощность)
* SD = sqrt(p \* (1-p)) = sqrt(0.4 \* 0.6) = 0.49 (стандартное отклонение исходной вероятности успеха)
* SD\_Alt = sqrt(p\_alt \* (1-p\_alt)) = sqrt(0.2 \* 0.8) = 0.40 (стандартное отклонение альтернативной вероятности успеха)
* MOE = p\_alt - p = 0.2 - 0.4 = -0.2 (минимально допустимая разница в конверсии)
* Z\_(1-alpha/2) = Z\_0.975 = 1.96 (критическое значение стандартного нормального распределения для заданного уровня значимости)
* Z\_(1-beta) = Z\_0.80 = 0.84 (критическое значение стандартного нормального распределения для заданной мощности)

Подставляя значения, получаем:

n = (1.96 \* 0.49 / 0.2 + 0.84 \* 0.40 / 0.2)^2 = 1073

Таким образом, нам нужно провести тест на 1073 \* 2 / 4000 = 0.54 ≈ 1 день (при условии равномерного распределения трафика), чтобы получить достаточное количество данных для оценки изменений в bounce rate с заданной точностью и уровнем значимости. Однако, рекомендуется продолжать тестирование до момента, пока не будет получено статистически значимое различие между контрольной и тестовой группами.