

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ  
РОССИЙСКОЙ ФЕДЕРАЦИИ**  
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ  
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ  
«Национальный исследовательский ядерный университет «МИФИ»  
**Обнинский институт атомной энергетики** –  
филиал федерального государственного автономного образовательного учреждения высшего  
образования «Национальный исследовательский ядерный университет «МИФИ»  
(ИАТЭ НИЯУ МИФИ)

Отделение Интеллектуальные кибернетические системы

**Выпускная квалификационная работа —  
бакалаврская работа**

по направлению подготовки **09.03.02 Информационные системы и технологии**

Направленность (профиль) **Информационные технологии**

**«Разработка приложения для автоматизированного  
статистического анализа данных, полученных при  
КГО стендовым методом на реакторах типа ВВЭР»**

Выполнил:

студент гр. ИС2-Б20

\_\_\_\_\_

Костевич А.Е.

Руководитель ВКР,

старший преподаватель ОИКС

\_\_\_\_\_

Радаев А.В.

Нормоконтроль

доцент отделения ИКС, к.ф.-м.н.

\_\_\_\_\_

Качанов Б.В.

Выпускная квалификационная

работа допущена к защите

\_\_\_\_\_

Руководитель

образовательной программы

09.03.02 Информационные системы

и технологии

канд. тех. наук

\_\_\_\_\_

Мирзеабасов О.А.

Обнинск, 2024 г

## РЕФЕРАТ

Работа 39 стр., 1 табл., 12 рис., 8 ист.

РАЗРАБОТКА ПРИЛОЖЕНИЯ ДЛЯ АВТОМАТИЗИРОВАННОГО СТАТИСТИЧЕСКОГО АНАЛИЗА ДАННЫХ, ПОЛУЧЕННЫХ ПРИ ПРОВЕДЕНИИ КГО СТЕНДОВЫМ МЕТОДОМ ДЛЯ РЕАКТОРОВ ТИПА ВВЭР.

Объектом разработки является настольное приложение с графическим интерфейсом, поддерживающее функции визуализации, экспорта, а также статистического анализа данных.

Целью данной работы является разработка программного обеспечения с целью повышения точности и упрощения процесса анализа данных, полученных при проведении КГО в пеналах СОДС для реакторов типа ВВЭР.

Задачами ВКР являются изучение и анализ существующего метода КГО, выбор критериев для анализа данных, выбор и освоение инструментов разработки, реализация алгоритмов анализа данных.

В работе производится подробный анализ существующего подхода к проведению процедуры и анализу данных КГО на реакторах типа ВВЭР, приводятся предложения по улучшению и автоматизации этого подхода, а также описывается процесс разработки программного обеспечения с учётом внесённых предложений. Разработанное приложение было успешно апробировано на Нововоронежской АЭС.

## ОПРЕДЕЛЕНИЯ

Реперный радионуклид — радионуклид, по выходу которого из твэла в теплоноситель первого контура судят о герметичности оболочки твэла, так как он обладает ядерно-физическими и химическими характеристиками, позволяющими надежно регистрировать его в условиях эксперимента.

Негерметичный твэл - твэл, в оболочке которого имеется повреждение, приводящее к выходу продуктов деления из него.

Негерметичная ТВС — ТВС, в составе которой имеются негерметичные твэлы.

## ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ

БВ — Бассейн выдержки.

ВВЭР — Водо-водяной энергетический реактор.

КГО — Контроль герметичности оболочек.

ЛПР — Лицо, принимающее решение.

ПД — Продукты деления.

ПК — Продукты коррозии.

СОДС — Система обнаружения дефектных сборок.

ТВС — Тепловыделяющая сборка.

# СОДЕРЖАНИЕ

<b>ВВЕДЕНИЕ</b>	<b>3</b>
<b>1 Обзор существующей методики проведения процедуры КГО</b>	<b>4</b>
стендовым методом	4
1.1 Основные положения . . . . .	4
1.2 Процедура проведения КГО стендовым методом . . . . .	4
1.3 Обработка результатов . . . . .	5
1.4 Учет выгорания топлива ТВС . . . . .	10
<b>2 Предложения по улучшению методики обработки данных</b>	<b>12</b>
2.1 Проблемы существующего подхода . . . . .	12
2.2 Предложения по улучшению . . . . .	13
2.2.1 Метод IQR . . . . .	13
2.2.2 Критерий Шапиро-Уилка . . . . .	15
2.2.3 U-критерий Манна-Уитни . . . . .	16
2.2.4 T-критерий Крамера-Уэлча . . . . .	19
<b>3 Проектирование приложения с учётом внесённых предло-</b>	<b>21</b>
<b>жений</b>	
3.1 Требования к проектируемому приложению . . . . .	21
3.2 Архитектура приложения . . . . .	22
3.3 Пользовательский сценарий использования . . . . .	22
<b>4 Разработка приложения</b>	<b>24</b>
4.1 Обзор инструментов разработки . . . . .	24
4.1.1 Python . . . . .	24
4.1.2 Pandas + NumPy . . . . .	24
4.1.3 SciPy . . . . .	25
4.1.4 Matplotlib+Seaborn . . . . .	25
4.1.5 PyQt . . . . .	25
4.2 Импорт данных . . . . .	26
4.3 Анализ данных . . . . .	28
4.3.1 Поиск выбросов . . . . .	28

4.3.2	Статистические тесты . . . . .	30
4.4	Пользовательский интерфейс . . . . .	31
4.4.1	Основные окна приложения . . . . .	31
4.4.2	Окно импорта файла . . . . .	32
4.4.3	Интеграция Matplotlib в качестве QWidget . . . . .	32
4.4.4	Главное окно приложения . . . . .	34
4.4.5	Окно анализа выборок . . . . .	34
<b>ЗАКЛЮЧЕНИЕ</b>		<b>37</b>
<b>СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ</b>		<b>38</b>

# ВВЕДЕНИЕ

Атомные электростанции играют ключевую роль в современной энергетике. Однако сопутствующие ядерной энергетике риски требуют непрерывного совершенствования методов контроля и обслуживания ядерных установок.

В частности, одним из значимых аспектов эксплуатации ядерных реакторов является контроль герметичности оболочек тепловыделяющих элементов. В настоящее время анализ данных, полученных при проведении КГО, частично осуществляется в ручном режиме, что требует значительных ресурсов времени и труда. Более того, этот подход подвержен человеческим ошибкам и может ограничивать возможности в проведении анализа данных с высокой точностью и скоростью.

Как известно, одним из недостатков реактора типа ВВЭР является невозможность перегрузки топлива без остановки реактора и ошибка, допущенная при принятии решения относительно герметичности ТВС, может потребовать незапланированную остановку реактора, что влечёт за собой существенные экономические издержки.

Цель настоящей работы заключается в разработке программного обеспечения, работа которого направлена на повышение эффективности и достоверности результатов КГО, а также снижение трудовых затрат.

В данной работе будет проведен обзор существующего метода обработки результатов КГО, приведены предложения по его автоматизации, а также описан процесс создания прототипа программного обеспечения.

# **1 Обзор существующей методики проведения процедуры КГО стендовым методом**

## **1.1 Основные положения**

В данной работе рассматривается метод КГО в пеналах СОДС [1], который является одним из наиболее надёжных способов определения негерметичных ТВС. СОДС входит в состав обязательного оборудования всех действующих и проектируемых АЭС с реактором ВВЭР.

Метод основан на измерении утечки ПД из-под оболочек твэлов путем гамма-спектрометрического анализа изотопного состава проб воды, отбираемых из контура циркуляции СОДС, по активности реперных радионуклидов  $^{131}\text{I}$ ,  $^{134}\text{Cs}$ ,  $^{136}\text{Cs}$ ,  $^{137}\text{Cs}$  и  $^{133}\text{Xe}$ . Инициирование выхода радионуклидов в воду стенда КГО осуществляется посредством изменения давления циркулирующей по контуру стенда воды в процессе выдержки ТВС в этой воде – настаивании.

## **1.2 Процедура проведения КГО стендовым методом**

1. Процедура проведения КГО начинается проведения испытаний для каждой ТВС в пеналах СОДС с последующим отбором проб воды.

Проверка ТВС проводится при циркуляции воды по контуру стенда КГО без ее замены и состоит из двух циклов:

- Настаивание ТВС при избыточном (верхнем) давлении в контуре от  $4,5 \cdot 10^5$  Па до  $6,0 \cdot 10^5$  Па продолжительностью 5 минут.
- Настаивание ТВС при избыточном (нижнем) давлении в контуре от  $1,0 \cdot 10^5$  Па до  $1,5 \cdot 10^5$  Па до полного перемешивания (не менее 15 минут).

С целью соблюдения одинаковых условий испытаний требуется, чтобы значения верхнего и нижнего избыточного давления были одинаковыми при проверке всех ТВС.

2. После завершения настаивания ТВС производится отбор пробы воды из контура стенда КГО.



3. В каждой  $j$ -ой пробе воды, взятой из стенда КГО при испытании  $j$ -ой ТВС, на спектрометрической установке измеряются значения удельной активности и приводятся на момент останова реактора:

- $A_{j,\text{КГО}}^i$  — реперных  $i$ -х радионуклидов продуктов деления ( $^{131}\text{I}$ ,  $^{134}\text{Cs}$ ,  $^{136}\text{Cs}$ ,  $^{137}\text{Cs}$  и  $^{133}\text{Xe}$ )
- $A_{j,\text{КГО}}^{i'}$  — радионуклида продуктов коррозии(ПК) ( $^{54}\text{Mn}$  или  $^{58}\text{Co}$ ,  $^{60}\text{Co}$ ,  $^{51}\text{Cr}$ ,  $^{59}\text{Fe}$ ).

4. Для учета фоновой активности радионуклидов йода, цезия и продуктов коррозии периодически производится измерение их активности в воде, подаваемой в стенд КГО (с каждой вновь приготовленной порцией раствора борной кислоты на СВО), и в бассейне выдержки (один раз в сутки).

5. Проверка фоновой составляющей за счет загрязнения стенда радиоактивными продуктами (холостая проба) производится перед началом работ по КГО, а также периодически (не реже одного раза в сутки). Для этого без загрузки ТВС в пенал проводятся все операции по промывке контура и настаиванию с отбором и анализом пробы.

6. Итогом проведения спектрометрического анализа проб воды является таблица значений, в которых для каждой  $j$ -ой ТВС приводятся в соответствие значения активности  $A_{j,\text{КГО}}^i$  каждого из регистрируемых реперных радионуклидов и  $A_{j,\text{КГО}}^{i'}$  продуктов коррозии. Статистический анализ результатов измерения проводится для ТВС, в пробах которых значимо регистрировались ПД. Результаты измерений ТВС, при проверке которых реперные ПД не регистрировались, из статистического расчета исключаются.

### 1.3 Обработка результатов

1. Анализ герметичности ТВС, согласно [1], основан на выборочном поиске выбросов методом "трёх сигм".

2. Основными реперными радионуклидами, по которым устанавливается наличие(отсутствие) негерметичных твэлов в ТВС являются  $^{131}\text{I}$ ,  $^{134}\text{Cs}$ ,  $^{137}\text{Cs}$ . Наличие в контролируемой пробе  $^{136}\text{Cs}$  и(или)  $^{133}\text{Xe}$ , значимо

превышающих их содержание в холостых пробах, является однозначным основанием для включения ТВС в список подозрительных, требующих как минимум дополнительной проверки.

3. Полученные значения представляются в графическом виде в такой хронологической последовательности, в какой ТВС проверялись в стенде КГО. Примеры графического представления результатов КГО приведены на рисунках 1 и 2. На основании визуального анализа этих данных на графике может быть сделано заключение, относятся ли они к одному статистическому распределению. Таким способом проводится оценка соблюдения одинаковых условий проверки всех ТВС. Если условия менялись с течением времени (на практике так происходит почти всегда), то производится разделение исходных данных на выборки, которые относятся к одному статистическому распределению.

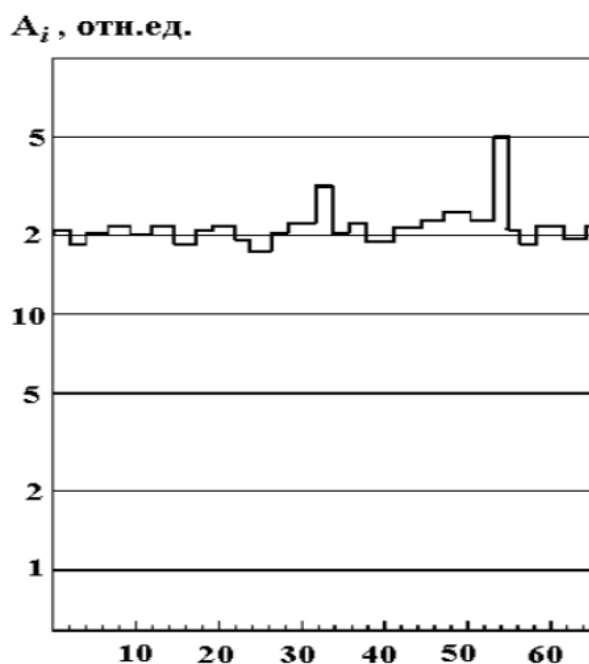


Рисунок 1 – Графическое хронологическое представление данных, принадлежащих к одному распределению. [1]

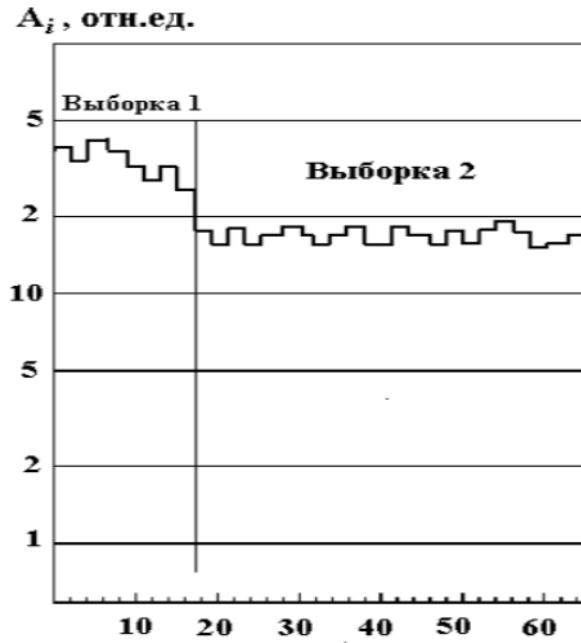


Рисунок 2 – Графическое хронологическое представление данных, принадлежащих к различным распределениям. [1]

4. Для каждой полученной совокупности данных, относящихся к одному и тому же статистическому распределению, вычисляются  $\bar{A}_{\text{кго}}^i$  — среднеарифметические значения удельной активности радионуклидов  $^{131}\text{I}$  (и  $^{134}\text{Cs}$ ,  $^{136}\text{Cs}$ ,  $^{137}\text{Cs}$ ,  $^{133}\text{Xe}$ ) и  $\bar{A}_{\text{кго}}^{i'}$  — среднеарифметическое значение удельной активности  $^{54}\text{Mn}$  (или  $^{58}\text{Co}$ ,  $^{60}\text{Co}$ ,  $^{51}\text{Cr}$ ,  $^{59}\text{Fe}$ ), по формулам 1 и 2:

$$\bar{A}_{\text{кго}}^i = \frac{1}{N} \sum_{j=1}^N A_{j,\text{кго}}^i. \quad (1)$$

$$\bar{A}_{\text{кго}}^{i'} = \frac{1}{N} \sum_{j=1}^N A_{j,\text{кго}}^{i'}. \quad (2)$$

Кроме того, рассчитывают соответствующие им среднеквадратичные отклонения (стандартные статистические неопределенности) по формулам 3 и 4:

$$\sigma_{A_{\text{кго}}^i} = \sqrt{\frac{1}{N-1} \sum_{j=1}^N (A_{j,\text{кго}}^i - \bar{A}_{\text{кго}}^i)^2}, \quad (3)$$

$$\sigma_{A_{\text{кго}}^{i'}} = \sqrt{\frac{1}{N-1} \sum_{j=1}^N (A_{j,\text{кго}}^{i'} - \bar{A}_{\text{кго}}^{i'})^2}, \quad (4)$$

где  $N$  — количество проверенных ТВС.

5. Если  $N > 10$ , то ТВС, для которых выполняется условие 5 являются герметичными.

$$A_{j, \text{кго}}^i \leq A_{\text{кго}}^i + 3 * \sigma_{A_{\text{кго}}^i}. \quad (5)$$

ТВС, для которых одновременно выполняются условия 6 и 7 являются негерметичными.

$$A_{j, \text{кго}}^i > A_{\text{кго}}^i + 3 * \sigma_{A_{\text{кго}}^i}. \quad (6)$$

$$A_{j, \text{кго}}^{i'} \leq A_{\text{кго}}^{i'} + 3 * \sigma_{A_{\text{кго}}^{i'}}. \quad (7)$$

Важно отметить, что активности радионуклидов ПК измеряются с целью учёта при анализе данных. ПК, образующиеся в конструкционных материалах реактора по мере эксплуатации, переносятся по теплоносителю и могут откладываться на ТВС, что влечёт за собой повышение активности в том числе и реперных ПД [2]. Именно поэтому повышение активности реперных ПД совместно с активностями ПК может являться признаком некачественной отмывки ТВС при подготовке к проведению испытаний.

6. Если количество ТВС в выборке  $N < 10$ , то в формулах 5-7 в качестве коэффициента при и вместо коэффициента 3 используются коэффициенты Стьюдента, приведенные в таблице 1, для доверительной вероятности 0,95.

Таблица 1 – Значения коэффициента Стьюдента в зависимости от количества проверенных ТВС и вероятности, с которой ТВС могут быть отнесены к разряду имеющих негерметичные твэлы

Кол-во ТВС	0,95	0,99	0,999
2	12,7	66,7	637
3	4,30	9,93	31,6
4	3,18	5,84	12,9
5	2,78	4,60	8,61
6	2,57	4,03	6,86
7	2,45	3,71	5,96
8	2,36	3,50	5,41
9	2,31	3,36	5,04
10	2,26	3,25	4,78

7. Заключение о герметичности ТВС, для которых при выполнении условия 7 условие 6 выполняется не для всех основных реперных ПД( $^{131}\text{I}$ ,  $^{134}\text{Cs}$ ,  $^{137}\text{Cs}$ ), производится с учетом дополнительной информации: наличие (величина удельной активности) в пробах КГО других реперных ПД( $^{136}\text{Cs}$ ,  $^{133}\text{Xe}$ ), а также соотношения активности  $^{134}\text{Cs}$  и  $^{137}\text{Cs}$  с учётом выгорания.

8. Если в результате вычислений выявлены ТВС, содержащие твэлы с негерметичными оболочками, то проводится повторный расчет величин по формулам 1-4 и проверка по условиям 6 и 7 для остальных ТВС.

9. Повторение расчетов и проверок производится до тех пор, пока все ТВС, включаемые в повторную проверку, не будут удовлетворять условию 5.

10. После завершения последовательно проведенных расчетов и проверок повторный КГО твэлов проводится для следующих ТВС:

- для которых выполняется условие 6 для основных реперных ПД и одновременно не выполняется условие 7;
- для которых выполняется условие 6, но проверенных сразу после ТВС, для которых также выполняется условие 6.
- для которых выполняются условие 6 и условие 7, но удельная активность  $^{134}\text{Cs}$  и  $^{137}\text{Cs}$  в пробе не превышала  $7,4 \cdot 10^4$  Бк/кг ( $2 \cdot 10^{-6}$  Ки/кг).

11. Результат повторной проверки включается в анализируемую совокупность вместо первичного, если он качественно отличается от него.

#### 1.4 Учет выгорания топлива ТВС

Радионуклид  $^{137}\text{Cs}$  ( $t_{1/2}=30$  лет) является конечным продуктом радиоактивного бета-распада предшественников в цепочке радиоактивного распада. Влияние других мод образования  $^{137}\text{Cs}$  несущественно, так что количество атомов этого радионуклида в топливе и под оболочкой ТВЭЛ оказывается прямо пропорционально выгоранию.

Радионуклид  $^{134}\text{Cs}$  ( $t_{1/2}=2,06$  лет) образуется преимущественно за счет реакции  $(n, \gamma)$  на ядрах стабильного изотопа  $^{133}\text{Cs}$ , который является последним членом цепочки радиоактивного распада и накапливается в топливе практически линейно с выгоранием.

Поскольку образование  $^{134}\text{Cs}$  из  $^{133}\text{Cs}$  в первом приближении линейно с выгоранием, общее количество  $^{134}\text{Cs}$  оказывается близким к квадратичной функции от выгорания, естественно, с учетом радиоактивного распада. Таким образом, соотношение активностей  $^{134}\text{Cs}$  и  $^{137}\text{Cs}$  в топливе и под оболочкой оказывается значимо зависящим от выгорания. При проведении КГО на остановленном реакторе в пеналах СОДС определение соотношения удельных активностей радионуклидов  $^{134}\text{Cs}$  и  $^{137}\text{Cs}$  в пробах КГО может быть использовано в качестве индикатора для дополнительной проверки факта обнаружения ТВС с негерметичными ТВЭЛами.

1. Для ТВС, в которых при выполнении условия (7) условие (6) выполняется не для всех контролируемых реперных ПД, строится коэффициент, равный отношению отношения удельных активностей  $^{134}\text{Cs}$  и  $^{137}\text{Cs}$  в j-ой пробе стенда КГО:

$$K_{\text{кго}}^j = \frac{A_{j,\text{кго}}^{134\text{Cs}}}{A_{j,\text{кго}}^{137\text{Cs}}}. \quad (8)$$

На основе сопоставления определенных по формуле (8) коэффициентов  $K_{\text{кго}}^j$  - соотношений измеренных удельных активностей  $^{134}\text{Cs}$  и  $^{137}\text{Cs}$  в пробах от ТВС, определенных как негерметичные, с кривой, приведенной на рисунке 3 определяются расчетные выгорания топлива ТВС с негерметичными ТВЭЛами.

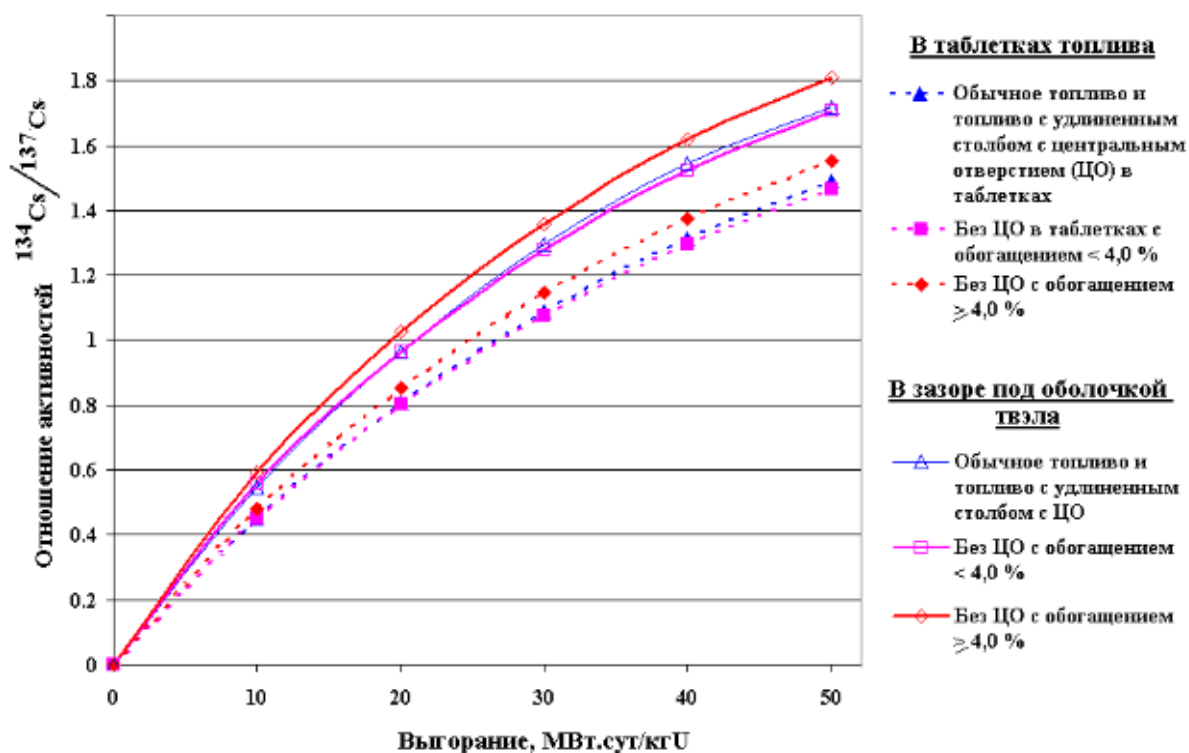


Рисунок 3 – Расчетные соотношения активностей радионуклидов  $^{134}\text{Cs}$  и  $^{137}\text{Cs}$  в топливе и в зазоре под оболочкой твэлов в функции выгорания для типичных историй облучения топлива реакторов ВВЭР-1000. [1]

2. Совпадение (с учетом неопределенностей) полученных таким образом выгораний топлива с выгораниями топлива этих же ТВС по данным физических расчетов (с учетом неравномерности энерговыделения и выгорания по высоте и радиусу кассеты) является дополнительным фактом, подтверждающим наличие негерметичных твэлов в составе контролируемых ТВС.

## 2 Предложения по улучшению методики обработки данных

### 2.1 Проблемы существующего подхода

Рассмотрев методику проведения процедуры КГО согласно[1] можно выделить несколько замечаний, которые можно пересмотреть:

1. Метод анализа данных, приведённый в параграфе 1.3, основывается на поиске выбросов по правилу "3 сигм". Данное правило утверждает, что абсолютная величина отклонения нормально распределённой случайной величины от её математического ожидания не превосходит трёх среднеквадратичных отклонений с вероятностью[3]:

$$P(|X - m| < 3\sigma) = 0,9973. \quad (9)$$

Проблема этого метода заключается в том, что он применим для выборок, значения которых извлечены из нормально распределённых генеральных совокупностей. Но в случае проведения КГО(согласно разделу 1) нет достаточных оснований утверждать, что все значения активностей будут распределены по нормальному закону. Следовательно, требуется проверить характер распределения для каждой выборки с целью установления корректности применения метода "3 сигм".

Кроме того, среднее и среднеквадратическое отклонение, рассчитываемые в данном методе, также изменяются под воздействием аномальных значений, что приводит к маскировке выбросов[4]. Следовательно, имеет смысл рассмотреть альтернативы, например метод межквартильного размаха(IQR).

2. Процедура КГО с учётом времени и объёма испытаний может проходить до нескольких недель. С течением времени в БВ, а также в воде, подаваемой на стенд КГО, может изменяться концентрация борной кислоты с целью борного регулирования, что негативно сказывается на однородности условий проведения испытаний. В связи с этим значения активностей ПД, полученные в разное время, могут принадлежать разным статистическим распределениям, следовательно, анализироваться должны отдельно.



Согласно пункту 1.3.3, разделение на выборки происходит "На основании визуального анализа" графических данных. Хочу отметить, что в изучаемой методике существует способ анализа полученных выборок на принадлежность к одному статистическому распределению с целью объединения нескольких выборок, но он не учитывает анализ на корректность разбиения исходных данных. В связи с вышеперечисленным возникает необходимость проверки данных в выборках, полученных на основании визуального анализа.

3. Описанный процесс анализа данных производится в ручном режиме с использованием программного комплекса Excel. Автоматизация этого процесса позволит снизить вероятность ошибок, ведущих к преждевременной остановке реактора, а также снизить потребность во временных и трудовых затратах.

## **2.2 Предложения по улучшению**

### **2.2.1 Метод IQR**

Учитывая изложенное в 2.1.1, в качестве альтернативы методу "3 сигм" имеет смысл рассмотреть метод поиска выбросов с помощью межквартильного размаха (Interquartile range/IQR).

Межквартильный размах (IQR) — это статистическая мера, равная разности между первым и третьим квартилями распределения (25-м и 75-м процентилями). Первый квартиль ( $Q_1$ ) — это значение, ниже которого находится 25% данных, а третий квартиль ( $Q_3$ ) — это значение, ниже которого находится 75% данных. Можно так же сказать, что интерквартильный размах это половина выборки, центрированная относительно медианы.

Этот показатель полезен для оценки изменчивости признака в асимметричных распределениях или наборах данных с выбросами. Интерквартильный размах является устойчивым (робастным) аналогом дисперсии, поскольку не подвержен влиянию аномальных значений. Именно поэтому он является наиболее перспективным аналогом методу "3 сигм". Таким образом, интерквартильный размах позволяет обнаруживать аномальные значения, являясь альтернативой среднеквадратическому отклонению, ко-

торое эффективно только для нормально распределенных данных. IQR — это непараметрическая оценка, что делает его применимым для анализа данных любого распределения.

В современных практиках поиска выбросов с использованием межквартильного размаха применяется следующий метод расчёта критических значений:

$$X^{min} = Q1 - 1.5 * IQR; \quad (10)$$

$$X^{max} = Q3 + 1.5 * IQR; \quad (11)$$

На рисунке 4 приведена иллюстрация коэффициента IQR для нормального распределения.

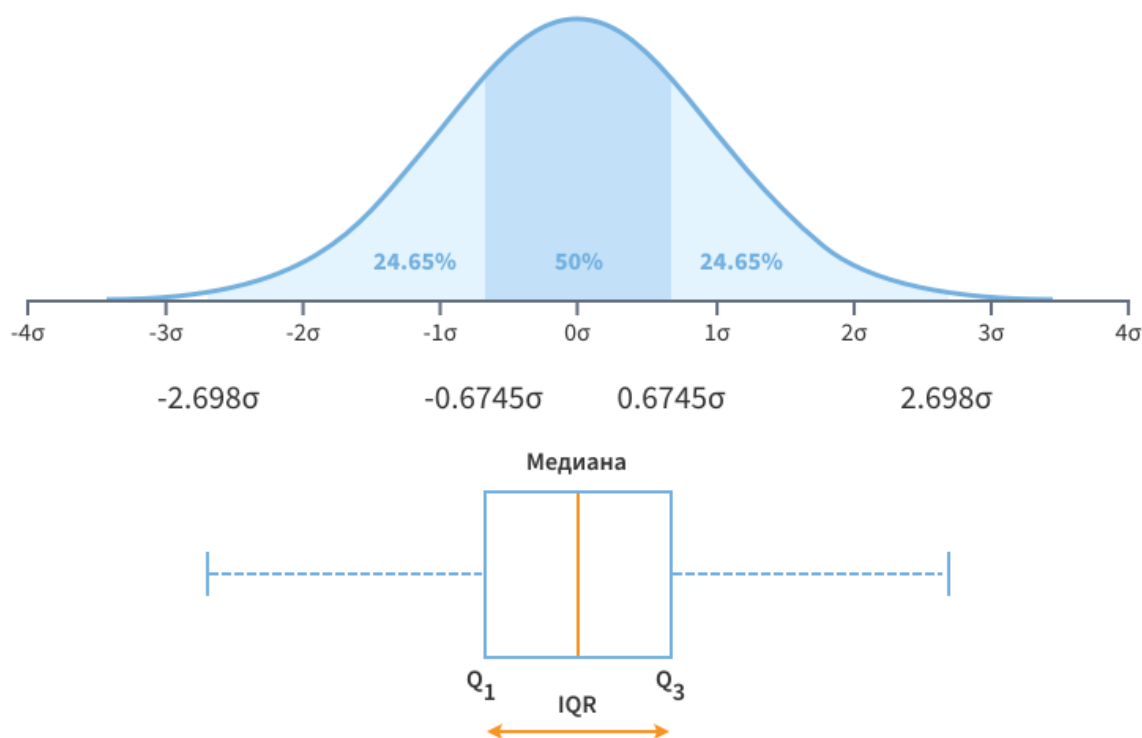


Рисунок 4 – Графическое представление межквартильного размаха на примере нормального распределения.

Из рисунка 4 видно, что данный подход в случае нормального рас-

пределения охватывает интервал:

$$P(|X^{min/max} - X| < 2.698\sigma) = 0,9545. \quad (12)$$

### 2.2.2 Критерий Шапиро-Уилка

Как было описано выше, применение метода "3 сигм" обосновано в случаях, когда распределение выборки близко к нормальному. Следовательно, требуется установить насколько сильно анализируемое распределение отклоняется от нормального закона. На основании результата данной проверки принимается решение относительно выбора метода поиска выбросов.

В качестве критерия проверки распределения на нормальность был выбран критерий Шапиро-Уилка. Данный критерий был выбран на основании его мощности для выборок низкого объёма ( $8 < n < 50$ ) [5].

Критерий Шапиро-Уилка, базируется на анализе линейной комбинации разностей ранговых статистик. При построении статистики для вариационного ряда  $X_{(1)} \leq X_{(2)} \leq X_{(3)} \leq X_{(4)} \leq \dots \leq X_{(n)}$ , полученного по наблюдаемой выборке  $X_1, X_2, \dots, X_n$ , вычисляют величину [6]:

$$S = \sum_k \alpha_k [X_{(n+1-k)} - X_{(k)}], \quad (13)$$

где индекс  $k$  изменяется от 1 до  $n/2$  или от 1 до  $(n-1)/2$  при четном и нечетном  $n$  соответственно. Коэффициенты  $\alpha_k$  приведены в . Статистика критерия вычисляется по формуле:

$$W = \frac{S^2}{\sum_{i=1}^n (X_i - \bar{X})^2}. \quad (14)$$

Таблица критических значений  $p$ -квантилей для статистики  $W$  приводятся в [7].

Данный критерий проверяет следующие гипотезы:

- Гипотеза  $H_0$ : "Случайная величина  $X$  распределена нормально".
- Гипотеза  $H_1$ : "Распределение случайной величины  $X$  не соответствует нормальному закону".

### 2.2.3 U-критерий Манна-Уитни

Помимо проверки распределения на нормальность требуется убедиться, что данные, объединённые в выборки "на основании визуального анализа"[1], принадлежат одному и тому же статистическому распределению. Существуют критерии, которые позволяют определить степень различий между двумя выборками. Например один из таких критериев — U-критерий Манна-Уитни.

U-критерий Манна-Уитни является непараметрическим критерием, поэтому, в отличие от t-критерия Стьюдента, не требует наличия нормального распределения сравниваемых совокупностей. U-критерий подходит для сравнения малых выборок: в каждой из выборок должно быть не менее 3 значений признака. Допускается, чтобы в одной выборке было 2 значения, но во второй тогда должно быть не менее пяти.

При вычислении статистики строится общий вариационный ряд  $X_{(1)} \leq X_{(2)} \leq X_{(3)} \leq X_{(4)} \leq \dots \leq X_{(n)}$ , полученный из выборок случайных величин  $X$  и  $Y$ , а также ранговые ряды  $r(X_i)$  и  $r(Y_i)$ . Вычисляют статистики  $U$  по формулам:

$$U_1 = n1 * n2 + \frac{1}{2}n1(n1 + 1) - R_x, \quad (15)$$

$$U_2 = n1 * n2 + \frac{1}{2}n2(n2 + 1) - R_y, \quad (16)$$

где  $n1$  — объём выборки  $X$ ,  $n2$  — объём выборки  $Y$ ,  $R_x$  и  $R_y$  соответственно:

$$R_x = \sum_{i=1}^{n1} r(x_i); \quad (17)$$

$$R_y = \sum_{i=1}^{n2} r(y_i); \quad (18)$$

Эмпирическую U-статистику Манна-Уитни вычисляют по формуле:

$$U = \min\{U_1, U_2\}. \quad (19)$$

Полученная статистика сравнивается с критическим(табличным) зна-

чением при определённом уровне значимости  $\alpha$ . Если эмпирическое значение меньше критического, то признаётся статистическая значимость различий между уровнями признака в рассматриваемых выборках (принимается альтернативная гипотеза). Достоверность различий тем выше, чем меньше значение  $U$ .

Если выборку, полученную по результатам разбиения исходных данных, описанного в 1.3, разделить при помощи генератора псевдослучайных чисел на 2 подвыборки, то можно сформулировать следующие гипотезы:

- $H_0$ : Между значениями активностей двух подвыборок нет статистических различий, т.е. значения в двух подвыборках распределены по одному и тому же закону.
- $H_1$ : Значения в двух подвыборках распределены по разным законам.

После разбиения на подвыборки выполняется проверка по критерию Манна-Уитни. Результатом этой проверки является  $p$ -значение, которое равно вероятности получения наблюдаемых эмпирических значений при принятии гипотезы  $H_0$ . Гипотеза  $H_0$  отклоняется, если  $p$ -значение меньше уровня значимости  $\alpha = 0,05$

Чтобы исключить влияние случайности, рекомендуется повторить разбиение на подвыборки с последующей проверкой выдвинутых гипотез. Результатом повторений будет выборка  $p$ -значений, которая, при большом количестве повторений, согласно центральной предельной теореме, будет распределена по нормальному закону. Описанный алгоритм проиллюстрирован на рисунке 5.

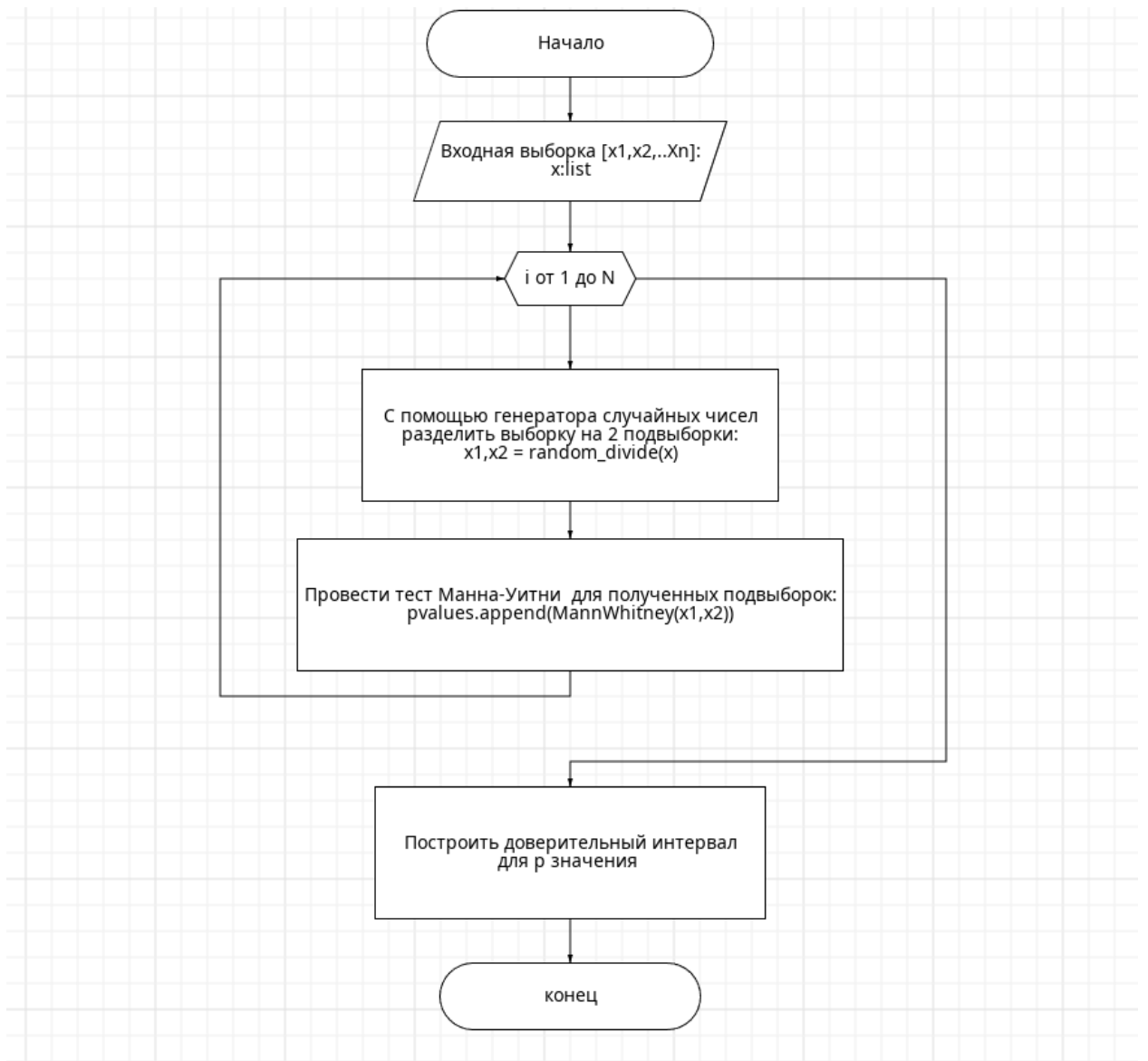


Рисунок 5 – Пример проверки значений выборки на принадлежность одному и тому же статистическому распределению. [1]

Если случайная величина  $X$  распределена по нормальному закону с параметрами  $N(a, \sigma)$ , тогда её математическое ожидание  $a$  с вероятностью  $P$  принадлежит интервалу:

$$a \in \left( \bar{X} - \frac{t_{1-\alpha/2} * S}{\sqrt{N}}, \bar{X} + \frac{t_{1-\alpha/2}(k) * S}{\sqrt{N}} \right), \quad (20)$$

где  $\bar{X}$  — оценка математического ожидания  $a$ ;  $S = \sqrt{S^2}$  — оценка среднего квадратического отклонения  $\sigma$ ;  $t_{1-\alpha/2}(k)$  — квантиль распределения Стьюдента с  $k$  степенями свободы;  $N$  — объём выборки;  $k$  - число степеней свободы при вычислении оценки  $S$ .

#### 2.2.4 Т-критерий Крамера-Уэлча

Т-критерий Крамера-Уэлча — тест, основанный на распределении Стьюдента и предназначенный для проверки статистической гипотезы о равенстве математических ожиданий случайных величин, имеющих необязательно равные известные дисперсии[8]. Является модификацией t-критерия Стьюдента.

Сравнение математических ожиданий проводится с целью проверки принадлежности одному статистическому распределению. Проверка производится в дополнение к проверке по U-критерию Манна-Уитни по алгоритму, представленному на рисунке 5. Данный критерий применяется для выборок, значения которых, в отличие от критерия Манна-Уитни, распределены по нормальному закону.

Пусть даны две независимые выборки нормально распределённых случайных величин  $X_1, X_2, \dots, X_n$  и  $Y_1, Y_2, \dots, Y_n$  с параметрами  $N(\mu_x, \sigma_x^2)$  и  $N(\mu_y, \sigma_y^2)$  соответственно. Тогда проверяемая гипотеза  $H_0: \mu_x = \mu_y$ . Критерий Крамера-Уэлча основывается на статистике, рассчитываемой по формуле:

$$T = \frac{(\bar{x} - \bar{y})\sqrt{mn}}{\sqrt{n\sigma_x^2 + m\sigma_y^2}}, \quad (21)$$

где  $m$  — объём выборки  $x$ ,  $n$  — объём выборки  $y$ .

Критерий Крамера-Уэлча имеет прозрачный смысл — разность выборочных средних арифметических для двух выборок делится на естественную оценку среднего квадратического отклонения этой разности. Естественность указанной оценки состоит в том, что неизвестные статистике дисперсии заменены их выборочными оценками. Из многомерной центральной предельной теоремы и из теорем о наследовании сходимости вытекает, что при росте объемов выборок распределение статистики Т Крамера-Уэлча сходится к стандартному нормальному распределению с математическим ожиданием 0 и дисперсией 1. Итак, при справедливости  $H_0$  и больших объемах выборок распределение статистики Т приближается с помощью стандартного нормального распределения  $\Phi(x)$ , из таблиц которого

следует брать критические значения.

Из асимптотической нормальности статистики  $T$  следует, что правило принятия решения для критерия Крамера-Уэлча выглядит так:

- если  $T \leq \Phi(1-\frac{\alpha}{2})$ , то  $H_0$  принимается на уровне значимости  $\alpha$ .
- если  $T > \Phi(1-\frac{\alpha}{2})$ , то  $H_0$  отклоняется на уровне значимости  $\alpha$ .



## 3 Проектирование приложения с учётом внесённых предложений

### 3.1 Требования к проектируемому приложению

С целью обеспечения наибольшей точности и минимизации рисков, связанных с возможностью некорректного принятия решений относительно герметичности ТВС, предлагаю разработать приложение для обработки и анализа данных, полученных при КГО стендовым водным методом на реакторе типа ВВЭР, которое призвано помочь лицу, принимающему решения относительно герметичности ТВС. Основной задачей данного приложения является автоматизированная(под контролем ЛПР) обработка и анализ данных. Исходные данные хранятся в табличном виде в формате ODS. Под обработкой и анализом данных понимается разделение исходных данных на выборки, принадлежащие к одному статистическому распределению, произведение всех необходимых расчётов, а также принятие решения относительно герметичности ТВС. После предварительного анализа методики, описанной в главе 1, а также учитывая изложенное в главе 2, могу выдвинуть следующие требования к приложению:

1. Приложение должно наглядно демонстрировать ЛПР основания принятия решения относительно герметичности ТВС, т.е. иметь обширный функционал для визуализации и экспорта графиков распределений, негерметичных ТВС и т.д. Исходя из этого был выбран формат настольного приложения с графическим интерфейсом.

2. Приложение должно иметь функционал для проведения статистических тестов с целью проверки выборок, заданных пользователем, на принадлежность одному статистическому распределению, а также на характер распределения.

В качестве статистических тестов предлагаю использовать:

- U-критерий Манна-Уитни — с целью установления принадлежности выборок к одному статистическому распределению
- Критерий Шапиро-Уилка — с целью проверки на отклонение рас-

пределения выборки от нормального закона.

- Критерий Крамера-Уэлча для независимых выборок — с целью установления статистически значимых различий между независимыми выборками, независимые выборки должны быть распределены нормально (Проверяется критерием Шапиро-Уилка).

3. Приложение должно поддерживать поиск выбросов методами "IQR" и "3 сигма". По усмотрению ЛПР может применяться любой метод в зависимости от ситуации.

4. Приложение должно иметь возможности для экспорта всех преобразованных данных, на основании которых принимались решения о герметичности, в исходный ODS формат.

### **3.2 Архитектура приложения**

На данном этапе проектирования выделяются независимые модули, из которых будет состоять проектируемое приложение:

1. Модуль пользовательского интерфейса, который отвечает за создание графического интерфейса, позволяющего пользователям взаимодействовать с другими модулями системы. Данный модуль поддерживает функции работы с графиками, в том числе редактирование, создание легенды, а также сохранение в формате картинки для последующего использования в отчётах о проведении КГО.

2. Модуль предварительной обработки(подготовки), который отвечает за импорт и очистку исходных данных.

3. Модуль обработки и анализа данных, который отвечает за проведение расчётов и статистического анализа, полученных после предварительной обработки. Включает в себя выполнение статистических тестов, поиск выбросов, а также экспорт полученных данных в .ods формат.

### **3.3 Пользовательский сценарий использования**

Ниже приведены шаги, которые будет выполнять пользователь при работе с проектируемым приложением:

1. Импорт данных: пользователь выбирает файл в формате .ods с входными данными для анализа.

2. Анализ входных данных: перед пользователем открывается окно, в котором имеется возможность построения и гибкого редактирования графиков по входным данным. Согласно, 1.3.3 на основании визуального анализа графиков, пользователь разделяет входные данные на выборки.

3. Анализ выборок: После разбиения входных данных возникает возможность выполнения статистических тестов для каждой выборки. Пользователь выбирает выборку и анализируемую величину, после чего запускает статистические тесты. Результаты тестов выводятся в отдельном окне для каждой выборки и анализируемой величины. На основании результатов тестов, пользователь делает заключение о корректности разбиения входных данных и возможности дальнейшего анализа.

В случае, если результаты статистических тестов не позволяют проводить дальнейший поиск выбросов в выборках, окно с выборками закрывается и шаги 2-3 повторяются до тех пор, пока все выборки не будут удовлетворять условиям, позволяющим проводить поиск выбросов.

4. После проведения статистических тестов для выборок пользователь запускает поиск выбросов в выборках согласно выбранному методу. Результатом поиска является 3 таблицы:

- Таблица негерметичных ТВС
- Таблица ТВС, для которых необходимо провести повторную проверку
- Таблица параметров, на основании которых принималось заключение о герметичности

5. После проведения анализа пользователь экспортирует полученные таблицы в .ods для составления отчёта о проведении КГО.

## 4 Разработка приложения

### 4.1 Обзор инструментов разработки

Для разработки приложения на основании требований, описанных в 3.1, были выбраны следующие инструменты:

#### 4.1.1 Python

Язык программирования Python - высокоуровневый, интерпретируемый, объектно-ориентированный язык программирования, который широко используется для разработки веб-приложений, научных вычислений, анализа данных, искусственного интеллекта, автоматизации задач и многих других областей. В совокупности с большим набором пользовательских библиотек, python предоставляет мощные инструменты для обработки, анализа и визуализации данных.

В качестве альтернативы Python рассматривался язык R. R предоставляет более широкий спектр функций по обработке и анализу данных, но набор инструментов для визуализации, а также создания интерфейса приложений ограничен. Именно этот фактор стал решающим в пользу Python.

#### 4.1.2 Pandas + NumPy

NumPy (Numerical Python) — это библиотека для работы с массивами и матрицами, а также для выполнения математических операций над ними. Преимущество NumPy заключается в дешевизне операций и скорости работы, что делает ее неотъемлемым инструментом для анализа данных, машинного обучения, статистики и других областей, где требуется работа с табличными данными.

Pandas — это библиотека для анализа данных, построенная на основе NumPy. Она предоставляет высокоуровневые структуры данных и инструменты для работы с табличными и временными рядами данных.

Совместное использование Pandas и NumPy позволяет эффективно решать широкий спектр задач в области научных исследований, финансо-

вого анализа, обработки больших данных и машинного обучения. NumPy обеспечивает высокую производительность при выполнении числовых вычислений, тогда как Pandas предоставляет удобные инструменты для обработки и анализа данных. Вместе эти библиотеки образуют мощный инструментарий для решения задач, связанных с анализом и обработкой данных в Python.

#### **4.1.3 SciPy**

SciPy — библиотека, основанная на расширении NumPy, которая применяется для более сложных научных и инженерных вычислений. SciPy в основном написана на Python и частично на языках C, C++ и Fortran, в связи с чем отличается высокой производительностью и скоростью работы. В рамках разработки приложения использовался модуль `scipy.stats`, который предоставляет обширный функционал для проведения статистических вычислений.

#### **4.1.4 Matplotlib+Seaborn**

Библиотеки Matplotlib и Seaborn. Эти библиотеки предоставляют возможности для визуализации данных в Python. Matplotlib является основной библиотекой для создания различных типов графиков, в то время как Seaborn предоставляет более высокоуровневый интерфейс для создания статистических графиков.

#### **4.1.5 PyQt**

PyQt — набор расширений кроссплатформенного графического фреймворка Qt, выполненный в виде библиотеки Python. Qt — фреймворк для разработки кроссплатформенного программного обеспечения с графическим интерфейсом, написанный на языке программирования C++.

Базовыми элементами графического интерфейса PyQt являются виджеты (Widgets) — основные строительные блоки для создания графических интерфейсов пользователя. PyQt предоставляет широкий ассортимент виджетов, которые можно использовать для построения сложных и функциональных приложений. Виджеты в PyQt охватывают разнообраз-

ные элементы, такие как кнопки, текстовые поля, метки, списки и многое другое. В этом тексте мы рассмотрим основные виды виджетов и их использование в PyQt.

PyQt использует систему сигналов и слотов для реализации взаимодействия с пользователем. Сигналы используются для уведомления об изменениях состояния, а слоты — для обработки этих изменений.

## 4.2 Импорт данных

На первом этапе разработки необходимо реализовать импорт входных данных из формата .ods. Пример модельных данных представлен на рисунках 6-7

№	№ п/п по программе	Код и номер кассеты	Индекс и номер ПС СУЗ	Коорд. выпр. яч. акт. зоны	Дата	к-во кампаний	Номер штрафа	I-131
		БВ			06.07.2019 7:43			7,10E+04
		Исходная			06.07.2019 8:30			3,13E+01
		Холостая			06.07.2019 8:50		3	1,52E+02
		Холостая			06.07.2019 9:10		4	н/н
		Бак			06.07.2019 10:00			н/н
1	1	N43012 11957	N033472	08-30	06.07.2019 13:15	3	3	2,84E+03
2	2	N43012 11954	N033474	09-27	06.07.2019 13:55	3	4	2,67E+03
3	3	N43012 11965	N033481	07-27	06.07.2019 14:55	3	3	2,82E+03
		Холостая			06.07.2019 21:49		4	7,48E+02
		Исходная			07.07.2019 0:47			8,20E+01
4	16	N43012 13845	N033476	05-33	07.07.2019 1:29	2	3	3,08E+03

Рисунок 6 – Пример входных данных

Cs-134	Cs-137	Mn-54	Cs-136	Xe-133	Ba-140	Ru-103	Ru-106	Ce-141	Ce-144	Ag-110m
4,15E+04	4,92E+04	4,49E+03	5,22E+03	3,33E+04	н/н	н/н	н/н	н/н	н/н	8,22E+02
3,89E+02	5,20E+02	н/н	н/н	н/н	н/н	н/н	н/н	н/н	н/н	н/н
3,21E+03	6,60E+03	3,56E+02	н/н	н/н	н/н	н/н	н/н	н/н	н/н	н/н
1,43E+04	1,67E+04	8,68E+02	н/н	н/н	н/н	н/н	н/н	н/н	н/н	н/н
4,27E+01	9,36E+01	1,56E+00	н/н	н/н	н/н	н/н	н/н	н/н	н/н	н/н
3,32E+02	4,59E+02	3,76E+02	н/н	н/н	н/н	н/н	н/н	н/н	н/н	1,49E+01
1,73E+03	2,12E+03	8,48E+02	н/н	н/н	н/н	н/н	н/н	н/н	н/н	2,24E+01
2,16E+02	2,84E+02	6,40E+02	н/н	н/н	н/н	н/н	н/н	н/н	н/н	2,93E+01
1,19E+03	1,37E+03	1,00E+03	н/н	н/н	н/н	н/н	н/н	н/н	н/н	н/н
6,25E+02	7,66E+02	6,40E+00	н/н	н/н	н/н	н/н	н/н	н/н	н/н	н/н
5,93E+02	6,88E+02	5,38E+02	н/н	н/н	н/н	н/н	н/н	н/н	н/н	3,43E+01

Рисунок 7 – Пример входных данных

Как видно из примера, входные данные содержат строки и столбцы, которые используются в других расчётах. Следовательно, для использования в приложении эти данные необходимо очистить.

Чтение файла из табличных файлов(таких как .xls, .xlsx, .ods) реализовано в методе `read_excel()` библиотеки `pandas`. Данная функция возвращает объект `DataFrame`, который передаётся в метод очистки под определённый шаблон входных данных. Фрагмент реализации метода, отвечающего за очистку `DataFrame` представлена на листинге 1.

Листинг 1 – Метод очистки данных

---

```

1 def _clear_data(self, df):
2     # На текущий момент данные очищаются под конкретный шаблон пробных данных.
3     # Делаем срез DataFrame, используем первые 14 столбцов
4     df = df.iloc[:,0:14]
5     # Переименовываем столбцы для удобного анализа
6     df.rename(columns = {
7         "№": "Id1",
8         "№ п/п по программе": "Id2",
9         "Код и номер кассеты": "Name",
10        "Индекс и номер ПС СУЗ": "Index",
11        "Коорд. выгр. яч. акт. зоны ": "Coordinates",
12        "Дата": "DateTime",

```

```

13         "к-во кампаний":"Age",
14         "Номер пенала":"IdPenal"}, inplace = True)
15
16     for i in range(0, len(df.columns.to_list())):
17         column = df.columns.to_list()[i]
18         df.replace({column:"-"}, np.NaN, inplace=True) # Заменяем спец. символы
19         df.replace({column:"н/н"}, np.NaN, inplace=True)
20
21     # Заполняем пропуски и приводим типы
22     df["Age"]=df["Age"].fillna(0.0).astype(int)
23     df["IdPenal"]=df["IdPenal"].fillna(0.0).astype(int)
24     df["Id2"]=df["Id2"].fillna(0.0).astype(int)
25
26     ...
27
28     return penalties, penalties_id, cassetes

```

---

## 4.3 Анализ данных

### 4.3.1 Поиск выбросов

Поиск выбросов в разрабатываемом приложении реализован в качестве метода у сущности, представляющей выборку. Метод основан на вычислении критического значения для выборки по заданному критерию. После вычисления критического значения объект DataFrame агрегируется по условию. Полученные строки удаляются из исходного DataFrame и производится повторный расчёт критического значения согласно параграфу 1.3.

Фрагмент реализации метода поиска выбросов, связанный с расчётом критических значений представлен в листинге 2.

Листинг 2 – Фрагмент метода поиска выбросов

---

```

1     ...

2     # method_type - аргумент функции, указывающих метод поиска выбросов
3     if method_type=="IQR":
4         Расчёт параметров для метода "IQR"
5         a_q1 = df[criteriums[i]].quantile(q=.25)

```



```

6         a_q3 = df[criteriums[i]].quantile(q=.75)
7         a_IQR = a_q3-a_q1
8
9         a_corosion_q1 = df[criteriums[i]].quantile(q=.25)
10        a_corosion_q3 = df[criteriums[i]].quantile(q=.75)
11        a_corosion_IQR = a_corosion_q3-a_corosion_q1
12
13        a_crit = (a_q3+1.7*a_IQR)
14        a_corosion_crit = (a_corosion_q3+1.7*a_corosion_IQR)
15    else:
16        #Расчёт параметров для метода "3 сигм"
17        a_mean = df[criteriums[i]].mean()
18        a_corosion_mean = df["Mn-54"].mean()
19        a_std = df[criteriums[i]].std()
20        a_corosion_std = df["Mn-54"].std()
21        a_crit = self.crit_value(a_mean, a_std, len(df))
22        a_corosion_crit = self.crit_value(a_corosion_mean, a_corosion_std, len(df))
23
24    crit_df = df[(df[criteriums[i]]>a_crit)].reset_index(drop=True)
25    non_hermetic = crit_df[crit_df["Mn-54"]<a_corosion_crit]
26    recheck = crit_df[crit_df["Mn-54"]>=a_corosion_crit]
27
28    ...

```

---

Результатом выполнения данного метода являются 3 таблицы DataFrame:

- Таблица негерметичных ТВС, содержит исходные данные о ТВС, а так же радионуклиды, на основании которых предполагается негерметичность.
- Таблицы ТВС для повторной проверки, содержит исходные данные о ТВС, а так же радионуклиды, на основании которых предполагается негерметичность.
- Таблица расчётных параметров, содержащая информацию о математическом ожидании, среднеквадратическом отклонении, а так же о КВАНТИЛЯХ.

#### 4.3.2 Статистические тесты

Все статистические тесты, описанные в 2.2.2-2.2.4, реализованы в библиотеке `scipy.stats`, что довольно сильно упрощает реализацию данного модуля. Основной задачей во время разработки данного модуля являлось определение достаточного количества повторений тестов (Манна-Уитни и Крамера-Уэлча) согласно алгоритму, предложенному в 2.2.3. Увеличение количества повторений существенно замедляет работу модуля, но в то же время малое число повторений не обеспечивает требуемой точности.

На основе модельных данных, я проводил тесты для 3 выборок и сравнивал длины доверительных интервалов для  $N$  от 10 до 1000 повторений. По результатам исследования было установлено, что:

- при изменении числа повторений с 10 до 100, длина доверительного интервала при уровне значимости  $\alpha = 0,05$  длина доверительного интервала изменяется в среднем на  $\approx 0,2 - 0,3$  и составляет  $\approx 0,05$ . Время выполнения метода при этом в среднем изменяется на 0,4 секунды
- при изменении числа повторений от 100 до 1000, длина доверительного интервала при уровне значимости  $\alpha = 0,05$  длина доверительного интервала изменяется в среднем на  $\approx 0,02 - 0,03$  и составляет  $\approx 0,018$ . Время выполнения метода при этом в среднем изменяется на 2,1-2,5 секунды.

В результате данного испытания я пришёл к выводу, что длина доверительного интервала  $\approx 0,05$  является приемлемой, поэтому выбрал количество повторений  $N=100$ . Реализация теста по критерию Манна-Уитни представлена в листинге 3.

Листинг 3 – Реализация теста Манна-Уитни

---

```
1 def mannwhitney_test(self, criterium):
2     pvalues = []
3
4     for i in range(0,100):
5         n1, n2 = random_generate(self.df, criterium)
```

```

6
7         statistic, pvalue = stats.mannwhitneyu(n1, n2, alternative='two-sided')
8
9         pvalues.append(pvalue)
10
11     pvalues = pd.Series(pvalues)
12     pvalue = pd.Series(pvalues).mean()
13     confidence = stats.norm.interval(confidence=0.95, loc=np.mean(pvalues),
14                                     scale=stats.sem(pvalues))
15
16     return (format(confidence[0], '.3f'), format(confidence[1], '.3f'))
17

```

---

## 4.4 Пользовательский интерфейс

Данный раздел содержит описание инструментов и трудностей, с которыми я столкнулся во время разработки приложения, а также примеры реализаций графических элементов, используемых в моём приложении.

### 4.4.1 Основные окна приложения

На основании анализа пользовательского сценария использования, представленного в 3.3, был сделан вывод, что проектируемое приложение будет состоять из 4 окон (2 основных и 2 диалоговых):

- Диалоговое окно импорта файла
- Окно анализа пеналов
- Окно анализа выборок
- Диалоговое окно с результатами статистических тестов

Хочу отметить, что разделение окна анализа выборок и окна анализа пеналов сделано намеренно, т.к. при переходе в окно анализа выборок должна сохраняться не только возможность выбрать выборки заново без изменения в логике. Кроме того подразумевается, что работа по пеналам может проводиться параллельно, следовательно, именно поэтому окно анализа выборок было решено сделать самостоятельным окном, а не интегрировать внутрь другого.

#### 4.4.2 Окно импорта файла

При открытии файла используется класс `QFileDialog`, который предоставляет окно выбора файла. Удобство этого класса заключается в том, что имеется возможность установить фильтры на расширения файлов, чтобы не вызывать исключений при попытке открыть файл с расширением, которое не поддерживается. Пример окна для выбора файла представлен на рисунке 8.

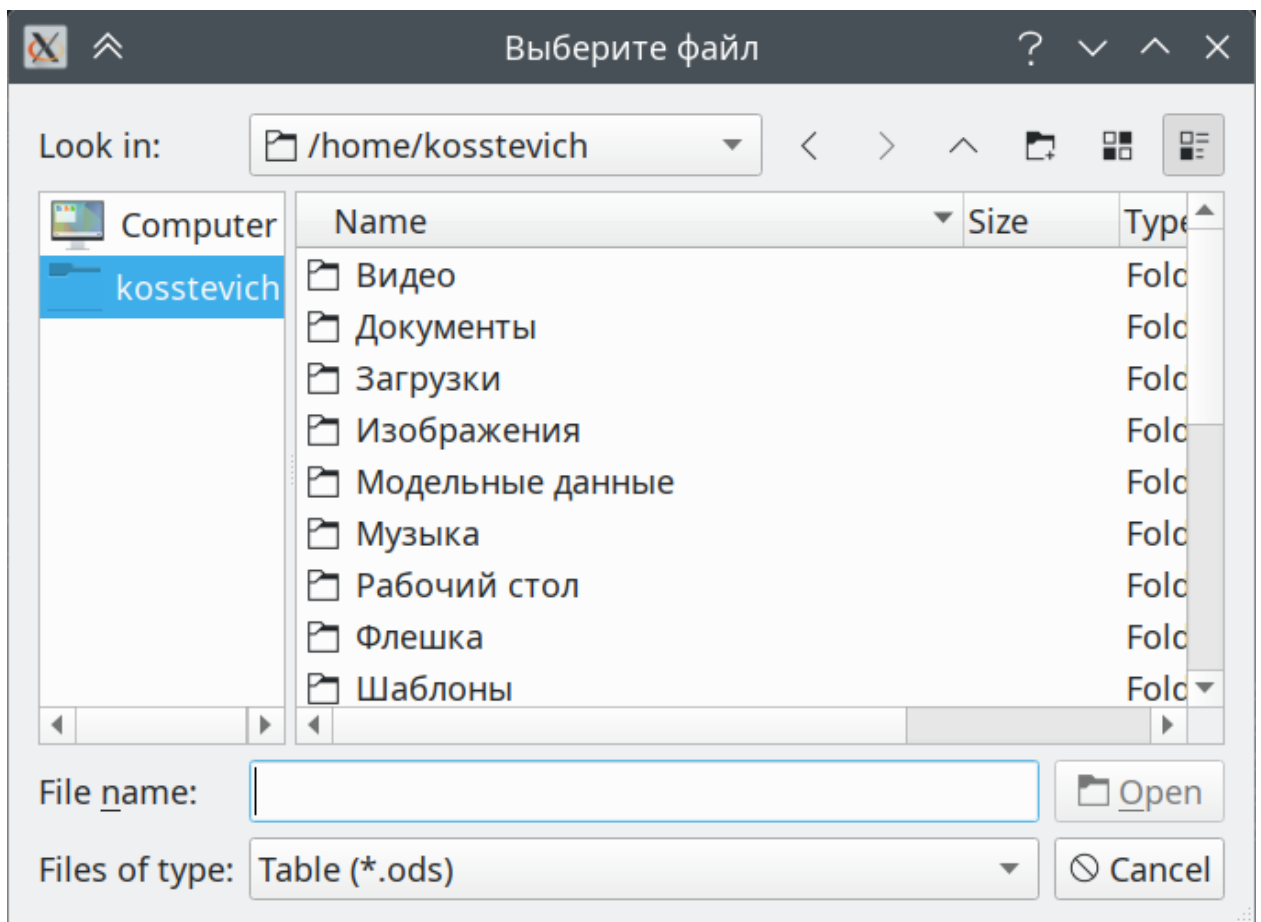


Рисунок 8 – Окно выбора файла

#### 4.4.3 Интеграция Matplotlib в качестве QWidget

Во время разработки приложения я столкнулся с трудностью интеграции графиков Matplotlib в `QWidget`. Проблема заключалась в том, что базовый класс Matplotlib создаёт для графиков независимые окна. Данная реализация мне не подходит, т.к. разрабатываемое приложение должно иметь целостный формат.

Изучив документацию библиотеки, я нашёл класс `FigureCanvasQTAgg`,

предоставляющий backend-реализацию холста Canvas для PyQt. На основе него был разработан собственный пользовательский виджет для использования в моём приложении. Реализация пользовательского виджета представлена в листинге 4

Листинг 4 – Пользовательский класс, интегрирующий Canvas в QWidget

---

```
1 class PlotData(FigureCanvasQTAagg):
2     # Виджет для отрисовки графиков, использующий matplotlib
3     def __init__(self, parent=None):
4         sns.set(style="whitegrid", context="paper")
5         self.fig = plt.figure(figsize=(15, 10))
6         self.axes = self.fig.add_subplot(111)
7         super(PlotData, self).__init__(self.fig)
8
9     def draw_plot(self, df, axe_x="Id1",
10                  axe_y="I-131", type = "barplot", name=None):
11         self.axes.cla()
12         # Метод позволяет рисовать разные виды графиков seaborn
13         # Выбор типа графика происходит по аргументу type
14         getattr(sns,type)(data=df, x=axe_x, y=axe_y, ax=self.axes)
15         plt.title(label= name, fontsize=16)
```

---

Кроме реализации холста я использовал класс NavigationToolbar2QT, который предоставляет возможности навигации, редактирования легенды, а также экспорт графика в формате изображения. На рисунке 9 красной рамкой выделены реализованные мною классы.

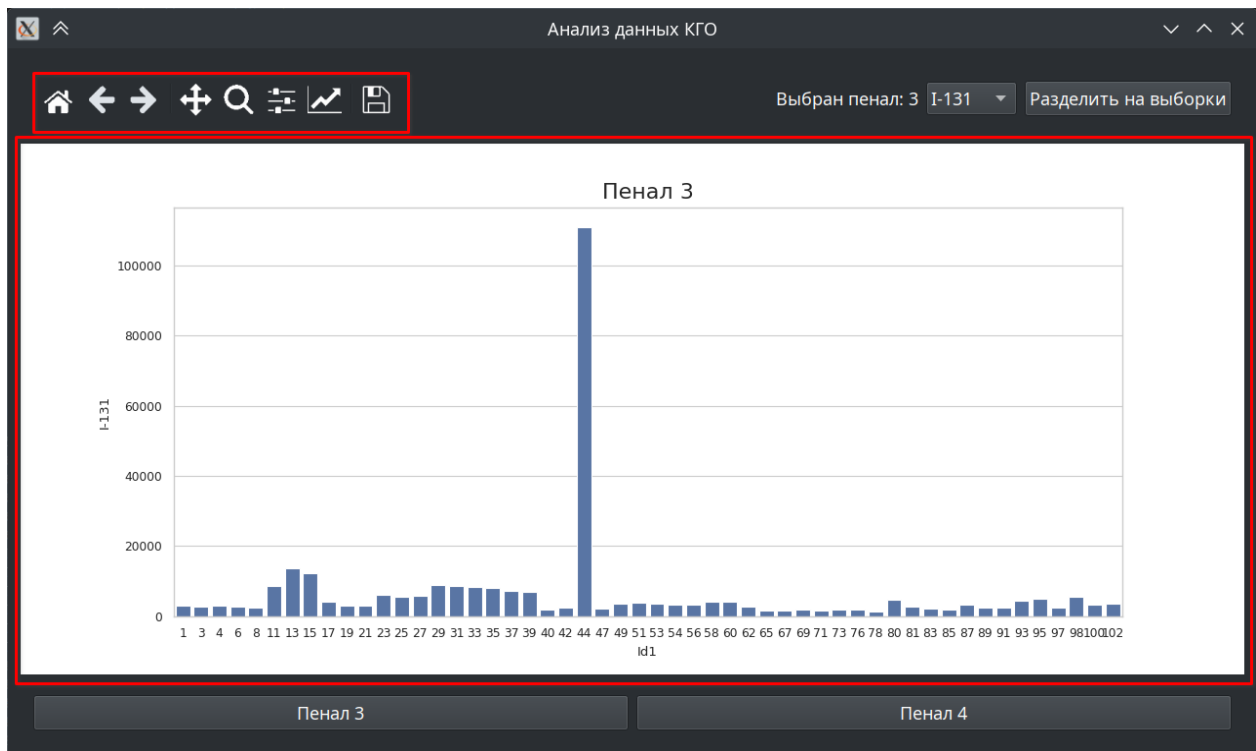


Рисунок 9 – Главное окно приложения

#### 4.4.4 Главное окно приложения

Пример внешнего вида главного окна приложения приведён на рисунке 9. Данное окно реализовано с использованием элементов Qt:

- QPushButton — класс, отвечающий за кнопку.
- QLabel — класс, позволяющий выводить текстовую информацию.
- QComboBox — класс, реализующий выпадающий список.
- QHBoxLayout — класс, позволяющий располагать виджеты в горизонтальной разметке
- QVBoxLayout — класс, позволяющий располагать виджеты в вертикальной разметке

Кнопки, отвечающие за переход по пеналам, реализованы динамически в зависимости от количества пеналов в исходных данных.

#### 4.4.5 Окно анализа выборок

Пример реализации окна анализа выборок представлен на рисунке 10.

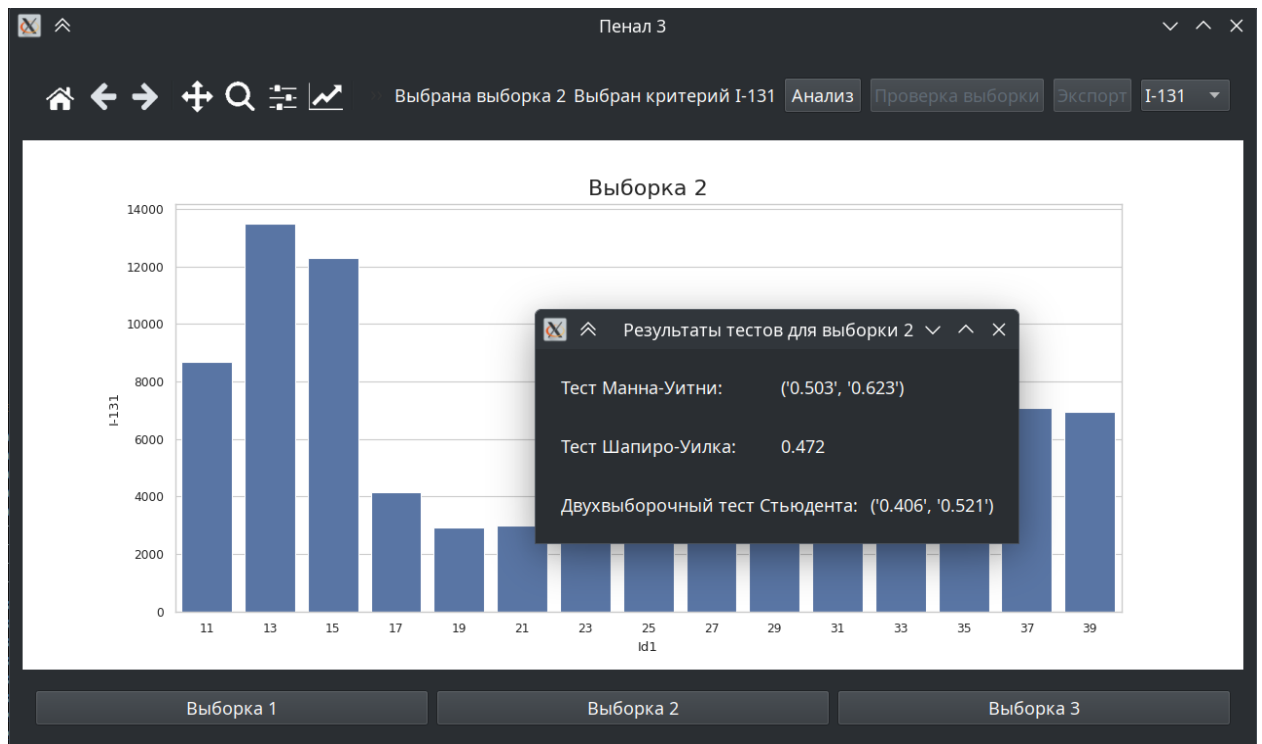


Рисунок 10 – Окно анализа выборок

Несмотря на схожесть с окном анализа выборок, данное окно имеет более сложную структуру. В частности, класс `PlotData`, описанный в 4.4.3, после проведения поиска выбросов заменяется на класс `QTableWidget`, который содержит выходную информацию в табличном виде. Данный функционал реализован с применением класса `QStackedWidget`, который позволяет отображать только один из нескольких дочерних объектов. Пример табличного вывода представлен на рисунке 11.





## ЗАКЛЮЧЕНИЕ

В ходе проделанной работы был проведён анализ существующего метода анализа данных КГО в пеналах СОДС на реакторах ВВЭР типа. Были приведены предложения по улучшению и автоматизации данного процесса. Было реализовано программное обеспечение с учётом внесённых предложений, в которой реализованы следующие функции:

- Поиск выбросов методом "3 сигм".
- Поиск выбросов методом "IQR".
- Анализ распределений выборок .
- Экспорт в табличный формат для составления отчётов КГО.
- Визуализация и экспорт графиков для составления отчётов КГО.

В дальнейшем планируется исследование данной области на предмет интеграции технологий машинного обучения с целью проверки ТВС на герметичность.

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Сборки тепловыделяющие ядерных реакторов типа ВВЭР-1000 ТИПОВАЯ МЕТОДИКА КОНТРОЛЯ ГЕРМЕТИЧНОСТИ ОБОЛОЧЕК ТЕПЛОВЫДЕЛЯЮЩИХ ЭЛЕМЕНТОВ / АО «Концерн Росэнергоатом». — 2009. — РД ЭО 1.1.2.10.0521-2009.
2. Поведение продуктов коррозии в первом контуре ЯЭУ с водным теплоносителем / Б. А. Гусев [и др.] // Вестник СПбГУ. Серия 4. Физика. Химия. — 2012. — № 4. — URL: <https://cyberleninka.ru/article/n/povedenie-produktov-korrozii-v-pervom-konture-yaeu-s-vodnym-teplonositelem>.
3. *Кремер. Н. Ш.* Теория вероятностей и математическая статистика — 5-е изд. — Москва : Издательство Юрайт, 2024. — С. 538.
4. *Дорофеев В. С., Волосатова Т. М.* Ансамблирование методов обнаружения выбросов при подготовке обучающей выборки данных. // Моделирование, оптимизация и информационные технологии. — 2022. — 10(3). — URL: <https://moitvivi.ru/ru/journal/pdf?id=1210>.
5. *Лемешко Б. Ю.* Критерии проверки отклонения распределения от нормального закона. Руководство по применению / Новосибирский государственный технический университет. — 2014.
6. *Александровская Л. Н., Кириллин А. В.* Рекомендации по применению ряда критериев проверки отклонения распределения вероятностей от нормального закона в практике инженерного статистического анализа // Известия Самарского научного центра РАН. — 2017. — № 1—1. — URL: <https://cyberleninka.ru/article/n/rekomendatsii-po-primeneniyu-ryada-kriteriev-proverki-otkloneniya-raspredeleniya-veroyatnostey-ot-normalnogo-zakona-v-praktike>.
7. ГОСТ Р ИСО 5479-2002 / Проверка отклонения распределения вероятностей от нормального распределения. — 2002.
8. *Орлов А. И.* Эконометрика. — Москва "Экзамен", 2002. — С. 437.