

## ДЗ АВ тесты Голубева

После первого прогона я поставила себе следующие вопросы о том, как устроена платформа, на которые искала ответы в следующем прогоне:

- 1) Сколько дней проходит до АВ тестирования у задач разного размера?
- 2) Сколько дней оптимально держать тест?
- 3) Есть ли разница между группами и сложностями?
- 4) АВ тесты, которые крутятся, требуют денег в нашей игре, почему их вообще надо останавливать?
- 5) Как устроена система начисления выгоды, какая выгода от задач разной сложности?

В итоге я выяснила, что задачи различаются по дням выполнения разработчиками:

Большой релиз ~8

Небольшое изменение функционала ~5

Мелкая доработка ~3

Вопрос сколько дней держать тест некорректен, длительность теста зависит от кол-ва участников, а кол-во участников набирается в тесты нелинейно.

Примерно 100 000 участников каждый день распределяется по активным тестам, этот факт и отвечает на вопрос, зачем вообще закрывать тесты которые не выпущены на прод. Чем меньше активных тестов тем быстрее наберется нужное кол-во участников.

А сколько вообще нужно участников? Исследовав как изменяются значения в ходе изменений на платформе я выяснила, что в самом начале конверсия в покупку приблизительно 5%, а наблюдаемые изменения около 0.2.

Воспользовавшись Калькулятором достоверности АВ-тестов. Получаем, что нам надо около 190к в каждую из групп.

До теста

### Рассчитать размер выборки

Проверьте, сколько людей должно участвовать в эксперименте для достоверных результатов. Подходит для тестирования open rate, click rate, конверсии в заказ.

Количество вариантов	Средний показатель	Ожидаемый абсолютный прирост
- 2 +	5 %	0.2 %

Достоверность 95%

Мощность 80%

Размер выборки, количество человек

Всего: **372,831**    В каждом варианте: **186,416**

Однако, видимо из-за особенностей сайта, как мне показалось (я записывала результат и следила за тенденцией) прирост 0.2 сохраняется только для первых нескольких тестов, а далее он увеличивается, причем так, что размер необходимой выборки фиксируется на уровне 70к – 80к для каждой группы.

Приведу примеры, когда тесты были значимы, успешны(на это указывает прибавка в прибыли)

Тестовая группа	Контрольная группа	Рассчитанное кол-во человек в 1 группе
6.32%	5.97%	75,871
7.66%	7.29 %	77,500
8.61%	8.13%	50,890

Я понимаю, что так нехорошо делать с точки зрения статистики, но в обычной жизни у нас есть, какая-то информация о том, что это за A/B тест и статистика о том, как в среднем ведут себя клиенты. Плюс, мне кажется, что в обычной жизни нет зависимости, что с увеличением конверсии в покупку увеличивается разница между контролем и тестом. По моему ощущению, она должна быть, как минимум, рандом, а как максимум, наоборот уменьшаться, потому что есть средние показатели конверсии, улучшить которые сложнее и сложнее.

Систему начисления выгоды я изучила поверхностно, прибавка за небольшое изменение функционала 140 – 340 руб пересекается по цене с мелкой доработкой 260 -270, а за большой релиз прибавляется 400 – 600 рублей. Скорее всего, это должно быть еще и связано с увеличением конверсии в каждом конкретном случае. Однако, наблюдение за тем как ведут себя тесты каждого из отделов, сделало бесполезным исследование влияния размера задачи на прибыль.

AB тесты от отдела Бэкэнд и Реклама с большой вероятностью оказываются успешными и выкатываются на прод. В AB тесты от отдела рекламы нет разницы между тестом и контролем. В AB тестах от рекомендаций и поиска почти всегда контроль оказывается статистически значимее, чем тест. (важно смотреть не только на p-value, но и на направление). Это открытие как-будто не про статистику и скорее всего не встретится в реальной жизни. (Хотя, наверное является неплохой практикой и в реальном мире собирать статистику успешности тестов и корректировать работу команд, выдающих плохой результат) Однако этот факт играет решающую роль в стратегии прохождения.

Предлагаю две стратегии прохождения.

**Первая больше с упором на статзнания и предположение, что платформа максимально повторяет реальный мир, как будто я ее прохожу впервые.**

Брать задачи равномерно из разных групп, исходя из предположения, что надо развивать все направления и несколько отдельных тестов параллельно в одном отделе это нехорошо, потому что не факт, что релиз обоих даст хороший результат.

Определить заранее пороговое  $p$ -value и кол-во дней, которые необходимо набрать для достоверного результата и пересчитать его со временем, учитывая, что показатель конверсии подрос. У меня нет готового математического ответа, как именно, но есть смысл останавливать тесты у которых наблюдается тенденция ухудшения (ну и вообще при принятии решений не забывать смотреть не только на  $p$ -value, но и на направление тренда) не дожидаясь всех пользователей, я это делала при достижении 50к. На данной платформе это необходимо, потому что пользователи распределяются по тестам и если их много, то результат ждать придется долго.

Распределять нагрузку по размеру задач, так чтобы АВ платформа не простаивала, для этого не стоит брать много тяжелых задач одновременно.

### Game Over

Your Results:

Дни в игре: 90

Итоговый доход: 522,350

Итоговая конверсия: 10.37%

Отправить результат

**Вторая именно чтобы получить максимальный выигрыш у платформы, учитывая ее особенности.**

С точки зрения статистики стратегия та же определить  $p$ -value и размер выборки.

Дальше тактика держать активными только тесты типа бэкэнд и реклама (полезные), а все остальные сразу закрывать, потому что мы заведомо знаем, что они неуспешные. Тактика брать с самого начала в задачи только бэкэнд и рекламу плохая, потому что сначала на платформе будет много активных полезных АВ тестов и пользователи по ним будут долго копиться, а потом будет долгий застой от других команд, потому что в бэклоге закончатся нужные категории.

Намного эффективнее брать несколько полезных АВ тестов и пару бесполезных, которые сразу останавливаются. Так в бэклоге всегда будут оставаться полезные АВ тесты и активных экспериментов будет мало, так что они быстро наберут необходимых пользователей.

Такая тактика позволяет набрать 766к

Game Over

Your Results:

Дни в игре: 90

Итоговый доход: 766,400

Итоговая конверсия: 14.17%

Голубева Ольга Витальевна

Отправить результат

P.S Спасибо большое за такое дз, было супер интересно и познавательно!