

Описание домашнего задания №3

Дано: файлы c1.txt и c2.txt, содержащие клики пользователей со сплита 1 и 2 соответственно

Формат: каждая строка соответствует одной сессии и содержит десять цифр: нулей или единиц, обозначающих был ли клик в документ или нет в порядке возрастания ранга документа в поисковой выдаче.

Например, начало c1.txt:

```
1 1 1 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0
1 0 0 0 0 0 0 0 0 0
```

Означает следующее. Представлена информация о трех поисковых сессиях. Сессия 1 – пользователь просмотрел (кликнул) в документы 1, 2, 3. В сессию 2 кликов не было, а в сессии 3 был клик только в первый документ.

Задача: методом бутстрепинга проверить гипотезу H_0 о том, что CTR на двух сплитах можно объяснить шумами. Выполнить данный подсчет для любой другой кликовой метрики, например: количество отказов (доля сессий без кликов), доля кликов в первый результат, позиция последнего клика, среднее количество кликов на сессию и т.д.

Как считать метрику: Обычно кликовые метрики считают следующим образом: выбирают за определенный промежуток времени сессии (например за 1 минуту) и считают метрику. Например, за 1 минуту на сплит пришло 100 запросов, из них 20 сессий без кликов, тогда CTR равен $(100-20)/100 = 0,8$. В вашем случае информации о времени нет, поэтому возможно просто разбить сессии на группы по 100 или по 1000 сессий, превратить каждую группу в точки метрики, CTR или другую. Количество сессий в одном сплите 1 000 000, тогда взяв в группу, например, 1000 сессий, получим $(1\,000\,000/1000) = 1000$ значений CTR.

Как проверять гипотезу: Считать t-статистику, и проверять условие $t > t^*$, где $t^* = 1,96$ для $\alpha=5\%$.

Что делать с бутстрепингом: На каждом шаге семплированием получать новое множество значений метрики (аналогично, как делали на семинаре для множества побед и ничьих), проверять гипотезу и считать количество итераций, на которых гипотеза H_0 может быть отвергнута. По результату посчитать ASL(уровень значимости) как count/N , где count — количество итераций при которых гипотезу H_0 была отвергнута, а N общее количество итераций.

В [этой статье](#) доступно описана схема эксперимента на двух сплитах и как считать t статистику

Про бутстрепинг можно почитать [тут](#) и [тут](#)