



ТЕМА ПРОЕКТА

*«Анализ сердечно-сосудистых заболеваний с применением методов
интеллектуального анализа данных»*

Выполнил(-а): студент НИУ ВШЭ

Группы КиМБ-211

Иманалиева Аминат

Проверил: Журавлева Светлана

Москва, 2025 г.

АБСТРАКТ

Отчет представляет собой описание, посвященное анализу заболеваний сердца среди мужчин и женщин, с учетом показателей здоровья и различных показателей образа жизни. Анализ направлен на изучение взаимосвязей между образом жизни и распространенностью заболеваний сердца.

Ключевые слова: сердечно-сосудистые заболевания, риск сердечно-сосудистых заболеваний, статистический анализ, логистическая регрессия.

ВВЕДЕНИЕ

Сердечно-сосудистые заболевания (ССЗ) представляют собой одну из наиболее распространенных по уровню заболеваемости в мире, являясь основной причиной смертности и инвалидности среди взрослых. Эти заболевания охватывают широкий спектр патологий, включая ишемическую болезнь сердца, сердечную недостаточность, гипертоническую болезнь и другие состояния, нарушающие нормальное функционирование сердца и кровеносных сосудов [1]. Факторы риска сердечно-сосудистых заболеваний включают как модифицируемые, такие как нездоровое питание, низкий уровень физической активности, курение и злоупотребление алкоголем, так и немодифицируемые, такие как возраст, пол и наследственность [2]. Понимание этих факторов имеет решающее значение для разработки эффективных стратегий профилактики и лечения ССЗ.

Актуальность работы. В последние годы наблюдается рост интереса к разработке новых методов прогнозирования и предотвращения сердечных заболеваний, включая интеллектуальный анализ данных и создание инструментов для оценки риска на основе различных факторов [3]. В этом контексте актуальность изучения взаимосвязей между образом жизни и распространенностью сердечно-сосудистых заболеваний становится доступной благодаря анализу представленных данных.

Цель работы. Изучить взаимосвязи между образом жизни и распространенностью сердечно-сосудистых заболеваний. Для достижения данной цели были поставлены следующие **задачи**:

- 1) Провести предварительную фильтрацию данных.
- 2) Проанализировать распределение данных при помощи базовой визуализации.
- 3) Применить статистические тесты для анализа зависимости между переменными; независимыми переменными и целевым признаком.
- 4) Применить регрессионный анализ для предсказания вероятности наличия сердечного заболевания.

МАТЕРИАЛЫ И МЕТОДЫ

Исходные данные содержали 319795 наблюдений и 18 признаков. Анализ типов данных показал, что 4 переменные являются числовыми (1 - double, 3 - integer), а 14 переменных - категориальными (11 бинарные номинальные, 1 мультиноминальная и 2 порядковые). Была проведена

предварительная фильтрация данных на содержание дубликатов и пропущенных значений, анализ выбросов проведен методом `boxplot`, удаление выбросов выполнено с помощью метода межквартильного размаха. После удаления выбросов осталось 255233 наблюдений. Все категориальные переменные были преобразованы в факторные с учетом порядка (для порядковых переменных).

Для количественных переменных были построены гистограммы распределения и проверка нормальности распределения данных с помощью Q-Q графиков. Для оценки сбалансированности наблюдений внутри категориальных переменных были построены `bar plot`. Для проверки нормальности распределения данных были применены тесты Андерсона-Дарлинга и Колмогорова-Смирнова (`library(nortest)`, `ad.test`, `ks.test`). Сравнение средних значений индекса массы тела между мужчинами и женщинами с использованием теста Манна-Уитни (`alternative = "two.sided"`).

Корреляционный анализ включал определение коэффициентов корреляции Спирмена между количественными переменными методом Спирмена (`method = 'spearman'`). Для оценки различий количественных признаков между тремя и более группами, определенными по расе, использовался непараметрический критерий Краскела-Уолиса.

Проверка связи категориальных переменных с заболеванием сердца определялся методом хи-квадрат согласия Пирсона (`chisq.test()`). Для описания распределения категориальных переменных в группах пациентов с наличием и отсутствием сердечно-сосудистых заболеваний использовались частотный анализ с вычислением процентных долей для каждой группы, с использованием функции `descrTable`.

Для изучения зависимости заболеваемости сердца от различных факторов были построены графики следующие методы визуализации: столбчатые диаграммы (`geom_bar()`) для отображения распределения категориальных переменных в зависимости от наличия сердечных заболеваний, `box plot` (`geom_boxplot()`) для оценки распределения количественных признаков, таких как индекс массы тела (BMI) и продолжительность сна, в зависимости от категориальных факторов (возраст, курение и др.).

Для предсказания вероятности наличия сердечного заболевания была построена логистическая регрессия, где в качестве целевой переменной использовался бинарный показатель наличия сердечно-сосудистого заболевания (HeartDisease). Первоначальная модель включала несколько предикторов, однако для оптимизации была применена процедура пошагового исключения переменных (`stepAIC`), позволившая выделить наиболее значимые факторы. Промежуточная модель включала в себя переменные: возрастная категория (AgeCategory), общее состояние здоровья (GenHealth), диабет (Diabetic), наличие заболеваний почек (KidneyDisease), инсульт (Stroke) и проблемы с передвижением (DiffWalking). Качество логистической модели оценивалось с помощью метрик, ROC-кривая и AUC (Area Under Curve).

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЯ

Часть первая. Знакомство с данными и предварительная фильтрация

Исходный датасет включал 319795 наблюдений и 18 признаков:

- **HeartDisease** - наличие заболеваний сердца
- **BMI** - индекс массы тела
- **Smoking** – курение
- **AlcoholDrinking** - употребление алкоголя
- **Stroke** - шемический инсульт
- **PhysicalHealth** - показатель, сколько дней в месяц плохое физическое самочувствие
- **MentalHealth** - показатель, сколько дней в месяц плохое ментальное самочувствие
- **DiffWalking** -трудности при подъеме по лестнице
- **Sex** - пол
- **AgeCategory** - категория возраста
- **Diabetic** - наличие диабета
- **PhysicalActivity** - взрослые, которые сообщили, что в течение последних 30 дней занимались физическими упражнениями помимо своей обычной работы
- **GenHealth** - уровень субъективного благополучия
- **SleepTime** - часы сна
- **Asthma**- наличие астмы
- **KidneyDisease** - наличие заболеваний почек
- **Skin Cancer** - рак кожи
- **Race** – раса

Числовые (4): BMI (double), PhysicalHealth, MentalHealth, SleepTime (integer)

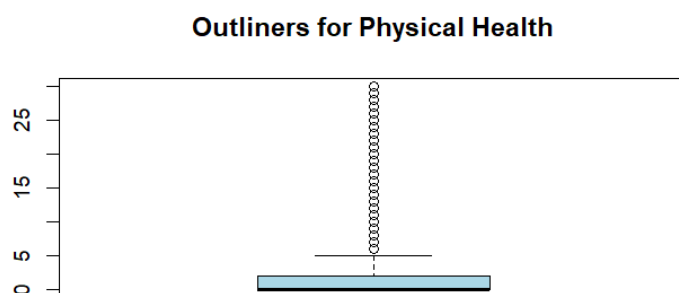
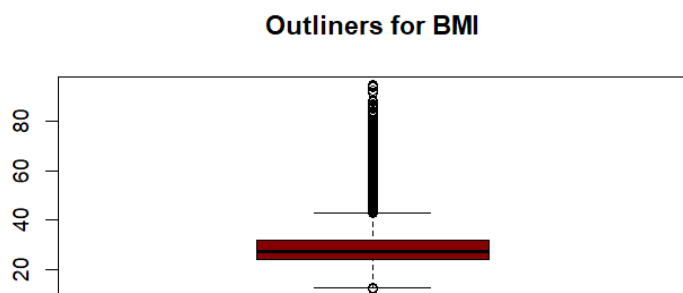
Категориальные (14): 11 бинарные номинальные, 1 мультиноминальные (Race) и 2 порядковые (AgeCategory, GenHealth)

Проверка на пропущенные значения изначально показала, что в данных отсутствуют пропущенные значения, имелись дубликаты 18078 значений, после удаления которых осталось 301717 наблюдений. Общая статистика количественных данных показала следующие данные, представленных в Таблице 1:

Таблица 1 — Описательная статистика количественных данных

#	BMI	PhysicalHealth	MentalHealth	SleepTime
# Min.	:12.02	Min. : 0.000	Min. : 0.000	Min. : 1.000
# 1st Qu.:	24.03	1st Qu.: 0.000	1st Qu.: 0.000	1st Qu.: 6.000
# Median	:27.41	Median : 0.000	Median : 0.000	Median : 7.000
# Mean	:28.44	Mean : 3.572	Mean : 4.121	Mean : 7.085
# 3rd Qu.:	31.65	3rd Qu.: 2.000	3rd Qu.: 4.000	3rd Qu.: 8.000
# Max.	:94.85	Max. : 30.000	Max. : 30.000	Max. : 24.000

Анализ данных на выбросы показал наличие выбросов по признаку “BMI” (7905 выбросов), “Physical Health” (6136 выбросов), “SleepTime” (3542 выбросов), “MentalHealth” (38713 выбросов), представленных на Рисунке:



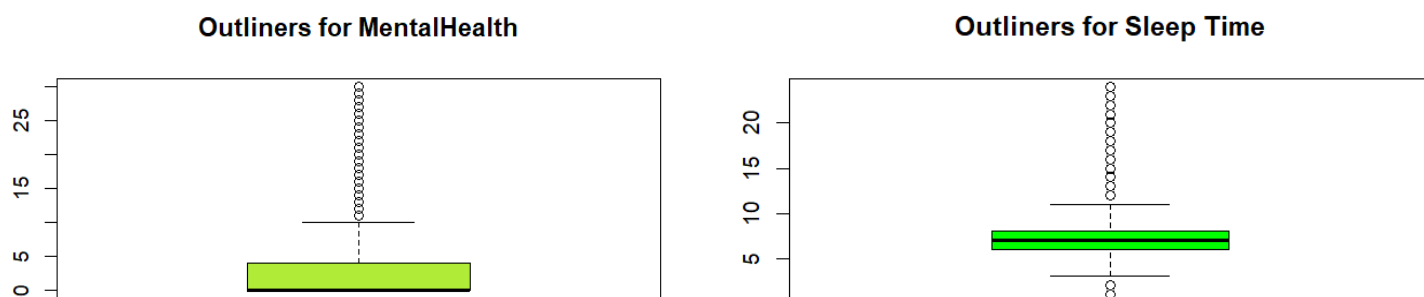


Рисунок. Анализ количественных переменных на выбросы

Результаты после удаления выбросов по признакам “BMI” и “Mental Heath” представлены на Рисунке:

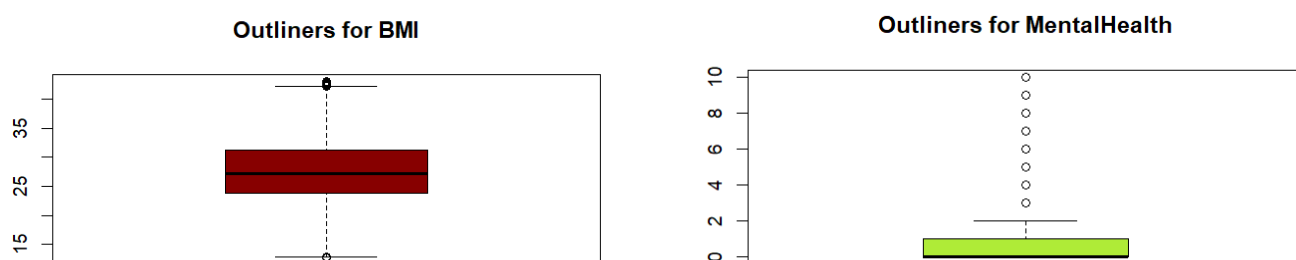


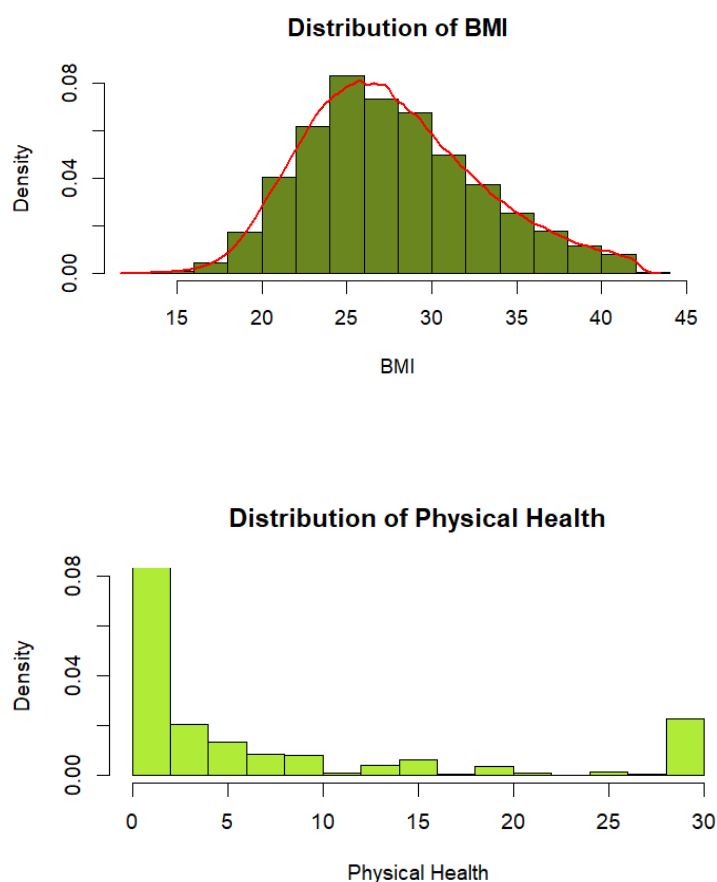
Рисунок. Box plot после удаления выбросов по признакам “BMI” и “Mental Heath”

После обработки выбросов была проведена проверка на наличие пропущенных значений. В результате этого процесса было исключено 48618 наблюдений, что привело к сокращению общего объема выборки до 255233 записей. Хотя удаление выбросов обычно не рекомендуется, поскольку трудно определить их природу (биологическую или техническую), в данном случае это решение было принято по нескольким причинам: значительный

объем оставшейся выборки, результаты предварительного анализа с учетом выбросов, большое количество обнаруженных аномальных значений. Учитывая эти факторы, было решено, что удаление выбросов улучшит качество анализа и повысит надежность результатов исследования.

Часть вторая. Базовая визуализация для оценки распределения данных

Гистограммы для числовых переменных для оценки распределения данных (первичная оценка по $f(x) = F'(x)$ и подчинении нормальному распределению). Результаты представлены на Рисунке:



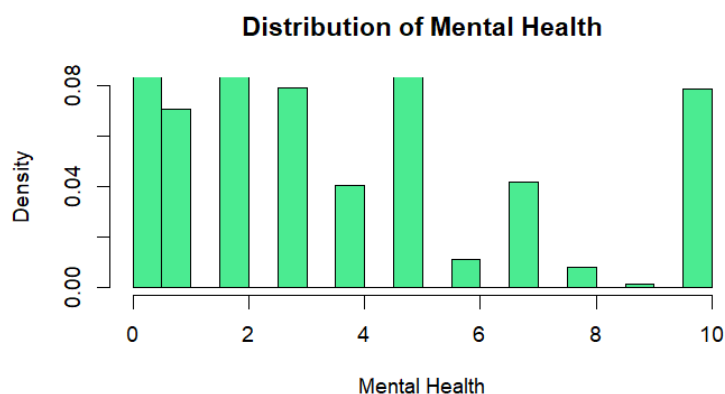
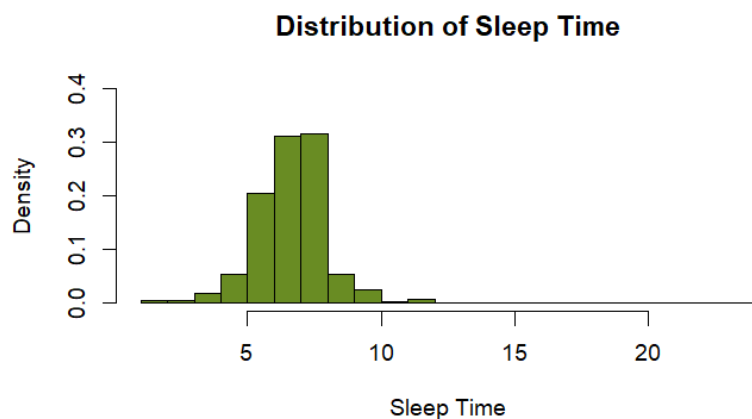


Рисунок. Анализ распределения числовых переменных по плотности

На основе проведенной визуализации данных можно сделать предварительное заключение о том, что распределение исследуемых переменных существенно отклоняется от нормального (гауссовского) распределения. Предположение требует дальнейшей верификации с помощью формальных статистических тестов на нормальность, таких как тест Шапиро-Уилка, Колмогорова-Смирнова или Андерсона-Дарлинга.

Результаты распределения данных по частоте представлены на Рисунке:

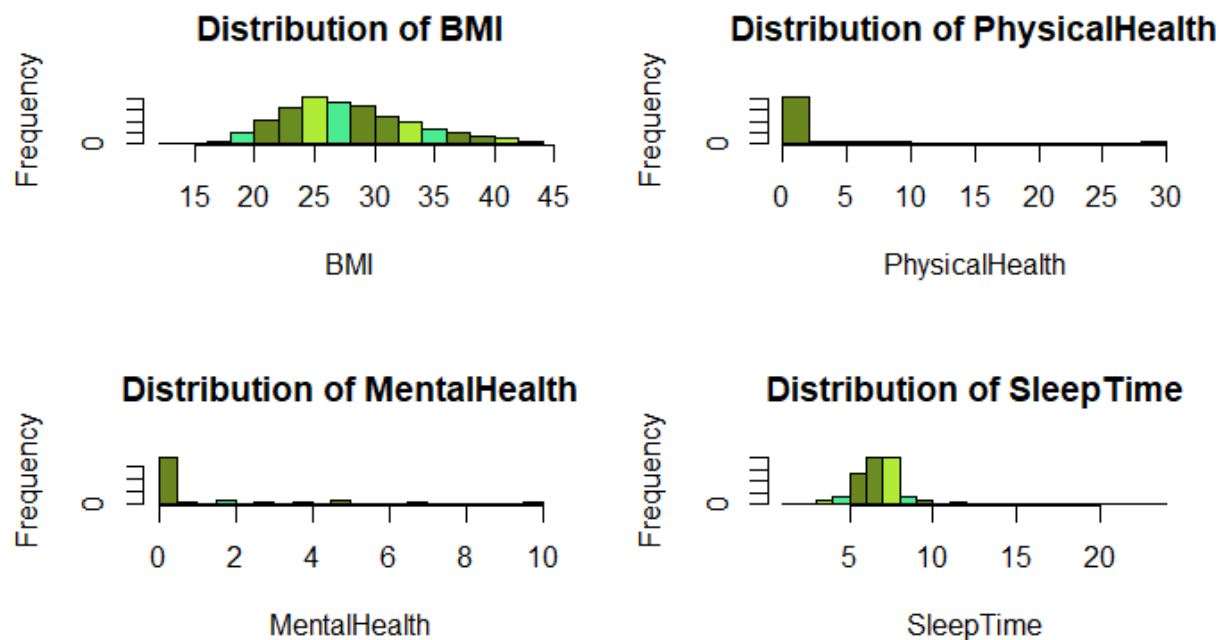


Рисунок. Анализ распределения числовых переменных по частоте

Для оценки нормальности распределения данных был применен метод визуализации с использованием графика квантиль-квантиль (QQ-plot), результаты которого представлены на Рисунке ниже.

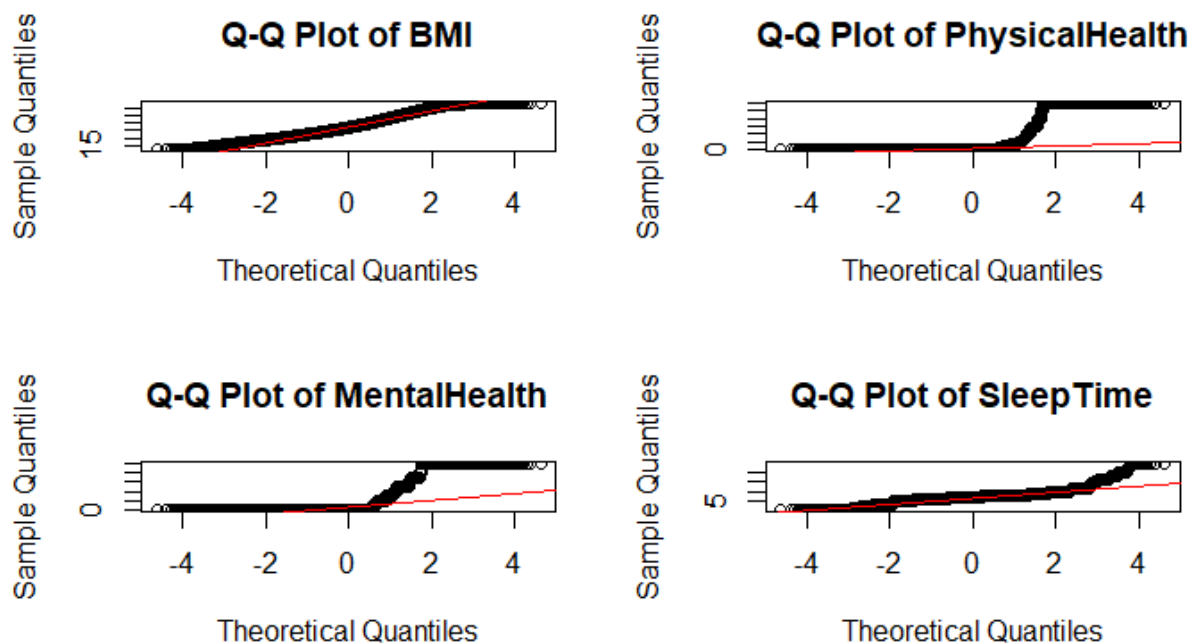


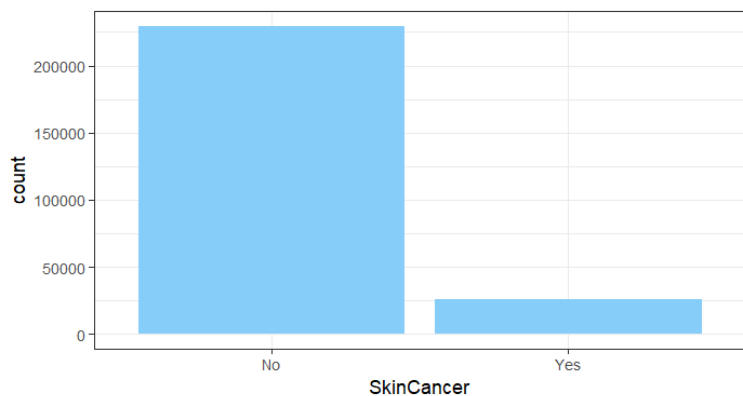
Рисунок. Q-Q-plot

Точки на графике демонстрируют систематическое отклонение от прямой линии, что указывает на несоответствие распределения данных нормальному закону. Наблюдаемые кривые имеют нелинейный характер, что дополнительно подтверждает отклонение от нормальности.

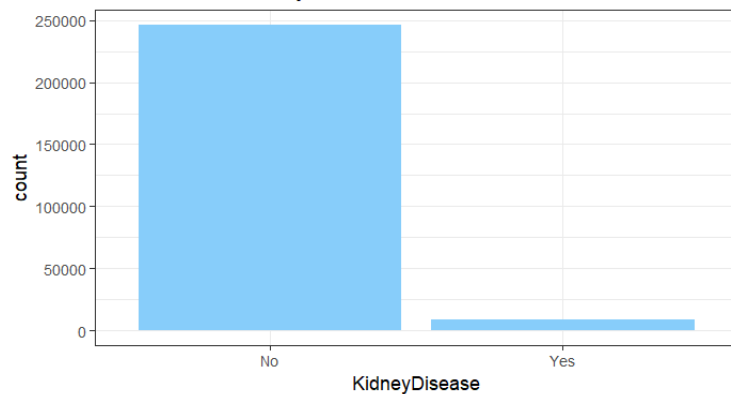
На основании этих визуальных признаков можно сделать предварительное заключение о том, что исследуемые данные с высокой вероятностью не соответствуют нормальному распределению.

Для оценки сбалансированности данных в рамках каждого категориального признака были построены столбчатые диаграммы (барплоты), представленные на Рисунке:

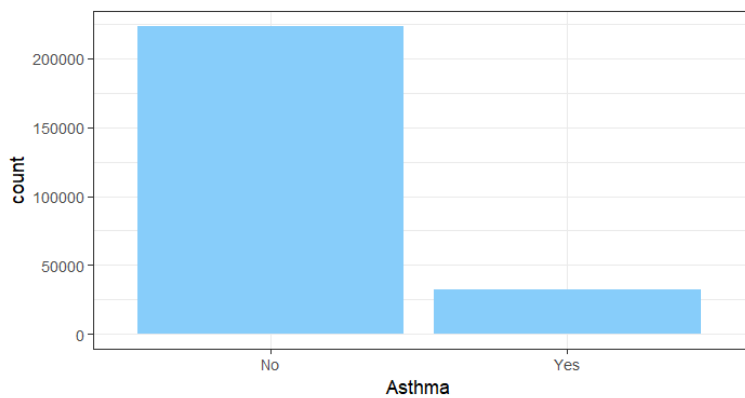
Distribution of SkinCancer



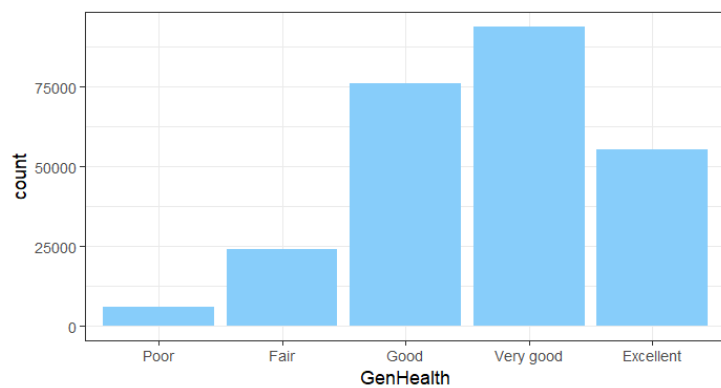
Distribution of KidneyDisease



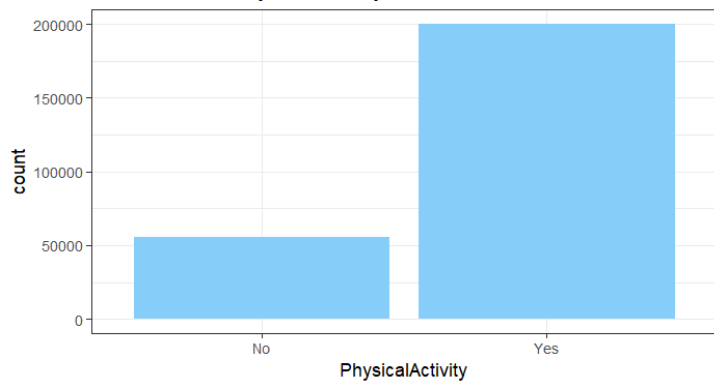
Distribution of Asthma



Distribution of GenHealth



Distribution of PhysicalActivity



Distribution of Diabetic

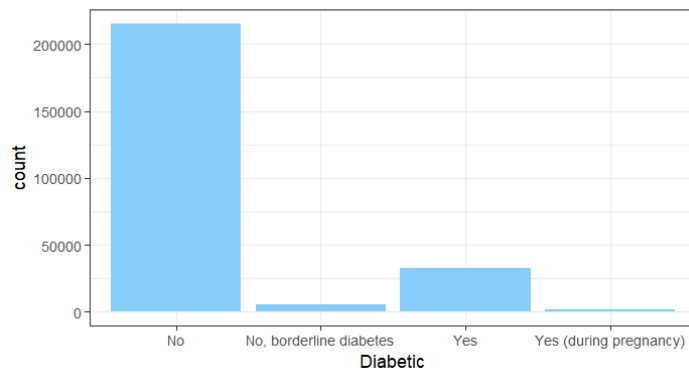
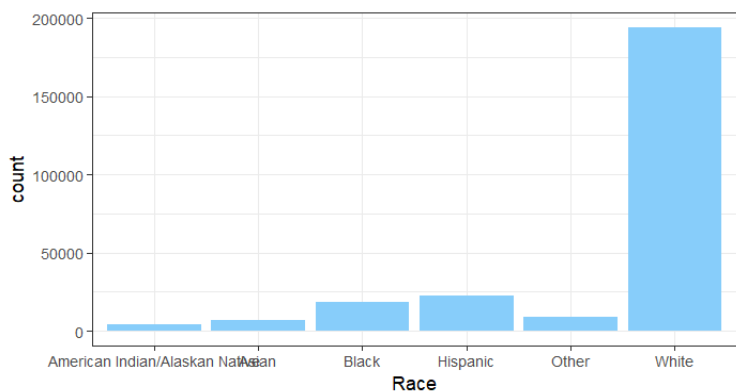


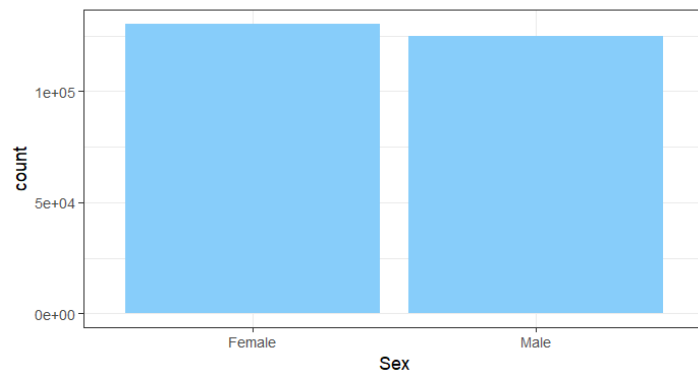
Рисунок. Визуализация категориальных данных



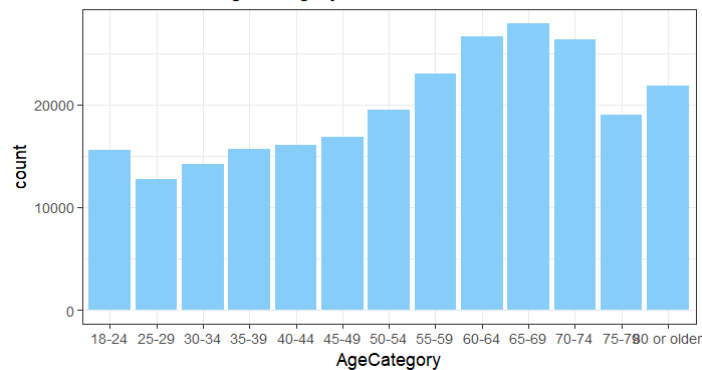
Distribution of Race



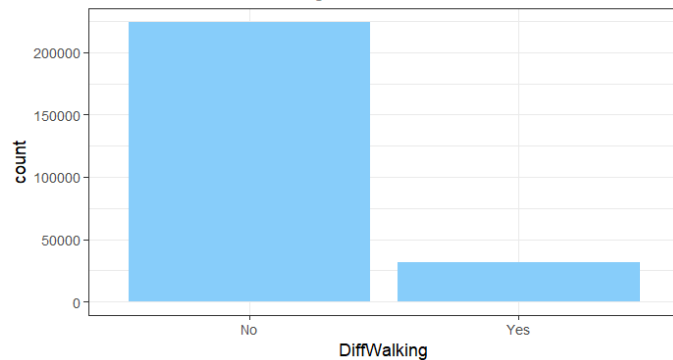
Distribution of Sex



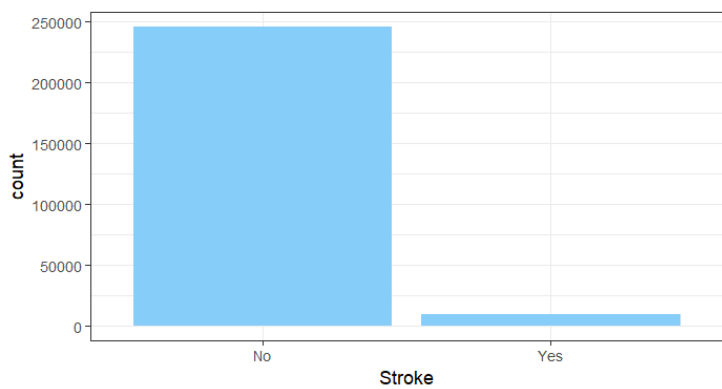
Distribution of AgeCategory



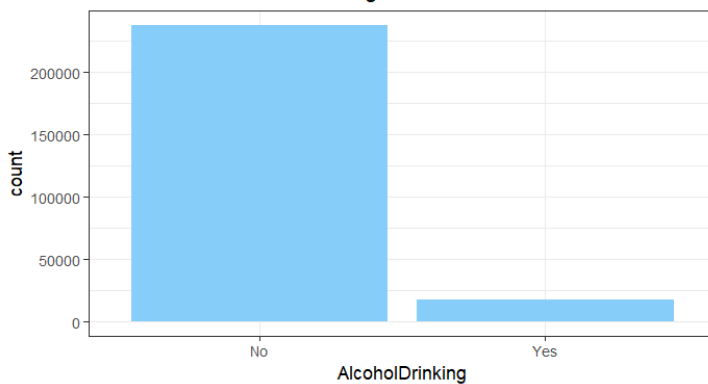
Distribution of DiffWalking



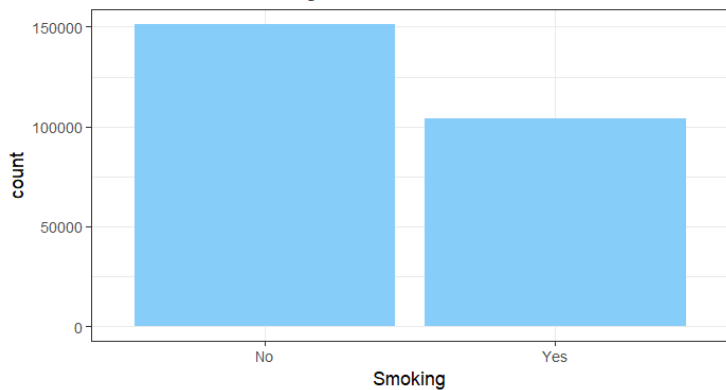
Distribution of Stroke



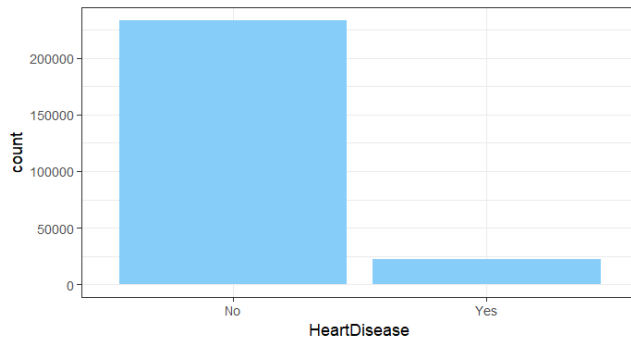
Distribution of AlcoholDrinking



Distribution of Smoking



Distribution of HeartDisease



Примечание: под базовой визуализацией я не имела ввиду наличие библиотеки `ggplot2`, а именно первичную оценку по одному признаку характера данных. Анализ представленных результатов выявляет существенную несбалансированность во всех исследуемых признаках, за исключением пола. Несбалансированность может привести к смещению в оценках параметров и снижению статистической мощности тестов.

Часть третья. Статистических анализ

При анализе числовых переменных первый этап оценки — соответствие данных нормальному распределению.

- 1) **Визуальная оценка** распределения с помощью **графических методов (QQ-plot и гистограмма)**. QQ-plot сравнивает квантили эмпирического и теоретического нормального распределений. Гистограмма показывает форму распределения данных.
- 2) **Статистические тесты** на нормальность, включая тест Шапиро-Уилка (мощный для малых выборок), тест Колмогорова-Смирнова (для больших выборок) и тест Андерсона-Дарлинга (эффективен для "хвостов" распределения).

Нулевая гипотеза (H_0): проверяемый признак (переменная) подчиняется закону нормального распределения.

Альтернативная гипотеза (H_1): переменная не подчиняется закону нормального распределения.

Тест Андерсона-Дарлинга показал результаты, представленные в Таблице 2:

Таблица 2 — Результаты теста Андерсона-Дарлинга

Признак	statistics, p-value	Принимаем/отклонить Н0
<i>BMI</i>	$A = 1390.7, p\text{-value} < 2.2e-16$	отклоняем
<i>PhysicalHealth</i>	$A = 61245, p\text{-value} < 2.2e-16$	отклоняем
<i>MentalHealth</i>	$A = 47931, p\text{-value} < 2.2e-16$	отклоняем
<i>SleepTime</i>	$A = 8389.7, p\text{-value} < 2.2e-16$	отклоняем

Тест Шапиро-Уилка не подошел из-за более 5к количества наблюдений в выборке, но подходит тест Колмагорова-Смирнова, результаты представлены в Таблице 3:

Таблица 3 — Результаты теста Колмагорова-Смирнова

Признак	statistics, p-value	Принимаем/отклонить Н0
<i>BMI</i>	$D = 0.061346, p\text{-value} < 2.2e-16$	отклоняем
<i>PhysicalHealth</i>	$D = 0.3824, p\text{-value} < 2.2e-16$	отклоняем
<i>MentalHealth</i>	$D = 0.17869, p\text{-value} < 2.2e-16$	отклоняем
<i>SleepTime</i>	$D = 0.17869, p\text{-value} < 2.2e-16$	отклоняем

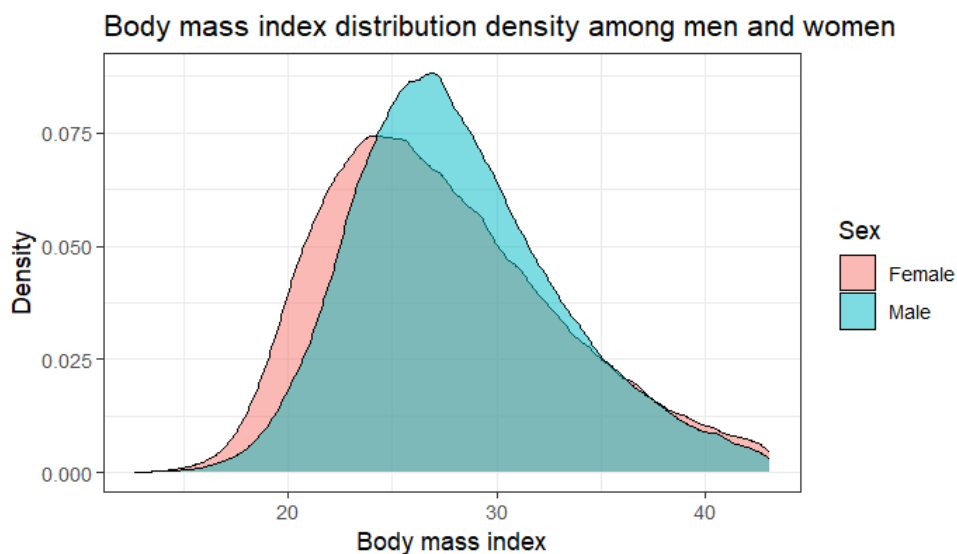
Согласно полученным результатам, мы отклоняем нулевую гипотезу во всех 4 случаях. Принимаем альтернативную гипотезу: все четыре признака не подчиняются закону нормального распределения.

Корреляционный анализ между числовыми переменными

Нулевая гипотеза (H0): среднее по BMI статистически не отличается между мужчинами и женщинами в выборке.

Альтернативная гипотеза (H1): среднее по BMI статистически отличается между мужчинами и женщинами.

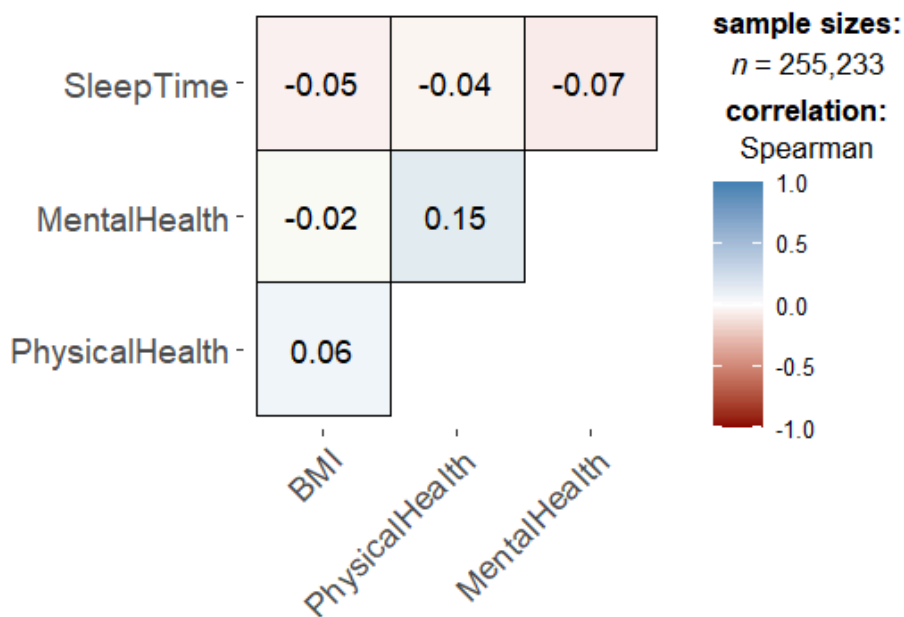
Большая выборка, непараметрический тест Манна-Уитни, группы считаются независимыми, получилась статистика $W = 7300283454$, $p\text{-value} < 2.2e-16$, что дает возможность отклонить нулевую гипотезу и принять альтернативную, средние между двумя группами статистически значимо отличимы. При рассмотрении средних в двух группах, было замечено, что с установленной задачей ответ не сошелся: средней индекс массы тела среди женщин составил 27.4, среди мужчин 28. Для оценки результатов был построен график плотности, представленный на Рисунке ниже:



Анализ полученных данных показывает: несмотря на статистически значимый результат, фактическая разница в средних значениях массы тела

между двумя группами оказывается минимальной. Я задумалась об интерпретации статистической значимости в контексте больших выборок, так как незначительные различия между группами могут быть определены как статистически значимые. Это явление я нашла как **"проблему большой выборки"** или **"парадокс р-значения"**.

Для коэффициента корреляции количественных переменных с индексом массы тела получила следующие результаты, оценивая как непараметрический по методу спирмена: BMI и MentalHealt = -0.0243326, BMI и SleepTime = -0.05401905, BMI и PhysicalHealth = 0.05941798. Общая таблица корреляции количественных переменных указана ниже:



X = non-significant at $p < 0.05$ (Adjustment: Holm)

Все три корреляции имеют значения, близкие к нулю, что говорит о слабых или практически отсутствующих монотонных зависимостях между индексом массы тела (BMI) и остальными переменными.

Анализ категориальных переменных по целевой переменной Heart Disease был проанализирован по следующим критериям: 2 сравниваемые группы, бинарные катег переменные с большой выборкой - лучше применить критерий хи-квадрат Пирсона.

Результаты показали во всех случаях очень низкие **p-value (меньше 2.2e-16)**, что означает, что все связи между сердечными заболеваниями и другими переменными считаются статистически значимыми.

Для анализа количественного признака в контексте трех и более групп было принято решение провести группировку по категориальному номинальному признаку Race. Учитывая специфику данных, а именно отсутствие нормального распределения исследуемой величины или неравенство дисперсий между группами, а также наличие трех и более сравниваемых групп, был выбран непараметрический метод анализа - критерий Краскела-Уоллиса. Этот статистический тест представляет собой непараметрический аналог однофакторного дисперсионного анализа (ANOVA) и особенно эффективен в ситуациях, когда: данные не соответствуют предположениям параметрических тестов, выборки независимы, необходимо сравнить три или более группы.

Полученные результаты представлены в Таблице 4:

Таблица 4 — критерий Краскела-Уоллиса по признаку расы

Variable	P_Value
1 BMI	0.000000e+00
2 PhysicalHealth	1.093667e-143
3 MentalHealth	1.125231e-63
4 SleepTime	2.144990e-159

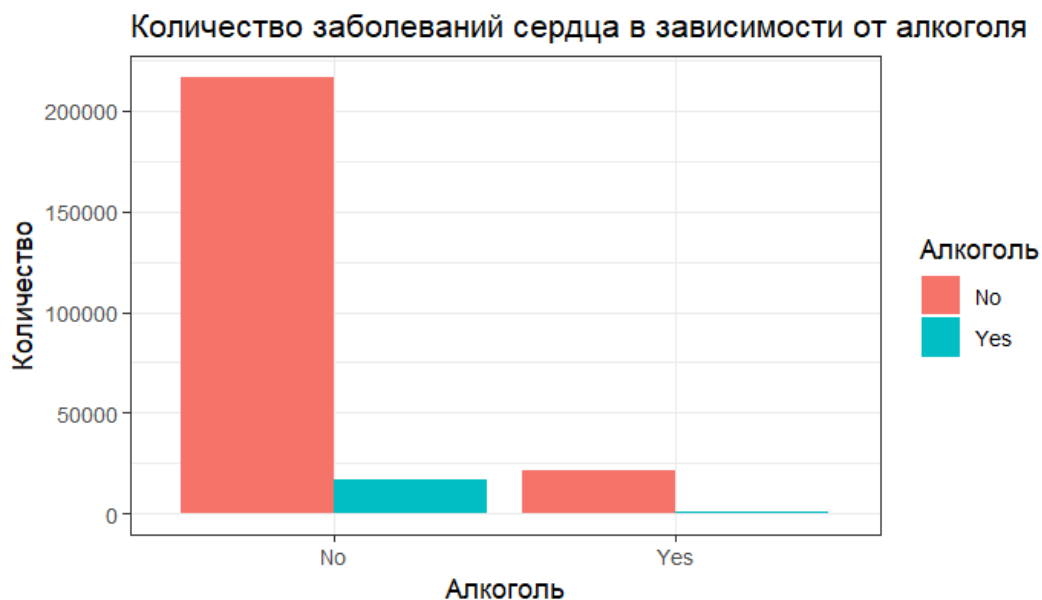
Согласно полученным результатам, каждый из признаков указывает на наличие значимых различий в показателях между группами по расе.

Сравнение групп в данных по признаку заболеваний сердца показали следующие **выводы**: люди с сердечными заболеваниями чаще курят (57.3%) по сравнению с теми, у кого нет заболеваний (39.2%), имеют более низкое общее состояние здоровья (24.1% против 9.8%) и чаще страдают от диабета (31.2% против 11%), мужчины чаще страдают от заболеваний (61.9% против 47.7% у женщин) - здесь может и правда, потому этот признак сбалансирован. Люди с сердечными заболеваниями имеют больше проблем с движением (31.6% против 10.4%), а также страдают от инсульта (14.9% против 2.61%) и заболеваний почек (11.6% против 2.84%).

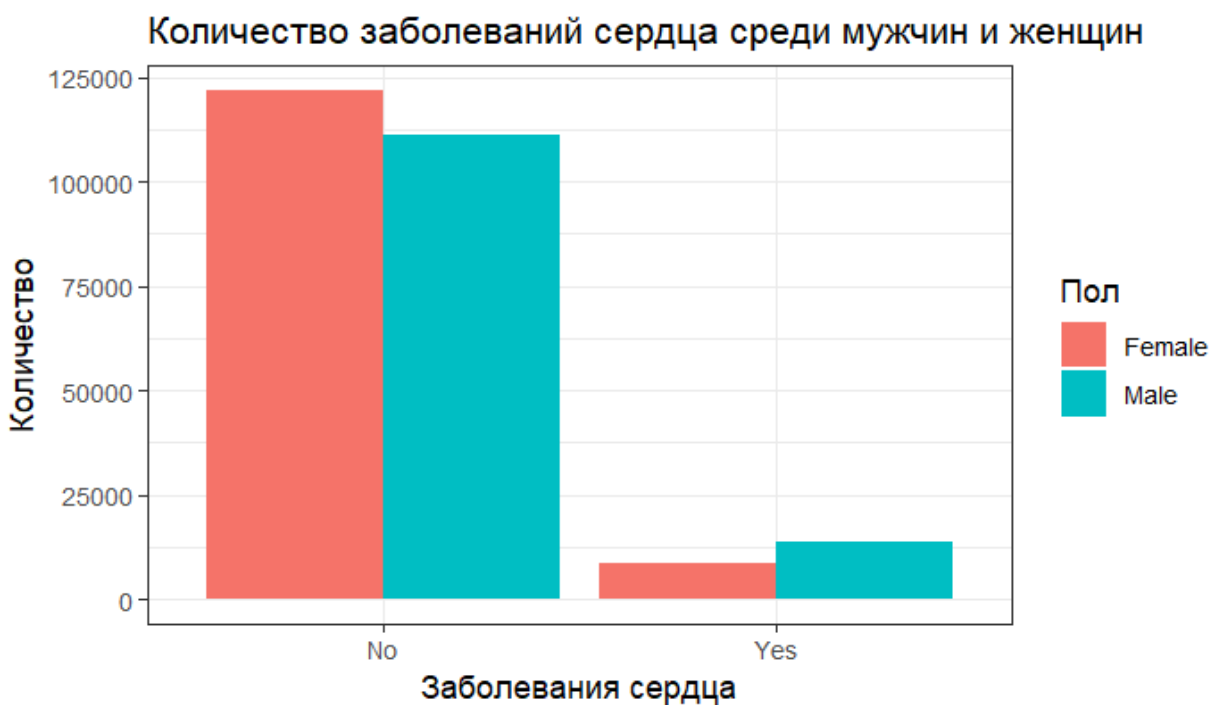
Часть четвертая. Продвинутая визуализация



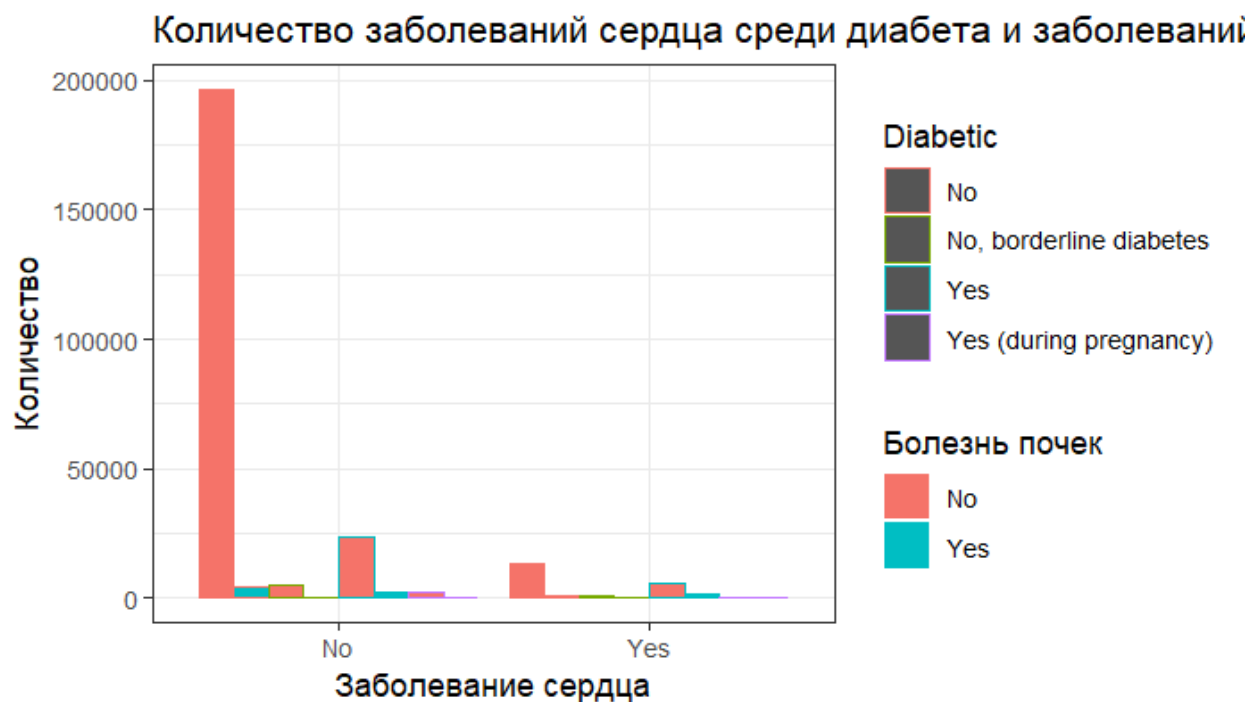
Преобладает группа отсутствующих диабет и без заболеваний сердца в обеих группах, наименьшие значения показаны во время беременности в обеих группах.

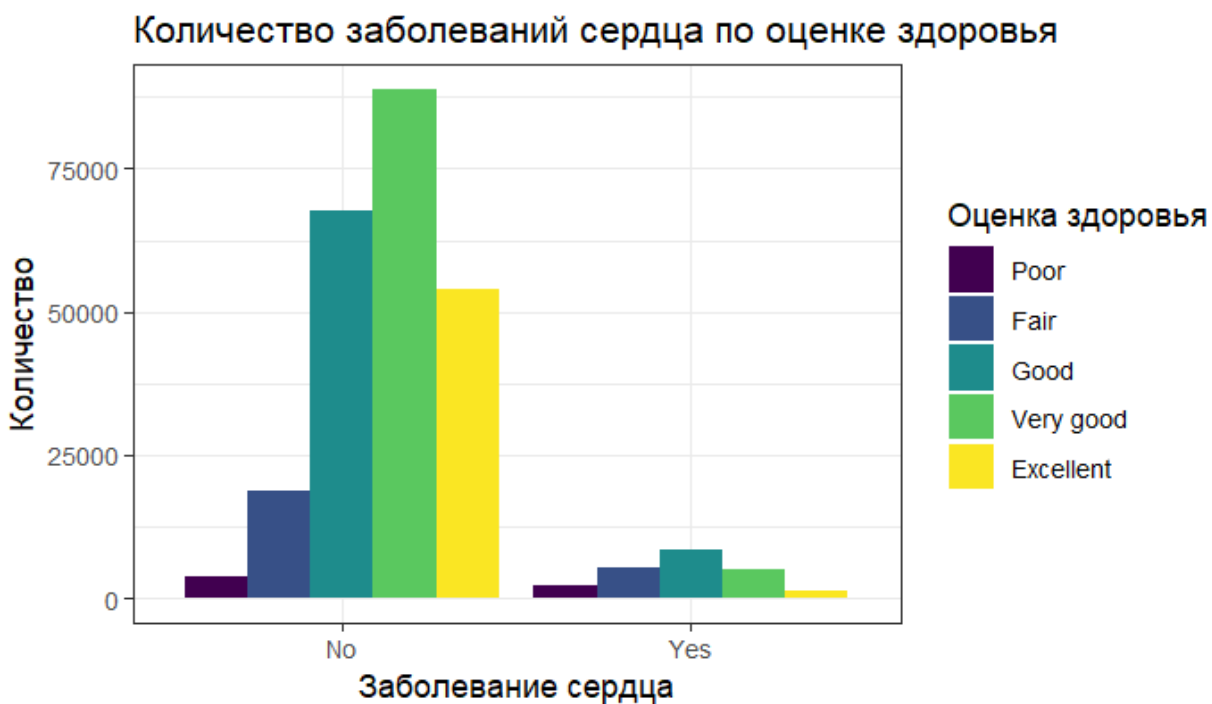


В обеих группах среди непьющих нет заболеваний сердца.

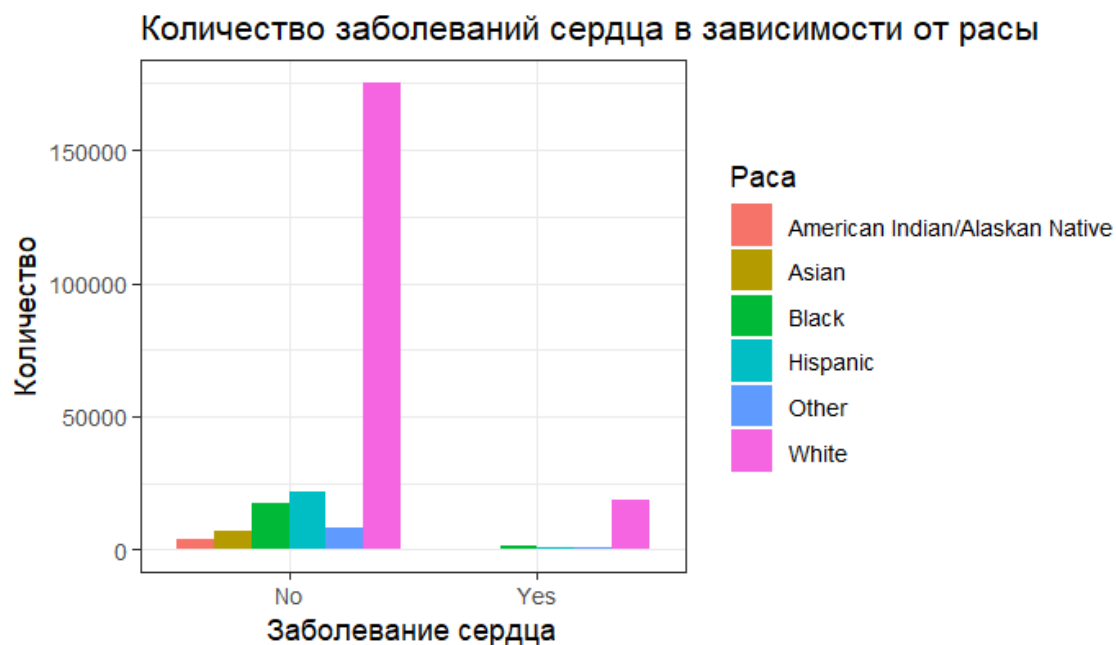


В полученных данных больше здоровых людей, практически одинаковое соотношение в двух случаях.

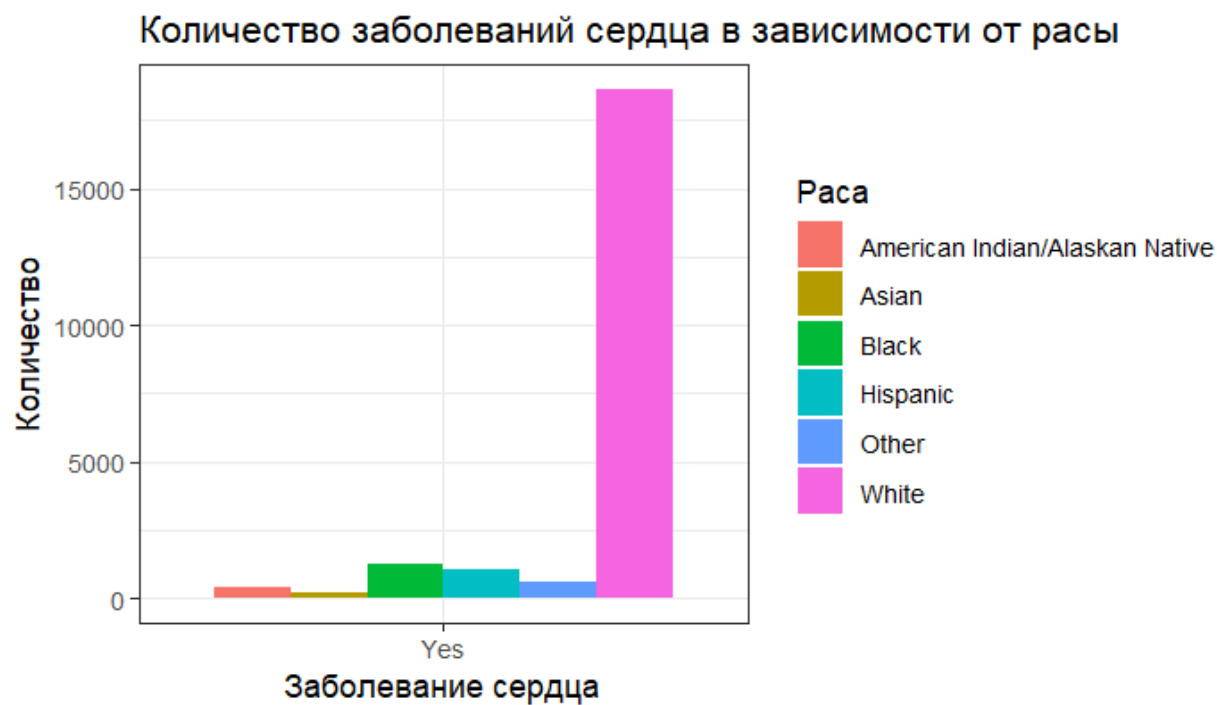


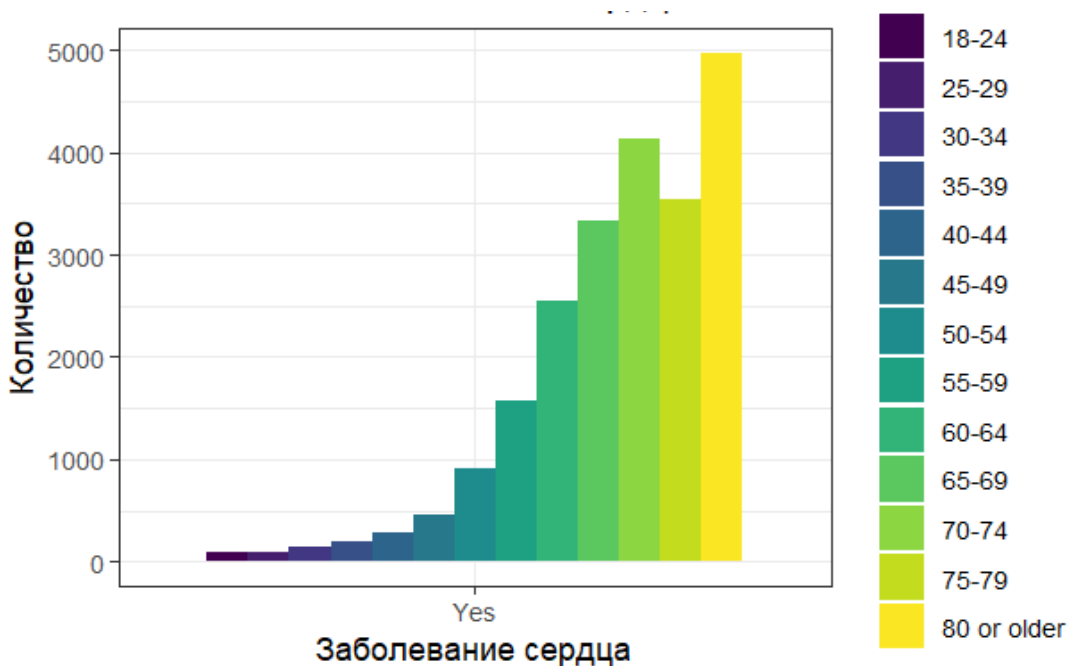


Люди с заболеваниями сердца в целом оценивают свое состояние здоровья хуже.



В обеих группах наиболее высокие показатели у белой расы. Если рассмотреть отдельно для заболеваний:





Прослеживается увеличение тенденции с возрастом к заболеванию сердца.

Анализ между категориальными и количественными переменными

Как было видно ранее, среди категориальных переменных близкое к равному число наблюдений представлено только у sex. В проверке гипотез было принято решение остановиться именно на нем, но в добавочной визуализации хотелось посмотреть распределение массы тела от возраста и курения (сначала по отдельности, а потом в группе М и Ж).

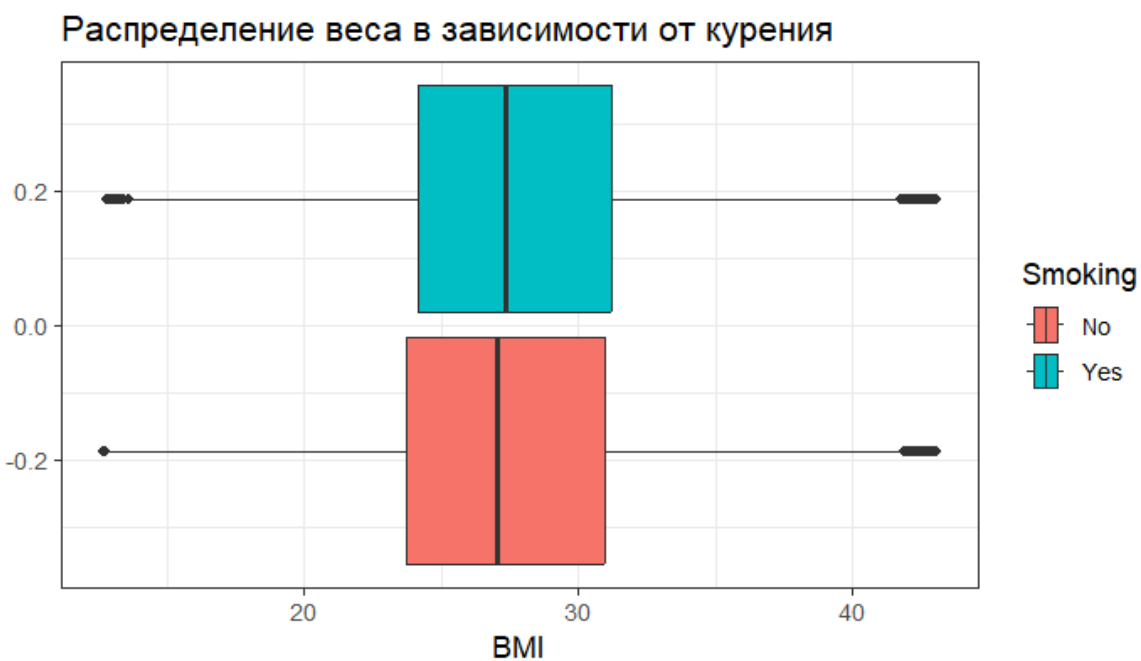


Рисунок. Распределение веса в зависимости от курения (использование box plot)

Видно, что средние примерно равны.

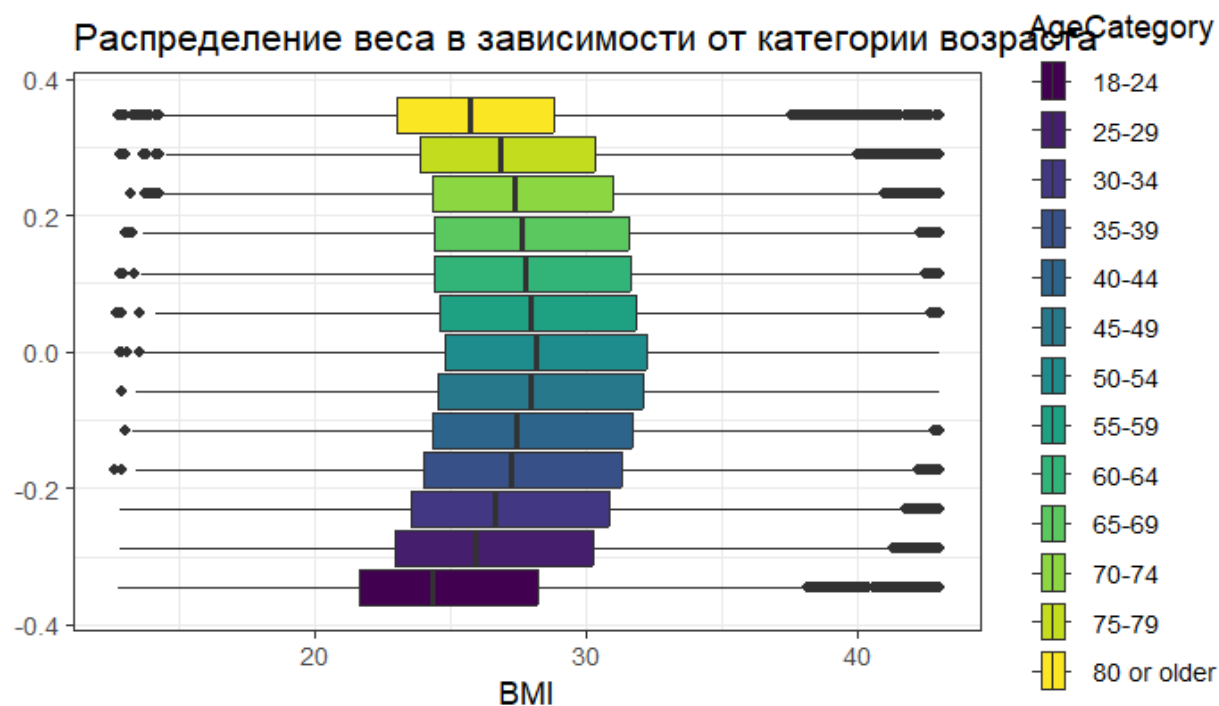


Рисунок. Распределение веса в зависимости от категории возраста

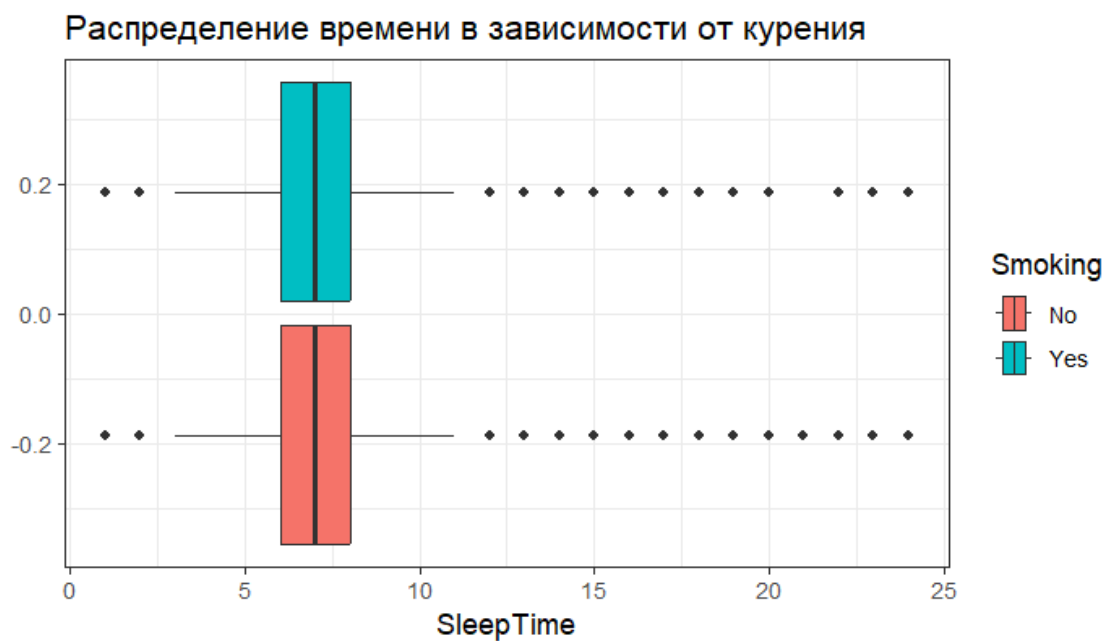


Рисунок. Распределение количества сна от курения

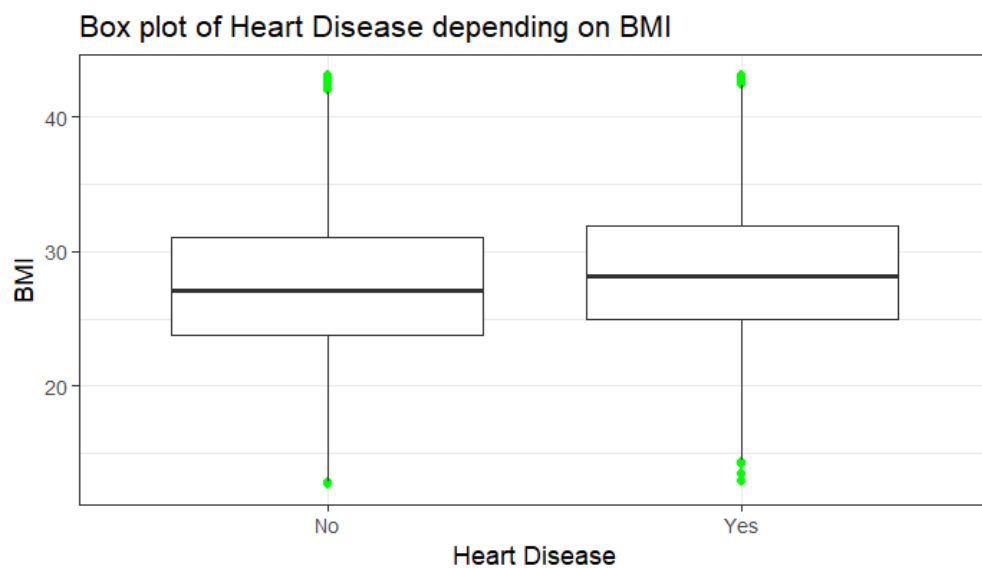


Рисунок. Зависимость индекса массы тела от заболеваний сердца

Часть шестая. Регрессионный анализ

Для анализа данных была выбрана модель логистической регрессии, что обусловлено природой целевой переменной. В данном случае, таргетная (зависимая) переменная является бинарной категориальной, то есть может принимать только два возможных значения. Путем перебора признаков, выбрала более оптимальные (перепроверила и по значимости, на всех был перебор). Репорт представлен ниже:

Coefficients:

	Estimate	Std. Error	z value	
Pr(> z)				
(Intercept)	-3.09456	0.01847	-167.501	<
2e-16 ***				
AgeCategory.L	3.98013	0.07212	55.187	<
2e-16 ***				
AgeCategory.Q	-0.14602	0.06559	-2.226	
0.0260 *				
AgeCategory.C	-0.32919	0.06373	-5.166	
2.4e-07 ***				
AgeCategory^4	0.05666	0.06375	0.889	
0.3741				
AgeCategory^5	0.09848	0.06545	1.505	
0.1324				
AgeCategory^6	-0.01721	0.06496	-0.265	
0.7911				
AgeCategory^7	-0.09802	0.06195	-1.582	
0.1136				
AgeCategory^8	0.13630	0.05770	2.362	
0.0182 *				



AgeCategory^9 0.5462	-0.03262	0.05406	-0.603
AgeCategory^10 0.5522	0.02994	0.05036	0.594
AgeCategory^11 0.7843	-0.01267	0.04627	-0.274
AgeCategory^12 0.2144	0.04932	0.03973	1.241
GenHealth.L 2e-16 ***	-1.63616	0.02855	-57.303 <
GenHealth.Q 0.1403	-0.03440	0.02333	-1.475
GenHealth.C 0.0454 *	0.03800	0.01899	2.001
GenHealth^4 0.4337	0.01142	0.01458	0.783
DiabeticNo, borderline diabetes 0.1362	0.06866	0.04607	1.490
DiabeticYes 2e-16 ***	0.48937	0.01783	27.440 <
DiabeticYes (during pregnancy) 0.0149 *	-0.29845	0.12260	-2.434
KidneyDiseaseYes 2e-16 ***	0.53332	0.02716	19.635 <
StrokeYes 2e-16 ***	1.02341	0.02509	40.786 <
DiffWalkingYes 2e-16 ***	0.16007	0.01904	8.407 <

Площадь под кривой составила 0.8199, но нулевое значение recall и переобучение модели. Поэтому после отфильтрованных признаков значимые

оказались AgeCat, где статистически значимыми признаками оказались

```
AgeCategory.L      4.810987      0.071236      67.536      < 2e-16 *** , AgeCategory.Q
-0.158519      0.065276      -2.428      0.0152 * , AgeCategory.C -0.399570      0.063445
-6.298 3.02e-10 *** , AgeCategory^5      0.129337      0.065171      1.985      0.0472 * ,
AgeCategory^8      0.126918      0.057389      2.212      0.0270 *
```

Точность вышла 0.9128639, но она не очень информативна, так как все признаки внутри данных не сбалансированы, recall = 0.13 - очень низкий.

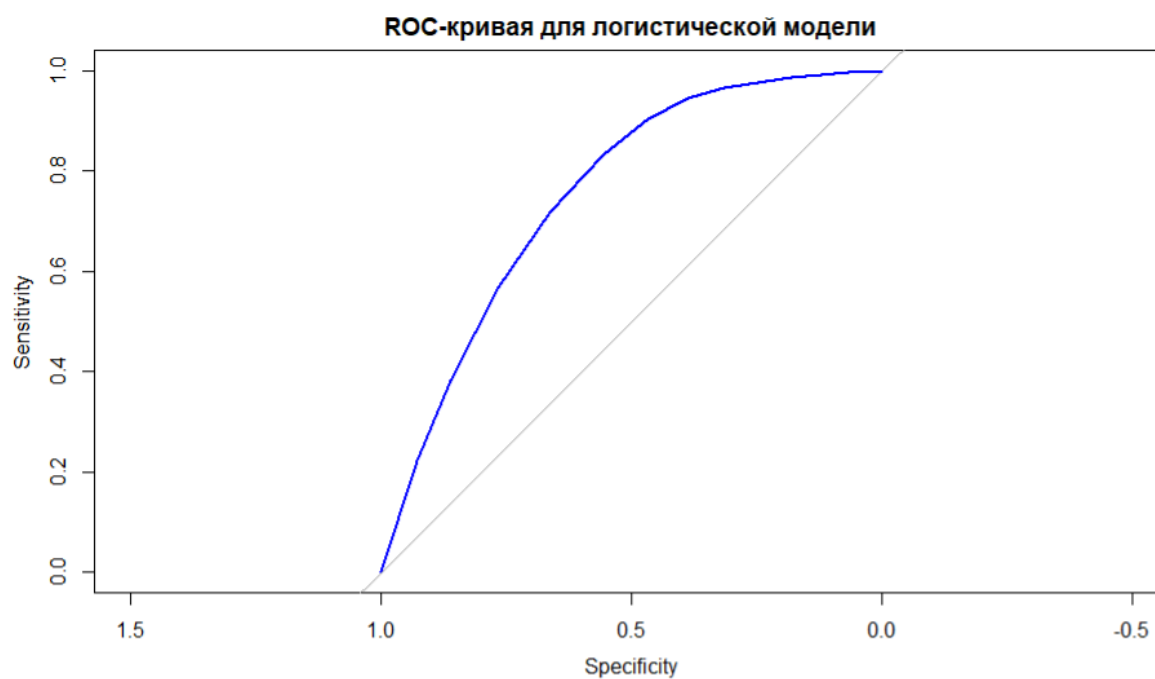


Рисунок. ROC-кривая предсказания сердечных заболеваний по категории возраста

Площадь под кривой: 0.7532393

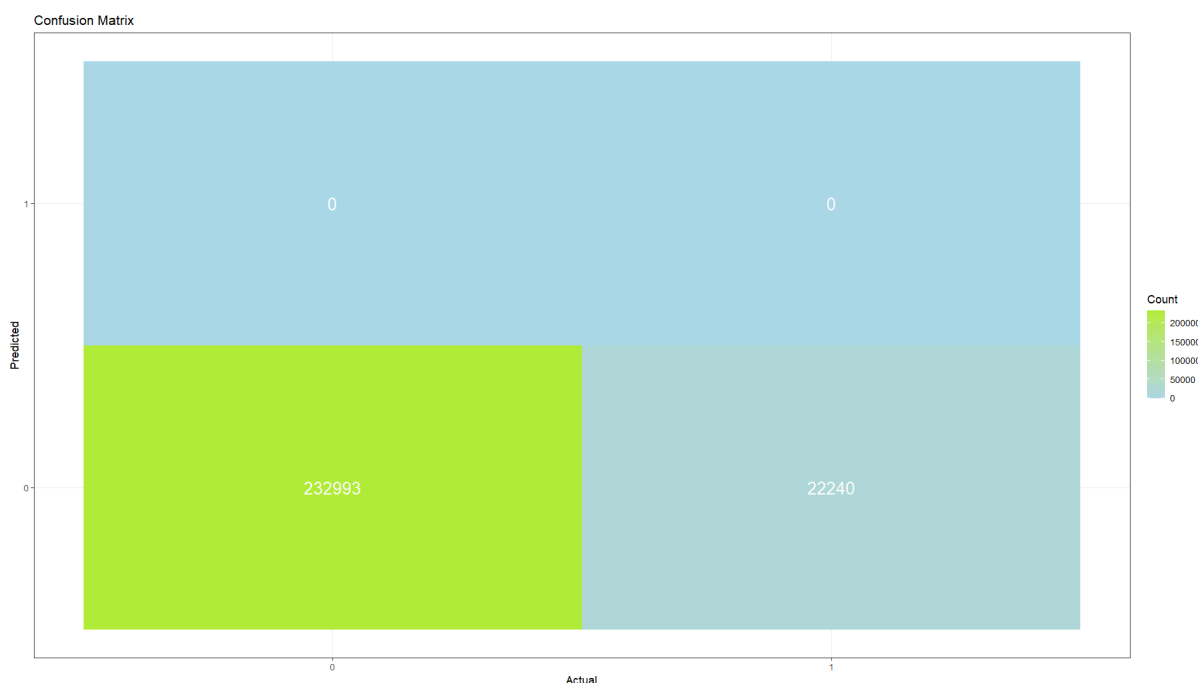


Рисунок. Тепловая карта

Результаты анализа показывают, что наиболее значимым предиктором является возрастная категория (AgeCategory), при этом модель демонстрирует $AUC = 0.753$, что свидетельствует о ее приемлемой предсказательной способности. Однако наблюдалась проблема дисбаланса классов, что отразилось на низком значении Recall (0.13), что требует дальнейшего улучшения модели.

Результаты логистической регрессии оказались неудовлетворительными, возможно по **несбалансированность классов, отсутствие гомоскедастичности и нарушение предпосылок**, может существенно снизить эффективность модели. **Или вариант логистическая регрессия** может оказаться недостаточно гибкой и может стоит подобрать

другую модель. Можно попробовать с различными алгоритмами, например, деревьями решений.

СРАВНЕНИЕ С АВТОРАМИ

Основные выводы авторов статей:

1. Из всех представленных графиков вывод: употребление алкоголя и курение не являются основными факторами развития сердечных заболеваний.
2. Нулевая гипотеза о нормальности данных отклонена.
3. Существует статистически значимая взаимосвязь между целевой переменной и категориальными признаками.
4. Нет корреляции между числовыми переменными в данных.
5. Оценка логистической регрессии по всем независимым признакам получила Recall_score: 0.08926080892608089, F1-score: 0.1532934131736527, Accuracy_score: 0.9115697310819262, Precision_score: 0.5423728813559322.

Высокая точность объясняется большим количеством здоровых людей. При работе с несбалансированными классами показатель бесполезен. Напомним, что более информативными показателями являются recall. Recall показывает, какую долю объектов положительного класса из всех объектов положительного класса нашел алгоритм.

Баланс между этими двумя показателями равен F1-score. Именно введение precision не позволяет нам записать все объекты в один класс, так как в этом случае мы получаем увеличение уровня ложных срабатываний.

СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ

1. Tsao C. W. et al. Heart disease and stroke statistics—2023 update: a report from the American Heart Association //Circulation. – 2023. – Т. 147. – №. 8. – С. e93-e621.
2. Lemieux A. et al. Meta-analysis of risk factors for congenital heart disease—part 2: maternal medication, reproductive technologies, and familial and fetal factors //Canadian Journal of Cardiology. – 2024.
3. Soni J. et al. Predictive data mining for medical diagnosis: An overview of heart disease prediction //International Journal of Computer Applications. – 2011. – Т. 17. – №. 8. – С. 43-48.