Санкт-Петербургский политехнический университет
Петра Великого

Физико-механический иститут

Кафедра «Прикладная математика»

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

Выполнил студент:
Куксенко Кирилл Сергеевич
группа: 5040102/20201

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

Санкт-Петербург
2023 г.

Содержание

Список иллюстраций

1 Постановка задачи

Построить линейную регрессию на данных с твинной неопределённостью.

2 Теория

2.1 Твины

В классической интервальной арифметике при любой операции над интервалами происходит расширение неопределённости. Эти операции дают внешнюю оценку результатов вычислений.

Для получения более содержательных результатов желательно иметь внутреннюю оценку результатов вычислений в виде интервала. Для решения этой задачи вместо обычных интервалов используется более сложная конструкция - твин. Твин представляет собой составной интервал, определённый следующим образом.

$$\mathbb{T} = [\mathbf{t}_{in}, \mathbf{T}_{ex}] = \{\mathbf{x} \in \mathbb{KR} | \mathbf{t}_{in} \subseteq \mathbf{x} \subseteq \mathbf{T}_{ex}\} \tag{1}$$

2.2 Линейная регрессия

Будем рассматривать линейную модель.

$$y = \beta_0 + \beta_1 x \tag{2}$$

Подставив в зависимость 2 данные для входных переменных x_i, $i \in \overline{1, n}$ и потребовав включения полученного значения в соответствующий интервал y_i, получим

$$\beta_0 + \beta_1 x_i \in \mathbf{y}_i \tag{3}$$

Что равносильно системе линейных алгебраических уравнений с интервальной правой частью

$$\begin{cases} \beta_0 + \beta_1 x_1 = \mathbf{y}_1 \\ \dots \\ \beta_0 + \beta_1 x_n = \mathbf{y}_n \end{cases} \tag{4}$$

В случае, когда в правой части твины имеем

$$\begin{cases} \beta_0 + \beta_1 x_1 = [\mathbf{t}_{in1}, \mathbf{T}_{ex1}] \\ \dots \\ \beta_0 + \beta_1 x_n = [\mathbf{t}_{inn}, \mathbf{T}_{exn}] \end{cases} \tag{5}$$

Решая две системы уравнений, для внутренних и внешних оценок, получим два коридора совместных решений. Каждый из коридоров является внутренней и внешней оценкой.

3 Реализация

Весь код написан на языке Python (версии 3.7.3). Ссылка на GitHub с исходным кодом.

4 Результаты

Данные S_X были взяты из файлов *data/dataset2/XV_spN.txt*,
где $X \in P = \{-0.45, -0.35, -0.25, -0.15, -0.05, 0.05, 0.15, 0.25, 0.35, 0.45\}$.
Набор δ_i получен из файла *data/dataset2/0.0V_sp443.txt*.

Внешняя оценка для каждого измерения была получена следующим образом. $\mathbf{T}_{exi} = [\min S_i, \max S_i]$, $i \in P$.

А внутренняя оценка была получена. $\mathbf{t}_{ini} = [median(S_i) - \varepsilon, median(S_i) + \varepsilon]$, $\varepsilon = 200.0$, $i \in P$.

Таким образом для i-го измерения имеем неопределённость
$\mathbb{T}_i = [[\min S_i, \max S_i], [median(S_i) - \varepsilon, median(S_i) + \varepsilon]]$.

Полученная выборка имеет вид.

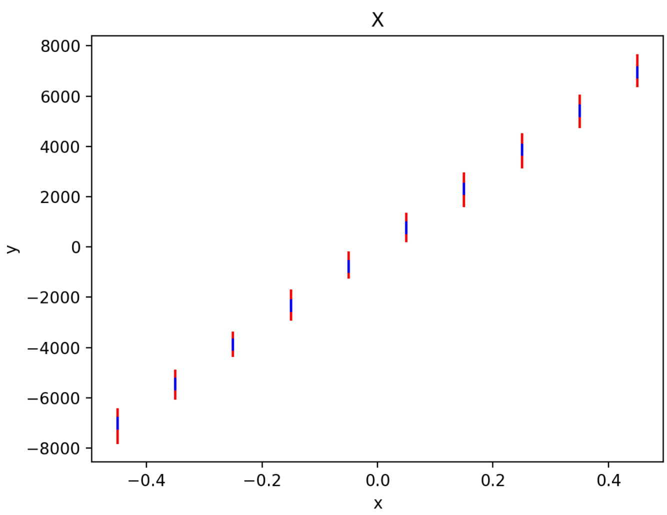

Рис. 1: Исходная выборка X

Построим коридор совместных значений для выборки X.

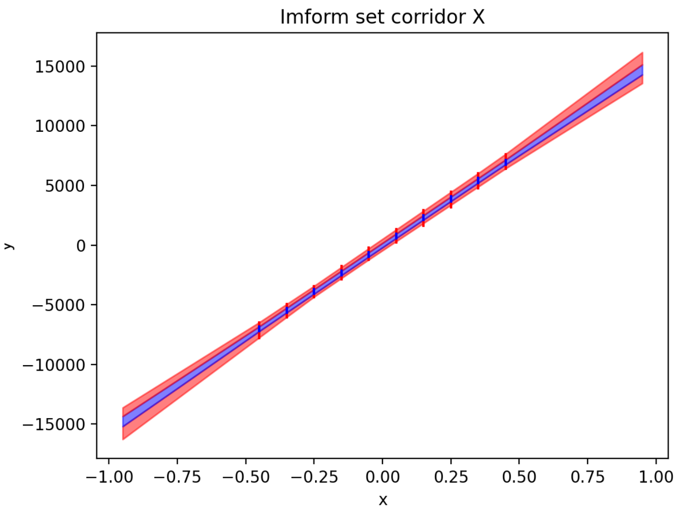

Рис. 2: Коридор совместных значений для X

Сравним информационные множества для внутренней и внешней оценок.

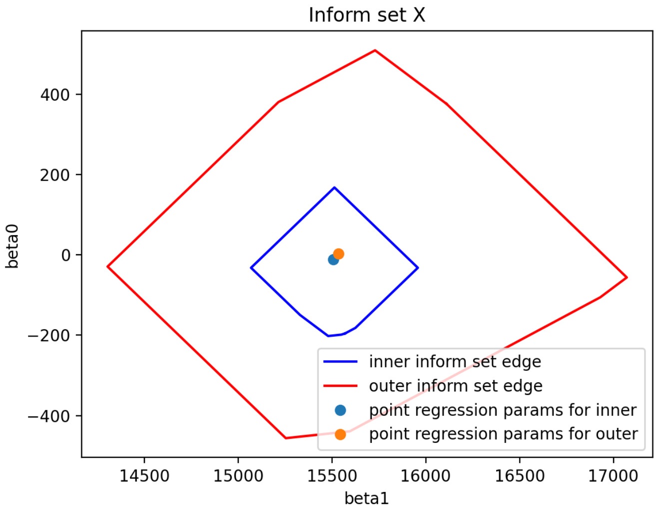

Рис. 3: Информационные множества для внутренней и внешней оценок для X

Для внутренней оценки измерений получили следующие оценки параметров модели 2: $\beta_0 = -11.33$, $\beta_1 = 15506.67$. А для внешней: $\beta_0 = 3.25$, $\beta_1 = 15535.0$.

Проведём аналогичные вычисления для выборки остатков $\mathcal{E}$, полученной следующим образом. $\mathbb{T}'_i = [\mathbf{t}'_{ini}, \mathbf{T}'_{exi}]$, где $\mathbf{t}'_{ini} = \mathbf{t}_{ini} - (\beta_{0in} + \beta_{1in} x_i)$, $\mathbf{T}'_{exi} = \mathbf{T}_{exi} - (\beta_{0ex} + \beta_{1ex} x_i)$.

Полученная выборка $\mathcal{E}$ имеет вид.

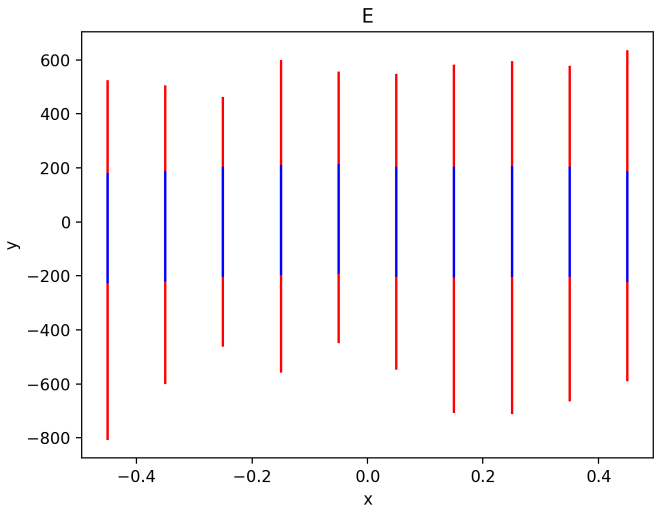

Рис. 4: Выборка остатков $\mathcal{E}$

Построим коридор совместных значений для выборки $\mathcal{E}$.

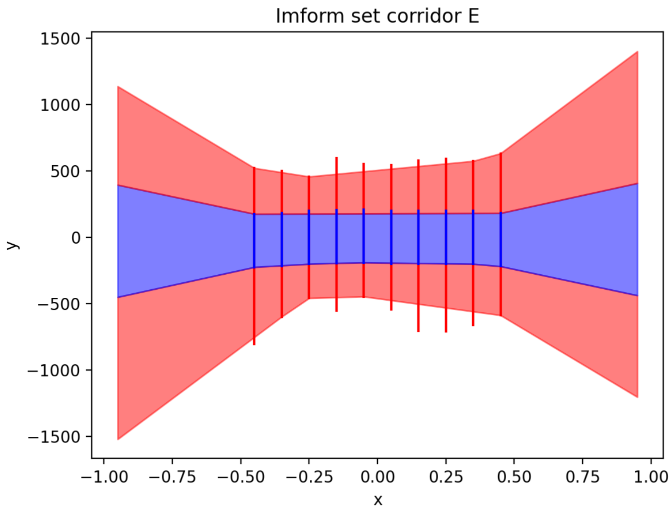

Рис. 5: Коридор совместных значений для $\mathcal{E}$

Также сравним информационныке множества для внутренней и внешней оценок.

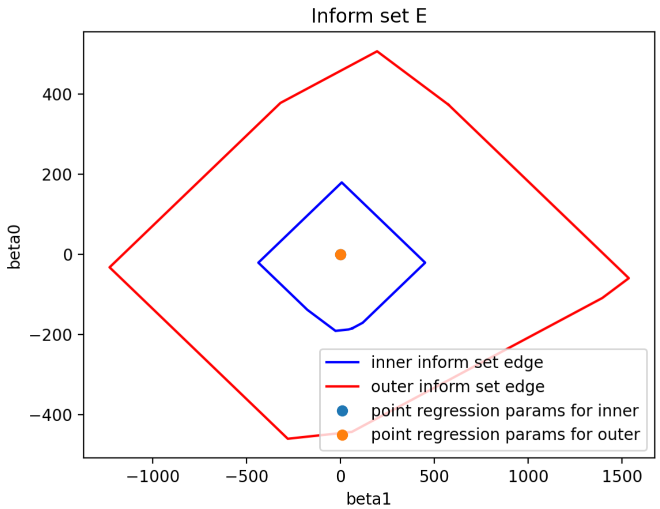

Рис. 6: Информационные множества для внутренней и внешней оценок для $\mathcal{E}$

Для наглядности рассмотрим ещё одну выборку X_1, полученную следующим образом. Внутренняя оценка неопределённости: $\mathbf{t}_{ini} = [median(S_i) - \varepsilon, median(S_i) + \varepsilon]$, $\varepsilon = 25.0$, $i \in P$ Внешняя оценка неопределённости: $\mathbf{T}_{exi} = [Q(S_i, 0.375), Q(S_i, 0.625)]$, $i \in P$, где $Q(S, q)$ - q квантиль выборки S.

Выборка и коридор совместных значений для выборки X_1 имеют вид.

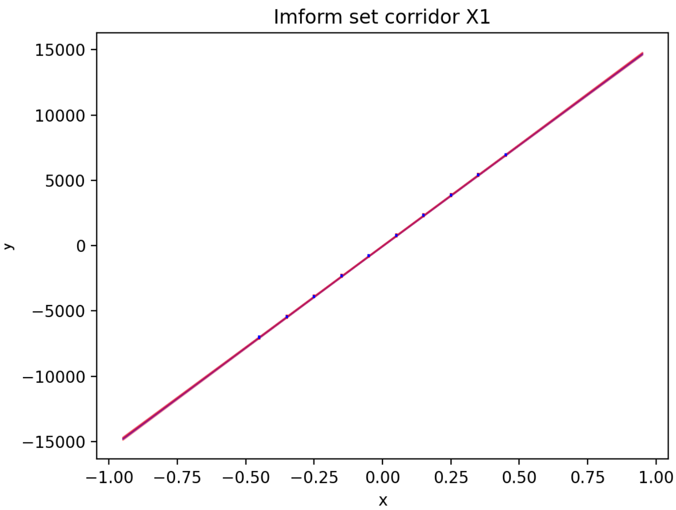

Рис. 7: Выборка и коридор совместных значений для X_1

Для внутренней оценки измерений получили следующие оценки параметров модели 2. $\beta_0 = -11.33, \beta_1 = 15506.67$. А для внешней: $\beta_0 = -9.59$, $\beta_1 = 15516.37$.

А информационные множества представлены на следующем рисунке.

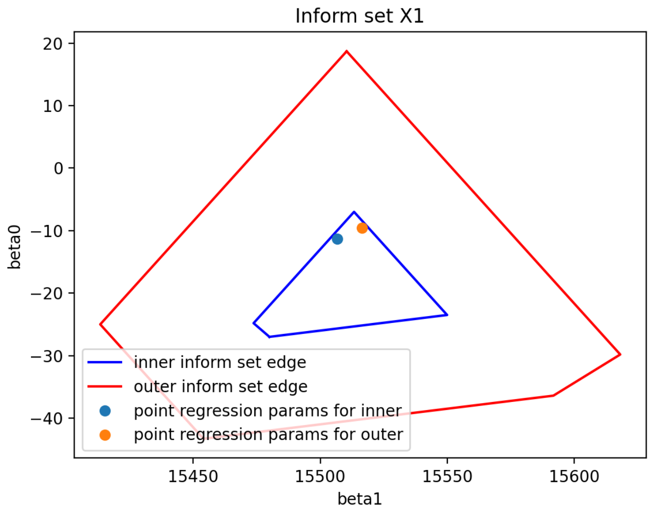

Рис. 8: Информационные множества для внутренней и внешней оценок для X_1

Проведём аналогичные вычисления для остатков. Выборка остатков и коридор совместных значений имеют следующий вид.

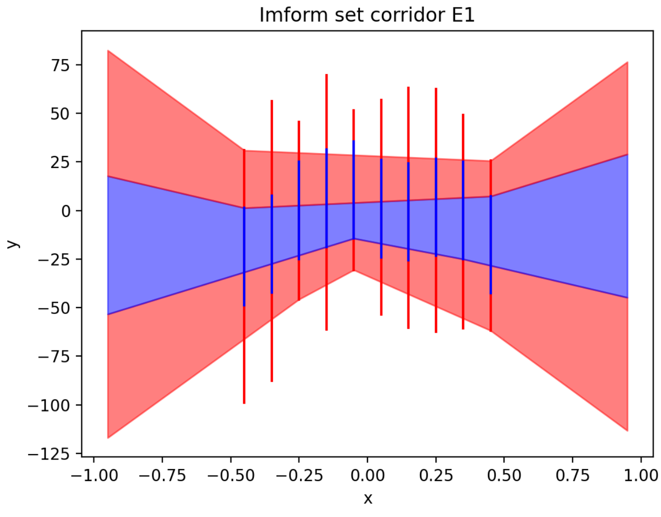

Рис. 9: Выборка и коридор совместных значений для $\mathcal{E}_1$

А информационные множества для $\mathcal{E}_1$ имеют вид.

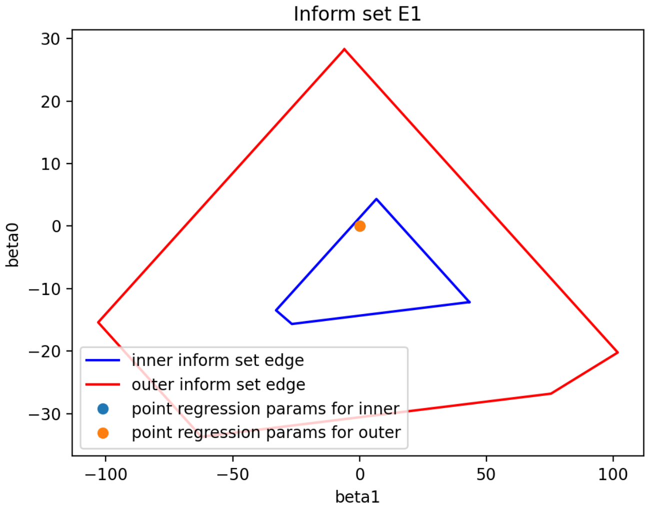

Рис. 10: Информационные множества для внутренней и внешней оценок для $\mathcal{E}_1$

5 Обсуждение

Из полученных результатов можно заметить следующее. Для всех рассмотренных выборок информационное множество, полученное для внутренней оценки, содержится в информационном множестве для внешней оценки, что неудивительно, ведь для всех измерений внутренняя оценка содержится во внешней.

Также можно отметить, что исходная выборка и полученная из неё выборка остатков имеют схожие по форме информационные множества для обоих оценок - внутренней и внешней (рис. 3, 6, 8, 10). Информационные множества отличаются только сдвигом вдоль осей β_0, β_1.

На рис. 3, 6 и 8, 10 видно, что информационное множество для внутренней оценки схоже по форме с информационным множеством для внешней оценки. При этом информационное множество, которое представляет из себя многоугольник, для внутренней оценки имеет не больше вершин, чем для внешней. Тогда можно предположить, что для внутренняя оценка измерений имеет не большее количество граничных измерений, чем внешняя.