Санкт-Петербургский политехнический университет
Петра Великого

Институт прикладной математики и механики
Кафедра «Прикладная математика»

Отчёт
по лабораторной работе №6
по дисциплине
«Математическая статистика»

Выполнил студент:
Самутичев Евгений Романович
группа: 3630102/70201

Проверил:
к.ф.-м.н., доцент
Баженов Александр Николаевич

Санкт-Петербург
2020 г.

Содержание

Список иллюстраций

1 Постановка задачи

Найти оценки коэффициентов a, b 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 $y_i = 2+2x_i+e_i$. При построении оценок коэффициентов использовать два критерия: критерий наименьших квадратов и критерий наименьших модулей. Проделать то же самое для выборки, у которой в значения y_1 и y_2 вносятся возмущения 10 и -10.

2 Теория

2.1 Простая линейная регрессия

Регрессионную модель описания данных называют *простой линейной регрессией*, если

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

, где $\{x_i\}_{i=1}^n$ - значения фактора, $\{y_i\}_{i=1}^n$ - наблюдаемые значения отклика, а $\{\varepsilon_i\}_{i=1}^n$ - независимые, нормально распределенные по закону $N(0, \sigma)$ случайные величины, а β_0, β_1 - оцениваемые параметры [1, стр. 507]. Для оценки применяются различные методы, в данной работе рассмотрен следующий подход: вводится критерий рассогласования отклика и регрессионной функции, после чего оценки параметров регресии выводятся из задачи минимизации критерия. Рассмотрим два таких критерия.

2.1.1 Критерий наименьших квадратов

Достаточно простые расчетные формулы для оценок получают при выборе критерия в виде суммы квадратов отклонений значений отклика от значений регрессионной функции:

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

Приведем сами расчетные формулы [1, стр. 509]:

$$\widehat{\beta}_1 = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - (\bar{x})^2} \quad \widehat{\beta}_0 = \bar{y} - \bar{x}\widehat{\beta}_1 \tag{3}$$

Важным свойством является несмещенность оценки, однако она чувствительна к выбросам и если нужна робастная оценка, то следует рассмотреть следующий критерий.

2.1.2 Критерий наименьших модулей

В отличие от задач метода наименьших квадратов, для этого критерия минимизацию на практике проводят численно, решая:

$$M\left(\beta_0, \beta_1\right) = \sum_{i=1}^{n} \left|y_i - \beta_0 - \beta_1 x_i\right| \to \min_{\beta_0, \beta_1} \tag{4}$$

В данной работе был использован метод Нелдера-Мида [2], применимый к негладким функциям (в том числе к $M\left(\beta_0, \beta_1\right)$). Подробнее см. реализация.

3 Реализация

Работа выполнена с использованием языка **Python** в интегрированной среде разработки **PyCharm**, были задействованы библиотеки:

- **NumPy** - векторизация вычислений, работа с массивами данных, вычисление выборочных характеристик
- **SciPy** - модуль **stats** для генерации данных по эталонной зависимости, оценок МНК, модуль **optimize** для метода Нелдера-Мида
- **Matplotlib** - построение графиков

Исходный код работы приведен в приложении.

4 Результаты

4.1 Выборка без выбросов

- Критерий наименьших квадратов:

$$\widehat{\beta}_0 = 2.47 \quad \widehat{\beta}_1 = 1.95 \quad Q(2) = 13.9637 \quad M(4) = 13.9182$$

- Критерий наименьших модулей:

$$\widehat{\beta}_0 = 2.49 \quad \widehat{\beta}_1 = 1.68 \quad Q = 15.9356 \quad M = 13.3737$$

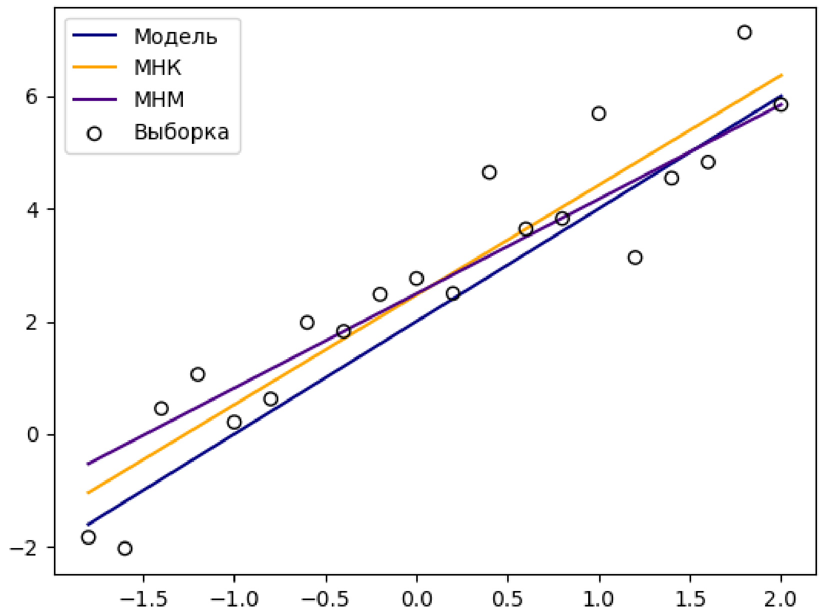

Рис. 1: Без выбросов

4.2 Выборка с выбросами

- Критерий наименьших квадратов:

$$\widehat{\beta}_0 = 2.61 \quad \widehat{\beta}_1 = 0.52 \quad Q = 154.2302 \quad M = 37.381$$

- Критерий наименьших модулей:

$$\widehat{\beta}_0 = 2.67 \quad \widehat{\beta}_1 = 1.35 \quad Q = 172.7536 \quad M = 29.9906$$

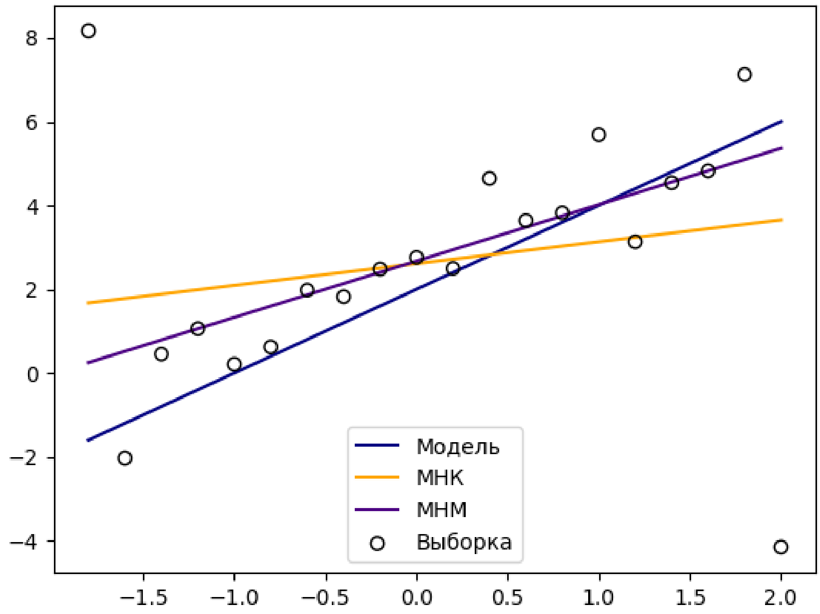

Рис. 2: С выбросами

5 Обсуждение

Из графиков видно, что оценка по критерию наименьших модулей значительно лучше приближает эталонную зависимость при наличии выбросов и это согласуется с теорией т.к. она является робастной. В тоже время, критерий наименьших квадратов дает более точное приближение в отсутствие выбросов и, к тому же, проще для вычислений. Полученные значения M, Q упорядочены как и ожидалось, для оценки МНК значение Q меньше, чем для любой другой, аналогично для оценки МНМ и значения M

6 Приложения

1. Исходный код лабораторной `https://github.com/zhenyatos/statlabs/tree/master/Lab6`

Список литературы

[1] **Вероятностные разделы математики.** Учебник для бакалавров технических направлений. // Под ред. Максимова Ю.Д. - СПб «Иван Федоров», 2001. - 592 с., илл

[2] Метод Нелдера — Мида // Википедия. [2019—2019]. Дата обновления: 11.09.2019. URL: https://ru.wikipedia.org/?oldid=102111276 (дата обращения: 11.09.2019).