

Санкт-Петербургский Политехнический университет  
Петра Великого

Физико-механический институт

Высшая школа прикладной математики и вычислительной физики

**Отчёт**  
**по лабораторной работе №3**  
**по дисциплине**  
**"Анализ данных с интервальной**  
**неопределенностью"**

Выполнил студент:

Иванов Андрей Игоревич,  
группа 5040102\20201

Проверил:

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

Санкт-Петербург, 2023

# Содержание

|          |   |          |
|----------|---|----------|
| <b>1</b> | <b>Постановка задачи</b>  | <b>3</b> |
| <b>2</b> | <b>Теория</b>   | <b>4</b> |
| 2.1      | Формирование интервальной выборки . . . . .                             | 4        |
| 2.2      | Формирование выборки остатков . . . . .                                 | 4        |
| 2.3      | Точечная линейная регрессия . . . . .                                   | 4        |
| 2.4      | Информационное множество . . . . .                                      | 4        |
| 2.5      | Классификация измерений . . . . .                                       | 5        |
| 2.6      | Взаимные отношения интервалов наблюдения и прогнозного интервала модели | 5        |
| <b>3</b> | <b>Реализация</b>   | <b>7</b> |
| <b>4</b> | <b>Результаты</b>   | <b>8</b> |
| 4.1      | Интервальная выборка . . . . .  | 8        |
| 4.2      | Выборка остатков . . . . .  | 10       |

## Список иллюстраций

|   |  |    |
|---|--|----|
| 1 | График интервальной выборки . . . . .                          | 8  |
| 2 | Точечная регрессия интервальной выборки . . . . .              | 8  |
| 3 | Информационное множество для интервальной выборки . . . . .    | 9  |
| 4 | Коридор совместных значений для интервальной выборки . . . . . | 9  |
| 5 | График выборки остатков . . . . .                              | 10 |
| 6 | Точечная регрессия выборки остатков . . . . .                  | 11 |
| 7 | Информационное множество для выборки остатков . . . . .        | 11 |
| 8 | Коридор совместных значений для выборки остатков . . . . .     | 12 |
| 9 | Диаграмма статусов для выборки остатков . . . . .              | 12 |

# 1 Постановка задачи

Дано множество вещественных выборок, соответствующих показаниям  $-0.45, -0.35, -0.25, -0.15, 0.05, 0.05, 0.15, 0.25, 0.35, 0.45$ . Необходимо:

- Сформировать интервальную выборку по имеющимся данным и выборку остатков по полученной интервальной;
- Найти точечную линейную регрессию для обеих выборок;
- Построить информационное множество коэффициентов регрессии (решить задачу восстановления зависимости) для обеих выборок
- Построить коридор совместных зависимостей задачи восстановления для обеих выборок
- Построить диаграмму статусов для выборки остатков

## 2 Теория

### 2.1 Формирование интервальной выборки

Дано множество из  $N$  выборок вещественных чисел  $\{\mathbf{X}_i\}_{i=1}^N$ . По этому множеству формируется интервальная выборка по следующему принципу:

$$X = \{(min(\mathbf{X}_i), max(\mathbf{X}_i)) | \mathbf{X}_i \in \{\mathbf{X}_i\}_{i=1}^N\} \quad (1)$$

### 2.2 Формирование выборки остатков

Дана интервальная выборка  $Y$  и ее регрессионная модель. Выборка остатков формируется следующим образом:

$$X_{res} = \{y_i - (\beta_0 + \beta_1 \cdot x_i) | y_i \in Y, x_i \in \{values\}\} \quad (2)$$

### 2.3 Точечная линейная регрессия

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

$$y = \beta_0 + \beta_1 x \quad (3)$$

Поставим задачу оптимизацию для нахождения точечных оценок параметров  $\beta_0, \beta_1$ .

$$\begin{aligned} \sum_{i=1}^m w_i &\rightarrow \min \\ \text{mid}\mathbf{y}_i - w_i \cdot \text{rad}\mathbf{y}_i &\leq \beta_0 + \beta_1 x \leq \text{mid}\mathbf{y}_i + w_i \cdot \text{rad}\mathbf{y}_i \\ w_i &\geq 0, i = 1, \dots, m \\ w_i, \beta_0, \beta_1 &-? \end{aligned} \quad (4)$$

Задачу можно решить методами линейного программирования.

### 2.4 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  $\Upsilon$ , сопоставляющее каждому значению аргумента  $x$  множество

$$\Upsilon(x) = \bigcup_{\beta \in \Omega} f(x, \beta) \quad (5)$$

, где  $\Omega$  - информационное множество,  $x$  - вектор переменных,  $\beta$  - вектор оцениваемых параметров.

## 2.5 Классификация измерений

Измерения можно классифицировать следующим образом. Измерения, добавление которых к выборке не приводит к модификации модели, называются *внутренними*. Те, которые изменяют модель, называются *внешними*. Измерения, которые определяют какую-либо границу информационного множества, называются *граничными*. *Выбросами* называются те измерения, которые делают информационное множество пустым. Граничные измерения - подмножество внутренних, выбросы - внешних.

Для удобства анализа взаимоотношения информационных множеств работу с ними заменяют на анализ взаимоотношения интересующего интервального измерения и интервального прогнозируемого значения модели (коридора совместных значений).

## 2.6 Взаимные отношения интервалов наблюдения и прогнозного интервала модели

Существует несколько характеристик, определяющих это взаимоотношение.

*Размахом (плечом)* называется следующее отношение:

$$l(x, \mathbf{y}) = \frac{\Upsilon(x)}{rad(\mathbf{y})} \quad (6)$$

*Относительным остатком* называется отношение:

$$r(x, \mathbf{y}) = \frac{mid(\mathbf{y}) - mid(\Upsilon(x))}{rad(\mathbf{y})} \quad (7)$$

здесь  $x$  - точечное значение,  $\mathbf{y}$  - интервальное значение интересующей величины (отклик  $x$ ),  $\Upsilon(x)$  - интервальная оценка интересующей величины (значение коридора совместных значений).

Для внутренних наблюдений выполняется неравенство:

$$|r(x, \mathbf{y})| \leq 1 - l(x, (y)) \quad (8)$$

В случае равенства

$$|r(x, \mathbf{y})| = 1 - l(x, (y)) \quad (9)$$

измерение будет граничным.

Выбросы определяются неравенством

$$|r(x, \mathbf{y})| > 1 + l(x, \mathbf{y}) \quad (10)$$

### 3 Реализация

Лабораторная работа выполнена на языке Python 3.10 с помощью загружаемых пакетов NumPy, Matplotlib, SciPy. Исходный код лабораторной работы находится на GitHub репозитории.



## 4 Результаты

### 4.1 Интервальная выборка

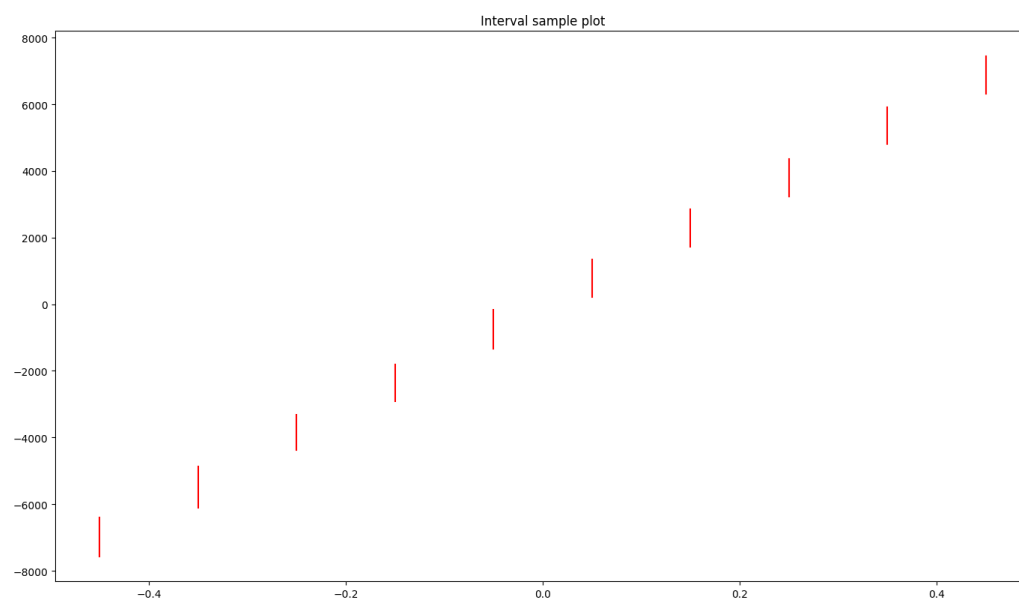


Рис. 1: График интервальной выборки

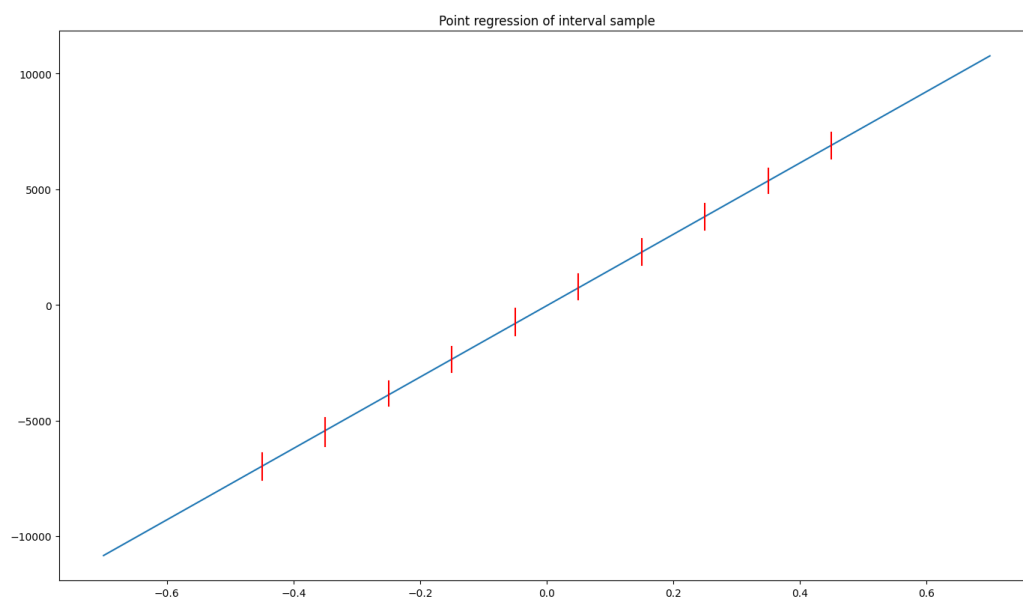


Рис. 2: Точечная регрессия интервальной выборки

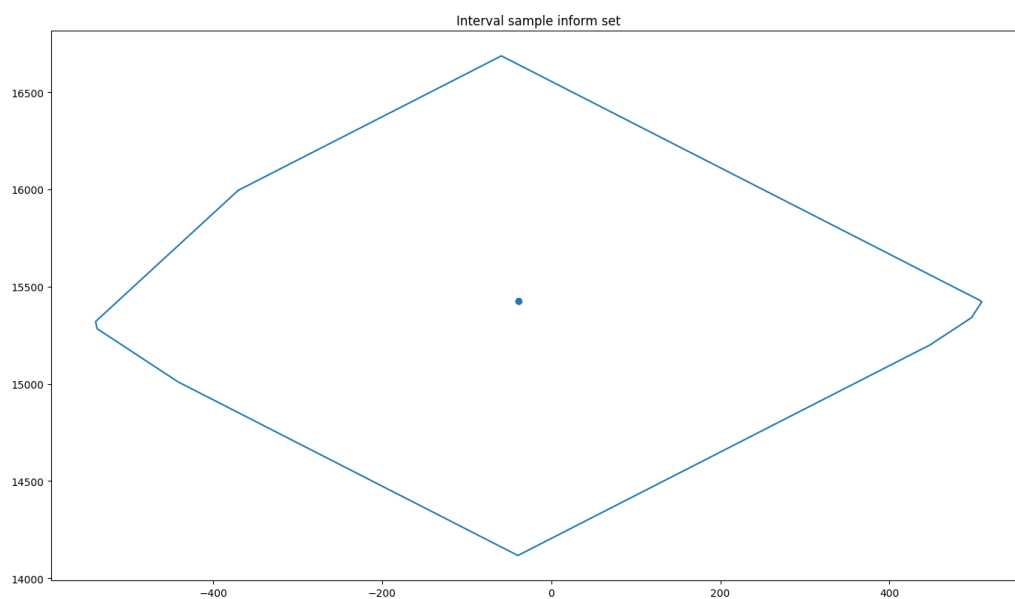


Рис. 3: Информационное множество для интервальной выборки

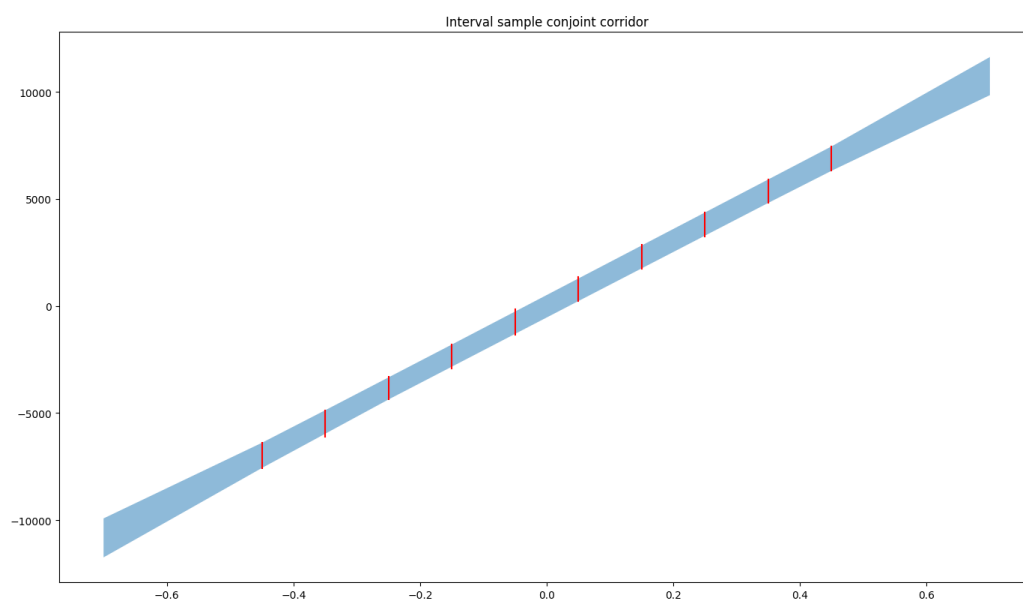


Рис. 4: Коридор совместных значений для интервальной выборки

Ниже приведена таблица, описывающая итоговую интервальную выборку

| Набор данных     | $\underline{x_i}$ | $\overline{x_i}$ |
|------------------|-------------------|------------------|
| -0.45V_sp115.dat | -7568             | -6392            |
| -0.35V_sp196.dat | -6117             | -4872            |
| -0.25V_sp403.dat | -4369             | -3307            |
| -0.15V_sp155.dat | -2906             | -1804            |
| -0.05V_sp465.dat | -1338             | -155             |
| 0.05V_sp321.dat  | 227               | 1335             |
| 0.15V_sp135.dat  | 1721              | 2843             |
| 0.25V_sp135.dat  | 3235              | 4368             |
| 0.35V_sp300.dat  | 4812              | 5907             |
| 0.45V_sp1014.dat | 6313              | 7450             |

Искомая модель принимает вид:

$$y = -39.03125 + 15424.374999 \cdot x \quad (11)$$

## 4.2 Выборка остатков

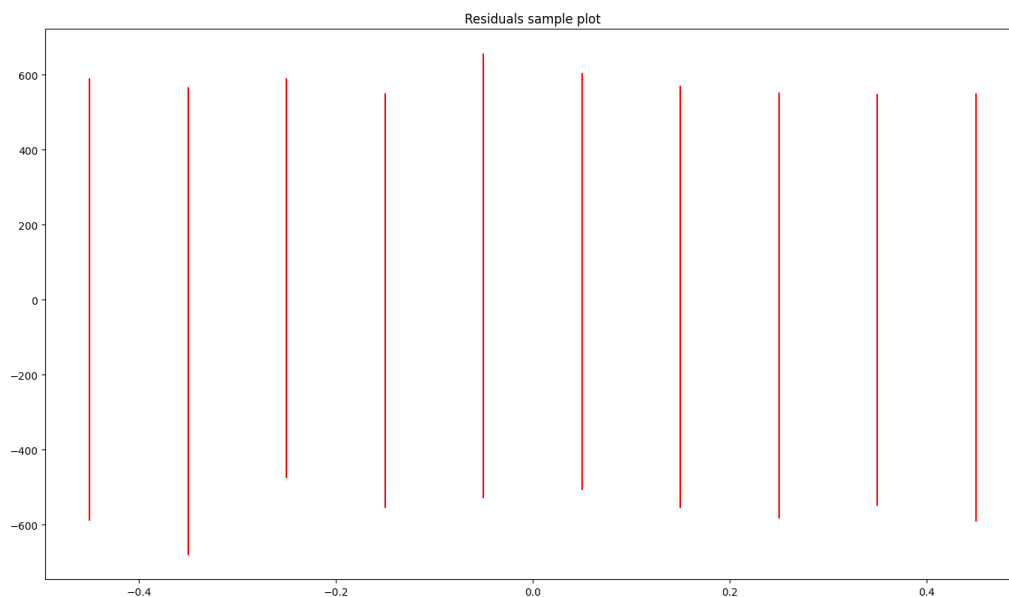


Рис. 5: График выборки остатков

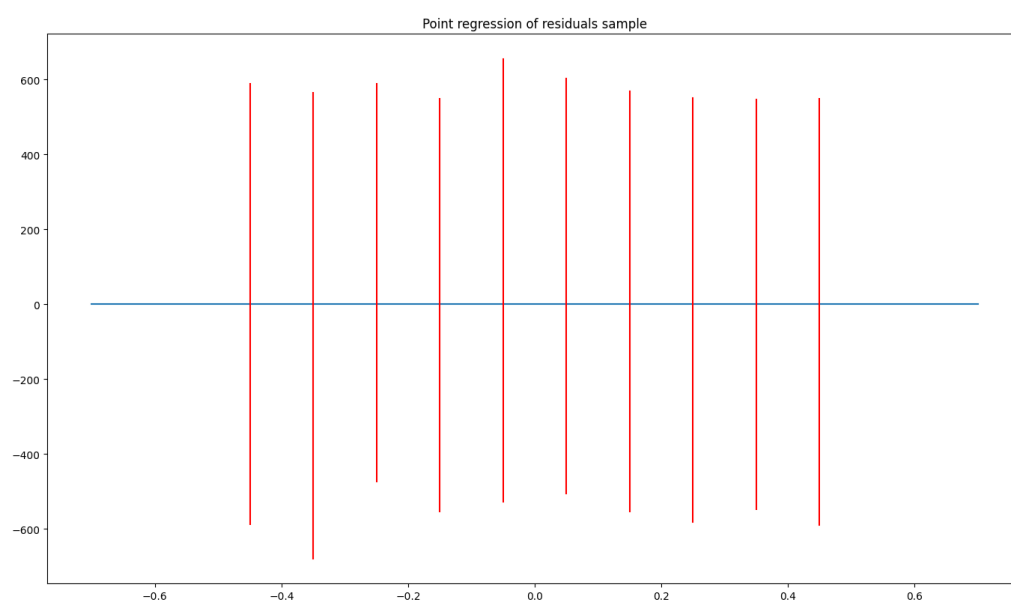


Рис. 6: Точечная регрессия выборки остатков

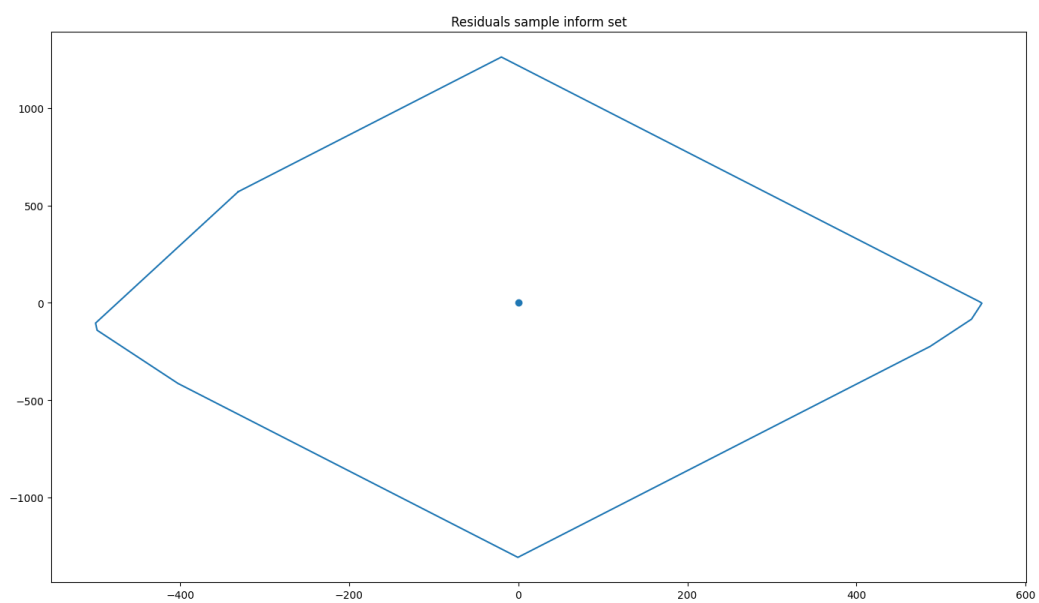


Рис. 7: Информационное множество для выборки остатков

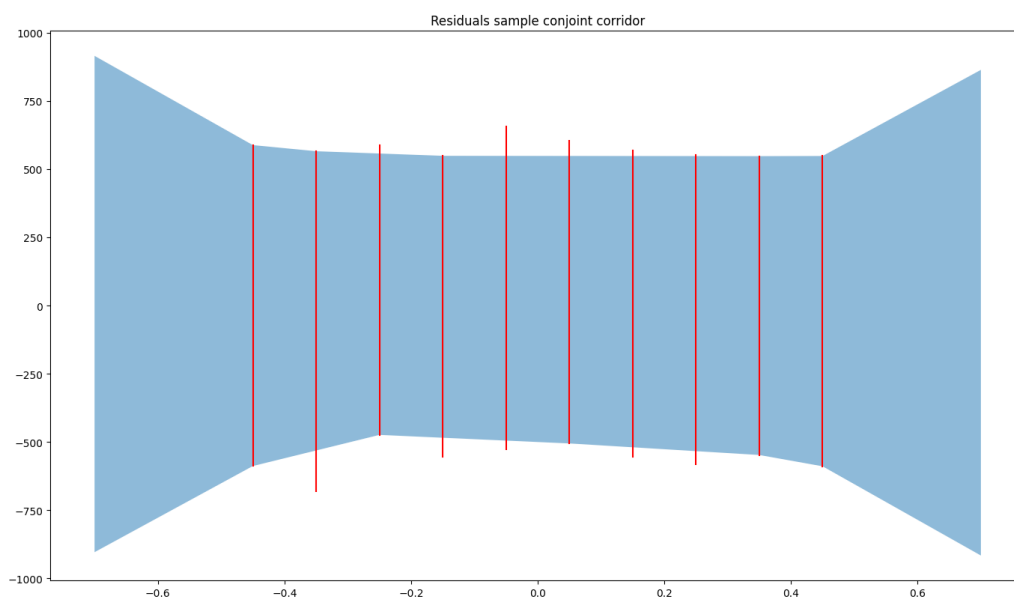


Рис. 8: Коридор совместных значений для выборки остатков

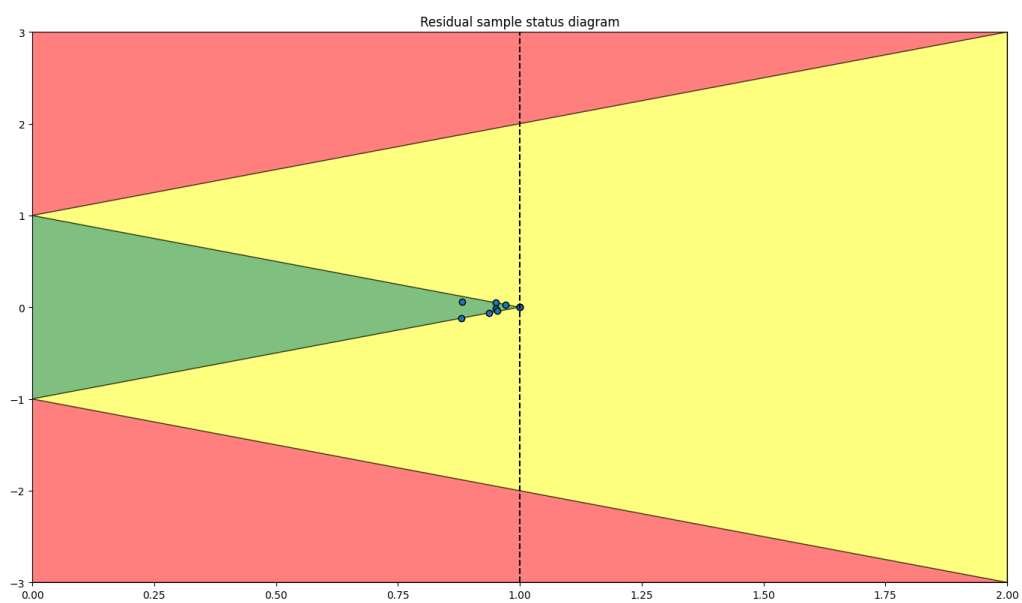


Рис. 9: Диаграмма статусов для выборки остатков

Ниже приведена таблица, описывающая выборку остатков:

| Набор данных     | $\underline{x}_i$  | $\overline{x}_i$  |
|------------------|--------------------|-------------------|
| -0.45V_sp115.dat | -588.000000538862  | 587.999999461138] |
| -0.35V_sp196.dat | -679.4375004133717 | 565.5624995866283 |
| -0.25V_sp403.dat | -473.8750002878801 | 588.1249997121199 |
| -0.15V_sp155.dat | -553.3125001623894 | 548.6874998376106 |
| -0.05V_sp465.dat | -527.7500000368989 | 655.2499999631011 |
| 0.05V_sp321.dat  | -505.1874999114084 | 602.8125000885916 |
| 0.15V_sp135.dat  | -553.6249997859177 | 568.3750002140823 |
| 0.25V_sp135.dat  | -582.062499660427  | 550.937500339573  |
| 0.35V_sp300.dat  | -547.4999995349363 | 547.5000004650637 |
| 0.45V_sp1014.dat | -588.937499409446  | 548.062500590554  |

Искомая модель принимает вид:

$$y = 0.000001 \cdot x \quad (12)$$