

RESEARCH BY EGOR ERMAKOV & ANDREY SAVARY | TEAM 5

# Medical Yandex Research

исследование факторов, влияющих на вероятность развития острого  
поражения почек

OVERVIEW

# structure

ПРЕДОБРАБОТКА      ОБЩИЙ АНАЛИЗ      СЧИТАЕМ МЕТРИКИ      ГИПОТЕЗЫ



we are here

## ПРЕДОБРАБОТКА

# initial research data



### Количественные

Возраст, длительность операции...

### Бинарные

Развитие ОПП, наличие различных заболеваний у пациента...

### Категориальные

Хроническая болезнь почек...

ПРЕДОБРАБОТКА

# обработка выбросов



## физиологически невозможные

Те значения, которые скорее всего попали по ошибке, в реальном мире они не возможны или очень редко попадаются, следовательно на них нельзя делать статистические выводы

## ПРЕДОБРАБОТКА

# ЧТО МЫ ЕЩЁ НАШЛИ

В ходе изучения данных мы смогли найти много неточностей и ошибок в датасете. Эти ошибки мы заменили или удалили из дальнейшего исследования.

КОЛОНКИ

ДУБЛИКАТЫ

ТИПЫ ДАННЫХ

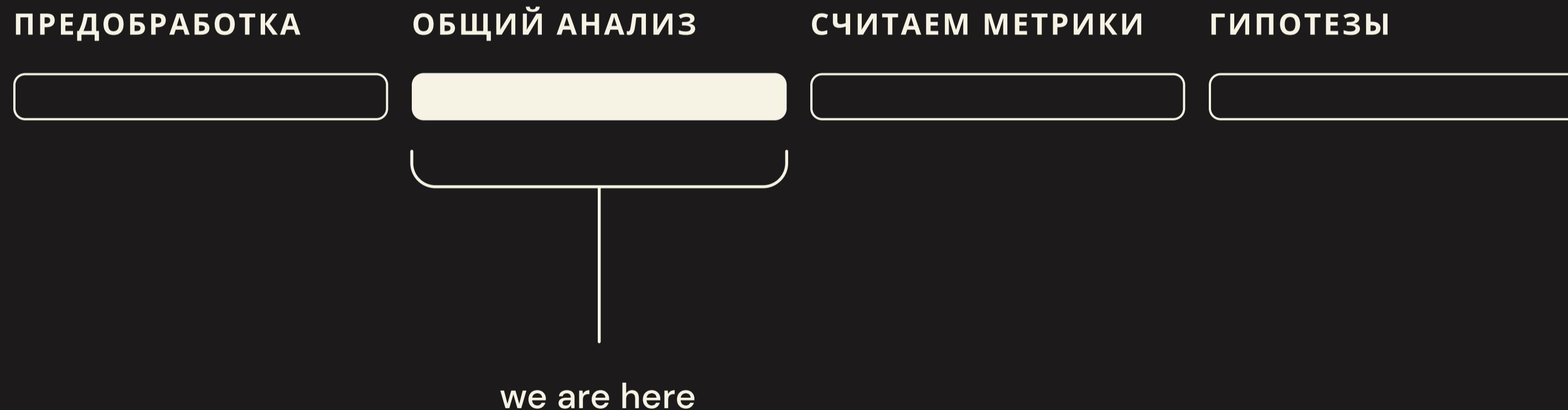
EXTRA DATA

МУЖЧИНЫ

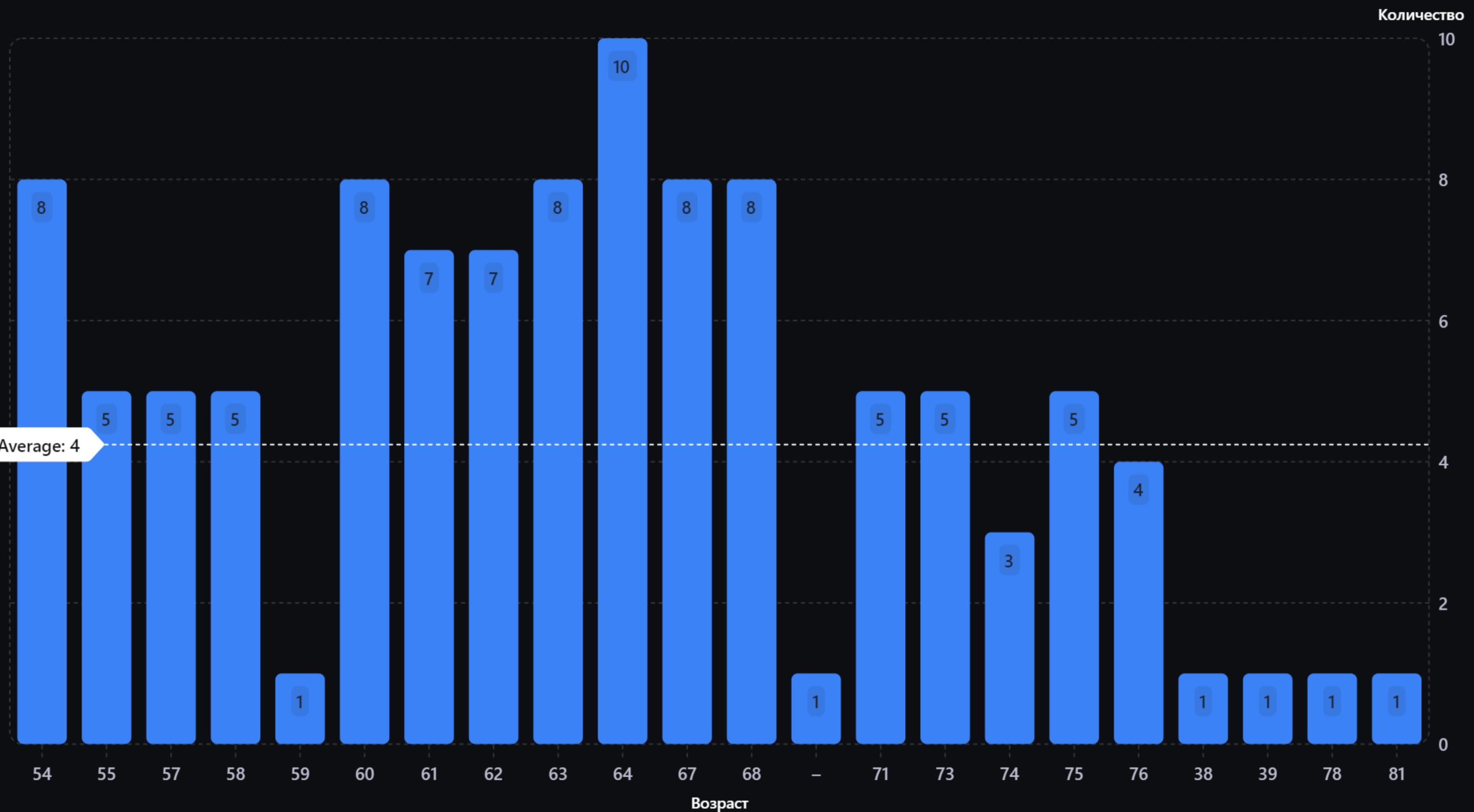


## OVERVIEW

# structure



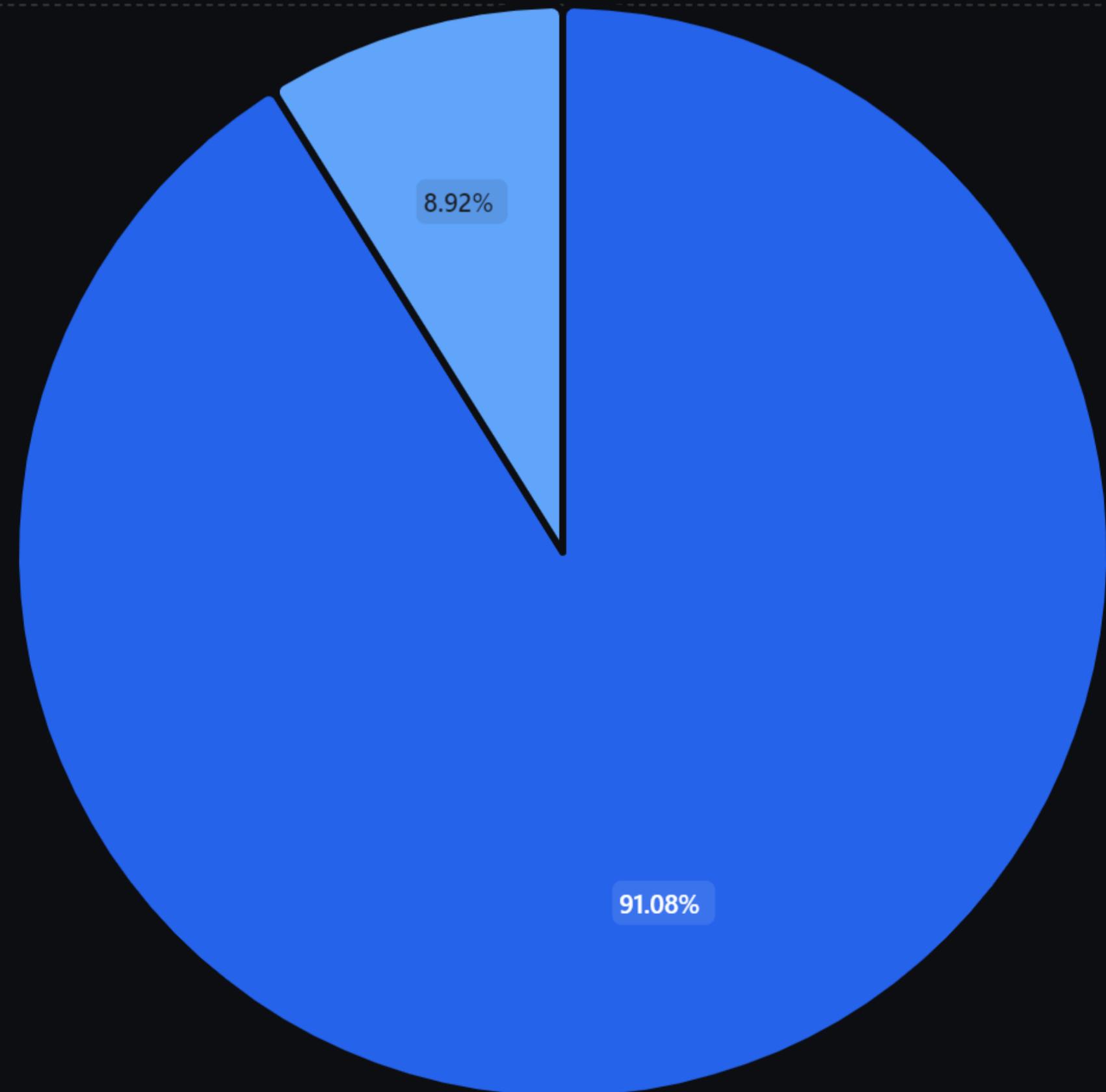
## Распределение возраста в исследовании



Количество мужчин и женщин в исследовании

**143**      **14**

Мужской      Женский

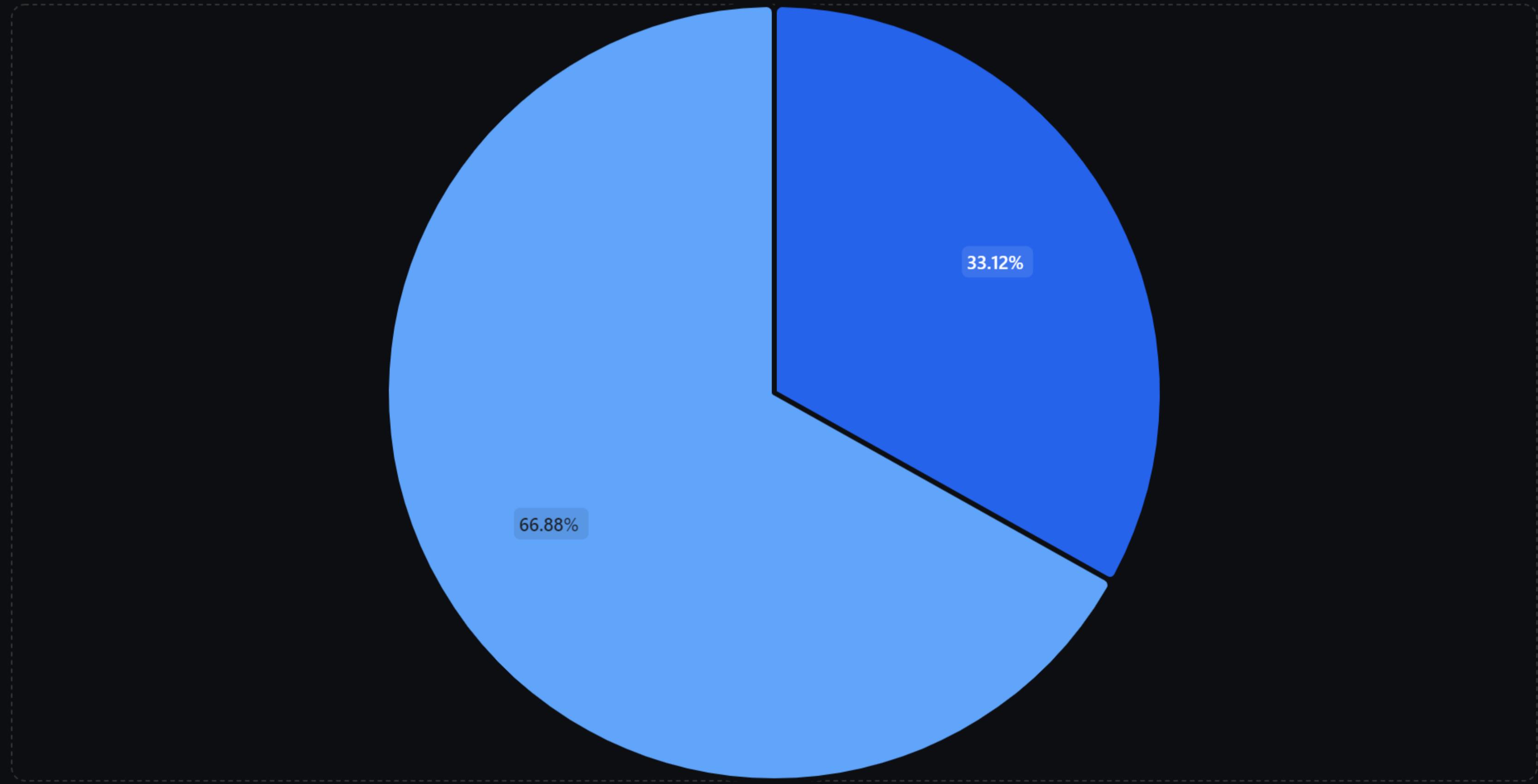




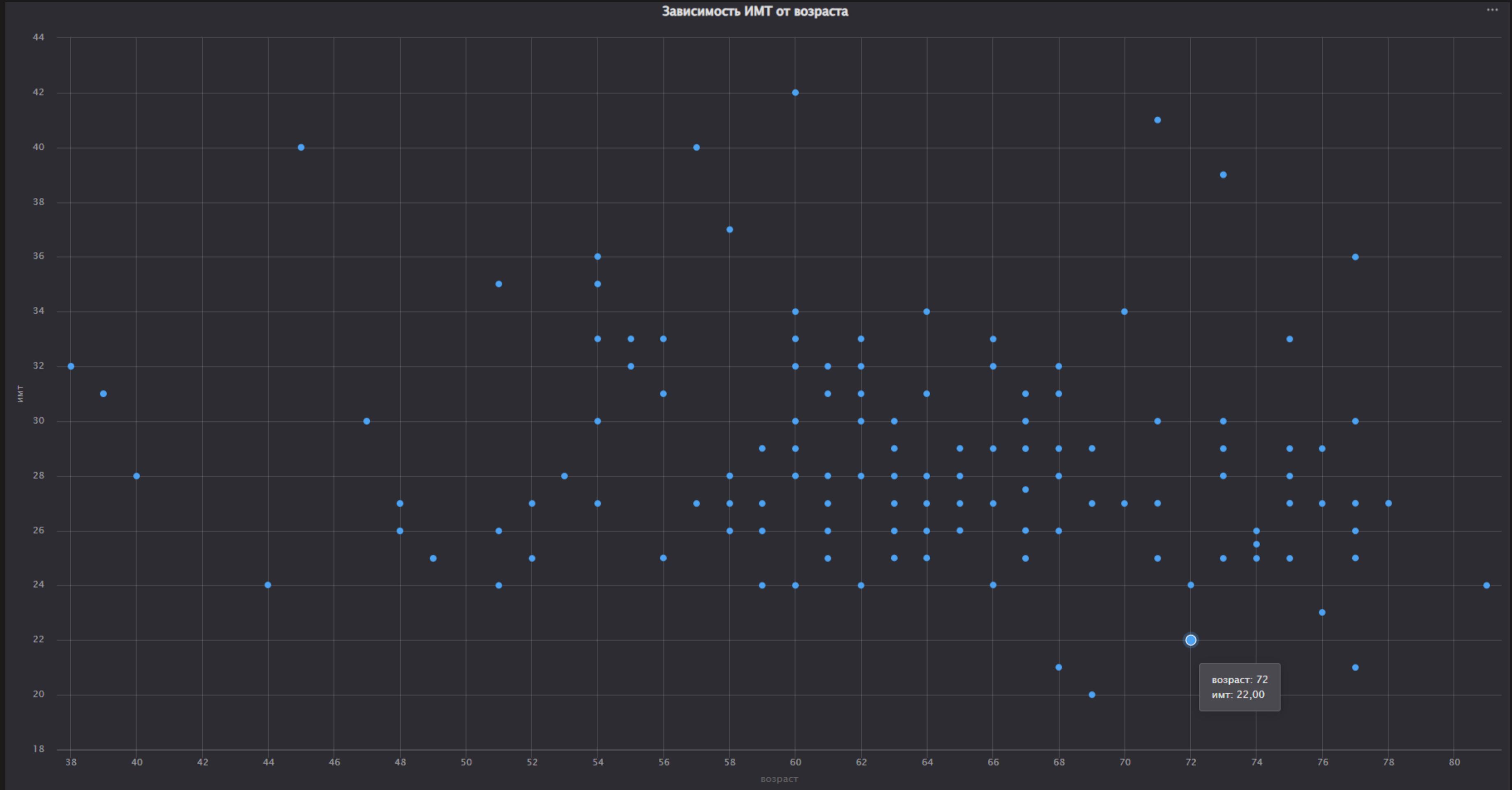
Распределение сахарного диабета в исследовании

**52      105**

Есть      Нет



### Зависимость ИМТ от возраста



## OVERVIEW

# structure

ПРЕДОБРАБОТКА      ОБЩИЙ АНАЛИЗ      СЧИТАЕМ МЕТРИКИ      ГИПОТЕЗЫ



we are here

■ Есть ОПП #1

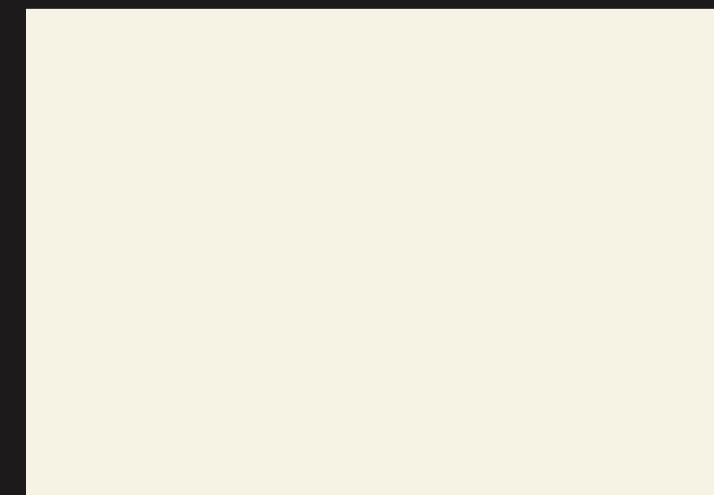
Те пациенты, у которых выявлено острое  
поражение почек

■ Нет ОПП #2

Те пациенты, у которых не выявлено острое  
поражение почек

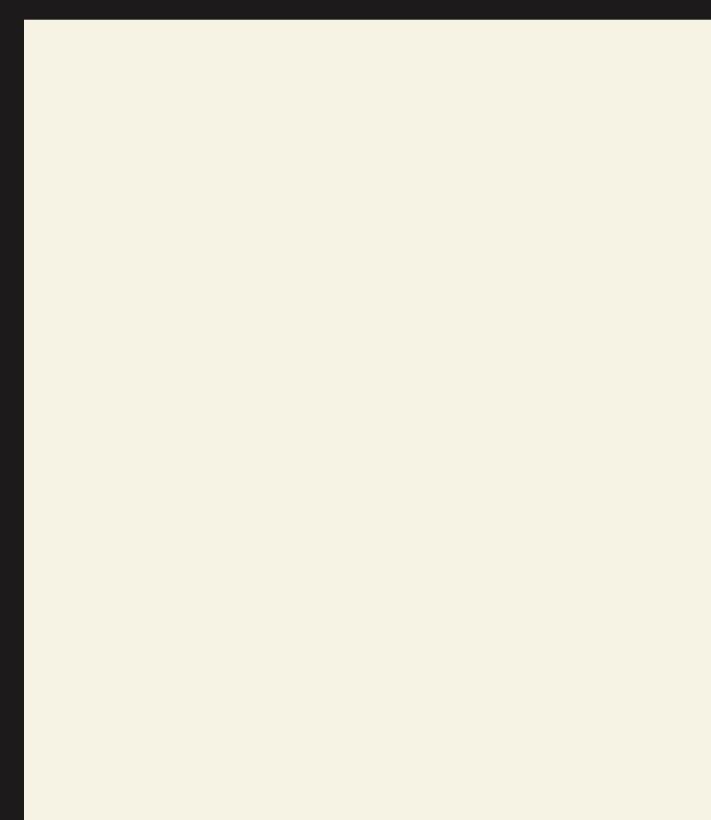
**36% | 26%**

Наличие сахарного  
диабета



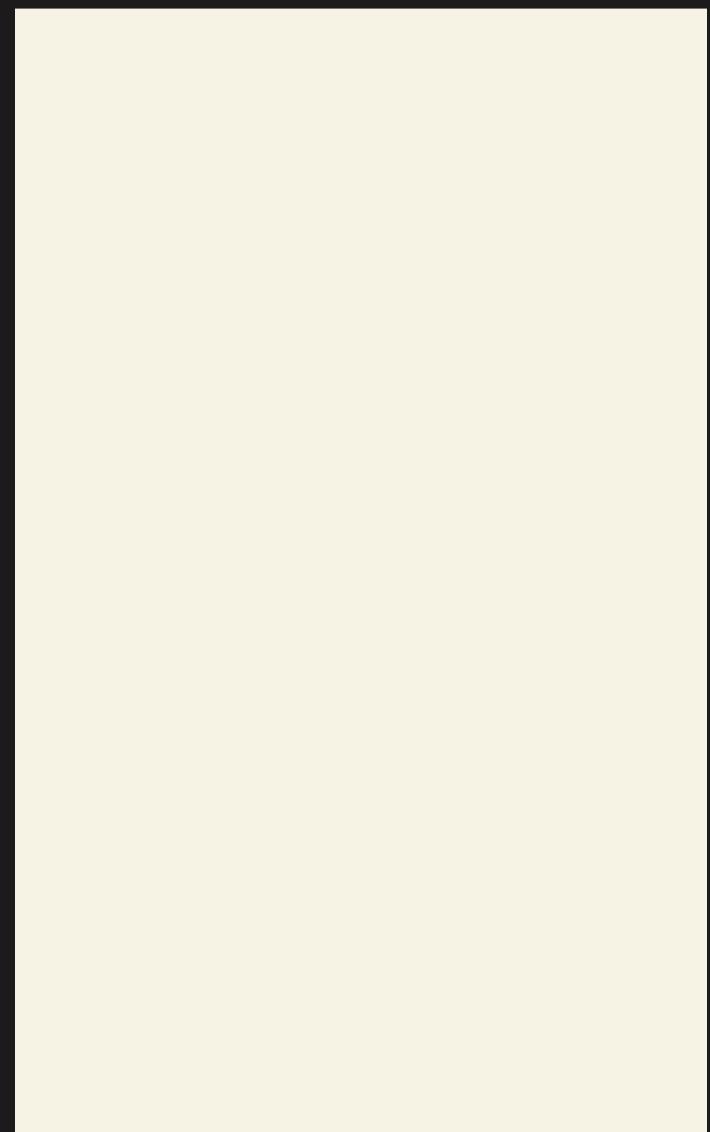
**88% | 96%**

Наличие  
гипертонии



**51% | 51%**

Наличие  
хронической  
болезни почек



**1.**

Сформулировать гипотезу

**2.**

Построить достаточное количество графиков, чтобы графически проанализировать данные

**3.**

Провести статистические тесты или найти коэффициенты корреляции

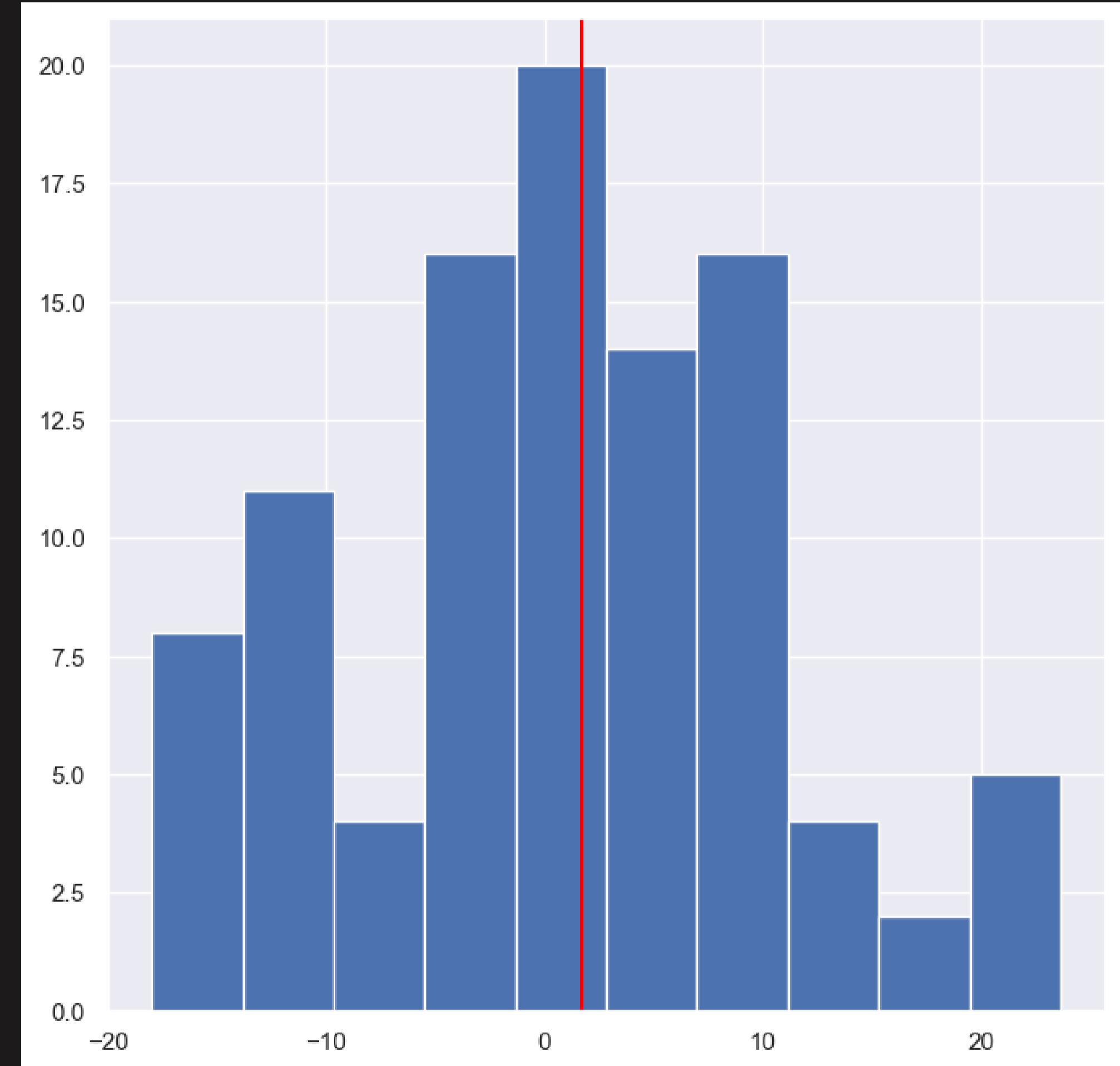
**Зависимость  
длительности  
операции от  
инфаркта миокарда**

## МЕТРИКИ

Исследовать зависимость  
длительности операции от  
факта перенесенного в  
прошлом инфаркта миокарда

Красная линия – Statistics

1.6910



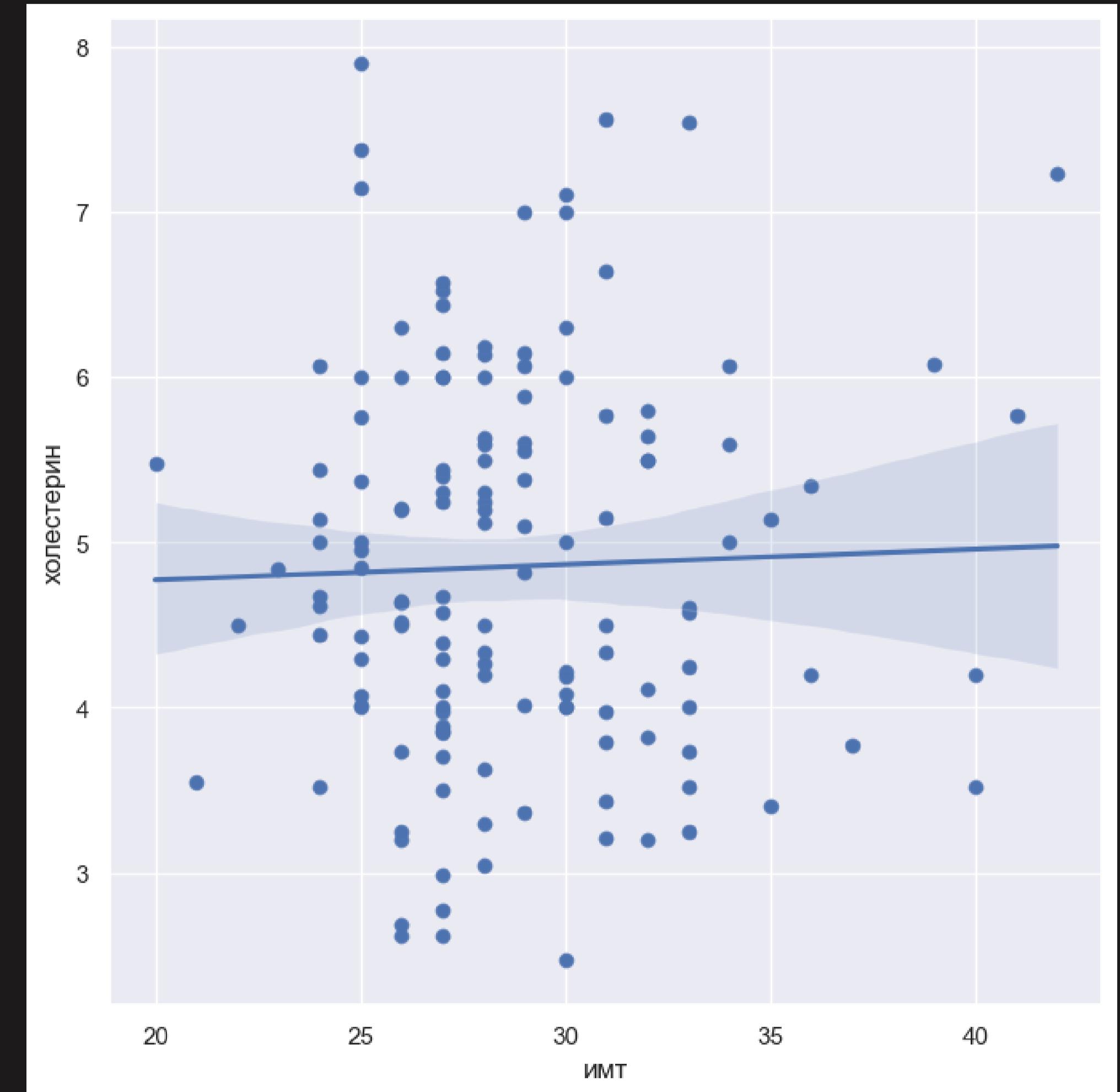
## МЕТРИКИ

Верно ли, что у пациентов с ИМТ выше нормы будет повышенный уровень холестерина?

Нельзя однозначно сказать

$\leq 25 - 115$

$>25 - 26$

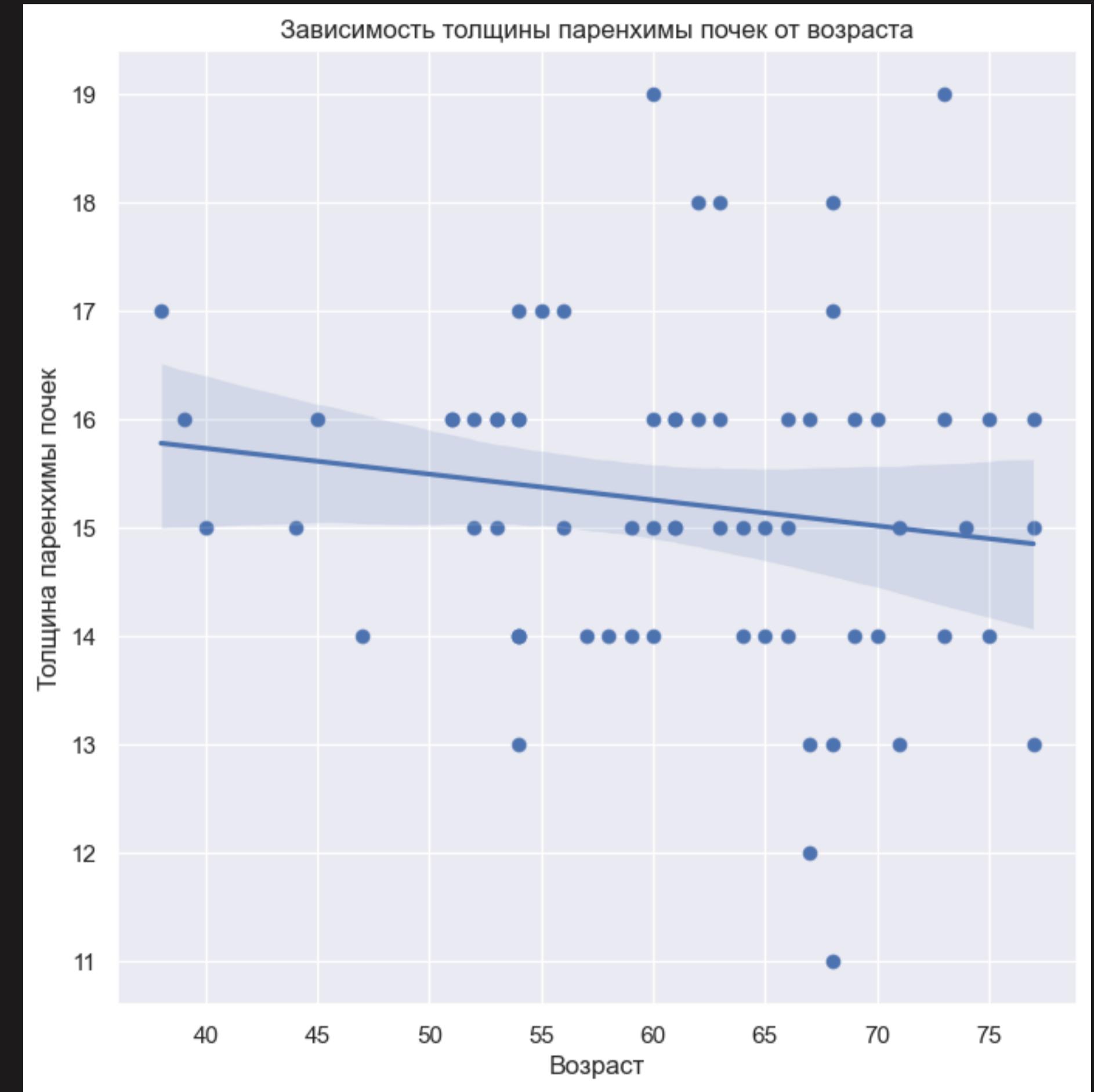


## МЕТРИКИ

Верно ли, что даже без хронических болезней почек с возрастом толщина паренхимы почек уменьшается

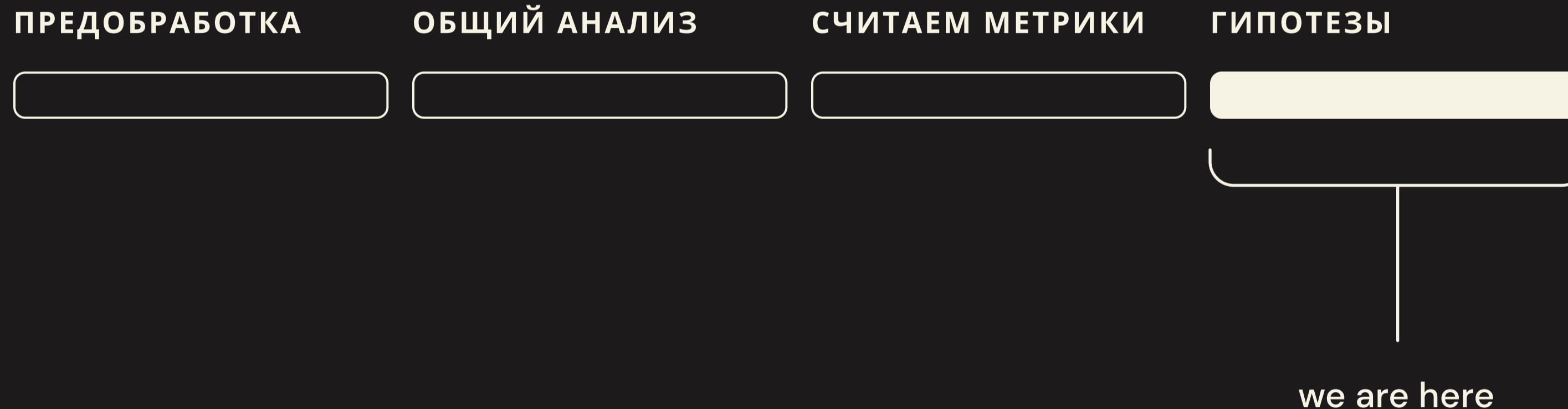
Коэффициент кореляции Спирмена

Коэффициент корреляции: -0.16493



## OVERVIEW

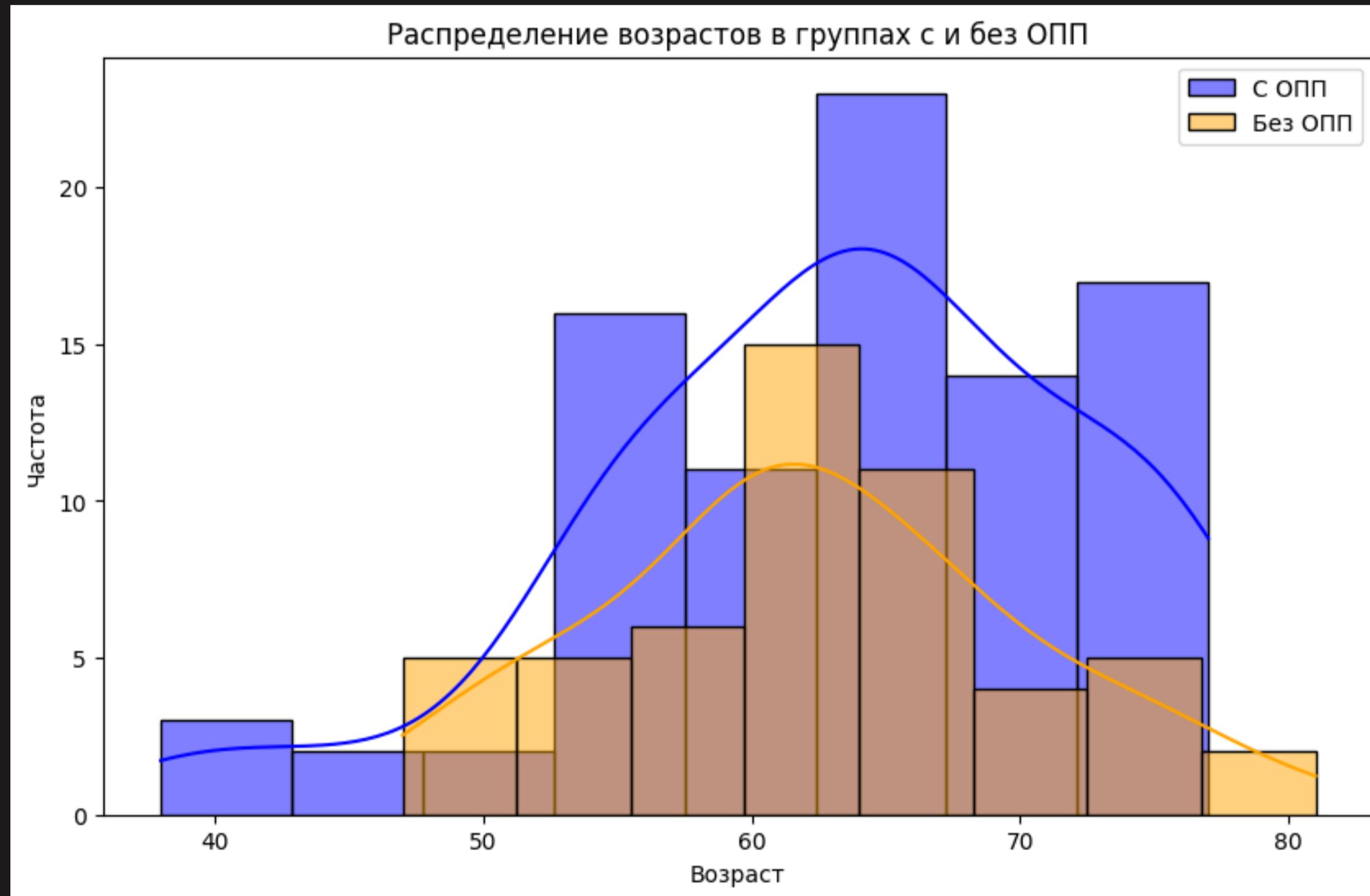
# structure



**Гипотеза:**  
**Возраст влияет**  
**на вероятность**  
**развития ОП**

## ГИПОТЕЗЫ

# Возраст влияет на вероятность развития ОПП



Mann-Whitney U-статистика: 2607.5

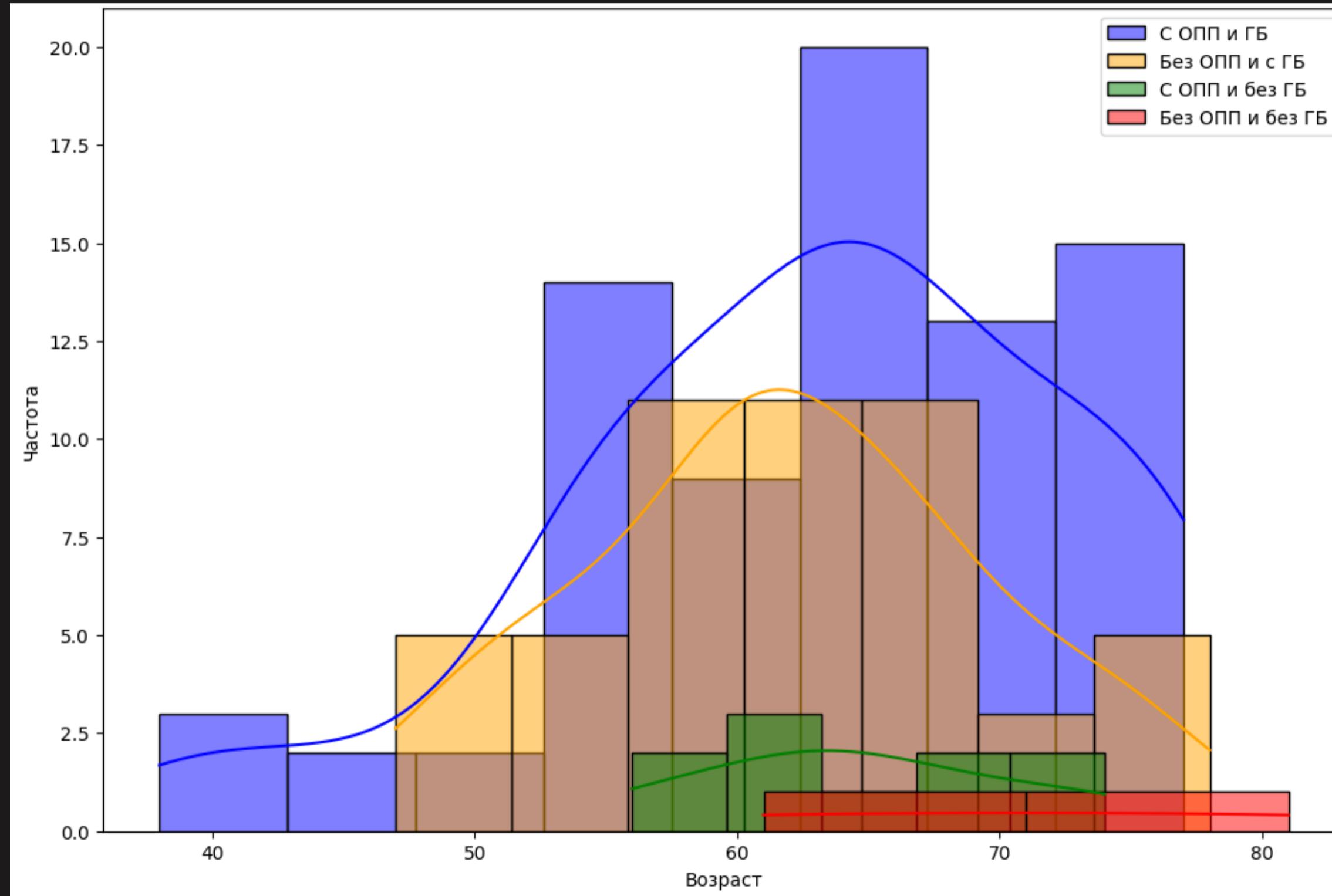
p-значение: 0.24

Нет статистически значимых различий в возрасте между группами с и без ОПП

**Гипотеза: Наличие  
гипертонической  
болезни (ГБ) связано  
с повышенной  
вероятностью  
развития ОПП**

## ГИПОТЕЗЫ

# Наличие гипертонической болезни (ГБ) связано с повышенной вероятностью развития ОПП



НО: Нет связи между ГБ и наличием ОПП.

Н1: У пациентов с ГБ выше вероятность наличия ОПП.

С ГБ:

U-статистика : 2235.0  
р-значение (с ГБ): 0.24

Без ГБ:

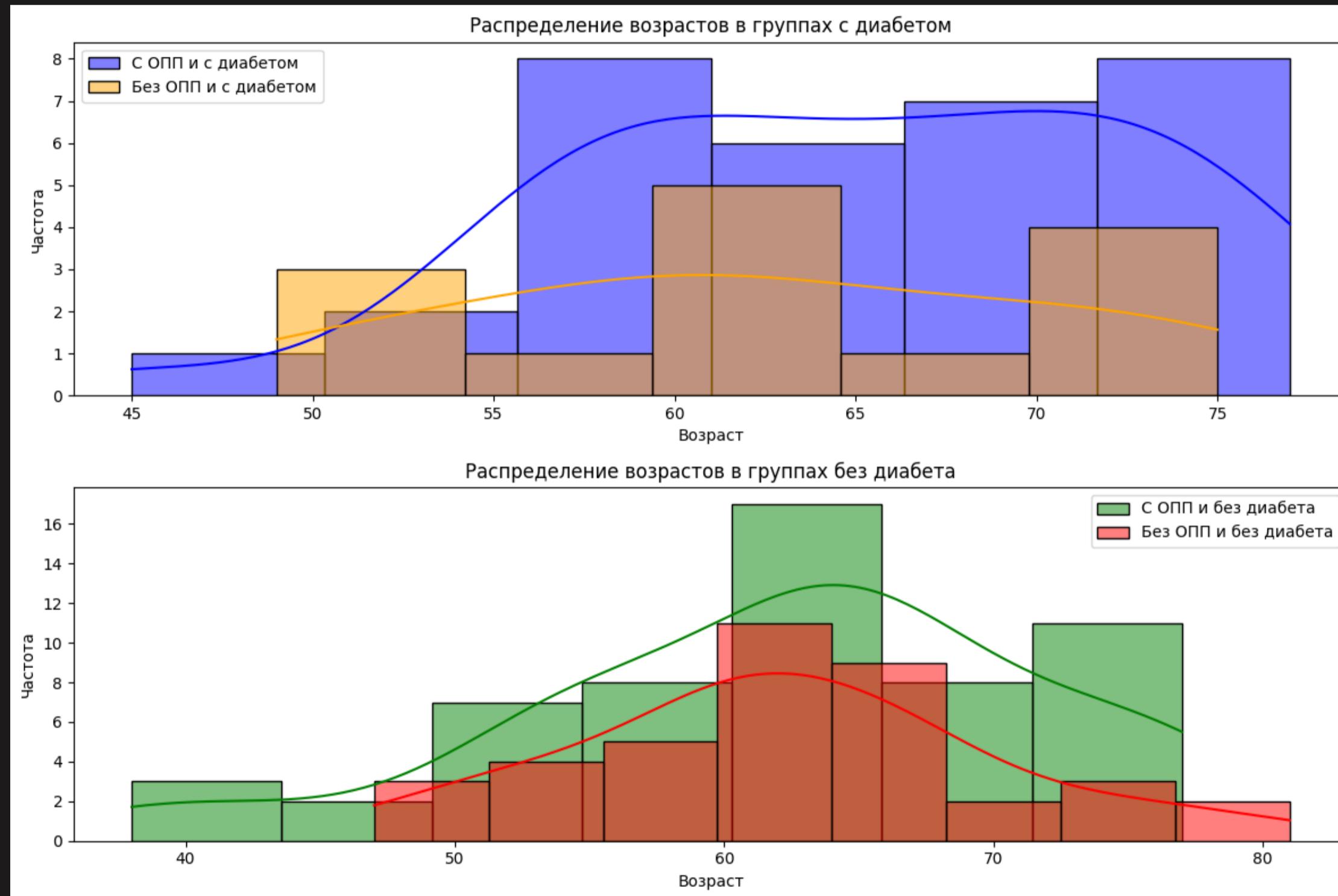
U-статистика : 8.0  
р-значение (без ГБ): 0.75

Нет статистически значимых различий ни в одной из групп

**Гипотеза: Факт  
развития ОПП зависит  
от наличия сахарного  
диабета**

## ГИПОТЕЗЫ

# Факт развития ОПП может зависеть от наличия сахарного диабета



С диабетом:

U-статистика: 265.5

p-значение: 0.32714948967213275

Без диабета:

U-статистика: 1166.5

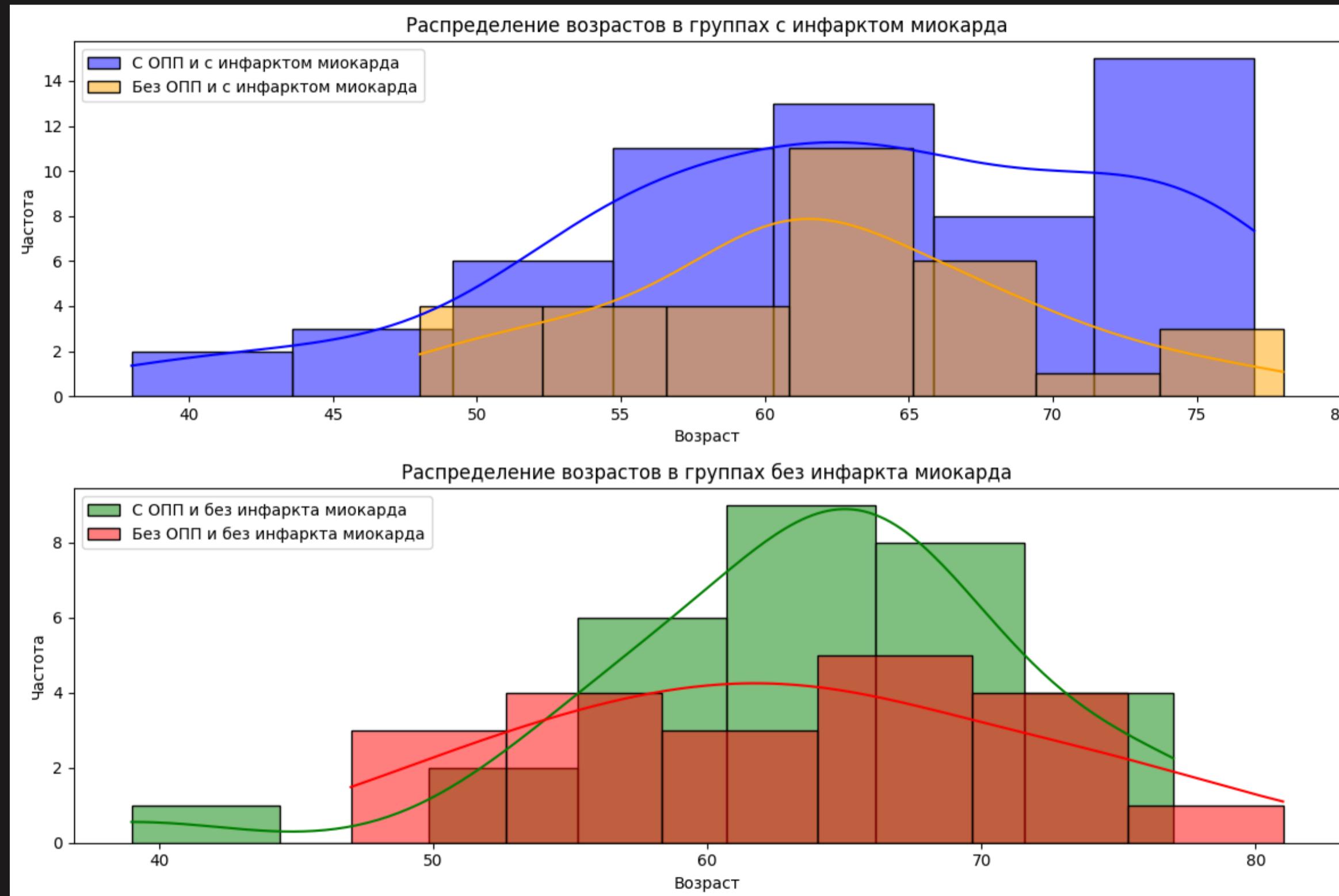
p-значение: 0.5752196590874028

Ни в одной из групп статистически значимых различий не обнаружено.

**Гипотеза: Наличие  
инфаркта  
миокарда  
увеличивает  
вероятность  
развития ОПП**

## ГИПОТЕЗЫ

# Наличие инфаркта миокарда увеличивает вероятность развития ОПП



С инфарктом миокарда

U-статистика: 1079.0

p-значение: 0.32

Без инфаркта миокарда:

U-статистика: 332.0

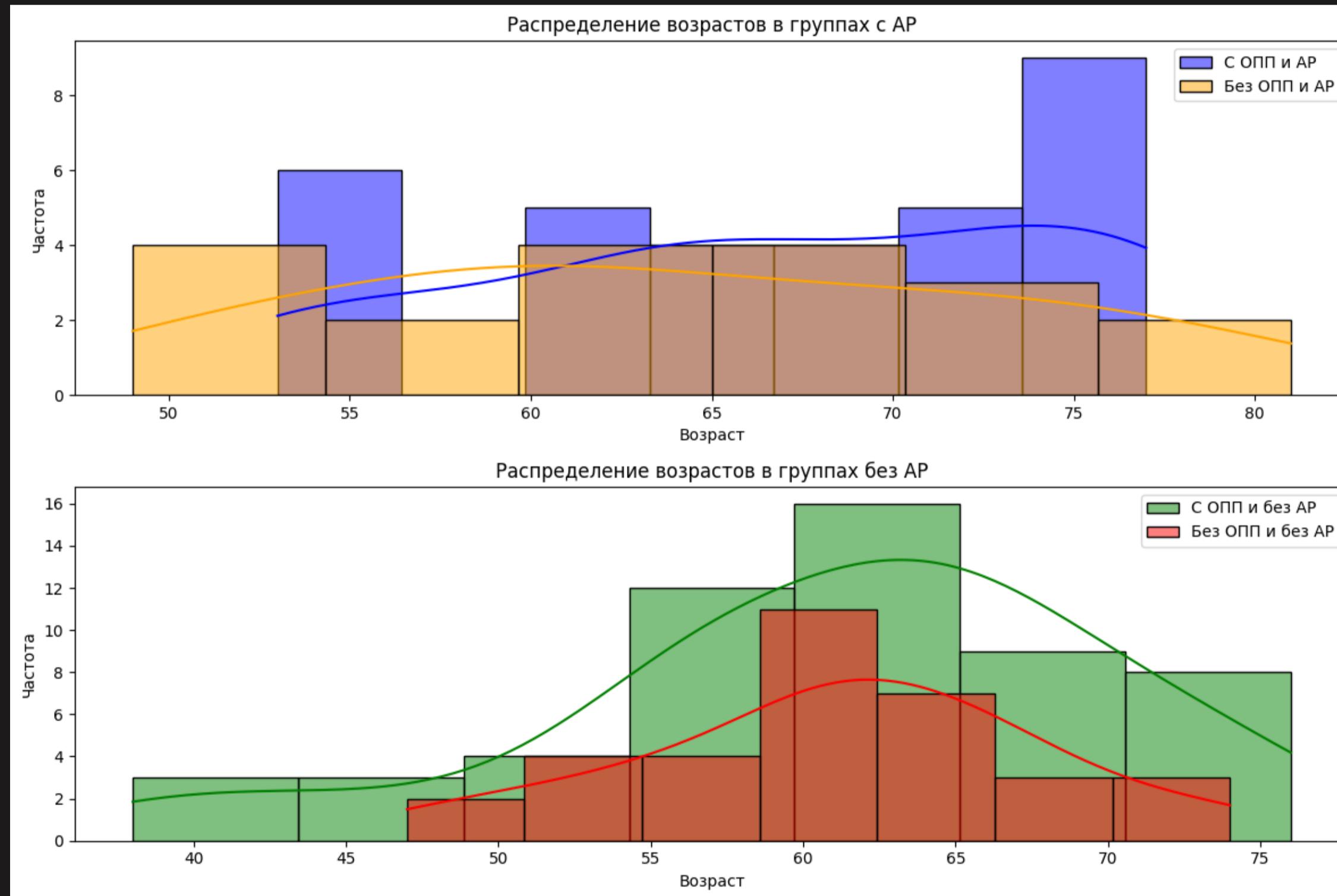
p-значение: 0.532211719139251

Нет статистически значимых различий ни в одной из групп

**Гипотеза: Повышенный  
уровень  
систолического  
давления (САД) связан  
с увеличением  
вероятности наличия  
ОПП**

## ГИПОТЕЗЫ

# Повышенный уровень систолического давления (САД) связан с увеличением вероятности наличия ОПП



САД выше нормы:

U-статистика: 616.0

p-значение: 0.06

САД в норме:

U-статистика: 677.5

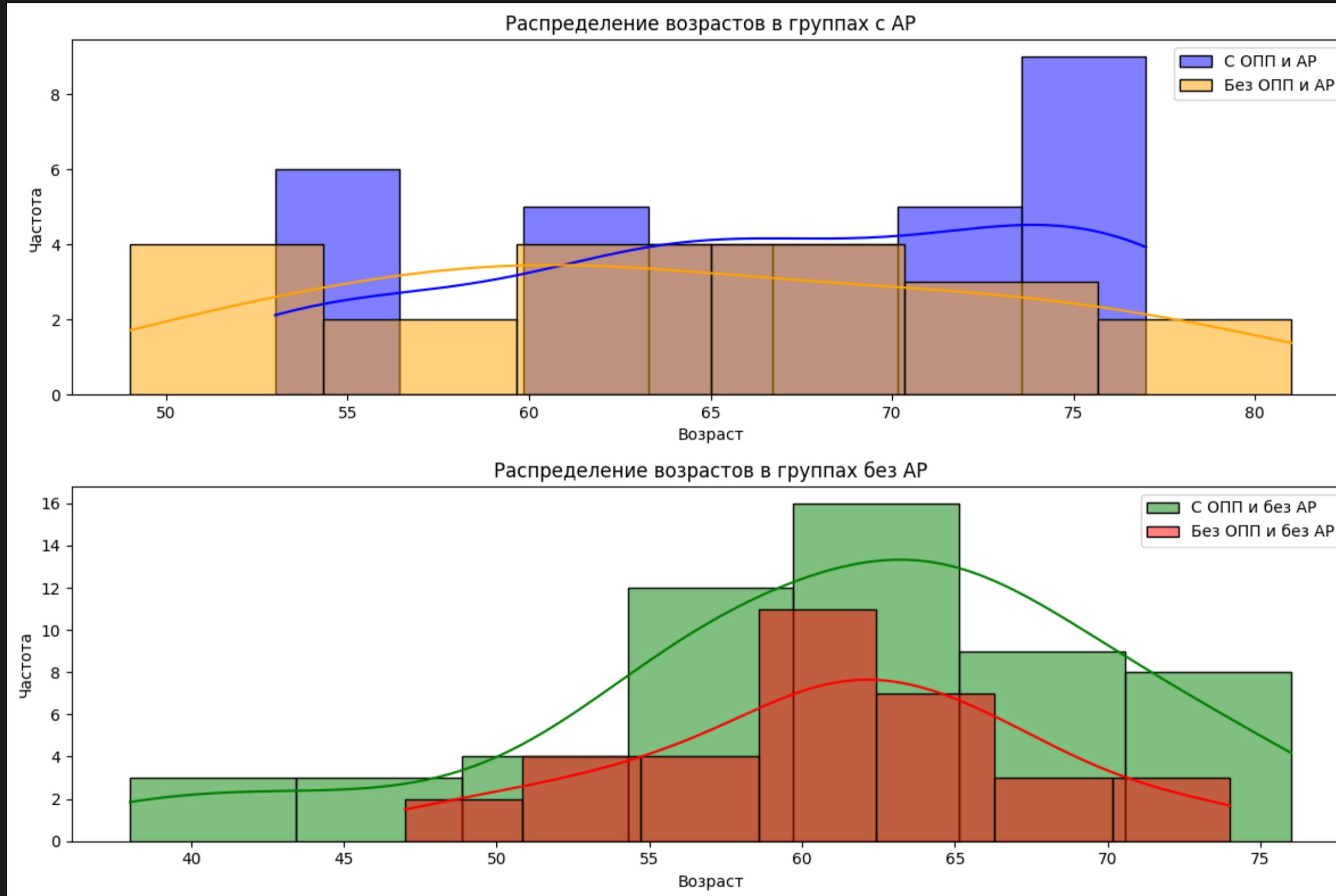
p-значение: 0.91

Нет статистически значимых различий ни в одной из групп

**Гипотеза: Наличие  
аллергического  
ринита связано с  
уменьшением  
вероятности  
развития ОПП**

## ГИПОТЕЗЫ

# Наличие аллергического ринита связано с уменьшением вероятности развития ОПП



Пациенты с АР:

U-статистика: 377.5

p-значение: 0.23

Пациенты с без АР:

U-статистика (без АР): 985.0

p-значение (без АР): 0.68

Нет статистически значимых различий ни в одной из групп.

OVERVIEW

# structure

ПРЕДОБРАБОТКА

ОБЩИЙ АНАЛИЗ

СЧИТАЕМ МЕТРИКИ

ГИПОТЕЗЫ

РЕШЕНИЕ ЗАДАЧИ ЛИНЕЙНОЙ РЕГРЕССИИ

we are here

## CREDITS

# meet the team

EGOR

**@ermakov\_copyright**

ANDREY

**@UnrealSavary**