Санкт-Петербургский
Политехнический университет Петра Великого

# Отчет по курсовой работе
# по дисциплине "Математическая статистика"
# по теме "Оценки коэффициентов линейной регрессии"

| | |
|---|---|
| Студенты: | Скворцов Владимир Сергеевич |
| | Горюнов Максим Юрьевич |
| | Голузин Егор Константинович |
| Преподаватель: | Баженов Александр Николаевич |
| Группа: | 5030102/10201 |

Санкт-Петербург
2024

# Содержание

# 1 Постановка задачи

На основе имеющихся данных о выходных значениях вольтметра, изменяющихся с течением времени, при известных входных значениях подаваемого напряжения необходимо получить оценку данной зависимости с использованием следующих методов:

1. Метод наименьших квадратов.
2. Метод наименьших модулей.

Для исследования предоставлен шестой столбец данных. Ссылка на исходный материал: https://disk.yandex.ru/d/OAQgCulS6NfbOQ.

# 2 Используемые теоретические понятия

## 2.1 Метод наименьших квадратов

При оценивании параметров регрессионной модели используют различные методы. Один из наиболее распрстранённых подходов заключается в следующем: вводится мера (критерий) рассогласования отклика и регрессионной функции, и оценки параметров регрессии определяются так, чтобы сделать это рассогласование наименьшим. Достаточно простые расчётные формулы для оценок получают при выборе критерия в виде суммы квадратов отклонений значений отклика от значений регрессионной функции (сумма квадратов остатков):

$$Q(\beta_0, \beta_1) = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 \to \min_{\beta_0, \beta_1}. \tag{1}$$

Задача минимизации квадратичного критерия (1) носит название задачи метода наименьших квадратов (МНК), а оценки $\beta_0$, $\beta_1$ параметров $\beta_0$, $\beta_1$, реализующие минимум критерия (1), называют МНК-оценками.

## 2.2 Метод наименьших модулей

Робастность оценок коэффициентов линейной регрессии (т.е. их устойчивость по отношению к наличию в данных редких, но больших по величине выбросов) может быть обеспечена различными способами. Одним из них является использование метода наименьших модулей вместо метода наименьших квадратов:

$$\sum_{i=1}^{n} |y_i - \beta_0 - \beta_1 x_i| \to \min_{\beta_0, \beta_1}. \tag{2}$$

## 2.3 Бокс-плот Тьюки

Боксплот (англ. box plot) — график, использующихся в описательной статистике, компактно изобрадающий одномерное распределение вероятностей. Такой вид диаграммы в удобной форме показывает медиану, нижний и верхний квартили и выбросы. Границами ящика служат первый и третий квартили, линия в середине ящика — медиана. Концы усов — края статистически значимой выборки (без выброса). Длину «усов» определяют разность первого квартиля и полутора межквартальных расстояний и сумма третьего квартиля и полутора межквартальных расстояний. Формула имеет вид

$$X_1 = Q_1 - \frac{3}{2}(Q_3 - Q_1),\ X_2 = Q_3 + \frac{3}{2}(Q_3 - Q_1), \tag{3}$$

где $X_1$ — нижняя граница уса, $X_2$ — верхняя граница уса, $Q_1$ — первый квартиль, $Q_3$ - третий квартиль. Данные, выходящие за границы усов (выбросы), отображаются на графике в виде маленьких кружков. Выбросами считаются величины , такие что:

$$\left[ \begin{array}{l} x < X_1^T \\ x > X_2^T \end{array} \right. \tag{4}$$

# 3 Описание работы

Лабораторные работы выполнены с использованием Python и сторонних библиотек `numpy`, `pandas`, `matplotlib`, `seaborn`. Для каждого значения напряжения данные были обработаны при помощи боксплота Тьюки, для обработанных данных также был построен график исследуемой зависимости.

Ссылка на GitHub репозиторий:
https://github.com/vladimir-skvortsov/math-stats-course-work

# 4 Результаты

## 4.1 Напряжение $U = -0.45V$

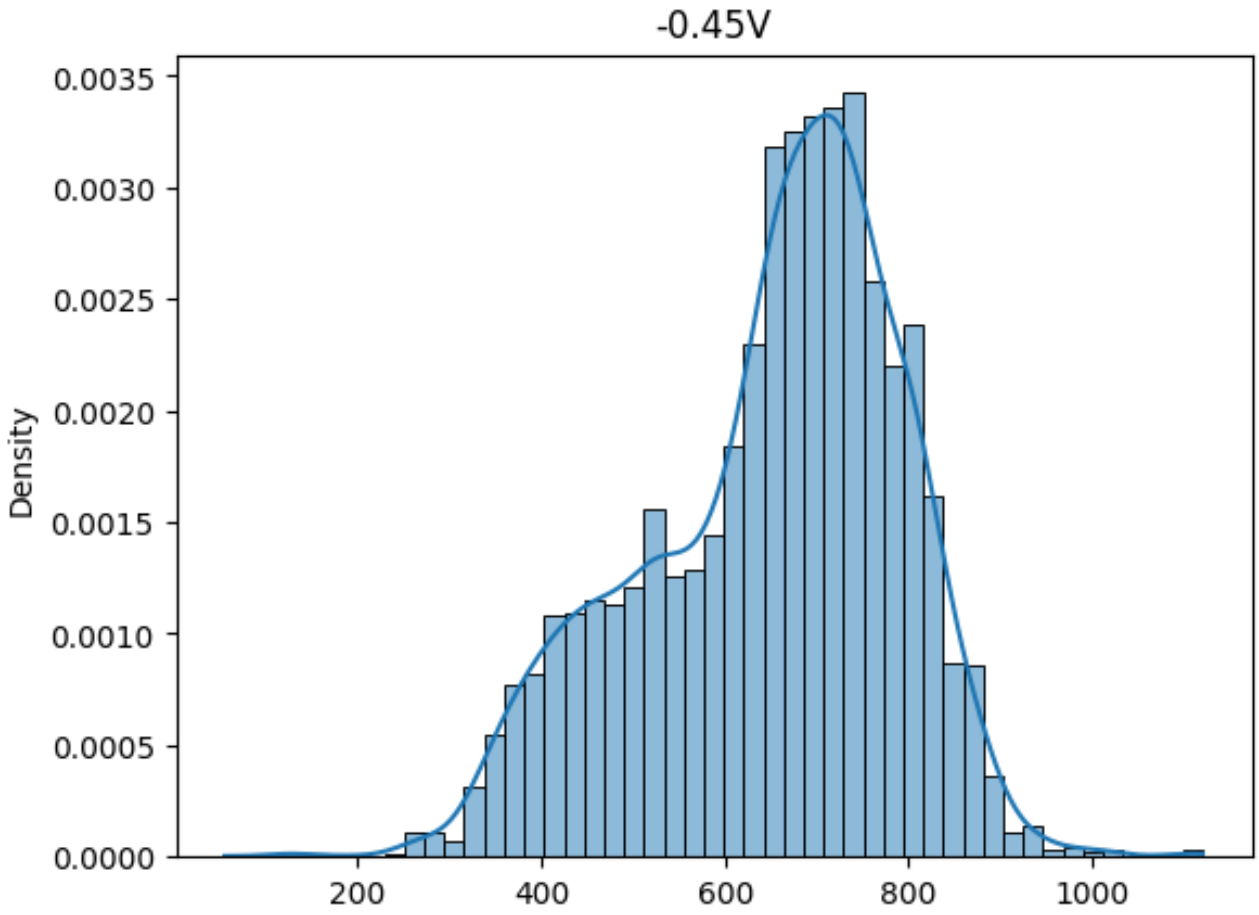


Рис. 1: Гистограмма распределения значений выходного напряжения

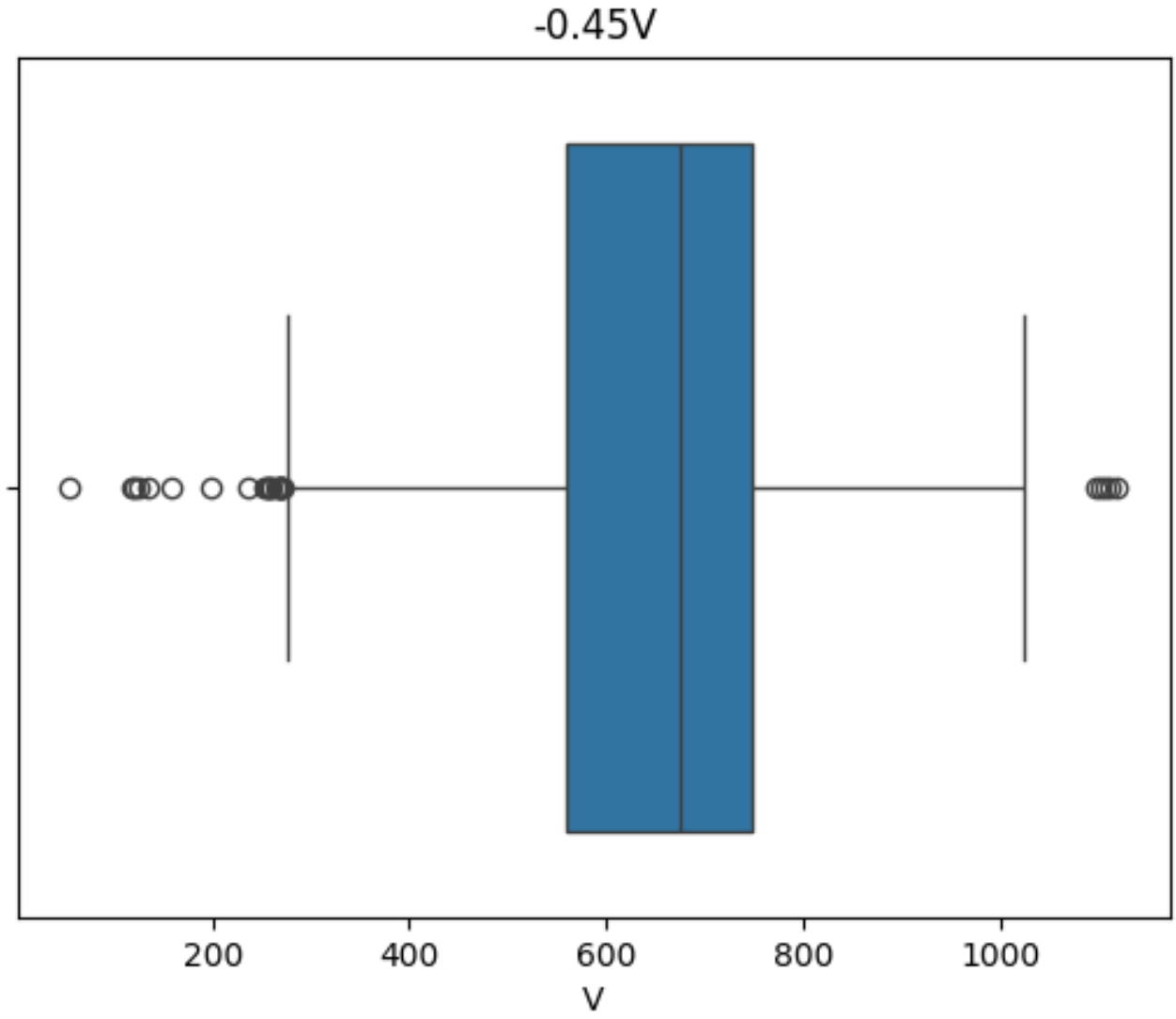


Рис. 2: Боксплот распределения значений выходного напряжения

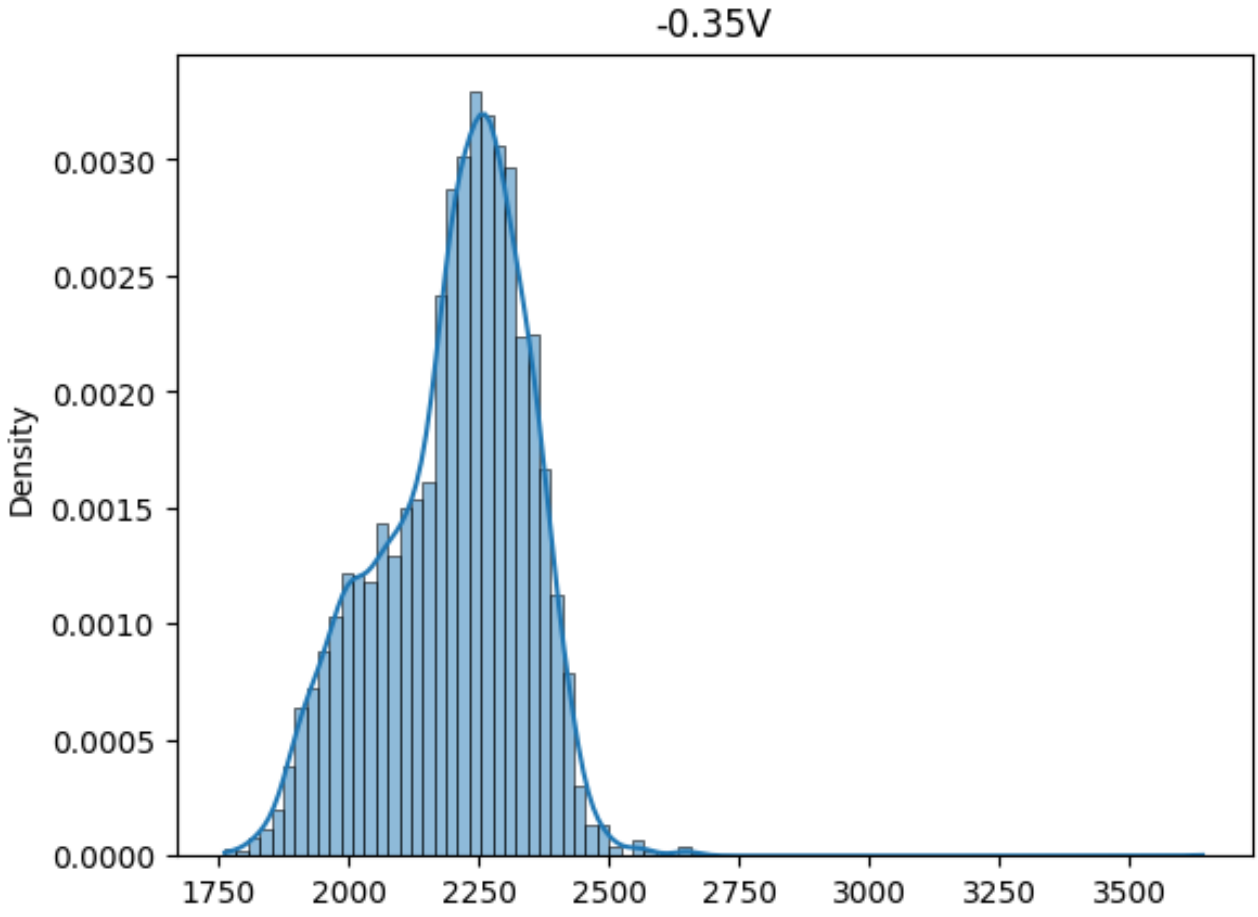


Рис. 3: Гистограмма распределения значений выходного напряжения

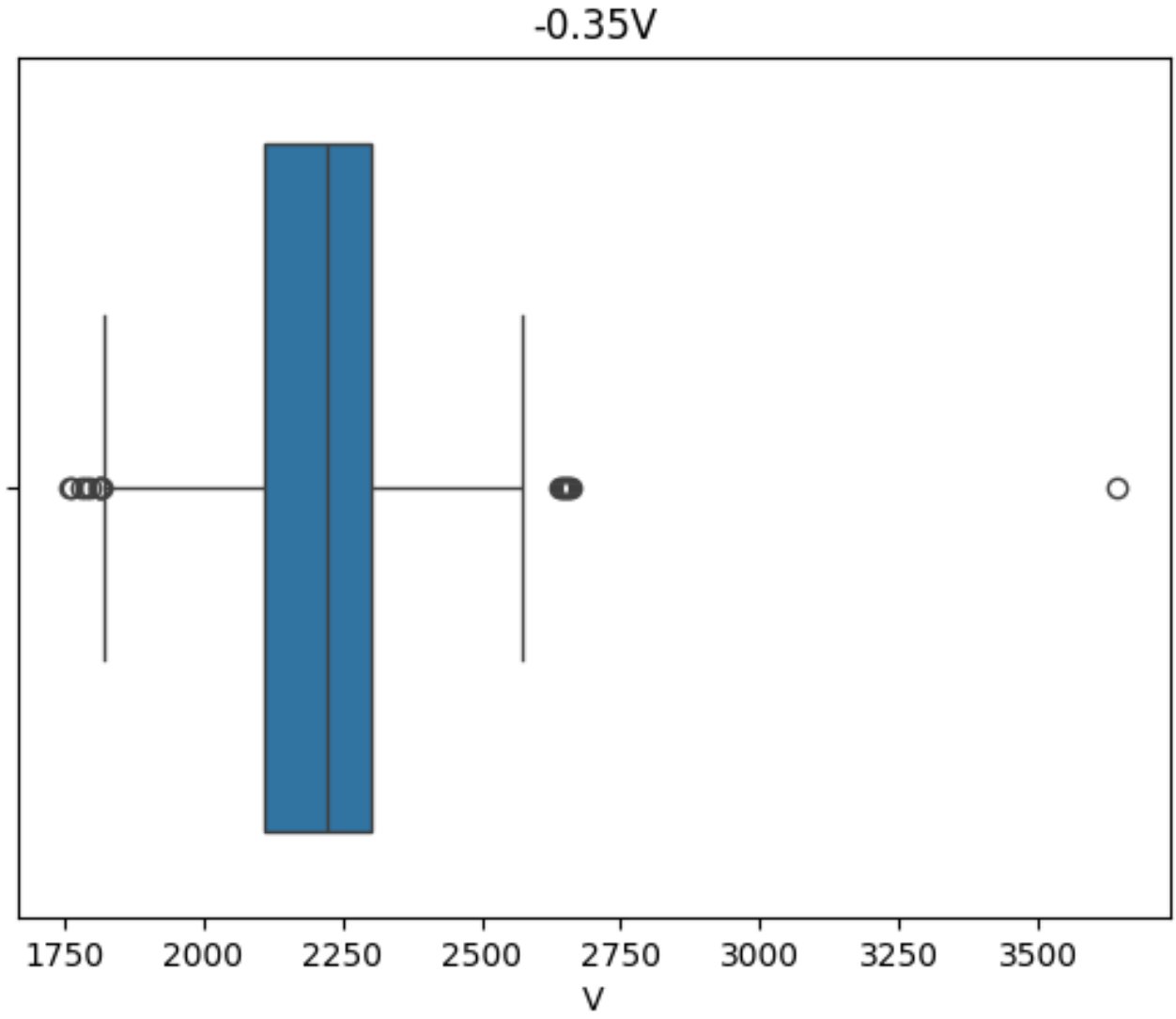


Рис. 4: Боксплот распределения значений выходного напряжения

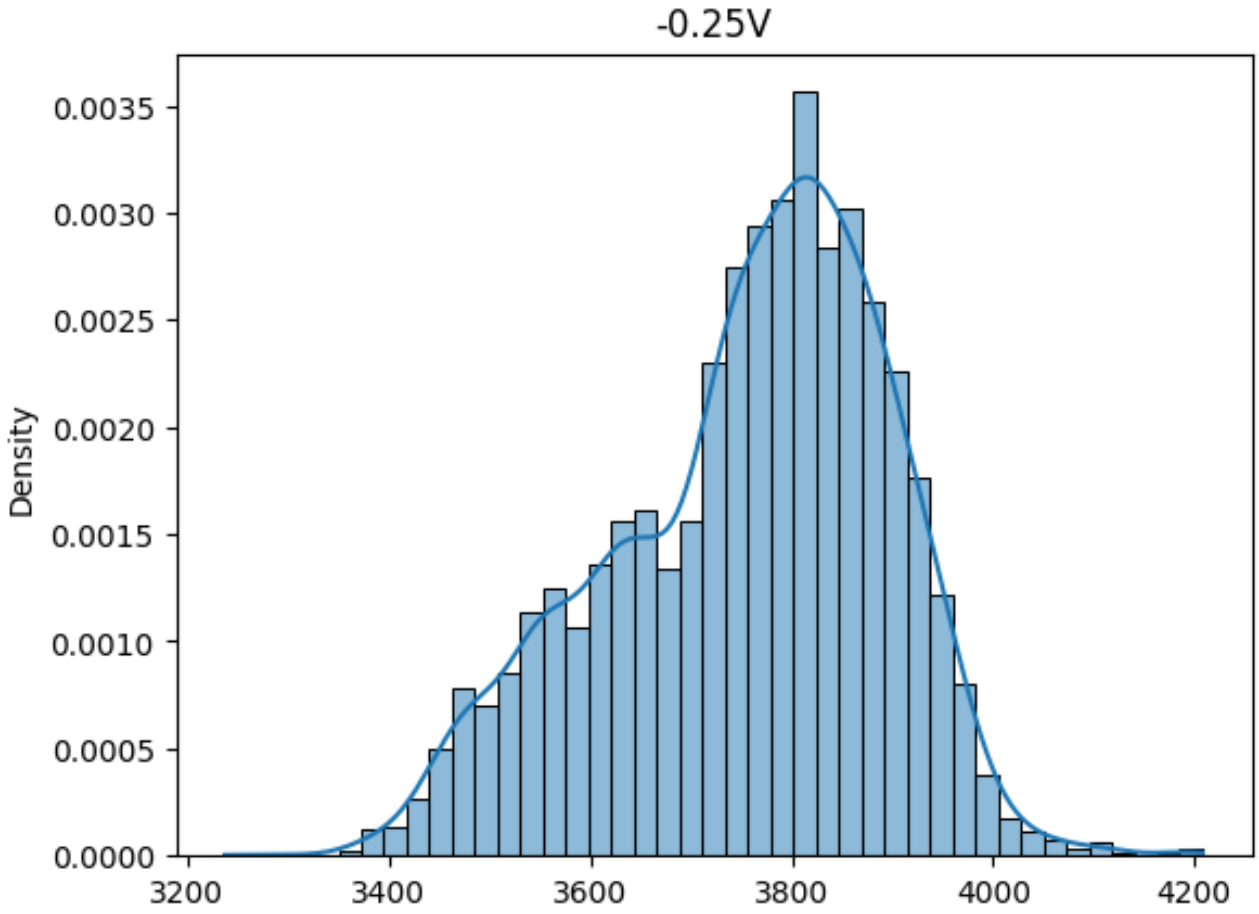


Рис. 5: Гистограмма распределения значений выходного напряжения

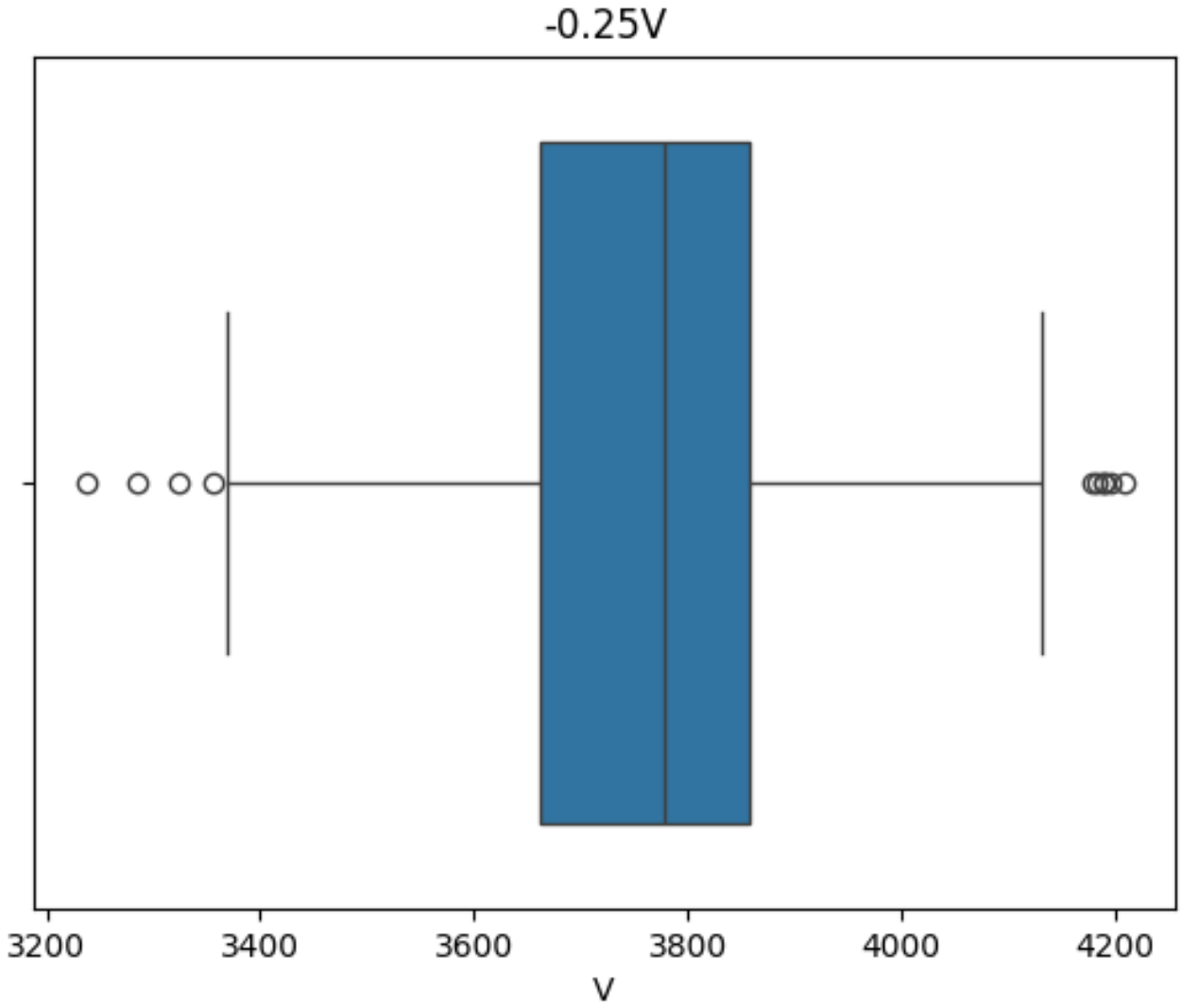


Рис. 6: Боксплот распределения значений выходного напряжения

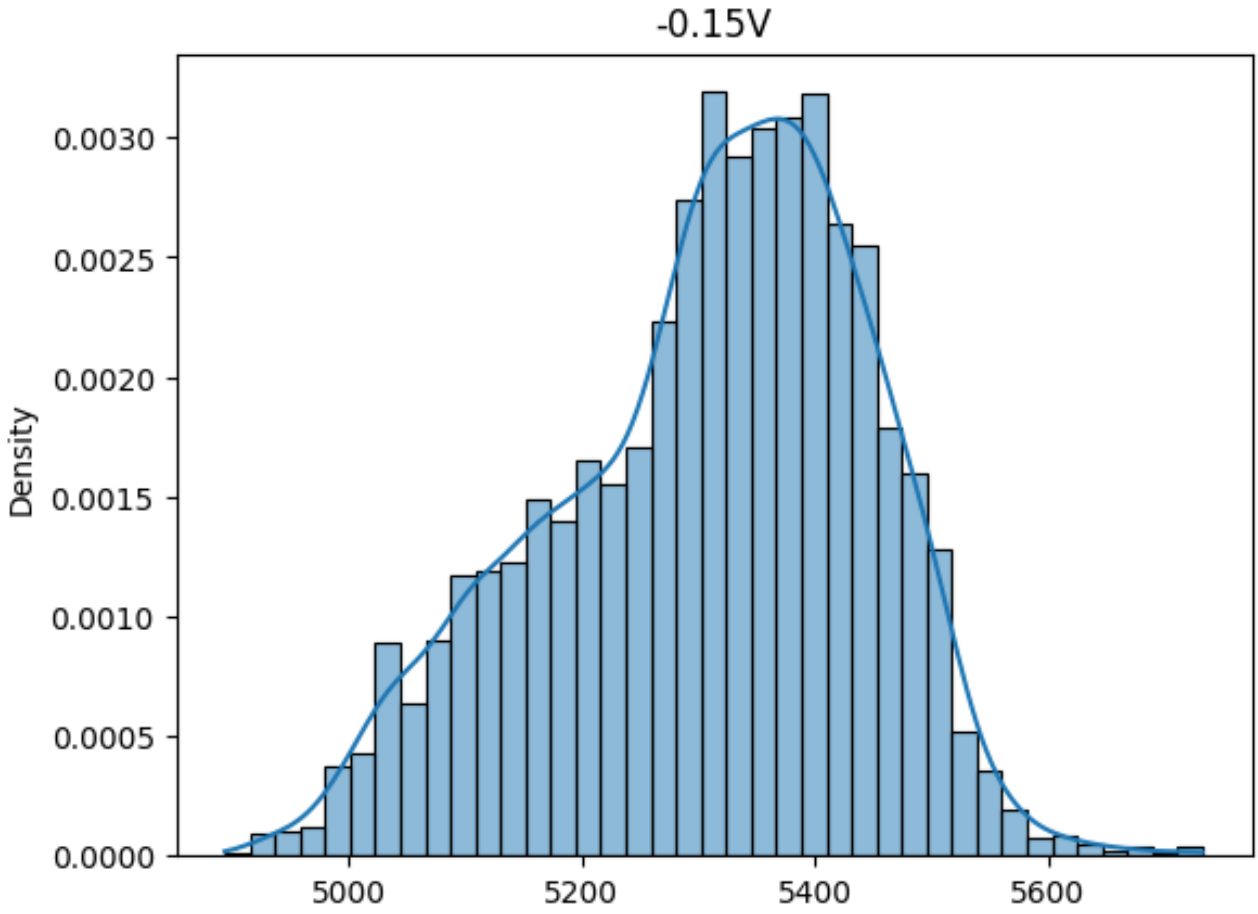


Рис. 7: Гистограмма распределения значений выходного напряжения

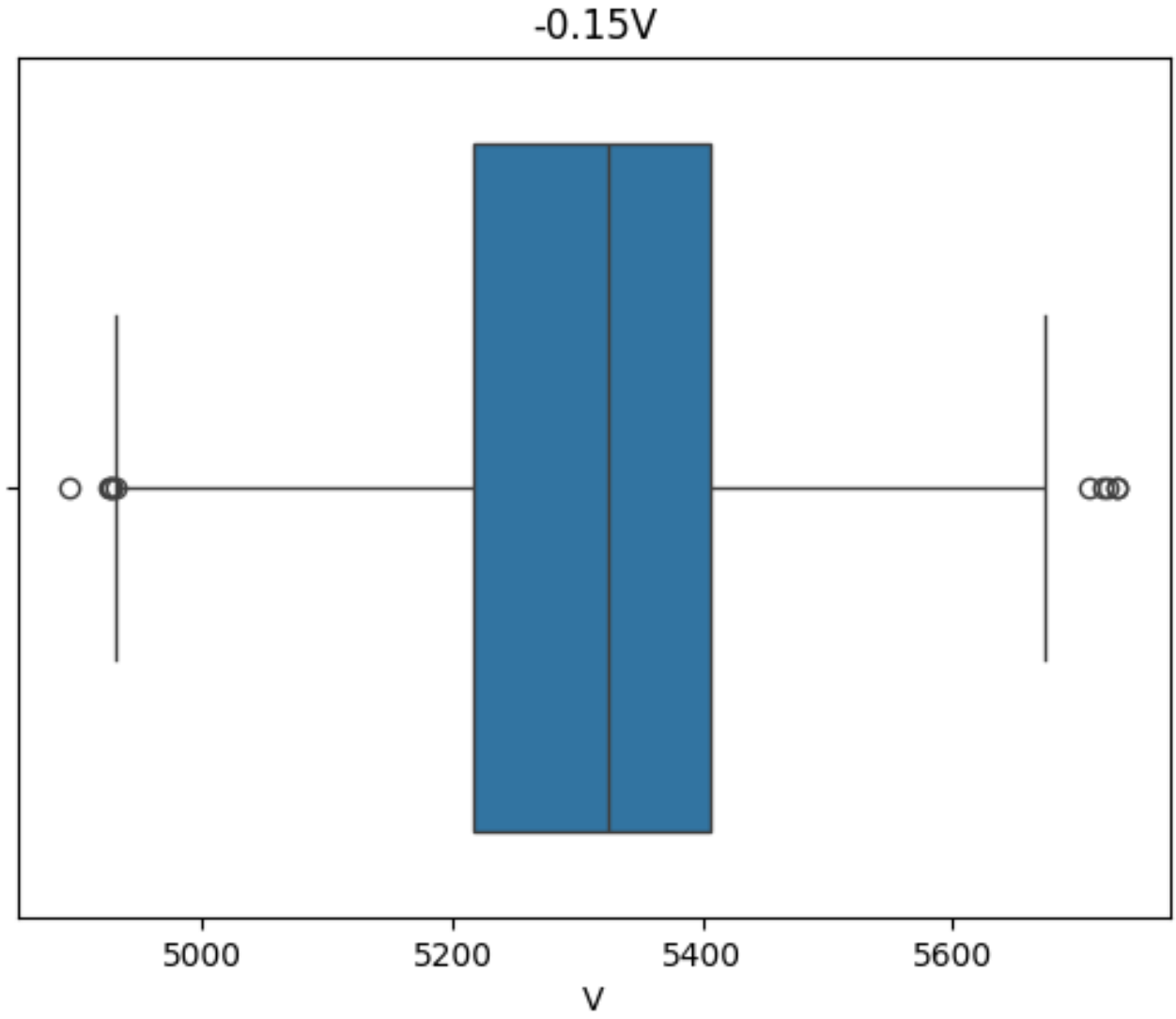


Рис. 8: Боксплот распределения значений выходного напряжения

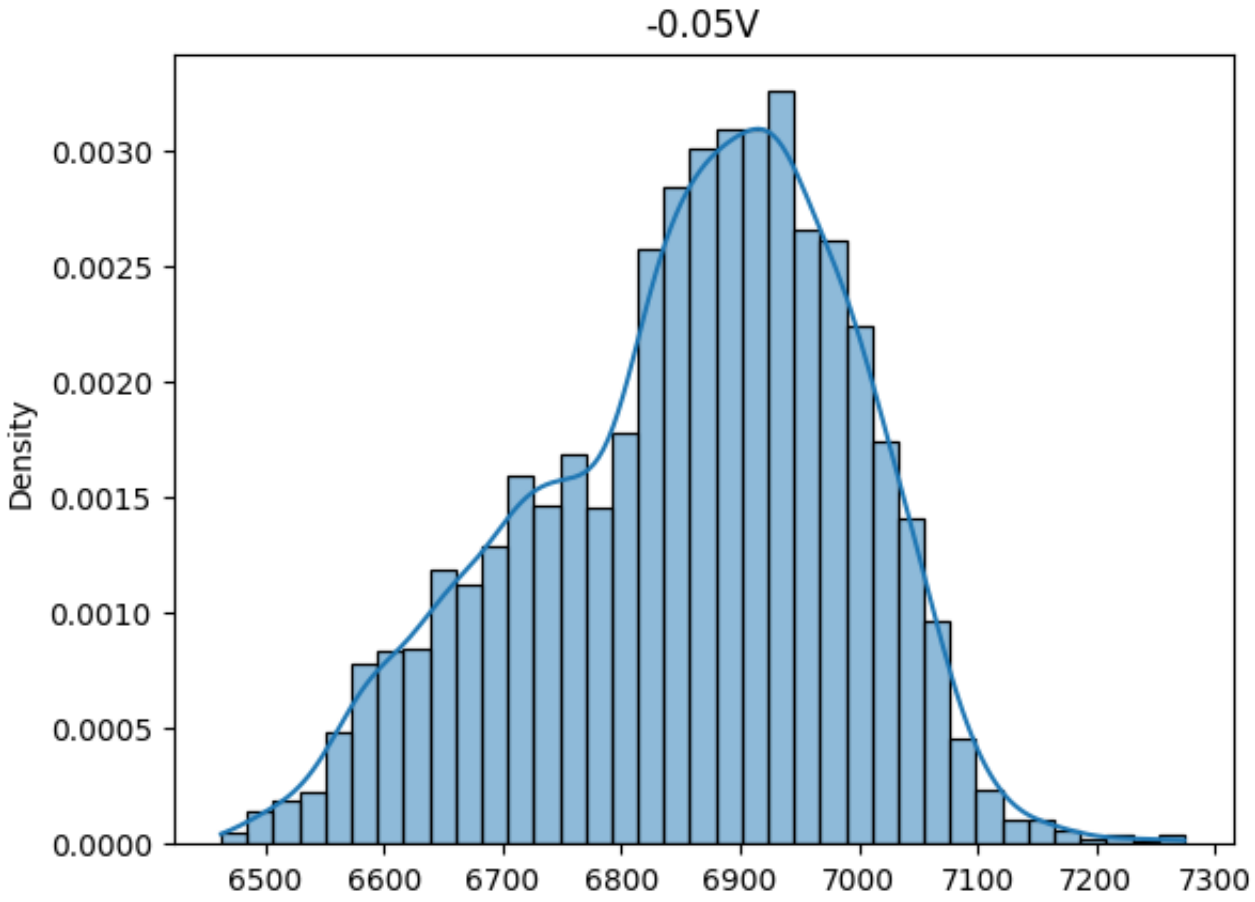


Рис. 9: Гистограмма распределения значений выходного напряжения

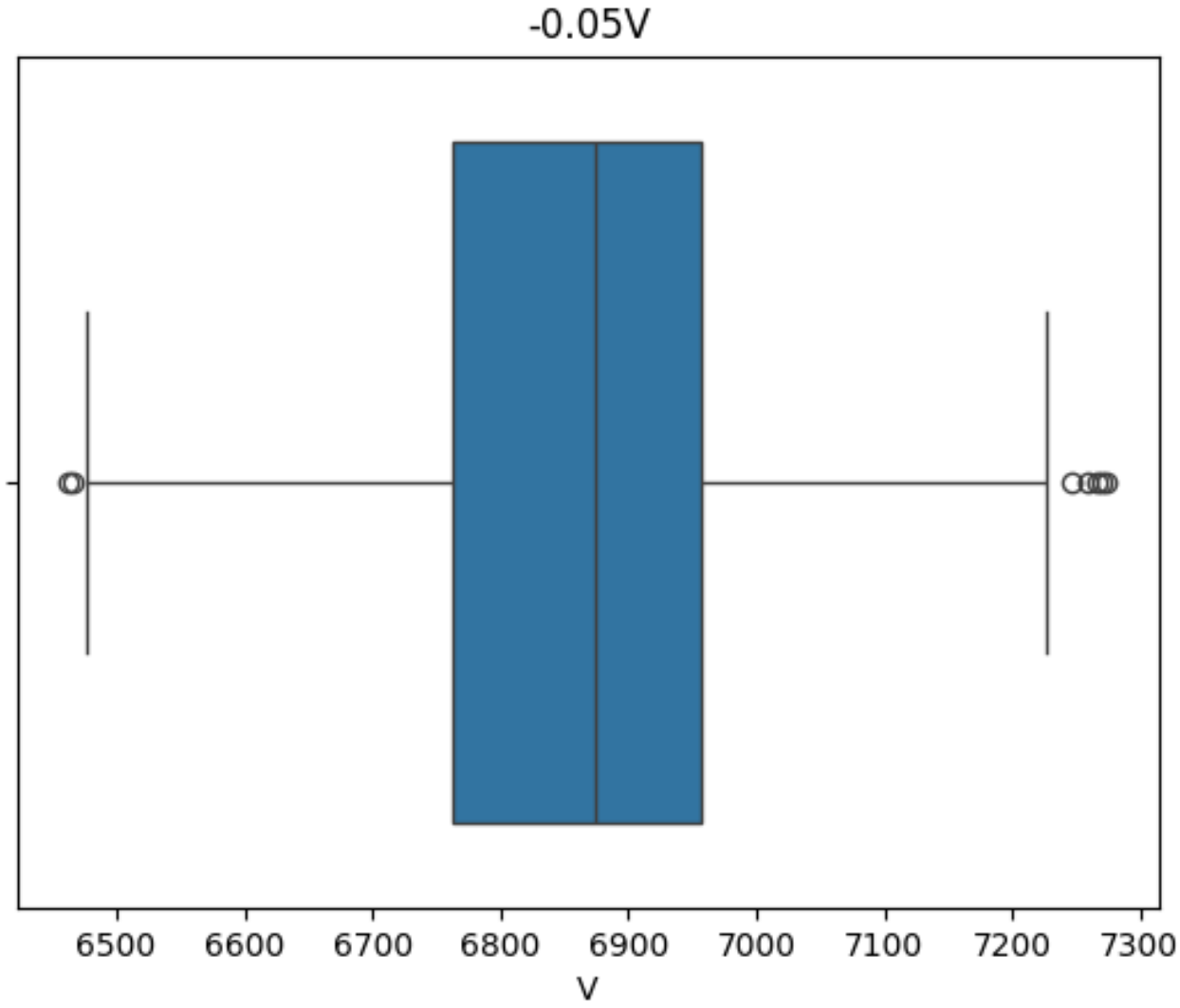


Рис. 10: Боксплот распределения значений выходного напряжения

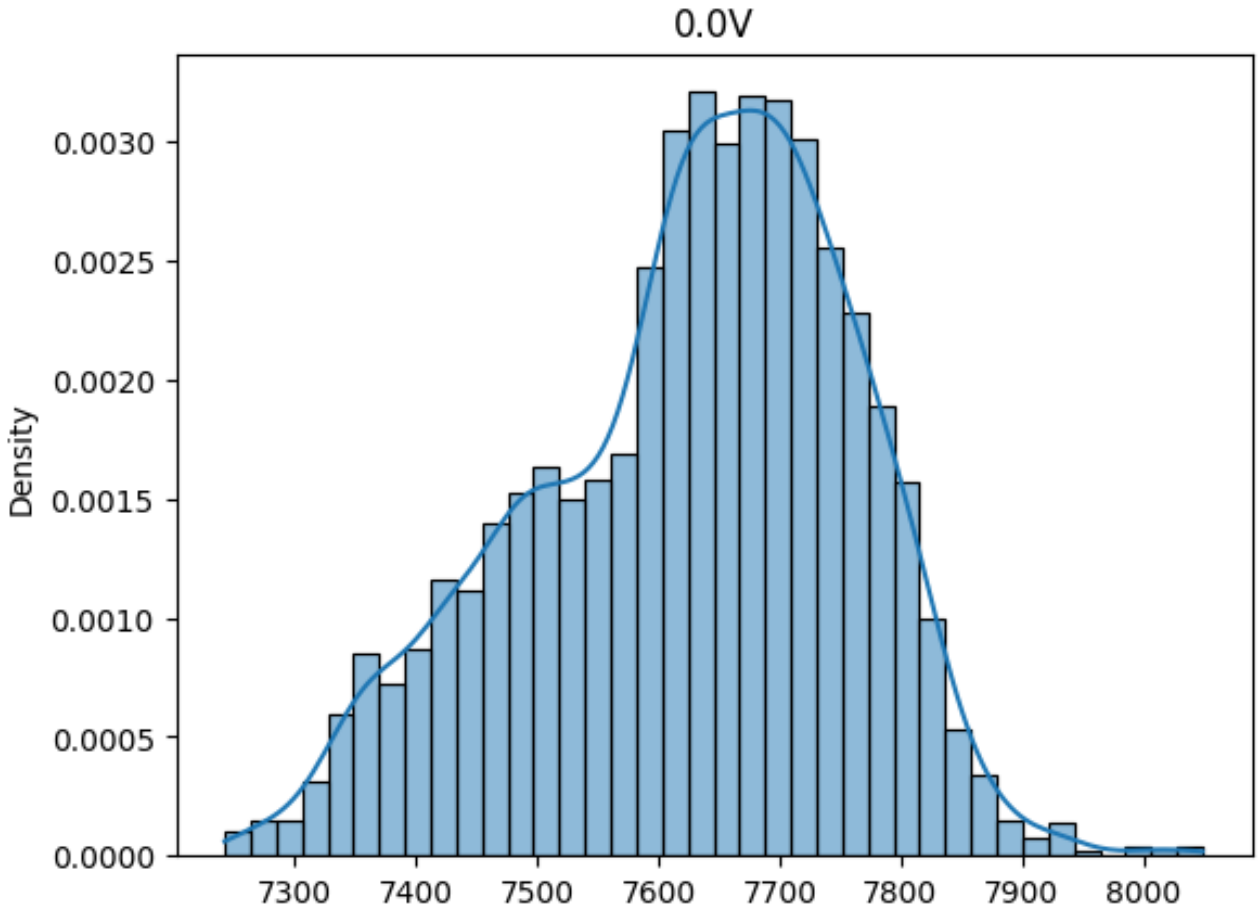


Рис. 11: Гистограмма распределения значений выходного напряжения

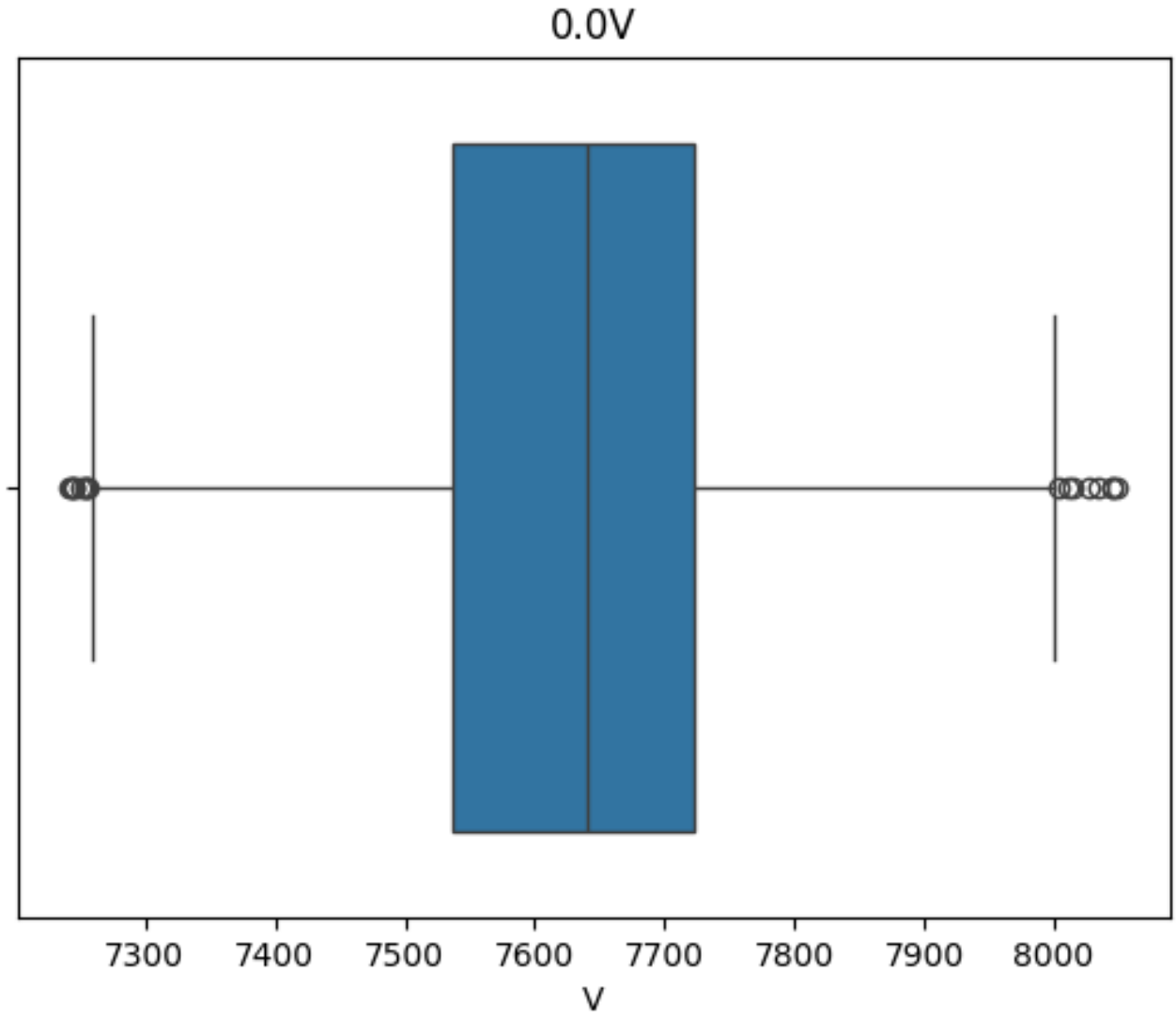


Рис. 12: Боксплот распределения значений выходного напряжения

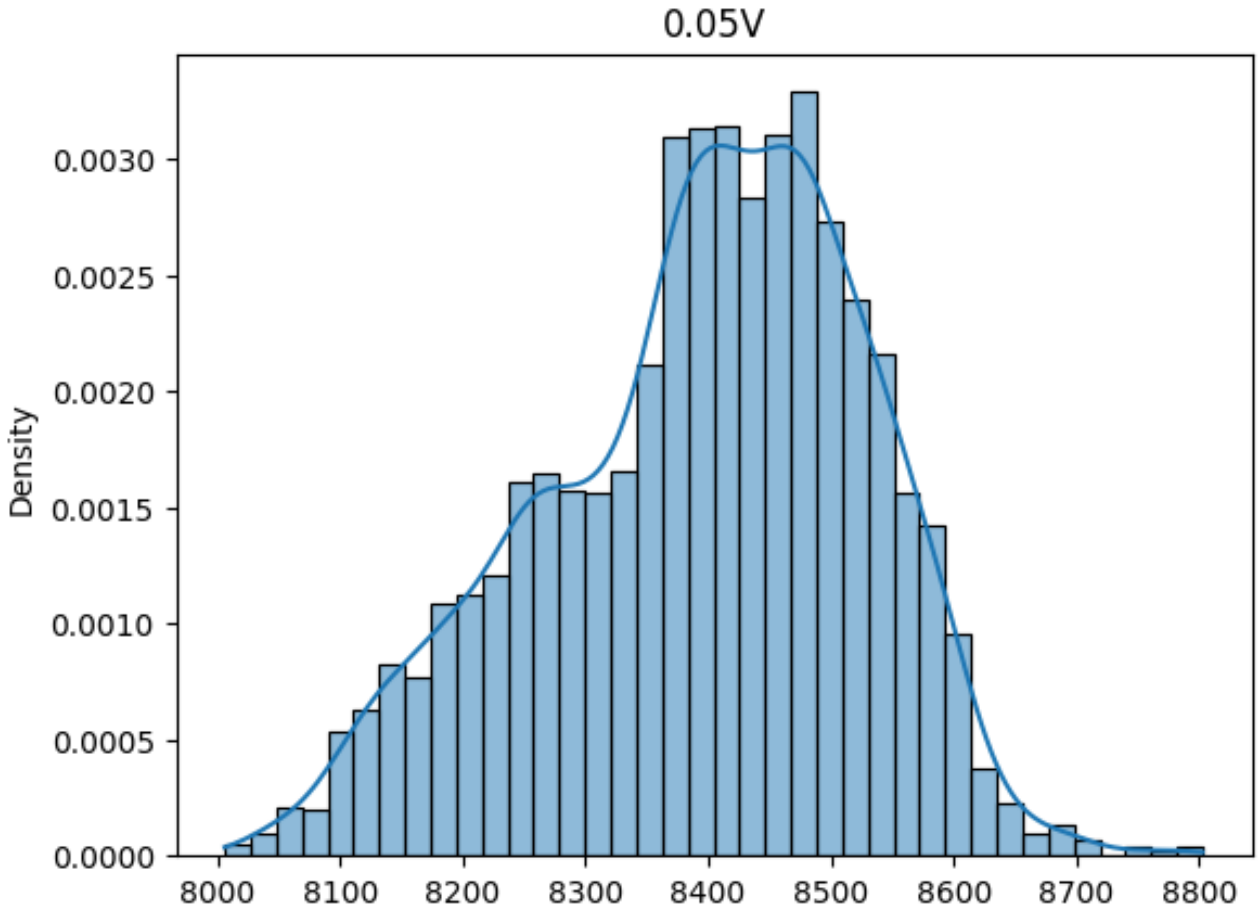


Рис. 13: Гистограмма распределения значений выходного напряжения

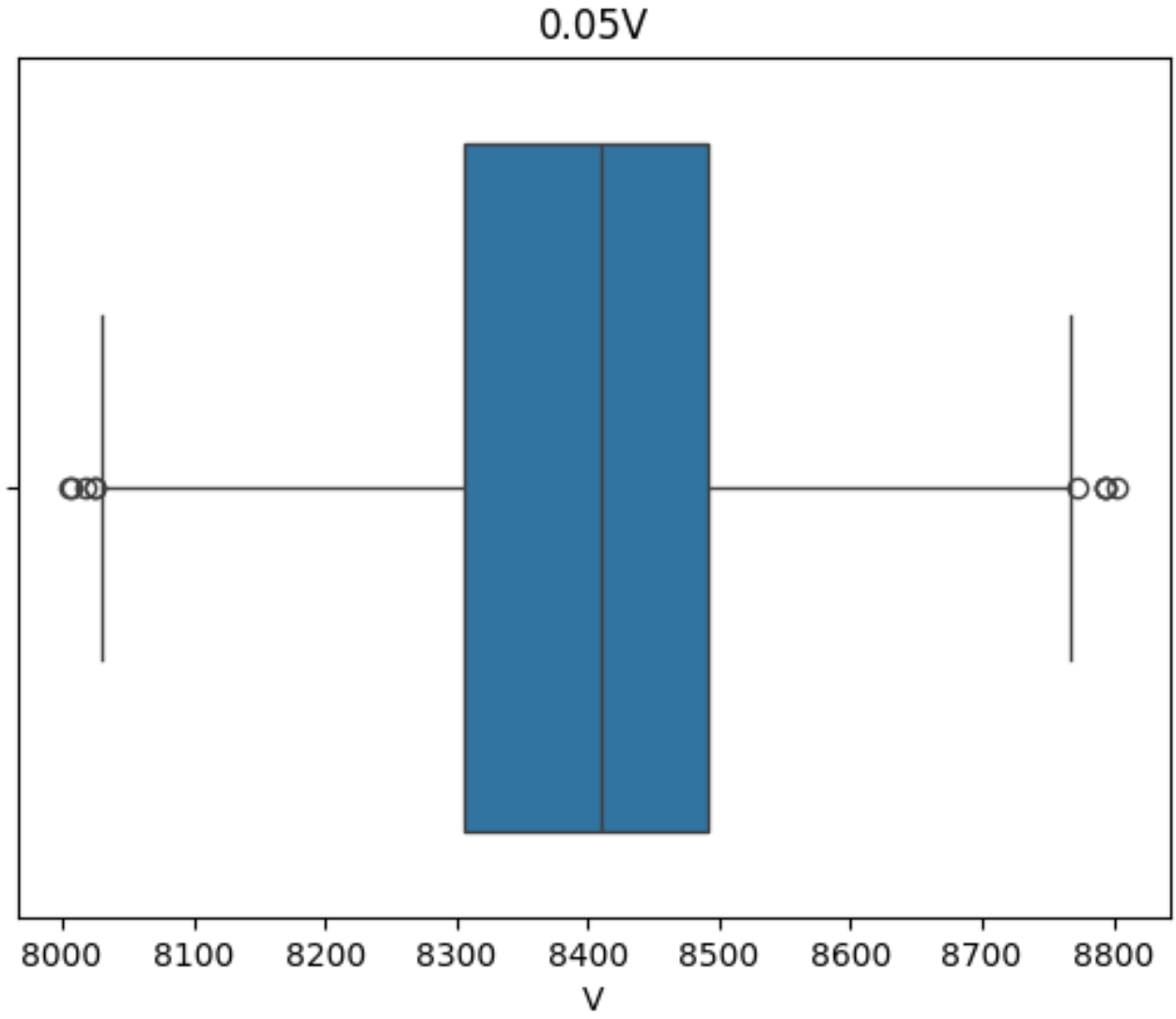


Рис. 14: Боксплот распределения значений выходного напряжения

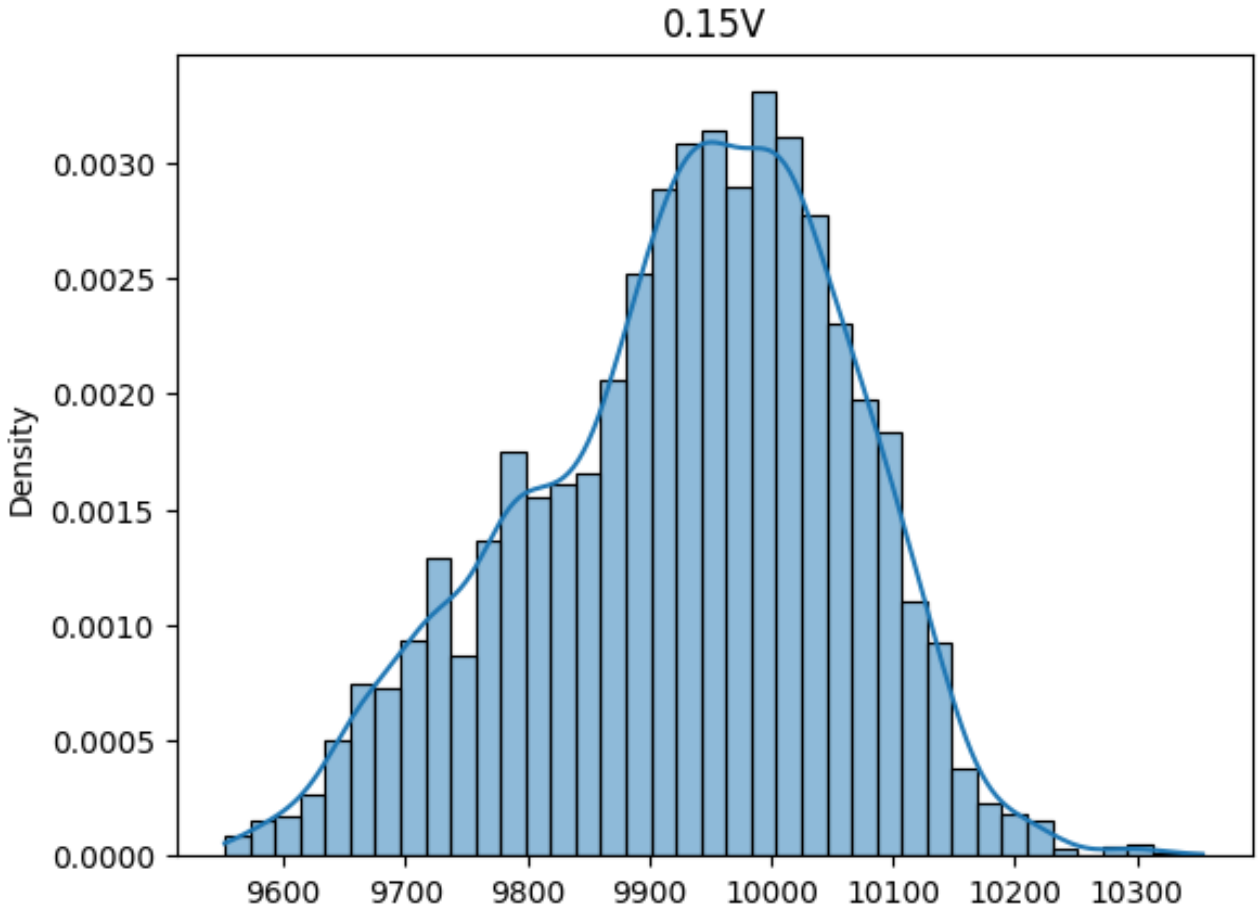


Рис. 15: Гистограмма распределения значений выходного напряжения

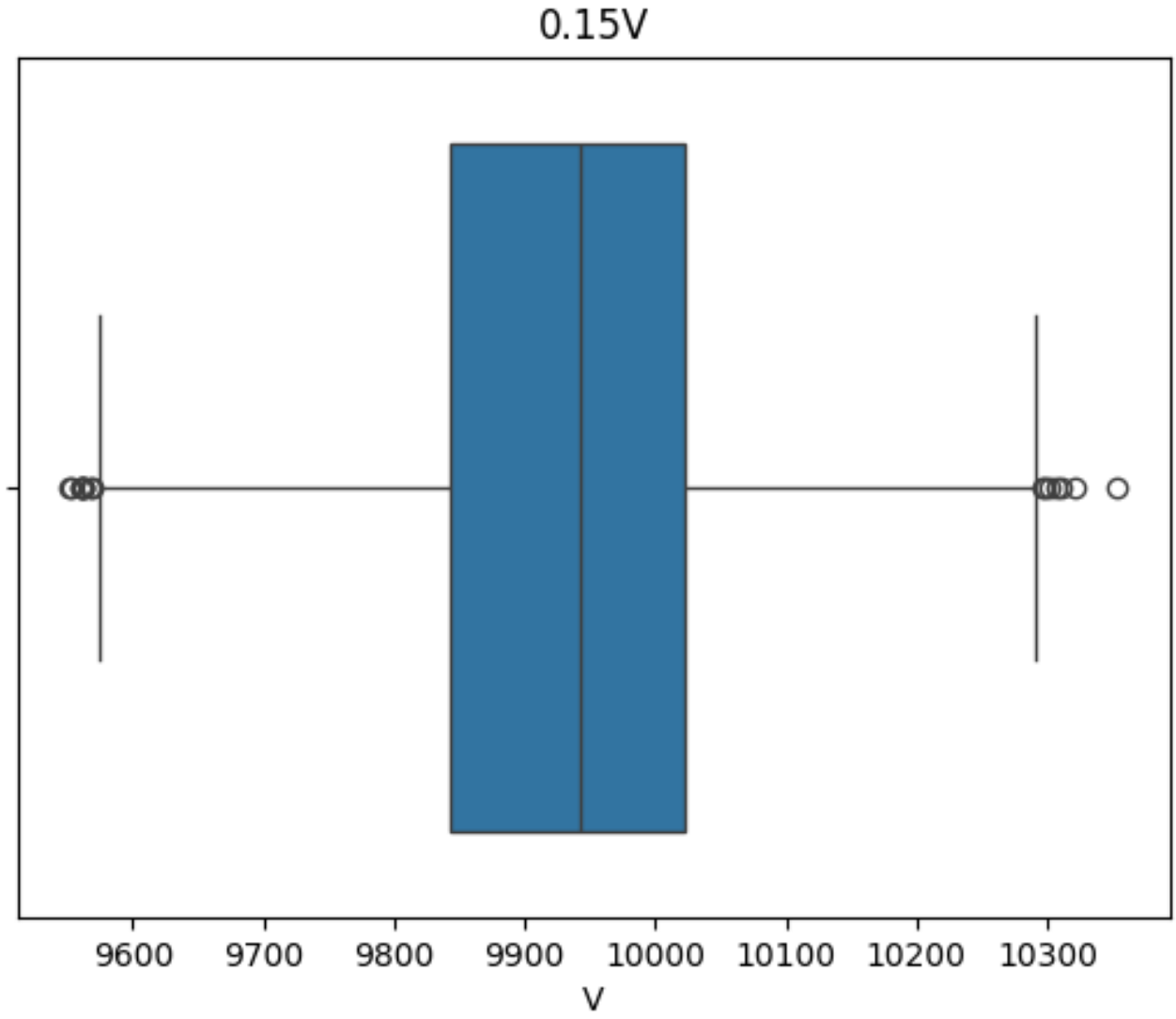


Рис. 16: Боксплот распределения значений выходного напряжения

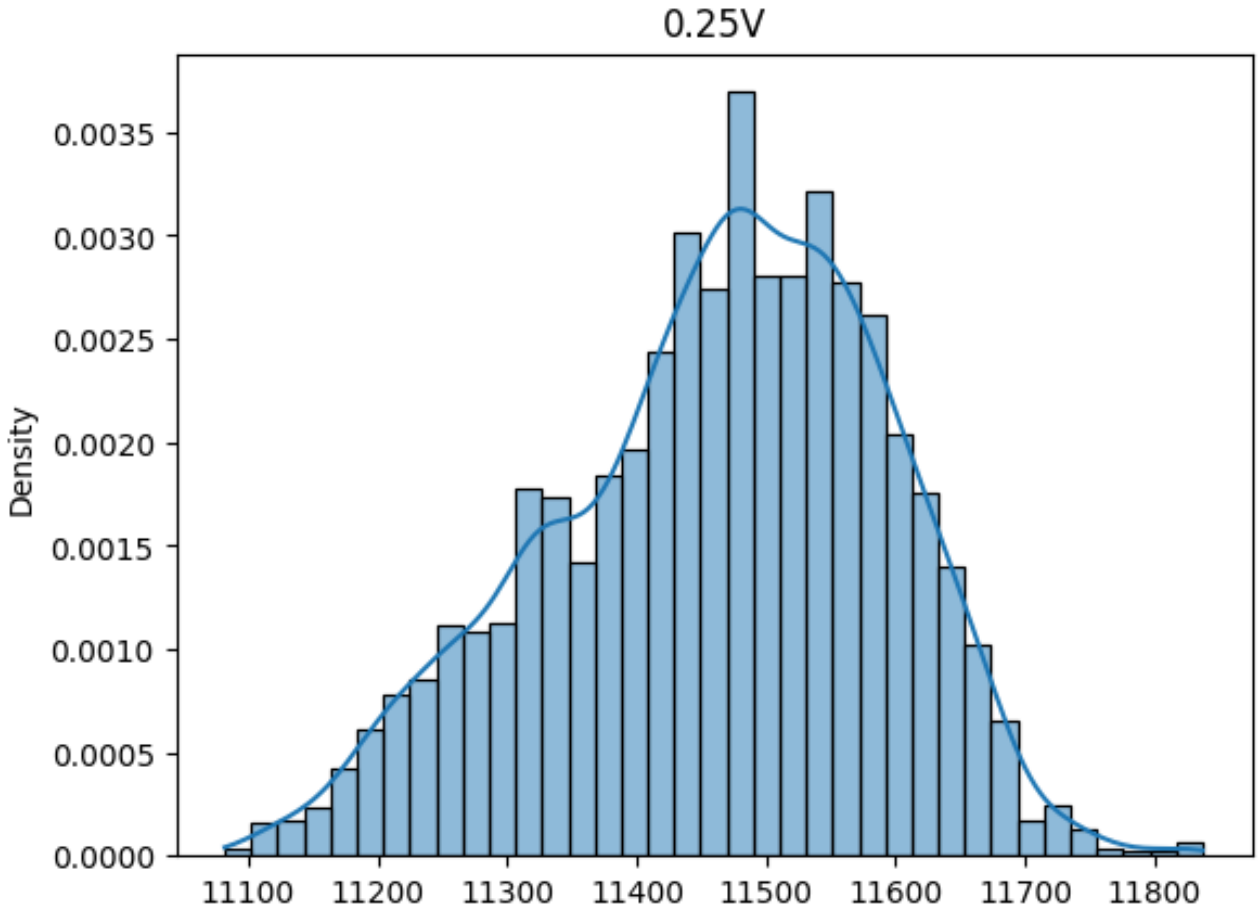


Рис. 17: Гистограмма распределения значений выходного напряжения

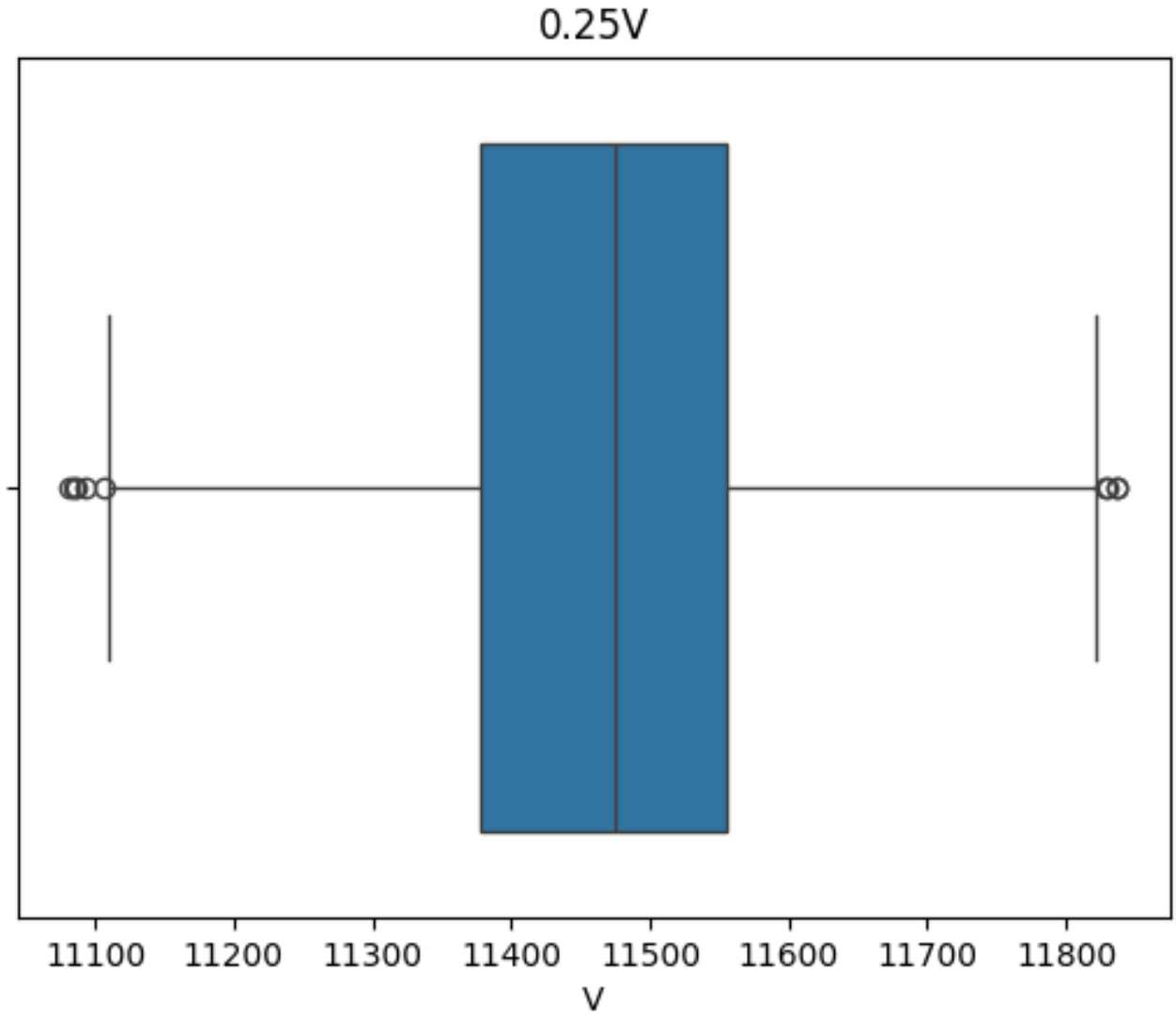


Рис. 18: Боксплот распределения значений выходного напряжения

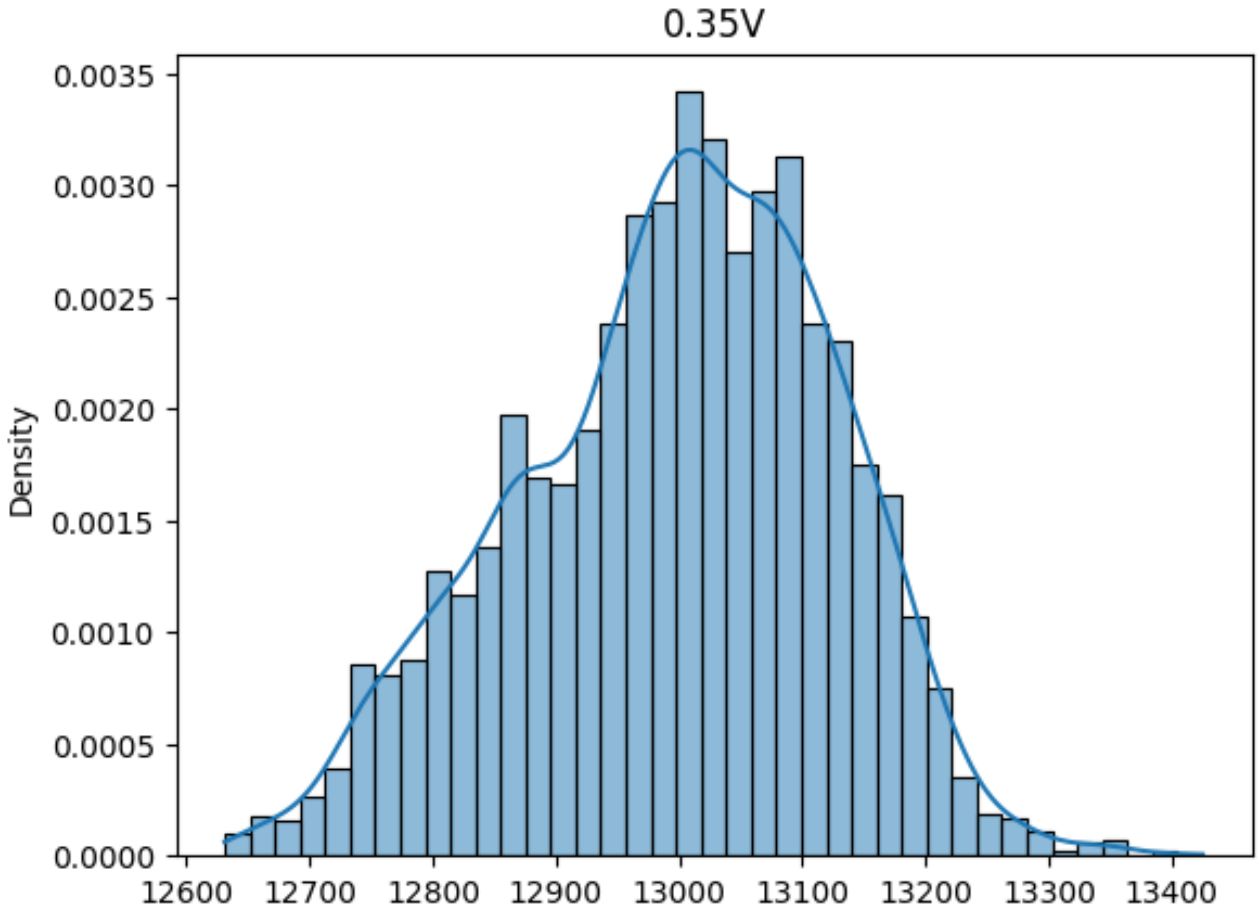


Рис. 19: Гистограмма распределения значений выходного напряжения

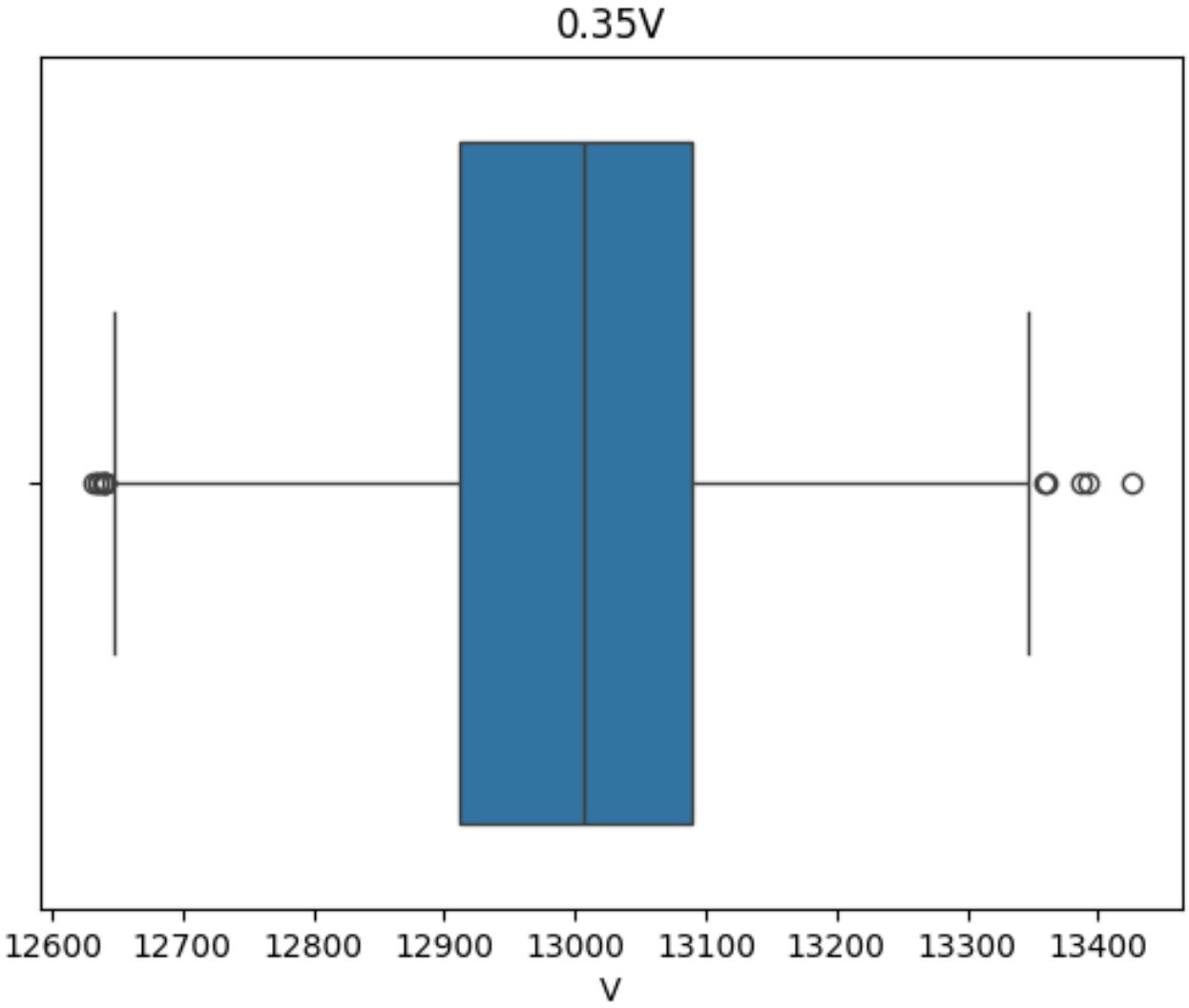


Рис. 20: Боксплот распределения значений выходного напряжения

**4.2 Напряжение $U = -0.35V$**

**4.3 Напряжение $U = -0.25V$**

**4.4 Напряжение $U = -0.15V$**

**4.5 Напряжение $U = -0.05V$**

**4.6 Напряжение $U = 0.0V$**

**4.7 Напряжение $U = 0.05V$**

**4.8 Напряжение $U = 0.15V$**

**4.9 Напряжение $U = 0.25V$**

**4.10 Напряжение $U = 0.35V$**

**4.11 Напряжение $U = 0.45V$**

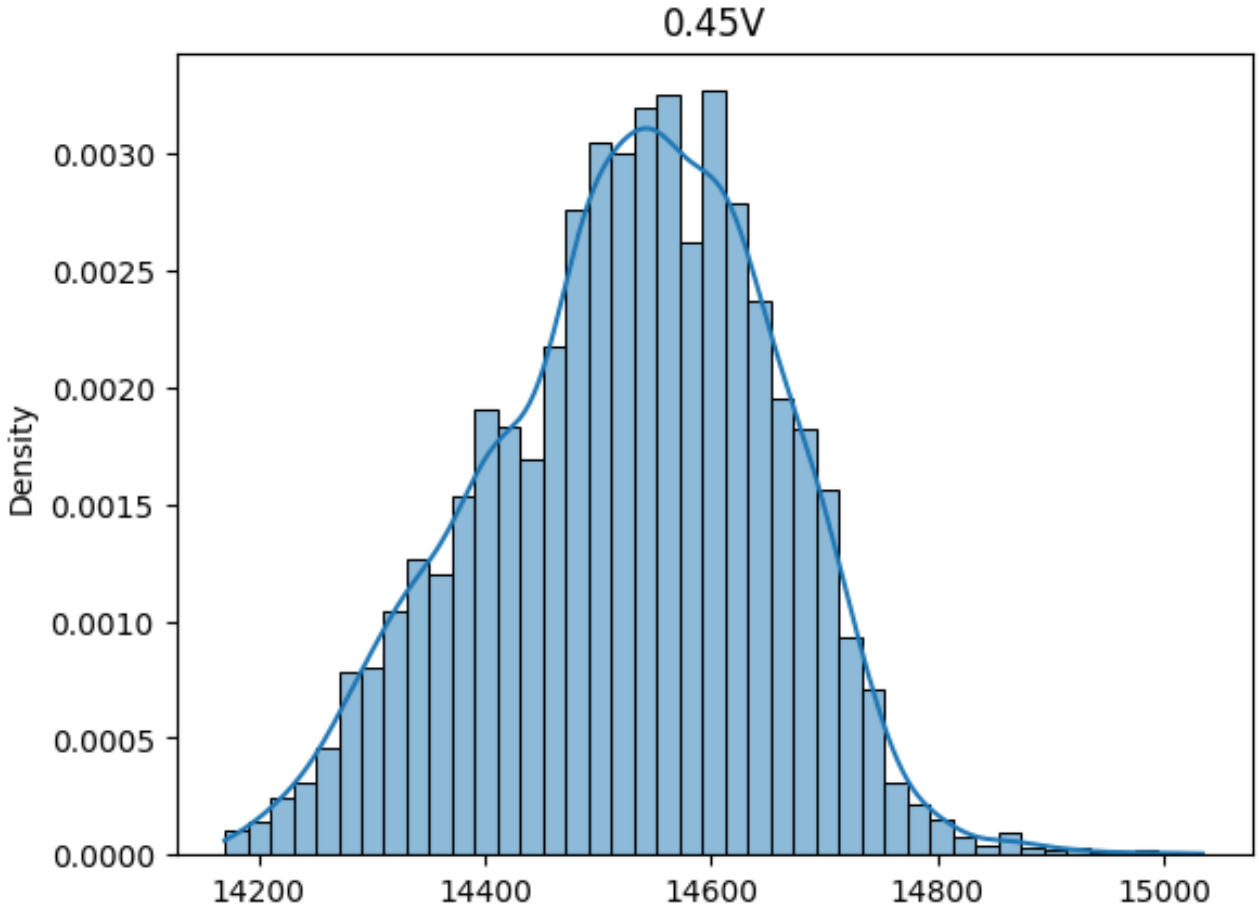


Рис. 21: Гистограмма распределения значений выходного напряжения

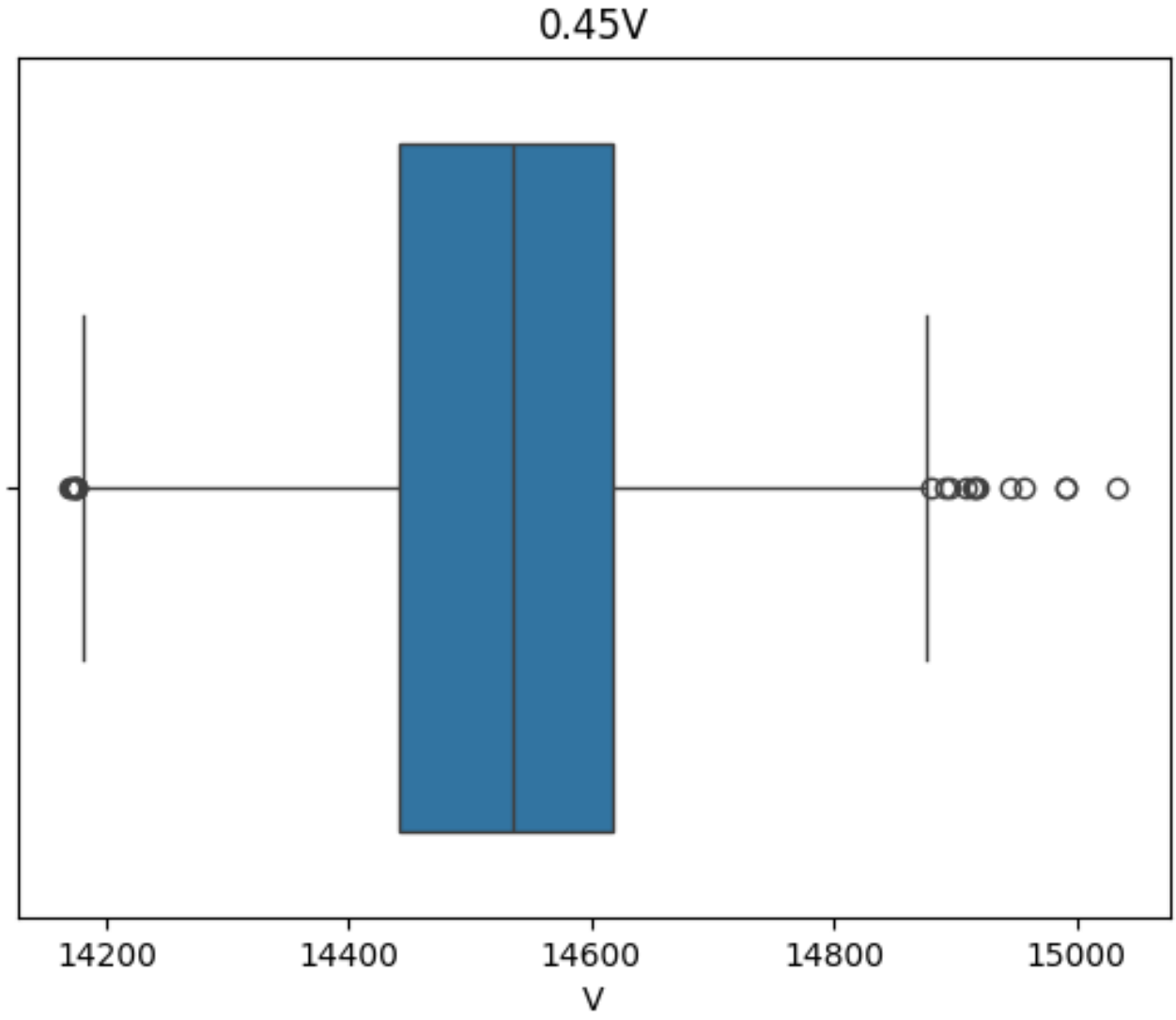


Рис. 22: Боксплот распределения значений выходного напряжения

| | | | |
|---|---|---|---|
| -0.05 | 190.0 | 4931.0 | 5691.0 |
| -0.45 | 190.0 | 4931.0 | 5691.0 |
| -0.35 | 190.0 | 4931.0 | 5691.0 |
| 0.35 | 190.0 | 4931.0 | 5691.0 |
| -0.25 | 190.0 | 4931.0 | 5691.0 |
| 0.05 | 190.0 | 4931.0 | 5691.0 |
| 0.15 | 190.0 | 4931.0 | 5691.0 |
| 0.45 | 190.0 | 4931.0 | 5691.0 |
| 0.0 | 190.0 | 4931.0 | 5691.0 |
| 0.25 | 190.0 | 4931.0 | 5691.0 |
| -0.15 | 190.0 | 4931.0 | 5691.0 |

Таблица 1: Боксплот Тьюки

## 4.12 Линейная регрессия до предобработки данных

## 4.13 Напряжение $U = 0.05V$

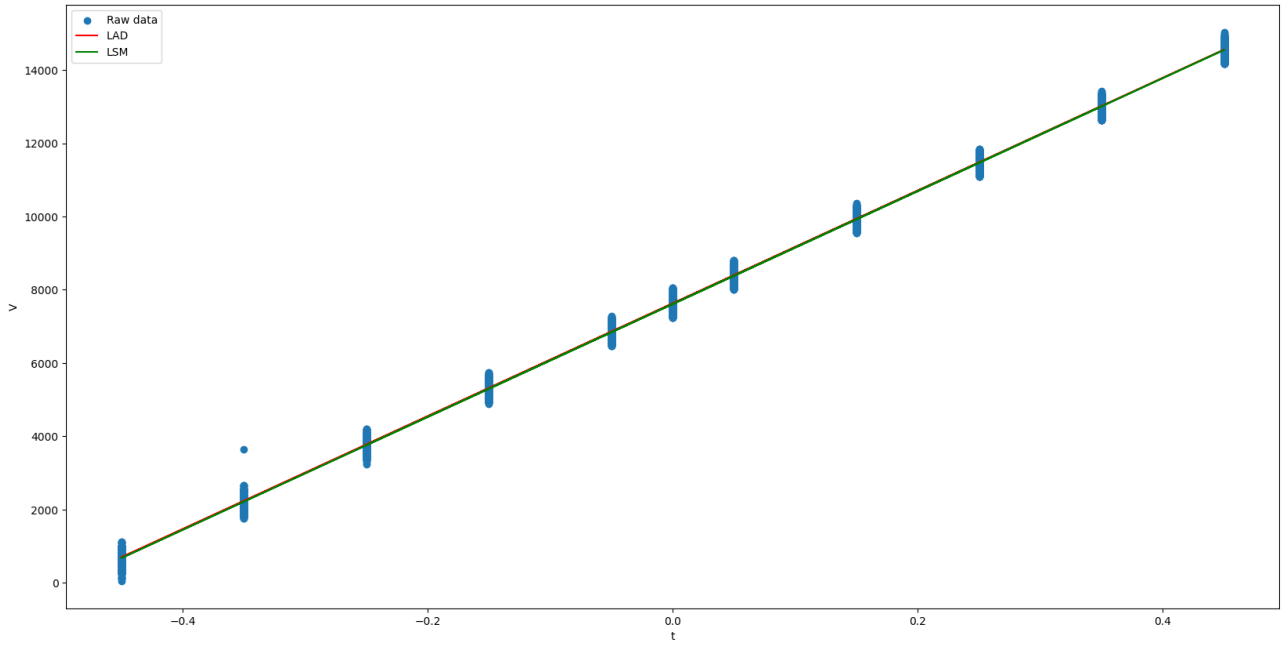


Рис. 23: Полученные прямые для необработанных данных

## 4.14 Линейная регрессия после предобработки данных

## 4.15 Напряжение $U = 0.05V$

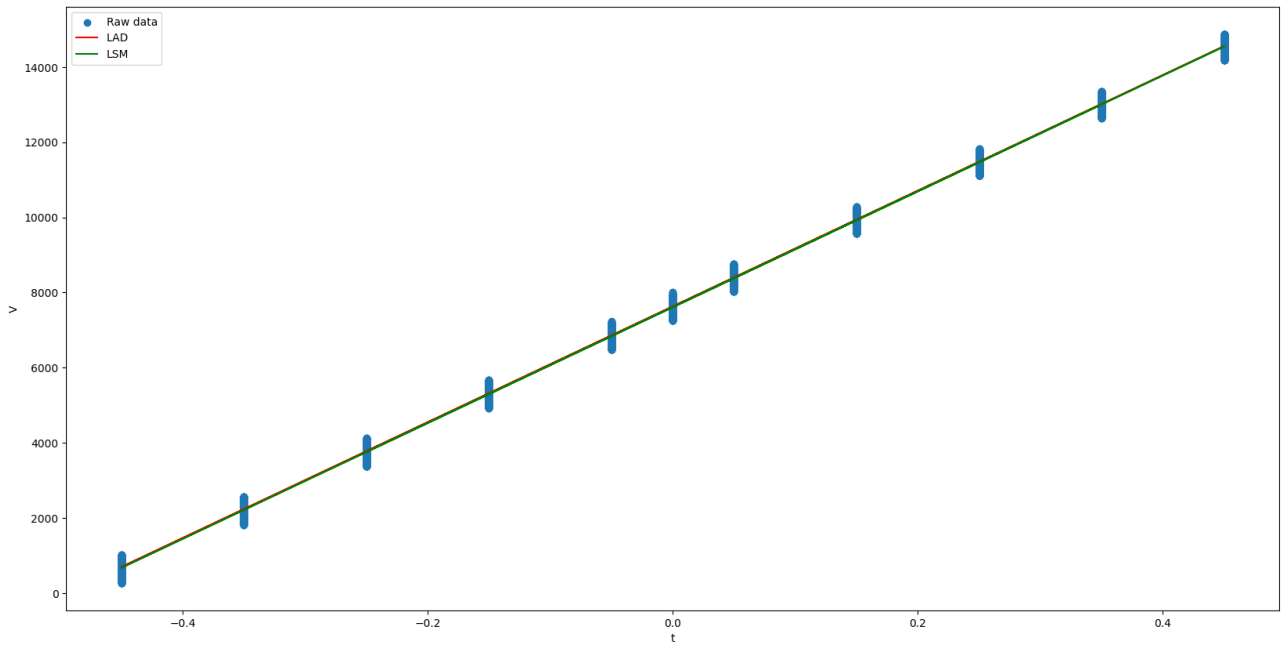


Рис. 24: Полученные прямые для обработанных данных

# 5 Выводы

1. Метод наименьших квадратов позволяет при помощи нетрудоемких вычислений получить коэффициенты линейной регрессии зависимых величин.

2. Метод наименьших модулей обеспечивает робастность оценок коэффициентов линейной регрессии.

3. При этом предобработка данных позволяет более точно оценить параметры линейной регрессии. Так, при использовании предобработки из исследуемых данных удаляются выбросы, которые понижают точность полученных без предобработки результатов.