

Санкт-Петербургский политехнический университет Петра Великого
Физико-механический институт

Направление подготовки
«01.04.02 Прикладная математика и информатика»

Отчёт
по лабораторной работе №3
по дисциплине
«Анализ данных с интервальной
неопределенностью»

Выполнила студентка гр. 5040102/20201

Харисова Т.А.

Преподаватель

Баженов А.Н.

Санкт-Петербург
2023

Оглавление

Постановка задачи.....	3
Теория.....	3
Реализация	3
Результаты	3
Анализ результатов	7

Постановка задачи

Дана интервальная выборка. Требуется провести анализ остатков.

Теория

Заданы $X = \{x_i\}_{i=1}^n$ – точечная выборка независимых входных данных, $Y = \{y_i\}_{i=1}^n$ – интервальная выборка выходных данных.

Коридор совместных зависимостей

$$Y_i = \beta_1 x_i + \beta_0, \quad i = 1..n, \quad (1)$$

где β_0, β_1 лежат в информационном множестве.

Для выполнения анализа остатков

$$[\underline{\varepsilon}_i, \overline{\varepsilon}_i] = y_i - Y_i, \quad i = 1..n \quad (2)$$

производится построение диаграммы статусов по следующим характеристикам.

Остаток:

$$r_i = \frac{\text{mid } y_i - \text{mid } Y_i}{\text{rad } y_i} \quad (3)$$

Размах:

$$l_i = \frac{\text{rad } Y_i}{\text{rad } y_i} \quad (4)$$

Границы статусов наблюдений задаются выражениями

$$|r_i| \leq 1 - l_i \quad (5)$$

$$|r_i| > 1 + l_i \quad (6)$$

$$l_i > 1 \quad (7)$$

Для внутренних наблюдений выполняется соотношение (5), для граничных наблюдений выполняется равенство в (5), для выбросов выполняется соотношение (6). При невыполнении условий (5) и (6) наблюдения считаются внешними, если к этому добавляется выполнение условия (7) – абсолютно внешними.

Реализация

Работа выполнена с помощью языка программирования Python в среде разработки Visual Studio Code. Ссылка на исходный код работы: [Lab 3 \(github.com\)](https://github.com).

Для построения выборки использовались данные из файлов “-0.45V_sp31.dat”, “-0.35V_sp670.dat”, “-0.25V_sp484.dat”, “-0.15V_sp831.dat”, “-0.05V_sp547.dat”, “0.05V_sp321.dat”, “0.15V_sp9.dat”, “0.25V_sp320.dat”, “0.35V_sp300.dat”, “0.45V_sp176.dat”. Значения погрешности δ_i брались из файла “0.0V_sp443.dat”.

Результаты

Исходные данные подвергаются предварительной коррекции: вместо y_i рассматриваются $\dot{y}_i - \delta_i, i = 1..n$, где δ_i – некоторая погрешность.

$x = [-0.45, -0.35, -0.25, -0.15, -0.05, 0.05, 0.15, 0.25, 0.35, 0.45]$, в качестве y берется мода данных из соответствующих файлов, обинтерваленных значением $\varepsilon = 300$. Назовем выборку Y_1 .

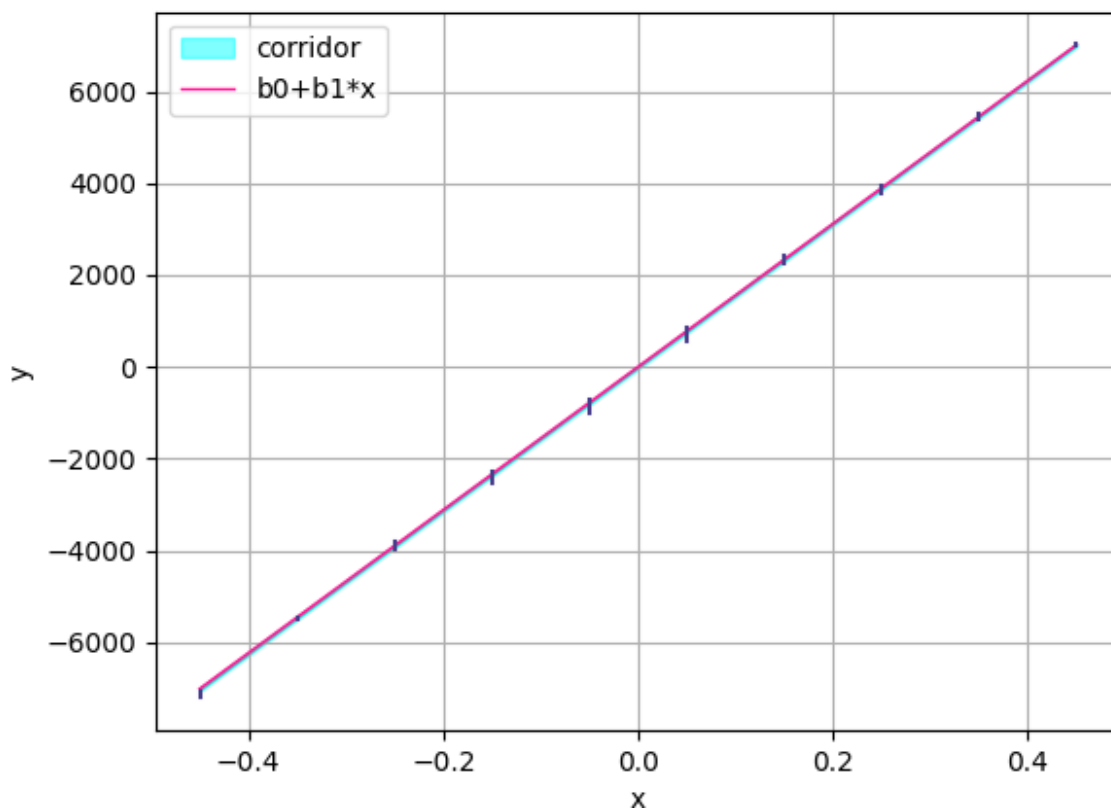


Рисунок 1. Выборка Y_1 и коридор совместных зависимостей

При построении линейной регрессии параметры весов для значений y_1 и y_2 были больше единицы.

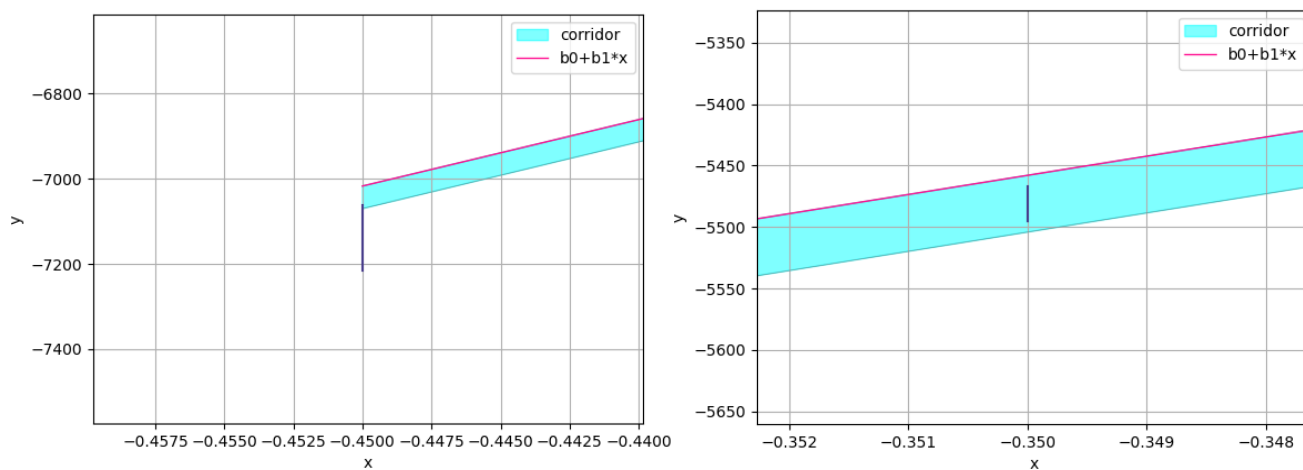


Рисунок 2. Коридор совместных зависимостей Y_1 в точках y_0 и y_1

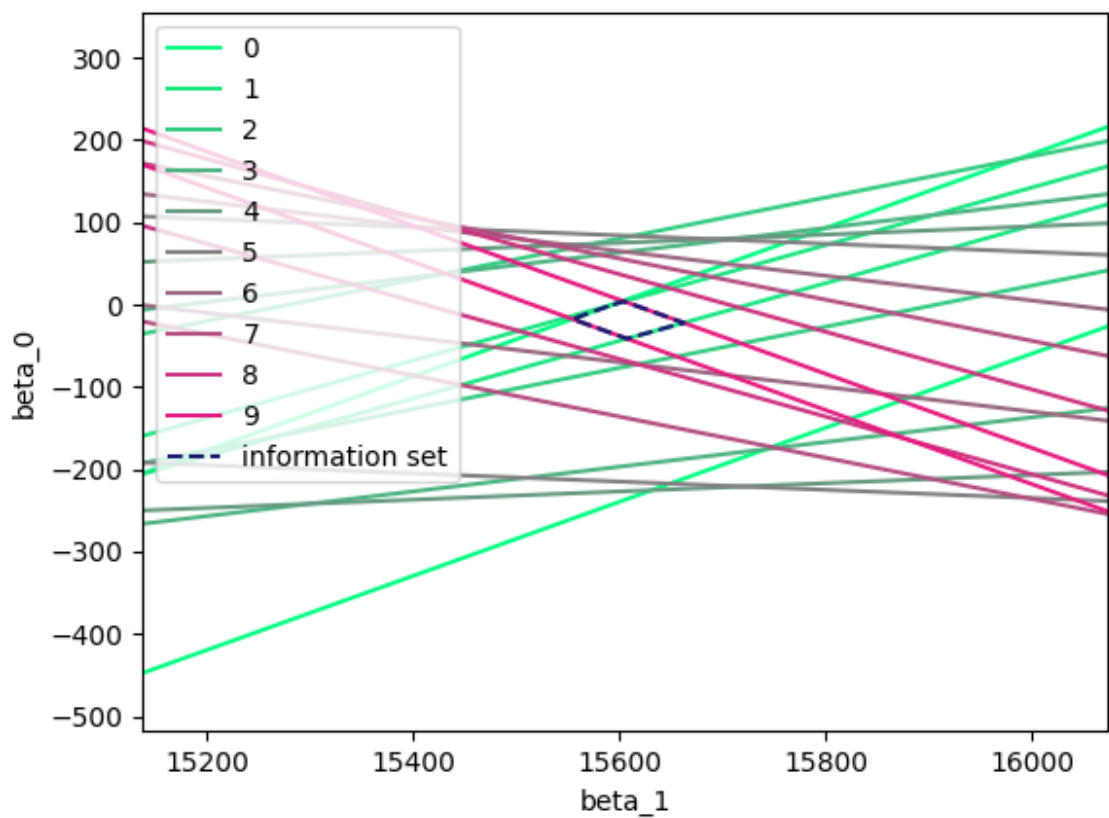


Рисунок 3. Информационное множество Y_1

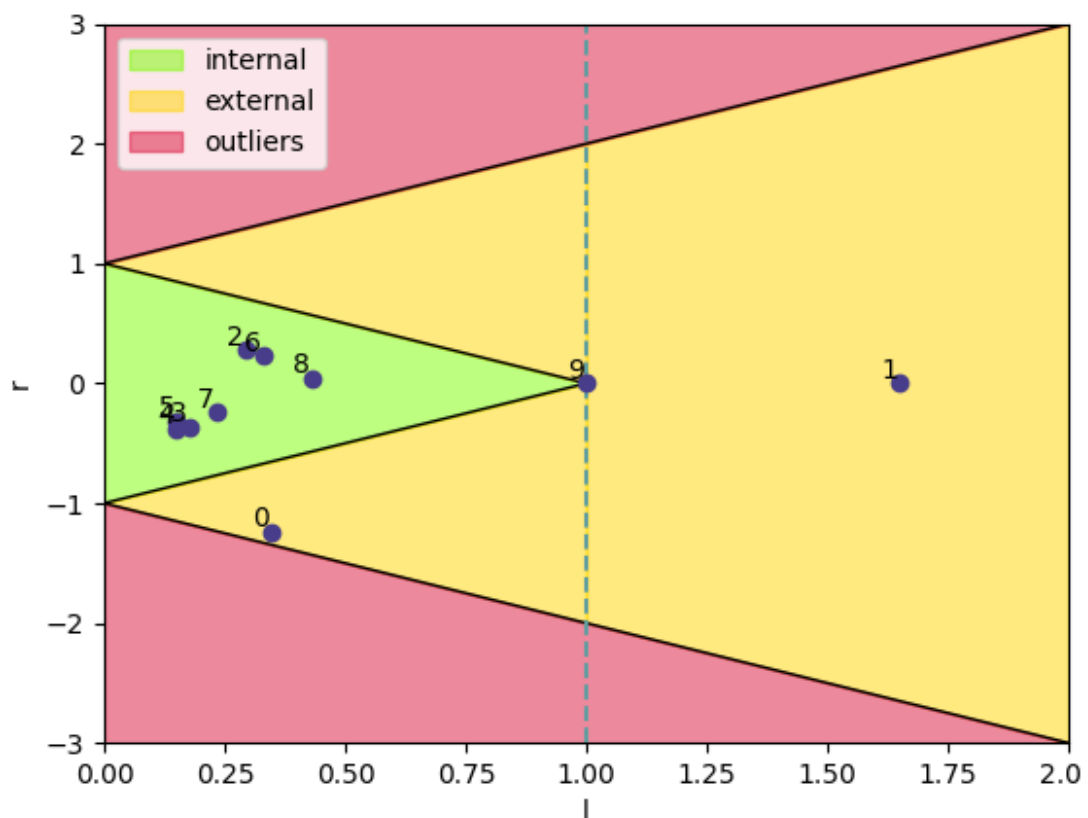


Рисунок 4. Диаграмма статусов для интервальных наблюдений Y_1

Наблюдение 9 является граничным, наблюдение 1 – абсолютно внешнее, наблюдение 0 внешнее, но близко к выбросам. Остальные наблюдения внутренние.

Рассмотрим для другой выборки, построенной аналогично по данным, обинтерваленным значением $\varepsilon = 450$. Назовем ее Y_2 .

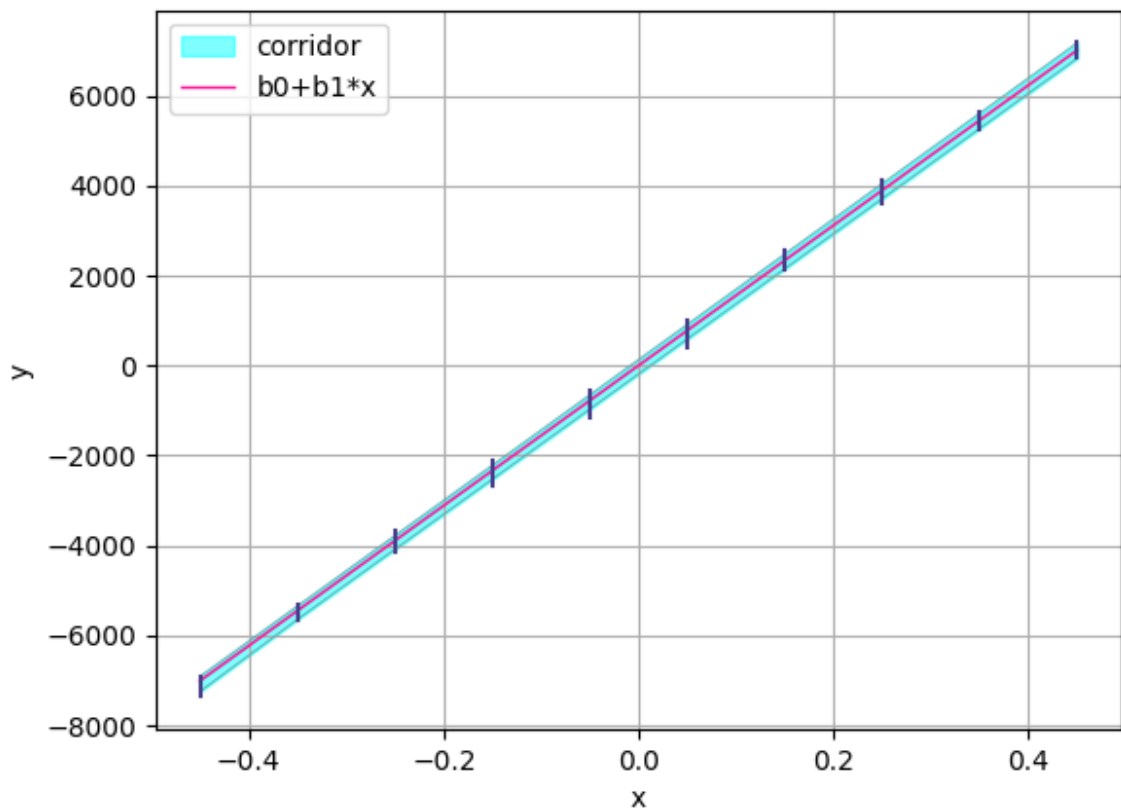


Рисунок 5. Выборка Y_2 и коридор совместных зависимостей

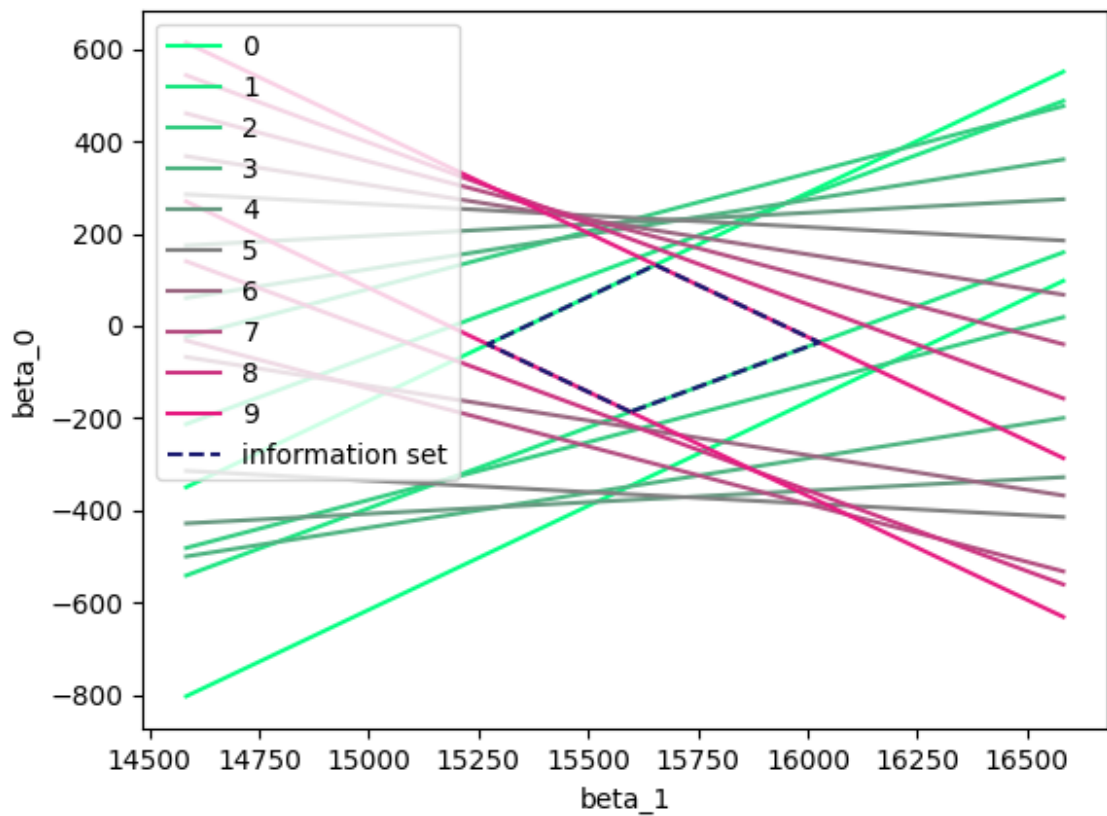


Рисунок 6. Информационное множество Y_2

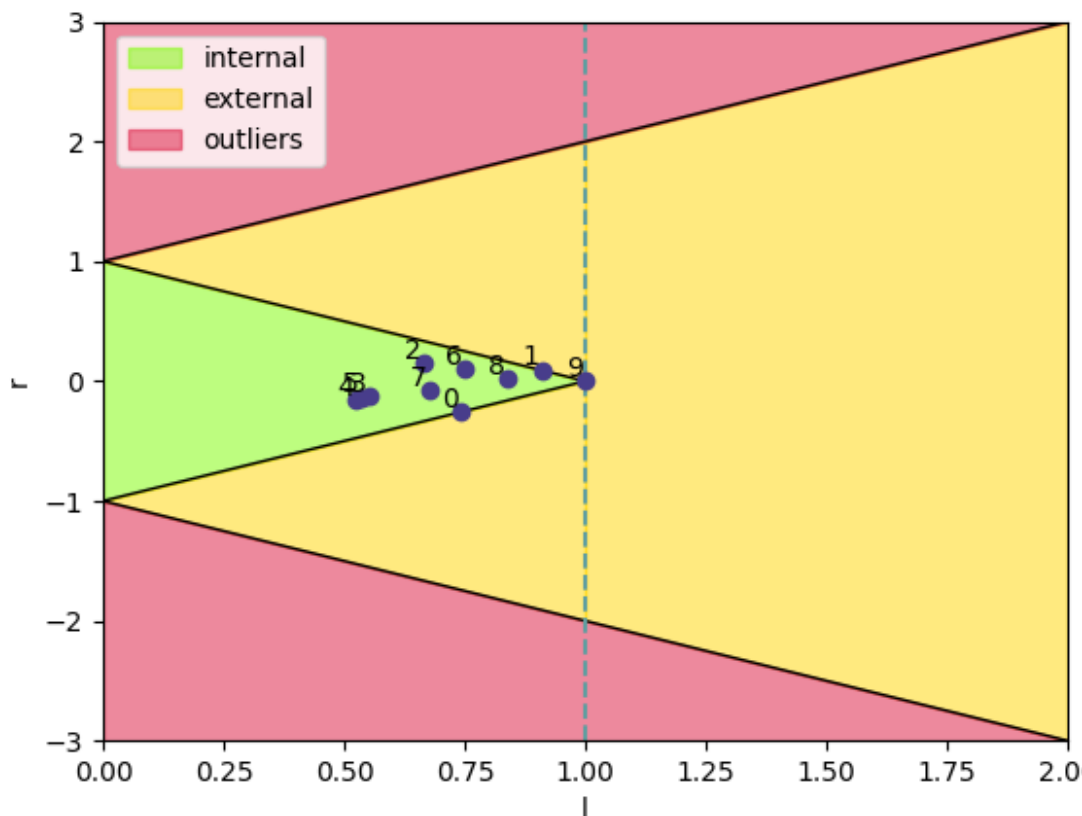


Рисунок 7. Диаграмма статусов для интервальных наблюдений Y_2

Наблюдения 0, 1 и 9 граничные, остальные наблюдения внутренние.

Анализ результатов

У второй выборки с более широкой интервальной неопределенностью исходных данных информационное множество и коридор совместных зависимостей также оказываются шире. Границы информационного множества задаются наблюдениями 0, 1 и 9 у обеих выборок. Наблюдение 9 задает две границы информационного множества и находится на угле линии граничных наблюдений. Наблюдения 0 и 1 по диаграмме статусов являются внешними для первой выборки и граничными для второй, что соответствует графику коридора совместных зависимостей.