

ПРАВИТЕЛЬСТВО РОССИЙСКОЙ ФЕДЕРАЦИИ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»

Факультет компьютерных наук
Департамент программной инженерии

СОГЛАСОВАНО

Доцент департамента математики
факультета экономических наук, кандидат
физико-математических наук

_____ Е. Р. Горяинова
« ____ » _____ 2025 г.

УТВЕРЖДАЮ

Академический руководитель
образовательной программы
«Программная инженерия»
профессор департамента программной
инженерии, канд. техн. наук

_____ Н. А. Павлочев
« ____ » _____ 2025 г.

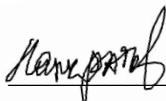
Подп.идата	
Инв.№дубл.	
Взам.инв.№	
Подп.идата	
Инв.№подл	

Разработка программного комплекса для исследования
влияния аномальных наблюдений на точность
прогнозирования в регрессионных моделях

Руководство оператора

ЛИСТ УТВЕРЖДЕНИЯ

RU.17701729.11.04-01 РО 01-1

Исполнители:
Студент БПИ-221
 / Панкратов С. Ю. /
«12» мая 2025 г

УТВЕРЖДЕН

RU.17701729.11.04-01 PO 01-1

**Разработка программного комплекса для исследования
влияния аномальных наблюдений на точность
прогнозирования в регрессионных моделях**

**Руководство программиста
RU.17701729.11.04-01 РП 01-1**

Листов 15

<i>Подп.идата</i>	
<i>Инв.№дубл.</i>	
<i>Взам.инв.№</i>	
<i>Подп.идата</i>	
<i>Инв.№подл</i>	

АННОТАЦИЯ

Настоящий документ «Программный комплекс для исследования влияния аномальных наблюдений на точность прогнозирования в регрессионных моделях. Руководство оператора» предназначен для ознакомления пользователей с графическим интерфейсом и функциональными возможностями программного комплекса "MSnOutliers".

Руководство оператора разработано в соответствии с ГОСТ 19.505-79 «Единая система программной документации. Руководство оператора» и содержит следующие разделы:

В разделе «Назначение программы» приведена информация о функциональном назначении программного комплекса, его возможностях и особенностях использования в области исследования влияния аномальных наблюдений на регрессионные модели.

В разделе «Условия выполнения программы» представлены требования к аппаратному и программному обеспечению, необходимому для корректной работы комплекса.

Раздел «Выполнение программы» содержит последовательное описание действий оператора при работе с программным комплексом, включая создание моделей, генерацию данных, запуск анализа и интерпретацию результатов.

В разделе «Сообщения оператору» приведены тексты информационных и предупреждающих сообщений, которые могут возникать в процессе работы с программой, а также рекомендации по действиям оператора в каждой из описанных ситуаций.

Настоящий документ предназначен для пользователей программного комплекса, имеющих базовые навыки работы с компьютером и общее представление о регрессионном анализе данных и методах обнаружения аномалий.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.11.04-01 РП 03-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

СОДЕРЖАНИЕ

АННОТАЦИЯ.....	2
1. НАЗНАЧЕНИЕ ПРОГРАММЫ.....	4
1.1. Функциональное назначение	4
1.2. Возможности программы	5
1.3. Целевая аудитория и область применения	5
2. УСЛОВИЯ ВЫПОЛНЕНИЯ ПРОГРАММЫ	6
2.1. Аппаратные средства, необходимые для функционирования программы	6
2.2. Программное обеспечение, необходимое для функционирования программы	6
2.3. Требования к персоналу	7
3. ВЫПОЛНЕНИЕ ПРОГРАММЫ.....	8
3.1. Установка продукта	8
3.2. Процесс сборки программного комплекса	8
3.3. Процесс работы с программным комплексом	8
3.4. Экраны программного комплекса	9
3.4.1. Главный экран	9
3.4.2. Экран создания модели	10
3.4.3. Экран генерации данных	10
3.4.4. Экран запуска первичного анализа данных	10
4. СООБЩЕНИЯ ОПЕРАТОРУ	11
СПИСОК ИСПОЛЬЗУЕМОЙ ЛИТЕРАТУРЫ	13
ЛИСТ РЕГИСТРАЦИИ ИЗМЕНЕНИЙ.....	15

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.11.04-01 РП 03-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

1. НАЗНАЧЕНИЕ ПРОГРАММЫ

1.1. Функциональное назначение

Программный комплекс для исследования влияния аномальных наблюдений на точность прогнозирования в регрессионных моделях (коротко, "MSnOutliers") предназначен для исследования влияния аномальных наблюдений на точность прогнозирования в различных регрессионных моделях. Комплекс позволяет проводить сравнительный анализ устойчивости регрессионных методов к наличию аномалий в данных, а также оценивать эффективность различных алгоритмов машинного обучения для обнаружения и удаления выбросов.

Программный комплекс обеспечивает выполнение следующих функций:

1) Работа с регрессионными моделями:

- 1.1) Создание и настройка классических методов регрессии (метод наименьших квадратов - LSM).
- 1.2) Создание и настройка робастных методов регрессии (метод Хьюбера - HUB, метод Тьюки - TUK).
- 1.3) Создание и настройка метода наименьших абсолютных отклонений (LAD).
- 1.4) Создание и настройка метода медианных наклонов Тейла-Сена (THS).

2) Работа с данными:

- 2.1) Генерация синтетических данных с заданным количеством признаков и наблюдений.
- 2.2) Добавление контролируемого уровня шума различных распределений и контролируемых искажений.
- 2.3) Загрузка внешних наборов данных в формате CSV.

3) Очистка данных от аномалий:

- 3.1) Применение алгоритма Isolation Forest (IForest).
- 3.2) Применение алгоритма плотностной кластеризации (DBSCAN).
- 3.3) Применение алгоритма оценки плотности ядра (KDE).
- 3.4) Применение алгоритма k-ближайших соседей (KNN).

4) Анализ результатов:

- 4.1) Проведение экспериментов с различными уровнями шума.
- 4.2) Визуализация зависимости ошибки от уровня шума для каждой модели.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.11.04-01 РП 03-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

4.3) Сравнение эффективности различных методов очистки данных.

1.2. Возможности программы

Программный комплекс "MSnOutliers" реализует следующие возможности:

- 1) Создание и настройка различных типов регрессионных моделей через удобный графический интерфейс.
- 2) Генерация синтетических данных с контролируемым уровнем и типом аномалий.
- 3) Применение алгоритмов машинного обучения для обнаружения и удаления аномальных наблюдений.
- 4) Параллельное выполнение экспериментов для ускорения процесса анализа.
- 5) Визуализация результатов в виде информативных графиков.
- 6) Сохранение конфигураций моделей и результатов анализа для последующего использования.

1.3. Целевая аудитория и область применения

Программный комплекс "MSnOutliers" предназначен для школьников и студентов, проходящих подготовку по дисциплине “Математическая статистика” или каким-либо смежным с ней, а также преподавателей, читающих вышеупомянутые курсы.

Комплекс может использоваться как для образовательных целей, так и для практического применения при выборе наиболее подходящих методов регрессии и очистки данных для конкретных задач анализа.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.11.04-01 РП 03-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

2. УСЛОВИЯ ВЫПОЛНЕНИЯ ПРОГРАММЫ

2.1. Аппаратные средства, необходимые для функционирования программы

Для эксплуатации программного комплекса "MSnOutliers" должны использоваться следующие средства вычислительной техники (СВТ):

- 1) СВТ индивидуального пользования - АРМ оператора, представляющее собой ПЭВМ со следующими минимальными характеристиками:
 - 1.1) Процессор: Intel Core i5 6-го поколения / AMD Ryzen 5 1-го поколения / Apple Silicon (M1/M2/M3) или аналогичный.
 - 1.2) Оперативная память: не менее 8 ГБ.
 - 1.3) Свободное дисковое пространство: не менее 500 МБ.
 - 1.4) Видеоадаптер: с поддержкой разрешения не менее 1280x720.
- 2) Рекомендуемые характеристики для комфортной работы при обработке больших наборов данных:
 - 2.1) Процессор с поддержкой многопоточности (4+ ядер).
 - 2.2) Оперативная память: 16 ГБ и выше.
 - 2.3) SSD-накопитель для ускорения операций чтения/записи файлов.

2.2. Программное обеспечение, необходимое для функционирования программы

Для эксплуатации программного комплекса "MSnOutliers" необходимо следующее программное обеспечение, устанавливаемое на АРМ оператора:

- 1) Операционные системы (один из вариантов)
 - 1.1) Microsoft Windows версии не ниже 10.
 - 1.2) Apple macOS версии не ниже 11 "Big Sur".
 - 1.3) Linux (дистрибутивы Ubuntu 20.04 LTS и выше, Fedora 34 и выше, или другие с поддержкой Qt 6.x и Qt 5.15).
- 2) Библиотеки и зависимости:
 - 2.1) Qt версии 5.15 или выше.
 - 2.2) Библиотека Eigen версии 3.4.0 или выше.
 - 2.3) Библиотека nlohmann/json версии 3.11.3 или выше.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.11.04-01 РП 03-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

2.4) Python версии 3.9 или выше с библиотеками matplotlib версии 3.10.3 или выше и numpy версии 1.25.0 или выше.

3) Инструменты сборки (для установки из исходного кода):

3.1) CMake версии 3.14 или выше.

3.2) Компилятор с поддержкой C++17 (в зависимости от операционной системы)

3.2.1) GCC 9.0 или выше (для Linux).

3.2.2) Clang 13.0.0 или выше (для macOS).

3.2.3) MSVC 2019 или выше (для Windows).

2.3. Требования к персоналу

К эксплуатации программного комплекса "MSnOutliers" допускаются операторы:

- 1) Имеющие навыки работы с ПЭВМ и полностью освоившие графический пользовательский интерфейс операционной системы Microsoft Windows, macOS или Linux.
- 2) Обладающие базовыми знаниями в области регрессионного анализа данных и понимающие принципы работы с различными регрессионными моделями.
- 3) Ознакомленные с документацией программного комплекса "MSnOutliers" и принципами его функционирования.
- 4) Понимающие базовые концепции обнаружения аномалий в данных и методы устойчивой регрессии.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.11.04-01 РП 03-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

3. ВЫПОЛНЕНИЕ ПРОГРАММЫ

3.1. Установка продукта

Для запуска разработанных программ, необходимо их загрузить по ссылке на репозиторий: <https://github.com/Octopupu5/MSnOutliers>.

Основное взаимодействие с программным комплексом происходит через пользовательский интерфейс. В процессе работы комплекса генерируются файлы типа .json, при работе с комплексом нужно убедиться, что файлы генерируются внутри каталога с исходным кодом программного комплекса.

3.2. Процесс сборки программного комплекса

Для сборки проекта используется система CMake. Подробная инструкция к тому, как именно собирать программный комплекс, содержится в файле README.md. В этом файле описано следующее:

1. Необходимо активировать виртуальную среду в Python.
2. В активированной среде необходимо скачать библиотеки numpy и matplotlib для корректной работы блока отрисовки данных.
3. Необходимо из корневой директории проекта запустить сборку с помощью CMake. В процессе сборки будут собраны бинарные файлы для каждого из блоков программного комплекса.
4. Далее, для взаимодействия с комплексом через интерфейс, необходимо запустить бинарный файл ui.

3.3. Процесс работы с программным комплексом

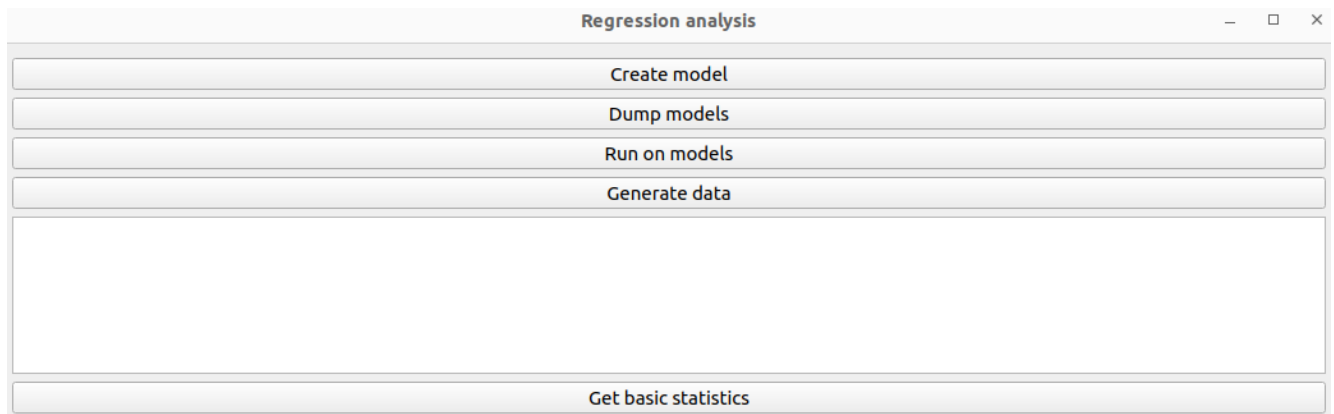
Взаимодействие с приложением происходит через UI. При запуске приложения пользователь попадает на главный экран. На главном экране доступно 4 кнопки. По нажатии на кнопку «Create model», пользователь переходит на экран генерации модели, где вводит желаемые параметры модели. Одним из параметров является путь до файла с данными, на котором предполагается запускать ту или иную модель — если файла по выбранному пути нет, это поле будет оставлено пустым. При этом будет создано окно с сообщением об ошибке поиска файла и уведомлением о том, что путь до данных будет оставлено пустым. После создания модели на главном экране в

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.11.04-01 РП 03-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

таблице моделей в текущей сессии появляется новая модель. По нажатии на кнопку «Dump models» происходит генерация JSON-файла с названием models.json с описанием моделей в текущей сессии. При этом будет выведено сообщение о том, что файл с моделями был сгенерирован и будет указан путь до этого файла. При нажатии на кнопку «Run on models» происходит запуск модуля Main, если файла с описанием моделей нет — появляется окно с сообщением об ошибке. Результатом работы модуля Main является генерация json-файлов с описанием результатов работы программного комплекса. Каждый файл описывает результаты работы модели из списка, полученного на предыдущем этапе. Далее эти файлы попадают в блок отрисовки, где по ним генерируются графики. Графики помещаются на экран визуализации данных и выводятся для ознакомления пользователю. При нажатии на кнопку «Generate data» пользователь переходит на окно генерации данных. После ввода необходимых параметров генерации, запускает процесс регуляции, результатом которого является файл sample.csv. Эти данные могут использоваться пользователем в дальнейшем анализе методов.

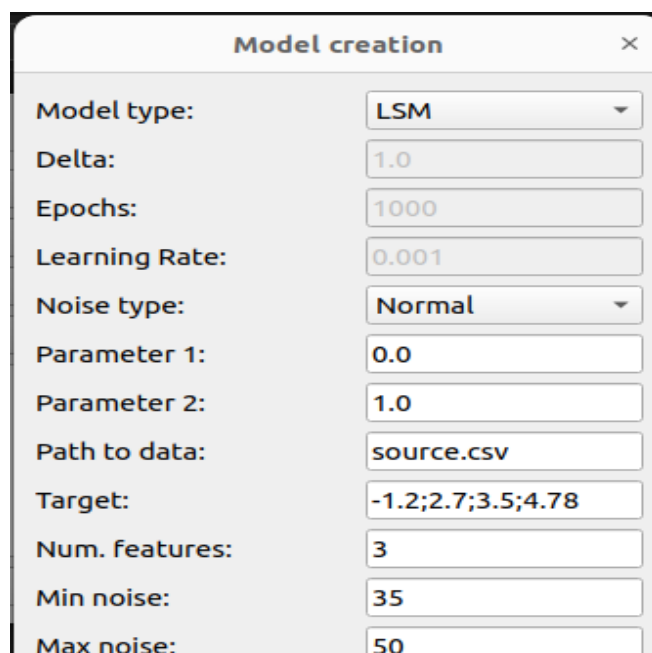
3.4. Экраны программного комплекса

3.4.1. Главный экран



Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.11.04-01 РП 03-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

3.4.2. Экран создания модели



Model creation

Model type: LSM

Delta: 1.0

Epochs: 1000

Learning Rate: 0.001

Noise type: Normal

Parameter 1: 0.0

Parameter 2: 1.0

Path to data: source.csv

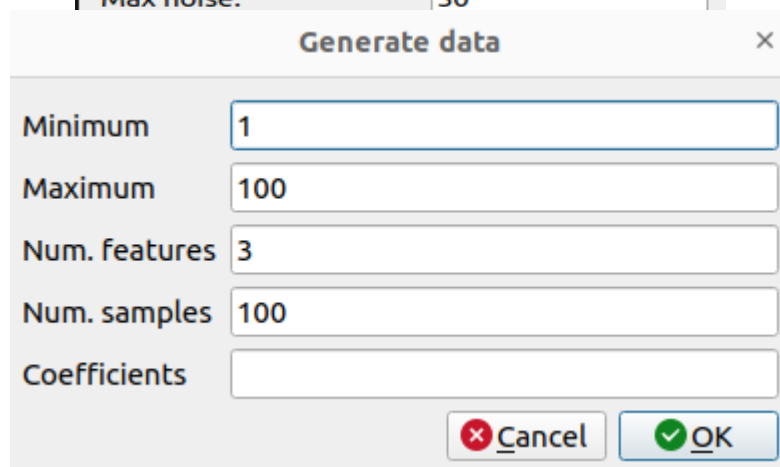
Target: -1.2;2.7;3.5;4.78

Num. features: 3

Min noise: 35

Max noise: 50

3.4.3. Экран генерации данных



Generate data

Minimum: 1

Maximum: 100

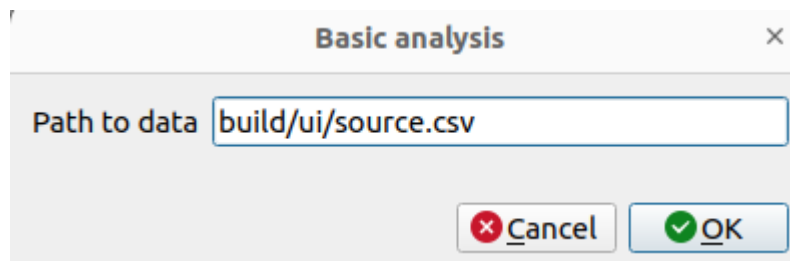
Num. features: 3

Num. samples: 100

Coefficients:

Cancel OK

3.4.4. Экран запуска первичного анализа данных



Basic analysis

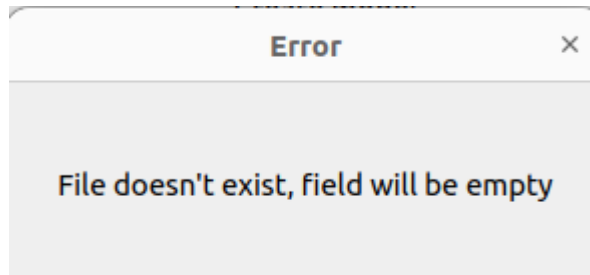
Path to data: build/ui/source.csv

Cancel OK

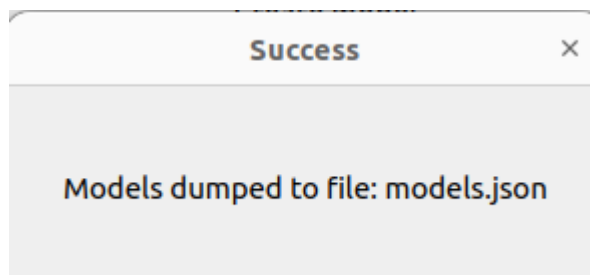
Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.11.04-01 РП 03-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

4. СООБЩЕНИЯ ОПЕРАТОРУ

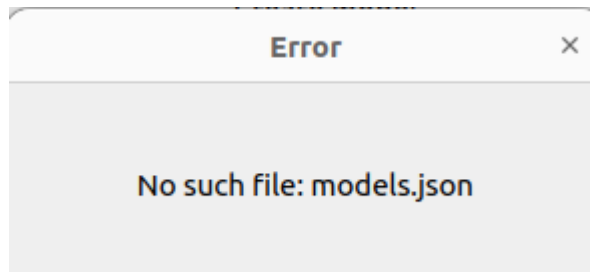
4.1. Сообщение о том, что файл, переданный при генерации модели, не был найден



4.2 Сообщение об успешной генерации файла с описанием выбранных пользователем моделей

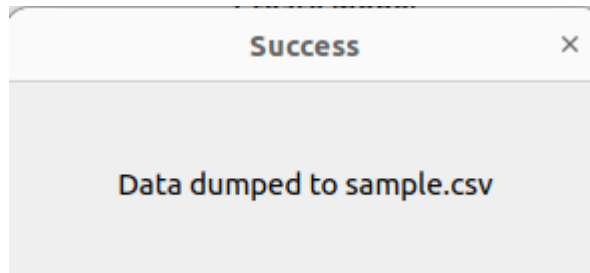


4.3 Сообщение о том, что файл с описанием моделей в рамках текущей сессии, не был найден

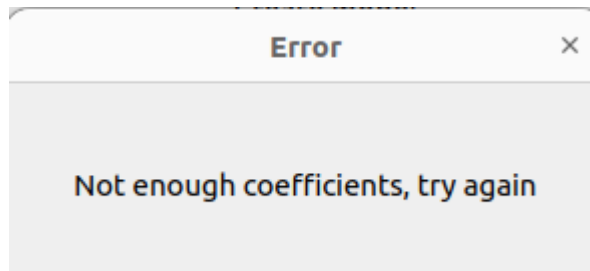


4.4 Сообщение об успешной генерации данных

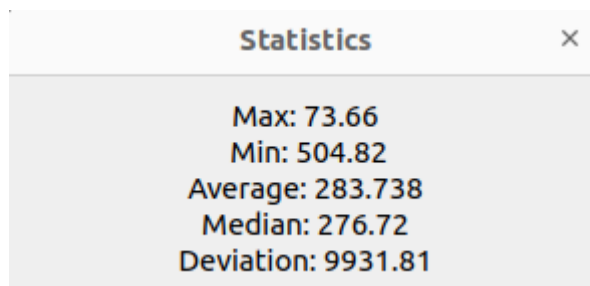
Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.11.04-01 РП 03-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата



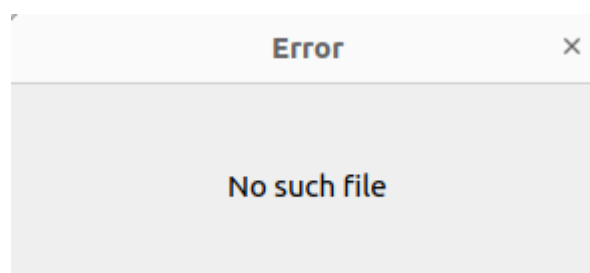
4.5 Сообщение об ошибке в коэффициентах зависимости при генерации данных



4.6 Сообщение с результатами первичной статистики

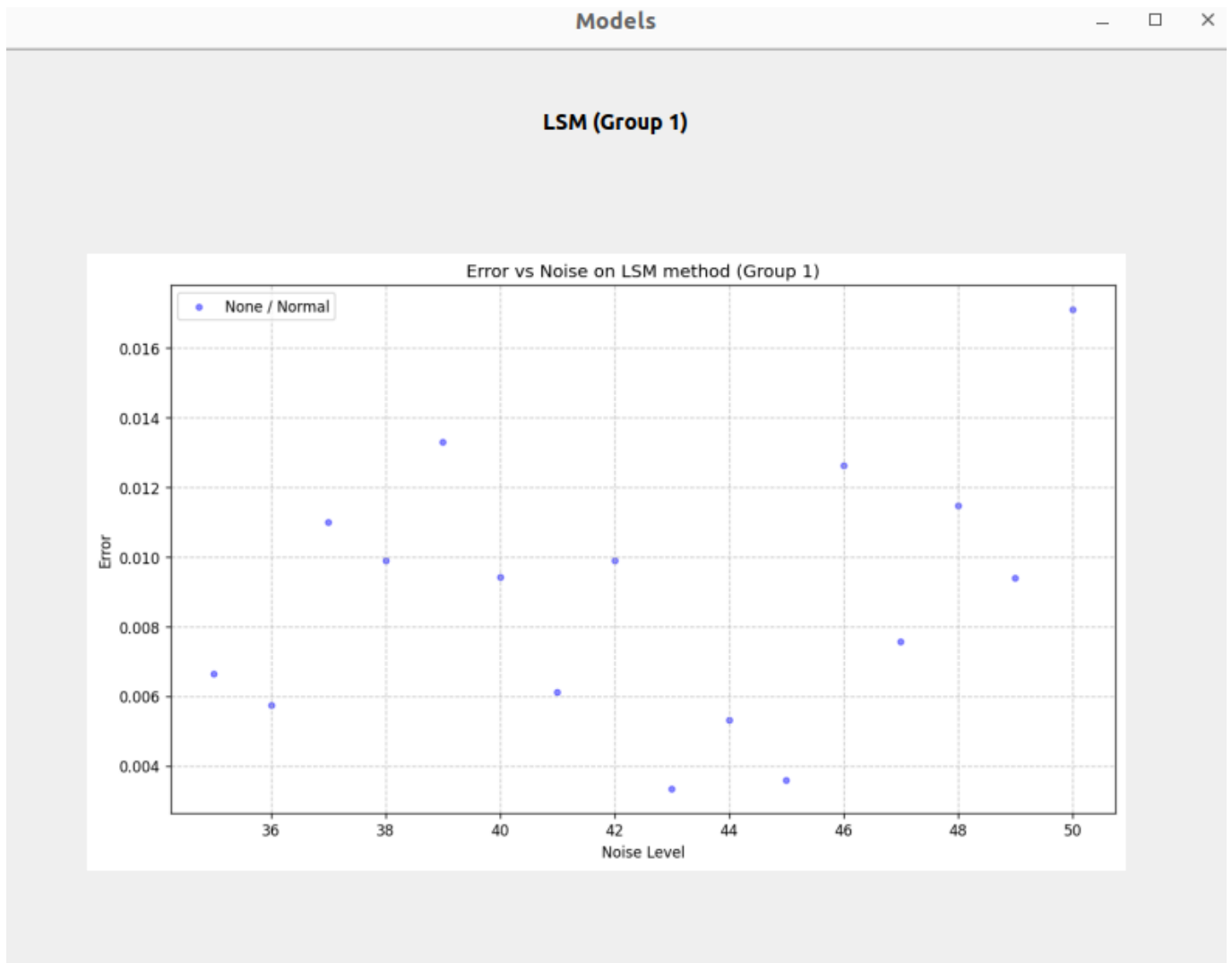


4.7 Сообщение о том, что файл для первичного анализа не был найден



4.8 Визуализация работы блока Main

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.11.04-01 РП 03-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата



СПИСОК ИСПОЛЬЗУЕМОЙ ЛИТЕРАТУРЫ

1. Библиотека "Qt" [Электронный ресурс] / The Qt Company – Режим доступа: <https://www.qt.io/>, свободный.
2. Библиотека "Eigen" [Электронный ресурс] / Eigen Development Team – Режим доступа: <https://eigen.tuxfamily.org/>, свободный.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.11.04-01 РП 03-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

3. Библиотека "nlohmann/json" [Электронный ресурс] / Niels Lohmann – Режим доступа: <https://github.com/nlohmann/json>, свободный.
4. Библиотека "matplotlib" [Электронный ресурс] / Matplotlib Development Team – Режим доступа: <https://matplotlib.org>, свободный.
5. Библиотека "numpy" [Электронный ресурс] / NumPy Development Team – Режим доступа: <https://numpy.org/>, свободный.
6. Система сборки "CMake" [Электронный ресурс] / Kitware Inc. – Режим доступа: <https://cmake.org/>, свободный.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.11.04-01 РП 03-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

ЛИСТ РЕГИСТРАЦИИ ИЗМЕНЕНИЙ

[illegible]