Санкт-Петербургский Политехнический университет

Петра Великого

Институт прикладной математики и механики

Высшая школа прикладной математики и вычислительной физики

Отчёт

по лабораторной работе 6

по дисциплине

"математическая статистика"

Выполнил студент:

Аникин Александр Алексеевич,

группа 3630102\80201

Проверил:

к.ф.-м.н., доцент

Баженов Александр Николаевич

Санкт-Петербург

2021

Содержание

Список иллюстраций

1 Постановка задачи

Найти оценки коэффициентов линейной регрессии $y_i = a + bx_i + e_i$ используя 20 точек на отрезке $[-1.8; 2]$ с равномерным шагом равным 0.2. Ошибку e_i считать нормально распределённой с параметрами $(0, 1)$. В качестве эталонной зависимости взять $y_i = 2 + 2x_i$. 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.

2 Теория

2.1 Модель простой линейной регрессии

Регрессионную модель описания данных называют простой линейной регрессией, если

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \tag{1}$$

где $x_1, ..., x_n$ — заданные числа (значения фактора); $y_1, ..., y_n$ — наблюдаемые значения отклика; $\epsilon_1, ..., \epsilon_n$ — независимые, нормально распределённые $N(0, \sigma)$ с нулевым математическим ожиданием и одинаковой (неизвестной) дисперсией случайные величины; β_0, β_1 — неизвестные параметры, подлежащие оценке.

2.2 Метод наименьших квадратов

Метод наименьших квадратов (МНК):

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

Расчётные формулы для МНК оценок:

$$\hat{\beta}_1 = \frac{\overline{xy} - \overline{x} * \overline{y}}{\overline{x^2} - \overline{x}^2} \tag{3}$$

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

2.3 Метод наименьших модулей

Метод наименьших модулей(МНМ):

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

Расчётные формулы для МНМ оценок:

$$r_Q = \frac{1}{n} \sum_{i=1}^{n} sgn(x_i - med(x)) sgn(y_i - med(y)) \tag{6}$$

$$q_y^* = \frac{y_j - y_l}{k_q(n)}, \quad q_x^* = \frac{x_j - x_l}{k_q(n)}, \tag{7}$$

$$l = \begin{cases} \frac{n}{4} + 1 & \text{при} \quad \frac{n}{4} \quad \text{дробном} \\ \frac{n}{4} & \text{при} \quad \frac{n}{4} \quad \text{целом} \end{cases}, \quad j = n - l + 1 \tag{8}$$

$$\hat{\beta}_1 = r_Q \frac{q_y^*}{q_x^*} \tag{9}$$

$$\hat{\beta}_0 = med(y) - \hat{\beta}_1 med(x) \tag{10}$$

3 Реализация

Лабораторная работа выполнена на языке Python 3.8 с помощью загружаемых пакетов SciPy, MatPlotLib, NumPy. Исходный код лабораторной работы находится на GitHub репозитории.

4 Результаты

4.1 Оценка коэффициентов линейной регрессии

4.1.1 Выборка без возмущений

- Метод наименьших квадратов:

$$a \approx 2.08, \quad b \approx 1.85$$

- Метод наименьших модулей:

$$a \approx 1.55, \quad b = 1.03$$

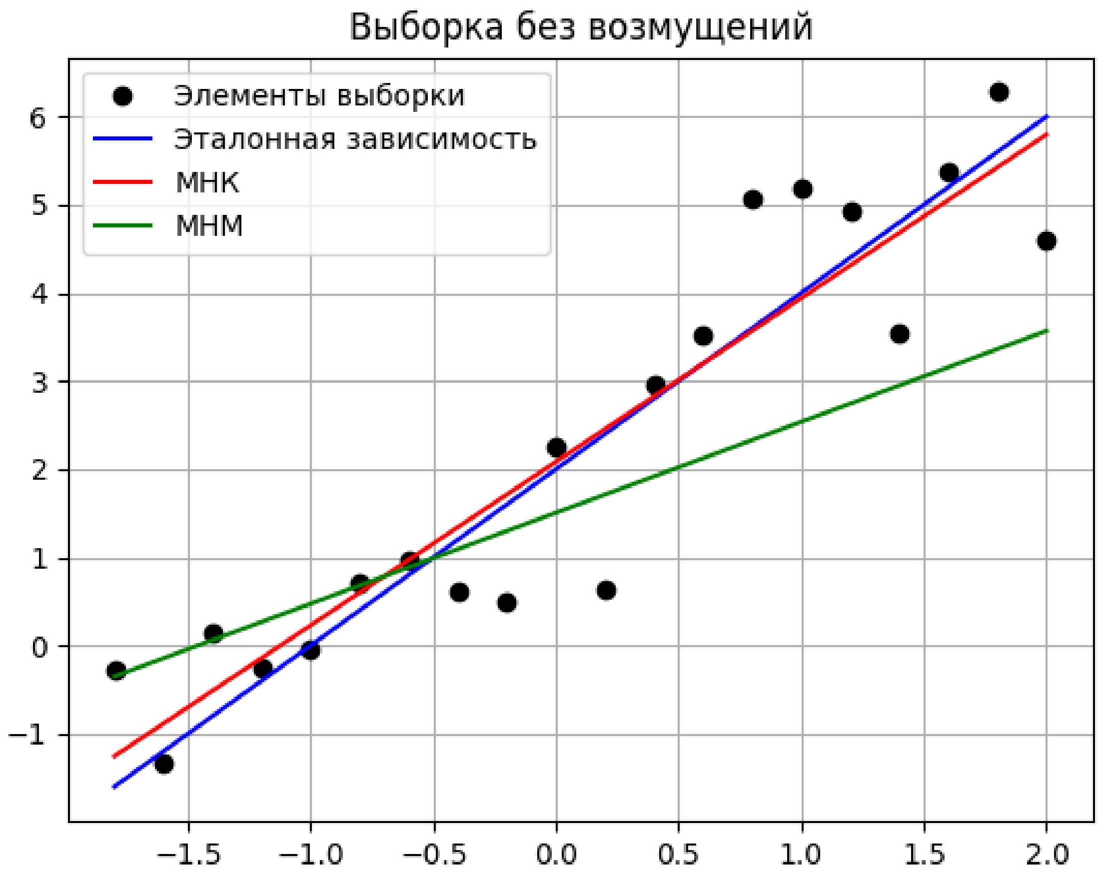

Рис. 1: Выборка без возмущений

4.1.2 Выборка с возмущениями

- Метод наименьших квадратов:

$$a \approx 2.22, \quad b \approx 0.42$$

- Метод наименьших модулей:

$$a \approx 1.56, \quad b = 1.77$$

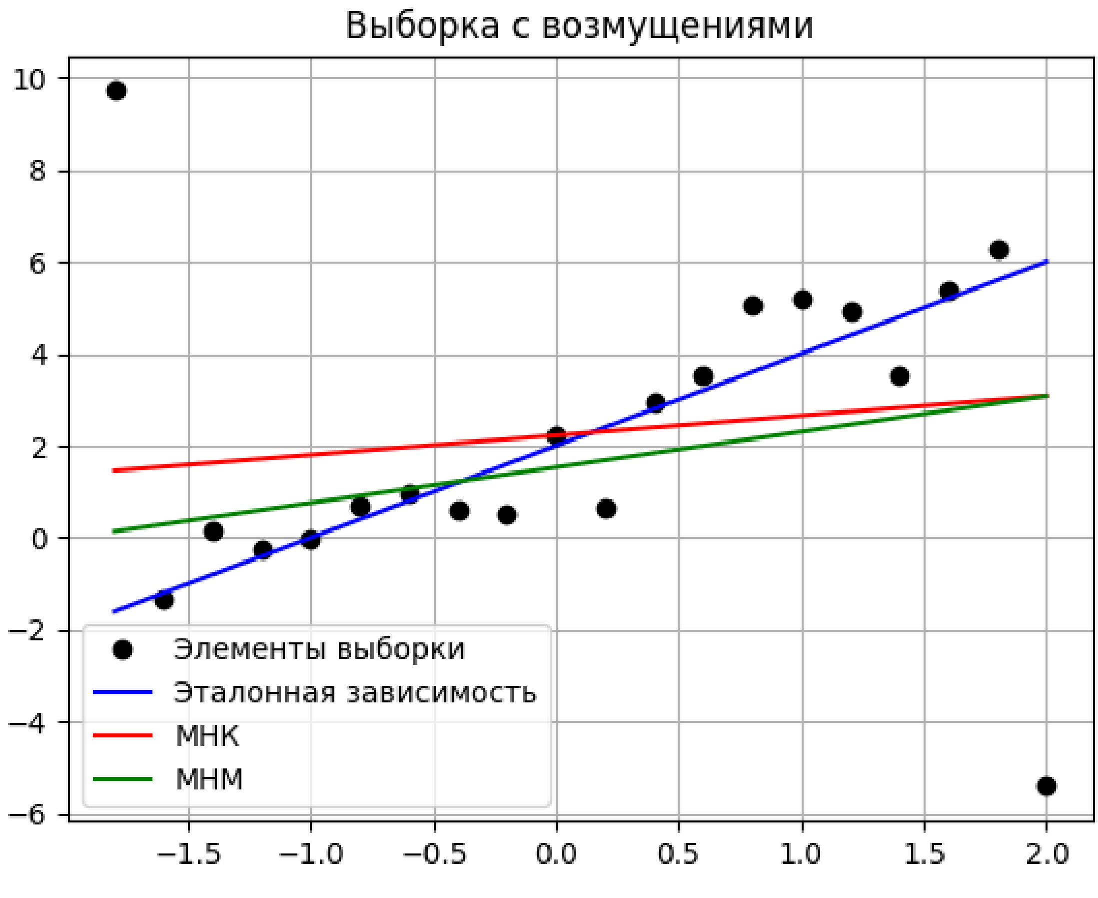

Рис. 2: Выборка с возмущениями

5 Обсуждение

Стоит заметить, что в случае выборки без значительных возмущений метод наименьших квадратов дает более точную оценку чем метод наименьших модулей, однако при внесении возмущений в краевые точки выборки МНК показывает довольно сильное отклонение, МНМ остается относительно близок к эталонной модели (ближе чем МНК). На основании проведенного исследования можно установить, что при малых отклонениях исходных данных целесообразнее использовать метод наименьших квадратов, в противном случае - метод наименьших модулей.