

САНКТ-ПЕТЕРБУРГСКИЙ ПОЛИТЕХНИЧЕСКИЙ
УНИВЕРСИТЕТ им. ПЕТРА ВЕЛИКОГО

ИНСТИТУТ ПРИКЛАДНОЙ МАТЕМАТИКИ И МЕХАНИКИ

КАФЕДРА ПРИКЛАДНОЙ МАТЕМАТИКИ

ЛАБОРАТОРНАЯ РАБОТА №6

3 КУРС, ГРУППА 33631/2

Студент

Д. А. Плаксин

Преподаватель

Баженов А. Н.

САНКТ-ПЕТЕРБУРГ
2019 г.

Содержание

1. Список иллюстраций	3
2. Список таблиц	3
3. Постановка задачи	4
4. Теория.....	4
4.1. Метод наименьших квадратов	4
4.2. Метод наименьших модулей	4
5. Реализация.....	4
6. Результаты	5
7. Выводы	5
8. Список литературы	6
9. Приложения	6

1 Список иллюстраций

1	Графики линейной регрессии	5
---	----------------------------------	---

2 Список таблиц

1	Таблица оценок коэффициентов линейной регрессии без возмущений . . .	5
2	Таблица оценок коэффициентов линейной регрессии с возмущениями . . .	5

3 Постановка задачи

Необходимо найти оценки линейной регрессии $y_i = a + bx_i + e_i$, используя 20 точек отрезка $[-1.8; 2]$ с равномерным шагом 0.2. Ошибку e_i считать нормально распределённой с параметрами $(0, 1)$. В качестве эталонной зависимости взять $y_i = 2 + 2x_i + e_i$. При построении оценок коэффициентов использовать два критерия: критерий наименьших квадратов и критерий наименьших модулей.

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

4 Теория

Простая линейная регрессия [4]:

$$y_i = ax_i + b + e_i, \quad i = \overline{1, n}, \quad (1)$$

где x_i – заданные числа, y_i – наблюдаемые значения, e_i – независимы и нормально распределены, a и b – неизвестные параметры, подлежащие оцениванию.

4.1 Метод наименьших квадратов

Критерий – минимизация функции [5]:

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

Оценки \hat{a} и \hat{b} параметров a и b , в которых достигается минимум $Q(a, b)$, называются МНК-оценками. В случае линейной регрессии их можно вычислить из формулы [6]:

$$\begin{cases} \hat{a} = \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - \bar{x}^2} \\ \hat{b} = \bar{y} - \hat{a}\bar{x} \end{cases} \quad (3)$$

Метод наименьших квадратов является несмещённой оценкой.

МНК чувствителен к выбросам (т.к. в вычислении используется выборочное среднее значение величин крайне неустойчивое к редким, но большим по величине выбросам)

4.2 Метод наименьших модулей

Критерий наименьших модулей – заключается в минимизации следующей функции [7]:

$$M(a, b) = \sum_{i=1}^n |y_i - ax_i - b| \rightarrow \min \quad (4)$$

МНМ-оценки обладают свойством робастности. Но на практике решение реализуется только численно.

5 Реализация

Работы была выполнена на языке *Python3.7*. Для генерации выборок использовался модуль [1]. Для построения графиков использовалась библиотека *matplotlib* [2]. Функции распределения обрабатывались при помощи библиотеки *scipy.stats* [3].

6 Результаты

Рис. 1: Графики линейной регрессии

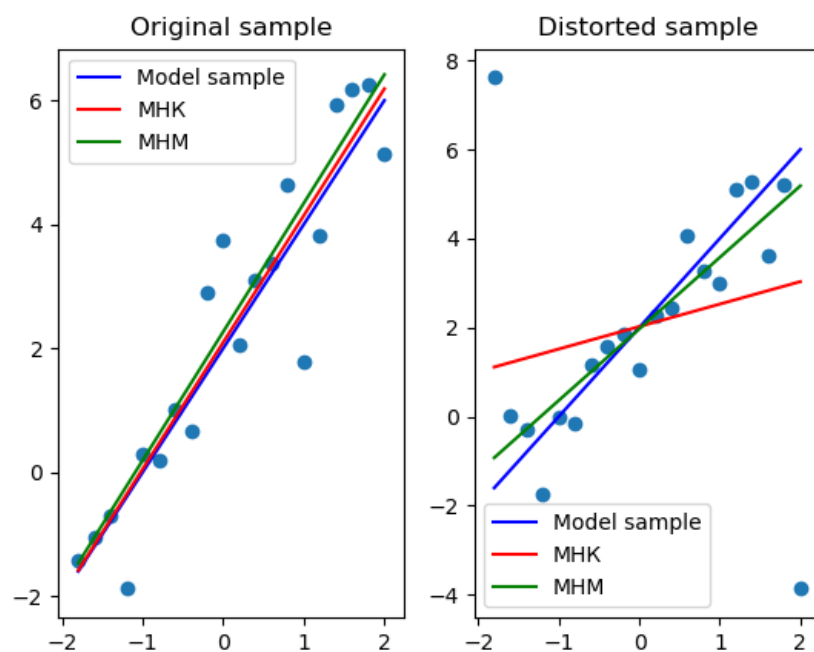


Таблица 1: Таблица оценок коэффициентов линейной регрессии без возмущений

	\hat{a}	\hat{b}
МНК	2.000	2.000
МНМ	2.045	2.098

Таблица 2: Таблица оценок коэффициентов линейной регрессии с возмущениями

	\hat{a}	\hat{b}
МНК	2.000	2.000
МНМ	0.505	2.023

7 Выводы

По графику 1 видно, что оба метода дают хорошую оценку коэффициентов линейной регрессии, если нет выбросов. Однако выбросы сильно влияют на оценки по

МНК.

Выбросы мало влияют на оценку по МНМ, но ценой за это является большая по сравнению с МНК сложность вычисления.

8 Список литературы

- [1] Модуль numpy - <https://physics.susu.ru/vorontsov/language/numpy.html>
- [2] Модуль matplotlib - <https://matplotlib.org/users/index.html>
- [3] Модуль scipy - <https://docs.scipy.org/doc/scipy/reference/>
- [4] https://en.wikipedia.org/wiki/Linear_regression
- [5] <http://www.cleverstudents.ru/articles/mnk.html>
- [6] Шевляков Г. Л. Лекции по математической статистике, 2019.
- [7] Вероятностные разделы математики. Учебник для бакалавров технических направлений. //Под ред. Максимова Ю.Д. - СПб.: "Иван Федоров 2001. - 592 с.

9 Приложения

Код отчёта: <https://github.com/MisterProper9000/MatStatLabs/blob/master/MatStatLab5/MatStatLab6.tex>

Код лабораторной: <https://github.com/MisterProper9000/MatStatLabs/blob/master/MatStatLab5/MatStatLab6.py>

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.optimize import minimize

def fit(X, params):
    return X.dot(params)

def cost_function(params, X, y):
    return np.sum(np.abs(y - fit(X, params)))

def generate(a, b, h):
    n = int((b - a) / h) + 2
    x = np.linspace(-1.8, 2, n)
    e_i = np.random.standard_normal(n)
    y = 2 + 2*x + e_i
    return x, y

def perturbation(y):
    res = []
    for i in range(0, len(y)):
        res.append(y[i])
    back = len(y) - 1
    res[0] = y[0] + 10
    res[back] = y[back] - 10
    return res

def mnk(x, y):
```

```

    a2 = (np.mean(x*y) - np.mean(x) * np.mean(y)) / (np.mean(x*x) - np.mean(x)
    **2)
    b2 = np.mean(y) - a2 * np.mean(x)
    return a2, b2

def mmm(x, y):
    X = np.asarray([np.ones(20, ), x]).T
    x0 = np.ones(2)
    output = minimize(cost_function, x0, args=(X, y))

    y_hat = fit(X, output.x)

    print("y hat = ", np.median(y_hat))

    return x, y_hat

l = -1.8
r = 2
h = 0.2
a = 2
b = 2

plt.figure()
plt.subplot(121)
plt.title("Original sample")
print("\t\t\tOriginal sample")

x, y = generate(l, r, h)
print("%12s:\t a = %lf, b = %lf" % ("Model sample", a, b))
plt.plot(x, a * x + b, 'b', label='Model sample')

plt.scatter(x, y)

m, c = mnk(x, y)
print("%12s:\ta = %lf, b = %lf" % ("", m, c))
plt.plot(x, m*x + c, 'r', label='')

m, c = mmm(x, y)
plt.plot(m, c, 'g', label='')

plt.legend()

print("\n")
plt.subplot(122)
plt.title("Distorted sample")
print("\t\t\tDistorted sample")

x, y = generate(l, r, h)
y = perturbation(y)

print("%12s:\ta = %lf, b = %lf" % ("Model sample", a, b))
plt.plot(x, a * x + b, 'b', label='Model sample')
plt.scatter(x, y)

m, c = mnk(x, y)
print("%12s:\ta = %lf, b = %lf" % ("", m, c))
plt.plot(x, m*x + c, 'r', label='')

xx, yy = mmm(x, y)
plt.plot(xx, yy, 'g', label='')

plt.legend()
plt.show()

```