

Санкт-Петербургский политехнический университет
Петра Великого

Физико-механический институт

Кафедра «Прикладная математика»

**Отчёт по курсовой работе
по дисциплине «Анализ данных с интервальной
неопределённостью»**

Выполнил студент:
Аникин Александр 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
2.1	Точечная линейная регрессия	2
2.2	Информационное множество	2
3	Результаты	3
4	Обсуждение	4

Список иллюстраций

1	Синим цветом представлены значения для различных элементов с их стандартными отклонениями. Оранжевым цветом показаны результаты регрессии.	3
---	--	---

1 Постановка задачи

2 Теория

2.1 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

$$y = \beta_0 + \beta_1 x \quad (1)$$

Поставим задачу оптимизации 2.1 для нахождения точечных оценок параметров β_0, β_1 .

$$\begin{aligned} \sum_{i=1}^m w_i &\rightarrow \min \\ \text{mid}\mathbf{y}_i - w_i \cdot \text{rad}\mathbf{y}_i &\leq X\beta \leq \text{mid}\mathbf{y}_i + w_i \cdot \text{rad}\mathbf{y}_i \\ w_i &\geq 0, i = 1, \dots, m \\ w, \beta &-? \end{aligned} \quad (2)$$

Задачу 2.1 можно решить методами линейного программирования.

2.2 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 Υ , сопоставляющее каждому значению аргумента x множество

$$\Upsilon(x) = \bigcup_{\beta \in \Omega} f(x, \beta) \quad (3)$$

, где Ω - информационное множество, x - вектор переменных, β - вектор оцениваемых параметров.

Информационное множество может быть построено как пересечение полос, заданных

$$\underline{\mathbf{y}}_i \leq \beta_0 + \beta_1 x_{i1} + \dots + \beta_m x_{im} \leq \overline{\mathbf{y}}_i \quad (4)$$

, где $i = \overline{1, ny_i} \in \mathbf{Y}, x_i \in X$, X - точечная выборка переменных, \mathbf{Y} - интервальная выборка откликов.

3 Результаты

Данные были взяты из таблицы S1 приложенной к работе Roederer, Ian U., et al. "Element abundance patterns in stars indicate fission of nuclei heavier than uranium." *Science* 382.6675 (2023)

В ней представлены определенные значения для различных элементов и их стандартные отклонения. В качестве x выступает значение $[Eu/Fe]$, а в качестве y $X \pm eX$, $X \in \{Se, Sr, Y, Nb, Mo\}$. Эти данные соответствуют части данных с *рис Fig. 2. (A) Abundance ratios of groups of elements that do or do not correlate with [Eu/Fe]*. Остальные же данные в материалах статьи представлены рядом ссылок, и автоматизировать их сбор не представляется возможным.

Далее рассмотрим сами данные и получившие результаты.

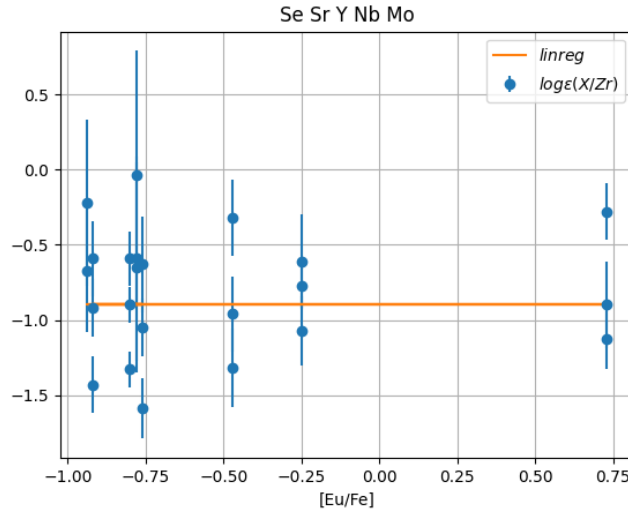


Рис. 1: Синим цветом представлены значения для различных элементов с их стандартными отклонениями. Оранжевым цветом показаны результаты регрессии.

Полученны $(\beta_1, \beta_2) = (-0.0, -0.9)$. Полученные результаты отличаются от результатов в статье из-за отсутствия сдвига, однако общий вывод будет одинаков из-за β_1 близкого к нулю.

Также стоит отметить что информационное множество пустое, так как настоящая зависимость отличается от линейной и сами интервалы представляют собой лишь стандартное отклонение от среднего.

4 Обсуждение

Можно сказать, что полученные результаты соответствуют результатам из исходной работы с учетом отсутствия сдвига и меньшего количества данных. Общий вывод об отсутствии линейной зависимости между значениями сохраняется.