

Интенсив

по анализу А/В-тестирований

Привет!

Вступайте в чат Телеграм

<https://t.me/joinchat/Cz4zNVhM2vKwJae1cJlM5g>



Опыт

Мы консультируем продуктовые команды, а также ведем обучающие мероприятия по анализу онлайн и офлайн экспериментам

Ключевые клиенты



Организация

Где мне найти материалы лекции?

Все материалы лекции (видео + презентации) доступны в Notion в тот же день после лекции

Когда ждать видео?

Видео появляется вечером того же дня, ссылки появятся в Notion. Уведомление будет в [чате](#)
[телеграмма](#)

Что если у меня появился вопрос к преподавателю во время лекции?

Вопросы можно задавать в зуме в разделе Questions. Почитать как это работает можно [тут](#)

Нужно ли мне знать Python или R, чтобы успешно пройти интенсив?

Мы разработали тренажер для тех, кто не умеет программировать

http://178.128.248.112:3838/exp_fest_calc/

Все домашние задания можно выполнить на нем (кроме ДЗ 4-го дня). Тренажер находится на этапе тестирования и может выдавать ошибки, но это бывает редко и поэтому не должно возникнуть проблем с выполнением ДЗ

Что сегодня в программе?

Об экспериментах

Быстрое введение в математическую статистику

e^xperiment fest

День 1

Об экспериментах

e^xperiment fest

ИДЕИ И ГИПОТЕЗЫ

07/2020

e^xperiment fest

Тезис 1

Проблематика

Все начинается с идеи

- Источники: анализ данных, анализ рынка, генерация
- Неподтвержденная идея является гипотезой

Об экспериментах

e^xperiment fest

Тезис 2

Проблематика

Каждой идее – свой метод проверки

- Глубинные интервью
- Опросы
- Юзабилити-тестирования

Об экспериментах

e^xperiment fest

Тезис 3

Проблематика

Проверке гипотезы необходимы условия

- Отсутствие влияющих факторов, кроме самого тестируемого изменения
- Репрезентативная оценка
- Точность

Об экспериментах

e^xperiment fest

ЧТО ТАКОЕ АБ

07/2020

e^xperiment fest

С чего все начиналось?

- **Биология, химия, медицина.** Статистика как инструмент использовалась для клинических исследований. Благодаря этим задачам математическая статистика развивалась.

Об экспериментах

e^xperiment fest

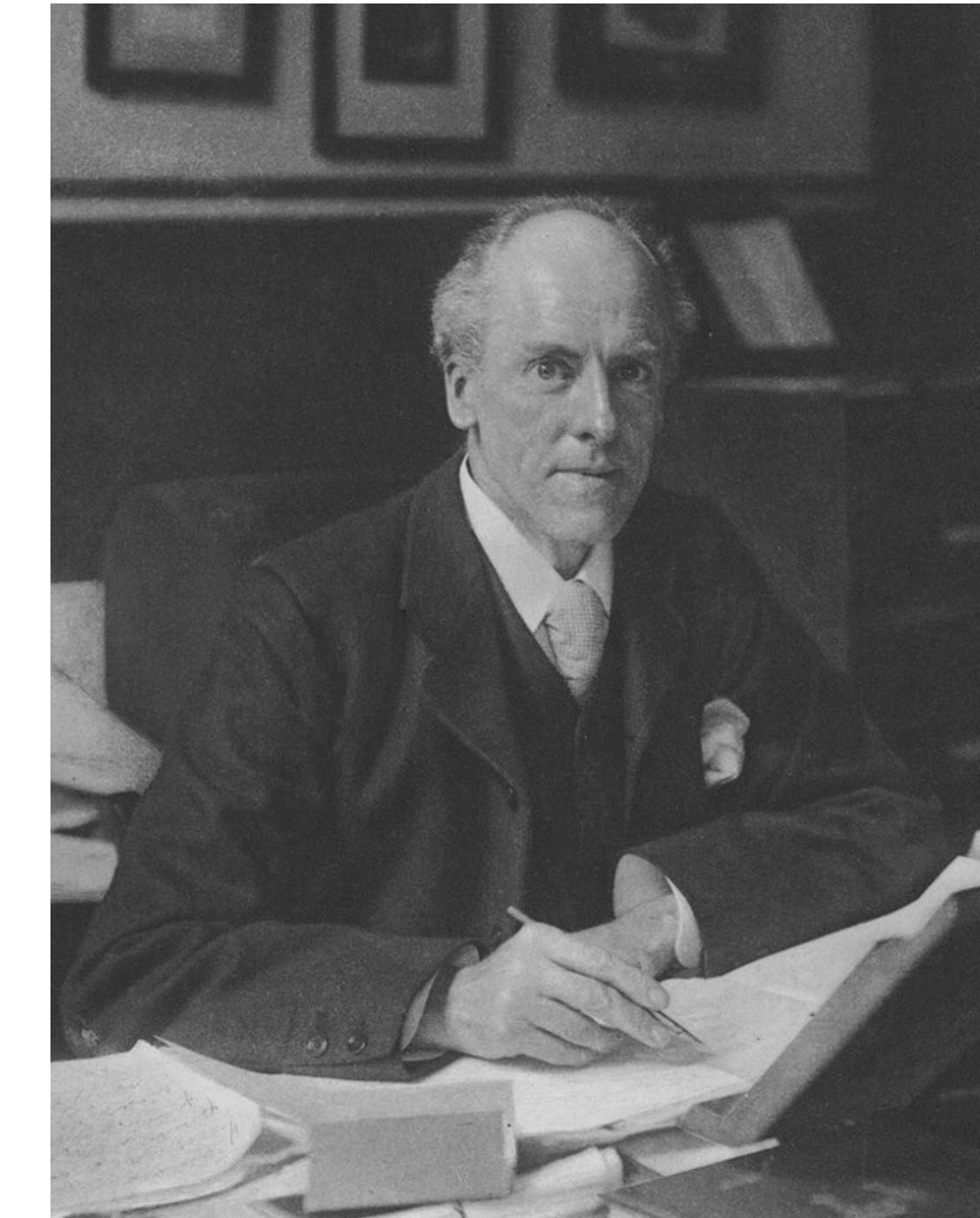
Фишер



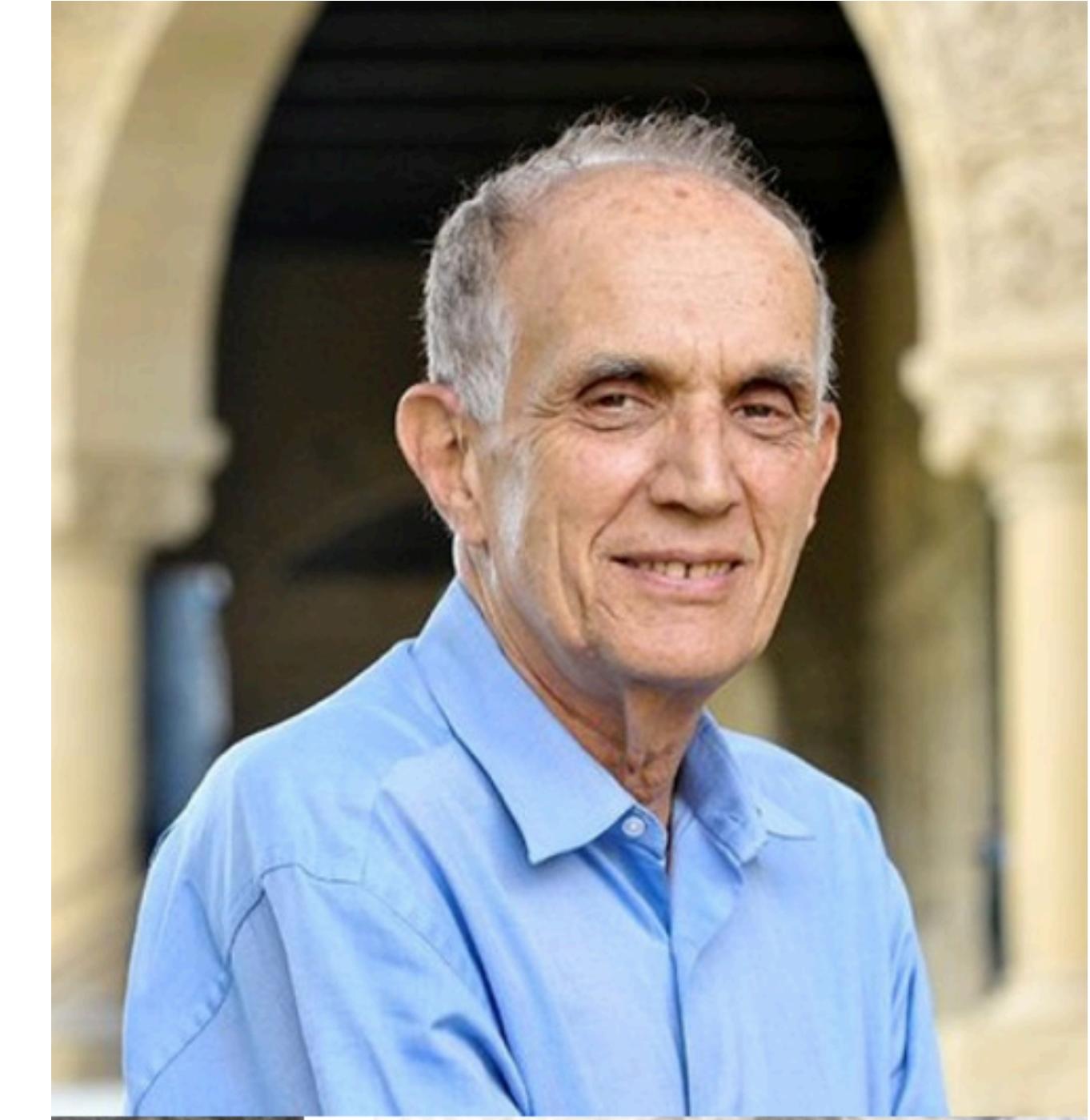
Госет



Пирсон



Эфрон



Об экспериментах

e^xperiment fest

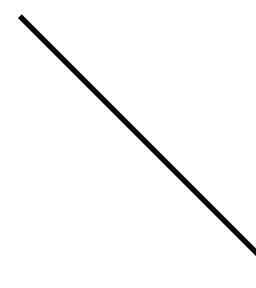
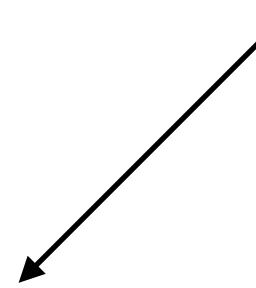
A/B-тестирование – это где...

- Проверяется два и более варианта (контроль и тест) с целью определения наиболее эффективного
- Степень эффективности «измеряется» с помощью посчитанных вероятностей ложноположительных и ложноотрицательных случаев

Об экспериментах

e^xperiment fest

Решения



Ложноотрицательные

Полезные изменения
упускаются из виду

Ложноположительные

Публикуем бесполезные
изменения, которые не
работают

Об экспериментах

e^xperiment fest

Ложноположительное решение

Гипотеза: повышение недельной цены подписки с 1\$ до 2\$

Итог:

На первых двух днях эксперимента был зафиксирован статистически значимый результат. Продакт и аналитик приняли решение принять результат как успешный. А вот после публикации изменения на всех пользователей – ключевой показатель изменился в худшую сторону.

Об экспериментах

e^xperiment fest

Ложноотрицательное решение

Проводили эксперимент 2 дня, не видели разницы, остановили

Итог:

Т.к. в эксперименте было охвачено только 2 дня, мы не
учитываем поведение аудитории в остальные дни
недели.

Возможно изменение имеет отложенный эффект:
пользователь в понедельник попал в тестовую группу, а
в пятницу принял решение

Об экспериментах

e^xperiment fest

Области применения A/B:

- Эксперименты в дизайне (UI / UX)
- Тестирование нового функционала в приложении и на сайте
- Операционные эксперименты
- Оптимизация back-end'a и алгоритмов (например, ранжирование)
- Эксперименты в ценообразовании

Об экспериментах

e^xperiment fest

Метрики эксперимента: уровни

Целевые

Показатели, на которые
направлено изменение

Опережающие

Показатели, хорошо
коррелируемые с целевыми,
дающие предикт и полезны
тогда, когда нет времени ждать
основную метрику

Guardrail

Показатели, на которые направлено
влияет изменение, но не являющиеся
целевыми. Рекомендуется за ними
наблюдать и на их основе в том числе
принимать решение (например,
каннибализация)

Пример системы уровней

Пример: e-commerce, тест нового UI корзины

Целевые

- Конверсия в покупку, средний чек, ARPU, ARPPU

Опережающие

- Добавления товара в корзину на сессию, просмотры товаров на сессию, отток чекаута, ошибки на чекауте

Guardrail

- Время от входа в корзину до ее прохождения, доля поисковых запросов из корзины, взаимодействие с рекомендательными блоками в корзине

Об экспериментах

e^xperiment fest

Пример системы уровней

Пример: образовательный продукт, тест нового образовательного контента

Целевые

- Продление обучения, Средний доход на платящего пользователя (ARPPU)

Опережающие

- Интенсивность обучения, кол-во ошибок в момент обучения, частота обращений в службу поддержки, технические характеристики качества видео

Guardrail

- Время проведенное за одним занятием, прерывание занятий, перемотка видео-контента

Об экспериментах

e^xperiment fest

Метрики: типы

Доли	регистрации, удержание на 7 день [0,1,1,0,0,0,1,1,1,0,1,0,1]
Непрерывные	время в сек./мин./т.п., чек в рублях [1123.32, 324.4, 823.21, 924.91]
Отношения	поездок на водителя, кликов на сессии, цена за 1000 показов [$10/123 = 0.081$, $4129.2/12488 = 0.33$, $1/100 = 0.01$]
Об экспериментах	<i>e^xperiment fest</i>

Типы экспериментов

Классические: А/В

А/А

А/В/С/...

TDI (team draft interleaving)

Diff-in-Diff

Synthetic control

Об экспериментах

e^xperiment fest

A/B

Чем полезен?

Измерить эффект от изменения

Ключевые особенности

- Каждая группа эксперимента видит свой вариант
- Группы независимы
- Группы взяты из одной ГС
- Распределение может быть неравномерным

Об экспериментах

e^xperiment fest

A/A

Чем полезен?

- Проверить сплит-систему
- Выбрать гомогенные группы

Ключевые особенности

- Группы независимы
- Группы взяты из одной ГС
- Часто используется для симуляций

Об экспериментах

e^xperiment fest

A/B/C/...

Чем полезен?

Тот же А/В, только проверяется от 2 и более изменений

Ключевые особенности

- Группы независимы
- Группы взяты из одной ГС
- Сопряжена с проблемой множественной проверки гипотез

Об экспериментах

e^xperiment fest

TDI (*team draft interleaving*)

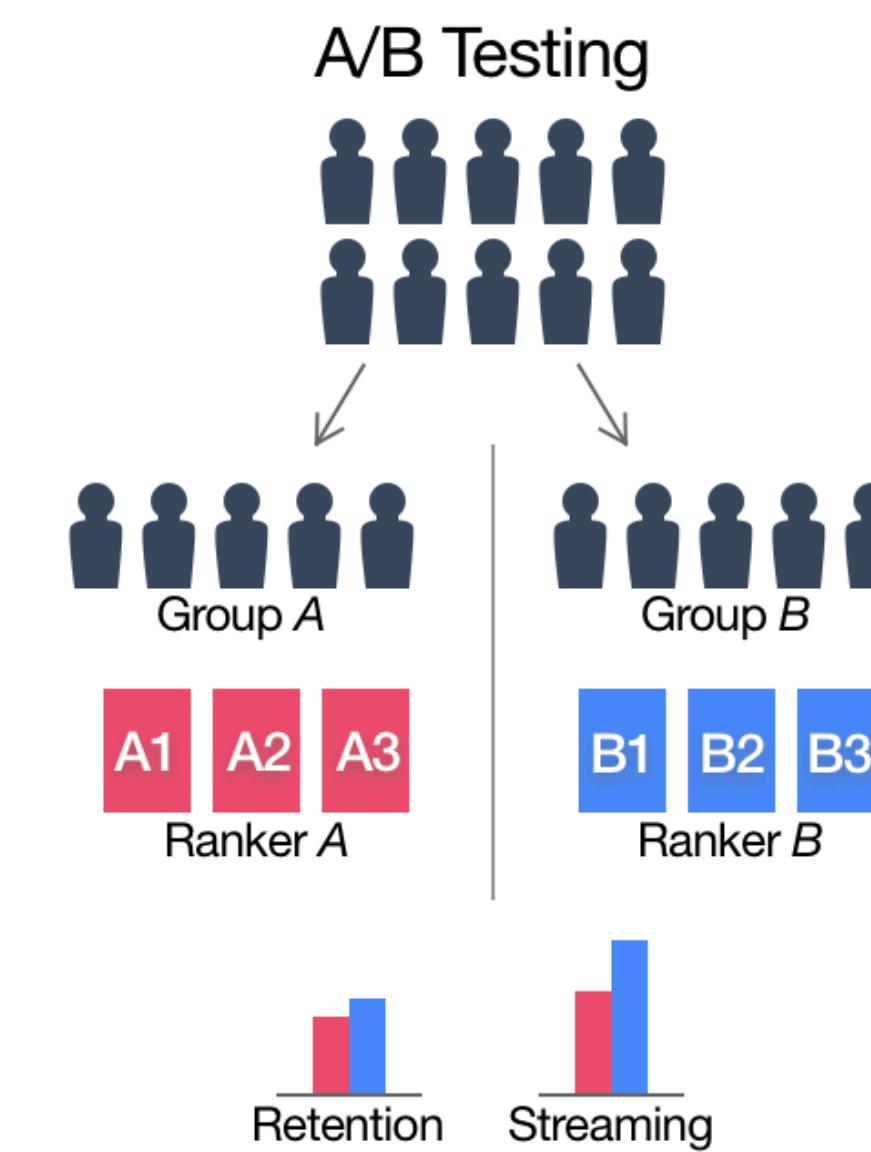
Чем полезен?

Изменение в ранжированных списках

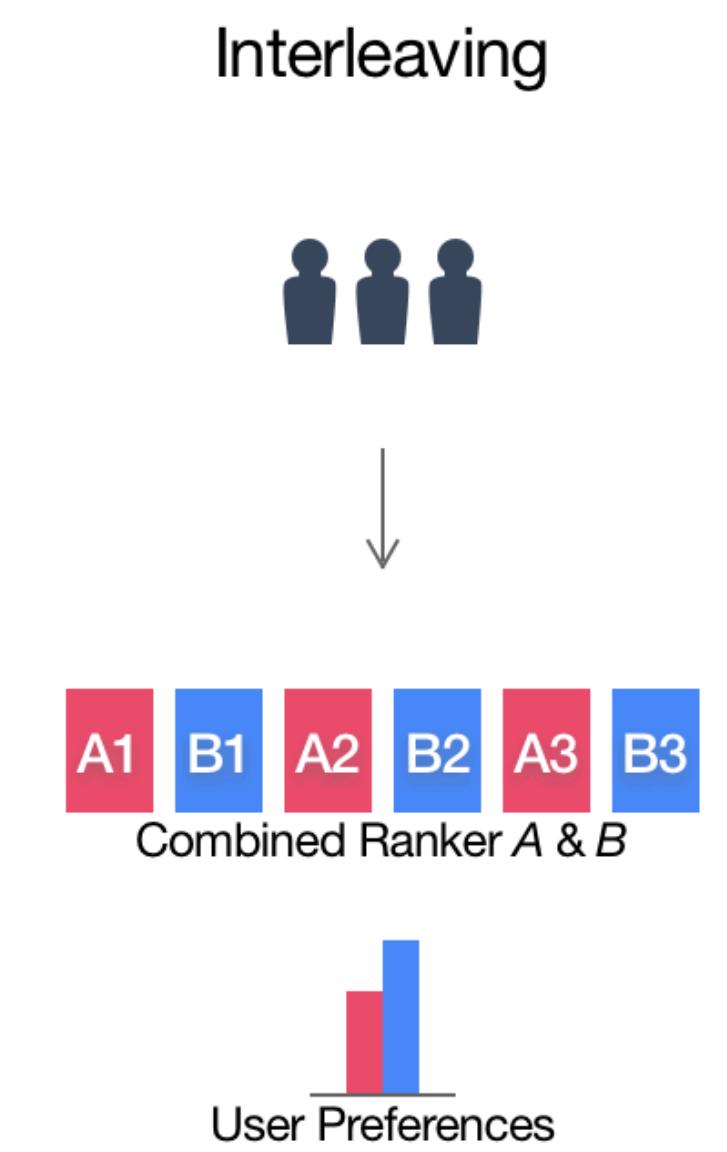
Ключевые особенности

- Один пользователь видит сразу несколько вариантов
- Чаще всего используется в поиске и рекомендациях
- Выборки зависимы – что накладывает особенности

Об экспериментах



e^xperiment fest



TDI

Задача: протестировать два поисковых алгоритма

Метод TDI позволяет показать пользователю результат выдачи пользовательского запроса с применением двух алгоритмов.

Пользователь решает кликом, какой же алгоритм дал релевантный ответ.

Разделения на варианты здесь нет

Об экспериментах

Search bar: что такое дисперсия

dic.academic.ru › dic.nsf › bse › Дисперсия ▾
Дисперсия - это... Что такое Дисперсия?
Дисперсия (от лат. *dispersio* рассеяние) в математической статистике и теории вероятностей, наиболее употребительная мера рассеивания, т. е.

www.matburo.ru › tvart_sub
Как найти дисперсию? Формула дисперсии, примеры ...
Как найти дисперсию случайной величины? Формула дисперсии, примеры вычисления дисперсии дискретной и непрерывной случайных величин.
2 сент. 2015 г. - Добавлено пользователем Университет СИНЕРГИЯ

wiki.loginom.ru › articles › variance ▾
Дисперсия · Loginom Wiki
variance — дисперсия). Пусть X — случайная величина, определённая на некотором вероятностном пространстве. Тогда дисперсией называется.

otvet.mail.ru › question ▾
что такое дисперсия? - Ответы Mail.ru
27 июн. 2008 г. - Пользователь Castiel задал вопрос в категории Наука, Техника, Языки и получил на него 8 ответов.
Что такое дисперсия? (определение) 27 июн. 2016 г.
что такое дисперсия 23 июн. 2007 г.
Что такое дисперсия? 8 июл. 2015 г.
Что такое дисперсия и каким образом она ... 1 июл. 2015 г.
Другие результаты с сайта otvet.mail.ru

neerc.ifmo.ru › wiki › title=Дисперсия_случайной_ве... ▾
Дисперсия случайной величины — Викиконспекты
Определение: Дисперсией случайной величины (англ. *variance*) называется математическое ожидание квадрата отклонения этой случайной величины ...
Вы посещали эту страницу несколько раз (3). Дата последнего посещения: 02.05.20

Diff-in-Diff

Чем полезен?

Когда нет возможности поделить пользователей
на группы в один момент времени

Ключевые особенности

- Группы зависимы и разнесены во времени
- Один из типов регрессий

Об экспериментах

e^xperiment fest

Diff-in-Diff

Экономическое изменение в государстве

Одна из групп подвержена воздействию, или участвует в некоторой программе, во втором периоде, но не в первом. Вторая группа не подвержена воздействию ни в одном из периодов.

Метод устраняет смещение при сравнении исходов в опытной и контрольной группах только во втором периоде, которое может быть следствием постоянных различий между этими группами



Об экспериментах

e^xperiment fest

Synthetic control

Чем полезен?

Не чувствителен к социальным (сетевым) эффектам

Ключевые особенности

- Группы отделены друг от друга географически или физически
- Группы схожи по описательным статистикам, но находятся далеко друг от друга
- Контроль регулярно версионирует

Об экспериментах

e^xperiment fest

Synthetic control

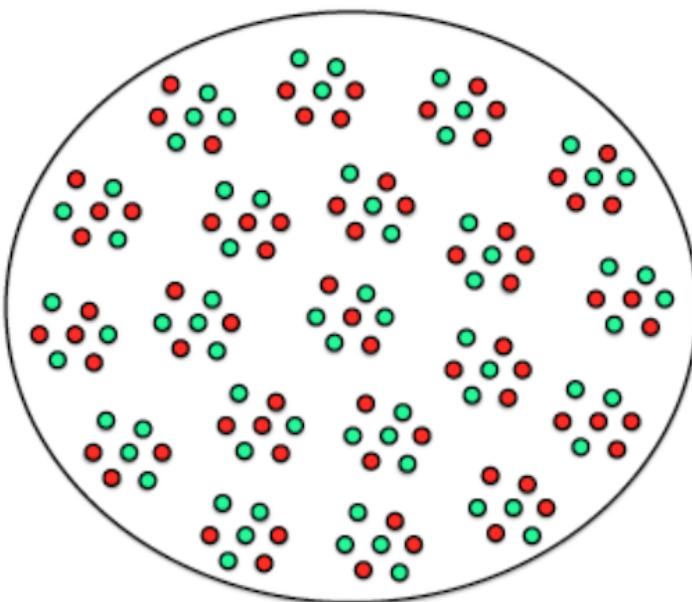
Мы хотим провести эксперимент в социальной сети – дать возможность пользователям отправить анимированные смайлики в сообщениях.

В подобных продуктах есть большая особенность – пользователи общаются между собой. И общение пользователей из групп А и Б могут оказывать сильное влияние на исход всего эксперимента, т.к. в процессе кто-то

BERNOULLI RANDOMIZATION

Assumes that members are independent.

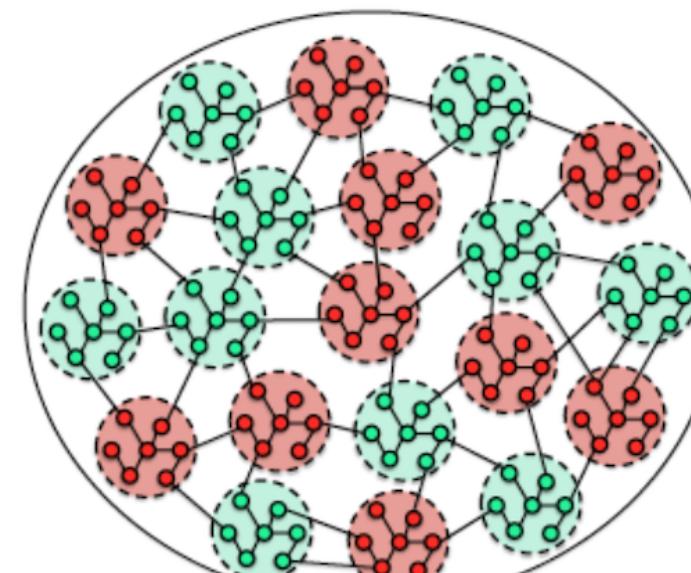
● Control (A) ● Treatment (B)



$$\Delta_{bernoilli} = \frac{\sum Y(\bullet)}{| \bullet |} / \frac{\sum Y(\circ)}{| \circ |}$$

CLUSTER-BASED RANDOMIZATION

Groups tightly connected members and assigns treatment at a group level.



$$\Delta_{cluster-based} = \frac{\sum Y(\bullet)}{| \bullet |} / \frac{\sum Y(\circ)}{| \circ |}$$

Самое важное в Synthetic control – это корректно выбрать группы для его формирования.

Важно, чтобы группы до получения «влияния» не отличались друг от друга по описательным статистикам и были репрезентативны друг другу.

Формирование групп не является разовым процессом, скорее перманентным поиском близких выборок для проведения эксперимента

Варианты для контроля №2



Об экспериментах

Варианты для контроля №1



experiment fest

Как устроено АБ-тестирование в продуктовых командах

- Каждый большой продукт имеет свои особенности как с точки зрения бизнеса так и с точки зрения метрик. Эти особенности накладывают определенные ограничения и дают творческий простор для развития методологии экспериментов

Об экспериментах

e^xperiment fest

Uber

Коммуникация водителей между собой

- Водители между собой общаются
- Общие водители создает сетевой эффект, который влияет на исход эксперимента
- Отслеживать факт общения по событиям крайне проблематично. Нет триггера, что коммуникация началась

Об экспериментах

e^xperiment fest

Uber

Вечный контроль (моделируемый)

- Избавление от сетевых эффектов по средствам разделения пользователей географически
- Используется на экспериментах для водительских мотиваций
- Вынужденная особенность продукта

Варианты для контроля №2



Об экспериментах

e^xperiment fest



Увеличение кол-во экспериментов на ограниченном трафике

- Культура компании обязывает все делать через А/В
- Трафик ограничен – команд много
- Ограничения на а/б создают проблемы для развития продукта

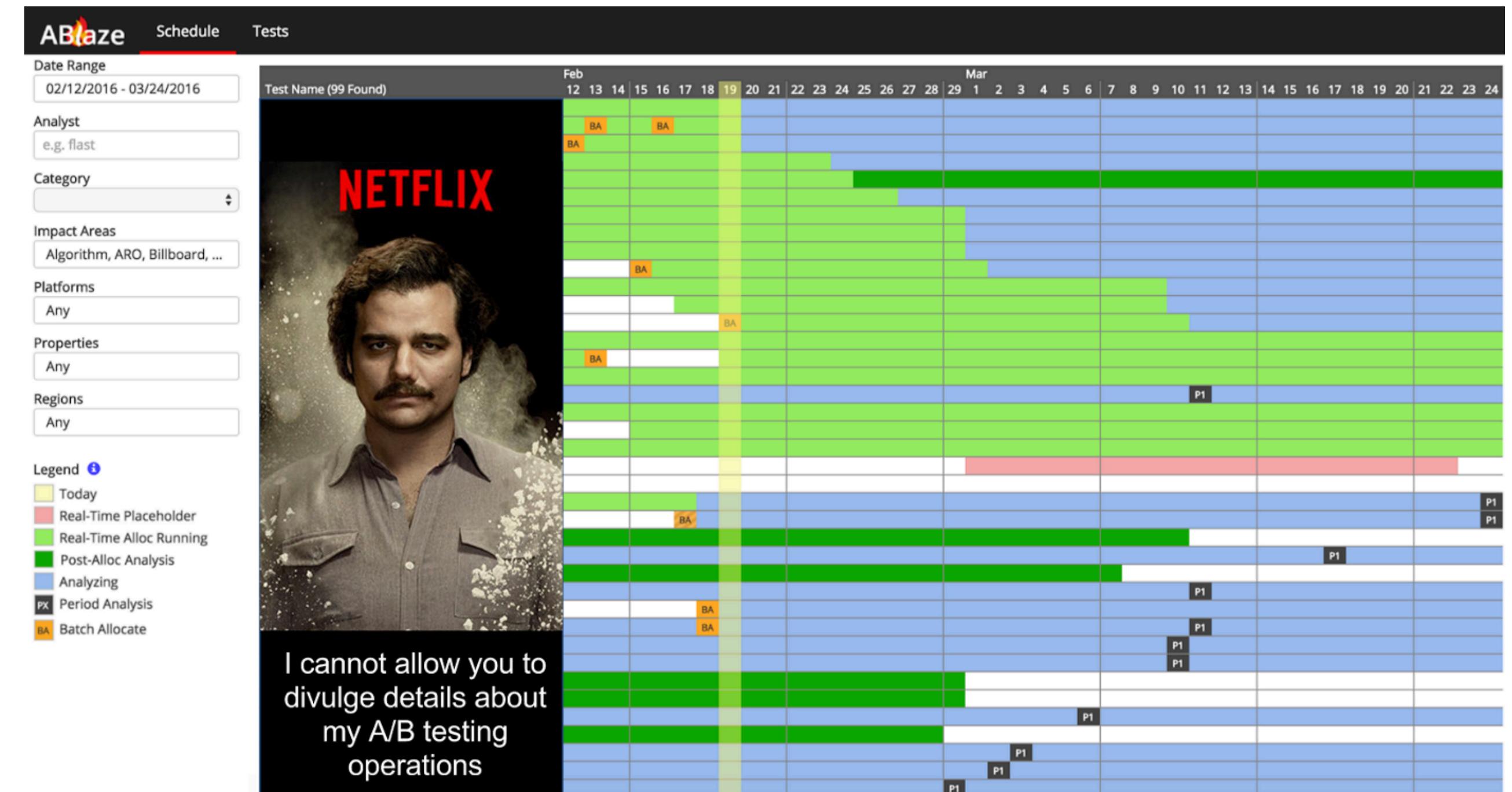
Об экспериментах

e^xperiment fest



Платформа экспериментов, внутренние инструменты и R&D

- Платформа – инструмент автоматизации а/b и развития R&D
- Внутренние инструменты ускоряют процесс анализа сложных экспериментов
- R&D позволяет находить новые методы ускорения а/b тестов



Об экспериментах

e^xperiment fest



Эксперимент может запустить каждый

- Эксперимент может запустить любой член команды на платформе
- Не каждая гипотеза чем то подтверждена
- Регулярно тестируются «гипотезы-пустышки»



Жесткая система модерации

- Каждый эксперимент модерируется
- Чтобы запустить эксперимент на платформе – нужно доказать его ценность и приложить основания для гипотезы
- Запустить эксперимент – сложно

Об экспериментах

e^xperiment fest

- Эксперименты – не просто разделение трафика или базовая статистика, это процесс с полноценным R&D
- Эксперименты – это методология управления гипотезами и результатами
- Чем больше продукт, тем больше людей выделяется под задачи экспериментов

Об экспериментах

e^xperiment fest

День 1

Быстрое введение в математическую статистику

e^xperiment fest

Что узнаем?

- Математическое ожидание случайной величины
- Показатели вариативности. Дисперсия и стандартное отклонение
- Вероятностное распределение

Аналитик Лёня и менеджер Петя решили в пятницу пойти в бар. В их городе есть два бара с разным меню. Но в среднем, за один бокал цена одинаковая. Петя делегировал принятие конечного решения Лёне со словами «ты же аналитик». Куда стоит им идти?

Когорта

№ позиции в меню	Цена, рубли
1	300
2	300
3	330
4	370

Воронка

№ позиции в меню	Цена, рубли
1	100
2	300
3	400
4	500

Чтобы ответить, нам поможет *описательная статистика*

Когорта

№ позиции в меню	Цена, рубли
1	300
2	300
3	330
4	370

Воронка

№ позиции в меню	Цена, рубли
1	100
2	300
3	400
4	500

Чтобы решить задачку, последовательно посчитаем 3 метрики:

1. *Математическое ожидание: среднее по выборке*
2. *Дисперсию:*
 - a. из мат. ожидания будем вычитать каждое значение выборки
 - b. возведем в квадрат
 - c. все просуммируем
 - d. разделим на количество наблюдений. Таким образом мы найдем сумму расстояний разностей
3. *Стандартное отклонение: возьмем корень из посчитанной дисперсии*

Когорта

№ позиции в меню	Цена, рубли
1	300
2	300
3	330
4	370

$$M_1 = 325$$

$$S^2_1 = ((325-300)^2 + (325-300)^2 + (325-330)^2 + (325-370)^2) / 4 = 1100$$

$$SD_2 = \sqrt{S_1} = 33.16$$

Воронка

№ позиции в меню	Цена, рубли
1	100
2	300
3	400
4	500

$$M_2 = 325$$

$$S^2_2 = (325-100)^2 + (325-300)^2 + (325-400)^2 + (325-500)^2) / 4 = 29166.67$$

$$SD_2 = \sqrt{S_2} = 170.78$$

Когорта

№ позиции в меню	Цена, рубли
1	300
2	300
3	330
4	370

$$M_1 = 325$$

$$S^2_1 = 1100$$

$$SD_1 = 33.16$$

Воронка

№ позиции в меню	Цена, рубли
1	100
2	300
3	400
4	500

$$M_2 = 325$$

$$S^2_2 = 29166.67$$

$$SD_2 = 170.78$$

Вывод

Оба бара имеют одинаковые мат.ожидания, однако разброс цен в меню в «Когорту» меньше ($33.16 < 170.78$). Лёне и Пете стоит зайти в «Воронку», если они хотят какого-то разнообразия

Еще один пример использования описательных статистик.

Можно оценить разброс каналов трафика по какой-то

$$cv = \frac{SD}{M} \times 100\%$$

метрике с помощью коэффициента вариации:

Органика	PPC
15%	35%

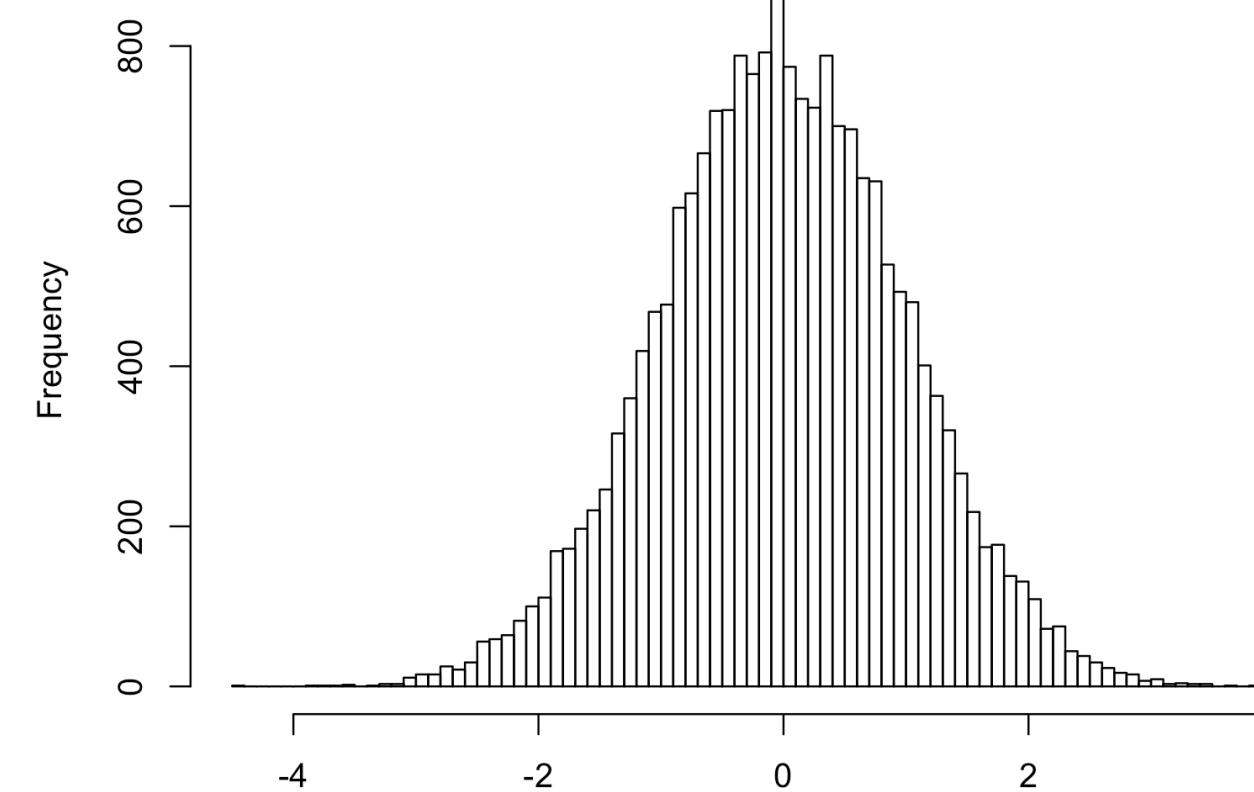
Органика явно устойчивее закупочного трафика

Демонстрация в R

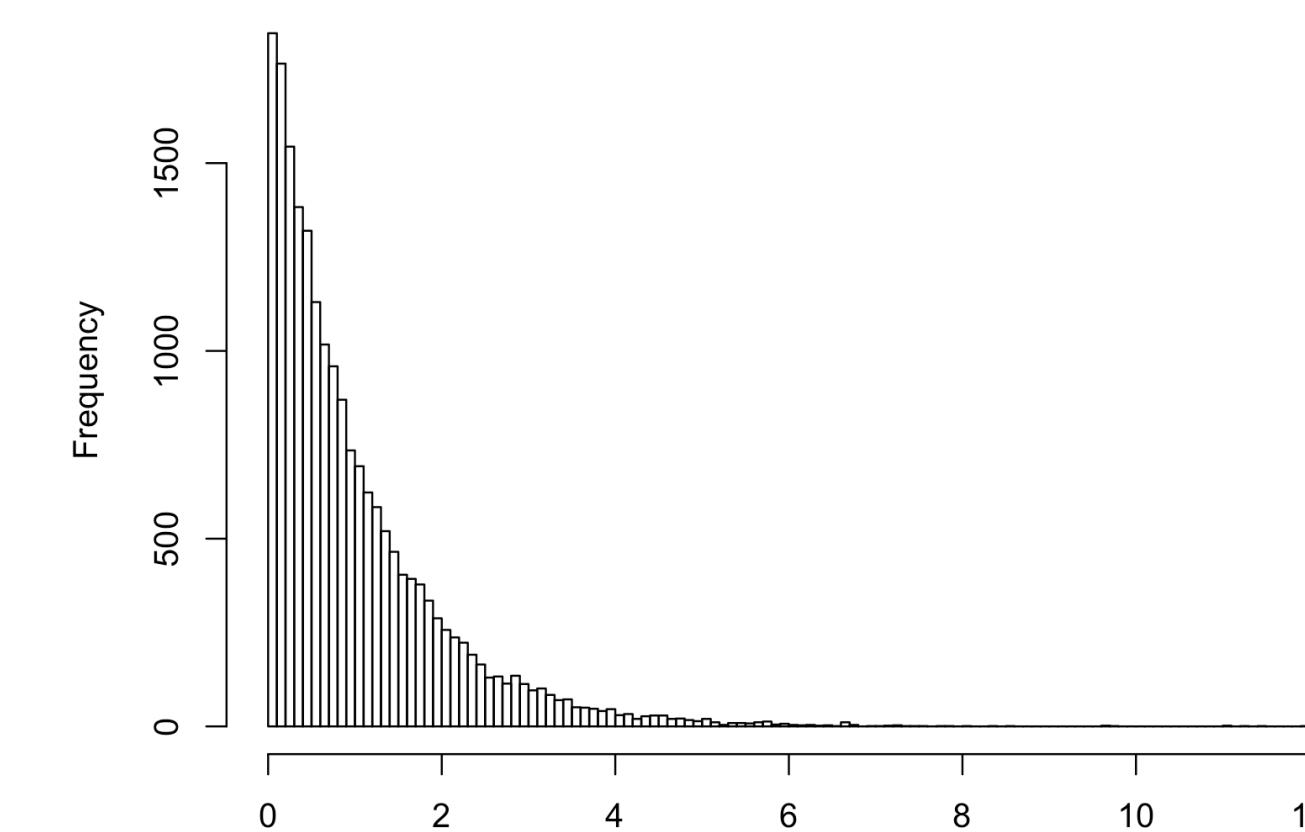
e^xperiment fest

Любые явления, характеристики и события можно описывать с помощью распределений. Стоит учесть, что их форма всегда зависит от природы их происхождения и вероятности встречаемости того или иного значения случайной величины

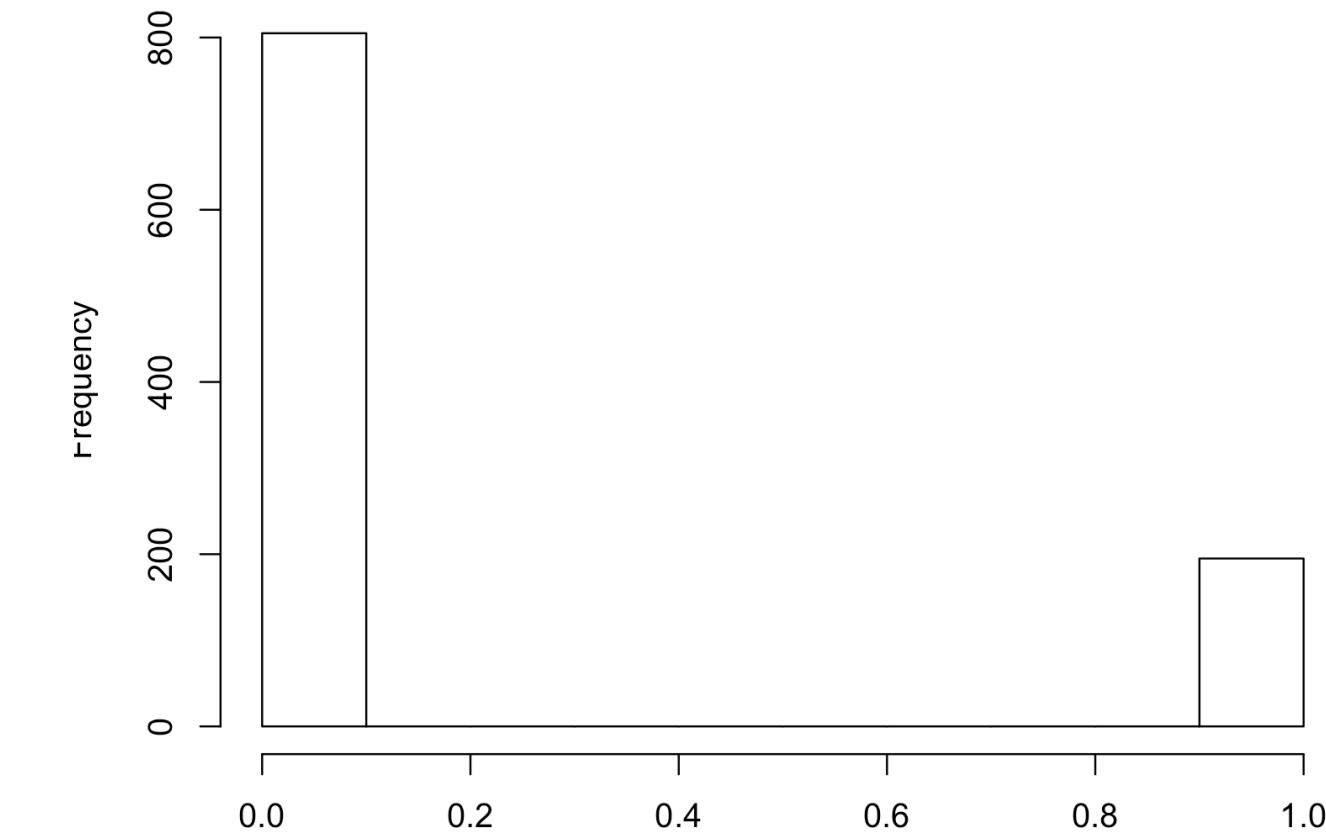
Нормальное распределение



Экспоненциальное распределение

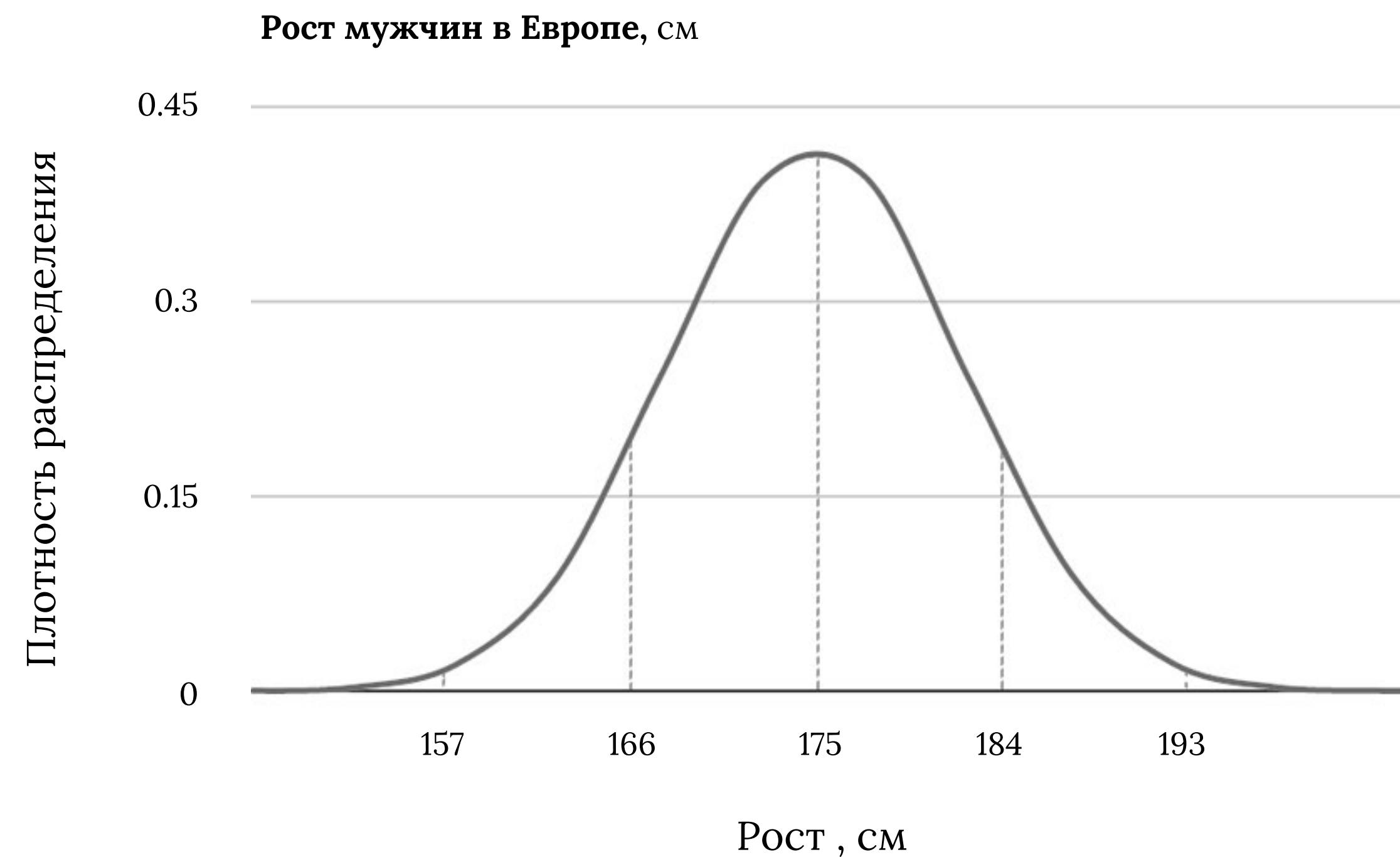


Распределение Бернулли

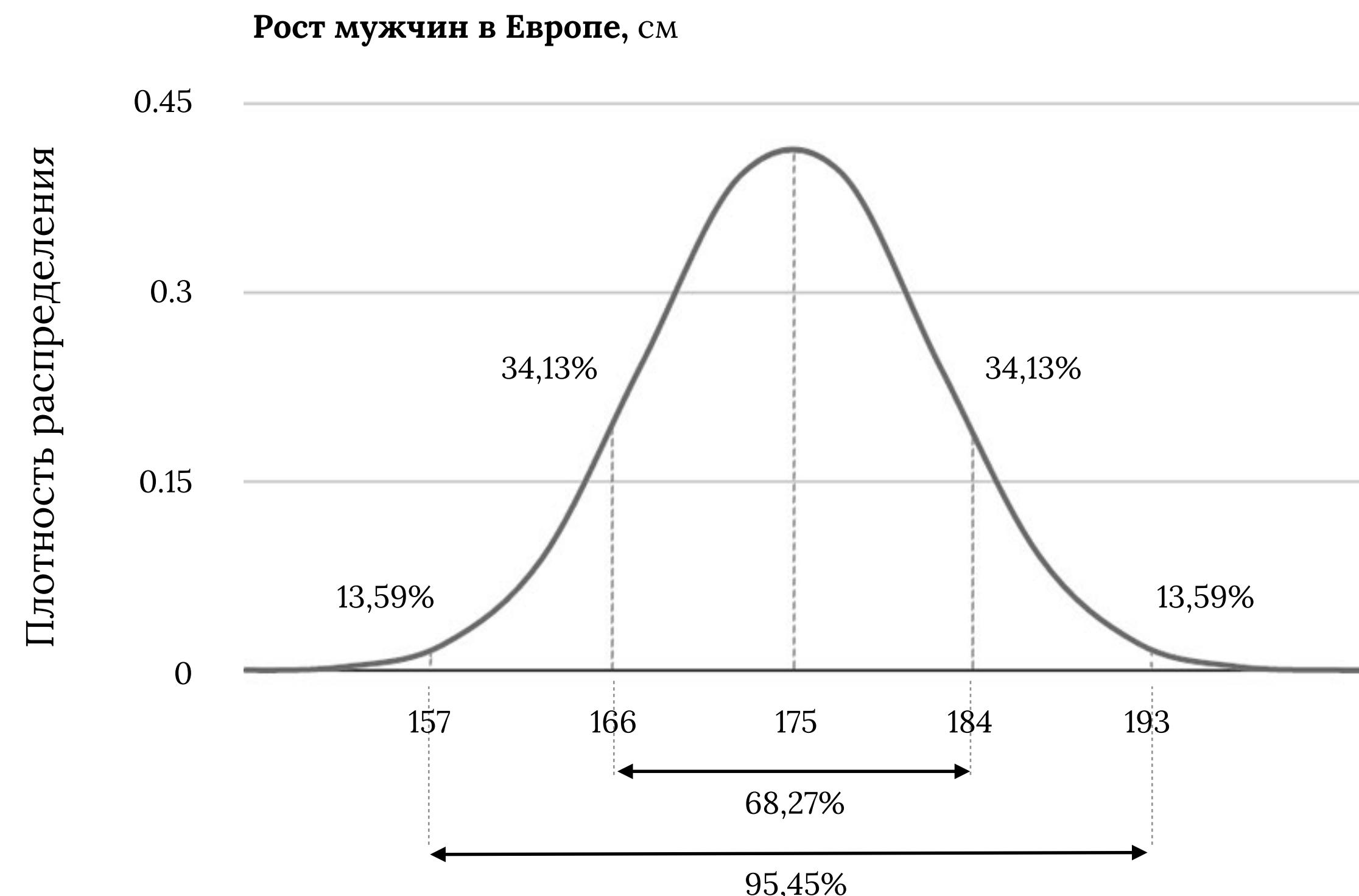


В статистике самым встречающимся распределением является **нормальное распределение**. Оно напоминает форму колокола и удивительным образом описывает многие явления в природе.

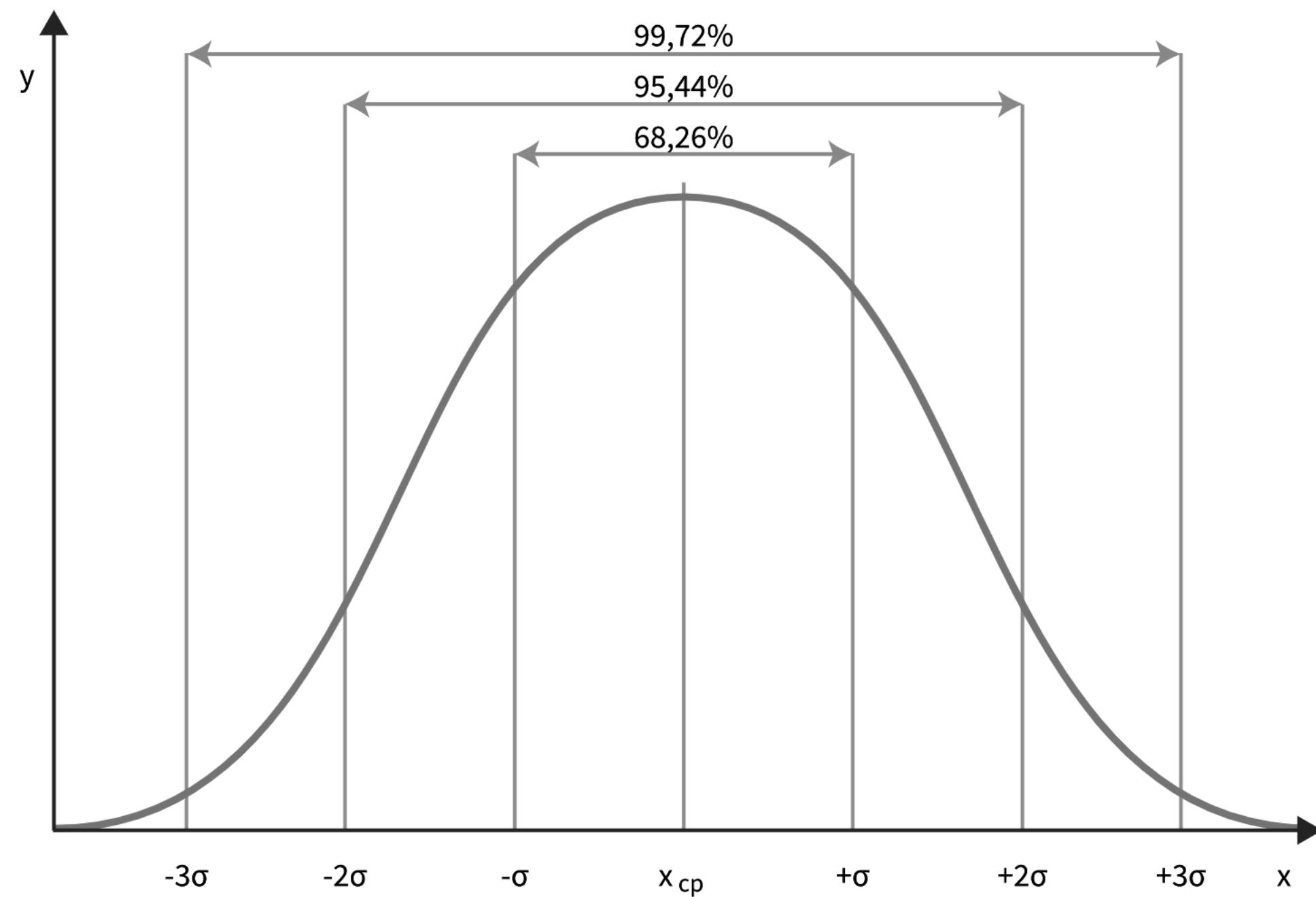
К примеру, рост мужчин в Европе:



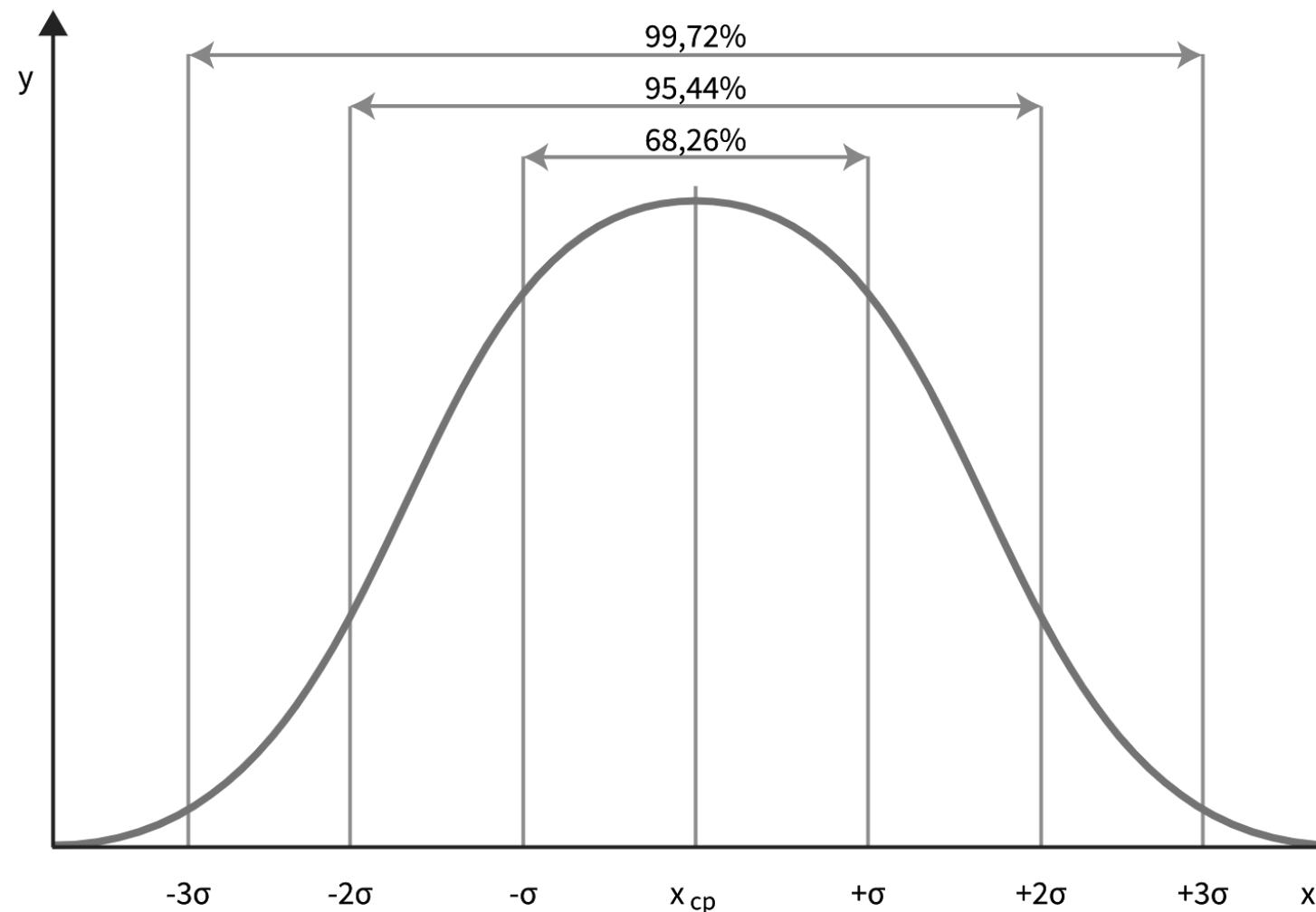
Большая часть значений расположена симметрично относительно среднего, а именно 95% всех значений случайной величины (или в пределах 1,96 стандартных отклонений). Это утверждается в «правиле 3-х СИГМ»



Оно утверждает, что вероятность того, что случайная величина отклонится от своего математического ожидания более, чем на три среднеквадратических отклонения, практически равна нулю. Правило 3-х сигм работает только для случайных величин, распределенных по нормальному закону.



- 68% значений находятся в пределах 1σ
- 95% значений находятся в пределах 2σ
- 99.7% значений находятся в пределах 3σ



№ позиции в меню	Цена, рубли
1	300
2	300
3	330
4	370

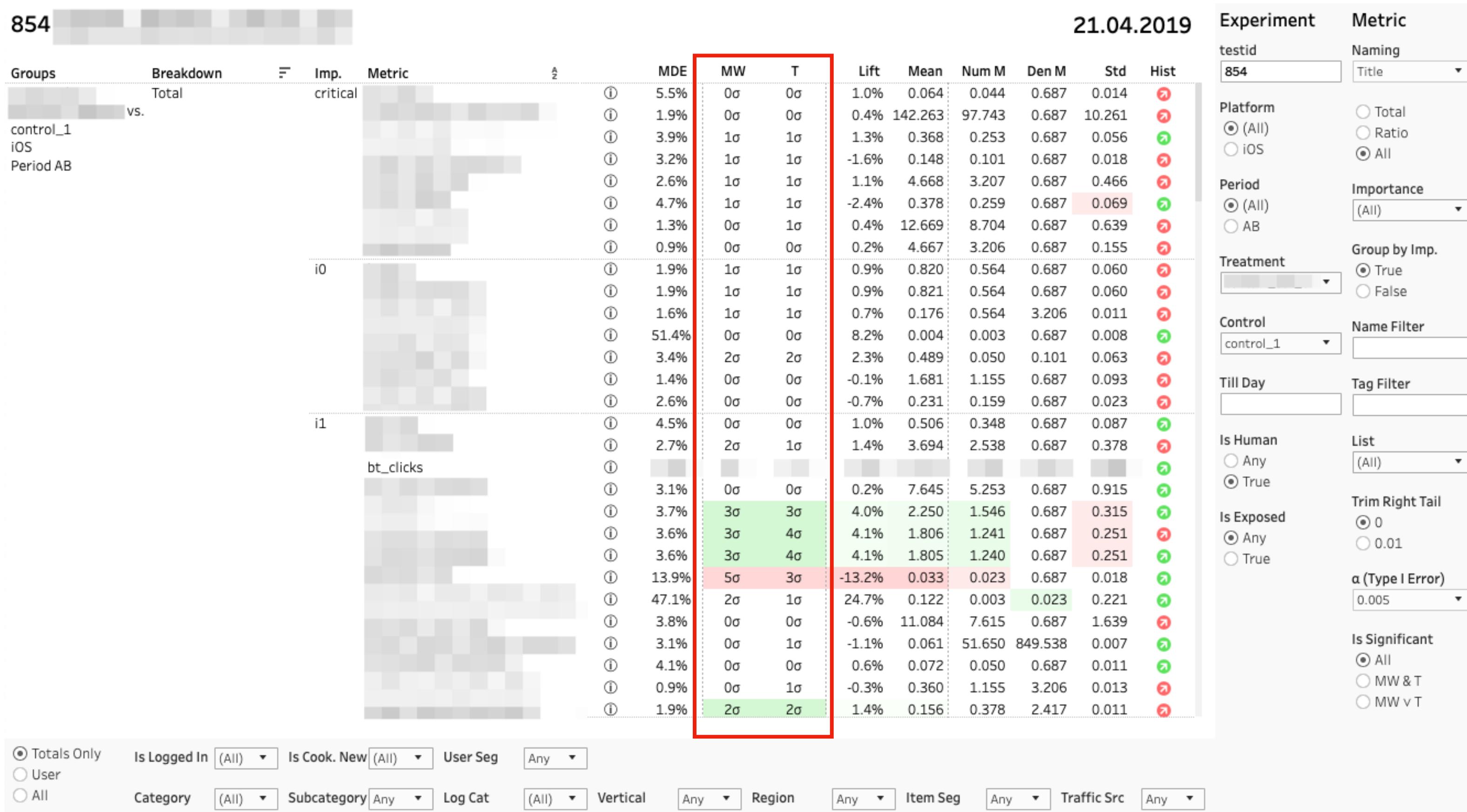
$$M_1 = 325 \text{ и } SD_2 = \sqrt{S_1} = 33.16$$

при 1 сигме $325 \pm 33.16 \cdot 1$ имеем отклонение от **291.84** до **358.16**

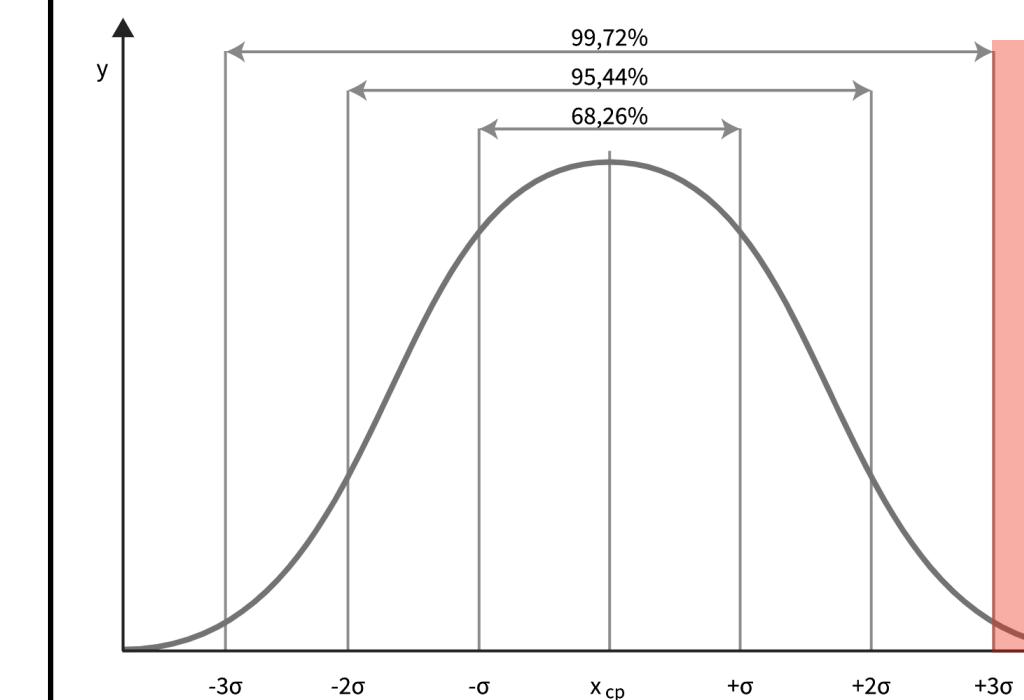
при 2 сигмах $325 \pm 33.16 \cdot 2$ имеем отклонение от **258.68** до **391.32**

при 3 сигмах $325 \pm 33.16 \cdot 3$ имеем отклонение от **225,52** до **424,48**

Не совсем к ЦПТ, но на сигмы ориентируется Avito при расчете метрик. Вышла ли метрика за какую-то сигму или нет можно на дашборде

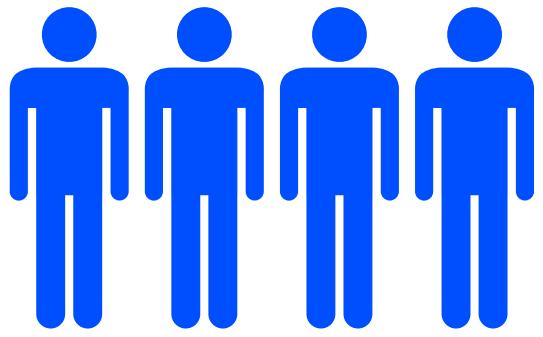


Например, если метрика прокрасилась на 3 сигме, то вероятность встретить случай, где нет разницы между вариантами = $1 - 0,9972 = 0,28\%$

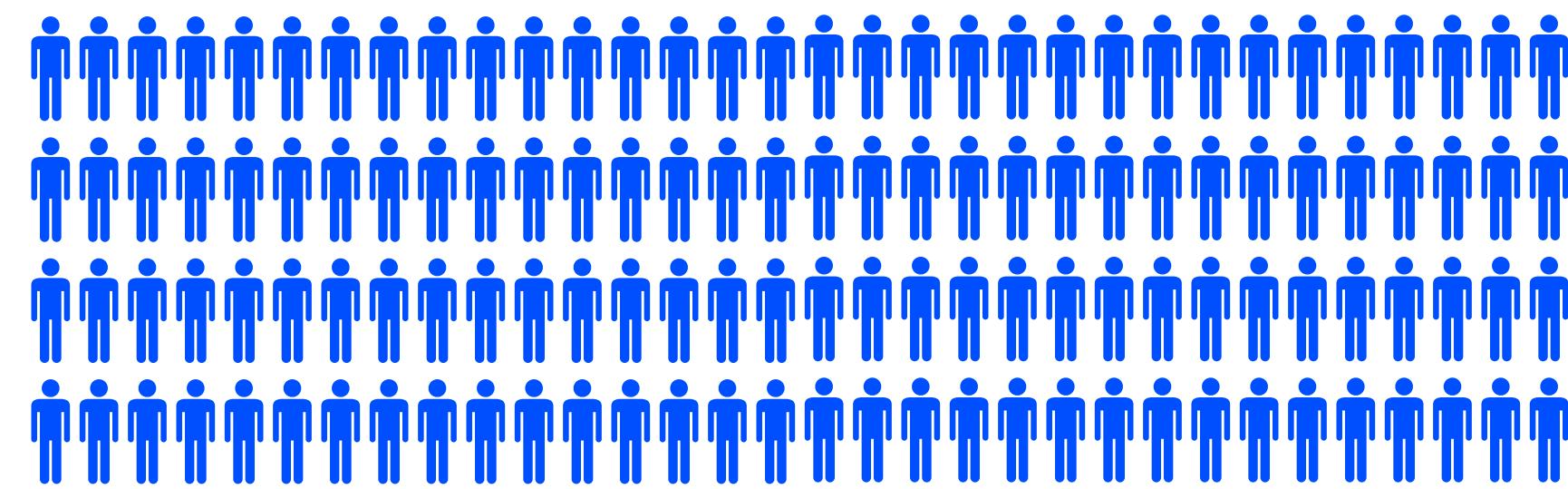


Нам приходится делать заключения на основе выборок
(например, данные по пользователям из эксперимента за
период 'x'), взятые из генеральной совокупности (все
возможные пользователи).

Выборка, размером $n = 10000$



Генеральная совокупность



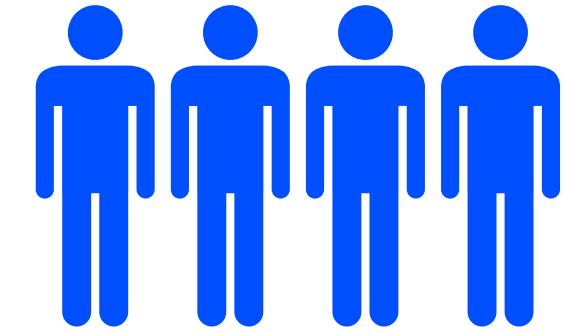
Демонстрация

e^xperiment fest

Статистики делают на основе выборок заключения о генеральной совокупности, чаще всего о ней ничего не известно (какая форма распределения, ее мат.ожидание и дисперсия).

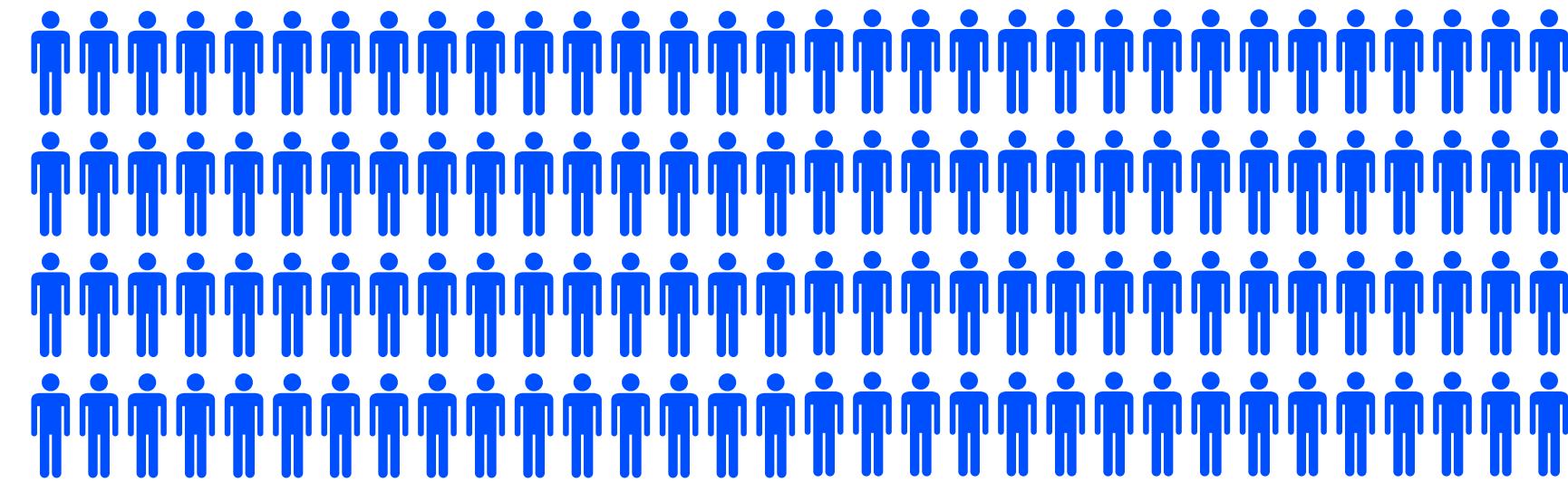
Выборка n

$$M = \dots, SD = \dots$$



Генеральная совокупность

$$\mu = ?, \sigma = ?$$



Выборка #1

$$M = 10, SD = 1.1$$

Выборка #2

$$M = 9, SD = 1$$

Выборка #3

$$M = 10.1, SD = 1$$

... и так возьмем 1000 выборок

И если мы возьмем средние всех этих выборок, то они скорее всего распределются нормально. И как мы уже знаем, 95% всех значений будут находиться в интервале $M \pm 1,96 \times (SD \div \sqrt{n})$. В этом нам помогает важное свойство *центральной предельной теоремы* (ЦПТ)

Выборка #1

$$M = 10, S^2 = 1.1$$

Выборка #2

$$M = 9, S^2 = 1$$

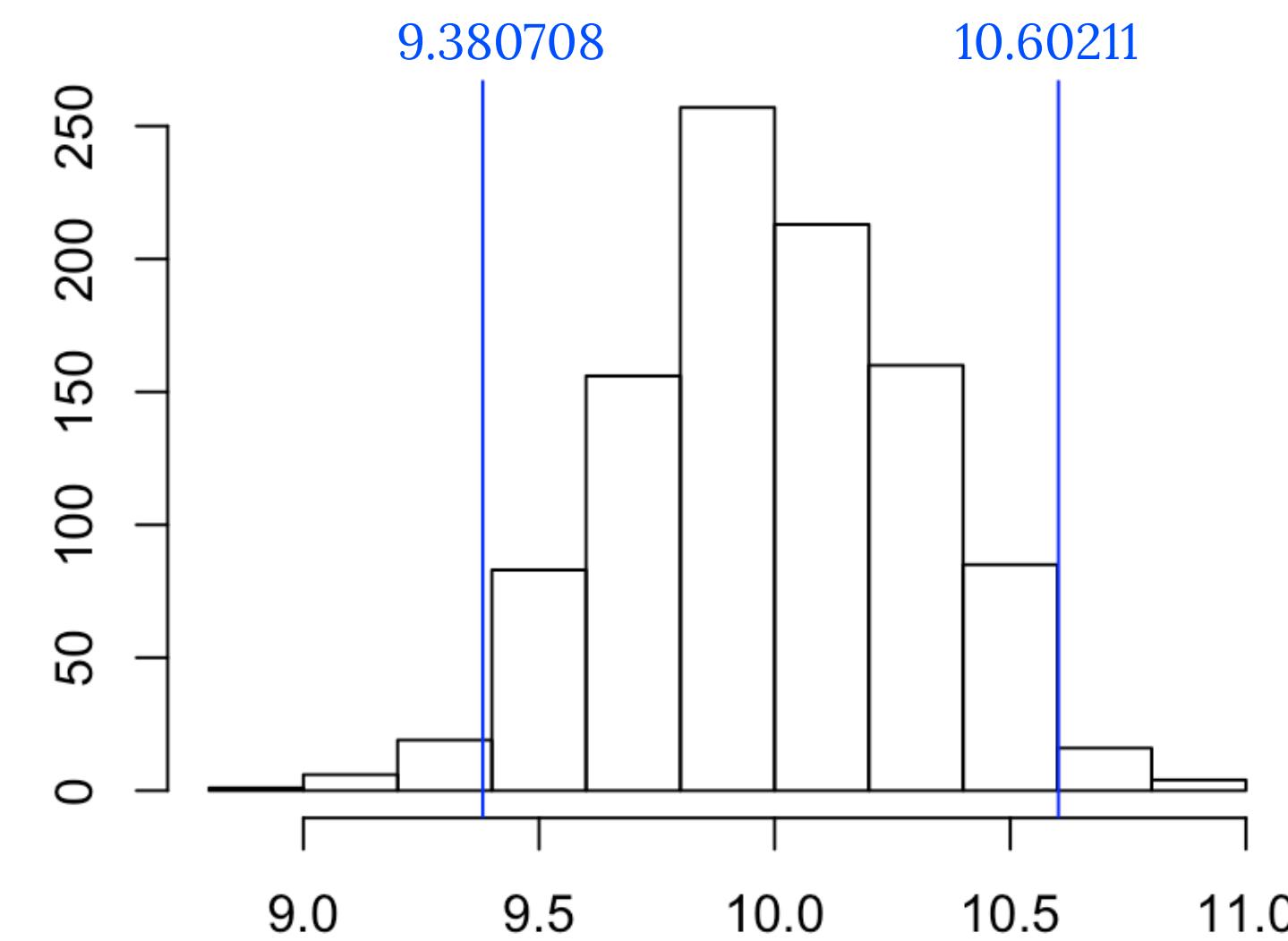
Выборка #3

$$M = 10.1, S^2 = 1$$

... и так возьмем 1000 выборок

Построим из средних
распределение

Распределение средних выборок, взятых из ген. совокупности
с параметрами $\mu = 10, \sigma = 1$



Для выборочного среднего можно посчитать 95% доверительный интервал (ДИ). Согласно ЦПТ, в этот интервал должно попадать среднее ген. совокупности.

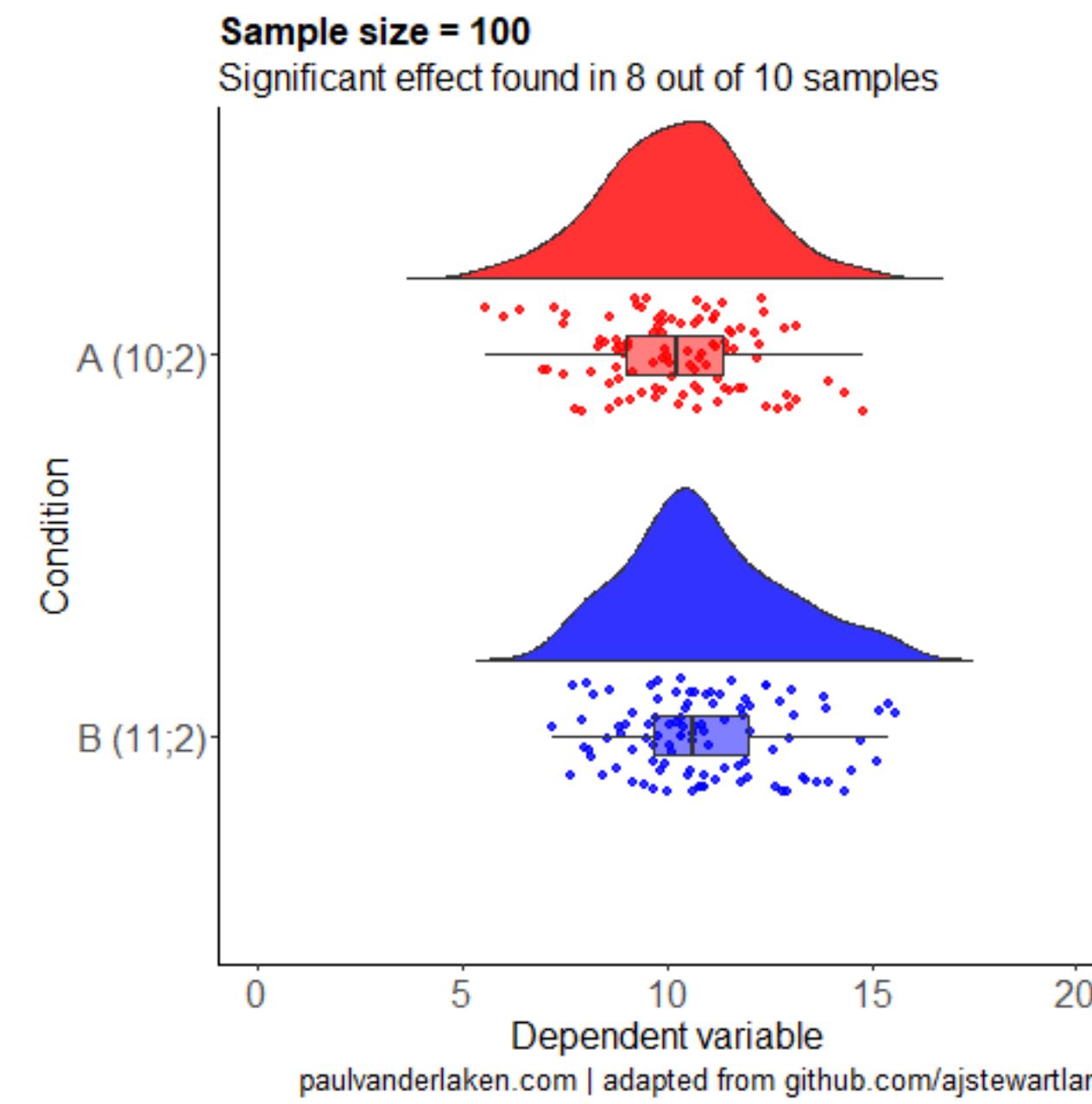
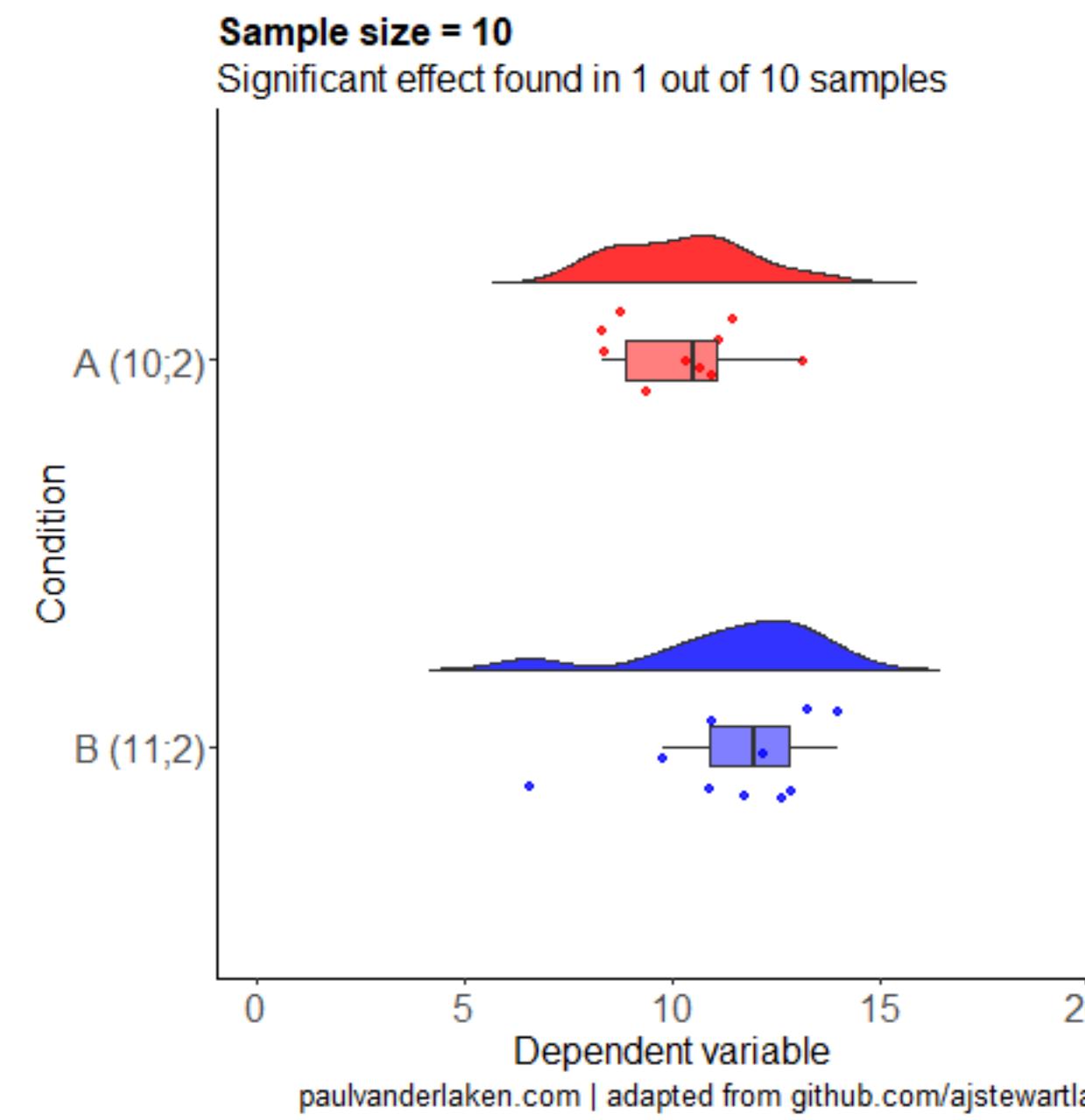
Чтобы посчитать 95% ДИ, можно воспользоваться ранее приведенной формулой для расчета *стандартной ошибки (Standard Error)*:

$$SE = \frac{SD}{\sqrt{n}}$$

$$M \pm 1,96 \times (SD \div \sqrt{n}) = 10 \pm 1,96 * (1 \div \sqrt{10000}) = [9.380708; 10.60211]$$

Это, кстати, тоже самое, если бы мы посчитали стандартное отклонение по всем средним выборок, взятых из ген. совокупности

А вот так будет влиять размер выборки на выборочное среднее



Демонстрация

e^xperiment fest

ЦПТ применяется в *Google Analytics* при построении отчетов



- Увеличивается скорость построения отчетов, при этом результат агрегатов сохраняется (конверсия у целей не сильно изменится)
- Экономия за счет выделения меньшего объема вычислений

ЦПТ очень сильный инструмент, но в то же время сложный.
ЦПТ нельзя просто так использовать для сравнения выборок,
необходимо соблюдать ряд предположений, например
наблюдения случайной величины должны быть независимы

Например, это предположение легко нарушается при
сравнении нового и старого сайта, потому что старая
аудитория увидит оба варианта

Итого

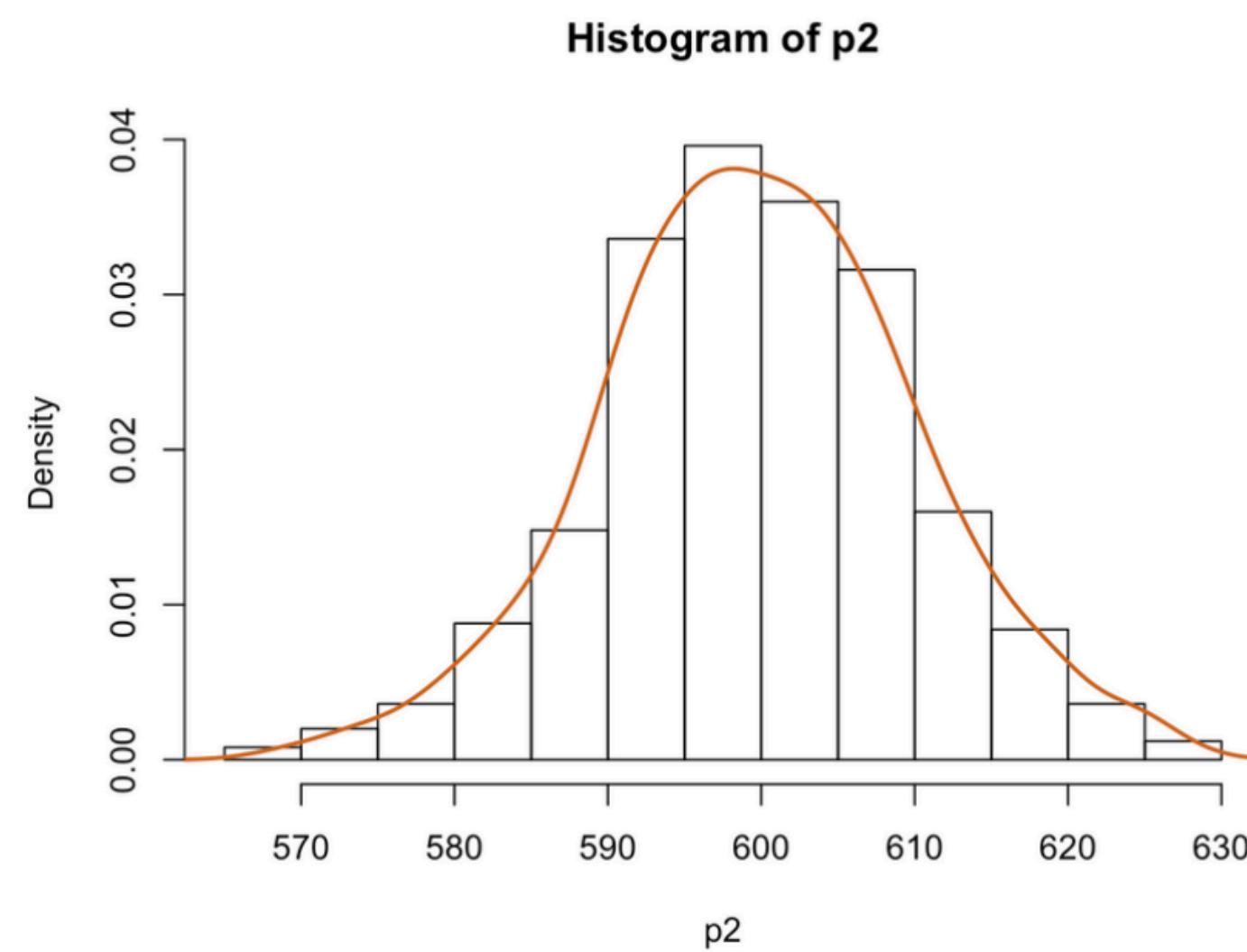
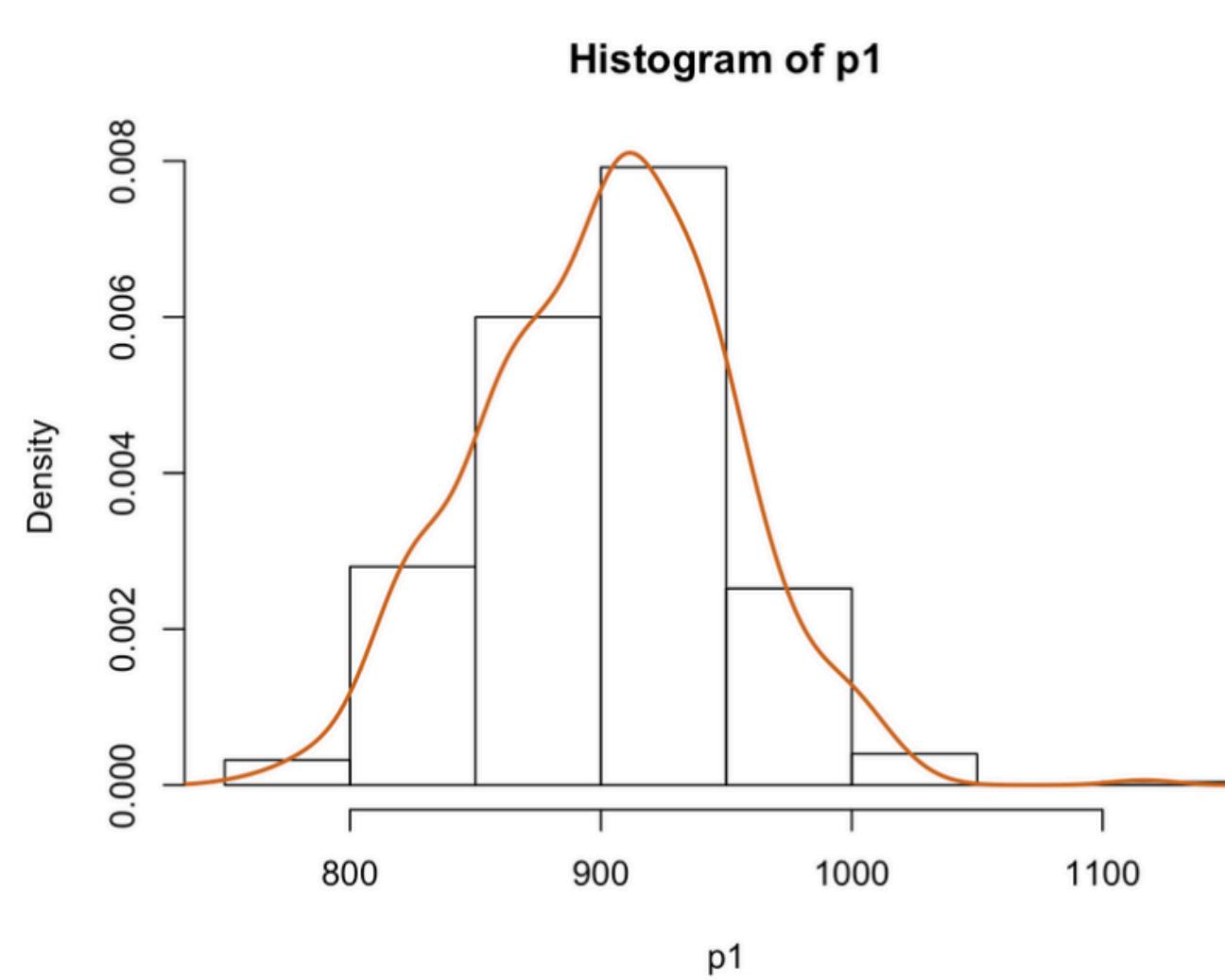
- Чтобы измерить изменчивость (шум) в данных, используют дисперсию, стандартное отклонение (а еще можно интерквартильный размах, куртозис и другие меры разброса)
- В статистике используют дискретные и непрерывные случайные величины с биномиальным и нормальным распределениями
- Чтобы оценить среднее ген. совокупности используют ЦПТ
- Доверительные интервалы используют, чтобы узнать как может распределиться выборочное среднее

Квиз

Быстрое введение в математическую статистику

e^xperiment fest

1. Где дисперсия больше?



Домашнее задание

Где найти описание ДЗ? – Описание заданий доступно по приватной ссылке в Notion (ссылка видна только тем, кому выслан доступ на почту, требуется регистрация)

Что если моя почта не позволяет зарегистрироваться в Notion? – Напишите организаторам интенсива на info@experiment-fest.ru ваши почты на изменение в формате: старая – новая

Где сдавать ДЗ? – Сдача ДЗ в Google Forms. В Notion доступны ссылки на формы ДЗ №1,2,3,4. В каждом ДЗ есть опрос (факт-чекинг) + кейс, который надо оформить в Google Docs и дать доступ с возможностью комментирования на info@experiment-fest.ru

Как решать кейс? – Ограничений по использованию инструмента нет. Вы можете воспользоваться специальным разработанным тренажером, если не умеете программировать или нет возможности использовать Excel

Как получить обратную связь по ДЗ? – Задать вопросы прямо в Google Forms + преподаватели комментируют кейсы в Google Docs (не забудьте открыть доступ на комментирование на почту info@experiment-fest.ru)

В течение какого времени проверяется ДЗ? – Любое ДЗ можно сдать до 14 августа 23:59 МСК. Проверка занимает 2-3 дня

Что если у меня остались вопросы по тому как выполнять ДЗ? – Обсуждение ДЗ в телеграм-чате интенсива

Демонстрация калькулятора

Конец первого дня

e^xperiment fest

Мирмахмадов Искандер

Черемисинов Виталий

07/2020

experiment-fest.ru