

Итоговая контрольная работа по блоку специализация

Урок 2. Итоговое задание

Задание 1.

Вам будут даны несколько кейсов для разных бизнесов с описанием бизнес задачи, которую необходимо решить:

Подведите результаты эксперимента в экселе по следующим данным: ab_stats.csv - Google Диск:

Стат значимо ли отличается ARPPU в двух группах? Какие рекомендации дадите менеджеру?

Решение:

Решение задачи в файле task_1.ipynb

Задание 2.

Мы хотим провести А/Б-тест для трех источников трафика.

Нынешняя конверсия равна 5%, мы ожидаем прирост в 0,2%. Уровень доверия 97% и уровень мощности 87%. Всего на наш продукт заходит 40 000 пользователей в месяц.

Задание 3.

За сколько дней мы сможем протестировать гипотезу? И что вы можете посоветовать по результатам подсчета?

Решение:

Воспользуемся онлайн-калькулятором <https://mindbox.ru/tools/ab-test-calculator/>

Получим следующие результаты:

mindbox

Продукты ▾Материалы ▾КлиентыЦены

Заявка

Калькулятор достоверности АВ-тестов

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

− 3 +

5 %

0.2 %

Достоверность

Мощность

97%

87%

Размер выборки, количество человек

Всего:В каждом варианте:

1,522,706507,569

Количество пользователей в месяц (человек):	40 000
Количество дней в месяце (дней):	30
Количество пользователей в день (человек):	1 333
Конверсия:	5%
Ожидаемый прирост:	0.2%
Уровень доверия:	97%
Уровень мощности:	87%
Размер выборки:	1 522 706
Источники трафика:	3
Количество человек в каждом варианте:	507 569
Количество дней для эксперимента при трафике в 1333 человек в день:	1 142

Расчет показал, что для проведения эксперимента необходимо 1142 дня, что чрезвычайно долго. Для уменьшения размера выборки (количества дней для проведения эксперимента) необходимо уменьшить уровень достоверности и мощности в пределах дефолтных настроек и увеличить абсолютный прирост, так как небольшой прирост в 0,2% сложнее обнаружить для этого нужна огромная выборка:

mindbox.ru

Калькулятор достоверности А/В-тестирования — Mindbox

mindbox

Продукты ▾Материалы ▾КлиентыЦены

Заявка

Калькулятор достоверности АВ-тестов

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

− 3 +

5 %

1.2 %

Достоверность

Мощность

95%

80%

Размер выборки, количество человек

Всего:

В каждом варианте:

32,121

10,707

Задание 3.1.

Допустим в задаче нет проблемы с количеством посетителей на сайт, тогда подведите результаты тестирования, если у нас следующие результаты по количеству конверсии:

- 1) 25
- 2) 30 000
- 3) 32 000

Решение:

Третий вариант тестирования лучше первого и второго, а второй - лучше первого. Примем первый вариант за контрольный, тогда второй вариант превосходит его на 20%, а третий - на 28%, что очень существенно. Разница между вторым и третьим вариантами - 6,67%. Для точного анализа надо знать, какое количество пользователей посетило сайт в каждой из трёх групп, чтобы сравнить конверсии уже в процентах от числа пользователей.

Если первая контрольная группа была с результатом 5% конверсии, значит число пользователей было 500 000. Тогда при таком же объеме выборок в остальных двух группах получим конверсии 6% и 6,4%, что гораздо выше ожидаемых в начале эксперимента показателей, а третья группа существенно лучше второй (на 0,4%) и первой (на 1,4%). В таком случае, имеет смысл переориентировать продукт полностью на третий вариант.

Задание 4.

Вы решили сравнивать метрику CPA в двух группах. Размер выборки - 2350 элементов в каждой группе.

Для проверки нормальности распределения на выборке в 2350 наблюдений применили, критерий Шапиро-Уилка и получили p-value, равный 0.00002, alpha = 5%.

Какой бы вывод мы могли сделать в данном случае?

В этом случае какой статистический критерий для проверки первоначальной гипотезы тут лучше всего подойдет и почему?

Решение:

Так как $p\text{-value}=0.00002$, что меньше $\alpha=0.05$, следовательно, данные распределены не нормально. Мы отвергаем нулевую гипотезу (H_0) на уровне значимости $\alpha=5\%$. Изменения, внесенные в метрику CPA, дали статистически значимый эффект. Рекомендуется использовать новую версию.

Метрика CPA (CPA — Cost per Action) является количественной. Так как данные распределены не нормально, поэтому для проверки первоначальной гипотезы при сравнении двух групп необходимо использовать критерий Манна-Уитни.

Задание 5.

Мы провели АБ-тест на увеличение average timespent per user. По итогам тестирования мы получили следующие данные.

Является ли результат статистически значимым с уровнем доверия 80%? Какую версию мы выкатим на продакшн?

А) Средняя - 360, отклонение - 40, количество – 9802

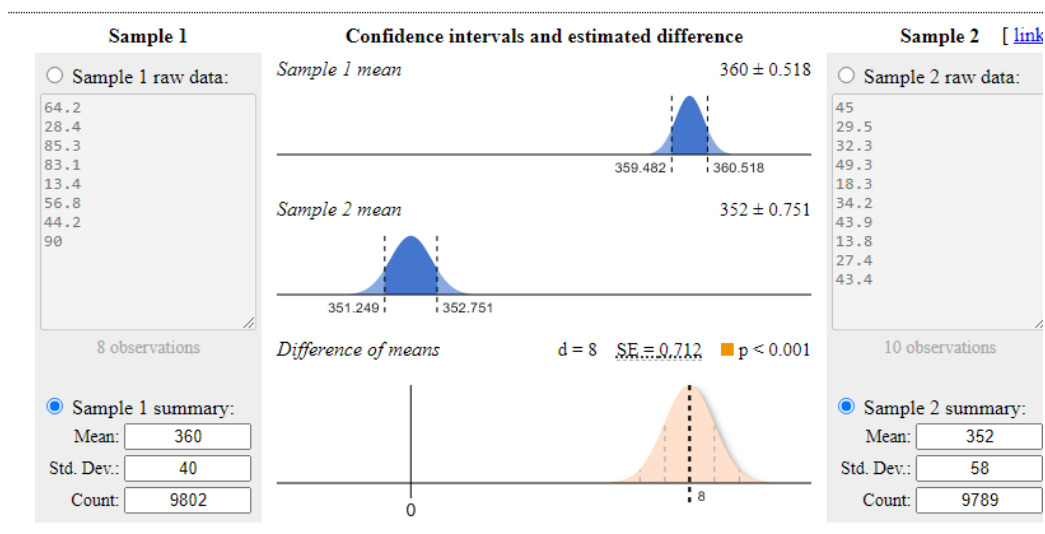
В) Средняя - 352, отклонение - 58, количество – 9789

Решение:

Воспользуемся онлайн-калькулятором <https://www.evanmiller.org/ab-testing/t-test.html>

Получим следующие результаты:

Question: Does the average value differ across two groups?



По графику видим, что вариант В показал себя хуже. В группе А среднее значение на 8 больше, чем в группе В и эта разница статистически значима при 80%-м уровне доверия.

Доверительные интервалы совсем не пересекаются, значит в продакшн выкатываем версию А.

Задание 6.

Создайте техническую архитектуру проекта по аб тестированию продукта он-лайн кинотеатра с учетом кросс-девайс аналитики по следующей гипотезе:

Если договориться с банком о 99% кэшбэке на подписку первого месяца, то это повысит конверсию в подписку на 30%, благодаря упрощенному принятию решения со стороны пользователя.

На схеме необходимо отобразить:

- 1) Управленческий процесс по договоренностям с внешними партнерами
- 2) Архитектуру данных с указанием систем, из которых будем скачивать данные
- 3) Внутриккомандное взаимодействие

Решение:

1. Управленческий процесс по договоренностям с банком:

- Обсуждение с банком условий о предоставлении 99% кэшбэка на подписку первого месяца.
- Согласование затрат на компенсацию 99%-го кэшбэка с финансовым отделом.
- Обсуждение с юристами компании, отделом продаж и маркетинга условий договоренности.

- Составление договора с условиями 99%-го кэшбэка и его подписка сторонами.
- Внесение информации о договоре и условиях кэшбэка в систему управления контентом.

2. Архитектура:

- Система управления подписками и контроля платежей.
- CRM система - данные о пользователях и их поведении.
- API банка - для получения информации о кэшбэке и транзакциях пользователей.
- Системы для сбора и хранения данных (например, Яндекс Метрика и др. аналоги).
- Система аналитики для сбора данных о конверсии в подписку и принятии решений (Яндекс Метрика, Excel, Power BI, Python).

Если наш продукт уже запущен и работает, то вышеуказанные системы будут существовать. С них уже собираем данные для анализа и проверки гипотезы.

3. Внутри командное взаимодействие:

- Менеджер по продукту/проекту - принимает решения на основе анализа данных от команды аналитики и информации о договоренности с банком о 99% кэшбэка, определяет экспериментальные группы и продолжение или прекращения кэшбэка на подписку первого месяца в зависимости от полученных результатов.
- UX/UI-дизайнеры - разработка дизайна.
- Аналитики - обеспечивают сбор данных о конверсии в подписку и анализируют их для оценки эффективности кэшбэка, а также предоставляют рекомендации по дальнейшим действиям на основе полученных результатов.
- Маркетологи - взаимодействуют с командой по аналитике для определения метрик, которые необходимо собирать и анализировать для оценки эффективности кэшбэка на подписку первого месяца.
- Разработчики - интегрируют систему управления контентом с системой управления клиентскими данными и банковской системой обработки транзакции, чтобы обеспечить автоматическое применение кэшбэка к подписке пользователей.