# Санкт-Петербургский Политехнический университет Петра Великого

#### Физико-механический институт

Высшая школа прикладной математики и вычислительной физики

# Отчёт по лабораторной работе №2 по дисциплине "Анализ данных с интервальной неопределенностью"

Выполнил студент:

Иванов Андрей Игоревич, группа  $5040102 \setminus 20201$ 

Проверил:

к.ф.-м.н., доцент

Баженов Александр Николаевич

# Содержание

1	Пос	становка задачи	3
2 Теория			
	2.1	Формирование интервальной выборки	4
	2.2	Точечная линейная регрессия	4
	2.3	Информационное множество	4
3	Pea	лизация	5
4	Рез	ультаты	6

# Список иллюстраций

1	График интервальной выборки	6
2	Точечная линейная регрессия	6
3	Информационное множество	7
4	Коридор совместных значений	7

## 1 Постановка задачи

Даны четыре вещественные выборки, соответствующие показаниям -0.5, -0.25, 0.25, 0.5. Необходимо:

- Сформировать интервальную выборку по имеющимся данным.
- Найти точечную линейную регрессию интервальной выборки
- Построить информационное множество коэффициентов регрессии (решить задачу восстановления зависимости)
- Построить коридор совместных зависимостей задачи восстановления

## 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:

$$X = \{ (min(\mathbf{X_i}), max(\mathbf{X_i})) \mid \mathbf{X_i} \in \{\mathbf{X_i}\}_{i=1}^N \}$$
 (1)

#### 2.2 Точечная линейная регрессия

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

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

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

$$\sum_{i=1}^{m} w_i \to \min$$

$$\operatorname{mid} \mathbf{y}_i - w_i \cdot \operatorname{rad} \mathbf{y}_i \le \beta_0 + \beta_1 x \le \operatorname{mid} \mathbf{y}_i + w_i \cdot \operatorname{rad} \mathbf{y}_i$$

$$w_i \ge 0, i = 1, ..., m$$

$$w_i, \beta_0, \beta_1 - ?$$

$$(3)$$

Задачу можно решить методами линейного программирования.

### 2.3 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

$$\Upsilon(x) = \bigcup_{\beta \in \Omega} f(x, \beta) \tag{4}$$

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

## 3 Реализация

Лабораторная работа выполнена на языке Python  $3.10\ c$  помощью загружаемых пакетов NumPy и MatPlotLib. Исходный код лабораторной работы находится на GitHub репозитории.

## 4 Результаты

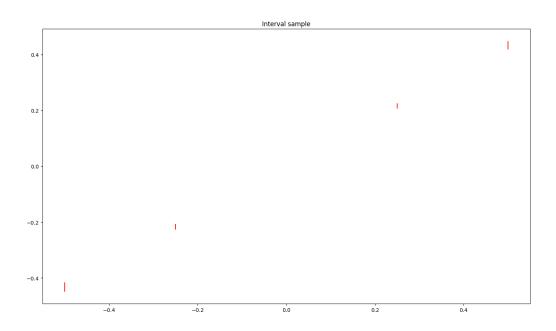


Рис. 1: График интервальной выборки

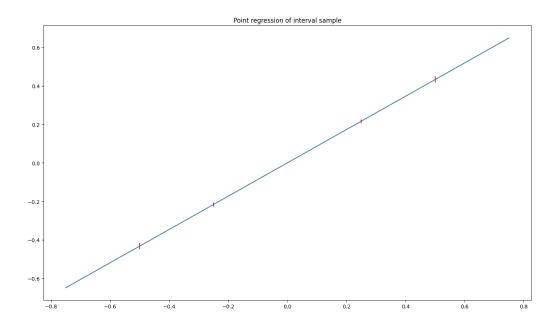


Рис. 2: Точечная линейная регрессия

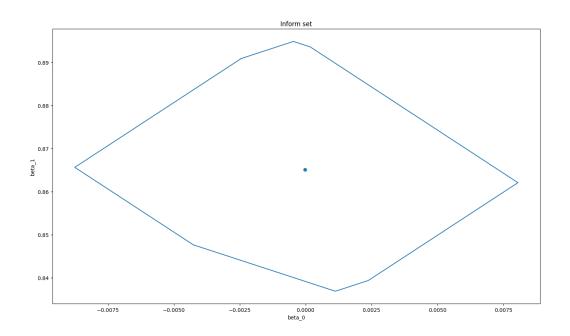


Рис. 3: Информационное множество

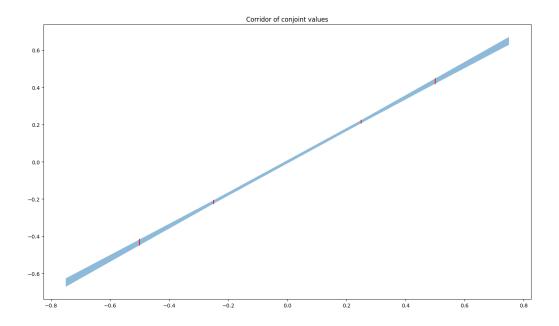


Рис. 4: Коридор совместных значений

Ниже приведена таблица, описывающая итоговую интервальную выборку

Набор данных	$\underline{\mathbf{x_i}}$	$\overline{\mathrm{x_i}}$
-0_5V_4	0.44788	0.4173
-0_25V_4	0.22516	0.20746
0_25V_4	0.20765	0.22357
0_5V_4	0.41956	0.44696

Искомая модель принимает вид:

$$y = -0.00003 + 0.865112 \cdot x \tag{5}$$