**План проведения А/В тестирования**

1. Гипотеза и подготовка к АВ тестированию:

Во-первых, необходимо определиться с силой эффекта, который мы хотим задетектировать при помощи А/В тестирования.

В зависимости от величины эффекта (процент изменения ключевых метрик) и общего трафика сайта нам необходимо подобрать количество выборки (допустим 10% от общего трафика, точное значение можно вычислить).

Стандартом индустрии считается мощность теста в 80%, то есть что мы в 80% случаев сможем получить значимые отличия при выбранной силе эффекта.

В итоге нулевой гипотезой будет ситуация, что мы не имеем значимых различий, которую мы можем отклонить с порогом вероятности 5%.

А альтернативной гипотезой будет ситуация, со значимыми различиями, которую мы можем подтвердить с тем большей вероятностью, чем выше будет эффект на метрики.

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, если нам необходимо оценить не только разницу средних значений метрик, но и какие-либо статистические параметры (например квантили) или в случае малых выборок или отсутствия конкретного 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. Оба теста лучше всего проводить минимум неделю, причем – не прерывать 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 все отлично.

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.