

Санкт-Петербургский политехнический университет
Петра Великого

Физико-механический институт

Высшая школа прикладной математики и вычислительной
физики

Курсовая работа
«Оценки коэффициентов линейной регрессии»
по дисциплине
«Математическая статистика»

Выполнили студенты:
Афанасьев Андрей Валерьевич
Аптуков Михаил Ильдусович
Корпусова Софья Дмитриевна
Марков Михаил Денисович
группа:
5030102/10201
Проверил:
доцент
Баженов Александр Николаевич

Санкт-Петербург
2024

Содержание

1	Постановка задачи	4
2	Теоретическое обоснование	4
2.1	Метод наименьших квадратов	4
2.2	Метод наименьших модулей	5
2.3	Квартили	6
2.4	Бокс-плот Тьюки	7
2.5	Правило трёх сигм	7
3	Описание работы	8
4	Предварительная обработка данных	8
4.1	Очистка данных от выбросов по правилу трёх сигм	8
4.2	Очистка данных от выбросов с использованием межквартильного размаха	9
5	Результаты	14
5.1	Исследование зависимости выходного напряжения от входного напряжения	14
5.1.1	Метод наименьших модулей	15
5.1.2	Метод наименьших квадратов	16
5.1.3	Сравнение результатов	17
5.2	Исследование зависимости выходного напряжения от времени	17
5.2.1	Сравнение результатов	21
6	Выводы	21

Список иллюстраций

1	Гистограмма распределения неотфильтрованных нормализованных данных	8
2	Гистограмма распределения отфильтрованных нормализованных данных . .	9
3	Распределение $U_{\text{вых}}$ для $U_{\text{вх}} = -0.45V$	12
4	График боксплотов $U_{\text{вых}}$ для различных $U_{\text{вх}}$ до фильтрации данных	13
5	Зависимость медианы $U_{\text{вых}}$ для различных $U_{\text{вх}}$ после фильтрации данных . .	14
6	Зависимость медианного значения $U_{\text{вых}}$ от $U_{\text{вх}}$ с линейной регрессией методом наименьших модулей без выбросов	15
7	Зависимость медианного значения $U_{\text{вых}}$ от $U_{\text{вх}}$ с линейной регрессией методом наименьших модулей с выбросами	15
8	Зависимость медианного значения $U_{\text{вых}}$ от $U_{\text{вх}}$ с линейной регрессией методом наименьших квадратов без выбросов	16
9	Зависимость медианного значения $U_{\text{вых}}$ от $U_{\text{вх}}$ с линейной регрессией методом наименьших квадратов с выбросами	16
10	Зависимость $U_{\text{вых}}$ от времени с линейной регрессией при $U_{\text{вх}} = -0.45V$. . .	17
11	Зависимость $U_{\text{вых}}$ от времени с линейной регрессией при $U_{\text{вх}} = -0.35V$. . .	17
12	Зависимость $U_{\text{вых}}$ от времени с линейной регрессией при $U_{\text{вх}} = -0.25V$. . .	18
13	Зависимость $U_{\text{вых}}$ от времени с линейной регрессией при $U_{\text{вх}} = -0.15V$. . .	18
14	Зависимость $U_{\text{вых}}$ от времени с линейной регрессией при $U_{\text{вх}} = -0.05V$. . .	18
15	Зависимость $U_{\text{вых}}$ от времени с линейной регрессией при $U_{\text{вх}} = 0.00V$	18
16	Зависимость $U_{\text{вых}}$ от времени с линейной регрессией при $U_{\text{вх}} = 0.05V$	19
17	Зависимость $U_{\text{вых}}$ от времени с линейной регрессией при $U_{\text{вх}} = 0.15V$	19
18	Зависимость $U_{\text{вых}}$ от времени с линейной регрессией при $U_{\text{вх}} = 0.25V$	19
19	Зависимость $U_{\text{вых}}$ от времени с линейной регрессией при $U_{\text{вх}} = 0.35V$	19
20	Зависимость $U_{\text{вых}}$ от времени с линейной регрессией при $U_{\text{вх}} = 0.45V$	20

Список таблиц

1	$U_{\text{BX}} = -0.45V$	10
2	$U_{\text{BX}} = -0.35V$	10
3	$U_{\text{BX}} = -0.25V$	10
4	$U_{\text{BX}} = -0.15V$	10
5	$U_{\text{BX}} = -0.05V$	10
6	$U_{\text{BX}} = 0.00V$	11
7	$U_{\text{BX}} = 0.05V$	11
8	$U_{\text{BX}} = 0.15V$	11
9	$U_{\text{BX}} = 0.25V$	11
10	$U_{\text{BX}} = 0.35V$	11
11	$U_{\text{BX}} = 0.45V$	12
12	Оценка коэффициентов линейной регрессии	17
13	Оценка коэффициентов линейной регрессии для зависимости $U_{\text{вых}}$ от времени	21

1 Постановка задачи

Требуется выполнить оценку параметров линейной регрессии с использованием различных методов:

1. Метод наименьших квадратов
2. Метод наименьших модулей

Провести исследования для 5 столбца данных (5 вариант).

Исследуемые данные: <https://disk.yandex.ru/d/OAQgCulS6NfbOQ>

2 Теоретическое обоснование

2.1 Метод наименьших квадратов

При оценивании параметров регрессионной модели используют различные методы. Один из наиболее распространённых 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):

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

Задача минимизации квадратичного критерия (1) носит название задачи метода наименьших квадратов (МНК), а оценки β_0, β_1 параметров β_0, β_1 , реализующие минимум критерия (1), называют МНК-оценками.

Расчётные формулы для МНК-оценок МНК-оценки параметров $\hat{\beta}_0, \hat{\beta}_1$ находятся из условия обращения функции $Q(\beta_0, \beta_1)$ в минимум.

Для нахождения МНК-оценок $\hat{\beta}_0, \hat{\beta}_1$ выпишем необходимые условия экстремума

$$\begin{cases} \frac{\partial Q}{\partial \beta_0} = -2 \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i) = 0 \\ \frac{\partial Q}{\partial \beta_1} = -2 \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i) x_i = 0 \end{cases} \quad (2)$$

Далее для упрощения записи сумм будем опускать индекс суммирования. Из системы (2) получим:

$$\begin{cases} n\hat{\beta}_0 + \hat{\beta}_1 \sum x_i = \sum y_i \\ \hat{\beta}_0 \sum x_i + \hat{\beta}_1 \sum x_i^2 = \sum x_i y_i \end{cases} \quad (3)$$

Разделим оба уравнения на n :

$$\begin{cases} \hat{\beta}_0 + \hat{\beta}_1 \left(\frac{1}{n} \sum x_i\right) = \frac{1}{n} \sum y_i \\ \hat{\beta}_0 \left(\frac{1}{n} \sum x_i\right) + \hat{\beta}_1 \left(\frac{1}{n} \sum x_i^2\right) = \frac{1}{n} \sum x_i y_i \end{cases} \quad (4)$$

и, используя известные статистические обозначения для выборочных первых и вторых начальных моментов

$$\bar{x} = \frac{1}{n} \sum x_i, \bar{y} = \frac{1}{n} \sum y_i, \bar{x}^2 = \frac{1}{n} \sum x_i^2, \bar{xy} = \frac{1}{n} \sum x_i y_i, \quad (5)$$

получим

$$\begin{cases} \hat{\beta}_0 + \hat{\beta}_1 \bar{x} = \bar{y} \\ \hat{\beta}_0 \bar{x} + \hat{\beta}_1 \bar{x}^2 = \bar{x} \bar{y}, \end{cases} \quad (6)$$

откуда МНК-оценку $\hat{\beta}_1$ наклона прямой регрессии находим по формуле Крамера

$$\hat{\beta}_1 = \frac{\bar{x} \bar{y} - \bar{x} \cdot \bar{y}}{\bar{x}^2 - (\bar{x})^2} \quad (7)$$

а МНК-оценку $\hat{\beta}_0$ определяем непосредственно из первого уравнения системы (6):

$$\hat{\beta}_0 = \bar{y} - \bar{x} \hat{\beta}_1 \quad (8)$$

Заметим, что определитель системы (6):

$$\bar{x}^2 - (\bar{x})^2 = \frac{1}{n} \sum (x_i - \bar{x})^2 = s_x^2 > 0, \quad (9)$$

если среди значений x_1, \dots, x_n есть различные, что и будем предполагать.

Доказательство минимальности функции $Q(\beta_0, \beta_1)$ в стационарной точке проведём с помощью известного достаточного признака экстремума функции двух переменных. Имеем:

$$\frac{\partial^2 Q}{\partial \beta_0^2} = 2n, \quad \frac{\partial^2 Q}{\partial \beta_1^2} = 2 \sum x_i^2 = 2n \bar{x}^2, \quad \frac{\partial^2 Q}{\partial \beta_1 \partial \beta_0} = 2 \sum x_i = 2n \bar{x} \quad (10)$$

$$\Delta = \frac{\partial^2 Q}{\partial \beta_0^2} \cdot \frac{\partial^2 Q}{\partial \beta_1^2} - \left(\frac{\partial^2 Q}{\partial \beta_1 \partial \beta_0} \right)^2 = 4n^2 \bar{x}^2 - 4n^2 (\bar{x})^2 = 4n^2 [\bar{x}^2 - (\bar{x})^2] = 4n^2 \left[\frac{1}{n} \sum (x_i - \bar{x})^2 \right] = 4n^2 s_x^2 > 0. \quad (11)$$

Этот результат вместе с условием $\frac{\partial^2 Q}{\partial \beta_0^2} = 2n > 0$ означает, что в стационарной точке функция Q имеет минимум.

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:

$$\sum_{i=1}^n |y_i - \beta_0 - \beta_1 x_i| \rightarrow \min_{\beta_0, \beta_1}. \quad (12)$$

Напомним, что использование метода наименьших модулей в задаче оценивания параметра сдвига распределений приводит к оценке в виде выборочной медианы, обладающей робастными свойствами. В отличие от этого случая и от задач метода наименьших квадратов, на практике задача (12) решается численно. Соответствующие процедуры представлены в некоторых современных пакетах программ по статистическому анализу.

Здесь мы рассмотрим простейшую в вычислительном отношении робастную альтернативу оценкам коэффициентов линейной регрессии по МНК. Для этого сначала запишем выражения для оценок (8) и (7) в другом виде:

$$\begin{cases} \hat{\beta}_1 = \frac{\bar{x} \bar{y} - \bar{x} \cdot \bar{y}}{\bar{x}^2 - (\bar{x})^2} = \frac{k_{xy}}{s_x^2} = \frac{k_{xy}}{s_x s_y} \cdot \frac{s_y}{s_x} = r_{xy} \frac{s_y}{s_x} \\ \hat{\beta}_0 = \bar{y} - \bar{x} \hat{\beta}_1 \end{cases} \quad (13)$$

В формулах (13) заменим выборочные средние \bar{x} и \bar{y} соответственно на робастные выборочные медианы $medx$ и $medy$, среднеквадратические отклонения s_x и s_y на робастные нормированные интерквартильные широты q_x^* и q_y^* , выборочный коэффициент корреляции r_{xy} — на знаковый коэффициент корреляции r_Q :

$$\hat{\beta}_{1R} = r_Q \frac{q_y^*}{q_x^*}, \quad (14)$$

$$\hat{\beta}_{0R} = medy - \hat{\beta}_{1R} medx, \quad (15)$$

$$r_Q = \frac{1}{n} \sum_{i=1}^n \text{sgn}(x_i - medx) \text{sgn}(y_i - medy), \quad (16)$$

$$\begin{aligned} q_y^* &= \frac{y_{(j)} - y_{(l)}}{k_q(n)}, \quad q_x^* = \frac{x_{(j)} - x_{(l)}}{k_q(n)}, \\ &\begin{cases} \left[\frac{n}{4} \right] + 1 \text{ при } \frac{n}{4} \text{ дробном,} \\ \frac{n}{4} \text{ при } \frac{n}{4} \text{ целом.} \end{cases} \\ &j = n - l + 1 \\ \text{sgn}(z) &= \begin{cases} 1 \text{ при } z > 0 \\ 0 \text{ при } z = 0 \\ -1 \text{ при } z < 0 \end{cases} \end{aligned} \quad (17)$$

Уравнение регрессии здесь имеет вид

$$y = \hat{\beta}_{0R} + \hat{\beta}_{1R}x \quad (18)$$

Статистики выборочной медианы и интерквартильной широты обладают робастными свойствами в силу того, что основаны на центральных порядковых статистиках, малочувствительных к большим по величине выбросам в данных. Статистика выборочного знакового коэффициента корреляции робастна, так как знаковая функция $\text{sgn}(z)$ чувствительна не к величине аргумента, а только к его знаку. Отсюда оценка прямой регрессии (18) обладает очевидными робастными свойствами устойчивости к выбросам по координате y , но она довольно груба.

2.3 Квартили

Квартили — это значения, которые делят упорядоченный набор данных на четыре равные части. Они представляют собой три точки данных, которые делят данные на четыре примерно равные группы.

Для набора данных, упорядоченного по возрастанию,

- Первый квартиль (Q_1) — это значение, ниже которого находится 25% наблюдений
- Второй квартиль (Q_2) — это медиана, значение, которое делит набор данных пополам
- Третий квартиль (Q_3) — это значение, ниже которого находится 75% наблюдений

Пусть n — количество наблюдений в наборе данных. Тогда для расчета квартилей можно использовать следующие формулы:

$$Q_1 = X_{(\lceil 0.25 \cdot n \rceil)} \quad (19)$$

$$Q_2 = X_{(\lceil 0.5 \cdot n \rceil)} \quad (20)$$

$$Q_3 = X_{(\lceil 0.75 \cdot n \rceil)} \quad (21)$$

2.4 Бокс-плот Тьюки

Боксплот (англ. box plot) — график, использующихся в описательной статистике, компактно изображающий 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 межквартильных расстояний и сумма третьего квартиля и полутора межквартильных расстояний. Формула имеет вид

$$X_1 = Q_1 - \frac{3}{2}(Q_3 - Q_1), \quad X_2 = Q_3 + \frac{3}{2}(Q_3 - Q_1), \quad (22)$$

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:

$$\begin{cases} x < X_1^T \\ x > X_2^T \end{cases} \quad (23)$$

2.5 Правило трёх сигм

Правило трёх сигм, является статистическим правилом, которое описывает распределение вероятностей для нормально распределённой случайной величины. Оно гласит, что:

- Приблизительно 68.2% значений случайной величины лежат в пределах одного стандартного отклонения от среднего.
- Приблизительно 95.4% значений лежат в пределах двух стандартных отклонений от среднего.
- Приблизительно 99.7% значений лежат в пределах трёх стандартных отклонений от среднего.

Для нормально распределённой случайной величины X с математическим ожиданием μ и стандартным отклонением σ , вероятность того, что X попадает в интервал $(\mu - k\sigma, \mu + k\sigma)$, где k — некоторая константа, равна $1 - \alpha$, где α — уровень значимости.

Поскольку нормальное распределение симметрично относительно своего среднего значения, мы можем сказать, что:

$$P(\mu - \sigma < X < \mu + \sigma) = P(-\sigma < Z < \sigma) = P(Z < \sigma) - P(Z < -\sigma) \quad (24)$$

Z — стандартная нормальная случайная величина.

Поскольку функция плотности вероятности для стандартной нормальной случайной величины симметрична, $P(Z < -\sigma) = P(Z > \sigma)$. Таким образом,

$$P(\mu - \sigma < X < \mu + \sigma) = 2P(Z < \sigma) - 1. \quad (25)$$

Из таблицы стандартного нормального распределения известно, что $P(Z < \sigma) \approx 0.8413$, следовательно,

$$P(\mu - \sigma < X < \mu + \sigma) \approx 2 \times 0.8413 - 1 \approx 0.6826. \quad (26)$$

Аналогично, для интервала $(\mu - 2\sigma, \mu + 2\sigma)$ и $(\mu - 3\sigma, \mu + 3\sigma)$ вероятности равны 0.9544 и 0.9974 соответственно, что подтверждает правило трёх сигм.

3 Описание работы

Курсовая работа выполнена на языке программирования Python в среде разработки PyCharm. В ходе работы были использованы следующие библиотеки: `numpy`, `matplotlib`, `math`, `scipy`.

GitHub репозиторий: <https://github.com/afafos/Statistics/tree/main/coursework>

4 Предварительная обработка данных

4.1 Очистка данных от выбросов по правилу трёх сигм

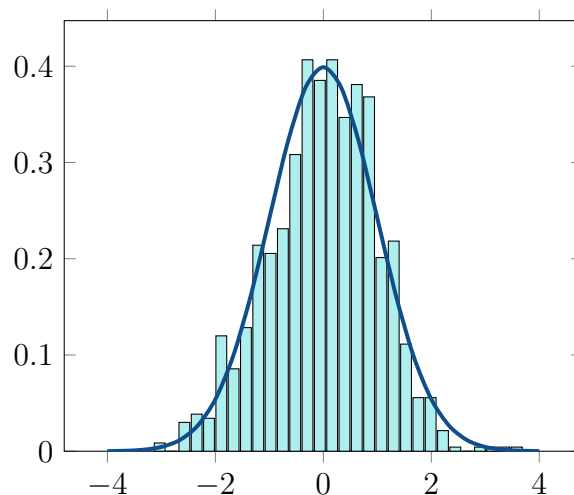


Рис. 1: Гистограмма распределения неотфильтрованных нормализованных данных

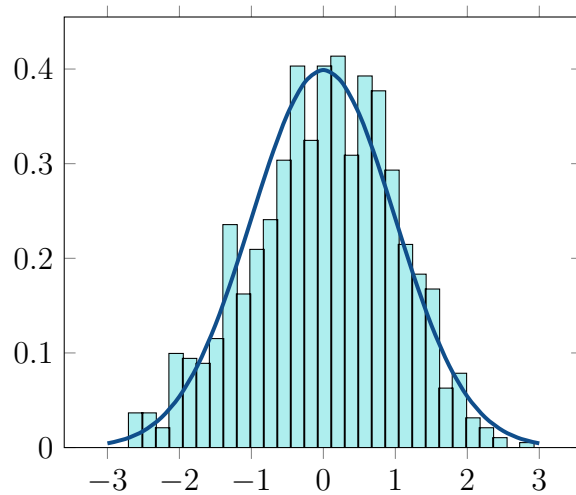


Рис. 2: Гистограмма распределения отфильтрованных нормализованных данных

Сначала для каждого столбика данных, содержащего выходные напряжения, рассчитывались среднее значение и стандартное отклонение. Затем нижняя и верхняя границы для фильтрации определялись как среднее значение \pm три стандартных отклонения σ . Далее данные фильтровались: оставлялись только те значения, которые попадали в этот диапазон.

4.2 Очистка данных от выбросов с использованием межквартильного размаха

Рассчитали квартили для столбца данных, содержащего выходное напряжение $U_{\text{вых}}$. Квартили делят данные на четыре равные части, где первый квартиль (25-й процентиль) определяет значение, ниже которого лежит 25% данных, а третий квартиль (75-й процентиль) определяет значение, ниже которого лежит 75% данных.

$$IQR = Q_3 - Q_1 \quad (27)$$

IQR - межквартильный размах. Он используется для измерения разброса в данных, содержащих выбросы.

После расчета квартилей и межквартильного размаха определим нижнюю и верхнюю границы для фильтрации выбросов. Нижняя граница интервала исследования определяется следующей формулой:

$$\text{Нижняя граница} = Q_1 - 1.5 \cdot IQR \quad (28)$$

Аналогично, верхняя граница интервала исследования определяется следующей формулой:

$$\text{Верхняя граница} = Q_3 + 1.5 \cdot IQR \quad (29)$$

Выбросы определяются как значения, которые находятся за пределами вычисленных границ.

Источник данных	IQR (27)	Нижняя граница (28)	Верхняя граница (29)
$sp = 115$	183.00	328.500	1060.500
$sp = 31$	185.50	333.500	1075.500
$sp = 442$	186.50	327.000	1073.000
$sp = 832$	190.25	317.375	1078.375
$sp = 913$	183.50	337.500	1071.500

Таблица 1: $U_{\text{вх}} = -0.45V$

Источник данных	IQR (27)	Нижняя граница (28)	Верхняя граница (29)
$sp = 196$	187.25	1880.125	2629.125
$sp = 443$	191.25	1869.875	2634.875
$sp = 567$	193.50	1866.500	2640.500
$sp = 670$	189.00	1880.500	2636.500
$sp = 852$	191.25	1875.125	2640.125

Таблица 2: $U_{\text{вх}} = -0.35V$

Источник данных	IQR (27)	Нижняя граница (28)	Верхняя граница (29)
$sp = 403$	186.50	3440.750	4186.750
$sp = 423$	185.25	3439.875	4180.875
$sp = 484$	195.00	3419.500	4199.500
$sp = 731$	191.00	3428.500	4192.500
$sp = 94$	184.00	3449.000	4185.000

Таблица 3: $U_{\text{вх}} = -0.25V$

Источник данных	IQR (27)	Нижняя граница (28)	Верхняя граница (29)
$sp = 155$	182.25	5003.375	5732.375
$sp = 649$	183.25	5003.125	5736.125
$sp = 831$	188.50	4985.000	5739.000
$sp = 892$	185.25	4993.125	5734.125
$sp = 993$	182.25	5002.625	5731.625

Таблица 4: $U_{\text{вх}} = -0.15V$

Источник данных	IQR (27)	Нижняя граница (28)	Верхняя граница (29)
$sp = 465$	189.00	6535.500	7291.500
$sp = 505$	182.00	6551.000	7279.000
$sp = 547$	190.00	6541.000	7301.000
$sp = 548$	181.00	6555.500	7279.500
$sp = 913$	181.00	6556.500	7280.500

Таблица 5: $U_{\text{вх}} = -0.05V$

Источник данных	IQR (27)	Нижняя граница (28)	Верхняя граница (29)
$sp = 443$	181.00	7326.500	8050.500
$sp = 812$	180.00	7329.000	8049.000
$sp = 813$	178.50	7334.750	8048.750
$sp = 892$	174.25	7339.375	8036.375
$sp = 953$	177.75	7330.375	8041.375

Таблица 6: $U_{\text{вх}} = 0.00V$

Источник данных	IQR (27)	Нижняя граница (28)	Верхняя граница (29)
$sp = 321$	178.25	8098.625	8811.625
$sp = 526$	182.00	8096.000	8824.000
$sp = 72$	179.25	8099.875	8816.875
$sp = 953$	174.25	8106.375	8803.375
$sp = 974$	179.25	8097.875	8814.875

Таблица 7: $U_{\text{вх}} = 0.05V$

Источник данных	IQR (27)	Нижняя граница (28)	Верхняя граница (29)
$sp = 135$	178.25	9634.375	10347.375
$sp = 298$	177.25	9638.875	10347.875
$sp = 465$	179.25	9639.125	10356.125
$sp = 9$	181.00	9633.500	10357.500
$sp = 93$	179.25	9635.125	10352.125

Таблица 8: $U_{\text{вх}} = 0.15V$

Источник данных	IQR (27)	Нижняя граница (28)	Верхняя граница (29)
$sp = 135$	174.25	11180.375	11877.375
$sp = 320$	178.50	11174.000	11888.000
$sp = 484$	180.25	11168.375	11889.375
$sp = 711$	176.75	11175.375	11882.375
$sp = 72$	170.25	11185.375	11866.375

Таблица 9: $U_{\text{вх}} = 0.25V$

Источник данных	IQR (27)	Нижняя граница (28)	Верхняя граница (29)
$sp = 300$	174.25	12717.375	13414.375
$sp = 361$	175.00	12712.500	13412.500
$sp = 608$	173.25	12722.875	13415.875
$sp = 670$	171.25	12723.125	13408.125
$sp = 933$	171.25	12720.125	13405.125

Таблица 10: $U_{\text{вх}} = 0.35V$

Источник данных	IQR (27)	Нижняя граница (28)	Верхняя граница (29)
$sp = 1014$	164.25	14270.375	14927.375
$sp = 176$	165.50	14266.250	14928.250
$sp = 361$	170.00	14261.000	14941.000
$sp = 832$	168.50	14259.000	14933.000
$sp = 873$	168.25	14259.375	14932.375

Таблица 11: $U_{\text{вх}} = 0.45V$

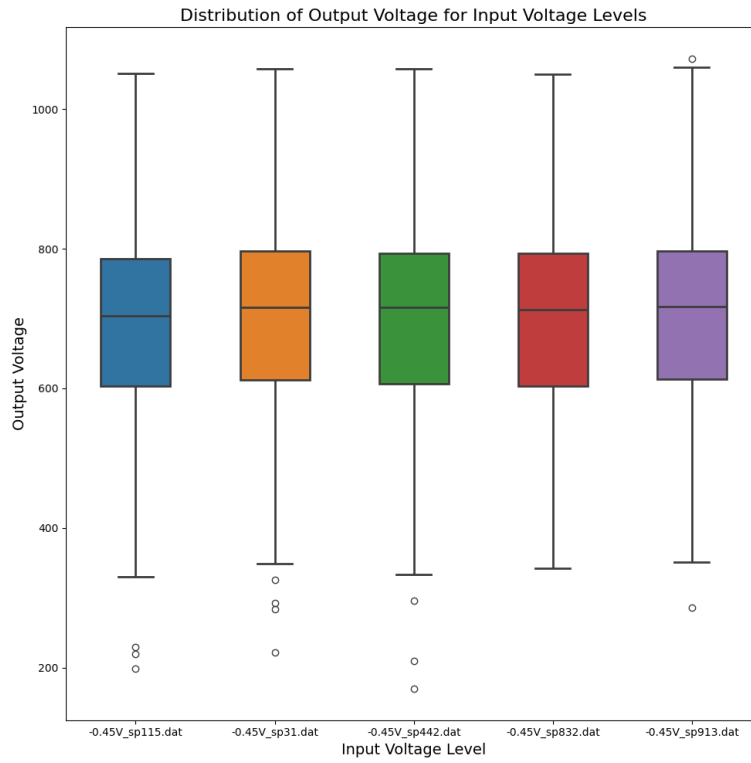


Рис. 3: Распределение $U_{\text{вых}}$ для $U_{\text{вх}} = -0.45V$

Стоит отметить, что на графике присутствует лишь небольшое количество выбросов за пределами 'усов', что свидетельствует о стабильности данных и отсутствии значительных аномалий в измерениях.

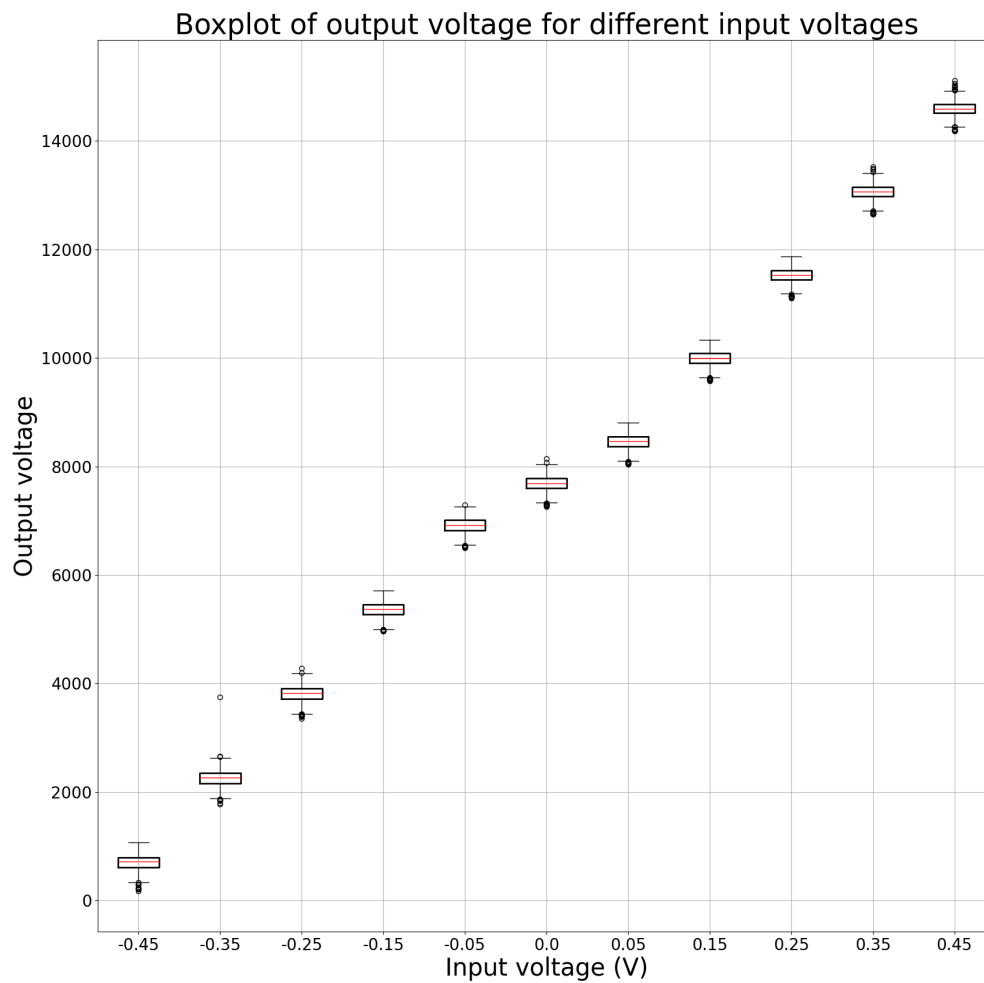


Рис. 4: График боксплотов $U_{\text{вых}}$ для различных $U_{\text{вх}}$ до фильтрации данных

График боксплотов иллюстрирует зависимость выходного напряжения от входного до прохождения через процесс фильтрации выбросов. Каждый боксплот соответствует определенному значению входного напряжения, отображенному на оси X. Вертикальная ось Y представляет выходное напряжение.

График позволяет оценить разброс значений выходного напряжения для различных входных значений. Внутри каждого боксплота горизонтальная линия обозначает медианное значение, а верхняя и нижняя границы прямоугольника представляют собой первый и третий квантили соответственно. Усы графика отображают диапазон данных за вычетом выбросов.

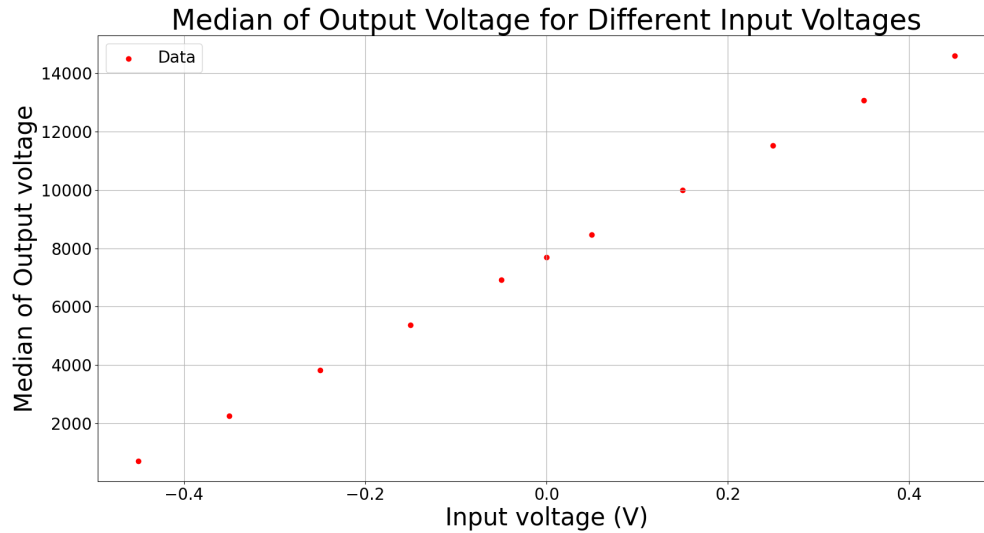


Рис. 5: Зависимость медианы $U_{\text{вых}}$ для различных $U_{\text{вх}}$ после фильтрации данных

График показывает зависимость медианы выходного напряжения от входного напряжения после фильтрации данных с использованием метода трёх сигм и удаления выбросов на основе межквартильного размаха.

Каждая точка на графике представляет собой медианное значение $U_{\text{вых}}$ для определённого $U_{\text{вх}}$. Точки распределены по входным напряжениям.

5 Результаты

5.1 Исследование зависимости выходного напряжения от входного напряжения

По заданию нужно было найти оценки коэффициентов линейной регрессии

$$y_i = \beta_0 + \beta_1 \cdot x_i \quad (30)$$

Ссылки на основные формулы для МНК: (8), (7).

Ссылка на основные формулы для МНМ: (13).

5.1.1 Метод наименьших модулей

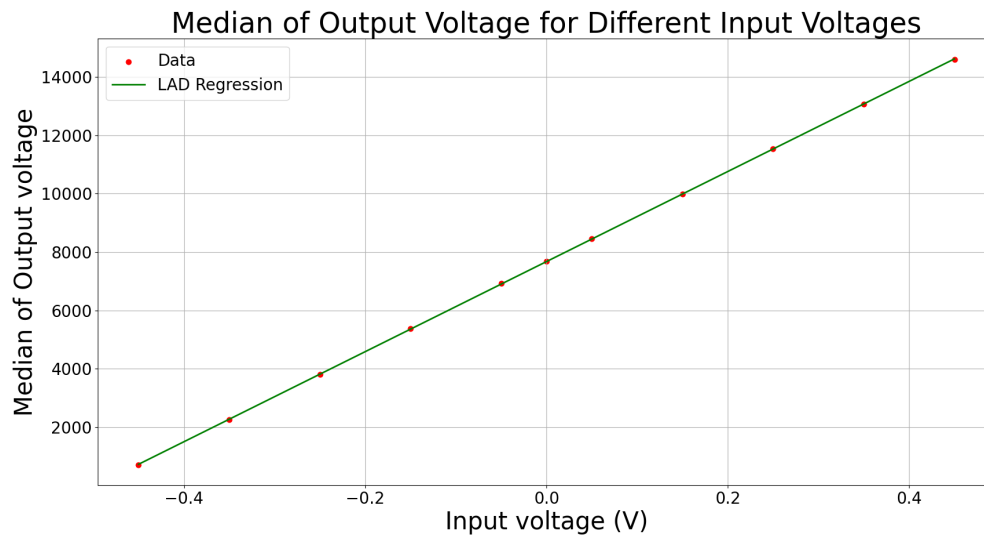


Рис. 6: Зависимость медианного значения $U_{\text{ВЫХ}}$ от $U_{\text{ВХ}}$ с линейной регрессией методом наименьших модулей без выбросов

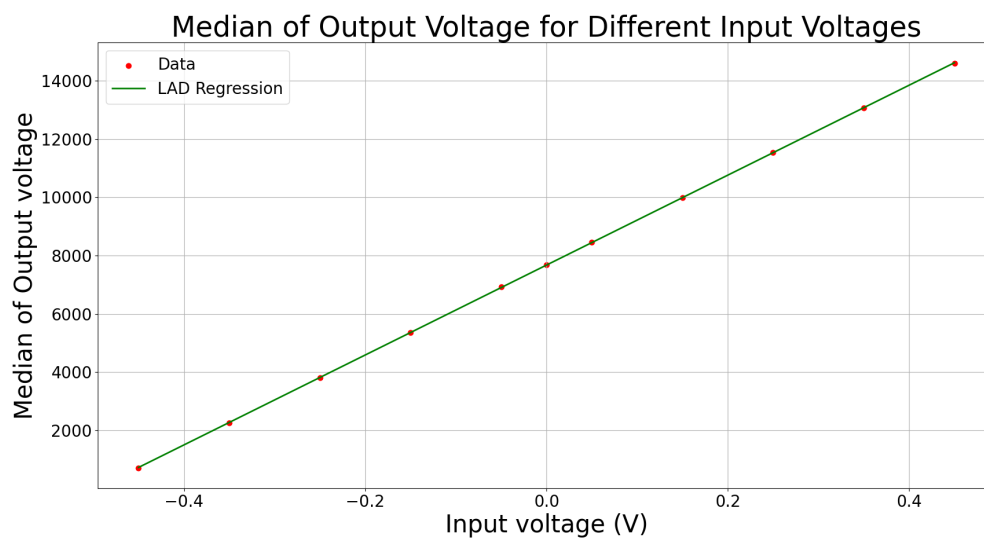


Рис. 7: Зависимость медианного значения $U_{\text{ВЫХ}}$ от $U_{\text{ВХ}}$ с линейной регрессией методом наименьших модулей с выбросами

5.1.2 Метод наименьших квадратов

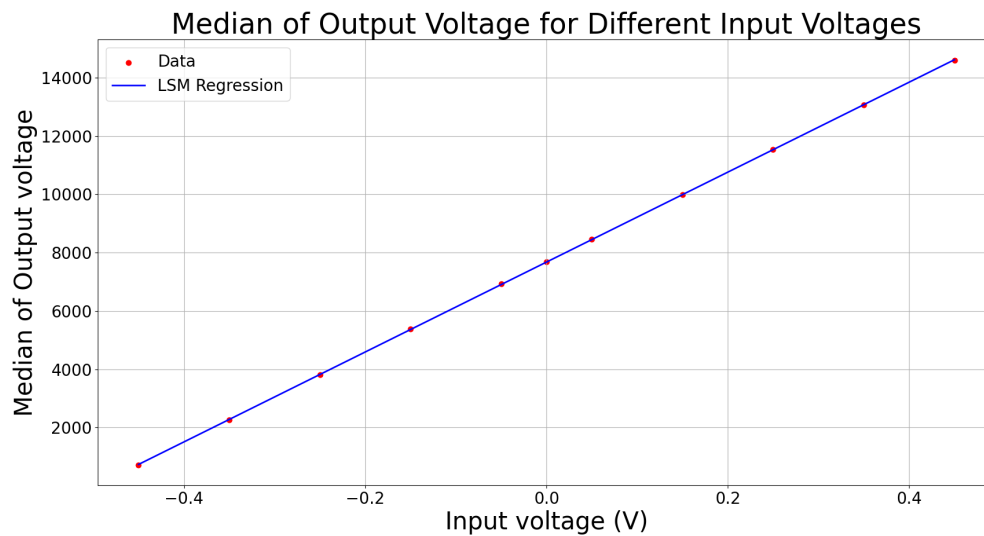


Рис. 8: Зависимость медианного значения $U_{\text{ВЫХ}}$ от $U_{\text{ВХ}}$ с линейной регрессией методом наименьших квадратов без выбросов

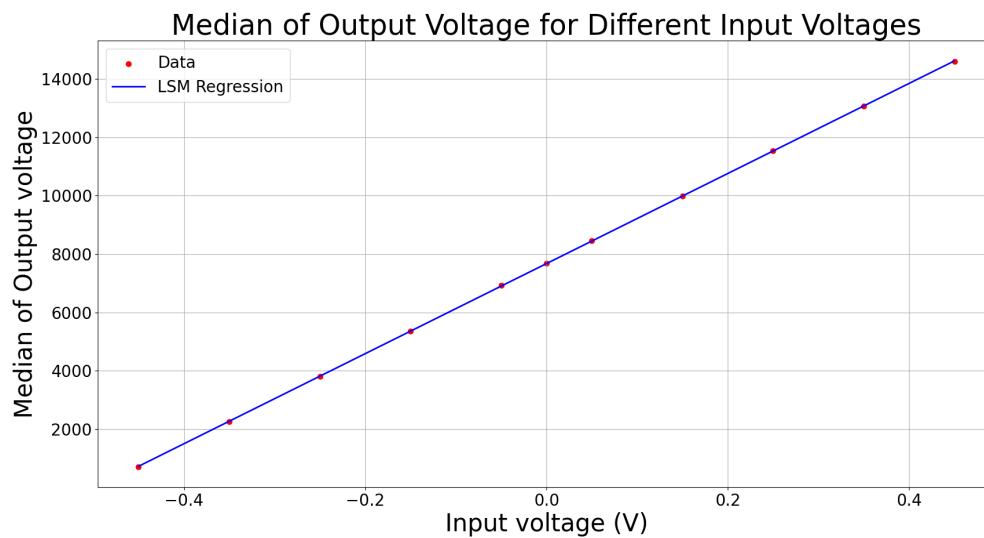


Рис. 9: Зависимость медианного значения $U_{\text{ВЫХ}}$ от $U_{\text{ВХ}}$ с линейной регрессией методом наименьших квадратов с выбросами

5.1.3 Сравнение результатов

$y = \beta_1 * x + \beta_0$ – уравнение линейной регрессии для методов МНК и МНМ без выбросов (31)

$y = \hat{\beta}_1 * x + \hat{\beta}_0$ – уравнение линейной регрессии для методов МНК и МНМ с выбросами (32)

$$\delta\beta_0 = \frac{|\beta_0 - \hat{\beta}_0|}{\beta_0} \quad (33)$$

$$\delta\beta_1 = \frac{|\beta_1 - \hat{\beta}_1|}{\beta_1} \quad (34)$$

МНМ	β_0	β_1	$\hat{\beta}_0$	$\hat{\beta}_1$	$\delta\beta_0$ (33)	$\delta\beta_1$ (34)
	7674.881	15424.047	7674.670	15423.919	$2.749 \cdot 10^{-5}$	$8.299 \cdot 10^{-6}$
МНК	β_0	β_1	$\hat{\beta}_0$	$\hat{\beta}_1$	$\delta\beta_0$ (33)	$\delta\beta_1$ (34)
	7674.864	15424.152	7674.636	15424.121	$2.971 \cdot 10^{-5}$	$2.010 \cdot 10^{-6}$

Таблица 12: Оценка коэффициентов линейной регрессии

5.2 Исследование зависимости выходного напряжения от времени

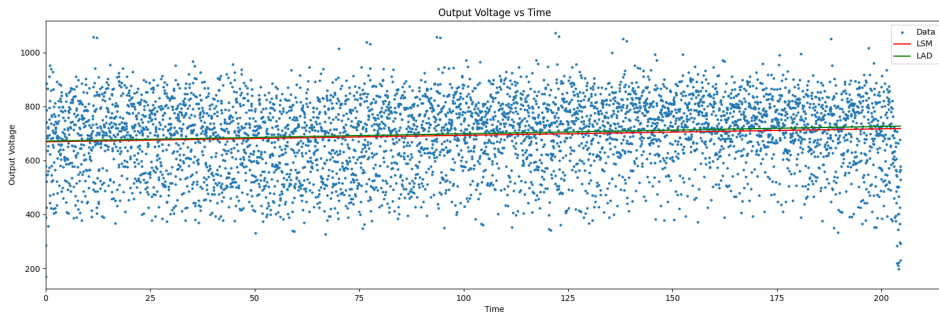


Рис. 10: Зависимость $U_{\text{вых}}$ от времени с линейной регрессией при $U_{\text{вх}} = -0.45V$

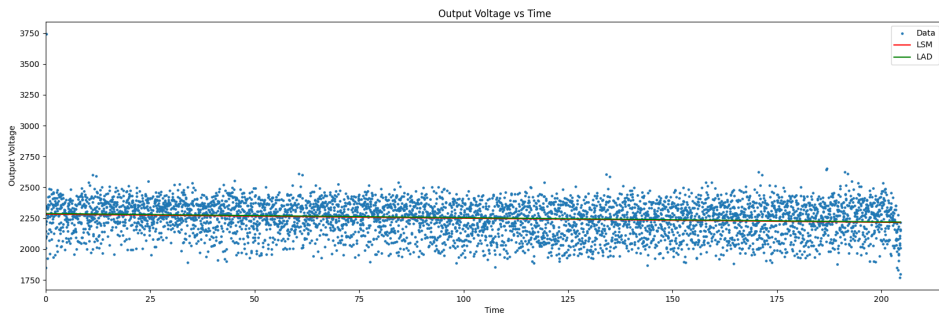


Рис. 11: Зависимость $U_{\text{вых}}$ от времени с линейной регрессией при $U_{\text{вх}} = -0.35V$

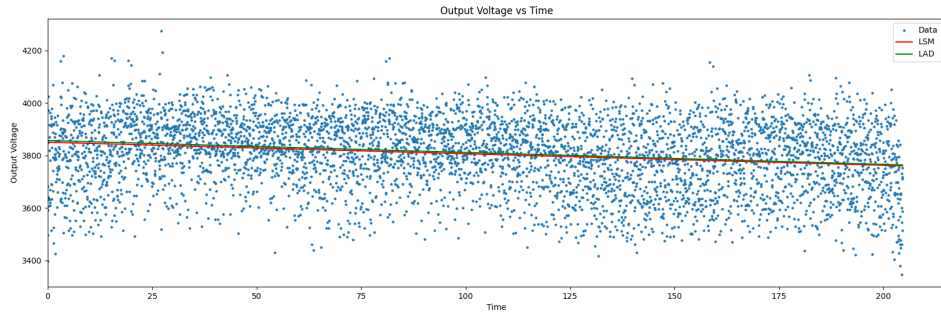


Рис. 12: Зависимость $U_{\text{ВЫХ}}$ от времени с линейной регрессией при $U_{\text{ВХ}} = -0.25V$

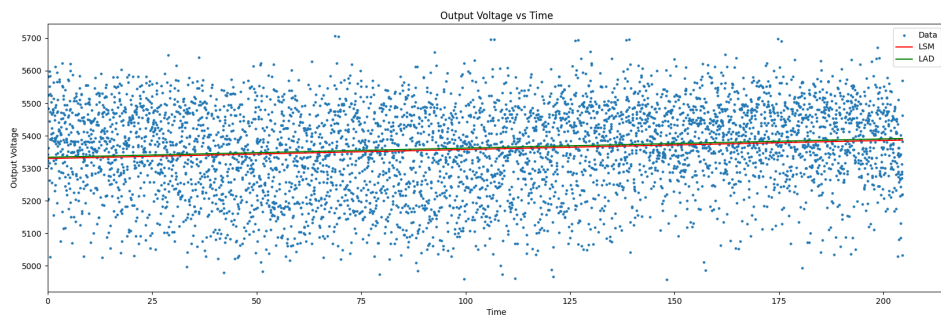


Рис. 13: Зависимость $U_{\text{ВЫХ}}$ от времени с линейной регрессией при $U_{\text{ВХ}} = -0.15V$

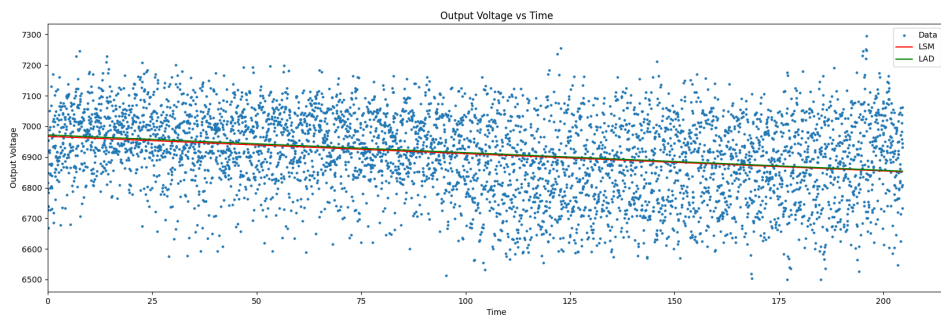


Рис. 14: Зависимость $U_{\text{ВЫХ}}$ от времени с линейной регрессией при $U_{\text{ВХ}} = -0.05V$

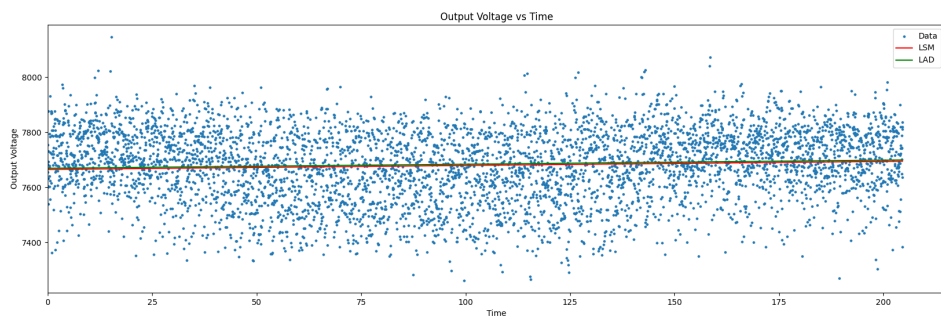


Рис. 15: Зависимость $U_{\text{ВЫХ}}$ от времени с линейной регрессией при $U_{\text{ВХ}} = 0.00V$

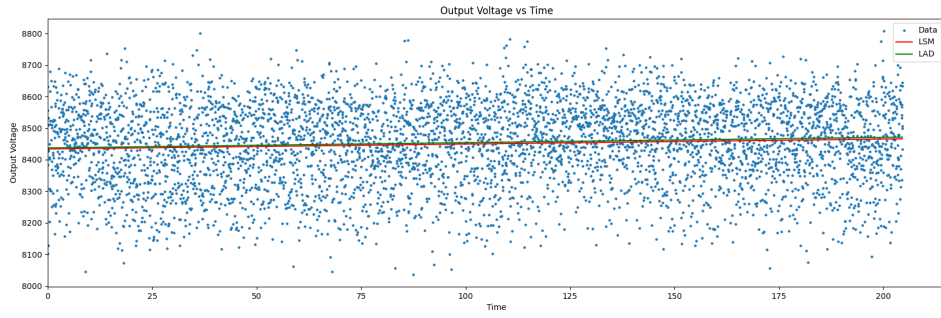


Рис. 16: Зависимость $U_{\text{ВЫХ}}$ от времени с линейной регрессией при $U_{\text{ВХ}} = 0.05V$

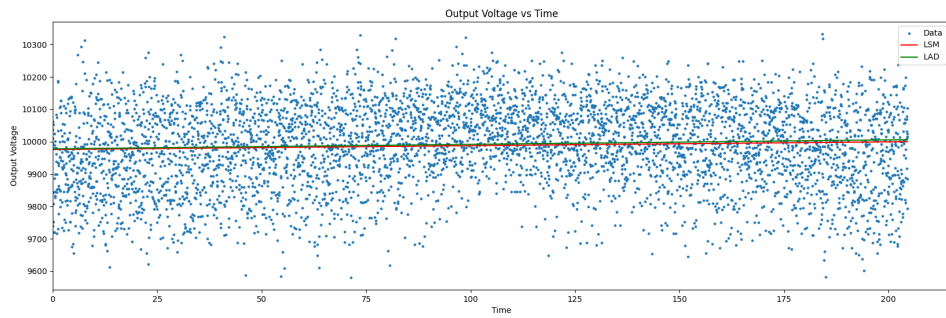


Рис. 17: Зависимость $U_{\text{ВЫХ}}$ от времени с линейной регрессией при $U_{\text{ВХ}} = 0.15V$

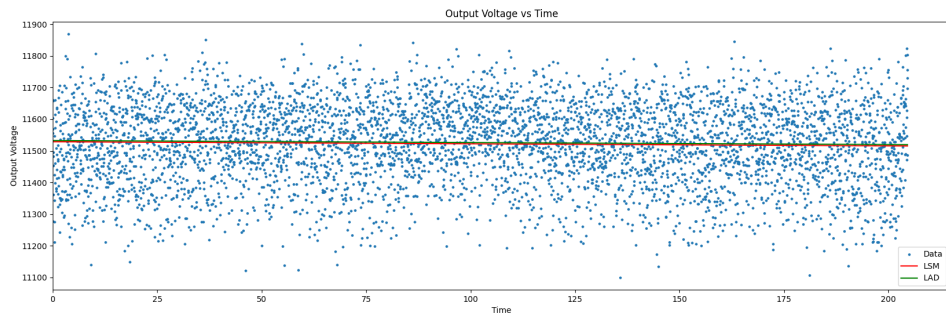


Рис. 18: Зависимость $U_{\text{ВЫХ}}$ от времени с линейной регрессией при $U_{\text{ВХ}} = 0.25V$

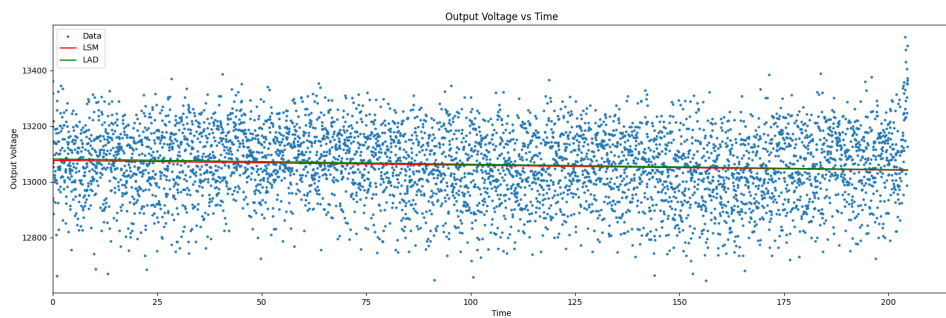


Рис. 19: Зависимость $U_{\text{ВЫХ}}$ от времени с линейной регрессией при $U_{\text{ВХ}} = 0.35V$

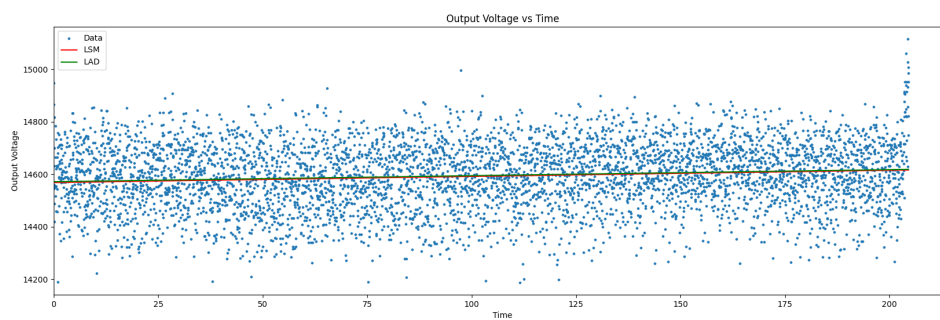


Рис. 20: Зависимость $U_{\text{вых}}$ от времени с линейной регрессией при $U_{\text{вх}} = 0.45V$

5.2.1 Сравнение результатов

$U_{\text{вх}}$	МНК		МНМ	
	β_0	β_1	β_0	β_1
-0.45V	668.912	0.240	671.554	0.271
-0.35V	2282.348	-0.329	2287.899	-0.338
-0.25V	3850.588	-0.440	3857.112	-0.455
-0.15V	5329.981	0.278	5333.946	0.282
-0.05V	6967.389	-0.565	6971.907	-0.577
0.00V	7665.136	0.146	7669.423	0.147
0.05V	8433.873	0.157	8437.205	0.167
0.15V	9975.730	0.118	9977.765	0.139
0.25V	11528.721	-0.065	11532.507	-0.063
0.35V	13077.159	-0.173	13082.194	-0.191
0.45V	14567.959	0.234	14571.145	0.230

Таблица 13: Оценка коэффициентов линейной регрессии для зависимости $U_{\text{вых}}$ от времени

6 Выводы

1. В процессе исследования было обнаружено, что данные имели незначительное количество выбросов. Это говорит о достаточной чистоте данных и их надёжности для проведения статистического анализа. Также о чистоте данных свидетельствуют малые значения изменения параметров линейной регрессии при удалении выбросов.
2. Оба метода, как метод наименьших модулей, так и метод наименьших квадратов, показали схожие результаты оценки коэффициентов регрессии.
3. Для анализа зависимости выходного напряжения от входного напряжения используются методы МНК и МНМ. Результаты показывают, что оценки коэффициентов линейной регрессии отличаются между двумя методами, но остаются близкими.
4. Анализ зависимости выходного напряжения от времени при различных уровнях входного напряжения позволяет увидеть динамику изменения выходного напряжения во времени, причем эта зависимость может быть описана линейной моделью. Для каждого уровня входного напряжения также дана оценка коэффициентов линейной регрессии с использованием методов МНК и МНМ.