

Описание бизнес кейса

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

1. 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?

```
In [2]: import pandas as pd
import numpy as np
import scipy.stats as stats
data_path = "ab_stats.csv"
data = pd.read_csv(data_path)
print("Размеры датасета", data.shape)
data.head()
```

Размеры датасета (23652, 5)

```
Out[2]:
```

	revenue	num_purchases	purchase	ab_group	av_site visit
0	0.0	0	0	A	9.040174
1	0.0	0	0	A	4.811628
2	0.0	0	0	A	7.342623
3	0.0	0	0	A	7.744581
4	0.0	0	0	A	10.511814

```
In [3]: # Создаем столбец, в котором вычисляем ARPPU
data['ARPPU'] = np.where(data['num_purchases'] > 0, data['revenue'] / data['num_pur']
data.head()
```

```
Out[3]:
```

	revenue	num_purchases	purchase	ab_group	av_site visit	ARPPU
0	0.0	0	0	A	9.040174	0.0
1	0.0	0	0	A	4.811628	0.0
2	0.0	0	0	A	7.342623	0.0
3	0.0	0	0	A	7.744581	0.0
4	0.0	0	0	A	10.511814	0.0

```
In [4]: dataA = data.drop(data[data['ab_group'] == 'A'].index) # Таблица для A группы
dataB = data.drop(data[data['ab_group'] == 'B'].index) # Таблица для B группы
choiceA = dataA['ARPPU'] # выборка ARPPU для группы A
choiceB = dataB['ARPPU'] # выборка ARPPU для группы B
```

```
In [5]: # Выполним t-тест для проверки значимости разницы ARPPU в группах A и B
t_stat, p_value = stats.ttest_ind(choiceA, choiceB)
print(f'stats = {t_stat:.3f}, p_value = {p_value:.3f}')
if p_value < 0.05:
    print(f'Результат статистически значим')
else:
    print(f'Результат не является статистически значимым')
```

stats = -0.525, p_value = 0.599

Результат не является статистически значимым

Так как ARPPU в двух группах статистически значимо не отличаются, данная гипотеза не работает.

Описание бизнес кейса

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

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

- 25 000
- 30 000
- 32 000

- Используем Калькулятор достоверности АБ-тестов на сайте mindbox.ru \

mindbox

Продукты ▾ Материалы ▾ Клиенты Цены

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

— 3 +

Средний показатель

5 %

Ожидаемый абсолютный прирост

.2 %

Размер выборки, количество человек

Всего: 1,522,706

В каждом варианте: 507,569

Достоверность

97%

Мощность

87%

Как пользоваться калькулятором ▾

Как мы видим из результатов выборка должна быть - 1,522,706 человек. Если в месяц заходит 40,000 пользователей, то нам понадобится примерно 1142 дня (это более 3-ех лет). Следовательно, от этого теста лучше отказаться, либо же менять какие-то параметры.

3.1 Используем Калькулятор достоверности АВ-тестов на сайте mindbox.ru

mindbox

Продукты ▾ Материалы ▾ Клиенты Цены

Заявка

Как пользоваться калькулятором ▾

После теста

Подвести итоги теста

Проверьте, какой из вариантов лучше и можно ли считать результаты значимыми.

Количество вариантов

— 3 +

	Число конверсий	Размер выборки	Конверсия	Доверительный интервал
Вариант А	25000	508000	4.9 %	4.8 – 5.0 %
Вариант В	30000	508000	5.9 %	5.8 – 6.0 %
Вариант С	32000	508000	6.3 %	6.2 – 6.4 %

Достоверность

97%

Как пользоваться калькулятором ^

Вывод

Вариант В лучше варианта А

Вариант С лучше варианта А, В

Как мы видим из результатов вариант В лучше варианта А. И вариант С лучше вариантов А и В. Следовательно, разница в конверсиях

статистически значимы и вариант надо отдать предпочтение варианту С (с конверсией 32000).

Описание бизнес кейса

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? \

Нулевая гипотеза в тесте Шапиро-Уилка заключается в том, что данные имеют нормальное распределение. Поскольку значение p-value (0.00002) значительно меньше уровня значимости (alpha = 0.05), мы отвергаем нулевую гипотезу о нормальности распределения и не можем использовать t-test или z-test. Следовательно, мы должны использовать Критерий Манна—Уитни (2 группы, независимые выборки, не нормальное распределение)

Описание бизнес кейса

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

```
In [13]: import math  
import scipy.stats as stats
```

Чтобы определить, является ли результат статистически значимым с уровнем доверия 80%, нам нужно провести тест на значимость разницы между двумя средними значениями. В данном случае мы будем использовать t-тест для независимых выборок. Вычислим t-статистику и сравним ее с t-критическим.

```
In [15]: m1 = 360  
sd1 = 40  
n1 = 9802  
m2 = 352  
sd2 = 58  
n2 = 9789  
alfa = 1 - 0.8  
df = n1 + n2 - 2
```

```
In [16]: t = (m1 - m2) / math.sqrt(sd1**2 / n1 + sd2**2 / n2)  
t_kr = stats.t.ppf(1 - alfa / 2, df)  
print(f't-статистика = {t:.2f}')  
print(f'Критерий Стюдента для alfa = {alfa:.1f} и {df} степеней свободы = {t_kr:.2f}')  
if t > t_kr:  
    print(f'Результат статистически значим с уровнем доверия 80%')  
else:  
    print(f'Результат не является статистически значимым с уровнем доверия 80%')
```

t-статистика = 11.24

Критерий Стюдента для alfa = 0.2 и 19589 степеней свободы = 1.28

Результат статистически значим с уровнем доверия 80%

Поскольку средняя для группы А (360) больше средней для группы В (352) и результат статистически значим, мы можем сделать вывод, что версия А (с средней 360) будет выкатана на продакшн.