Федеральное государственное автономное образовательное учреждение высшего образования
Санкт-Петербургский Политехнический университет Петра Великого
Физико-Механический институт

Лабораторная 3

Выполнил студент гр. 5030102/20101:	Бугайцев М.В.	____________
Преподаватель:	Баженов А. Н.	____________
Работа принята:	Дата	____________

Санкт-Петербург
2025

Содержание

1 Введение

Постановка задачи

Простая линейная регрессия

Модель простой линейной регрессии

Регрессионную модель описания данных называют простой линейной регрессией, если

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, \ldots, n, \tag{1}$$

где $x_1, \ldots, x_n$ — заданные числа (значения фактора); $y_1, \ldots, y_n$ — наблюдаемые значения отклика; $\varepsilon_1, \ldots, \varepsilon_n$ — 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

Метод наименьших квадратов (МНК):

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

Расчётные формулы для МНК-оценок

МНК-оценки параметров β_0 и β_1:

$$\hat{\beta}_1 = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\overline{x^2} - (\overline{x})^2}, \tag{3}$$

$$\hat{\beta}_0 = \overline{y} - \overline{x}\hat{\beta}_1. \tag{4}$$

Робастные оценки коэффициентов линейной регрессии

Метод наименьших модулей:

$$\sum_{i=1}^{n} |y_i - \beta_0 - \beta_1 x_i| \to \min_{\beta_0, \beta_1}. \tag{5}$$

Робастные оценки параметров:

$$\hat{\beta}_1^R = r_Q \frac{q_y^*}{q_x^*}, \tag{6}$$

$$\hat{\beta}_0^R = \text{med}\ y - \hat{\beta}_1^R\ \text{med}\ x, \tag{7}$$

где

$$r_Q = \frac{1}{n} \sum_{i=1}^{n} \text{sgn}(x_i - \text{med}\ x)\ \text{sgn}(y_i - \text{med}\ y), \tag{8}$$

$$q_y^* = \frac{y_{(j)} - y_{(l)}}{k_q(n)}, \quad q_x^* = \frac{x_{(j)} - x_{(l)}}{k_q(n)}. \tag{9}$$

Индексы l и j определяются следующим образом:

$$l = \begin{cases} [n/4] + 1 & n/4 \text{ дробное}, \\ n/4 & n/4 \text{ целое}, \end{cases} \quad j = n - l + 1.$$

Функция знака:

$$\text{sgn}(z) = \begin{cases} 1 & z > 0, \\ 0 & z = 0, \\ -1 & z < 0. \end{cases}$$

Уравнение регрессии имеет вид:

$$y = \hat{\beta}_0^R + \hat{\beta}_1^R x. \tag{10}$$

Задание

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. Проделать то же самое для выборки, у которой в значения y_1 и y_{20} вносятся возмущения 10 и -10.

Исходный код лабораторной работы привести на GitHub.

Оформить таблицы:

Таблица 1: Результаты для невозмущенной выборки

Метод	$\hat{a}$	a	$\hat{b}$	b
МНК				
МНМ				

Таблица 2: Результаты для возмущенной выборки

Метод	$\hat{a}$	a	$\hat{b}$	b
МНК				
МНМ				

2 Практическая часть

Таблица 1: Результаты для невозмущенной выборки

Метод	$\hat{a}$	a	$\hat{b}$	b
МНК	2.6	2	1.7	2
МНМ	2.3	2	1.7	2

Таблица 2: Результаты для возмущенной выборки

Метод	$\hat{a}$	a	$\hat{b}$	b
МНК	2.7	2	0.3	2
МНМ	2.3	2	1.3	2

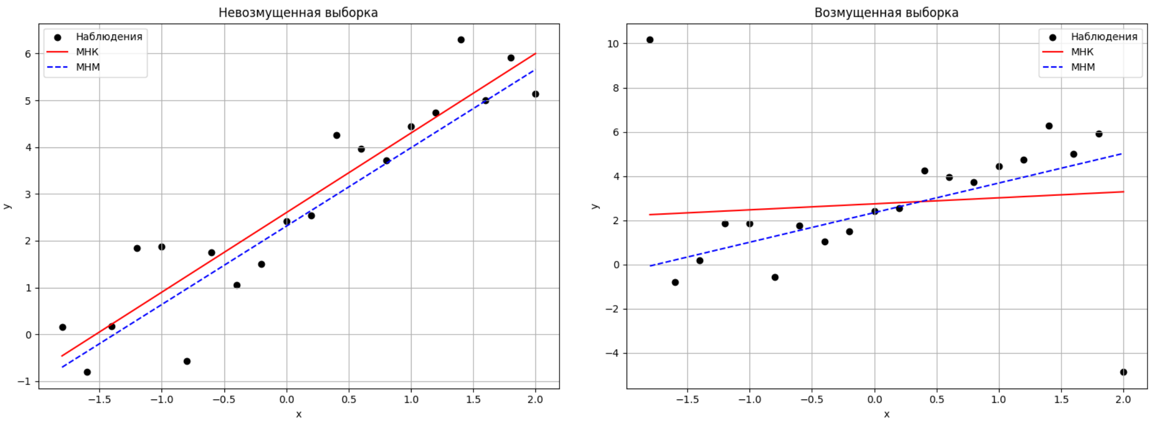

Рис. 1: Графическое представление результатов: слева — невозмущённая выборка, справа — возмущённая выборка.

3 Заключение

В данной работе были рассмотрены два метода оценки параметров линейной регрессии: метод наименьших квадратов (МНК) и метод наименьших модулей (МНМ). Проведены вычисления для двух выборок: без возмущений и с внесёнными возмущениями в крайние значения.

Результаты показывают, что метод наименьших квадратов даёт более точные оценки параметров при отсутствии выбросов, но чувствителен к экстремальным значениям в данных. В свою очередь, метод наименьших модулей демонстрирует устойчивость к выбросам.

Дополнительно, в ходе работы был написан код для генерации выборок, расчёта оценок параметров и их визуализации. Исходный код лабораторной работы размещён на GitHub.