

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	13

Список иллюстраций

1	Визуализация выборки Y_1	4
2	Визуализация выборки Y_2	4
3	Визуализация выборки Y_3	5
4	Визуализация выборки Y_4	5
5	Визуализация выборки Y_5	6
6	Визуализация выборки Y_6	6
7	Визуализация выборки Y_7	7
8	Визуализация выборки Y_8	7
9	Визуализация выборки Y_9	8
10	Точечная линейная регрессия для Y_1	8
11	Точечная линейная регрессия для Y_2	9
12	Точечная линейная регрессия для Y_3	9
13	Точечная линейная регрессия для Y_4	10
14	Точечная линейная регрессия для Y_5	10
15	Точечная линейная регрессия для Y_6	11
16	Точечная линейная регрессия для Y_7	11
17	Точечная линейная регрессия для Y_8	12
18	Точечная линейная регрессия для Y_9	12

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. $y = [\min_{t \in S_i} S_i, \max_{t \in S_i} S_i]$, $y_i \in Y_1$.

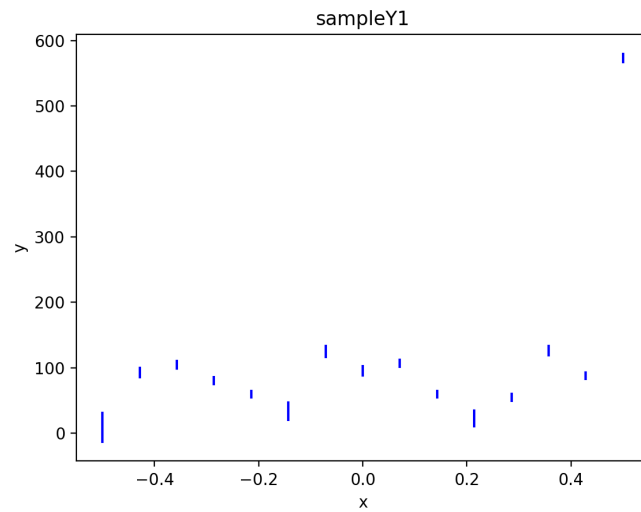


Рис. 1: Визуализация выборки Y_1

Индекс Жаккара выборки равен $JK(Y_1) = -0.052$

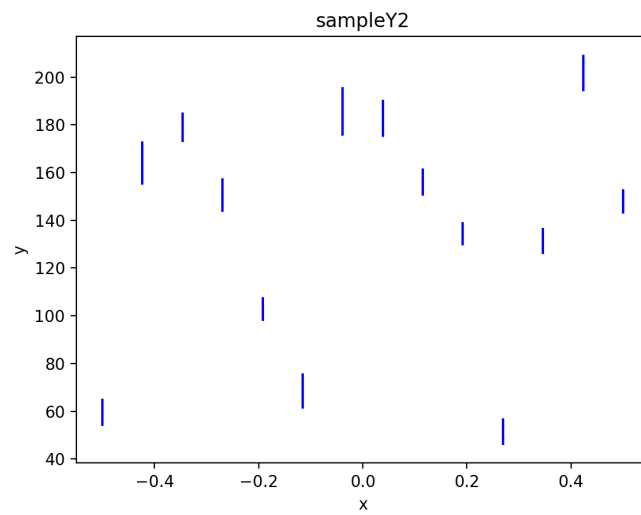


Рис. 2: Визуализация выборки Y_2

Индекс Жаккара выборки равен $JK(Y_2) = -0.088$

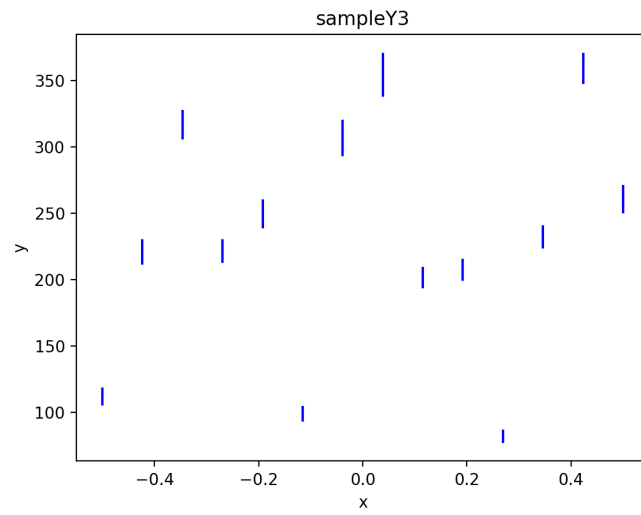


Рис. 3: Визуализация выборки Y_3

Индекс Жаккара выборки равен $JK(Y_3) = -0.057$

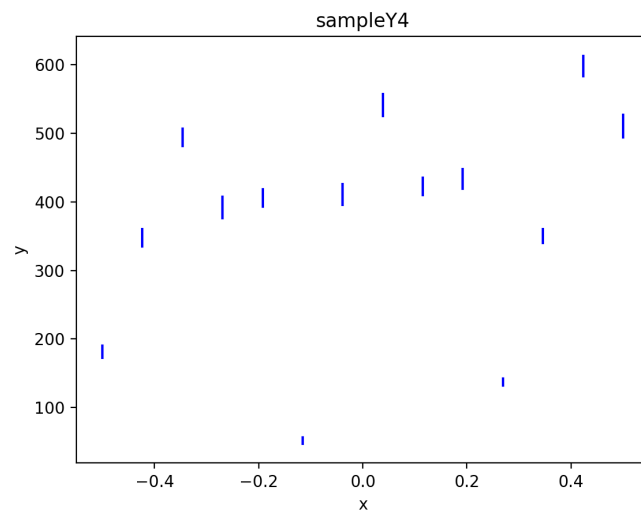


Рис. 4: Визуализация выборки Y_4

Индекс Жаккара выборки равен $JK(Y_4) = -0.036$

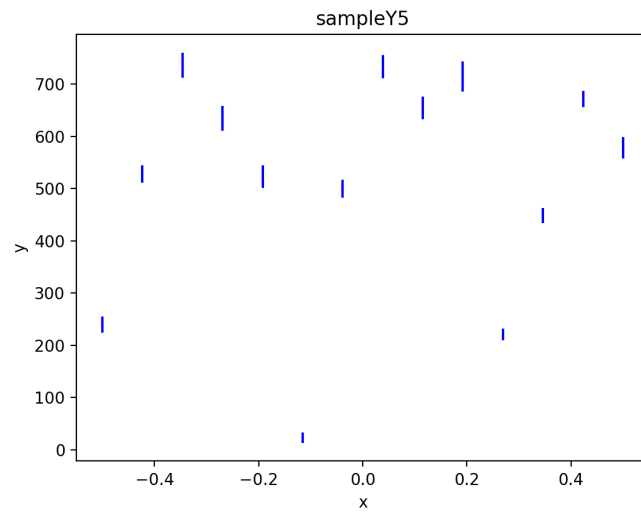


Рис. 5: Визуализация выборки Y_5

Индекс Жаккара выборки равен $JK(Y_5) = -0.04271$

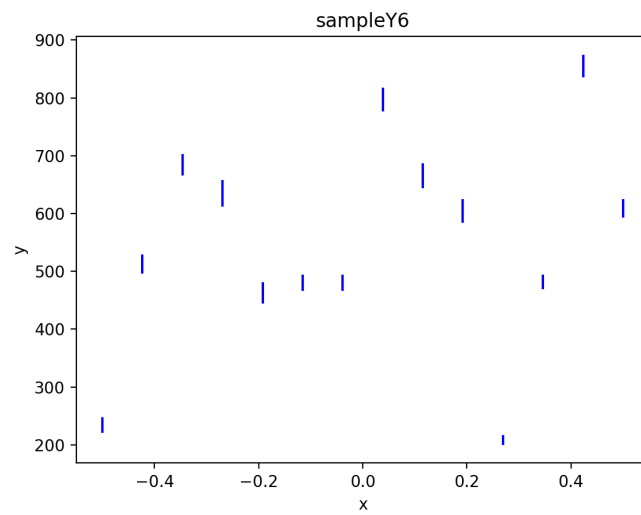


Рис. 6: Визуализация выборки Y_6

Индекс Жаккара выборки равен $JK(Y_6) = -0.0381$

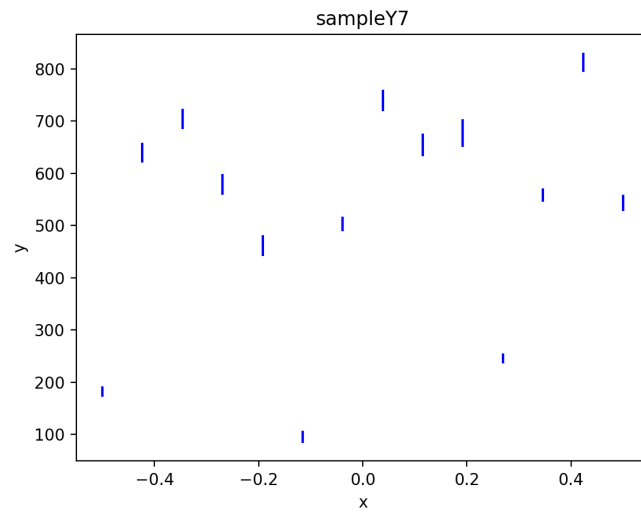


Рис. 7: Визуализация выборки Y_7

Индекс Жаккара выборки равен $JK(Y_7) = -0.0369$

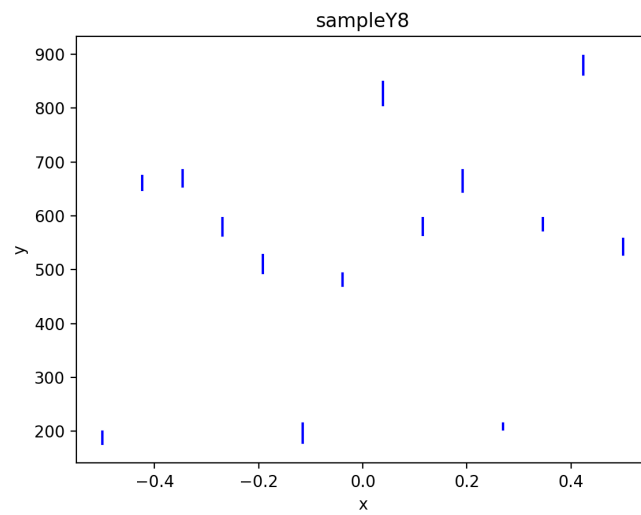


Рис. 8: Визуализация выборки Y_8

Индекс Жаккара выборки равен $JK(Y_8) = -0.0431$

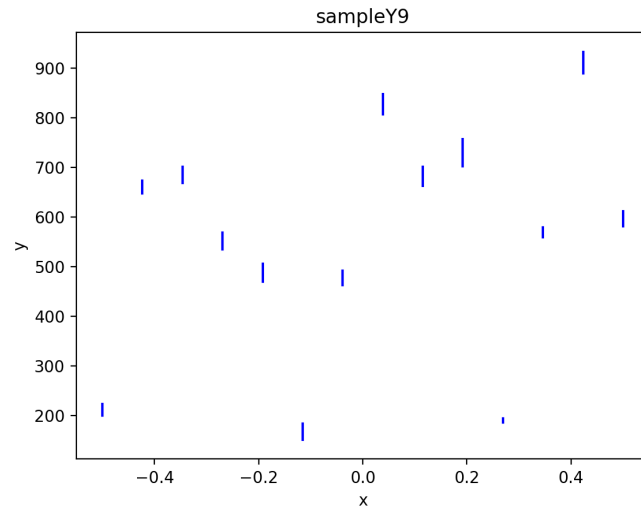


Рис. 9: Визуализация выборки Y_9

Индекс Жаккара выборки равен $JK(Y_9) = -0.0537$

Построим линейную регрессию, решив задачу 2 для выборок.

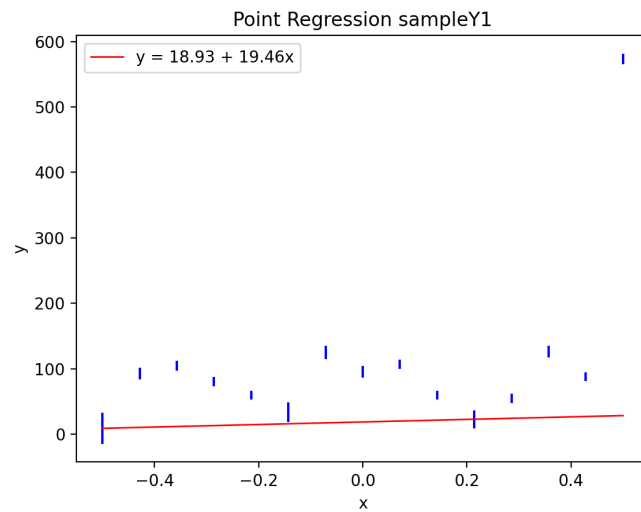


Рис. 10: Точечная линейная регрессия для Y_1

Для выборки Y_1 были получены оценки: $\beta_0 = 18.333, \beta_1 = 19.467$. Тогда полученная модель имеет вид $y = 18.333 + 19.467x$.

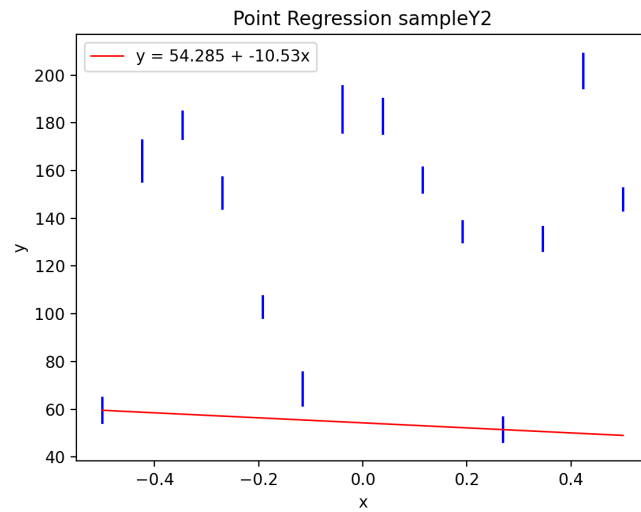


Рис. 11: Точечная линейная регрессия для Y_2

Для выборки Y_2 были получены оценки: $\beta_0 = 54.28, \beta_1 = -10.53$. Тогда полученная модель имеет вид $y = 54.28 - 10.53x$.

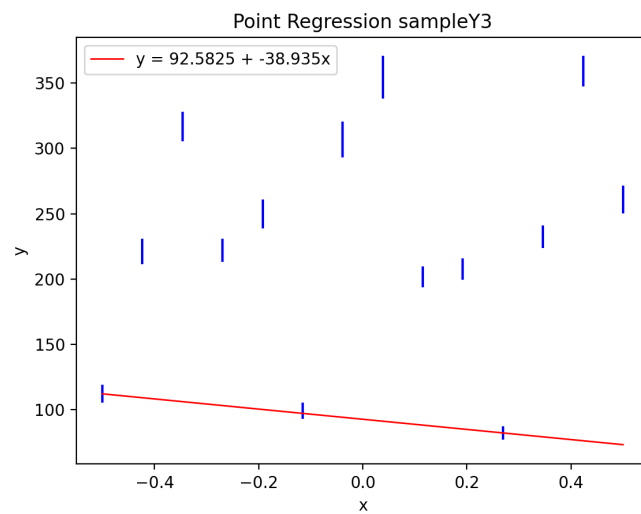


Рис. 12: Точечная линейная регрессия для Y_3

Для выборки Y_3 были получены оценки: $\beta_0 = 92.58, \beta_1 = -38.9$. Тогда полученная модель имеет вид $y = 92.58 - 38.93x$.

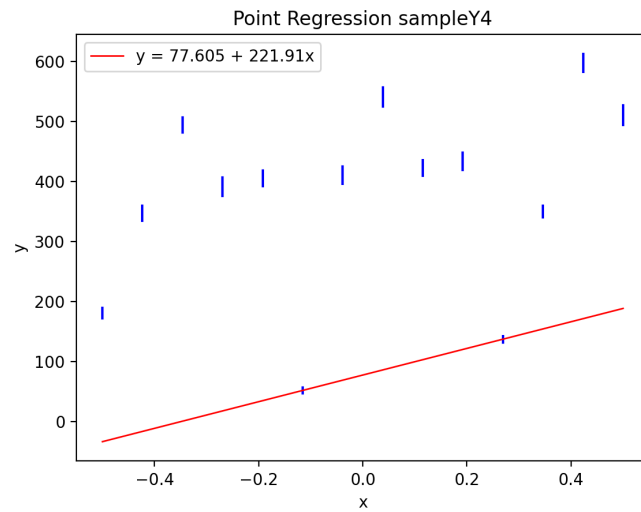


Рис. 13: Точечная линейная регрессия для Y_4

Для выборки Y_4 были получены оценки: $\beta_0 = 77.60, \beta_1 = 221.90$. Тогда полученная модель имеет вид $y = 77.60 + 221.90x$.

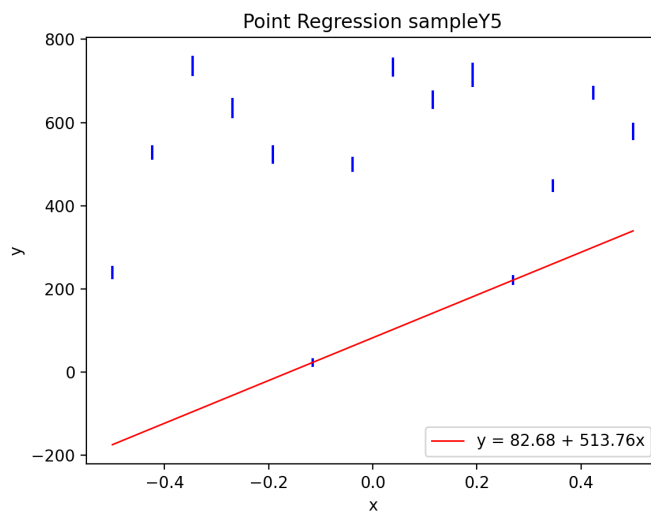


Рис. 14: Точечная линейная регрессия для Y_5

Для выборки Y_5 были получены оценки: $\beta_0 = 82.68, \beta_1 = 513.76$. Тогда полученная модель имеет вид $y = 82.68 + 513.76x$.

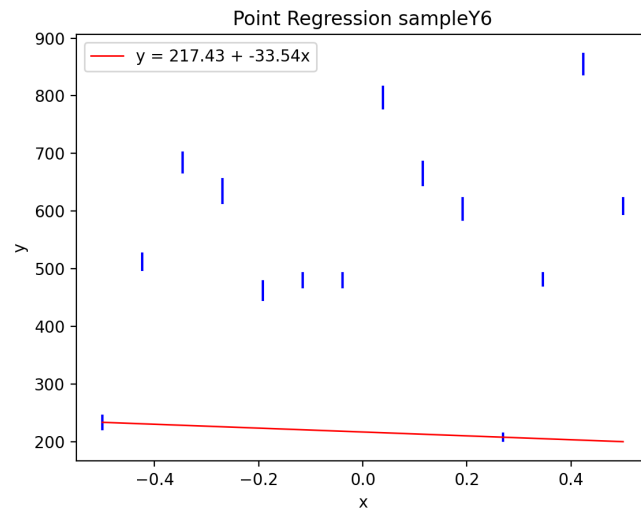


Рис. 15: Точечная линейная регрессия для Y_6

Для выборки Y_6 были получены оценки: $\beta_0 = 217.43, \beta_1 = -33.54$. Тогда полученная модель имеет вид $y = 217.43 - 33.54x$.

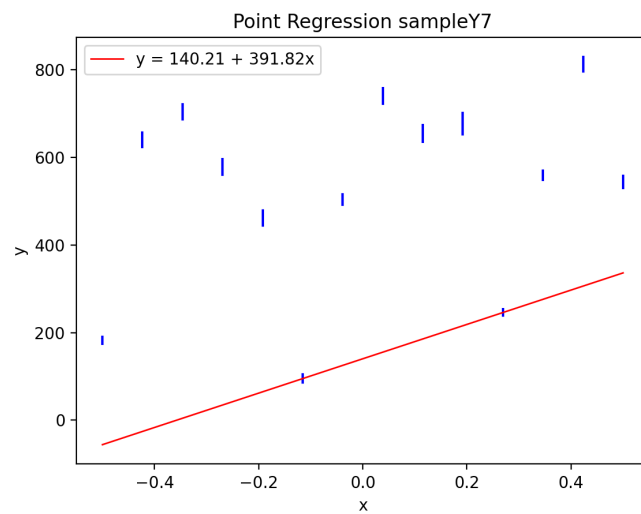


Рис. 16: Точечная линейная регрессия для Y_7

Для выборки Y_7 были получены оценки: $\beta_0 = 140.21, \beta_1 = 391.8$. Тогда полученная модель имеет вид $y = 140.21 + 391.8x$.

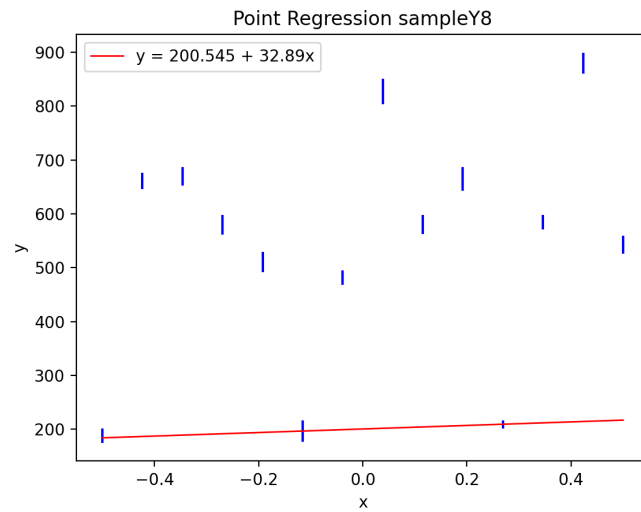


Рис. 17: Точечная линейная регрессия для Y_8

Для выборки Y_8 были получены оценки: $\beta_0 = 200.545, \beta_1 = 32.89$. Тогда полученная модель имеет вид $y = 200.545 + 32.89x$.

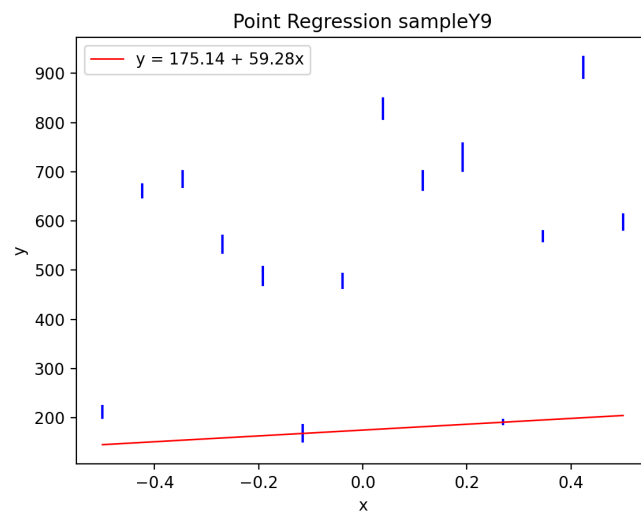


Рис. 18: Точечная линейная регрессия для Y_9

Для выборки Y_9 были получены оценки: $\beta_0 = 175.14, \beta_1 = 59.28$. Тогда полученная модель имеет вид $y = 175.14 + 59.28x$.

Информационные множества для данных выборок построить не получилось в виду несовместимости данных.

5 Обсуждение

Из полученных результатов в виде отрицательных индексов Жаккара для всех выборок и больших параметров β_0, β_1 можно сделать выводы о несовместимости данных.