

**Министерство науки и высшего образования Российской Федерации
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИТМО**

Факультет безопасности информационных технологий

Дисциплина:

«Математическая статистика»

ПРОЕКТНАЯ РАБОТА

«Статистический анализ взаимосвязи между вероятностью возникновения сердечно-сосудистых заболеваний и клиническими особенностями пациента»

Участники и их роли:

Баркова Александра - постановка задач участникам проекта, финальное объединение и проверка согласованности всех разделов, подготовка итоговой версии отчёта.

Звонарёва Елизавета - вступление, цели и задачи, информация о датасете, проведение t-тестов и построение визуализаций.

Зиновьев Тимофей - проработка теоретической части, составление и описание результатов, помощь с технической частью.

Юрицина Мария - идея и выбор датасета, реализация χ^2 -тестов и корреляционного анализа.

Санкт-Петербург

2025 г.

СОДЕРЖАНИЕ

Введение	3
1 Постановка задачи	4
1.1 Цели и задачи работы.....	4
1.2 Используемые программные средства	4
1.3 Математическая формулировка	5
1.4 Информация о датасете [Блок 1].....	5
2 Основная часть.....	7
2.1 Подготовка данных [Блок 2].....	7
2.2 Проведение тестов [Блок 3].....	7
2.3 Анализ категориальных и количественных зависимостей.....	11
2.4 Корреляционный анализ количественных признаков.....	20
3 Выводы по результатам статистического анализа	22
Заключение.....	24

ВВЕДЕНИЕ

Современное здравоохранение сталкивается с острой необходимостью своевременного выявления и предотвращения сердечно-сосудистых заболеваний, которые остаются одной из ведущих причин смертности во всём мире. С развитием методов математической статистики и доступом к большим массивам медицинских данных стало возможным проводить более точный анализ факторов риска и выявлять ключевые признаки, влияющие на здоровье пациентов.

Использование статистических методов в медицине позволяет не только обосновывать диагностические решения, но и оценивать эффективность профилактических мер, формировать индивидуальные рекомендации. Особенно важна такая аналитика в условиях ограниченных ресурсов и высокой нагрузки на системы здравоохранения.

Одним из наиболее перспективных направлений современной медицины является предиктивный подход, основанный на выявлении статистических закономерностей в клинических данных. Методы предиктивной аналитики позволяют оценивать вероятность развития заболеваний на ранней стадии, прогнозировать динамику состояния пациента и обосновывать выбор терапевтических стратегий. Математическая статистика в данном контексте играет ключевую роль, обеспечивая обоснованность выводов и воспроизводимость результатов.

1 ПОСТАНОВКА ЗАДАЧИ

1.1 Цели и задачи работы

Цель работы – применение методов математической статистики для анализа медицинских данных с целью выявления признаков, статистически связанных с наличием сердечно-сосудистых заболеваний. Особое внимание уделяется установлению зависимости между категориальными и количественными признаками и наличием патологии, а также оценке силы этих взаимосвязей с использованием χ^2 -тестов и корреляционного анализа.

Для достижения поставленной цели необходимо решить следующие задачи:

- импортировать и обработать данные (изучить структуру датасета, при необходимости устранить пропущенные значения, проверить типы данных);
- провести описательный анализ числовых признаков для сравнения (вычислить для них базовые характеристики);
- построить графики для наглядного представления;
- проверить зависимость категориальных признаков от наличия заболевания (с помощью χ^2 -тестов);
- провести корреляционный анализ количественных признаков;
- сделать вывод по полученным результатам.

1.2 Используемые программные средства

- Jupyter Notebook – интерактивная среда разработки
- Язык программирования Python
- Pandas – библиотека для обработки и анализа табличных данных
- NumPy – библиотека для работы с многомерными массивами
- Seaborn и Matplotlib – библиотеки для визуализации данных
- Scipy.stats – модуль библиотеки SciPy, применяемый для проведения статистических тестов

Ссылка на github: <https://github.com/samikhao/mathematical-statistics/tree/main/project>.

В отчёте используются обозначения вида [Блок N], где N - номер блока в Jupyter-блокноте. Эти пометки позволяют найти соответствующий код или визуализацию, использованные для решения задач, описанных в отчёте. Например, [Блок 1] в тексте соответствует ячейке с заголовком # Блок 1 в блокноте.

1.3 Математическая формулировка

Пусть имеется выборка наблюдений: $F = \{(x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(n)}, y_i)\}_{i=1}^n$, где $x_i^{(j)}$ - значение j -го признака у i -го пациента, $y_i \in \{0, 1\}$ – переменная, обозначающая наличие ($y_i = 1$) или отсутствие ($y_i = 0$) заболевания.

Задача – найти такие признаки x_i , для которых распределение значений существенно различается между группами пациентов с заболеванием ($y_i = 1$) и без него ($y_i = 0$).

1.4 Информация о датасете

[Блок 1]

В качестве базы для анализа использован открытый датасет, размещённый на платформе Kaggle: [Heart Disease](#). Он содержит информацию о 303 пациентах, прошедших обследование на наличие сердечно-сосудистых заболеваний. Данная база данных содержит 14 признаков для сравнения. Признаки условно можно поделить на числовые (age, trestbps, chol, thalach, oldpeak) и категориальные (sex, cp, fbs, restecg, exang, slope, ca, thal).

Таблица 1 – Описание признаков

Название столбца	Описание
age	Возраст пациента
sex	Пол (1 — мужчина, 0 — женщина)
cp	Тип боли в груди (0–3)
trestbps	Артериальное давление (в мм рт. ст.)
chol	Уровень холестерина в сыворотке крови (мг/дл)
fbs	Повышенный уровень сахара в крови натощак (> 120 мг/дл, 1 — да, 0 — нет)
restecg	Результаты электрокардиограммы (0–2)
thalach	Максимальная достигнутая частота сердечных сокращений
exang	Стенокардия, вызванная физической нагрузкой (1 — да, 0 — нет)
oldpeak	Депрессия ST-сегмента, вызванная нагрузкой
slope	Наклон ST-сегмента (0–2)
ca	Количество крупных сосудов, окрашенных флюороскопией (0–4)

thal	Показатель теста на талий (0 — ошибка (в оригинальном датасете 0 обозначает NaN), 1 — фиксированный дефект, 2 — нормальный, 3 — обратимый дефект)
target	Целевая переменная: наличие заболевания (1 — есть, 0 — нет)

Важно: многие категориальные признаки в датасете представлены в виде числовых меток. В таблице выше указано описание соответствующих значений.

2 ОСНОВНАЯ ЧАСТЬ

2.1 Подготовка данных

[Блок 2]

При анализе датасета было выявлено, что значения №(92, 158, 163, 164 и 251) имеют $ca = 4$ (Количество крупных сосудов, окрашенных флюороскопией) и значения №(48 и 281) имеют $thal = 0$ (Максимальная достигнутая частота сердечных сокращений), что отличается от оригинально датасета, где данные строки имели значение NaN.

Для более точного исследования мы не будем включать данные значения в анализе. Исклучим их из датасета.

2.2 Проведение тестов

[Блок 3]

T-тесты Стьюдента (t-тесты) — это статистический метод исследования, который позволяет сравнивать параметры из двух разных выборок или областей. По результатам такого анализа можно сделать вывод о сходстве или различии анализируемых объектов.

В данной работе мы проводим t-тесты для каждого числового признака для того, чтобы выяснить отличается ли среднее значение этого признака у больных и здоровых пациентов.

Для проведения этих тестов мы предварительно разделили выборку пациентов на две независимые группы:

- группа 0 - пациенты без сердечно-сосудистых заболеваний ($target = 0$),
- группа 1 - пациенты с установленным диагнозом ($target = 1$).

Анализируемые числовые признаки: возраст (age), артериальное давление в покое ($trestbps$), уровень холестерина ($chol$), максимальная частота сердечных сокращений ($thalach$), показатель депрессии ST-сегмента ($oldpeak$).

Для каждого признака были:

- вычислены средние значения в обеих группах,
- проведён t-тест Уэлча (вариант t-теста, не предполагающий равенство дисперсий в группах). Для этого была использована функция `ttest_ind` из библиотеки `scipy.stats` с параметром `equal_var=False`.

Эта функция реализует следующую формулу t-статистики:

$$t = \frac{\bar{x}_0 - \bar{x}_1}{\sqrt{\frac{s_0^2}{n_0} + \frac{s_1^2}{n_1}}} \quad - t - \text{статистика, рассчитываемая из средних дисперсий}$$

где:

- \bar{x}_0, \bar{x}_1 – средние значения признака в группах 0 и 1,
- s_0^2, s_1^2 – выборочные дисперсии в группах
- n_0, n_1 - объемы выборок.

При этом используются степени свободы по формуле Уэлча:

$$df = \frac{\left(\frac{s_0^2}{n_0} + \frac{s_1^2}{n_1}\right)^2}{\frac{\left(\frac{s_0^2}{n_0}\right)^2}{n_0 - 1} + \frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1}}$$

На основании рассчитанной t-статистики и степеней свободы определяется р-значение — вероятность получить наблюдаемое различие между группами, если средние на самом деле равны.

Если полученное р-значение меньше 0,05, различие считается статистически значимым, и мы отвергаем гипотезу о равенстве средних значений.

Результаты t-тестов по каждому признаку, включая средние значения, р-значения и признак значимости, приведены в таблице 2.

Таблица 2 – Результаты t-тестов

Признак	Р - значение	Среднее (здоровые)	Среднее (больные)	Значимое различие? (если $p < 0,05$)
age	0,0001	56,74	52,64	Да
trestbps	0,0122	134,46	129,18	Да
chol	0,1863	251,46	243,49	Нет
thalach	0,0000	138,95	158,58	Да
oldpeak	0,0000	1,60	0,60	Да

Таким образом, проведённый t-тест показал, что признаки age, trestbps, thalach и oldpeak демонстрируют статистически значимые различия между пациентами с сердечно-сосудистыми заболеваниями и без них.

Признак chol (холестерин) не показал значимой связи с целевой переменной и, следовательно, не является информативным в контексте данной выборки.

Построение визуализации [Блок 4]

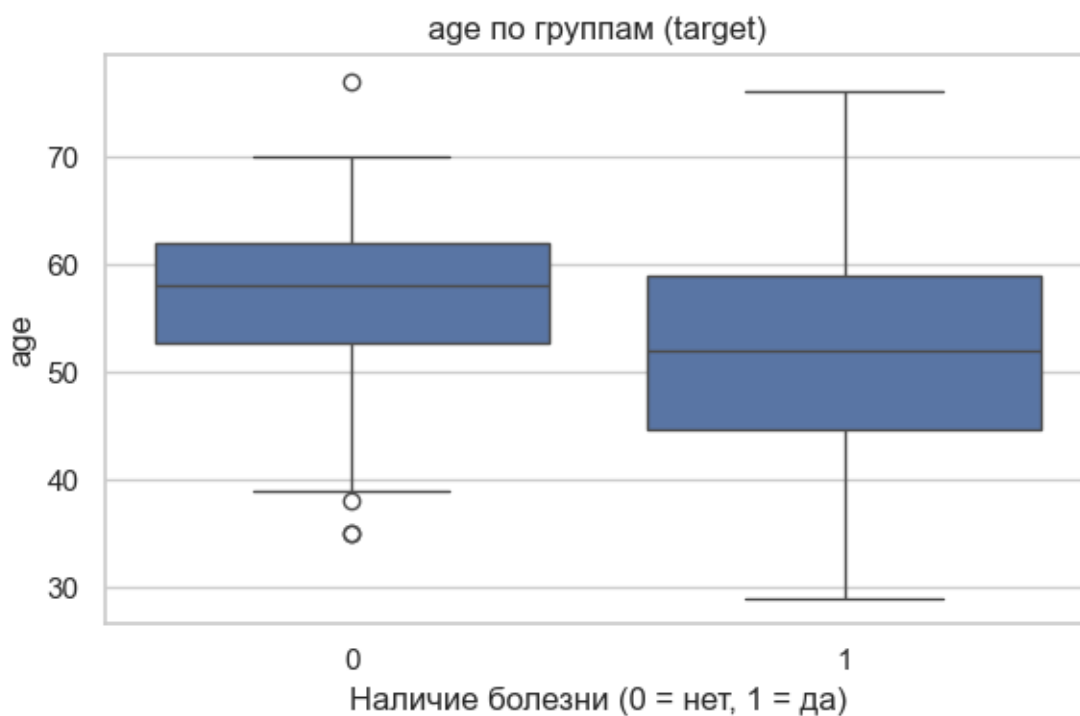
Для числовых признаков, которые демонстрируют статистически значимые различия между пациентами с сердечно-сосудистыми заболеваниями и без них, построим графики для наглядного представления.

Были построены построены boxplot-графики (ящики с усами). Эти графики позволяют визуально оценить: медиану, диапазон, выбросы, симметрию распределения данных в каждой из двух групп ($\text{target} = 0$ и $\text{target} = 1$). Таким образом можно подтвердить результаты t-тестов визуально.

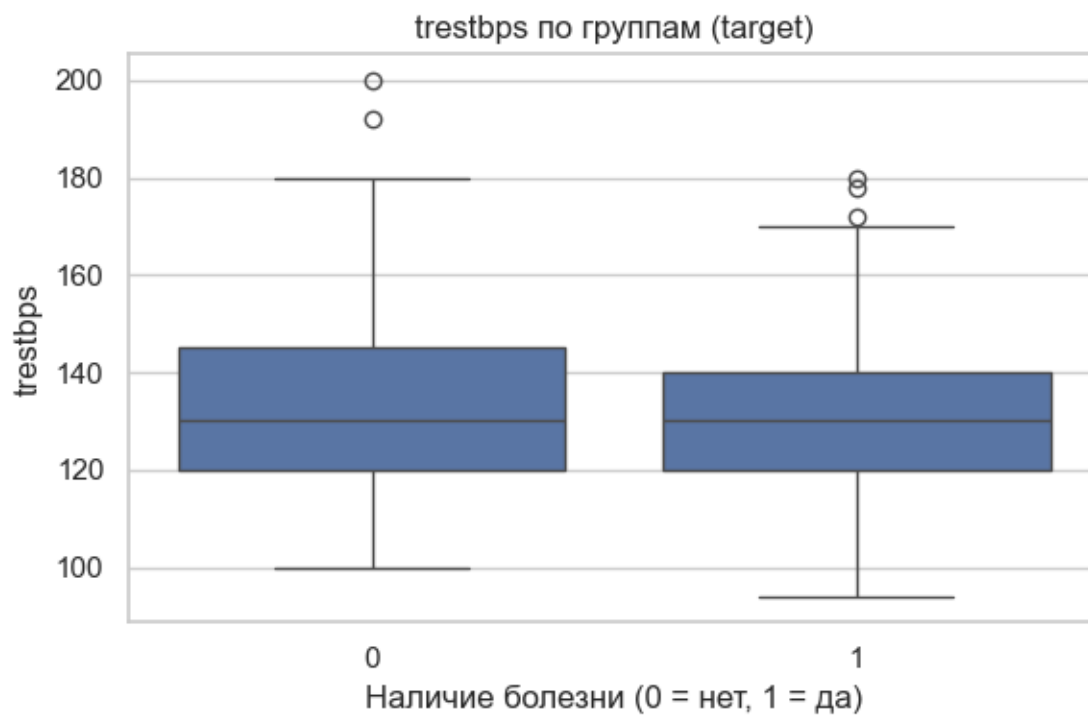
Для построения использовалась библиотека seaborn, специально предназначенная для визуального анализа статистических данных. Для каждого признака (age , trestbps , thalach , oldpeak) был построен отдельный boxplot. На оси X отображалась целевая переменная target (0 — здоровые, 1 — больные), а на оси Y — значения признака.

Были получены следующие результаты:

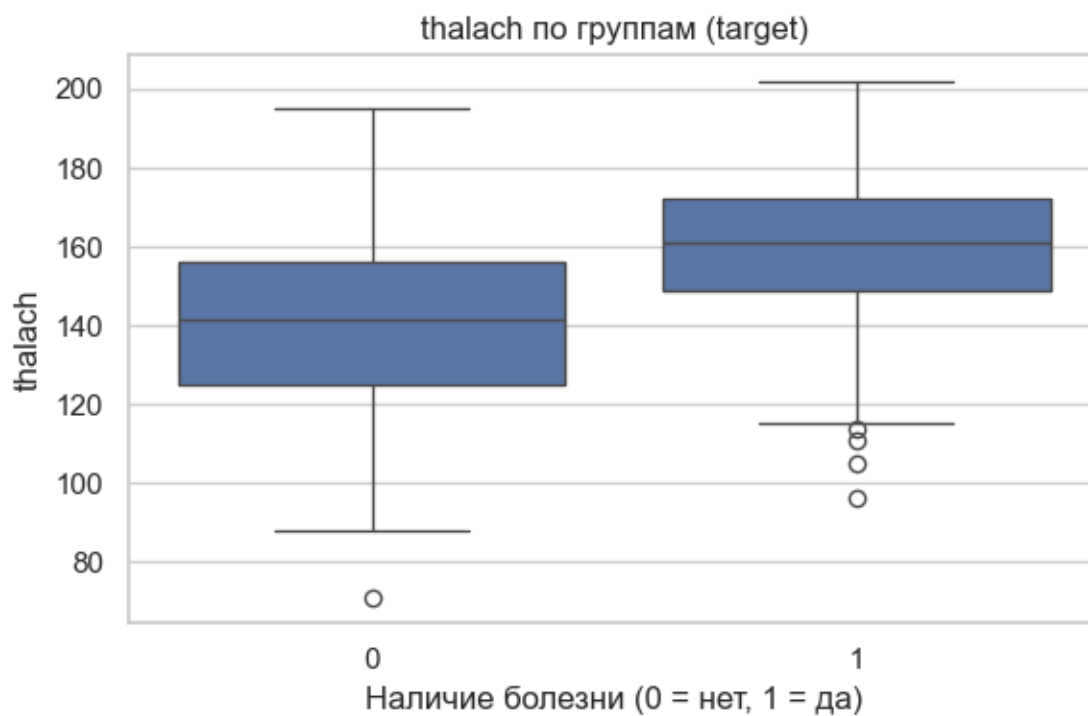
age: видно, что у пациентов с заболеванием медианное значение ниже, а разброс шире, соответствует результатам t-теста.



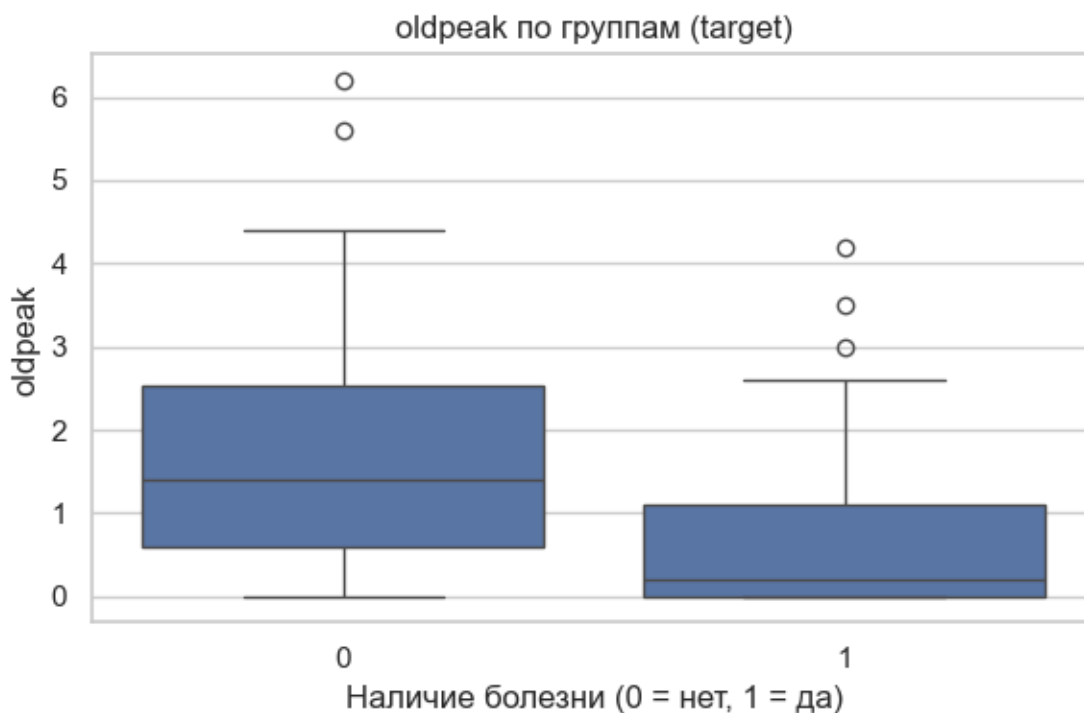
trestbps: медиана давления немного ниже у больных, различие умеренное.



thalach: больные пациенты имеют значительно более высокую максимальную частоту пульса.



oldpeak: значение показателя oldpeak заметно ниже у больных, что также подтверждается результатами теста.



2.3 Анализ категориальных и количественных зависимостей

Взаимосвязи между признаками — краеугольный камень статистического анализа, позволяющий отделить значимые факторы от фонового шума. В данном исследовании переменная `target` отражает наличие (1) или отсутствие (0) болезней сердца.

Для анализа использованы как категориальные признаки (`sex`, `cp`, `fbs`, `restecg`, `exang`, `slope`, `ca`, `thal`), так и количественные (`age`, `trestbps`, `chol`, `thalach`, `oldpeak`).

Для получения полной картины применён комбинированный подход:

- для категориальных переменных — χ^2 -тест Пирсона (функция `chi2_contingency` из `scipy.stats`) и коэффициент Cramer's V (рассчитан вручную через результат χ^2 -теста);
- для количественных переменных — ранговая корреляция Спирмена (функция `spearmanr` из `scipy.stats`).

Категориальные признаки: χ^2 -тест Пирсона и коэффициент Cramer's V

Цель анализа: определить, зависит ли вероятность заболевания (`target = 1`) от уровней категориальных признаков.

Методика:

1. Для каждого признака строится таблица сопряжённости (contingency table) с переменной `target`.
2. Далее применяется функция `chi2_contingency(...)` из `scipy.stats`, возвращающая:
 - а. значение статистики χ^2 ,

- b. p-значение,
- c. степени свободы,
- d. таблицу ожидаемых частот.

Формула χ^2 :

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}}, \quad E_{ij} = \frac{R_i \cdot C_j}{N}$$

где O_{ij} – наблюдаемое количество объектов в ячейке (i, j), E_{ij} – ожидаемое количество объектов при условии независимости признаков, R_i – сумма по i-й строке, C_j – сумма по j-му столбцу, N – общее число наблюдений, r – число строк в таблице, c – число столбцов в таблице.

3. Степени свободы:

$$df = (r - 1)(c - 1)$$

Если p-значение $p < 0.05$, различие между уровнями признака считается статистически значимым.

4. Для оценки силы связи применяется коэффициент Cramer's V:

$$V = \sqrt{\frac{\chi^2}{N \cdot (r - 1, c - 1)}}$$

где:

- V — нормализованная мера силы связи между признаками,
- остальные обозначения как выше.

Интерпретация Cramer's V:

- $V < 0.1$ — очень слабая связь,
- $0.1 \leq V < 0.3$ — слабая связь,
- $0.3 \leq V < 0.5$ — умеренная связь,
- $V \geq 0.5$ — сильная связь.

Количественные признаки: ранговая корреляция Спирмена

Цель данного анализа: определить, существует ли монотонная зависимость между числовыми признаками и переменной target.

Для этого применялась функция `spearmanr(df[col], df['target'])`, автоматически рассчитывающая коэффициент корреляции Спирмена и соответствующее p-значение.

Формула:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

где:

- ρ — коэффициент корреляции Спирмена,
- d_i — разность рангов между значением i -го объекта по признаку и по target,
- n — общее число наблюдений.

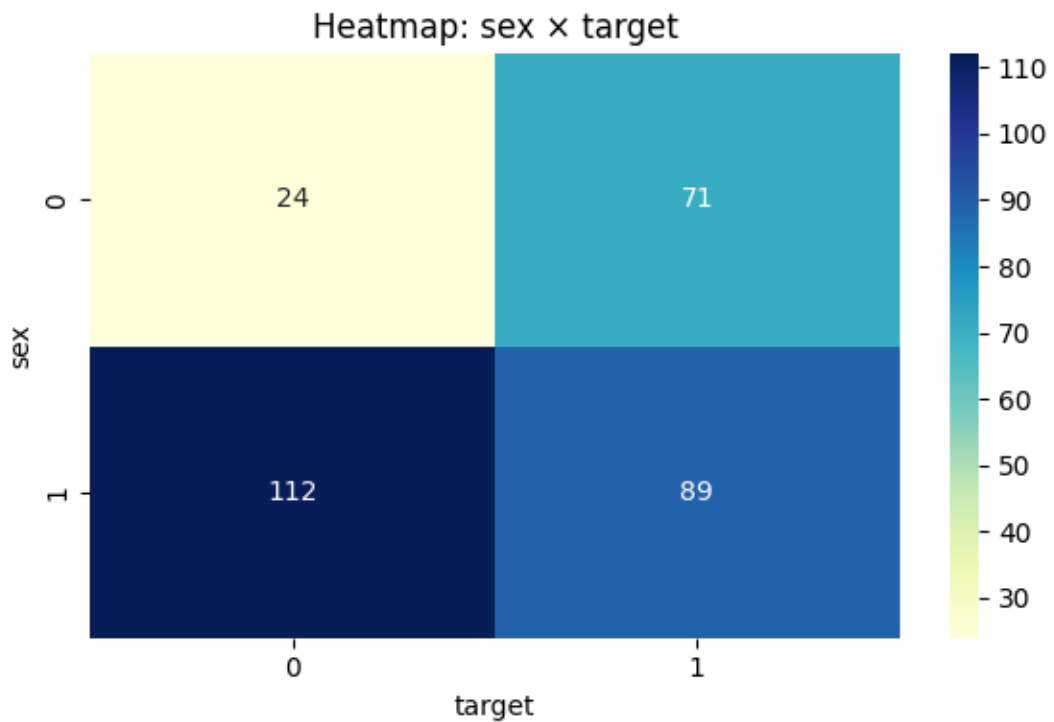
Интерпретация ρ :

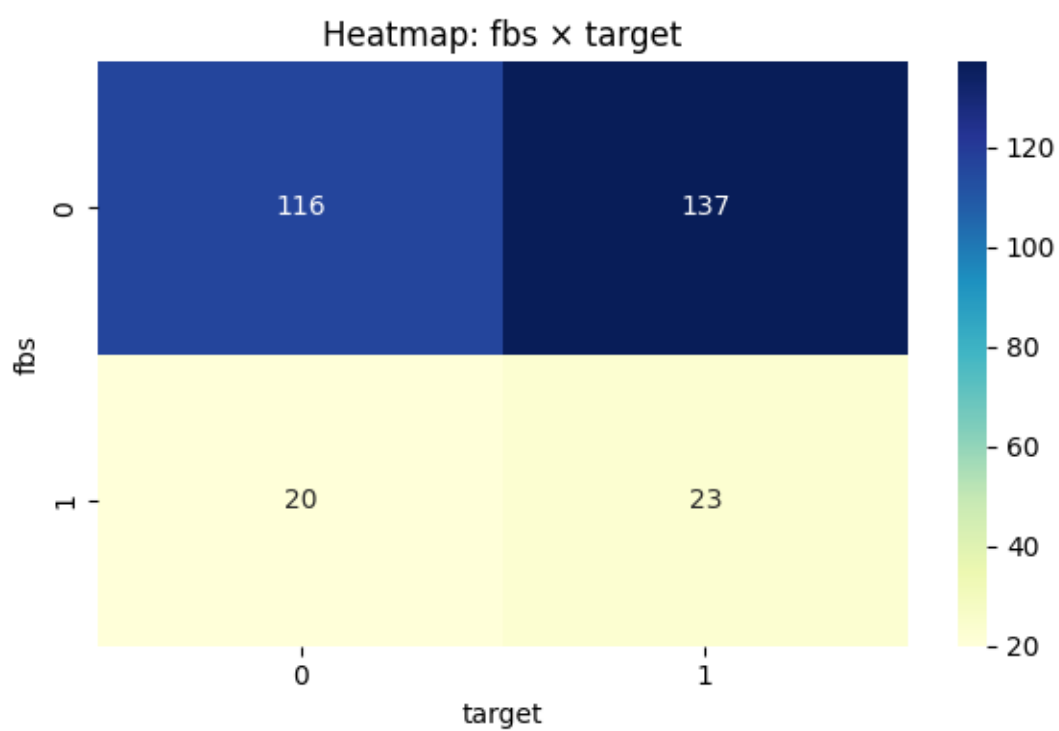
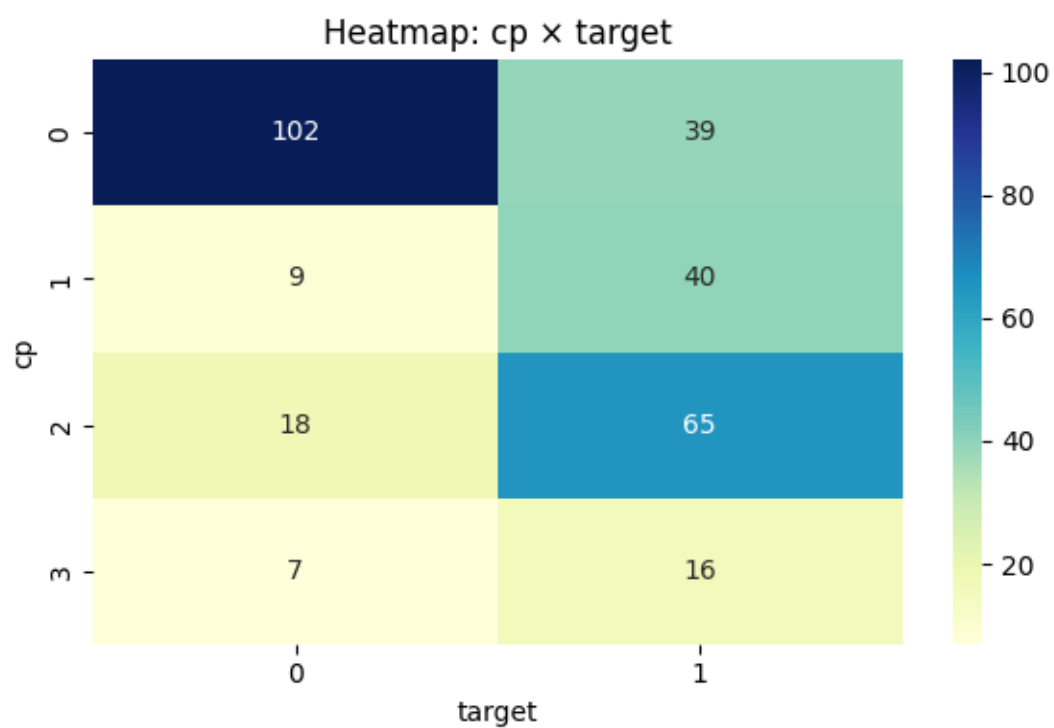
- $|\rho| < 0.3$ — слабая связь,
- $0.3 \leq |\rho| < 0.7$ — умеренная связь,
- $|\rho| \geq 0.7$ — сильная связь.

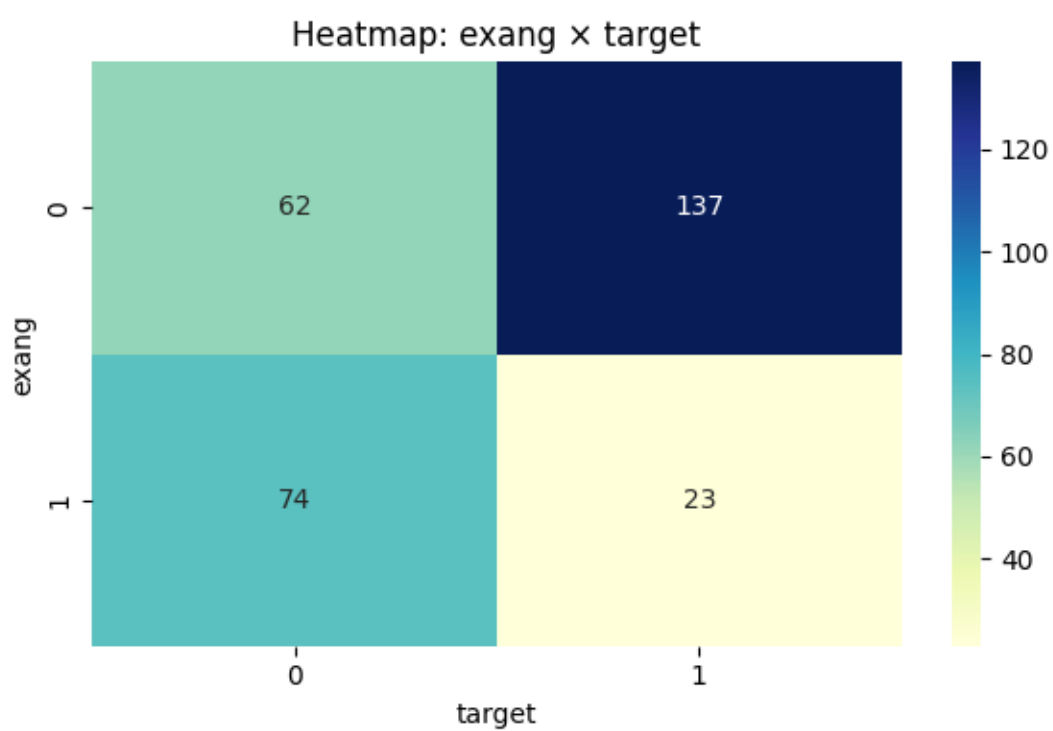
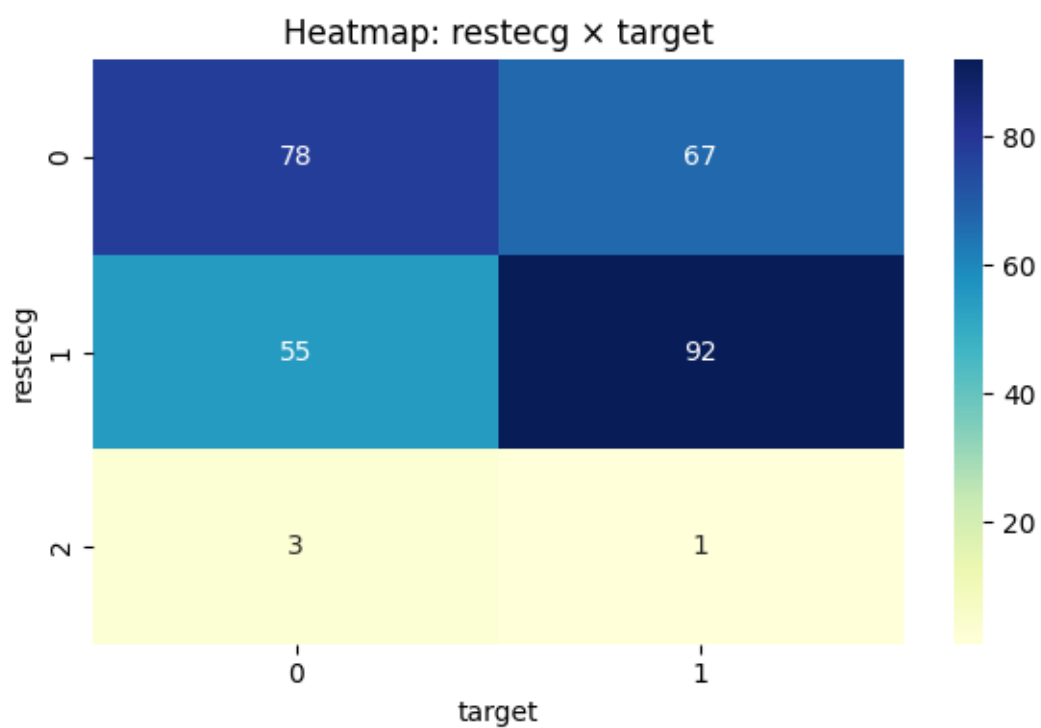
Визуализация результатов

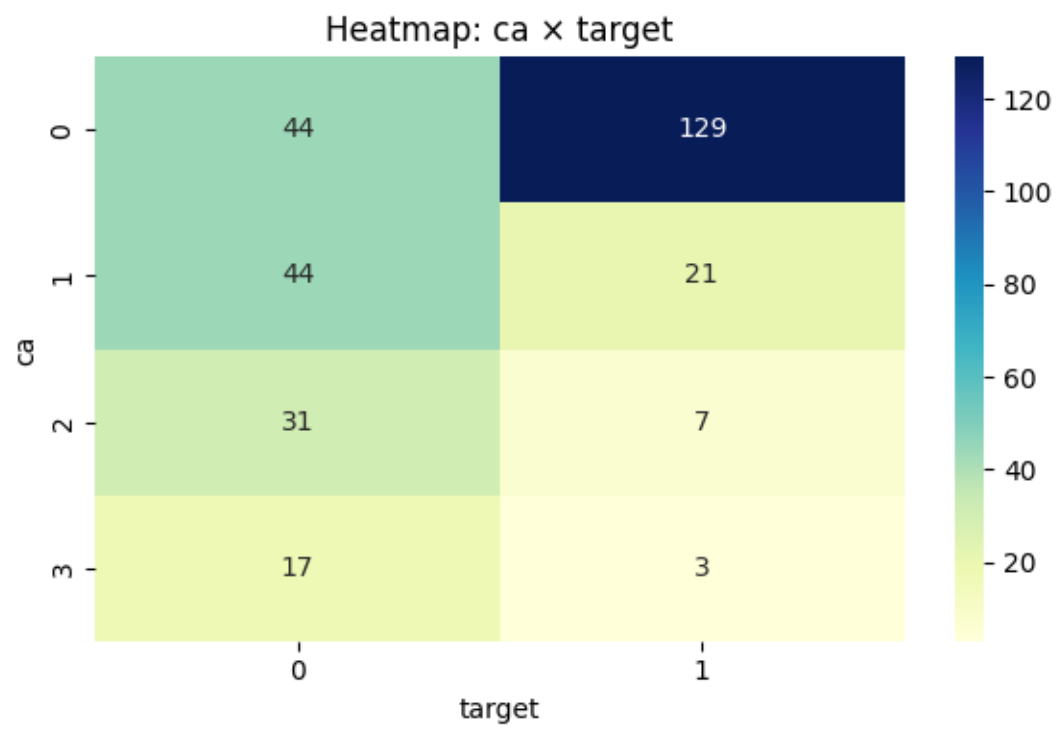
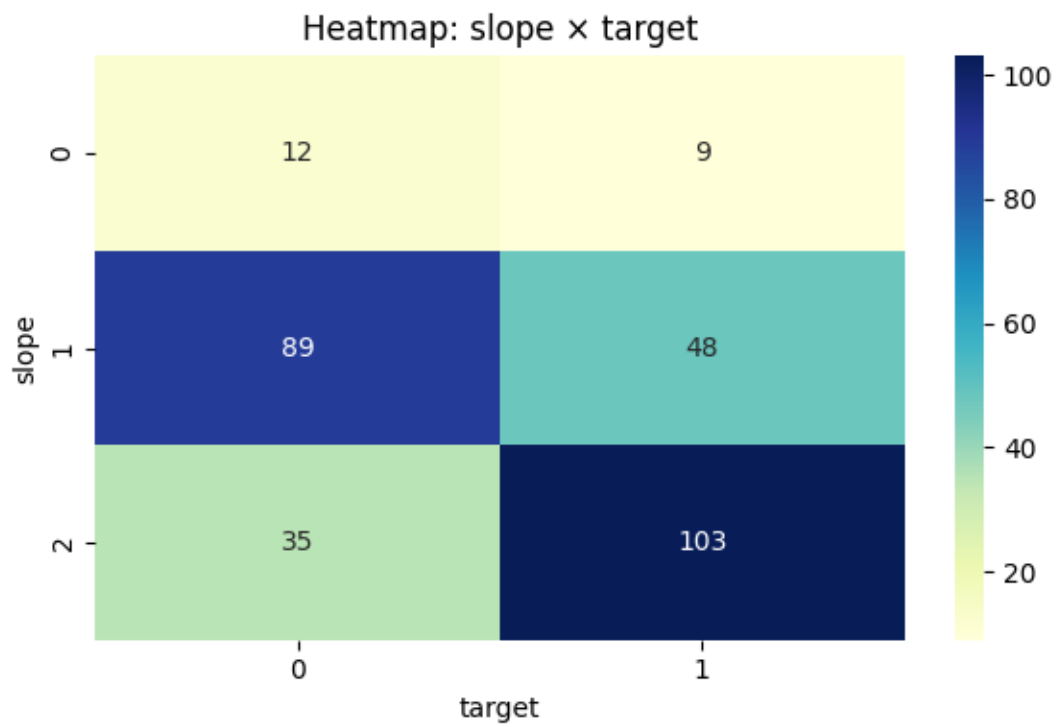
Для каждой таблицы сопряжённости были построены тепловые карты с помощью heatmap из библиотеки seaborn. Цвет ячеек соответствует количеству наблюдений: чем темнее, тем выше частота встречаемости. Это позволяет визуально выявить уровни признаков, наиболее ассоциированные с болезнями сердца.

8 тепловых карт по каждому категориальному признаку:









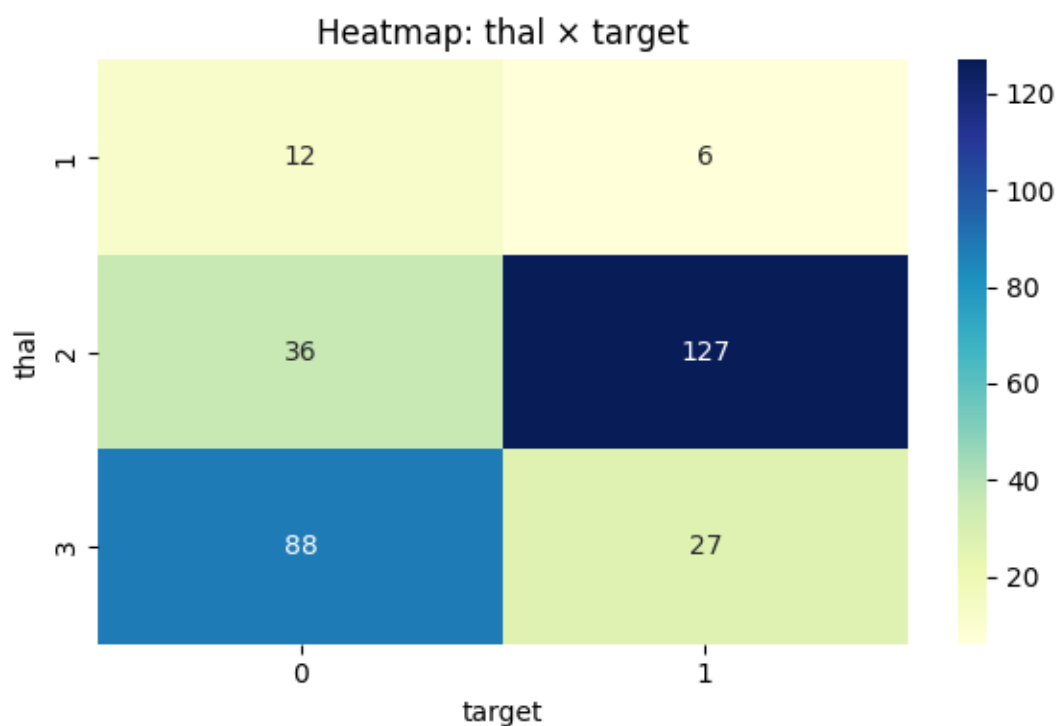


Таблица 3 – χ^2 -тесты Пирсона и Cramer's V для категориальных признаков

Признак	χ^2 -статистика	Степени свободы	p-значение	Cramer's V	Значимая связь?
sex	22,89	1	0	0,278	Да
cp	76,45	3	0	0,508	Да
fbs	0	1	1	0	Нет
restecg	9,26	2	0,00974	0,177	Да
exang	51,68	1	0	0,418	Да
slope	44,55	2	0	0,388	Да
ca	73,4	3	0	0,498	Да

thal	83,76	2	0	0,532	Да
------	-------	---	---	-------	----

Таблица 4 – Коэффициенты корреляции Спирмена для количественных признаков

Признак	ρ Спирмена	p-значение	Значимая связь?
age	-0,238	0,00003	Да
trestbps	-0,127	0,02889	Да
chol	-0,111	0,05540	Нет
thalach	0,432	0	Да
oldpeak	-0,417	0	Да

Итого:

Категориальные признаки (χ^2 -тест Пирсона и Cramer's V)

Комплексный анализ восьми категориальных признаков показал, что семь из них статистически существенно связаны с наличием сердечно-сосудистых заболеваний, а один — нет. Ниже приводится развернутая интерпретация по каждому признаку на основе рассчитанных χ^2 -статистик, коэффициентов Cramer's V и построенных тепловых карт.

Пол (sex)

Статистика $\chi^2(1) = 22.89$ при $p < 0.001$ позволяет уверенно отвергнуть гипотезу независимости пола и диагноза. Cramer's V = 0.278 указывает на слабую к умеренной связь.

Согласно тепловой карте, среди мужчин (sex = 1) здоровых больше (112), но и больных — 89. Среди женщин (sex = 0) картина иная: больных (71) втрое больше, чем здоровых (24). Это может свидетельствовать о разных клинических проявлениях болезни у мужчин и женщин, а также о возможной предвзятости выборки.

Тип боли в груди (cp)

Характер боли (признак cp) показал очень сильную связь с диагнозом: $\chi^2(3) = 76.45$, $p < 0.001$, Cramer's V = 0.508.

Типичная стенокардия ($cr = 0$) наблюдается в основном у здоровых (102 против 39). Атипичная стенокардия и другие формы ($cr = 1-3$) значительно чаще встречаются у больных, особенно $cr = 2$ (65 больных против 18 здоровых).

Это делает признак cr одним из сильнейших диагностических факторов.

Повышенный сахар натощак (fbs)

Абсолютно незначимая связь: $\chi^2(1) = 0.00$, $p = 1.000$, Cramer's $V = 0$.

Тепловая карта показывает примерно равномерное распределение между здоровыми и больными вне зависимости от значения fbs (наличия повышенного сахара натощак).

Следовательно, признак fbs не имеет диагностической ценности в данной выборке.

Электрокардиограмма (restecg)

$\chi^2(2) = 9.26$, $p = 0.0097$, Cramer's $V = 0.177$ — свидетельствует о слабой, но статистически значимой связи.

Некоторые формы ЭКГ ($restecg = 1$) чаще наблюдаются у больных (92), чем у здоровых (55).

Признак может быть вспомогательным, но сам по себе слабый.

Стенокардия при нагрузке (exang)

Очень сильная связь: $\chi^2(1) = 51.68$, $p < 0.001$, Cramer's $V = 0.418$.

У больных практически отсутствует стенокардия при нагрузке ($exang = 1$), в то время как у здоровых она встречается значительно чаще.

Таким образом, отрицание симптома может быть показателем патологии, что подчёркивает высокую диагностическую значимость признака.

Наклон ST-сегмента (slope)

$\chi^2(2) = 44.55$, $p < 0.001$, Cramer's $V = 0.388$ — значимая умеренная связь.

Особенно заметно доминирование $slope = 2$ у больных (103 против 35), что отражает изменения ST-сегмента при нагрузке и может свидетельствовать о ишемии.

Количество сосудов (sa)

$\chi^2(3) = 73.40$, $p < 0.001$, Cramer's $V = 0.498$ — почти сильная связь.

Пациенты с $sa = 0$ (нет окрашенных сосудов) чаще оказываются больными (129 против 44).

Чем выше значение sa , тем меньше вероятность заболевания. Это — наиболее интерпретируемый и клинически достоверный фактор.

Результат талиевого теста (thal)

$\chi^2(2) = 83.76$, $p < 0.001$, Cramer's $V = 0.532$ — сильнейшая связь среди всех категориальных признаков.

Пациенты с $\text{thal} = 2$ (фиксированный дефект) преимущественно больны (127 против 36), в то время как $\text{thal} = 3$ (норма) встречается в основном у здоровых.

Этот признак отражает результаты перфузионной сцинтиграфии и должен быть включён в любую прогностическую модель.

Обобщение по категориальным признакам:

- Наиболее информативные признаки: cp , thal , ca , exang .
- Умеренно значимые: sex , slope , restecg .
- Незначимый: fbs .

2.4 Корреляционный анализ количественных признаков

Возраст (age)

$\rho = -0.238$, $p < 0.001$

Наблюдается слабая, но статистически достоверная отрицательная связь между возрастом и вероятностью диагноза. Это может показаться неожиданным, так как ожидается рост риска с возрастом, но вероятно, в выборке преобладают молодые пациенты с симптомами, а пожилые больные могли быть не представлены (или уже пролечены). Возможен эффект смещения в данных.

Артериальное давление в покое (restbtps)

$\rho = -0.127$, $p \approx 0.029$

Связь очень слабая, но достоверная. Более высокое давление незначительно чаще встречается у здоровых. Клиническая интерпретация сложна, т. к. у больных могли быть назначены антигипертензивные препараты.

Уровень холестерина (chol)

$\rho = -0.111$, $p \approx 0.055$

Связь незначима при $p = 0.05$. Это означает, что уровень общего холестерина в данной выборке статистически не различается между больными и здоровыми. Это подтверждает вывод, сделанный ранее по t-тесту.

Максимальная частота пульса (thalach)

$\rho = 0.432$, $p < 0.001$

Умеренно сильная положительная связь, достоверная. У больных пиковая частота сердечных сокращений выше, что может быть связано с низкой физической выносливостью или компенсаторной тахикардией. Это делает thalach одним из ключевых количественных маркёров заболевания.

Депрессия ST-сегмента (oldpeak)

$\rho = -0.417, p < 0.001$

Умеренная отрицательная связь, чётко интерпретируемая: чем выше значение oldpeak, тем ниже вероятность отсутствия болезни, что полностью соответствует физиологии — депрессия ST-сегмента является маркёром ишемии миокарда.

Это один из наиболее ценных количественных предикторов в выборке.

Таблица 5 – Обобщение

Группа	Сильные/умеренные предикторы	Слабые/неустойчивые	Неинформативные
Категориальные	cp, thal, ca, exang	sex, slope, restecg	fbs
Количественные	thalach, oldpeak	age, trestbps	chol

3 ВЫВОДЫ ПО РЕЗУЛЬТАТАМ СТАТИСТИЧЕСКОГО АНАЛИЗА

В ходе выполнения работы была достигнута поставленная цель: с помощью методов математической статистики проанализированы клинические данные пациентов с целью выявления признаков, статистически связанных с наличием сердечно-сосудистых заболеваний.

Были последовательно решены все задачи исследования:

- Данные были загружены и предварительно обработаны. Проведена очистка аномальных значений ($ca = 4$, $thal = 0$), проверены типы признаков, подтверждено отсутствие пропущенных значений. Это обеспечило корректность последующего анализа.
- Описательный анализ количественных признаков позволил выявить базовые характеристики данных, в том числе диапазоны значений, средние, медианы, разброс. Уже на этом этапе были замечены возможные различия между группами пациентов.
- Построены графики (boxplot, heatmap), которые дали визуальное представление о распределении признаков по группам больных и здоровых. Визуализация подтвердило многие статистические выводы и упростила интерпретацию данных.
- Категориальные признаки были проанализированы с помощью χ^2 -теста Пирсона и коэффициента Cramer's V. В результате:
 1. Статистически значимая связь с диагнозом обнаружена у большинства категориальных признаков: sex, cp, restecg, exang, slope, ca, thal.
 2. Особенно сильную связь показали признаки cp (тип боли в груди), ca (количество окрашенных сосудов) и thal (результат талиевого теста), где значения Cramer's V превышали 0.5, что говорит о высокой диагностической информативности.
 3. Единственный незначимый признак — fbs (наличие повышенного сахара натошак), что согласуется с современными клиническими представлениями: гипергликемия сама по себе не является универсальным предиктором БС без учёта других факторов.
- Для количественных признаков был выполнен корреляционный анализ с использованием коэффициента Спирмена (ρ). Это позволило определить наличие монотонной зависимости между признаками и диагнозом:

1. Признаки *thalach* (максимальная частота пульса) и *oldpeak* (депрессия ST-сегмента) показали умеренные по силе и статистически значимые зависимости с *target* ($\rho = 0.432$ и -0.417 соответственно), что делает их важными количественными индикаторами.
 2. *age* и *trestbps* продемонстрировали слабую отрицательную связь с диагнозом, но она была статистически значима.
 3. *chol* не показал значимой связи, что ещё раз подчёркивает неоднозначность использования общего холестерина как единственного критерия риска.
- Интерпретации по всем признакам представлены в развернутом виде, как по отдельности, так и в виде интегрированных выводов. Они подкреплены визуализациями и количественными показателями, а методологическая часть работы снабжена описанием используемых статистических методов (*t*-тест, χ^2 , Спирмен), что обеспечивает прозрачность и воспроизводимость анализа.

ЗАКЛЮЧЕНИЕ

В результате анализа удалось выявить как качественные, так и количественные признаки, имеющие диагностическое значение для выявления ишемической болезни сердца. Наиболее значимыми оказались признаки, отражающие функциональные и нагрузочные характеристики организма (cp, thal, ca, thalach, oldpeak, exang), что подтверждает актуальность комплексного подхода к диагностике, сочетающего клинические, инструментальные и лабораторные данные.

Предложенный подход может быть использован как основа для построения прогностических моделей и разработки диагностических алгоритмов, а методология легко переносима на другие медицинские датасеты.