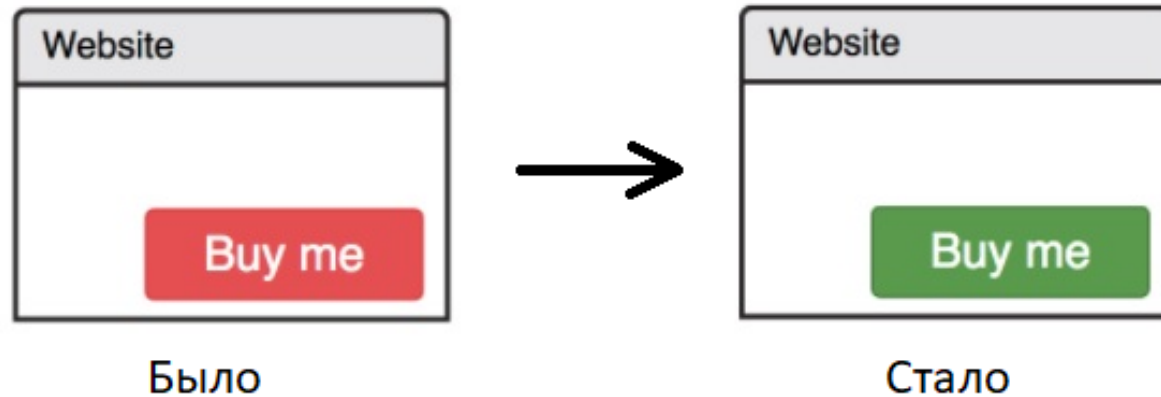


A/B-тестирование

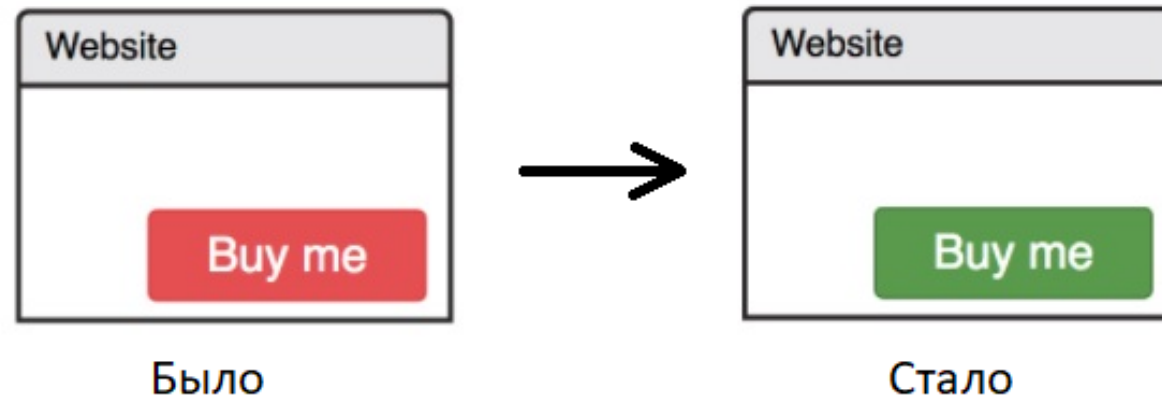
# Зачем нужны A/B-тесты

- Представим ситуацию, что пользователи активно используют наш ресурс. И в один прекрасный день мы *решили что-то поменять*, например, *поменять цвет кнопки*, отвечающей за покупку продукта.



# Зачем нужны А/В-тесты

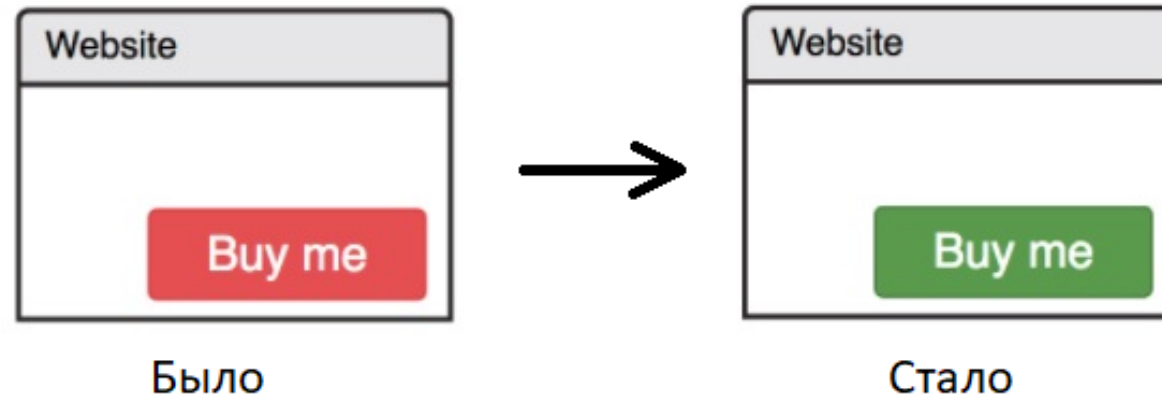
- Представим ситуацию, что пользователи активно используют наш ресурс. И в один прекрасный день мы *решили что-то поменять*, например, *поменять цвет кнопки*, отвечающей за покупку продукта.



- Наше решение — это *интуитивное предположение* о том, что пользователям ресурса приятнее видеть зеленую кнопку, и мы ожидаем увеличение числа кликов по кнопке.

# Зачем нужны А/В-тесты

- Представим ситуацию, что пользователи активно используют наш ресурс. И в один прекрасный день мы *решили что-то поменять*, например, *поменять цвет кнопки*, отвечающей за покупку продукта.



- Наше решение — это *интуитивное предположение* о том, что пользователям ресурса приятнее видеть зеленую кнопку, и мы ожидаем увеличение числа кликов по кнопке.
- Но **наше предположение не означает, что после внесения изменений мы получим желаемый эффект**. Для проверки таких гипотез мы и проводим А/В тесты.

# Что такое А/В-тестирование

- Делим пользователей на две группы (А и В):
  - Группа А – контрольная – ей показывается страница без изменений
  - Группа В – тестовая – ей показывается страница с внесенным изменением
- Эффект от изменений оценивается путем сравнения целевых метрик в группах



# Где используются А/В-тесты: реклама


**Infusionsoft**   
about a week ago

Boost email marketing and improve lead scoring with our automated solution. Save up to 40% off your first 6 months. Offer expires May 15, experience our live demo today.

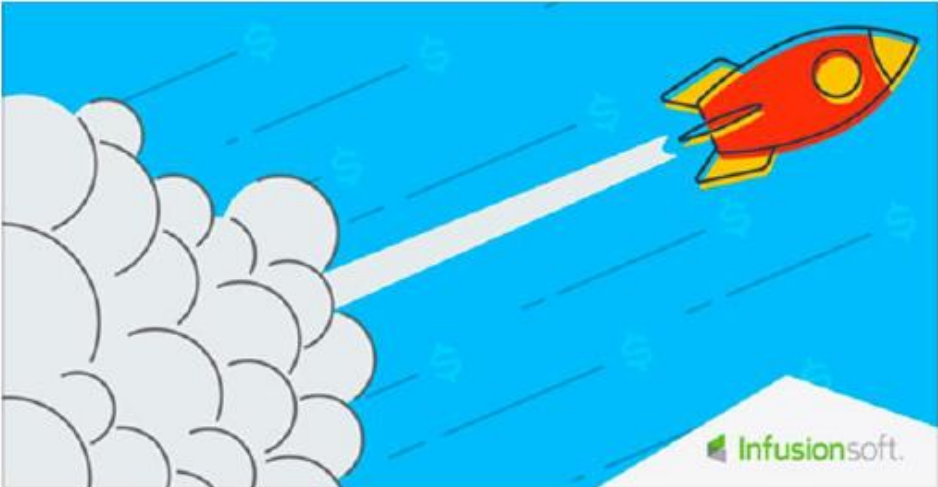


Personalized, automated marketing.  
125,000+ Users. "It's a game-changer in my business!"  
[WWW.INFUSIONSOFT.COM](http://WWW.INFUSIONSOFT.COM) [Request Time](#)

Like Comment Share

**Infusionsoft**   
about a month ago

Discover the business marketing tool helping 125,000 users achieve small business success.

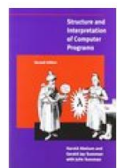


Do what you love. Automate the rest.  
"It's a game-changer in my business!"  
[WWW.INFUSIONSOFT.COM](http://WWW.INFUSIONSOFT.COM) [Learn More](#)

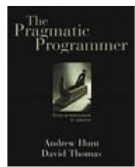
197 8 47

# Где используются А/В-тесты: интернет-магазин

## Frequently Bought Together



+



Total price: **\$83.09**

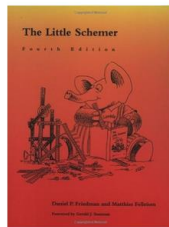
Add both to Cart

Add both to List

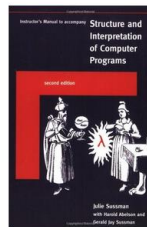
- ✓ **This item:** Structure and Interpretation of Computer Programs - 2nd Edition (MIT Electrical Engineering and... by Harold Abelson Paperback **\$50.50**
- ✓ **The Pragmatic Programmer: From Journeyman to Master** by Andrew Hunt Paperback **\$32.59**

## Customers Who Bought This Item Also Bought

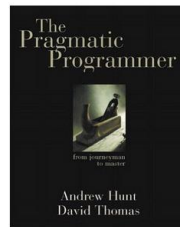
Page 1 of 13



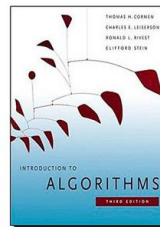
**The Little Schemer - 4th Edition**  
› Daniel P. Friedman  
★★★★☆ 64  
Paperback  
**\$36.00** ✓Prime



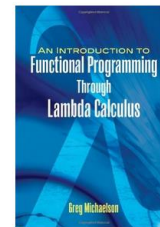
**Instructor's Manual t/a Structure and Interpretation of Computer Programs...**  
› Gerald Jay Sussman  
★★★★☆ 5  
Paperback  
**\$28.70** ✓Prime



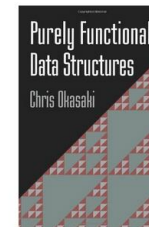
**The Pragmatic Programmer: From Journeyman to Master**  
› Andrew Hunt  
★★★★☆ 328  
Paperback  
**\$32.59** ✓Prime



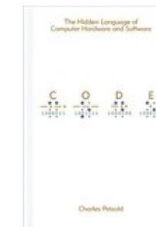
**Introduction to Algorithms, 3rd Edition (MIT Press)**  
› Thomas H. Cormen  
★★★★☆ 313  
#1 Best Seller in Computer Algorithms  
Hardcover  
**\$66.32** ✓Prime



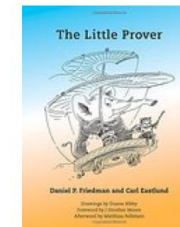
**An Introduction to Functional Programming Through Lambda...**  
› Greg Michaelson  
★★★★☆ 23  
Paperback  
**\$20.70** ✓Prime



**Purely Functional Data Structures**  
› Chris Okasaki  
★★★★☆ 19  
Paperback  
**\$40.74** ✓Prime



**Code: The Hidden Language of Computer Hardware and Software**  
› Charles Petzold  
★★★★☆ 334  
#1 Best Seller in Machine Theory  
Paperback  
**\$17.99** ✓Prime

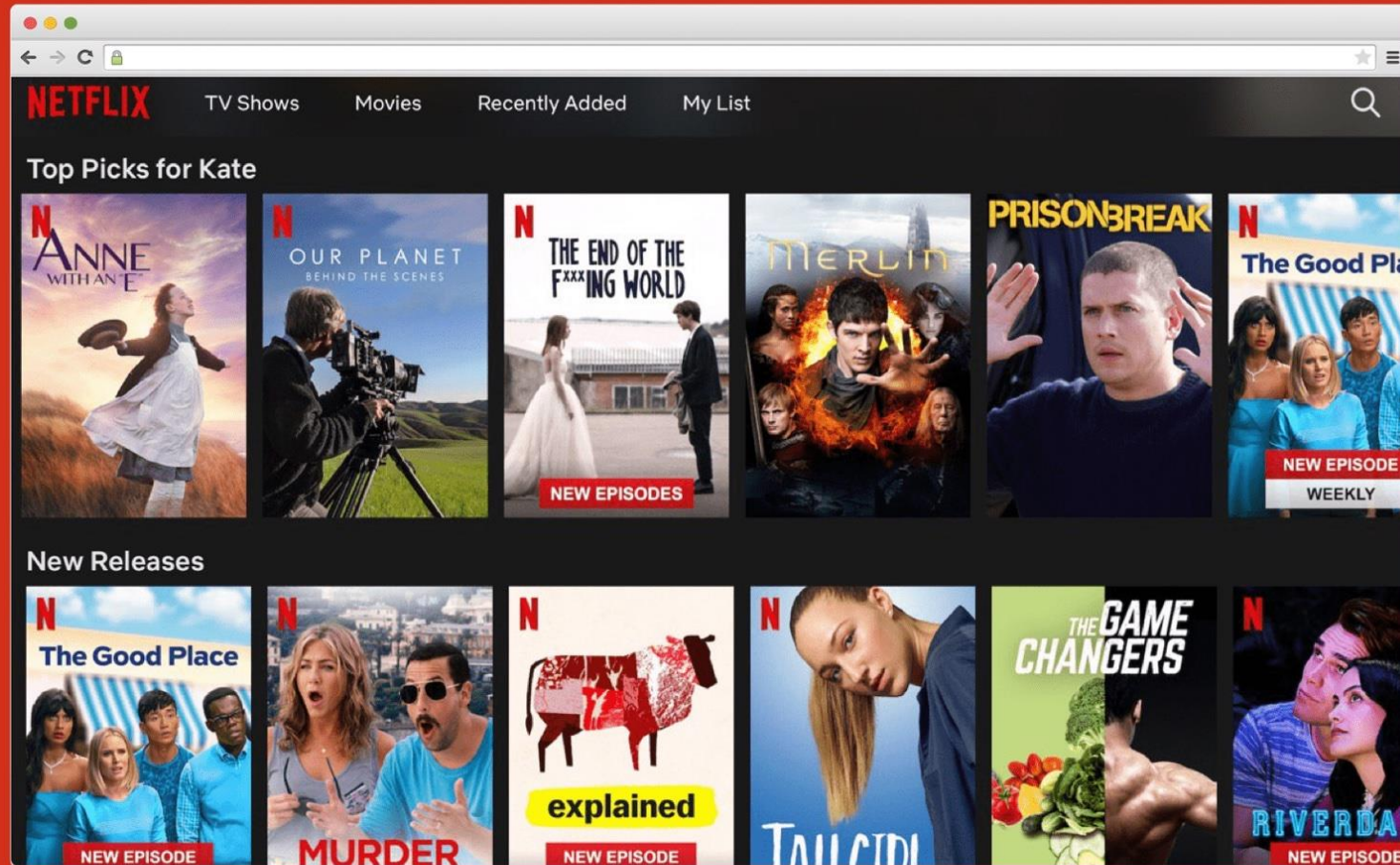


**The Little Prover (MIT Press)**  
Daniel P. Friedman  
★★★★☆ 4  
Paperback  
**\$31.78** ✓Prime



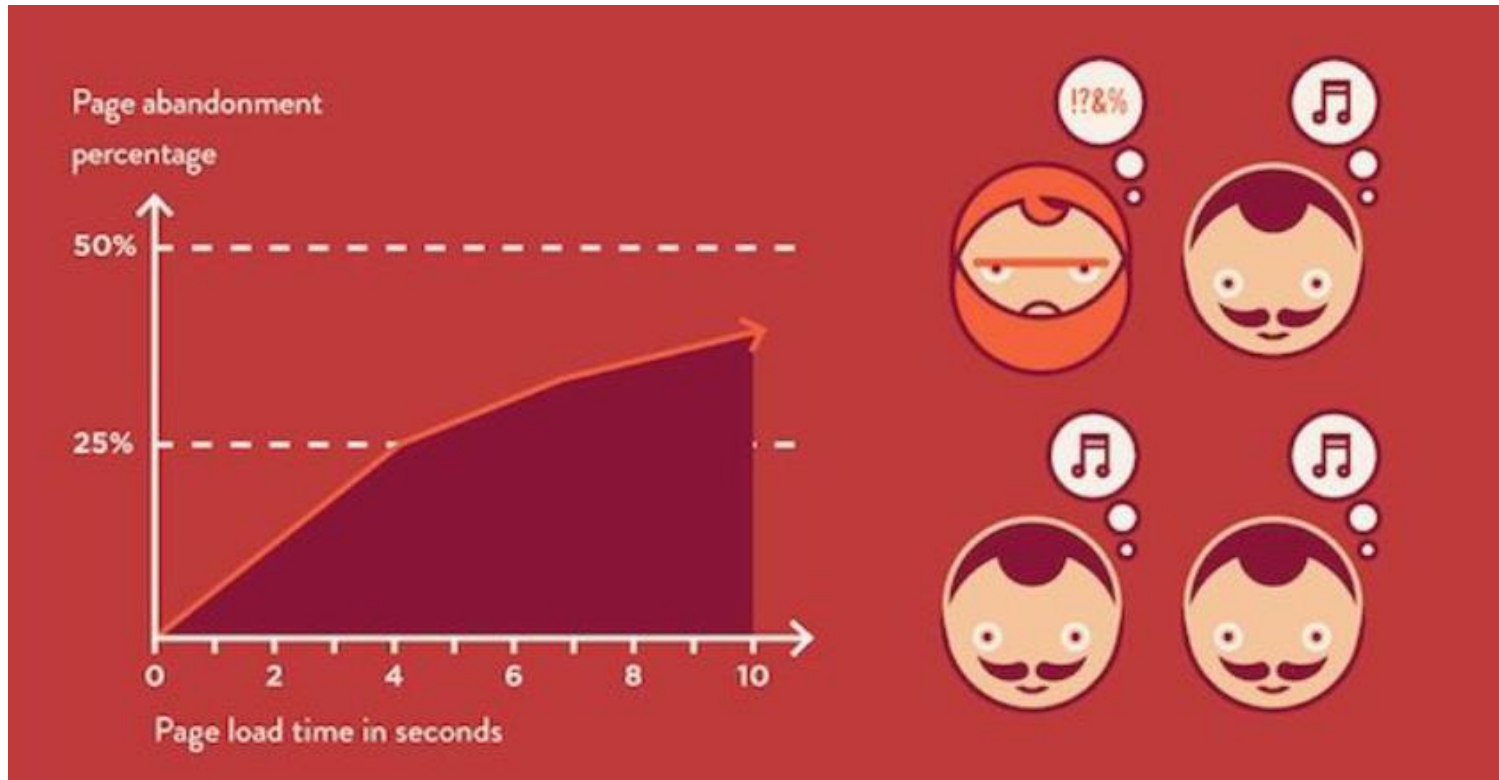


# Где используются A/B-тесты: Netflix





# Где используются А/В-тесты: алгоритмы



<https://www.fastcompany.com/1825005/how-one-second-could-cost-amazon-16-billion-sales>

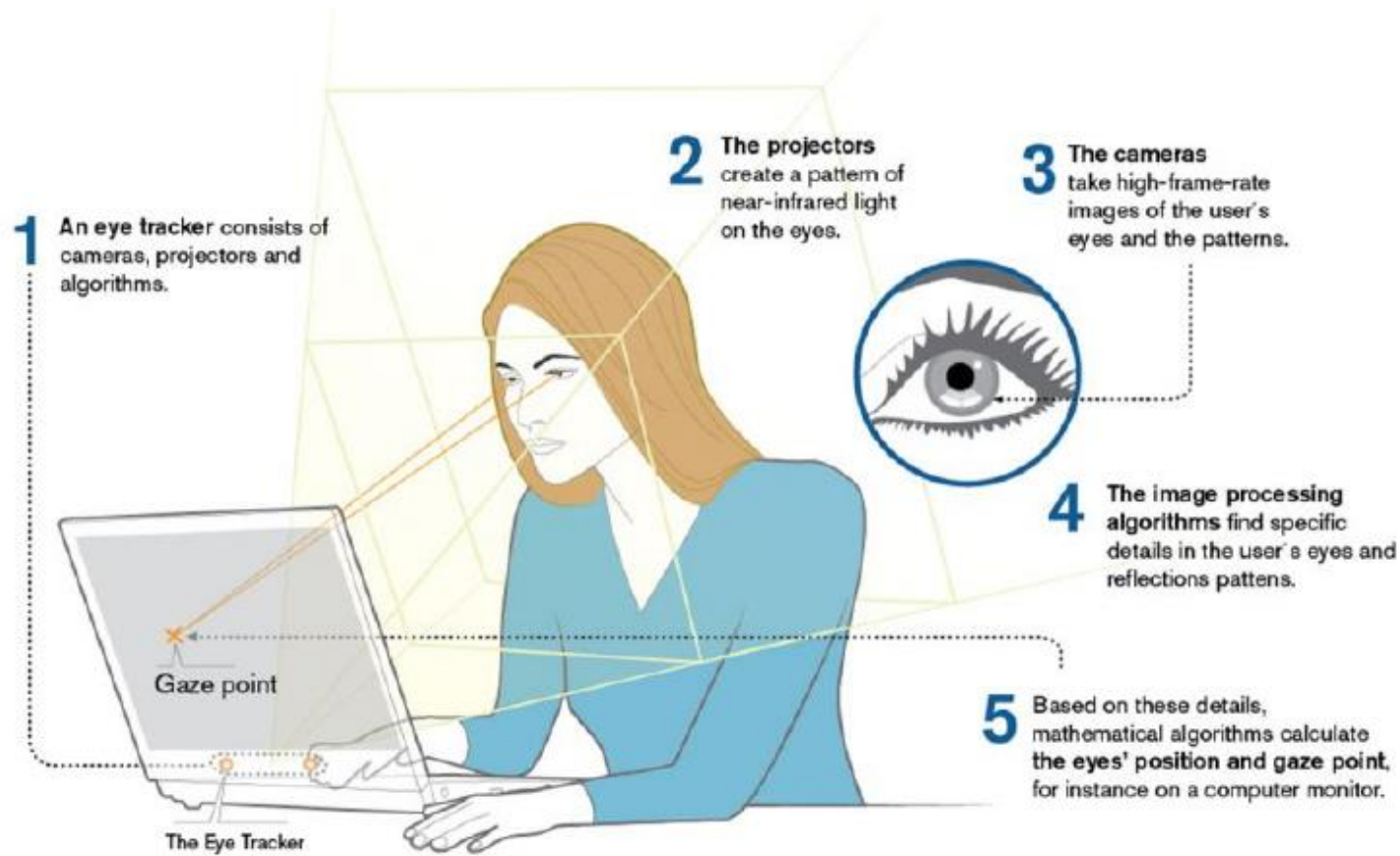
# Ресурсы на проведение А/В-тестов

- Для проведения теста нужна инфраструктура:
  - ✓ Разбиение пользователей на группы
  - ✓ Сбор и хранение необходимых данных
- Результат нельзя получить мгновенно
  - ✓ Необходимо время для накопления достаточного объема данных
- В некоторых случаях можно отказаться от проведения А/В-теста
  - ✓ Если изменение будет в любом случае внесено

# Альтернативы A/B-тестам

- Qualitative analysis (качественный анализ)
- UX исследования, фокус группы
- Eye tracking
- Опросы

# Eye tracking



# Оценка эффекта

По красной кнопке “купить” кликало 5% тестовых пользователей, по зелёной – 8%.

	<b>A</b>	<b>B</b>
<b>Conversion Rate</b>	5%	8%

Какой вывод?

# Оценка эффекта

По красной кнопке “купить” кликало 5% тестовых пользователей, по зелёной – 8%.

	<b>A</b>	<b>B</b>
<b>Conversion Rate</b>	5%	8%

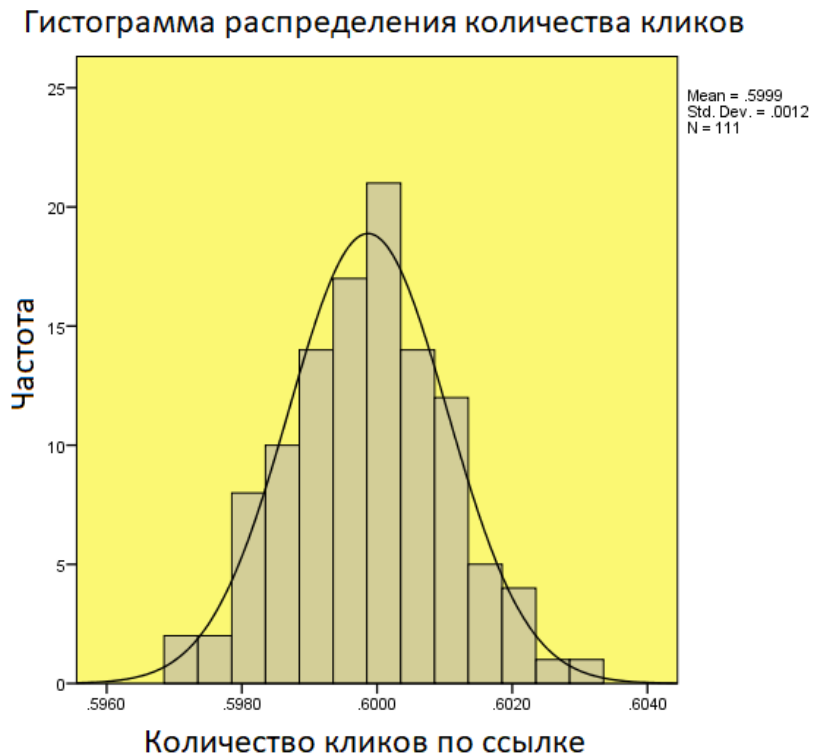
Какой вывод?

Это может быть случайность, а может быть наше изменение увеличило конверсию. Как проверить?



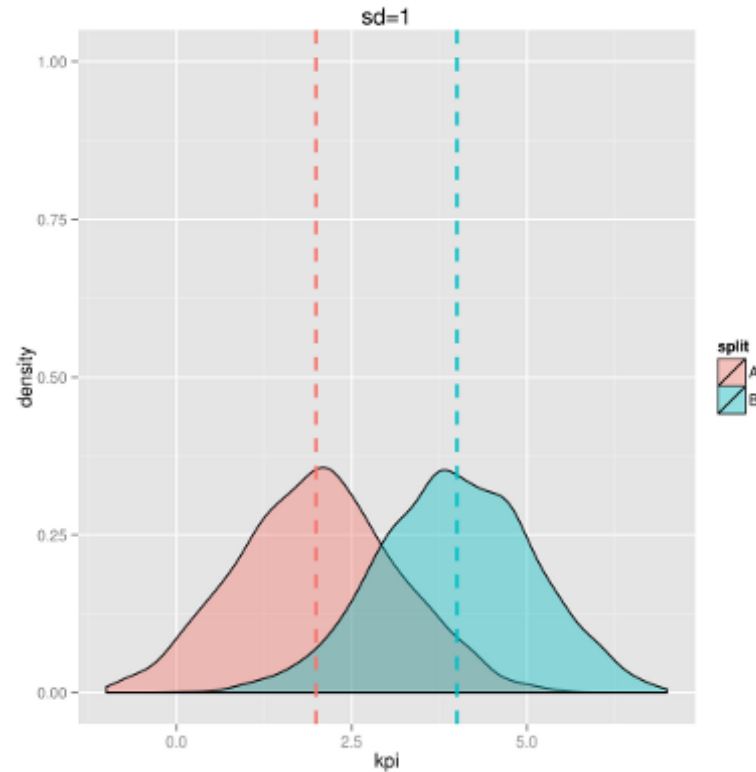
# Оценка эффекта

- значения показателей могут меняться день ото дня — это значит, что мы имеем дело со *случайными величинами*.
- для сравнения случайных величин оценивают *средние значения*, а для оценки среднего значения требуется некоторое время, чтобы накопить историю.



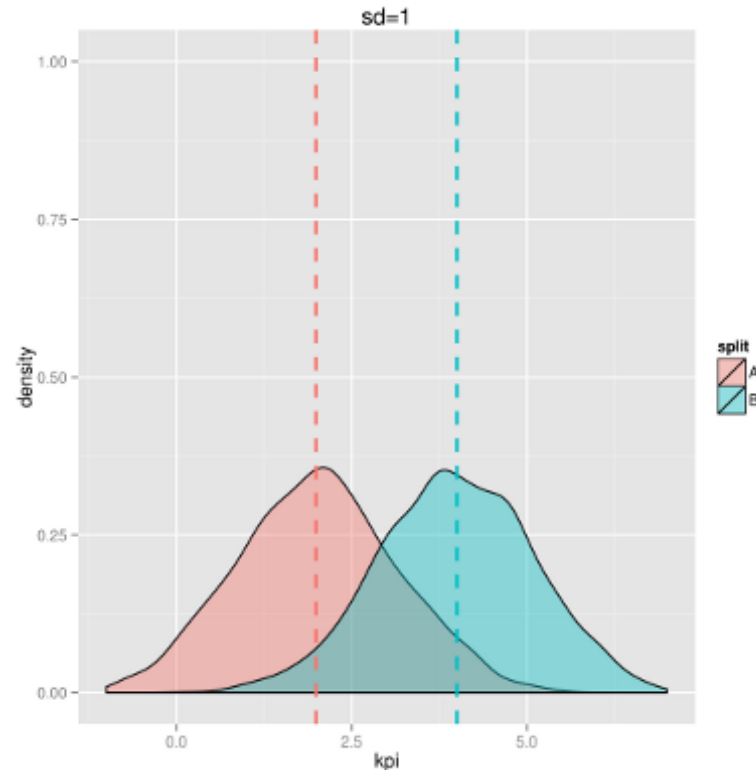
# Оценка эффекта

- эффект от внесения изменения определяют как *разность между средними значениями* показателя в сегментах.



# Оценка эффекта

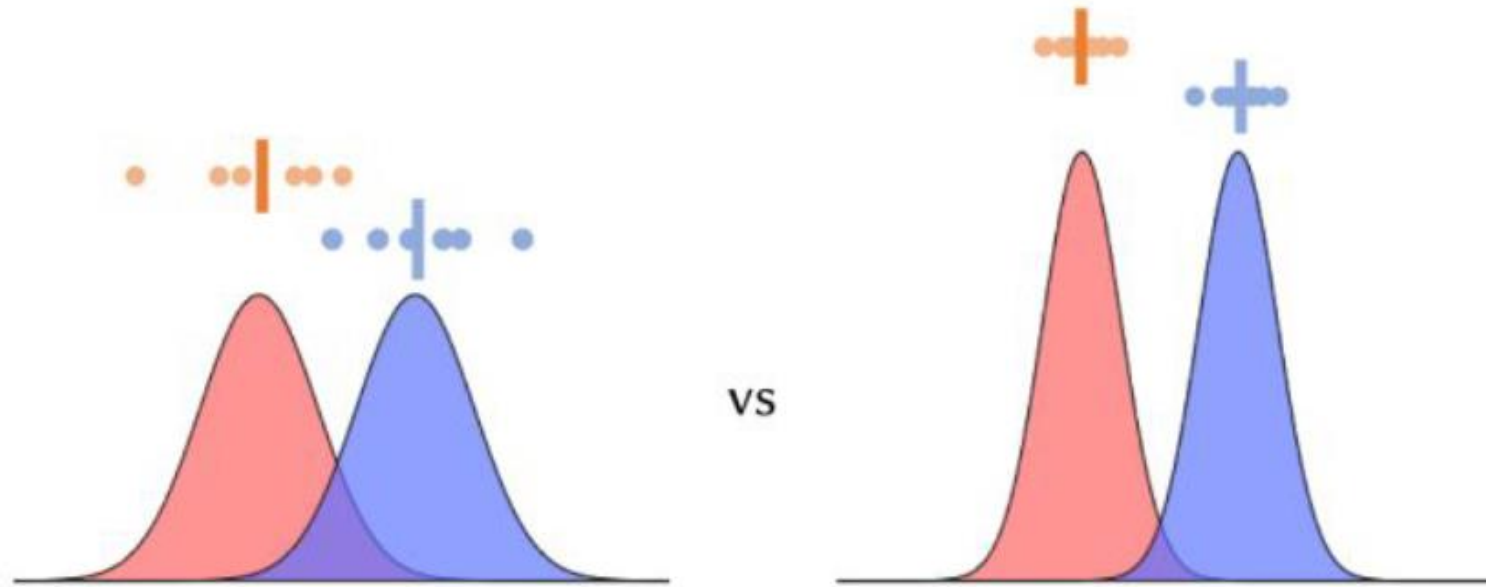
- эффект от внесения изменения определяют как *разность между средними значениями* показателя в сегментах.



Насколько мы уверены в достоверности полученного результата?  
Если мы еще раз проведем тест, то какова вероятность того, что мы сможем повторить результат?

# Оценка эффекта

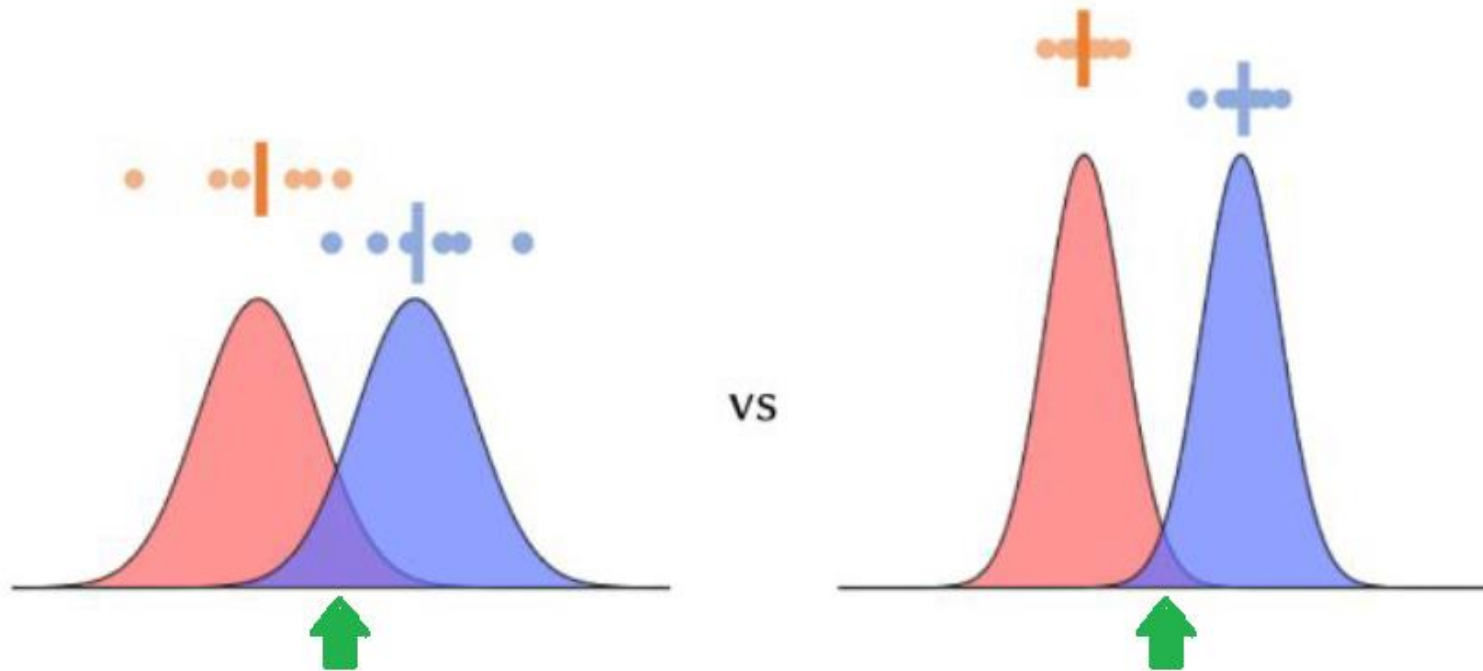
Разность между средними одинаковая.



В чём же разница?

# Оценка эффекта

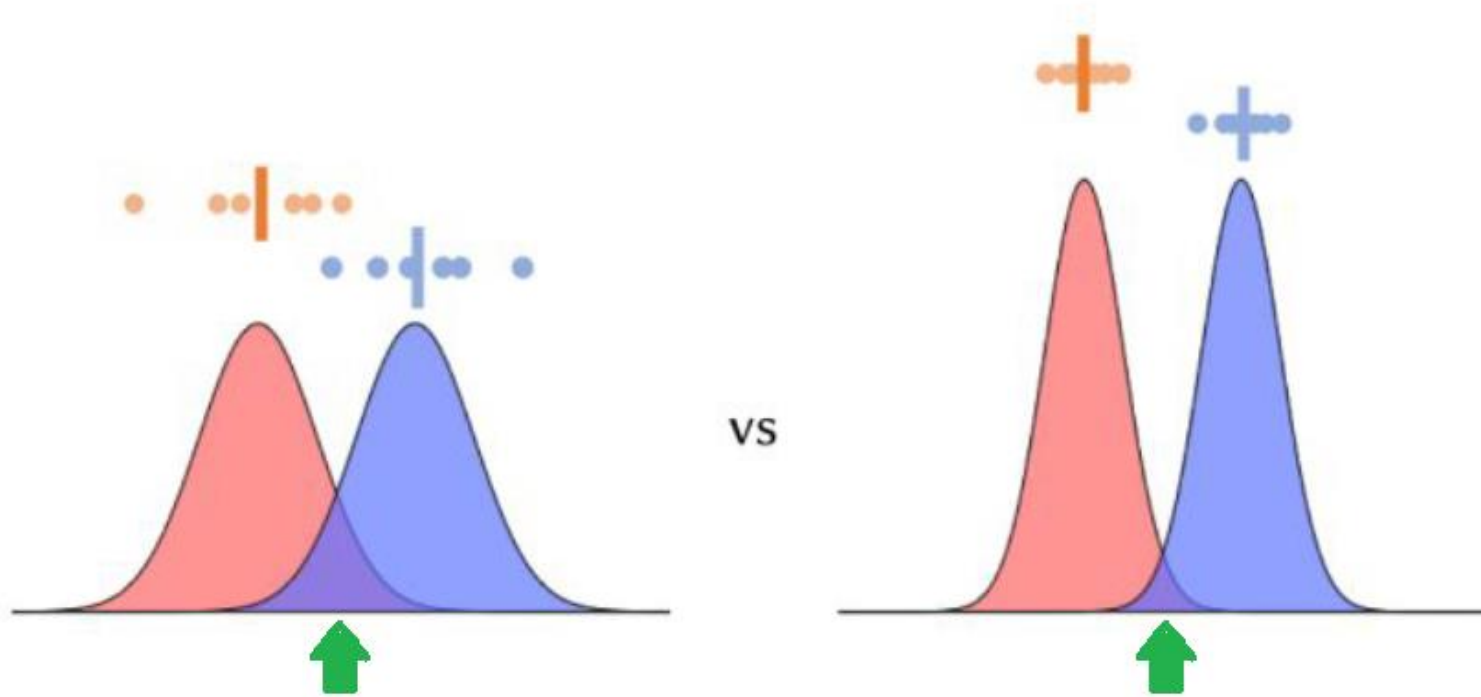
Разность между средними одинаковая.



В чём же разница?

В разбросе значений и в площади пересечения распределений!

# Оценка эффекта



- чем меньше пересечение, тем с большей уверенностью мы можем сказать, что эффект действительно значим.
- эта “уверенность” в статистике называется *значимостью результата*.



# Выбор уровня значимости

- как правило, для принятия положительного решения об эффективности изменений уровень значимости выбирают равным 90%, 95% или 99%.
- пересечение распределений при этом равно соответственно 10%, 5% или 1%.
- при невысоком уровне значимости существует опасность сделать *ошибочные выводы* об эффекте, полученном в результате изменения.

# Проверка статистических гипотез

Для оценки значимости результатов необходимо провести *проверку гипотезы*. Что это?

- Нулевая гипотеза: статистической разницы между средними нет
- Альтернативная гипотеза: разница есть

Нам необходимо применить алгоритм проверки гипотез из математической статистики для того, чтобы решить: принять нулевую гипотезу или отвергнуть её.

# Проверка статистических гипотез

Для оценки значимости результатов необходимо провести *проверку гипотезы*. Что это?

- Нулевая гипотеза: статистической разницы между средними нет
- Альтернативная гипотеза: разница есть

Нам необходимо применить алгоритм проверки гипотез из математической статистики для того, чтобы решить: принять нулевую гипотезу или отвергнуть её.

# Алгоритм проверки статистической гипотезы: тест Стьюдента

- задаемся порогом уверенности, например, 95%
- по выбранному порогу, пользуясь готовыми таблицами, находим критическое значение  $t_{\text{крит}}$
- рассчитываем значение статистики критерия  $t$  по нашим данным

Если  $t > t_{\text{крит}}$ , то отвергаем нулевую гипотезу, то есть, наблюдаемые различия статистически значимы.

# Алгоритм проверки статистической гипотезы: тест Стьюдента

Пример:

Мы провели эксперимент по клику на баннеры. В группе А результаты составили в среднем  $115 \pm 1$  клик в день, в группе В –  $104 \pm 2$  клика в день. Данные имеют нормальное распределение. Необходимо сделать вывод о статистической значимости полученных различий.

# Алгоритм проверки статистической гипотезы: тест Стьюдента

Пример:

Мы провели эксперимент по клику на баннеры. В группе А результаты составили в среднем 115+-1 клик в день (всего 37 человек), в группе В – 104+-2 клика в день (всего 37 человек). Данные имеют нормальное распределение. Необходимо сделать вывод о статистической значимости полученных различий.

- Задаем порог уверенности: 95%, для него определяем по таблице критическое значение:  $t_{\text{крит}} = 1,993$ .

Значения  $t$ -критерия Стьюдента при уровне значимости 0,05

Число степеней свободы, $f$	$t$ -критерий	Число степеней свободы, $f$	$t$ -критерий	Число степеней свободы, $f$	$t$ -критерий
1	12,71	13	2,16	25	2,06
2	4,3	14	2,15	26	2,06
3	3,18	15	2,13	27	2,05
4	2,78	16	2,12	28	2,05
5	2,57	17	2,11	29	2,04
6	2,45	18	2,10	30	2,04
7	2,37	19	2,09	40	2,02
8	2,31	20	2,09	50	2,01
9	2,26	21	2,08	60	2,00
10	2,23	22	2,07	100	1,98
11	2,20	23	2,07	120	1,98
12	2,18	24	2,06	$\infty$	1,96



# Алгоритм проверки статистической гипотезы: тест Стьюдента

Пример:

Мы провели эксперимент по клику на баннеры. В группе А результаты составили в среднем  $115 \pm 1$  клик в день (всего 37 человек), в группе В –  $104 \pm 2$  клика в день (всего 37 человек). Данные имеют нормальное распределение. Необходимо сделать вывод о статистической значимости полученных различий.

- Задаем порог уверенности: 95%, для него определяем по таблице критическое значение:  $t_{\text{крит}} = 1,993$ .

- Вычисляем значение статистики критерия:  $t = \frac{115,4 - 103,7}{\sqrt{1,2^2 + 2,3^2}}$

# Алгоритм проверки статистической гипотезы: тест Стьюдента

Пример:

Мы провели эксперимент по клику на баннеры. В группе А результаты составили в среднем  $115 \pm 1$  клик в день (всего 37 человек), в группе В –  $104 \pm 2$  клика в день (всего 37 человек). Данные имеют нормальное распределение. Необходимо сделать вывод о статистической значимости полученных различий.

- Задаем порог уверенности: 95%, для него определяем по таблице критическое значение:  $t_{\text{крит}} = 1,993$ .

- Вычисляем значение статистики критерия:  $t = \frac{115,4 - 103,7}{\sqrt{1,2^2 + 2,3^2}}$

- $t = 4.51 > t_{\text{крит}} = 1.993 \Rightarrow$  есть статистически значимые различия.

# A/B-тест онлайн

<https://abtestguide.com/calc/>