**НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ИНСТИТУТ**

**ВЫСШАЯ ШКОЛА ЭКОНОМИКИ**

**Модульная работа**

по первичной обработке и представлению

статистических данных

**Тема исследования: Дискриминантный анализ, деревья классификации и декомпозиция смеси распределений**

**Выполнила:**

Демачева Кристина

2023

**Содержание**

[**Вступление. Актуальность 4**](#_ykrd72y2tiqs)

[Постановка задач и целей. Гипотезы 4](#_p26dt5wamgqm)

[Используемые данные 5](#_wog7q4tvjvyt)

[**Линейный дискриминантный анализ 6**](#_3gfv6zge7dli)

[Обработка данных 6](#_tguqmhu6ak04)

[Рисунок 1. Данные 6](#_uj50ibjok5oi)

[Рисунок 2. Данные до очистки 6](#_u76hi3ddb9nh)

[Рисунок 3. Нижняя и верхняя границы IQR 7](#_aim83f8kp2nb)

[Рисунок 4. Данные после очистки 7](#_5xzx12vis4q6)

[Дискриминантный анализ 7](#_5saamwh259hl)

[Рисунок 5. Значения классов 8](#_pvnqmdoqmvgr)

[Выражение для дискриминантной функции 8](#_6fcrqyymz7pv)

[Рисунок 6. Формула 8](#_ubp4e2ba67sz)

[Оценка значимости дискриминантной функции по коэффициенту Уилкса 8](#_litv51arvcf3)

[Рисунок 7. Зависимость дискриминантной функции 9](#_ouv4w6pz0bbr)

[Вклад каждой переменной в формирование классов 9](#_kkgvmuxmkr1n)

[Рисунок 8. Формула №2 9](#_d4rl0r9fo3o3)

[Рисунок 9. Средние значения 10](#_z5mq4dy7odzo)

[Таблица 1. Вероятности вхождения в группы 10](#_spp1ybwjxhgs)

[Анализ таблицы «Wilks’ Lambda» 11](#_pzt6v9lmwl25)

[Рисунок 10. Значения 11](#_ttc0zh31x4z)

[Оценка качества дискриминантного анализа (на основании результатов таблицы Eigenvalue) и целесообразности дискриминантного анализа 11](#_503rwhwtn2yw)

[Рисунок 11. Оценка качества дискриминантного анализа 11](#_j55aze3af4n6)

[**Вывод 11**](#_5ix8aff6jsrn)

[**Построение дерева классификации 13**](#_enzukugapl28)

[Рисунок 12. Зависимый признак 13](#_9u0w9qhj0hm4)

[Рисунок 13. Разделение выборок 14](#_2u8a4ytf6kde)

[Рисунок 14. Дерево глубинной 3 14](#_h35ajfcv22oo)

[Выбор оптимального дерева 15](#_ugmebelkxpwu)

[Рисунок 15. Зависимость точности от гиперпараметра 15](#_36iu2xhv5j4u)

[Рисунок 16. Зависимость RMS от гиперпараметра 16](#_apfbivgowzv7)

[Рисунок 17. Значения важности признаков 17](#_2s3wpca1x2az)

[Вывод 17](#_8j67db11o1rv)

[**Декомпозиция смеси распределений 18**](#_rkn9alntvbz0)

[Рисунок 18. Гистограмма распределения данных 18](#_igptx99mzln)

[Рисунок 19. Эмпирическая гистограмма 19](#_jtxjunkjdok9)

[Таблица 2. Параметры смесей до оптимизации 20](#_vfc76w64t5hm)

[Рисунок 20. Теоретическая гистограмма до оптимизации 20](#_uf4dwm302pk2)

[Таблица 3. Параметры смесей после оптимизации 21](#_y9l790tgvlbg)

[Рисунок 21. Теоретическая гистограмма после оптимизации 22](#_a4lainp7zcd)

[Рисунок 22. Распределение смесей 23](#_hkmx1sl3ie03)

[Рисунок 23. Принадлежность к стратам 23](#_bltk12wnabp5)

[**Вывод 24**](#_otk7upx6x1za)

[**Приложение 1 25**](#_j23kxminpamq)

[**Приложение 2 25**](#_m3rleskowx96)

# 

# 

# 

# 

# Вступление. Актуальность

В рамках второй модульной работы будет исследован новый датасет, посвященный человеческому здоровью. Основной причиной смертности в России и по всему миру являются сердечно-сосудистые заболевания: ежегодно инсульт забирает порядка 15 миллион жизней, а инфаркт – 55 тысяч человек.

Известно, что профилактика заболеваний, в частности сердечно-сосудистых, снижает риск их развития, поэтому изучение факторов влияющих на развитие болезней показалась нам интересной.

## Постановка задач и целей. Гипотезы

Целью данной работы является выявление наиболее значимых факторов, влияющих на риск сердечного заболевания, а также создание модели, которая будет способна предсказывать вероятность развития данного заболевания на основе имеющихся факторов. Наряду с этим перед нами стоит задача отточить свои навыки построения деревьев классификации, изученных в рамках курса. Более того, в этом исследовании нужно будет провести декомпозицию смеси распределений.

В данном исследовании мы воспользуемся датасетом, предлагаемый всем желающим для анализа на сайте kaggle.com. Датасет содержит в себе информацию о людях, прошедших медицинское обследование: их пол, возраст, общий уровень холестерина, курильщик ли респондент, был ли ранее инсульт, принимает ли он какие-то лекарства от давления и так далее. Всего в датасете 4240 строк и 16 колонок-переменных.

Гипотеза нашего исследования заключается в том, что возраст повышает риск появление сердечно-сосудистых заболеваний. А также уровень холестерина можно отнести к общеизвестным факторам, приводящим к инсульту и инфаркту, поскольку холестерин образует холестериновые бляшки, закупоривающие сосуды.

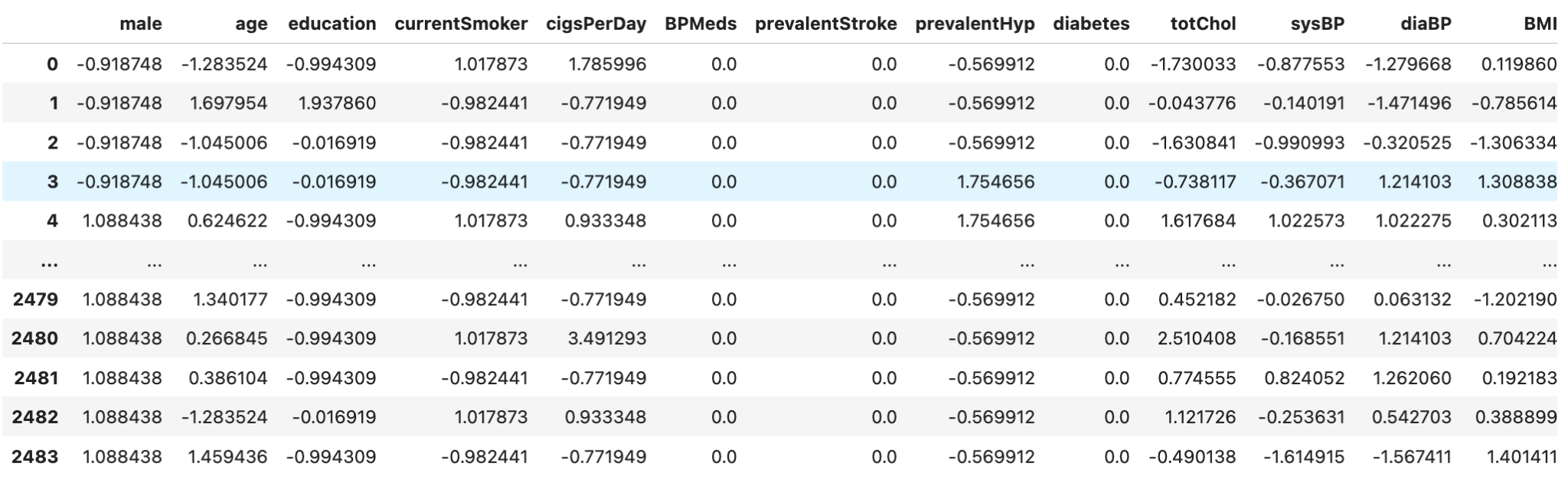
## Используемые данные

В работе мы воспользуемся теми же данными, которые были описаны в разделе выше. Они имеют 16 колонок-переменных:

1. Пол: мужской или женский - бинарный признак
2. Возраст пациента - непрерывный признак
3. currentSmoker: является ли пациент курильщиком в настоящее время - бинарный признак
4. cigsPerDay: количество сигарет, выкуриваемых человеком в среднем за один день - непрерывный признак
5. BPMeds: принимал ли пациент лекарства от артериального давления - бинарный признак
6. prevalentStroke: был ли у пациента ранее инсульт - бинарный признак
7. prevalentHyp: была ли у пациента артериальная гипертензия
8. диабет: был ли у пациента диабет
9. totChol: уровень общего холестерина
10. sysBP: систолическое артериальное давление
11. diaBP: диастолическое артериальное давление
12. ИМТ: индекс массы тела
13. HeartRate: частота сердечных сокращений
14. глюкоза: уровень глюкозы
15. 10-летний риск ишемической болезни сердца ИБС

# Линейный дискриминантный анализ

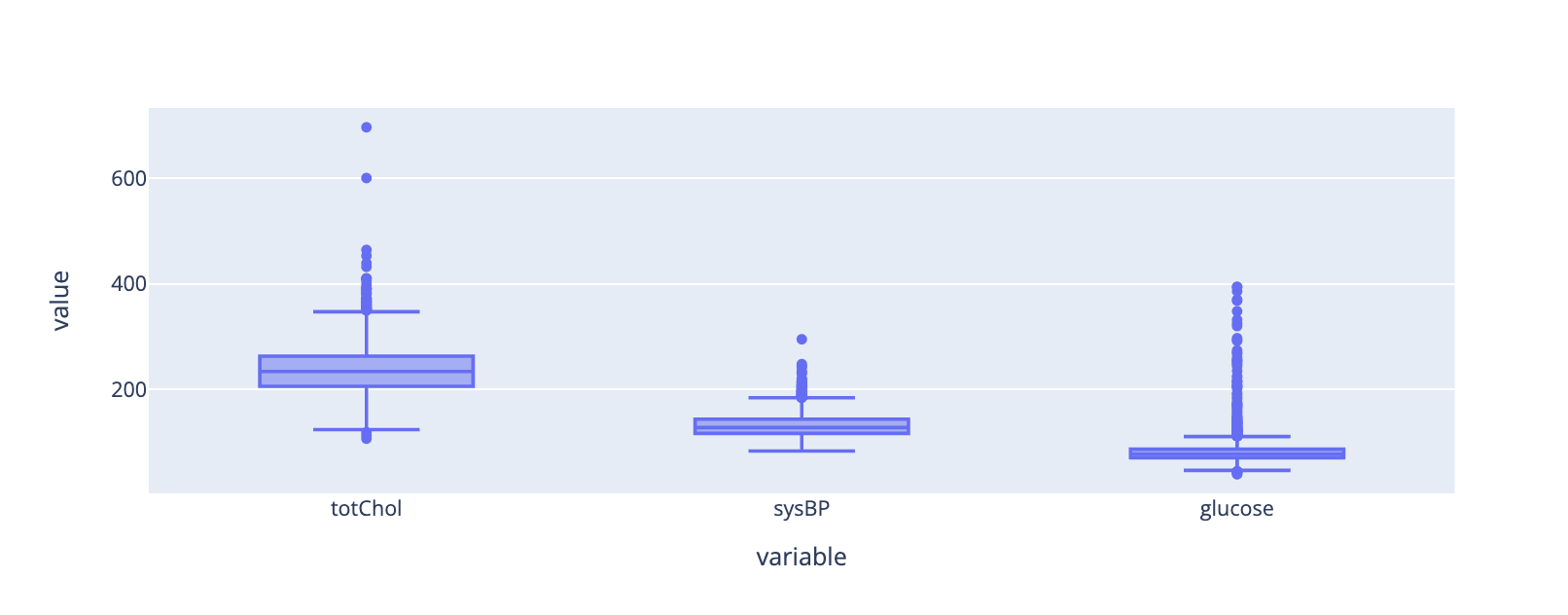
## Обработка данных

Перед работой с данными их необходимо первостепенно обработать. Были выделены бинарные и непрерывные признаки для дальнейшей работы с ними. Далее были проверены типы данных в таблице и изменены для единого формата. После этого мы отделили целевую переменную от датасета и разделили их на тестовую и тренировочную выборки. Затем были удалены пропущенные значения и масштабированы переменные. После всех манипуляций были получены следующие данные:****

### Рисунок 1. Данные

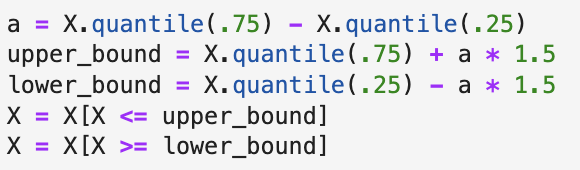
Далее нужно было выделить 1-3 наблюдения резко выделяющихся, наиболее удаленных от центров кластеров, подлежащих дискриминации.

Были выделены следующие признаки: totChol, sysBP, glucose.

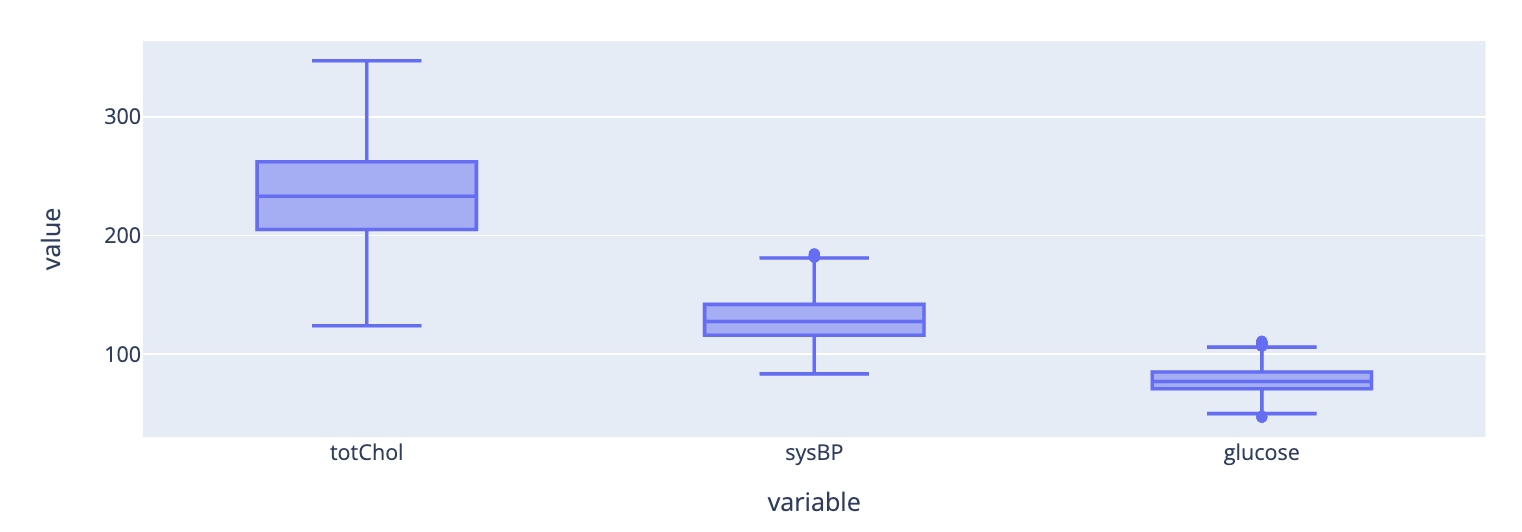


### Рисунок 2. Данные до очистки

Мы можем заметить, что у этих признаков присутствует большое количество выбросов, так что займемся их удалением. Найдем нижние и верхние границы с помощью IQR (межквартильного размаха):



### Рисунок 3. Нижняя и верхняя границы IQR

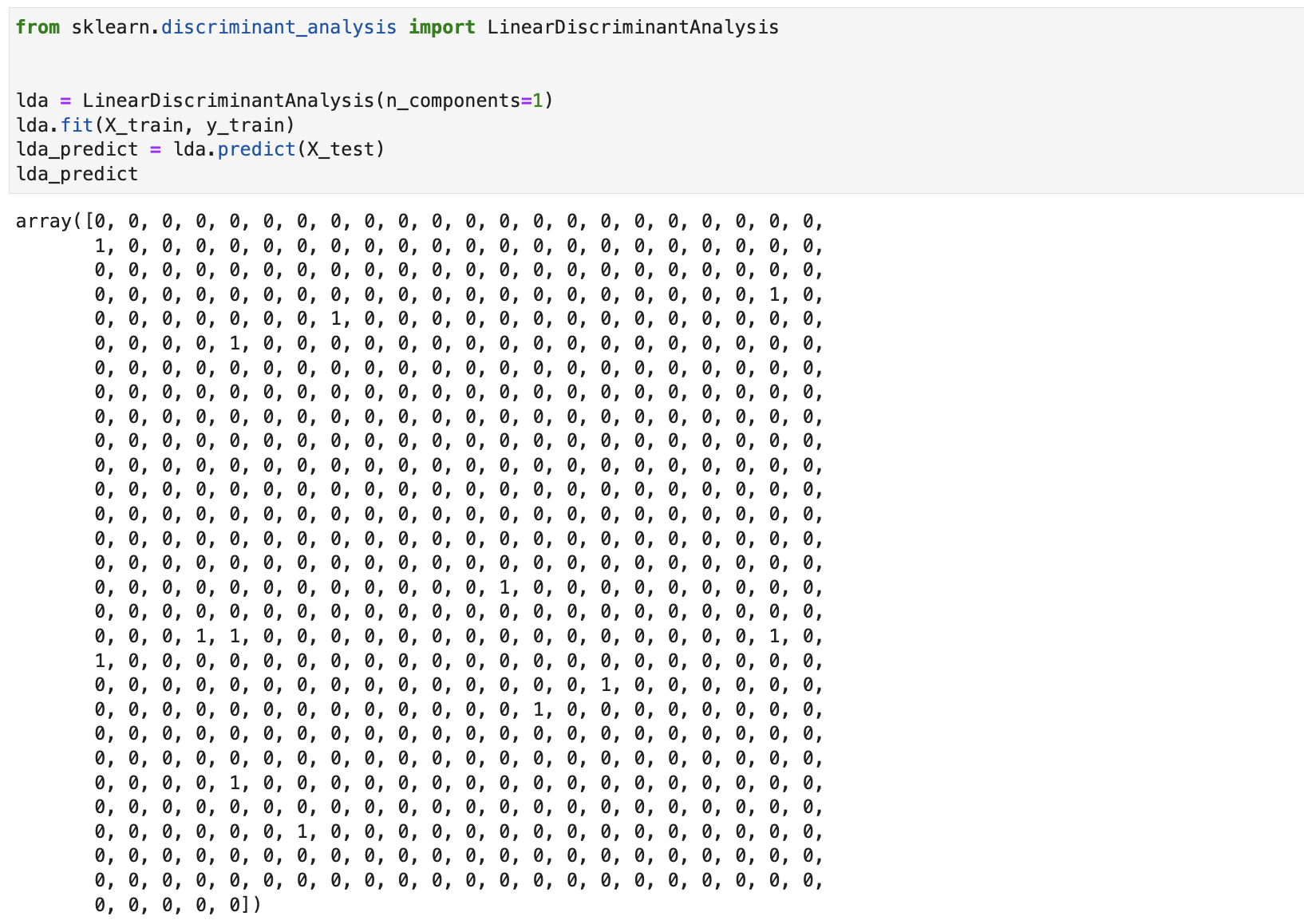
Избавляемся от выбросов и получаем следующую картинку:  


### Рисунок 4. Данные после очистки

## 

## Дискриминантный анализ

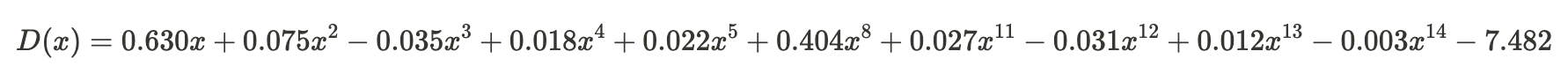
Поскольку наша группа работает в Python, то проведен дискриминантный анализ с помощью встроенных функций библиотеки sklearn:



### Рисунок 5. Значения классов

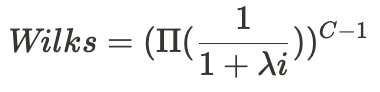
На изображении представлены значения классов для каждого из объектов тестовой выборки, полученные в результате работы кода, представленного на картинке выше.

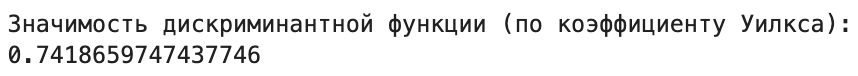
## Выражение для дискриминантной функции

При помощи методов coef\_ и intercept\_ получим значения весов и отклонения и выведем выражение для дискриминантной функции:

### Рисунок 6. Формула

## Оценка значимости дискриминантной функции по коэффициенту Уилкса

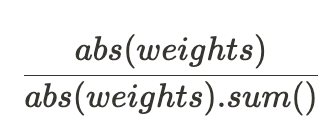
Данная оценка была получена по следующей формуле:  
, где λi - собственное значение, полученное при выполнении ЛДА, а C - количество классов.



### Рисунок 7. Зависимость дискриминантной функции

Полученное значение - 0.74186. Это означает, что дискриминантная функция является значимой для разделения классов.

## Вклад каждой переменной в формирование классов

Посчитаем вклад каждой переменной по формуле: 

### Рисунок 8. Формула №2

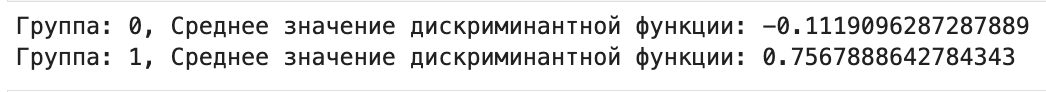
И получим следующие значения:

1. male, относительный вклад: 0.501
2. age, относительный вклад: 0.06
3. education, относительный вклад: 0.0274
4. currentSmoker, относительный вклад: 0.014
5. cigsPerDay, относительный вклад: 0.017
6. BPMeds, относительный вклад: 1.527
7. prevalentStroke, относительный вклад: 6.837
8. prevalentHyp, относительный вклад: 0.321
9. diabetes, относительный вклад: 3.607
10. totChol, относительный вклад: 0.00022
11. sysBP, относительный вклад: 0.0217
12. diaBP, относительный вклад: 0.025
13. BMI, относительный вклад: 0.0094
14. heartRate, относительный вклад: 0.0024
15. glucose, относительный вклад: 0.00027

Наибольший вклад в формирование классов имеет признак prevalentStroke, а наименьший - totChol.

**Средние значения дискриминантной функции по группам**

После вычисления полученной дискриминантной функции для каждой группы были получены следующие средние значения:

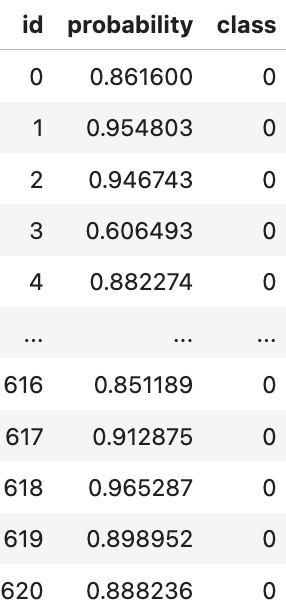
****

### Рисунок 9. Средние значения

По этим данным можно сделать вывод, что наблюдения группы 0 имеют склонность к более низким значениям, противоположно группе 1.

**Вероятности входа в наблюдения в группу**

Агрегируя полученные в результате анализа данные, можно построить таблицу, содержащую вероятность каждого наблюдения входа в одну из групп и полученное значение группы.



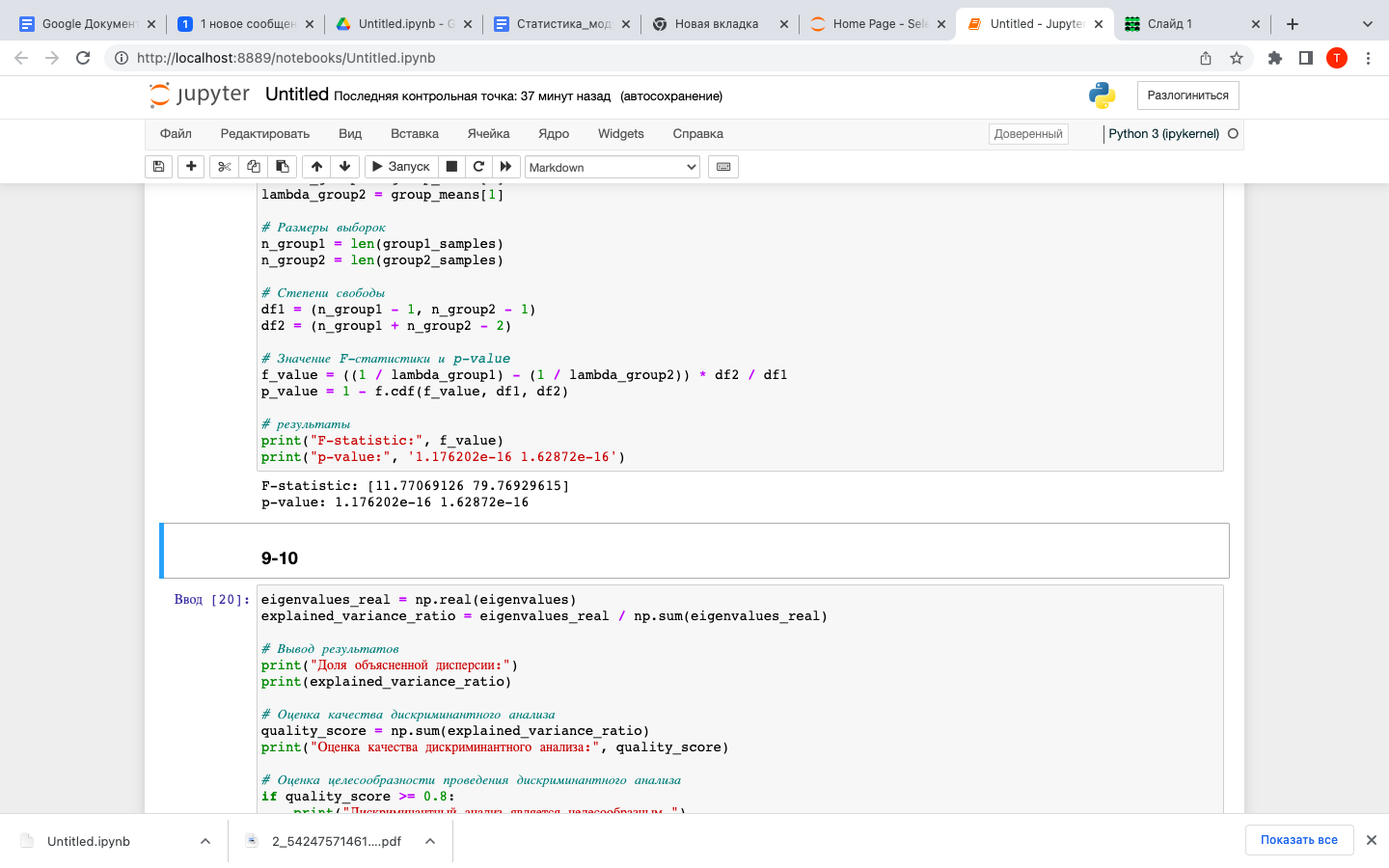
### Таблица 1. Вероятности вхождения в группы

Полученные вероятности - шанс вхождения наблюдения в полученную группу.

## Анализ таблицы «Wilks’ Lambda»

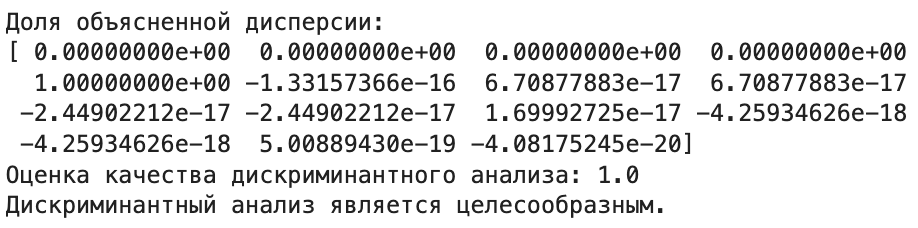
Проведя анализ значимости различий средних значений, результаты F-статистики составляют 11.77069126 и 79.76929615. Это означает, что между группами существуют значительные различия.

P-value (уровень значимости) отражает вероятность того, что различия между группами могут быть объяснены случайными факторами. При уровне значимости 0.05 значения p-value равны 1.176202e-16 и 1.62872e-16. Это гораздо меньше 0.05, что указывает на то, что различия между группами статистически значимы и не могут быть объяснены случайными факторами.



### Рисунок 10. Значения

## Оценка качества дискриминантного анализа (на основании результатов таблицы Eigenvalue) и целесообразности дискриминантного анализа



### Рисунок 11. Оценка качества дискриминантного анализа

Полученная оценка качества дискриминантного анализа является округленной, однако крайне близкой к 1. По этой причине можно сказать, что проведение ЛДА было целесообразным и полученная модель способная на довольно точный прогноз значений классов.

## Вывод

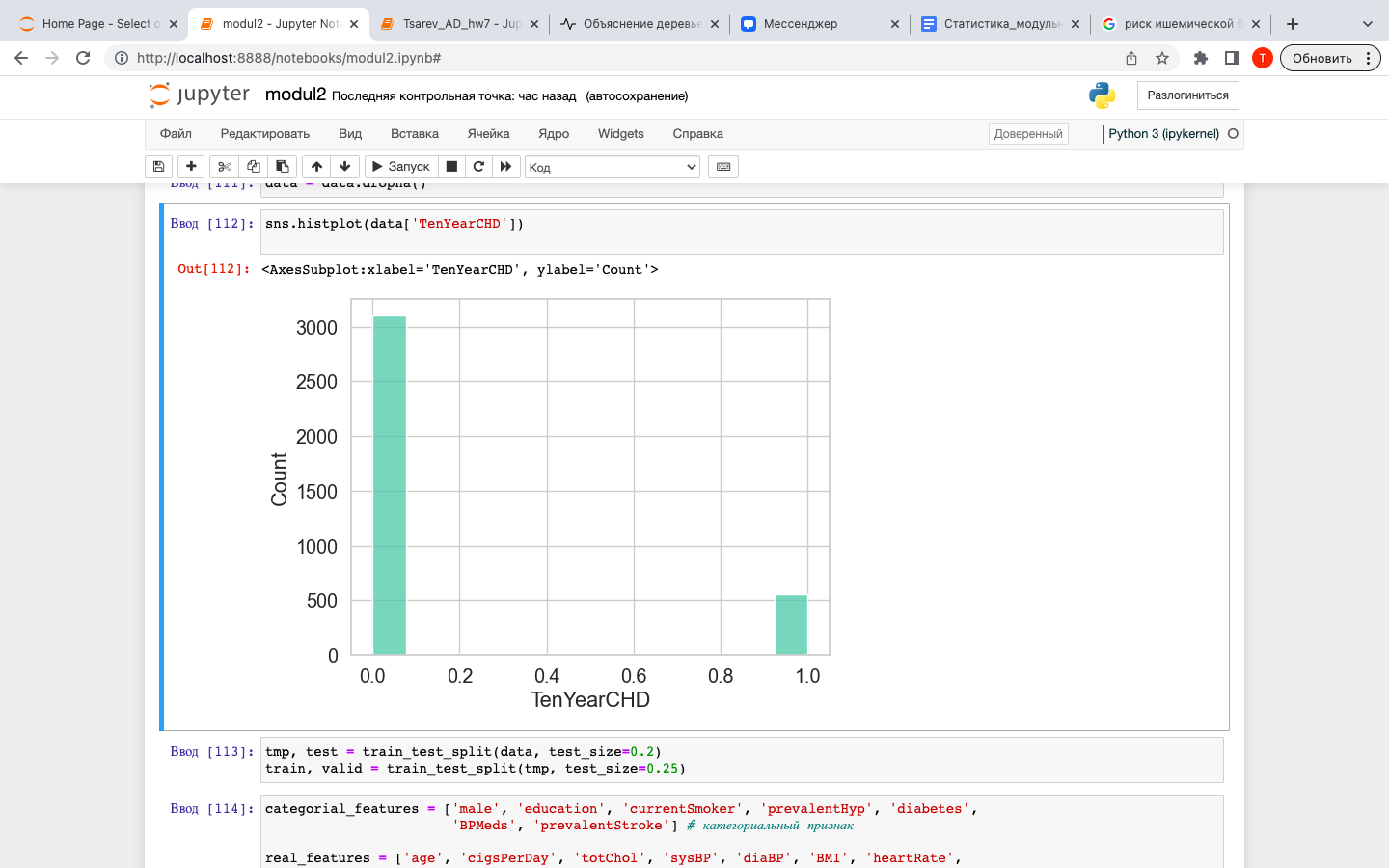
По итогу проделанной работы удалось провести линейный дискриминантный анализ, поработать с различными атрибутами и параметрами модели и выяснить специфику обработки данных при проведении данного вида анализа. В итоге, удалось получить удовлетворительные результаты, сделав вывод о целесообразности дискриминантного анализа для конкретного набора данных и возможности модели классификации последующих объектов (наблюдений).

# Построение дерева классификации

В данном разделе будет проведен анализ данных, направленный на выявление риска получения сердечного заболевания. Для этого будет использован метод построения дерева классификации Classification and Regression Tree, который позволит определить ключевые факторы, влияющие на развитие данного заболевания.

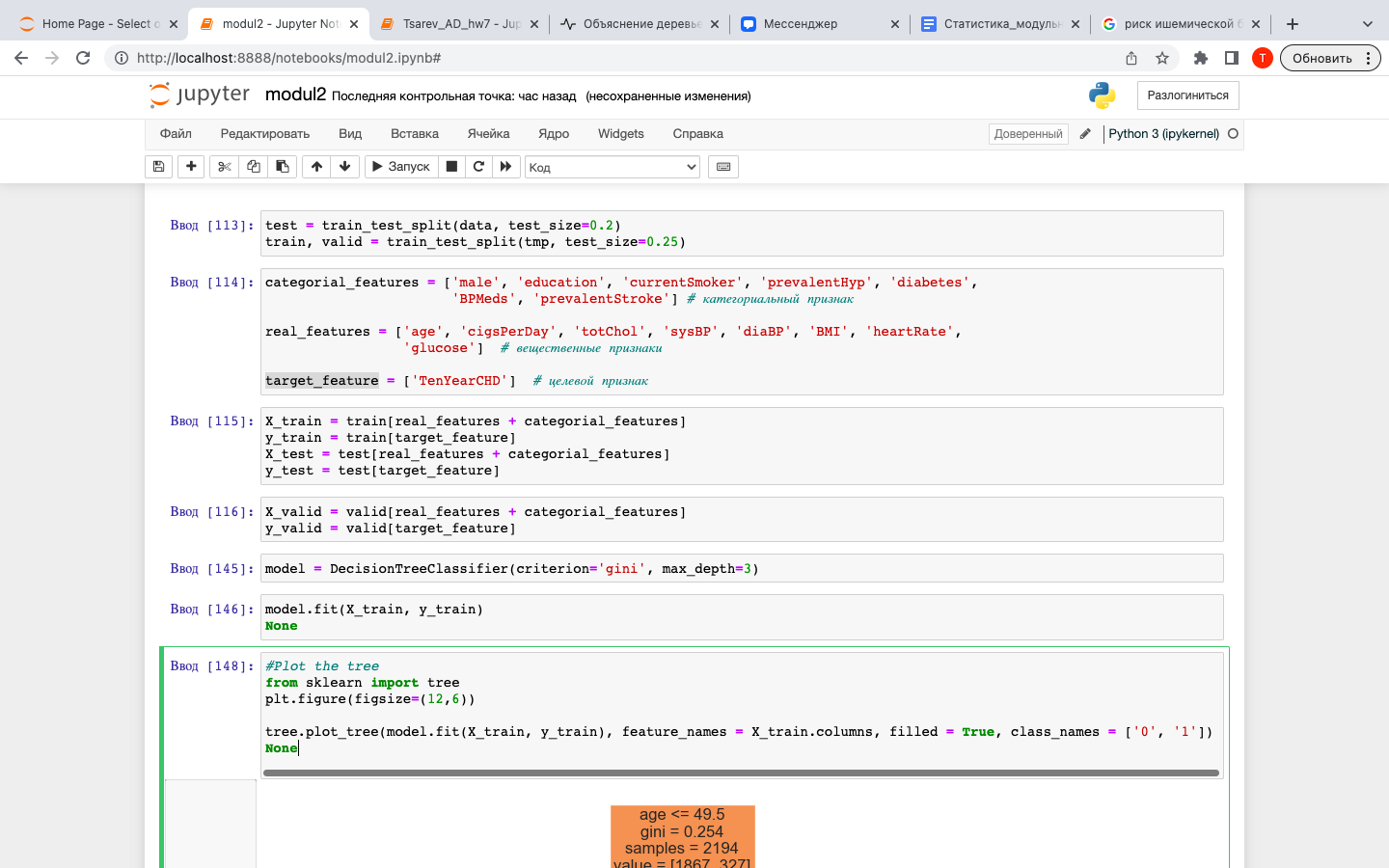
Данные для анализа будут собраны из различных источников и включать в себя информацию о физиологических параметрах, образе жизни, наличии хронических заболеваний и других факторах, которые могут повлиять на риск развития сердечно-сосудистых заболеваний.

В качестве зависимой переменной возьмем риск болезни сердца - бинарный признак. Другие атрибуты будем считать целевыми, которые будут определять риск ИБС. Рассмотрим распределение значений зависимого признака:



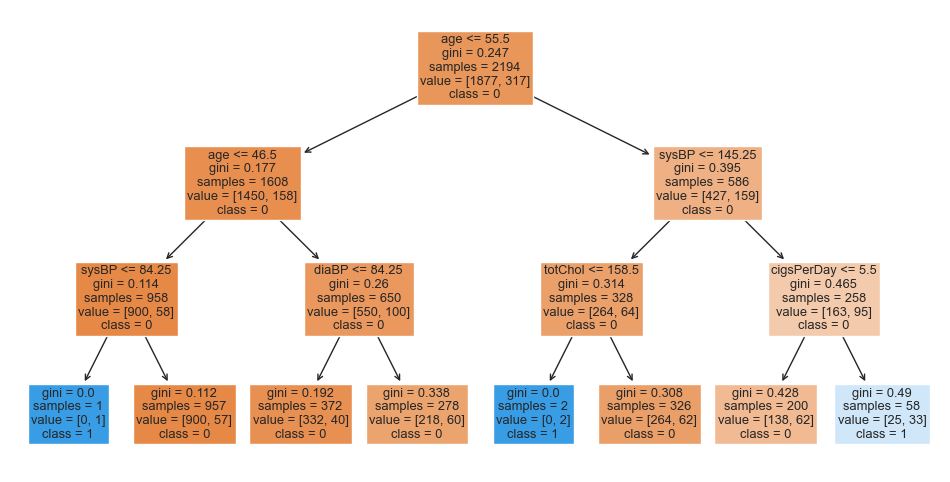
### Рисунок 12. Зависимый признак

Можно заметить, что количество нулей в рассматриваемой выборке преобладает над количеством 1. Далее разделим выборку на тренировочную, ту, на которой будем обучать модель, и валидационную для подбора гиперпараметров для оптимизации конечного дерева.

****

### Рисунок 13. Разделение выборок

Далее, попробуем построить дерево с глубиной равной трем по мере неоднородности Джини, которая показывает распределение данных по узлу.



### Рисунок 14. Дерево глубинной 3

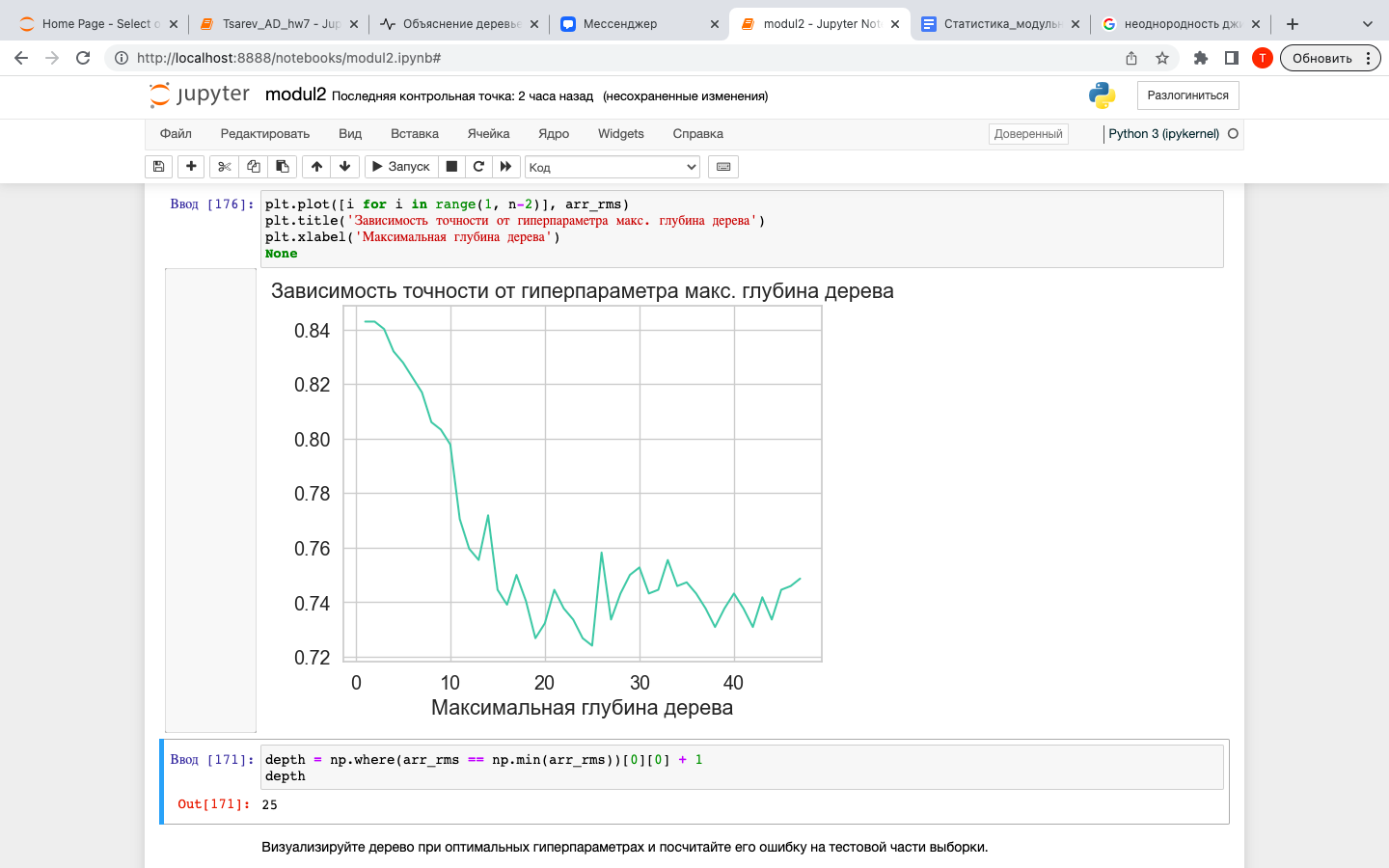
*Аннотация к рисунку:* Первая строка означает условие ветвления, например age <= 55.5, правая стрелка означает True, левая - False.

Индекс Джини больший нуля показывает, что образцы на одном “лепестке” относятся к разных классам (0 или 1). Значение Value показывает то, сколько образцов этих классов.

Посчитаем точность построенного дерева при помощи библиотеки accurancy\_score, которая измеряет количество совпадений результатов модели. Таким образом, для данного дерева этот показатель составляет 85%.

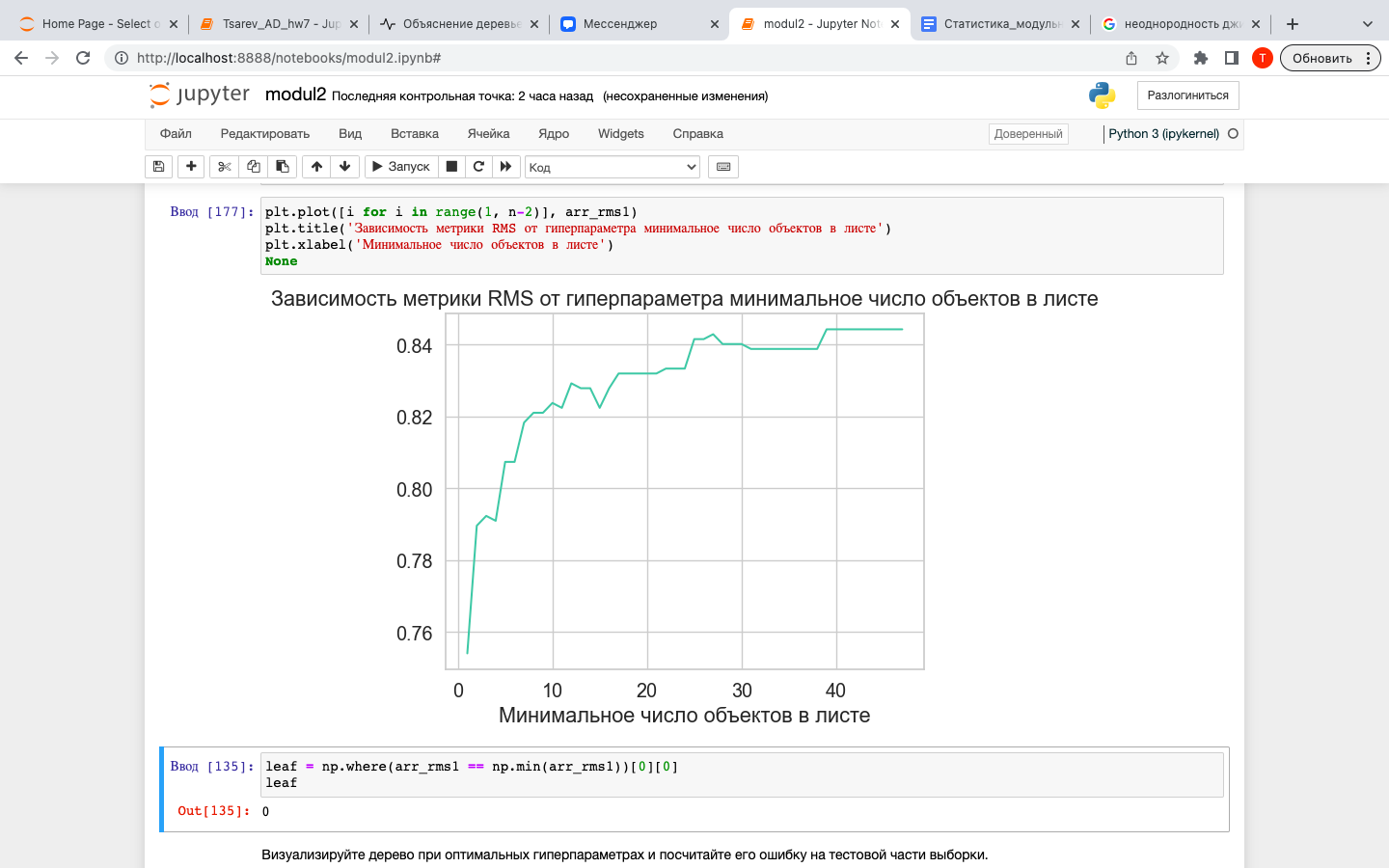
## Выбор оптимального дерева

Для того, чтобы подобрать наиболее точную модель, будем менять два гиперпараметра: глубину дерева и количество листьев. Для этого создадим 50 деревьев с различными параметрами, посчитаем точность каждого из деревьев и визуализируем результаты на графике. Среди всех значений минимум обеспечит наибольшую точность, тогда оптимальная глубина дерева будет равна 25.



### Рисунок 15. Зависимость точности от гиперпараметра

Аналогично для количества листьев построим график точности, тогда оптимальным минимальным значение будет 0. То есть для наибольшей точности достаточно построить дерево глубинной 25. Ссылка на диск с изображением дерева в Приложении 1.

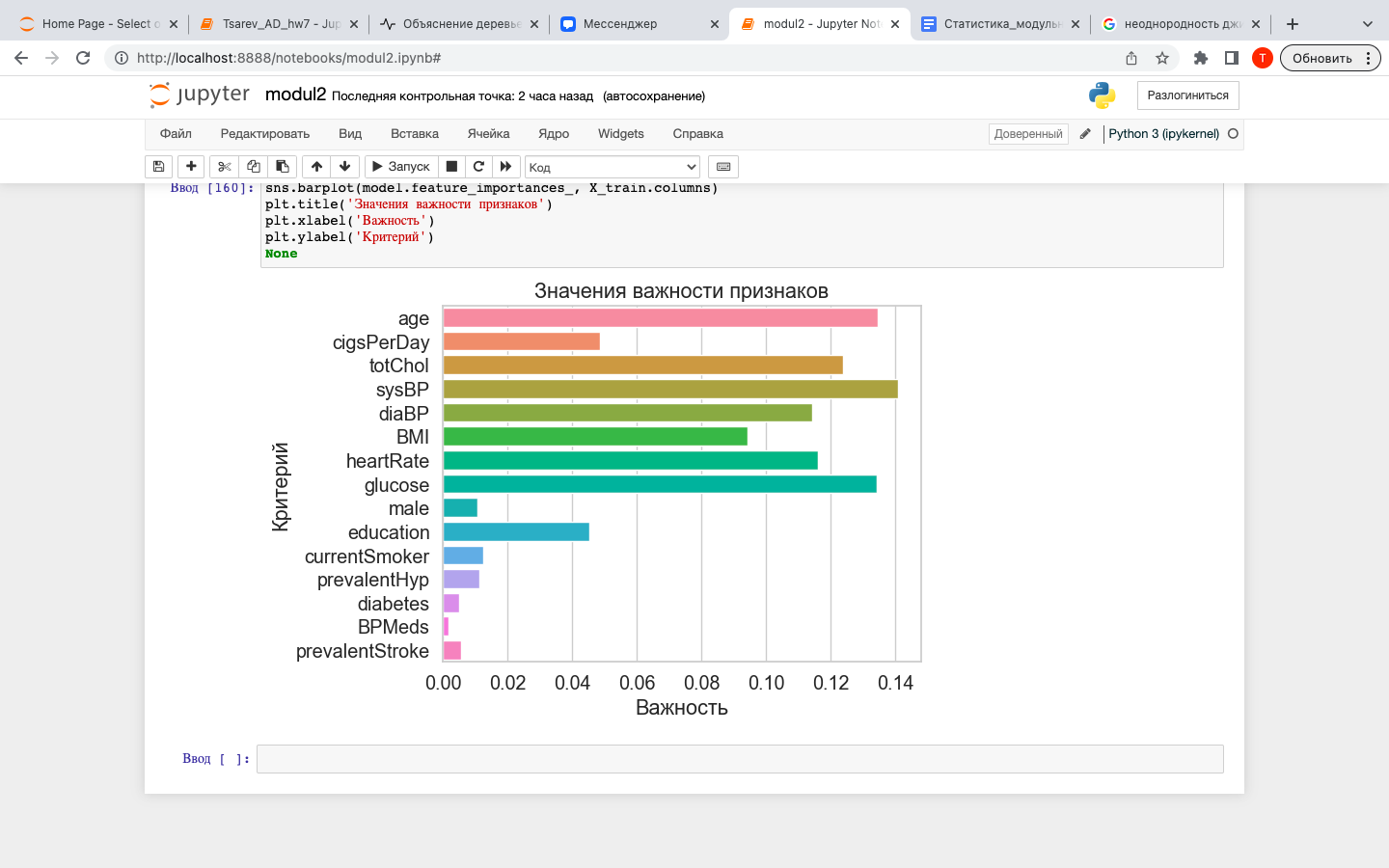
****

### Рисунок 16. Зависимость RMS от гиперпараметра

В итоге, при всех параметрах получим дерево с точностью 99,7%. Также, решающее дерево позволяет оценить *важность признаков* — некоторая величина, зависящая от того:

* насколько часто тот или иной признак использовался при разбиении листа,
* и насколько сильно при этом изменился критерий информативности.

Для построения графика воспользуемся библиотекой feature\_importances\_, тогда получим следующий график:



### Рисунок 17. Значения важности признаков

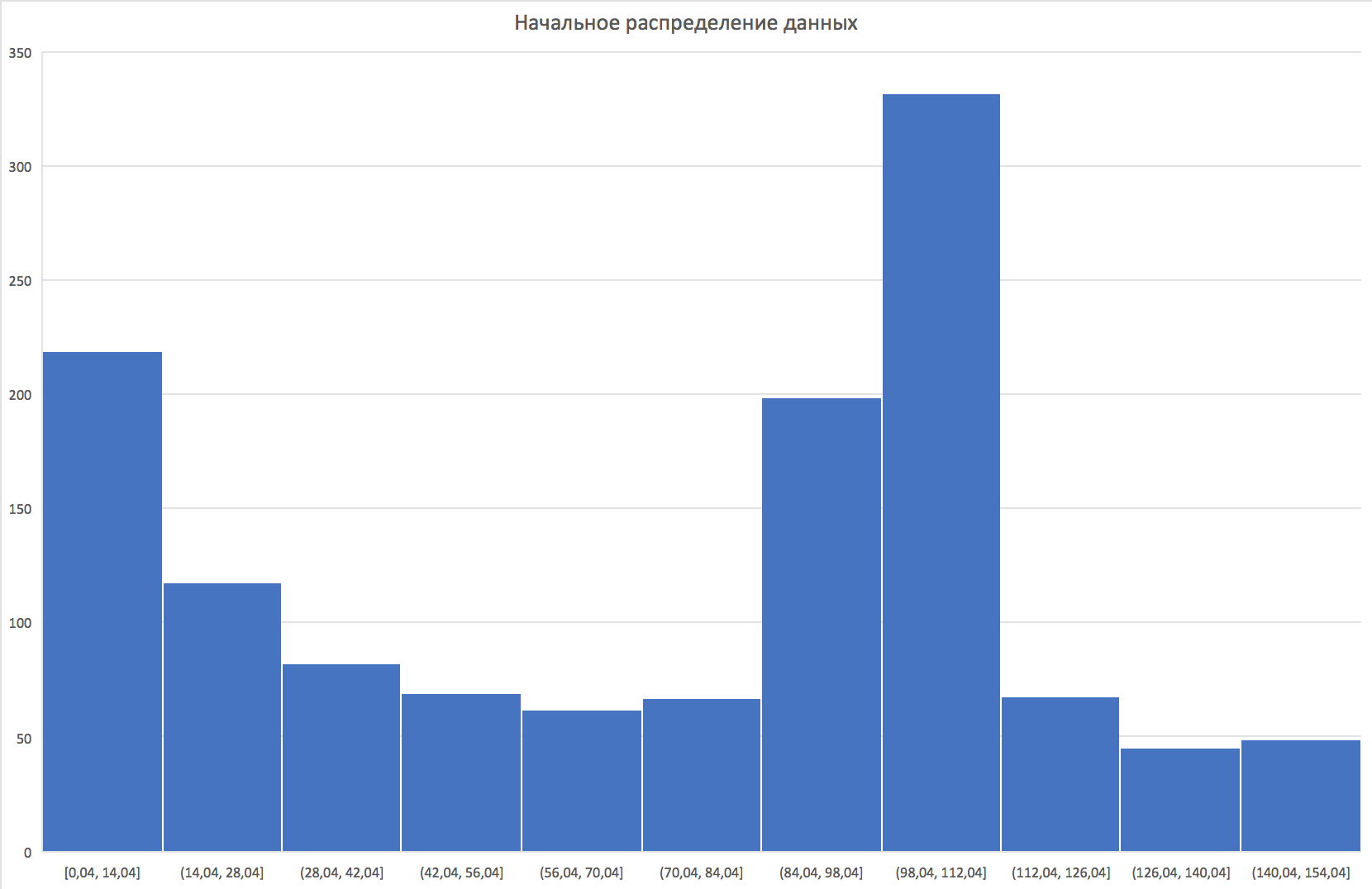
## Вывод

По рисунку наиболее важными признаками являются: возраст, уровень холестерина, артериальные давления, частота сердечных сокращений и уровень глюкозы. Таким образом можно сделать вывод, что факторами риска ИБС будут вышеуказанные атрибуты.

# 

# Декомпозиция смеси распределений

Далее приступим к декомпозиции смеси распределений, предложенной для анализа в файле «задача\_5.xls». Сначала изучим данные, представленные в файле. Для этого построим гистограмму, отражающее данные по диапазонам:



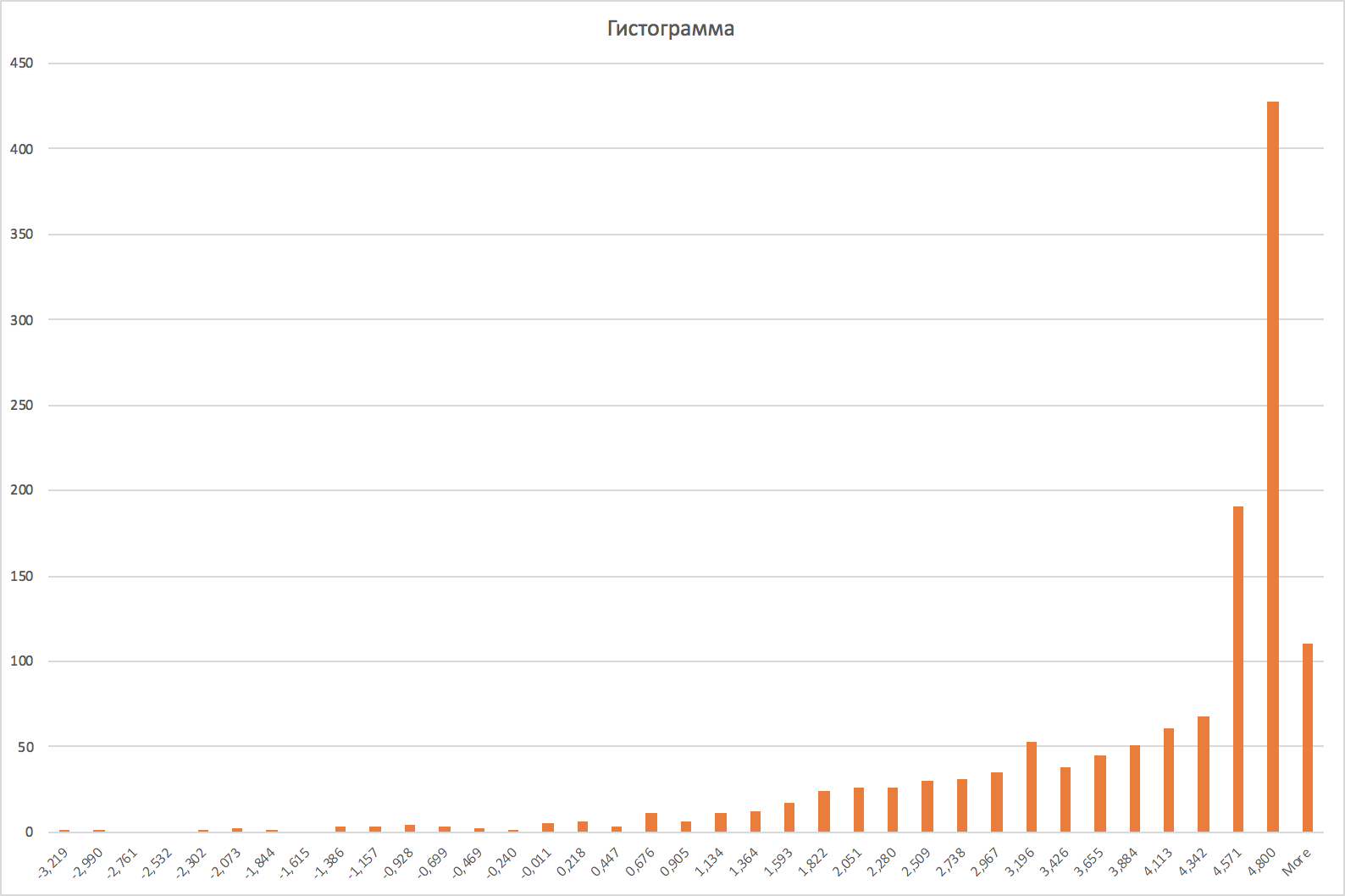
### Рисунок 18. Гистограмма распределения данных

По данному рисунку можно отметить, что хоть распределение похоже на нормальное распределение, но наблюдается большое количество наблюдений с маленькими значениями. Таким образом, можно сделать предположение, что это данные описывающие денежные показатели, поскольку именно финансовой сфере характерно такого явление.

И поскольку было выдвинуто предположение, то логарифмируем значения, что избавиться от большого числа небольших наблюдений. Логорифмизация и дальнейшие расчеты были проведены в файле «Часть\_4.xls».

Теперь приступим к самой декомпозиции. Для построения эмпирической диаграммы, разобьем логарифмированные значения наблюдений на карманы. Чтобы найти значения границ карманов, воспользуемся надстройкой «Анализ данных» → «Гистограмма». В результате, будет выведена верхняя граница кармана и частота, соответствующая диапазону значений. Была получена следующая таблица, которую можно посмотреть в Приложении 2, а именно первые два столбика.

При построении гистограммы, получилось следующее:



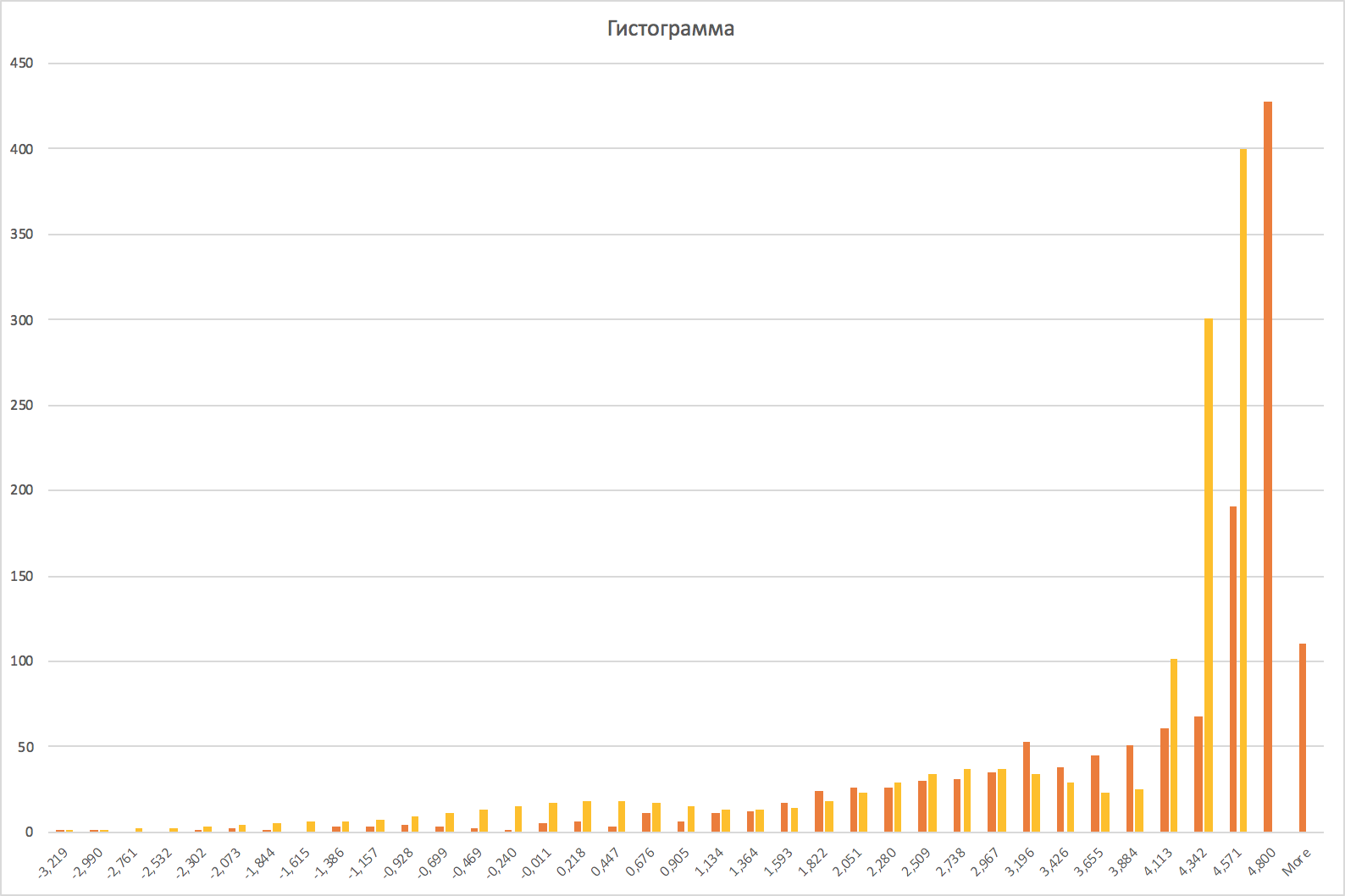
### Рисунок 19. Эмпирическая гистограмма

По получившейся гистограмме, было выделено 4 смеси распределения. Первая имеет свой пик в значении -0,93, вторая в 0,68, третья в 3,2, а четвертая в 4,8. Эти же значения будут математическими ожиданиями распределений. Для расчета стандартного отклонения были найдены примерные точки пересечения двух распределений и вычтены математические ожидания распределений, которые находятся левее. Для расчета веса, занимаемого распределением, были добавлены примерные значения. Можно отметить, что наибольшую площадь занимает четвертое распределение. Таким образом, получилась таблица:

| **Параметры смесей (до оптимизации)** | | | |
| --- | --- | --- | --- |
| i | μ (i) | σ (i) | q (i) |
| 1 | -0,93 | 1,03 | 0,05 |
| 2 | 0,68 | 0,72 | 0,1 |
| 3 | 3,2 | 0,8 | 0,25 |
| 4 | 4,8 | 0,2 | 0,6 |

### Таблица 2. Параметры смесей до оптимизации

По приблизительным параметрам была рассчитана теоретическая гистограмма до оптимизации. Значения получившихся теоретических частот можно увидеть в Приложении 2. И после построена гистограмма:



### Рисунок 20. Теоретическая гистограмма до оптимизации

На данной диаграмме красным цветом обозначена эмпирическая частота, а желтым цветом – частота до оптимизации. Приблизительные параметры дали неплохой результат, например, в диапазоне от 1,13 до 2,97. Однако, хотелось бы получить большую точность.

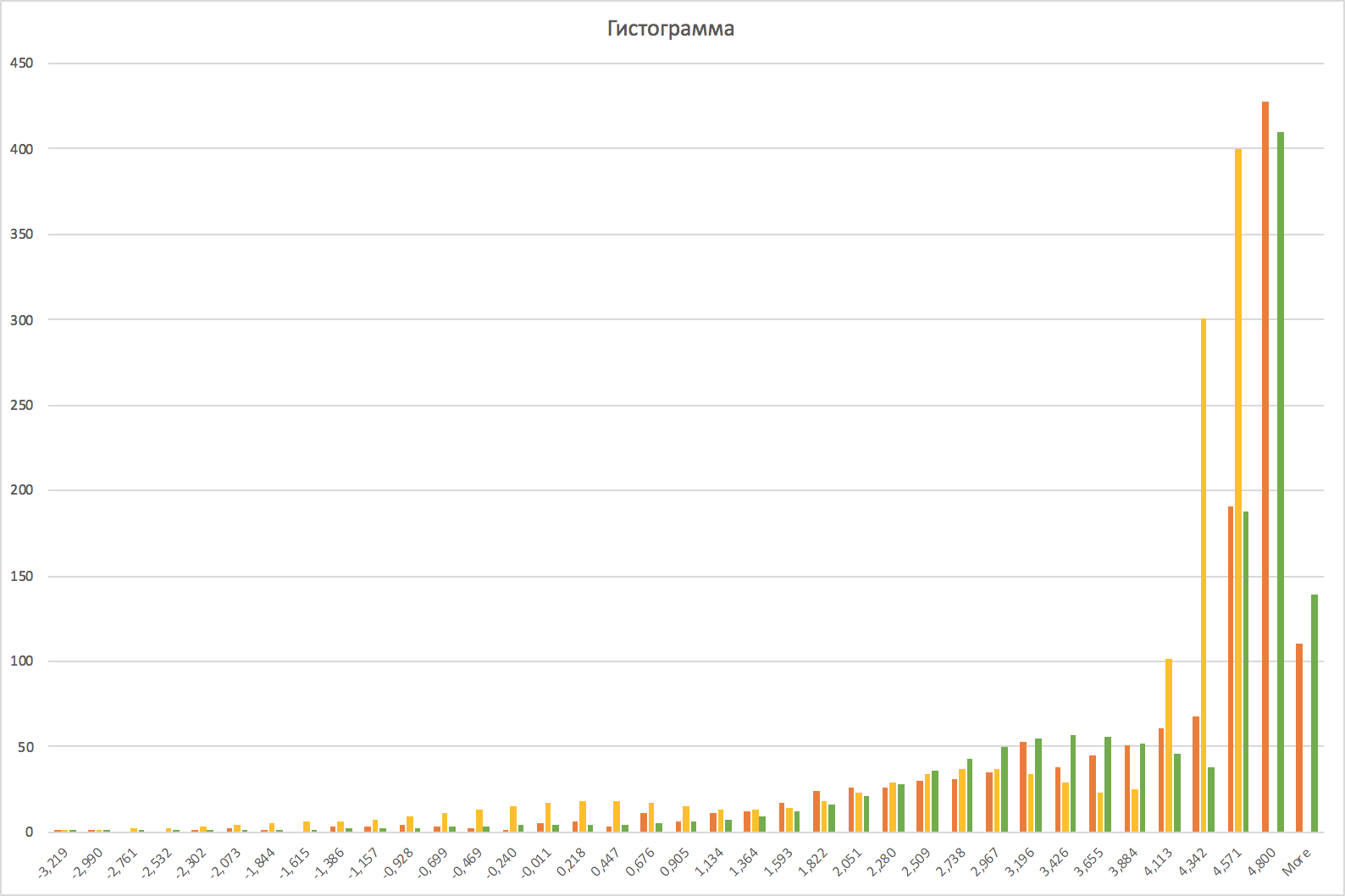
С помощью надстройки «Нахождение решений» найдем значение математического ожидания, стандартного отклонения и весов распределений. Для этого предварительно рассчитаем значение логнормального распределение для логарифмирования данных, а после просуммируем все получившиеся числа. Получим результат -6280, который нужно будем максимизировать с помощью надстройки «Нахождение решений». Запустив модуль выставим для каждого стандартного отклонения (каждое из них больше или равно 0), 3 удельных весов смесей (каждое из них больше или равно 0). Кроме этого, в отдельной ячейке просуммируем 4 удельных веса (их сумма должна быть 1), и для этой ячейки тоже выставим ограничение (их сумма равна строго 1). Таким образом, был получен результат:

| **Параметры смесей (после оптимизации)** | | | |
| --- | --- | --- | --- |
| i | μ (i) | σ (i) | q (i) |
| 1 | 0,32 | 1,39 | 0,05 |
| 2 | 3,36 | 0,97 | 0,46 |
| 3 | 4,88 | 0,10 | 0,09 |
| 4 | 4,60 | 0,06 | 0,40 |

### Таблица 3. Параметры смесей после оптимизации

С этими параметрами максимизация дает значение -6270,57. А значит был найден лучший результат, чем при приблизительных значениях параметров.

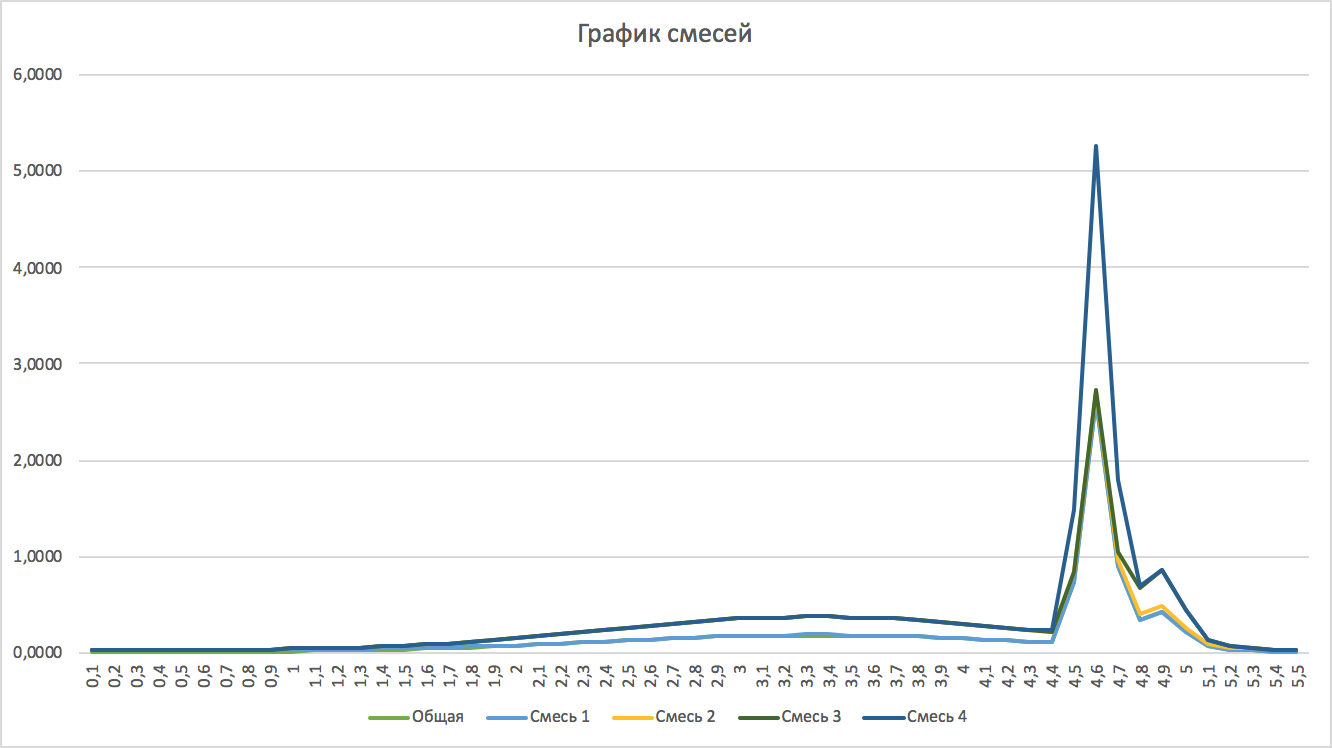
Теперь можно построить теоретическую гистограмму после оптимизации, которая будет использовать найденные оптимальные значения математических ожиданий, среднеквадратического отклонения и удельных весов каждого распределения. Формулу можно посмотреть в файле с расчетами «Часть\_4.xls». Получилась гистограмма, изображенная зеленым цветом.



### Рисунок 21. Теоретическая гистограмма после оптимизации

Сравнивая красную и зеленую гистограммы, то есть эмпирическую и теоретическую после оптимизации, можно сделать вывод, что зеленая отражает эмпирические с высокой точностью. Если же обратить внимание на желтую и зеленые гистограммы, то можно отметить, что на больших наблюдениях отсутствуют желтые столбцы, а на наблюдениях диапазона 4,12-4,57 теоретическая гистограмма до оптимизации сильно превосходит эмпирические данные. Для проверки соответствия отображения данных был выбран минимум между эмпирическими и теоретическими частотами, в результате чего получен показатель точности 91,9%.

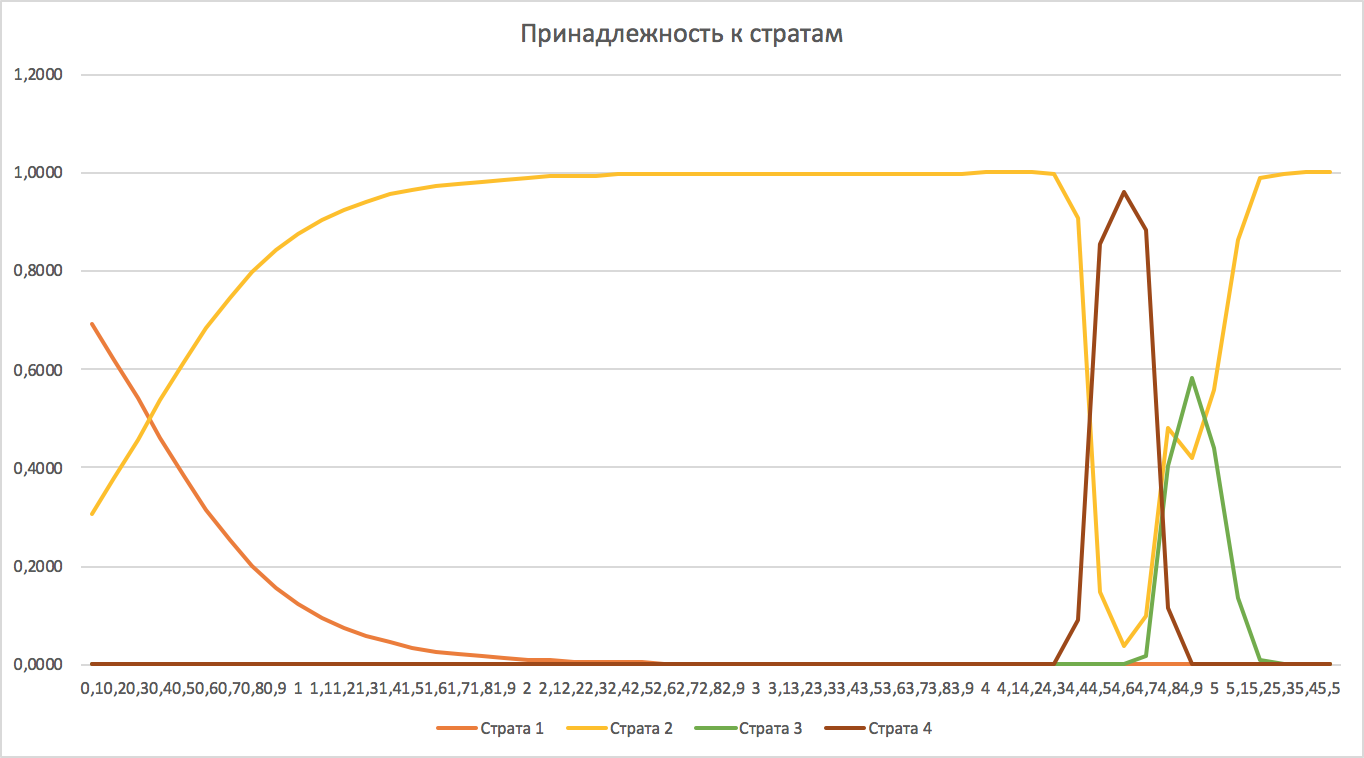
Далее была построена таблица с интервалами для подсчета плотности распределения четырех смесей. Ниже привожу получившийся график:



### Рисунок 22. Распределение смесей

По рисунку выше можно отметить, смесь 4 имеет наибольшую плотность. Остальные три примерно равны между собой.

После были подсчитаны затраты и построен соответствующий график, принадлежности к ним:



### Рисунок 23. Принадлежность к стратам

Анализируя рисунок, можно сделать вывод, что страта 1 имеет много маленьких значений, страта 2 включает в себя наибольший объем наблюдений, страта 3 имеет узкий диапазон значений [4,3-4,9], а страта 4 аналогично 3 страте имеет диапазон [4,7-5,2].

Порядковые номера участников нашего проекта - 5, 7 и 28. Посмотрим значения соответствующие значение наблюдений – 4,45, 2,77 и 3,12. Получается, что 7 и 28 наблюдения относятся к страте 2, то есть ко второй смеси, а 5 наблюдение к страте 4 или же к 4 смеси.

# Вывод

Таким образом, нами был проведен дискриминантный анализ, построены деревья классификации и проведена декомпозиция смесей распределения. Мы смогли отточить навыки работы с данными статистическими инструментами, что поможет на контрольной работе и профессиональной деятельности.

# 

# Приложение 1

Ссылка на дерево: ​​https://disk.yandex.ru/d/-5mIsTEo3bNEnA

# Приложение 2

| **Эмпирическая гистограмма** | | **Теоретическая гистограмма (до оптимизации)** | | **Теоретическая гистограмма (после оптимизации)** | | |
| --- | --- | --- | --- | --- | --- | --- |
|
| ***Карман*** | ***Частота*** | ***Промежуточные*** | ***Частота теоретическая (до оптимизации)*** | ***Промежуточные*** | ***Частота теоретическая (после оптимизации)*** | ***Минимумы для проверки соотвествия*** |
| **-3,219** | **1** | **0,860** | **0,860** | **0,361** | **0,361** | **0,361** |
| **-2,990** | **1** | **1,491** | **0,631** | **0,571** | **0,210** | **0,210** |
| **-2,761** | **0** | **2,473** | **0,982** | **0,881** | **0,311** | **0,000** |
| **-2,532** | **0** | **3,930** | **1,457** | **1,328** | **0,447** | **0,000** |
| **-2,302** | **1** | **5,986** | **2,057** | **1,954** | **0,626** | **0,626** |
| **-2,073** | **2** | **8,753** | **2,767** | **2,807** | **0,853** | **0,853** |
| **-1,844** | **1** | **12,304** | **3,551** | **3,938** | **1,131** | **1,000** |
| **-1,615** | **0** | **16,665** | **4,361** | **5,398** | **1,460** | **0,000** |
| **-1,386** | **3** | **21,820** | **5,154** | **7,232** | **1,835** | **1,835** |
| **-1,157** | **3** | **27,746** | **5,926** | **9,477** | **2,244** | **2,244** |
| **-0,928** | **4** | **34,483** | **6,737** | **12,149** | **2,672** | **2,672** |
| **-0,699** | **3** | **42,211** | **7,727** | **15,248** | **3,099** | **3,000** |
| **-0,469** | **2** | **51,288** | **9,077** | **18,751** | **3,503** | **2,000** |
| **-0,240** | **1** | **62,203** | **10,915** | **22,618** | **3,867** | **1,000** |
| **-0,011** | **5** | **75,392** | **13,189** | **26,799** | **4,181** | **4,181** |
| **0,218** | **6** | **90,978** | **15,586** | **31,259** | **4,460** | **4,460** |
| **0,447** | **3** | **108,541** | **17,563** | **36,008** | **4,749** | **3,000** |
| **0,676** | **11** | **127,091** | **18,550** | **41,150** | **5,142** | **5,142** |
| **0,905** | **6** | **145,316** | **18,226** | **46,943** | **5,793** | **5,793** |
| **1,134** | **11** | **162,052** | **16,736** | **53,867** | **6,924** | **6,924** |
| **1,364** | **12** | **176,773** | **14,721** | **62,675** | **8,808** | **8,808** |
| **1,593** | **17** | **189,897** | **13,124** | **74,411** | **11,736** | **11,736** |
| **1,822** | **24** | **202,772** | **12,876** | **90,355** | **15,943** | **15,943** |
| **2,051** | **26** | **217,357** | **14,585** | **111,874** | **21,519** | **21,519** |
| **2,280** | **26** | **235,682** | **18,325** | **140,186** | **28,311** | **26,000** |
| **2,509** | **30** | **259,252** | **23,571** | **176,048** | **35,863** | **30,000** |
| **2,738** | **31** | **288,534** | **29,282** | **219,470** | **43,422** | **31,000** |
| **2,967** | **35** | **322,684** | **34,150** | **269,512** | **50,041** | **35,000** |
| **3,196** | **53** | **359,644** | **36,960** | **324,270** | **54,758** | **53,000** |
| **3,426** | **38** | **396,604** | **36,960** | **381,086** | **56,816** | **38,000** |
| **3,655** | **45** | **430,701** | **34,096** | **436,939** | **55,854** | **45,000** |
| **3,884** | **51** | **459,705** | **29,004** | **488,937** | **51,998** | **51,000** |
| **4,113** | **61** | **482,676** | **22,972** | **534,766** | **45,829** | **45,829** |
| **4,342** | **68** | **507,546** | **24,869** | **573,008** | **38,242** | **38,242** |
| **4,571** | **191** | **609,126** | **101,580** | **761,060** | **188,051** | **188,051** |
| **4,800** | **428** | **910,085** | **300,959** | **1170,577** | **409,518** | **409,518** |
| **More** | **110** | **#VALUE!** | **399,915** | **#VALUE!** | **139,423** | **110,000** |
|  |  |  |  |  | **Сумма** | **91,90%** |