



# Особенности формирования базы участников исследования в электронной таблице, загрузка базы в статистические программы

**Аглиуллина Саида Тахировна**  
доцент кафедры эпидемиологии и доказательной  
медицины ФГБОУ ВО Казанский ГМУ  
Минздрава России, к.м.н.

[saida.agliullina@kazan-gmu.ru](mailto:saida.agliullina@kazan-gmu.ru)

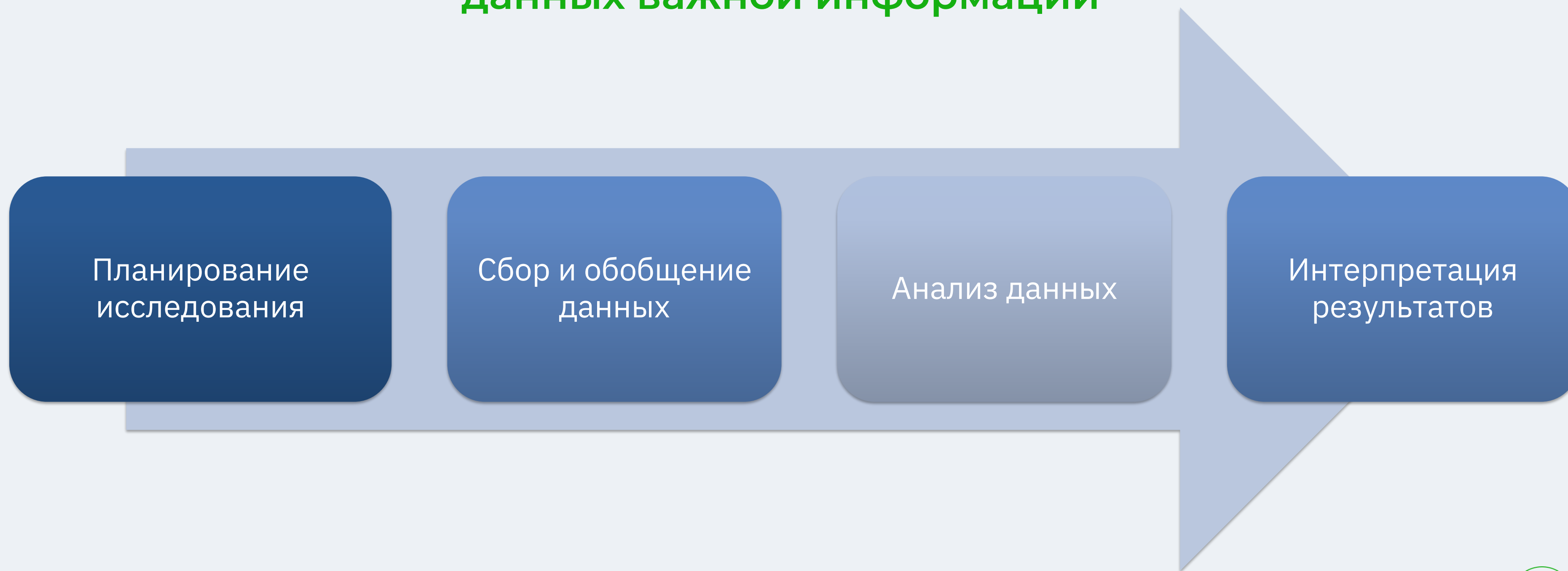


## План

1. Типы переменных в исследовании
2. Особенности формирования базы данных исследования
3. Ресурсы для создания базы данных исследования (электронные таблицы, Yandex Forms и др.)
4. Загрузка базы в статистические программы (StatTech и др.)



## Цель исследований – получение данных из выборки индивидуумов и извлечение из этих данных важной информации



# Планирование исследования

Обоснование актуальности проведения исследования

Формулирование конечной и промежуточных целей



Выбор объекта и единицы исследования



Составление программы исследования

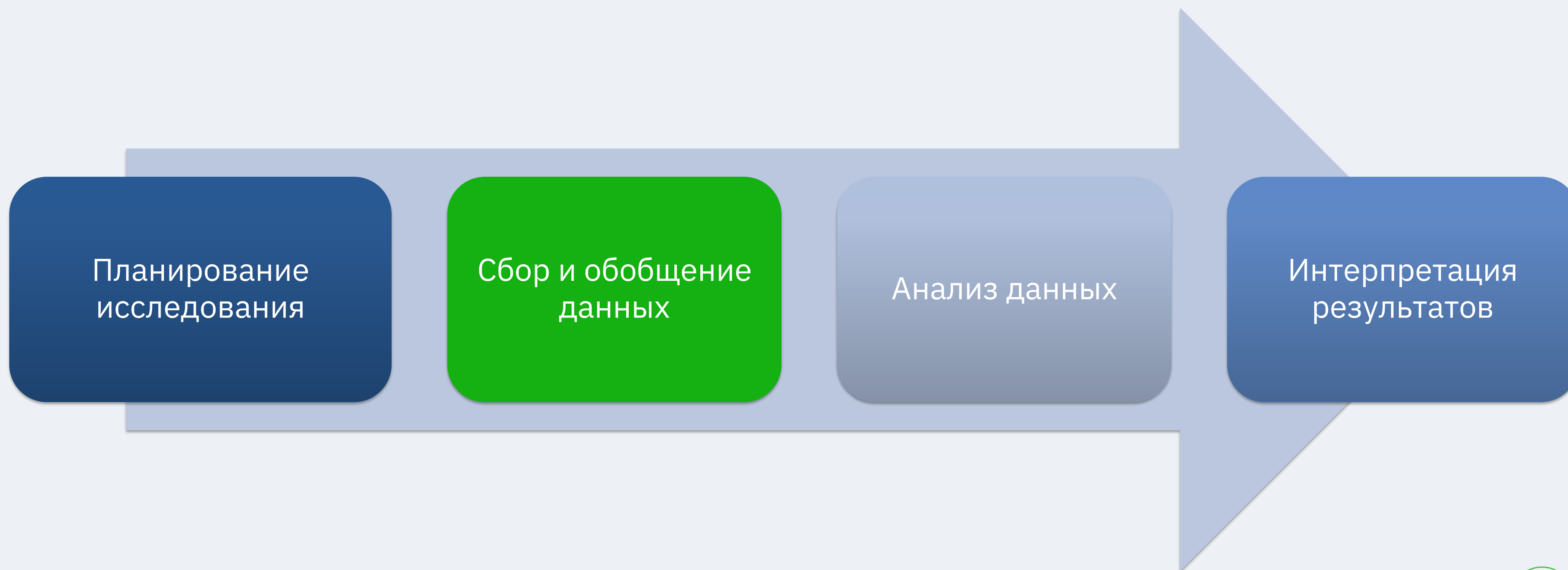


Составление плана исследования





## Проведение исследования



# Типы переменных в исследовании





## Количество обострений хронического холецистита в течение последнего года

- а. мультиномиальные
- б. порядковые
- в. дискретные
- г. непрерывные
- д. бинарные







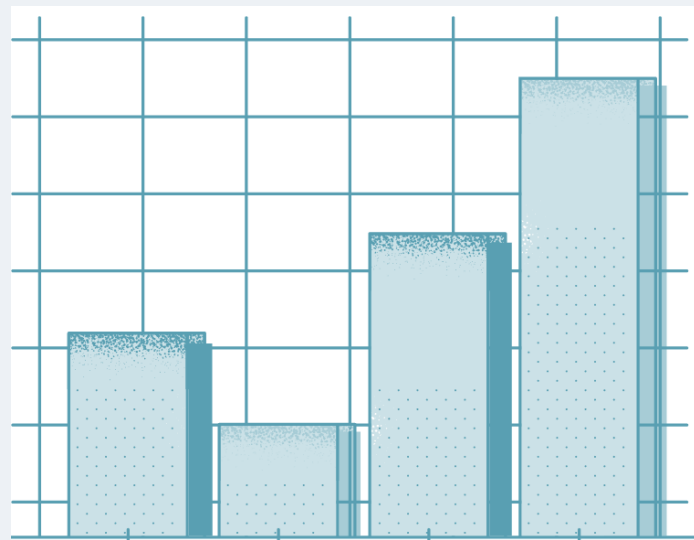
## Факт возникновения или отсутствия случая заболевания у вакцинированного лица в исследуемой группе за обозначенный период времени

- а. мультиномиальные
- б. порядковые
- в. дискретные
- г. непрерывные
- д. бинарные









# Особенности формирования базы данных исследования



1. Первая переменная (первый столбец) – номер пациента/респондента
2. Каждый столбец = отдельная переменная  
*(пол/возраст/уровень глюкозы и др.)*
3. Каждая строка данных = отдельный индивидуум (пациент/респондент)
4. Кодировать переменные в виде чисел,  
*за исключением ФИО, буквенный код пациента и дат.*
5. Кодировка бинарных данных – использовать значения 0 и 1 (но допускается 1 и 2).  
*Например, женский пол-0, мужской пол-1.*
6. Кодировка мультиномиальных и порядковых данных – использовать значения 1, 2, 3 и тд.
7. 1 ячейка – 1 число.
  - *Не допускать перечисления значений в одной ячейке (например, при множественном ответе на вопрос: «1,3,7,10»). Необходимо создать отдельные столбцы для каждого варианта ответа.*
  - *Значения количественных переменных без указания единиц измерения.*
8. Нет информации о пациенте/респонденте – пропуск (не 0!)



## При кодировке рекомендовано представить мультиномиальные переменные в виде порядковых

### Степень ИМТ (мультиномиальная)

1. Дефицит массы тела
2. Избыточная масса тела
3. Ожирение
4. Норма



### Степень ИМТ (порядковая)

1. Дефицит массы тела
2. Норма
3. Избыточная масса тела
4. Ожирение

# Примеры кодировки категорий номинальных и порядковых данных



## Пол (номинальная бинарная)

- 0-женщины
- 1-мужчины

## Степень ИМТ (порядковая)

- 1-Норма
- 2-Избыточная масса тела
- 3-Ожирение

## Семейный статус

### (мультиномиальная)

- 1-В официальном браке
- 2-В гражданском браке
- 3-Не замужем/холост
- 4-Разведен(а)
- 5-Вдов(ц)ы

# Ресурсы для создания базы данных исследования



Электронные  
таблицы

	A	B	C	D	E	F	G	H
№	фамилия	м-1, ж-0	возраст	рост	вес	Курение	ЯК лег -1,	
1	Иванова	ж	48 лет	162 см	75	курит	2	
2	Иванов	муж	39 лет	190 см	82	не курит	2	
3	Петрова	ж	40	162 см	75	курит	2	
4	Иванов	муж	58	174 см	94	не курит	4	
5	Иванов	м	50	176 см	68	курит	2	
6	Сидорова	ж	34 г.	160 см	52	не курит	2	
7	Павлов	муж.	41	166 м	74	не курит	2	
8	Иванов	м	58 лет	167 см	62	не курит	2	



Yandex Forms

Как выглядит шаблон В конструкторе У пользователя Создать форму из шаблона

Страница 1

- \* Ваш пол  
 Один вариант
- \* Сколько вам лет?  
 Один вариант
- \* Вы учитесь или работаете?  
 Один вариант
- \* Сколько у вас свободного времени?  
 Один вариант
- Что вы обычно делаете в свободное время?  
 Несколько вариантов



Статистические  
программы

	A	B	C							
1										
2										
3										
4										
5										
6										
7										
8										
9										
10										
11										



# Пример плохой базы данных исследования

	A	B	C	D	E	F	G	H	I	J	K	L
1	№	фамил	Пол	возрас	рост	вес	Курени	Степень тяжести холецистита лег -1, сред 2 , тяж 3	вздутие	боль в	Частота стула : с	Сопут. Заб. 1-ССЗ, 2-ЖКТ, 3-эндок., 4-ОДА
2	1	Иванова	ж	48 лет	162 см	75	курит	сред	1	3	1	1,2,3
3	2	Иванов	муж	39 лет	190 см	82	не курит	ср.	3	4	2	1
4	3	Петрова	ж	40	162 см	75	курит	сред	7	8	3	
5	4	Иванов	муж	58	174 см	94	не курит	тяжелая	0	1	1	1
6	5	Иванов	м	50	176 см	68	курит	сред.	2	3	2	
7	6	Сидорова	ж	34 г.	160 см	52	не курит	сред	6	5	2	3
8	7	Павлов	муж.	41	166 м	74	не курит	сред	4	5	3	
9	8	Иванов	м	58 лет	167 см	62	не курит	сред	4	4	2	
10	9	Петров	муж.	30 лет	172	66	не курит	легкая	4	4	1	1
11	10	Дроздова	жен.	31	53	164	не курит	средней	4	2	1	
12	11	Хасанов	муж.	45	171	64	не курит	сред	2	3	1	
13	12	Иванов	м	42	172	75	не курит	сред	9	6	3	1,2
14	13	Петров	муж.	32	174	75	не курит	сред	3	4	1	
15	14	Закирова	ж	32	162	56	не курит	сред	0	2	1	
16	15	Иванов	муж.	50	173	70	не курит	сред	3	2	1	
17	16	Иванова	ж	50	162	55	не курит	легк	3	5	2	3,4
18	17	Петрова	м	25	183	75	кур.	сред	3	4	3	
19	18	Сидорова	ж	59	164	64	курит	сред	3	4	1	4
20	19	Петрова	ж	65	160	77	курит	сред	5	0	1	4



# Коррекция базы данных исследования

	A	B	C	D	E	F	G	H
1	№	фамил	Пол	возрас	рост	вес	Курени	Степень тяжести холецистита (лег -1, сред 2 , тяз 3)
10	9	Петров	муж.	30 лет				
17	16	Иванова	ж	50				
21	20	Сидорова	ж	67				
29	28	Иванов	муж.	49				
30	29	Петров	м	30				
31	30	Дмитрие	м	19				
34	33	Петрова	м	61	170			
38	37	Пьянова	Жен	50				
42	41	Петров	м	30				

Сортировка от А до Я

Сортировка от Я до А

Сортировка по цвету

Удалить фильтр из столбца "Степень тяжести х..."

Фильтр по цвету

Текстовые фильтры

Поиск

ср

ср.

сред

сред.

средней

OK Отмена

	A	B	C	D	E	F	G	H
1	№	фамил	Пол	возрас	рост	вес	Курени	Степень тяжести холецистита (лег -1, сред 2 , тяз 3)
2	1	Иванова	ж	48 лет	162 см	75	курит	2
3	2	Иванов	муж	39 лет	190 см	82	не курит	ср.
4	3	Петрова	ж	40	162 см	75	курит	сред
6	5	Иванов	м	50	176 см	68	курит	сред.
7	6	Сидорова	ж	34 г.	160 см		не курит	сред
8	7	Петрова	ж	44	166 см		курит	сред
9	8	Иванова	ж	38	160 см		курит	сред
13	12	Иванов	м	42	172		не курит	сред
14	13	Петров	муж.	32	174		не курит	сред
15	14	Закирова	ж	32	162	56	не курит	сред
16	15	Иванов	муж.	50	173	70	не курит	сред
18	17	Петрова	м	25	183	75	кур.	сред
19	18	Сидорова	ж	59	164	64	курит	сред
20	19	Петрова	ж	65	160	37	курит	сред
22	21	Иванова	жен	52	157	67	н/к	ср
23	22	Петрова	ж	46	170	52	курит	ср



# Пример хорошей базы данных исследования

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	№	Фамилия	Пол (ж-0, м-1)	Возраст, полных лет	Рост, см	Масса тела, кг	Курение	Степень тяжести холецистита (лег -1, сред 2, тяж 3)	Вздутие живота, баллы по ВАШ	Боль в животе, баллы по ВАШ	Частота стула (стул до 4 -1, больше или равно 4 -2, больше или равно 6 -3)	ССЗ	ЖКТ	эндок	ОДА
2	1	Иванова	0	48	162	75	1	2	1	3	1	1	1	1	0
3	2	Иванов	1	39	190	82	0	2	3	4	2	1	0	0	0
4	3	Петрова	0	40	162	75	1	2	7	8	3	0	0	0	0
5	4	Иванов	1	58	174	94	0	3	0	1	1	1	0	0	0
6	5	Иванов	1	50	176	68	1	2	2	3	2	0	0	0	0
7	6	Сидорова	0	34	160	52	0	2	6	5	2	0	0	1	0
8	7	Павлов	1	41	166	74	0	2	4	5	3	0	0	0	0
9	8	Иванов	1	58	167	62	0	2	4	4	2	0	0	0	0
10	9	Петров	1	30	172	66	0	1	4	4	1	1	0	0	0
11	10	Дроздова	0	31	164	53	0	2	4	2	1	0	0	0	0
12	11	Хасанов	1	45	171	64	0	2	2	3	1	0	0	0	0
13	12	Иванов	1	42	172	75	0	2	9	6	3	1	1	0	0
14	13	Петров	1	32	174	75	0	2	3	4	1	0	0	0	0
15	14	Закирова	0	32	162	56	0	2	0	2	1	0	0	0	0
16	15	Иванов	1	50	173	70	0	2	3	2	1	0	0	0	0
17	16	Иванова	0	50	162	55	0	1	3	5	2	0	0	1	1



# Перед выгрузкой базы в статистическую программу **проверьте базу на наличие ошибок, опечаток и пропусков**

№	фамили	Пол (0-	возрас	рост	в
17	Петрова	1	25	183	
18	Сидорова	0	59	164	
19	Петрова	0	65	160	

№	фамили	Пол (0-	возрас	рост	вес	Курени
10	Дроздова	0	31	53	164	2

# База данных исследования готова к загрузке в статистические программы

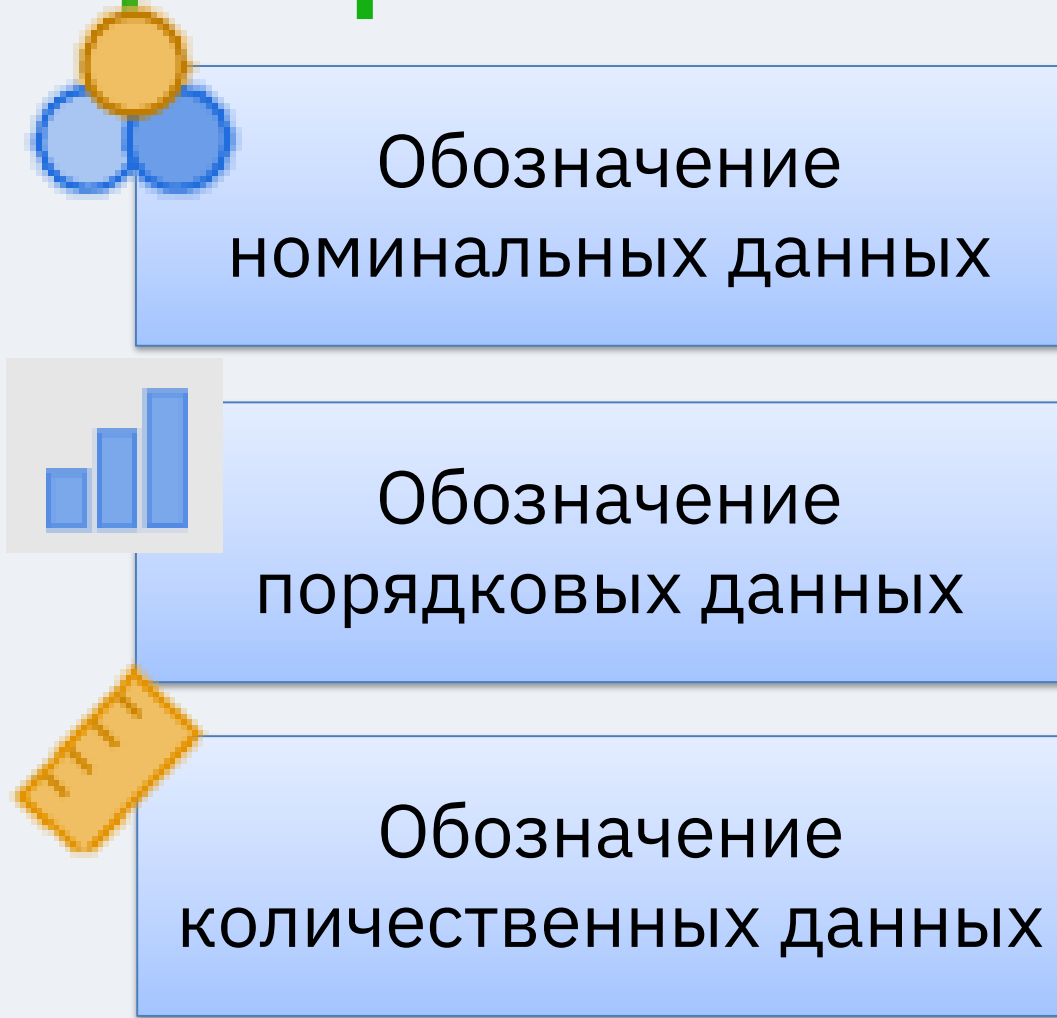
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	№	Фамилия	Пол (ж-0, м-1)	Возраст, полных лет	Рост, см	Масса тела, кг	Курение	Степень тяжести холецистита (лег -1, сред 2, тяж 3)	Вздутие живота, баллы по ВАШ	Боль в животе, баллы по ВАШ	Частота стула (стул до 4 -1, больше или равно 4 -2, больше или равно 6 -3)	ССЗ	ЖКТ	эндок	ОДА
2	1	Иванова	0	48	162	75	1	2	1	3	1	1	1	1	0
3	2	Иванов	1	39	190	82	0	2	3	4	2	1	0	0	0
4	3	Петрова	0	40	162	75	1	2	7	8	3	0	0	0	0
5	4	Иванов	1	58	174	94	0	3	0	1	1	1	0	0	0
6	5	Иванов	1	50	176	68	1	2	2	3	2	0	0	0	0
7	6	Сидорова	0	34	160	52	0	2	6	5	2	0	0	1	0
8	7	Павлов	1	41	166	74	0	2	4	5	3	0	0	0	0
9	8	Иванов	1	58	167	62	0	2	4	4	2	0	0	0	0
10	9	Петров	1	30	172	66	0	1	4	4	1	1	0	0	0
11	10	Дроздова	0	31	164	53	0	2	4	2	1	0	0	0	0
12	11	Хасанов	1	45	171	64	0	2	2	3	1	0	0	0	0
13	12	Иванов	1	42	172	75	0	2	9	6	3	1	1	0	0
14	13	Петров	1	32	174	75	0	2	3	4	1	0	0	0	0
15	14	Закирова	0	32	162	56	0	2	0	2	1	0	0	0	0
16	15	Иванов	1	50	173	70	0	2	3	2	1	0	0	0	0
17	16	Иванова	0	50	162	55	0	1	3	5	2	0	0	1	1



# Обозначение типов данных в различных статистических программах

1

A	B	C



2

	Имя	Тип	Ширина	Десятич...	Метка	Значения	Пропущенн...	Столбцы	Выравнивание	Шк
1	Возраст	Числовой	8	2		Нет	Нет	8	По право...	Шкалы
2										Шкалы
3										Порядковые
4										Номинальны
5										

Шкала

Шкалы

Порядковые

Номинальны

STATTECH Задать вопрос Отправить письмо Видео инструкции Мои базы SA Saida A. Individual v2.6.5


Поиск

**Demo** Нет базы данных **+** ЗАГРУЗИТЬ БАЗУ ДАННЫХ

**Individual** База: База данных ВИЧ (Своевременность диагностики)\_PT для статтех2 (до 15/09/2022) **ПРОДЛИТЬ** **УДАЛИТЬ БАЗУ** **ОТКРЫТЬ**

**Загрузка базы данных** X

Новая база данных



Нажмите или перетащите файл

Отмена **OK**





# Обозначение типов данных в StatTech

База данных ВИЧ (Своевременность диагностики)\_РТ для статтех2

Сохранить базу в файл Всего в базе: 13 столбцов и 1073 строк

Отменить действие (Ctrl+Z) Вернуть действие (Ctrl+Y) Сбросить все зависимости Описательная статистика

Настройки анализа

Выполнить анализ

№	Идентификатор	Цель обследования	Пол	Место проживания	Возрастная группа	Возраст	Стадия ВИЧ-инфекции	Уровень С
1	1	4 (Обследованные по с...)	1 (Мужчины)	1 (Город)	1 (Лица 30-49 лет)	Количественная	ВИЧ	726
2	2	1 (Профилактическая ц...)	0 (Женщины)	1 (Город)	2 (Лица 18-29 лет )	Не определено	3	255
3	3	1 (Профилактическая ц...)	1 (Мужчины)	1 (Город)	1 (Лица 30-49 лет)	Идентификатор	3	
4	4	2 (Обследованные по э...)	1 (Мужчины)	1 (Город)	2 (Лица 18-29 лет )	Количественная	ВИЧ	486
5	5	2 (Обследованные по э...)	0 (Женщины)	1 (Город)	2 (Лица 18-29 лет )	Категориальная	3	632
6	6	3 (Обследованные по к...)	0 (Женщины)	1 (Город)	3 (Лица 50 лет и старше)	Дата	4	



## Список литературы

1. Эпидемиология: учебник/ В.В. Власов. – 3-е изд., перераб. и доп. – М.: ГЭОТАР-Медиа, 2021. – 496 с.: ил.
2. Наглядная медицинская статистика: учеб. пособие/ Петри А., Сэбин К.; перевод с англ. под ред. В.П. Леонова. – 3-е изд., перераб. и доп. – М.: ГЭОТАР-Медиа, 2015. – 216 с.: ил.
3. Создание базы данных. Интернет-портал «Медицинская статистика». Доступно по ссылке: <https://www.medstatistic.ru/statistics/statistics1.html>