

Санкт-Петербургский Политехнический университет Петра
Великого

Отчет по лабораторной работе №7

Определение систематического сдвига в данных

Студент:	Швачко Никита Андреевич
Преподаватель:	Баженов Александр Николаевич
Группа:	5030102/20202

Санкт-Петербург 2025

1 Постановка задачи

Цель лабораторной работы — проверить гипотезу о законе распределения генеральной совокупности. Для этого были сгенерированы две выборки X_1 и X_2 , обладающие различными средними и стандартными отклонениями:

$$X_1 = N(0, 0.95), \quad X_2 = N(1, 1.05)$$

Задача заключалась в том, чтобы варьировать параметр сдвига a таким образом, чтобы проверить изменения индексов Жаккара для ****внутренних**** и ****внешних**** интервалов. Для каждой выборки были рассчитаны интервалы:

- Внутренний интервал: $[Q_{1/4}, Q_{3/4}]$ — первый и третий квартили.
- Внешний интервал: $[\min X_i, \max X_i]$ — минимальное и максимальное значение выборки.

В результате работы были найдены оценки сдвига a , при которых индексы Жаккара $J_{\text{Inn}}(a)$ и $J_{\text{Out}}(a)$ достигают максимума.

2 Описание используемых методов

Для решения задачи был использован следующий подход:

1. **Генерация выборок:** С использованием библиотеки NumPy были сгенерированы две выборки X_1 и X_2 по нормальному распределению с заданными параметрами.
2. **Расчет интервалов:** Для каждой выборки были рассчитаны внутренние и внешние интервалы, включающие первый и третий квартили для внутреннего интервала и минимальное и максимальное значение для внешнего интервала.
3. **Индекс Жаккара:** Для определения степени схожести выборок использовался индекс Жаккара. Этот индекс вычисляется как отношение длины пересечения интервалов к длине их объединения.
4. **Определение сдвига:** Параметр сдвига a изменялся в диапазоне от -2 до 4, и для каждого значения сдвига рассчитывались индексы Жаккара для внутренних и внешних интервалов. Искались значения a , при которых эти индексы достигают максимума.
5. **Визуализация:** Для представления результатов был построен график зависимостей индексов Жаккара от значения сдвига.

3 Результаты эксперимента

В результате эксперимента были получены следующие данные:

- **График:** На графике отображены зависимости $J_{\text{Inn}}(a)$ и $J_{\text{Out}}(a)$ от значения сдвига a . Максимальные значения индексов Жаккара соответствуют значениям сдвига a_{Inn} и a_{Out} .

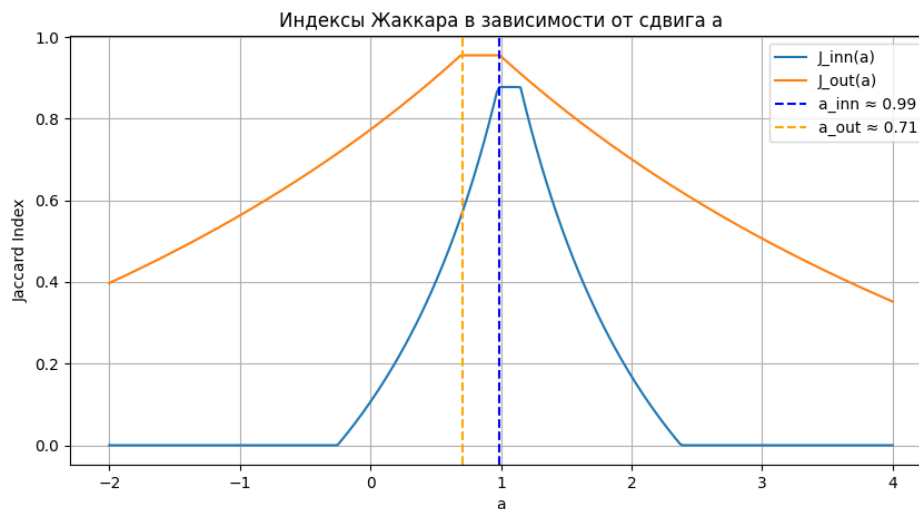


Рис. 1: Возмущённая выборка: влияние выбросов

- **Оценка сдвига:**

- Оценка сдвига для внутреннего интервала: $a_{inn} \approx 0.99$.
- Оценка сдвига для внешнего интервала: $a_{out} \approx 0.71$.

Таким образом, на основе индекса Жаккара были определены параметры сдвига для двух выборок.

4 Выводы

В ходе лабораторной работы были исследованы изменения индексов Жаккара при варьировании сдвига между выборками, что позволило точно оценить параметры сдвига a . Полученные результаты показывают, что метод индекса Жаккара эффективно позволяет измерить степень схожести между выборками и определить оптимальный сдвиг для их согласования.