

Санкт-Петербургский политехнический университет
Петра Великого

Физико-механический институт

Кафедра «Прикладная математика»

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

Выполнил студент:
Бочкарев Илья Алексеевич
группа: 5040102/20201

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

Санкт-Петербург
2023 г.

Содержание

1	Постановка задачи	2
2	Теория	2
3	Реализация	2
4	Результаты	3
5	Обсуждение	9

Список иллюстраций

1	Исходная интервальная выборка $X, (Y_1)$	3
2	Точечная линейная регрессия выборки $X, (Y_1)$	4
3	Информационное множество выборки $X, (Y_1)$	4
4	Коридор совместных зависимостей выборки $X, (Y_1)$	5
5	Диаграмма статусов выборки $X, (Y_1)$	5
6	Исходная интервальная выборка $X, (Y_2)$	6
7	Точечная линейная регрессия выборки $X, (Y_2)$	6
8	Информационное множество выборки $X, (Y_2)$	7
9	Коридор совместных зависимостей выборки $X, (Y_2)$	7
10	Диаграмма статусов выборки $X, (Y_2)$	8
11	Диаграмма статусов выборки $X, (Y_2)$ приближение	8

1 Постановка задачи

Имеется выборка $(X, (Y))$. X – множество вещественных чисел, Y – множество интервалов. Необходимо по построенному коридору совместных зависимостей провести анализ остатков.

2 Теория

Для выборки $(X, (Y))$, $X = \{x_i\}_{i=1}^n$, $Y = \{y_i\}_{i=1}^n$ (x_i – точный, y_i – интервальный) и $\Upsilon(x)$ – коридора совместных зависимостей определены следующие отношения.

Размах:

$$l(x, y) = \frac{\Upsilon(x)}{rad(y)} \quad (1)$$

Относительный остаток:

$$r(x, y) = \frac{mid(y) - mid(\Upsilon(x))}{rad(y)} \quad (2)$$

Границы статусов наблюдений задаются выражениями:

$$|r(x, y)| \leq 1 - l(x, y) \quad (3)$$

3 выполнено для внутренних наблюдений. Если достигается равенство, то наблюдение граничное.

$$|r(x, y)| > 1 + l(x, y) \quad (4)$$

4 выполнено для выбросов.

$$l(x, y) > 1 \quad (5)$$

При совместном невыполнении 3 и 4 наблюдение считается внешним. Если, кроме этого, выполнено еще и 5, то наблюдение – абсолютно внешнее.

3 Реализация

Весь код написан на языке Python (версии 3.9.5). [Ссылка на GitHub с исходным кодом](#).

4 Результаты

Данные были взяты из файлов *rawData/0.05V_sp321.dat*, *rawData/-0.05V_sp547.dat*, *rawData/0.15V_sp9.dat*, *rawData/-0.15V_sp831.dat*, *rawData/0.25V_sp320.dat*, *rawData/-0.25V_sp484.dat*, *rawData/0.35V_sp300.dat*, *rawData/-0.35V_sp670.dat*, *rawData/-0.45V_sp31.dat* и *rawData/0.45V_sp176.dat*. С коррекцией при помощи вспомогательных данных из файла *rawData/0.0V_sp812.dat*. Набор значений $X = [-0.45, -0.35, -0.25, -0.15, -0.05, 0.05, 0.15, 0.25, 0.35, 0.45]$. Набор значений Y_1 определяется как интервальная мода данных из соответствующих файлов (изначальные данные обынтерваливаются с $eps = 600$). Набор значений Y_2 определяется как обынтерваленное среднее из соответствующих файлов ($eps = 125$).

Начнем с Y_1 . Итоговая выборка:

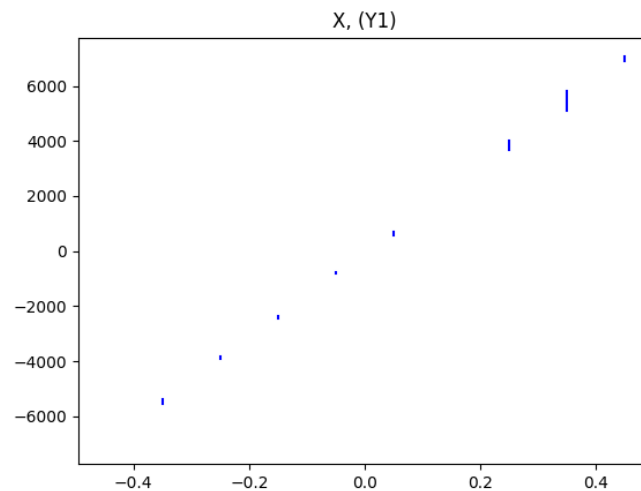


Рис. 1: Исходная интервальная выборка $X, (Y_1)$

Следует отметить, что 1 и 7 интервалы малы настолько, что отсутствуют на графике. Однако учитываются в дальнейших вычислениях.

Точечная линейная регрессия имеет вид:

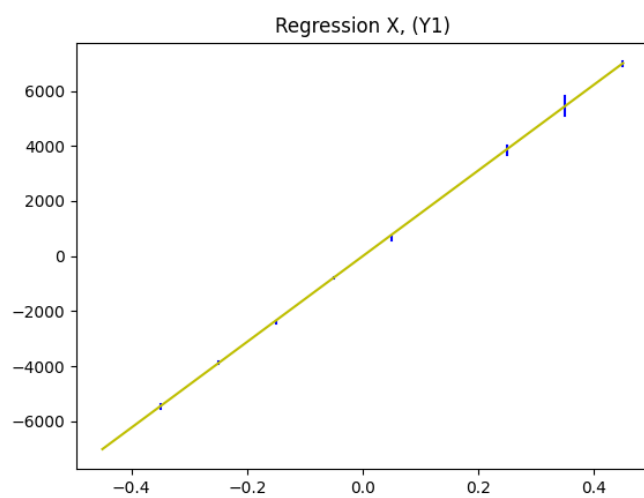


Рис. 2: Точечная линейная регрессия выборки $X, (Y_1)$

Построим информационное множество:

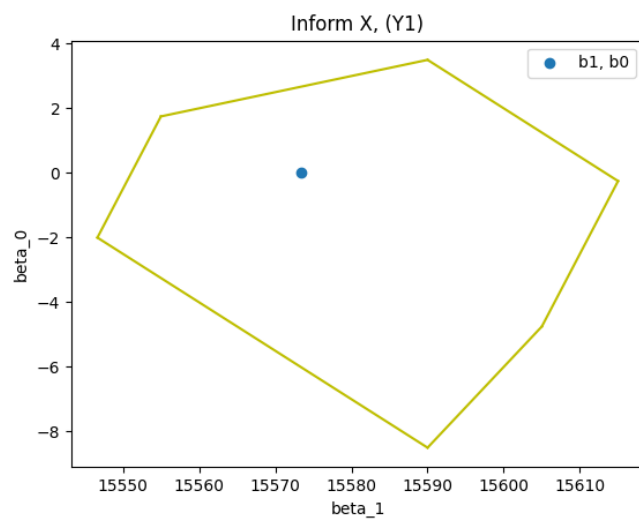


Рис. 3: Информационное множество выборки $X, (Y_1)$

Коридор совместных зависимостей:

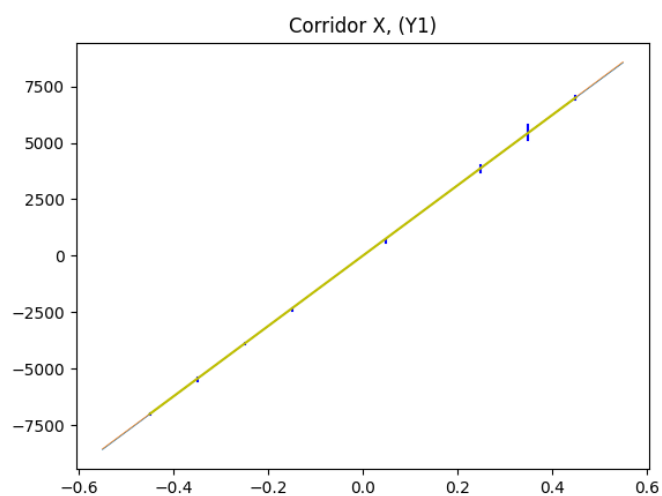


Рис. 4: Коридор совместных зависимостей выборки $X, (Y_1)$

4 и 6 наблюдения лежат вне коридора совместных зависимостей.
Строим диаграмму статусов:

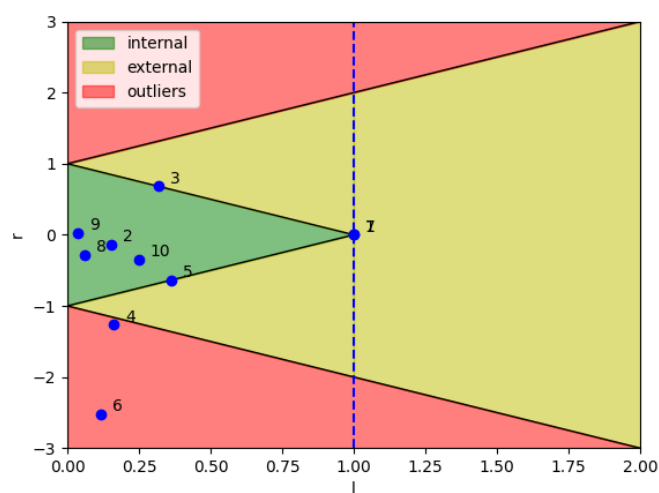


Рис. 5: Диаграмма статусов выборки $X, (Y_1)$

Наблюдения 4, 6 являются выбросами. 1, 3, 5, 7 наблюдения граничные, остальные внутренние.

Теперь Y_2 . Итоговая выборка:

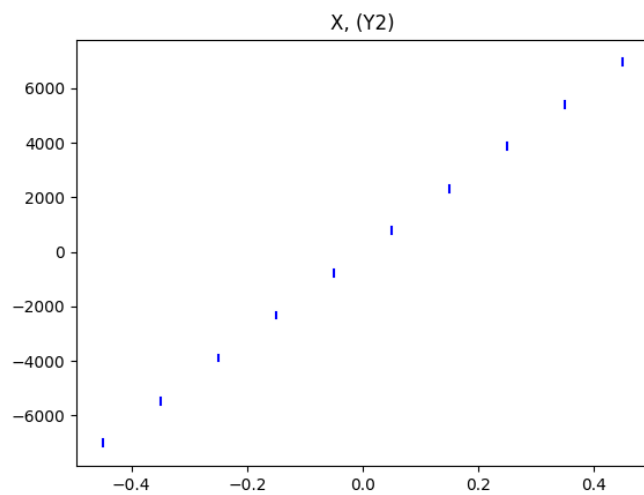


Рис. 6: Исходная интервальная выборка $X, (Y_2)$

Точечная линейная регрессия имеет вид:

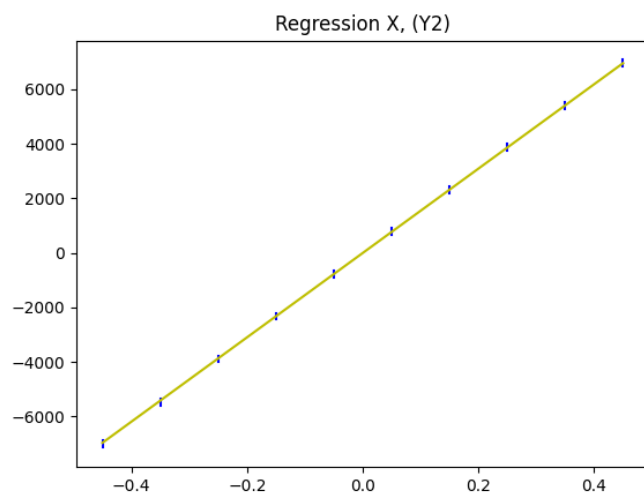


Рис. 7: Точечная линейная регрессия выборки $X, (Y_2)$

Построим информационное множество:

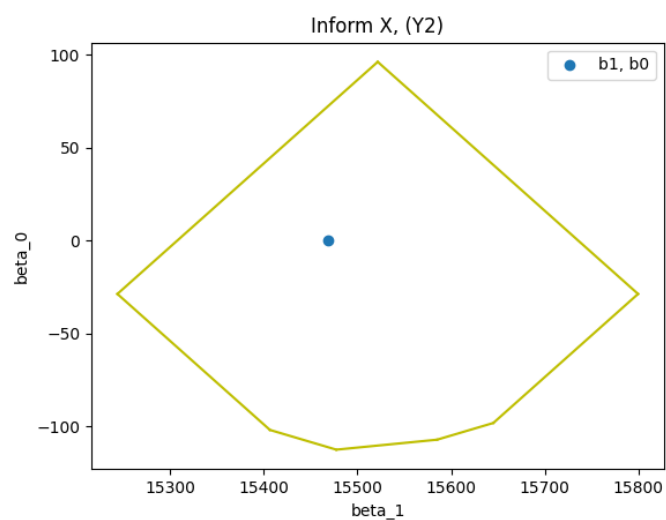


Рис. 8: Информационное множество выборки $X, (Y_2)$

Коридор совместных зависимостей:

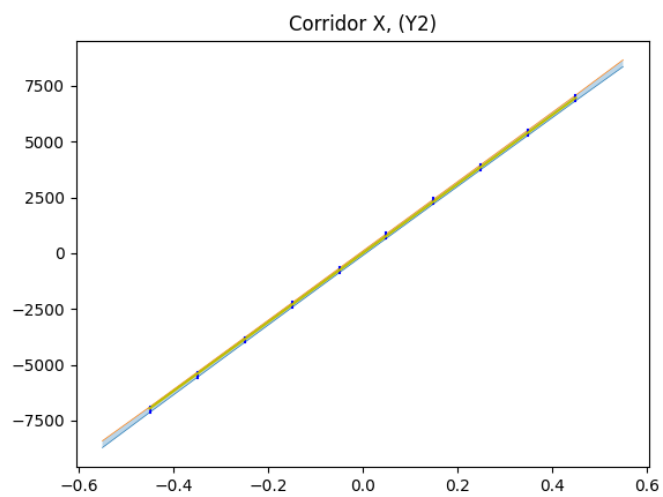


Рис. 9: Коридор совместных зависимостей выборки $X, (Y_2)$

Строим диаграмму статусов:

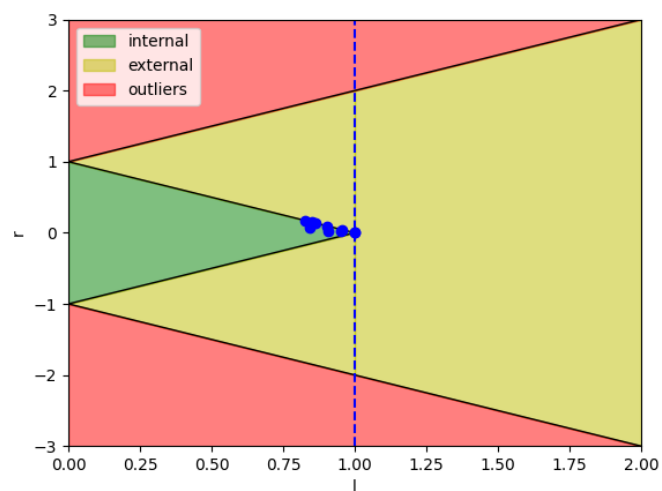


Рис. 10: Диаграмма статусов выборки $X, (Y_2)$

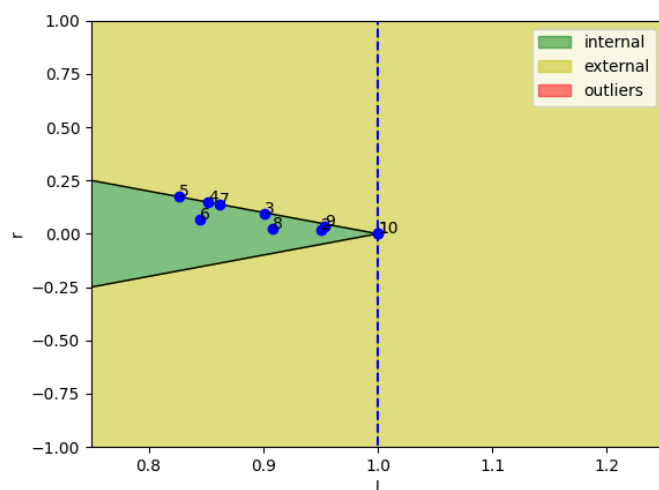


Рис. 11: Диаграмма статусов выборки $X, (Y_2)$ приближение

Наблюдения 1, 4, 5, 7, 10 являются граничными, остальные внутренние.

5 Обсуждение

Из полученных результатов можно заметить, что диаграммы статусов соотносятся с построенными информационными множествами: количество наблюдений, задающих информационное множество, совпадает с количеством наблюдений со статусом "граничное". Также статусы совпадают с положением наблюдений в коридоре совместных зависимостей.