Санкт-Петербургский политехнический университет Петра Великого
Физико-механический институт

Направление подготовки

«01.04.02 Прикладная математика и информатика»

# Отчёт
по лабораторной работе №2
по дисциплине
«Анализ данных с интервальной неопределенностью»

Выполнила студентка гр. 5040102/20201 Харисова Т.А.

Преподаватель Баженов А.Н.

Санкт-Петербург

2023

# Оглавление



# Список иллюстраций



# Список таблиц

## Постановка задачи

Дана интервальная выборка. Требуется восстановить для нее функциональную зависимость.

## Теория

Заданы $X = \{x_i\}_{i=1}^n$ – точечная выборка независимых входных данных, $Y = \{y_i\}_{i=1}^n$ – интервальная выборка выходных данных.

Задача поиска функциональной зависимости выглядит следующим образом:

$$y_i = f(x_i, \beta), \qquad i = 1..n, \tag{1}$$

требуется для заданных $(X, Y)$ найти вектор параметров $\beta$.

При построении линейной регрессионной модели набор данных аппроксимируется прямой:

$$y_i = \beta_1 x_i + \beta_0 \tag{2}$$

Для оценки параметров $\beta_1$ и $\beta_2$ необходимо найти решение системы вида:

$$\underline{y_i} \le \beta_1 x_i + \beta_0 \le \overline{y_i}, \qquad i = 1..n. \tag{3}$$

В случае несовместных данных применяется метод вариации неопределенности, с учетом которого система (3) преобразуется в задачу минимизации

$$\sum_{i=1}^{n} w_i \to \min \tag{4}$$

с условиями

$$\begin{gathered} \operatorname{mid} y_i - w_i \operatorname{rad} y_i \le \beta_1 x_i + \beta_0 \le \operatorname{mid} y_i + w_i \operatorname{rad} y_i \\ w_i \ge 0, \qquad i = 1..n, \end{gathered} \tag{5}$$

где $w_i$ – искомые веса.

Информационное множество строится как пересечение полос, заданных (3), с учетом весов $w_i$, если выборка несовместна.

Внешняя интервальная оценка параметра определяется минимальным и максимальным значениями, которых может достигать значение параметра в информационном множестве.

Коридор совместных зависимостей – множество всех функций, совместных с интервальными данными задачи восстановления зависимости.

## Реализация

Работа выполнена с помощью языка программирования Python в среде разработки Visual Studio Code. Ссылка на исходный код работы: Lab_2 (github.com)

Для построения выборки использовались данные из файлов “+0_5V_1.txt”, “+0_25V_1.txt”, “-0_25V_1.txt”, “-0_5V_1.txt”.

## Результаты

Исходные данные подвергаются предварительной коррекции: вместо $\dot{y}_i$ рассматриваются $\dot{y}_i - \delta_i, i = 1..n$, где $\delta_i$ – некоторая погрешность.

$x = [0.5, 0.25, -0.25, -0.5]$, в качестве $y$ берется среднее значение данных из соответствующих файлов, обынтерваленное значением $\varepsilon = 0.005$. Назовем выборку $Y_1$.

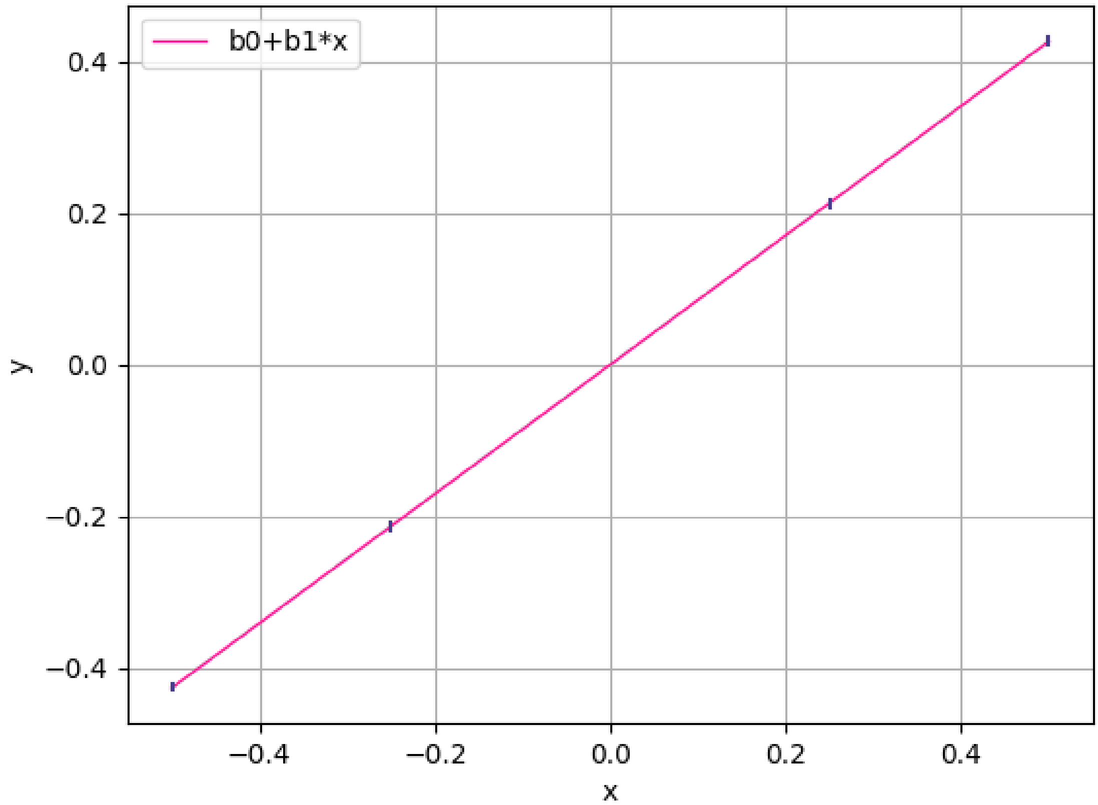


*Рисунок 1. Линейная регрессия для $Y_1$*

$$\beta_0 = 0.0001 \qquad \beta_1 = 0.8538$$

$$y = 0.8538x + 0.0001$$

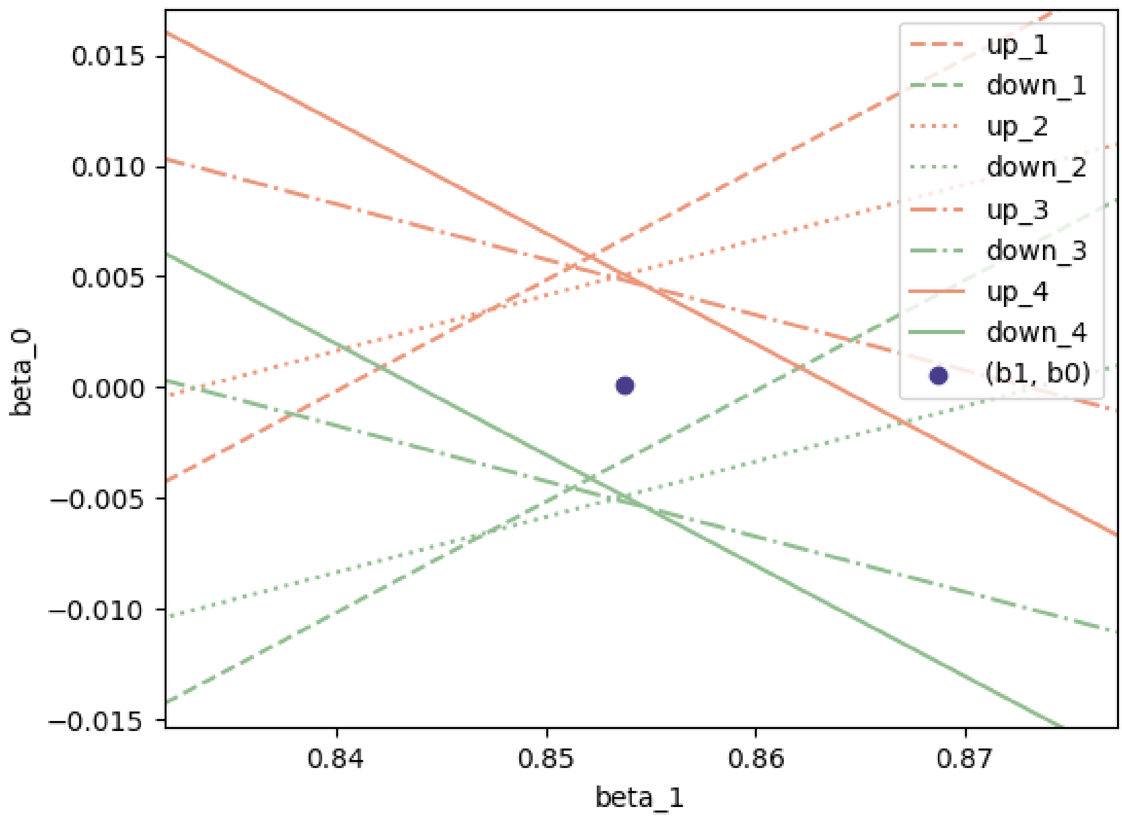


*Рисунок 2. Информационное множество для $Y_1$*

Интервальная оценка параметров $\beta_0 = [-0.0041, 0.0050]$ , $\beta_1 = [0.8421,0.8621]$.

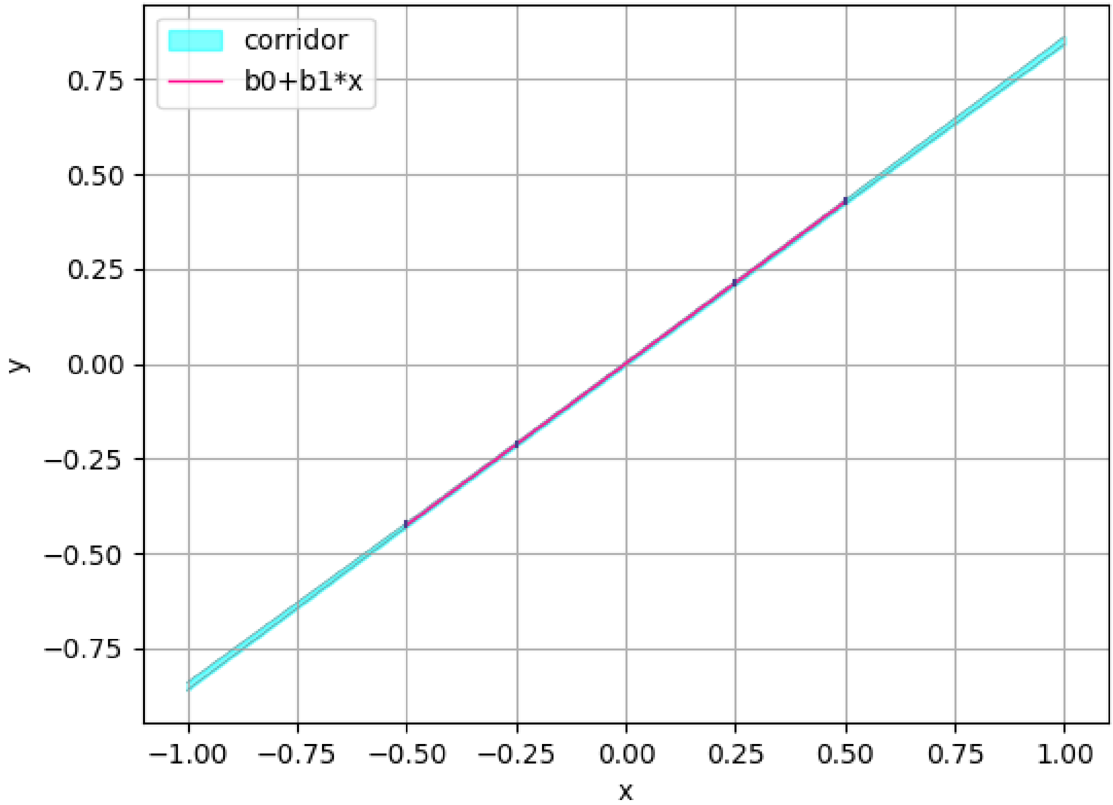


*Рисунок 3. Коридор совместных зависимостей для $Y_1$*

Рассмотрим построение линейной регрессии для интервалов, построенных по данным без вычитания погрешности $\delta_i$. Назовем выборку $Y_2$.

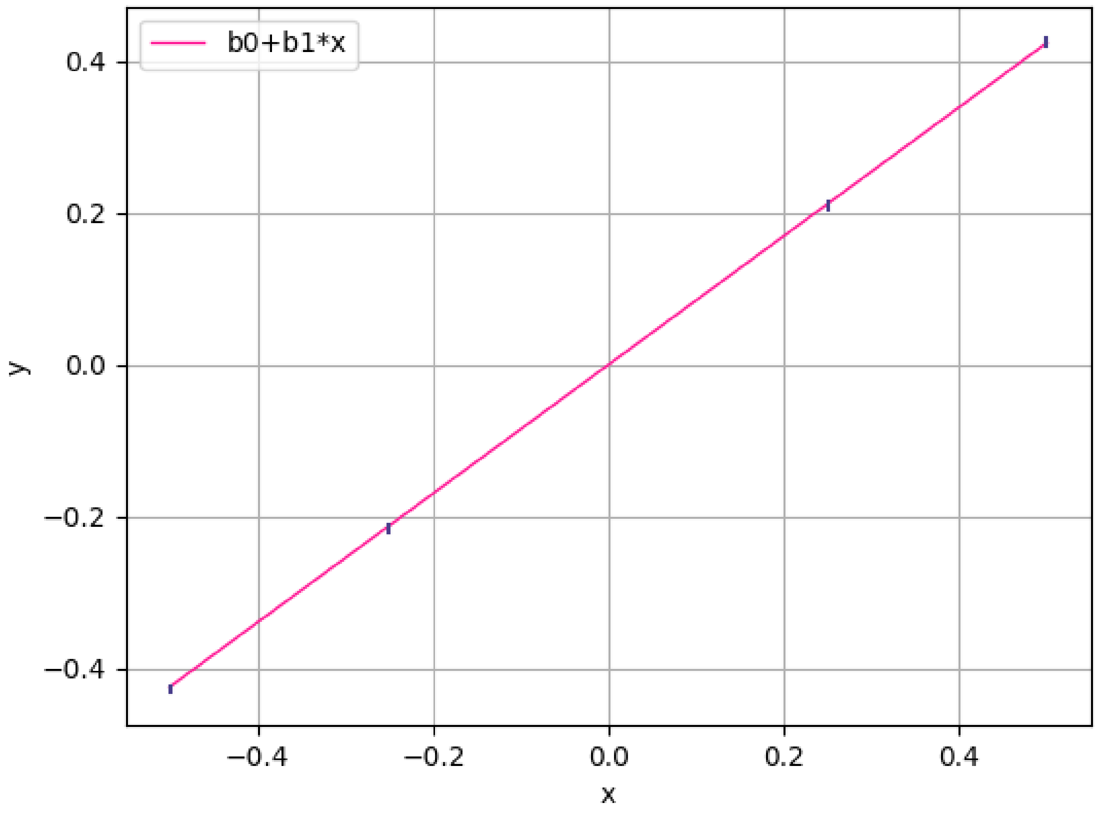


*Рисунок 4. Линейная регрессия для $Y_2$*

$$\beta_0 = 0.0 \qquad \beta_1 = 0.8492$$

$$y = 0.8492x$$

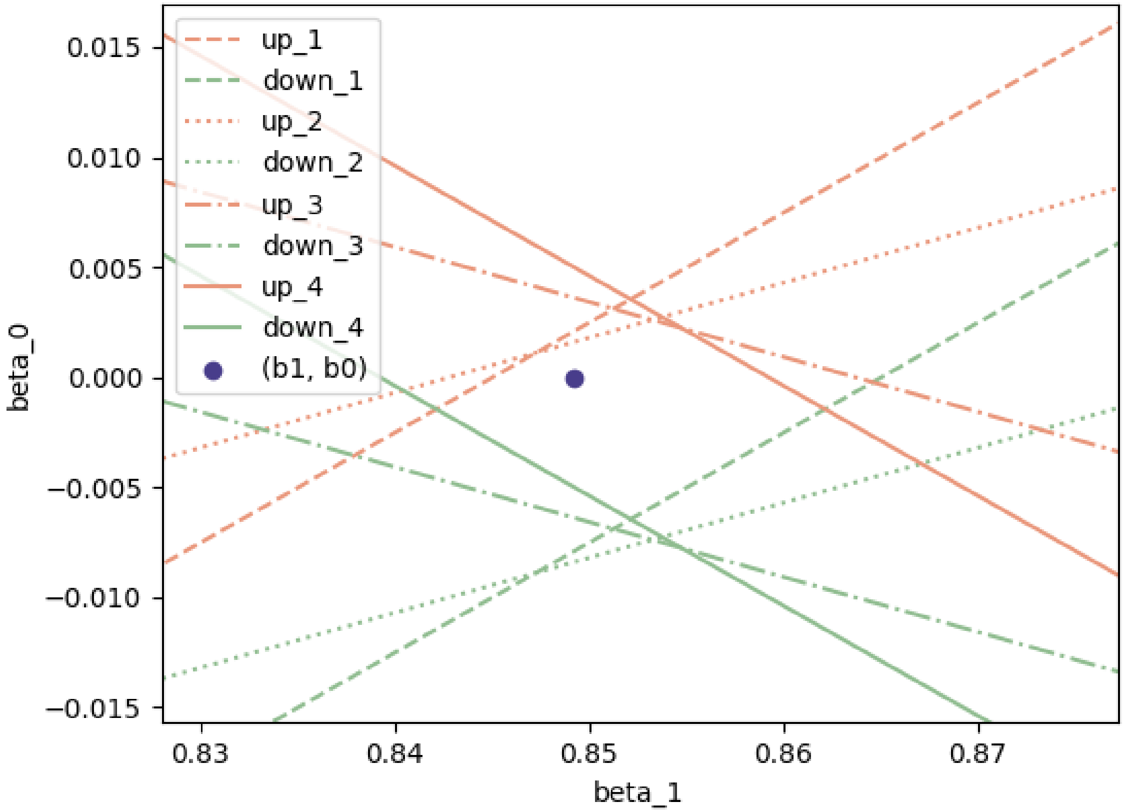


*Рисунок 5. Информационное множество для $Y_2$*

Интервальная оценка параметров $\beta_0 = [-0.0065, 0.0026]$ , $\beta_1 = [0.8421, 0.8621]$.

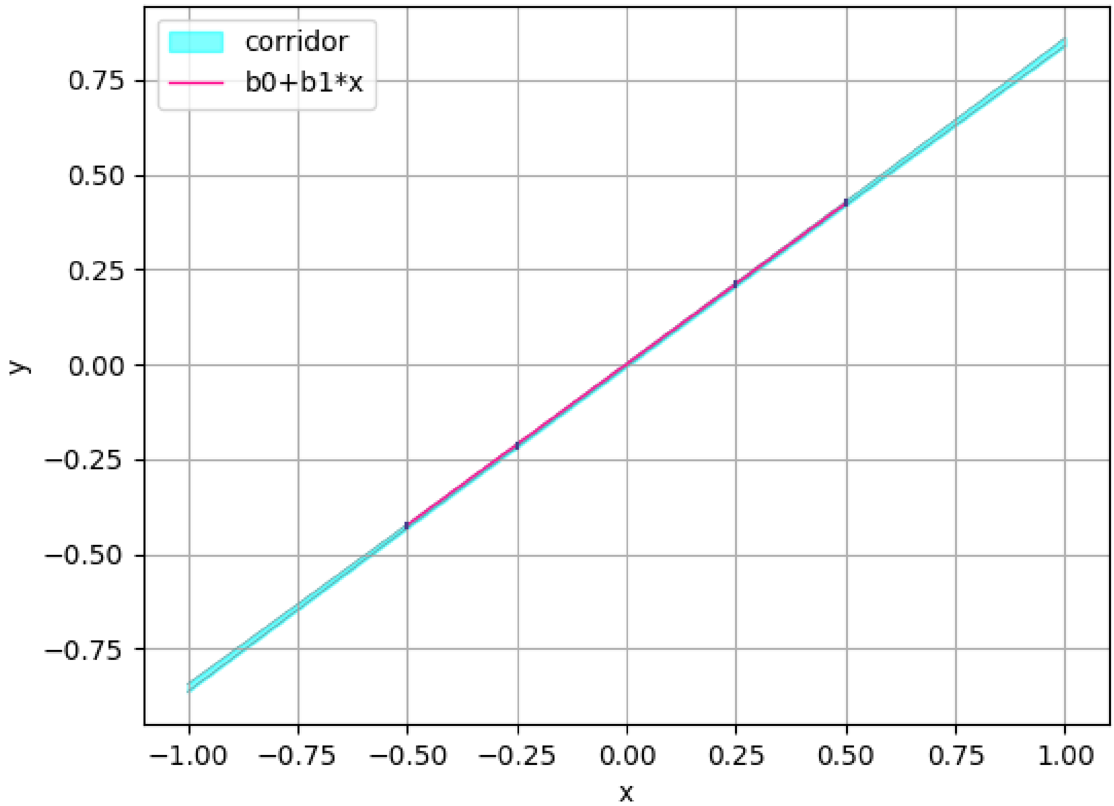


*Рисунок 6. Коридор совместных зависимостей для $Y_2$*

Рассмотрим построение регрессии для другой интервальной выборки. Вектор $x$ остается тем же, в качестве $y$ берется мода данных из соответствующих файлов, данные предварительно обынтерваливаются значением $\varepsilon = 0.05$. Назовем выборку $Y_3$.

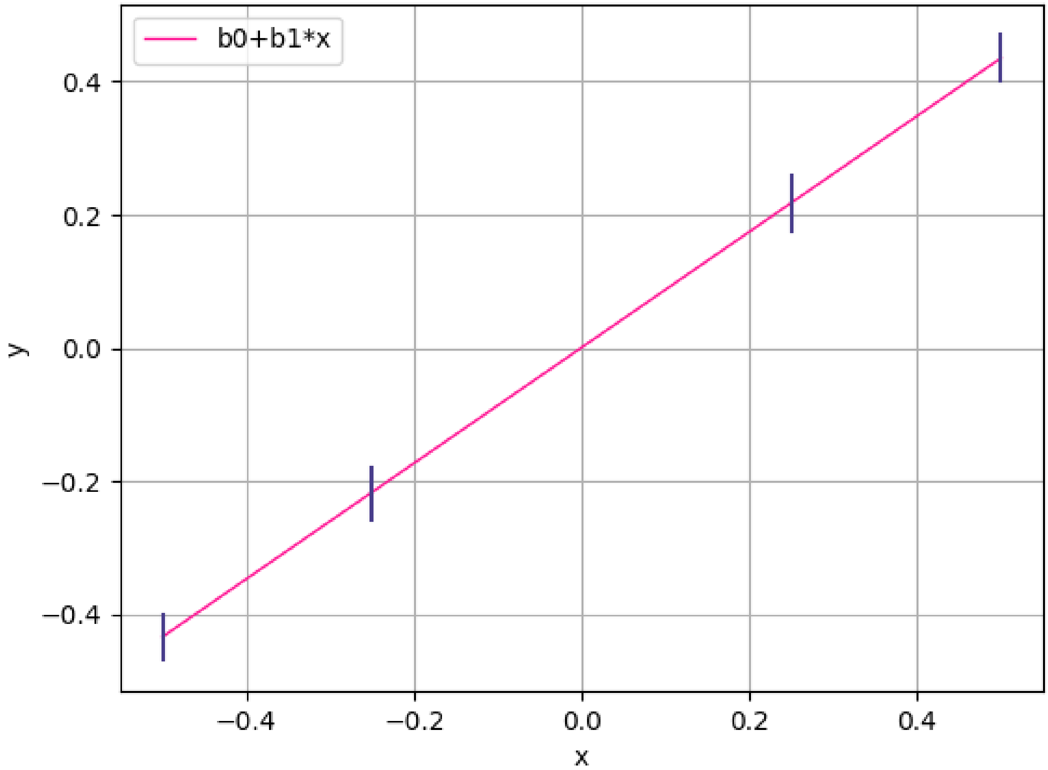


*Рисунок 7. Линейная регрессия для $Y_3$*

$$\beta_0 = 0.0005 \qquad \beta_1 = 0.8702$$

$$y = 0.8702x + 0.0005$$

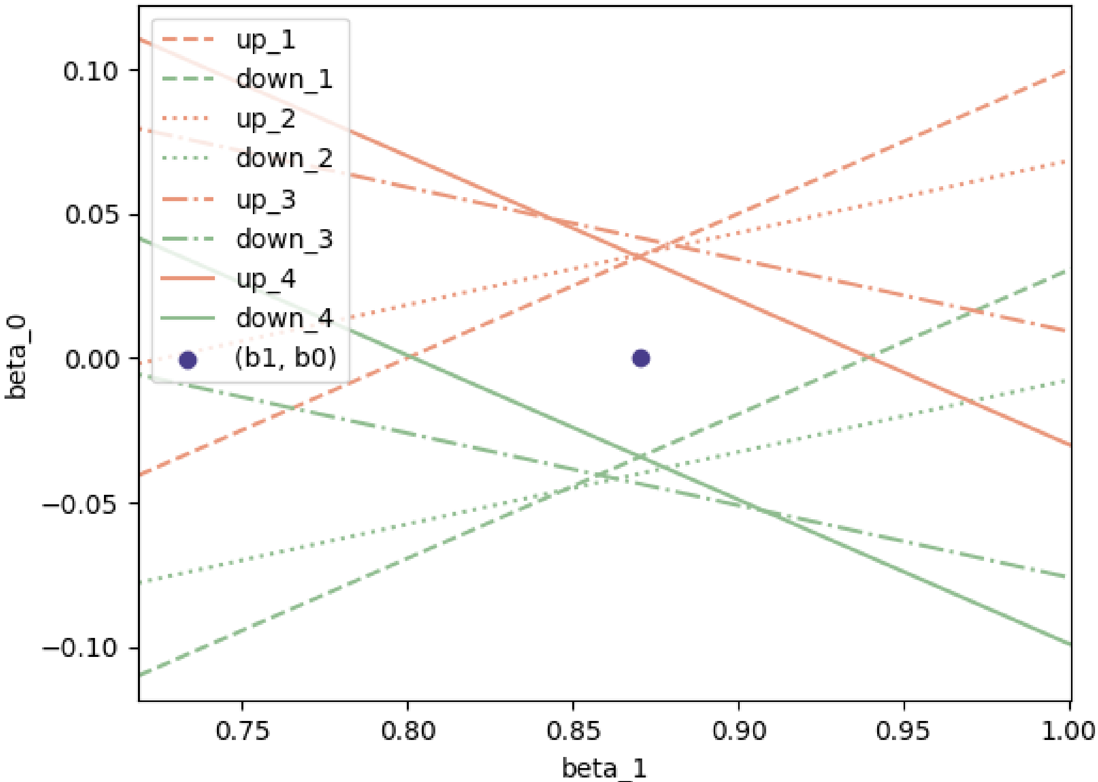


*Рисунок 8. Информационное множество для $Y_3$*

Интервальная оценка параметров $\beta_0 = [-0.0342, 0.0351]$ , $\beta_1 = [0.8009{,}0.9395]$.

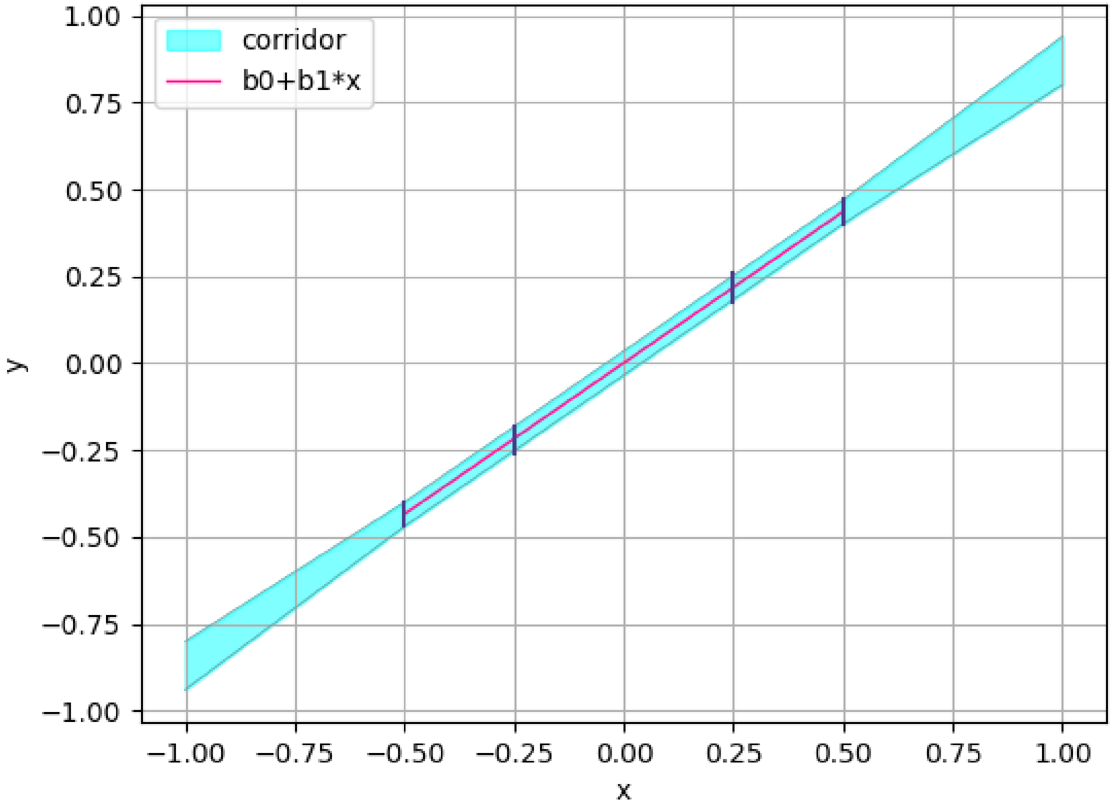


*Рисунок 9. Коридор совместных зависимостей для $Y_3$*

Рассмотрим построение линейной регрессии для интервалов, построенных по данным без вычитания погрешности $\delta_i$. Назовем выборку $Y_4$.

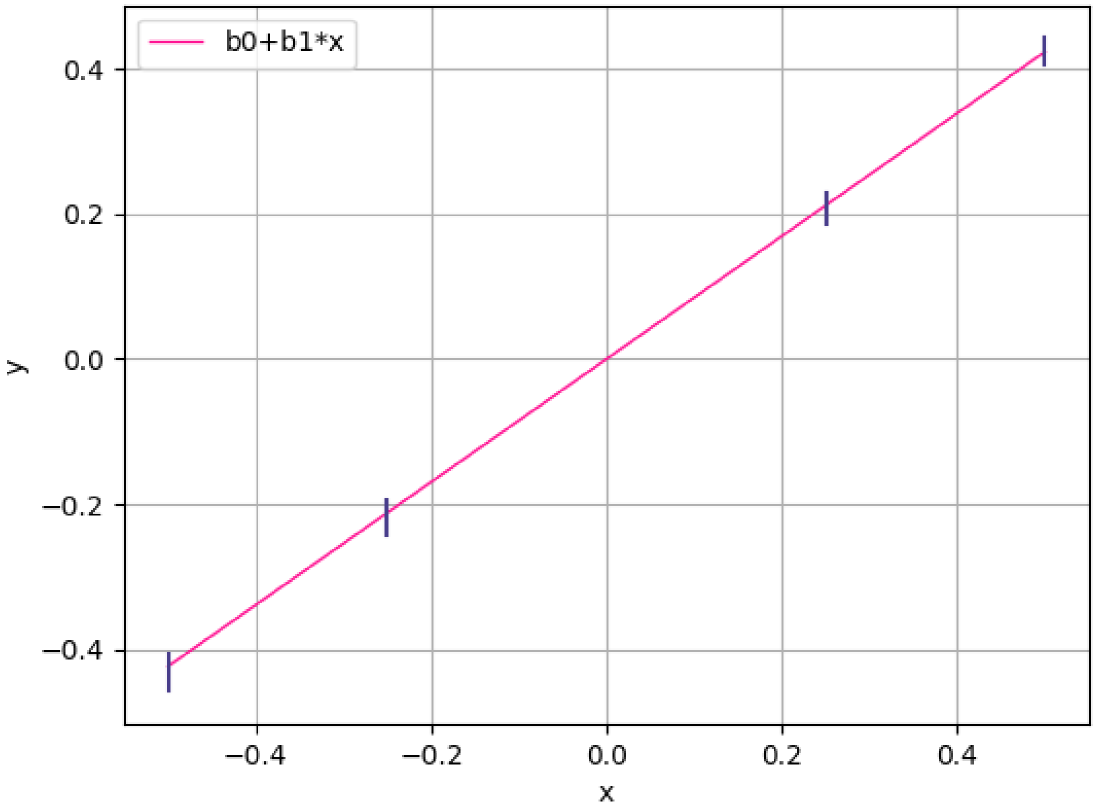


*Рисунок 10. Линейная регрессия для $Y_4$*

$$\beta_0 = 0.0 \qquad \beta_1 = 0.8468$$

$$y = 0.8468x$$

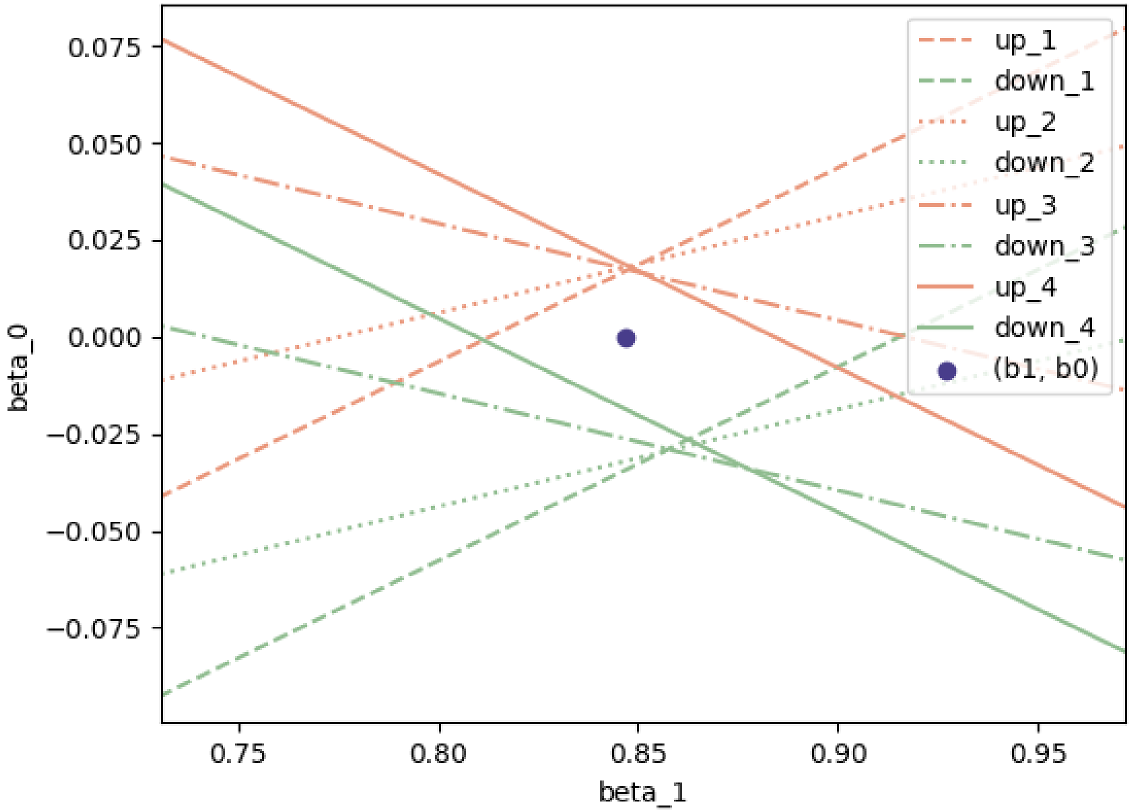


*Рисунок 11. Информационное множество для $Y_4$*

Интервальная оценка параметров $\beta_0 = [-0.0265, 0.0173]$, $\beta_1 = [0.8113, 0.8998]$.

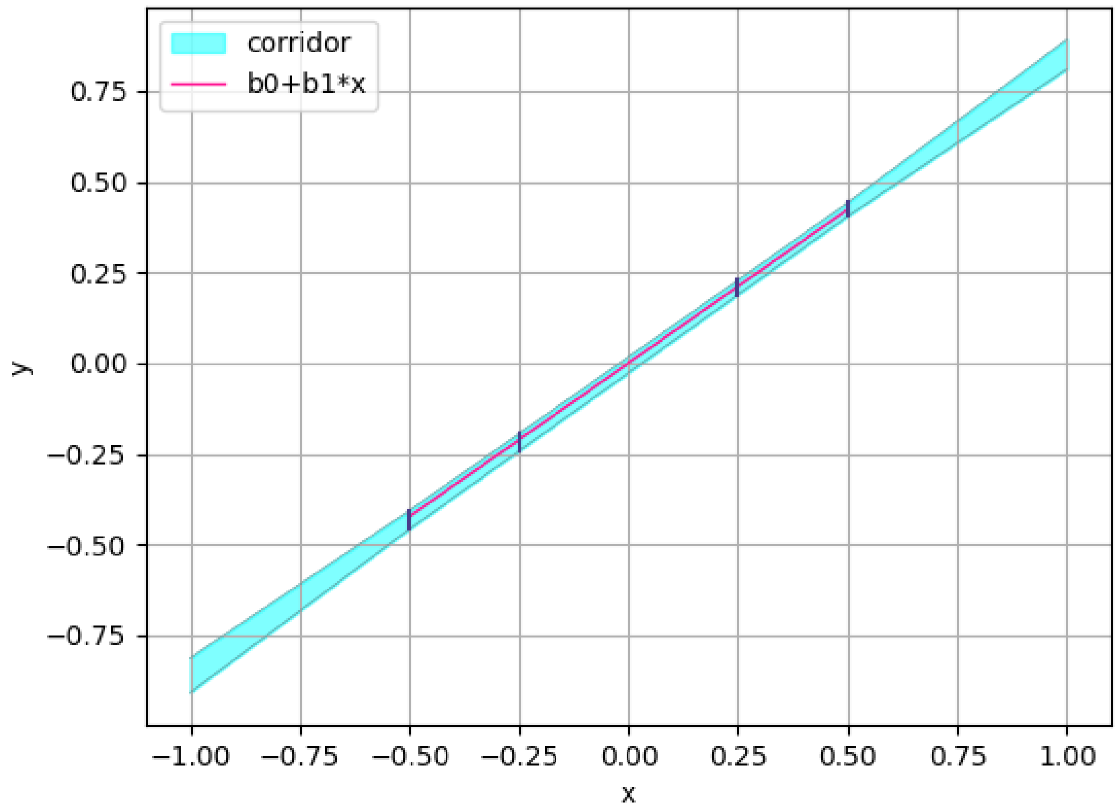


*Рисунок 12. Коридор совместных зависимостей для $Y_4$*

*Таблица 1. Параметры линейной регрессии для каждой выборки*

| | | |
|---|---|---|
| $Y_1$ | $y = 0.8538x + 0.0001$ | $\beta_0 = [-0.0041, 0.0050]$<br>$\beta_1 = [0.8421, 0.8621]$ |
| $Y_2$ | $y = 0.8492x$ | $\beta_0 = [-0.0065, 0.0026]$<br>$\beta_1 = [0.8421, 0.8621]$ |
| $Y_3$ | $y = 0.8702x + 0.0005$ | $\beta_0 = [-0.0342, 0.0351]$<br>$\beta_1 = [0.8009, 0.9395]$ |
| $Y_4$ | $y = 0.8468x$ | $\beta_0 = [-0.0265, 0.0173]$<br>$\beta_1 = [0.8113, 0.8998]$ |

## Анализ результатов

Первая и вторая интервальные выборки имеют одинаковую ширину интервалов, ширина коридора совместных зависимостей также примерно одинакова. Ширина интервалов третьей выборки больше ширины интервалов четвертой выборки, аналогично с шириной коридора совместных зависимостей. Оценка параметров регрессии для третьей выборки произведена с большей интервальной неопределенностью, чем в случае других выборок. Для всех выборок найденные точечные значения параметров регрессии лежат внутри информационного множества, а линия регрессии лежит внутри коридора совместных зависимостей.