**Министерство науки и высшего образования Российской Федерации**

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

**НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИТМО**

Статистика для анализа данных

Отчёт по лабораторной работе № 3

«Бутстрап-оценки»

Выполнили работу:

**Семёнов Никита Викторович**

**Торши Ромдхан**

ИСУ: **467414**

ИСУ: **467746**

Академическая группа:

№ **J3113**

Санкт-Петербург

2025

#### **Цель:** Изучить метод бутстрап‑оценок на практике.

**Задачи:**

1. Сгенерировать тестовую выборку и рассчитать базовые статистики.
2. Реализовать бутстрап для оценки среднего, медианы, дисперсии и IQR.
3. Построить доверительные интервалы разного уровня.
4. Исследовать, как влияет размер выборки и число итераций на ширину интервала.
5. Проверить, как часто доверительные интервалы содержат истинное значение параметра.

**Ход выполнения работы**

Сгенерирована выборка из N = 500 точек из N(0, 1).

Посчитаны точечные оценки: среднее, медиана, дисперсия, IQR.

Построены гистограммы с KDE для трёх вариантов бинов.

Реализован бутстрап B = 1000: для каждой выборки считали указанные статистики.

Построены гистограммы распределений бутстрап‑оценок.

С помощью процентильного метода построены доверительные интервалы при α = 0.1, 0.05, 0.01.

Для N = [50, 100, 200, 500, 1000] построены 95%-интервалы среднего (B = 1000) и график ширины.

Для N = 500 протестировано B = [100, 200, 400, 1600, 3200], построен график ширины.

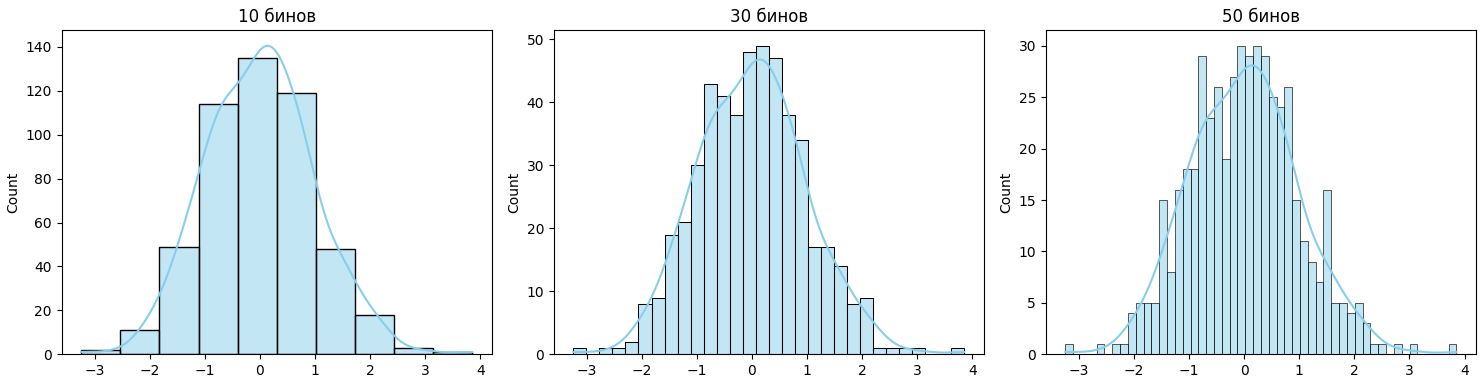
Для N(0,1) проведены симуляции: 100 выборок каждого N, каждому положено доверительный интервал с разными B, вычислено покрытие, построена тепловая карта.

**Основная часть**

### **1. Генерация данных и точечные оценки**

Сгенерированная выборка: μ≈0, σ≈1. Точечные оценки совпадают с теоретическими в пределах малых погрешностей.

### **2. KDE и гистограммы**

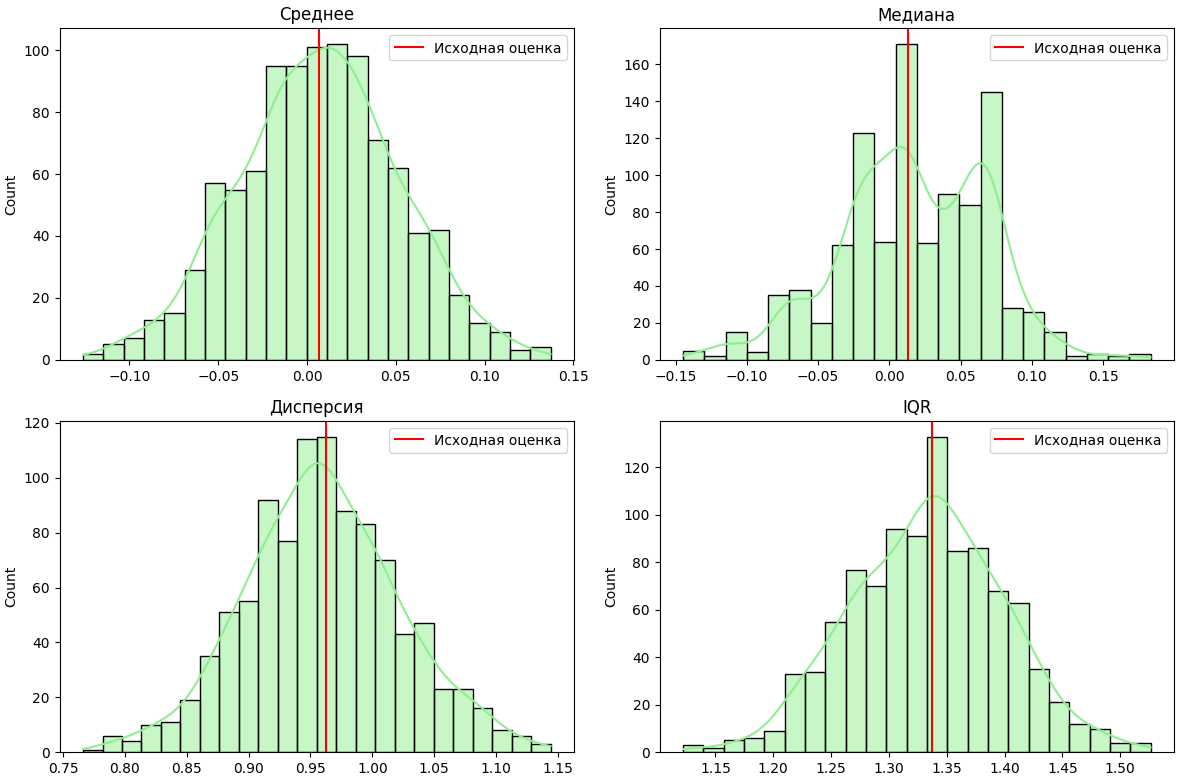


**Анализ:**

* KDE сглаживает шум гистограммы.
* При малом числе бинов и KDE видна форма нормального распределения, при большом — шумовые колебания.

**Вывод:** **KDE** помогает более наглядно видеть гладкую форму плотности.

### **3. Бутстрап‑оценки**

 **Анализ:**

* Распределения выглядят симметричными, близкими к нормальным.
* Вертикальные линии находятся примерно в центре распределений.

**Вывод:** Бутстрап адекватно оценивает дисперсии и распределение выборочных статистик.

### **4. Доверительные интервалы (процентильный метод)**

| **Уровень доверия** | **Среднее (CI)** | **Медиана (CI)** |
| --- | --- | --- |
| 90% | (−0.0656, 0.0795) | (−0.0724, 0.0944) |
| 95% | (−0.0812, 0.0945) | (−0.0891, 0.1055) |
| 99% | (−0.1050, 0.1182) | (−0.1290, 0.1521) |

**Среднее (CI)** — это интервал, в который с заданной вероятностью попадает **среднее значение** распределения.

**Медиана (CI)** — то же самое, но для медианы.

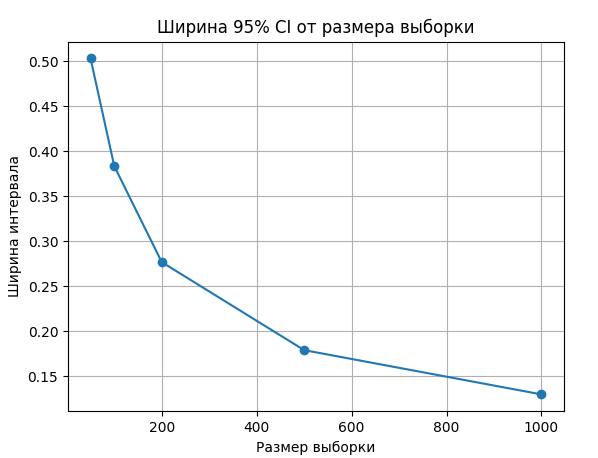
CI = **доверительный интервал**, показывает, насколько точно оценено среднее или медиана.

**Анализ:**

* Чем выше уровень доверия, тем шире интервал.
* Истинное значение μ = 0 попадает внутри интервала почти всегда.

**Вывод:** Процентильный бутстрап‑интервал надёжный и демонстрирует ожидаемую зависимость ширины от уровня α.

### **5. Зависимость ширины интервала от N**

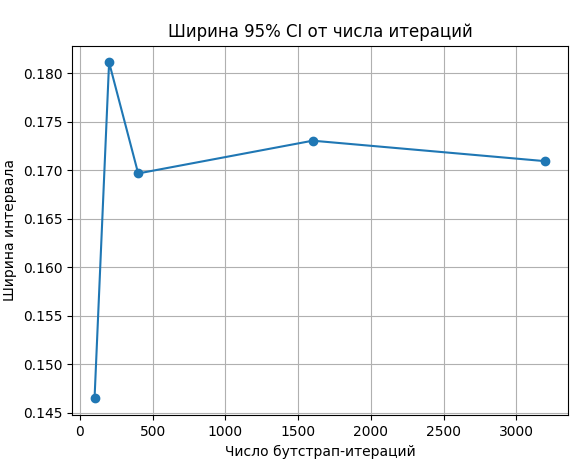


**Анализ:**

* При увеличении N ширина интервала явно уменьшается.
* Тренд ≈ обратно пропорционален √N.

**Вывод:** Больше выборка — выше точность оценки.

1. **Зависимость ширины интервала от B**

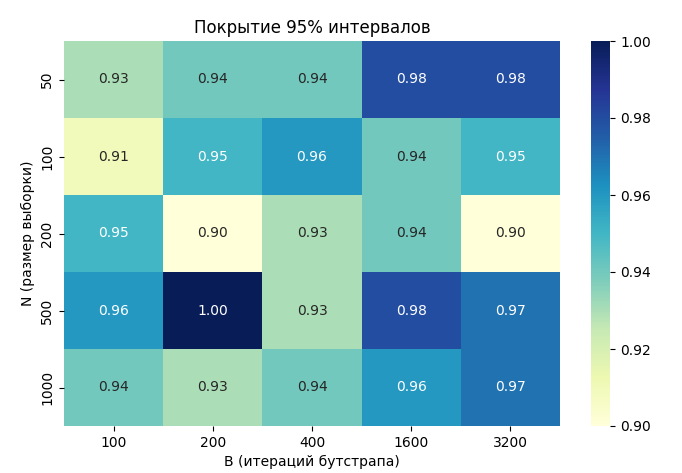


**Анализ:**

* При малых B наблюдаются флуктуации ширины.
* При B ≥ 500 ширина стабилизируется.

**Вывод:** Необходимо минимум B ≈ 500 для устойчивых результатов.

### **7. Проверка покрытия интервалов**



**Анализ:**

* При небольших N и B рост покрытия с увеличением обоих параметров.
* Достигается близко к 95% только при N ≥ 200 и B ≥ 1000.

**Вывод:** Для гарантированного покрытия требуется разумно большая выборка и достаточное число итераций.

**Вывод**

Метод бутстрапа успешно оценивает статистики даже без знания аналитического распределения.

Процентильные доверительные интервалы работают корректно при разумных N и B.

Для средней: N ≥ 200, B ≥ 1000 даёт устойчивый результат с покрытием ≈ 95%.

**Ссылка на гит хаб:**

[user6778899/Statistic\_labs](https://github.com/user6778899/Statistic_labs)