Санкт-Петербургский политехнический университет
Петра Великого
Институт прикладной математики и механики
Кафедра «Прикладная математика»

**Отчёт
по лабораторной работе №5
по дисциплине
«Интервальный анализ»**

Выполнил студент:
Величко Арсений Юрьевич

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

Санкт-Петербург
2022 г.

Содержание

1 Постановка задачи

Для задачи линейной регрессии вида

$$\mathbf{y} = \mathbf{X}\beta \tag{1}$$

такой что для точечного $x = (2, 5, 7, 11)^{\mathrm{T}}$

$$\mathrm{mid}\,\mathbf{y} = (5, 8, 10, 13)^{\mathrm{T}}, \quad \mathrm{rad}\,\mathbf{y} = (1.7, 1.3, 2.4, 2.2)^{\mathrm{T}} \tag{2}$$

необходимо:

- построить интервальное множество решений β, сделать точечные оценки параметров
- построить коридор совместных зависимостей
- задать набор точек предсказания внутри и вне интервала измерений, построить соответсвующий им набор значений $\mathbf{y}$

2 Теория

2.1 Информационное множество

Информационным множеством задачи восстановления зависимостей в нашем случае будем называть

$$\Omega = \{\beta \in \mathbb{R}^2 \mid X\beta \in \mathbf{y}\}, \;\; \text{где } X = \begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix} \tag{3}$$

2.2 Коридор совместных значений

Пусть в задаче восстановления зависимостей информационное множество Ω параметров зависимостей $y = f(x, \beta)$ совместных с данными является непустым. *Коридором совместных значений* рассматриваемой задачи называется многозначное отображение Υ, сопоставляющее каждому значению аргумента x множество

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

2.3 Точечные оценки

Центр наибольшей диагонали информационного множества

$$\hat{\theta}_{\mathrm{maxdiag}} = 0.5(b_1 + b_2) \tag{5}$$

Центр тяжести информационного множества

$$\hat{\theta}_{\mathrm{gravity}} = \frac{1}{N}\sum_{i=1}^{N} b_i \tag{6}$$

В вышеприведённых оценках оценках b_i - i-ая вершина информационного множества, а N - общее кол-во вершин.
Оценка метода наименьших квадратов - решение точечной задачи, составленной из центров интервалов, методом наименьших квадратов.

2.4 Граничны измерения

Граничными называют измерения, определяющие какой-либо фрагмент границы информационного множества.
Подмножество всех граничных измерений в выборке S_n играет особую роль, поскольку оно является *минимальной подвыборкой, полностью определяющей модель.*

3 Реализация

Лабораторная работа выполнена при помощи пакета Octave с использованием функций из данного репозитория.
Ссылка на репозиторий с исходный кодом:
`https://github.com/ArsenyVelichko/IntervalAnalysis`

4 Результаты

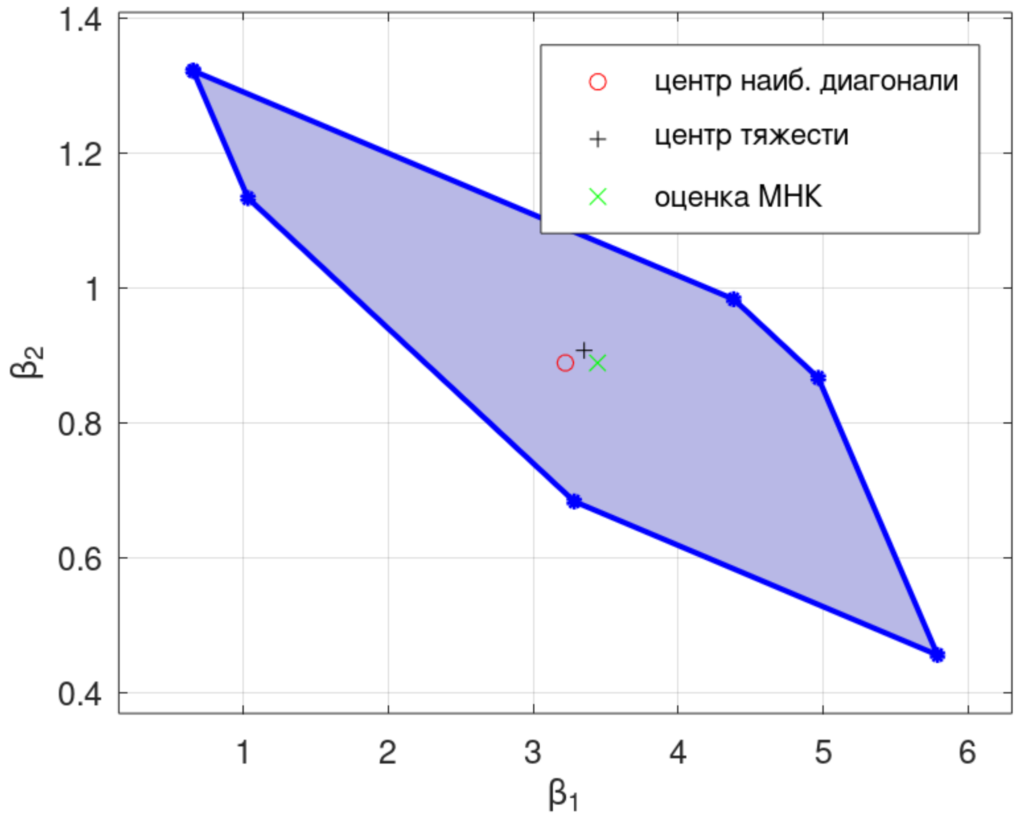

Рис. 1: Информационное множество и точечные оценки задачи (1)

Положения оценок:

$$\hat{\beta}_{\text{maxdiag}} = [3.2222, 0.8889], \quad \hat{\beta}_{\text{gravity}} = [3.3519, 0.9074], \quad \hat{\beta}_{\text{lsm}} = [3.4444, 0.8889] \quad (7)$$

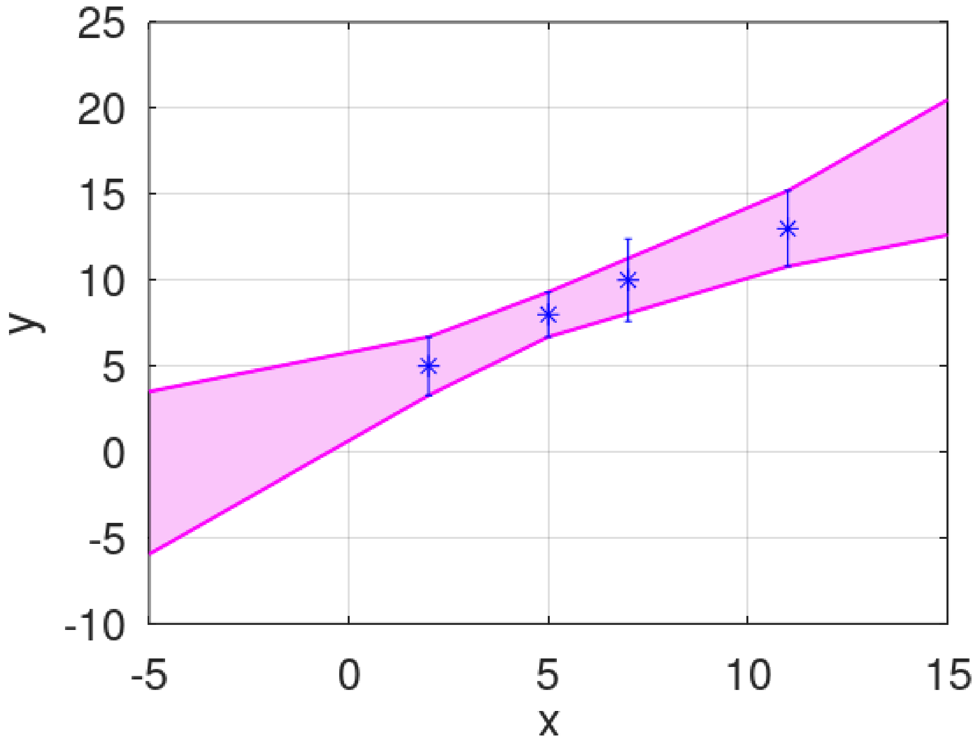

Рис. 2: Коридор совместных решений задачи (1)

Множество граничных измерений

$$X_{\text{boundary}} = \{2, 5, 11\} \tag{8}$$

В качестве точек для предсказаний был выбран вектор $x = (-3, 1, 4, 9, 14)^{\mathrm{T}}$

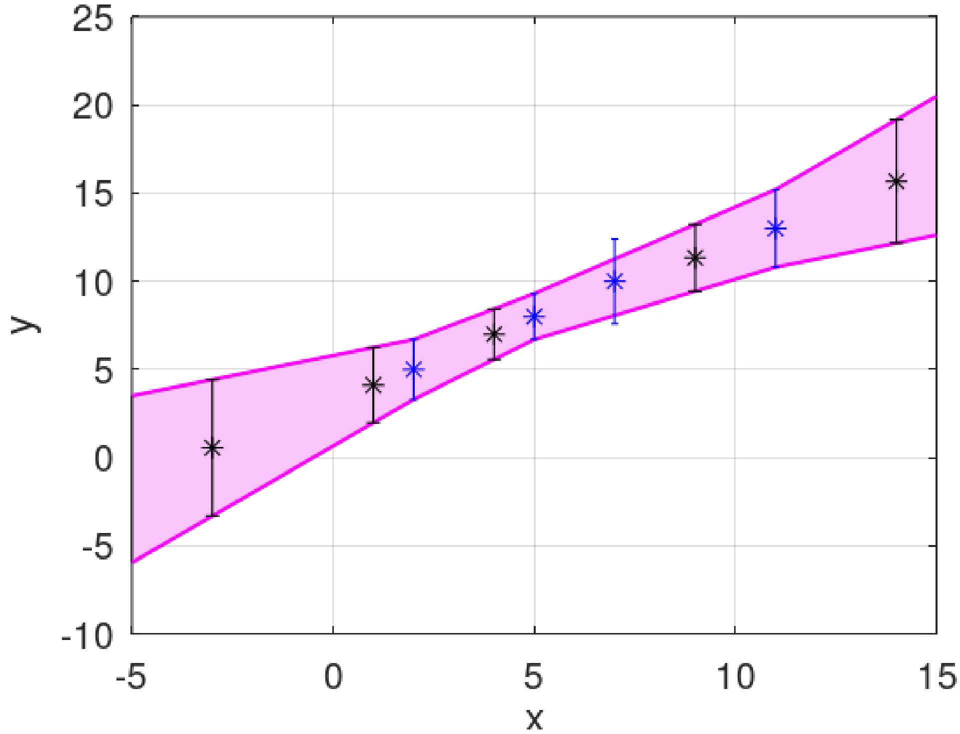

Рис. 3: Значения предсказаний задачи (1)

5 Обсуждение

1. Информационное множество (4) представляет из себя шестигранник. Это означает, что 2 из 8 вершин брусов погрешности не оказывают влияния на регрессионную модель.

2. Все точечные оценки находятся внутри информационного множества и расположены близко друг к другу.

3. По графику коридора совместных значений (4) мы можем увидеть те 2 вершины, которые не оказывают влияния. Они обе принадлежат интервалу, соответствующему третьему измерению. Таким образом, третье измерение $x_3 = 7$ не является граничным и его удаление из выборки не изменит модель. Как и следовало ожидать, коридор начинает сильно расширяться при отдалении от множества измерений.

4. Обратившись к графику предсказаний (4) мы видим, что те предсказания, которые находяться вблизи измерений, действительно могли бы неплохо описывать исходную зависимость. В то же время, крайние предсказания имеют большой интервал неопределённости, что говорит нам о том, что экстраполяция исходной зависимости на большой интервал весьма затруднительна.