**Анализ результатов АБ-теста**

В вопросе 6: "Для этого примените для каждого из срезов (по каждому из уникальных значений столбца browser) критерий Манна-Уитни между control и exp группами и сделайте поправку Холма-Бонферрони на множественную проверку с α=0.05."

Правильно ли я понимаю, делаем как и в задании про гены: чтобы сделать эту поправку надо указать method = 'holm' и alpha = 0.05/6 (т.е. alpha = обычный уровень значимости деленный на количество групп браузеров) ?

Ответ в этом вопросе у меня принимается, но хотелось бы разобраться, так как и с делением на 6 и без деления на 6, результат функции multipletests одинаковый.

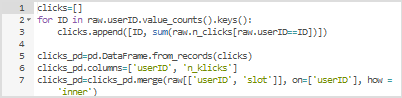
По вопросу 5: там предлагается выбрать ВСЕ правильные ответы, при этом заданы 4 взаимоисключающие альтернативы + 1 вопрос. Т.е. как бы не было рассчитано присвоение баллов, больше чем на 2 я ответить не могу... или я что-то неправильно понимаю?

попробуйте ответить меньше, чем на 2 :)

Здравствуйте. У меня вопрос по второму заданию. В данном случае метод бутстрапа должен долго производить вычисления или нет? Так как у меня ужасно долго все работает.

поставьте сразу размер псевдовыборки 500

В пункте 5 требуется в соответствие каждому пользователю поставить количество кликов, которые он сделал во всех браузерах. Я сделал это таким способом:



Проблема в первых трех строчках. В них я для каждого уникального значения userID в исходном датасете нахожу и суммирую все значения n\_clicks в строчках с совпадающим userID. То есть для каждого из 565к пользователей я выполняю поиск совпадающих значений среди 566к элементов. Время выполнения кода - 12 минут, никуда не годится.

Попробуйте через dict (один проход + сохранение в dict, сложность ~n)



Выше описанная функция **mannwhitneyu** возвращает одно значение **p\_value**, тогда как для определения поправки Холма-Бонферрони требуется передать в функцию массив из **p\_values**. Не понимаю, где взять этот массив?

Надо разбить выборку на подвыборки по различным браузерам, см. задание. Соответственно у вас будет число сравнений по числу различных браузеров. там шесть различных браузеров (колонка browser). И две группы пользователей (колонка slot). Вам нужно сравнить контрольную группу пользователей с экспериментальной для каждого вида браузера. Это шесть сравнений.