

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	5

Список иллюстраций

1	Синим цветом представлены значения для различных элементов с их стандартными отклонениями. Оранжевым цветом показаны результаты регрессии.	3
2	Синим цветом представлены значения для различных элементов с тремя стандартными отклонениями. Оранжевым цветом показаны результаты регрессии.	4
3	Информационное множество для выборки с радиусом в 3σ	5

1 Постановка задачи

Решить задачу восстановления зависимости на экспериментальных данных. Данные взять из материалов (таблица) статьи Roederer, Ian U., et al. "Element abundance patterns in stars indicate fission of nuclei heavier than uranium." Science 382.6675 (2023).

2 Теория

2.1 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 - точечный, y_i - интервальный. Пусть иско-мая 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 .

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

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

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

, где $i = \overline{1, n}$, $\underline{y}_i \in \mathbf{Y}$, $x_i \in X$, 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.

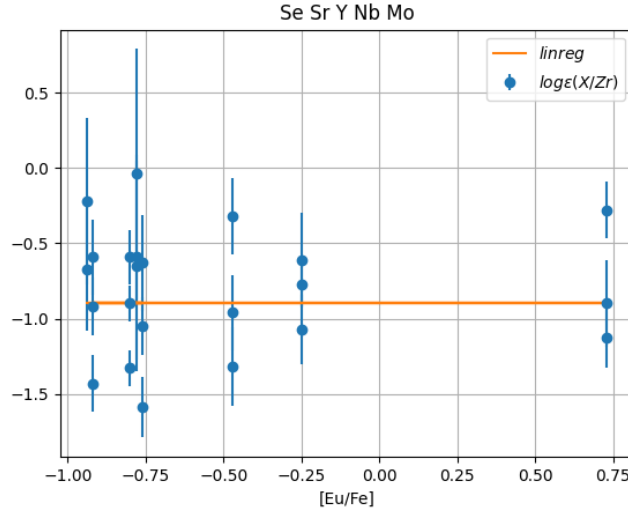


Рис. 1: Синим цветом представлены значения для различных элементов с их стандартными отклонениями. Оранжевым цветом показаны результаты регрессии.

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.

Чтобы сделать выборку совместной, произведем обинтервваливание следующим образом: вместо $(a, b) = (\mu - \sigma, \mu + \sigma)$ по правилу трех сигм будем брать $(a, b) = (\mu - 3\sigma, \mu + 3\sigma)$, что расширит интервалы и при нормальном распределении значения окажутся внутри интервала с вероятностью 0.9973 вместо 0.6826 для радиуса в одно стандартное отклонение. Самое главное, что такое обинтерваливание позволит сделать выборку совместной, а информационное множество не пустым.

Таким образом получаем следующую выборку:

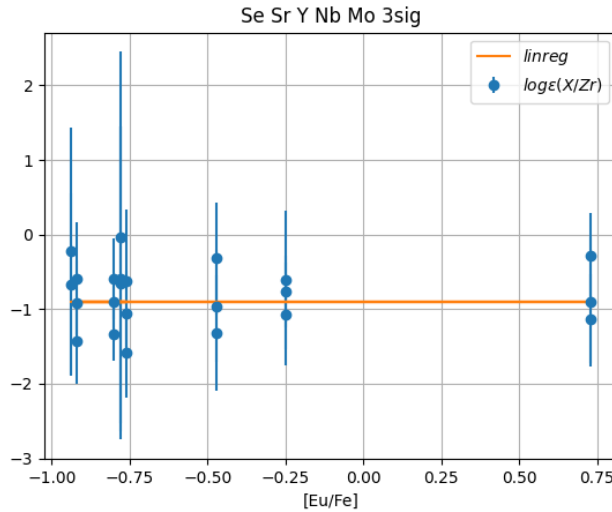


Рис. 2: Синим цветом представлены значения для различных элементов с тремя стандартными отклонениями. Оранжевым цветом показаны результаты регрессии.

В этом случае результаты регрессии не поменяются $(\beta_1, \beta_2) = (-0.0, -0.9)$, а информационное множество будет выглядеть следующим образом:

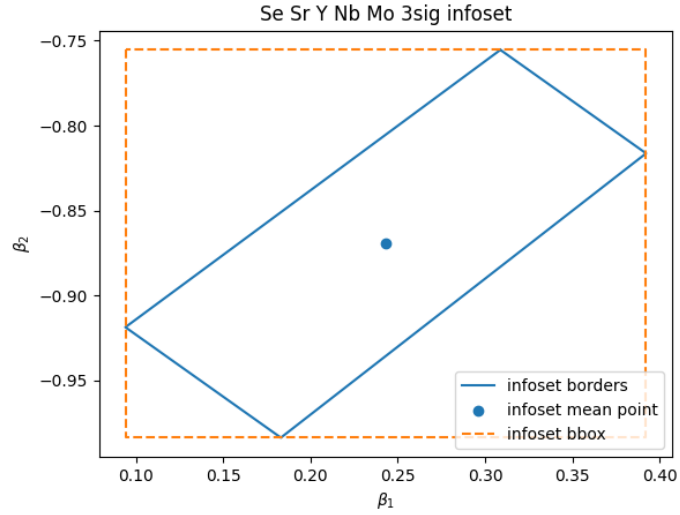


Рис. 3: Информационное множество для выборки с радиусом в 3σ

Итого получим следующие интервалы для $(\beta_1, \beta_2) = ([0.093, 0.392], [-0.983, -0.755])$. А центр информационного множества будет следующим $(\beta_1, \beta_2) = (0.243, -0.869)$.

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.