**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 № 4

«Дисперсионный анализ»

Выполнили работу:

**Семёнов Никита Викторович**

**Торши Ромдхан**

ИСУ: **467414**

ИСУ: **467746**

Академическая группа:

№ **J3113**

Санкт-Петербург

2025

#### **Цель работы**

Освоить методы парного тестирования и однофакторного дисперсионного анализа (ANOVA) для проверки гипотез о равенстве средних значений между группами данных. Научиться интерпретировать результаты статистических тестов и анализировать их применимость и информативность.

**Ход выполнения работы**

1. Были сгенерированы две выборки из нормального распределения, каждая из которых включала три группы:

* В первой выборке группы имели **близкие** математические ожидания и одинаковую дисперсию.
* Во второй выборке группы имели сильно **отличающиеся** математические ожидания и одинаковую дисперсию.

1. Для каждой выборки были построены графики ядерной оценки плотности распределения (KDE) по группам и для объединённых данных.
2. Для каждой пары групп были выполнены два вида парных тестов:

* Z-тест (при известных дисперсиях).
* T-тест (при неизвестных, но равных дисперсиях).

1. Для всей выборки был проведён однофакторный дисперсионный анализ (ANOVA) с использованием F-критерия.
2. Было выполнено сравнение методов по вычислительной эффективности и полноте предоставляемой информации.

**Основная часть**

### **Краткие определения:**

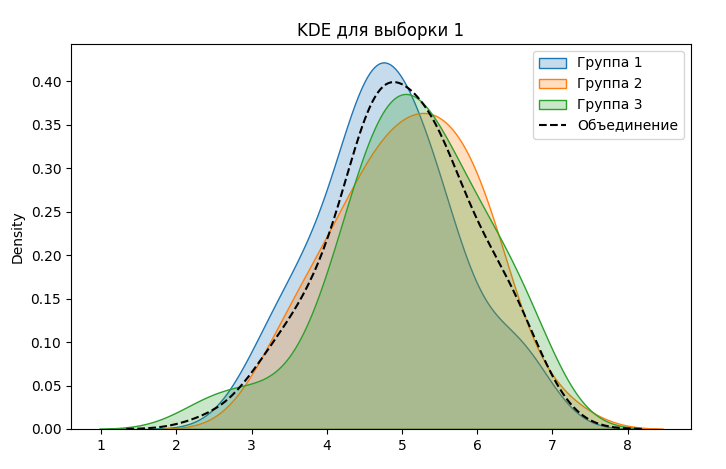
**Парные тесты** — это методы статистического сравнения двух групп с целью проверки, отличаются ли их средние значения. Тесты бывают:

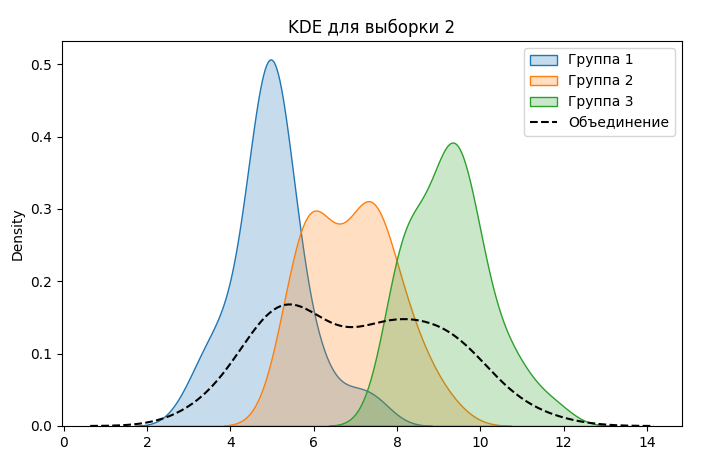
* **Z-тест** — применяется, когда известны дисперсии сравниваемых выборок.
* **T-тест** — применяется, если дисперсии неизвестны, но считаются равными. Более универсален и часто используется в реальных задачах.

**ANOVA (однофакторный дисперсионный анализ)** — это метод проверки гипотезы о равенстве средних значений в нескольких группах. Вместо сравнения пар групп, ANOVA позволяет сразу определить, есть ли **хотя бы одно** статистически значимое отличие между группами. Основан на сравнении **внутригрупповой** и **межгрупповой** изменчивости (дисперсий).

**Нулевая гипотеза:**  
**H₀: μ₁ = μ₂**  
(то есть средние значения в *двух* группах **одинаковы**)

**Графики и Анализ**

  
На графике представлены три сглаженных распределения для групп выборки 1. Их пики и форма ***почти*** совпадают, что визуально подтверждает схожесть средних значений. Линия объединения (пунктир) показывает общую форму распределения всех данных.  
Различия между группами ***незначительны***, ожидания близки.

  
Здесь пики KDE-графиков заметно различаются — каждая группа имеет своё выраженное математическое ожидание.  
Между группами явные различия в средних значениях.

### **Результаты парных тестов**

**Выборка 1:**

| **Пары** | **Z** | **p(Z)** | **t** | **p(t)** | **Вывод** |
| --- | --- | --- | --- | --- | --- |
| 1 vs 2 | –1.03 | 0.3011 | –1.13 | 0.2634 | нет различий (малые) |
| 1 vs 3 | –1.17 | 0.2437 | –1.23 | 0.2233 | нет различий (малые) |
| 2 vs 3 | –0.13 | 0.8951 | –0.14 | 0.8915 | нет различий |

**Выборка 2:**

| **Пары** | **Z** | **p(Z)** | **t** | **p(t)** | **Вывод** |
| --- | --- | --- | --- | --- | --- |
| 1 vs 2 | –7.46 | <0.001 | –7.70 | <0.001 | значимые различия |
| 1 vs 3 | –16.65 | <0.001 | –17.94 | <0.001 | значимые различия |
| 2 vs 3 | –9.19 | <0.001 | –9.32 | <0.001 | значимые различия |

**p(Z)** - вероятность получить наблюдаемое (или более экстремальное) значение Z-статистики.

**Результаты ANOVA**

| **Выборка** | **F** | **p-value** | **Вывод** |
| --- | --- | --- | --- |
| 1 | 0.92 | 0.4027 | нет оснований отвергать H₀ |
| 2 | 150.54 | <0.001 | средние существенно различны |

В выборке 1 ни парные тесты, ни ANOVA не нашли различий (p > 0.05).

В выборке 2 все тесты дали p < 0.001: группы статистически различаются.

ANOVA удобно проверяет сразу все группы, парные тесты — для уточнения конкретных пар.

Для практики: используйте методы из библиотек для квантилей N(0,1), t и F‑распределений, а альтернативная гипотеза делайте двусторонней.

**Вывод**

В работе проверили равенство средних в двух выборках с помощью парных тестов и ANOVA. В первой выборке группы не отличаются, а во второй — различия значимые. Это подтвердило, что дисперсионный анализ и парные тесты помогают находить отличия между группами.

**Ссылка на гит хаб:**

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