МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ

РОССИЙСКОЙ ФЕДЕРАЦИИ

ФГАОУ ВО «СЕВЕРО–КАВКАЗСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

ИНСТИТУТ ЦИФРОВОГО РАЗВИТИЯ

МЕЖИНСТИТУТСКАЯ БАЗОВАЯ КАФЕДРА

**КУРСОВОЙ ПРОЕКТ**

по дисциплине «Междисциплинарный проектный практикум»

на тему:

«Разработка приложения, реализующего двухфакторный дисперсионный анализ»

**Выполнил:**

Пуценко Иван Алексеевич

студент 2 курса

группы ПИЖ-б-о-21-1

направление подготовки

09.03.04 Программная инженерия

направленность (профиль)

«Разработка и сопровождение

программного обеспечения»

очной формы обучения

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(подпись)

**Руководитель проекта:**

Свистунов И.В., доцент

межинститутской базовой кафедры

Проект допущен к защите \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(подпись руководителя) (дата)

Проект выполнен и

защищен с оценкой \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Дата защиты \_\_\_\_\_\_\_\_\_\_\_\_\_

|  |  |  |
| --- | --- | --- |
| Члены комиссии: |  |  |
| доцент кафедры инфокоммуникаций канд. физ.-мат. наук |  | Братченко Н.Ю. |
| доцент кафедры инфокоммуникаций канд. техн. наук | (подпись) | Воронкин Р.А. |
| доцент межинститутской базовой кафедры | (подпись) | Свистунов И.В. |

(подпись)

Ставрополь, 2023 г.

|  |  |
| --- | --- |
|  | УТВЕРЖДАЮ  Зав. кафедрой инфокоммуникаций  докт. техн. наук, профессор  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Г.И. Линец  «\_\_\_\_» \_\_\_\_\_\_\_\_\_\_\_2023 г. |

|  |  |
| --- | --- |
| Институт | цифрового развития |
| Кафедра | инфокоммуникаций |
| Направление подготовки | 09.03.04 Программная инженерия |
| Профиль подготовки | Разработка и сопровождение программного обеспечения |

**ЗАДАНИЕ**

**на курсовой проект**

|  |  |
| --- | --- |
| студента | Пуценко Иван Алексеевич |
| по дисциплине | Междисциплинарный проектный практикум |
| **1. Тема проекта:**  разработка приложения, реализующего проверку гипотез на основе U-критерия Манна-Уитни. | | |
| **2. Цель:** U-критерий Манна-Уитни, разработка приложения, реализующего проверку гипотез на основе U-критерия Манна-Уитни. | | |
| **3. Задачи:**  1) Провести теоретическое описание и анализ U-критерия Манна-Уитни.  2) Программная реализация приложения, реализующего проверку гипотез на основе U-критерия Манна-Уитни. | | |
| **4. Перечень подлежащих разработке вопросов:**  **а) по теоретической части:**  1: теоретические основы и применение  1.1 Основные понятия U-критерия Манна-Уитни  1.2 Примеры применения U-критерия Манна-Уитни  1.3 Выводы  2 Анализ предметной области и постановка задачи  2.1 Предметная область  2.2 Постановка задачи  2.3 Подход к разработке  2.3.1 Выбор языка программирования  2.4 Области применения  2.5 Алгоритмы работы приложения  2.6 Выводы  **б) по практической части:**  3 Разработка программного приложения  3.1 Обоснование выбора модели и среды разработки для программы  3.2 Описание используемых данных  3.3 Анализ и визуализация данных  3.4 Функциональность программы и их описание  3.5 Запуск программы и описание выполнения  3.6 Разработка и реализация алгоритмов  3.7 Выводы | | |
| **5. Исходные данные:**  а) по литературным источникам:  - использовать приведенное в рекомендуемых источниках описание различных алгоритмов поиска в неупорядоченных последовательностях.  - использовать известные в науке результаты исследований для сравнительной оценки полученных результатов и подтверждения адекватности реализуемого алгоритма.  б) по вариантам, разработанным преподавателем:  - самостоятельно с учетом степени подготовленности выбрать язык программирования для реализации алгоритма;  - использовать известные из учебного курса методические рекомендации для исследования алгоритмов поиска.  **6. Список рекомендуемой литературы:**   1. Шелонцев В.А., Шелонцева Л.Н. Непараметрические методы статистики: Учебное пособие. – 3-е изд. – Омск: Полиграфический центр КАН, 2016. – 60 с. 2. М. Холлендер, Д.А. Вулф, пер. с англ. Д. С. Шмерлинга. Непараметрические методы статистики. – Москва: Финансы и статистика, 1983. – 518с. 3. Горяинова Е.Р., Панков А.Р., Платонов Е.Н. Прикладные методы анализа статистических данных. - М.: Издательский дом Высшей школы экономики, 2012. - 310 с. 4. Лаврищева Е. М., Петрухин В.А. Методы и средства инженерии программного обеспечения: Учебное пособие. – Москва: Московский физико-технический институт (государственный университет), 2007. – 415 с. 5. Гайдышев И. Анализ и обработка данных. – Санкт-Петербург: Питер, 2001. – 752 с. 6. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика, Основы моделирования и первичная обработка данных. – Москва: Финансы и статистика, 1983. – 471 с.   7. tkinter — Python interface to Tcl/Tk // Python Documentation URL: https://docs.python.org/3/library/tkinter.html (дата обращения: 20.03.2023).  8. io — Core tools for working with streams // Python Documentation URL: https://docs.python.org/3/library/io.html (дата обращения: 20.03.2023).  9. pandas documentation // Pandas URL: https://pandas.pydata.org/docs/ (дата обращения: 20.03.2023). | | |

**7. Контрольные сроки представления отдельных разделов курсового проекта:**

25 % - Первый раздел « » \_\_\_\_\_\_\_\_\_\_\_\_ 2023 г.

50 % - Второй раздел « » \_\_\_\_\_\_\_\_\_\_\_\_ 2023 г.

75 % - Третий раздел « » \_\_\_\_\_\_\_\_\_\_\_\_ 2023 г.

100 % - Оформленная работа, прошедшая нормоконтроль « » \_\_\_\_\_\_\_\_\_2023 г.

**8. Срок защиты студентом курсового проекта « » 2023г.**

Руководитель курсового проекта

доцент межинститутской базовой

кафедры

Задание принял(а) к исполнению студент(ка) очной формы обучения

2 курса ПИЖ-б-о-21-1 группы \_\_\_\_\_\_\_\_\_\_\_\_ И.А.Пуценко

СОДЕРЖАНИЕ

[ВВЕДЕНИЕ 5](#_Toc136403296)

[1 ТЕОРЕТИЧЕСКИЕ ОСНОВЫ U-КРИТЕРИЯ МАННА-УИТНИ 6](#_Toc136403297)

[1.1 U-критерий Манна-Уитни 6](#_Toc136403298)

[1.2 Алгоритм критерия 6](#_Toc136403299)

[1.3 Статистика U 7](#_Toc136403300)

[1.4 Пример задачи с использованием U-критерия Манна-Уитни 9](#_Toc136403301)

[1.5 Примеры использования U-критерия 11](#_Toc136403302)

[1.6 Выводы 13](#_Toc136403303)

[2 АНАЛИЗ ПРЕДМЕТНОЙ ОБЛАСТИ И ПОСТАНОВКА ЗАДАЧИ 14](#_Toc136403304)

[2.1 Предметная область 14](#_Toc136403305)

[2.2 Постановка задачи 15](#_Toc136403306)

[2.3 Подход к разработке 15](#_Toc136403307)

[2.4 Выбор языка программирования 15](#_Toc136403308)

[2.5 Области применения 16](#_Toc136403309)

[2.6 Алгоритмы 17](#_Toc136403310)

[2.7 Выводы 19](#_Toc136403311)

[3 РАЗРАБОТКА ПРОГРАММНОГО ПРИЛОЖЕНИЯ 20](#_Toc136403312)

[3.1 Ведение 20](#_Toc136403313)

[3.2 Назначение и область применения 20](#_Toc136403314)

[3.3 Обоснование вида разрабатываемой модели 20](#_Toc136403315)

[3.4 Выбор компьютерной среды моделирования 21](#_Toc136403316)

[3.5 Реализация приложения 22](#_Toc136403317)

[3.6 Выводы 24](#_Toc136403318)

[ЗАКЛЮЧЕНИЕ 25](#_Toc136403319)

[СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ 26](#_Toc136403320)

[ПРИЛОЖЕНИЕ А 27](#_Toc136403321)

[ПРИЛОЖЕНИЕ Б 31](#_Toc136403322)

## ВВЕДЕНИЕ

Математическая статистика является наукой, которая изучает методы анализа закономерностей в случайных явлениях и процессах на основе ограниченного количества наблюдений. Ее целью является разработка вероятностно-статистических моделей для описания случайных явлений.

В нашей жизни многие аспекты состоят из множества данных. Примерами могут служить количество дефектных изделий в производстве при определенных условиях или результаты испытаний материалов на прочность. Для того чтобы получить новые знания, человеку необходимо уметь обрабатывать эти данные. Математическая статистика позволяет делать выводы о поведении различных процессов, а также оценивать надежность и точность полученных результатов на основе ограниченной статистической информации. Таким образом, статистические представления являются неотъемлемой частью современного человеческого познания.

Цель данной работы заключается в исследовании и программной реализации U-критерия Манна-Уитни. Для достижения этой цели необходимо выполнить следующие задачи:

1. Изучить теорию, связанную с U-критерием Манна-Уитни.
2. Реализовать код на одном из языков программирования.

В рамках данного курсового проекта описаны три раздела. Первый раздел включает в себя основные понятия, связанные с исследуемым критерием. Второй раздел содержит анализ предметной области, постановку задачи и обоснование выбора языка программирования. Работа также включает два приложения, которые содержат исходный код программы и демонстрацию ее работы.

1. ТЕОРЕТИЧЕСКИЕ ОСНОВЫ U-КРИТЕРИЯ МАННА-УИТНИ

1.1 U-критерий Манна-Уитни

U-критерий Манна-Уитни – это тест, который позволяет определить, существует ли статистически значимая разница между двумя независимыми выборками. Этот тест был разработан в 1947 году Генри Манном и Дональдом Уитни и используется в различных областях, таких как медицина, биология, экономика и др.

С появлением компьютеров и развитием статистических программ, U-критерий Манна-Уитни стал одним из наиболее популярных и эффективных тестов. В настоящее время U-критерий Манна-Уитни используется для широкого спектра задач, таких как сравнение эффективности лекарственных препаратов, оценка качества продукции и др. Сегодня существуют различные программы, которые позволяют использовать U-критерий Манна-Уитни, что делает его доступным для использования в разных областях и задачах.

## 1.2 Алгоритм критерия

Критерий Манна-Уитни – это непараметрический статистический тест, который используется для сравнения двух выборок. Этот тест основан на рангах выборок и позволяет определить, значимо ли различие между ними.

Для использования критерия Манна-Уитни необходимо выполнить следующие шаги:

1. Сформулировать нулевую гипотезу, которая заключается в том, что медианы двух выборок равны.
2. Вычислить ранги для всех элементов двух выборок, объединив их в одну выборку.
3. Вычислить сумму рангов для каждой выборки.
4. Вычислить статистику U, которая является минимальным значением из сумм рангов двух выборок.
5. Вычислить критическое значение статистики U, используя таблицы критических значений. Если вычисленное значение статистики U меньше критического значения, то нулевая гипотеза не отвергается. Если же вычисленное значение статистики U больше критического значения, то нулевая гипотеза отвергается в пользу альтернативной гипотезы, которая заключается в том, что медианы двух выборок не равны.
6. Интерпретировать результаты и сделать выводы.

Критерий Манна-Уитни часто используется в случаях, когда данные не имеют нормального распределения или когда выборки имеют разный размер. Этот тест также позволяет сравнивать выборки с разной дисперсией. Вместо сравнения средних значений выборок, как в t–тесте, критерий Манна-Уитни сравнивает суммы рангов выборок, что делает его более устойчивым к выбросам в данных.

## 1.3 Статистика U

Для применения критерия Манна-Уитни необходимо вычислить статистику U. Эта статистика используется для оценки различий между двумя выборками, которые могут иметь различный размер и не обязательно имеют нормальное распределение.

Статистика U вычисляется по формуле 1.1, которая содержит несколько компонентов. Рассмотрим ее подробнее:

Формула 1.1

1. n1 и n2 – размеры выборок
2. R1 – сумма рангов выборки с размером n1.

в формуле U означает максимально возможное значение суммы рангов для двух выборок, если бы они были объединены в одну выборку и отсортированы по возрастанию.

– это поправка, которая учитывает порядок ранжирования элементов в каждой выборке и дает более точную оценку статистики U.

– это наибольшая сумма рангов элементов в одной из выборок. Ранги элементов выборки могут быть определены несколькими способами, но обычно используется метод присвоения среднего ранга повторяющимся значениям.

Важно отметить, что статистика U может быть определена только для случая, когда выборки имеют разные значения. Если в выборках есть повторяющиеся значения, то необходимо использовать поправку Бонферрони, чтобы скорректировать статистику U.

Поправка Бонферрони – это метод корректировки уровня значимости при множественных сравнениях. Она заключается в том, чтобы умножить уровень значимости (обычно обозначаемый как α) на общее количество сравнений, которые проводятся.

Для применения поправки Бонферрони при использовании U-критерия Манна-Уитни необходимо выполнить следующие шаги:

1. Определить общее количество сравнений, которые вы собираетесь выполнить. Например, если у вас есть три группы и вы планируете сравнить все возможные комбинации, то общее количество сравнений составит три (1 группа – 2 группа, 1 группа – 3 группа, 2 группа – 3 группа).
2. Умножьте уровень значимости α на общее количество сравнений. Например, если ваш исходный уровень значимости составляет 0,05, то после применения поправки Бонферрони новый уровень значимости будет равен .
3. Сравните значение статистики U с критическим значением U, соответствующим скорректированному уровню значимости. Если значение статистики U превышает критическое значение, то различия между выборками можно считать статистически значимыми.

Применение поправки Бонферрони может снизить вероятность ошибки первого рода при множественных сравнениях. Однако, следует помнить, что она также может снизить мощность статистического теста, то есть вероятность обнаружить статистически значимые различия, если они действительно есть. Поэтому, перед применением поправки Бонферрони, важно тщательно обдумать, насколько она необходима для вашего конкретного исследования.

Таким образом, вычисление статистики U основывается на оценке рангов элементов выборок и учитывает размеры выборок. Она позволяет определить, есть ли статистически значимые различия между двумя выборками, что делает ее полезной для анализа различных экспериментальных данных.

## 1.4 Пример задачи с использованием U-критерия Манна-Уитни

У нас есть две группы студентов: группа A и группа B. Мы хотим сравнить их успеваемость в математике, используя результаты экзаменов.

У нас есть следующие данные:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Группа A | 10 | 8 | 12 | 9 | 7 | 11 | 8 | 7 | 10 | 11 |
| Группа B | 13 | 15 | 12 | 14 | 16 | 10 | 12 | 13 | 11 | 14 |

Мы хотим проверить гипотезу о том, что нет статистически значимой разницы в успеваемости между двумя группами.

Шаги решения:

1. Формулируем нулевую и альтернативную гипотезы:

H0: Нет статистически значимой разницы в успеваемости между группой A и группой B.

H1: Существует статистически значимая разница в успеваемости между группой A и группой B.

1. Вычисляем ранги для каждой выборки:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| № | Выборка 1 | Ранг 1 | Выборка 2 | Ранг 2 |
| 1 | 10 | 7 | 13 | 15.5 |
| 2 | 8 | 3.5 | 15 | 19 |
| 3 | 12 | 13 | 12 | 13 |
| 4 | 9 | 5 | 14 | 17.5 |
| 5 | 7 | 1.5 | 16 | 20 |
| 6 | 11 | 10 | 10 | 7 |
| 7 | 8 | 3.5 | 12 | 13 |
| 8 | 7 | 1.5 | 13 | 15.5 |
| 9 | 10 | 7 | 11 | 10 |
| 10 | 11 | 10 | 14 | 17.5 |
| Суммы: |  | 62 |  | 148 |

1. Вычисляем сумму рангов для каждой выборки:
2. Вычисляем статистику U по формуле 1.1:
3. Находим критическое значение для уровня значимости α (обычно α = 0.05) по таблице (рисунок 1.1):



Рисунок 1.1 – таблица критических значений

Для выбранного размера выборки и уровня значимости α = 0.05 критическое значение равно 27.

1. Сравниваем статистику U со значением критической области:

7 < 27, следовательно, мы отвергаем нулевую гипотезу и принимаем альтернативную гипотезу о том, что существует статистически значимая разница в успеваемости между группой A и группой B.

Таким образом, мы можем заключить, что группа B в целом имеет более высокие результаты в математике, чем группа A.

## 1.5 Примеры использования U-критерия

1. Исследование эффективности нового лекарства: Метод U-критерия может использоваться для сравнения эффективности нового лекарства с уже существующим лекарством. Для этого необходимо провести случайную выборку из двух групп пациентов: одна группа получит новое лекарство, а другая – существующее лекарство. Затем проводится сравнение показателей эффективности лекарств между двумя группами.
2. Оценка эффективности рекламы: Метод U-критерия может использоваться для оценки эффективности разных видов рекламы на продажи товаров. Необходимо выбрать две группы потенциальных покупателей: одна группа будет просматривать один вид рекламы, а вторая – другой вид. Затем проводится сравнение объема продаж между двумя группами.
3. Исследование влияния факторов на успеваемость: Метод U-критерия может использоваться для определения влияния различных факторов на успеваемость студентов. Необходимо выбрать две группы студентов: одна группа будет обучаться по одной образовательной методике, а вторая группа – по другой. Затем проводится сравнение успеваемости между двумя группами.
4. Исследование влияния диеты на здоровье: Метод U-критерия может использоваться для оценки влияния различных диет на здоровье людей. Необходимо выбрать две группы людей: одна группа будет следовать одной диете, а вторая – другой диете. Затем проводится сравнение показателей здоровья между двумя группами.

Примеры использования этого теста в IT могут включать:

1. Сравнение времени выполнения двух различных алгоритмов сортировки: Можно сравнить время, которое занимает выборка данных для сортировки алгоритмом быстрой сортировки с временем, которое занимает ту же выборку данных для сортировки алгоритмом слияния. Критерий Манна-Уитни может использоваться для определения того, различается ли время выполнения двух алгоритмов.
2. Сравнение показателей производительности двух разных моделей компьютеров: Можно сравнить производительность двух разных моделей компьютеров по разным критериям, таким как скорость загрузки приложений, время отклика и т. д. Критерий Манна-Уитни может использоваться для определения того, различаются ли производительность двух моделей компьютеров.
3. Сравнение уровня удобства пользования двух разных программ: Можно сравнить уровень удобства пользования двух разных программ по разным критериям, таким как количество кликов мыши, необходимых для выполнения определенной задачи, и т. д. Критерий Манна-Уитни может использоваться для определения того, различается ли уровень удобства пользования двух программ.
4. Сравнение конверсии двух различных версий веб–страницы: Можно сравнить конверсию двух разных версий веб–страницы по разным критериям, таким как количество кликов на определенную кнопку, время, затраченное на заполнение формы и т. д. Критерий Манна-Уитни может использоваться для определения того, различается ли конверсия двух версий веб–страницы.

## 1.6 Выводы

Критерий Манна-Уитни является мощным инструментом статистического анализа, который позволяет сравнивать две выборки и определять, существует ли статистически значимая разница между ними. Этот критерий широко применяется в различных областях, включая медицину, биологию, экономику, социологию, и т.д.

Кроме того, критерий Манна-Уитни имеет ряд преимуществ по сравнению с другими методами статистического анализа, такими как t–тест. В частности, этот критерий не требует предположений о нормальности распределения данных, что делает его более универсальным и применимым для различных типов выборок.

* 1. АНАЛИЗ ПРЕДМЕТНОЙ ОБЛАСТИ И ПОСТАНОВКА ЗАДАЧИ

## 2.1 Предметная область

Предметная область критерия Манна-Уитни - это математическая статистика. В настоящее время термин "статистика" используется в двух основных значениях. Во-первых, как особая отрасль практической деятельности, связанная со сбором, обработкой и анализом массовых количественных данных. Во-вторых, как наука, которая разрабатывает теоретические положения и методы, используемые в статистической практике. Статистика основывается на анализе должным образом собранных и обработанных цифровых данных, и выводы делаются на основе этого анализа.

Критерий Манна-Уитни является статистическим тестом, используемым для сравнения двух независимых выборок и проверки гипотезы о равенстве средних значений между двумя группами. Он является непараметрическим аналогом двухвыборочного t-теста, который позволяет проводить сравнения, когда данные не удовлетворяют условиям нормальности или равенства дисперсий.

Цель использования критерия Манна-Уитни заключается в определении, существует ли статистически значимая разница между двумя группами. Он позволяет исследователям выявить различия между двумя наборами данных, основываясь на ранговом порядке значений, вместо самого значения. Это особенно полезно, когда данные имеют отклонения от нормальности или когда выборки малы.

Автоматизация процесса статистического анализа с использованием критерия Манна-Уитни может облегчить задачу исследователя и позволить ему получить статистические результаты без необходимости вникать в сложности математической статистики.

* 1. Постановка задачи

Нужно разработать приложение, которое позволит оценивать наличие или отсутствие статистически значимой разницы между двумя выборками элементов. Это приложение должно предоставлять следующие функции:

1. Возможность ввода результатов эксперимента в виде csv-файла.
2. Отображение результатов эксперимента.
3. Расчеты с использованием формул, соответствующих U-критерию Манна-Уитни.
   1. Подход к разработке

Цель работы – разработка приложения, реализующего алгоритм расчета U-критерия Манна-Уитни. Для этого необходимо:

* 1. изучить предметную область и алгоритм расчета U-критерия Манна-Уитни;
  2. составить схему алгоритма;
  3. оформить программную документацию;
  4. написать программный код на языке программирования Python;
  5. протестировать приложение.

## 2.4 Выбор языка программирования

Определение языка программирования Python в качестве основного инструмента для разработки приложения, реализующего алгоритм расчета U-критерия Манна-Уитни может быть обосновано следующими факторами:

1. Бесплатность и открытый исходный код: Python распространяется по лицензии с открытым исходным кодом, что делает его доступным для всех. Это также позволяет легко распространять и модифицировать язык согласно нуждам разработчика.
2. Встраиваемость и платформонезависимость: Python обладает интерактивной природой и хорошей переносимостью, что делает его удобным для динамической семантики и быстрого прототипирования. Он может быть легко встроен в различные приложения, даже те, которые разрабатываются на других языках программирования.
3. Динамическая типизация: Python является языком с динамической типизацией, что означает, что тип переменной определяется автоматически во время выполнения программы. Разработчику не нужно предварительно объявлять переменные и их типы, что упрощает процесс программирования и повышает гибкость.
4. Богатство фреймворков: Python имеет обширное сообщество разработчиков и большое количество фреймворков, которые упрощают процесс разработки приложений. Эти фреймворки предлагают готовые решения и инструменты для реализации различных функциональностей, что может сэкономить время и упростить разработку.

Использование языка программирования Python в разработке приложения для реализации алгоритма расчета U-критерия Манна-Уитни может обеспечить гибкость, удобство и эффективность процесса разработки.

## 2.5 Области применения

Реализуемое приложение будет применяться для оценки наличия или отсутствия статистически значимой разницы между двумя выборками по заданному критерию на основе данных, введенных в виде csv файла.

Например, приложение может применяться для сравнения успеваемости групп студентов.

## 2.6 Алгоритмы

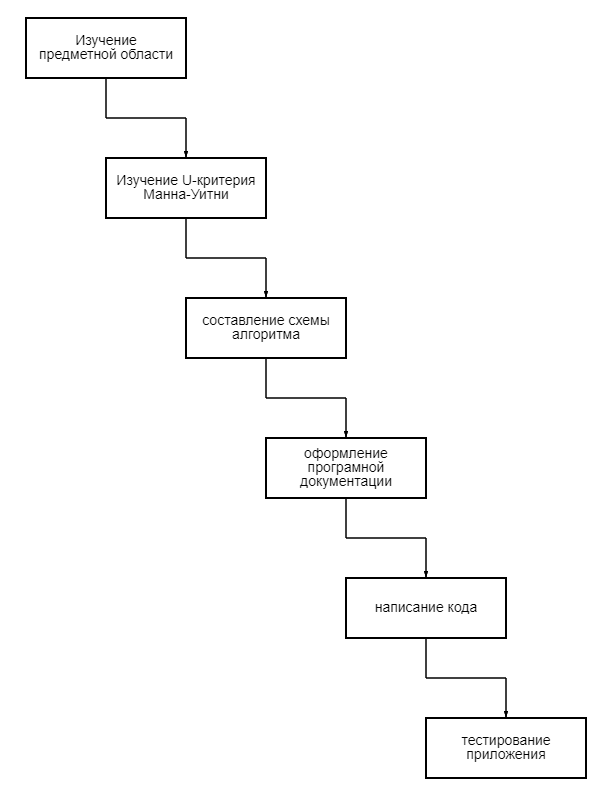


Рисунок 2.1 – Модель разработки приложения

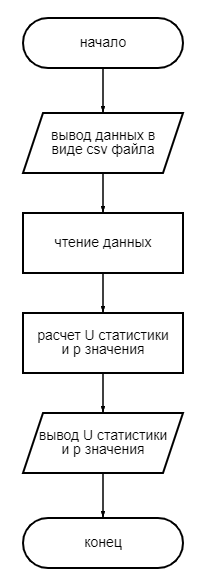


Рисунок 2.2 – Блок-схема решения задачи

## 2.7 Выводы

Проведение анализа в заданной предметной области привело к выводу, что внедрение автоматизированного инструмента для статистической обработки данных способствует значительному улучшению процесса анализа, сделав его более эффективным, автономным и точным. Благодаря автоматизации расчетов и обработки данных, исследователи могут получить более надежные и точные результаты, а также сэкономить время и ресурсы, ранее затрачиваемые на ручные расчеты.

Таким образом, автоматизация процесса статистической обработки данных с использованием предлагаемого приложения на основе алгоритма расчета критерия Манна-Уитни представляет значительный потенциал для улучшения исследовательских и аналитических процессов в различных научных областях, повышая точность и эффективность анализа данных.

* 1. РАЗРАБОТКА ПРОГРАММНОГО ПРИЛОЖЕНИЯ

## 3.1 Ведение

Программа «Критерий Манна-Уитни» предназначена для реализации расчета U-критерия Манна-Уитни.

## 3.2 Назначение и область применения

Реализуемое приложение будет применяться для оценки наличия или отсутствия статистически значимой разницы между двумя выборками по заданному критерию на основе данных, введенных в виде csv файла.

Например, приложение может применяться для сравнения успеваемости групп студентов.

## 3.3 Обоснование вида разрабатываемой модели

При выборе темы "Разработка приложения, реализующего алгоритм расчета U-критерия Манна-Уитни" для курсовой работы было учтено несколько факторов, которые обосновывают его выбор.

Во-первых, этот критерий широко применяется в различных областях исследований, включая экономику, медицину, биологию, социологию и другие. Это связано с необходимостью сравнения нескольких групп по определенному признаку. Таким образом, изучение этого метода позволит студенту освоить инструмент, который широко применяется в научных исследованиях.

Во-вторых, U-критерий Манна-Уитни представляет собой сложный статистический метод, требующий определенных знаний и навыков в области математической статистики. Изучение этого метода позволит студенту углубить свои знания в этой области и развить навыки анализа данных. Разбираясь в деталях данного метода, студент получит возможность более глубоко понять основы статистики и применять их в будущих исследованиях.

В-третьих, U-критерий Манна-Уитни может быть применен для решения реальных задач, что делает данную тему курсовой работы актуальной. Например, можно использовать этот метод для оценки эффективности нового лекарства по сравнению с существующими.

Таким образом, выбор темы "Разработка приложения, реализующего алгоритм расчета U-критерия Манна-Уитни " для курсовой работы является обоснованным и актуальным, учитывая широкое применение этого критерия в научных исследованиях, его сложность, которая позволяет студенту расширить свои знания и навыки в области статистики, а также возможность применения данного метода для решения реальных задач.

## 3.4 Выбор компьютерной среды моделирования

Выбор компьютерной среды PyCharm для выполнения курсовой работы обоснован рядом факторов, которые подтверждают ее преимущества и удобство использования.

Во-первых, PyCharm представляет собой мощный интегрированный редактор кода, специально разработанный для разработки программного кода на языке Python. Он предлагает обширный набор инструментов, включая автодополнение кода, проверку синтаксиса, интеграцию с системой контроля версий Git, отладку и множество других функций. Простой и удобный интерфейс редактора позволяет как новичкам, так и опытным программистам быстро освоиться с ним и повысить эффективность работы.

Во-вторых, PyCharm обладает широкой поддержкой плагинов, которые позволяют расширить его функциональность. Например, существуют плагины для интеграции с различными системами контроля версий, подсветки кода, интеграции с системами непрерывной интеграции и доставки, а также множество других плагинов, которые удовлетворяют разнообразным потребностям программистов. Это позволяет использовать PyCharm для решения различных задач в области программирования и адаптировать его к индивидуальным предпочтениям и потребностям пользователей.

В-третьих, PyCharm является кроссплатформенным редактором кода, что означает его совместимость с различными операционными системами, такими как Windows, Mac OS и Linux. Это делает PyCharm универсальным инструментом разработки, который можно использовать на разных платформах, обеспечивая непрерывность работы и гибкость выбора операционной системы.

Таким образом, выбор компьютерной среды моделирования PyCharm для выполнения курсовой работы обоснован мощным функционалом, удобством использования, широкой поддержкой плагинов и кроссплатформенной совместимостью. Этот инструмент позволяет программистам эффективно разрабатывать программный код на языке Python и адаптировать его под свои индивидуальные потребности и предпочтения.

## 3.5 Реализация приложения

1. В начале программы, подключаются необходимые библиотеки: **tkinter** для создания графического интерфейса, **filedialog** для открытия диалогового окна выбора файла, **csv** для работы с файлами CSV и **scipy.stats** для расчета критерия Манна-Уитни.
2. Затем, определены функции:

**program\_data()** - создает новое окно (Toplevel) с информацией о программе.

**instruction()** - создает новое окно (Toplevel) с приветствием и инструкцией по использованию приложения.

**open\_csv\_file()** - открывает диалоговое окно для выбора CSV-файла, загружает данные из файла и отображает их в соответствующих текстовых полях на графическом интерфейсе.

**calculate\_mann\_whitney()** - получает данные из текстовых полей, выполняет расчет критерия Манна-Уитни с использованием функции **mannwhitneyu** из библиотеки **scipy.stats**, и обновляет метку на графическом интерфейсе с результатами расчета.

1. Создается главное окно (Tk) с заголовком "Критерий Манна-Уитни".
2. Создается меню для главного окна, где определены опции:

"Данные о программе" - при выборе открывается окно с информацией о программе (**program\_data()**).

"Инструкция" - при выборе открывается окно с приветствием и инструкцией (**instruction()**).

"Открыть CSV" - при выборе открывается диалоговое окно для выбора CSV-файла (**open\_csv\_file()**).

"Выход" - при выборе закрывается главное окно и программа завершается.

1. Создаются метки и текстовые поля для ввода данных:

**label\_data1** - метка для описания первой выборки данных.

**entry\_data1** - текстовое поле для ввода данных первой выборки.

**label\_data2** - метка для описания второй выборки данных.

**entry\_data2** - текстовое поле для ввода данных второй выборки.

1. Создается кнопка "Рассчитать", которая вызывает функцию **calculate\_mann\_whitney()** при нажатии.
2. Создается метка **label\_result** для вывода результатов расчета.
3. Запускается цикл обработки событий (mainloop()), который отображает графический интерфейс и обрабатывает события взаимодействия пользователя с программой.

## 3.6 Выводы

В процессе разработки третьего раздела курсовой работы были определены цель и область применения программного приложения. Была обоснована выбранная модель разработки и предложена компьютерная среда для моделирования. Кроме того, в разделе был описан исходный код на языке программирования Python, который реализует функциональность приложения.

# ЗАКЛЮЧЕНИЕ

В ходе выполнения курсового проекта были решены следующие задачи:

1. Было проведено теоретическое исследование и анализ расчета U-критерия Манна-Уитни.
2. Был разработан программный метод расчета U-критерия Манна-Уитни.

Для решения этих задач были изучены следующие вопросы:

1. Основные понятия U-критерия Манна-Уитни.
2. Области применения U-критерия Манна-Уитни.
3. Анализ предметной области.
4. Постановка задачи курсового проекта.

В рамках первого раздела было представлено теоретическое описание расчета U-критерия Манна-Уитни. Были изучены ключевые понятия и области применения U-критерия, а также его этапы и основные аспекты. Также был рассмотрен пример задачи для расчета U-критерия Манна-Уитни.

Во втором разделе был проведен анализ предметной области, сформулирована задача и выбран подход к разработке.

В третьем разделе было обосновано выбор модели U-критерия Манна-Уитни. Затем был описан выбор используемых программных инструментов, аргументированы их существенные преимущества. Также был представлен алгоритм работы приложения, включающий его функции и возможности.

# СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ

1. Использование математики в статистике и основы математической статистики // российский индекс научного цитирования url: https://elibrary.ru/item.asp?Id=49536778 (дата обращения: 12.04.2023).

2. Статистика: теория статистики // российский индекс научного цитирования url: https://elibrary.ru/item.asp?Id=50379905 (дата обращения: 12.04.2023).

3. Наука о данных в R для программы Цифровых гуманитарных исследований url: https://agricolamz.github.io/DS\_for\_DH/ (дата обращения: 12.04.2023).

4. U-тест Манна-Уитни | кодкамп URL: https://www.codecamp.ru/blog/mann-whitney-u-test// (дата обращения: 12.04.2023).

5. Методы статистики URL: https://medstatistic.ru/methods/methods2.html (дата обращения: 12.04.2023).

6. Гайдышев И. Анализ и обработка данных. – Санкт-Петербург: Питер, 2001. – 752 с. (дата обращения: 12.04.2023).

7. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика, Основы моделирования и первичная обработка данных. – Москва: Финансы и статистика, 1983. – 471 с. (дата обращения: 12.04.2023).

8. Критерий Манна-Уитни — самый главный враг A/B-тестов url: https://habr.com/ru/companies/avito/articles/709596/ (дата обращения: 12.04.2023).

# ПРИЛОЖЕНИЕ А

import tkinter as tk

from tkinter import filedialog

import csv

from scipy.stats import mannwhitneyu

def program\_data():

welcome\_window = tk.Toplevel(window)

welcome\_window.title("Данные о программе")

label\_welcome = tk.Label(welcome\_window, text="Программа составлена студентом группы ПИЖ-б-о-21-1 Пуценко Иваном Алексеевичем")

label\_welcome.pack()

def instruction():

large\_text = """Добро пожаловать в приложение реализующее расчет U-критерия Манна-Уитни!

В меню "Файл" вы можете выбрать следующие опции:

"Данные о программе": Откроет новое окно с информацией о программе и ее авторе.

"Инструкция": Откроет новое окно с приветственным сообщением.

"Открыть CSV": Позволяет выбрать CSV-файл с данными для анализа.

После выбора опции "Открыть CSV" появится диалоговое окно, в котором вы можете выбрать CSV-файл с данными.

Выберите нужный файл и нажмите "Открыть".

После открытия файла, данные из первых двух строк будут загружены и отображены в текстовых полях "Данные 1 выборки" и "Данные 2 выборки".

Введите необходимые значения в текстовые поля или оставьте значения по умолчанию.

Нажмите кнопку "Рассчитать", чтобы выполнить критерий Манна-Уитни для введенных данных.

Результаты анализа будут отображены под кнопкой в виде статистики и p-значения.

Если p-значение больше 0.05, будет выведено сообщение, указывающее на отсутствие статистически значимой разницы между выборками.

Если p-значение меньше или равно 0.05, будет выведено сообщение о наличии статистически значимой разницы между выборками.

Для выхода из программы выберите опцию "Выход" в меню "Файл" или закройте окно программы."""

welcome\_window = tk.Toplevel(window)

welcome\_window.title("Инструкция")

label\_welcome = tk.Label(welcome\_window, text=large\_text)

label\_welcome.pack()

def open\_csv\_file():

file\_path = filedialog.askopenfilename(filetypes=[("CSV Files", "\*.csv")])

if file\_path:

with open(file\_path, "r") as file:

reader = csv.reader(file)

separator = ": "

data1 = "".join(next(reader)[0].split(separator)[1:]).split(' ')

data2 = "".join(next(reader)[0].split(separator)[1:]).split(' ')

entry\_data1.delete(0, tk.END)

entry\_data1.insert(0, ", ".join(data1))

entry\_data2.delete(0, tk.END)

entry\_data2.insert(0, ", ".join(data2))

def calculate\_mann\_whitney():

# Получение данных из текстовых полей

data1 = [float(x.strip()) for x in entry\_data1.get().split(",")]

data2 = [float(x.strip()) for x in entry\_data2.get().split(",")]

# Расчет критерия Манна-Уитни

statistic, p\_value = mannwhitneyu(data1, data2)

# Обновление метки с результатами

if p\_value > 0.05:

label\_result.config(

text="Статистика: {:.2f}, p-значение: {:.2f}\nМежду данными выборками нет статистически значимой разницы.".format(

statistic, p\_value))

else:

label\_result.config(

text="Статистика: {:.2f}, p-значение: {:.2f}\nМежду данными выборками есть статистически значимая разница.".format(

statistic, p\_value))

# Создание графического интерфейса приложения

window = tk.Tk()

window.title("Критерий Манна-Уитни")

# Создание меню

menu = tk.Menu(window)

window.config(menu=menu)

file\_menu = tk.Menu(menu)

menu.add\_cascade(label="Файл", menu=file\_menu)

file\_menu.add\_command(label="Данные о программе", command=program\_data)

file\_menu.add\_command(label="Инструкция", command=instruction)

file\_menu.add\_command(label="Открыть CSV", command=open\_csv\_file)

file\_menu.add\_command(label="Выход", command=window.quit)

# Создание меток и текстовых полей для ввода данных

label\_data1 = tk.Label(window, text="Данные 1 выборки:")

label\_data1.pack()

entry\_data1 = tk.Entry(window)

entry\_data1.pack()

label\_data2 = tk.Label(window, text="Данные 2 выборки:")

label\_data2.pack()

entry\_data2 = tk.Entry(window)

entry\_data2.pack()

# Кнопка для запуска расчета

button\_calculate = tk.Button(window, text="Рассчитать", command=calculate\_mann\_whitney)

button\_calculate.pack()

# Метка для вывода результатов

label\_result = tk.Label(window, text="")

label\_result.pack()

# Запуск цикла обработки событий

window.mainloop()

# ПРИЛОЖЕНИЕ Б

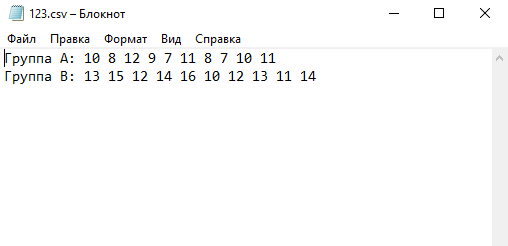


Рисунок Б.1 – Вид файла с входными данными

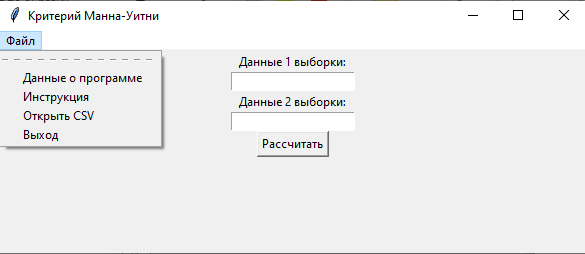


Рисунок Б.2 – Вид окна приложения после запуска

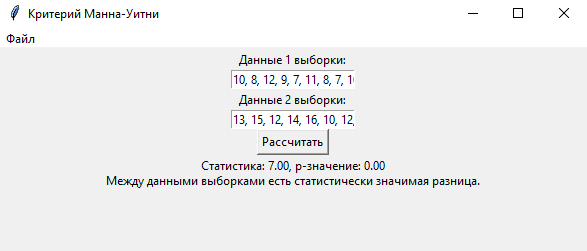


Рисунок Б.3 – Вид окна приложения после вычисления значений по входным данным