

Санкт-Петербургский Политехнический Университет Петра
Великого
Физико-механический институт

Высшая школа прикладной математики и вычислительной
физики

Лабораторная работа №4
по дисциплине “Математическая статистика”

Обучающаяся:

А.Д. Балакшина
(группа 5030102/20101)

Преподаватель:

А.Н. Баженов

Санкт-Петербург

2025

Содержание

| | | |
|----------|--|----------|
| 1 | Постановка задачи | 3 |
| 2 | Формализация | 3 |
| 3 | Выполнение работы | 3 |
| 4 | Результаты | 4 |
| 5 | Анализ результатов | 4 |
| 5.1 | Сравнение методов на данных без выбросов | 4 |
| 5.2 | Сравнение методов на данных с выбросами | 4 |
| 6 | Теоретическое обоснование | 5 |
| 7 | Выводы | 5 |

1 Постановка задачи

Найти оценки коэффициентов линейной регрессии $y_i = a + bx_i + \varepsilon_i$, используя 20 точек на отрезке $[-1.8; 2]$ с равномерным шагом равным 0.2. Ошибку ε_i считать нормально распределённой с параметрами $(0, 1)$.

В качестве эталонной зависимости взять $y_i = 2 + 2x_i + \varepsilon_i$. При построении оценок коэффициентов использовать два критерия: критерий наименьших квадратов и критерий наименьших модулей. Прodelать то же самое для выборки, у которой в значения y_1 и y_{20} вносятся возмущения 10 и -10.

2 Формализация

Простая линейная регрессия:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, \dots, n,$$

где x_1, \dots, x_n — заданные числа (значения фактора); y_1, \dots, y_n — наблюдаемые значения отклика; $\varepsilon_1, \dots, \varepsilon_n$ — независимые, нормально распределённые $N(0, \sigma)$ с нулевым математическим ожиданием и одинаковой (неизвестной) дисперсией случайные величины (ненаблюдаемые); β_0, β_1 — неизвестные параметры, подлежащие оцениванию.

Метод наименьших квадратов (МНК):

$$\sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2 \rightarrow \min_{\beta_0, \beta_1}. \quad (1)$$

Метод наименьших модулей:

$$\sum_{i=1}^n |y_i - \beta_0 - \beta_1 x_i| \rightarrow \min_{\beta_0, \beta_1}. \quad (2)$$

3 Выполнение работы

Лабораторная работа выполнена на языке программирования Python 3.12 с использованием библиотек `numpy`, `scipy`, `sklearn`, `matplotlib`.

Были построены модели на данных без и с возмущениями, выполнены вычисления, построены графики (сохранялись в виде картинок `png`). Программа отработала корректно.

4 Результаты

Таблица 1: Результаты для невозмущенной выборки

| | Метод | \hat{a} | $\frac{\hat{a}}{a}$ | \hat{b} | $\frac{\hat{b}}{b}$ |
|-----|-------|-----------|---------------------|-----------|---------------------|
| (1) | МНК | 1.87 | 0.94 | 1.96 | 0.98 |
| (2) | МНМ | 1.65 | 0.83 | 2.12 | 1.06 |

Таблица 2: Результаты для возмущенной выборки

| | Метод | \hat{a} | $\frac{\hat{a}}{a}$ | \hat{b} | $\frac{\hat{b}}{b}$ |
|-----|-------|-----------|---------------------|-----------|---------------------|
| (1) | МНК | 2.01 | 1.01 | 0.53 | 0.26 |
| (2) | МНМ | 1.61 | 0.81 | 2.10 | 1.05 |

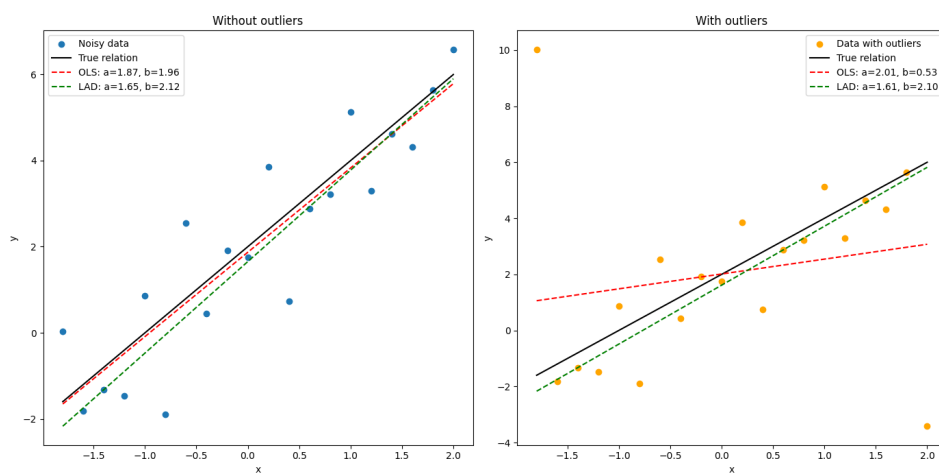


Рис. 1: Результаты вычислительного эксперимента

5 Анализ результатов

5.1 Сравнение методов на данных без выбросов

- МНК показал лучшую точность, чем МНМ.

5.2 Сравнение методов на данных с выбросами

- МНК показал крайнюю чувствительность к выбросам. Оценка b имеет значительное смещение
- МНМ продемонстрировал устойчивость. Оценки изменились незначительно по сравнению со случаем без выбросов.

6 Теоретическое обоснование

Различие в поведении методов объясняется их целевыми функциями:

- МНК минимизирует **сумму квадратов отклонений**, что делает его:
 - Эффективным при нормальном распределении ошибок
 - Чувствительным к выбросам (квадратичный штраф усиливает влияние больших отклонений)
- МНМ минимизирует **сумму абсолютных отклонений**, что:
 - Делает его устойчивым к выбросам
 - Менее эффективным для нормально распределенных ошибок

7 Выводы

- Для данных **без выбросов** предпочтительнее использовать МНК
- Для данных **с выбросами** или для распределений с тяжёлыми хвостами следует применять МНМ.