**План проведения А/В тестирования**

1. Гипотеза и подготовка к АВ тестированию:

Во-первых, необходимо определиться насколько сильные различия мы хотим задетектировать при помощи А/В тестирования.

В зависимости от величины желаемых различий подобрать размер выборки, необходимый для достижения адекватной мощности теста (вероятность отвергнуть нулевую гипотезу, когда она реально ошибочна).

В итоге нулевой гипотезой будет ситуация, что мы не имеем значимых различий, которую мы можем отклонить с p\_value 0,05 (в случае двусторонних критериев и расчета в одну сторону 0,025).

Альтернативной гипотезой будет ситуация, со значимыми различиями, величина которых будет не менее выбранной нами.

Далее необходимо разбить выбранную выборку на случайные, неизменные, детерменированные тестовые группы (допустим при помощи соленого хэширования) – для того, чтобы отслеживать эффект, была возможность многократно его повторять для каждого пользователя и при этом параллельно проводить несколько тестов по разным изменениям.

1. Метрики:

* конверсия в клики, в регистрацию, в оформление платной версии;
* время сессии, проведенное на сайте;
* показатель отказов – количество пользователей, покинувших сайт, ничего не используя;
* retention – показатель устойчивости когорт пользователей;
* количество новых пользователей;
* средний чек на пользователя.

Данные метрики всесторонне позволят изучить влияние на систему, даже если одни не изменятся или получат отрицательное влияние, другие могут вырасти – по этому, необходимо учитывать сбалансированное влияние.

В случае когда часть метрик улучшилась, а часть ухудшилась – необходимо сгруппировать их по влиянию и проанализировать, какой эффект для нас в приоритете и сильнее сказывается на доходе за интервал времени, далее делать выбор.

В случае, если нам необходимо оценить не только разницу средних значений метрик, но и какие-либо статистические параметры (например квантили) или ненормального распределения величин – необходимо использовать бутстреп, который позволит сгенерировать распределение необходимой метрики, распределенное нормально.

1. Срок:

Перед проведением АВ теста необходимо проверить однородность разбиения на группы с помощью проведения АА теста.

Продолжительность тестов зависит от редкости события и от размера различий, которые мы должны задетектировать, так как данные должны накопиться до уровня достаточного для принятия решений.

Важно не прерывать тестирование в процессе и не продлять намеренно, иначе может случиться так называемая проблема подглядывания и результаты будут дискредитированы.

1. Решение:

При помощи статистических тестов (т-тест, манна-уитни) и визуализации сравниваем группы, если появились значимые отличия, то можно явно сказать – хуже или лучше стала каждая из метрик.

В случае если при выбранном значении величины эффекта и мощности теста в 80% у нас не появились значимые отличия – это говорит о том, что различия даже если и есть, то их величина меньше нашего выбранного граничного условия, а значит не имеет экономического смысла в их принятии.