

**Санкт-Петербургский политехнический университет Петра Великого**

**Институт прикладной математики и механики**

**Кафедра «Телематика (при ЦНИИ РТК)»**

**Отчет по лабораторной работе**

**Оценки коэффициентов линейной регрессии**

**По дисциплине «Теория вероятностей и математическая статистика»**

Выполнил

Студент гр.3630201/80101

\_\_\_\_\_

В.Н. Сеннов

Руководитель

доцент к.ф.-м.н.

\_\_\_\_\_

А.Н. Баженов

«\_\_\_» \_\_\_\_\_ 202\_\_г.

Санкт-Петербург  
2020

# Содержание

<b>1</b>	<b>Постановка задачи</b>	<b>4</b>
<b>2</b>	<b>Математическое описание</b>	<b>5</b>
2.1	Простая линейная регрессия . . . . .	5
2.2	Метод наименьших квадратов . . . . .	5
2.3	Метод наименьших модулей . . . . .	5
<b>3</b>	<b>Особенности реализации</b>	<b>6</b>
<b>4</b>	<b>Результаты работы программы</b>	<b>7</b>
4.1	Выборка без возмущений . . . . .	7
4.2	Выборка с возмущениями . . . . .	7
	<b>Заключение</b>	<b>9</b>
	<b>Список литературы</b>	<b>10</b>
<b>A</b>	<b>Репозиторий с исходным кодом</b>	<b>11</b>

## Список иллюстраций

1	Графики для выборки без возмущений. . . . .	7
2	Графики для выборки с возмущениями. . . . .	8

# 1 Постановка задачи

В рамках данной работы необходимо сгенерировать выборку для зависимости

$$y_i = 2 + 2x_i + e_i,$$

где  $x_1 = -1.8$ ,  $x_2 = -1.6$ ,  $\dots$ ,  $x_{20} = 2$ ,  $e_i \sim N(0, 1)$ .

Необходимо найти оценки коэффициентов линейной регрессии методом наименьших квадратов и наименьших модулей. Найти те же оценки для выборки, в которой элементы  $y_1$  и  $y_{20}$  возмущены на 10 и -10.

## 2 Математическое описание

### 2.1 Простая линейная регрессия

Регрессионную модель описания данных называют простой линейной регрессией, если

$$y_i = a + bx_i + e_i, \quad (1)$$

где  $x_1, \dots, x_n$  — заданные числа;  $y_1, \dots, y_n$  — наблюдаемые значения отклика;  $e_i \sim N(0, \sigma)$ ;  $a$ ,  $b$  — неизвестные параметры регрессии.

### 2.2 Метод наименьших квадратов

Оценки  $a'$ ,  $b'$  для параметров  $a$ ,  $b$  в формуле (1) можно найти минимизируя функционал (2).

$$Q(a', b') = \sum_{i=1}^n (y_i - a' - b'x_i)^2 \quad (2)$$

Оценки  $a'$  и  $b'$  можно найти по следующим формулам:

$$b' = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - (\bar{x})^2} \quad (3)$$

$$a' = \bar{y} - \bar{x} \cdot b' \quad (4)$$

### 2.3 Метод наименьших модулей

Оценки  $a'$ ,  $b'$  для параметров  $a$ ,  $b$  в формуле (1) можно найти минимизируя функционал (5).

$$Q'(a', b') = \sum_{i=1}^n |y_i - a' - b'x_i| \quad (5)$$

Расчетных формул для метода наименьших модулей нет, минимизировать функционал можно численно.

Метод наименьших модулей позволяет сделать робастную оценку, то есть устойчивую к случайным возмущениям.

### 3 Особенности реализации

Программа для выполнения лабораторной была написана на языке Python 3.8.2. Для генерации выборки был использован модуль **stats** библиотеки `scipy`, для минимизации функционала был использован модуль **optimize** из той же библиотеки. Для построения графиков использовалась библиотека `Matplotlib`.

Оценки коэффициентов линейной регрессии были оценены методом наименьших квадратов по формулам (4) и (3) и методом наименьших модулей. Функционал (5) был минимизирован функцией **optimize.minimize**.

В приложении А приведена ссылка на репозиторий с исходным кодом.

## 4 Результаты работы программы

### 4.1 Выборка без возмущений

Оценки, полученные методом наименьших квадратов:

$$a' \approx 2.2400, \quad b' \approx 2.2019$$

Оценки, полученные методом наименьших модулей:

$$a' \approx 2.6183, \quad b' \approx 2.2312$$

На рис. 1 изображены графики для выборки без возмущений.

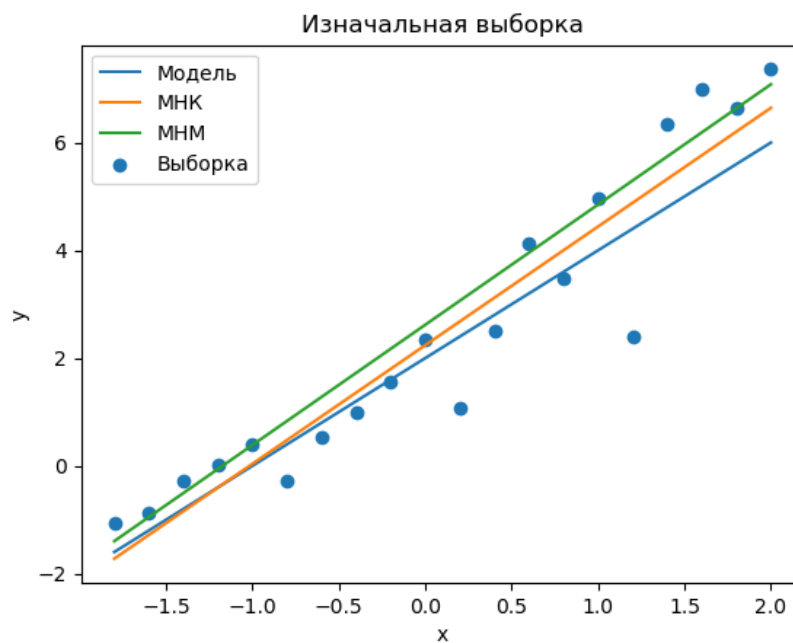


Рис. 1: Графики для выборки без возмущений.

### 4.2 Выборка с возмущениями

Оценки, полученные методом наименьших квадратов:

$$a' \approx 2.3828, \quad b' \approx 0.7733$$

Оценки, полученные методом наименьших модулей:

$$a' \approx 2.0569, \quad b' \approx 1.8281$$

На рис. 2 изображены графики для выборки с возмущениями.

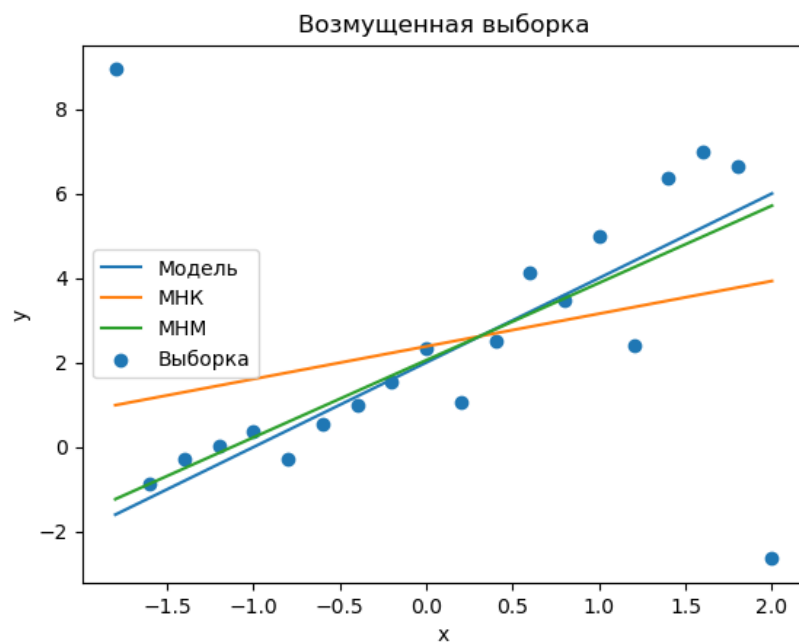


Рис. 2: Графики для выборки с возмущениями.



## Заключение

В рамках лабораторной работы была сгенерирована выборка для линейной зависимости с нормально распределенной ошибкой, были оценены оценки коэффициентов линейной регрессии методом наименьших квадратов и методом наименьших модулей для полученной выборки и выборки с возмущениями. Для полученных оценок были построены оценки.

Видно, что для выборки без возмущений и метод наименьших квадратов, и метод наименьших модулей достаточно точно приближают исходную зависимость. Но найти коэффициенты методом наименьших квадратов проще, чем методом наименьших модулей.

Для выборки с возмущениями метод наименьших квадратов дают гораздо менее точную оценку, очень сильно отличается коэффициент наклона. При этом оценка, полученная методом наименьших модулей, практически не меняется.

Программа для лабораторной была написана языке Python 3.8.2, для генерации выборок была использована библиотека `scipy`, для построения графиков использовалась библиотека `Matplotlib`.

## Список литературы

- [1] Теоретическое приложение к лабораторным работам №5-8 по дисциплине «Математическая статистика». — Спб.: Санкт-Петербургский политехнический университет, 2020. — 12 с.

## А Репозиторий с исходным кодом

Исходный код программы для данной лабораторной размещен на сервисе GitHub.

Ссылка на репозиторий: <https://github.com/Vovan-S/TV-Lab1>.