

Санкт-Петербургский политехнический университет
Петра Великого

Физико-механический институт

Кафедра «Прикладная математика»

**Отчёт по курсовой работе
по дисциплине «Анализ данных с интервальной
неопределённостью»**

Выполнил студент:
Игнатъев Даниил Дмитриевич
группа: 5040102/20201

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

Санкт-Петербург
2023 г.

Содержание

1	Постановка задачи	2
2	Теория	2
2.1	Точечная линейная регрессия	2
2.2	Информационное множество	2
3	Реализация	3
4	Результаты	3
5	Обсуждение	6

Список иллюстраций

1	Визуализация выборки, Y_1	4
2	Точечная линейная регрессия для Y_1	4
3	Информационное множество для Y_1	5
4	Коридор совместных значений для Y_1	6

1 Постановка задачи

Необходимо для заданных выборок найти точную линейную регрессию, информационные множества и коридоры решений. Сравнить полученные результаты.

2 Теория

2.1 Точечная линейная регрессия

Рассматривается задача восстановления зависимости для выборки $(X, (Y))$, $X = \{x_i\}_{i=1}^n$, $Y = \{y_i\}_{i=1}^n$, x_i - точечный, y_i - интервальный. Пусть искомая модель задана в классе линейных функций

$$y = \beta_0 + \beta_1 x \quad (1)$$

Поставим задачу оптимизацию 2 для нахождения точечных оценок параметров β_0, β_1 .

$$\begin{aligned} \sum_{i=1}^m w_i &\rightarrow \min \\ \text{mid}y_i - w_i \cdot \text{rad}y_i &\leq X\beta \leq \text{mid}y_i + w_i \cdot \text{rad}y_i \\ w_i &\geq 0, i = 1, \dots, m \\ w, \beta &-? \end{aligned} \quad (2)$$

Задачу 2 можно решить методами линейного программирования.

2.2 Информационное множество

Информационным множеством задачи восстановления зависимости будем называть множество значений всех параметров зависимости, совместных с данными в каком-то смысле.

Коридором совместных зависимостей задачи восстановления зависимости называется многозначное множество отображений Υ , сопоставляющее каждому значению аргумента x множество

$$\Upsilon(x) = \bigcup_{\beta \in \Omega} f(x, \beta) \quad (3)$$

, где Ω - информационное множество, x - вектор переменных, β - вектор оцениваемых параметров.

Информационное множество может быть построено, как пересечение полос, заданных

$$\underline{y}_i \leq \beta_0 + \beta_1 x_{i1} + \dots + \beta_m x_{im} \leq \overline{y}_i \quad (4)$$

, где $i = \overline{1, n}$, $\underline{y}_i \in \mathbf{Y}$, $x_i \in X$, X - точечная выборка переменных, \mathbf{Y} - интервальная выборка откликов.

3 Реализация

Проект реализован на языке Python v. 3.2.5. [GitHub](#).

4 Результаты

Данные S_X были взяты из файлов *data/poly_i.csv*, где $i \in \{0, 1, \dots, 9\}$.

Набор значений X точечный и одинаков для всех выборок.
 $X = [-0.5, -0.35714286, -0.21428571, -0.07142857, 0.07142857, 0.21428571, 0.35714286, 0.5]$.
 Набор значений отклика Y - интервальный.

Построим линейную регрессию и найдём информационное множество.

Рассмотрим выборку Y следующим образом. $y = [\min_{t \in S_i} S_i, \max_{t \in S_i} S_i]$, $y_i \in Y_1$.

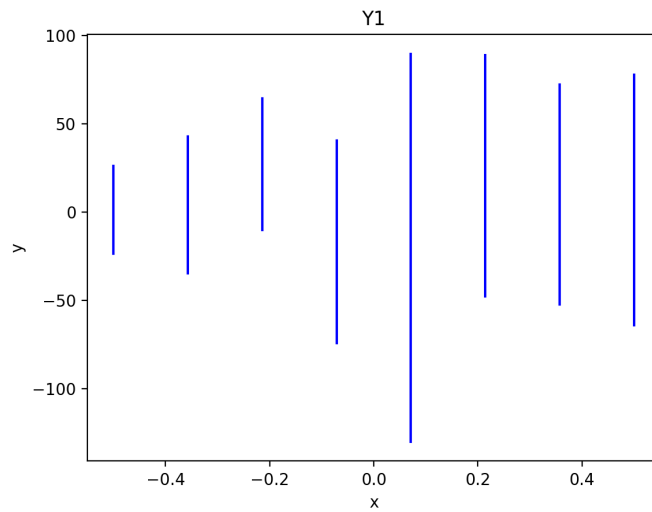


Рис. 1: Визуализация выборки, Y_1

Построим линейную регрессию, решив задачу 2 для выборки Y_1 .

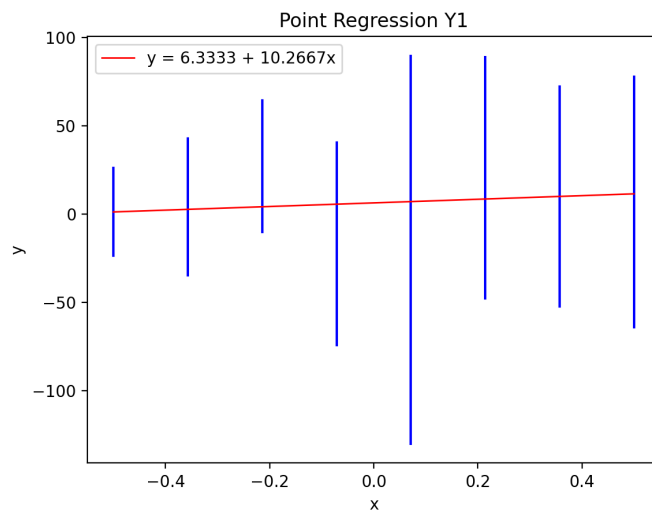


Рис. 2: Точечная линейная регрессия для Y_1

Получим следующие оценки для параметров: $\beta_0 = 6.333$, $\beta_1 = 10.267$. Тогда полученная модель имеет вид $y = 6.333 + 10.267x$.

Найдём для данной выборки информационное множество.

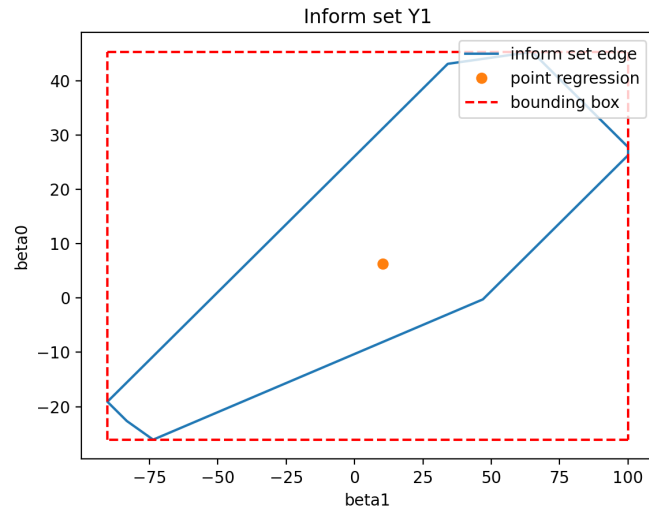


Рис. 3: Информационное множество для Y_1

На рис. 3 можно заметить, что найденные параметры β_0, β_1 решением задачи 2 лежат внутри информационного множества.

Построим коридор совместных значений для выборки Y_1 и информационного множества 3 и оценим значения выходной переменной y вне пределов значений входной переменной x .

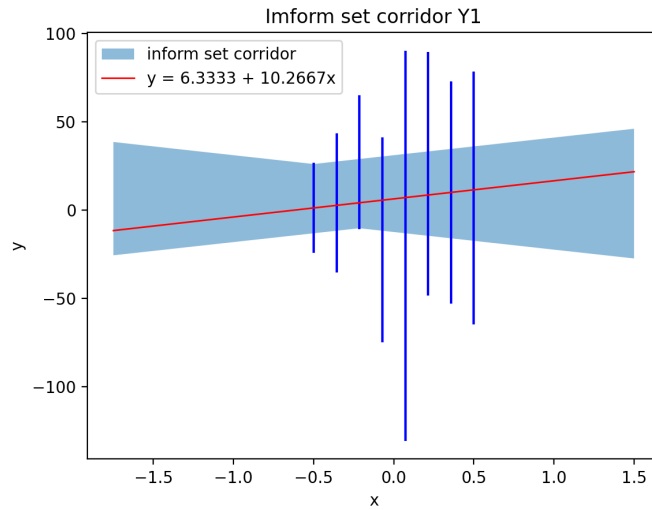


Рис. 4: Коридор совместных значений для Y_1

На рис. 4 видно, что построенная точечная регрессия лежит внутри коридора совместных значений, что согласуется с рис. 3.

5 Обсуждение

Полученные результаты в виде больших параметров β_0, β_1 , а также широкого коридора совместных решений обусловлены большой интервальной неопределённостью. Также стоит отметить, что точечная линейная регрессия попала в информационное множество.