

Изучение конверсии пользователей

Постановка задачи

Однажды я проходил собеседование на модную должность «DataScientist»-а в одну известную компанию - электронную торговую площадку. И вот одно из тестовых заданий звучало следующим образом: маркетологи решили увеличить посещения своего сайта за счет изменения дизайна. Они провели эксперимент на пользователях, часть из них видела старый интерфейс, часть - новый. Как и всегда, пользователи совершали покупки. Результаты их действий записывались в базу данных. Ее дамп приложен.

Требуется воссоздать базу данных и проанализировать изменение конверсии пользователей.

Решение

Для нас самой интересной частью является именно расчет и анализ изменения конверсии, а также оценка ее значимости.

Сама конверсия вычисляется по формуле $\delta = \frac{n_+}{n_+ + n_-}$, где n_+ - количество положительных исходов (пользователь купил товар), n_- - количество отрицательных исходов (пользователь посмотрел на товар, но не приобрел его). Как можно заметить, конверсия эквивалентна вероятности купить товар.

Изменение же конверсии при АВ тестировании находится так:

$$\Delta = \delta_A - \delta_B$$

Итак, вернемся к нашим баранам, а точнее к результатам АВ-тестирования

Исход	А	В
Конверсия	7339	6645
Нет конверсии	19843	16530

Вычисляем конверсию для случая А:

$$\delta_A = \frac{7339}{7339 + 19843} = 0.27$$

а для случая В:

$$\delta_B = \frac{6645}{6645 + 16530} = 0.29$$

Изменение конверсии составило

$$\Delta = 2.0\%$$

Ага, в случае В, когда кнопки были другого цвета, конверсия возросла аж на 2%. Теперь нам нужно оценить, значимо ли это изменение, либо оно могло произойти случайным образом. Для этого сформируем нулевые гипотезы. Их у нас может быть три:

1. $H_0: p_A = p_B$
2. $H_0: p_A \leq p_B$
3. $H_0: p_A \geq p_B$

и соответствующие альтернативные гипотезы:

1. $H_1: p_A \neq p_B$
2. $H_1: p_A > p_B$
3. $H_1: p_A < p_B$

По сути, нам нужно показать, что изменение конверсии не случайно, то есть опровергнуть гипотезу $H_0: p_A = p_B$.

Так как у нас серия успехов и провалов, то результаты этого эксперимента подчиняются распределению Бернулли с вероятностями успеха (p) и провала ($q=1-p$).

В итоге Z статистика вычисляется по следующей формуле:

$$z = \frac{p_A - p_B}{\sqrt{pq/n_A - pq/n_B}}$$

n_A - размер выборки A

n_B - размер выборки B

p_A - вероятность успеха по выборке A

p_B - вероятность успеха по выборке B

p, q - вероятности успехи и провала по всей совокупности

Расчеты способ 1

Определим вероятности

$$p_A = 0.27,$$

$$p_B = 0.29,$$

$$p = 13984/50357 = 0.28,$$

$$q = 1 - p = 0.72$$

$$n_A = 27182$$

$$n_B = 23175$$

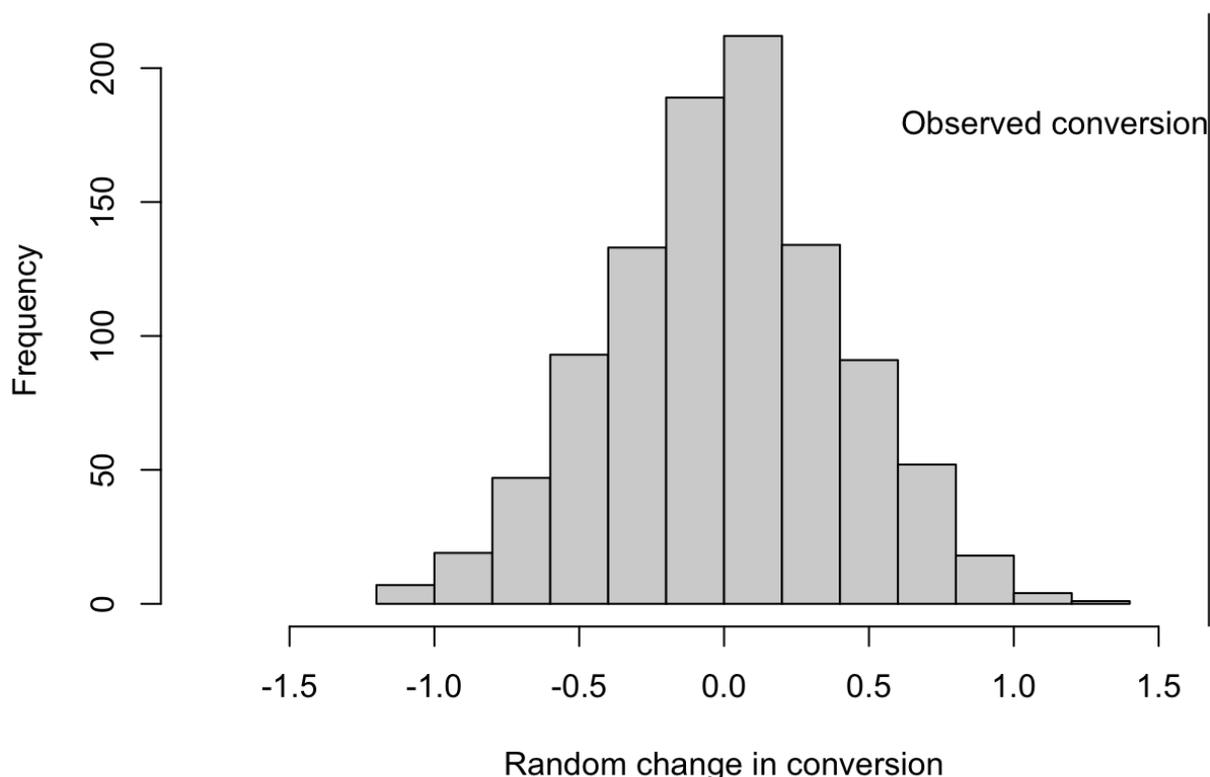
Вычислим Z - значение:

$$Z = \frac{0.29 - 0.27}{\sqrt{\frac{0.2016}{23175} - \frac{0.2016}{27182}}} = 17.6$$

Получили, что $|Z| > 1.96$, значит разница в конверсии является значимой.

Расчеты способ 2

Второй способ заключается в использовании бутстрап подхода. Мы моделируем все исходы исходя из знаний об откликах и количестве экспериментов. Делаем случайную выборку из n_A элементов из общего числа экспериментов двух исходов. Отставшаяся часть будет выборкой n_B с количеством элементов n_B . Разность между двумя выборками и будет изменение конверсии. Эту процедуру повторяем большое количество раз (в нашем случае я выбрал $N=1000$). Полученное в каждом эксперименте изменение конверсии формирует распределение, которое является нормальным. Оно будет иметь некоторое математическое ожидание. Если мы сравним полученное в результате бутстрапа распределение с измеренным изменением конверсии (2%)



На рисунке вертикальная линия представляет измеренную конверсию в 2%. Вероятность того, что она не случайная равна вероятности площади распределения справа от вертикальной линии. Она равна $p_value = 3.05 \times 10^{-5}$. Она меньше, чем уровень значимости $\alpha = 0.05$, следовательно нулевая гипотеза отвергается. И значит наблюдаемое изменение конверсии не случайно. Хотя мы оставляем вероятность случайного изменения конверсии в 5% случаев.

From:

<http://lidarbackup.dvo.ru/dokuwiki/> - **Записки репетитора**

Permanent link:

<http://lidarbackup.dvo.ru/dokuwiki/doku.php/posts:conversion>



Last update: **2022/04/27 04:17**