



---

# ИНСТРУКЦИЯ ДЛЯ РАБОТЫ С СЕРВИСОМ

Январь, 2024

# СОДЕРЖАНИЕ

ГЛАВА 1. ПОДГОТОВКА БАЗЫ ДАННЫХ.....	4
ГЛАВА 2. НАЧАЛО РАБОТЫ В СТАТТЕСН. РЕГИСТРАЦИЯ, АВТОРИЗАЦИЯ, ПРИОБРЕТЕНИЕ СЛОТА, ЗАГРУЗКА БД .....	8
2.1. Регистрация и авторизация .....	8
2.2. Личный кабинет .....	10
2.2.1. Личная информация.....	12
2.2.2. Реферальная программа .....	17
2.2.3. Управление базами .....	19
2.2.4. Настройки аккаунта .....	21
2.2.5. Предложить идею и оставить отзыв.....	22
2.2.6. Выйти из аккаунта.....	23
2.3. Приобретение слота .....	25
2.4. Загрузка базы данных .....	27
2.5. Создание пустой базы данных .....	30
2.6. Удаление и восстановление базы данных .....	31
ГЛАВА 3. ВЫБОР ТИПА ПЕРЕМЕННЫХ .....	33
ГЛАВА 4. НАСТРОЙКА КОЛОНОК. СОЗДАНИЕ ГРУППЫ ПОКАЗАТЕЛЕЙ. ПРЕОБРАЗОВАНИЕ КОЛОНКИ.....	37
4.1. Настройка колонок .....	37
4.2 Изменение названия колонки и категорий для графиков .....	41
4.3 Создание группы показателей .....	43
4.3. Изменение данных .....	48
4.3.1 Функции «Вставить...», «Сортировка...», «Очистить все значения», «Удалить», добавить «N строк» / «N столбцов».....	48
4.3.2. Функция «Преобразовать колонку».....	52
4.3.3. Функция «Отображение колонок».....	61
ГЛАВА 5. НАСТРОЙКИ АНАЛИЗА.....	62
ГЛАВА 6. ОПИСАТЕЛЬНАЯ СТАТИСТИКА .....	66
ГЛАВА 7. ПОСТРОЕНИЕ ЗАВИСИМОСТЕЙ МЕЖДУ ПОКАЗАТЕЛЯМИ .....	70
7.1. Сравнение двух групп по количественному признаку .....	70
7.2. Общая структура результатов анализа .....	74
7.3. Сравнение двух групп по категориальному признаку .....	76
7.4. Сравнение трех и более групп по количественному признаку .....	79
7.5. Сравнение трех и более групп по категориальному признаку.....	81
7.6. Корреляционный анализ .....	83
7.7. Рос – Анализ.....	85
7.8. Анализ данных в формате «До - После» .....	90

7.8.1. Сравнение показателей в формате «До – После» без деления на группы.....	90
7.8.2. Сравнение показателей в формате «До – После» с делением на группы .....	94
7.8.3. Сравнение трех и более показателей в формате «До – После».....	99
<b>ГЛАВА 8. ПОСТРОЕНИЕ МОДЕЛЕЙ.....</b>	<b>101</b>
8.1. Линейная регрессия.....	101
8.2. Логистическая регрессия.....	108
8.3. Анализ выживаемости .....	114
8.4. Динамические ряды.....	124
8.5. Вычисление переменных .....	128
<b>ГЛАВА 9. СОЗДАНИЕ ОПРОСА.....</b>	<b>132</b>
<b>ГЛАВА 10. КАЛЬКУЛЯТОР.....</b>	<b>140</b>

## ГЛАВА 1. ПОДГОТОВКА БАЗЫ ДАННЫХ

Для того чтобы выполнить статистическую обработку данных с помощью нашего онлайн-сервиса, Вам нужно сделать всего 3 шага:



Оформление базы данных для сервиса StatTech не отличается от принципов подготовки базы для любой другой статистической программы и предполагает выполнение нескольких несложных общепринятых правил:

1) Данные в таблице базы (Рис. 1.1) располагаются следующим образом:

- Строки – это пациенты (клиенты, животные, другие единицы наблюдения). Данные о каждом пациенте уместаются в одну строку.
- Столбцы – это показатели (возраст, пол, артериальное давление и т.д.)
- Верхняя строчка базы данных – названия показателей без указания единиц измерения или других дополнительных сведений. Для названия показателей используется только первая строка в базе данных.

	A	B	C	D
1	Номер	Пол	Профессия	Возраст
2	1	0	2	24
3	2	0	2	34
4	3	1	1	35
5	4	1	2	25
6	5	1	3	19

Рисунок 1.1 – Пример таблицы базы в MS Excel

Нельзя использовать слияние каких-либо ячеек в базе (Рис. 1.2):

	A	B	C	D
1	Номер	Возраст	Масса: до-после	
2	1	24	60	59
3	2	34	66	64
4	3	35	70	73
5	4	25	85	82
6	5	19	80	78

Рисунок 1.2 – Пример неправильного оформления таблицы базы - слияние ячеек

Нельзя использовать более одной строки для названия показателей (Рис. 1.3):

	A	B	C	D	E	F
1			Масса		Показатели крови	
2	Номер	Возраст	До диеты	После диеты	Глюкоза	Холестерин
3	1	24	60	59	5,2	3,22
4	2	34	66	64	5,6	4,31
5	3	35	70	73	6,2	3,76
6	4	25	85	82	3,5	5,01
7	5	19	80	78	4,7	3,42

Рисунок 1.3 – Пример неправильного оформления таблицы базы

2) Все показатели должны быть числовыми. Не допускается использование текстовых переменных. Категориальные переменные кодируются в соответствии с их значениями, каждому из которых присваивается свое число. Например, в столбце «Пол» следует указывать не «мужчина» или «женщина», а числовые обозначения: 1 и 2, соответственно.

Для показателей, значения которых подразумевают ранжирование, числовые коды присваивайте в порядке возрастания. Например, степени тяжести заболевания: легкая – 1, средняя – 2, тяжелая – 3.

Не забудьте отдельно записать и сохранить для себя ключ к используемым кодировкам, он пригодится при загрузке базы данных на портал StatTech!

**Пример исходной базы с текстовыми обозначениями категориальных переменных (Рис. 1.4):**

	A	B	C	D
1	Пол	Профессия	Наличие АГ	Степень АГ
2	женский	служащий	наличие	третья
3	женский	служащий	отсутствует	АГ отсутствует
4	мужской	рабочий	наличие	вторая
5	мужской	служащий	наличие	первая
6	мужской	руководитель	отсутствует	АГ отсутствует

Рисунок 1.4 – Пример базы с текстовыми обозначениями категориальных переменных

**Та же база данных после перекодировки категориальных переменных (Рис. 1.5):**

	A	B	C	D
1	Пол	Профессия	Наличие АГ	Степень АГ
2	0	2	1	3
3	0	2	0	0
4	1	1	1	2
5	1	2	1	1
6	1	3	0	0

Рисунок 1.5 – Пример базы с перекодировкой текстовых категориальных переменных

**Ключ к базе:**

Пол: 0 – женский, 1 - мужской

Профессия: 1 – рабочий, 2 – служащий, 3 – руководитель

Наличие АГ: 0 – отсутствует, 1 – наличие

Степень АГ: 0 – АГ отсутствует, 1 – первая, 2 – вторая, 3 – третья

3) Количественные показатели указываются в числовом формате без указания единиц наблюдения. В качестве разделителя разрядов используйте либо запятую, либо точку (Рис. 1.6).

Правильное обозначение					Неправильное обозначение				
	A	B	C	D		A	B	C	D
1	Возраст	ИМТ	Глюкоза	АД	1	Возраст	ИМТ	Глюкоза	АД
2	24	19,59	4,1	114	2	24 лет	19,59	4,1 ммоль/л	114
3	34	21,95	5	118	3	34 лет	21.95	5 ммоль/л	118
4	35	23,41	4,9	120	4	35 лет	23.41	4,9 ммоль/л	120
5	25	26,23	5,4	121	5	25 лет	26,23	5,4 ммоль/л	121
6	19	26,23	4,3	125	6	19 лет	26,23	4,3 ммоль/л	125

В качестве разделителя использована запятая

В ячейках базы – только цифры

Используются разные разделители – и точка, и запятая

Значения указаны вместе с единицами измерения

Рисунок 1.6 – Пример оформления количественных переменных

4) Некоторые категориальные показатели принимают у одной и той же единицы наблюдения сразу несколько значений. Например, это относится к сопутствующим заболеваниям, которых у одного пациента может быть несколько. В таких случаях недопустимо указывать в одной ячейке несколько значений через запятую. Необходимо создать несколько столбцов с отдельными заболеваниями, в которые данные заносим в бинарном, двоичном виде: 1 (что означает «Есть данное заболевание») или 0 («Нет данного заболевания»).

5) Присваивайте названия показателям в наиболее полном варианте, так, как Вам бы хотелось видеть их в таблицах и графиках. В неизменном виде они будут использоваться при оформлении результатов анализа Ваших данных. **Например, вместо «Глюкоза», лучше написать «Содержание глюкозы в сыворотке крови».**

## ГЛАВА 2. НАЧАЛО РАБОТЫ В STATTECH. РЕГИСТРАЦИЯ, АВТОРИЗАЦИЯ, ПРИОБРЕТЕНИЕ СЛОТА, ЗАГРУЗКА БД

### 2.1. Регистрация и авторизация

Если пользователь не зарегистрирован в программе, то необходимо нажать соответствующую кнопку и пройти регистрацию (Рис. 2.1).

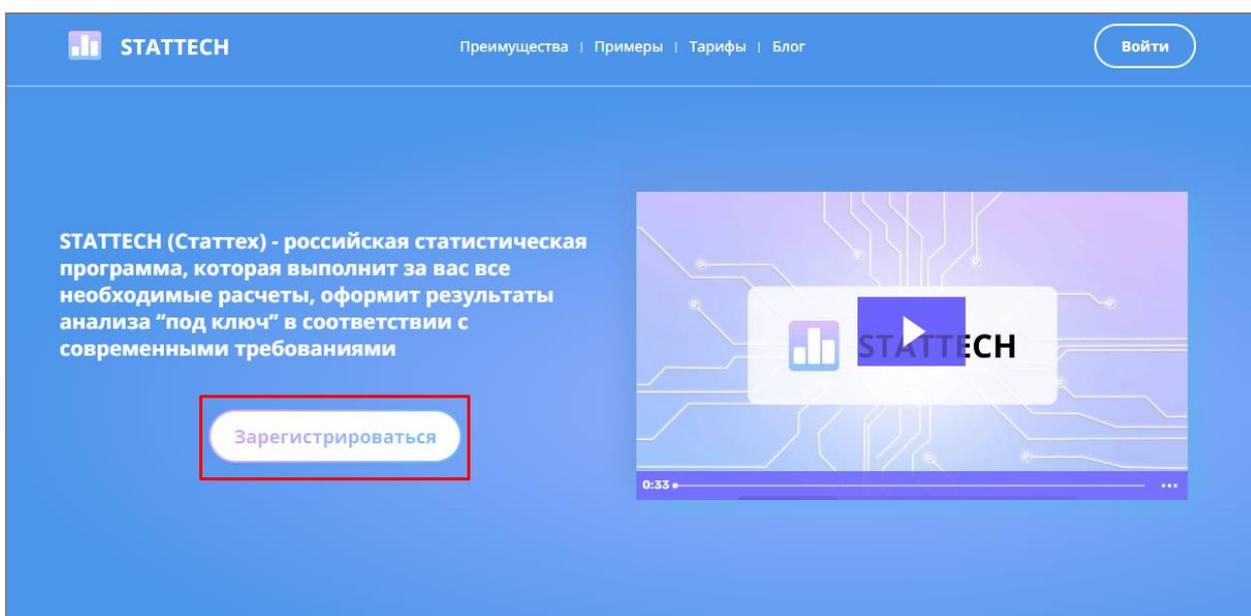


Рисунок 2.1 – Кнопка «Зарегистрироваться» на лэндинге (веб-странице) сайта

Открывается форма, где необходимо заполнить email, имя пользователя, придумать пароль и ввести его повторно (Рис. 2.2). После чего нажать на кнопку «Зарегистрироваться». Далее на введенную почту придет письмо, на ссылку которой надо будет перейти для завершения регистрации.

The image shows a registration form with the following elements:

- Input field: "Адрес эл. почты" (Email address)
- Input field: "Имя пользователя" (Username)
- Input field: "Пароль" (Password) with a visibility toggle icon.
- Input field: "Повтор пароля" (Repeat password) with a visibility toggle icon.
- Text: "Продолжая регистрацию, Вы подтверждаете свое согласие с"
- Form element: A checkbox followed by the text "Пользовательским соглашением, Политикой конфиденциальности и Правилами пользования сайтом".
- Button: "Зарегистрироваться" (Register).

Рисунок 2.2 – Форма регистрации нового пользователя

При последующих входах нужно будет проходить только авторизацию. В данной форме вводится email и пароль или же можно проходить авторизацию через аккаунт Google (Рис. 2.3).

The image shows an authorization form on the STATTECH website. The form includes:

- Header: "STATTECH" logo and navigation links: "Войти" (Login), "Зарегистрироваться" (Register), and "Меню" (Menu).
- Input field: "Email".
- Input field: "Пароль" (Password).
- Buttons: "Войти" (Login) and "Забыли пароль?" (Forgot password?).
- Option: "Sign in with Google" button.

Рисунок 2.3 – Форма авторизации входа в личный кабинет

## 2.2. Личный кабинет

После регистрации и авторизации на сайте переходим в «Личный кабинет» (Рис. 2.4).

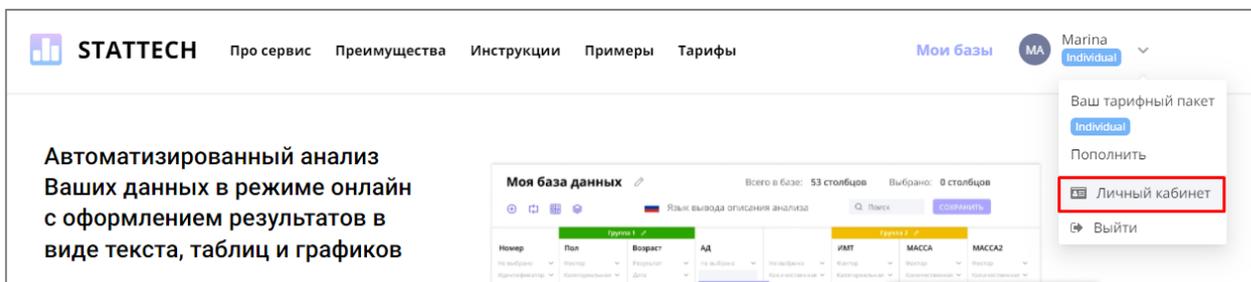


Рисунок 2.4 - Кнопка «Личный кабинет» на лэндинге (веб-странице) сайта

Личный кабинет состоит из разделов (Рис. 2.5):

- «Личная информация» – в данном разделе хранится персональная информация
- «Реферальная программа» – условия и ссылки по партнерской программе
- «Управление базами» – отображение всех загруженных баз данных и функция «Добавить слот»
- «Настройки аккаунта» – содержит функционал по смене пароля от аккаунта
- «Предложить идею» - возможность написать свою идею по улучшению и модернизации работы сайта
- «Оставить отзыв» – возможность написать свой отзыв о работе сайта
- Возможность «Выйти из аккаунта»

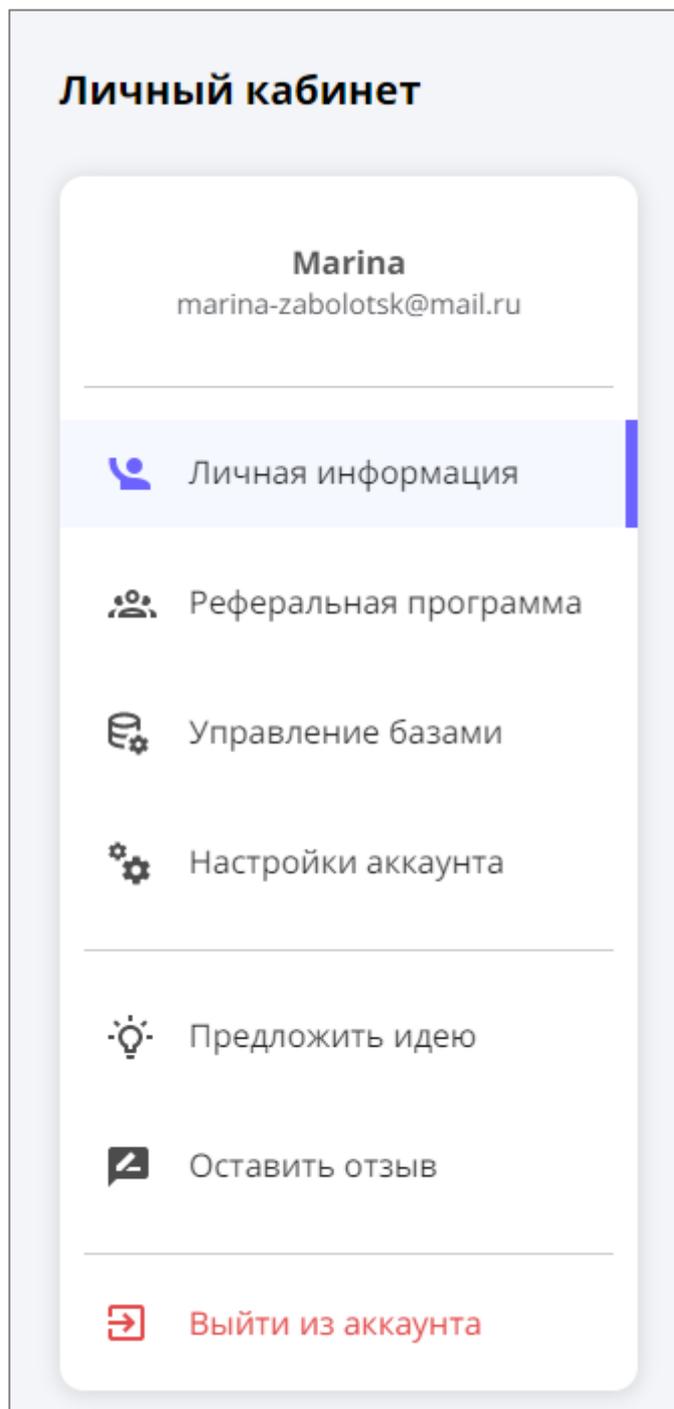
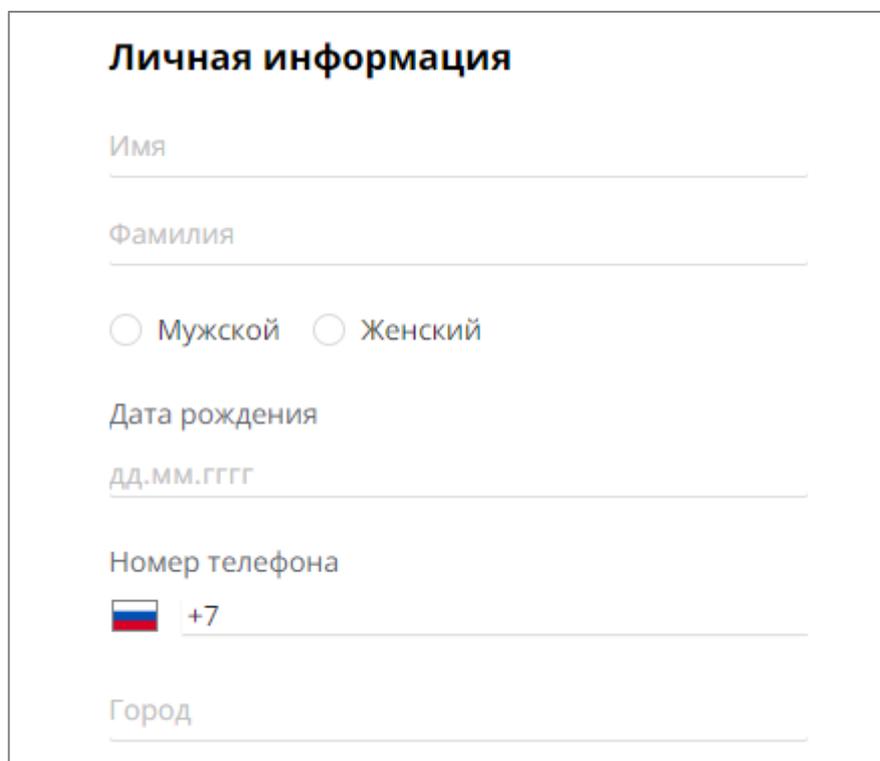


Рисунок 2.5 – Разделы личного кабинета

### 2.2.1. Личная информация

Данный раздел личного кабинета хранит в себе личную информацию аккаунта (Рис. 2.6).



The image shows a web form titled "Личная информация" (Personal Information). It contains several input fields: "Имя" (Name), "Фамилия" (Surname), a gender selection with radio buttons for "Мужской" (Male) and "Женский" (Female), "Дата рождения" (Date of Birth) with a placeholder "дд.мм.гггг", "Номер телефона" (Phone Number) with a Russian flag icon and "+7" prefix, and "Город" (City).

Рисунок 2.6 – Форма ввода персональных данных в личном кабинете

Заполнение данной формы подразумевает под собой заполнение таких полей, как «Имя», «Фамилия», пол, «Дата рождения», «Номер телефона», «Город».

Также в данном разделе есть возможность заполнить данные о Вашей профессиональной и научной деятельности (Рис. 2.7).

The image shows a web form with four dropdown menus and two buttons. The first dropdown is labeled 'Сфера деятельности' and contains the text 'Выберите...'. The second dropdown is labeled 'Род научной деятельности' and contains 'Не выбрано'. The third dropdown is labeled 'Наличие ученой степени' and contains 'Не выбрано'. The fourth dropdown is labeled 'Наличие ученого звания' and contains 'Не выбрано'. At the bottom, there are two buttons: 'ОТМЕНИТЬ' (cancel) and 'СОХРАНИТЬ' (save).

Рисунок 2.7 – Форма ввода данных о профессиональной и научной деятельности в личном кабинете

- «Сфера деятельности» - здесь Вы выбираете сферу своей работы, например, если Вы работаете врачом - терапевтом, то выбираете «Медицина, ветеринария» (Рис. 2.8). Нажав на поле «Сфера деятельности» откроется выпадающий список, с возможностью выбора оптимального заполнения данной строки.

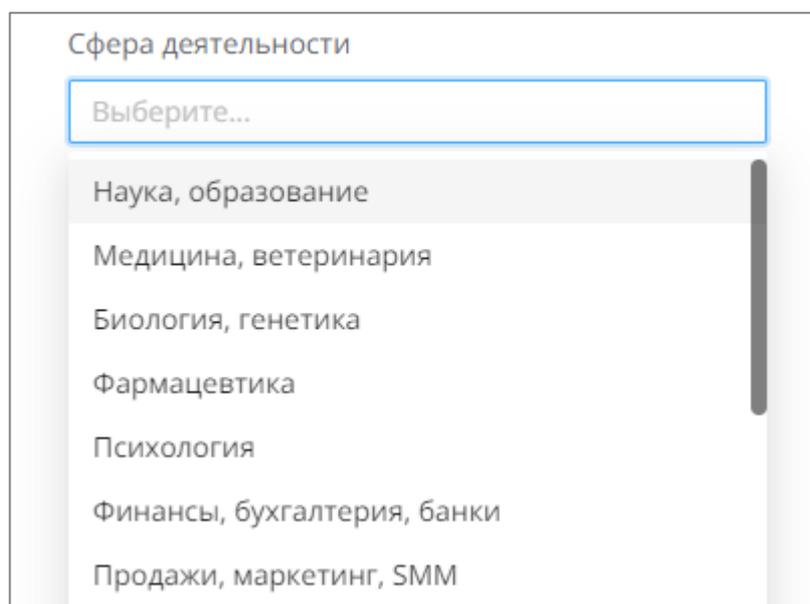


Рисунок 2.8 – Выпадающий список вариантов заполнения строки «Сфера деятельности»

- «Род научной деятельности» подразумевает под собой выбор уровня занятий научной деятельности. Данная строка также имеет варианты ответов по типу выпадающего списка, путем нажатия на пиктограмму  в конце строки (Рис. 2.9).

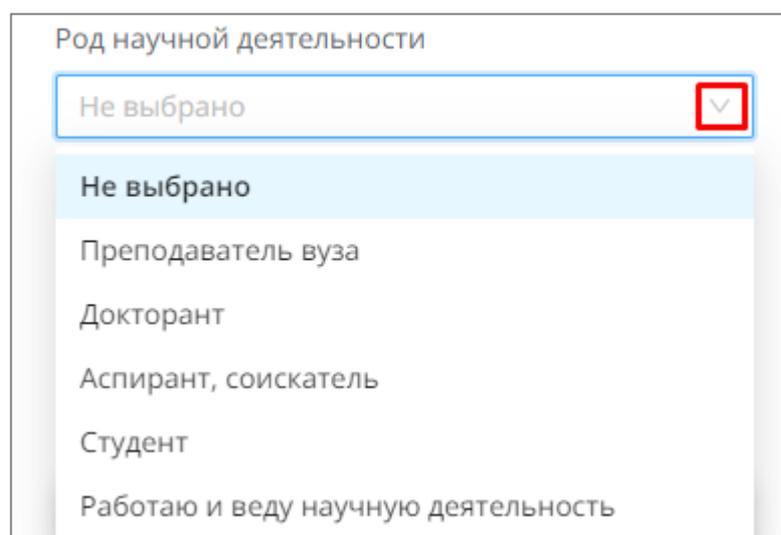


Рисунок 2.9 – Выпадающий список вариантов заполнения строки «Род научной деятельности»

- «Наличие ученой степени» подразумевает под собой варианты заполнения данной строки сведениями о наличии или отсутствие научной степени. Нажав на пиктограмму  в конце строки, можно увидеть варианты для выбора ответа (Рис. 2.10).

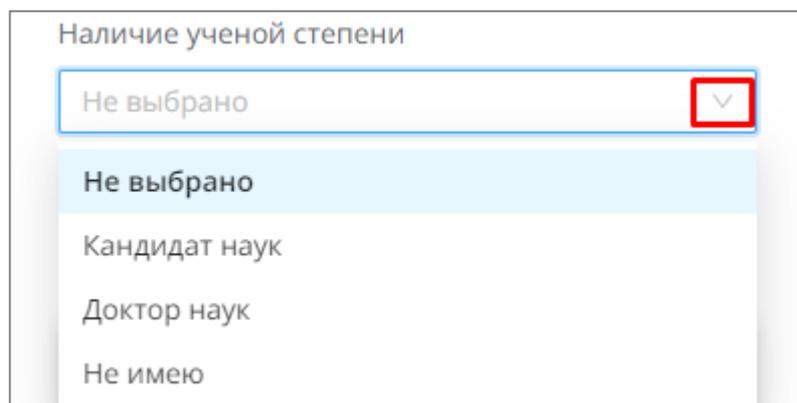


Рисунок 2.10 - Выпадающий список вариантов заполнения строки «Наличие ученой степени»

- «Наличие ученого звания» предполагает заполнение данной строки либо значением «не имею», либо выбором ученого звания. Как и в предыдущих строках, функционал заполнения остается прежним: знак  в конце строки (Рис. 2.11).

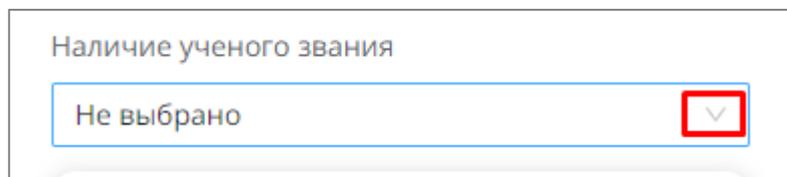
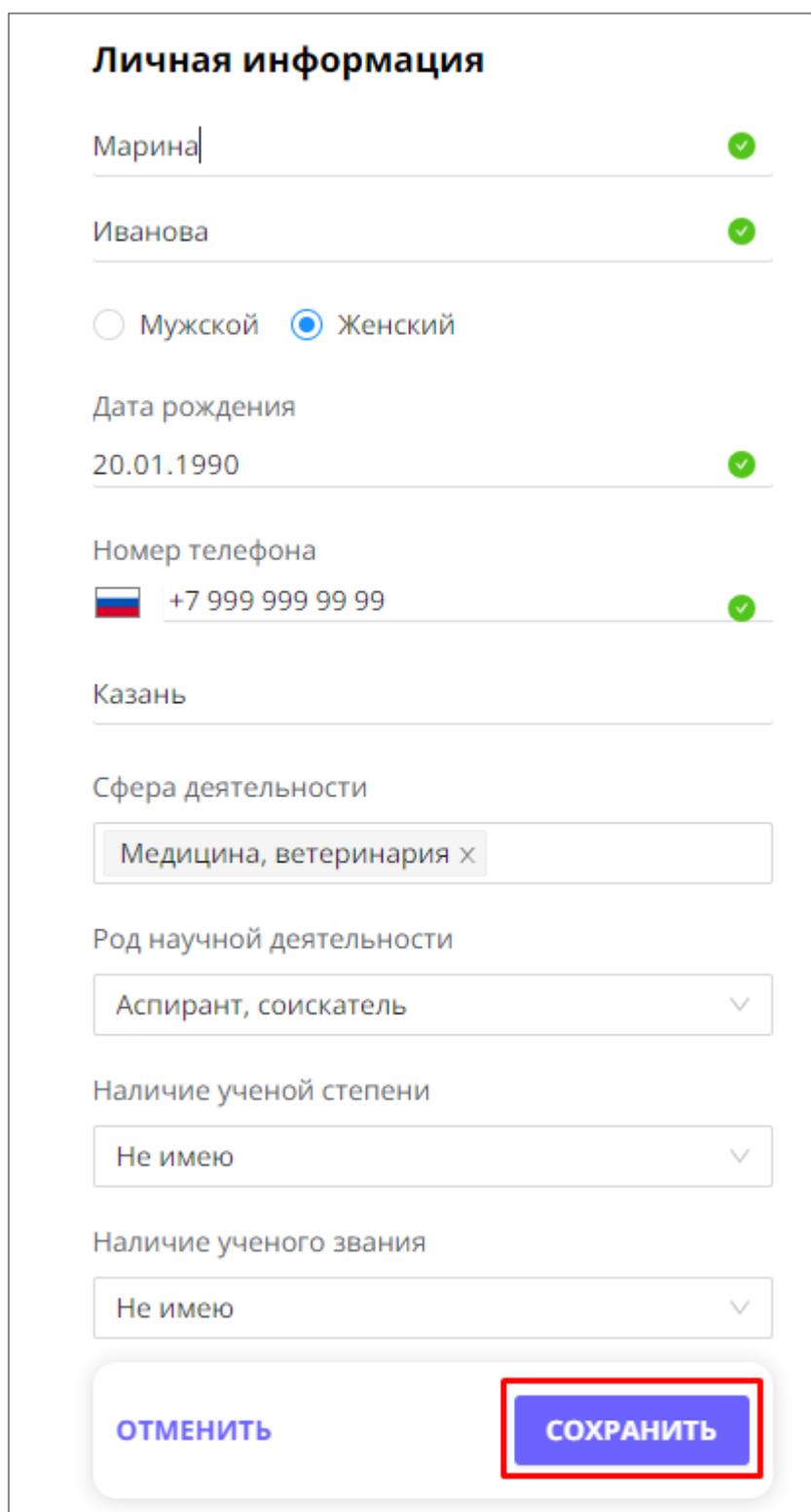


Рисунок 2.11 – Кнопка выпадающего списка в строке «Наличие ученого звания»

После заполнения полей с персональными данными и строк о профессиональной и научной деятельности необходимо нажать на кнопку «Сохранить» (Рис. 2.12).



**Личная информация**

Марина ✓

Иванова ✓

Мужской  Женский

Дата рождения  
20.01.1990 ✓

Номер телефона  
+7 999 999 99 99 ✓

Казань

Сфера деятельности  
Медицина, ветеринария x

Род научной деятельности  
Аспирант, соискатель ▾

Наличие ученой степени  
Не имею ▾

Наличие ученого звания  
Не имею ▾

**ОТМЕНИТЬ** **СОХРАНИТЬ**

Рисунок 2.10 – Кнопка «Сохранить» в разделе «Личная информация» и пример заполнения формы

## 2.2.2. Реферальная программа

Данный раздел личного кабинета отображает что такое реферальная программа (Рис. 2.11), как ей можно воспользоваться и сколько баллов Вы за это получите.

### Что такое реферальная программа?



Рисунок 2.11 – Что такое реферальная программа

При оплате подписки новым пользователем (ранее никогда не совершавшим покупку в своём аккаунте) с указанием в поле «Промокод» вашего реферального кода он получит скидку 10% на всю сумму покупки. В течение 24 часов после этого вы получите 10% от суммы покупки (с учётом скидки) на свой счёт в виде баллов. Накопленные баллы можно использовать для генерации скидочного купона на равнозначную сумму (1 балл = 1 рубль), который можно применить для оплаты своей подписки либо отдать его другим пользователям. (Рис. 2.12).

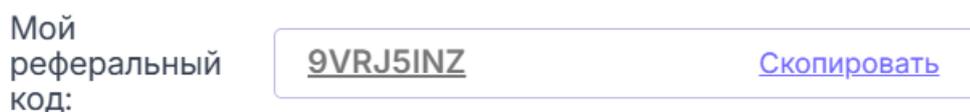


Рисунок 2.12 – Персональный реферальный код

Так же в данном разделе личного кабинета есть возможность просмотра количества накопленных баллов и возможность «Сгенерировать скидочный купон» (рис. 2.13). В дальнейшем скидочный купон можно использовать при покупке или продление тарифного пакета.

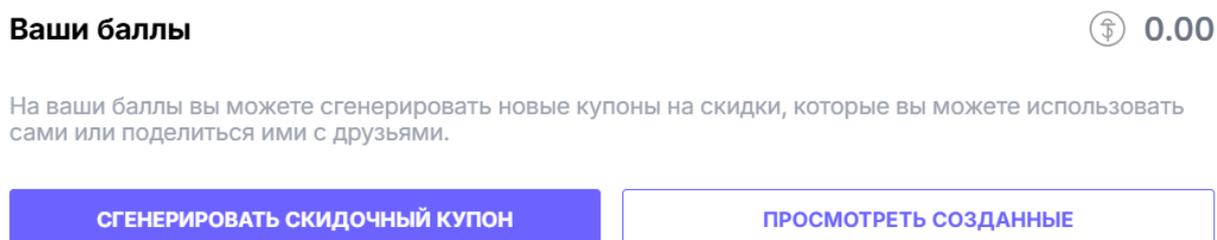


Рисунок 2.13 – Кнопка «Сгенерировать скидочный купон» в разделе «Реферальная программа» личного кабинета

Еще одна возможность в данном разделе – просмотр сколько человек воспользовались Вашей реферальной ссылкой и сколько из них оформило подписку (рис. 2.14).

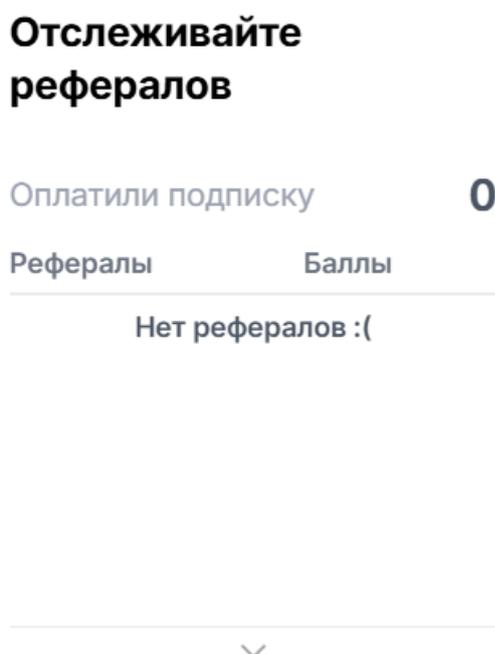


Рисунок 2.14 – Кнопка «Просмотреть» для отслеживания рефералов

### 2.2.3. Управление базами

Раздел «Управление базами» отображает количество купленных слотов и количество загруженных баз. Так же в данном разделе есть возможность «Добавить слот» (Рис. 2.15).

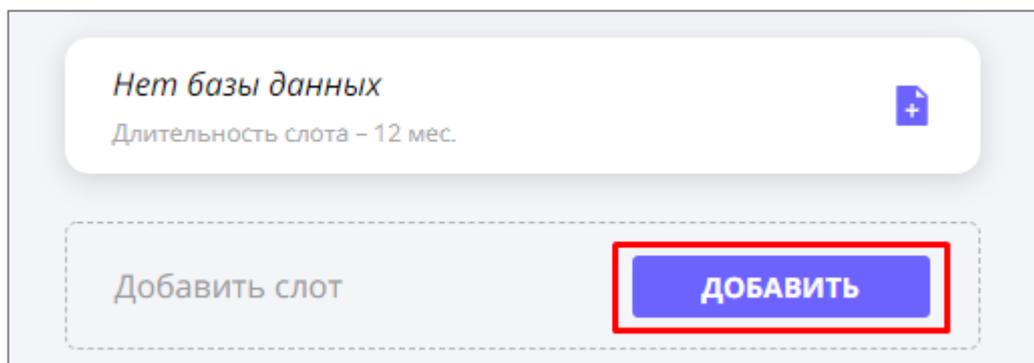


Рисунок 2.15 – Кнопка «Добавить слот» в разделе «Управление базами» в личном кабинете

После нажатия на кнопку «Добавить» сайт направит Вас для выбора тарифного пакета и его оплаты (Рис. 2.16).

Тарифы			
	ДЕМО бесплатно	ПОПУЛЯРНЫЙ ИНДИВИДУАЛЬНЫЙ от 8 000,00 Р	КОРПОРАТИВНЫЙ от 50 000,00 Р
Описание			
Размер базы	30 строк	∞	∞
Техподдержка	—	✓	✓
Доступ к обучающим материалам	—	✓	✓
Минимальный срок доступа к базе данных	1 месяц	12 месяцев	12 месяцев
		<input type="button" value="ВЫБРАТЬ"/>	<input type="button" value="ВЫБРАТЬ"/>

Рисунок 2.16 – Выбор подходящего тарифа

После нажатия кнопки «Выбрать» откроется форма ввода данных, которая предусматривает выбор количества баз и количества месяцев использования (рис. 2.18). После заполнения этих данных необходимо нажать кнопку «Оплатить» и сайт перенаправит Вас на оплату.

**Индивидуальный** ×

Выберите количество баз и срок их действия

Количество баз

Длительность (мес.)

Промокод

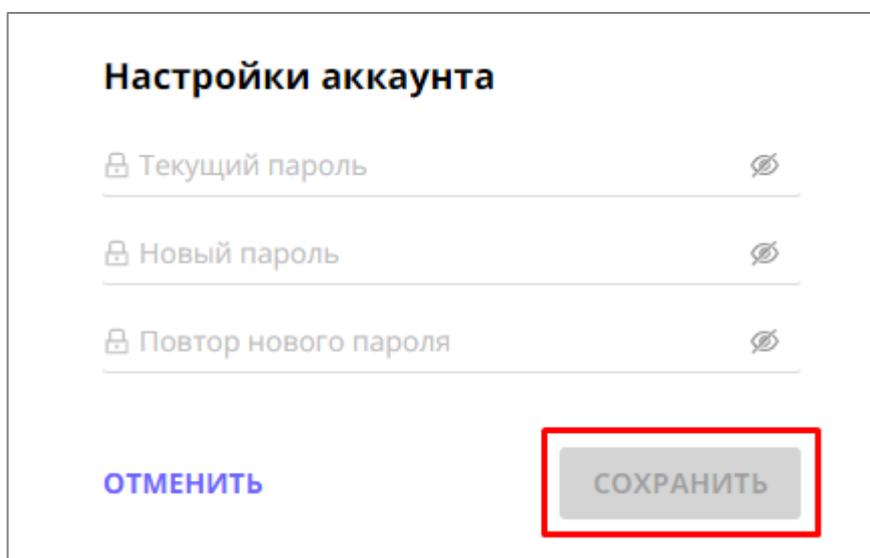
Стоимость 8 000,00 Р

Рисунок 2.18 – Интерфейс настройки тарифного пакета

## 2.2.4. Настройки аккаунта

В разделе «Настройки аккаунта» расположена функция смены пароля от личного кабинета.

Для смены пароля Вам необходимо сначала в строке «Текущий пароль» ввести уже имеющийся пароль от личного кабинета. В строке «Новый пароль» ввести придуманный Вами пароль, а в строке «Повтор нового пароля» продублировать то, что Вы ввели в строке «Новый пароль». Затем нажать на кнопку «Сохранить» для сохранения изменений (Рис. 2.19).



The image shows a web form titled "Настройки аккаунта" (Account Settings). It contains three input fields for passwords, each with a lock icon on the left and an eye icon on the right to toggle visibility. The fields are labeled "Текущий пароль" (Current password), "Новый пароль" (New password), and "Повтор нового пароля" (Repeat new password). At the bottom left is a blue button labeled "ОТМЕНИТЬ" (Cancel), and at the bottom right is a grey button labeled "СОХРАНИТЬ" (Save), which is highlighted with a red rectangular border.

Рисунок 2.19 – Кнопка «Сохранить» в разделе «Настройки аккаунта» в личном кабинете

### 2.2.5. Предложить идею и оставить отзыв

Разделы «Предложить идею» и «Оставить отзыв» в личном кабинете предусматривают под собой переход на почту, чтобы оставить отзыв, идею или пожелание в чат технической поддержки сайта StatTech.

На ОС Windows перенаправление идет на Microsoft Outlook, где Вы можете добавить свой электронный ящик, если он еще не был добавлен и написать нам (Рис. 2.20).

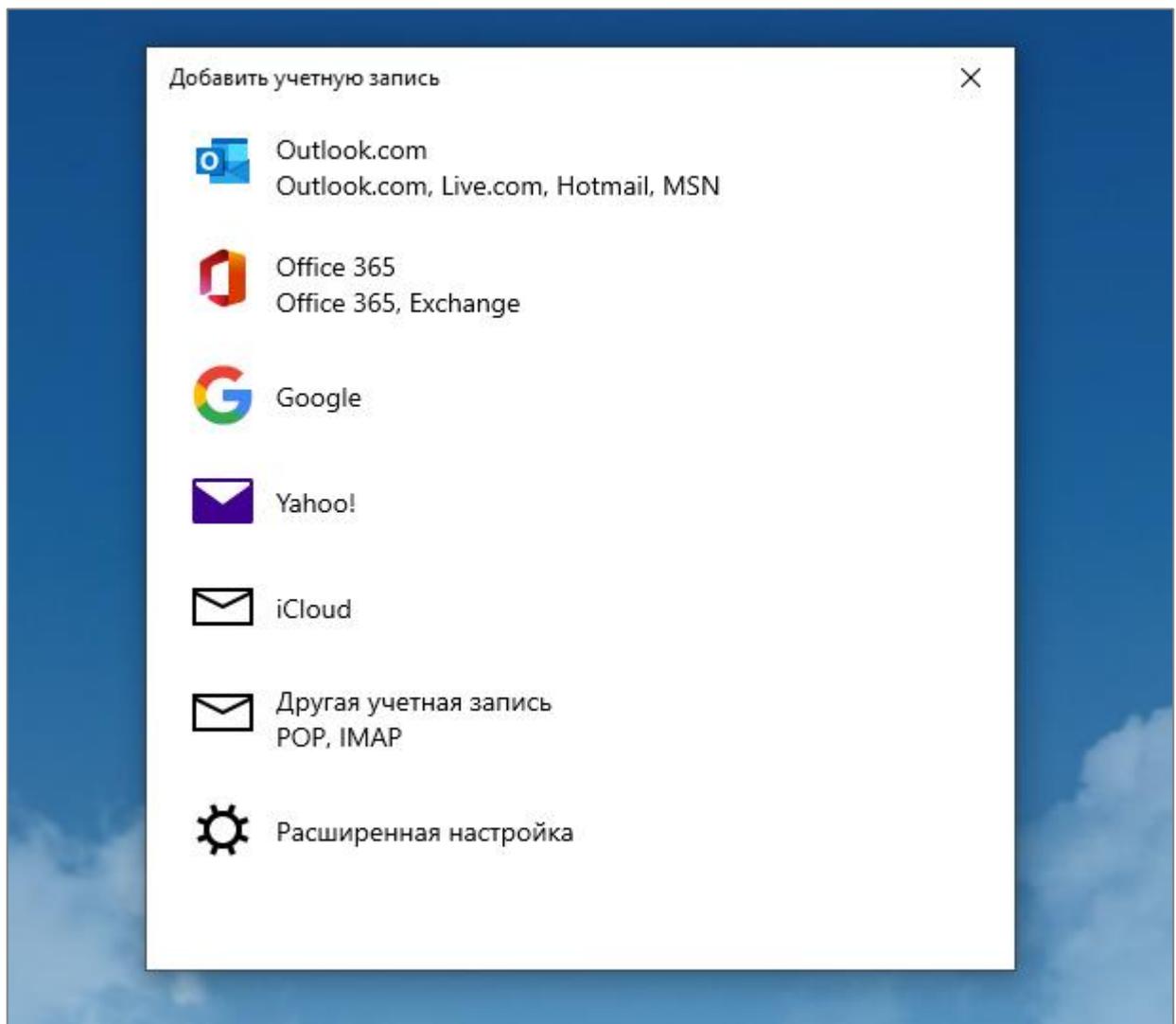


Рисунок 2.20 – Окно перенаправления на почту через приложение Microsoft Outlook

## 2.2.6. Выйти из аккаунта

Так же в личном кабинете есть функция «Выйти из аккаунта» (Рис. 2.21).

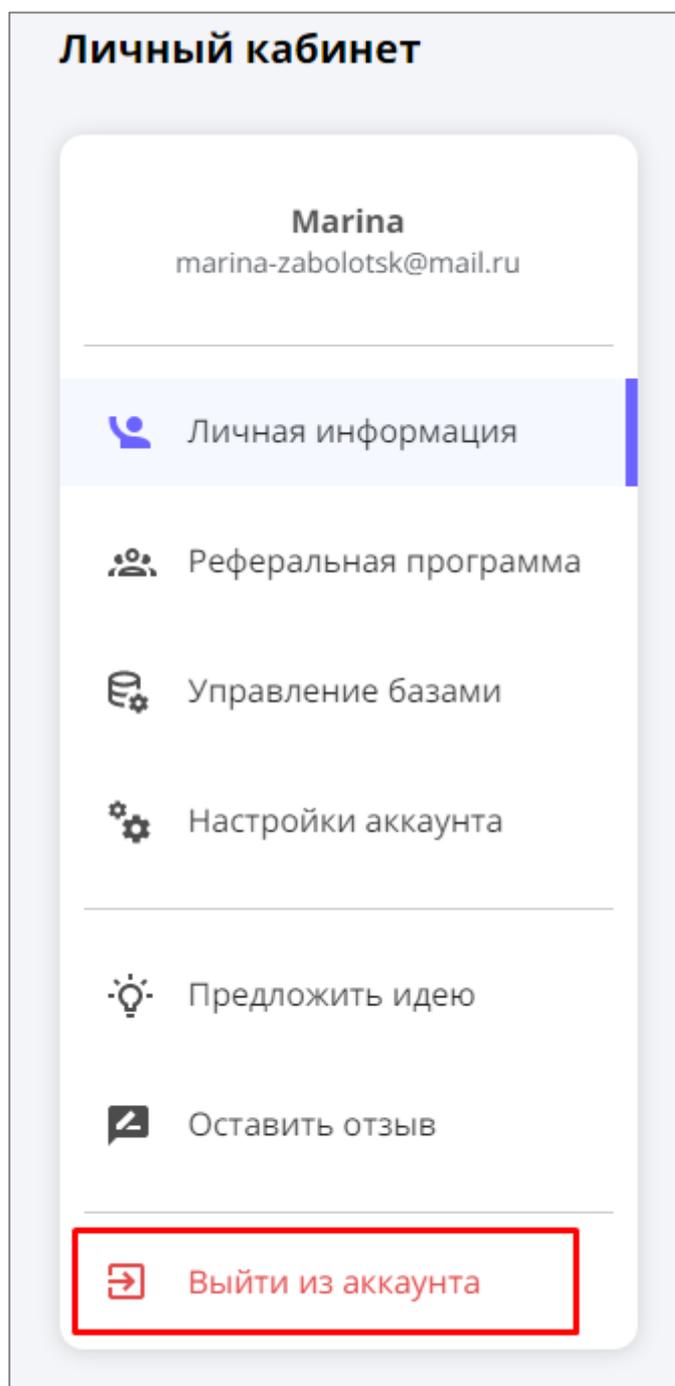


Рисунок 2.21 – Кнопка «Выйти из аккаунта» в личном кабинете

Так же выйти из аккаунта можно на стартовой веб-странице сайта (Рис. 2.22). Вам необходимо нажать знак  рядом со своим именем в правом верхнем углу стартовой страницы сайта.

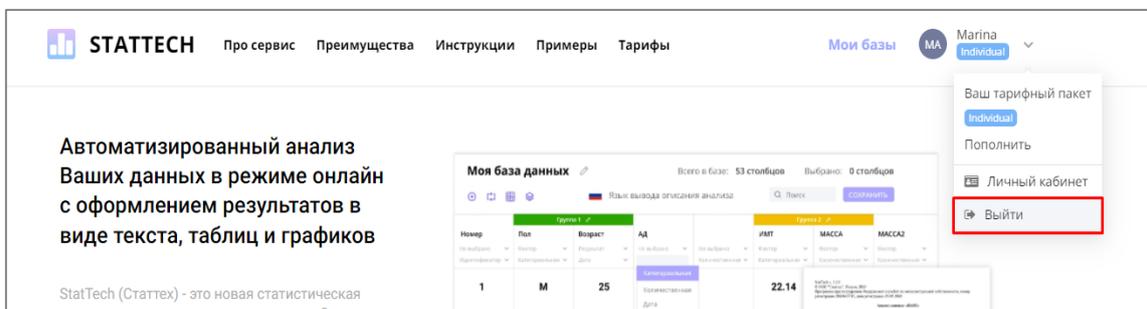


Рисунок 2.22 - Кнопка «Выйти» на лэндинге (веб-странице) сайта

### 2.3. Приобретение слота

После входа следует выбрать режим работы: «Демо» или «Индивидуальный» («Корпоративный» режим активируется только для организаций).

При выборе бесплатного режима «Демо» разрешается загрузка базы данных, в которой не более 30 строк. Срок доступа к режиму – неограничен. Базу из слота режима «Демо» можно удалить и заменить на новую неограниченное количество раз.

Платный режим «Индивидуальный» отличается от режима «Демо» возможностью загрузки базы любого объема, возможностью приобретения дополнительных слотов для одновременной загрузки нескольких баз. Срок действия данного тарифа определяется пользователем, минимальный срок – 12 месяцев. После окончания срока действия тарифа пользователь может его продлить.

Для покупки слота необходимо в личном кабинете нажать кнопку «Пополнить» (рис. 2.23) и выбрать подходящий тариф (рис. 2.24).

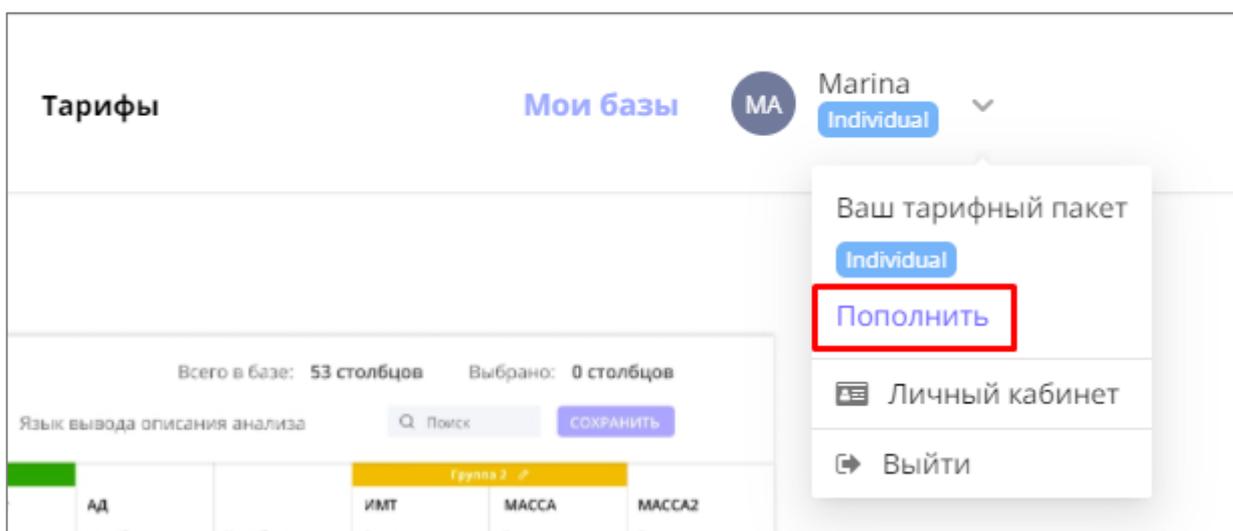


Рисунок 2.23 – Кнопка «Пополнить» для перехода на форму для приобретения / продления доступа к программе

Тарифы			
	ДЕМО бесплатно	популярный ИНДИВИДУАЛЬНЫЙ от 8 000,00 Р	КОРПОРАТИВНЫЙ от 50 000,00 Р
Описание			
Размер базы	30 строк	∞	∞
Техподдержка	—	✓	✓
Доступ к обучающим материалам	—	✓	✓
Минимальный срок доступа к базе данных	1 месяц	12 месяцев	12 месяцев
		<b>ВЫБРАТЬ</b>	<b>ВЫБРАТЬ</b>

Рисунок 2.24 – Страница выбора подходящего тарифа

После нажатия кнопки «Выбрать» откроется форма ввода данных, которая предусматривает выбор количества баз и количества месяцев использования (рис. 2.25). После заполнения этих данных необходимо нажать кнопку «Оплатить» и сайт перенаправит Вас на оплату.

### Индивидуальный ✕

Выберите количество баз и срок их действия

Количество баз

Длительность (мес.)

Промокод

Стоимость 8 000,00 Р

Отмена
Оплатить

Рисунок 2.25 – Интерфейс настройки тарифного пакета

## 2.4. Загрузка базы данных

После приобретения слота нажимаем кнопку «Мои базы», чтобы загрузить данные (Рис. 2.26).

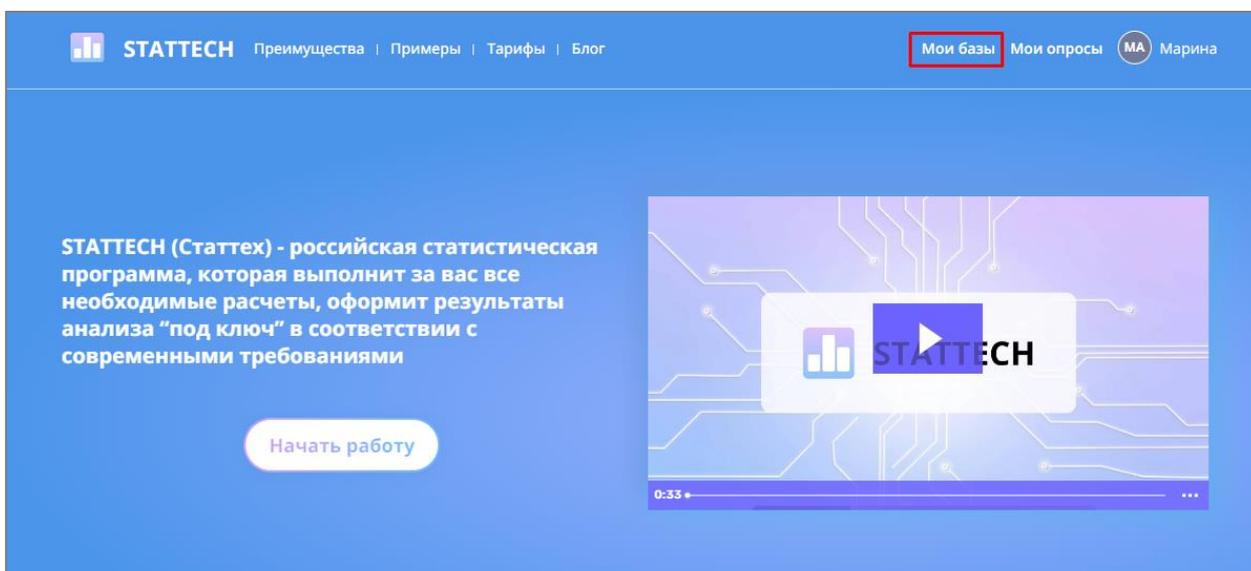


Рисунок 2.26 – Кнопка «Мои базы» на стартовой странице (лендинге) сайта

При необходимости можно удалить базу данных, чтобы освободить место для загрузки другой базы данных.

Выбираем «Загрузить базу» (Рис. 2.27), после чего выбираем базу формата .XLSX (MS Excel) или .STDB из файлов компьютера и подтверждаем ее загрузку (Рис. 2.28 и Рис. 2.29). На экран выводится сообщение об успешной загрузке базы данных.

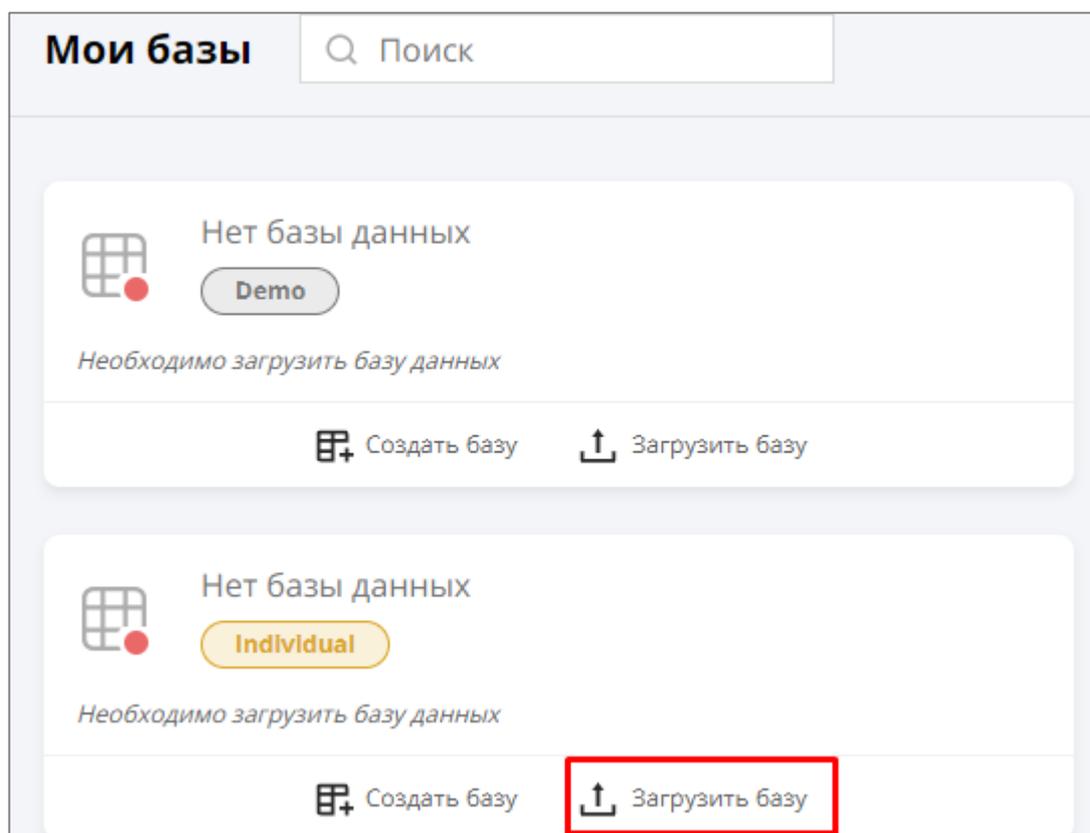


Рисунок 2.27 – Кнопка «Загрузить базу»

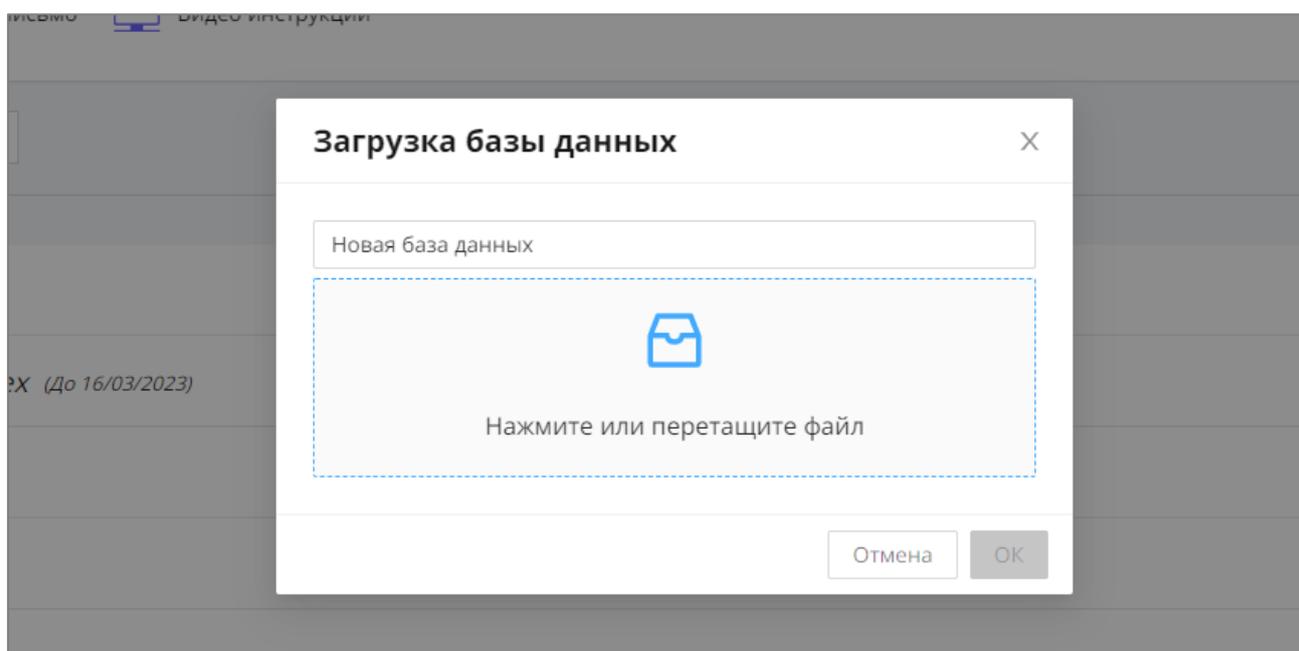


Рисунок 2.28 – Интерфейс загрузки базы данных

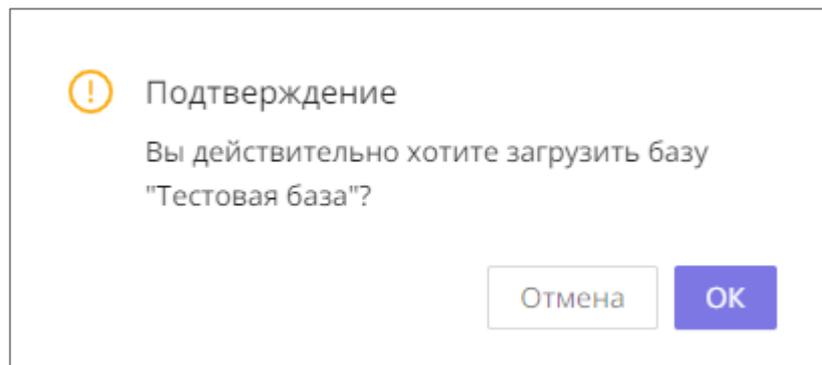


Рисунок 2.29 – Окно подтверждения загрузки базы данных

## 2.5. Создание пустой базы данных

Если у Вас нет готовой БД, то есть возможно создать пустую БД, нажав на кнопку «Создать базу» на свободном слоте (Рис. 2.30).

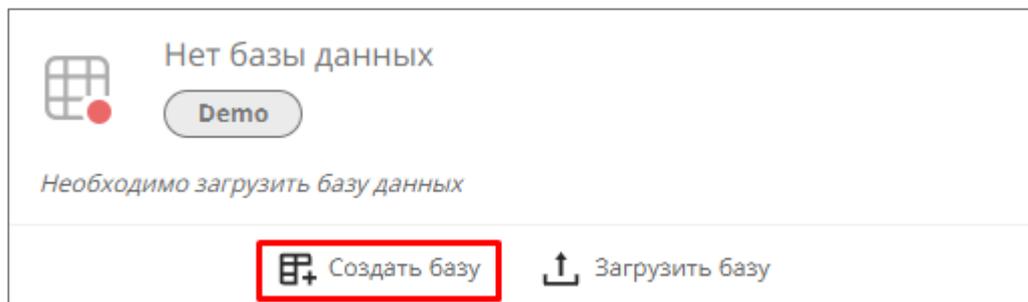


Рисунок 2.30 – Кнопка «Создать базу»

После нажатия на кнопку откроется форма для заполнения, где необходимо указать «Название базы», «Число строк» и «Число столбцов». После заполнения полей нажать на кнопку «Создать базу» (Рис. 2.31).

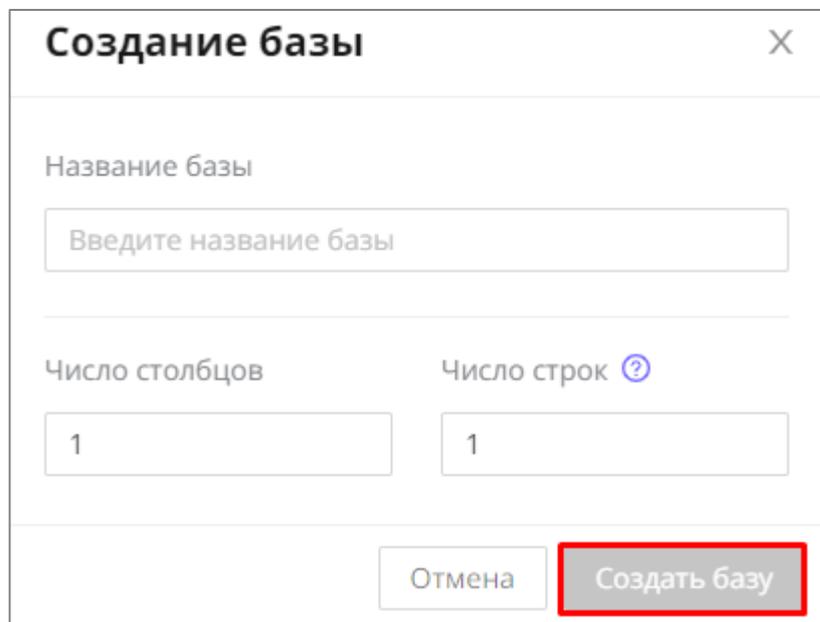


Рисунок 2.31 – Форма заполнения при создании пустой БД

Возможность создания пустой базы данных на ДЕМО-слоте предусмотрена с количеством строк НЕ БОЛЕЕ 30.

## 2.6. Удаление и восстановление базы данных

Для удаления базы данных необходимо нажать на кнопку «Удалить базу» на слоте в разделе «Мои базы» (Рис. 2.32).

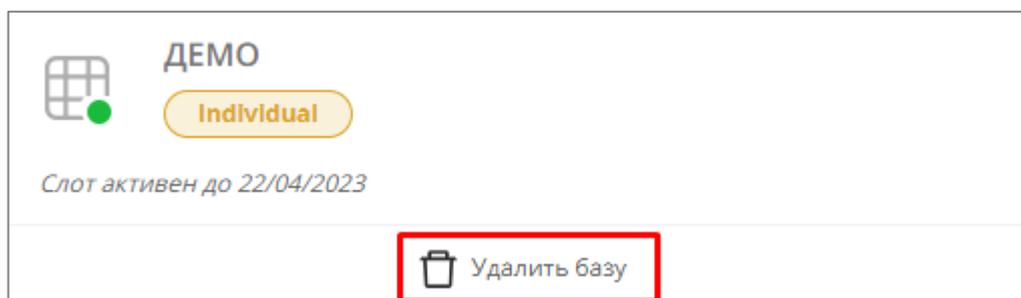


Рисунок 2.32 – Кнопка «Удалить базу» на слоте

После удаления базы она перемещается в корзину, где хранится на протяжении 30 дней. «Корзина» находится в правом верхнем углу в разделе «Мои базы» (Рис. 2.33).

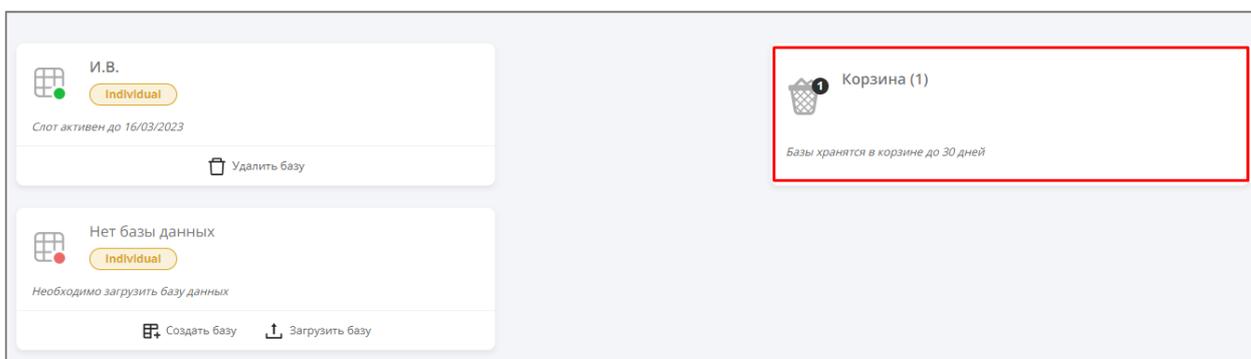


Рисунок 2.33 – «Корзина» в разделе «Мои базы»

Кликнув на «Корзину», откроется форма, где можно увидеть удаленные базы данных. На данной форме можно «Вернуть базу» и «Очистить корзину» (Рис. 2.34).

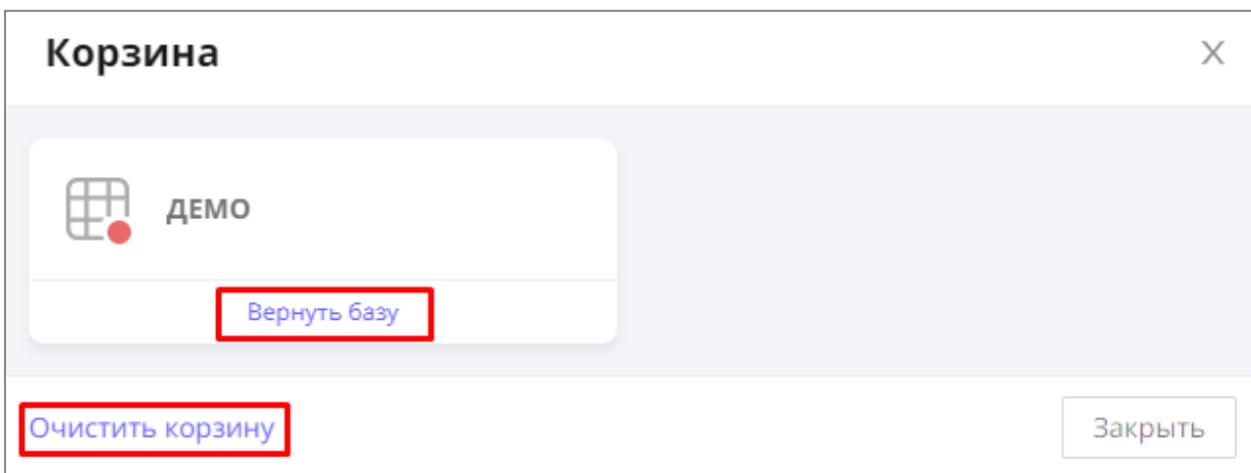


Рисунок 2.34 – Кнопки «Вернуть базу» и «Очистить корзину»

## ГЛАВА 3. ВЫБОР ТИПА ПЕРЕМЕННЫХ

Открываем необходимую базу данных (Рис. 3.1). Для этого наводим «мышкой» на загруженную базу и нажимаем левой кнопкой мыши.

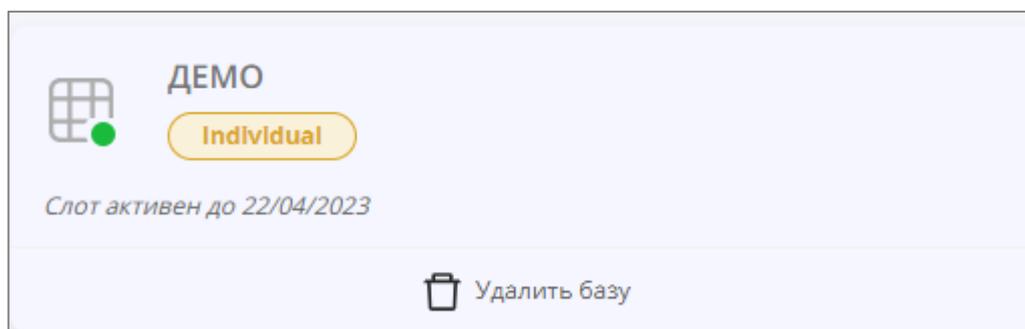


Рисунок 3.1 – «Открыть» в веб-интерфейсе

Необходимо определить тип переменной для каждого показателя в колонках. Автоматически, во всех колонках, на момент загрузки базы, проставлено «Не определено» (Рис. 3.2).

	Без группы				
	☑ Номер	☑ Фамилия	☑ Пол	☑ Возраст	☑ Группа
	Не определено				
1	1	Семенова		44	
2	2	Савченкова		34	
3	3	Землянкин	1	45	
4	4	Романов	1	59	
5	5	Сергеев	1	66	
6	6	Седова		45	

Рисунок 3.2 – Пример колонок в базе данных

Для того, чтобы задать тип переменной в колонке, Вам необходимо нажать на знак  в каждой колонке, для выбора необходимого типа переменной (Рис. 3.3).

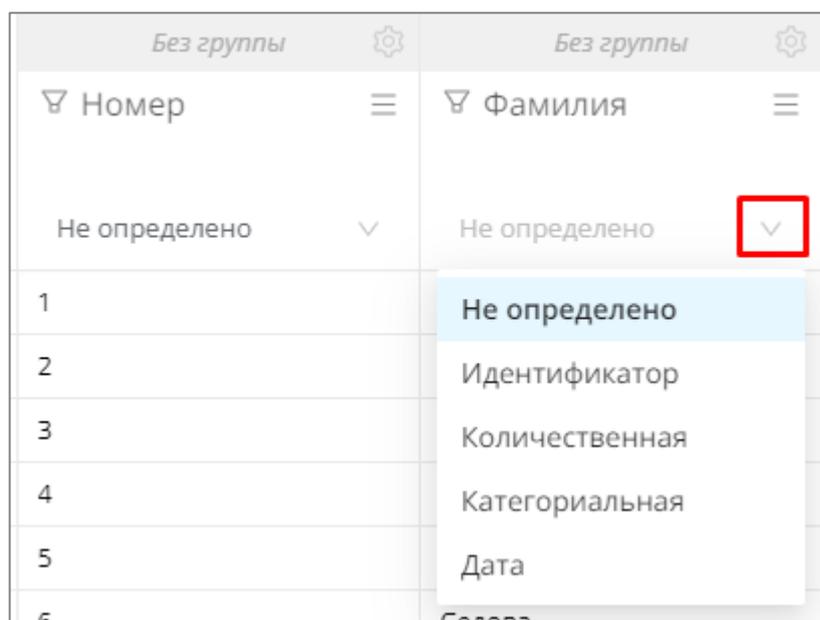


Рисунок 3.3 – Пиктограмма выбора типа переменных

Есть 4 типа переменных:

1. *Идентификатор*. Переменная, которая помогает идентифицировать. Например, фамилия.
2. *Количественная*. Применяется там, где использованы только числовые значения. Например, возраст или показатели анализа крови такие, как лейкоциты, эритроциты и т.д.
3. *Категориальная*. При выборе категориальной переменной будет задаваться категория для каждого типа. Например, 0 – женщины, 1 – мужчины. В поле «Категория» необходимо вводить названия в именительном падеже и с заглавной буквы, а в поле «Категория в род. падеже» - с маленькой буквы и в родительном падеже (Рис. 3.4).

**Задать категории** X

Числовое значение	Категория	Категория в род. падеже
0	<input type="text" value="Мужчины"/>	<input type="text" value="мужчин"/>
1	<input type="text" value="Женщины"/>	<input type="text" value="женщин"/>

Рисунок 3.4– Форма заполнения категориальной переменной

После заполнения формы необходимо нажать кнопку «OK» для применения данных и тип переменной задается в колонке (Рис. 3.5).

<i>Без группы</i>		<i>Без группы</i>		<i>Без группы</i>	
∇ Фамилия	≡	∇ Пол	≡	∇ Возраст	≡
Идентификатор	∇	Категориальная	∇	Количественная	∇
Семенова		0 (Мужчины)	∇	44	
Савченкова		0 (Мужчины)	∇	34	
Земляникин		1 (Женщины)	∇	45	
Романов		1 (Женщины)	∇	59	
Сергеев		1 (Женщины)	∇	66	
Седова		0 (Мужчины)	∇	45	

Рисунок 3.5 – Пример категориальной переменной

4. *Дата*. Используется для колонок с датами. Например, дата рождения, дата операции (Рис. 3.6).

<i>Без группы</i> ⚙		<i>Без группы</i> ⚙	
☑ № ИБ	☰	☑ Дата операции	☰
Идентификатор	▼	Дата	▼
1680		2007-03-15T00:00:00+00:00	
2889		2007-05-21T00:00:00+00:00	
5448		2008-08-21T00:00:00+00:00	
6726		2007-10-17T00:00:00+00:00	
7322		2007-11-08T00:00:00+00:00	
8669		2008-11-14T00:00:00+00:00	

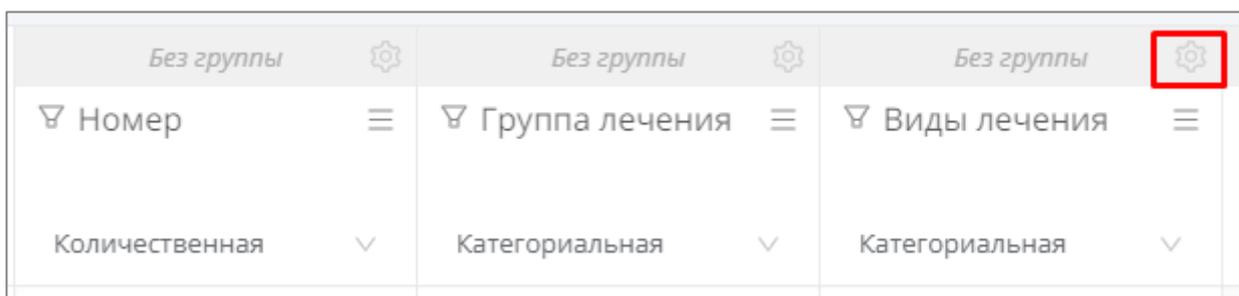
Рисунок 3.6 – Пример типа переменных «Дата»

## ГЛАВА 4. НАСТРОЙКА КОЛОНОК. СОЗДАНИЕ ГРУППЫ ПОКАЗАТЕЛЕЙ. ПРЕОБРАЗОВАНИЕ КОЛОНКИ

### 4.1. Настройка колонок

Настройка колонок включает в себя возможность редактирования поля «Наименование», «Наименование (в родительном падеже)», «Единица измерения», «Число десятичных знаков».

Для того, чтобы открыть форму редактирования колонки необходимо нажать на знак шестеренки вверху колонки (Рис. 4.1).



Без группы	Без группы	Без группы
☑ Номер	☑ Группа лечения	☑ Виды лечения
Количественная	Категориальная	Категориальная

Рисунок 4.1 – Пиктограмма открытия формы настройки колонки

После открытия формы, заполняем необходимые поля:

- «Наименование» - заполняется название колонки в именительном падеже с заглавной (большой) буквы.
- «Наименование (в родительном падеже)» - прописываем название колонки с маленькой буквы, в родительном падеже.
- «Наименование для отображения на графиках» - прописываем название для отображения на графике.
- «Единицы измерения» заполняются для количественных переменных (ммоль/л, г/л, полных лет и т.д.).

- «Число десятичных знаков» - число знаков после запятой, с какой точностью будут выводиться данные после расчета. Указываем для количественных переменных.

Ниже приведен пример заполнения данной формы для количественного показателя (Рис. 4.2).

The image shows a configuration window with the following elements:

- Navigation tabs: **Основные**, Группы, Зависимости, Модели.
- Field: Наименование (Leukoциты) with a green checkmark icon.
- Field: Наименование (в родительном падеже) (содержание лейкоцитов до операции).
- Field: Наименование для отображения на графиках (Ley).
- Field: Единица измерения (10<sup>9</sup>/л).
- Field: Число десятичных знаков (2).
- Checkbox:  Временной срок ⓘ.
- Buttons: Отмена, ОК.

Рисунок 4.2 – Пример заполнения формы настройки колонки

Настройка колонки с иными видами типа переменных имеет более упрощенную форму редактирования, которая содержит только «Наименование» и «Наименование (в родительном падеже)» (Рис. 4.3).

**Настройки колонки "Группа"** X

Основные Группы Зависимости Модели

Наименование  
Группа ✓

Наименование (в родительном падеже)  
группы

Отмена ОК

Рисунок 4.3 – Пример заполнения формы настройки колонки

Так же в настройках колонки есть функция «Временной срок». Её необходимо использовать если показатель имеет временный характер для анализа выживаемости (Рис. 4.4).

**Настройки колонки "Гемоглобин (до)"** X

Основные Группы Зависимости Модели

Наименование  
Гемоглобин (до) ✓

Наименование (в родительном падеже)

Единица измерения

Число десятичных знаков 0

Временной срок ⓘ

Отмена ОК

Рисунок 4.4 – Кнопка «Временной срок» в настройках колонки

Для подключения функции «Временной срок» необходимо проставить галочку в окошке (Рис. 4.5).

Настройки колонки "Гемоглобин (до)"

Основные Группы Зависимости Модели

Наименование  
Гемоглобин (до) ✓

Наименование (в родительном падеже)

Единица измерения

Число десятичных знаков 0

Временной срок ⓘ

Тип выживаемости  
Общая ▼

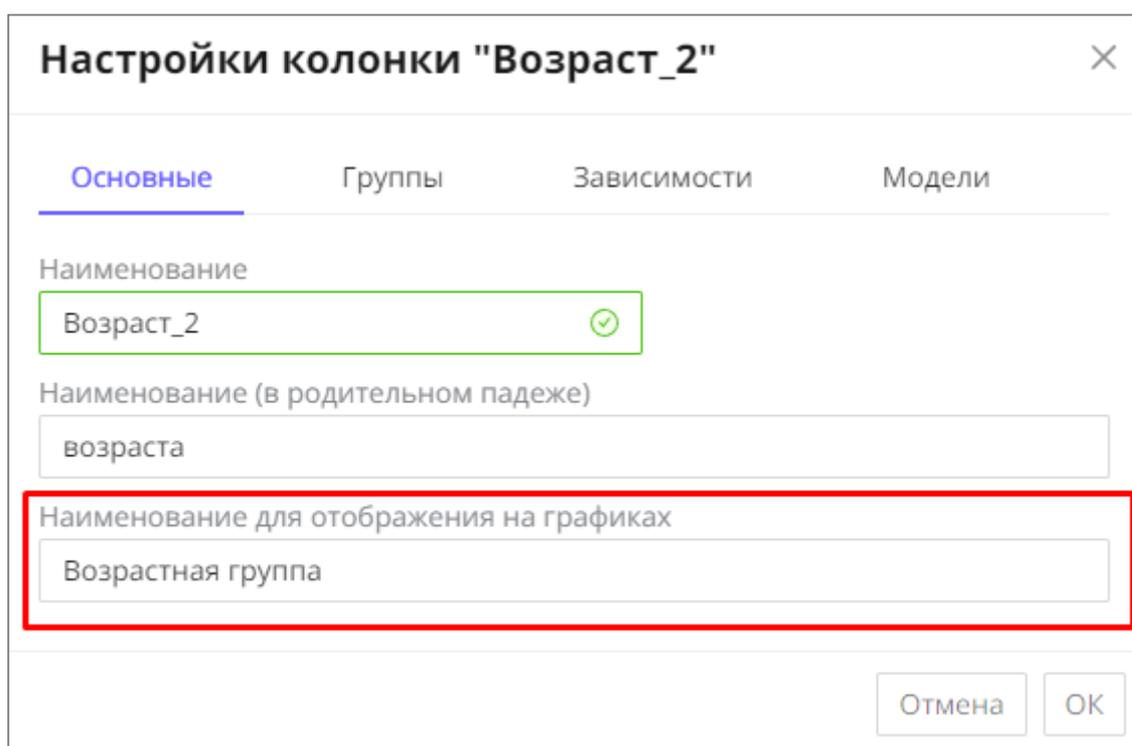
Отмена ОК

Рисунок 4.5 – Кнопка проставления «Временной срок» и активный функционал «Тип выживаемости»

При проставлении учета «Временной срок» становится активным выбор «Тип выживаемости» (Рис. 4.5).

## 4.2 Изменение названия колонки и категорий для графиков

Для того, чтобы изменить названия колонки, необходимо нажать на пиктограмму . Откроется форма настройки колонки, где в строке «Наименование для отображения на графиках» можно ввести название, которое будет использоваться в легенде на графике (Рис. 4.6).



Настройки колонки "Возраст\_2" ✕

Основные Группы Зависимости Модели

Наименование

Возраст\_2 ✓

Наименование (в родительном падеже)

возраста

Наименование для отображения на графиках

Возрастная группа

Отмена ОК

Рисунок 4.6 – Строка «Наименование для отображения на графиках»

Для изменения категорий необходимо зайти в форму настройки категориальных переменных и заполнить столбец «Наименования для отображения на графиках» (Рис. 4.7).

**!!Постановка галочки на целевой категории даёт возможность при анализе выживаемости, ОШ, ОР или логистической регрессии другие категории объединять в «Другие», а одну категорию сделать целевой!! При условии, что у переменной больше, чем 2 категории.**

### Задать категории ×

Числовое значение	Категория	Категория в род. падеже	Название для графиков	Целевая
1	<input type="text" value="мышцы шейно-пл"/>	<input type="text" value="мышцы шейно-пл"/>	<input type="text" value="Мышцы ШПП"/>	<input type="checkbox"/>
2	<input type="text" value="мышцы поясницы"/>	<input type="text" value="мышцы поясницы"/>	<input type="text" value="Мышцы ККО"/>	<input type="checkbox"/>
3	<input type="text" value="мышцы спины (тс"/>	<input type="text" value="мышцы спины (тс"/>	<input type="text" value="Мышцы спины"/>	<input type="checkbox"/>
4	<input type="text" value="комбинированнь"/>	<input type="text" value="комбинированнь"/>	<input type="text" value="Симбиоз"/>	<input type="checkbox"/>

Рисунок 4.7 – Настройка названий для графиков у категорий переменной

### 4.3 Создание группы показателей

Группы показателей используются для объединения нескольких показателей. **В группу объединяются либо только категориальные, либо только количественные показатели отдельно, но не вместе, так как в результатах могут появиться ошибки.** Каждая создаваемая группа имеет свой цвет.

По умолчанию, все показатели при загрузке базы определены «Без группы». Чтобы создать группу показателей или выбрать уже созданную, Вам необходимо зайти в настройки колонки через знак , далее перейти в раздел «Группы» (Рис. 4.8).

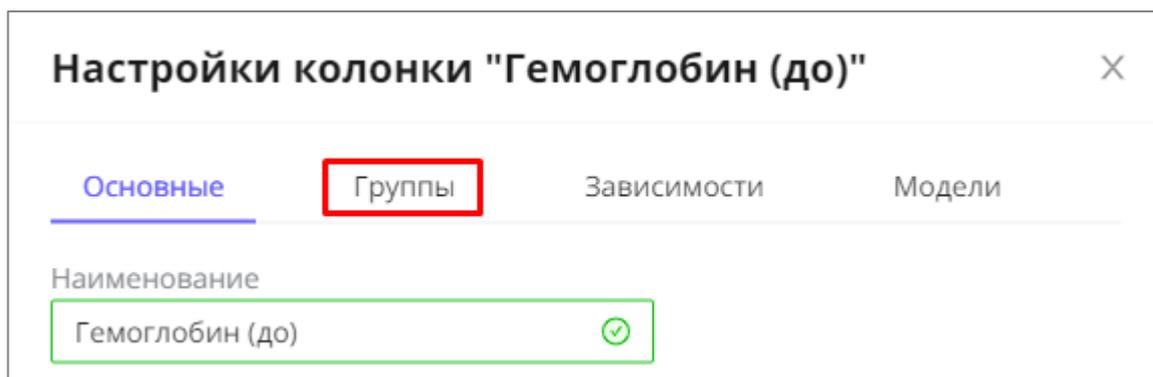


Рисунок 4.8 – Раздел «Группы» в настройках колонки

В разделе «Группы» нажимаем кнопку «Создать новую группу» и заполняем поля. В первой строке пишем наименование группы с заглавной буквы в именительном падеже, ниже пишем наименование группы в родительном падеже и нажимаем знак галочки для сохранения изменений (Рис 4.9).

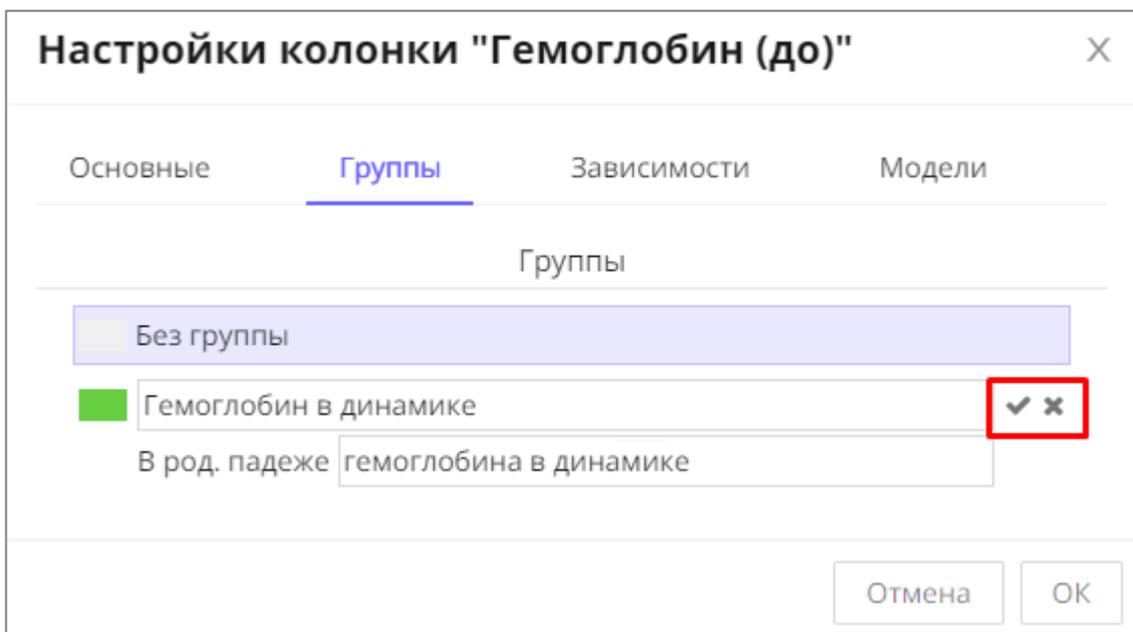


Рисунок 4.9 – Форма заполнения новой группы значений

Для выбора созданной группы для колонки нужно выбрать группу и потом нажать на кнопку «Ок» (Рис. 4.10).

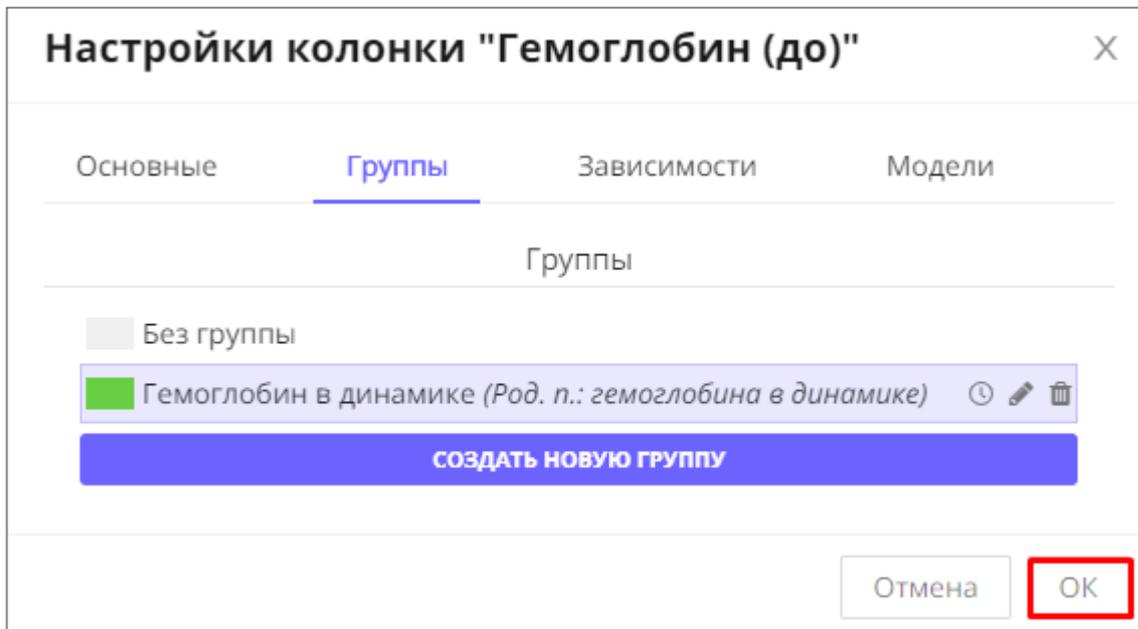


Рисунок 4.10 – Форма выбора группы значений

При необходимости сравнения показателей в формате «До - После» необходимо при создании группы нажать на знак часиков  (Рис. 4.11).

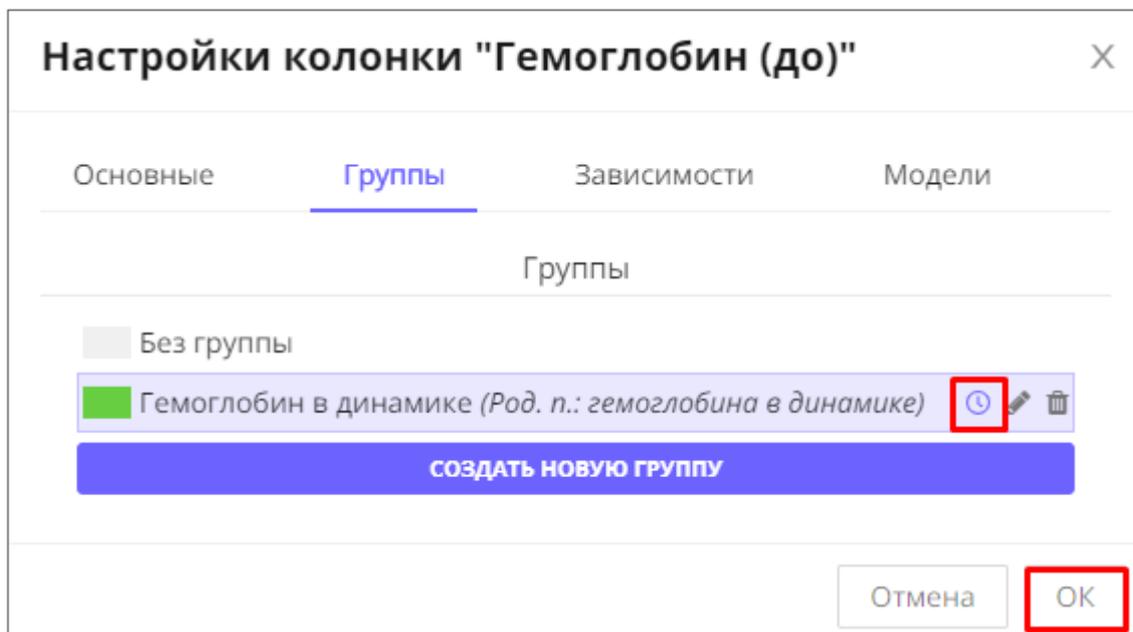


Рисунок 4.11 – Кнопка настройки «До – После»

По умолчанию, колонка, на которой создается группа, уже входит в эту группу. Удалить ее из группы можно путем нажатия на кнопку «Без группы» (Рис. 4.12).



Рисунок 4.12 – Кнопка «Без группы» в настройках колонки

Для включения в созданную группу других показателей необходимо выбрать колонку с нужным показателем, перейти в «Настройки колонки» -> «Группы». На экране отобразится список всех созданных групп. Необходимо выбрать нужную группу, нажав на ее название (Рис. 4.13).

Без группы	Гемоглобин в дина...	Гемоглобин в дина...	Без группы
Группа	Гемоглобин (до)	Гемоглобин (после)	Осложнения ранние
Категориальная	Количественная	Количественная	Категориальная
0 (Контрольная группа)	130	103	1 (Наличие)
0 (Контрольная группа)	118	95	1 (Наличие)

Рисунок 4.13 – Пример объединения нескольких колонок в одну группу

Ранее созданную группу можно редактировать. Для этого необходимо зайти в «Настройки колонки» -> «Группы». Ищем необходимую Вам группу для редактирования и нажимаем на знак карандаша  (Рис. 4.14).

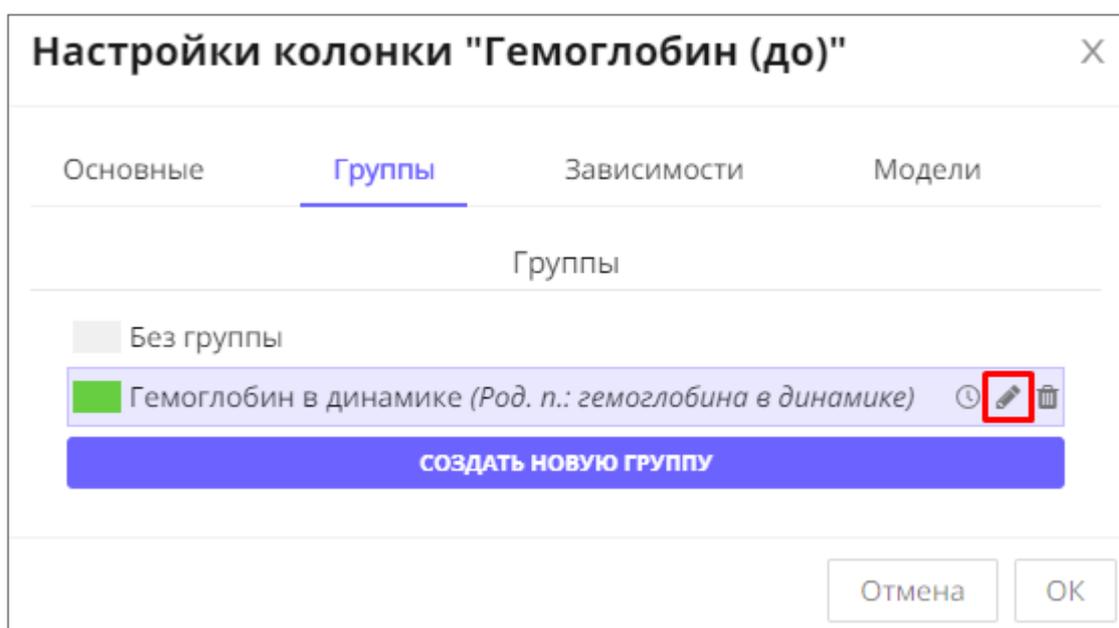


Рисунок 4.14 – Кнопка редактирования группы

Также, при необходимости, ненужную группу можно удалить. Для этого необходимо зайти в «Настройки колонки» -> «Группы». На ненужной Вам группе нажимаем на знак корзины  и группа удалится (Рис. 4.15).

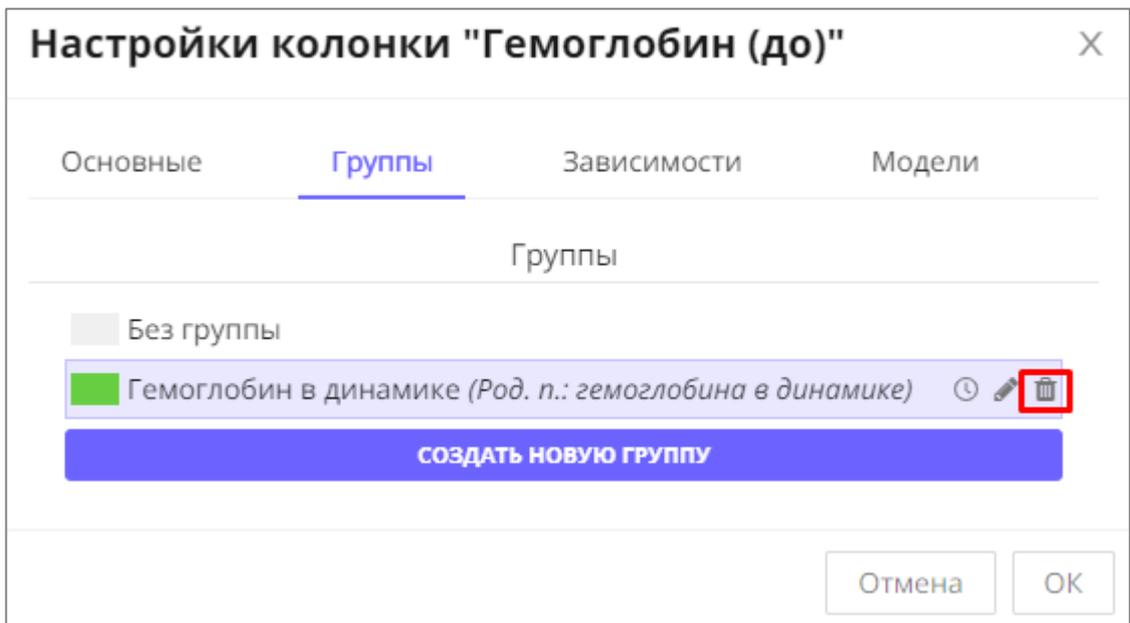


Рисунок 4.15 – Кнопка удаления группы

## 4.3. Изменение данных

### 4.3.1 Функции «Вставить...», «Сортировка...», «Очистить все значения», «Удалить», добавить «N строк» / «N столбцов»

На колонке под знаком  хранится функционал по изменению данных, включая вставку, сортировку, преобразование и удаление колонок (Рис. 4.16).

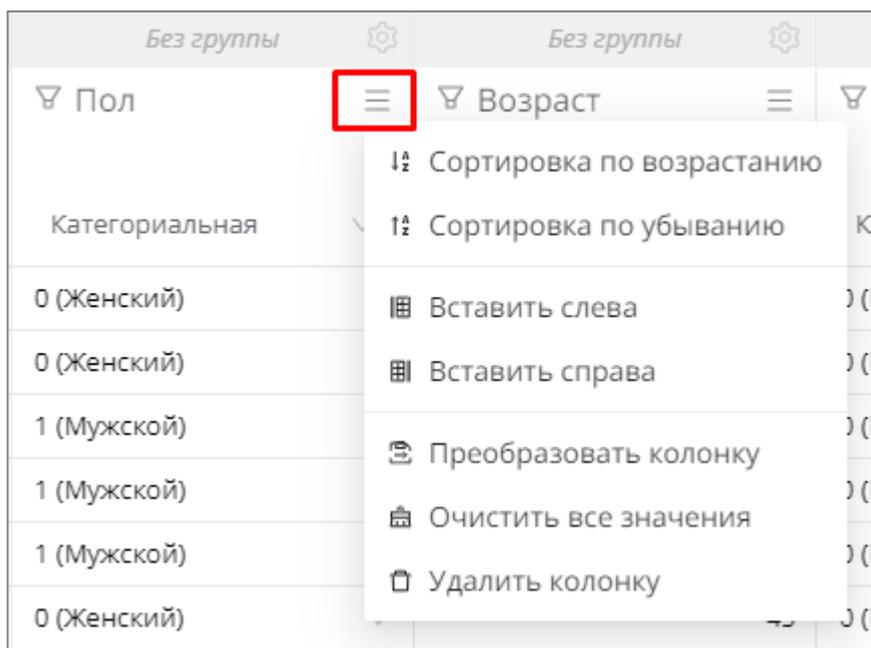


Рисунок 4.16 – Кнопка дополнительного функционала на колонке

«Сортировка по возрастанию» и «Сортировка по убыванию» подразумевают под собой сортировку показателей, обозначенных числами, от меньшего значения к большему и наоборот, соответственно.

Колонки с текстовыми значениями, например, фамилии пациентов, будут сортироваться по алфавиту. Если применить «Сортировка по возрастанию», то последовательность данных в колонке будет от А до Я. Если использовать «Сортировка по убыванию», то от Я до А.

Кнопки «Вставить слева» и «Вставить справа» предполагают появление слева или справа от колонки, на которой используете эту кнопку, новой пустой (не заполненной данными) колонки (Рис. 4.17).

	Без группы	Без группы
	▾ Новая колонка    ≡ Не определено    ▾	▾ Номер    ≡ Не определено    ▾
1		1
2		2
3		3
4		4
5		5
6		6
7		7
8		8

Рисунок 4.17 – Результаты действия кнопки «Вставить слева»

Функция «Очистить все значения» предполагает удаление всех значений в колонке, но сама колонка останется в базе (Рис.4.18).

	Без группы
	▾ Номер    ≡ Не определено    ▾
1	
2	
3	
4	
5	
6	

Рисунок 4.18 – Вид колонки после нажатия на кнопку «Очистить все значения»

Функция «Удалить колонку» представляет собой удаление всей колонки со всеми данными в ней.

Кнопка «Преобразовать колонку» подразумевает под собой копирование колонки со всеми её значениями или преобразование её данных. Функционал преобразования данных рассмотрен в отдельной подглаве.

Добавление нескольких строк возможно путём нажатия на кнопку «Добавить» внизу, слева страницы. По умолчанию при нажатии на кнопку добавляется одна строка. При необходимости добавления нескольких строк нужно ввести число в окне рядом, а потом нажать на кнопку «Добавить» (Рис. 4.19).

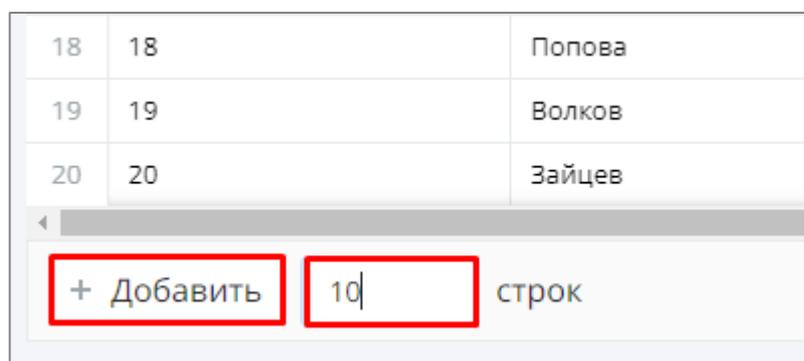


Рисунок 4.19 – Кнопка «Добавить» и окно ввода количества строк

Добавление нескольких столбцов доступно при нажатии на кнопку  в конце всех столбцов базы данных (Рис. 4.20).

STATTECH | Задать вопрос | Отправить письмо | Инструкции | Мои базы | Мария | МАР |

ДЕМО | Сохранить базу в файл | Всего в базе: 17 столбцов и 20 строк

Отменить действие (Ctrl+Z) | Вернуть действие (Ctrl+Y) | Сбросить все зависимости | Описательная статистика | Настройки анализа | Выполнить анализ

Курение	Глюкоза	Наличие артер гипертонии	Отсутствие АГ	Наличие АГ	АД	Препарат	АД после лечения
3 (стание)	4.7	0 (Отсутствие гипертонии)	1 (Отсутствие АГ)	0 (Отсутствие)	121	0 (Без лечения)	121
4 (ие)	4.9	0 (Отсутствие гипертонии)	1 (Отсутствие АГ)	0 (Отсутствие)	125	0 (Без лечения)	123
5 (стание)	6.5	0 (Отсутствие гипертонии)	1 (Отсутствие АГ)	0 (Отсутствие)	126	0 (Без лечения)	125
6 (стание)	6.5	0 (Отсутствие гипертонии)	1 (Отсутствие АГ)	0 (Отсутствие)	132	0 (Без лечения)	130
7 (стание)	5.1	0 (Отсутствие гипертонии)	1 (Отсутствие АГ)	0 (Отсутствие)	135	0 (Без лечения)	132
8 (стание)	7.2	0 (Отсутствие гипертонии)	1 (Отсутствие АГ)	0 (Отсутствие)	136	0 (Без лечения)	133
9 (ие)	5.8	1 (Наличие гипертонии)	0 (Наличие АГ)	1 (Наличие)	144	1 (Препарат А)	125
10 (стание)	5.2	1 (Наличие гипертонии)	0 (Наличие АГ)	1 (Наличие)	145	2 (Препарат В)	150
11 (ие)	6.4	1 (Наличие гипертонии)	0 (Наличие АГ)	1 (Наличие)	162	1 (Препарат А)	145
12 (ие)	5.7	1 (Наличие гипертонии)	0 (Наличие АГ)	1 (Наличие)	165	2 (Препарат В)	170
13 (стание)	6.9	1 (Наличие гипертонии)	0 (Наличие АГ)	1 (Наличие)	167	1 (Препарат А)	125
14 (ие)	5.9	1 (Наличие гипертонии)	0 (Наличие АГ)	1 (Наличие)	176	2 (Препарат В)	138
15 (стание)	8.2	1 (Наличие гипертонии)	0 (Наличие АГ)	1 (Наличие)	185	1 (Препарат А)	133
16 (ие)	9.5	1 (Наличие гипертонии)	0 (Наличие АГ)	1 (Наличие)	185	2 (Препарат В)	169
17 (ие)	7.8	1 (Наличие гипертонии)	0 (Наличие АГ)	1 (Наличие)	189	1 (Препарат А)	158
18 (ие)	7.3	1 (Наличие гипертонии)	0 (Наличие АГ)	1 (Наличие)	190	2 (Препарат В)	191
19 (ие)	8.9	1 (Наличие гипертонии)	0 (Наличие АГ)	1 (Наличие)	202	1 (Препарат А)	137
20 (ие)	8.5	1 (Наличие гипертонии)	0 (Наличие АГ)	1 (Наличие)	203	2 (Препарат В)	195

Рисунок 4.20 – Функционал по добавлению нескольких столбцов

При необходимости добавления нескольких столбцов нужно ввести число в строке «Число столбцов», а потом нажать на кнопку «Добавить» (Рис. 4.21). По умолчанию в строке «защито» число 1 для добавления одного столбца.

☰+

**Добавить новые столбцы**

Число столбцов

1

Отмена **Добавить**

Рисунок 4.21 – Кнопка «Добавить» и окно ввода количества столбцов

### 4.3.2. Функция «Преобразовать колонку»

Нажимаем кнопку «Преобразовать колонку» (Рис. 4.22).

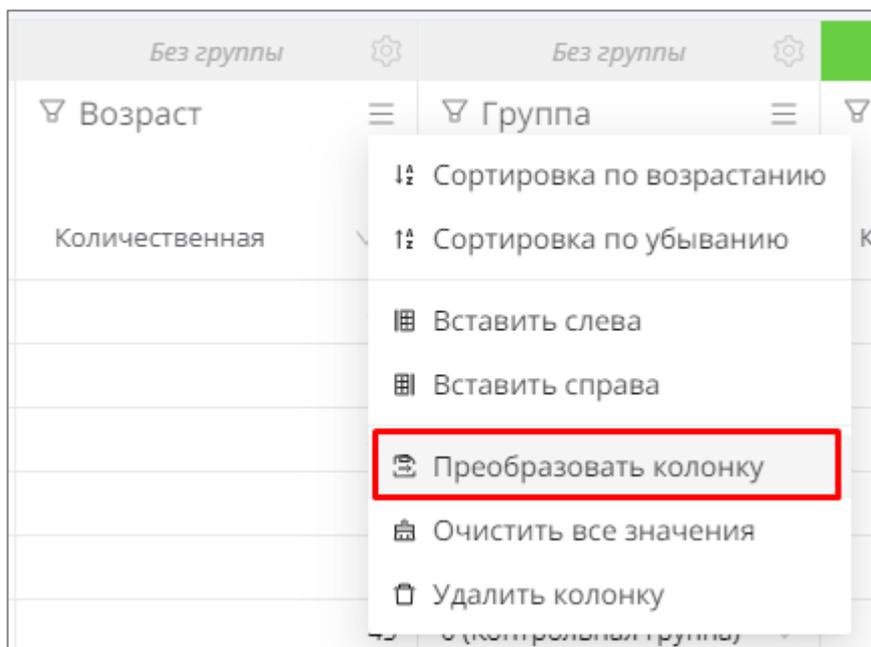


Рисунок 4.22 – Кнопка «Преобразовать колонку»

После нажатия на кнопку появится окно преобразования колонки. Здесь нужно заполнить поле «Наименование» – это наименование новой колонки. Также необходимо выбрать «Режим». (Рис.4.23).

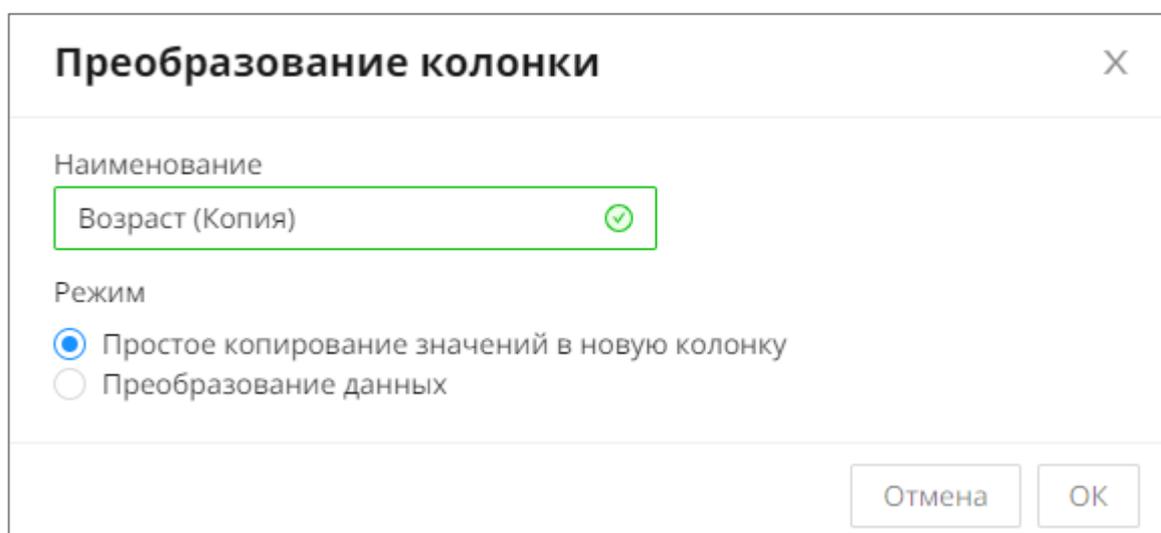


Рисунок 4.23 – Окно преобразования колонки

Выбор режима «Простое копирование значений в новую колонку» предполагает копирование колонки с данными без каких-либо изменений (Рис. 4.24).

<i>Без группы</i>	⚙	<i>Без группы</i>	⚙
∨ Возраст	≡	∨ Возраст (Копия)	≡
Количественная	∨	Количественная	∨
	44		44
	34		34
	45		45
	59		59
	66		66
	45		45

Рисунок 4.24 – Пример простого копирования значений в новую колонку

Выбор режима «Преобразование данных» откроет дополнительный функционал (Рис. 4.25).

**Преобразование колонки** [X]

Наименование  
Возраст (Копия) ✓

Режим  
 Простое копирование значений в новую колонку  
 Преобразование данных

*Правила преобразования (приоритет – сверху вниз)*

Если значение = ПУСТО, то [ 0 ] [🗑️]

**ДОБАВИТЬ НОВОЕ ПРАВИЛО**

Значение по умолчанию Исходное значение ▾

[Отмена] [ОК]

Рисунок 4.25 – Функционал кнопки «Преобразование данных»

В окне «Правила преобразования» необходимо нажать на кнопку «Добавить новое правило» (Рис. 4.26). Под правилом подразумевается определение условия и результата перекодировки значений по принципу «ЕСЛИ <условие>, ТО <результат>».

Рисунок 4.26 – Кнопка «Добавить новое правило»

В StatTech доступны различные правила преобразования.

1) Присвоение исходным значениям до определенного числа нового значения. Например, «Если возраст до 40 лет – то присвоить значение 1». Значения в исходной колонке до 40 включительно будут перекодированы в новой колонке числом 1 (Рис. 4.27).

Рисунок 4.27 – Пример присвоения исходным данным до определенного числа нового значения

2) Замена пустых значений определенным числом. Например, «Если в колонке возраст есть пустые ячейки – то присвоить им значение 35» (Рис. 4.28).

The image shows a configuration window for a data transformation rule. At the top, the field 'Наименование' (Name) contains 'Возраст (Копия)' (Age (Copy)) with a green checkmark icon. Below this, the 'Режим' (Mode) section has two radio buttons: 'Простое копирование значений в новую колонку' (Simple copying of values to a new column) and 'Преобразование данных' (Data transformation), with the latter being selected. The 'Правила преобразования (приоритет – сверху вниз)' (Transformation rules (priority – top to bottom)) section contains a rule: 'Если значение = ПУСТО, то' (If value = EMPTY, then) followed by a text input field containing '35' and a trash icon. A blue button labeled 'ДОБАВИТЬ НОВОЕ ПРАВИЛО' (ADD NEW RULE) is positioned below the rule. At the bottom, there are labels 'Значение по умолчанию' (Default value) and 'Исходное значение' (Original value) with a dropdown arrow.

Рисунок 4.28 – Пример замены пустых значений определенным числом.

3) Замена определенных значений или диапазонов на пустые ячейки. Например, «Если возраст от 1 до 17 лет – то заменить значения на пустые» (Рис. 4.29).

Наименование  
Возраст (Копия) ✓

Режим  
 Простое копирование значений в новую колонку  
 Преобразование данных

Правила преобразования (приоритет – сверху вниз)

⋮ Если значение от 1 до 17 , то  ПУСТО 🗑

**ДОБАВИТЬ НОВОЕ ПРАВИЛО**

Значение по умолчанию Исходное значение ▾

Рисунок 4.29 – Пример замены диапазона значений на пустые ячейки

4) Замена остальных значений, не определенных другими правилами, на конкретное значение или пустые ячейки. Для этого необходимо выбрать в строке «Значение по умолчанию» -> «Исходное значение» (Рис. 4.30). Нажав на знак  увидим выпадающий список, где необходимо выбрать остальные значения будут равняться «Пусто» (Рис. 4.31) или «Значение» (Рис.4.32).

**Преобразование колонки** X

Наименование  
 ✓

Режим  
 Простое копирование значений в новую колонку  
 Преобразование данных

*Правила преобразования (приоритет – сверху вниз)*

⋮ Если значение от  до , то   🗑️

**ДОБАВИТЬ НОВОЕ ПРАВИЛО**

Значение по умолчанию Исходное значение ▼

Исходное значение

Пусто <

Значение

9	2008-11-14T00:00:00+00:00	
5	2008-03-24T00:00:00+00:00	Мигунова Лариса Михайлов

Рисунок 4.30 – Кнопка выбора значения по умолчанию

Наименование  
 ✓

Режим  
 Простое копирование значений в новую колонку  
 Преобразование данных

*Правила преобразования (приоритет – сверху вниз)*

⋮ Если значение от  до , то   🗑️

**ДОБАВИТЬ НОВОЕ ПРАВИЛО**

Значение по умолчанию Пусто ▼

Рисунок 4.31 – Замена значений, не определенных правилами, на пустые ячейки

Рисунок 4.32 – Замена значений, не определенных правилами, на конкретное значение

Правила преобразования колонки применяются последовательно, сверху вниз. Поэтому при сложных вариантах преобразования можно оперировать открытыми диапазонами. Например, перекодировку переменной «Индекс массы тела» (ИМТ) в «Степени избытка массы тела» по схеме:

- Если ИМТ < 25 кг/м<sup>2</sup>, то 1 (норма или дефицит массы тела),
- если  $25 \leq \text{ИМТ} < 30$ , то 2 (избыточная масса тела),
- если ИМТ  $\geq 30$ , то 3 (ожирение),

можно описать следующими правилами:

- Если ИМТ  $\geq 30$ , то 3,
- если ИМТ  $\geq 25$ , то 2,
- если ИМТ < 25, то 1.

В таком порядке обеспечивается правильное определение границ диапазонов с учетом включения ( $\geq$ ,  $\leq$ ) или не включения ( $>$ ,  $<$ ) пороговых значений. Пример перекодировки представлен на Рис. 4.33.

### Преобразование колонки X

Наименование  
 ✓

Режим  
 Простое копирование значений в новую колонку  
 Преобразование данных

*Правила преобразования (приоритет – сверху вниз)*

Если значение = ПУСТО, то		<input type="text" value="0"/>	
⋮ Если значение от <input type="text" value="30"/> до <input type="text"/>	, то	<input type="checkbox"/> <input type="text" value="3"/>	
⋮ Если значение от <input type="text" value="25"/> до <input type="text"/>	, то	<input type="checkbox"/> <input type="text" value="2"/>	
⋮ Если значение от <input type="text"/> до <input type="text" value="25"/>	, то	<input type="checkbox"/> <input style="border: 1px solid blue;" type="text" value="1"/>	

**ДОБАВИТЬ НОВОЕ ПРАВИЛО**

Значение по умолчанию  Исходное значение  ▼

Рисунок 4.33 – Пример заполнения формы непрерывных данных

Знак корзины дает возможность удалить созданное правило. После заполнения всех правил преобразования данных необходимо нажать на кнопку «OK». И рядом с колонкой, на которой настраивалось преобразование появляется новая колонка с преобразованными данными (Рис. 4.34).

Без группы	Без группы
ИМТ	ИМТ (Копия)
Количественная	Количественная
19.1	1
20.6	1
25.8	2
30.3	3
42.5	3
22.4	1
15.6	1
31.2	3

Рисунок 4.34 – Пример преобразования колонки ИМТ

### 4.3.3. Функция «Отображение колонок»

Функция отображения колонок позволяет скрыть колонки / переменные, у которых не задан тип данных. Для этого необходимо нажать на кнопку «Отображение переменных» и снять галочку «Отображать колонки с неопределенным типом данных» (Рис.4.35).

Без группы	Без группы	Без группы	Без группы	Без группы	Без группы
Номер	Фамилия	Пол	Трудовой статус	Возраст	Масса тела
Не определено	Не определено	Категориальная	Не определено	Количественная	Количественная
1	Семенова	0 (Женщины)	1	24	60
2	Савченкова	0 (Женщины)	3	34	68

Рисунок 4.35 – Функция «Отображение колонок»

## ГЛАВА 5. НАСТРОЙКИ АНАЛИЗА

Функционал по индивидуальной настройке анализа находится в правом верхнем углу базы данных и обозначен знаком шестерёнки  (Рис.5.1).

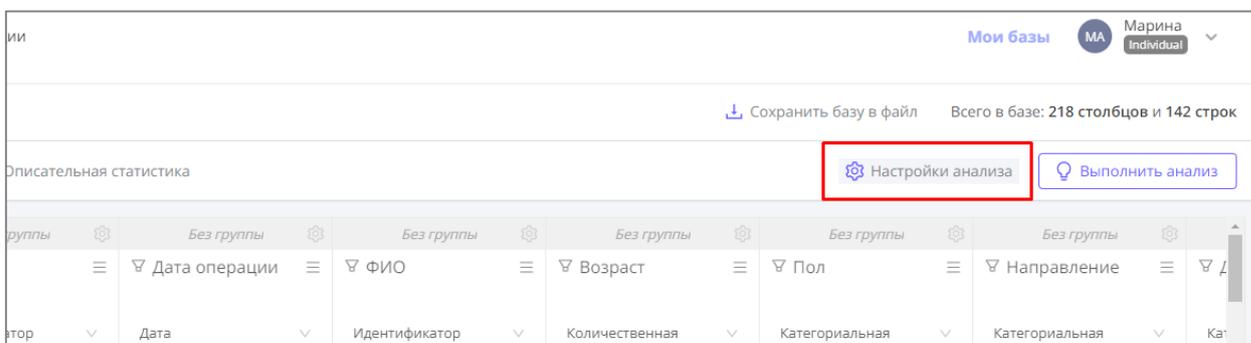


Рисунок 5.1 – Кнопка «Настройки анализа»

При нажатии на «Настройки» откроется форма редактирования настроек анализа, которая включает в себя:

- «Проверка на нормальность» подразумевает возможность выбора пользователем одного из 3 вариантов обработки количественных показателей:
  - «Выполняется» (по умолчанию) – предварительно данные проверяются на нормальность распределения с помощью критериев Шапиро-Уилка (применяется при  $N < 50$ ) и Колмогорова-Смирнова (при  $N > 50$ ). В случае соответствия нормальному распределению описание и анализ выполняются параметрическими методами. При распределении, отличном от нормального, описание и анализ выполняются непараметрическими методами.
  - «Всегда нормальное» – проверка на нормальность распределения не проводится. Независимо от распределения данных, их описание и анализ выполняются **параметрическими** методами.
  - «Всегда “ненормальное”» – проверка на нормальность распределения не проводится. Независимо от распределения данных, их описание и анализ выполняются **непараметрическими** методами.

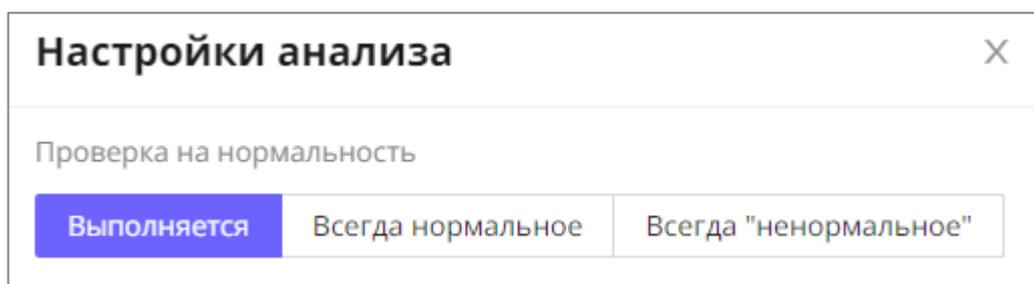


Рисунок 5.2 – Функционал «Проверка на нормальность» в настройках анализа

- Отображение диаграмм – представляет собой возможность формирования анализа данных:
  - без диаграмм (переключатель «Показывать диаграммы» выключен),
  - со всеми диаграммами (переключатель «Показывать диаграммы» включен, кнопка «Показывать все диаграммы» активна),
  - с диаграммами только для статистически значимых результатов (переключатель «Показывать диаграммы» включен, кнопка «Только для значимых результатов» активна) (Рис. 5.3).

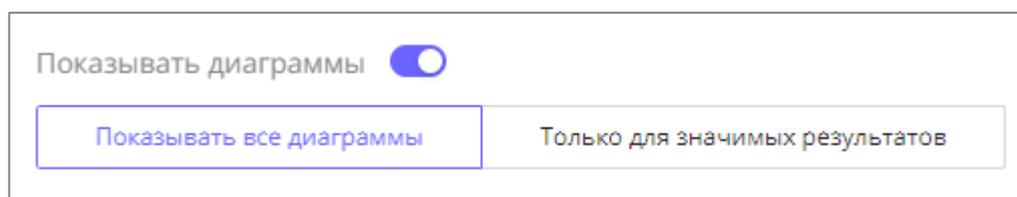


Рисунок 5.3 – Функционал по отображению диаграмм в настройках анализа

- «Цветовая схема графиков» предоставляет возможность выбора диапазона цветов для диаграмм и графиков из 9 вариантов (Рис. 5.4).

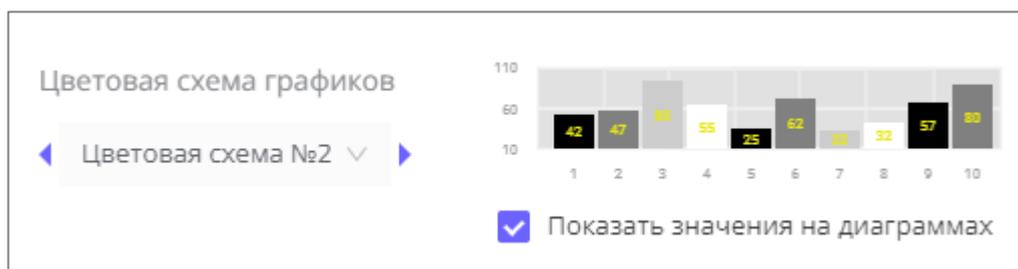


Рисунок 5.4 – «Цветовая схема графиков» в настройках анализа

- Функция «Показать значения на диаграммах» включает и выключает отображение цифровых значений на диаграммах (Рис. 5.5).

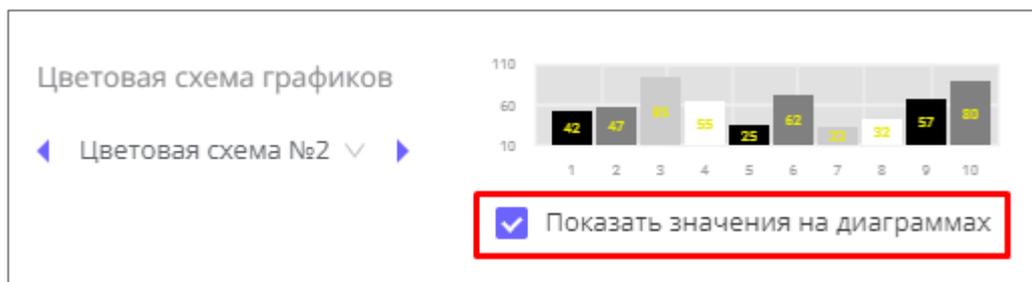


Рисунок 5.5 – Функция «Показать значения на диаграммах» в настройках анализа

- «Язык результата анализа» - выбор языка, на котором формируется анализ статистических данных в MS Word (Рис. 5.6).

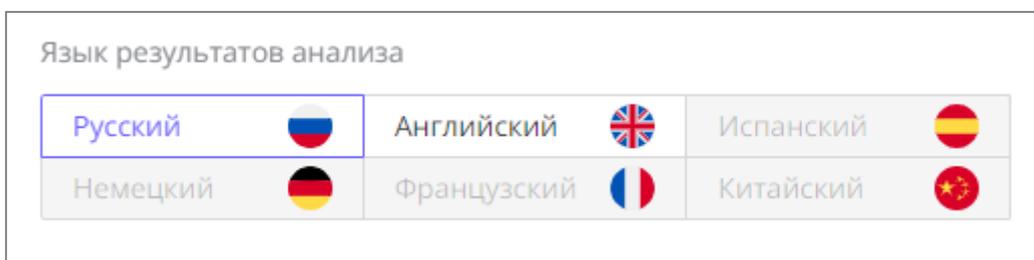


Рисунок 5.6 – Функция выбора «Язык результата анализа» в настройках анализа

- «Точность измерения p-value» - здесь указывается количество знаков после запятой (Рис. 5.7).

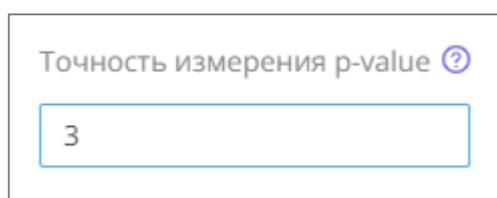
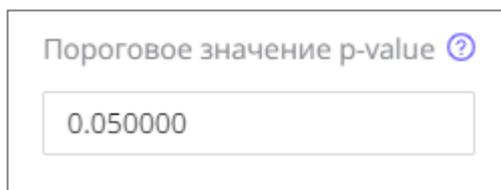


Рисунок 5.7 – «Точность измерения p-value» в настройках анализа

- «Пороговое значение p-value» - значение порога p-value, ниже которого показатели будут считаться статистически значимыми (Рис. 5.8).



Пороговое значение p-value ?

0.050000

Рисунок 5.8 – «Пороговое значение p-value» в настройках анализа

## ГЛАВА 6. ОПИСАТЕЛЬНАЯ СТАТИСТИКА

Описательная статистика – оценка выборки данных. Строится в зависимости от типа статистических данных.

1. Категориальные показатели описываются в таблице с указанием абсолютных значений и их процентных долей (Рис. 6.1).

Показатели	Категории	Абс.	%
Пол	Женский	13	65,0
	Мужской	7	35,0
Группа	Контрольная группа	10	50,0
	Основная группа	10	50,0

Рисунок 6.1 – Пример описательной статистики категориальных показателей

2. Количественные показатели описываются в зависимости от распределения. При нормальном распределении переменные описываются с помощью средней арифметической ( $M$ ), стандартного отклонения ( $\pm SD$ ), 95% ДИ (доверительного интервала). Также для описания переменных используются min-max значения (Рис. 6.2).

Показатель	$M \pm SD$	95% ДИ	n	min	max
ИМТ (полных лет)	$38,05 \pm 16,41$	30,37 – 45,73	20	15,60	65,00

Рисунок 6.2 – Пример описательной статистики количественных переменных с нормальным распределением

3. При распределении, отличном от нормального, переменные описываются с помощью медианы ( $Me$ ) и интерквартильного размаха ( $Q_1-Q_3$ ). Также для описания переменных используются min-max значения (Рис. 6.3).

Показатель	Me	$Q_1 - Q_3$	n	min	max
Возраст	48	44 – 56	99	22	79

Рисунок 6.3 – Пример описательной статистики количественных переменных с распределением, отличным от нормального

Кнопка описательной статистики находится над колонками (Рис. 6.4).

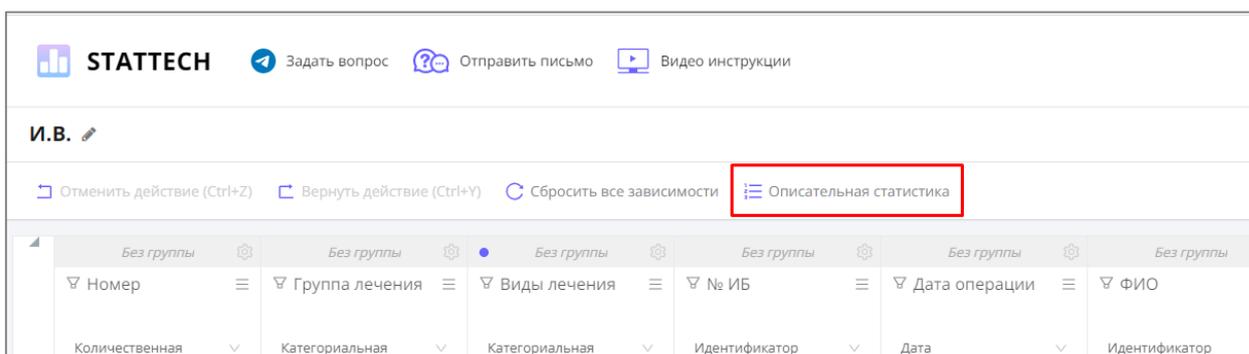


Рисунок 6.4 – Кнопка «Описательная статистика»

Нажав на кнопку «Описательная статистика» откроется перечень всех колонок. В данном разделе имеется возможность выбрать для описательной статистики либо несколько показателей, либо выбрать все (Рис. 6.5 и Рис. 6.6).

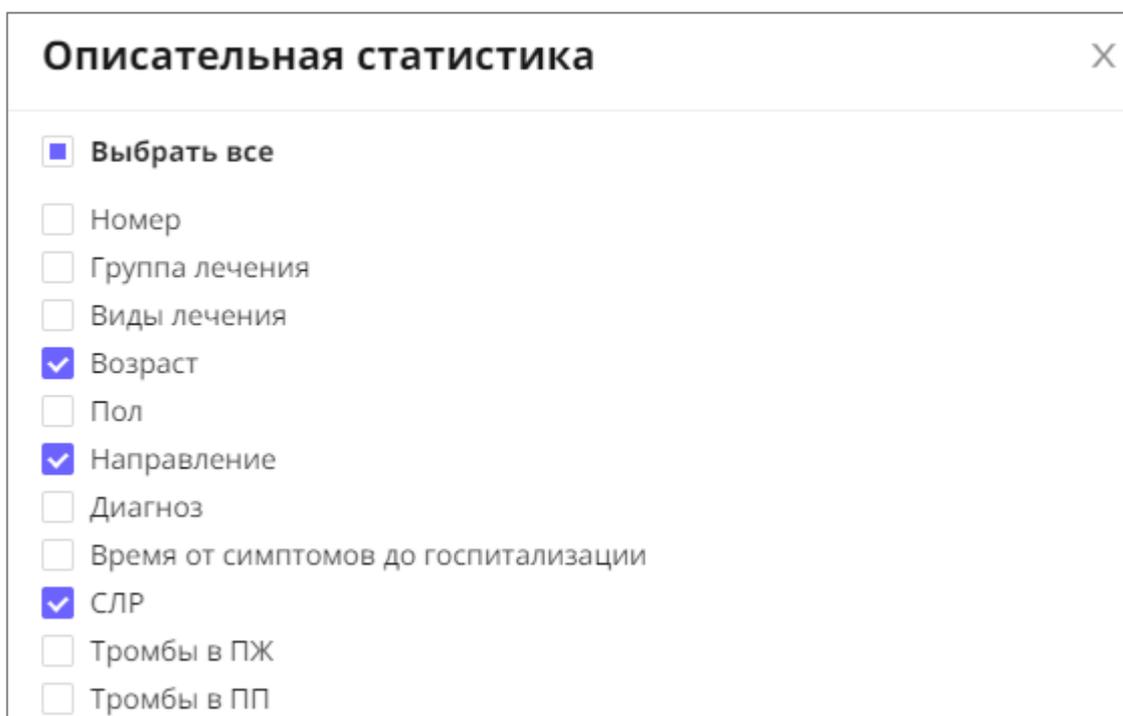


Рисунок 6.5 – Пример выбора несколько показателей для описательной статистики

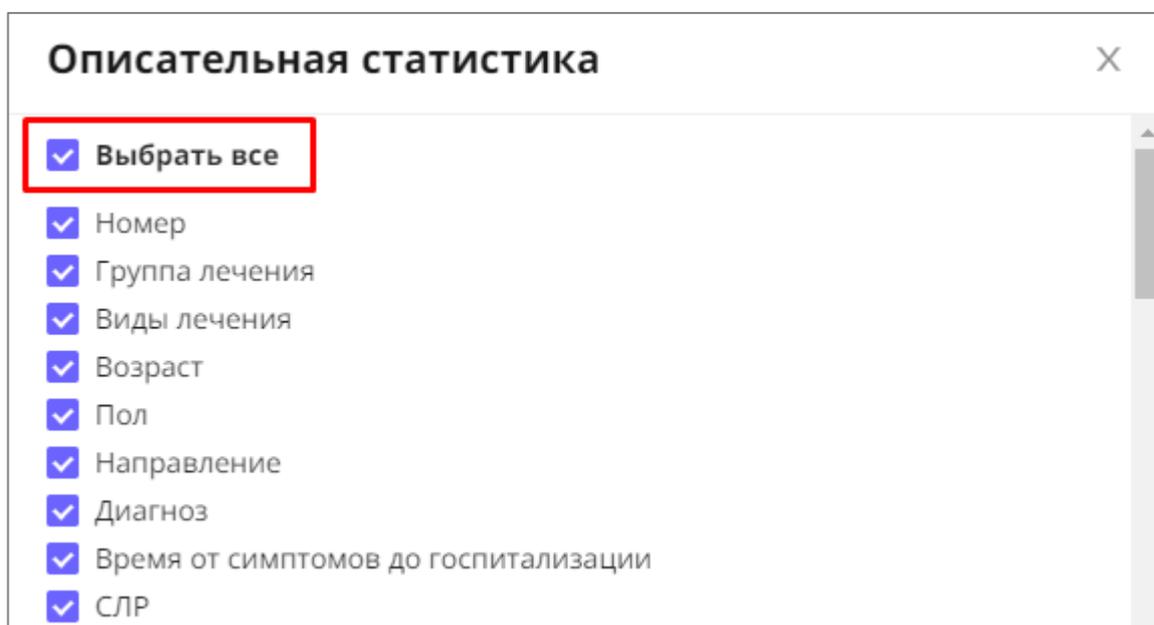


Рисунок 6.6 – Кнопка «Выбрать все» в разделе «Описательная статистика»

Если необходимо снять выбор показателей, то нужно нажать «Выбрать все» еще раз. И тогда все «галочки» снимутся.

После выбора всех необходимых показателей для описательной статистики нажимаем на кнопку «Ок». Далее необходимо нажать на кнопку «Выполнить анализ» в правом верхнем углу страницы (Рис. 6.7).

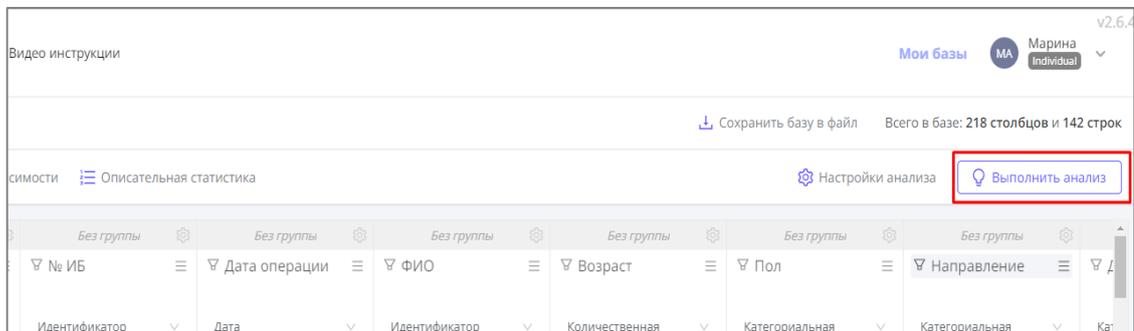


Рисунок 6.7 – Кнопка «Выполнить анализ»

## ГЛАВА 7. ПОСТРОЕНИЕ ЗАВИСИМОСТЕЙ МЕЖДУ ПОКАЗАТЕЛЯМИ

### 7.1. Сравнение двух групп по количественному признаку

Для построения зависимостей необходимо определить зависимый и независимый показатели. Например, если задача состоит в сравнении артериального давления у курящих и некурящих людей, то зависимый показатель – уровень систолического артериального давления (САД), независимый – наличие курения.

Анализ начинается от независимого показателя, в нашем случае – колонки «Курение». В базе данных на необходимой колонке нужно нажать знак шестеренки , и перейти во вкладку «Зависимости» (Рис. 7.1).

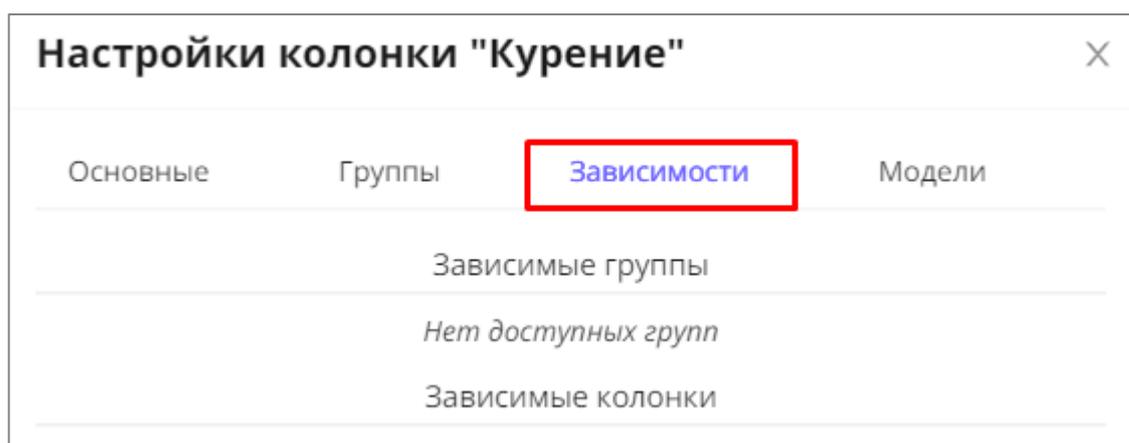


Рисунок 7.1 – Вкладка «Зависимости» в настройках колонки

В данной вкладке будут отображаться 2 раздела: «Зависимые группы», если группы созданы, и «Зависимые колонки».

Для того, чтобы сравнить «САД» в зависимости от наличия курения, необходимо в разделе «Зависимые колонки» проставить галочку на колонке «САД». После выбора зависимой колонки нажмите кнопку «ОК». (Рис. 7.2).

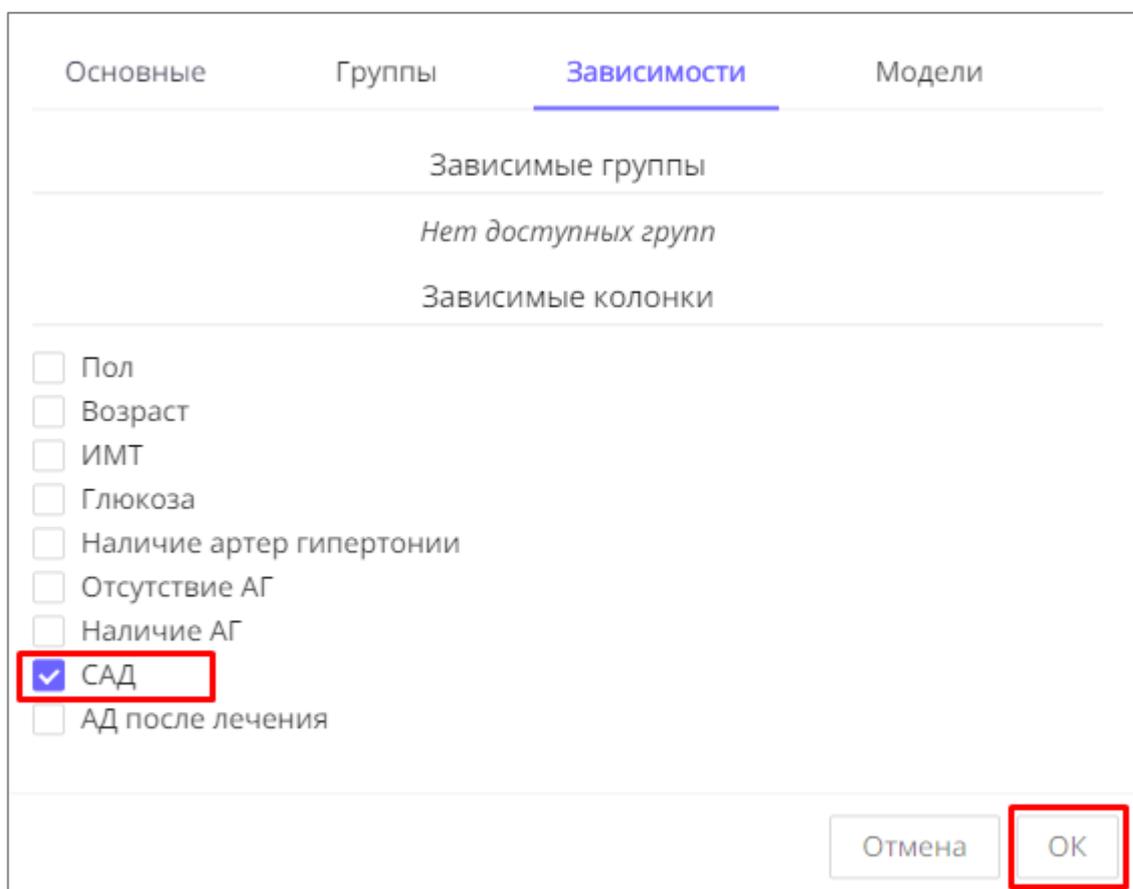


Рисунок 7.2 – Функционал выбора зависимого показателя

Для получения результатов анализа нажмите в верхнем правом углу кнопку «Выполнить анализ» (Рис. 7.3).

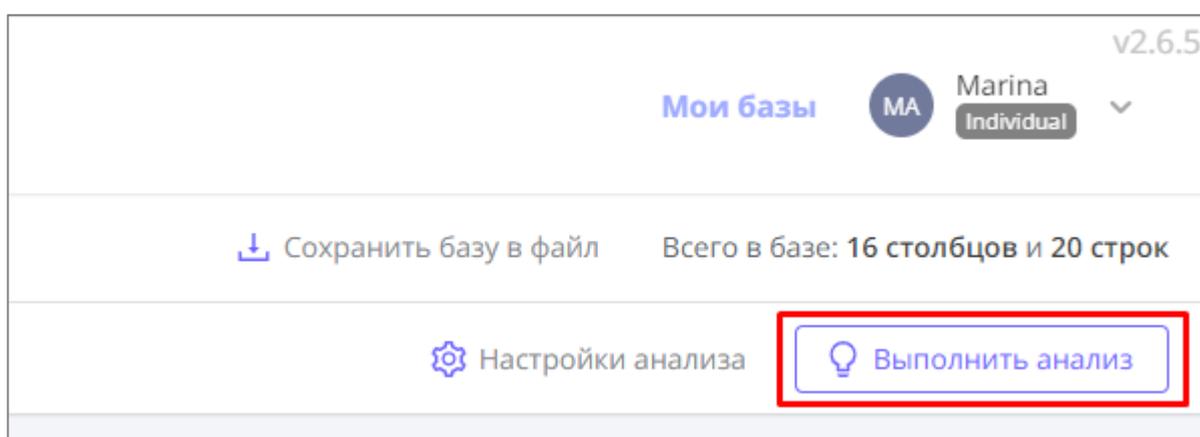


Рисунок 7.3 – Кнопка «Выполнить анализ»

После этого с правой стороны экрана появится выполненный анализ (Рис. 7.4).

STATTECH [Задать вопрос](#) [Отправить письмо](#) [Видео инструкции](#) v2.6.5

[Сохранить базу в файл](#) Всего в базе: 16 столбцов и 20 строк

[Выйти](#) [Настройки анализа](#) [Обновить](#) [Закрыть анализ](#)

ИМТ	Курение	Глюкоза	Наличие артер гипертонии	Отсутствие АГ	Наличие АГ
19.591836734693878	0 (Отсутствие)	4.7	0 (Отсутствие гипертон...	1 (Отсутствие АГ)	0 (Отсутствие)
21.952479338842977	0 (Отсутствие)	4.7	0 (Отсутствие гипертон...	1 (Отсутствие АГ)	0 (Отсутствие)
26.234567901234566	0 (Отсутствие)	4.7	0 (Отсутствие гипертон...	1 (Отсутствие АГ)	0 (Отсутствие)
26.234567901234566	1 (Наличие)	4.9	0 (Отсутствие гипертон...	1 (Отсутствие АГ)	0 (Отсутствие)
34.06360930832945	0 (Отсутствие)	6.5	0 (Отсутствие гипертон...	1 (Отсутствие АГ)	0 (Отсутствие)
28.834579099129364	0 (Отсутствие)	6.5	0 (Отсутствие гипертон...	1 (Отсутствие АГ)	0 (Отсутствие)
21.258503401360546	0 (Отсутствие)	5.1	0 (Отсутствие гипертон...	1 (Отсутствие АГ)	0 (Отсутствие)
21.258503401360546	0 (Отсутствие)	7.2	0 (Отсутствие гипертон...	1 (Отсутствие АГ)	0 (Отсутствие)
27.513384889946465	1 (Наличие)	5.8	1 (Наличие гипертонии)	0 (Наличие АГ)	1 (Наличие)
27.548209366391188	0 (Отсутствие)	5.2	1 (Наличие гипертонии)	0 (Наличие АГ)	1 (Наличие)
30.693877551020407	1 (Наличие)	6.4	1 (Наличие гипертонии)	0 (Наличие АГ)	1 (Наличие)
19.692665289256198	1 (Наличие)	5.7	1 (Наличие гипертонии)	0 (Наличие АГ)	1 (Наличие)
33.984374999999999	0 (Отсутствие)	6.9	1 (Наличие гипертонии)	0 (Наличие АГ)	1 (Наличие)
26.296566837107374	1 (Наличие)	5.9	1 (Наличие гипертонии)	0 (Наличие АГ)	1 (Наличие)

Анализ выполнен успешно [Скачать графики](#) [Скачать Word](#)

StatTech 2.6.5  
© ООО "Статтех", Россия, 2020  
Программа зарегистрирована Федеральной службой по интеллектуальной собственности, номер регистрации 2020615715, дата регистрации 29.05.2020

Пользователь: Marina  
Почта: marina-zabolotsk@mail.ru

**Анализ данных: «ДЕМО»**

**Описание методов статистического анализа**

Статистический анализ проводился с использованием программы StatTech v. 2.6.5 (разработчик - ООО "Статтех", Россия). Количественные показатели оценивались на предмет соответствия нормальному распределению с помощью критерия Шапиро-Уилка (при числе исследуемых менее 50) или критерия Колмогорова-Смирнова (при числе исследуемых более 50). В случае отсутствия нормального распределения количественные данные описывались с помощью медианы (Me) и нижнего и верхнего квартилей (Q1 – Q3).

Рисунок 7.4 – Сформированный анализ данных в StatTech

Полученные результаты можно скачать в формате .DOCX (MS Word), нажав на соответствующую кнопку (Рис.7.5).

[Сохранить базу в файл](#) Всего в базе: 16 столбцов и 20 строк

[Настройки анализа](#) [Обновить](#) [Закрыть анализ](#)

✓ Анализ выполнен успешно [Скачать графики](#) [Скачать Word](#)

StatTech 2.6.5  
© ООО "Статтех", Россия, 2020  
Программа зарегистрирована Федеральной службой по интеллектуальной собственности, номер регистрации 2020615715, дата регистрации 29.05.2020

Рисунок 7.5 – Кнопка «Скачать Word» в сформированном анализе данных

Также есть возможность скачать графики в высоком разрешении, нажав соответствующую кнопку. Диаграммы скачиваются в одном архиве с расширением .ZIP (Рис. 7.6).

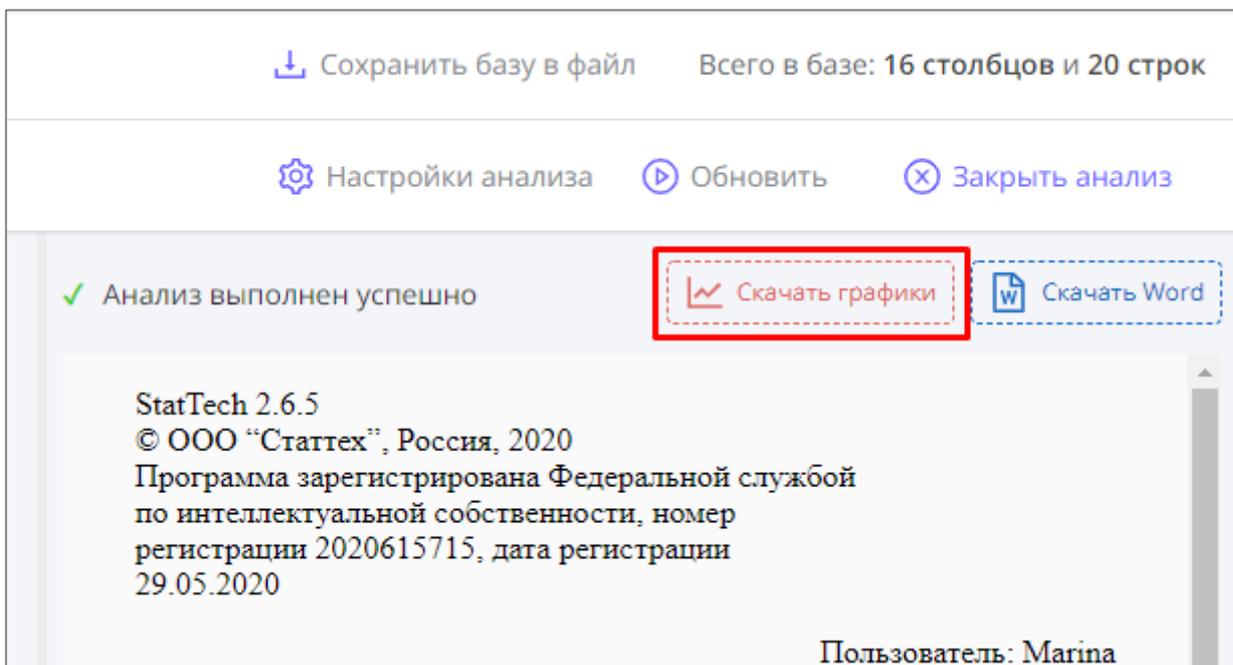


Рисунок 7.6 – Кнопка «Скачать графики» в сформированном анализе данных

## 7.2. Общая структура результатов анализа

Результат анализа данных состоит из нескольких частей:

- Описание методов статистического анализа (Рис. 7.7);
- Таблица с анализом зависимостей показателей;
- Вывод по результатам анализа зависимости показателей с указанием используемого метода;
- Диаграмма.

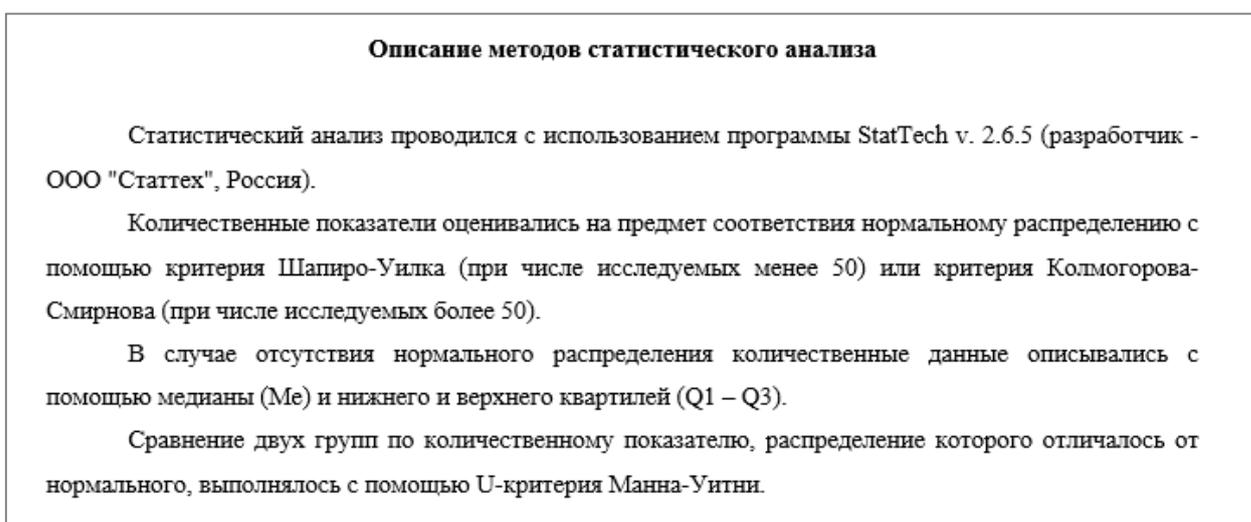


Рисунок 7.7 – Пример описания методов статистического анализа

Данные по статистическому анализу выводятся в таблицу (Рис. 7.8). Нумерация таблиц и диаграмм в сформированном анализе идёт последовательно.

Был проведен анализ показателя "САД" в зависимости от курения.

Таблица 1 – Анализ показателя "САД" в зависимости от курения

Показатель	Категории	САД			p
		Me	Q <sub>1</sub> – Q <sub>3</sub>	n	
Курение	Отсутствие	134	122 – 143	10	0,009*
	Наличие	180	163 – 190	10	

\* – различия показателей статистически значимы ( $p < 0,05$ )

В соответствии с представленной таблицей при сравнении показателя "САД" в зависимости от курения, были установлены существенные различия ( $p = 0,009$ ) (используемый метод: U-критерий Манна-Уитни).

Рисунок 7.8 – Пример вывода статистического анализа показателя «САД» в зависимости от курения

Под таблицей выводится диаграмма (Рис. 7.9). В зависимости от распределения можно увидеть или внутрискладовую диаграмму, или ящичную диаграммы.

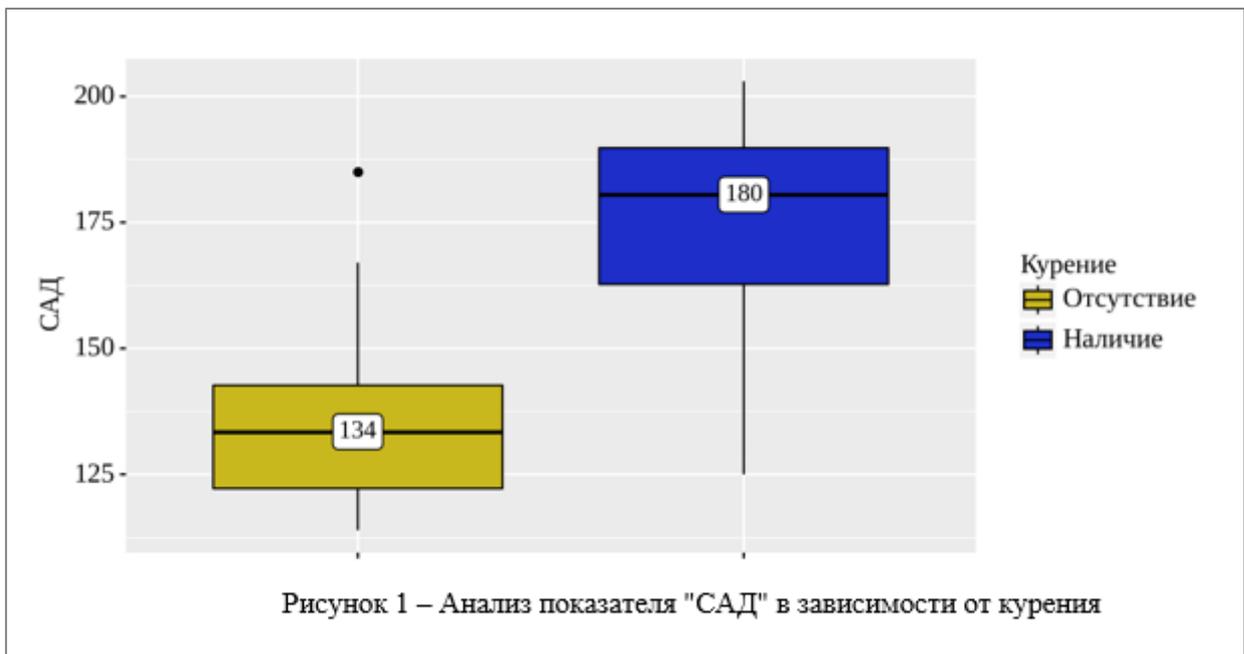


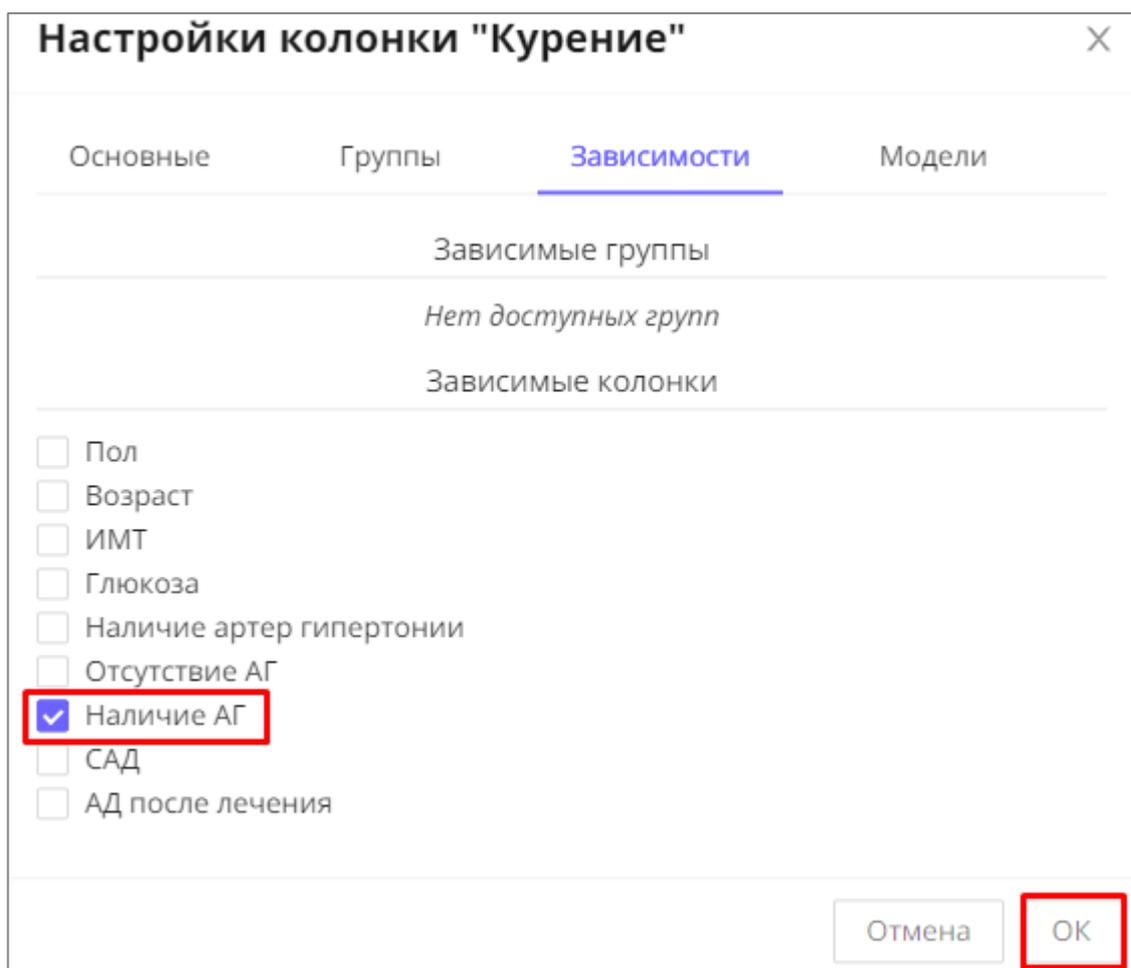
Рисунок 7.9 – Пример диаграммы в статистическом анализе

### 7.3. Сравнение двух групп по категориальному признаку

Для начала надо определить зависимый и независимый показатель. Например, если рассматривать наличие артериальной гипертензии (АГ) в зависимости от курения, то зависимый показатель – наличие АГ, а независимый – наличие курения.

Анализ начинается от независимого показателя. В базе данных на необходимой колонке нужно нажать знак шестеренки , и перейти во вкладку «Зависимости» (Рис. 7.1).

Для того, чтобы сравнить «Наличие АГ» в зависимости от наличия курения, необходимо в разделе «Зависимые колонки» проставить галочку на колонке «Наличие АГ». После выбора зависимой колонки нажмите кнопку «ОК». (Рис. 7.10).



## Рисунок 7.10 - Функционал выбора зависимого показателя

Для получения результатов анализа нажмите в верхнем правом углу кнопку «Выполнить анализ» (Рис. 7.3).

После этого с правой стороны экрана появится выполненный анализ (Рис. 7.4).

В связи с тем, что мы анализировали зависимость одного категориального признака от другого, то используемые методы и способы описания будут отличны от анализа количественных показателей (Рис. 7.11).

Статистический анализ проводился с использованием программы StatTech v. 2.6.5 (разработчик - ООО "Статтех", Россия).

Категориальные данные описывались с указанием абсолютных значений и процентных долей.

Сравнение процентных долей при анализе четырехпольных таблиц сопряженности выполнялось с помощью точного критерия Фишера (при значениях ожидаемого явления менее 10)

Был выполнен анализ показателя "Наличие АГ" в зависимости от курения.

Таблица 1 – Анализ показателя "Наличие АГ" в зависимости от курения

Показатель	Категории	Курение		p
		Отсутствие	Наличие	
Наличие АГ	Отсутствие	7 (70,0)	1 (10,0)	0,020*
	Наличие	3 (30,0)	9 (90,0)	

\* – различия показателей статистически значимы ( $p \leq 0,05$ )

Согласно полученным данным при сравнении показателя "Наличие АГ" в зависимости от курения, были установлены существенные различия ( $p = 0,020$ ) (используемый метод: Точный критерий Фишера).

Рисунок 7.11 – Пример вывода статистического анализа «Наличия АГ» в зависимости от курения

Под выводом можно увидеть диаграмму с выводом расчета шансов «наличия АГ» в зависимости от курения (Рис. 7.12).

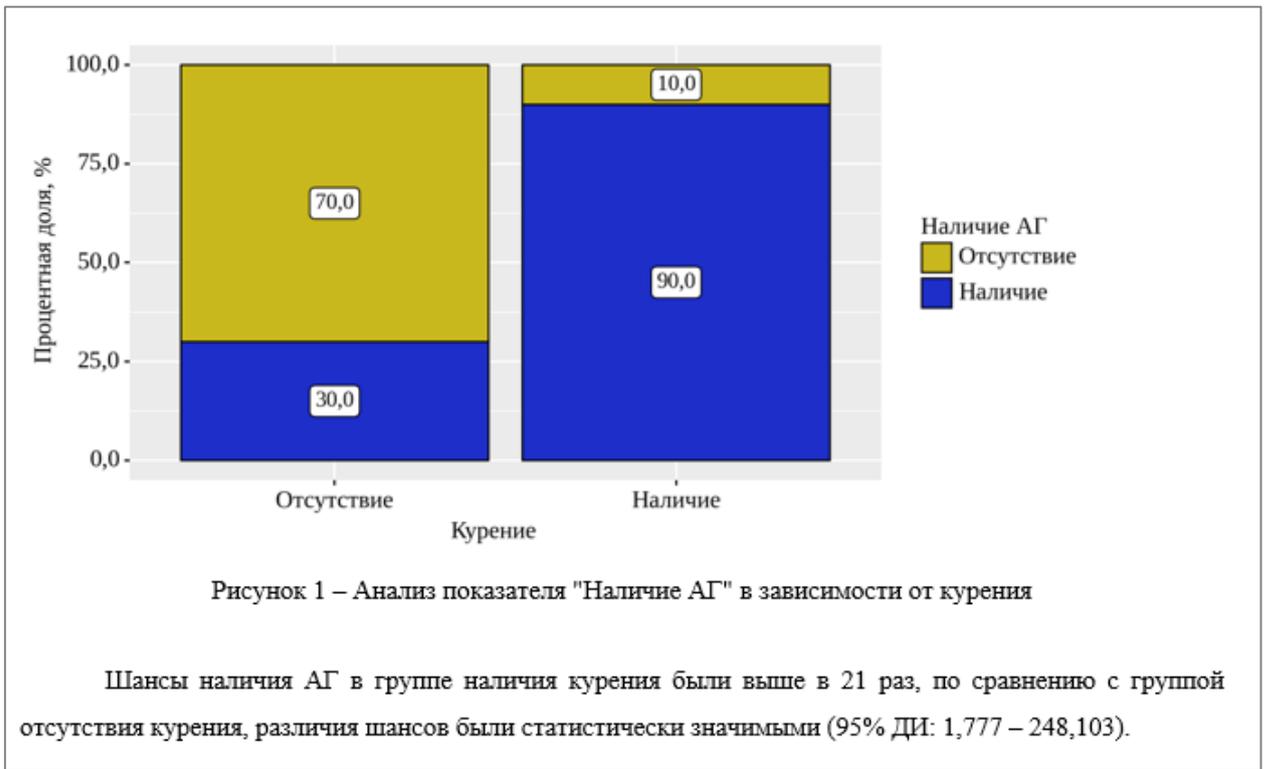


Рисунок 7.12 – Пример диаграммы и вывода в сформированном анализе

#### 7.4. Сравнение трех и более групп по количественному признаку

Сначала надо определить зависимый и независимый показатель. Например, если рассматривать степень индекса масса тела (ИМТ) в зависимости от уровня артериального давления, то зависимый показатель – уровень систолического артериального давления (САД), а независимый – степень ИМТ.

Для того, чтобы построить зависимость степени индекса массы тела, колонку со значениями ИМТ надо преобразовать, присвоив каждому значению свою степень (см. Раздел 4.3.2).

Анализ начинается от независимого показателя. В базе данных на необходимой колонке нужно нажать знак шестеренки , и перейти во вкладку «Зависимости» (Рис. 7.1).

Для того, чтобы сравнить «САД» в зависимости от степени ИМТ, необходимо в разделе «Зависимые колонки» проставить галочку на колонке «САД». После выбора зависимой колонки нажмите кнопку «ОК» (Рис. 7.2).

Для получения результатов анализа нажмите в верхнем правом углу кнопку «Выполнить анализ» (Рис. 7.3).

После этого с правой стороны экрана появится выполненный анализ (Рис. 7.4).

В связи с тем, что мы анализировали зависимость нескольких групп по количественному признаку, то используемые методы и способы описания будут отличны от анализа других показателей (Рис. 7.13).

Таблица 1 – Анализ показателя "САД" в зависимости от показателя "ИМТ"

Показатель	Категории	САД			p
		Me	Q <sub>1</sub> – Q <sub>3</sub>	n	
ИМТ	Норма или дефицит МТ	135	118 – 136	5	0,010* p <sub>3-1</sub> = 0,021 p <sub>3-2</sub> = 0,041
	Избыточная МТ	138	127 – 145	6	
	Ожирение	185	167 – 190	9	

\* – различия показателей статистически значимы ( $p \leq 0,05$ )

При сопоставлении показателя "САД" в зависимости от показателя "ИМТ", нами были выявлены статистически значимые различия ( $p = 0,010$ ) (используемый метод: Критерий Краскела–Уоллиса).

Рисунок 7.13 – Пример сравнения трёх групп по количественному признаку

После результатов анализа можно увидеть диаграмму, отображающую статистические данные (Рис. 7.14).

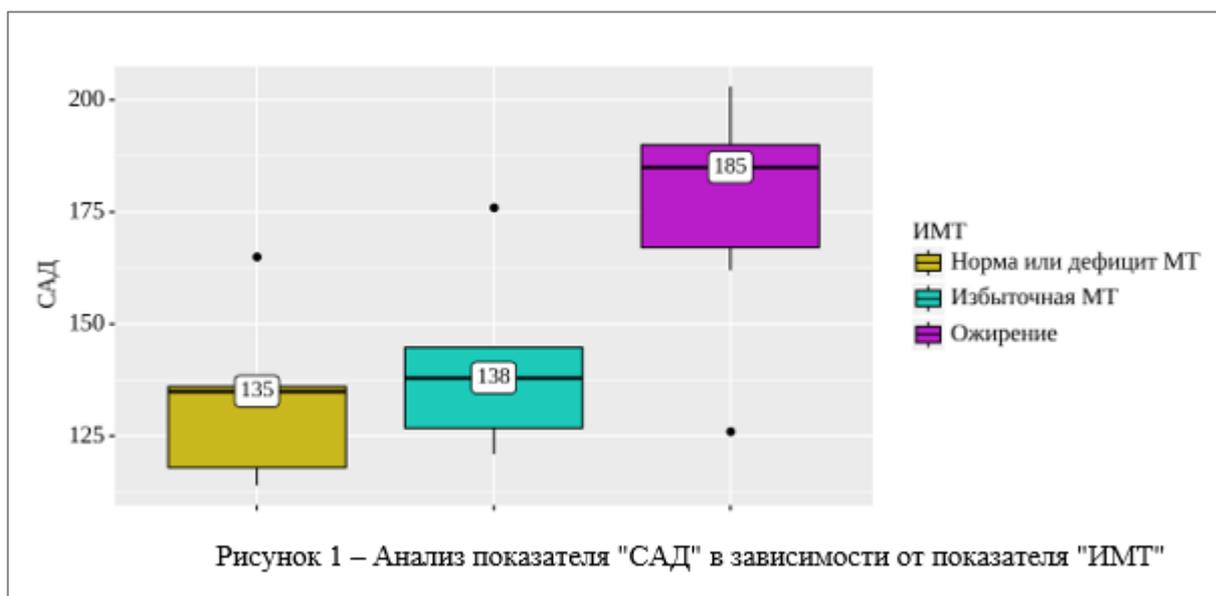


Рисунок 7.14 – Пример графика сравнения трёх групп по количественному признаку

## 7.5. Сравнение трех и более групп по категориальному признаку

Сначала надо определить зависимый и независимый показатель. Например, если рассматривать степень индекса масса тела (ИМТ) в зависимости от наличия артериальной гипертензии, то зависимый показатель – наличие артериальной гипертензии («Наличие АГ»), а независимый – степень ИМТ.

Для того, чтобы построить зависимость степени индекса массы тела, колонку со значениями ИМТ надо преобразовать, присвоив каждому значению свою степень (см. Раздел 4.3.2).

Анализ начинается от независимого показателя. В базе данных на необходимой колонке нужно нажать знак шестеренки , и перейти во вкладку «Зависимости» (Рис. 7.1).

Для того, чтобы сравнить «Наличие АГ» в зависимости от степени ИМТ, необходимо в разделе «Зависимые колонки» проставить галочку на колонке «Наличие АГ». После выбора зависимой колонки нажмите кнопку «ОК» (Рис. 7.10).

Для получения результатов анализа нажмите в верхнем правом углу кнопку «Выполнить анализ» (Рис. 7.3).

После этого с правой стороны экрана появится выполненный анализ (Рис. 7.4).

В связи с тем, что мы анализировали зависимость нескольких групп по категориальному признаку, то используемые методы и способы описания будут отличны от анализа других показателей (Рис. 7.15).

Таблица 1 – Анализ показателя "Наличие АГ" в зависимости от показателя "ИМТ"

Показатель	Категории	ИМТ			p
		Норма или дефицит МТ	Избыточная МТ	Ожирение	
Наличие АГ	Отсутствие	4 (80,0)	3 (50,0)	1 (11,1)	0,035* p <sub>1-3</sub> = 0,030
	Наличие	1 (20,0)	3 (50,0)	8 (88,9)	

\* – различия показателей статистически значимы ( $p \leq 0,05$ )

При сопоставлении показателя "Наличие АГ" в зависимости от показателя "ИМТ", были выявлены статистически значимые различия ( $p = 0,035$ ) (используемый метод: Хи-квадрат Пирсона).

Рисунок 7.15 – Пример сравнения трёх групп по категориальному признаку

После результатов анализа можно увидеть диаграмму, отображающую статистические данные (Рис. 7.16).

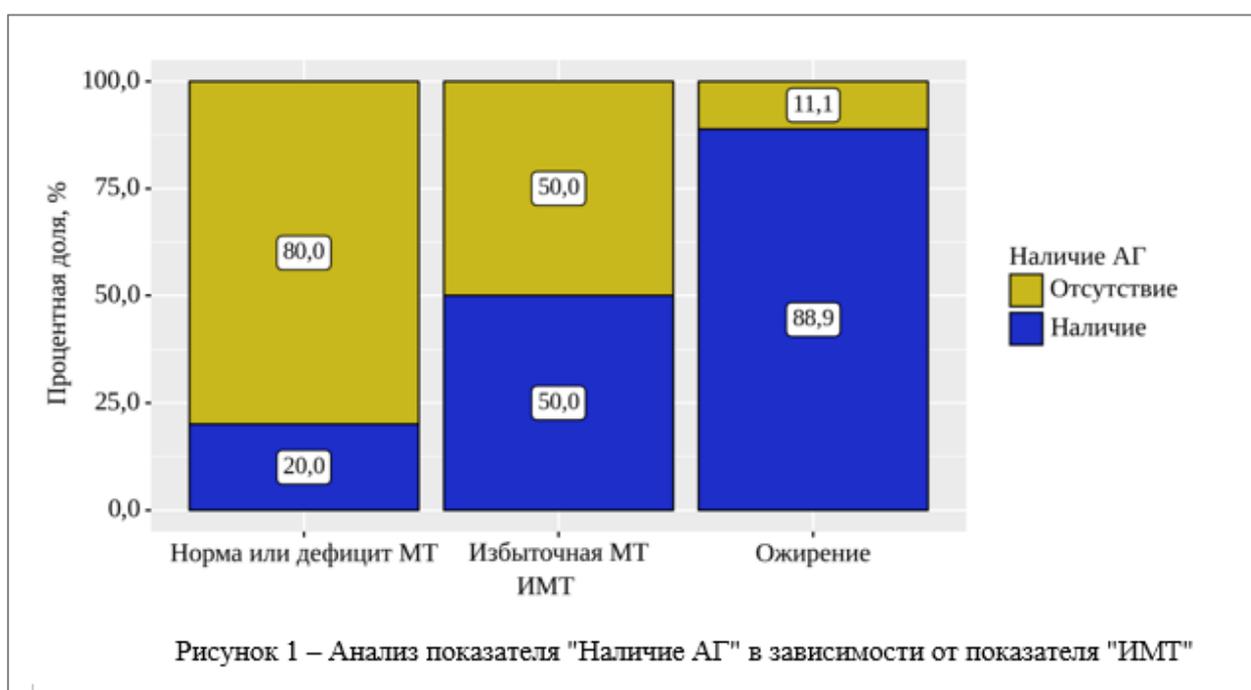


Рисунок 7.16 – Пример графика сравнения трёх групп по категориальному признаку

## 7.6. Корреляционный анализ

Для проведения корреляционного анализа возьмем построение зависимости уровня артериального давления (САД) от возраста. Оба показателя являются количественными переменными.

Для построения зависимостей необходимо определить зависимый и независимый показатели. В данном случае зависимый показатель – уровень систолического артериального давления (САД), независимый – возраст.

В базе данных на зависимой колонке нужно нажать знак шестеренки , и перейти во вкладку «Зависимости» (Рис. 7.1).

Для того, чтобы сравнить «САД» в зависимости от возраста, необходимо в разделе «Зависимые колонки» проставить галочку на колонке «САД». После выбора зависимой колонки нажмите кнопку «ОК». (Рис. 7.2).

Для получения результатов анализа нажмите в верхнем правом углу кнопку «Выполнить анализ» (Рис. 7.3).

После этого с правой стороны экрана появится выполненный анализ (Рис. 7.4).

Вывод данных по корреляционному анализу будет зависеть от распределения показателей (нормальное и «ненормальное») (Рис. 7.17 и Рис. 7.18).

Показатель	Характеристика корреляционной связи		
	ρ	Теснота связи по шкале Чеддока	ρ
САД – Возраст	0,626	Заметная	0,003*

\* – различия показателей статистически значимы ( $p < 0,05$ )

Наблюдаемая зависимость возраста от показателя "САД" описывается уравнением парной линейной регрессии:

$$Y_{\text{Возраст}} = 0,317 \times X_{\text{САД}} - 10,806$$

При увеличении показателя "САД" на 1 мм.рт.ст следует ожидать увеличение возраста на 0,317 полных лет. Полученная модель объясняет 37,9% наблюдаемой дисперсии возраста.

Рисунок 7.17 – Пример результатов корреляционного анализа при «ненормальном» распределении

Таблица 1 – Результаты корреляционного анализа взаимосвязи показателя "САД" и возраста

Показатель	Характеристика корреляционной связи		
	$r_{xy}$	Теснота связи по шкале Чеддока	p
САД – Возраст	0,616	Заметная	0,004*

\* – различия показателей статистически значимы ( $p < 0,05$ )

Наблюдаемая зависимость возраста от показателя "САД" описывается уравнением парной линейной регрессии:

$$Y_{\text{Возраст}} = 0,317 \times X_{\text{САД}} - 10,806$$

При увеличении показателя "САД" на 1 мм.рт.ст следует ожидать увеличение возраста на 0,317 полных лет. Полученная модель объясняет 37,9% наблюдаемой дисперсии возраста.

Рисунок 7.17 – Пример результатов корреляционного анализа при нормальном распределении

После результатов анализа можно увидеть диаграмму, отображающую статистические данные (Рис. 7.18).

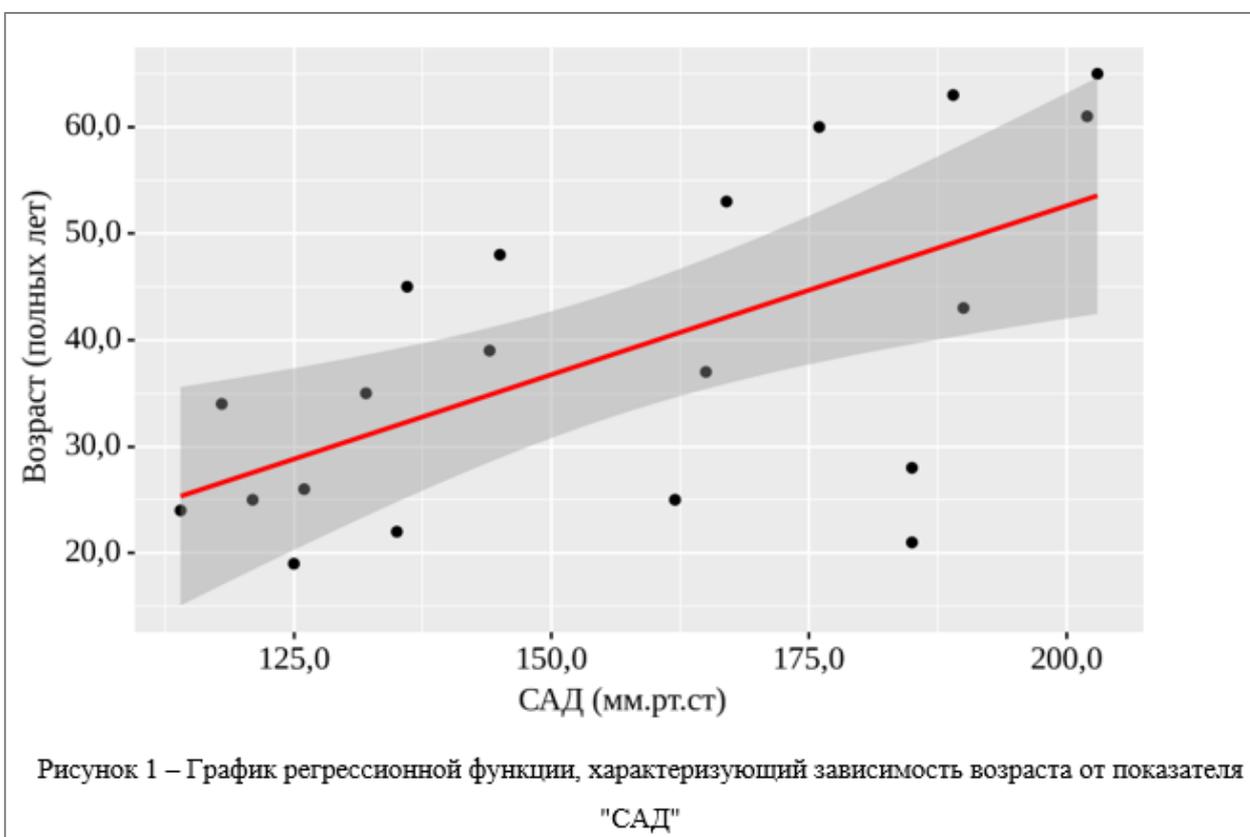


Рисунок 7.18 – Пример графика регрессионной функции

## 7.7. Roc – Анализ

ROC – анализ используется для оценки диагностической значимости *количественных признаков* при прогнозировании определенного исхода.

Для примера, возьмем зависимость уровня артериального давления (САД) от пола.

Для построения зависимостей необходимо определить зависимый и независимый показатели. В данном случае зависимый показатель – уровень систолического артериального давления (САД), независимый – пол.

В базе данных на зависимой колонке нужно нажать знак шестеренки , и перейти во вкладку «Зависимости» (Рис. 7.1).

Для того, чтобы сравнить «САД» в зависимости от пола, необходимо в разделе «Зависимые колонки» проставить галочку на колонке «САД». После выбора зависимой колонки нажмите кнопку «ОК». (Рис. 7.2).

Для получения результатов анализа нажмите в верхнем правом углу кнопку «Выполнить анализ» (Рис. 7.3).

После этого с правой стороны экрана появится выполненный анализ (Рис. 7.4).

Шаблон ROC-анализа отличается по своей структуре от других видов построения зависимостей между показателями. Структура выполненного ROC – анализа состоит из:

- Описание методов статистического анализа (Рис. 7.19).

Количественные показатели оценивались на предмет соответствия нормальному распределению с помощью критерия Шапиро-Уилка (при числе исследуемых менее 50) или критерия Колмогорова-Смирнова (при числе исследуемых более 50).

Количественные показатели, имеющие нормальное распределение, описывались с помощью средних арифметических величин (M) и стандартных отклонений (SD), границ 95% доверительного интервала (95% ДИ).

Сравнение двух групп по количественному показателю, имеющему нормальное распределение, при условии равенства дисперсий выполнялось с помощью t-критерия Стьюдента.

Для оценки диагностической значимости количественных признаков при прогнозировании определенного исхода, применялся метод анализа ROC-кривых. Разделяющее значение количественного признака в точке cut-off определялось по наивысшему значению индекса Юдена.

Рисунок 7.19 – Пример описания методов статистического анализа данных

- Таблица анализа показателей (Рис. 7.20).

Таблица 1 – Анализ показателя "САД" в зависимости от показателя "Пол"

Показатель	Категории	САД (мм.рт.ст)			p
		M ± SD	95% ДИ	n	
Пол	Женщины	150 ± 28	131 – 169	11	0,325
	Мужчины	163 ± 31	139 – 188	9	

При анализе показателя "САД" в зависимости от показателя "Пол", нам не удалось установить статистически значимых различий (p = 0,325) (используемый метод: t-критерий Стьюдента).

Рисунок 7.20 – Пример табличной формы анализа показателей

- Диаграмма анализа показателей (Рис. 7.21).

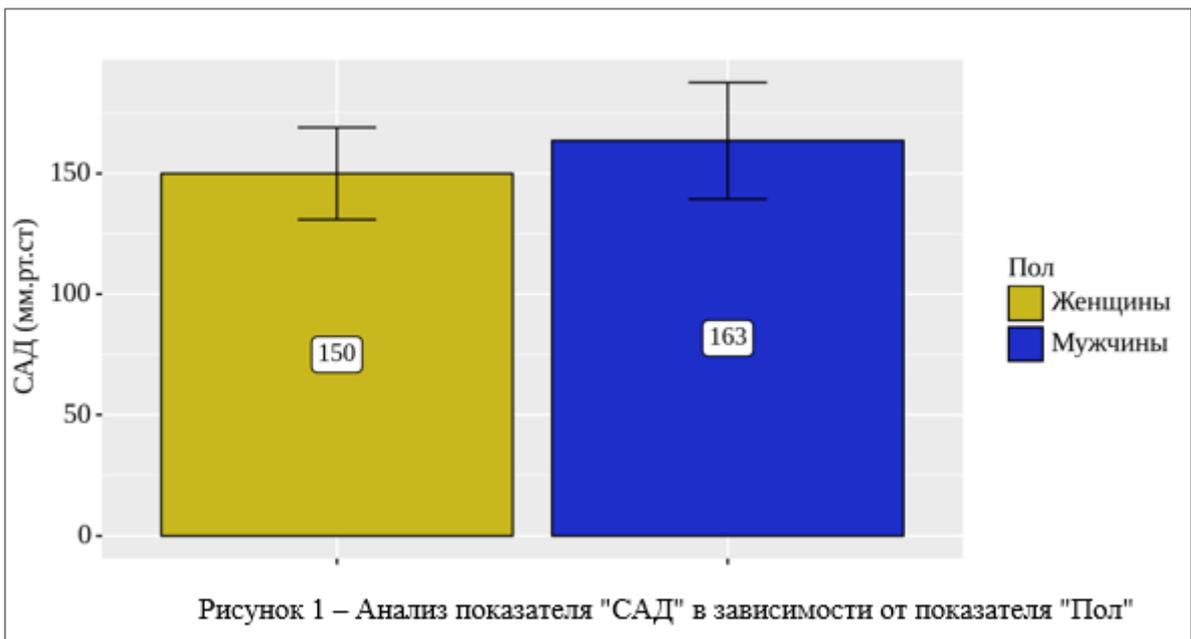


Рисунок 7.21 – Пример диаграммы анализа показателей

- ROC – кривая (Рис. 7.22).

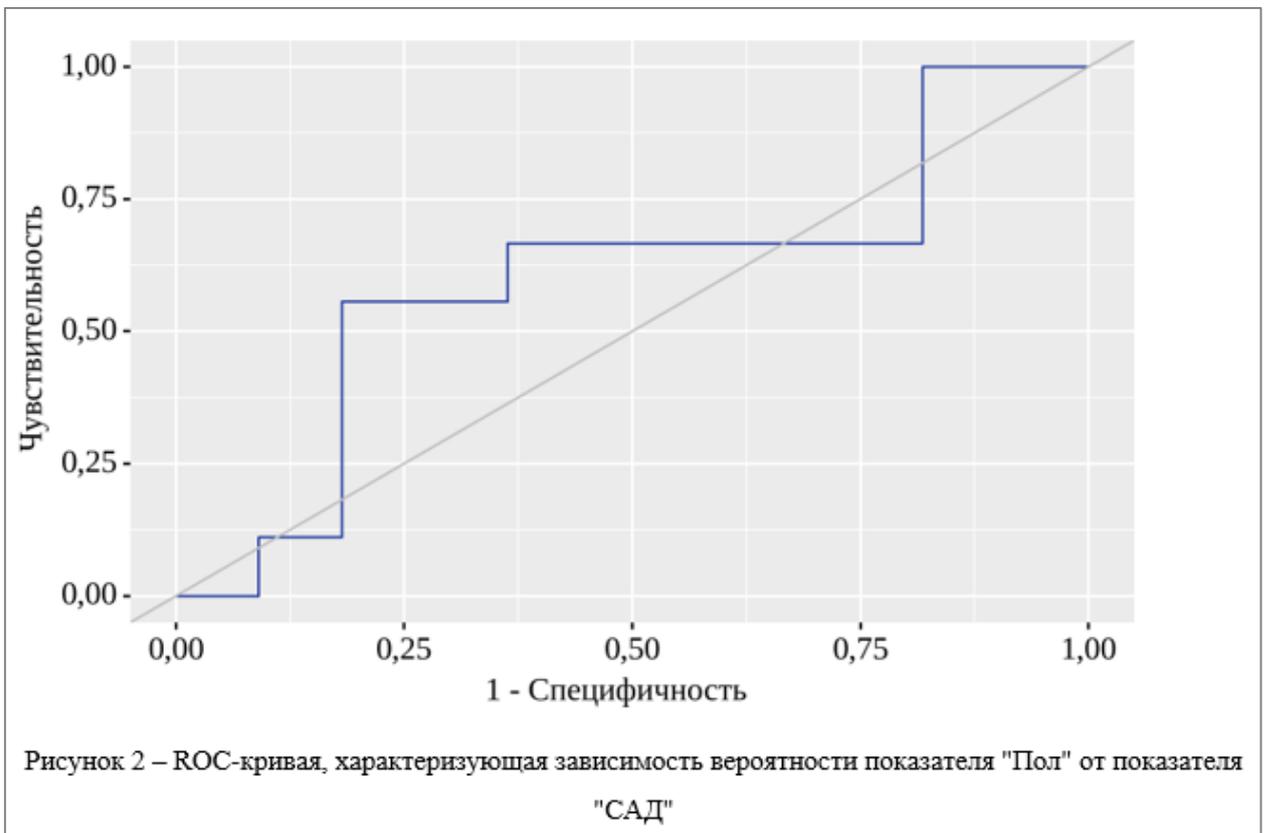


Рисунок 7.22 – Пример ROC – кривой

- График специфичности и чувствительности (Рис. 7.23).

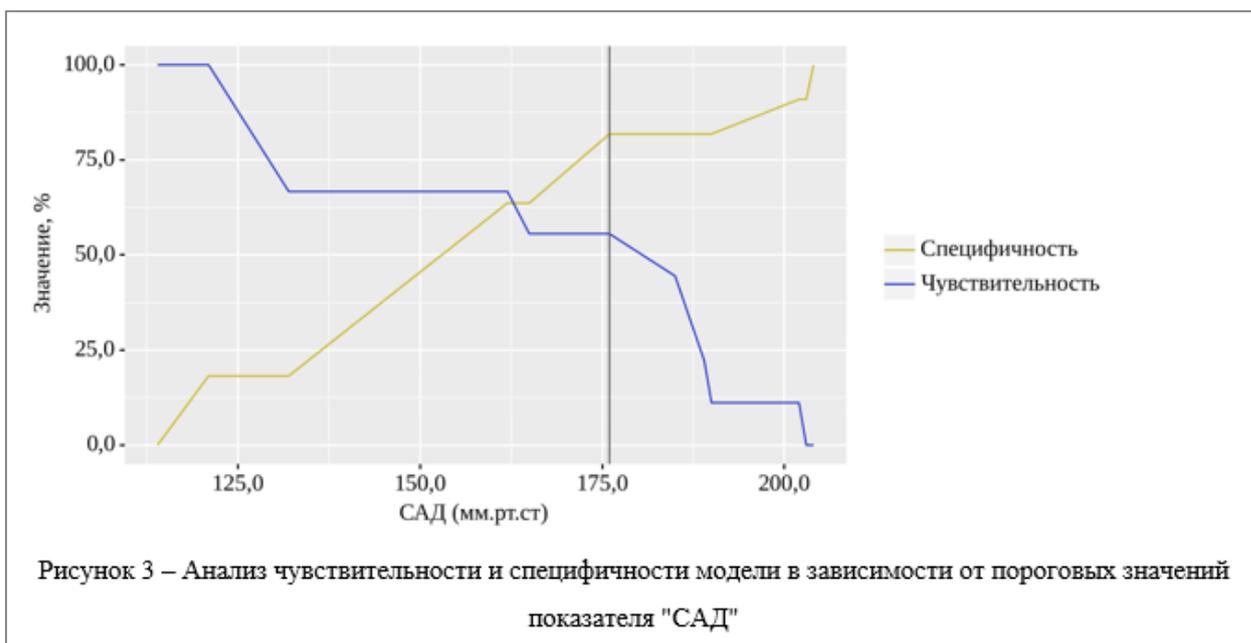


Рисунок 7.23 – Примера графика анализа специфичности и чувствительности модели

- Таблица расчета пороговых значений. PPV. NPV. (Рис. 7.24).

Таблица 2 – Пороговые значения показателя "САД"

Порог	Чувствительность (Se), %	Специфичность (Sp), %	PPV	NPV
176	55,6	81,8	71,4	69,2
165	55,6	63,6	55,6	63,6
<b>162</b>	<b>66,7</b>	<b>63,6</b>	<b>60,0</b>	<b>70,0</b>

Рисунок 7.24 – Пример пороговых значений

- Вывод (Рис. 7.25).

Площадь под ROC-кривой составила  $0,596 \pm 0,131$  с 95% ДИ: 0,340 – 0,852. Полученная модель не была статистически значимой ( $p = 0,470$ ).

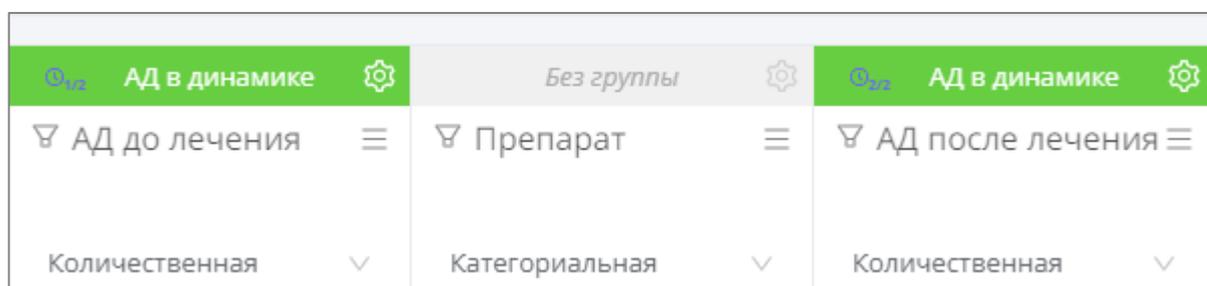
Пороговое значение показателя "САД" в точке cut-off, которому соответствовало наивысшее значение индекса Юлена, составило 176 мм.рт.ст. Мужчины прогнозировалось при значении показателя "САД" выше данной величины или равном ей. Чувствительность и специфичность модели составили 55,6% и 81,8%, соответственно.

Рисунок 7.25 – Пример вывода ROC-анализа

## 7.8. Анализ данных в формате «До - После»

### 7.8.1. Сравнение показателей в формате «До – После» без деления на группы

Для того, чтобы проанализировать два показателя в формате «До – После» необходимо создать группу показателей (см. Раздел 4.2). Необходимые колонки отнести к этой группе, для примера, возьмем показатели уровня артериального давления до и после лечения (Рис. 7.26).



АД в динамике	Без группы	АД в динамике
АД до лечения	Препарат	АД после лечения
Количественная	Категориальная	Количественная

Рисунок 7.26 – Пример распределения показателей к группам в формате «До – После»

После того как необходимые показатели отнесены к группам необходимо нажать кнопку «Выполнить анализ» в правом верхнем углу экрана (Рис. 7.27).

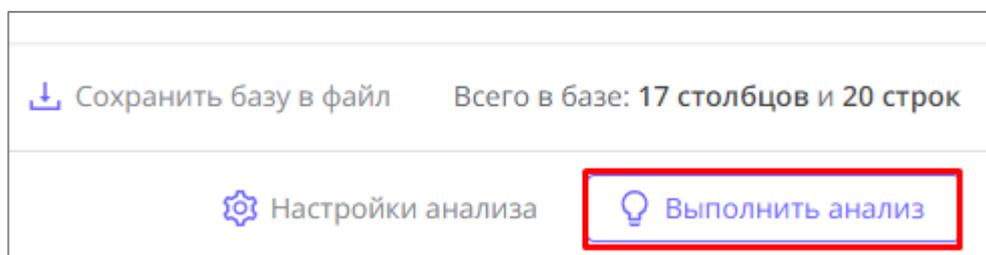


Рисунок 7.27 – Кнопка «Выполнить анализ»

После этого с правой стороны экрана появится выполненный анализ (Рис. 7.28).

Без группы	Без группы	Без группы	Без группы	Без группы	Без группы
ИМТ	Курение	Глюкоза	Наличие артер гипертонии	Отсутствие АГ	Наличие АГ
Категориальная	Категориальная	Количественная	Категориальная	Категориальная	Категориальная
1 (Норма или дефицит ...)	0 (Отсутствие)	4.7	0 (Отсутствие гипертон...	1 (Отсутствие АГ)	0 (Отсутствие)
1 (Норма или дефицит ...)	0 (Отсутствие)	4.7	0 (Отсутствие гипертон...	1 (Отсутствие АГ)	0 (Отсутствие)
2 (Избыточная МТ)	0 (Отсутствие)	4.7	0 (Отсутствие гипертон...	1 (Отсутствие АГ)	0 (Отсутствие)
2 (Избыточная МТ)	1 (Наличие)	4.9	0 (Отсутствие гипертон...	1 (Отсутствие АГ)	0 (Отсутствие)
3 (Ожирение)	0 (Отсутствие)	6.5	0 (Отсутствие гипертон...	1 (Отсутствие АГ)	0 (Отсутствие)
2 (Избыточная МТ)	0 (Отсутствие)	6.5	0 (Отсутствие гипертон...	1 (Отсутствие АГ)	0 (Отсутствие)
1 (Норма или дефицит ...)	0 (Отсутствие)	5.1	0 (Отсутствие гипертон...	1 (Отсутствие АГ)	0 (Отсутствие)
1 (Норма или дефицит ...)	0 (Отсутствие)	7.2	0 (Отсутствие гипертон...	1 (Отсутствие АГ)	0 (Отсутствие)
2 (Избыточная МТ)	1 (Наличие)	5.8	1 (Наличие гипертонии)	0 (Наличие АГ)	1 (Наличие)
2 (Избыточная МТ)	0 (Отсутствие)	5.2	1 (Наличие гипертонии)	0 (Наличие АГ)	1 (Наличие)
3 (Ожирение)	1 (Наличие)	6.4	1 (Наличие гипертонии)	0 (Наличие АГ)	1 (Наличие)
1 (Норма или дефицит ...)	1 (Наличие)	5.7	1 (Наличие гипертонии)	0 (Наличие АГ)	1 (Наличие)
3 (Ожирение)	0 (Отсутствие)	6.9	1 (Наличие гипертонии)	0 (Наличие АГ)	1 (Наличие)
2 (Избыточная МТ)	1 (Наличие)	5.9	1 (Наличие гипертонии)	0 (Наличие АГ)	1 (Наличие)
3 (Ожирение)	0 (Отсутствие)	8.2	1 (Наличие гипертонии)	0 (Наличие АГ)	1 (Наличие)
3 (Ожирение)	1 (Наличие)	9.5	1 (Наличие гипертонии)	0 (Наличие АГ)	1 (Наличие)
3 (Ожирение)	1 (Наличие)	7.8	1 (Наличие гипертонии)	0 (Наличие АГ)	1 (Наличие)

СтатТех 2.6.7  
© ООО "СтатТех", Россия, 2020  
Программа зарегистрирована Федеральной службой по интеллектуальной собственности, номер регистрации 2020615715, дата регистрации 29.05.2020

Пользователь: Marina  
Почта: marina-zabolotsk@mail.ru

**Анализ данных: «ДЕМО»**

**Описание методов статистического анализа**

Статистический анализ проводился с использованием программы StatTech v. 2.6.7 (разработчик - ООО "СтатТех", Россия). Количественные показатели оценивались на предмет соответствия нормальному распределению с помощью критерия Шапиро-Уилка (при числе исследуемых менее 50) или критерия Колмогорова-Смирнова (при числе исследуемых более 50). Количественные показатели, имеющие нормальное распределение, описывались с помощью средних арифметических величин (M) и стандартных отклонений (SD), границы 95% доверительного интервала (95% ДИ). При сравнении нормально распределенных количественных показателей, рассчитанных для двух связанных выборок, использовался парный t-критерий Стьюдента.

Активация Windows  
Был выполнен анализ динамики АД.  
После завершения анализа Windows перейдите в раздел "Параметры".

Таблица 1 – Анализ динамики АД

Рисунок 7.28 – Пример сформированного анализа данных

Варианты сформированного анализа показателей в формате «До – После» будут зависеть от нормальности распределения, в связи с использованием разных критериев анализа показателей. Структура самого анализа остается неизменной: описание методов статистического анализа, таблица анализа данных, вывод по результатам анализа показателей с указанием используемого метода, диаграмма.

- 1) При «ненормальном» распределении данных таблица анализа данных будет выглядеть следующим образом (Рис. 7.29).

Этапы наблюдения				p
АД до лечения		АД после лечения		
Me	Q <sub>1</sub> – Q <sub>3</sub>	Me	Q <sub>1</sub> – Q <sub>3</sub>	
176 (n=13)	162 – 189	145 (n=13)	133 – 169	0,008*

\* – различия показателей статистически значимы (p < 0,05)

В процессе анализа у АД отмечались статистически значимые изменения (p = 0,008) (используемый метод: критерий Уилкоксона).

Рисунок 7.29 – Пример табличной формы анализа данных «До – После» при «ненормальном» распределении

Диаграмма же при данном распределении будет иметь следующий вид (Рис. 7.30).

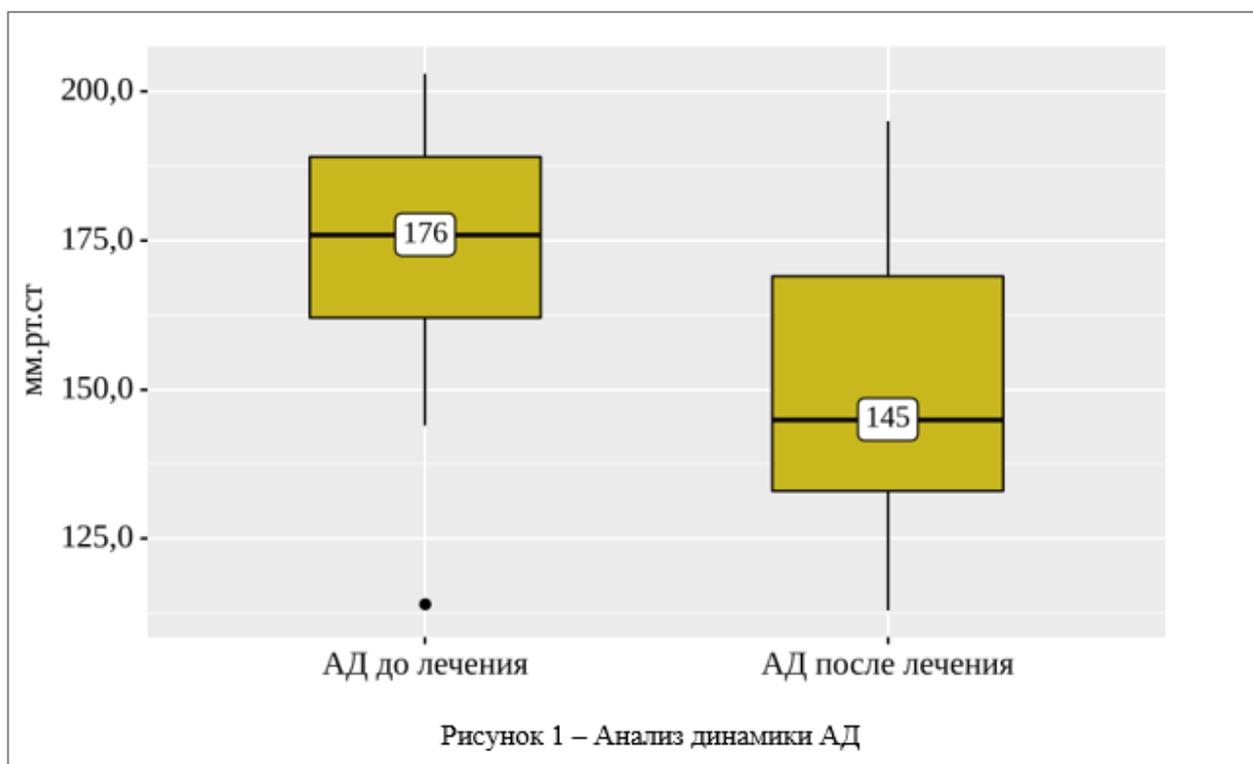


Рисунок 7.30 – Пример диаграммы анализа показателей «До – После» при «ненормальном» распределении

2) При **нормальном** распределении данных таблица анализа данных будет выглядеть следующим образом (Рис. 7.31).

Этапы наблюдения				p
АД до лечения		АД после лечения		
M ± SD	95% ДИ	M ± SD	95% ДИ	
171 ± 26 (n=13)	156 – 187	150 ± 25 (n=13)	135 – 165	0,005*

\* – различия показателей статистически значимы (p < 0,05)

Анализ показал, что у АД отмечались статистически значимые изменения (p = 0,005) (используемый метод: парный t-критерий Стьюдента).

Рисунок 7.31 – Пример табличной формы анализа данных «До – После» при нормальном распределении

Диаграмма же при данном распределении будет иметь следующий вид (Рис. 7.32).

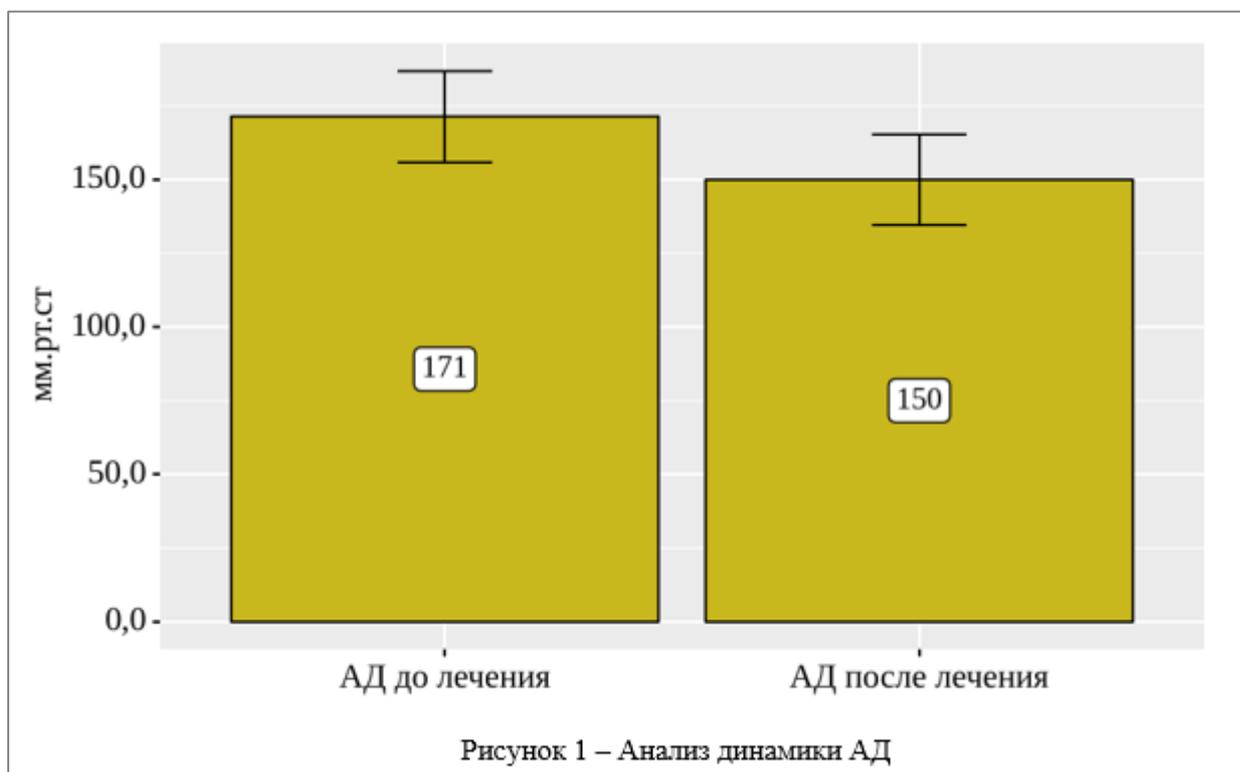
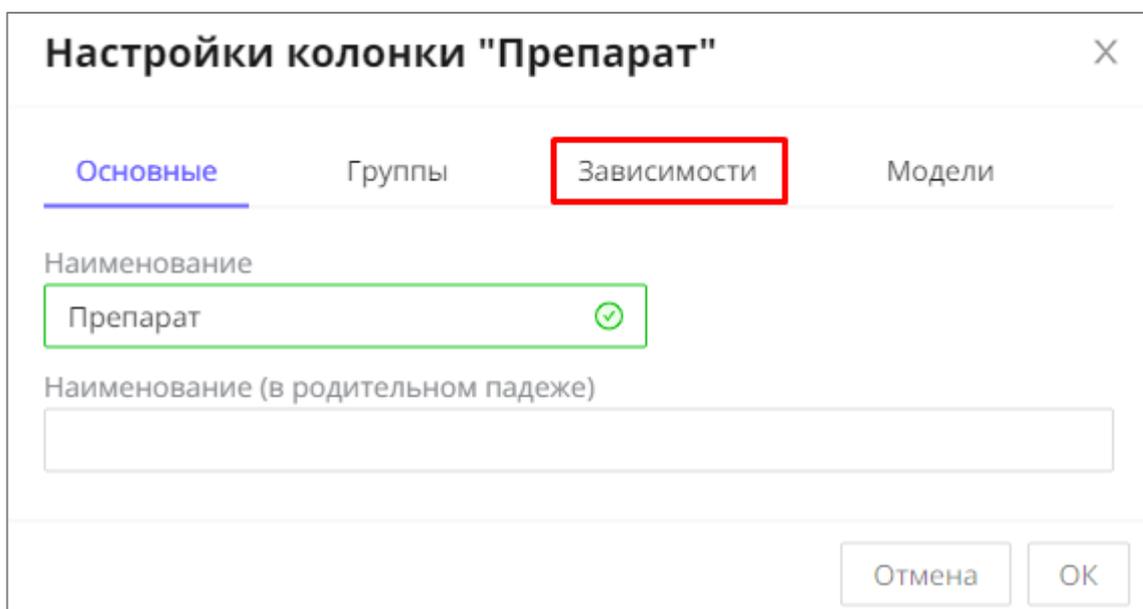


Рисунок 7.32 – Пример диаграммы анализа показателей «До – После» при нормальном распределении

## 7.8.2. Сравнение показателей в формате «До – После» с делением на группы

Для того, чтобы проанализировать два показателя в формате «До – После» необходимо создать группу показателей (см. Раздел 4.2). Необходимые колонки отнести к этой группе, для примера, возьмем показатели уровня артериального давления (АД) до и после лечения (Рис. 7.26). Построим зависимость уровня АД от лечения препаратами.

Для этого нам необходимо на столбце «Препарат» зайти в настройки колонки, нажав на знак . Перейти на вкладку «Зависимости». (Рис. 7.33).



Препарат 7.33 – Вкладка «Зависимости» в настройках колонки

На вкладке «Зависимости» выбрать «Зависимые группы», проставить галочку на необходимой группе – «АД в динамике» и нажать на кнопку «ОК» (Рис. 7.34).

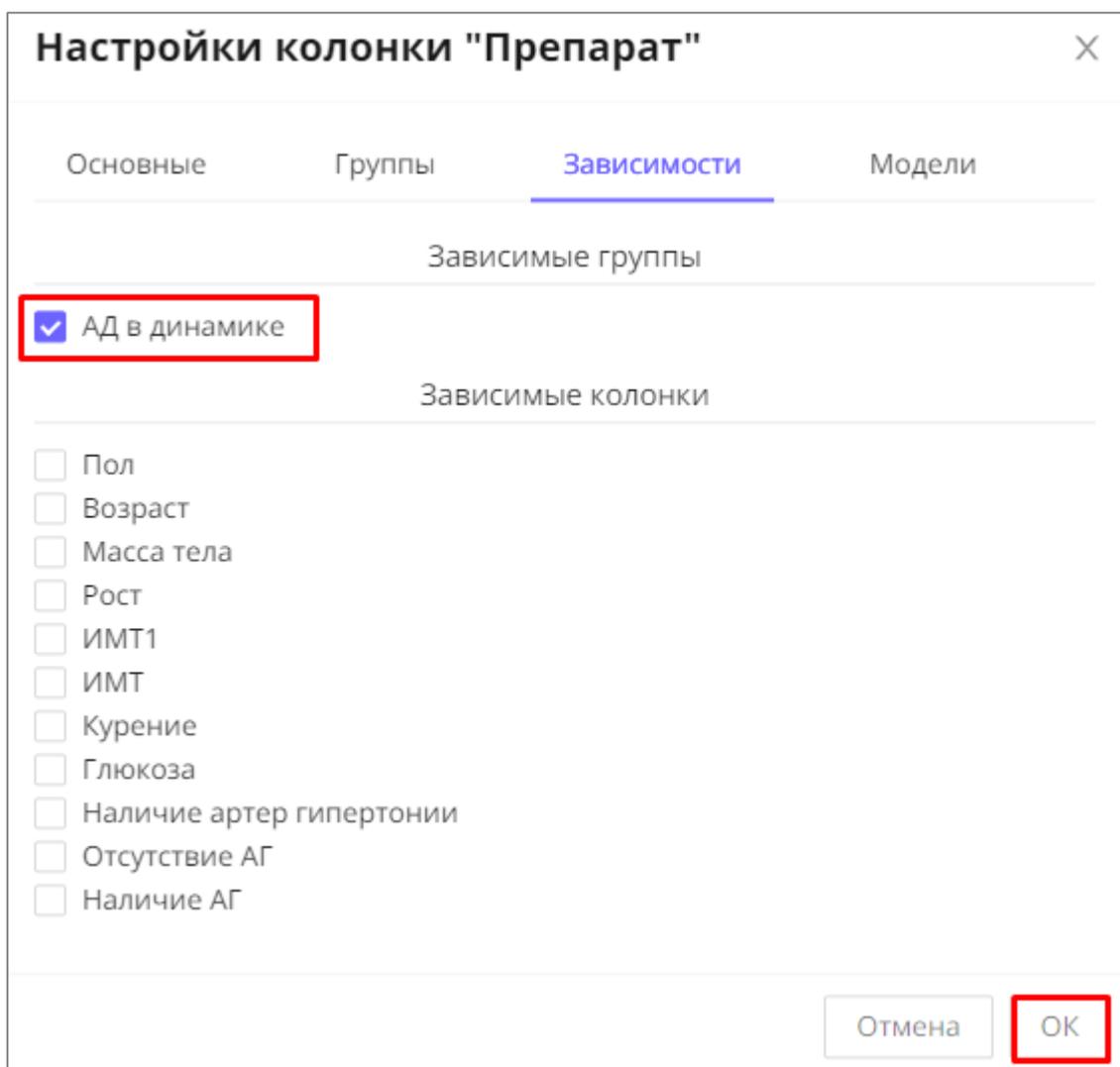


Рисунок 7.34 – Выбор зависимой группы в настройках колонки

После этих действий необходимо нажать на кнопку «Выполнить анализ» в левом верхнем углу экрана (Рис. 7.27).

Структура анализа данных также будет зависеть от распределения данных в анализируемых колонках.

- 1) Пример анализа данных при «ненормальном» распределении представлен ниже. Структура анализа данных будет также состоять из описания методов статистического анализа (Рис. 7.35), таблицы анализа данных (Рис. 7.36), вывод по результатам анализа показателей с указанием используемого метода (Рис.7.37), диаграммы (Рис. 7.38).

### Описание методов статистического анализа

Статистический анализ проводился с использованием программы StatTech v. 2.6.7 (разработчик - ООО "Статтех", Россия).

Количественные показатели оценивались на предмет соответствия нормальному распределению с помощью критерия Шапиро-Уилка (при числе исследуемых менее 50) или критерия Колмогорова-Смирнова (при числе исследуемых более 50).

В случае отсутствия нормального распределения количественные данные описывались с помощью медианы (Me) и нижнего и верхнего квартилей (Q1 – Q3).

Сравнение трех и более групп по количественному показателю, распределение которого отличалось от нормального, выполнялось с помощью критерия Краскела-Уоллиса, апостериорные сравнения – с помощью критерия Данна с поправкой Холма.

При сравнении количественных показателей, распределение которых отличалось от нормального, в двух связанных группах, использовался критерий Уилкоксона.

Рисунок 7.35 – Пример описания методов статистического анализа

Таблица 1 – Анализ динамики АД в зависимости от показателя "Препарат"

Препарат	Этапы наблюдения				p
	АД до лечения		АД после лечения		
	Me	Q <sub>1</sub> – Q <sub>3</sub>	Me	Q <sub>1</sub> – Q <sub>3</sub>	
Без лечения	114 (n=1)	114 – 114	113 (n=1)	113 – 113	1,000
Препарат А	176 (n=6)	163 – 188	135 (n=6)	127 – 143	0,031*
Препарат В	180 (n=6)	168 – 189	170 (n=6)	155 – 186	0,438

Рисунок 7.36 – Пример табличной формы анализа показателей при «ненормальном» распределении

Проведенный анализ показал, что при анализе показателя на этапе показателя "АД до лечения" нам не удалось установить статистически значимых различий ( $p = 0,257$ ) (используемый метод: Критерий Краскела–Уоллиса). В ходе анализа показателя на этапе показателя "АД после лечения" были выявлены статистически значимые различия ( $p = 0,023$ ) (используемый метод: Критерий Краскела–Уоллиса).

Проведенный анализ показал, что в группе без лечения не удалось выявить статистически значимых изменений ( $p = 1,000$ ) (используемый метод: критерий Уилкоксона). Проведенный анализ показал, что в группе препарата А нами были установлены статистически значимые изменения ( $p = 0,031$ ) (используемый метод: критерий Уилкоксона). Анализ показал, что в группе препарата В нам не удалось выявить статистически значимых изменений ( $p = 0,438$ ) (используемый метод: критерий Уилкоксона).

Рисунок 7.37 – Пример вывода по результатам анализа данных при «ненормальном» распределении

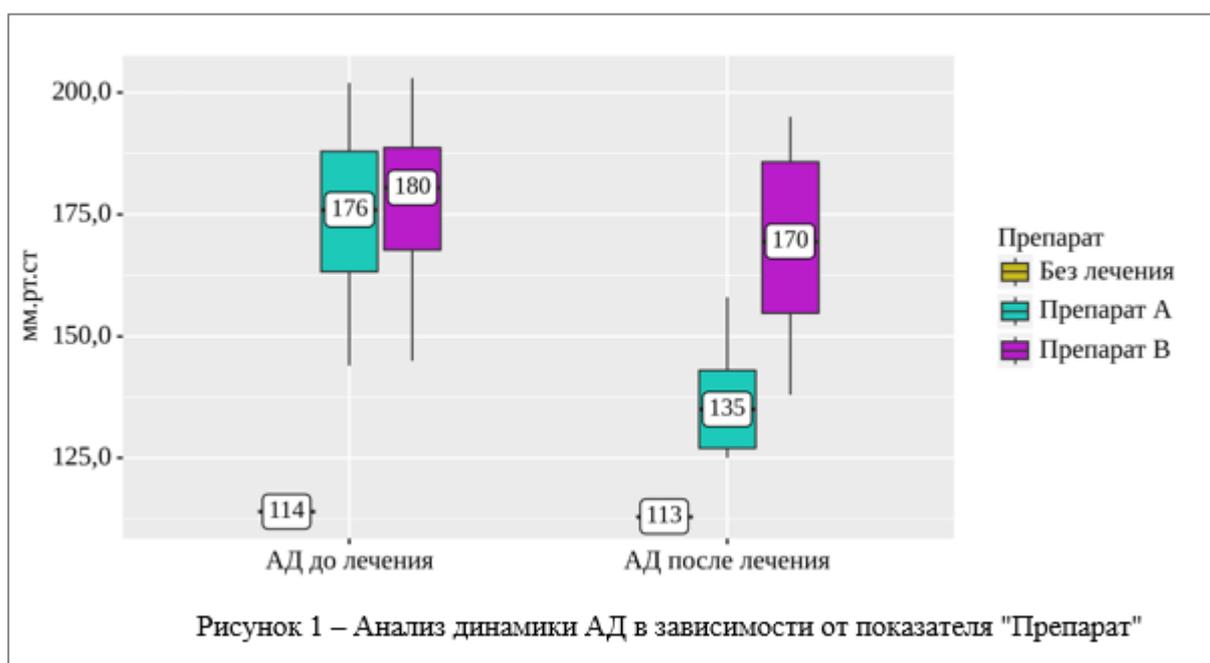


Рисунок 7.38 – Пример диаграммы анализа данных при «ненормальном» распределении

2) Анализ данных при нормальном распределении будет отличаться используемыми критериями для сравнения показателей. Видом диаграммы (Рис. 7.40) и таблицей анализа данных (Рис. 7.39).

Таблица 1 – Анализ динамики АД в зависимости от показателя "Препарат"

Препарат	Этапы наблюдения				p
	АД до лечения		АД после лечения		
	M ± SD	95% ДИ	M ± SD	95% ДИ	
Без лечения	126 ± 8 (n=8)	119 – 133	124 ± 7 (n=8)	118 – 130	0,009*
Препарат А	175 ± 21 (n=6)	153 – 197	137 ± 13 (n=6)	124 – 151	0,005*
Препарат В	177 ± 20 (n=6)	156 – 199	169 ± 22 (n=6)	145 – 192	0,266
p	< 0,001* РБез лечения – Препарат В = 0,001 РБез лечения – Препарат А = 0,001		< 0,001* РБез лечения – Препарат В = 0,001 РПрепарат В – Препарат А = 0,004		–

\* – различия показателей статистически значимы ( $p \leq 0,05$ )

Рисунок 7.39 – Пример табличной формы анализа данных при нормальном распределении

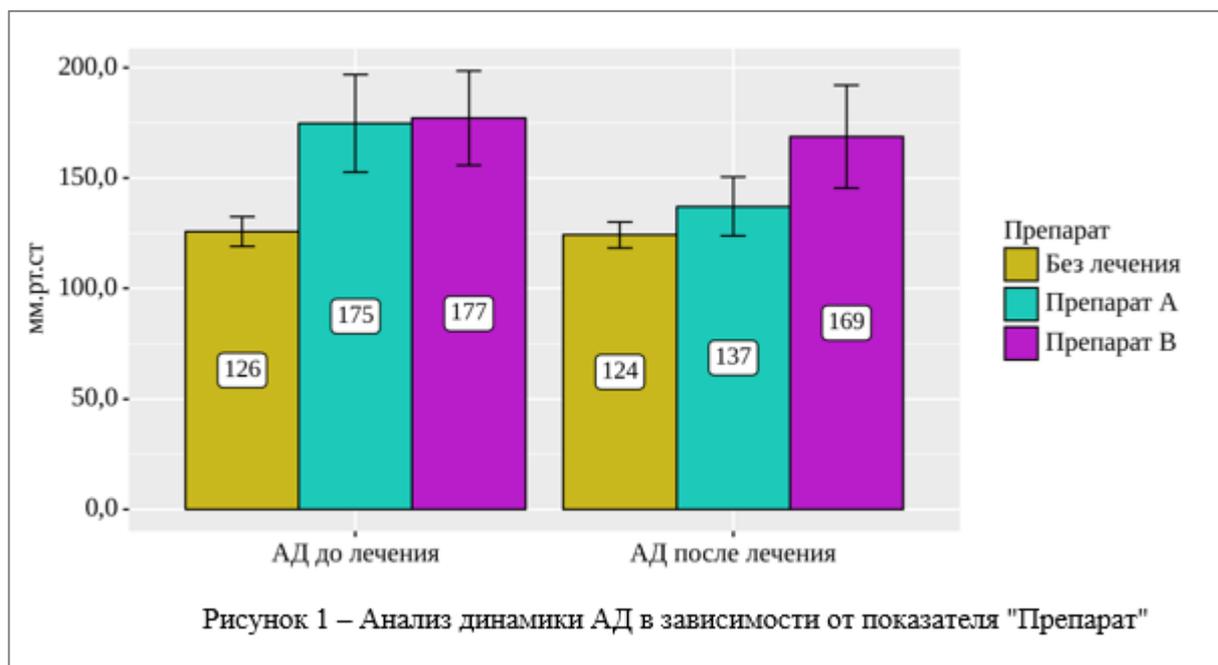


Рисунок 1 – Анализ динамики АД в зависимости от показателя "Препарат"

Рисунок 7.40 – Пример диаграммы анализа данных при нормальном распределении

### 7.8.3. Сравнение трех и более показателей в формате «До – После»

Для того, чтобы проанализировать три и более показателя в формате «До – После» необходимо создать группу показателей (см. Раздел 4.2). Необходимые колонки отнести к этой группе, для примера, возьмем показатели уровня артериального давления до / после лечения / через 3 месяца (Рис. 7.41).

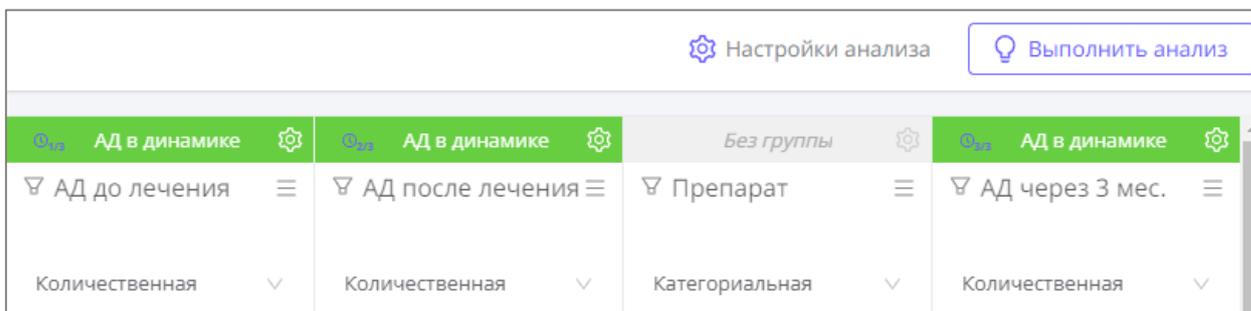


Рисунок 7.41 - Пример распределения показателей к группам в формате «До – После»

После того как необходимые показатели отнесены к группам необходимо нажать кнопку «Выполнить анализ» в правом верхнем углу экрана (Рис. 7.42).

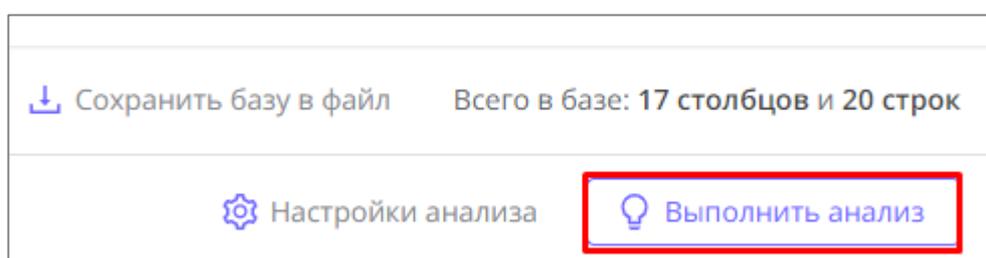


Рисунок 7.42 – Кнопка «Выполнить анализ»

Структура выполненного анализа будет иметь стандартные разделы:

- Описание методов статистического анализа;
- Таблица анализа показателей с расчётами;
- Выводы к таблице;
- Диаграмма.

При анализе нормального распределения количественных данных в динамике при оценке трех и более показателей строится точечная диаграмма вместо столбиковой (Рис. 7.43).

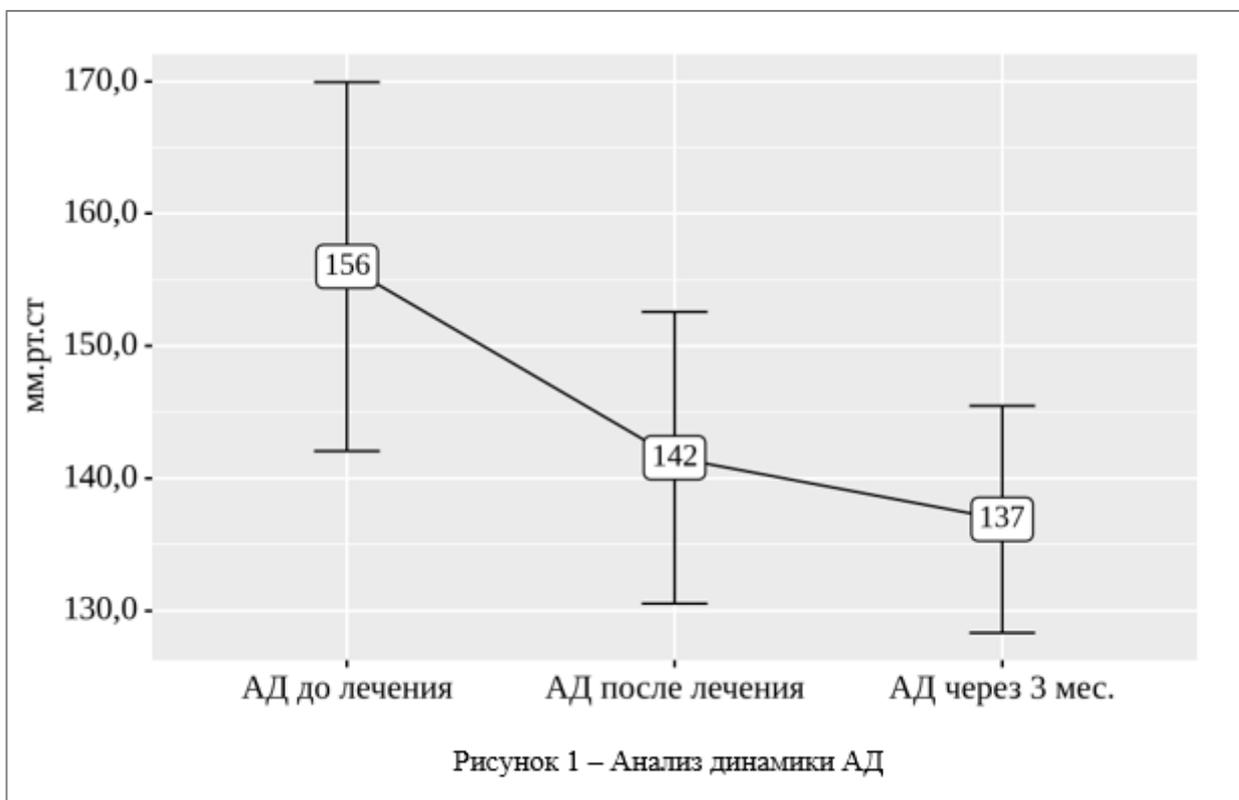
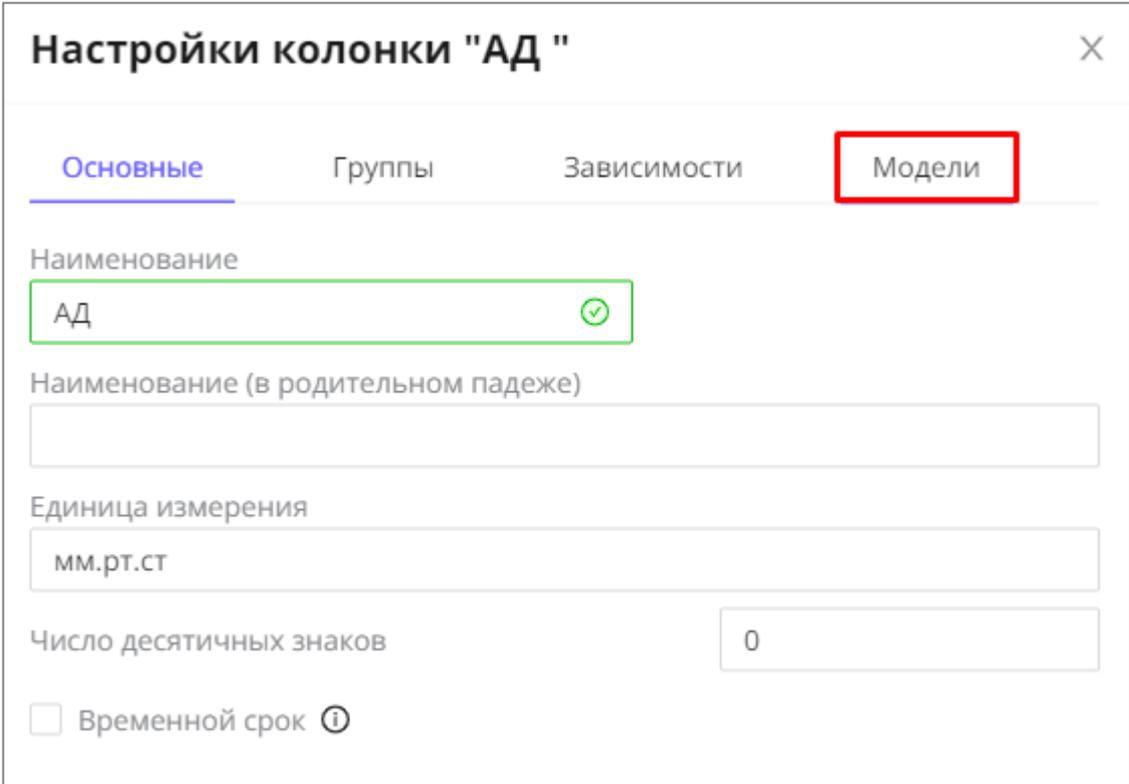


Рисунок 7.43 – Пример точечной диаграммы при анализе трех показателей в динамике

## ГЛАВА 8. ПОСТРОЕНИЕ МОДЕЛЕЙ

### 8.1. Линейная регрессия

Для построения линейной регрессии нужно зайти в настройки колонки, нажав на знак . Далее перейти на вкладку «Модели» (Рис.8.1).



The image shows a dialog box titled "Настройки колонки "АД "" with a close button (X) in the top right corner. It has four tabs: "Основные", "Группы", "Зависимости", and "Модели". The "Модели" tab is selected and highlighted with a red border. Below the tabs, there are several input fields and a checkbox:

- "Наименование": A text input field containing "АД" with a green checkmark icon on the right.
- "Наименование (в родительном падеже)": An empty text input field.
- "Единица измерения": A text input field containing "мм.рт.ст".
- "Число десятичных знаков": A numeric input field containing "0".
- A checkbox labeled "Временной срок" with an information icon (i) to its right.

Рисунок 8.1 – Вкладка «Модели» в настройках колонки

На данной вкладке расположен функционал построения моделей, который состоит из «Отбора предикторов» (Рис. 8.2). и выбора «Независимых колонок» (Рис. 8.3).

Зависимая переменная

АД

Отбор предикторов

Критерий отбора

P-value меньше

Минимальный критерий Акаике

Рисунок 8.2 – Функционал «Отбор предикторов»

Независимые колонки

Пол

Возраст

Масса тела

Рост

ИМТ1

ИМТ

Курение

Глюкоза

Наличие артер гипертонии

Отсутствие АГ

Наличие АГ

Препарат

АД после лечения

Рисунок 8.3 – Выбор независимых колонок для построения регрессии

В отборе предикторов есть возможность выбора с помощью «Пошагового исключения» - включение в анализ только значимых

переменных. Критериями отбора для «Пошагового исключения» могут быть либо «р -Value», либо «Минимальный критерий Акаике» (Рис. 8.4). И «Принудительное включение» - все выбранные переменные (колонки) будут входить в анализ (Рис. 8.5).

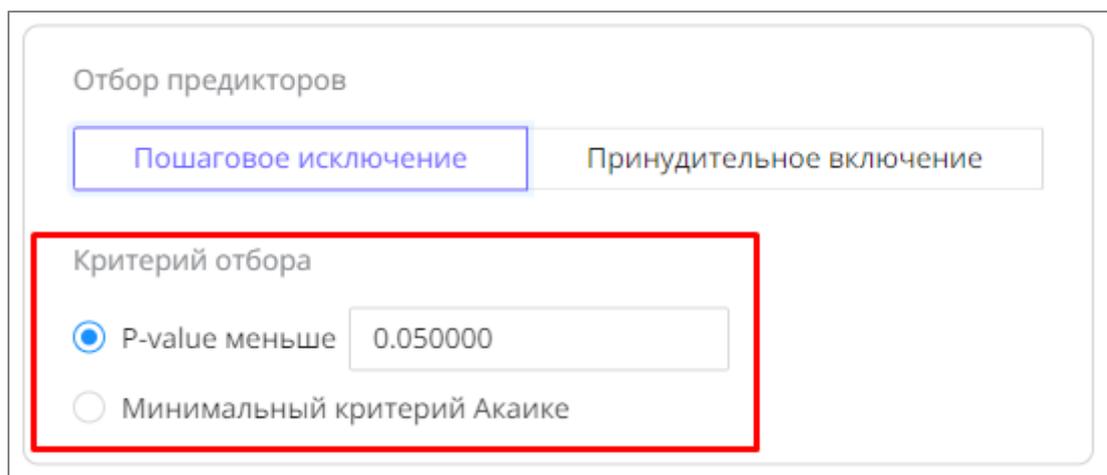


Рисунок 8.4 – «Критерий отбора» во вкладке «Модели»

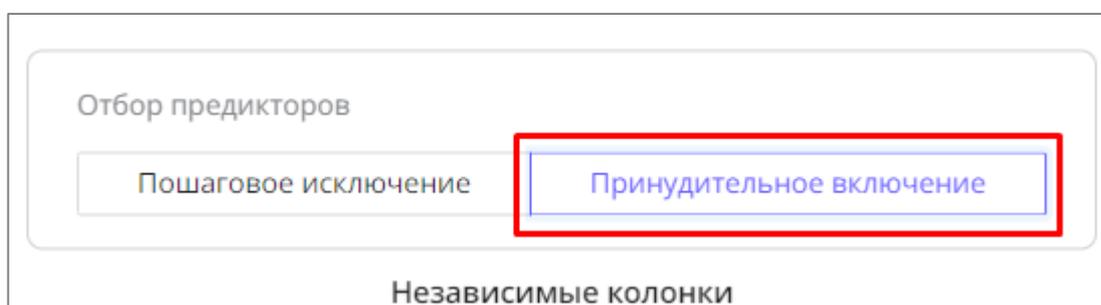


Рисунок 8.5 – «Принудительное включение» во вкладке «Модели»

Для построения логистической регрессии необходимо на вкладке «Модели» в блоке «Независимые колонки» выбрать необходимую переменную для анализа. Для примера, возьмем построение прогностической модели, характеризующей зависимость уровня артериального давления (АД) от возраста пациента, с помощью метода линейной регрессии.

Для выполнения данного анализа необходимо во вкладке «Модели» на колонке «АД» выбрать в блоке «Независимые колонки» показатель «Возраст». Затем нажать на кнопку «ОК» (Рис. 8.6).

Настройки колонки "АД"

Основные Группы Зависимости **Модели**

Зависимая переменная

АД

Отбор предикторов

Пошаговое исключение **Принудительное включение**

Независимые колонки

- Пол
- Возраст**
- Масса тела
- Рост
- ИМТ1
- ИМТ
- Курение
- Глюкоза
- Наличие артер гипертонии
- Отсутствие АГ
- Наличие АГ
- Препарат
- АД после лечения

Отмена **ОК**

Рисунок 8.6 – Выбор показателя во вкладке «Модели»

Далее необходимо нажать на кнопку «Выполнить анализ» в правом верхнем углу экрана (Рис.8.7).

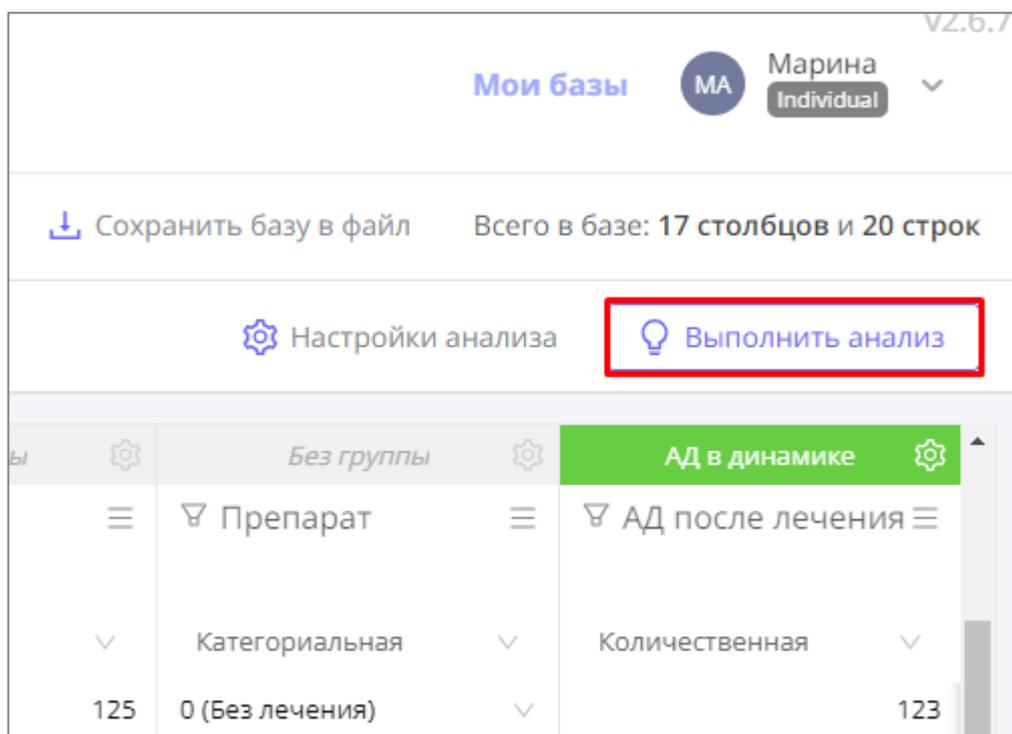


Рисунок 8.7 – Кнопка «Выполнить анализ»

Структура выполненного анализа будет выглядеть следующим образом (Рис. 8.8).

Оценка зависимости показателя "АД" от количественных факторов была выполнена с помощью метода линейной регрессии. Число наблюдений составило 20.

Таблица 1 – Анализ показателя "АД" в зависимости от возраста

	B	Стд. ошибка	t	p
Intercept	109,752	14,949	7,342	< 0,001*
Возраст	1,197	0,361	3,317	0,004*

\* – различия показателей статистически значимы ( $p < 0,05$ )

Наблюдаемая зависимость показателя "АД" от возраста описывается уравнением линейной регрессии:

$$Y_{\text{АД}} = 109,752 + 1,197X_{\text{Возраст}}$$

где Y – величина показателя "АД",  $X_{\text{Возраст}}$  – Возраст (полных лет)

При увеличении возраста на 1 полных лет. следует ожидать увеличение показателя "АД" на 1,197 мм.рт.ст.

Полученная регрессионная модель характеризуется коэффициентом корреляции  $r_{xy} = 0,616$ , что соответствует заметной тесноте связи по шкале Чеддока. Модель была статистически значимой ( $p = 0,004$ ). Полученная модель объясняет 37,9% наблюдаемой дисперсии показателя "АД".

Рисунок 8.8 – Пример вывода анализа с помощью метода линейной регрессии.

Также данный анализ можно выполнить от нескольких факторов. Тогда в «Независимых колонках» во вкладке «Модели» необходимо выбрать несколько факторов для анализа. (Рис. 8.9)

Независимые колонки

- Пол
- Возраст
- Масса тела
- Рост
- ИМТ1
- ИМТ
- Курение
- Глюкоза
- Наличие артер гипертонии
- Отсутствие АГ
- Наличие АГ
- Препарат
- АД после лечения

Отмена ОК

Рисунок 8.9 – Пример выбора несколько показателей для анализа

Далее мы нажимаем на кнопку «Выполнить анализ» (Рис.8.7). Структура сформированного анализа выглядит следующим образом (Рис. 8.10).

Оценка зависимости показателя "АД" от количественных факторов была выполнена с помощью метода линейной регрессии. Число наблюдений составило 20.

Таблица 1 – Анализ показателя "АД" в зависимости от возраста, содержания глюкозы в сыворотке крови, ИМТ

	B	Стд. ошибка	t	p
Intercept	37,884	20,347	1,862	0,081
Возраст	0,662	0,254	2,606	0,019*
Глюкоза	13,135	3,346	3,926	0,001*
ИМТ1	0,246	0,972	0,253	0,803

\* – различия показателей статистически значимы ( $p \leq 0,05$ )

Наблюдаемая зависимость показателя "АД" от возраста, содержания глюкозы в сыворотке крови, ИМТ описывается уравнением линейной регрессии:

$$Y_{\text{АД}} = 37,884 + 0,662X_{\text{Возраст}} + 13,135X_{\text{Глюкоза}} + 0,246X_{\text{ИМТ1}}$$

где Y – величина показателя "АД",  $X_{\text{Возраст}}$  – Возраст (полных лет),  $X_{\text{Глюкоза}}$  – Глюкоза (ммоль/л),  $X_{\text{ИМТ1}}$  – ИМТ1 (кг/м<sup>2</sup>)

При увеличении возраста на 1 полных лет следует ожидать увеличение показателя "АД" на 0,662 мм.рт.ст. при увеличении содержания глюкозы в сыворотке крови на 1 ммоль/л. следует ожидать увеличение показателя "АД" на 13,135 мм.рт.ст. при увеличении ИМТ на 1 кг/м<sup>2</sup>. следует ожидать увеличение показателя "АД" на 0,246 мм.рт.ст.

Полученная регрессионная модель характеризуется коэффициентом корреляции  $r_{\text{xy}} = 0,879$ , что соответствует высокой тесноте связи по шкале Чеддока. Модель была статистически значимой ( $p < 0,001$ ). Полученная модель объясняет 77,3% наблюдаемой дисперсии показателя "АД".

Рисунок 8.10 – Пример сформированного многофакторного анализа с помощью метода линейной регрессии

## 8.2. Логистическая регрессия

Для построения логистической регрессии нужно зайти в настройки колонки, нажав на знак . Далее перейти на вкладку «Модели» (Рис.8.1).

На данной вкладке расположен функционал построения моделей, который состоит из «Отбора предикторов» (Рис. 8.2). и выбора «Независимых колонок» (Рис. 8.3).

В отборе предикторов есть возможность выбора с помощью «Пошагового исключения» - включение в анализ только значимых переменных. Критериями отбора для «Пошагового исключения» могут быть либо «р -Value», либо «Минимальный критерий Акаике» (Рис. 8.4). И «Принудительное включение» - все выбранные переменные (колонки) будут входить в анализ (Рис. 8.5).

Для построения логистической регрессии необходимо на вкладке «Модели» в блоке «Независимые колонки» выбрать необходимую переменную для анализа. Для примера, возьмем построение прогностической модели, характеризующей зависимость уровня артериального давления (АД) от возраста пациента, наличия курения и индекса массы тела (ИМТ), с помощью метода бинарной логистической регрессии.

Для выполнения данного анализа необходимо во вкладке «Модели» на колонке «АД» выбрать в блоке «Независимые колонки» показатель «Возраст», «Курение» и «ИМТ». Затем нажать на кнопку «ОК» (Рис. 8.11).

### Настройки колонки "Наличие АГ" X

Основные      Группы      Зависимости      **Модели**

Зависимая переменная

Наличие АГ

Отбор предикторов

**Пошаговое исключение**      Принудительное включение

Критерий отбора

P-value меньше

Минимальный критерий Акаике

Независимые колонки

- Пол
- Трудовой статус
- Возраст**
- Масса тела
- Рост
- ИМТ**
- Степень ИМТ
- Курение**
- Глюкоза
- Наличие артериальной гипертонии
- Отсутствие АГ
- АД до лечения
- Препарат

Рисунок 8.11 – Выбор переменных во вкладке «Модели»

Далее необходимо нажать на кнопку «Выполнить анализ» в правом верхнем углу экрана (Рис.8.7).

Структура выполненного анализа будет выглядеть следующим образом (Рис. 8.12 - Рис. 8.15).

Статистический анализ проводился с использованием программы StatTech v. 2.8.2 (разработчик - ООО "Статтех", Россия).

Построение прогностической модели вероятности определенного исхода выполнялось при помощи метода логистической регрессии. Мерой определенности, указывающей на ту часть дисперсии, которая может быть объяснена с помощью логистической регрессии, служил коэффициент  $R^2$  Найджелкерка.

Для оценки диагностической значимости количественных признаков при прогнозировании определенного исхода, применялся метод анализа ROC-кривых. Разделяющее значение количественного признака в точке cut-off определялось по наивысшему значению индекса Юдена.

Была разработана прогностическая модель для определения вероятности показателя "Наличие АГ" в зависимости от возраста, ИМТ, курения методом бинарной логистической регрессии. Число наблюдений составило 20. Наблюдаемая зависимость описывается уравнением:

$$P = 1 / (1 + e^{-z}) \times 100\%$$
$$z = -52,867 + 0,599X_{\text{Возраст}} + 1,084X_{\text{ИМТ}} + 11,241X_{\text{Наличие}}$$

где P – вероятность наличия АГ,  $X_{\text{Возраст}}$  – Возраст (полных лет),  $X_{\text{ИМТ}}$  – ИМТ (кг/м<sup>2</sup>),  $X_{\text{Наличие}}$  – Курение (0 – Отсутствие, 1 – Наличие)

Полученная регрессионная модель является статистически значимой ( $p < 0,001$ ). Исходя из значения коэффициента детерминации Найджелкерка, модель объясняет 88,2% наблюдаемой дисперсии показателя "Наличие АГ".

Рисунок 8.12 – Пример разработанной прогностической модели и описание методов статистического анализа

Таблица 1 – Характеристики связи предикторов модели с вероятностью выявления показателя "Наличие АГ"

Предикторы	Unadjusted		Adjusted	
	COR; 95% ДИ	p	AOR; 95% ДИ	p
Возраст	1,109; 1,006 – 1,224	0,038*	1,820; 0,691 – 4,792	0,226
ИМТ	1,276; 1,010 – 1,613	0,041*	2,956; 0,464 – 18,803	0,251
Курение: Наличие	21,000; 1,777 – 248,142	0,016*	76160,634; 0,001 – 4014765276602,423	0,215

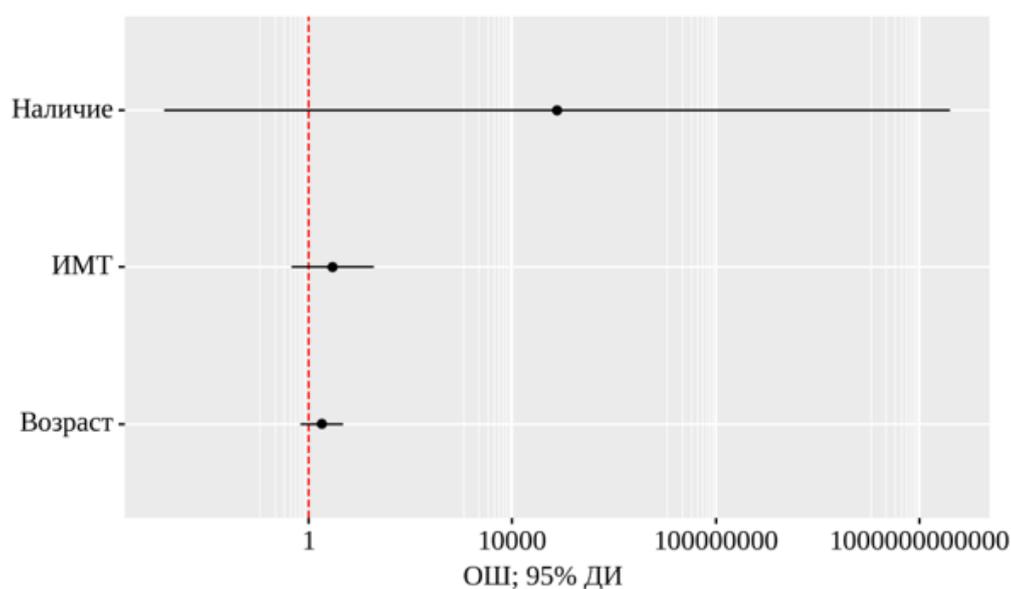


Рисунок 1 – Оценки отношения шансов с 95% ДИ для изучаемых предикторов показателя "Наличие АГ"

При оценке зависимости вероятности наличия АГ от значения логистической функции P с помощью ROC-анализа была получена следующая кривая.

Рисунок 8.13 – Пример вывода бинарной логистической регрессии

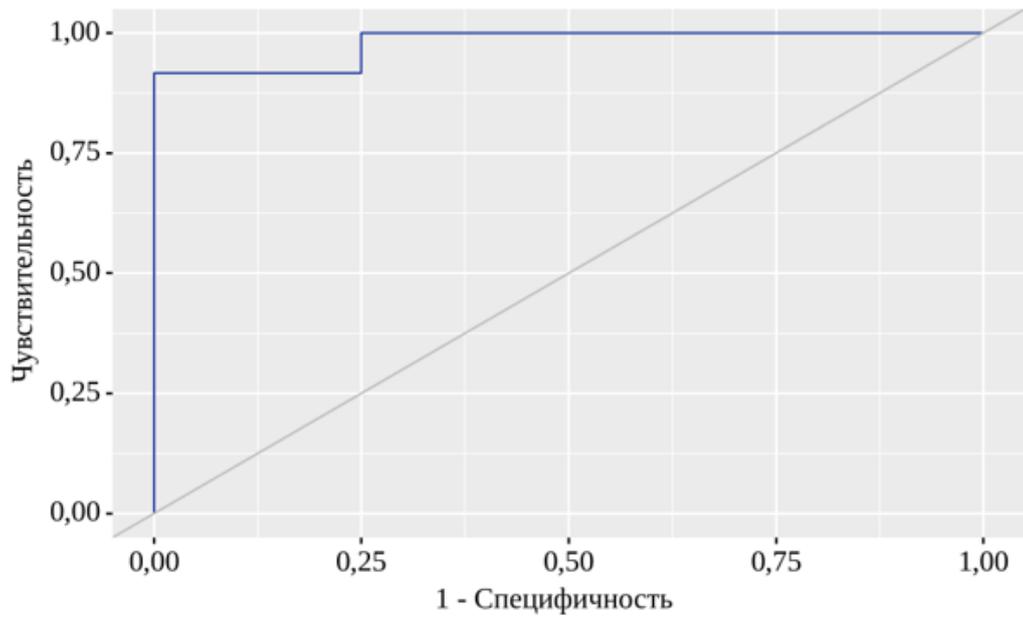


Рисунок 2 – ROC-кривая, характеризующая зависимость вероятности показателя "Наличие АГ" от значения логистической функции

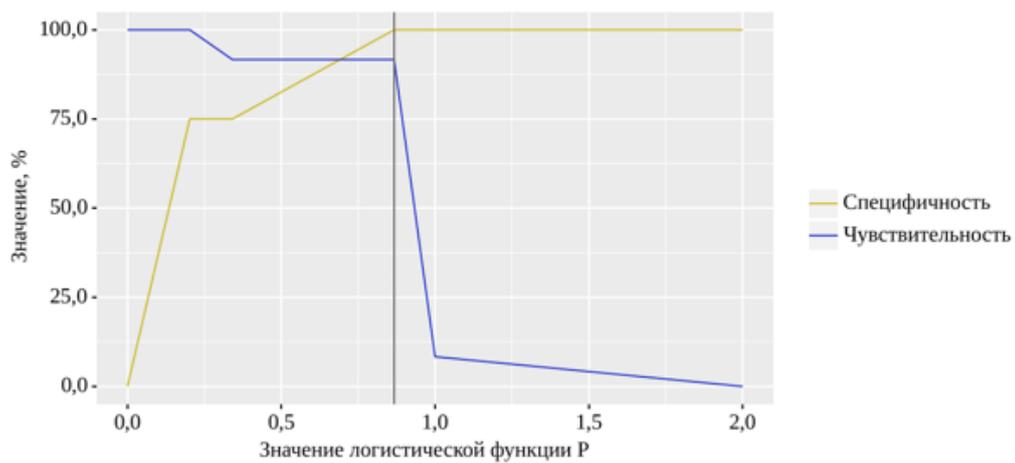


Рисунок 3 – Анализ чувствительности и специфичности модели в зависимости от пороговых значений логистической функции |

Рисунок 8.14 – Пример ROC- кривой и анализ чувствительности и специфичности построенной модели

Таблица 2 – Пороговые значения логистической функции

Порог	Чувствительность (Se), %	Специфичность (Sp), %	PPV	NPV
<b>0,867</b>	<b>91,7</b>	<b>100,0</b>	<b>100,0</b>	<b>92,3</b>
0,341	91,7	75,0	78,6	90,0
0,203	100,0	75,0	80,0	100,0

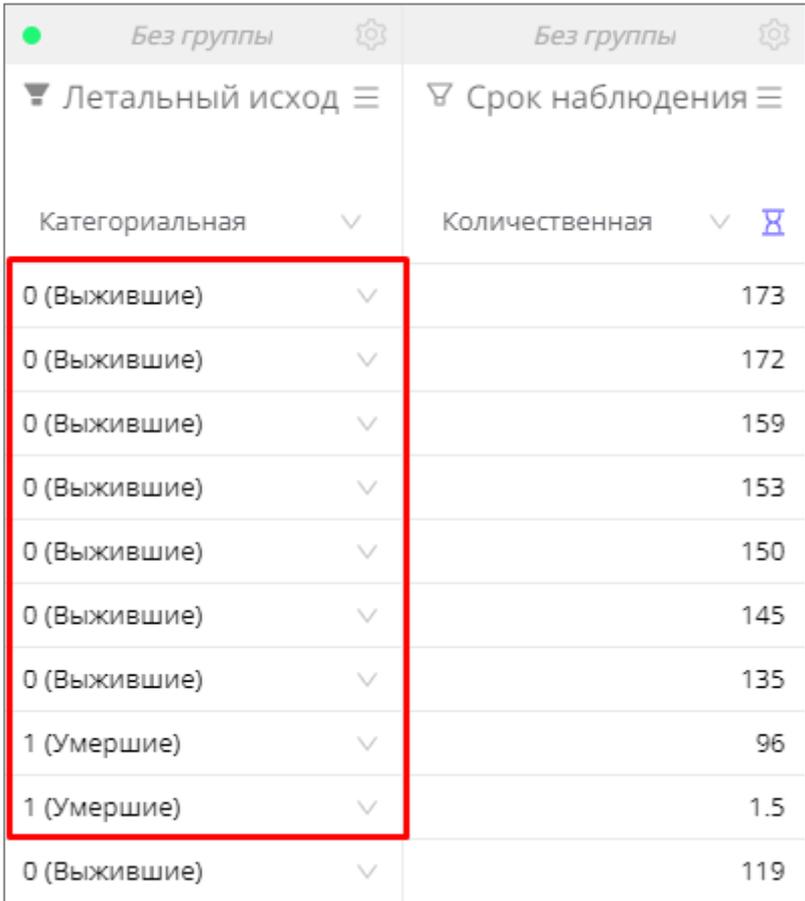
Площадь под ROC-кривой составила  $0,979 \pm 0,032$  с 95% ДИ: 0,917 – 1,000. Полученная модель была статистически значимой ( $p < 0,001$ ).

Пороговое значение логистической функции P в точке cut-off, которому соответствовало наивысшее значение индекса Юдена, составило 0,867. Наличие прогнозировалось при значении логистической функции P выше данной величины или равном ей. Чувствительность и специфичность модели составили 91,7% и 100,0%, соответственно.

Рисунок 8.15 – Пример таблицы и вывода пороговых значений логистической функции

### 8.3. Анализ выживаемости

Для анализа выживаемости необходимо, чтобы в базе данных был столбец «Событие». Например, признак летального исхода. Столбец должен иметь бинарное распределение (0 – событие произошло, 1 – событие отсутствовало) (Рис. 8.1).



Летальный исход	Срок наблюдения
0 (Выжившие)	173
0 (Выжившие)	172
0 (Выжившие)	159
0 (Выжившие)	153
0 (Выжившие)	150
0 (Выжившие)	145
0 (Выжившие)	135
1 (Умершие)	96
1 (Умершие)	1.5
0 (Выжившие)	119

Рисунок 8.1 – Пример заполнения столбца «Событие» - «Летальный исход»

Также для проведения анализа выживаемости необходимо создать столбец «Срок наблюдения» - это выражение от даты до «События» или от даты до последнего контакта с пациентом, если событие не произошло. Измеряется в неделях, месяцах, годах или других временных отрезках (Рис. 8.2).

Без группы	⚙	Без группы	⚙
☑ Смертность	☰	☑ Срок наблюдения	☰
Категориальная	∨	Количественная	∨ ⌚
0 (Показатель 1)	∨		173
0 (Показатель 1)	∨		172
0 (Показатель 1)	∨		159
0 (Показатель 1)	∨		153
0 (Показатель 1)	∨		150
0 (Показатель 1)	∨		145
0 (Показатель 1)	∨		135
1 (Показатель 2)	∨		96
1 (Показатель 2)	∨		15

Рисунок 8.2 – Пример заполнения столбца «Срок наблюдения»

Для того, чтобы проанализировать выживаемость на столбце «Срок наблюдения» необходимо поставить галочку в настройках колонки на функции «Временной срок». После активации функции «Временной срок» появится строка с выбором «Тип Выживаемости» (Рис. 8.3).

Настройки колонки "Срок наблюдения" X

Основные Группы Зависимости Модели

Наименование  
Срок наблюдения ✓

Наименование (в родительном падеже)

Единица измерения  
Месяцев

Число десятичных знаков  
0

Временной срок ⓘ

Тип выживаемости  
Общая ▼

Рисунок 8.3 – Функция «Временной срок» в настройках колонки

Далее необходимо на колонке «Событие» зайти в «Настройки колонки» - «Модели» и выбрать необходимые показатели для анализа в блоке «Независимые колонки». «Срок наблюдения» - должен быть указан всегда. Далее нажать кнопку «ОК» (Рис. 8.4).

### Настройки колонки "Смертность"

Основные      Группы      Зависимости      **Модели**

Зависимая переменная  
Смертность

Отбор предикторов  
**Пошаговое исключение**      Принудительное включение

Критерий отбора  
 P-value меньше   
 Минимальный критерий Акаике

Независимые колонки

- Номер
- Группа лечения
- Виды лечения
- Дата операции
- Срок наблюдения ⌚
- Возраст
- Пол
- Направление
- Диагноз
- Время от симптомов до госпитализации
- СЛР
- Тромбы в ПЖ
- Тромбы в ПП

Отмена      **OK**

Рисунок 8.4 – Выбор показателей для анализа «События»

Затем необходимо нажать кнопку «Выполнить анализ» в верхнем правом углу страницы (Рис. 8.5).

↓ Сохранить базу в файл    Всего в базе: 220 столбцов и 142 строк

⚙️ Настройки анализа    💡 Выполнить анализ

Без группы	Без группы	Без группы	Без группы	Без группы
☑️ Смертность	☑️ Срок наблюдения	☑️ Возраст	☑️ Пол	☑️ Категория
Категориальная	Количественная	Количественная	Категориальная	Кат
▼	▼ ⌚	▼	▼	▼
▼		40	0 (Женский)	1 (С
▼		19	0 (Женский)	4 (Д
▼		66	1 (Мужской)	1 (С
▼		35	1 (Мужской)	2 (С
▼		88	0 (Женский)	2 (С
1 (Показатель 2)	12	79	0 (Женский)	3 (М

Рисунок 8.5 – Кнопка «Выполнить анализ»

Анализ выживаемости можно сделать однофакторным, то есть посмотреть влияние на выживаемость лишь одного фактора. Выбирается один показатель во вкладке «Модели» (Рис. 8.6). И многофакторным, когда выбирается несколько показателей (Рис. 8.7).

**Независимые колонки**

- Номер
- Группа лечения
- Виды лечения
- Дата операции
- Срок наблюдения ⌚
- Возраст
- Пол
- Направление
- Диагноз
- Время от симптомов до госпитализации
- СЛР
- Тромбы в ПЖ
- Тромбы в ПП

Рисунок 8.6 – Пример выбора одного фактора для анализа

Независимые колонки

- Номер
- Группа лечения
- Виды лечения
- Дата операции
- Срок наблюдения ⌚
- Возраст
- Пол
- Направление
- Диагноз
- Время от симптомов до госпитализации
- СЛР
- Тромбы в ПЖ
- Тромбы в ПП

Отмена    ОК

Рисунок 8.7 – Пример выбора нескольких факторов для анализа  
(многофакторный анализ)

Структура анализа выживаемости состоит из:

- 1) Описания методов статистического анализа (Рис. 8.8);

**Описание методов статистического анализа**

Статистический анализ проводился с использованием программы StatTech v. 2.6.7 (разработчик - ООО "Статтех", Россия).

Оценка функции выживаемости пациентов проводилась по методу Каплана-Мейера. График оценки функции выживаемости представляет из себя убывающую ступенчатую линию, значения функции выживаемости между точками наблюдений считаются константными. Метод Каплана-Мейера позволяет выполнять анализ цензурированных данных, т.е. оценивать выживаемость с учетом того, что пациенты могут выбывать в ходе эксперимента или иметь разные сроки наблюдения.

Анализ выживаемости пациентов проводился по методу регрессии Кокса, подразумевающему прогнозирование риска наступления события для рассматриваемого объекта и оценку влияния заранее определенных независимых переменных (предикторов) на этот риск. Риск рассматривается как функция, зависящая от времени. Базовые предположения, лежащие в основе метода, состоят в том, что все объясняющие переменные независимы, линейно влияют на риск наступления события, а также что риски наступления события для любых двух объектов в любой отрезок времени пропорциональны.

Рисунок 8.8 – Пример описания методов статистического анализа

2) Кривой выживаемости (Рис. 8.9). Она отображается только для категориальных показателей;

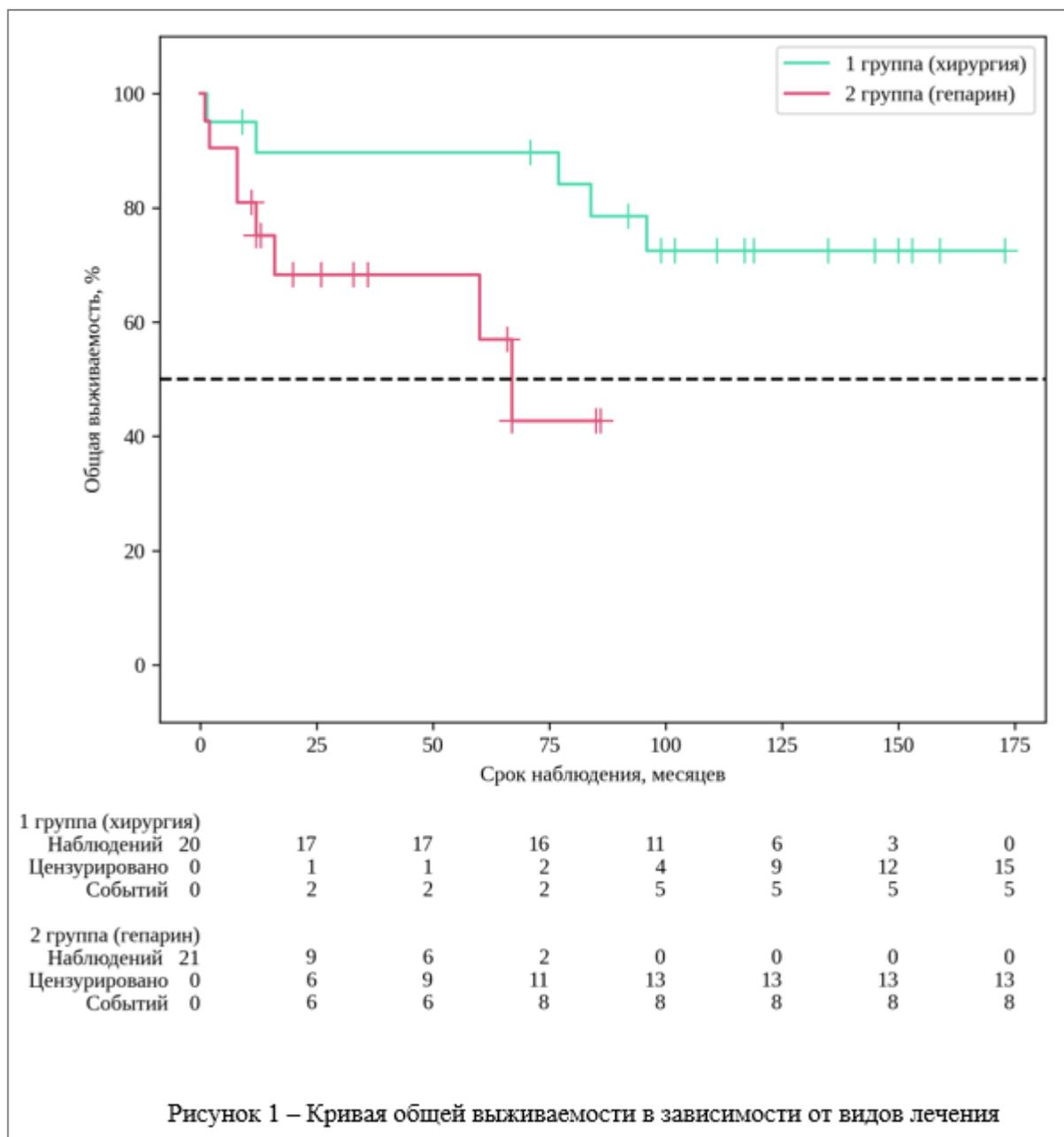


Рисунок 8.9 – Пример кривой выживаемости

### 3) Вывод по кривой выживаемости (Рис. 8.10);

Анализ показал, что медиана срока дожития в группе хирургического лечения не была достигнута, медиана срока дожития в группе лечения гепарином составила 67 месяцев от начала наблюдения (95% ДИ: 12 – ∞ месяцев), медиана срока дожития в группе тромболизиса не была достигнута.

Различия общей выживаемости, оцененные с помощью теста отношения правдоподобия, не были статистически значимы ( $p = 0,051$ ).

Рисунок 8.10 – Пример вывода к кривой выживаемости

### 4) Оценка взаимосвязи выживаемости с изучаемыми факторами с помощью метода регрессии Кокса (Рис.8.11);

При оценке взаимосвязи общей выживаемости с изучаемыми факторами с помощью метода регрессии Кокса была получена следующая модель пропорциональных рисков.

$$h_i(t) = h_0(t) \times \exp(1,445 \times X_{\text{Виды лечения: 2 группа (гепарин)}})$$

где  $h_i(t)$  – прогнозируемый риск Умершие для  $i$ -того элемента наблюдения (в %),  $h_0(t)$  – базовый риск Умершие за определенный временной период  $t$  (в %),  $X_{\text{Виды лечения: 2 группа (гепарин)}}$  – Виды лечения: 2 группа (гепарин)

Рисунок 8.11 – Пример оценки взаимосвязи выживаемости с изучаемыми факторами

### 5) Таблица базового риска в зависимости от временных периодов (Рис.8.12);

Таблица 1 – Значения базового риска летального исхода для разных временных периодов

Временные периоды	Значения базового риска $h_0(t)$ , %
1	0,908
2	1,853
2	2,806
8	4,794
9	4,794
11	4,794
12	7,345

Рисунок 8.12 – Пример таблицы базового риска

б) Таблица изменения рисков в зависимости от изучаемых факторов (Рис. 8.13);

Таблица 2 – Изменение рисков летального исхода в зависимости от влияния отдельных факторов

Фактор риска	Unadjusted		Adjusted	
	HR; 95% ДИ	p	HR; 95% ДИ	p
Виды лечения: 2 группа (гепарин)	4,243; 1,166 – 15,446	0,028*	4,243; 1,166 – 15,446	0,028*

\* – влияние предиктора статистически значимо (p < 0,05)

Рисунок 8.13 – Пример таблицы рисков в зависимости от влияния факторов

7) График оценки отношения рисков с 95% ДИ для изучаемых факторов (Рис. 8.14);

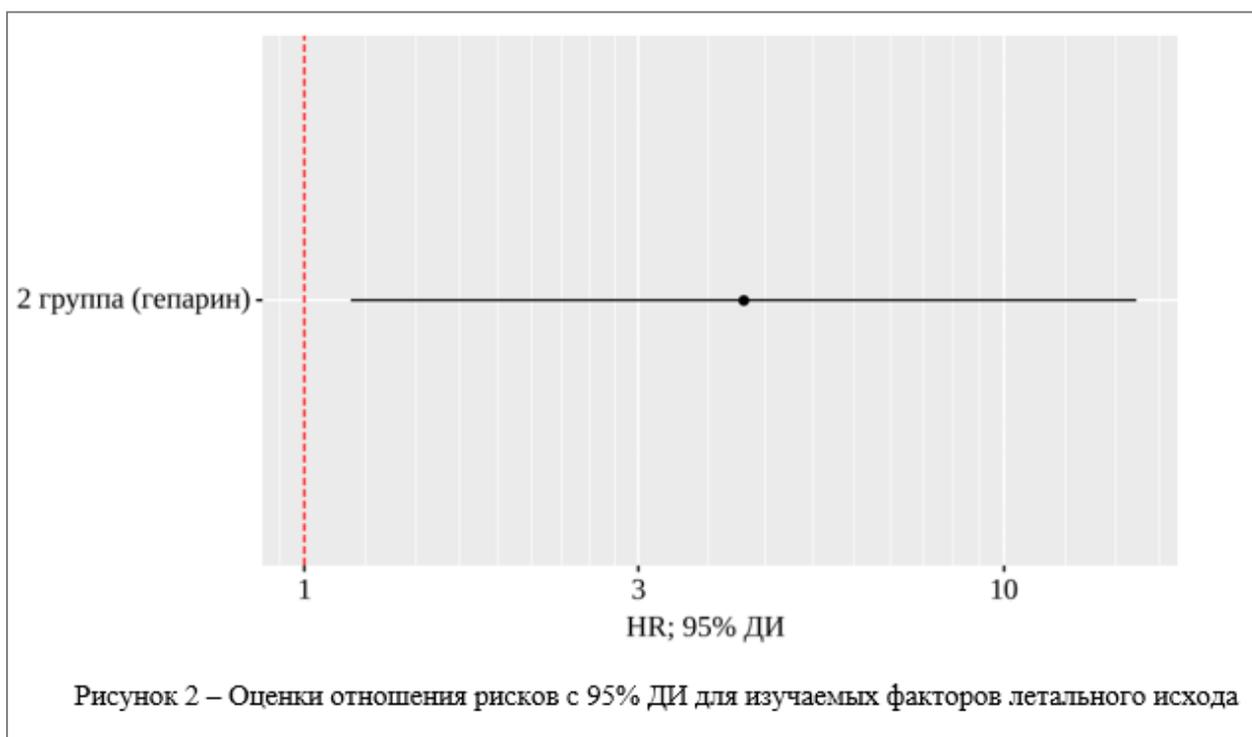


Рисунок 8.14 – Пример графика оценки отношения рисков

8) Таблицы значений безрецидивной выживаемости (Рис. 8.15).

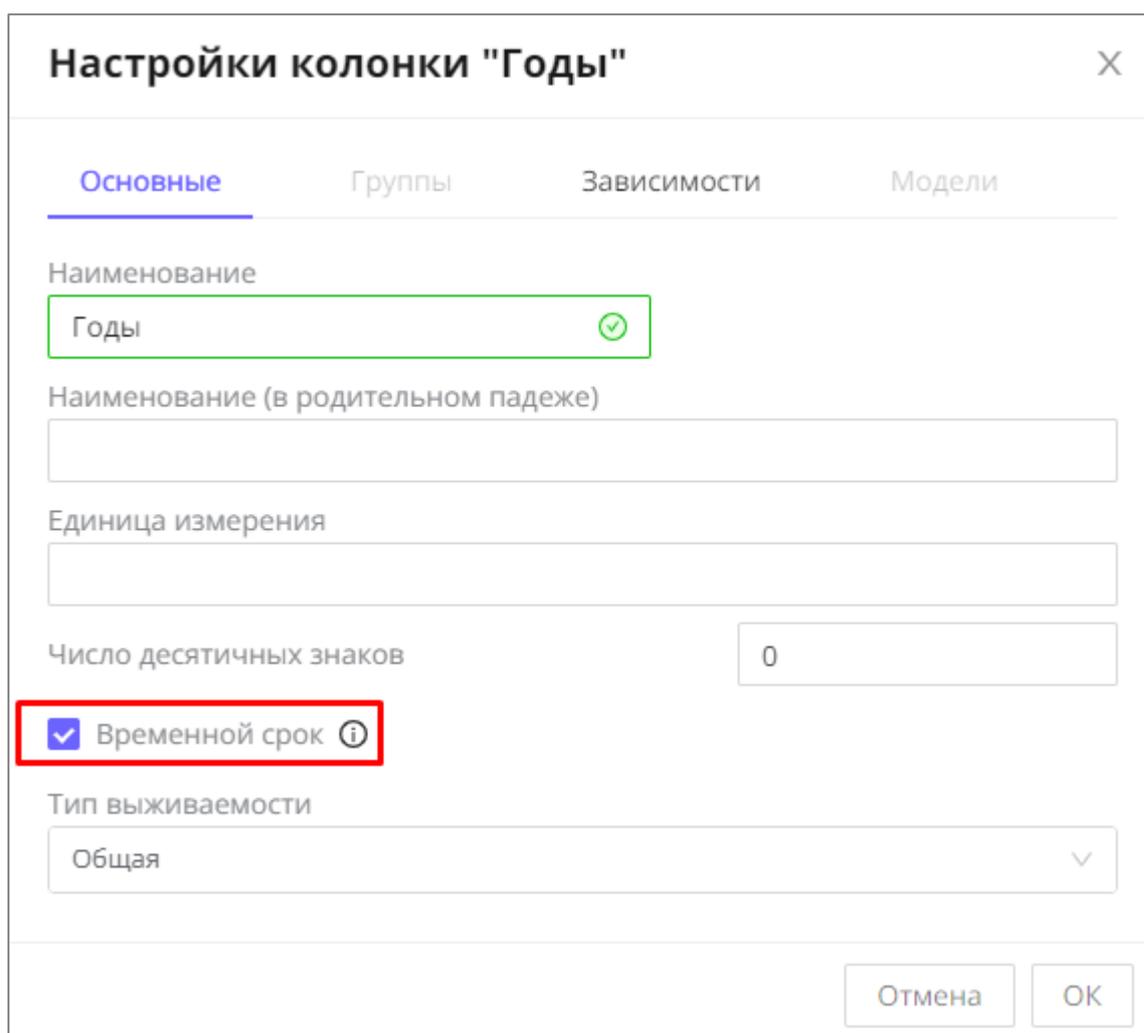
Срок наблюдения	Безрецидивная выживаемость	95% ДИ
0,0	100,0	100,0 – 100,0
5,0	83,3	61,5 – 93,4
10,0	61,7	39,1 – 78,0
15,0	49,8	27,1 – 69,0
20,0	12,5	0,8 – 40,7
25,0	0,0	0,0 – 0,0

Рисунок 8.15 – Таблица значений безрецидивной выживаемости

## 8.4. Динамические ряды

Для анализа и построения динамических рядов необходимо, чтобы в базе данных был столбец с явлением (например, число умерших) и столбец с периодом времени (например, года).

На столбце «Годы» заходим в настрой колонки и проставляем признак временного срока (Рис. 8.1).



Настройки колонки "Годы" X

Основные Группы Зависимости Модели

Наименование  
Годы ✓

Наименование (в родительном падеже)  
[Empty field]

Единица измерения  
[Empty field]

Число десятичных знаков 0

Временной срок ⓘ

Тип выживаемости  
Общая ▾

Отмена ОК

Рисунок 8.1 – Функция «Временной срок»

Затем переходим на вкладку «Зависимости» и проставляем галочку на колонке с событием, в данном случае, «Число умерших» и нажимаем кнопку «ОК» (Рис. 8.2).

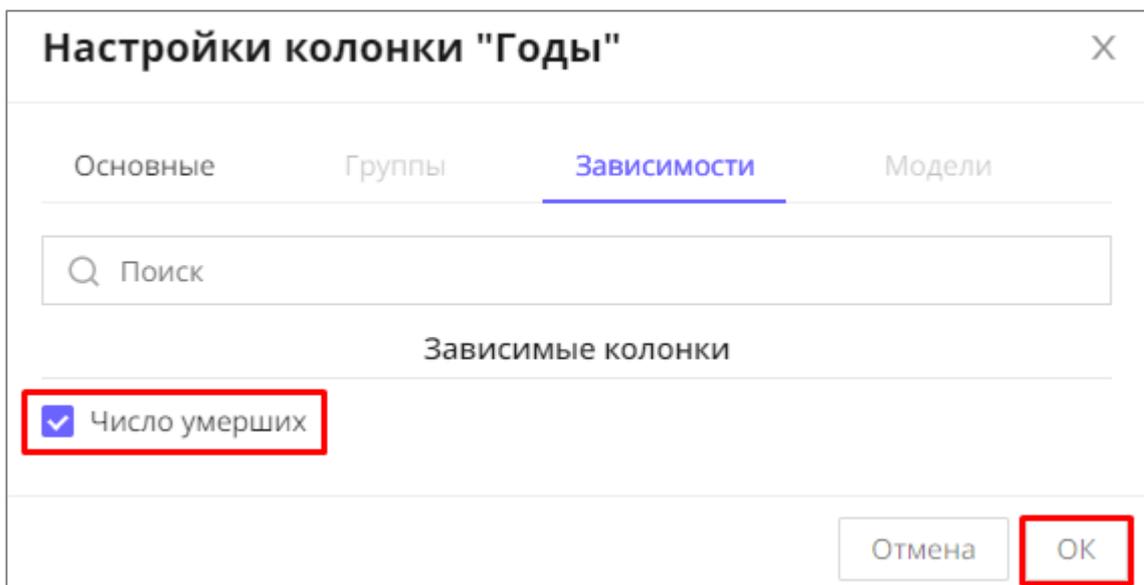


Рисунок 8.2 – Выбор зависимой колонки для построения динамических рядов

Далее необходимо нажать кнопку «Выполнить анализ» в верхнем правом углу страницы (Рис. 8.3).

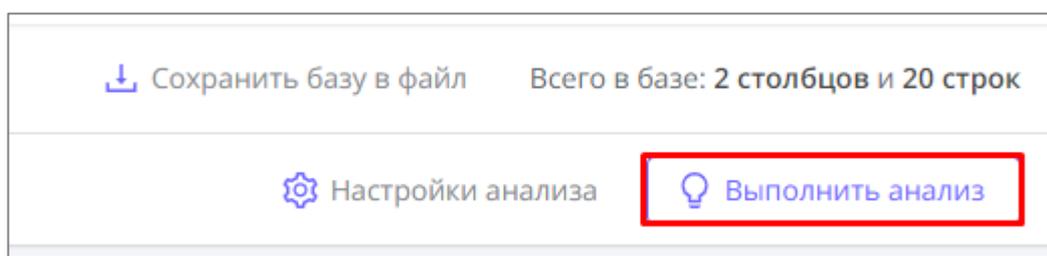


Рисунок 8.3 – Кнопка «Выполнить анализ»

Выполненный анализ будет состоять из таблицы анализа динамики показателя с событием (Рис. 8.4), текстового вывода (Рис. 8.5) и диаграммы (Рис. 8.6).

Таблица 1 – Анализ динамики показателя "Число умерших"

Годы	Число умерших	Абсолютный прирост (убыль)	Показатель наглядности, %	Показатель роста (снижения), %	Темп роста (снижения), %	Значение 1% прироста	Метод укрупнения интервала	Метод скользящей средней
2006	49218	–	100	–	–	–	45055	46613
2007	40892	-8326	83	83	-17	492		44800
2008	44290	3398	90	108	8	409	46136	44388
2009	47982	3692	97	108	8	443		47328
2010	49713	1731	101	104	4	480	48392	48255
2011	47071	-2642	96	95	-5	497		47714
2012	46358	-713	94	98	-2	471	46275	46540
2013	46192	-166	94	100	-0	464		46490
2014	46921	729	95	102	2	462	46700	46530
2015	46478	-443	94	99	-1	469		46116
2016	44949	-1529	91	97	-3	465	44504	45162
2017	44058	-891	90	98	-2	449		44538
2018	44608	550	91	101	1	441	43728	43838
2019	42848	-1760	87	96	-4	446		47244
2020	54276	11428	110	127	27	428	57288	52475
2021	60301	6025	123	111	11	543		61502

Рисунок 8.4 – Табличная форма анализа данных

При анализе динамики показателя "Число умерших" отмечался рост показателя с 2007 по 2010 - с 40892 до 49713 человек, с 2013 по 2014 - с 46192 до 46921 человек, с 2017 по 2018 - с 44058 до 44608 человек, с 2019 по 2021 - с 42848 до 60301 человек. Снижение показателя отмечалось с 2006 по 2007 - с 49218 до 40892 человек, с 2010 по 2013 - с 49713 до 46192 человек, с 2014 по 2017 - с 46921 до 44058 человек, с 2018 по 2019 - с 44608 до 42848 человек. Максимальное значение показателя "Число умерших" отмечалось в 2021 и составило 60301 человек. Минимальное значение показателя "Число умерших" отмечалось в 2007 и составило 40892 человек. Наивысший абсолютный прирост отмечался в 2020, когда он составил 11428 человек. Наивысшая в абсолютном значении убыль отмечалась в 2007, когда она составляла -8326 человек. Наивысший темп прироста отмечался в 2020, когда он составил 26,7%. Наивысший в абсолютном значении темп убыли отмечался в 2007, когда он составил -16,9%.

Рисунок 8.5 – Текстовый вывод к анализу данных

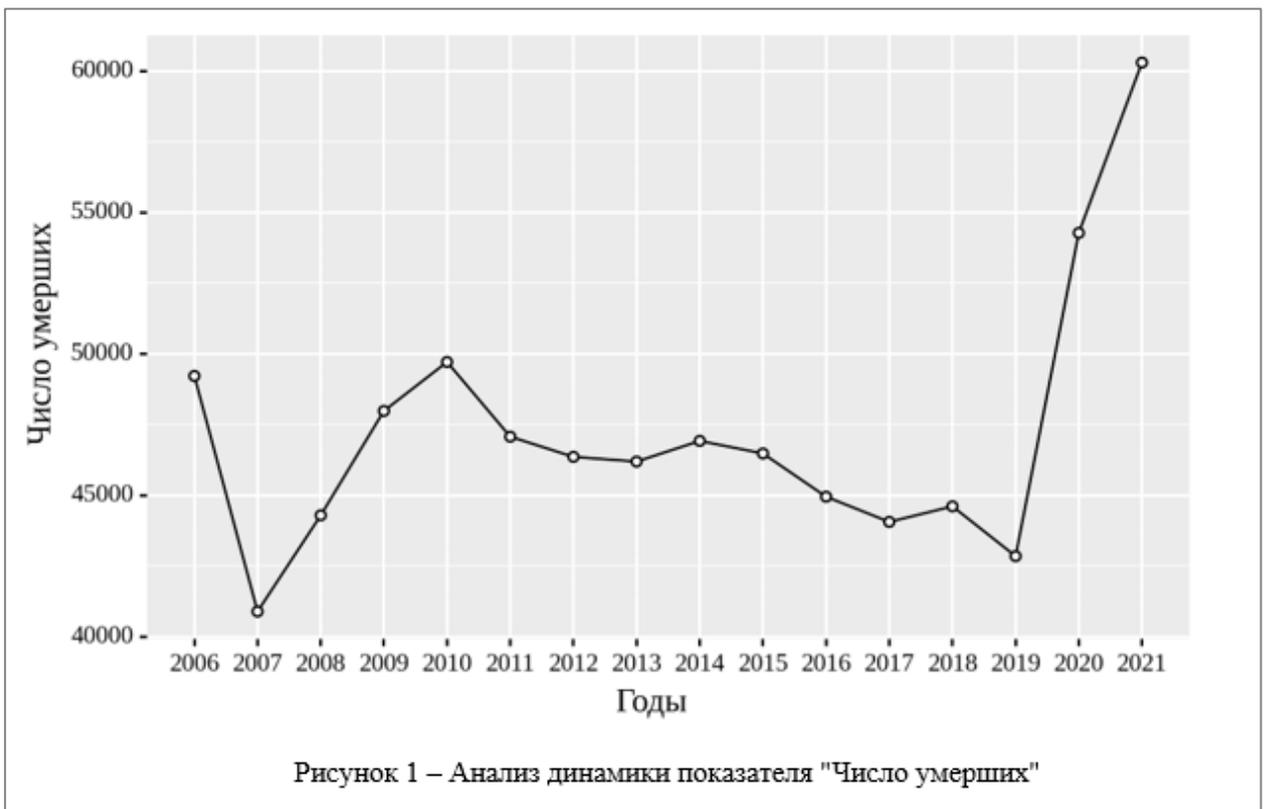


Рисунок 8.6 – Диаграмма анализа данных

## 8.5. Вычисление переменных

Для вычисления переменных необходимо нажать кнопку в верхней строке окна базы данных (Рис. 8.7).

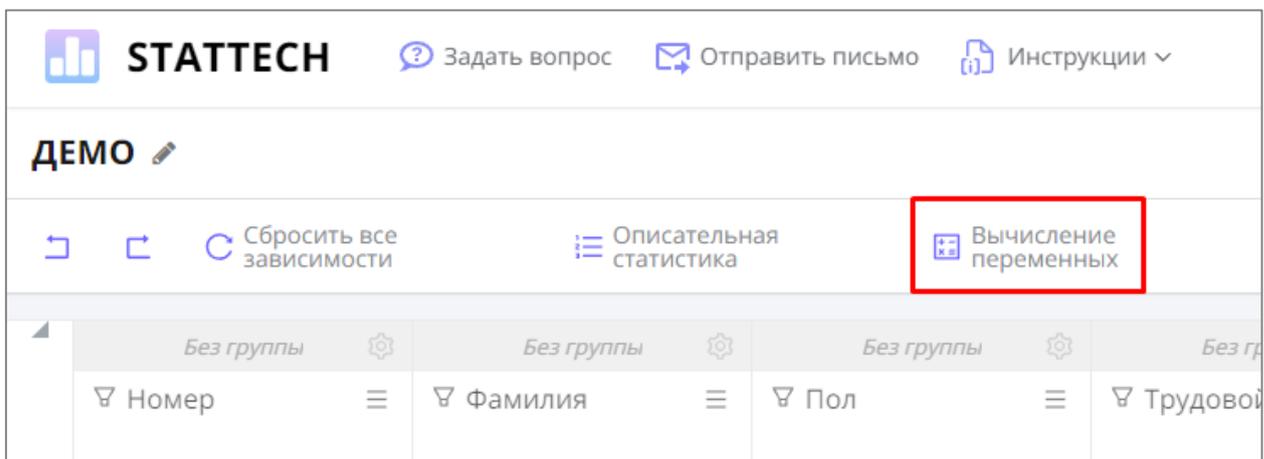


Рисунок 8.7 – Кнопка «Вычисление переменных»

После нажатия открывается окно, напоминающее калькулятор, где слева расположены «Переменные» (столбцы из базы данных), а справа функционал вычисления (Рис. 8.8).

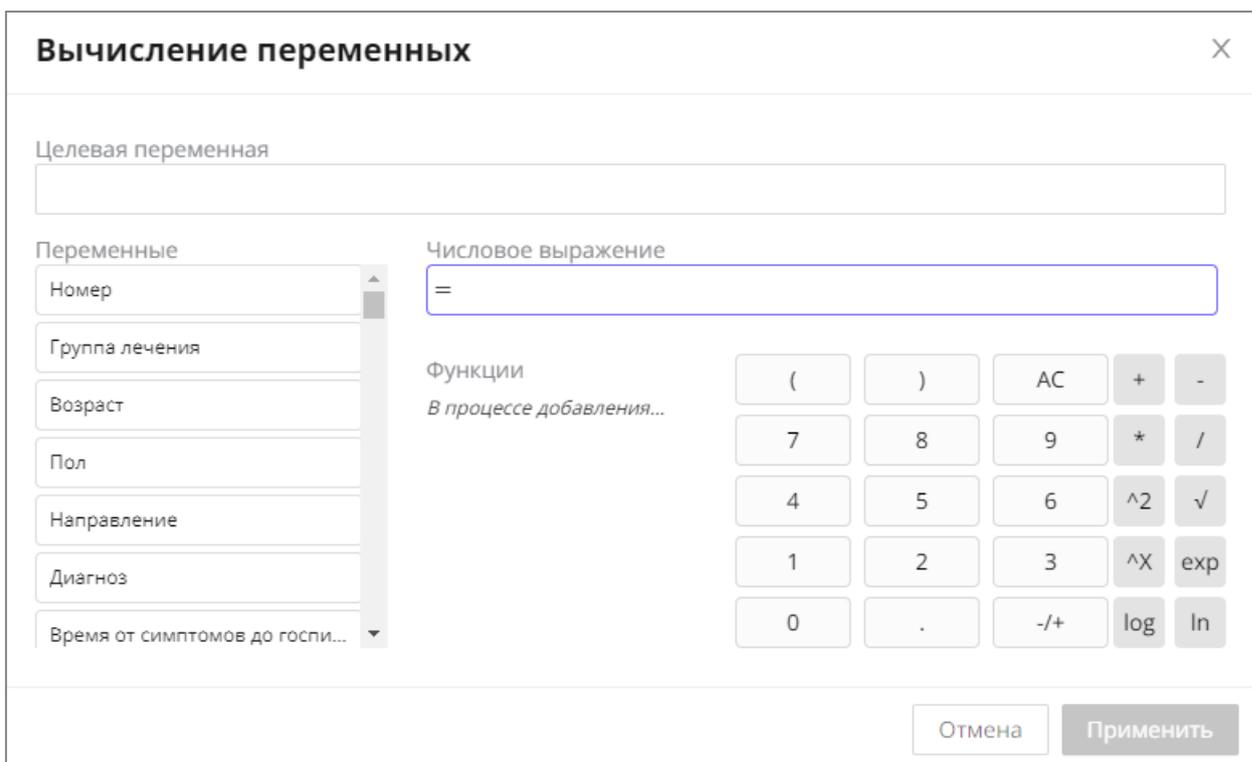


Рисунок 8.8 – Окно «Вычисления переменных»

Таблица клавиш вычисления переменных

AC	Кнопка очистки строки «Числовое выражение»
+	Сложение переменных (чисел)
-	Вычитание переменных (чисел)
*	Умножение переменных (чисел)
/	Деление переменных (чисел)
$\wedge 2$	Квадрат переменной (числа)
$\sqrt{\quad}$	Квадратный корень из переменной (числа)
$\wedge x$	Возведение в определенную степень
exp	Вычисление экспоненты
log	Функция десятичного логарифма
ln	Функция натурального логарифма
-/+	Позволяет сменить текущее положительное число на отрицательное, и наоборот
Разность дат	<p>Функция «Разность дат» (Правила использования отображены на рисунке ниже - Рис.8.8.1)</p> <p>Year – Год</p> <p>Quarter – Квартал</p> <p>Month – Месяц</p> <p>Week - Неделя</p> <p>Day – День</p> <p>Hour – Час</p> <p>Minute – Минута</p> <p>Second – Секунда</p> <p>Millisecond - Миллисекунда</p>

Использование:

**dateDiff("дата конца", "дата начала", year | quarter | month | week | day | hour | minute | second | millisecond)**

Например, выражение **dateDiff("01.01.2002", "01.01.2001", day)** вернет 365

Важно:

- В первом и втором аргументах можно использовать колонки и составные (вычисляемые) выражения. Можно использовать результат функции **dateDiff** в качестве аргумента.
- Все даты должны быть в формате **дд.мм.гггг**
- Все даты должны быть в кавычках. В случае, если аргументом не является строка с датой (т.е. это колонка и/или составное выражение, например, **dateDiff(<col1>, "01.01.2000", day)**), то кавычки не нужны.
- Все даты должны быть в одном формате
- Все даты должны быть в одной локали

Рисунок 8.8.1 – Правила использования функции «Разность дат»

В строке «Целевая переменная» необходимо написать название колонки, в которую будет рассчитана новая переменная. В строке «Числовое выражение» необходимо написать формулу для расчета. Например, для вычисления роста в метрах, берется уже имеющаяся в базе данных колонка «Рост» и делится на 100. Для завершения вычисления необходимо нажать на кнопку «Применить» (Рис. 8.9).

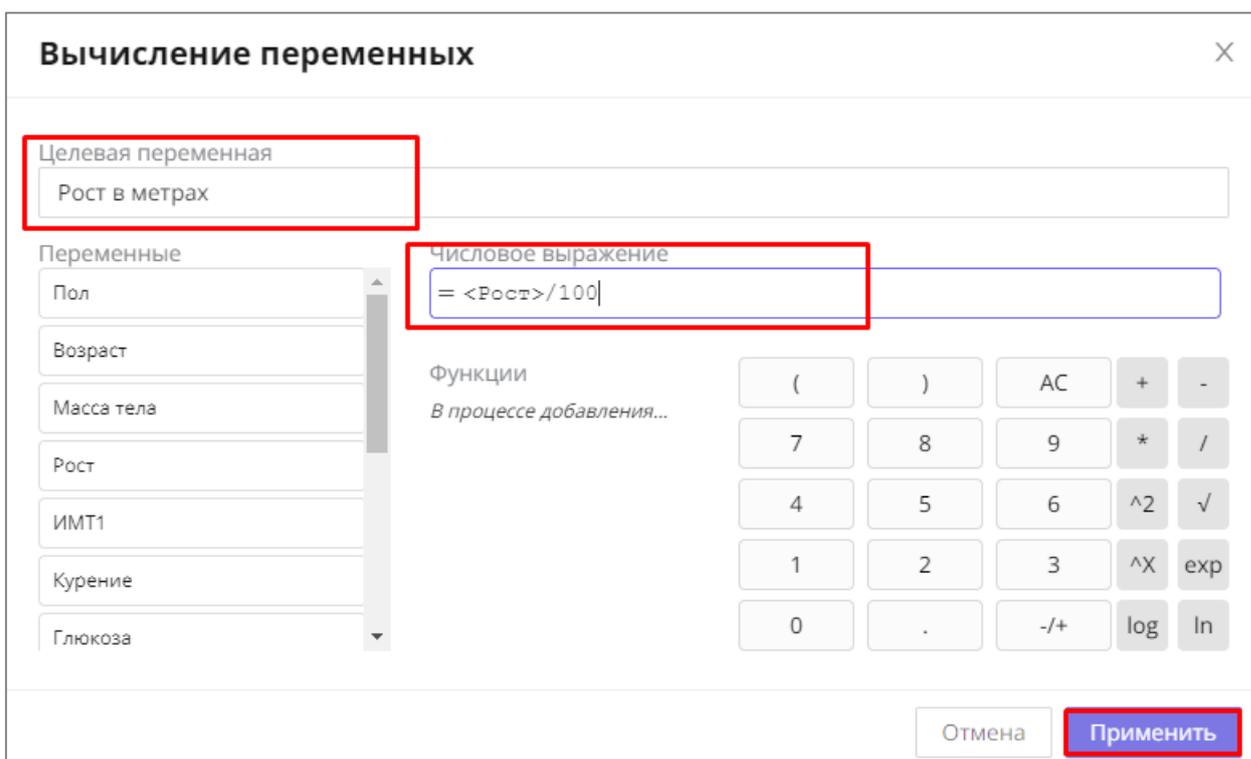


Рисунок 8.9 – Пример расчета новой переменной

Новая колонка добавляется в конец базы данных (Рис.8.10). У данных колонок всегда будет задан количественный тип переменной.

Без группы	АД в динамике	Без группы
▾ Препарат Категориальная	▾ АД после лечения Количественная	▾ Рост в метрах Количественная
0 (Без лечения)	113	1.75
0 (Без лечения)	118	1.76
0 (Без лечения)	121	1.8
0 (Без лечения)	123	1.8
0 (Без лечения)	125	1.67
0 (Без лечения)	130	1.89
0 (Без лечения)	132	1.68
0 (Без лечения)	133	1.68

Рисунок 8.10 – Пример созданной вычисляемой колонки в базе данных

## ГЛАВА 9. СОЗДАНИЕ ОПРОСА

Кнопка «Мои опросы» находится рядом с кнопкой «Мои базы» в личном кабинете (Рис.9.1).

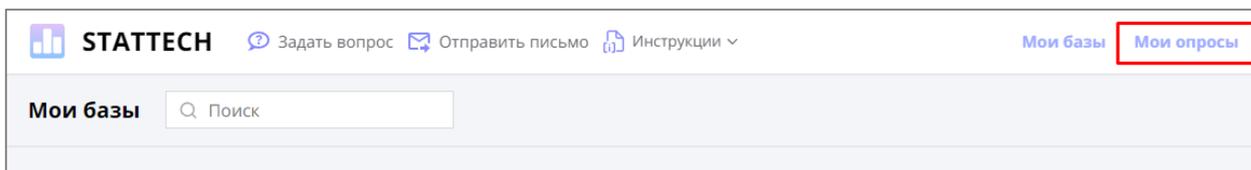


Рисунок 9.1 – Кнопка «Мои опросы»

В разделе «Мои опросы» находятся слоты на опросы. Новый опрос можно либо создать самим, либо загрузить из файла, нажав соответствующую кнопку (Рис. 9.2).

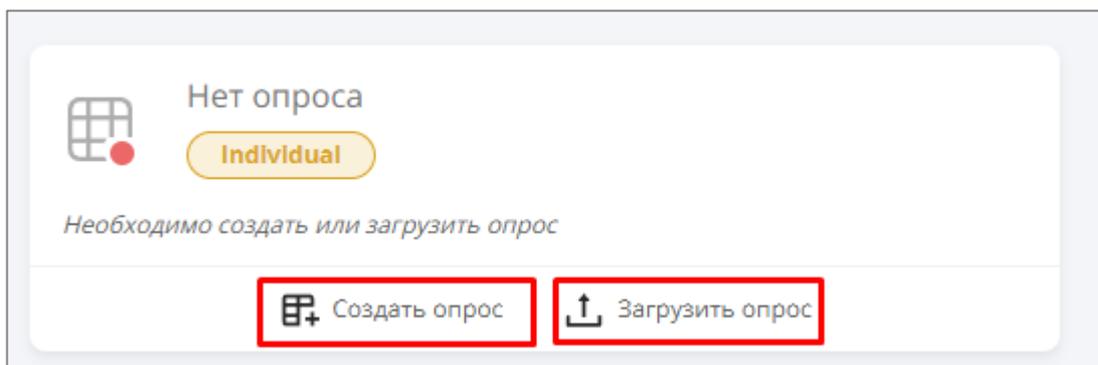


Рисунок 9.2 – Кнопки «Создать опрос» и «Загрузить опрос»

После нажатия кнопки «Создать опрос» откроется форма опросника, состоящего из 3х разделов: «Настройки», «Вопросы» и «Ответы». В разделе «Настройки» находится блок с наименованием опроса, возможность описания опросника и функция по сохранению опросника в файл (Рис. 9.3).

Опрос активен до: не установлено

## Оценка работы врачей УЗ диагностики

Уважаемые коллеги, просьба ответить на вопросы.

**Настройки**      Вопросы      Ответы

Наименование опроса

Оценка работы врачей УЗ диагностики

Описание опроса

Уважаемые коллеги, просьба ответить на вопросы.

Сохранить опрос в файл

Рисунок 9.3 – Раздел «Настройки» в опроснике

Раздел «Вопросы» включает в себя создание вопросов путем нажатия на соответствующую кнопку (Рис. 9.4).

Опрос активен до: не установлено

## Оценка работы врачей УЗ диагностики

Уважаемые коллеги, просьба ответить на вопросы.

Настройки      **Вопросы**      Ответы

+ Добавить вопрос

Рисунок 9.4 – Кнопка «Добавить вопрос»

После нажатия на кнопку откроется форма создания вопроса (Рис. 9.5).

Вопрос №1

Описание вопроса

Наименование колонки в базе

Название колонки в родительном падеже

Тип данных

Единица измерения

Число десятичных знаков

Количественная

0

Пользователи внесут свой ответ в виде числового значения

Показывать вопрос  Обязательный вопрос

Рисунок 9.5 – Форма создания вопроса

В верхней строке необходимо написать вопрос. В поле «Описание вопроса» можно дать развернутое пояснение / описание к вопросу. В строке «Наименование колонки в базе» и «Наименование колонки в родительном падеже» - заполняем так, как хотим, чтобы эти данные выводились в анализе. В поле «Тип данных» указываем тип в зависимости от вопроса. Например, если ответ подразумевает под собой число, то это количественный тип данных, если необходимо ответ выбрать из каких-то вариантов, то категориальный тип, если ответ на вопрос подразумевает какую-то дату, то тип данных будет дата. Примеры заполнения формы вопроса представлены ниже (Рис. 9.6, Рис. 9.7 и Рис. 9.8).

Вопрос №1

Укажите Ваш возраст

Описание вопроса

Необходимо указать количество полных лет на текущий момент

Наименование колонки в базе: Возраст

Название колонки в родительном падеже: возраста

Тип данных: Количественная

Единица измерения: полных лет

Число десятичных знаков: 0

Пользователи внесут свой ответ в виде числового значения

Показывать вопрос
  Обязательный вопрос

Рисунок 9.6 – Пример создания вопроса с ответами в формате количественного типа данных

Значок  внизу вопроса даёт возможность удалить весь вопрос, а знак  - скопировать. Снятие галочки «Обязательный вопрос»  «Обязательный вопрос» или «Показывать вопрос»  «Показывать вопрос» даёт возможность не отвечать или не показывать данный вопрос, соответственно.

Вопрос №2

Проводилась ли Вам компьютерная томография легких?

Описание вопроса

Наименование колонки в базе: КТ

Название колонки в родительном падеже: наличие КТ

Тип данных: Категориальная

0	Нет	не проводилась КТ	
1	Да	проводилась КТ	

Рисунок 9.7 – Пример создания вопроса с ответами в формате категориального типа данных (с выбором ответа)

Вопрос №3

Дата проведения обследования

Описание вопроса

Наименование колонки в базе: Дата КТ

Название колонки в родительном падеже: Дата КТ

Тип данных: Дата

Формат ввода даты: ДД.ММ.ГГГГ

Пользователи внесут свой ответ в виде выбора даты

Показывать вопрос  Обязательный вопрос

[Сохранить](#)

Рисунок 9.8 – Пример создания вопроса с ответами в формате даты

Раздел «Ответы» отображает счетчик с количеством ответов, выгруженными ответами в базу и ответами, ожидающими выгрузки. **В данном разделе также находится функция «Принимать ответы», которая делает опросник активным** (Рис. 9.9).

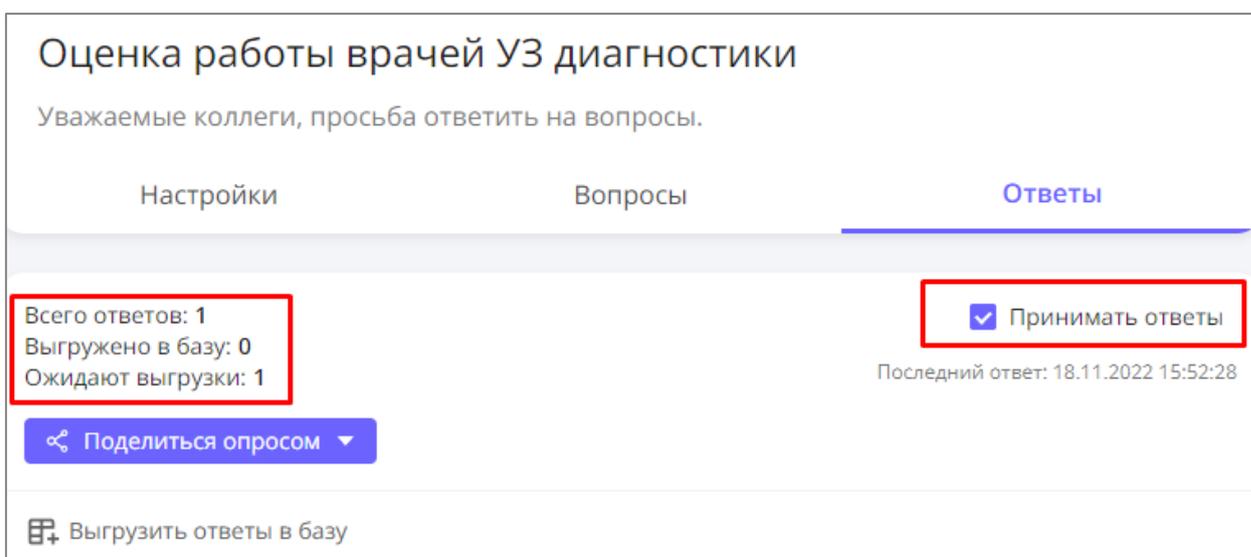


Рисунок 9.9 – Счетчик ответов и функция «Принимать ответы»

В данном разделе находится функция «Поделиться опросом». Можно поделиться либо ссылкой на опрос, либо QR-кодом (Рис. 9.10).



Рисунок 9.10 – Функция «Поделиться опросом»

Функция «Выгрузить ответы в базу» позволяет выгрузить ответы в имеющуюся базу данных в StatTech или в новый слот базы данных, либо скачать ответы в MS Excel (Рис.9.11).

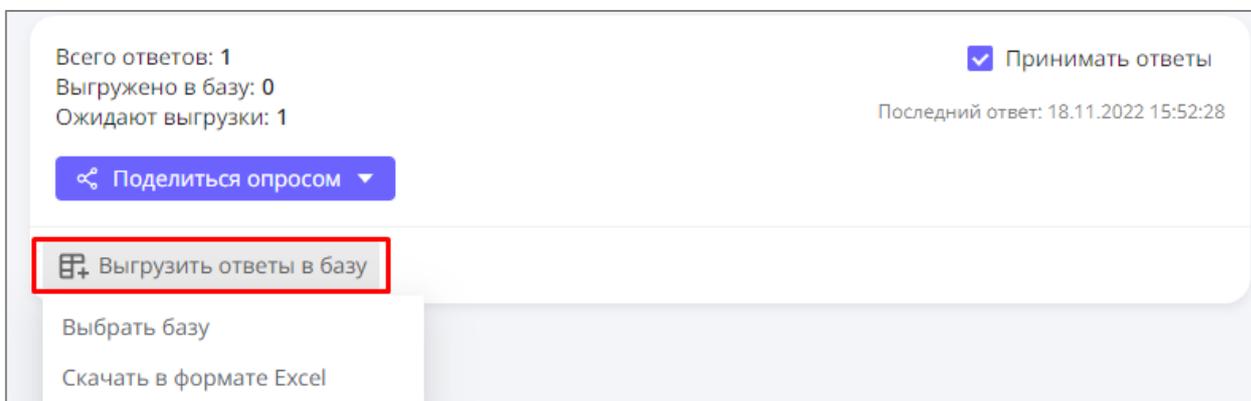


Рисунок 9.11 – Кнопка «Выгрузить ответы в базу»

При нажатии кнопки «Скачать в формате Excel» опросник выгрузится в таблицы Excel, где в шапке столбцов будут данные, которые написаны в строке «Наименование колонки в базе данных». При нажатии на кнопку «Выбрать базу» откроется окно с выбором слота базы данных. После выбора пустого слота необходимо нажать кнопку «Дополнить» (Рис. 9.12).

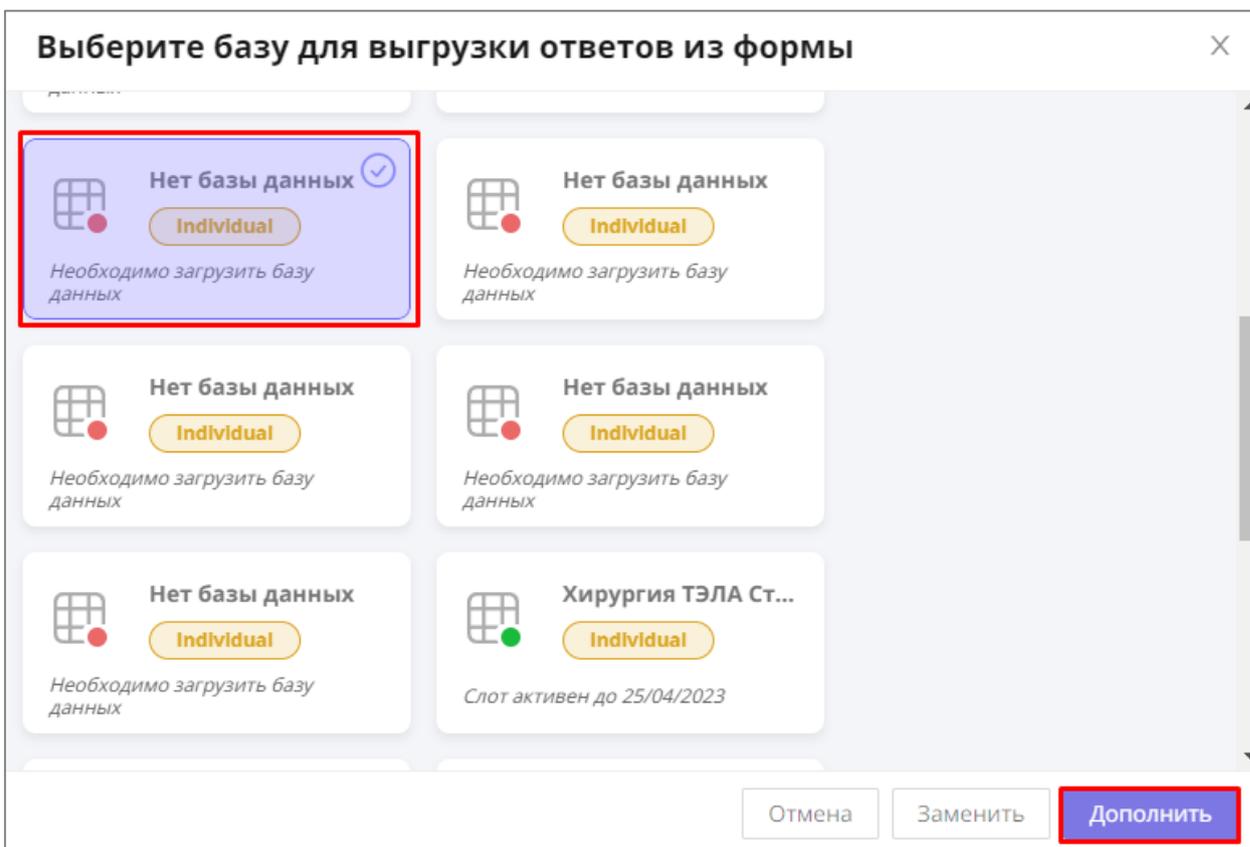


Рисунок 9.12 – Функционал выгрузки ответов в базу данных

При выборе слота базы данных, где уже имеются данные, необходимо нажать кнопку «Заменить». Старая база данных из слота будет перемещена в корзину и её можно будет восстановить в течение 30 дней (Рис. 9.13).

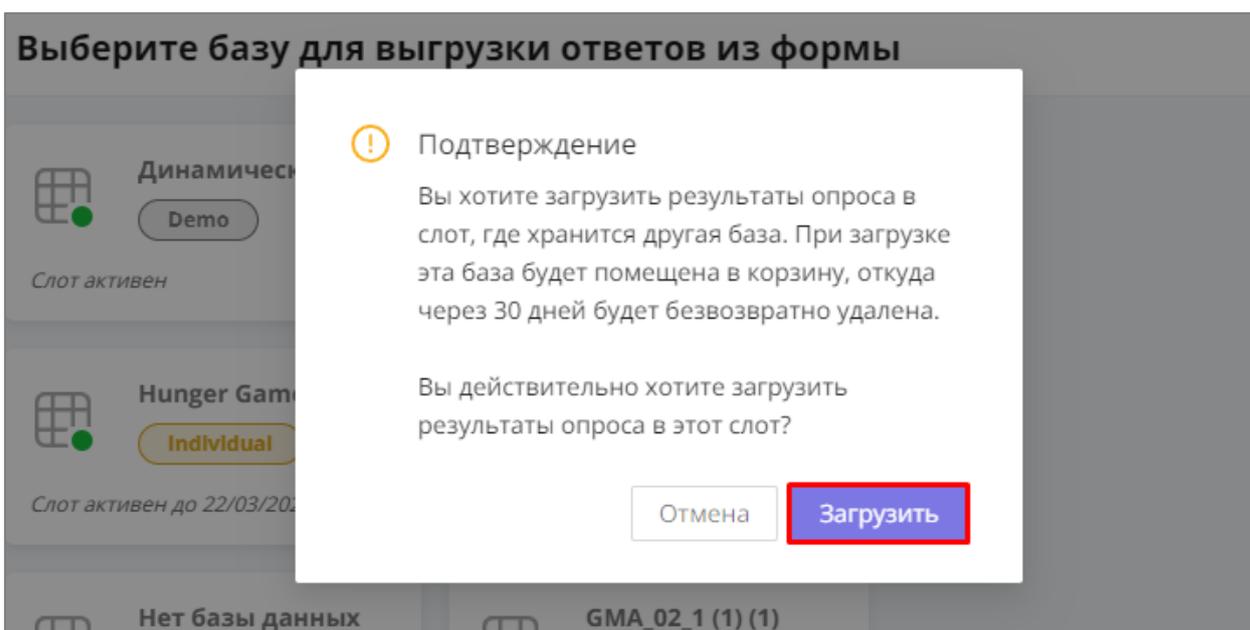


Рисунок 9.13 – Подтверждение загрузки ответов в заполненную базу данных

## ГЛАВА 10. КАЛЬКУЛЯТОР

Для того, чтобы зайти в калькулятор для расчёта объема выборки необходимо нажать в правом верхнем углу кнопку «Калькулятор» (Рис.10.1).

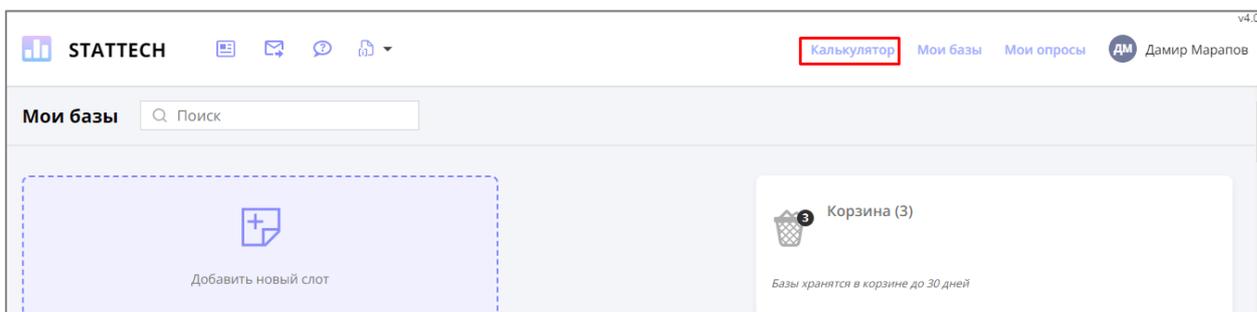


Рисунок 10.1 – Кнопка «Калькулятор»

После нажатия на кнопку откроется функционал работы с калькулятором (Рис.10.2).

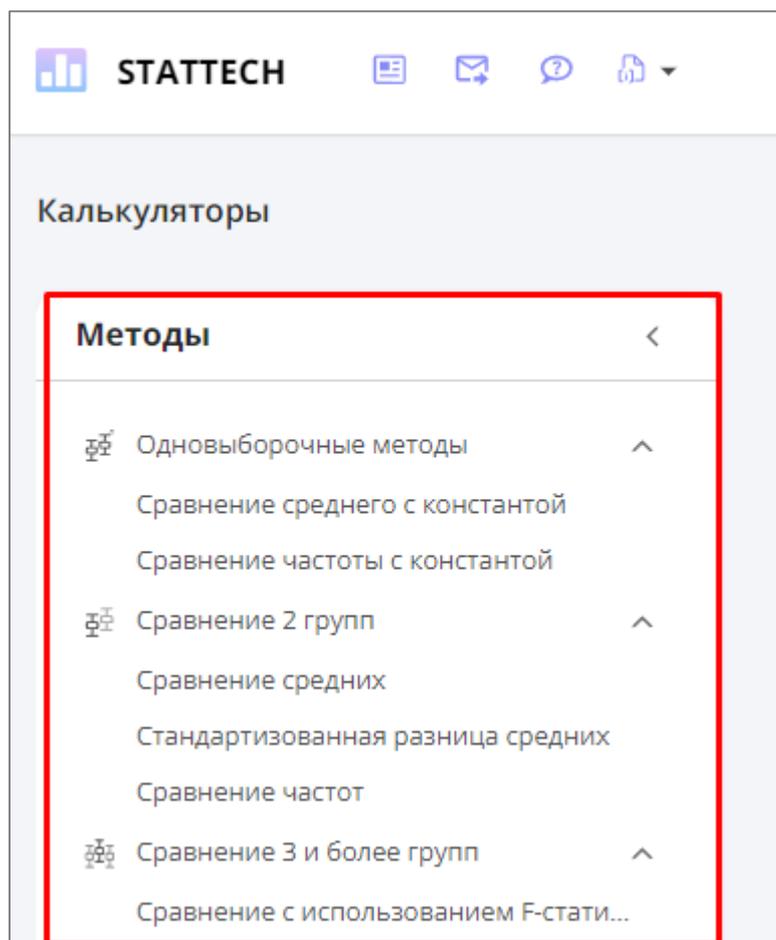


Рисунок 10.2 – Функционал калькулятора расчёта объема выборки

Выбираем необходимый метод расчета выборки, например, сравнение 2х групп, кликнув левой кнопкой мыши по методу. После этого откроется окно с параметрами, где необходимо заполнить данные для калькулятора (Рис.10.3).

The screenshot shows the 'STATTECH' application interface. On the left, a sidebar titled 'Методы' (Methods) lists various statistical methods. The 'Сравнение 2 групп' (Comparison of 2 groups) method is selected and highlighted in blue. The main panel, titled 'Параметры' (Parameters), contains the following settings:

- Вероятность ошибки I рода (Type I error probability): A slider set to 0.25.
- Мощность (Power): A slider set to 0.8.
- Среднее в группе 1 (Mean in group 1): Input field with value 30.
- Среднее в группе 2 (Mean in group 2): Input field with value 28.
- Стандартное отклонение в группе 1 (Standard deviation in group 1): Input field with value 12.00.
- Стандартное отклонение в группе 2 (Standard deviation in group 2): Input field with value 8.00.
- Отношение размеров групп 1 и 2 (Ratio of group sizes 1 and 2): Input field with value 1.00.
- Альтернативная гипотеза (Alternative hypothesis): Radio button selected for 'Двусторонняя' (Two-sided).

Рисунок 10.3 – Параметры метода расчёта объема выборки

После заполнения параметров необходимо нажать на кнопку «Рассчитать» внизу параметров (Рис.10.4).

The screenshot shows the 'STATTECH' application interface. On the left, a sidebar titled 'Методы' (Methods) lists various statistical methods. The 'Корреляция' (Correlation) method is selected and highlighted in blue. The main panel, titled 'Параметры' (Parameters), contains the following settings:

- Практически незначимая разница (Practically insignificant difference): Input field with value 0.
- Альтернативная гипотеза (Alternative hypothesis): Radio buttons for 'Эквивалентность' (Equivalence), 'Неменьшая эффективность' (Not less effective), and 'Превосходство' (Superiority).

The 'Рассчитать' (Calculate) button is highlighted with a red box.

Рисунок 10.4 – Кнопка «Рассчитать»

После этого отобразятся «Результаты» полученных расчетов объема выборки (Рис.10.5).

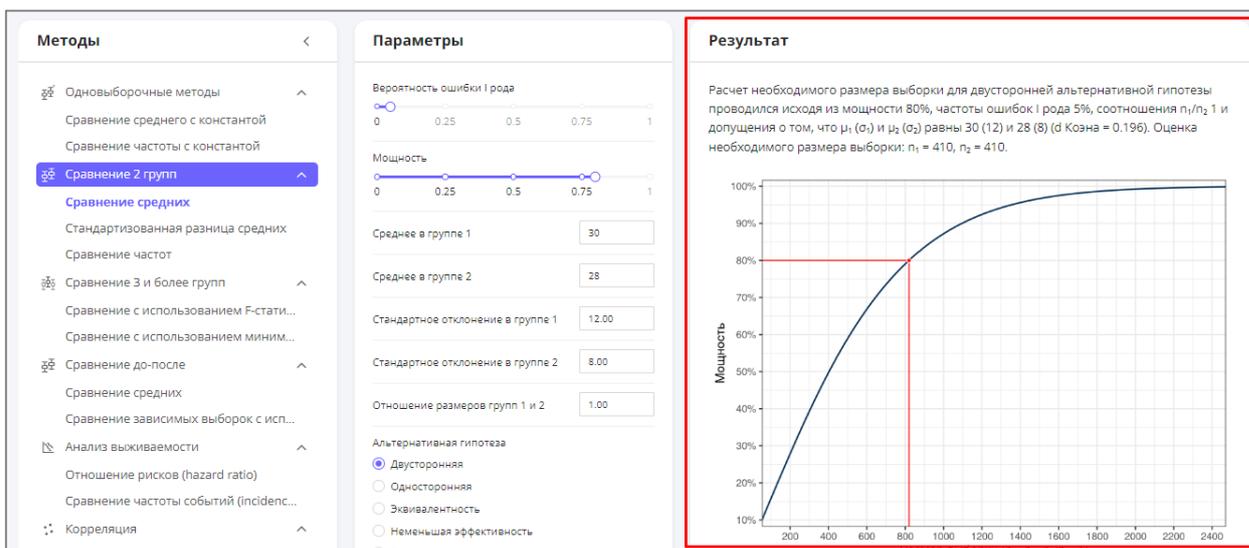


Рисунок 10.5 – Результат полученных расчётов