

Статистические инструменты для проведения A/B-теста

Продвинутые инструменты для работы с ненормальным распределением

Продвинутые инструменты для работы
с ненормальным распределением

Цель урока

Изучить продвинутые инструменты
для работы с ненормальными
распределениями и структурировать
знания об изученных ранее критериях.

Продвинутые инструменты для работы
с ненормальным распределением

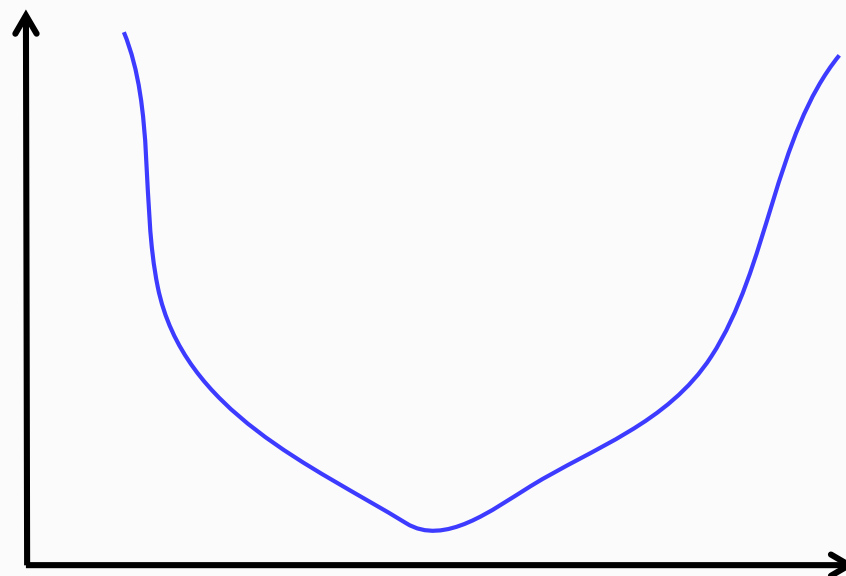
Задачи урока

- ✓ Рассмотреть метод «бакетинг»
- ✓ Рассмотреть метод «бутстрап»
- ✓ Разобрать, на каких распределениях данных лучше работает каждый критерий
- ✓ Узнать, какой объём данных подавать для критерия
- ✓ Понять, какая метрика подходит под критерий
- ✓ Сравнить критерии

Продвинутые инструменты для работы
с ненормальным распределением

Пример, когда предыдущие способы не помогают и только бакетинг и бутстрап спасают

Такое распределение
не получится так просто привести
к нормальному. Мы убедились
в этом на прошлой практике.



Продвинутые инструменты для работы
с ненормальным распределением

Бакетинг

Допущения: независимые
наблюдения между и внутри
бакетов.

Ограничения:

- необходимо минимум 2 500 юзеров в каждой группе
- обязателен user_id



Бакетинг

- 1 Рассчитываем метрику по пользователям
- 2 Разбиваем случайно пользователей на бакеты
- 3 Рассчитываем агрегированное значение метрики для каждой — это новая метрика



Бакетинг

- 4 Если новая метрика имеет **ненормальное** распределение в каждой из групп, то повторяем шаги 2–4
- 5 Сравниваем средние значения новой метрики по группам t-тестом



Продвинутые инструменты для работы
с ненормальным распределением

Бакетинг: сколько бакетов брать?

- Слишком мало бакетов — плохо. Тест на нормальность будет с маленькой мощностью
- Слишком много бакетов — плохо. Меньше шанс получить нормальное распределение

Но чем больше бакетов, тем меньше информации теряется!

Бакетинг: сколько бакетов брать?

Алгоритм выбора количества бакетов:

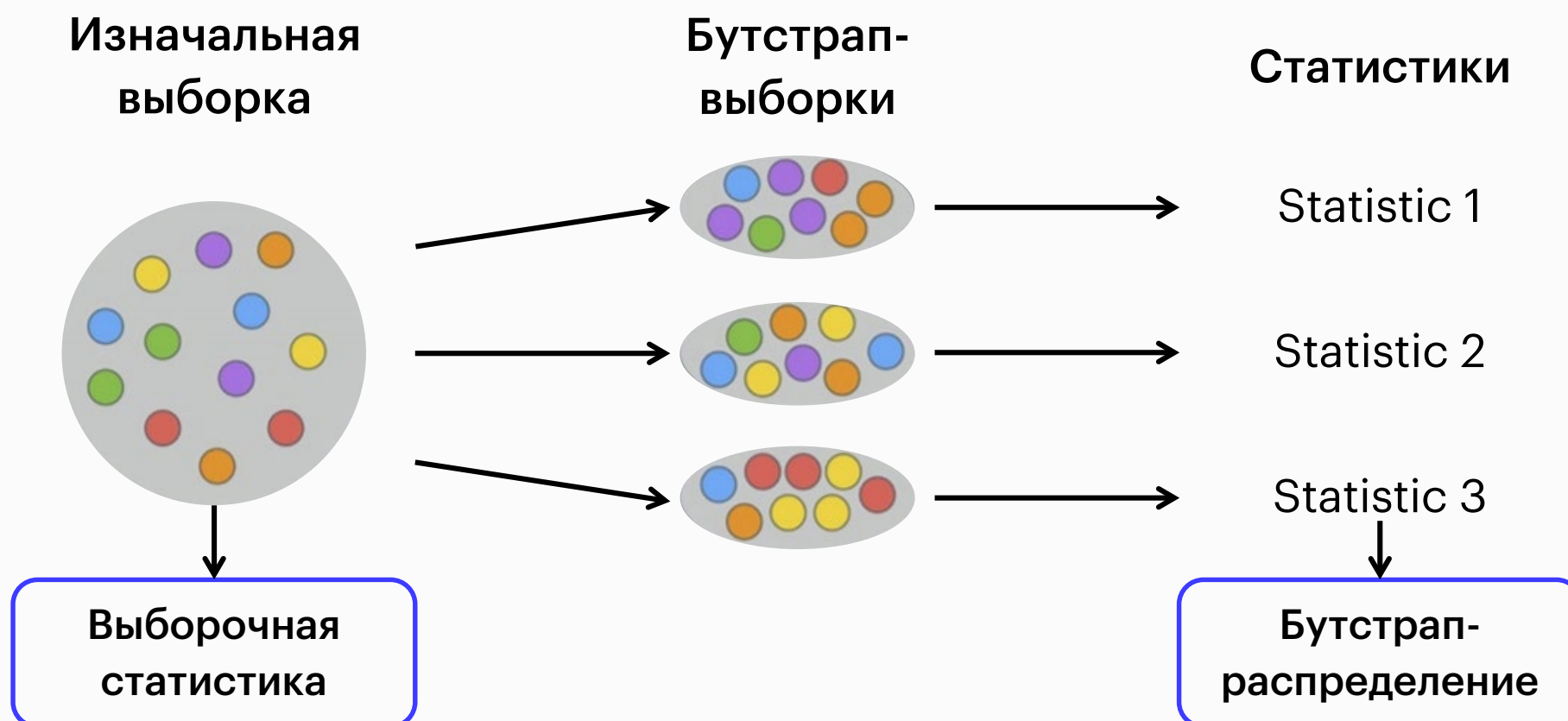
- 1 Создаём много выборок, между которыми:
 - не должно быть значимой разницы
 - должна быть значимая разница
- 2 Проводим тесты с разным количеством бакетов
- 3 Измеряем точность для каждого количества бакетов
- 4 Выбираем оптимальное количество бакетов

Продвинутые инструменты для работы
с ненормальным распределением

Бутстрап

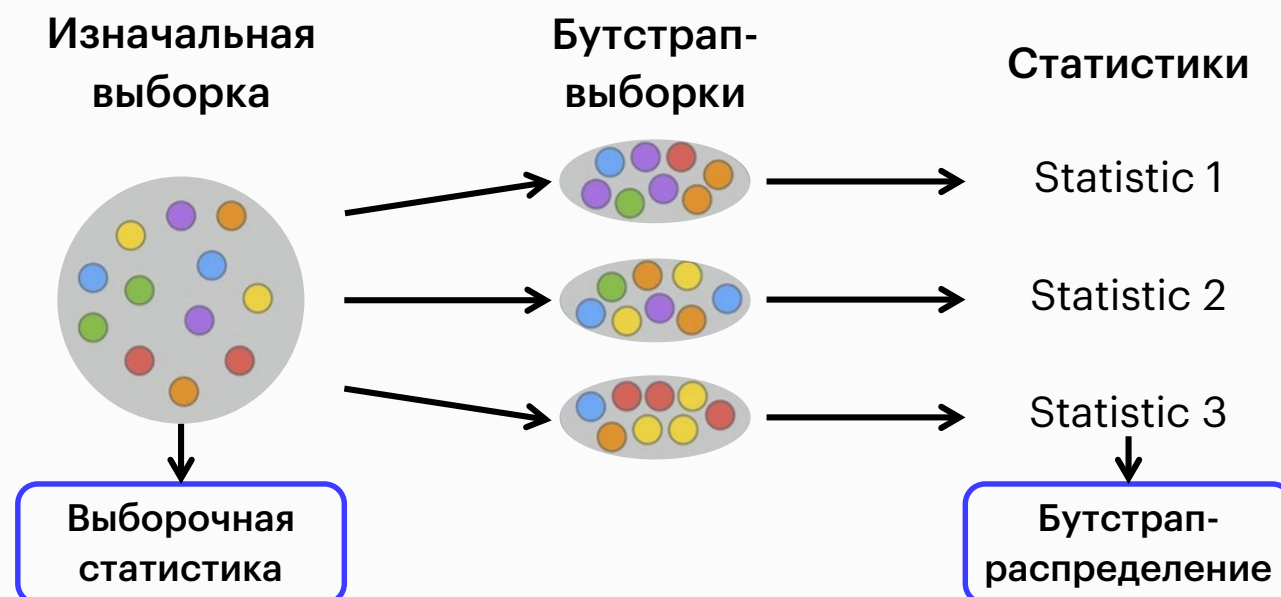
Допущения: независимые наблюдения
между и внутри групп.

Ограничения: менее надёжный при тяжёлых
или длинных хвостах распределения.



Бутстрап

- 1 Рассчитываем наблюдаемое различие между группами А и Б
- 2 Преобразовываем значения метрик в группах так, чтобы у них было одинаковое среднее
- 3 Создаём случайные выборки из преобразованных метрик
- 4 Считаем метрику на каждой выборке
- 5 Строим распределение метрики после бутстрапа для каждой группы
- 6 Сравниваем распределения



Бутстрап: кейс

Есть результаты тестирования методом «бутстрап» при разных α для каждой из групп.

Группа А: интервал 95 %: [0.7; 0.75]

Группа Б: интервал 99 %: [0.71; 0.79]

Различие не значимо?

Бутстрап: кейс

Есть результаты тестирования методом «бутстрап» при разных α для каждой из групп.

Группа А: интервал 95 %: [0.7; 0.75]

Группа Б: интервал 99 %: [0.71; 0.79]

Различие не значимо?

Ответ: сравнивать интервалы, соответствующие разным α , нет смысла.

Продвинутые инструменты для работы
с ненормальным распределением

Бутстрап: кейс

Теперь берём одинаковые α . Из интервалов видно, что конверсия в тестовой группе всё-таки больше.

Группа А: интервал 95 %: [0.7; 0.75]

Группа Б: интервал 95 %: [0.74; 0.77]

Релизим фичу?

Бутстрап: кейс

Теперь берём одинаковые α . Из интервалов видно, что конверсия в тестовой группе всё-таки больше.

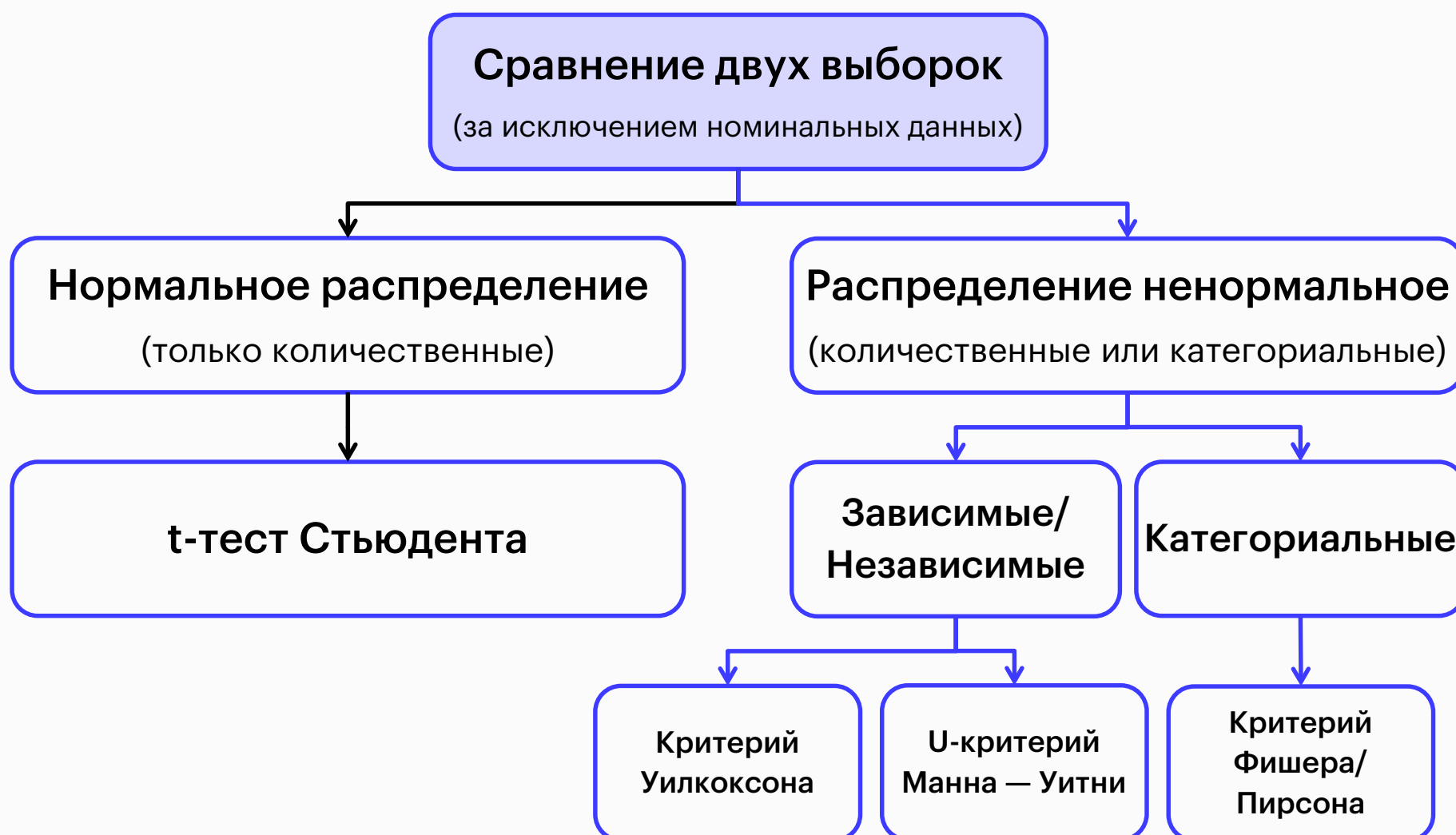
Группа А: интервал 95 %: [0.7; 0.75]

Группа Б: интервал 95 %: [0.74; 0.77]

Релизим фичу?

Ответ: нет. Результаты не значимые, так как есть пересечение.

Карта статистических тестов гипотез



Когда и какие методы используем?

Распределение	Метод агрегации значений метрики				
	График	Размер выборки	Общее	Среднее по юзерное	
	Симметричное, лёгкие хвосты		< 2 500	Бутстрап	Бутстрап ~ t-тест, ранговый
	Симметричное, лёгкие хвосты	> 2 500	Бакетинг, бутстрап	t-тест, бакетинг, бутстрап, ранговый	
	Тяжёлые/длинные хвосты (перекошенное распределение)		< 2 500	Бутстрап	Ранговый, бутстрап
	Тяжёлые/длинные хвосты (перекошенное распределение)	> 2 500	Бакетинг, бутстрап	Бакетинг, ранговый бутстрап, t-тест	
	Мультимодальное		< 2 500	Бутстрап	Бутстрап, ранговый
Мультимодальное	> 2 500	Бакетинг, бутстрап	Бакетинг, бутстрап, ранговый, ~ t-тест		

Продвинутые инструменты для работы
с ненормальным распределением

Итоги урока

- ✓ Мы познакомились ещё с парой современных методов, хорошо показывающих себя в работе с нестандартными распределениями в данных
- ✓ Разобрали основной набор критериев
- ✓ Обсудили область применения каждого из них