

Skillbox

# Урок 1

## Как работает сервис

Специалист



Заказы



## Как работает сервис



## Как работает сервис






# Постановка АБ-теста

## А

### Бытовая техника (Анастасия Сергеевна)





Заказ №17888302

Дата создания	только что	<a href="#">Пожаловаться на заказ</a>
Имя кл.	Анастасия Сергеевна	
Услуга	 бытовая техника	
Ставка	 2000 руб. / усл.	
Район	 Тимирязевская	
Формат	сегодня	
Объект	подключить стиральную машину	
Цена заявки	70 руб. <a href="#">Это дорого</a>	

## Б

### Бытовая техника (Анастасия Сергеевна)

Заказ №17888302

Дата создания	только что	<a href="#">Пожаловаться на заказ</a>
Клиент	 Анастасия Сергеевна	
Телефон	подтверждён	
Услуга	 бытовая техника	
Ставка	 2000 руб. / усл.	
Район	 Тимирязевская	
Формат	сегодня	
Объект	подключить стиральную машину	
Цена заявки	70 руб. <a href="#">Это дорого</a>	

# Поиск метрики доверия специалистов к сервису

Доля специалистов, отправивших отклики

# Поиск метрики доверия специалистов к сервису

Откликнувшиеся специалисты

---

Все специалисты

Доля специалистов, отправивших отклики

## Поиск метрики доверия специалистов к сервису

Откликнувшиеся специалисты


---

Все специалисты

Доля специалистов, отправивших отклики



## Поиск метрики доверия специалистов к сервису

- Доля специалистов, отправивших отклики 
- Среднее число откликов на специалистов

# Поиск метрики доверия специалистов к сервису

$$\sum \text{Число откликов}(1) + \text{Число откликов}(2) + \dots$$

---

Все специалисты

среднее число откликов на специалистов

## Поиск метрики доверия специалистов к сервису



$\sum \text{Число откликов}(1) + \text{Число откликов}(2) + \dots$

---

Все специалисты

среднее количество откликов на специалистов

## Поиск метрики доверия специалистов к сервису

- Доля специалистов, отправивших отклики 
- Среднее число откликов на специалистов 
- Доля заказов, на которое специалист откликнулся, относительно всех просмотренных заказов

# Поиск метрики доверия специалистов к сервису




Все отклики

---

Все просмотры

конверсия из просмотра в отклик

## Поиск метрики доверия специалистов к сервису

- Доля специалистов, отправивших отклики 
- Среднее число откликов на специалистов 
- Доля заказов, на которое специалист откликнулся, относительно всех просмотренных заказов, или **конверсия из просмотра в отклик.** 

## **Гипотеза об конверсии из просмотра в отклик**

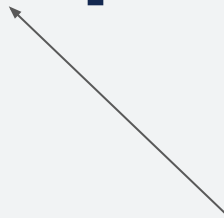
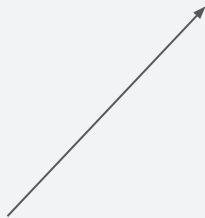
Специалисты будут отправлять больше откликов, если будут видеть в заказах аватарки и отметку подтвержденного номера - увеличится конверсия из просмотра в отклик.

**Конверсия будет базироваться на числе просмотров.  
На их основе будем считать длительность эксперимента.**

# Длительность эксперимента

Скорость накопления  
данных/трафик

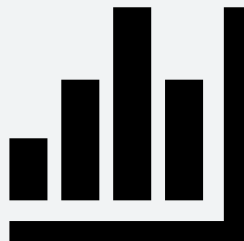
Размер выборки





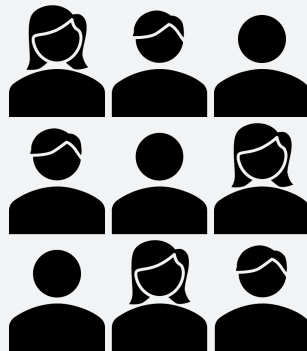
# Длительность эксперимента

Скорость накопления  
данных/трафик



10 000 просмотров в день

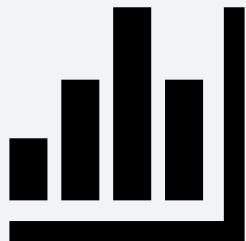
Размер выборки



?

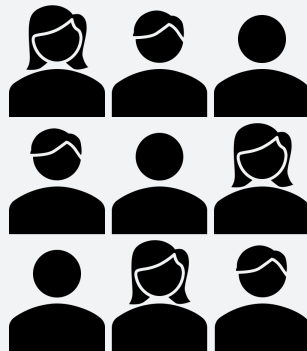
# Длительность эксперимента

Скорость накопления  
данных/трафик





10 000 просмотров в день

Размер выборки









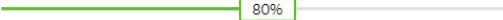
# Расчет размера выборки

Калькулятор достоверности A/B-тестирования

Что тестируем	Значение показателей	Размер выборки (чел.)
Показатель, который хочу протестировать <div>Конверсия в заказы ▾</div>	Средняя Конверсия по истории <div>5,0 %</div>	Вариант А  29 827
Количество вариантов тестирования <div>— 2 +</div>	Ожидаемый прирост Конверсии (абсолютный) <div>0,5 %</div>	Вариант В  29 827
Достоверность <div><div></div><div>95%</div></div>	Мощность <div><div></div><div>80%</div></div>	





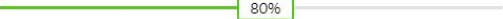
# Расчет размера выборки

Калькулятор достоверности A/B-тестирования

Что тестируем	Значение показателей	Размер выборки (чел.)
Показатель, который хочу протестировать	Средняя Конверсия по истории	Вариант А  29 827
Конверсия в заказы 	<input type="text" value="5,0"/> %	Вариант В  29 827
Количество вариантов тестирования	Ожидаемый прирост Конверсии (абсолютный)	
 2 	<input type="text" value="0,5"/> %	
Достоверность	Мощность	
 95%	 80%	

# Расчет размера выборки





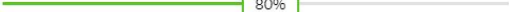
Калькулятор достоверности A/B-тестирования

Что тестируем	Значение показателей	Размер выборки (чел.)
Показатель, который хочу протестировать	Средняя Конверсия по истории	Вариант А  29 827
Конверсия в заказы 	<input type="text" value="5,0"/> %	Вариант В  29 827
Количество вариантов тестирования — 2 +	Ожидаемый прирост Конверсии (абсолютный)	
	<input type="text" value="0,5"/> %	
Достоверность	Мощность	
 95%	 80%	

Наше историческое значение конверсии из просмотра в отклик — 20%








# Расчет размера выборки

Калькулятор достоверности A/B-тестирования

Что тестируем	Значение показателей	Размер выборки (чел.)
Показатель, который хочу протестировать	Средняя Конверсия по истории	Вариант A  251
Конверсия в заказы 	<input type="text" value="20"/> %	Вариант B  251
Количество вариантов тестирования - 2 +	Ожидаемый прирост Конверсии (абсолютный)	
	<input type="text" value="10"/> %	
Достоверность	Мощность	
 95%	 80%	



# Расчет размера выборки

Калькулятор достоверности A/B-тестирования

Что тестируем	Значение показателей	Размер выборки (чел.)
Показатель, который хочу протестировать	Средняя Конверсия по истории	Вариант А  2 511 706
Конверсия в заказы 	<input type="text" value="20"/> %	Вариант В  2 511 706
Количество вариантов тестирования	Ожидаемый прирост Конверсии (абсолютный)	
 2 	<input type="text" value="0,1"/> %	
Достоверность	Мощность	
 95%	 80%	

# Расчет размера выборки

Калькулятор достоверности A/B-тестирования

Что тестируем	Значение показателей	Размер выборки (чел.)
Показатель, который хочу протестировать <div>Конверсия в заказы ▾</div>	Средняя Конверсия по истории <div>20 %</div>	Вариант А  2 511 706
Количество вариантов тестирования <div>— 2 +</div>	Ожидаемый прирост Конверсии (абсолютный) <div>0,1 %</div>	Вариант В  2 511 706
Достоверность <div><div></div>95%</div>	Мощность <div><div></div>80%</div>	

По расчетам конверсия изменяется в пределах 1%.

Попробуем взять 0,5% для надежности.



# Расчет размера выборки

Калькулятор достоверности A/B-тестирования

Что тестируем	Значение показателей	Размер выборки (чел.)
Показатель, который хочу протестировать	Средний Open Rate по истории	Вариант А
Open Rate ▾	20,0 %	100 468
Количество вариантов тестирования	Ожидаемый прирост Open Rate (абсолютный)	Вариант В
— 2 +	0,5 %	100 468
Достоверность	Мощность	
95%	80%	

## Длительность эксперимента

**100 000**  
просмотров



**2**  
варианта

---

**10 000**  
просмотров в день

## Длительность эксперимента

**100 000**  
просмотров



**2**  
варианта



**20**  
дней

---

**10 000**  
просмотров в день

## **А зачем проводить АБ-тест?**

Почему нельзя просто сравнить средние метрики по заказам с дополнительной информацией и по заказам без этого?

# Практика

Skillbox

# Урок 2

## **А зачем проводить АБ-тест?**

Заказы могут отличаться чем-то еще!

АБ-тест нужен, чтобы это учесть.

В АБ-тесте единственная разница между заказами — только в форме отображения.

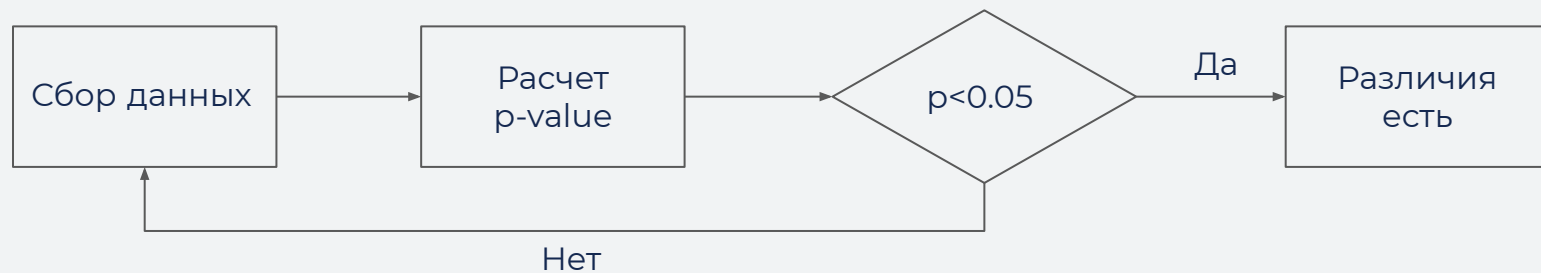
# Проблема подглядывания

- Периодически различие между вариантами теста может становиться статистически значимым, хотя различий нет
- Хочется принять решение при первых признаках  $p\text{-value} < 0.05$

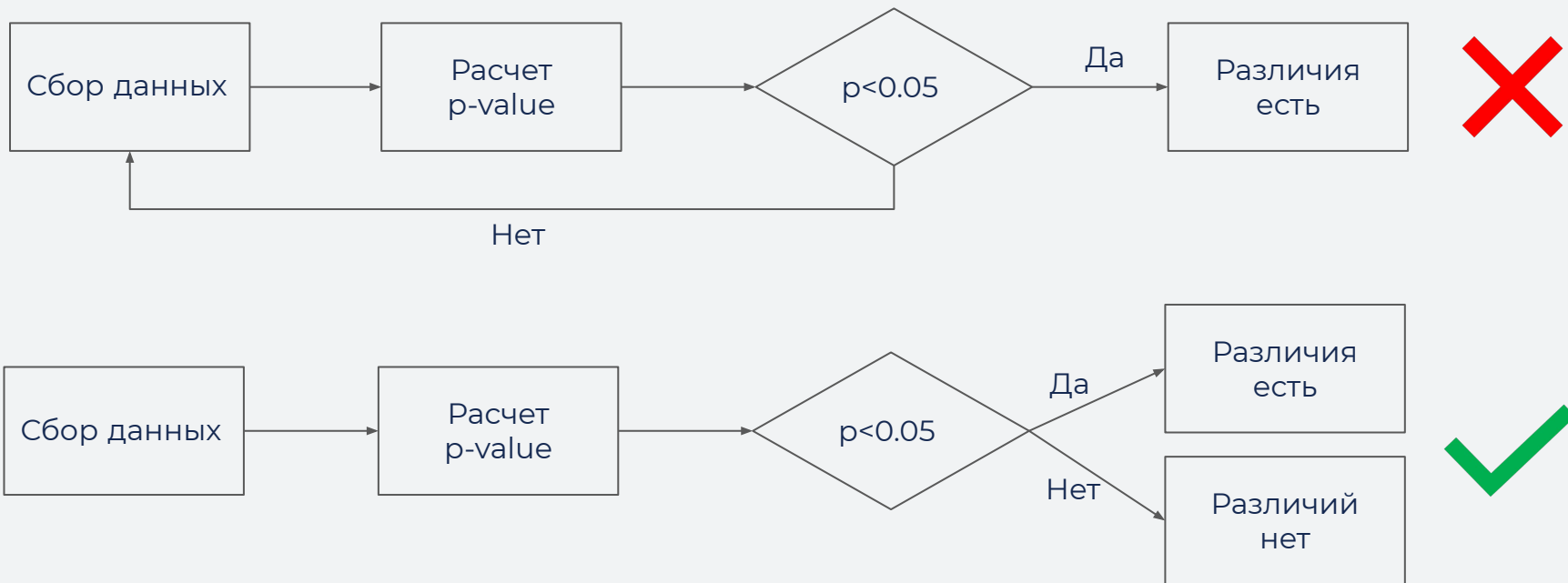




## Схема проведения АБ-теста



## Схема проведения АБ-теста



# Методы борьбы с подглядыванием

- Расчет выборки перед экспериментом

# Методы борьбы с подглядыванием

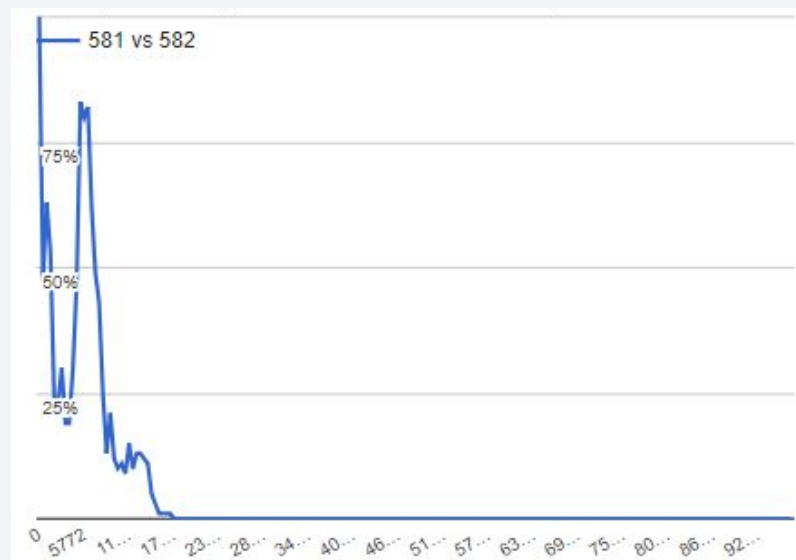
- Расчет выборки перед экспериментом
- “Умное” подглядывание за графиком p-value

# Примеры

Нет значимых различий



Есть значимые различия



# Построим график p-value по дням

Так выглядит наш датафрейм

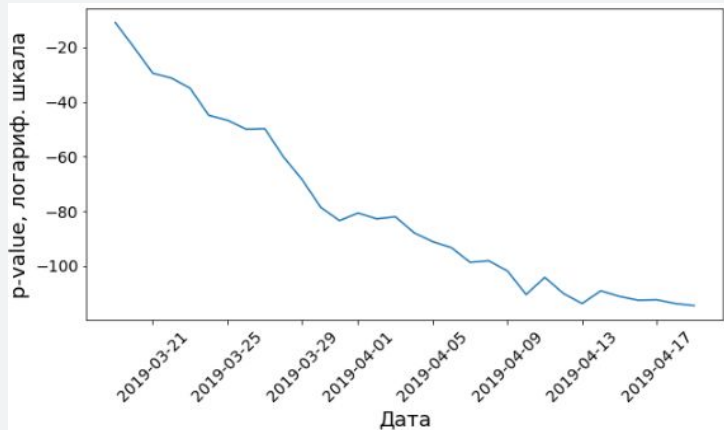
	case_id	spec_id	order_id	first_dt	e_dt	d_view
0	B	ffa2dd59dc131d92994c2f267c19251dbb836ecbb11eec...	NaN	NaT	NaT	NaT
1	B	e800fca7e8c2e643c56572d66933bfeaf2cc66a914328d...	NaN	NaT	NaT	NaT
2	B	c28ed784882c00fff28a0b0fdd434127bb5de905ca2318...	NaN	NaT	NaT	NaT
3	B	f651bd1ee3f165513b5096aa0e858529da005ae656563e...	NaN	NaT	NaT	NaT
4	B	002fa07fb4f6c54fcc8934855e7bd5fe2167f8526c23f6...	NaN	NaT	NaT	NaT
...	...	...	...	...	...	...
302359	A	b6255234f9ac0a8c610b2c332e34c6a9f9ae4d6e0d109d...	18252119.0	2019-04-01 16:18:43	2019-04-01 16:26:39	2019-04-01
302360	A	b6255234f9ac0a8c610b2c332e34c6a9f9ae4d6e0d109d...	18353035.0	2019-04-08 14:53:00	NaT	2019-04-08
302361	A	d90a88dc6b6428d9c13d12666d6955703a1e0a77947053...	NaN	NaT	NaT	NaT
302362	A	131eaa2b170fb675b5b37bc464aaa9fed5147b8e6dd333...	NaN	NaT	NaT	NaT
302363	A	40a30ca0c487ed1a424257e51f8bb0e33780f33b641c45...	NaN	NaT	NaT	NaT

Считаем кумулятивное число просмотров и откликов по вариантам, изменение конверсии по дням и ее стат.значимость с помощью z-test

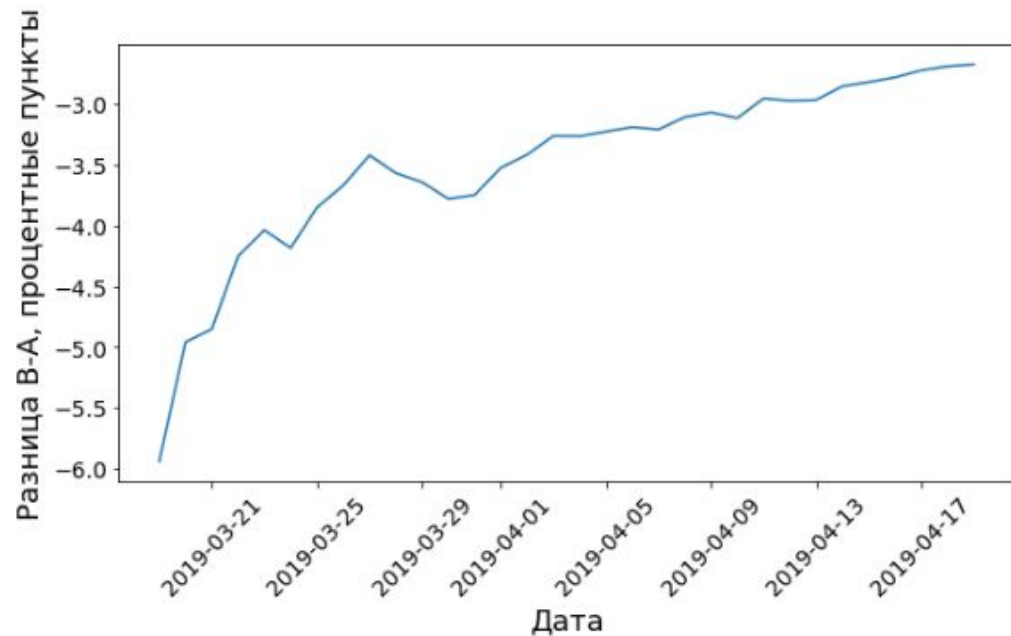
# Построим график p-value по дням

Считаем кумулятивное число просмотров и откликов по вариантам, изменение конверсии по дням и ее стат.значимость с помощью z-test

```
a_views = a_web.groupby(['case_id', 'd_view']).agg({'view': 'nunique', 'response': 'nunique'})  
c = a_views.groupby('case_id').cumsum()  
c['conversion'] = c.response / c.view
```



## Абсолютная разница в конверсии






# Практика

# Урок 3

## Почему конверсия падает?

$$\frac{\text{Все отклики}}{\text{Все просмотры}}$$


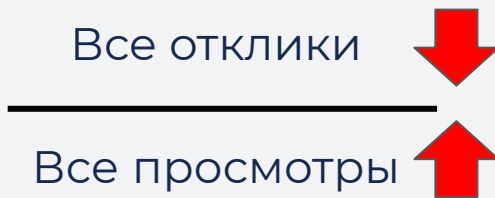
уменьшилось число откликов

## Почему конверсия падает?

$$\frac{\text{Все отклики}}{\text{Все просмотры} \uparrow}$$

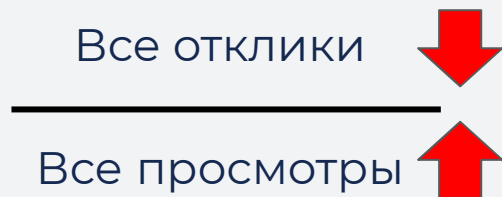
увеличилось число просмотров

## Почему конверсия падает?

$$\frac{\text{Все отклики}}{\text{Все просмотры}}$$


и то, и другое одновременно

## На что обратить внимание дальше?



среднее число откликов на специалиста

среднее число просмотров на специалиста

# Изменение средних в эксперименте

```
a_views_by_spec.groupby('case_id').view.describe()
```

	count	mean	std	min	25%	50%	75%	max
case_id								
A	56818.0	1.592594	7.271583	0.0	0.0	0.0	1.0	764.0
B	57512.0	1.762119	7.458915	0.0	0.0	0.0	1.0	363.0

```
a_views_by_spec.groupby('case_id').response.describe()
```

	count	mean	std	min	25%	50%	75%	max
case_id								
A	56818.0	0.323665	3.16577	0.0	0.0	0.0	0.0	563.0
B	57512.0	0.310961	2.06235	0.0	0.0	0.0	0.0	209.0

## Изменение средних в эксперименте

```
a_views_by_spec.groupby('case_id').view.describe()
```

	count	mean	std	min	25%	50%	75%	max
case_id								
A	56818.0	1.592594	7.271583	0.0	0.0	0.0	1.0	764.0
B	57512.0	1.762119	7.458915	0.0	0.0	0.0	1.0	363.0

```
a_views_by_spec.groupby('case_id').response.describe()
```

	count	mean	std	min	25%	50%	75%	max
case_id								
A	56818.0	0.323665	3.16577	0.0	0.0	0.0	0.0	563.0
B	57512.0	0.310961	2.06235	0.0	0.0	0.0	0.0	209.0



## Кто в хвосте?

```
a_views_by_spec[a_views_by_spec.view>500]
```

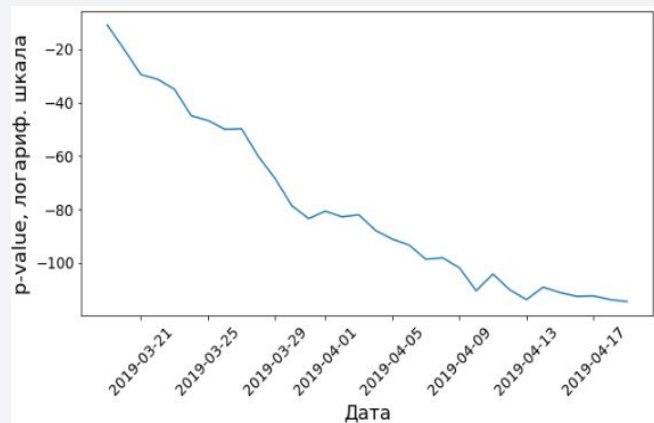
case_id	spec_id	view	response
A	e33a77a7a3af3609b838d77341c00856c76bf761ecea5b477e0cddfc	764	563

**Это бот!**

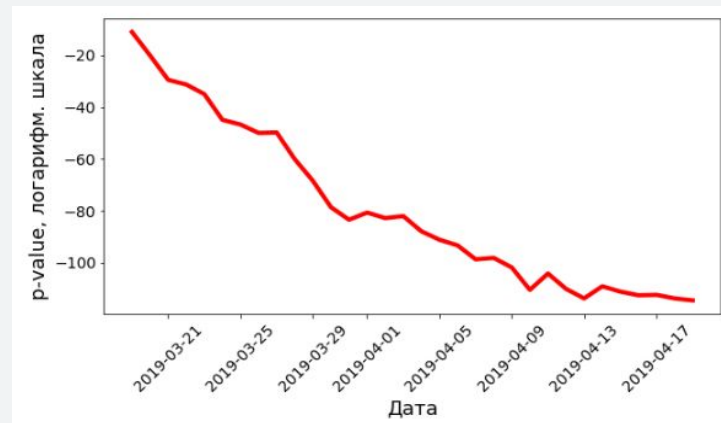


# Перепроверяем результаты

ДО:



ПОСЛЕ:



# Перепроверяем результаты

**ДО:**

**ПОСЛЕ:**

Вариант	Просмотры	Отклики	Конверсия
А	90 488	18 390	20,3%
Б	101 343	17 884	17,6%

Вариант	Просмотры	Отклики	Конверсия
А	89 724	17 827	19,9%
Б	101 343	17 884	17,6%

# Перепроверяем результаты

ДО:

ПОСЛЕ:

```
a_views_by_spec.groupby('case_id').view.describe()
```

	count	mean	std	min	25%	50%	75%	max
case_id								
A	56818.0	1.592594	7.271583	0.0	0.0	0.0	1.0	764.0
B	57512.0	1.762119	7.458915	0.0	0.0	0.0	1.0	363.0

```
a_views_by_spec.groupby('case_id').response.describe()
```

	count	mean	std	min	25%	50%	75%	max
case_id								
A	56818.0	0.323665	3.16577	0.0	0.0	0.0	0.0	563.0
B	57512.0	0.310961	2.06235	0.0	0.0	0.0	0.0	209.0

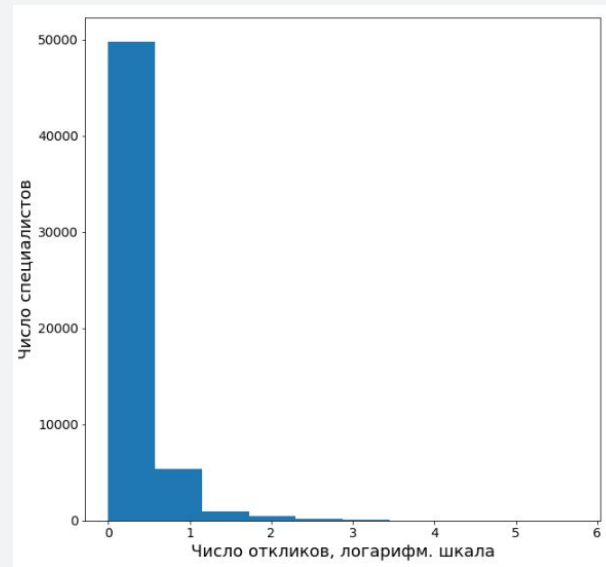
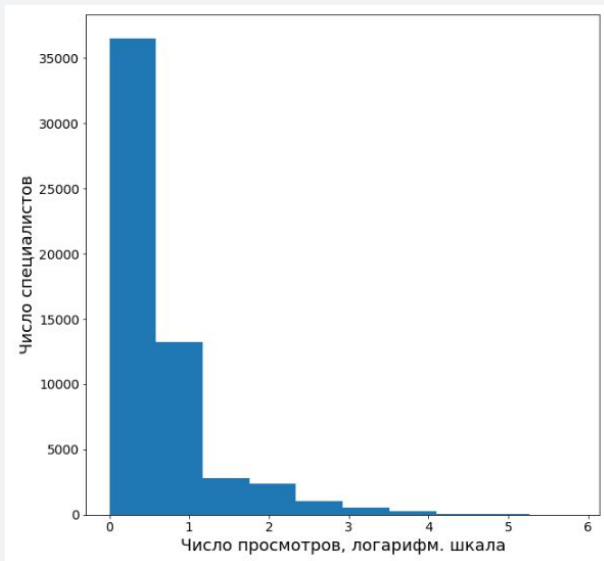
```
a_views_by_spec_clean.groupby('case_id').view.describe()
```

	count	mean	std	min	25%	50%	75%	max
case_id								
A	56817.0	1.579175	6.530392	0.0	0.0	0.0	1.0	346.0
B	57512.0	1.762119	7.458915	0.0	0.0	0.0	1.0	363.0

```
a_views_by_spec_clean.groupby('case_id').response.describe()
```

	count	mean	std	min	25%	50%	75%	max
case_id								
A	56817.0	0.313762	2.109437	0.0	0.0	0.0	0.0	313.0
B	57512.0	0.310961	2.062350	0.0	0.0	0.0	0.0	209.0

# Активность специалистов, как выглядят метрики?



# Используем бутстрап, проверяем, есть ли различия

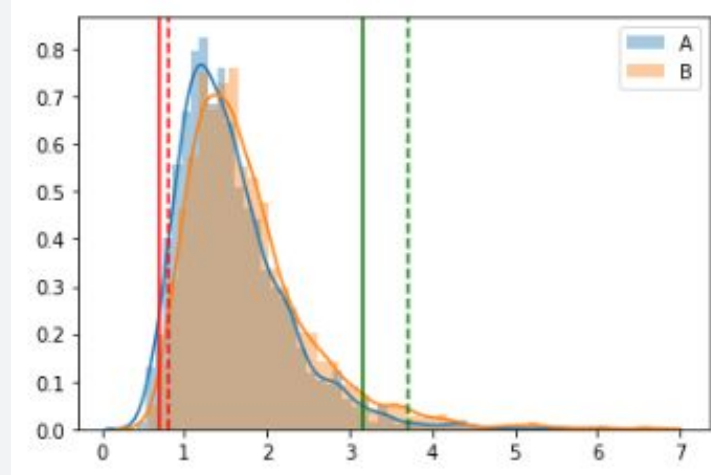
```
def bootstrap(arr,n=100,N_TRIAL = 1000,func = np.mean):
    res = []
    n = min(len(arr),n)
    # Инициализируем цикл попыток
    for _ in range(N_TRIAL):
        # Делаем подвыборку размера n
        if len(arr.shape)>1:
            ind = np.random.choice(np.arange(arr.shape[0]), size = (n,))
            subsample = arr.iloc[ind]
        else:
            subsample = np.random.choice(arr, size = (n,))
        # рассчитываем статистику
        stat = func(subsample)
        res.append(stat)
    return res # shape (N_TRIAL,)
```

```
samples_a = bootstrap(a_views_by_spec_clean.query("case_id=='A'")['view'],N_TRIAL = 3000)
samples_b = bootstrap(a_views_by_spec_clean.query("case_id=='B'")['view'],N_TRIAL = 3000)
```

# Визуализируем выборки и доверительные интервалы

```
import seaborn as sns
ax = sns.distplot(samples_a, label = 'A')
sns.distplot(samples_b, ax = ax, label = 'B')
ax.legend(['A', 'B'])
ax.axvline(np.percentile(samples_a, 2.5), linestyle = '--', color='r')
ax.axvline(np.percentile(samples_a, 97.5), linestyle = '--', color='g')

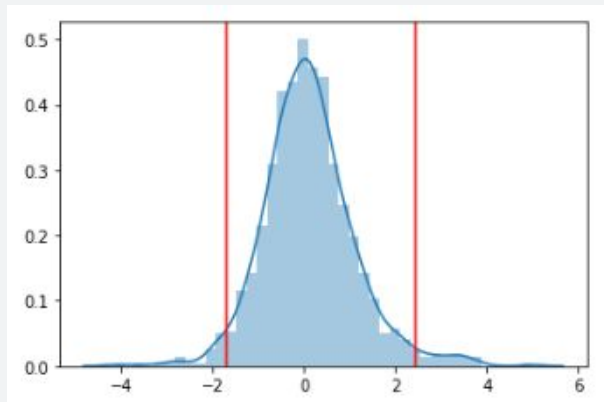
ax.axvline(np.percentile(samples_b, 2.5), linestyle = '--', color='r')
ax.axvline(np.percentile(samples_b, 97.5), linestyle = '--', color='g')
```



# Визуализируем разницу средних по вариантам

```
buf = []  
for i in range(1000):  
    a_sample = a_views_by_spec_clean.query("case_id=='A'")['view'].sample(100)  
    b_sample = a_views_by_spec_clean.query("case_id=='B'")['view'].sample(100)  
    buf.append(np.mean(b_sample)-np.mean(a_sample))
```

```
ax_buf = sns.distplot(buf)  
ax_buf.axvline(np.percentile(buf,2.5),linestyle = '-',color='r')  
ax_buf.axvline(np.percentile(buf,97.5),linestyle = '-',color='r')  
print('min =',np.percentile(buf,2.5),'max =',np.percentile(buf,97.5))
```





# Практика

# Урок 4

# Итоги теста

## Итоги теста

- **Гипотеза** “Специалисты будут отправлять больше откликов, если будут видеть в заказах аватарки и отметку подтвержденного номера - увеличится конверсия из просмотра в отклик”

## Итоги теста

- **Гипотеза** “Специалисты будут отправлять больше откликов, если будут видеть в заказах аватарки и отметку подтвержденного номера - увеличится конверсия из просмотра в отклик” **не подтвердилась.**

## Итоги теста

- **Гипотеза** “Специалисты будут отправлять больше откликов, если будут видеть в заказах аватарки и отметку подтвержденного номера - увеличится конверсия из просмотра в отклик” **не подтвердилась.**
- Наличие аватарки и подтвержденного номера ухудшает целевую метрику, которую мы сформулировали как доверие специалистов к заказам.

## Итоги теста

- **Гипотеза** “Специалисты будут отправлять больше откликов, если будут видеть в заказах аватарки и отметку подтвержденного номера - увеличится конверсия из просмотра в отклик” **не подтвердилась.**
- Наличие аватарки и подтвержденного номера ухудшает целевую метрику, которую мы сформулировали как доверие специалистов к заказам.
- Активность специалистов значимо не различается между вариантами.

# Разбор АБ-теста из практики

**А**

Йога

**Подходящая стоимость занятия**

☐ Пусть тренер предложит цену

**Б**

Йога

**Подходящая стоимость за час занятия**

☐ Пусть тренер предложит цену



## Длительность АБ-теста: входные данные

**2600**

Пользователей в день

**27%**

Конверсия

**~1.5%**

Изменение конверсии

## Длительность АБ-теста: расчет

**2600**

Пользователей в день

**27%**

Конверсия

**~1.5%**

Изменение конверсии

**10**

дней

## Длительность АБ-теста: расчет

**2600**

Пользователей в день

**27%**

Конверсия

**~1.2%**

Изменение конверсии

**16**

дней

## Изменение p-value в тесте



# Изменение конверсии по результатам теста

Вариант	Пользователи	Заказы	Конверсия
А	21 226	5612	26,4%
Б	20 965	5599	26,7%

## Изменение средней стоимости заказа

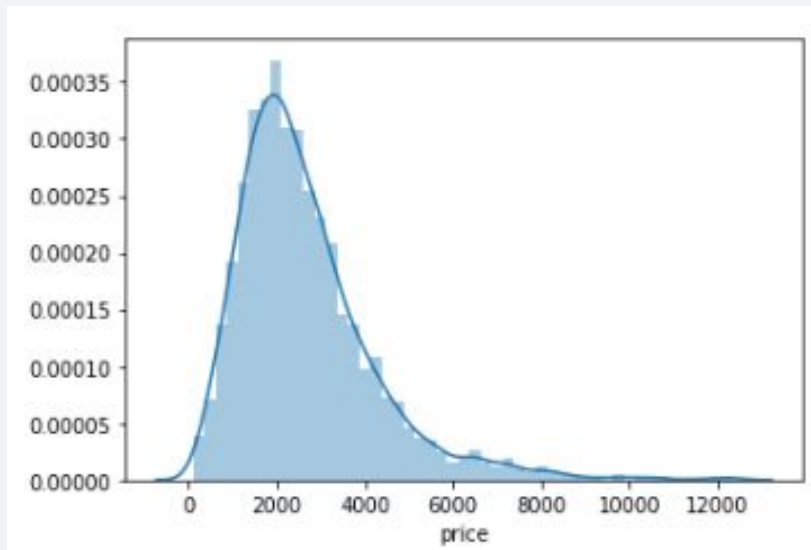
**2664 р.**

В варианте А

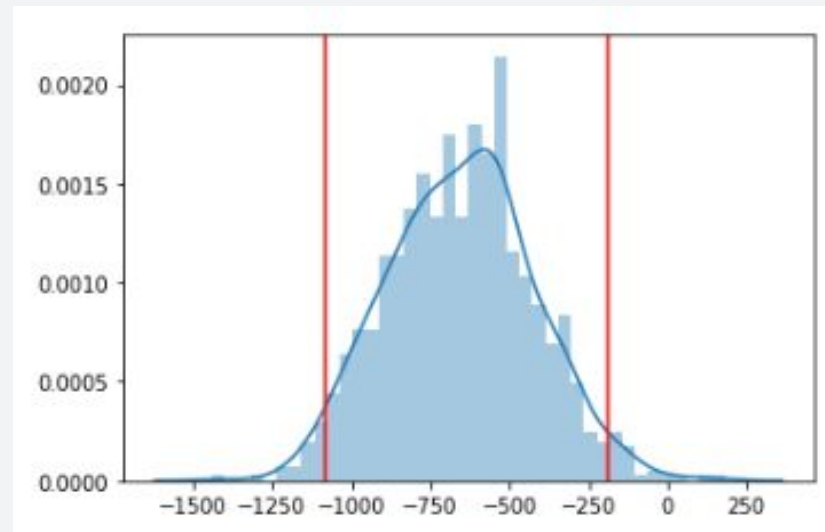
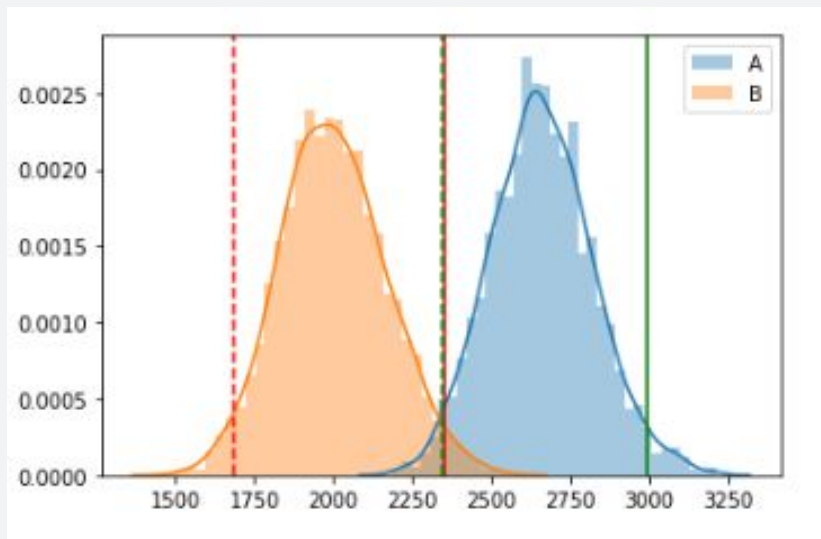
**2003 р.**

В варианте Б

# Распределение средней стоимости заказа



# Проверка значимости различий





## **Теперь мы знаем, что такое:**

- Постановка АБ-теста
- Расчет длительности теста и проблема подглядывания
- Расчеты значимости для средних и пропорций
- Работа с выбросами и очистка данных
- Корректные выводы по результатам анализа

# Практика