**Задание 1.**

|  |  |  |
| --- | --- | --- |
|  | Послали email | Открыли email |
| Группа А | 10 000 | 2 200 |
| Группа B | 10 000 | 2 000 |

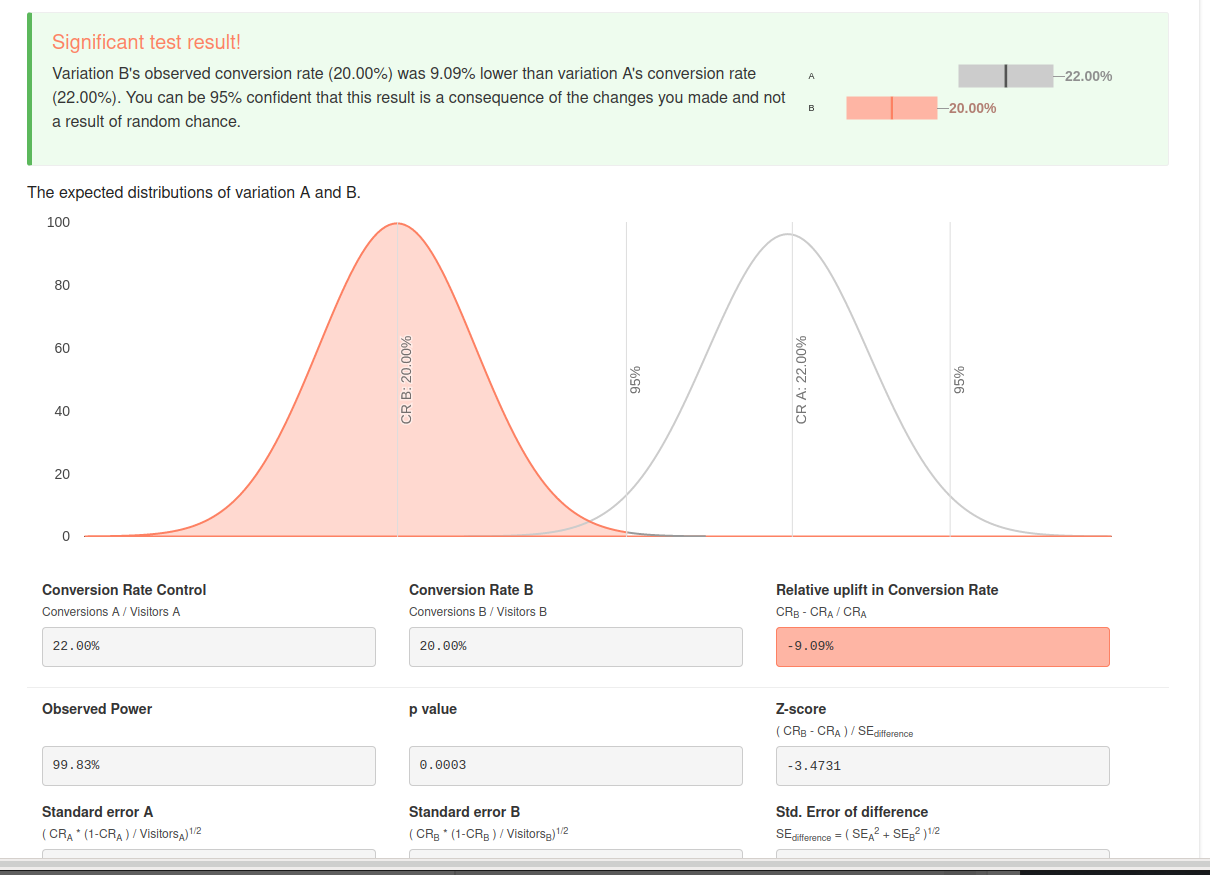
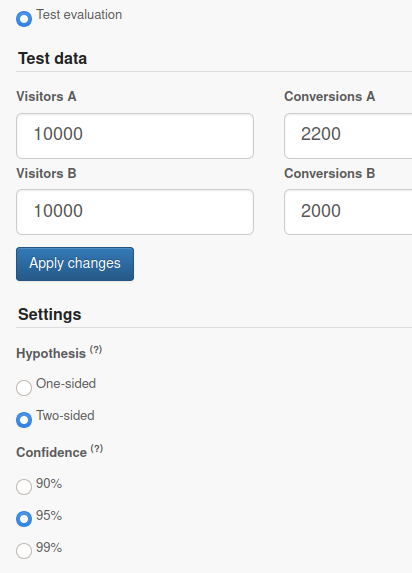
Маркетологи тестировали тему email-письма (разным группам послали email с разными темами). Целевой метрикой было процент открытий.

Какой теме стоит отдать предпочтение и почему? Объясните свое решение

Решение:

Результаты two-side теста статистически значимы.

На основе статистической значимости мы можем опровергнуть или не опровергнуть нулевую гипотезу. Решение о выборе темы тоже зависит от первоначальной гипотезы. Если H0 — нет разницы между темами для пользователей, то мы можем ее опровергнуть, разница есть. Разница может быть и плохой, и хорошей. Демонстрация из калькулятора:

Обычно two-side тест применяется, когда нет уверенности в том, что наши изменения улучшат результат, а не ухудшат, поэтому для гипотезы без точной формулировки о предпочтительности какой-то из тем я выставила такие параметры.

Можно выбрать тему A, если это действие изначально подразумевается после опровержения нулевой гипотезы.

**Задание 2.**

|  |  |  |
| --- | --- | --- |
|  | Послали email | Открыли email |
| Группа А | 10 000 | 2 070 |
| Группа B | 10 000 | 2 000 |

Кейс тот же, но итоги теста отличаются.

Решение:

Результаты не являются статистически значимыми (ни для one-side, ни для two-side). Если нулевая гипотеза — темы одинаково интересные, то мы не можем ее опровергнуть. Если заранее было определено, что при «равенстве тем» оставляется тема по умолчанию и эта тема A/B, тогда оставляем тему А/B соответственно. Если заранее было определено, что при «равенстве тем» пользователям рандомно будут отправляться письма на обе темы, то поступаем так. Нельзя точно сказать, какой теме отдать предпочтение, если непонятно, нужно ли это вообще делать.

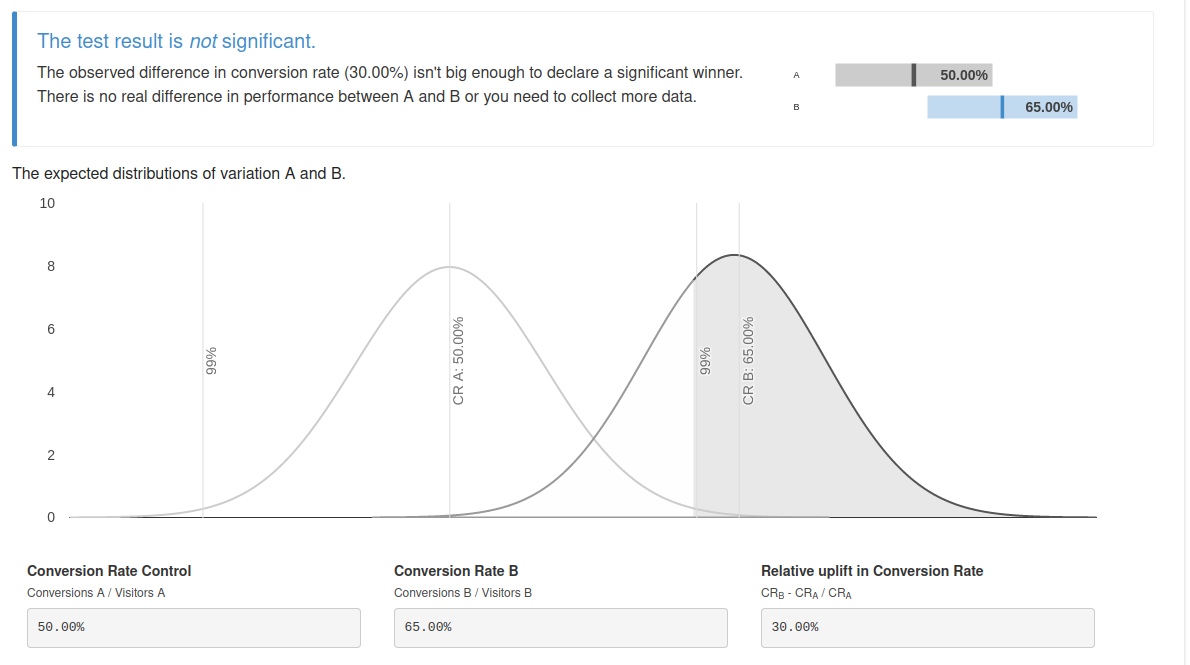
**Задание 3.**

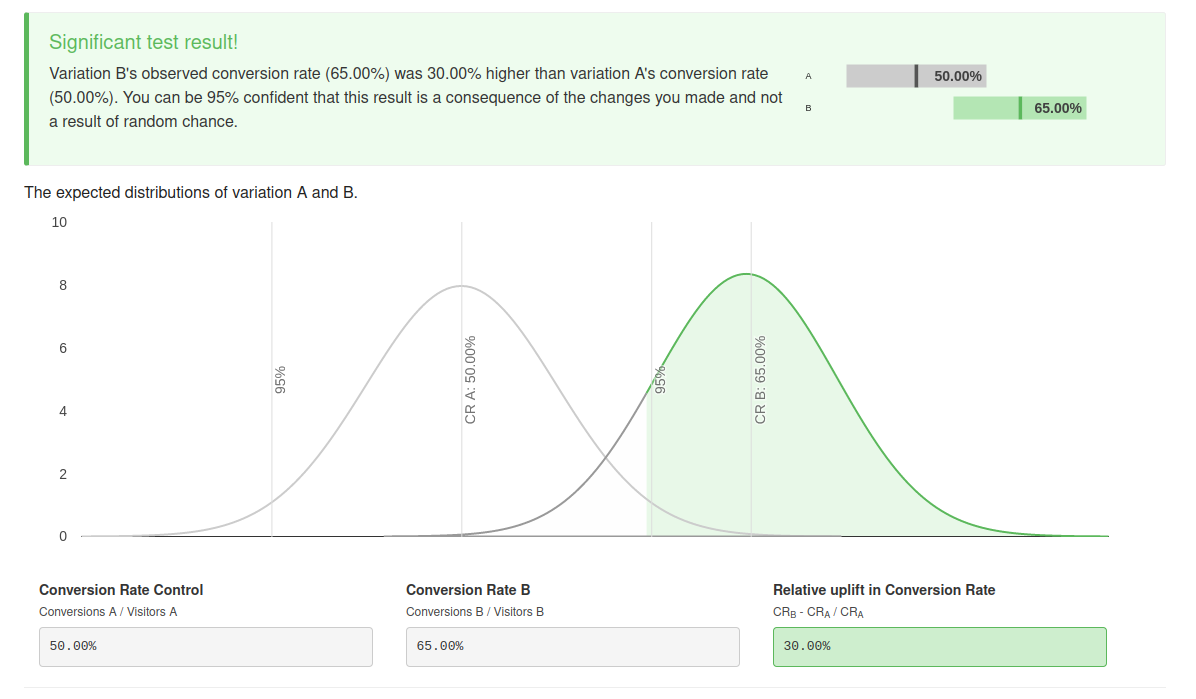
|  |  |  |
| --- | --- | --- |
|  | Размер группы | Совершили оплату |
| Тест | 100 | 65 |
| Контроль | 100 | 50 |

Тестовой группе отправляется пуш с призывом совершить оплату, контрольной группе не отправляли. Является ли различие статистически значимым при уровне значимости 1%? А при 5%?

Решение:

При уровне значимости 1 % различие не является статистически значимым, при 5% - является (в обоих видах тестирования, разумеется).



****

**Задание 4.**

Хотим протестировать новую версию сайта. Ожидаем, что вырастет число регистраций.   
Обычно у нас дневная конверсия из захода на сайт в регистрацию 50% и хотим поймать изменение на 5 %-пунктов при уровне значимости 1% и мощности 80%.

Так же известно, что каждый день к нам на сайт приходит 100 уникальных пользователей.

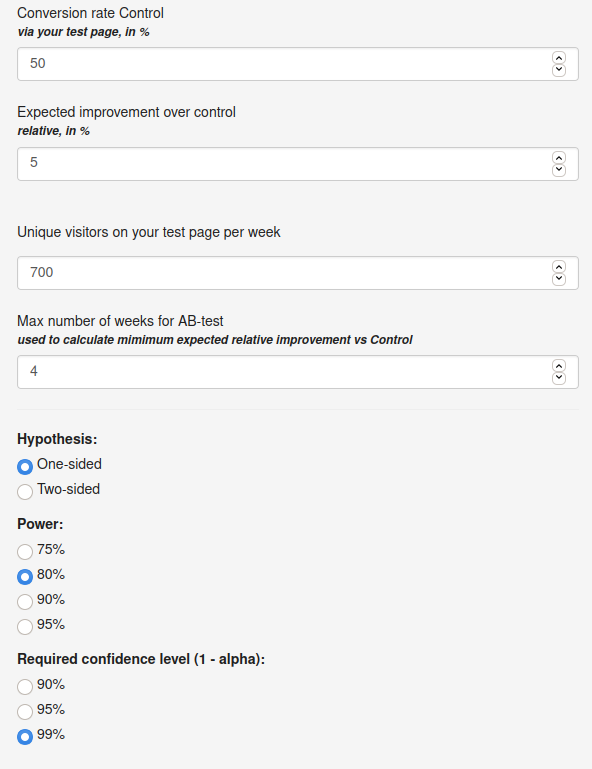
Какой размер выборки нам нужен? Сколько дней требуется проводить АВ тест?

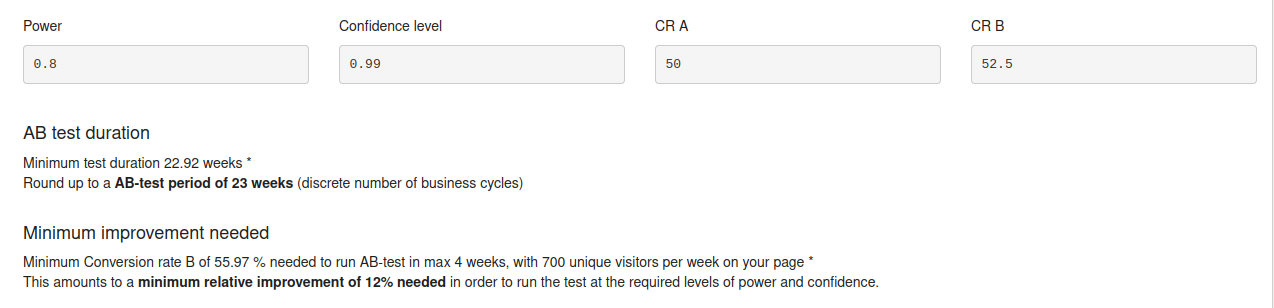
Решение:

8022 — размер выборки

23 недели (округленное значение)

Демонстрация из калькулятора:





**Задание 5.**

|  |  |  |
| --- | --- | --- |
|  | Размер группы | Оставили заявку |
| Тест | 1000 | 600 |
| Контроль | 1000 | 500 |

Проводим обзвон клиентов (тест – звоним, контроль – нет). Цель – чтоб клиент оставил заявку на продукт. Мы знаем, что средняя стоимость одного звонка 50 рублей (дозвонились или нет не важно), с одной заявки мы зарабатываем в среднем 180 рублей.

Проинтерпретируйте результаты АВ теста

Решение:

Если альтернативная гипотеза состоит в том, что тестовая группа более эффективная, а нулевая — в том, что обе группы дают одинаковую эффективность, то можно опровергнуть нулевую гипотезу и выбрать более эффективный тестовый вариант.

Мы в группе тестовой потратили по 50000 рублей, заработали 108000 рублей, в контрольной заработали 90000 рублей. Итого прибыль для тестовой группы — 58к, для контрольной — 90к. Ни в одном из случаев мы не остались в убытке.

Тест показывает статистически значимый результат. Способ действительно можно считать более эффективным, но у данного способа нет практической значимости, так как стоимость звонка слишком дорогая и фактически при применении более эффективного метода сбора заявок компания наоборот будет меньше зарабатывать.