

САНКТ-ПЕТЕРБУРГСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ ИМЕНИ ПЕТРА ВЕЛИКОГО

ФИЗИКО-МЕХАНИЧЕСКИЙ ИНСТИТУТ

ВЫСШАЯ ШКОЛА ПРИКЛАДНОЙ МАТЕМАТИКИ И ФИЗИКИ

### **Интервальный анализ**

**Курсовая работа по теме "Оценочная функция Тейла-Сена"**

Выполнил:

Студент: Рубанова Валерия

Группа: 5030102/00201

Принял:

к. ф.-м. н., доцент

Баженов Александр Николаевич

Санкт-Петербург  
2023 г.

## Содержание

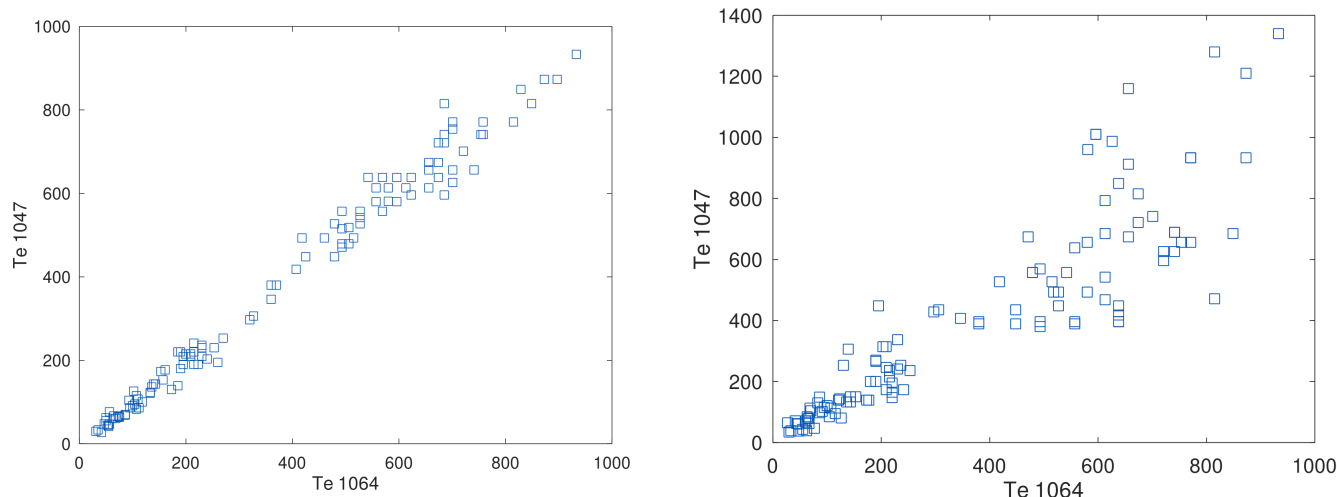
1	Постановка задачи	2
2	Теория	2
3	Реализация	2
4	Результаты	2
5	Выводы	5

## 1 Постановка задачи

Необходимо вычислить на двух наборах данных значение оценочной функции Тейла-Сена, а также сделать внутренние и внешние оценки наклона.

Построить гистограмму наклонов, задать каким-либо доверительный уровень и найти по нему оценку вариабельности медианы.

Исходные данные выглядят следующим образом:



## 2 Теория

В непараметрической статистике существует метод для робастного линейного сглаживания множества точек (простая линейная регрессия), в котором выбирается медиана наклонов всех прямых, проходящих через пары точек выборки на плоскости. Метод называется оценочной функцией Тейла — Сена.

Эта оценочная функция может быть эффективно вычислена и она нечувствительна к выбросам. Она может быть существенно более точна, чем неробастный метод наименьших квадратов для несимметричных и гетероскедастичных данных и хорошо конкурирует с неробастным методом наименьших квадратов даже для нормально распределенных данных в терминах статистической мощности. Метод признан «наиболее популярной непараметрической техникой оценки линейного тренда».

Как определил Тейл, ценочная функция Тейла — Сена множества точек на плоскости  $(x_i, y_i)$  — это медиана  $m$  коэффициентов наклона  $(y_j - y_i)/(x_j - x_i)$  по всем парам точек выборки. Сен расширил это определение для обработки случая, когда две точки имеют одинаковые координаты  $x$ . По определению Сена медиана коэффициентов наклона берётся только по парам точек, имеющих различные координаты  $x$ .

## 3 Реализация

Реализация метода Тейла-Сена выполнена на языке python с использованием библиотек numpy, pandas, matplotlib, scipy.stats.

## 4 Результаты

Для 1 и 2 столбца исходных данных получены следующие оценки:

Из этих данных можно сделать вывод, что наклоны в выборке изменяются достаточно существенно, от -15.6

Таблица 1: Результаты для 1 измерения

Оценки	Значения
Median Slope	1.006840390879479
Min Slope	-15.600000000000001
Max Slope	23.466666666666667
Mode Slope	1.0

до 23.467. Медианное значение близко к 1.0, что означает, что наклоны вокруг этого значения имеют примерно равное распределение. Наличие отрицательного наклона и высоких положительных наклонов может указывать на наличие различных уклонов в данных.

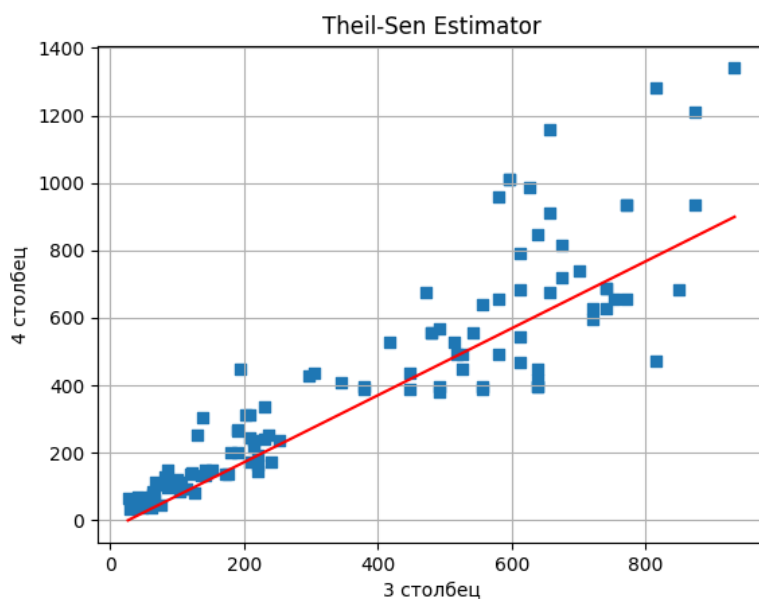
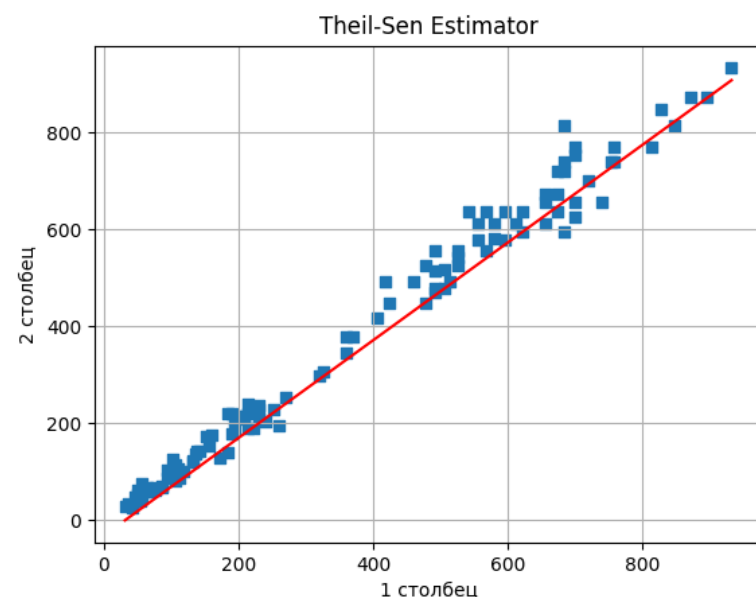
Для 3 и 4 столбца исходных данных получены следующие оценки:

Таблица 2: Результаты для 2 измерения

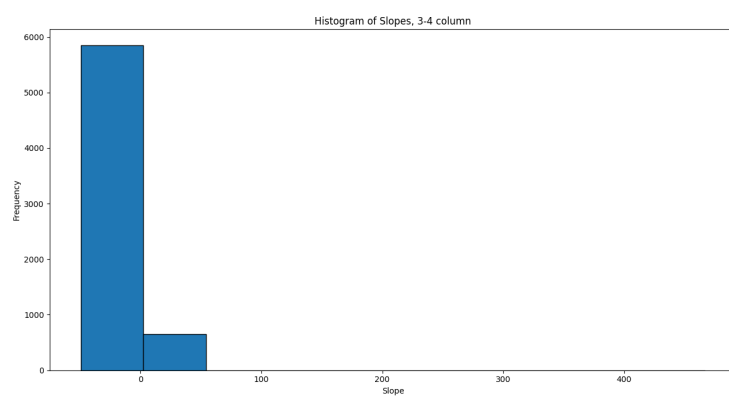
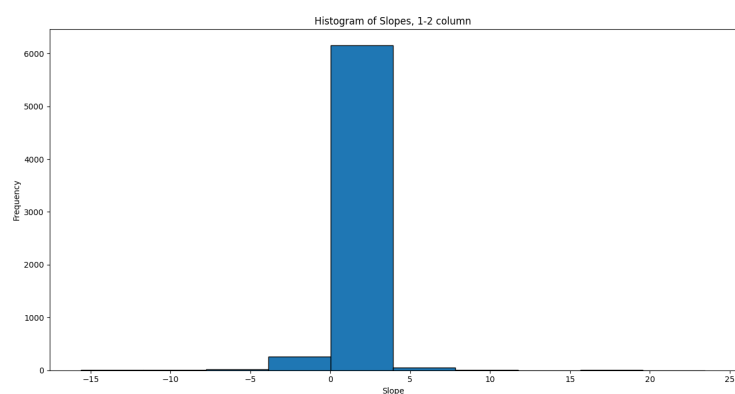
Оценки	Значения
Median Slope	0.9926894959599847
Min Slope	-49.25
Max Slope	467.0
Mode Slope	0.0

Этот анализ даёт общее представление о распределении наклонов, полученных методом Theil-Sen. Median Slope указывает на типичное значение, Min Slope и Max Slope указывают на диапазон значений, а Mode Slope предоставляет информацию о наиболее часто встречающемся значении.

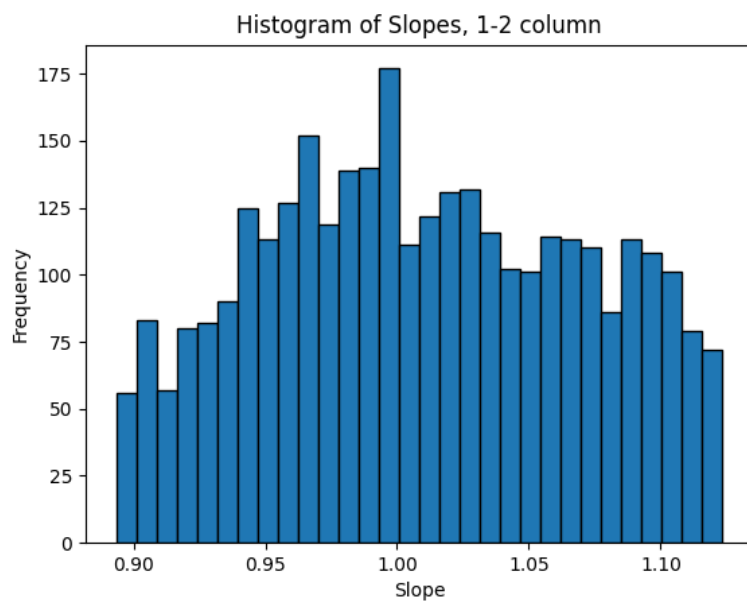
Результат вычисления оценочной функции Тейла-Сена следующий:

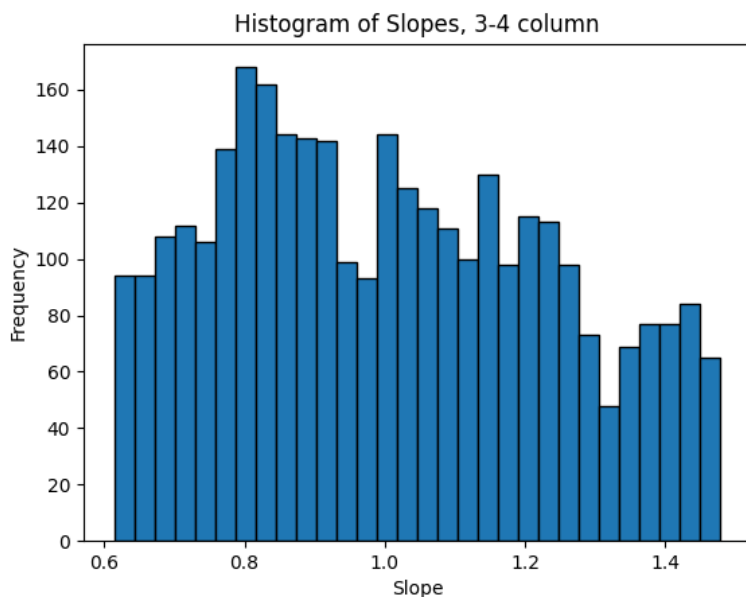


Гистограммы наклонов выглядят следующим образом:

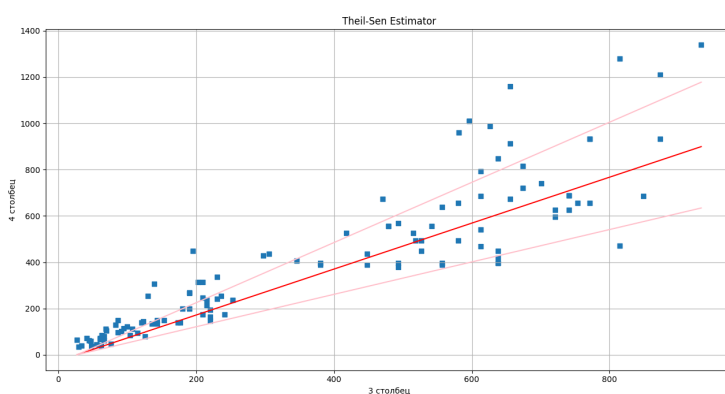
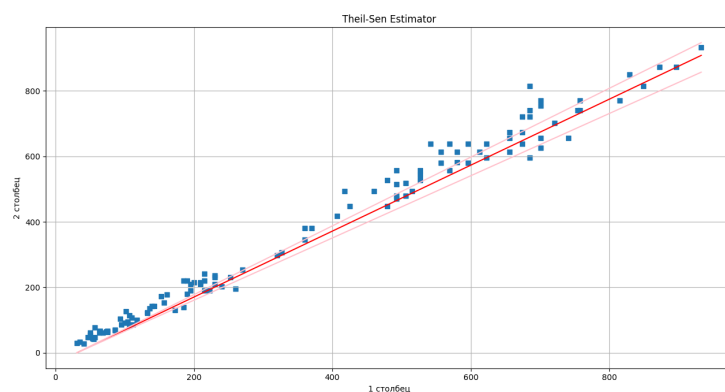


При откидывании 1 и последнего квартилей:





Оценка вариабельности медианы (стандартная ошибка медианы) с 95% доверительным уровнем для 1-2 колонок равна 139.0, для 3-4 – 94.5.  
Внутренние оценки следующие:



## 5 Выводы

В данной курсовой работе был рассмотрен метод оценочной функции Тейла-Сена для робастного линейного сглаживания множества точек. Была выполнена реализация данного метода на языке Python с использованием библиотек numpy, pandas, matplotlib, scipy.stats.

Были получены оценки для двух наборов данных. Для первого набора данных были получены следующие оценки: медианное значение наклона равно 1.006840390879479, минимальное значение наклона составляет -15.6, максимальное значение наклона - 23.46666666666667, модальное значение наклона - 1.0. Для второго набора данных были получены следующие оценки: медианное значение наклона равно 0.9926894959599847, минимальное значение наклона составляет -49.25, максимальное значение наклона - 467.0, модальное значение наклона - 0.0.

Полученные оценки дают представление о распределении наклонов в выборке. Медианное значение близко к единице, что говорит о примерно равном распределении наклонов вокруг этого значения. Наличие отрицательных и высоких положительных значений наклонов может указывать на наличие различных уклонов в данных.

Также был построен график оценочной функции Тейла-Сена для обоих наборов данных. Графики показывают тренд и изменение значений внутри выборки данных.

Таким образом, метод оценочной функции Тейла-Сена является эффективным и робастным способом для анализа данных и оценки линейного тренда.