Министерство науки и высшего образования Российской Федерации
Санкт-Петербургский политехнический университет Петра Великого
Физико-механический институт
Высшая школа теоретической механики и математической физики

Работа допущена к защите
Директор ВШТМиМФ,
д.ф.-м.н., чл.-корр. РАН
_____ А. М. Кривцов
«___» _____2023 г.

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА БАКАЛАВРА

Исследование влияния различных факторов на риск возникновения и течение ишемической болезни сердца методами математической статистики.

по направлению подготовки
01.03.03 «Механика и математическое моделирование»
Направленность
01.03.03_03 Математическое моделирование процессов нефтегазодобычи

Выполнила
Студентка гр. 5030103/90301 А. Р. Курмакаева
Руководитель
Профессор ВШТМиМФ, д.ф-м.н В. А. Кузькин

Санкт-Петербург
2023

САНКТ-ПЕТЕРБУРГСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ ПЕТРА ВЕЛИКОГО
Физико-механический институт
Высшая школа теоретической механики и математической физики

УТВЕРЖДАЮ

Директор ВШТМиМФ

А. М. Кривцов

«__»_____20___г.

ЗАДАНИЕ

на выполнение выпускной квалификационной работы

студенту Курмакаевой Алсу Рашитовне, гр. 5030103/90301

1. Тема работы: Исследование влияния различных факторов на риск возникновения и течение ишемической болезни сердца методами математической статистики

2. Срок сдачи студентом законченной работы: 05.06.2023

3. Исходные данные по работе: справочная литература, актуальные научные публикации по теме работы.

4. Содержание работы (перечень подлежащих разработке вопросов): постановка задачи квалификационной работы, анализ подходов к исследованию влияния различных факторов, предварительный анализ и подготовка данных, вычисление эффекта влияния различных факторов, построение предиктивной модели.

5. Перечень графического материала (с указанием обязательных чертежей): не предусмотрено.

6. Консультанты по работе: отсутствуют

7. Дата выдачи задания: 27.02.2023

Руководитель ВКР:_____ В. А. Кузькин, д.ф.-м.н., профессор ВШТМиМФ

Задание принял к исполнению: 27.02.2023

Студент: _____ А. Р. Курмакаева

РЕФЕРАТ

На 65 с., 16 рисунков, 8 таблицы, 3 приложения

КЛЮЧЕВЫЕ СЛОВА: СТАТИСТИЧЕСКИЕ ТЕСТЫ, ИШЕМИЧЕСКАЯ БОЛЕЗНЬ СЕРДЦА, ИБС, МАТЕМАТИЧЕСКАЯ МОДЕЛЬ.

Тема выпускной квалификационной работы: «Исследование влияния различных факторов на риск возникновения и течение ишемической болезни сердца методами математической статистики».

В данной работе проанализированы различные возможные факторы влияния на риск возникновения ишемической болезни сердца. Основной анализ производился с помощью статических тестов на основе критериев Манна-Уитни и ANOVA. Для выявления относительной степени влияния признаков была построена модель логистической регрессии. Для реализации вышеперечисленных источников был использован Python.

THE ABSTRACT

65 pages, 16 pictures, 8 tables, 3 appendices

KEY WORDS: STATISTICAL TESTS, ISCHEMIC HEART DISEASE, CHD, MATHEMATICAL MODEL.

The topic of the graduate qualification work: "Research of different factors influence on the risk of coronary heart disease appearance and course by methods of mathematical statistics".

Different possible factors of influence on the risk of coronary heart disease were analyzed in this work. The main analysis was done using Mann-Whitney and ANOVA statistical tests. A logistic regression model was built to reveal the relative degree of influence of the traits. Python was used to implement the above sources.

СОДЕРЖАНИЕ

ОСНОВНЫЕ ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ 5

ВВЕДЕНИЕ 6

ГЛАВА 1. ОПИСАНИЕ ДАННЫХ 7

1.1. Исследование данных на наличие пропусков 8

1.2. Выявление корреляций между параметрами исследования 11

ГЛАВА 2.ВЫЯВЛЕНИЕ ФАКТОРОВ ВЛИЯНИЯ 16

2.1. Описание методов исследования 16

2.2. Описание результатов исследования 20

ГЛАВА 3. ОТНОСИТЕЛЬНАЯ ОЦЕНКА ФАКТОРОВ ВЛИЯНИЯ 27

3.1. Описание регерссионной модели 28

3.2. Описание модели решающих деревьев 30

3.3. Описание модели catboost 32

3.4. Исследование полученных результатов 32

ЗАКЛЮЧЕНИЕ 41

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ 43

ПРИЛОЖЕНИЕ 45

ОСНОВНЫЕ ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ

ТГ – триглицериды

Адсист – Артериальное давление систолическое

Аддиас – Артериальное давление диасталическое

ОХС – общий холестерин

ТГ – триглицериды

АСТ – аспартатаминотрансфераза

PS – пульс

ХСЛПНП – холестерин липопротеинов низкой плотности

ХСЛПВП – холестерин липопротеинов высокой плотности

КА – коэффициент атерогенности

ВОЗ – Всемирная организация здравоохранения

ВВЕДЕНИЕ

Ишемическая болезнь сердца (ИБС) – термин, предложенный Комитетом экспертов ВОЗ в 1962 г. Сегодня ИБС используется для описания острых и хронических заболеваний сердца, полученных в результате недостаточности снабжения миокардом кровью. По статистическим исследованиям Всемирной организации здравоохранения сердечно-сосудистые являются наиболее частой причиной смерти людей к моменту 2022 года, при этом большая часть этих заболеваний приходится на ишемическую болезнь сердца [15]. На данный момент проводится множество различных исследований, позволяющих установить взаимосвязь между различными факторами жизни человека и вероятностью его заболевания [14]. При этом при проведении подобных экспериментов важно учитывать особенности данных, на которых они будут проводиться.

В данной работе исследованы параметры влияния на развитие ишемической болезни сердца. Целью данной работы является выявление с помощью статистических тестов основных факторов, внесших наибольший вклад на возникновение ИБС, а также сравнительный анализ степени влияния каждого. В ходе работы были поставлены следующие задачи:

1) Провести статистическую обработку данных.
2) Провести статистические тесты и выявить основные факторы влияния.
3) Построить предиктивную математическую модель на основе машинного обучения, выявляющую риск возникновения ИБС для оценки степени влияния каждого фактора.
4) Сравнить полученные результаты.

ГЛАВА 1. ОПИСАНИЕ ДАННЫХ

На сегодняшний день статистические исследования проводятся довольно часто для того, чтобы зафиксировать какой-либо эффект. При этом для различных сфер применения существуют разные подходы к исследованиям. Эти подходы отличаются друг от друга требованиями к подготовке теста, стандартами качества, а также предобработкой данных.

В случае исследования различных факторов влияния на риск возникновения ИБС будет использовано кагорное (обсервационное) исследование. При проведении подобного анализа лица – объекты исследования распределяются в две группы в соответствии с наличием выбранного фактора. Далее исходя из истории их болезни проводят сравнение частоты развития клинического исхода.

Проведение статистического исследования включает в себя несколько этапов:

1) Планирование исследования
2) Сбор данных
3) Предобработка полученных материалов
4) Статистический анализ

В исследуемой выборке 125 наблюдений генеральной совокупности – 125 наблюдаемых пациентов (данные обезличенны).

Предобработка полученных материалов включает в себя следующие задачи:

1. Очистка данных: удаление дубликатов - повторяющихся записей из набора данных, обработка пропущенных значений, обработка выбросов – аномальных значений.
2. Преобразование переменных: преобразования переменных для улучшения их распределений, создание новых переменных на основе имеющихся для введения новых аспектов исследования.
3. Фильтрация данных: удаление ненужных переменных, удаление от нерепрезентативных переменных.

4. Кодирование данных: преобразование категориальных переменных в числовые значения, обработка текстовых данных.
5. Масштабирование данных: приведение значений переменных к определенному диапазону или стандартному формату.

При статистическом анализе необходимо в соответствии условиями сбора генеральной совокупности выбрать критерий сравнения. В данном исследовании будут рассматриваться независимые выборки с учетом многих факторов. Для выполнения описательной статистики используются следующие методы: вычисление медиан и интерквантильных интервалов, пропорций. При сравнении двух независимых групп по одному признаку применяется критерий Манна-Уитни. При одновременном анализе трех и более признаков задействуются такие методы как: регрессионный анализ, логистический регрессионный анализ, анализ древовидных диаграмм [2].

1.1 Исследование данных на наличие пропусков

При проведении статистических исследований, а также построении моделей машинного обучения необходимо учесть пропуски, так как модель на них не сможет обучиться, а также удалить дубликаты. Необходимость удаления дубликатов связана с вычислением общей ошибки в подобных математических моделях, формула которой представлена ниже

$$\Delta_{\mathrm{ml}} = \mathrm{bias} + \mathrm{variance} + \mathrm{noise}, \quad (3)$$

где Δ_{ml} – ошибка модели, bias – смещение, $\mathrm{variance}$ – разброс, noise – шум.

Смешение показывает, насколько модель смещена от истинного значения исходной функции. Дубликаты вносят вклад в эту часть ошибки. Разброс в свою очередь отражает вариативность предсказания модели при обучении на различных тренировочных выборках. Шум – это ошибка, связанная с нерегулярностью и непредсказуемостью данных, их чаще всего сокращают с помощью поиска квантильных значений и удалением тех, что выходят за рамки 1 и 3 квантиля или 25 и 75 перцентиля.

Методы работы с пропусками:

1) Удаление пропусков
2) Заполнение пропусков его статистической характеристикой (модой, средним значением, медианой и тд.)
3) Заполнение новым значением, не встречавшимся ранее в выборке
4) Восстановление данных с помощью моделей машинного обучения.

Для того чтобы оценить количество пропусков была построена матрица разреженности (рис. 1.1):

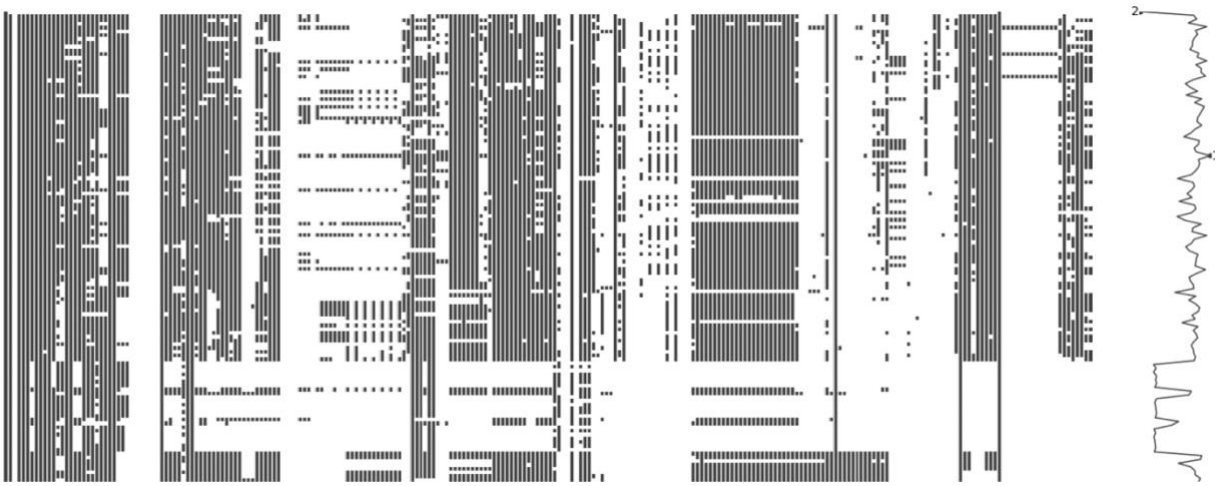


Рис. 1.1. Матрица разреженности исходного датасета.

Датасет состоит из 125 записей – 125 наблюдаемых пациентов и 221 признака – 221 исследуемых характеристик каждого пациента. Исходя из матрицы разреженности можно сделать вывод о том, что не все колонки являются информативными, большинство из них содержат большое число пропусков, также есть те, что и вовсе не заполнены. Для дальнейшего рассмотрения были выбраны следующие факторы: возраст, пол, номер визита, наследственность, адсист, аддиас, курение, рост, вес, ИМТ, менопауза, ОХС, ТГ, глюкоза, креатинин, билирубин, АСТ, артерия стентирования, aortic SP, конечные точки, FABP-4, ng/ml (кровь), FABP-4 epi mRNA, УЕЭ (эпикард. жир), FABP-4 мРНК в подкожном жире (УЕЭ), FABP4 rs16909192 (AA-1; AC-2; CC-3), FABP4 rs2290201 (GG-1; GA-2; AA-3), PS, Возраст развития СД, ХСЛПНП, ХСЛПВП, КА, эритроциты, гемоглобин, лейкоциты, лимфоциты, тромбоциты, моноциты.

Для сравнительного анализа из них подошли признаки, заполненность которых ниже 16% (рис. 1.2). Те признаки, заполненность которых была выше 16% были удалены. После описанного способа обработки был получен датесет, размерность которого: 125×18.

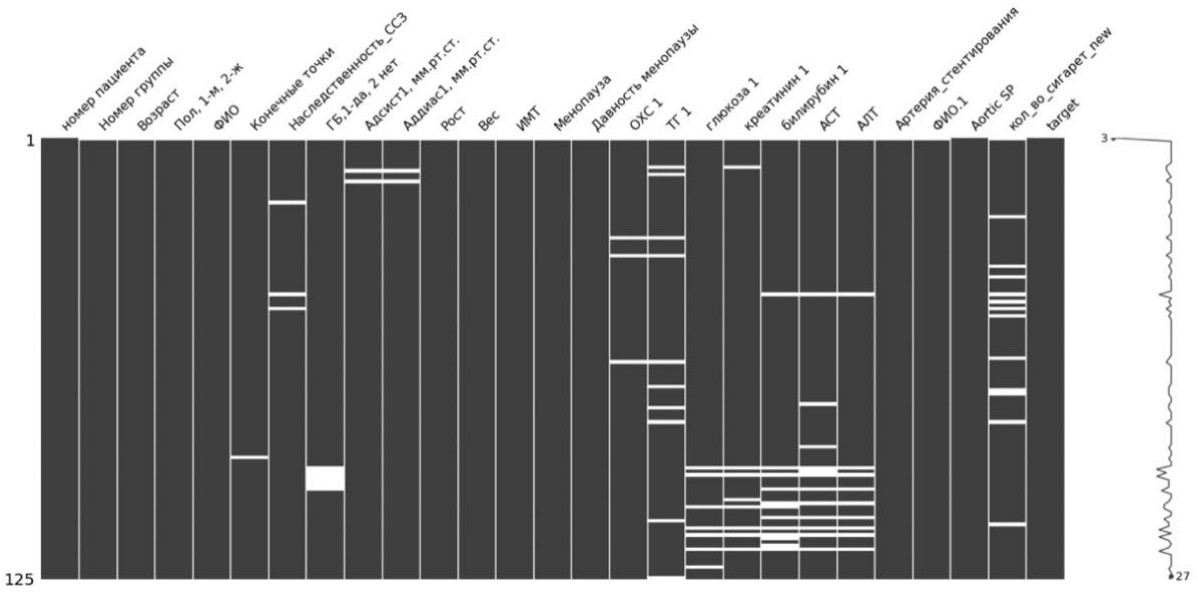


Рис. 1.2. Матрица разреженности нового датасета.

Далее были удалены все записи о пациенте, где в каком-либо описании его характеристик есть пропуски. Размерность полученного датасета 80×19, где 80 – количество записей или объектов исследования, а 19 – количество признаков.

Полученный первичный учетный документ для изучения влияния различных факторов на риск возникновения ИБС представлен в таблице 1.1:

№ п/п	Учетные признаки	Градация признака	Шифр
1	Возраст	28 – 78	-
2	Пол	женский мужской	2 1
3	Наследственность	да нет	1 2
4	Адсист	100 – 180	-
5	Аддиас	60 – 100	-
6	Курение	да нет	1 2

7	Рост	150 – 186	-
8	Вес	54 – 124	-
9	ИМТ	18.0 – 46.1	-
10	Менопауза	да нет	1 2
11	ОХС	2.2 – 8.58	
12	ТГ	0.35 – 5.23	-
13	Глюкоза	4.3 – 9.0	-
14	Креатинин	0.042 – 0.720	-
15	Билирубин	4.55 – 39.0	-
16	АСТ	12 - 167	-
17	Артерия стентирования	1-13	-
18	Конечные точки	0 >0	0 1
19	FABP-4, ng/ml (кровь)	11.49 – 439.06	-
20	FABP-4 epi mRNA, УЕЭ (эпикард. жир)	0.078 – 1.908	-
21	FABP-4 мРНК в подкожном жире (УЕЭ)	0.38 – 3.24	-
22	FABP4 rs16909192 (AA-1; AC-2; CC-3)	GG GA AA	1 2 3
23	FABP4 rs2290201 (GG-1; GA-2; AA-3)	AA AC CC	1 2 3
24	PS	58 – 115	-
25	Возраст развития СД	0 – 73	-
26	ХСЛПНП	0.09 – 5.83	-
27	ХСЛПВП	0.49 – 4.09	-
28	КА	0.8 – 9.6	-
29	Эритроциты	3.32 – 6.5	-
30	Гемоглобин	91 – 180	-
31	Лейкоциты	3.4 – 17.8	-
32	Лимфоциты	0.8 – 3.7	-
33	Тромбоциты	108 – 390	-
34	Моноциты	0.2 – 5.0	-

Таблица 1.1. Генеральная совокупность.

1.2 Выявление корреляций между параметрами исследования

При проведении исследований будут применены разные методы, в том числе и линейные, поэтому необходимо провести проверку на наличие взаимосвязей между параметрами. Проверка будет произведена с помощью определения корреляции Пирсона, которая показывает степень линейной зависимости между двумя непрерывными значениями. Вычисляется по следующей формуле [1]:

$$r = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}}, \quad (4)$$

где X_i и Y_i – значения переменных X и Y для наблюдения i, $\overline{X}$ и $\overline{Y}$ – средние значения, n – количество наблюдений.

Результат проверки представлен на рис. 1.3:

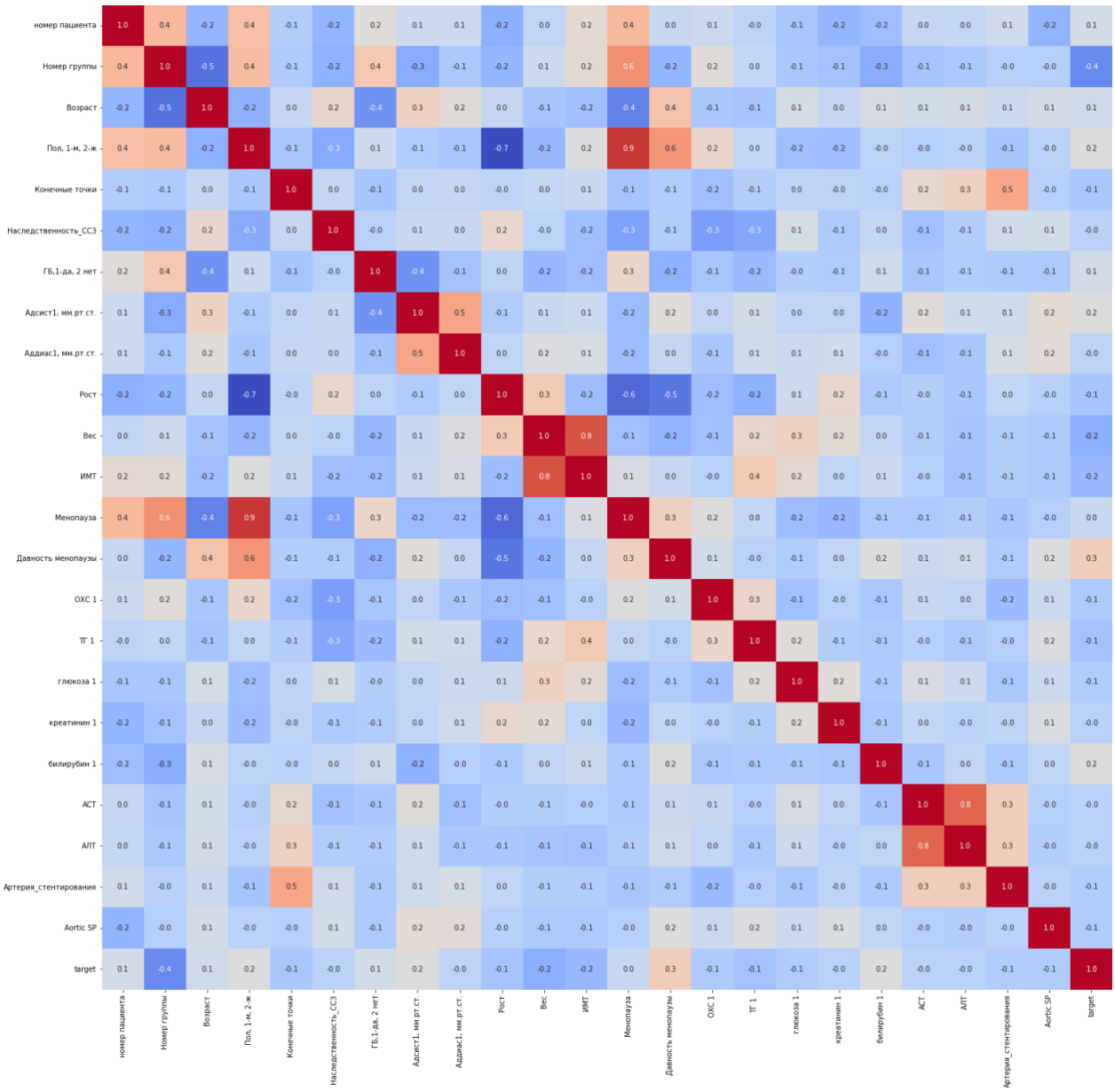


Рис. 1.3. Матрица корреляции исследуемых признаков.

Исходя из графика можно сделать следующие выводы: общая матрица корреляций показывает отсутствие линейной зависимости между большей частью рассматриваемых факторов влияния, однако есть также сильно скоррелированные между собой факторы:

1) Менопауза, пол, давность менопаузы (при попарном рассмотрении)
2) Стаж курения, курение
3) Вес, ИМТ
4) Рост, пол

При проведении статистических тестов будут рассмотрены все факторы, но при проведении исследования с помощью линейной логистической модели, из

каждой выделенной группы высокой корреляции будет оставлен лишь один признак (пол, курение, ИМТ).

Избавление от высоко скоррелированных признаков помогает избежать проблемы мультиколлинеарности, которая в свою очередь усложняет модель, также удаление этих признаков поможет сделать модель более интерпретируемой и лучшей с точки зрения производительности.

Для дополнительного изучения для группы, с наличием ИБС, были также построены матрицы корреляций зависимости различных факторов с FABP4 [10] в крови и в эпидуральном жире (рис. 1.4, рис. 1.5):

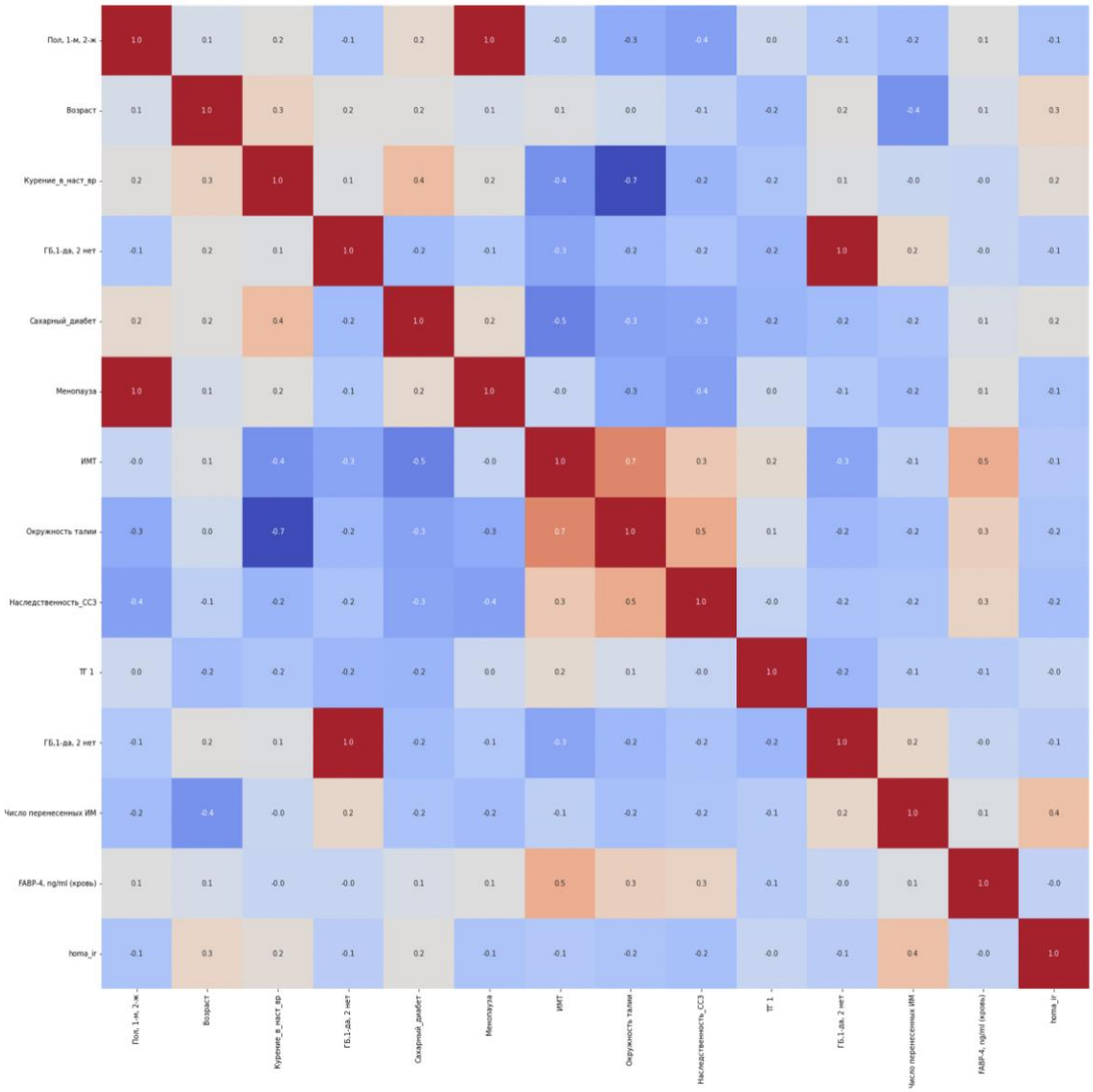


Рис. 1.4. Матрица корреляции для FABP4 в крови.

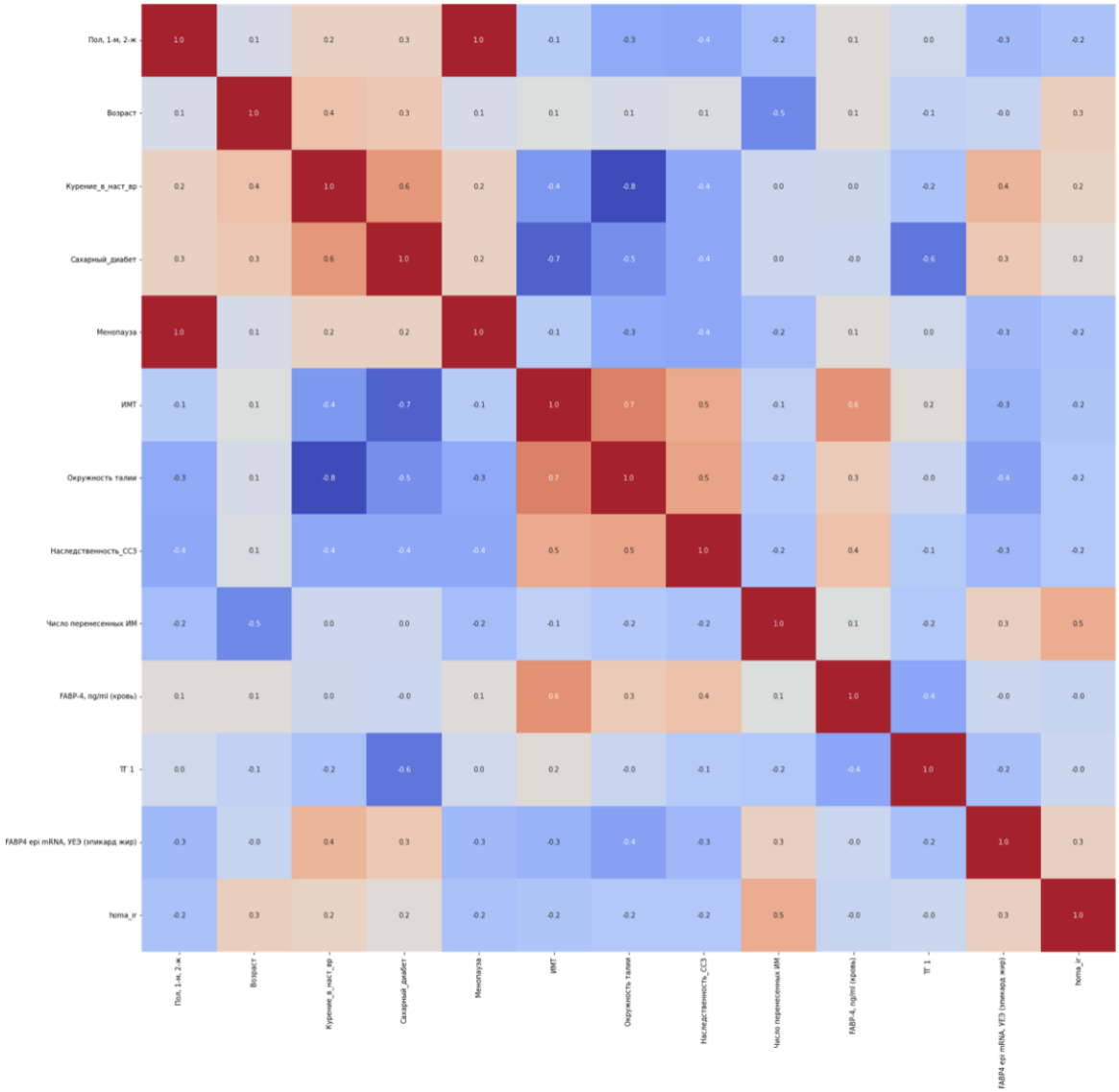


Рис. 1.5. Матрица корреляции для FABP4 в эпидуральном жире.

ГЛАВА 2. ВЫЯВЛЕНИЕ ФАКТОРОВ ВЛИЯНИЯ

2.1 Описание методов исследования

На этапе планирования исследования необходимо поставить с цель и задачи эксперимента. После определиться с генеральной совокупностью, а также с выборочной статистической совокупностью с учетом ее репрезентативности. Для определения генеральной совокупности используется расчет необходимого числа наблюдений, допустимой ошибки (p-value).

Формула вычисления количества необходимых наблюдений:

$$n = \frac{Nt^2pq}{N\Delta^2 + t^2pq}, \quad (1)$$

где $\Delta -$ предельная ошибка показателя, $p -$ величина исследуемого показателя, $q = (100 - p)$ или $(1 - p)$, $N -$ число наблюдений в генеральной совокупности, $t -$ коэффициент, показывающий вероятность достоверности полученного результат (Обычно берется равным 2).

Формула вычисления предельной ошибки показателя:

$$\Delta = t\sqrt{\frac{pq}{n}}, \quad (2)$$

Расчет размера подвыборки нужен при моделировании метода Bootstrap [5]. Использование этого метода помогает не учитывать ограничение по распределению исследуемого параметра, которое необходимо брать во внимание при выборе статистического теста. При проверке данных на нормальность с помощью теста Шапиро-Уилка [13] было выявлено, что не все данные проходят эту проверку (таблица 2.1). Поэтому использование метода Bootstrap – необходимость для соблюдения ограничений для использования выбранных статистических критериев.

№ п/п	Учетные признаки	Нормальность распределения признаки
1	Возраст	ненормальное
2	Пол	-
4	Наследственность	-
5	Адсист	ненормальное
6	Аддиас	нормальное
7	Курение	-
8	Рост	нормальное
9	Вес	нормальное
10	ИМТ	нормальное
11	Менопауза	-
12	ОХС	нормальное
13	ТГ	ненормальное
14	Глюкоза	ненормальное
15	Креатинин	ненормальное
16	Билирубин	ненормальное
17	АСТ	ненормальное
18	Артерия стентирования	-
19	Конечные точки	-
20	FABP-4, ng/ml (кровь)	ненормальное
21	FABP-4 epi mRNA, УЕЭ (эпикард. жир)	ненормальное
22	FABP-4 мРНК в подкожном жире (УЕЭ)	ненормальное
23	FABP4 rs16909192 (AA-1; AC-2; CC-3)	-
24	FABP4 rs2290201 (GG-1; GA-2; AA-3)	-
25	PS	ненормальное
26	Возраст развития СД	ненормальное
27	ХСЛПНП	ненормальное
28	ХСЛПВП	ненормальное
29	КА	ненормальное
30	Эритроциты	ненормальное
31	Гемоглобин	ненормальное
32	Лейкоциты	ненормальное
33	Лимфоциты	ненормальное
34	Тромбоциты	ненормальное
35	Моноциты	ненормальное

Таблица 2.1. Результат проверки нормальности распределени.

Основные шаги при моделировании метода Bootstrap описаны ниже [3]:

1. Исходные данные.

На этом этапе формируется генеральная совокупность наблюдений о факторах риска ИБС и соответствующих результатов.

2. Ресемплирование.

При ресемплировании из исходной выборки выбирается n случайных наблюдений с возвращением. По полученной выборке находится среднее значение. Этот процесс повторяется от 1000 до 10000 раз. При таком моделировании процесса верна центральная предельная теорема, по условиям которой полученное распределение выборочных средних является нормальным.

3. Анализ распределения.

На этапе анализа производится оценка статистик с помощью различных критериев.

При рассмотрении количественной переменной в качестве генеральной совокупность принимаются сами значения фактора, а при анализе категориальной переменной используется вероятность наблюдения ИБС у объекта исследования.

Критерий для статистического исследования численных факторов – критерий Манна-Уитни [4]. Необходимые условия применимости критерия – независимость выборки, а также нормальное распределение исследуемого параметра, выполнены. Основные шаги использования данного критерия:

1. Определение гипотез.

Нулевая гипотеза – распределение риска ишемической болезни сердца одинаково для обеих групп.

$$H0: F_1(x) = F_2(x), \quad (5)$$

где $F_1(x)$ и $F_2(x)$ –функции распределения риска для исследуемого признака соответственно.

Альтернативная гипотеза – распределение риска ишемической болезни сердца различно для двух групп.

$$H1: F_1(x) \neq F_2(x), \quad (6)$$

2. Ранжирование.

Проведение ранжирования по риску ишемической болезни сердца в каждой группе. Присвоения рангов наблюдения в обеих группах (U_1 и U_2).

3. Распределение статистики U.

Расчет суммы рангов для каждой группы (U_1 и U_2) [9]. Сравнение значений для U_1 и U_2.

4. Распределение статистики U.

При достаточно большой выборке, статистика U приближается к нормальному распределению. При малых выборках используется таблица критических значений для проверки статистической значимости.

5. Проверка статистического критерия.

Сравниваем значения статистики U с критическим значением для заданного уровня значимости – $p_{value} = 0.01$. Если значения статистики U меньше p_{value}, отвергаем нулевую гипотезу и считаем различия в распределении риска между группами статистически значимыми.

Критерий для исследования категориальных факторов – ANOVA [6]. Этот критерий подходит для описания и сравнения двух и более групп. Необходимые условия применимости критерия проверены. Принцип работы ANOVA заключается в выполнении следующих шагов:

1. Определение гипотез.

Этап аналогичен этапу 1 при описании предыдущего критерия. Основное различие в объектах выборки, если ранее рассматривались средние значения подвыборок из самих объектов начальной выборки, то на этом этапе рассматривается вероятность возникновения ИБС.

2. Разбиение данных на группы.

Исходные данные разбиваются на две или более группы в зависимости от исследуемого фактора.

3. Вычисление статистик.

По полученной выборки из вероятностей для каждой группы вычисляется оценки дисперсии от общего среднего значения и от средних внутри группы:

$$MS_B = \sum \frac{\left(\overline{X_j} - \overline{X_G}\right)^2}{k-1} n, \quad (7)$$

$$MS_W = \frac{s_1^2 + s_2^2 + \ldots + s_k^2}{k}, \quad (8)$$

где MS_B – оценка общей дисперсии по разбросу между группами, $\overline{X_j}$ – среднее группы j, $\overline{X_G}$ – общее среднее, k – число групп, n – размер группы,

MS_W – оценка общей дисперсии по разбросу внутри группы, $s_1^2 + s_2^2 + \ldots + s_k^2$ – сумма квадратов стандартных отклонений внутри группы.

Далее идет расчет F-статистики:

$$F = \frac{MS_B}{MS_W}, \quad (9)$$

4. Проверка статистического критерия.

Проводится аналогично предыдущему тесту.

2.2 Описание результатов исследования

Для описания результатов исследования были построены графики «Ящики с усами» для группы объектов с ИБС – группа 1 и для группы объектов без ИБС – группа 0. Пример построенных графиков представлен на рисунке 2.1, остальные графике представлены в приложении:

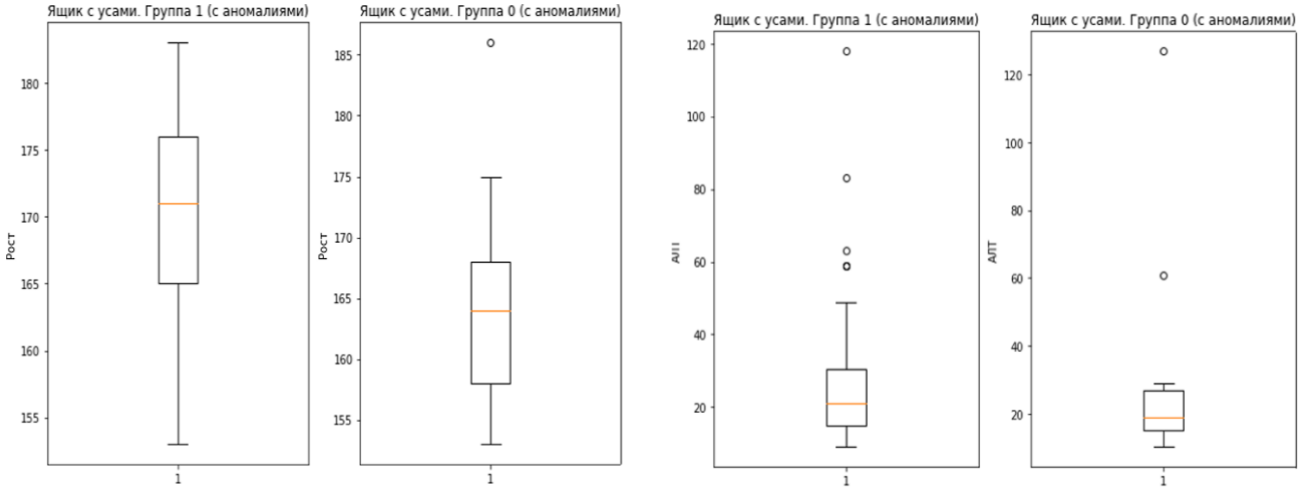


Рис. 2.1. Ящики с усами для количественных признаков.

Полученные графики помогают сравнить относительное расположение медиан (отмечены оранжевой линией на графике) исследуемого признака, а также указывают на 1 и 3 квантиль (границы «усов» графика), помимо этого на графике показаны аномальные значение – выпуклые точки.

При детальном рассмотрении графика видно, что есть факторы, медианы распределения которых значительно отличаются друг от друга, а также те, значения которых не сильно различимы.

В группу сильных различий попали такие факторы как: возраст, Адсист1, рост, вес, пол, давность менопаузы, ОХС, креатинин, билирубин, АСТ. Именно эти факторы являются первыми кандидатами на роль тех факторов, что сильно влияют на риск возникновения ишемической болезни сердца.

Результаты оценки влияния каждого признака с помощью статистического теста с заданной статистической значимостью p-value = 0.01 представлена в таблице 2.2, где alpha – полученная статистическая значимость, которая в последствии сравнивалась с p-value:

Учетные признаки	alpha
Возраст	0.000000
Пол	0.000000
Номер визита	0.000000
Наследственность	0.000000
Адсист	0.000000

Аддиас	0.186201
Курение	0.000000
Рост	0.000000
Вес	0.000000
ИМТ	0.000000
Менопауза	0.000000
ОХС	0.000020
ТГ	0.000010
Глюкоза	0.740582
Креатинин	0.001923
Билирубин	0.000010
АСТ	0.037024
Артерия стентирования	0.000000
Aortic SP	0.000000
Конечные точки	0.000000
FABP-4, ng/ml (кровь)	0.003354
FABP-4 epi mRNA, УЕЭ (эпикард. жир)	0.000000
FABP-4 мРНК в подкожном жире (УЕЭ)	0.000000
FABP4 rs16909192 (AA-1; AC-2; CC-3)	0.000000
FABP4 rs2290201 (GG-1; GA-2; AA-3)	0.000000
PS	0.000001
Возраст развития СД	0.000000
ХСЛПНП	0.853627
ХСЛПВП	0.190666
КА	0.000000
Эритроциты	0.001039
Гемоглобин	0.000000
Лейкоциты	0.001332
Лимфоциты	0.297920
Тромбоциты	0.069957
Моноциты	0.000000

Таблица 2.2. Значения p-value.

Исходя из полученных результатов в группу влияния вошли признаки, статистическая значимость alpha которых меньше значения 0.01: возраст, пол, наследственность, адсист, курение, рост, вес, ИМТ, менопауза, ОХС, ТГ, креатинин, билирубин, артерия стенирвания, aortic SP, FABP-4, ng/ml (кровь),

FABP-4 epi mRNA, УЕЭ (эпикард. жир), FABP-4 мРНК в подкожном жире (УЕЭ), FABP4 rs16909192 (AA-1; AC-2; CC-3), FABP4 rs2290201 (GG-1; GA-2; AA-3), PS, Возраст развития СД, КА, эритроциты, гемоглобин, лейкоциты, моноциты.

Для факторов, значения p_value которых больше 0.01 нет достаточных оснований для отклонения нулевой гипотезы.

Аналогично для группы людей с ИБС были получены таблицы влияния различных факторов на значения FABP4 в крови и эпидуральном жире:

	Alpha - FABP-4, ng/ml (кровь)	Alpha - FABP4 epi mRNA, УЕЭ (эпикард жир)
Пол	0.000000	0.081698
Возраст	0.000000	0.000599
Курение	0.000000	0.000000
Количество инфарктов миокарда	0.000000	0.000000
Артериальная гипертензия	0.000000	0.024372
Сахарный диабет	0.000000	0.003124
Индекс HOMA-IR	0.000000	0.000000
Менопауза	0.000000	0.060416
ИМТ_1	0.000000	0.000000
ИМТ_2	0.000000	0.000000
Обхват талии для мужчин	0.000000	0.021807
Наследственность	0.000000	0.000000
Триглицериды	0.163332	0.000000

Таблица 2.3. Значения p-value для FABP4.

Помимо этого, для более детального исследования и рассмотрения как могут повлиять на риск возникновения ишемической болезни сердца два

фактора, прошедших проверку по критическому значению, были построены дополнительные графики исследования зависимости относительно пола.

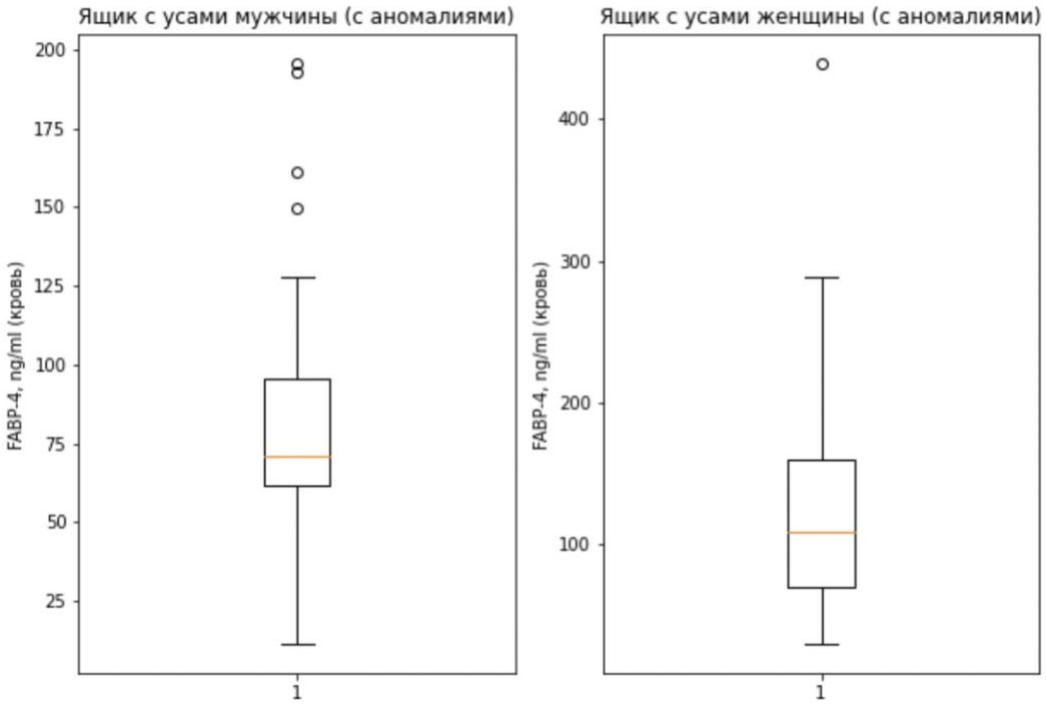


Рис. 2.2. Зависимость распределения фактора «FABP4-кровь» от пола.

Первый график (рис. 2.2) показывает наличие расхождений по параметру «FABP4-кровь» в двух группах – мужчины и женщины. Это также видно по разным значениям медиан (оранжевая линия) и относительным смещениям значений 1 и 3 квантилей.

Далее группа мужчин и женщин была поделена еще две подгруппы по наличию и отсутствию ИБС соответственно. Фактор «FABP4-кровь» был рассмотрен повторно для каждой подгруппы.

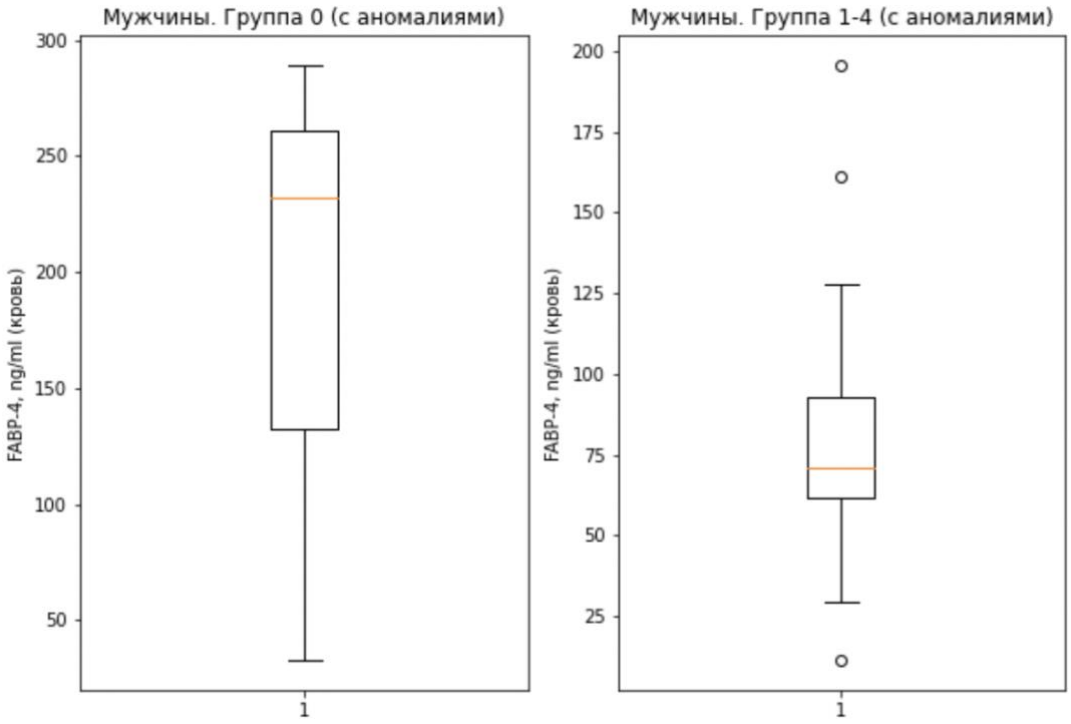


Рис. 2.3. Зависимость распределения фактора FABP4-кровь для мужчин с ИБС и без.

Полученный график для группы мужчин говорит об еще большем увеличении разрыва в статистически значимых величинах. Аналогичный график построен для группы женщин (рис. 2.4.):

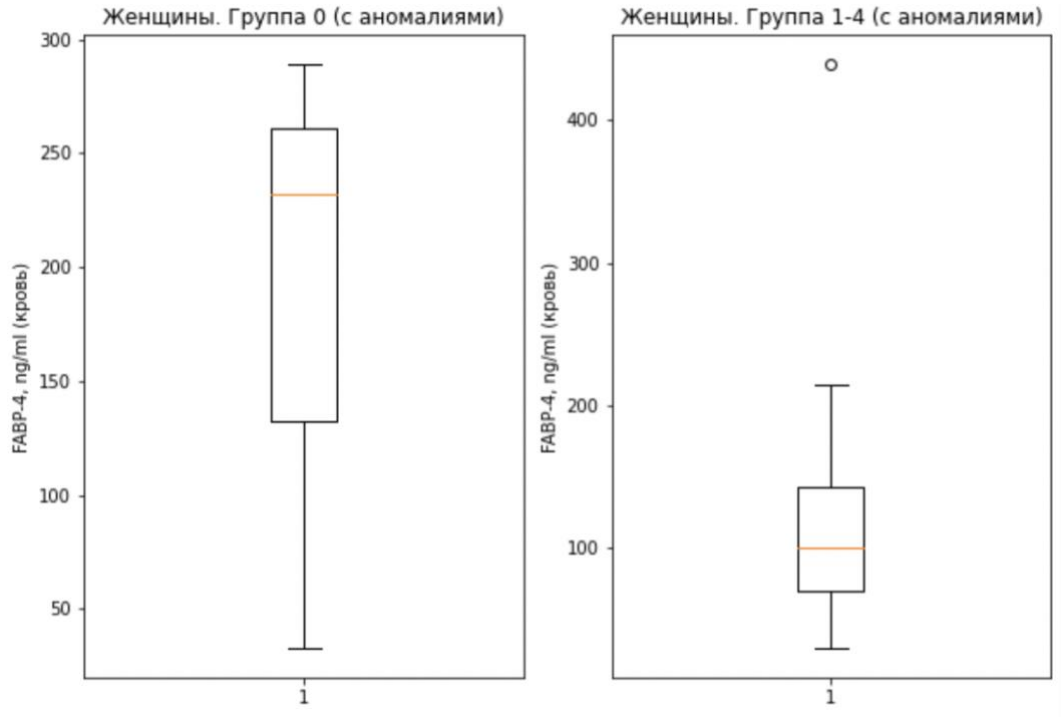


Рис. 2.4. Зависимость распределения фактора «FABP4-кровь» для женщин с ИБС и без.

На рисунке 2.4 также видно достоверное отличие в медианном значении.

Статистическая значимость отличия мужчин от женщин в группе людей с наличием ИБС по значению FABP4 в крови равна 0.000000, по значению FABP4 в эпидуральном жире равна 0.081698, в группе людей с отсутствием ИБС по значению FABP4 в крови равна 0.000000, по значению FABP4 в эпидуральном жире равна 0.000000.

Получившиеся результаты свидетельствуют о том, что наличие нескольких статистически значимых факторов влияния могут усиливать вероятность наличия или отсутствия ишемической болезни сердца. Детальный анализ относительной и совокупной зависимости факторов будет произведен с помощью предиктивной математической модели.

ГЛАВА 3. ОТНОСИТЕЛЬНАЯ ОЦЕНКА ФАКТОРОВ ВЛИЯНИЯ

Для относительной оценки степени влияния каждого фактора будут использованы 3 математические модели на основе машинного обучения: модель логистической регрессии, модель решающих деревьев, catboost. Основная задача моделей – провести классификацию объектов по наличию ИБС. Далее будет выбрана модель с наибольшей предиктивной способностью, которая будет оценена по таким метрикам как: $\mathrm{ROC}-\mathrm{AUC}$, Precision, Recall, F1, также будет выведена метрика accuracy – доля верно определенных объектов. Формулы для вычисления:

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}, \tag{7}$$

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}, \tag{8}$$

$$\mathrm{F1} = 2\,\frac{\mathrm{Recall}*\mathrm{Precision}}{\mathrm{Recall}+\mathrm{Precision}}, \tag{9}$$

$$\mathrm{ROC}-\mathrm{AUC} = \text{площадь под графиком в осях } \mathrm{TPR}, \mathrm{FPR}, \tag{10}$$

где TP – верно угаданный 0 класс (без ИБС), FP – неверно предсказанный 0 класс, FN – неверно предсказанный 1 класс (с ИБС), TPR – доля верно определенных положительных объектов, FPR – доля верно определенных отрицательных объектов.

Выбор метрик совершался исходя из дисбаланса классов в обучающей выборке модели, количество объектов, болеющих ИБС в рассматриваемой выборке гораздо больше (рис. 3.1):

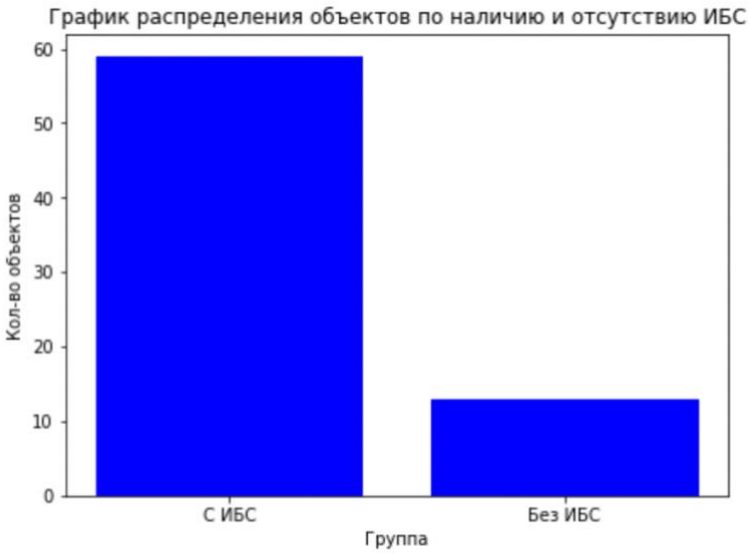


Рис. 3.1. Количество объектов в исследуемых группах.

После выбора наилучшей модели будут выведены коэффициенты – «веса» факторов, которые покажут относительную степень влияния каждого фактора.

3.1 Описание регрессионной модели

Исследование с помощью логистической регрессии состоит из следующих основных шагов:

1. Логистическая функция (сигмоидальная функция):

$$\sigma(z) = \frac{1}{1 + e^{-z}}, \tag{11}$$

где z – линейная комбинация входных признаков и их весов

2. Линейная комбинация входных признаков и их весов [12]:

$$z = w_0 + w_1 x_1 + w_2 x_2 + \cdots + w_n x_n, \tag{12}$$

где w_0, w_1, w_2, …, w_n– веса признаков, x_1, x_2, …, x_n – значения входных признаков.

3. Функция отклика – вероятность отнесения к первому и второму классу, соответствующие наличию и отсутствию ИБС:

$$P(y = 1|x) = \sigma(z), \tag{13}$$

$$P(y = 0|x) = 1 - \sigma(z), \tag{14}$$

где $P(y = 1|x)$ – вероятность отнесения объекта с данными признаками к классу, соответствующему отсутствию ИБС при условии x, $P(y = 0|x)$ – вероятность отнесения объекта с данными признаками к классу, соответствующему наличию ИБС при условии x.

4. Функция правдоподобия (likelihood):

$$L(w) = \prod_{i=1}^{N} P(y_i|x_i), \quad (15)$$

где $L(w)$ – функция правдоподобия, N – количество наблюдений, y_i – истинные метки классов для наблюдения i, x_i – значения признаков для наблюдения i.

5. Функция потерь (logloss):

$$J(w) = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log\left(P(y = 1|x_i)\right) + (1 - y_i)\log\left(P(y = 0|x_i)\right)\right], \quad (16)$$

где y_i – истинные метки классов для наблюдения i, $P(y = 1|x_i)$ – вероятность отнесения наблюдения i к классу, соответствующему отсутствию ИБС при условии x, $P(y = 0|x_i)$ – вероятность отнесения объекта с данными признаками к классу, соответствующему наличию ИБС при условии x.

6. Обучение модели:

Нахождение оптимальных весов w_i путем минимизации функции потерь $J(w)$ с использованием метода градиентного спуска.

7. Регуляризация модели:

Для улучшения результатов разделяющей функции будут также использованы L1 (лассо) и L2 (гребневая) регуляризация. После будет выбран оптимальный вариант. Основная суть регуляризации заключается в добавлении штрафного члена в функцию потерь для ограничения величины коэффициентов модели.

Функция потерь при L1 регуляризации:

$$J(\theta) = J(w) + \lambda \sum_{j=1}^{n} |\theta_j|, \quad (17)$$

Функция потерь при L2 регуляризации:

$$J(\theta) = J(w) + \lambda \sum_{j=1}^{n} \theta_j^{\,2}, \quad (18)$$

где $J(\theta)$ – функция потерь, θ – вектор параметров модели (весов), λ – параметр регуляризации, $J_\theta(w)$ – вычисленная функция потерь (10).

Результатом работы данной функции является вероятность отношения объекта к классу.

Данная модель является относительно простой, что говорит о ее высокой степени интерпретируемости. Также она имеет широкое применение в таких сферах как медицина, маркетинг, реклама, финансы и так далее.

3.2 Описание модели решающих деревьев

Решающие деревья относятся к древовидным моделям машинного обучения, которые в свою очередь применяются в статистических исследованиях при одновременном анализе трех и более признаков. Также виды древовидных моделей, такие как случайный лес, градиентный бустинг и так далее, обладают рядом преимуществ, делающие их полезным инструментом для анализа медицинских данных. Нами будет рассмотрено две древовидные модели: модель решающих деревьев и catboost для прогнозирования болезни.

Основные этапы реализации решения с помощью деревьев решений очень похожи на человеческий процесс принятия решений. Основной принцип работы алгоритма – нахождение последовательности простых решающих правил. Для определения качества разбиения выборки по выбранным правилам решающие деревья используют критерии оценки однородности. Алгоритм работы:

1. Определение узла.

   Выбор признака и трешхолда - порога по нему для. Разбиение набора данных по полученному признаку и его трешхолду – определение данных в «листы» [7].
2. Оценка однородности данных, может быть осуществлена по следующим формулам:

   Индекс Джинни:

$$\mathrm{Gini}(p) = 1 - \sum_{i=1}^{C} p_i^2 \tag{19}$$

   Энтропия:

$$\mathrm{Entropy}(p) = -\sum_{i=1}^{C} p_i \log_2(p_i), \tag{20}$$

   где $\mathrm{Gini}(p)$ – значение индекса Джинни для узла, $\mathrm{Entropy}(p)$ – значение энтропии для узла, C – количество классов, p_i – доля примеров класса i в узле.
3. Расчет прироста информации для оптимального разделения признака в узле.

   Information Gain:

$$\mathrm{IG}(D, A) = H(D) - \sum_{v=1}^{V} \frac{|D_v|}{|D|} H(D_v), \tag{21}$$

   где $\mathrm{IG}(D, A)$ – прирост информации при разделении по признаку A в узле D, $H(D_v)$ – энтропия или индекс Джини для узла D, V – количество различных значений признака A
4. Поиск следующего узла для оптимального разбиения, далее возвращаемся на этап 1. Процесс повторяется до тех пор, пока мы не достигнем заданной точности в оценке однородности.

При этом модели решающих деревьев склонны к переобучению, для избежания этой проблемы применена регуляризация, основанную на

следующих моментах: ограничение по максимальной глубине дерева, ограничение на минимальное количество объектов в листе, ограничение на максимальное количество листьев в дереве, требование улучшения качества при делении текущей подвыборки на две не менее чем на заданное число процентов. Выбор наиболее подходящего метода производился экспериментально.

3.3 Описание модели catboost

Модель catboost основана на градиентном бустинге [8]. Она предназначена для решения задач машинного обучения. Ее главной особенностью является метод работы с категориальными признаками, которые встречаются и в исследуемой нами выборке. В отличие от других моделей catboost не требует дополнительной предварительной обработки категориальных признаков [11].

Алгоритм работы модели:

1. Построение базовых моделей.

   В качестве базовых моделей принимаются деревья решений. На этом этапе происходит реализация стандартного алгоритма дерева решений для.

2. Вычисление функции потерь.

   Для решения задачи бинарной классификации берется стандартная функция потерь log loss, описанная ранее (16).

3. Обучение.

   Обучение модели является итеративным процессом, в котором происходит последовательное построение деревьев с учетом градиента функции потерь предыдущего дерева.

При необходимости также используется регуляризация модели.

3.4 Исследование полученных результатов

В результате было рассмотрено 4 математические модели на основе машинного обучения: модель логистической регрессии, модель

логистической регрессии с регуляризацией, модель решающих деревьев, catboost.

Все модели склонны к переобучению из-за небольшого количества объектов в обучающей выборке. Поэтому для достоверной оценки был применен метод кросс-валидации. При его использовании весь датасет делился на 5 равных частей. После проводилось поочередное построение 5 разных моделей каждого метода. При этом для тестовой выборки бралась 1 часть из полученных пяти, остальные 4 шли на обучающую выборку, после тестовая часть менялась, обучалась новая модели и так далее. Для каждой модели вычислялись заранее выбранные метрики. Также при окончании эксперимента было вычислено стандартное отклонение метрик, которое также необходимо для учета корректности работы моделей. Чем ниже этот показатель, тем лучше модель отвечает поставленной задачи.

Полученные результаты моделей были представлены в виде графика функции ROC-AUC, рис. 3.2:

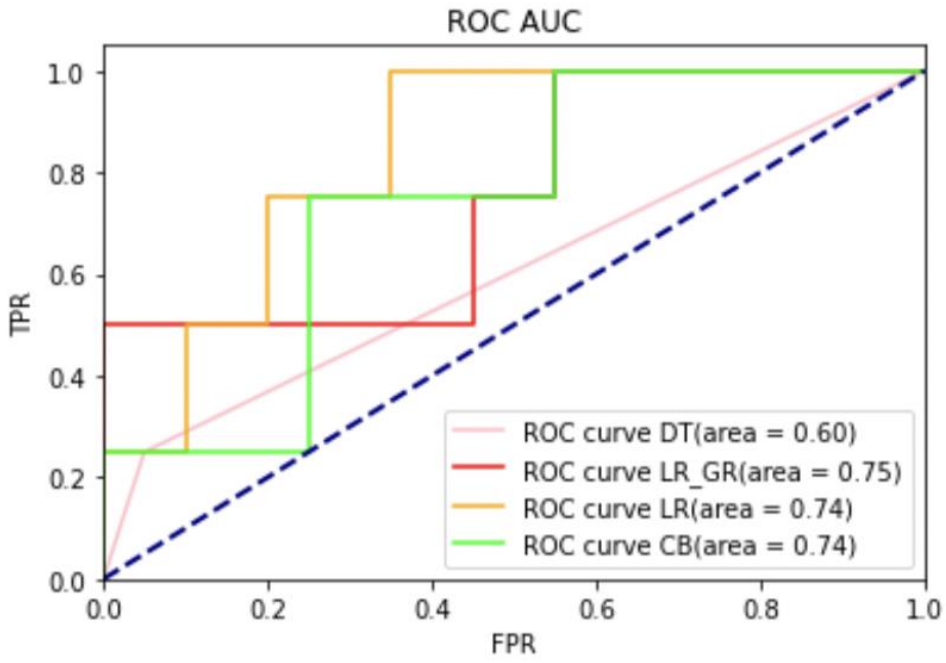


Рис. 3.2. Распределение количества объектов в исследуемых группах.

Наилучший результат показывает модель логистической регрессии, при регуляризационном коэффициенте $\lambda = 0$.

Помимо этого, была выведена таблица метрик, рис. 3.3:

	CatBoost	Logreg	Logreg_GR	Decision Tree
accuracy	0.833333	0.750000	0.750000	0.833333
precision	0.000000	0.796296	0.796296	0.803030
recall	0.000000	0.750000	0.750000	0.833333
roc_auc	0.680586	0.666117	0.645604	0.527106
f1	0.000000	0.621053	0.621053	0.619048

Рис. 3.3. Метрики качества моделей.

Исходя из таблицы, наилучший результат работы по метрике ROC-AUC имеет модель catboost, однако ее Precision и Recall равны нулю, что говорит о том, что она вовсе не предсказывает нулевой класс, то есть модель считает все объекты объектами первого класса (с ИБС), что говорит о ее плохой предиктивной способности. При этом наилучший результат также показывает логистическая регрессия. Её значение точности и полноты для определения каждого класса представлены на рисунке 3.4:

	precision	recall	f1-score	support
0	0.89	0.80	0.84	20
1	0.33	0.50	0.40	4
accuracy			0.75	24
macro avg	0.61	0.65	0.62	24
weighted avg	0.80	0.75	0.77	24

Рис. 3.4. Метрики качества модели логистической регрессии.

Значение метрики recall, равное 0.50, говорит, что модель верно нашла 2 объекта нулевого класса из 4 необходимых. При кросс-валидации стандартное отклонение метрик получилось равным 0.15, что говорит о дополнительных возможностях для улучшения модели, которое можно реализовать с помощью создания синтетической выборки или добавлением новых объектов в рассмотрение.

Для сравнительного анализа были выведены коэффициенты логистической регрессии, таблица 3.1:

Менопауза	-4.2444	x_1
ГБ,1-да, 2 нет	-4.0702	x_2
Конечные точки	-2.8613	x_3
Адсист1, мм.рт.ст.	-1.4117	x_4
билирубин 1	-1.1107	x_5
Возраст	-1.0734	x_6
АСТ	-1.0413	x_7
Рост	-0.5423	x_8
ОХС 1	-0.5386	x_9
глюкоза 1	-0.5030	x_{10}
Аддиас1, мм.рт.ст.	-0.2021	x_{11}
кол_во_сигарет_new	-0.1183	x_{12}
креатинин 1	-0.0915	x_{13}
ИМТ	1.2755	x_{14}
ТГ 1	2.2534	x_{15}
Наследственность_ССЗ	2.9108	x_{16}
Aortic SP	3.4545	x_{17}
Артерия_стентирования	4.4683	x_{18}

Таблица 3.1. Значения коэффициентов.

Исходя из полученных коэффициентов можно восстановить функцию логистической регрессии. Вид общей функции при полученном исследовании:

$$\sigma(z) = \frac{1}{1 + e^{-z}}, \tag{22}$$

$$\begin{aligned} z = {} & -4.2444x_1 - 4.0702x_2 - 2.8613x_3 - 1.4117x_4 - 1.1107x_5 \\ & -1.0734x_6 - 1.0413x_7 - 0.5423x_8 - 0.5386x_9 - 0.5030x_{10} \\ & -0.2021x_{11} - 0.1183x_{12} - 0.0915x_{13} + 1.2755x_{14} + 2.2534x_{15} \\ & +2.9108x_{16} + 3.4545x_{17} + 4.4683x_{18} \end{aligned} \tag{23}$$

где x_i – соответствующие значения переменных.

Также по коэффициентам можно оценить относительную важность исследуемых факторов. Степень влияния признака оценивается по абсолютному значению коэффициента, таблица 3.2:

креатинин 1	0.0915
кол_во_сигарет_new	0.1183
Аддиас1, мм.рт.ст.	0.2021
глюкоза 1	0.5030
ОХС 1	0.5386
Рост	0.5422
АСТ	1.0413
Возраст	1.0734
билирубин 1	1.1107
ИМТ	1.2755
Адсист1, мм.рт.ст.	1.4117
ТГ 1	2.2534
Конечные точки	2.8613
Наследственность_ССЗ	2.9108
Aortic SP	3.4545
ГБ,1-да, 2 нет	4.0702
Менопауза	4.2444
Артерия_стентирования	4.4683

Таблица 3.2. Абсолютные значения коэффициентов.

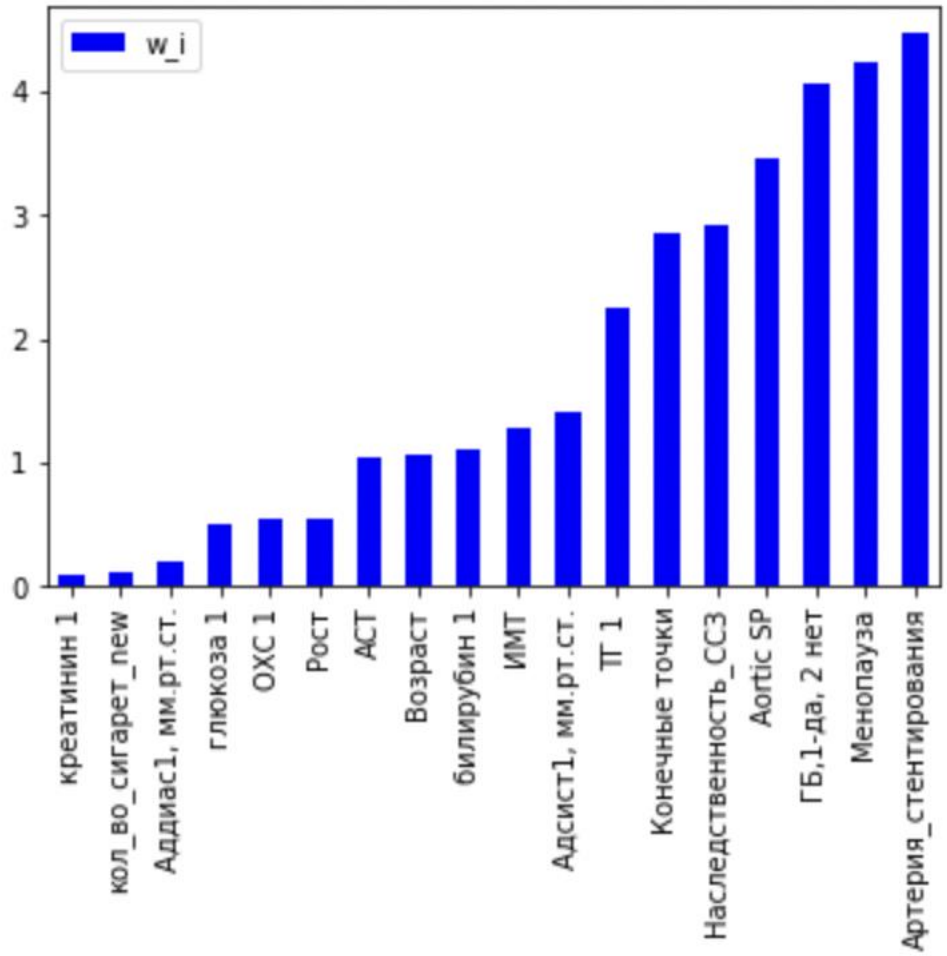


Рис. 3.5. График степени влияния факторов на риск возникновения ИБС.

Чем больше значение коэффициента, тем сильнее он влияет на риск возникновения ИБС. Наибольший вес в принятии решения моделью об отсутствии или наличии ишемической болезни сердца внесли признаки: артерия стентирования, наличие менопаузы, которая в свою очередь пропорциональна полу с точки зрения линейной корреляции, что также говорит о сильном влиянии пола в данном решении математической модели линейной регрессии. Также есть признаки, которые повлияли на результат логистической регрессии несильно: креатинин, количество сигарет, глюкоза, аддиас. Из них два признака: глюкоза, аддиас не имеют достаточных обоснований предполагать, что эти факторы влияют на риск возникновения ИБС со статистической значимостью, равной 0.01.

Стоит также отметить, что при исследовании были найдены группы высокой корреляции, из каждой группы был оставлен лишь один фактор для дальнейшего обучения модели. Поэтому при рассмотрении какого-либо фактора и его коэффициента, в случае если этот фактор входит в одну из групп высокой корреляции, его коэффициент и степень влияния распространяется на все факторы группы.

Аналогично были построены регрессионные модели для предсказания значений FABP4 в крови и эпидуральном жире в группе людей с ИБС для проведение многофакторного анализа, а также относительных степеней влияния:

ТГ 1	-28.3547
Курение_в_наст_вр	-26.4444
ИМТ	-23.9488
homa_ir	-0.8356
Пол, 1-м, 2-ж	16.2227
Менопауза	16.2227
Число перенесенных ИМ	20.0496
Наследственность_ССЗ	27.6155
Возраст	56.7007
Сахарный_диабет	61.0786
Окружность талии	160.8378

Таблица 3.3. Значения коэффициентов влияния факторов на значение FABP4 в крови.

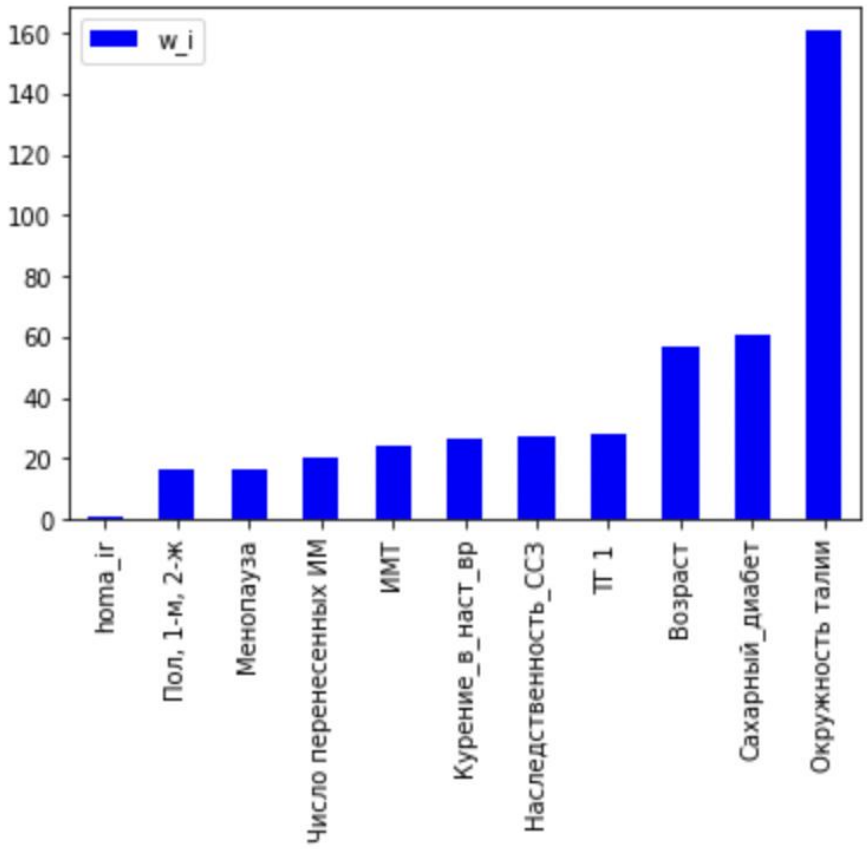


Рис. 3.6. График степени влияния факторов на значение FABP4 в крови.

Наследственность_ССЗ	-0.6130
Возраст	-0.3369
Число перенесенных ИМ	-0.2547
Пол, 1-м, 2-ж	-0.2476
Менопауза	-0.2476
homa_ir	-0.1826
Окружность талии	-0.1281
ИМТ	-0.0185
ТГ 1	0.0272
Сахарный_диабет	0.0386
FABP-4, ng/ml (кровь)	0.0869
Курение_в_наст_вр	0.1602

Таблица 3.4. Значения коэффициентов влияния факторов на значение FABP4 в эпидуральном жире.

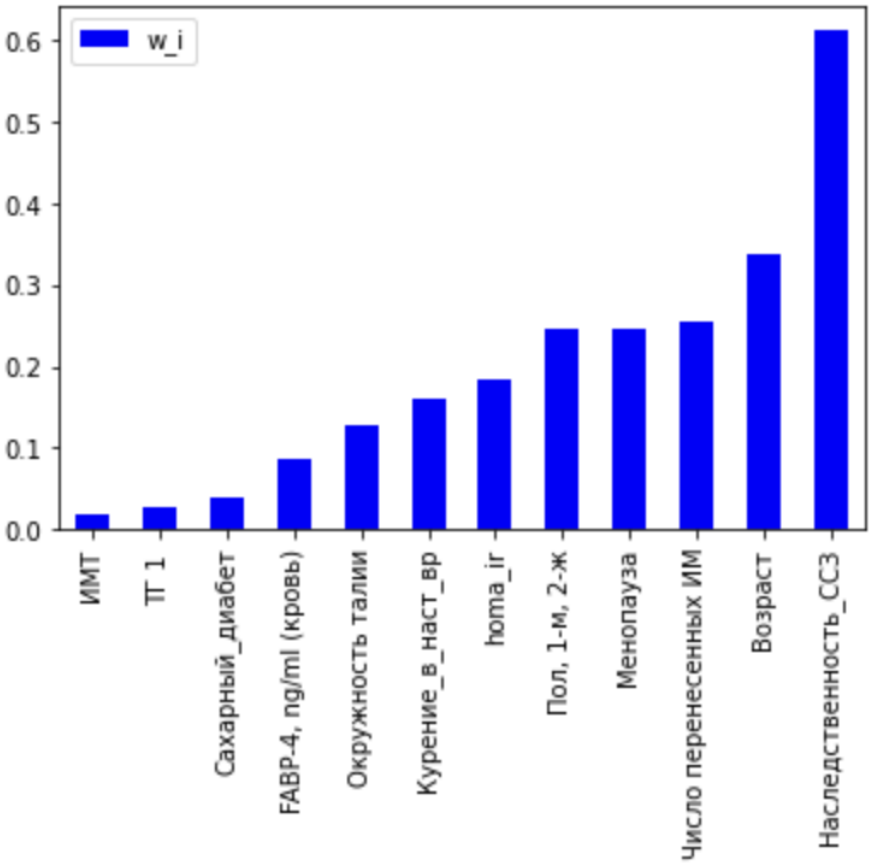


Рис. 3.7. График степени влияния факторов на значение FABP4 в эпидуральном жире.

ЗАКЛЮЧЕНИЕ

В ходе исследовательской работы был проведен анализ 221 различных факторов влияния на риск возникновения ишемической болезни сердца, при этом количество пациентов, вошедших в исследуемую группу, составило 125 человек. Данные были исследованы в обезличенном виде. Из общего набора факторов на основе анализа разреженности данных была выявлена группа наиболее репрезентативных признаков – 34 различных факторов.

Были сформулированы гипотезы о возможном влиянии каждого из этих факторов на возникновение болезни. Из них 29 факторов показали статистически значимое влияние на риск возникновения ИБС с заданной точностью 0.01.

Для проверки гипотез были выбраны следующие статистические критерии: Манна-Уитни для количественных факторов и ANOVA для категориальных. Критерий Манна-Уитни позволяет выявлять различия в значении параметра между малыми выборками. Дисперсионный анализ ANalysis Of Variance (ANOVA) дает возможность проводить множественные сравнения. Условия применимости обоих критериев были соблюдены – независимость наблюдений в выборке была учтена при сборе данных, нормальность распределений была реализована при помощи метода bootstrap и проверена с помощью критерия Шапиро-Уилка.

Также с помощью математических моделей, основанных на алгоритмах машинного обучения, была произведена оценка влияния одновременного наличия 18 различных факторов. Помимо этого, было найдено относительное влияние каждого признака. Для анализа были рассмотрены 3 модели: логистическая регрессия, решающие деревья и catboost. Все три модели были обучены решать задачу бинарной классификации – предсказания наличия и отсутствия ИБС у пациента по имеющимся факторам. Далее была выбрана модель с наилучшими предиктивными способностями с точки зрения метрик

машинного обучения, модель логистической регрессии. На основе коэффициентов полученной модели создан график степени влияния каждого фактора.

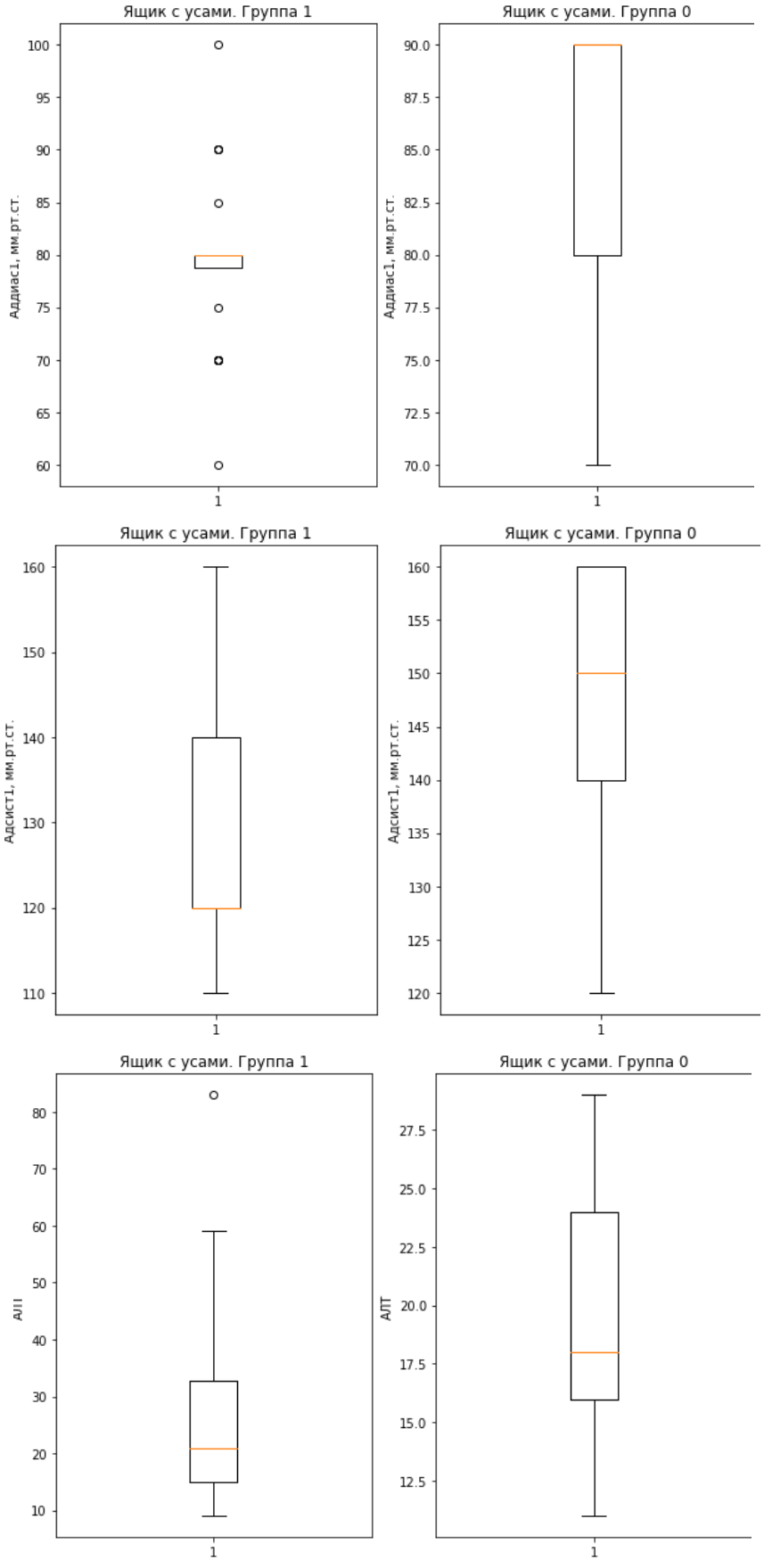
СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

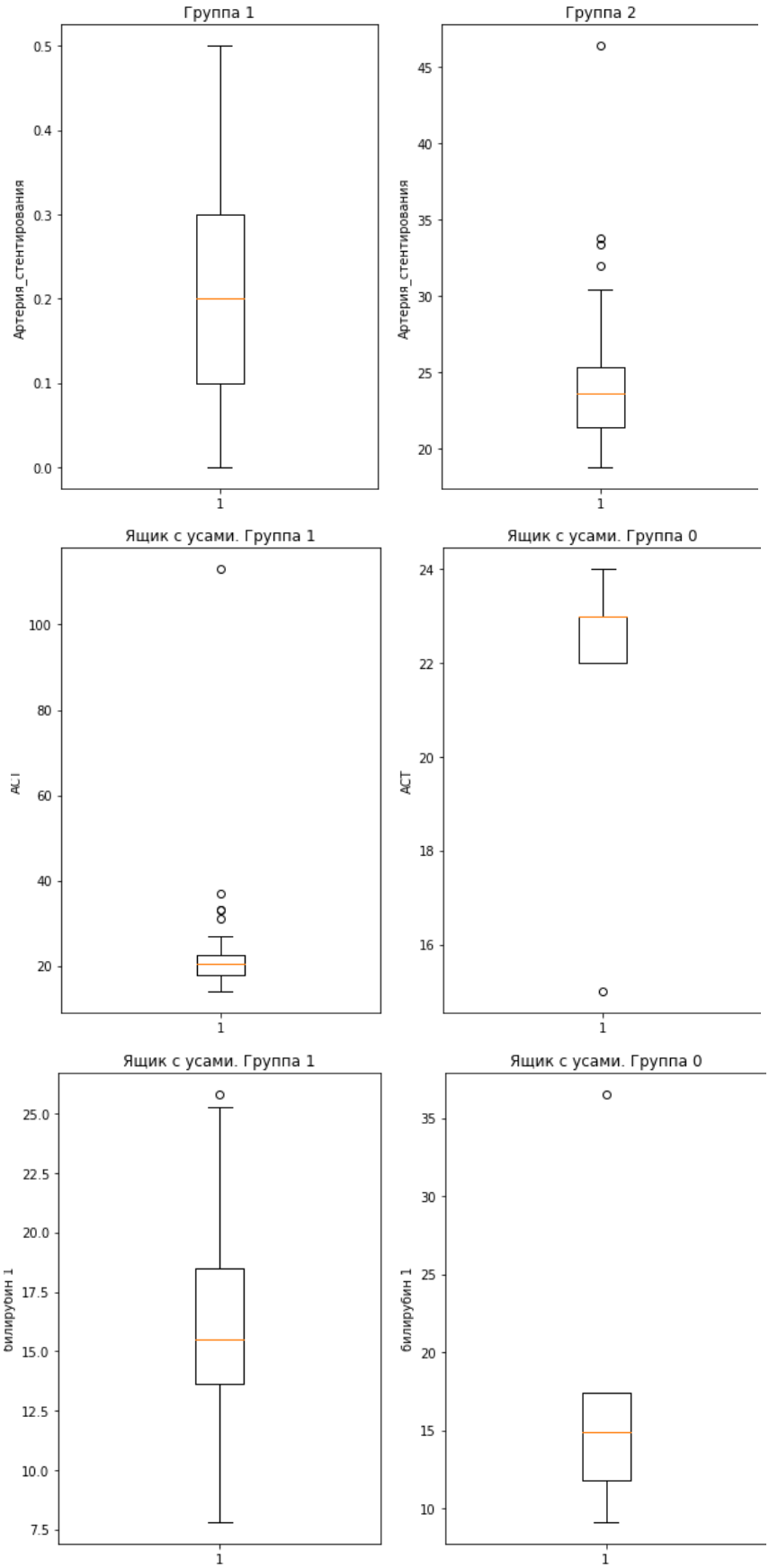
1. Сидняев Н.И. Теория вероятностей и математическая статистика: учебник для вузов, Москва: издательство Юрайт, 2023. 219 с.
2. Шарафутдинова Н.Х., Киреева Э.Ф., Николаева И.Е., Павлова М.Ю., Халфин Р.М., Шарафутдинов М.А., Борисова М.В., Латыпов А.Б., Галикеева А.Ш. – Статистические методы в медицине и здравоохранении, учеб. пособие, Уфа: ФГБОУ ВО БГМУ Минздрава России, 2018. – 131 с.
3. Chernick M. R. Bootstrap methods: A guide for practitioners and researchers. – John Wiley & Sons, 2011.
4. Conover W. J. Practical nonparametric statistics. – john wiley & sons, 1999. – Т. 350.
5. Efron B. Tibshirani RJ. An introduction to the bootstrap //Monographs on Statistics and Applied Probability. – 1993. – Т. 57. – С. 1-177.
6. Hinkelmann K., Kempthorne O. Design and analysis of experiments, volume 1: Introduction to experimental design. – John Wiley & Sons, 2007. – Т. 1.
7. Breiman L. Random forests //Machine learning. – 2001. – Т. 45. – С. 5-32.
8. Friedman J. H. Greedy function approximation: a gradient boosting machine //Annals of statistics. – 2001. – С. 1189-1232.
9. Mann H. B., Whitney D. R. On a test of whether one of two random variables is stochastically larger than the other //The annals of mathematical statistics. – 1947. – С. 50-60.
10. Miroshnikova V. V. et al. FABP4 and omentin-1 gene expression in epicardial adipose tissue from coronary artery disease patients //Genetics and molecular biology. – 2021. – Т. 44.
11. Prokhorenkova L. et al. CatBoost: unbiased boosting with categorical features //Advances in neural information processing systems. – 2018. – Т. 31.
12. Seber G. A. F., Lee A. J. Linear regression analysis. – John Wiley & Sons, 2003. – Т. 330.

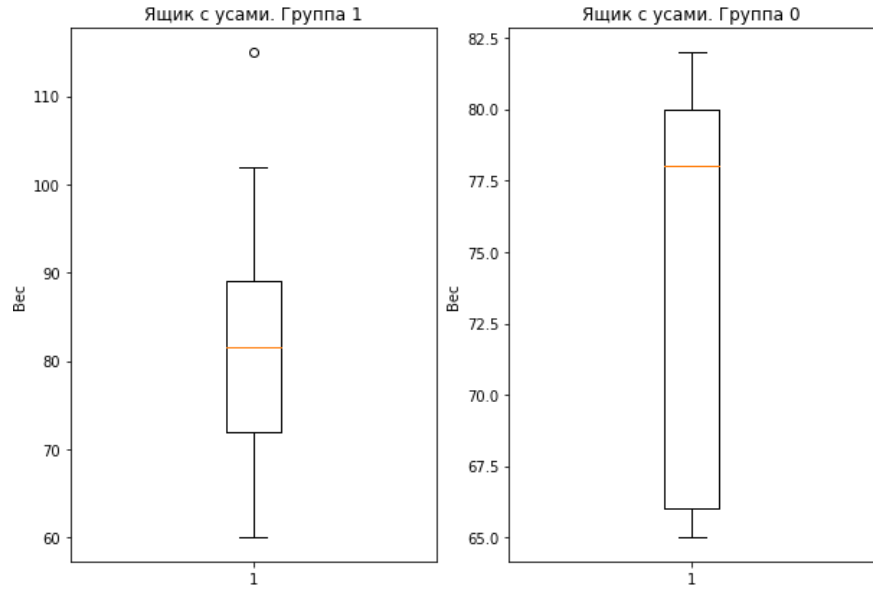
13. Shapiro S. S., Wilk M. B. An analysis of variance test for normality (complete samples) //Biometrika. – 1965. – Т. 52. – №. 3/4. – С. 591-611.
14. Yusuf S. et al. Global burden of cardiovascular diseases: part I: general considerations, the epidemiologic transition, risk factors, and impact of urbanization //Circulation. – 2001. – Т. 104. – №. 22. – С. 2746-2753.
15. Cardiovascular diseases (CVDs) // World Health Organization official site 2020. – URL: https://www.who.int/en/news-room/fact-sheets/detail/cardiovas- cular-diseases-(cvds) — (дата обращения: 18.03.2023).

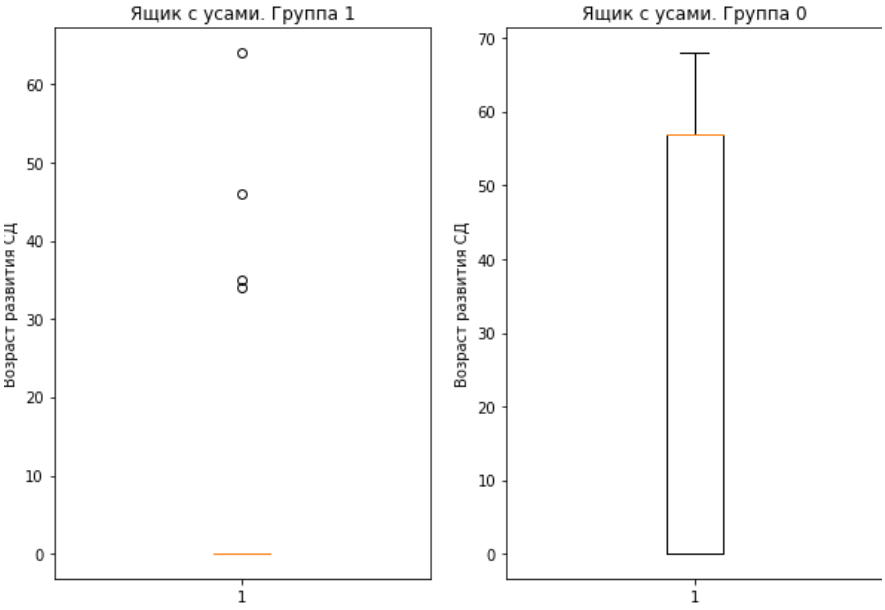
ПРИЛОЖЕНИЕ

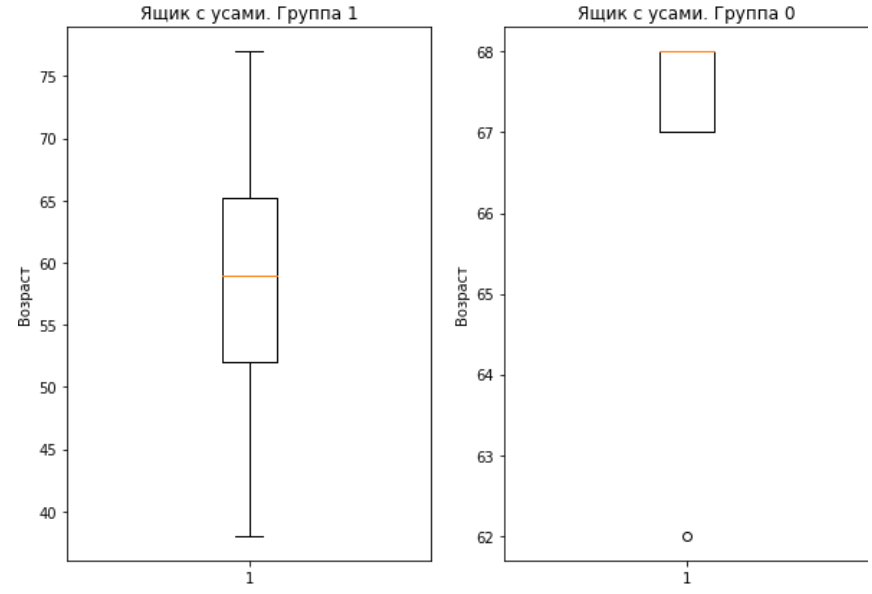
Графики ящиков с усами для репрезентативных факторов влияния на риск возникновения ИБС, группа 1 – люди с ИБС, группа 0 – люди без ИБС:

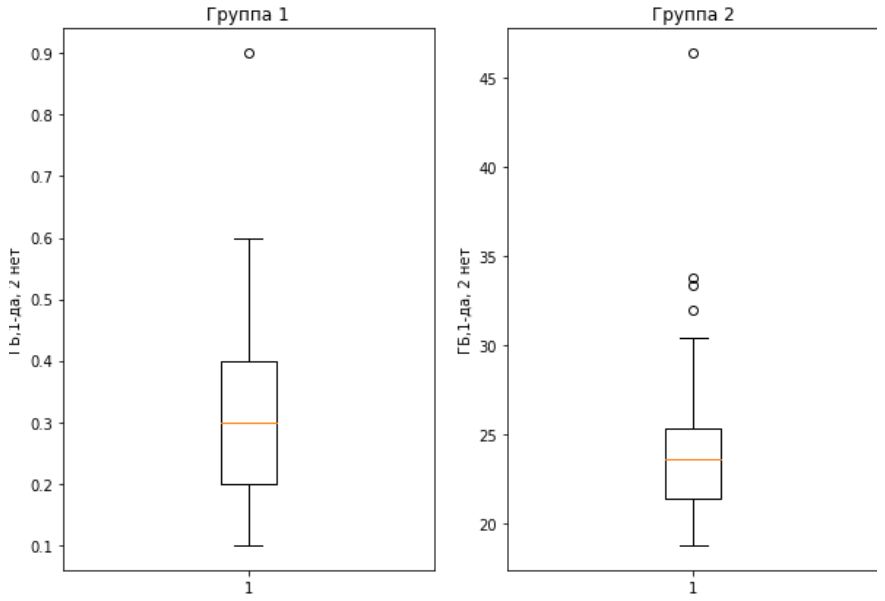



Группа 1
Артерия_стентирования
Группа 2
Артерия_стентирования
Ящик с усами. Группа 1
АСТ
Ящик с усами. Группа 0
АСТ
Ящик с усами. Группа 1
билирубин 1
Ящик с усами. Группа 0
билирубин 1

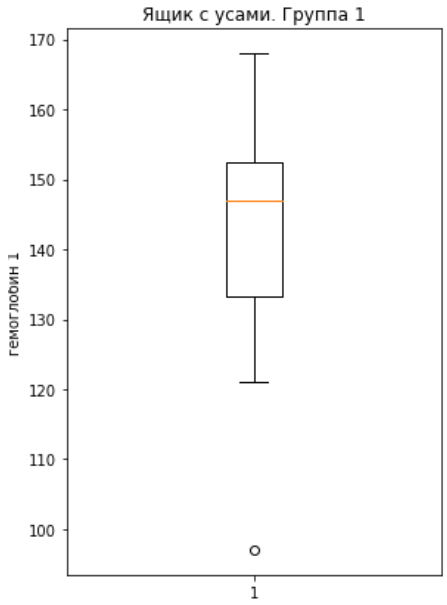

Ящик с усами. Группа 1
Вес
110
100
90
80
70
60
1
Ящик с усами. Группа 0
Вес
82.5
80.0
77.5
75.0
72.5
70.0
67.5
65.0
1

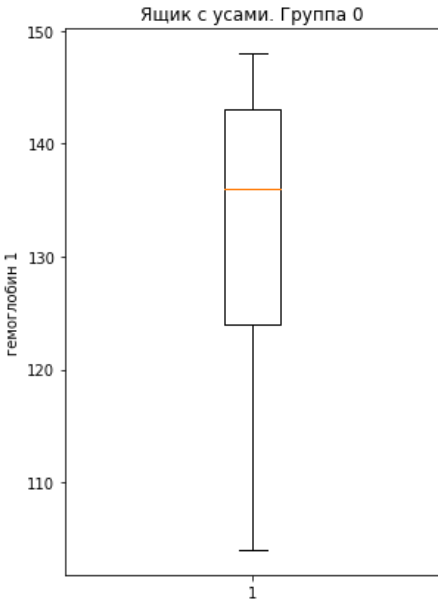

Ящик с усами. Группа 1
Возраст развития СД
60
50
40
30
20
10
0
1
Ящик с усами. Группа 0
Возраст развития СД
70
60
50
40
30
20
10
0
1

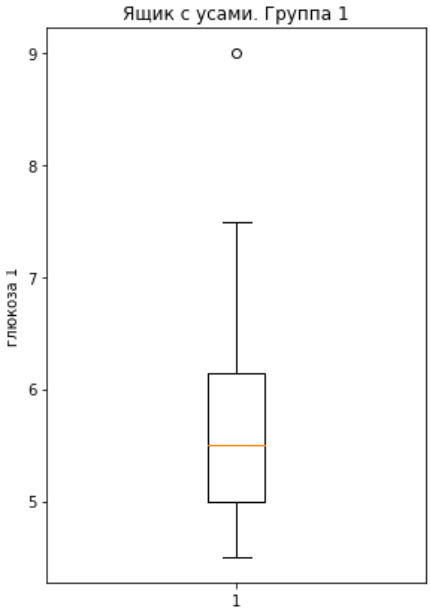

Ящик с усами. Группа 1
Возраст
75
70
65
60
55
50
45
40
1
Ящик с усами. Группа 0
Возраст
68
67
66
65
64
63
62
1

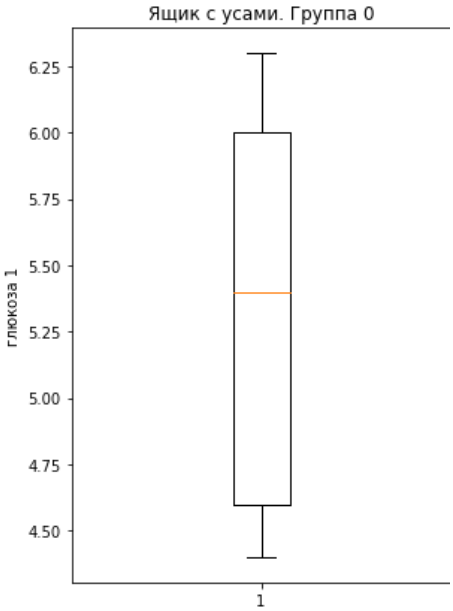

Группа 1
I b,1-да, 2 нет
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
1

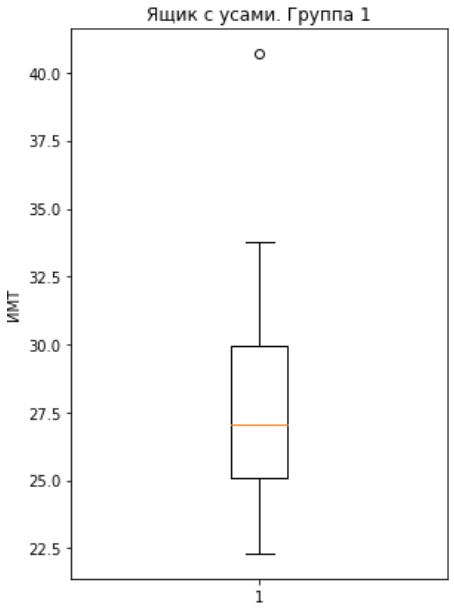
Группа 2
ГБ,1-да, 2 нет
45
40
35
30
25
20
1

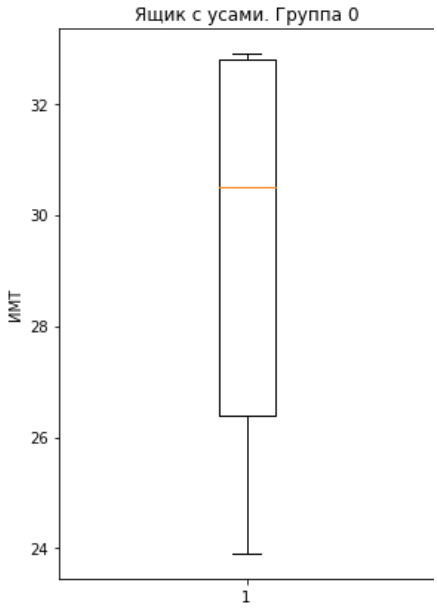

Ящик с усами. Группа 1
гемоглобин 1
170
160
150
140
130
120
110
100
1

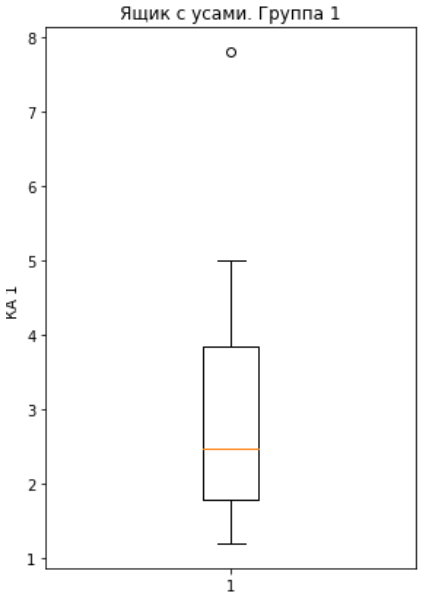

Ящик с усами. Группа 0
гемоглобин 1
150
140
130
120
110
1

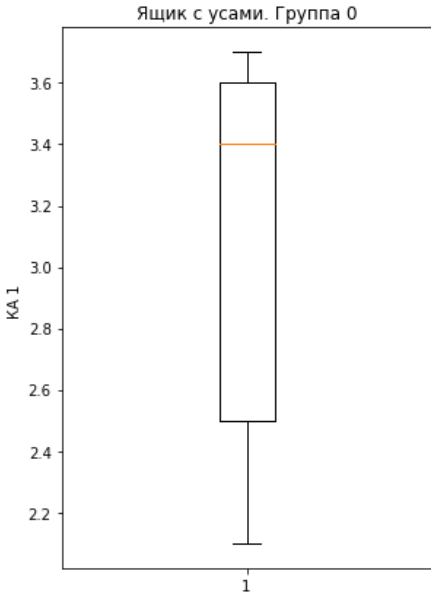

Ящик с усами. Группа 1
глюкоза 1
9
8
7
6
5
1

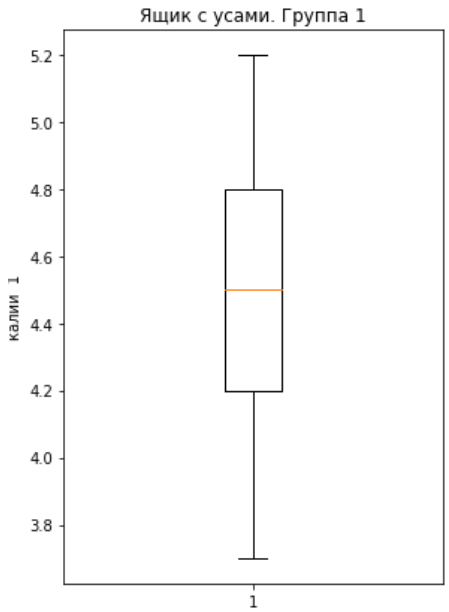

Ящик с усами. Группа 0
глюкоза 1
6.25
6.00
5.75
5.50
5.25
5.00
4.75
4.50
1

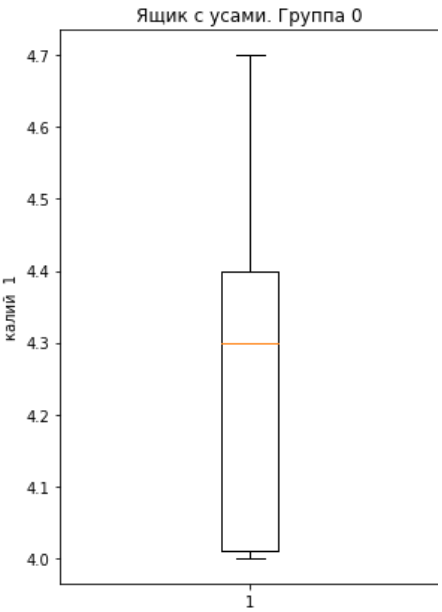

Ящик с усами. Группа 1
ИМТ
40.0
37.5
35.0
32.5
30.0
27.5
25.0
22.5
1

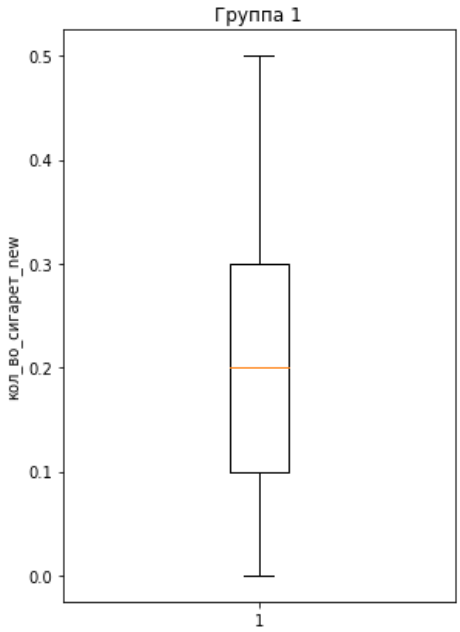

Ящик с усами. Группа 0
ИМТ
32
30
28
26
24
1

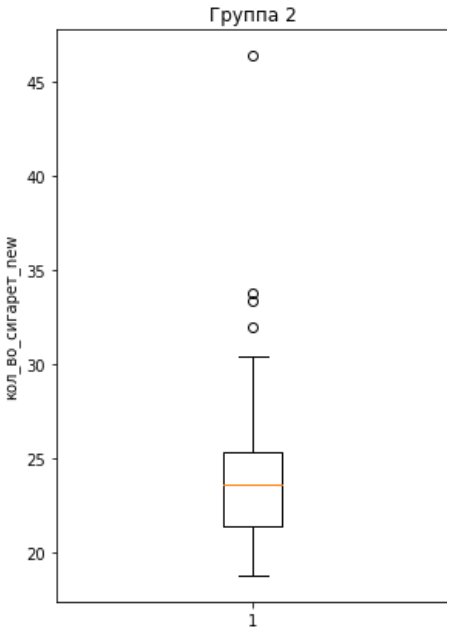

Ящик с усами. Группа 1
КА 1
8
7
6
5
4
3
2
1
1

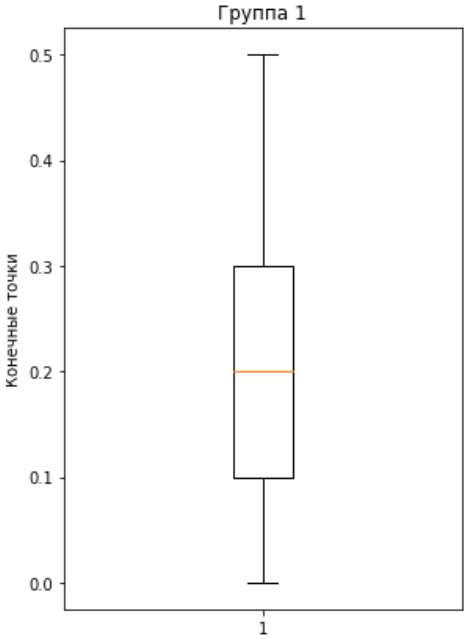

Ящик с усами. Группа 0
КА 1
3.6
3.4
3.2
3.0
2.8
2.6
2.4
2.2
1

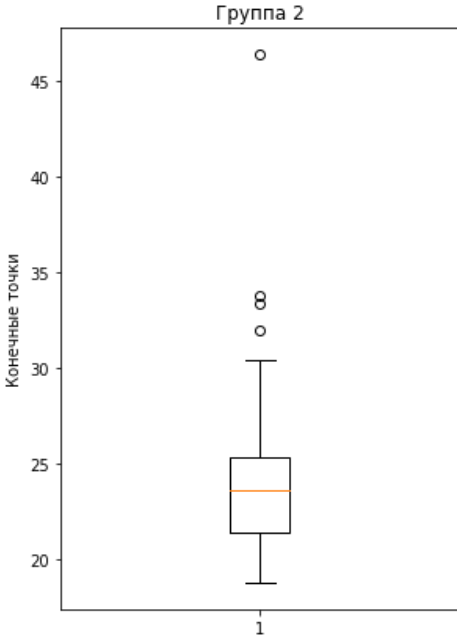

Ящик с усами. Группа 1
калии 1
5.2
5.0
4.8
4.6
4.4
4.2
4.0
3.8
1

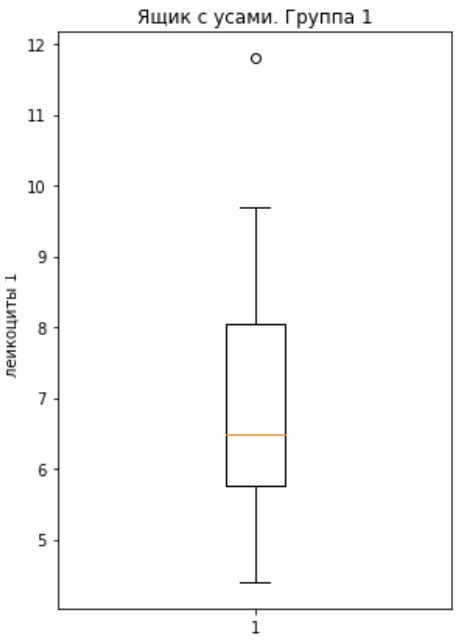

Ящик с усами. Группа 0
калий 1
4.7
4.6
4.5
4.4
4.3
4.2
4.1
4.0
1

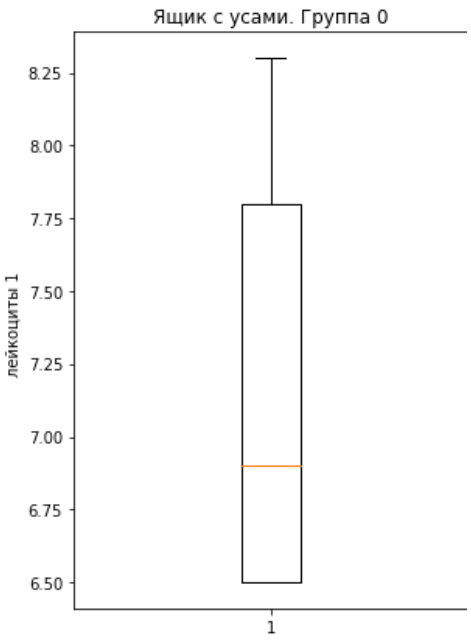

Группа 1
кол_во_сигарет_new
0.5
0.4
0.3
0.2
0.1
0.0
1

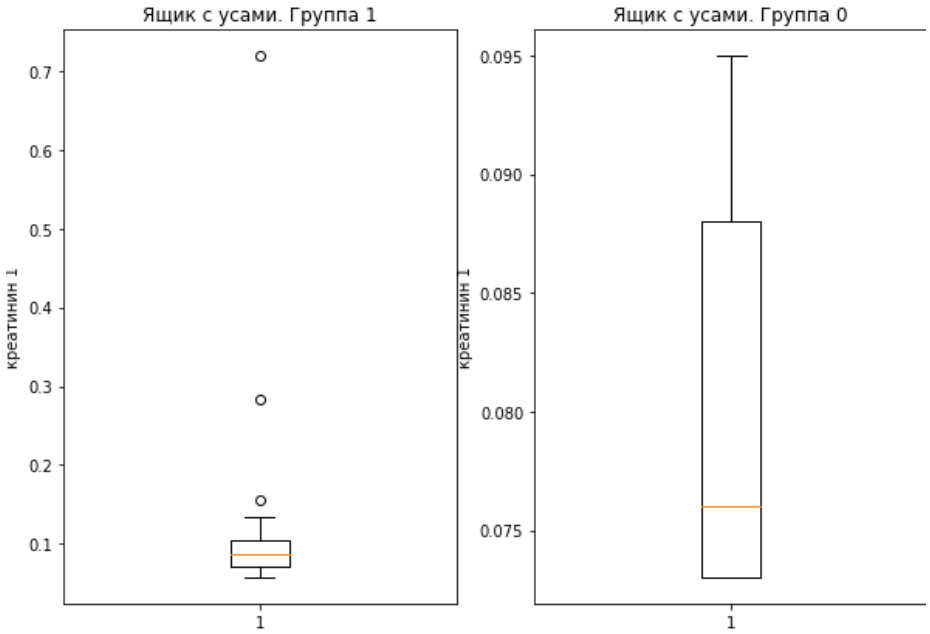

Группа 2
кол_во_сигарет_new
45
40
35
30
25
20
1


Группа 1
Конечные точки
0.5
0.4
0.3
0.2
0.1
0.0
1

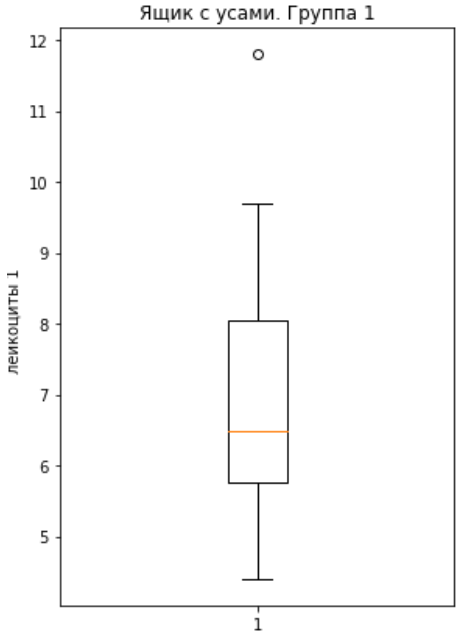

Группа 2
Конечные точки
45
40
35
30
25
20
1

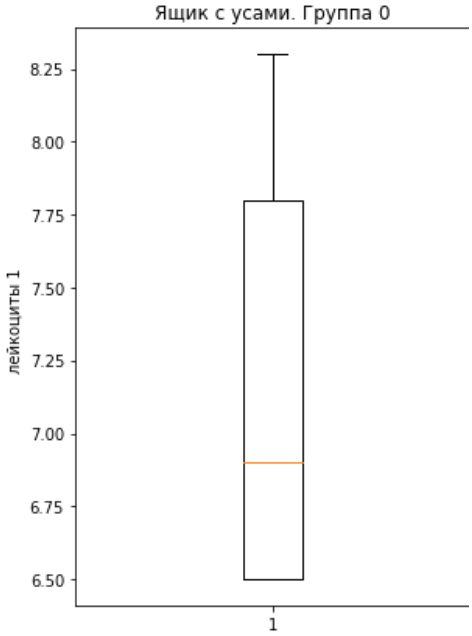

Ящик с усами. Группа 1
лейкоциты 1
12
11
10
9
8
7
6
5
1

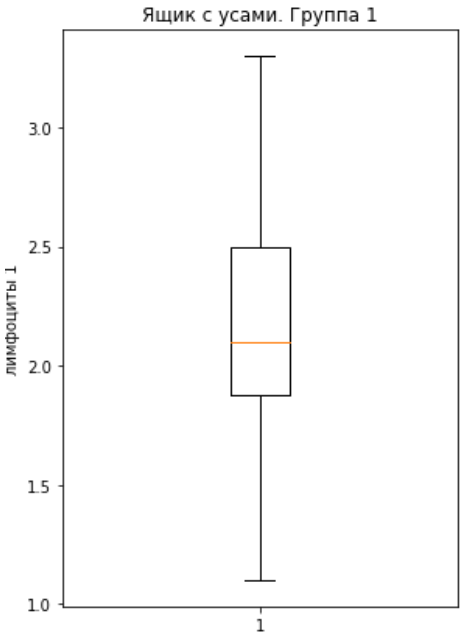

Ящик с усами. Группа 0
лейкоциты 1
8.25
8.00
7.75
7.50
7.25
7.00
6.75
6.50
1

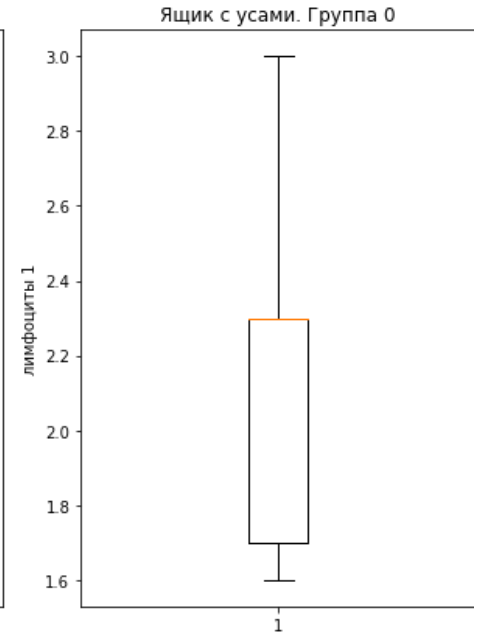

Ящик с усами. Группа 1
креатинин 1
0.7
0.6
0.5
0.4
0.3
0.2
0.1
1

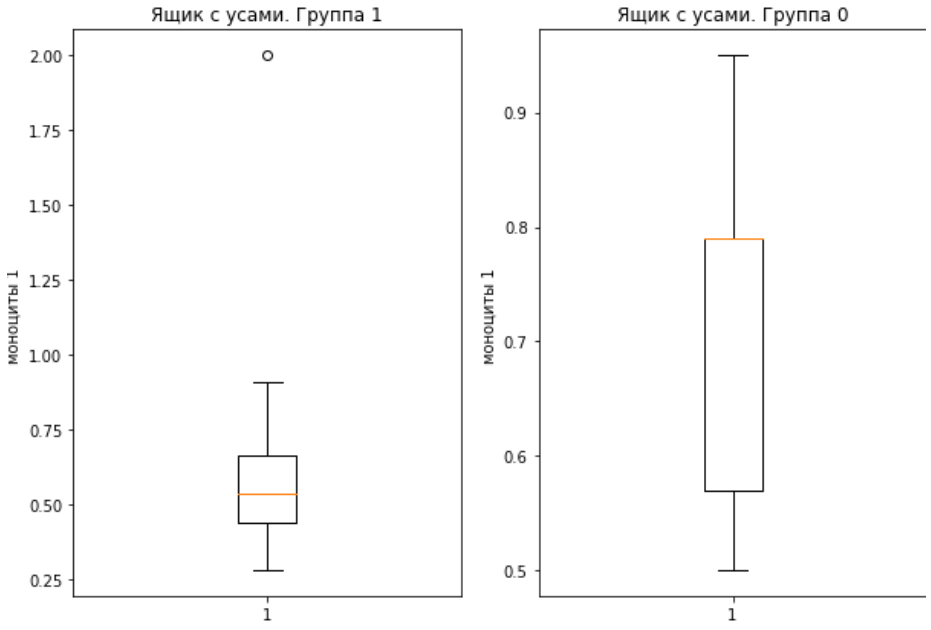
Ящик с усами. Группа 0
креатинин 1
0.095
0.090
0.085
0.080
0.075
1


Ящик с усами. Группа 1
лейкоциты 1
12
11
10
9
8
7
6
5
1

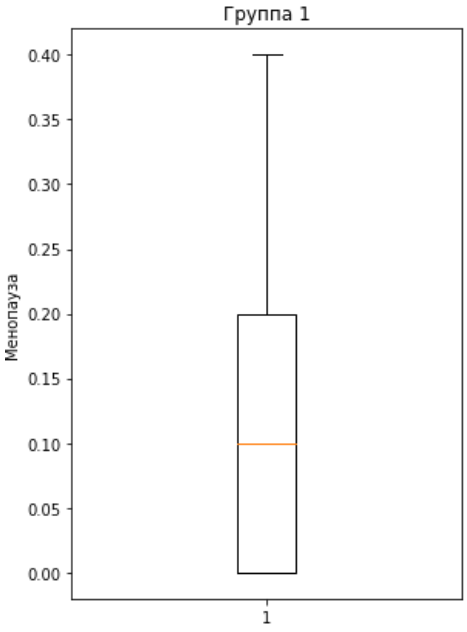

Ящик с усами. Группа 0
лейкоциты 1
8.25
8.00
7.75
7.50
7.25
7.00
6.75
6.50
1

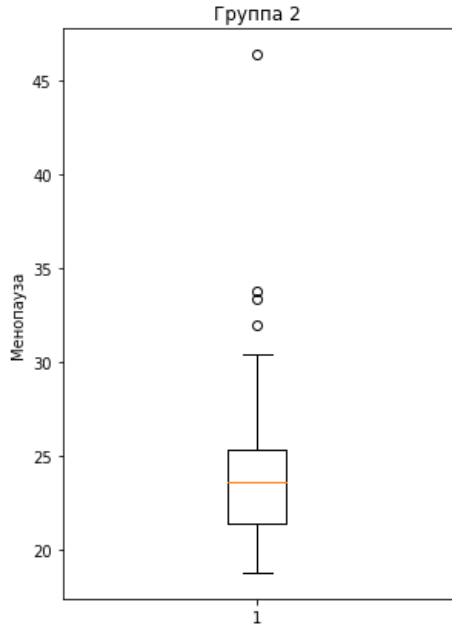

Ящик с усами. Группа 1
лимфоциты 1
3.0
2.5
2.0
1.5
1.0
1

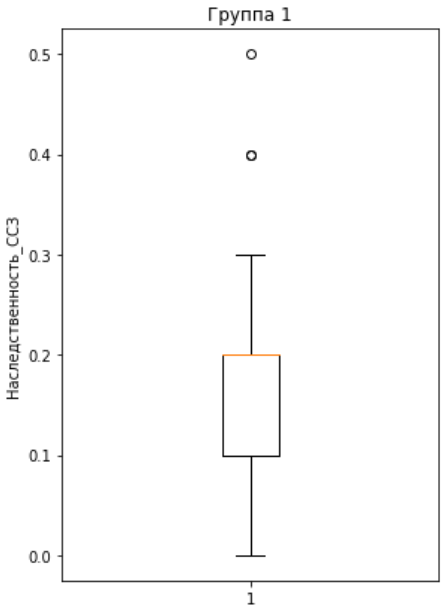

Ящик с усами. Группа 0
лимфоциты 1
3.0
2.8
2.6
2.4
2.2
2.0
1.8
1.6
1

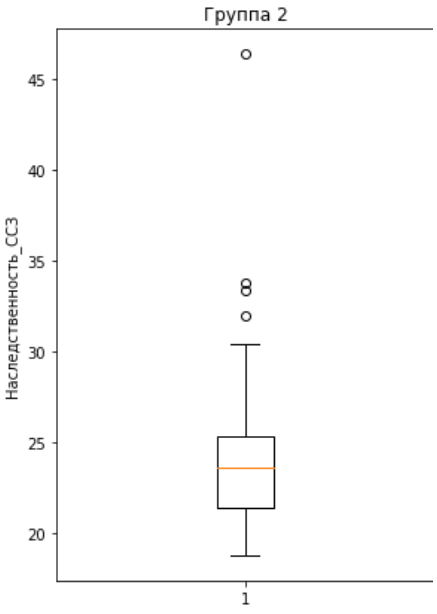

Ящик с усами. Группа 1
моноциты 1
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
1

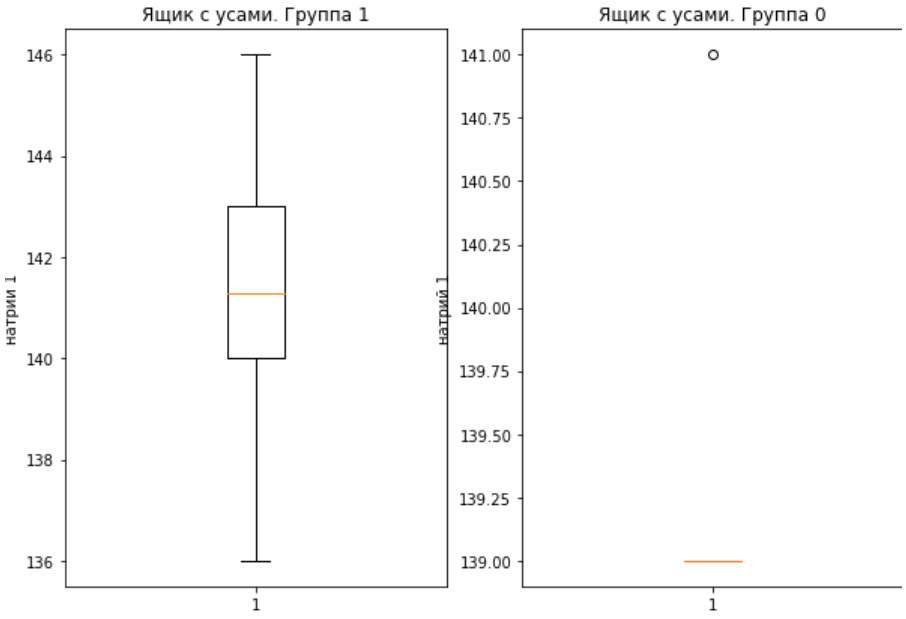
Ящик с усами. Группа 0
моноциты 1
0.9
0.8
0.7
0.6
0.5
1

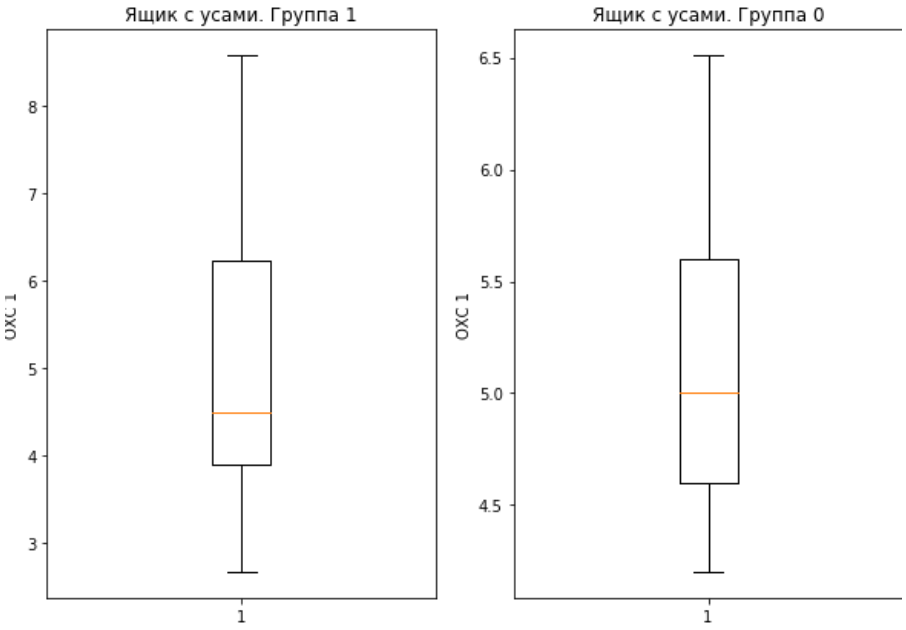

Группа 1
Менопауза
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
1

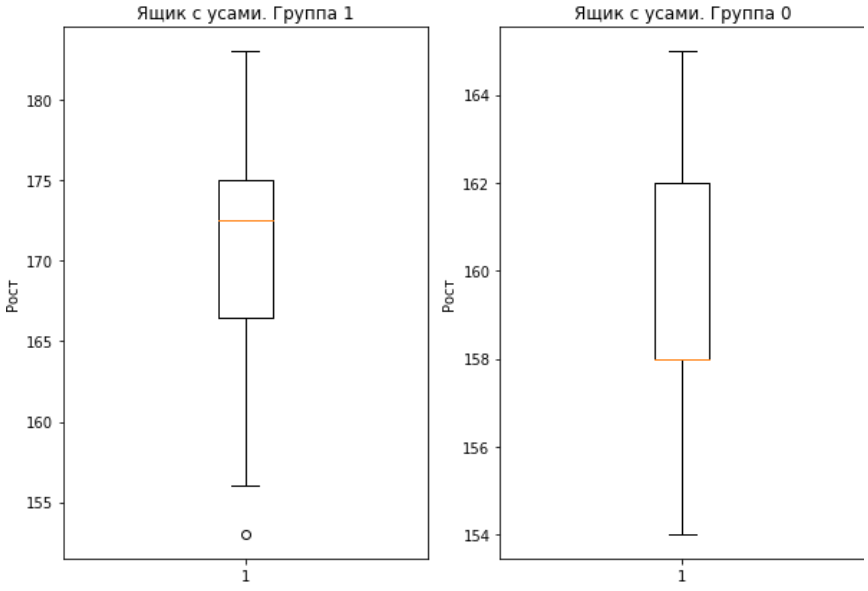

Группа 2
Менопауза
45
40
35
30
25
20
1

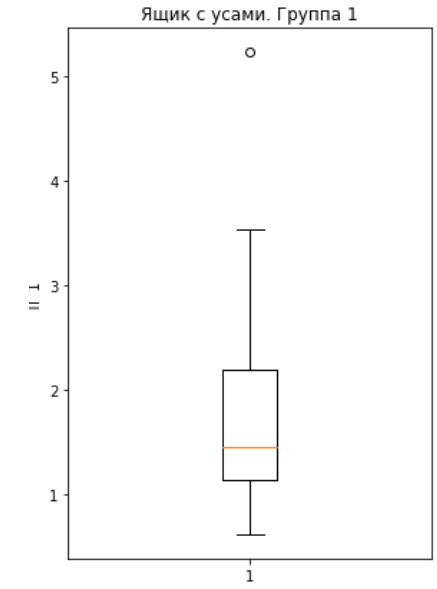

Группа 1
Наследственность_ССЗ
0.5
0.4
0.3
0.2
0.1
0.0
1

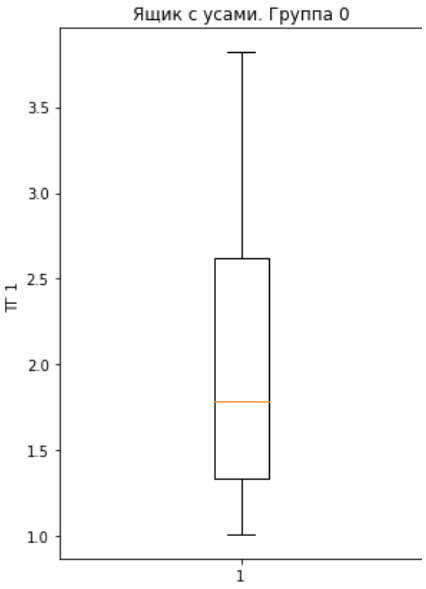

Группа 2
Наследственность_ССЗ
45
40
35
30
25
20
1

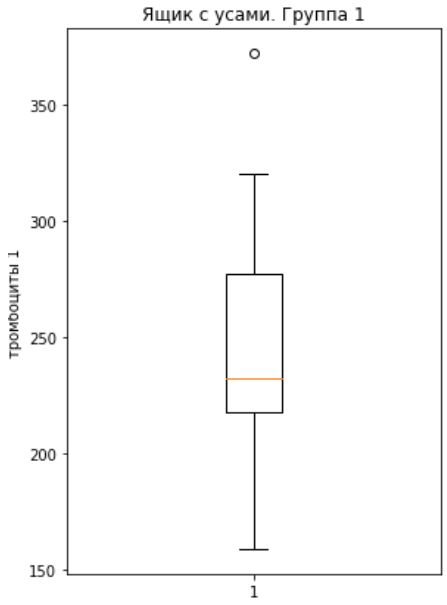

Ящик с усами. Группа 1
натрии 1
146
144
142
140
138
136
1
Ящик с усами. Группа 0
натрий 1
141.00
140.75
140.50
140.25
140.00
139.75
139.50
139.25
139.00
1

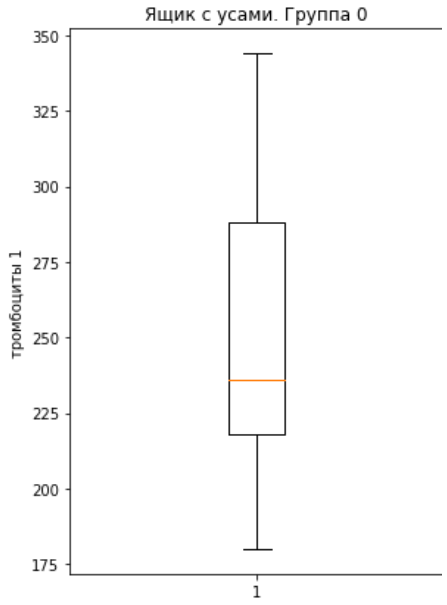

Ящик с усами. Группа 1
ОХС 1
8
7
6
5
4
3
1
Ящик с усами. Группа 0
ОХС 1
6.5
6.0
5.5
5.0
4.5
1

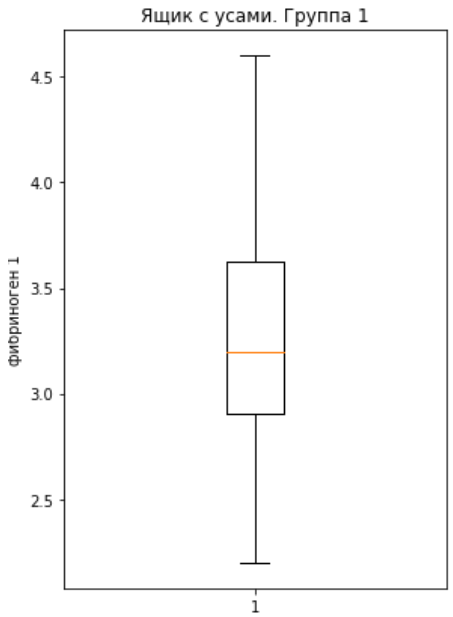

Ящик с усами. Группа 1
Рост
180
175
170
165
160
155
1
Ящик с усами. Группа 0
Рост
164
162
160
158
156
154
1

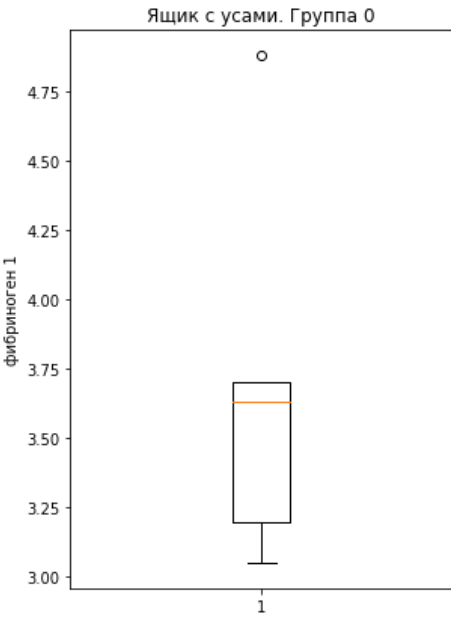

Ящик с усами. Группа 1
II 1
1

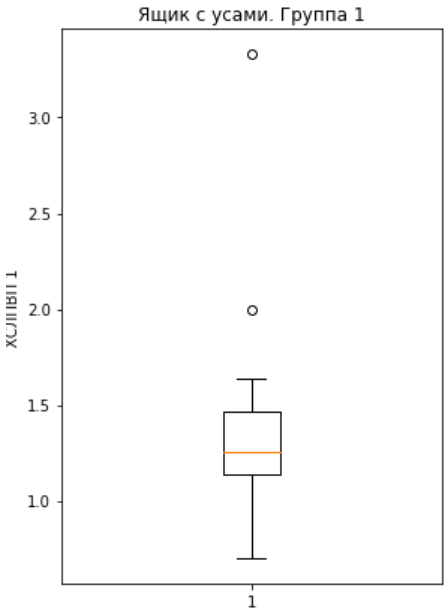

Ящик с усами. Группа 0
TT 1
1

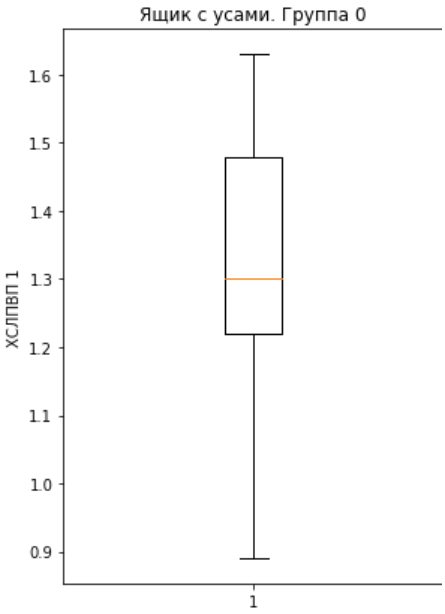

Ящик с усами. Группа 1
тромбоциты 1
1

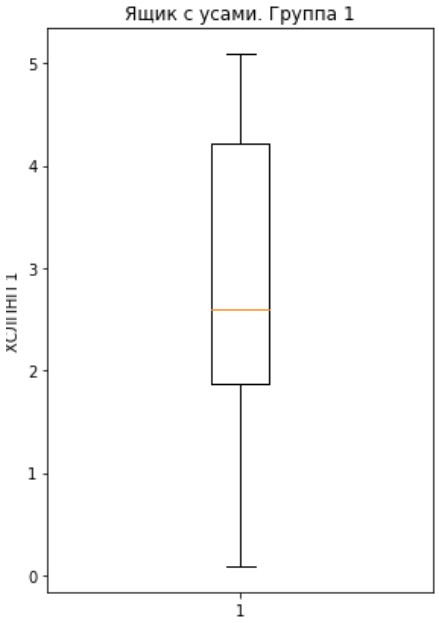

Ящик с усами. Группа 0
тромбоциты 1
1

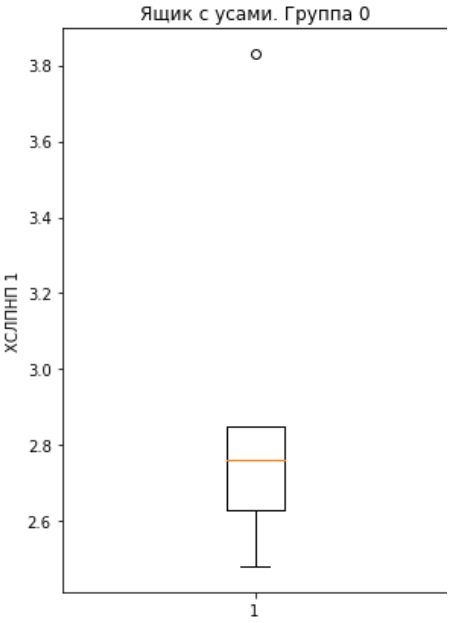

Ящик с усами. Группа 1
фибриноген 1
1

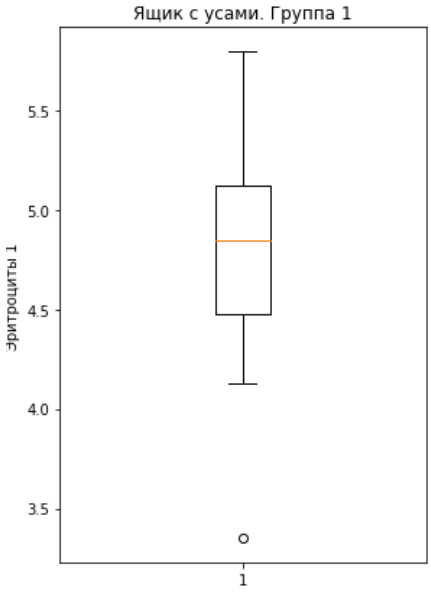

Ящик с усами. Группа 0
фибриноген 1
1

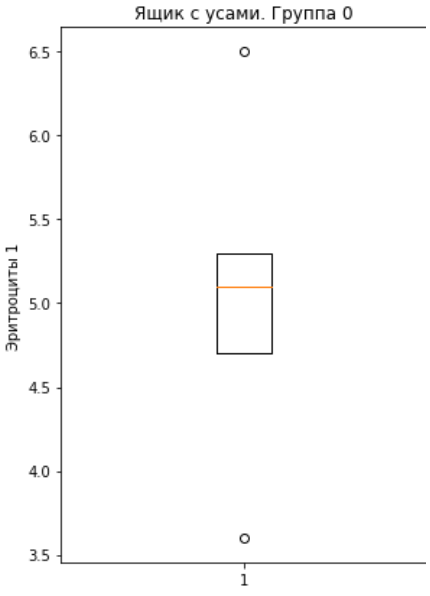

Ящик с усами. Группа 1
ХСЛПВП 1
3.0
2.5
2.0
1.5
1.0
1

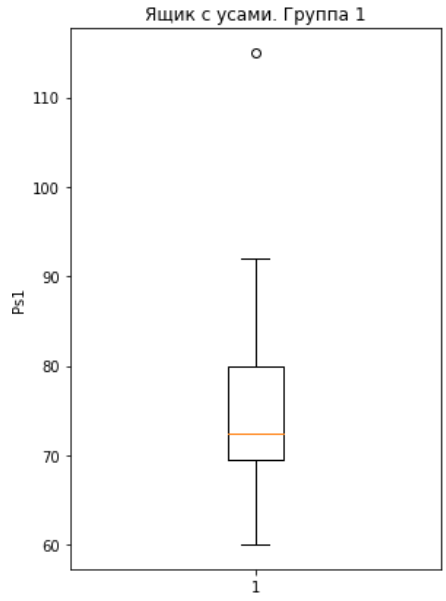

Ящик с усами. Группа 0
ХСЛПВП 1
1.6
1.5
1.4
1.3
1.2
1.1
1.0
0.9
1

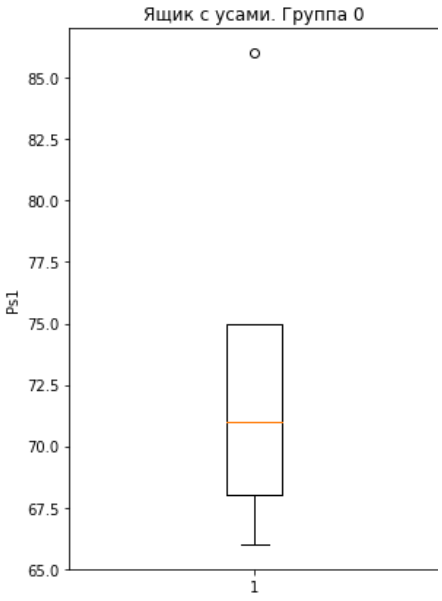

Ящик с усами. Группа 1
ХСЛПНП 1
5
4
3
2
1
0
1

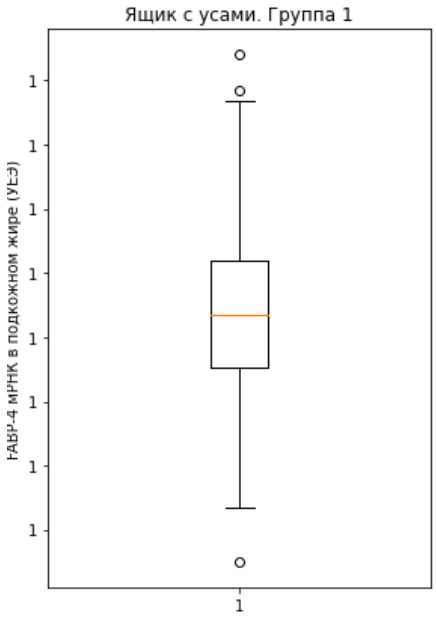

Ящик с усами. Группа 0
ХСЛПНП 1
3.8
3.6
3.4
3.2
3.0
2.8
2.6
1

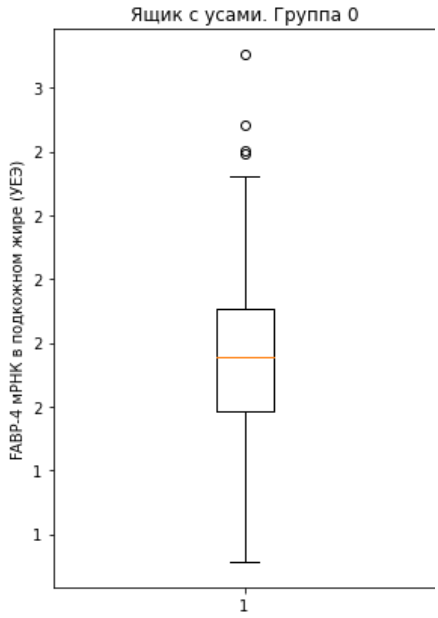

Ящик с усами. Группа 1
Эритроциты 1
5.5
5.0
4.5
4.0
3.5
1

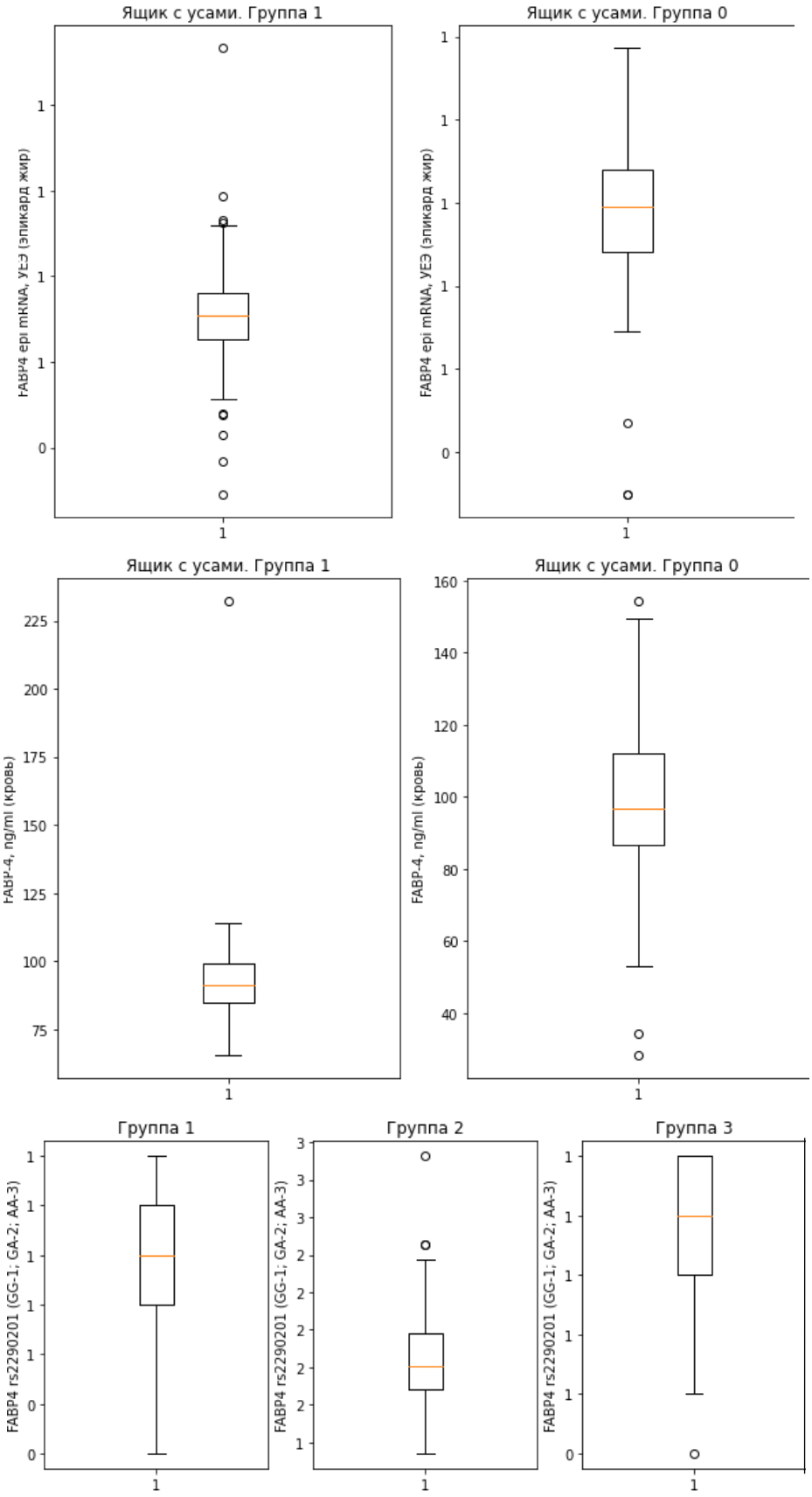

Ящик с усами. Группа 0
Эритроциты 1
6.5
6.0
5.5
5.0
4.5
4.0
3.5
1

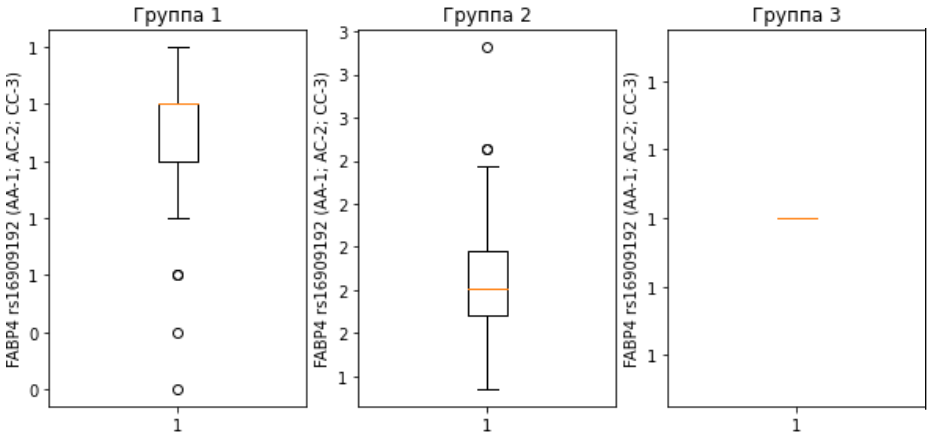

Ящик с усами. Группа 1
Ps1
110
100
90
80
70
60
1


Ящик с усами. Группа 0
Ps1
85.0
82.5
80.0
77.5
75.0
72.5
70.0
67.5
65.0
1


Ящик с усами. Группа 1
FABP-4 мРНК в подкожном жире (УЕЭ)
1
1
1
1
1
1
1
1
1
1

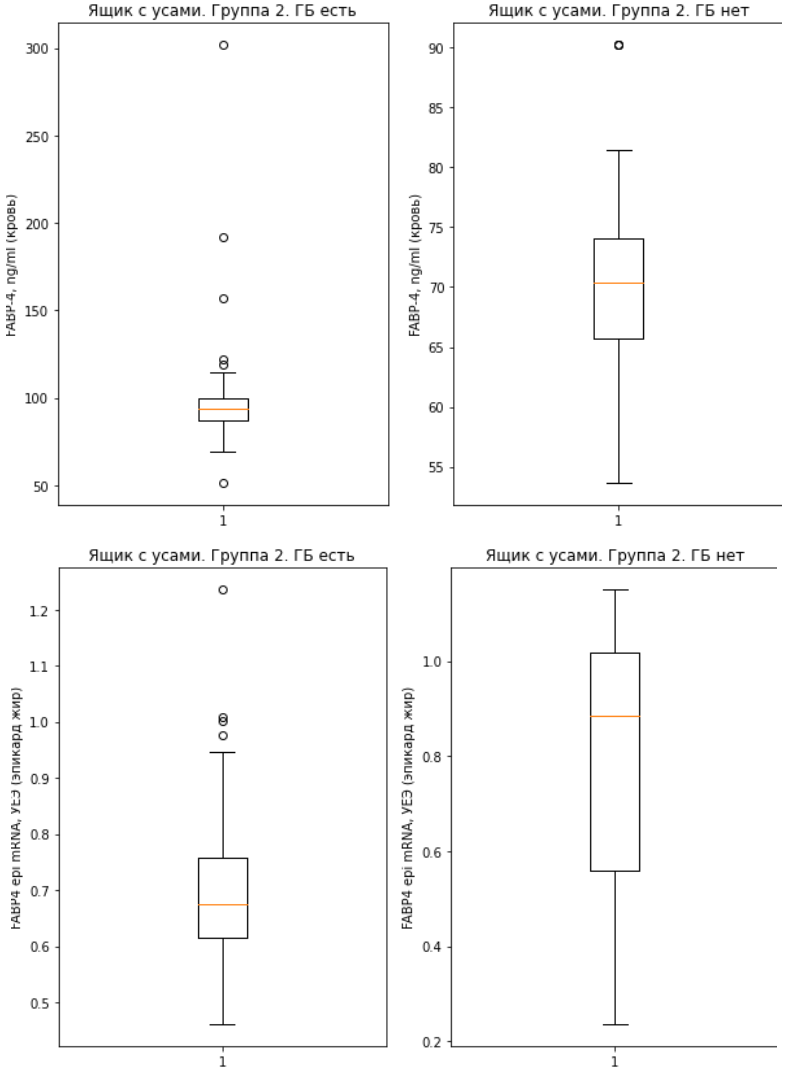

Ящик с усами. Группа 0
FABP-4 мРНК в подкожном жире (УЕЭ)
3
2
2
2
2
2
1
1
1

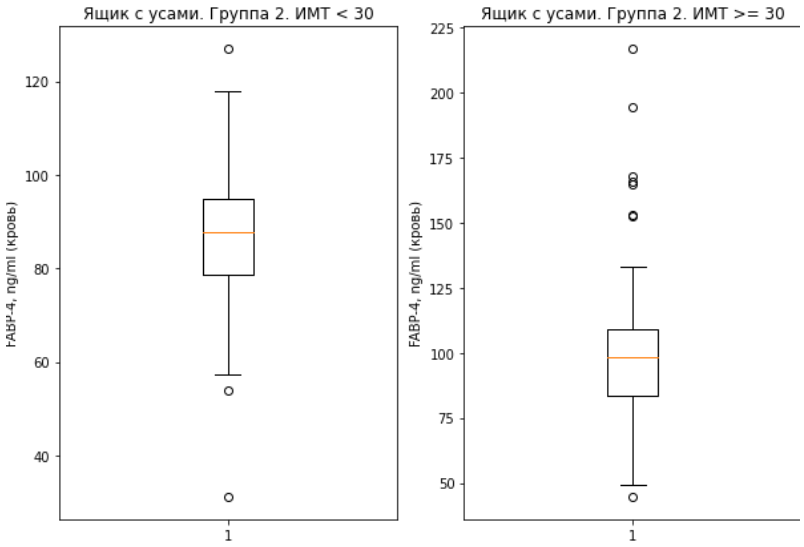

Ящик с усами. Группа 1
FABP4 epi mRNA, УЕЭ (эпикард жир)
Ящик с усами. Группа 0
FABP4 epi mRNA, УЕЭ (эпикард жир)
Ящик с усами. Группа 1
FABP-4, ng/ml (кровь)
Ящик с усами. Группа 0
FABP-4, ng/ml (кровь)
Группа 1
FABP4 rs2290201 (GG-1; GA-2; AA-3)
Группа 2
FABP4 rs2290201 (GG-1; GA-2; AA-3)
Группа 3
FABP4 rs2290201 (GG-1; GA-2; AA-3)



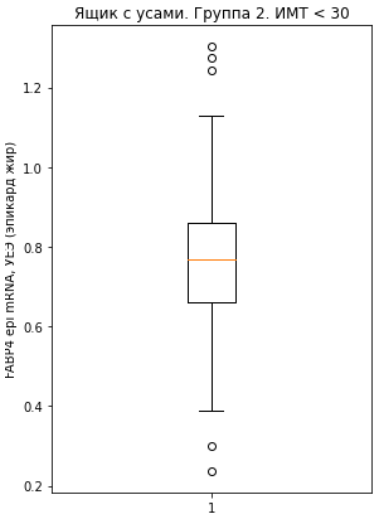
Приложение 1.

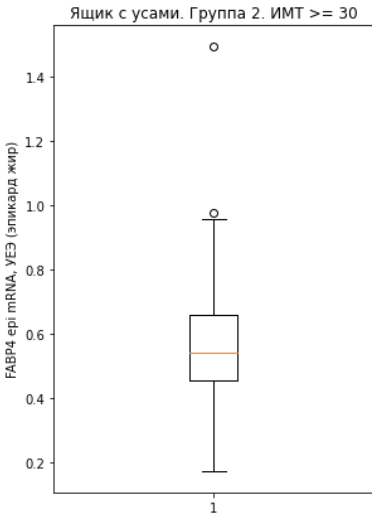
Графики ящиков с усами для репрезентативных факторов влияния на уровень FABP4 в крови и эпидуральном жире для группы людей с ИБС:

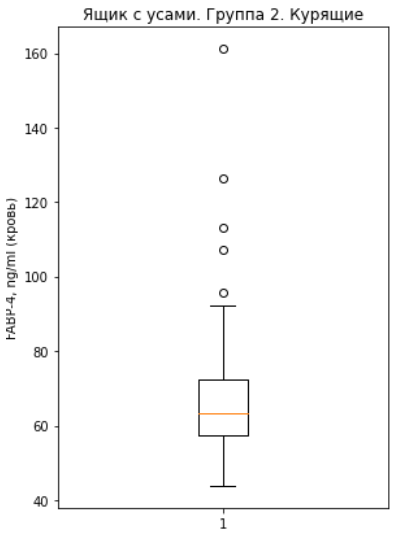


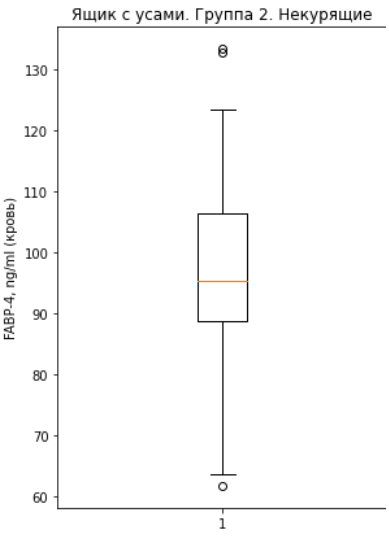

Ящик с усами. Группа 2. ИМТ < 30
FABP-4, ng/ml (кровь)
120
100
80
60
40
1

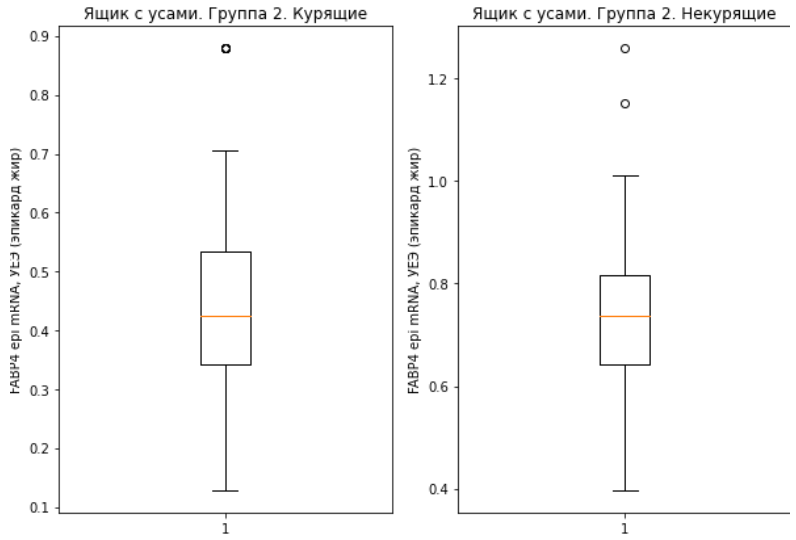
Ящик с усами. Группа 2. ИМТ >= 30
FABP-4, ng/ml (кровь)
225
200
175
150
125
100
75
50
1


Ящик с усами. Группа 2. ИМТ < 30
FABP4 epi mRNA, УЕЭ (эпикард жир)
1.2
1.0
0.8
0.6
0.4
0.2
1

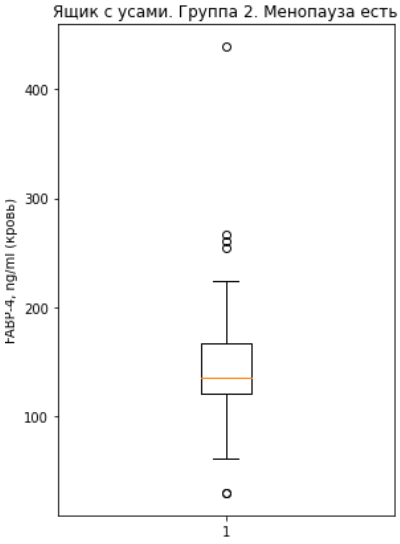

Ящик с усами. Группа 2. ИМТ >= 30
FABP4 epi mRNA, УЕЭ (эпикард жир)
1.4
1.2
1.0
0.8
0.6
0.4
0.2
1

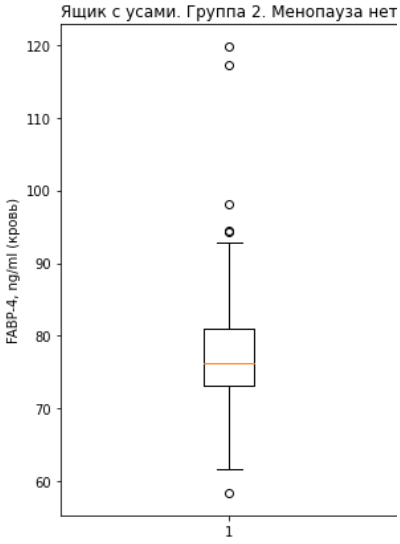

Ящик с усами. Группа 2. Курящие
FABP-4, ng/ml (кровь)
160
140
120
100
80
60
40
1

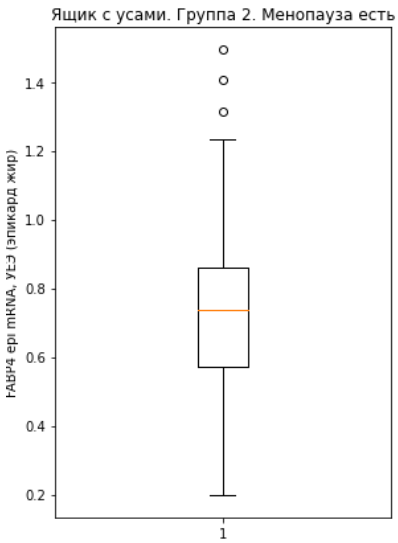

Ящик с усами. Группа 2. Некурящие
FABP-4, ng/ml (кровь)
130
120
110
100
90
80
70
60
1

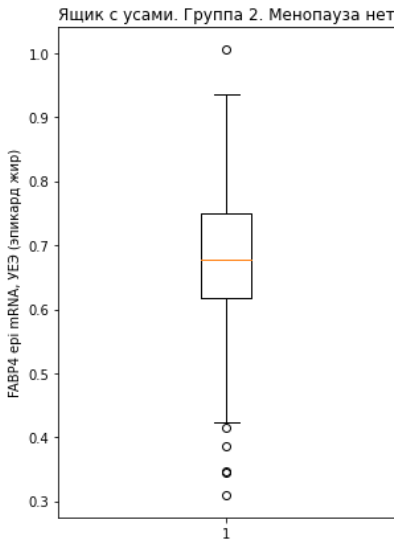

Ящик с усами. Группа 2. Курящие
FABP4 epi mRNA, УЕЭ (эпикард жир)
1

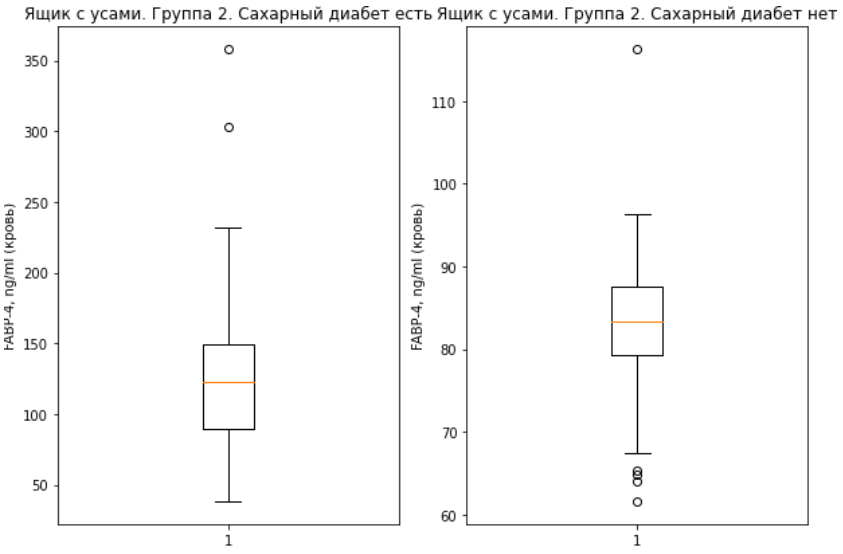
Ящик с усами. Группа 2. Некурящие
FABP4 epi mRNA, УЕЭ (эпикард жир)
1

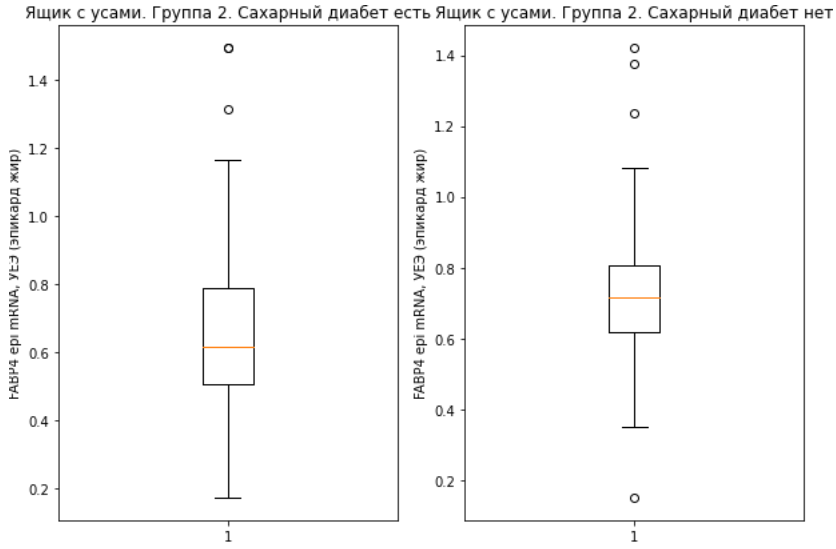

Ящик с усами. Группа 2. Менопауза есть
FABP-4, ng/ml (кровь)
1

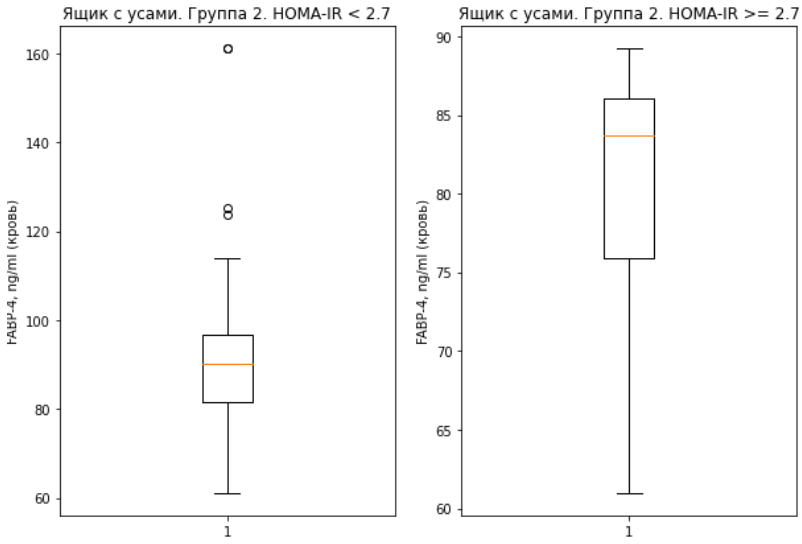

Ящик с усами. Группа 2. Менопауза нет
FABP-4, ng/ml (кровь)
1

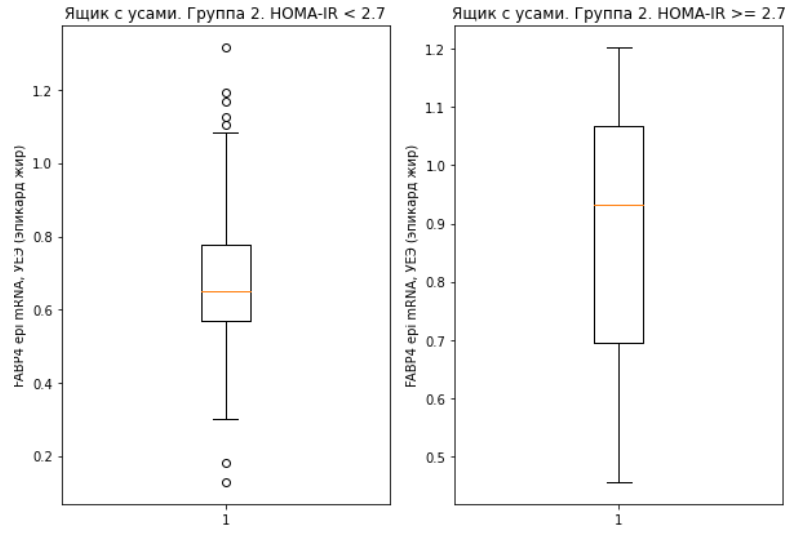

Ящик с усами. Группа 2. Менопауза есть
FABP4 epi mRNA, УЕЭ (эпикард жир)
1

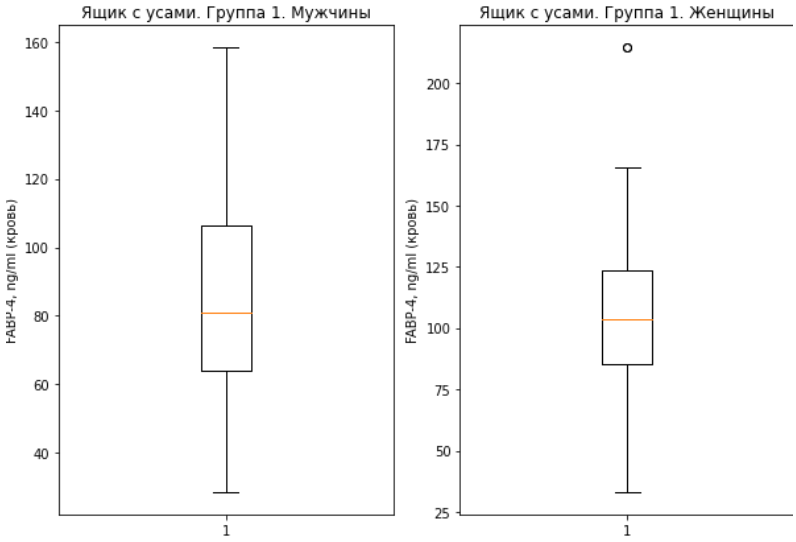

Ящик с усами. Группа 2. Менопауза нет
FABP4 epi mRNA, УЕЭ (эпикард жир)
1

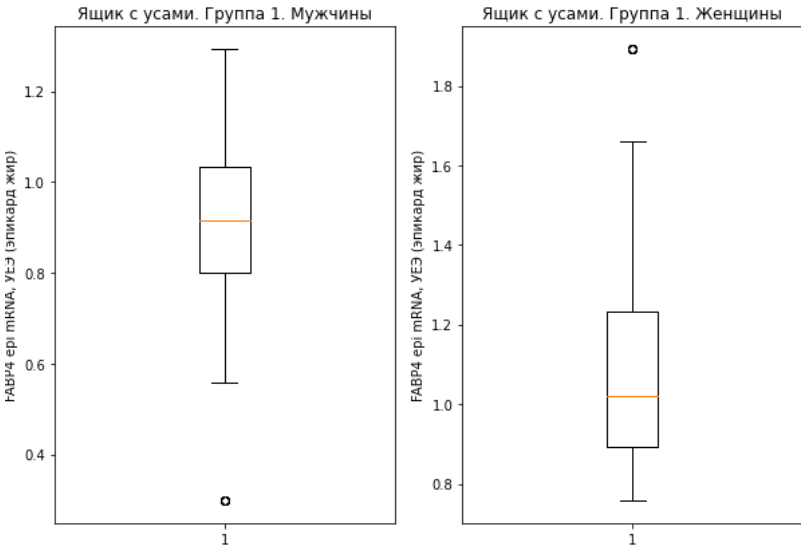

Ящик с усами. Группа 2. Сахарный диабет есть
FABP-4, ng/ml (кровь)
Ящик с усами. Группа 2. Сахарный диабет нет
FABP-4, ng/ml (кровь)

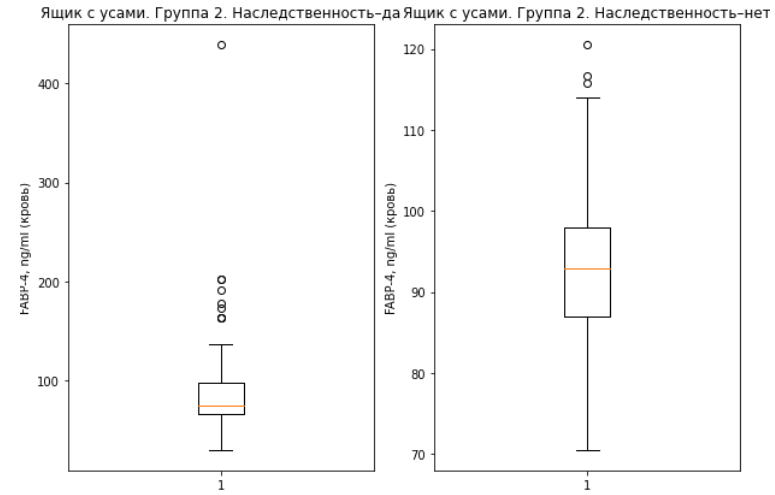

Ящик с усами. Группа 2. Сахарный диабет есть
FABP4 epi mRNA, УЕЭ (эпикард жир)
Ящик с усами. Группа 2. Сахарный диабет нет
FABP4 epi mRNA, УЕЭ (эпикард жир)

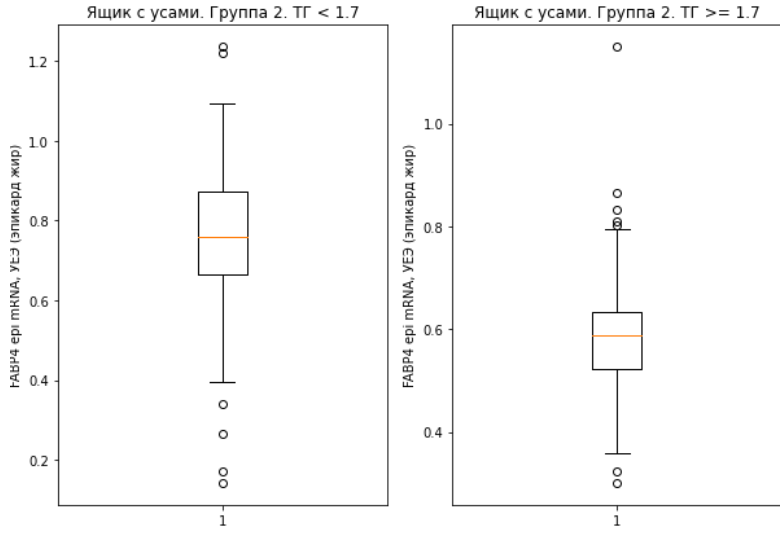

Ящик с усами. Группа 2. HOMA-IR < 2.7
FABP-4, ng/ml (кровь)
Ящик с усами. Группа 2. HOMA-IR >= 2.7
FABP-4, ng/ml (кровь)

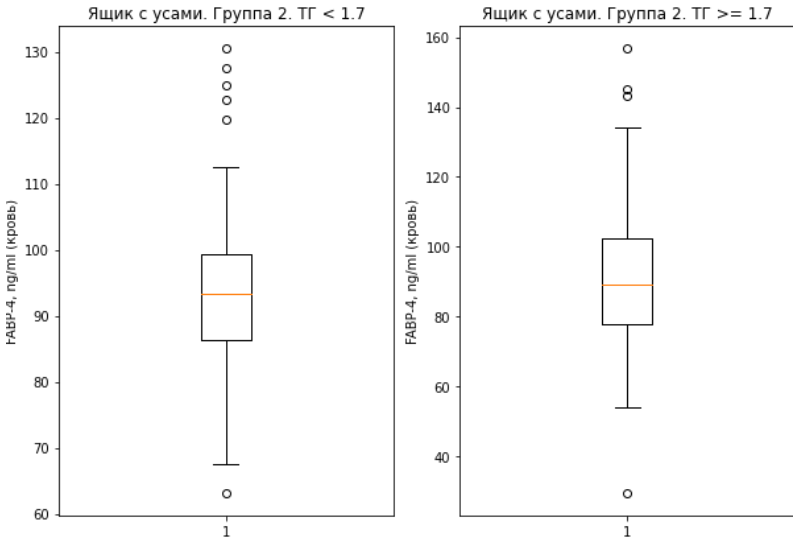

Ящик с усами. Группа 2. HOMA-IR < 2.7
FABP4 epi mRNA, УЕЭ (эпикард жир)
1
Ящик с усами. Группа 2. HOMA-IR >= 2.7
FABP4 epi mRNA, УЕЭ (эпикард жир)
1

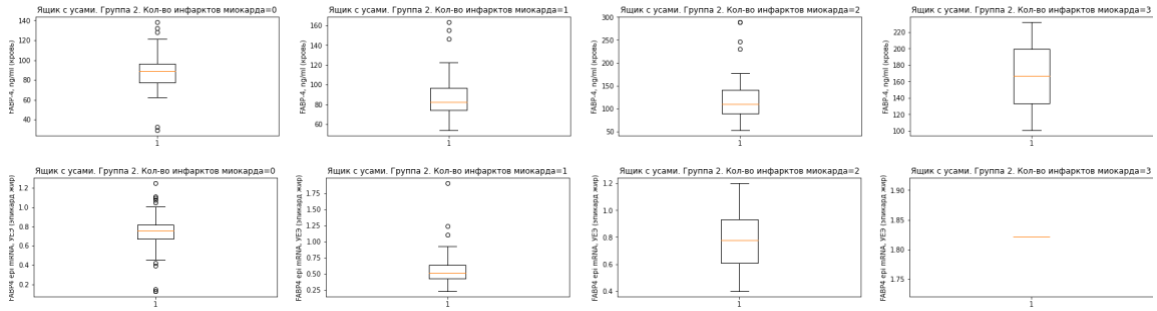

Ящик с усами. Группа 1. Мужчины
FABP-4, ng/ml (кровь)
1
Ящик с усами. Группа 1. Женщины
FABP-4, ng/ml (кровь)
1

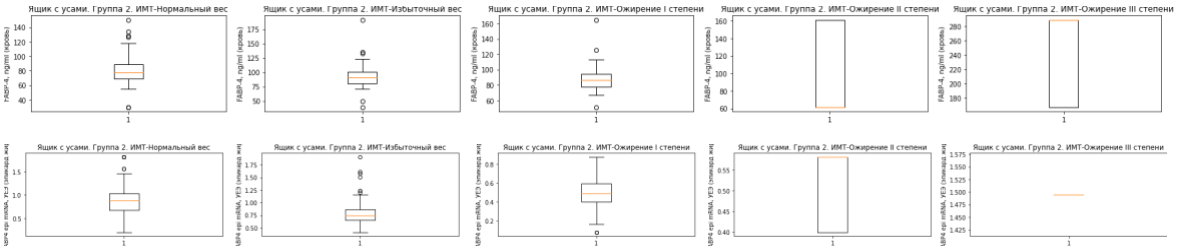

Ящик с усами. Группа 1. Мужчины
FABP4 epi mRNA, УЕЭ (эпикард жир)
1
Ящик с усами. Группа 1. Женщины
FABP4 epi mRNA, УЕЭ (эпикард жир)
1

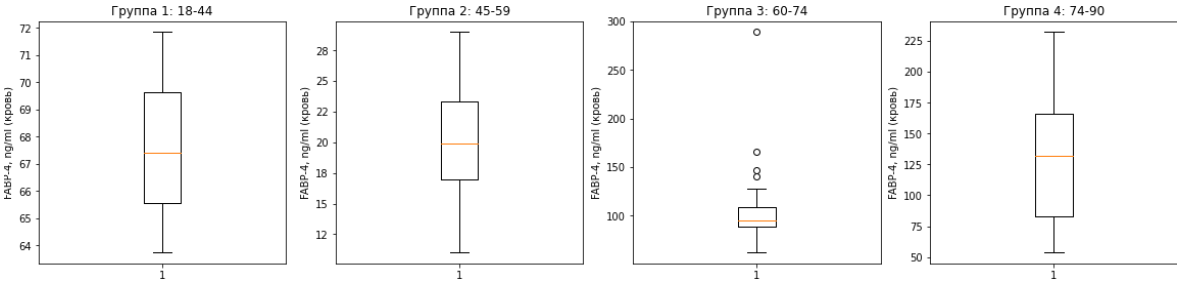

Ящик с усами. Группа 2. Наследственность-да
FABP-4, ng/ml (кровь)
Ящик с усами. Группа 2. Наследственность-нет
FABP-4, ng/ml (кровь)

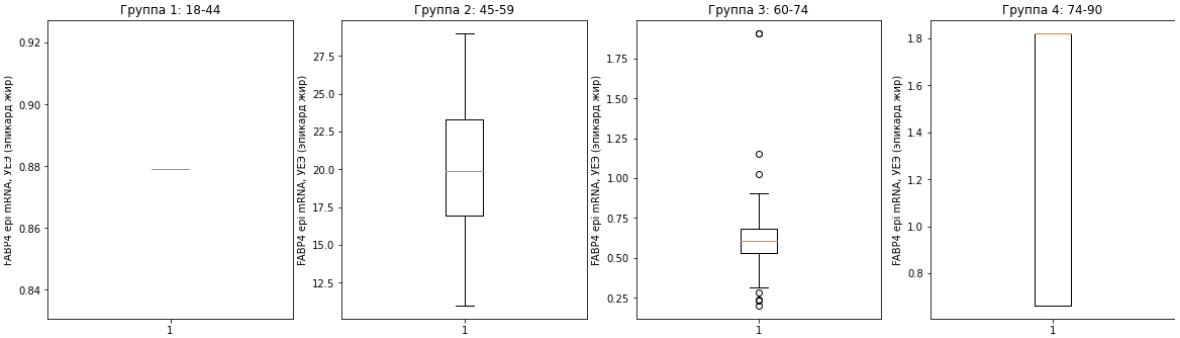

Ящик с усами. Группа 2. ТГ < 1.7
FABP4 epi mRNA, УЕЭ (эпикард жир)
Ящик с усами. Группа 2. ТГ >= 1.7
FABP4 epi mRNA, УЕЭ (эпикард жир)

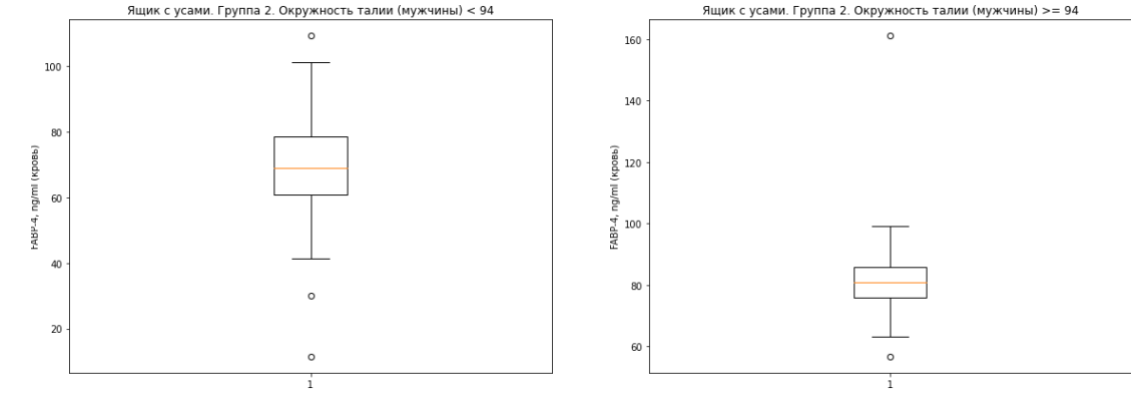

Ящик с усами. Группа 2. ТГ < 1.7
FABP-4, ng/ml (кровь)
Ящик с усами. Группа 2. ТГ >= 1.7
FABP-4, ng/ml (кровь)

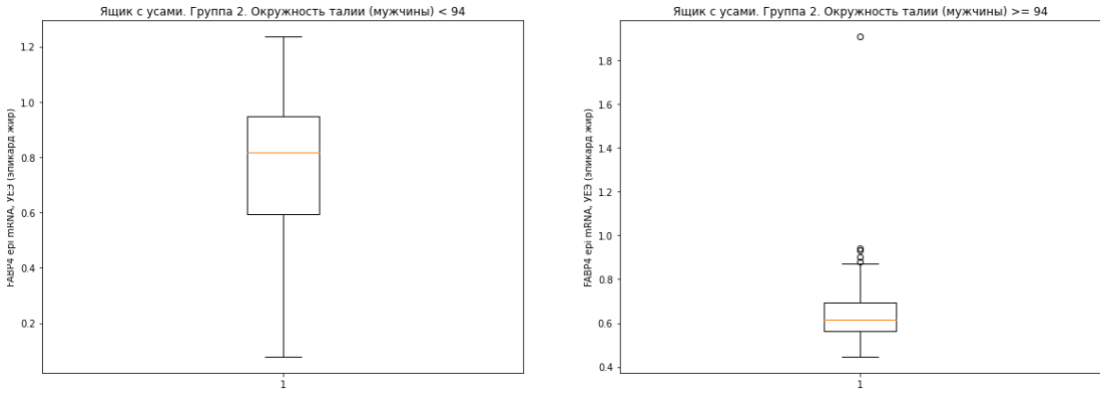

Ящик с усами. Группа 2. Кол-во инфарктов миокарда=0
Ящик с усами. Группа 2. Кол-во инфарктов миокарда=1
Ящик с усами. Группа 2. Кол-во инфарктов миокарда=2
Ящик с усами. Группа 2. Кол-во инфарктов миокарда=3
FABP-4, ng/ml (кровь)
FABP4 epi mRNA, YC3 (эпикард жир)
FABP4 epi mRNA, YE3 (эпикард жир)


Ящик с усами. Группа 2. ИМТ-Нормальный вес
Ящик с усами. Группа 2. ИМТ-Избыточный вес
Ящик с усами. Группа 2. ИМТ-Ожирение I степени
Ящик с усами. Группа 2. ИМТ-Ожирение II степени
Ящик с усами. Группа 2. ИМТ-Ожирение III степени
FABP-4, ng/ml (кровь)
UBP4 epi mRNA, YE3 (эпикард жир)


Группа 1: 18-44
Группа 2: 45-59
Группа 3: 60-74
Группа 4: 74-90
FABP-4, ng/ml (кровь)

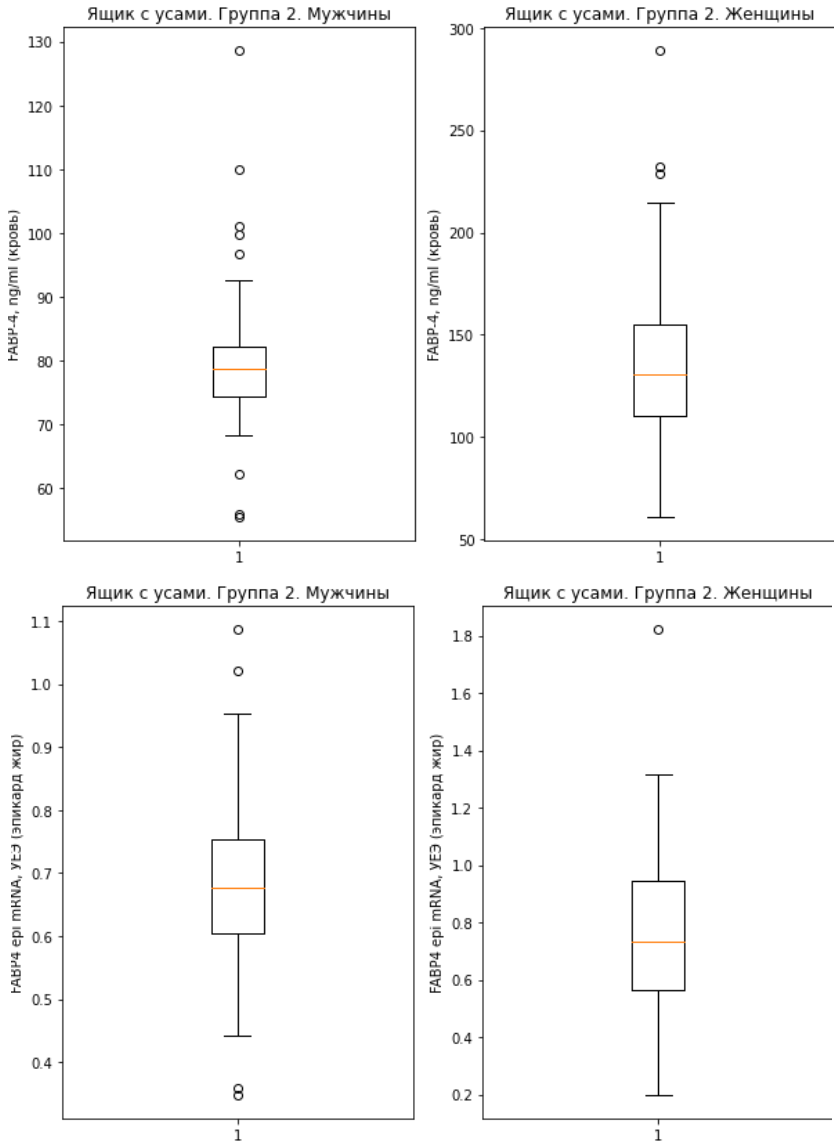

Группа 1: 18-44
Группа 2: 45-59
Группа 3: 60-74
Группа 4: 74-90
FABP4 epi mRNA, YE3 (эпикард жир)


Ящик с усами. Группа 2. Окружность талии (мужчины) < 94
Ящик с усами. Группа 2. Окружность талии (мужчины) >= 94
FABP-4, ng/ml (кровь)



Приложение 2.

Графики ящиков с усами для репрезентативных факторов влияния на уровень FABP4 в крови и эпидуральном жире для группы людей без ИБС:



Приложение 3.