|  |
| --- |
| Санкт-Петербургский политехнический университет Петра Великого |
| Высшая школа прикладной математики и вычислительной физики |
| Кафедра прикладной математики |

**Курсовая работа**

по дисциплине «Стохастические модели и анализ данных»

на тему

**Восстановление зависимостей**

|  |
| --- |
|  |

Выполнил студент гр. 5040102/00201

Чепулис М.А.

Преподаватель

Баженов А.Н.

Санкт-Петербург

2021 год

Оглавление

[Постановка задачи 3](#_Toc87994849)

[Параметры модели 5](#_Toc87994850)

[Коридор совместных зависимостей 7](#_Toc87994851)

[Прогноз за пределы интервала: 8](#_Toc87994852)

[Граничные точки множества совместности 8](#_Toc87994853)

[Заключение 9](#_Toc87994854)

[Приложение: 9](#_Toc87994855)

[Использованная литература 9](#_Toc87994856)

# Постановка задачи

Необходимо выбрать массив данных и восстановить линейную зависимость с учётом интервальной неопределённости данных.

Модель данных будем искать в классе линейных функций:

С неотрицательной первой производной:

Ниже показан график исходных данных

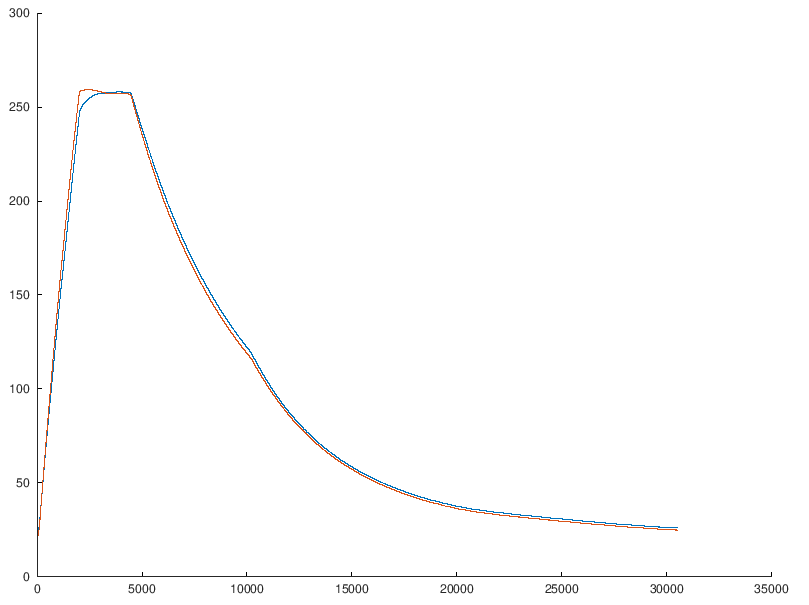


Рисунок 1 Исходные данные

# Решение

## Выбор рассматриваемой области

Выберем хорошо представимый линейной моделью участок:

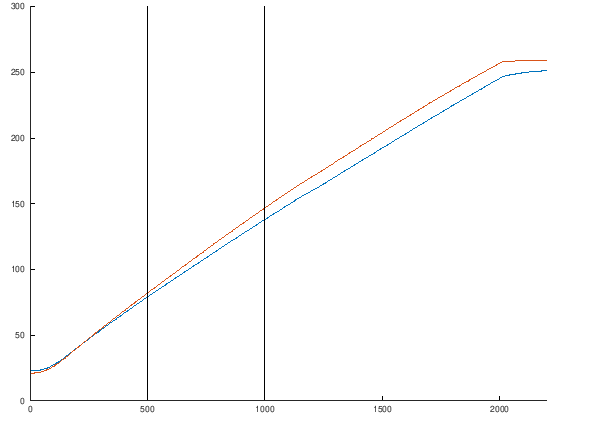


Рисунок 2 Уточнённый рассматриваемый участок

Оставим только нижнюю линию, и выберем на ней 5 точек:

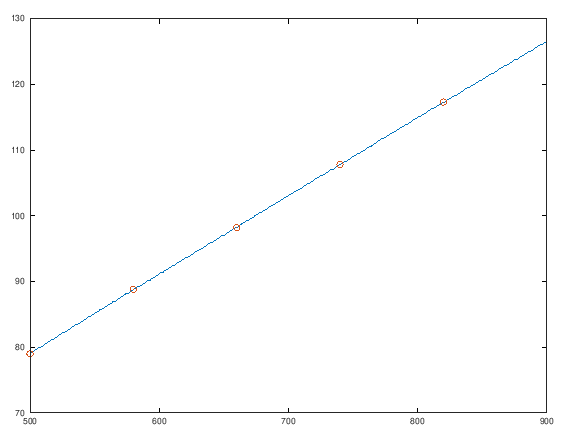


Рисунок 3 Выбранные точки из исходных данных

Посмотрим на выбранные значения:

x = 500 580 660 740 820

y = 79.0 88.8 98.2 107.8 117.3

В качестве начальной погрешности зададим , одинаковую для всех наблюдений. Этот выбор связан с последним значащим разрядом в данных.

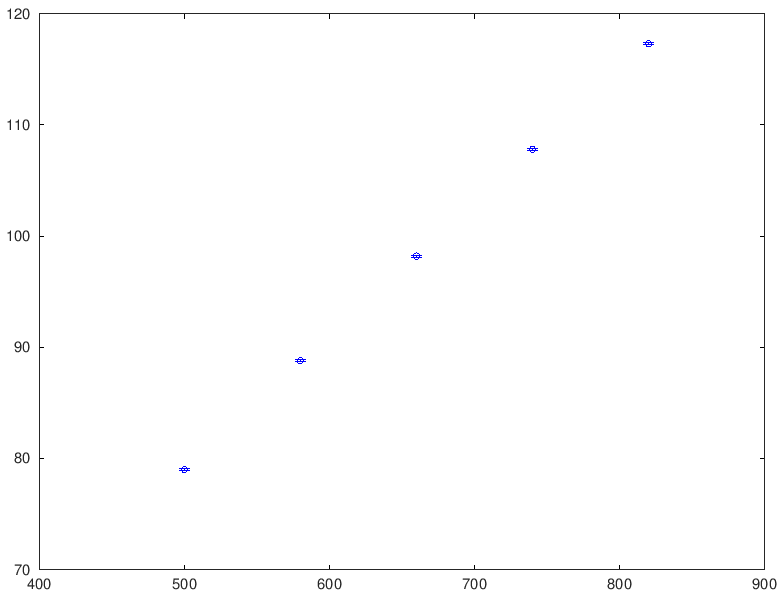


Рисунок 4 Входные данные с интервальной неопределённостью

## Параметры модели

Сперва построим линейную модель методом МНК как на точечных значениях:

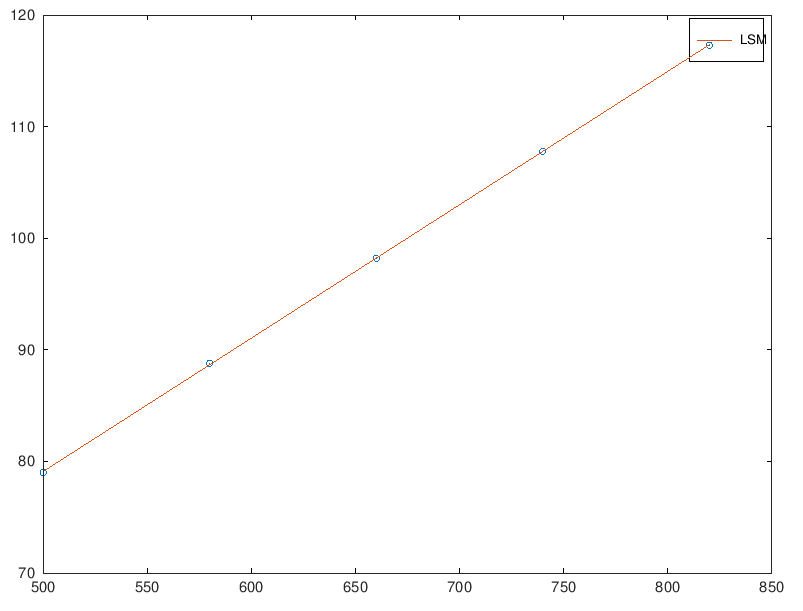


Рисунок 5 МНК линейная регрессия

При переходе к интервальному случаю, при попытке определить информационное множество мы обнаруживаем, что оно пусто. Предположим, что погрешность была недооценена. Для согласования с данными поставим задачу оптимизации и решим её методом линейного программирования [[1]](#ref_1):

где – матрица , в первом столбце которой элементы равные 1, во втором – значения .

В качестве значений

Значение весов в задаче оптимизации:

Как мы видим, требуются небольшие корректировки погрешности, потому не будем считать второе наблюдение выбросом.

Увеличим погрешность всех измерений:

Построим новое информационное множество параметров модели. Поскольку информационное множество задачи построения линейной зависимости по интервальным данным задаётся системой линейных неравенств, то оно представляет собой выпуклый многогранник [[2]](#ref_2).

Сразу обозначим на графике несколько точечных оценок:

* Центр наибольшей диагонали информационного множества:

где

* Центр тяжести информационного множества:

где – вершина многогранника, – их количество.

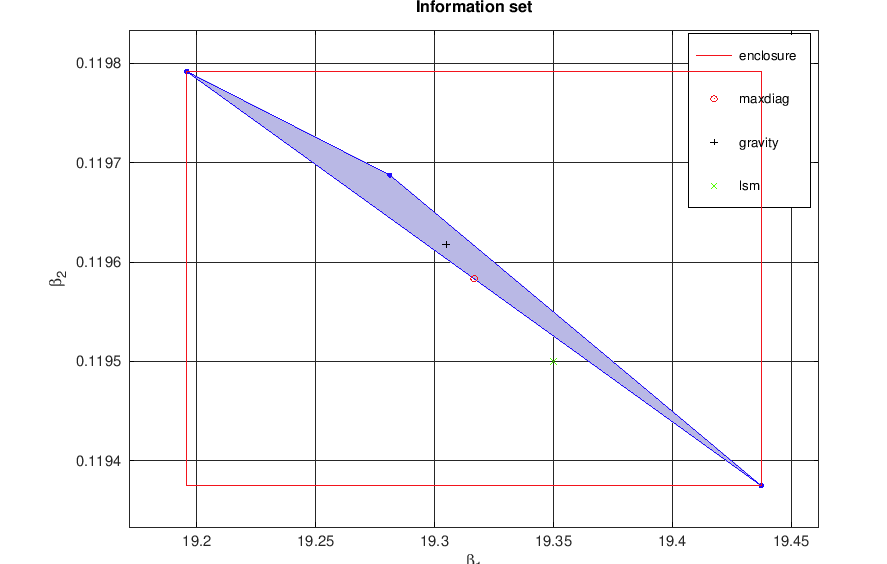


Рисунок 6 информационное множество линейной модели

Заметим, что значения, полученные при помощи МНК оказались за границами информационного множества.

## Коридор совместных зависимостей

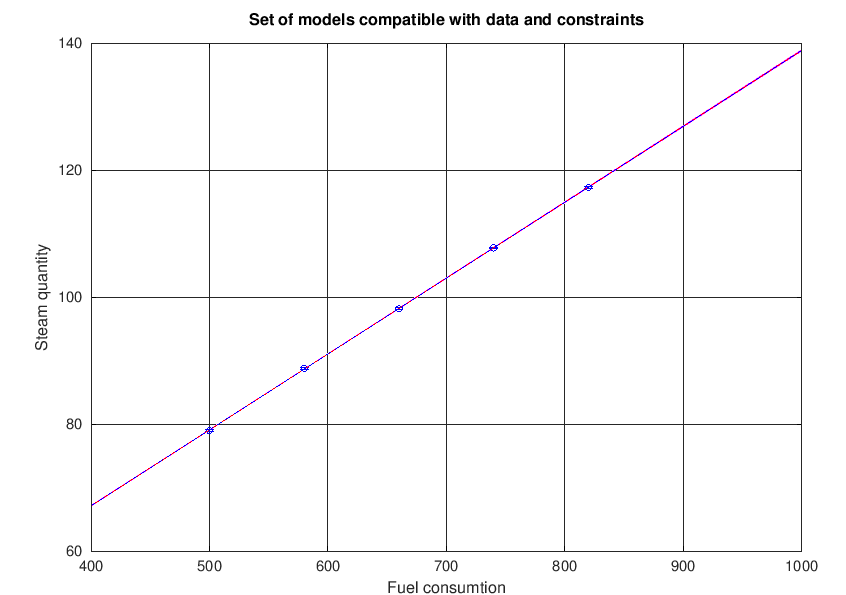


Рисунок 7 Коридор совместных зависимостей, весь диапазон

Однако коридор совместных событий слился в одну прямую. Рассмотрим подробнее, что происходит возле первой точки:

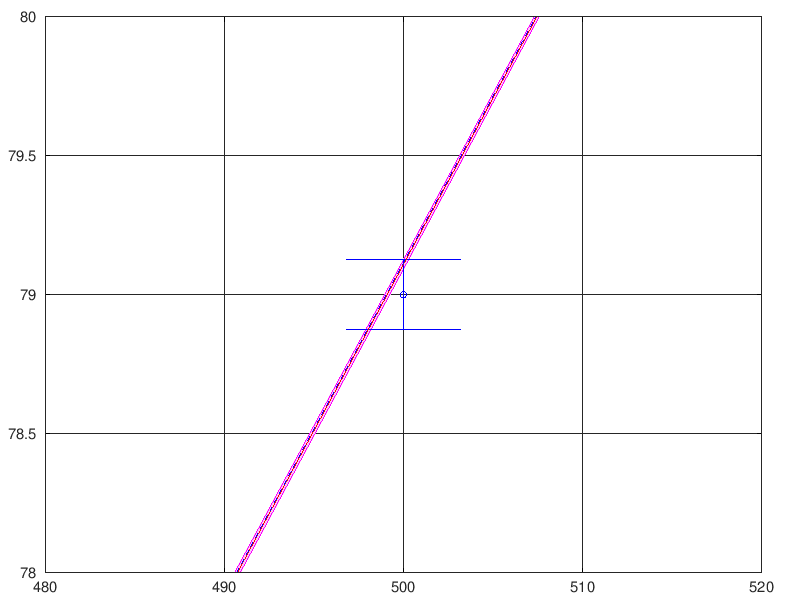


Рисунок 8 Коридор совместных событий в окрестности первого наблюдения

## Прогноз за пределы интервала:

С помощью построенной выше модели

Можно получить прогнозные значения выходной переменной:

Возьмём 3 точки:

Тогда

|  |  |  |
| --- | --- | --- |
|  |  |  |
| 450 | [73.102, 73.156 ] | 0.02 |
| 600 | [91.063, 91.094] | 0.01 |
| 950 | [132.844, 132.998] | 0.07 |

Неопределённость прогноза растёт по мере удаления от области, в которой производились исходные измерения. Это обусловлено видом коридора зависимости, расширяющимся за пределами области измерений.

## Граничные точки множества совместности

В данном случае граничными оказались точки с номерами 1, 2, 5.

Убедимся в этом посмотрев детально на каждую из точек подробнее

|  |  |  |  |
| --- | --- | --- | --- |
|  | 1 |  | 2 |
|  | 5 |  |  |

Как мы видим, точки 1 и 5 касаются верхней границы множества.

Точка 2 – нижней.

Убедимся, в том, что остальные точки (3 и 4) не являются граничными.

|  |  |  |  |
| --- | --- | --- | --- |
|  | 3 |  | 4 |

Тем самым набор точек [1, 2, 5] может полностью определить модель.

# Заключение

В ходе работы была построена линейная модель данных. Наблюдения рассматривались сначала как просто точечные, далее – как значения с интервальной неопределённостью.

Была задана погрешность наблюдений, однако выборка оказалась несовместной. Было принято решение, что в выборке отсутствуют выбросы и причина несовместности – недооценённая погрешность.

Для улучшения оценки погрешности была сформирована и решена задача линейного программирования. После корректировки выборка стала совместной.

Было получено информационное множество для параметров линейной модели, построен коридор совместности и обнаружены граничные точки коридора совместности.

По полученной модели были вычислены прогнозы за пределами области измерений.

# Приложение:

Ссылка на проект с кодом реализации:

<https://github.com/MChepulis/Lab-Stochastic-models-and-data-analysis>

# Использованная литература

1. А.Н. Баженов, С.И. Жилин, С.И. Кумков, С.П. Шарый. Обработка и анализ данных с интервальной неопределённостью. РХД. Cерия «Интервальный анализ и его приложение». Ижевск. 2021. с.200.
2. С.И.Жилин. Примеры анализа интервальных данных в Octave https://github.com/szhilin/octave-interval-examples