Санкт-Петербургский политехнический университет
Петра Великого

Физико-механический институт

Высшая школа прикладной математики и вычислительной
физики

Отчет по лабораторной работе №4
"Интервальный анализ"

Выполнили студент группы 5030102/10201: Скворцов Владимир Сергеевич

Преподаватель: Баженов Александр Николаевич

Санкт-Петербург
2024

Содержание

1 Постановка задачи

Определить параметры линейной регрессии

$$\mathbf{y} = \beta_0 + \beta_1 \mathbf{x}, \tag{1}$$

где $\mathbf{x}$ — входные данные, $\mathbf{y}$ — интервальные выходные данные, β_0, β_1 — параметры линейной регрессии.

Для калибровки измерителя, на вход подаётся набор постоянных напряжений

$$X = \{x_i\}. \tag{2}$$

Для надёжности, для каждого значения x проводится 100 измерений.

Получается набор интервальных выборок

$$\mathbf{Y} = \{\mathbf{y}_k\}_{k=1}^{100}. \tag{3}$$

$\mathrm{rad}\mathbf{y} = \frac{1}{2^N}$ В, $N = 14$.

Связь кодов данных и В:

$$V = \mathrm{Code}/16384 - 0.5. \tag{4}$$

Сделать оценки значений $\mathbf{Y}$ двумя способами:

- in: как интервал между первым и третьим квартилем
- ex: как границы бокс-плота

Решить ИСЛАУ 1 для внутренних и внешних оценок $\mathbf{y}$ Построить множество решений β_0, β_1. Построить коридор совместных зависимостей.

2 Необходимая теория

2.1 Интервальная мода

Пусть имеется интервальная выборка

$$\mathbf{X} = \{\mathbf{x}_i\}.$$

Сформируем массив интервалов $\mathbf{z}$ из концов интервалов $\mathbf{X}$.

Для каждого интервала $\mathbf{z}_i$ подсчитываем число μ_i интервалов из выборки $\mathbf{X}_i$, включающих $\mathbf{z}_i$. Максимальные $\mu_i = \max \mu$ достигаются для индексного множества K. Тогда можно найти интервальную моду как мультиинтервал

$$\mathrm{mode}\mathbf{X} = \bigcup_{k \in K} \mathbf{z}_k. \tag{5}$$

3 Реализаця

Лабораторная работа выполнена на языке программирования Python. В ходе работы были также использованы библиотеки `numpy` и `matplotlib`.

Ссылка на GitHub репозиторий: https://github.com/vladimir-skvortsov/spbstu-interval-anylysis

3.1 Алгоритм поиска оценок параметров линейной регрессии

Каждый из файлов содержит 100 фреймов, каждый из которых включает 1024 массива, состоящих из 8 двухбайтовых значений. В результате обработки этих данных было сформировано $1024 \times 8 = 8192$ интервальных систем линейных алгебраических уравнений (ИСЛАУ), представленных в следующем виде:

$$\begin{pmatrix} [x_1 - \mathrm{rad}\mathbf{y}, x_1 + \mathrm{rad}\mathbf{y}] & [1 - \mathrm{rad}\mathbf{y}, 1 + \mathrm{rad}\mathbf{y}] \\ \vdots & \vdots \\ [x_8 - \mathrm{rad}\mathbf{y}, x_8 + \mathrm{rad}\mathbf{y}] & [1 - \mathrm{rad}\mathbf{y}, 1 + \mathrm{rad}\mathbf{y}] \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_0 \end{pmatrix} = \begin{pmatrix} \hat{\mathbf{y}}_{1i} \\ \vdots \\ \hat{\mathbf{y}}_{8i} \end{pmatrix}, \; i \in \overline{1, 8192}$$

Для каждого отдельного пикселя фрейма x_j обозначает вольтаж, определяемый по названию файла, $\hat{\mathbf{y}}_{ji}$ — оценка значения, соответствующее каждому пикселю, по всем 100 фреймам, j — порядковый номер файла, а i — номер пикселя внутри файла. Параметры β_0 и β_1 представляют собой искомые параметры линейной регрессии.

Каждая система линейных алгебраических уравнений была решена с использованием метода Дж. Рона [1]. В результате были получены два множества интервалов оценок: $\mathbf{B}_0 = \{\beta_0\}_{i=1}^{8192}$ и $\mathbf{B}_1 = \{\beta_1\}_{i=1}^{8192}$. Оценка каждого из параметров линейной регрессии производится следующим образом:

$$\hat{\beta}_0 = \mathrm{mode}\mathbf{B}_0,$$
$$\hat{\beta}_1 = \mathrm{mode}\mathbf{B}_1.$$

Таким образом, конечные значения $\hat{\beta}_0$ и $\hat{\beta}_1$ служат наиболее вероятными оценками параметров регрессии, что позволяет более точно анализировать зависимость между переменными в исследуемых данных.

4 Результаты

4.1 Внутренняя оценка

Для внутренней оценки были получены следующие результаты:

$$\mathrm{mode}\mathbf{B}_0 = \{[8086.35, 8086.42], [8086.42, 8086.43], [8086.46, 8086.47], [8086.88, 8086.88]\},$$

$$\mathrm{mode}\mathbf{B}_1 = [13070.5, 13072.5].$$

4.2 Внешняя оценка

Для внешней оценки были получены следующие результаты:

$$\mathrm{mode}\mathbf{B}_0 = \bigcap_{i=1}^{8192} \beta_{0i} = [7927.51, 8224.58],$$
$$\mathrm{mode}\mathbf{B}_1 = \bigcap_{i=1}^{8192} \beta_{1i} = [13097.9, 13573.8].$$

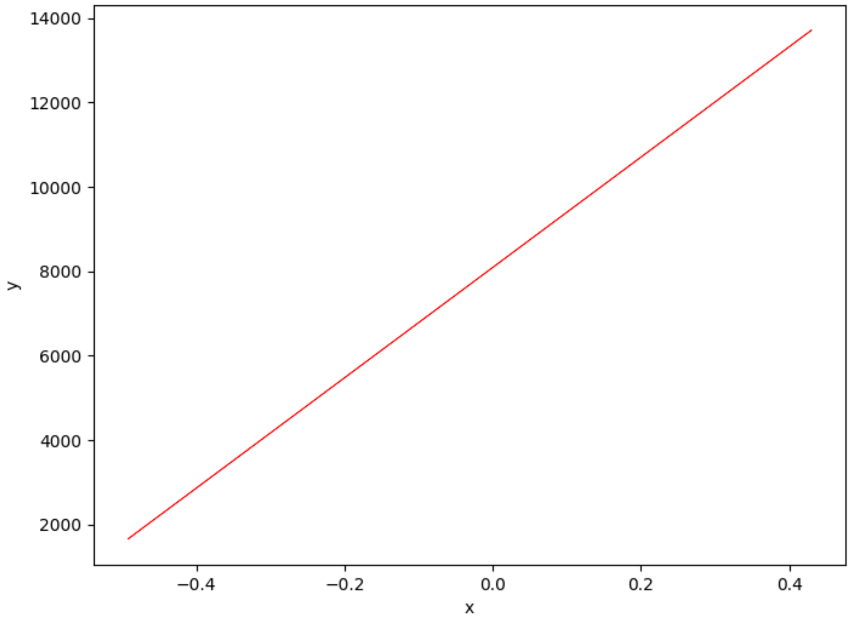

Рис. 1: Коридор совместных зависимостей для внутренней оценки.

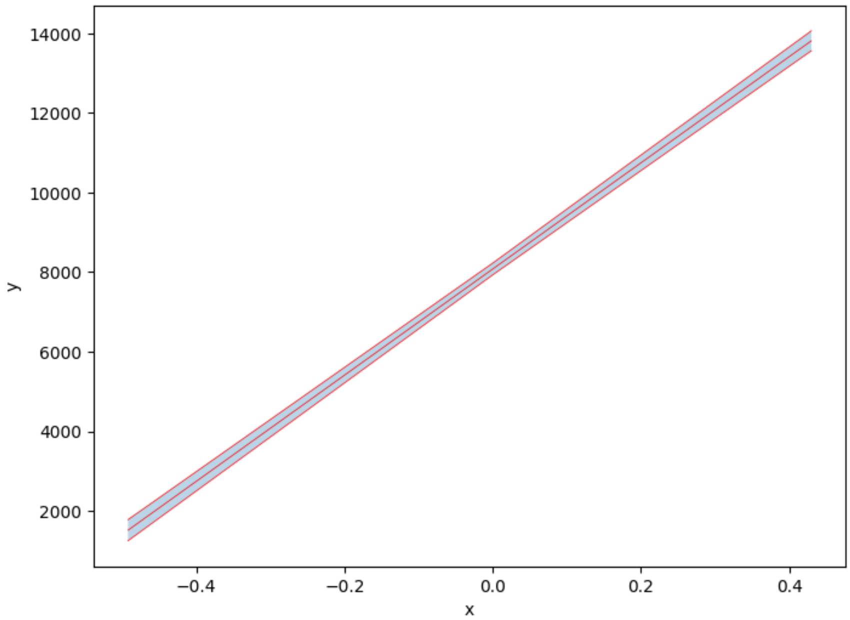

Рис. 2: Коридор совместных зависимостей для внешней оценки.

5 Вывод

В ходе выполнения лабораторной работы была реализована методика оценки параметров линейной регрессии на основе интервальных данных. Основные результаты включают в себя:

- Разработан алгоритм для нахождения внутренних и внешних оценок параметров линейной регрессии, что позволяет учитывать неопределённость в данных.
- Получены интервальные оценки параметров β_0 и β_1, которые демонстрируют диапазон возможных значений параметров регрессии.
- Построены коридоры совместных зависимостей, которые визуализируют интервальные решения и помогают в анализе устойчивости модели.

Результаты показывают, что предложенный подход позволяет более точно моделировать зависимости в данных, учитывая возможные вариации и ошибки. Это особенно полезно в приложениях, где точность измерений может варьироваться, и требуется надёжная оценка параметров модели.

Список литературы

[1] J. Rohn — «Enclosing solutions of overdetermined systems of linear interval equations», Reliable Computing 2 (1996), 167-171.