

Выборка в социологическом исследовании

- 1. Введение в количественные исследования**
 - 1.1. Обзор основных понятий статистики
 - 1.2. Типы шкал

- 2. Программный пакет SPSS и его возможности**
 - 2.1. Запуск программы, интерфейс, принципы работы
 - 2.2. Создание файлов данных. Настройка переменных

1. Введение в количественные исследования

ЧТО ИЗУЧАЮТ СОЦИАЛЬНЫЕ НАУКИ?

Общество
(Population)



Выборка
(Sample)



Человек
(Case)

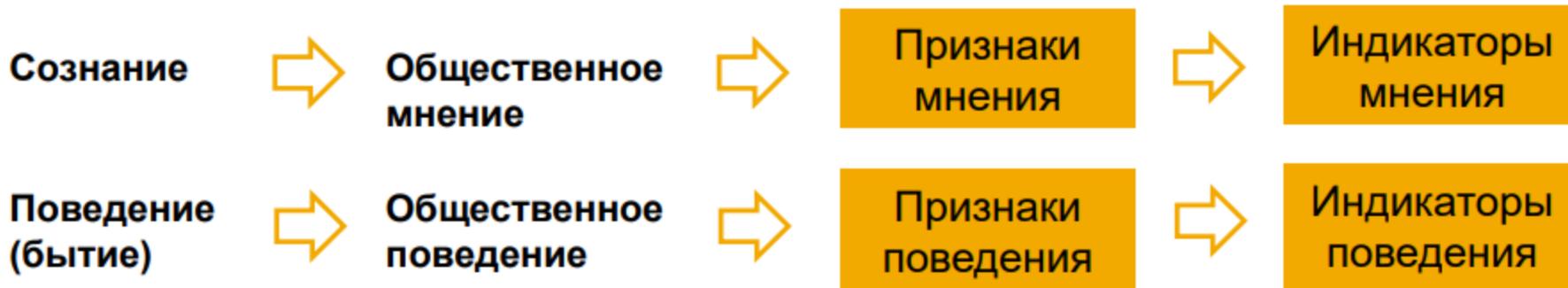


СОЗНАНИЕ

ПОВЕДЕНИЕ

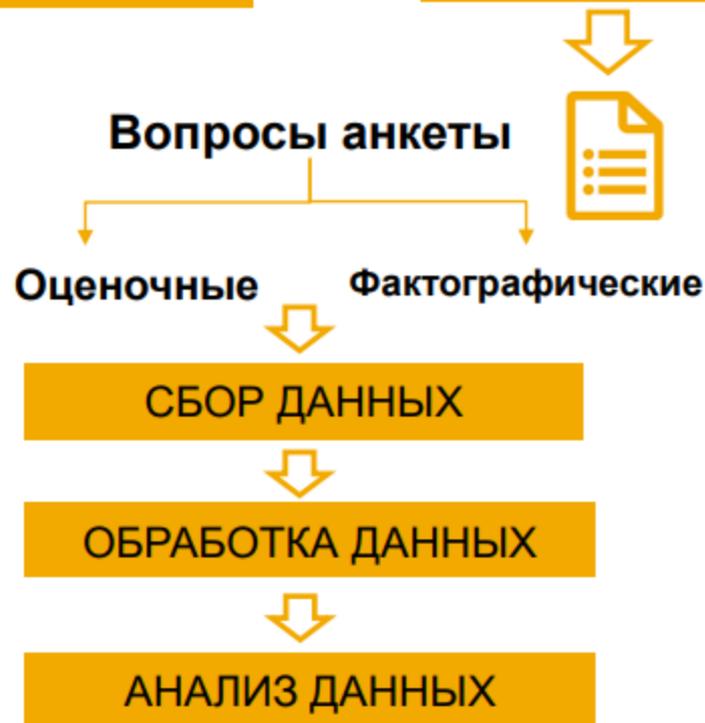
1. Введение в количественные исследования

ОТ ТЕОРЕТИЧЕСКОЙ СОЦИОЛОГИИ К ЭМПИРИЧЕСКОЙ



Свойства людей – это их физические, психические и социальные характеристики (**признаки**).

Социологов и маркетологов преимущественно интересуют **социальные характеристики: сознание (мнение) и бытие (поведение) людей, в случае маркетинга – потребителей.**



1. Введение в количественные исследования

ЧТО ТАКОЕ АНАЛИЗ ДАННЫХ?



Анализ данных является одним из этапов исследования и включает проверку соответствия между эмпирическими данными и теоретической моделью изучаемого явления.



Переменная (признак) – некоторое общее для всех изучаемых объектов, например людей, свойство, конкретные проявления которого могут меняться от объекта к объекту.



Различные проявления признака для разных объектов называют **значениями**. Значения переменной, которые она принимает для различных изучаемых объектов, приводят нас к необходимости рассматривать **распределение переменной**.

Наблюдение
(case)

Респондент (num_ank)	Возраст (AGE)	Пол (GENDER)	Образование (EDU)	Семейное положение (FAMILY)
1	21	1	3	1
2	34	2	2	2
3	19	1	3	3
4	52	1	4	2
5	46	2	5	3

Переменная
(variable)

Значение

1 = мужской
2 = женский

1. Введение в количественные исследования

ПРИМЕР АНАЛИЗА ДАННЫХ О РАСПРЕДЕЛЕНИИ ПЕРЕМЕННОЙ

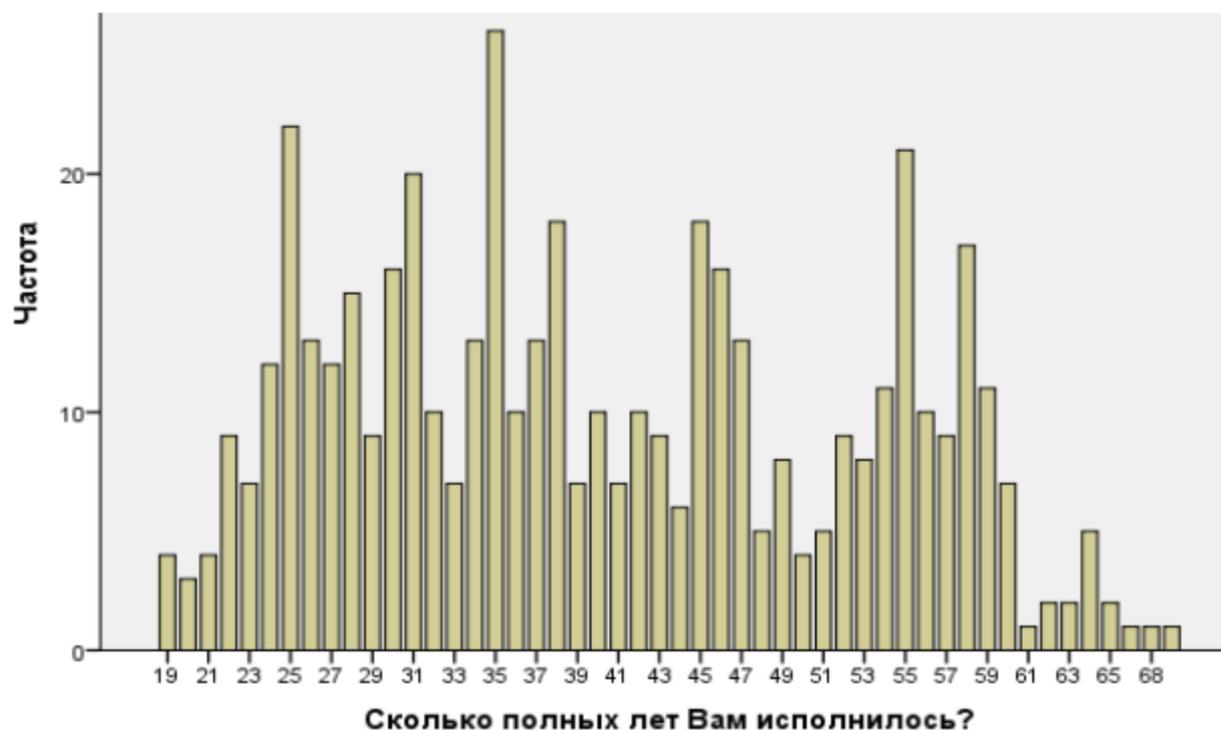
Задача: проанализировать возраст работающего населения

Объект исследования: работающее (полная занятость) население страны

Признак (переменная): возраст (age)

Значения переменной: 18...70 лет

Распределение значений переменной (distribution)



1. Введение в количественные исследования

КАКИЕ ДАННЫЕ АНАЛИЗИРУЕТ СОЦИОЛОГ ИЛИ МАРКЕТОЛОГ?

Данные – это результаты наблюдений, испытаний, накапливаемые с целью последующего изучения и анализа.

ДИСКРЕТНЫЕ ДАННЫЕ

представляют собой отдельные значения признака, общее число которых конечно или счетно (может быть подсчитано)

Пример:

Пол респондента (GENDER):
1 = Мужской 2 = Женский

НЕПРЕРЫВНЫЕ ДАННЫЕ

в отличие от дискретных данных, могут принимать любое значение в некотором интервале

Пример:

Доход работника (INCOME):
100\$.....100 000\$+

1. Введение в количественные исследования

ГЕНЕРАЛЬНАЯ СОВОКУПНОСТЬ И ВЫБОРКА

«Чтобы понять вкус супа, не обязательно съесть всю кастрюлю – достаточно одной ложки»

Генеральная совокупность (population) – полная совокупность изучаемых объектов.

Выборка (sample) – часть генеральной совокупности, отбираемая специальным образом и исследуемая с целью получения репрезентативных выводов о свойствах генеральной совокупности.

Репрезентативность выборки – это свойство выборки отражать генеральную совокупность с определенной погрешностью (ошибкой выборки).

Ошибка выборки — отклонение характеристик выборочной совокупности от характеристик генеральной совокупности.

Генеральная совокупность

Все занятое население России
72,3 млн человек



Выборочная совокупность (выборка)

500 чел.

1.1

ОБЗОР ОСНОВНЫХ
ПОНЯТИЙ
СТАТИСТИКИ



1.1. Обзор основных понятий статистики

В количественных исследованиях признаки изучаются на основе статистики их распределения, распространенности в обществе или среди отдельных групп.

- **Частотное распределение признака (frequency distribution)** – закономерность встречаемости разных его значений.
- **Частота (frequency)** – количество наблюдений, в которых признак принимает определенное значение или находится в определенном интервале.

Частотное распределение переменной (frequency distribution)



1. МЕРЫ СРЕДНЕГО УРОВНЯ

- Среднее
- Мода
- Медиана

2. МЕРЫ РАССЕЯНИЯ (ДИСПЕРСИИ)

- Дисперсия
- Средне-квадратическое (стандартное) отклонение
- Стандартная ошибка
- Размах

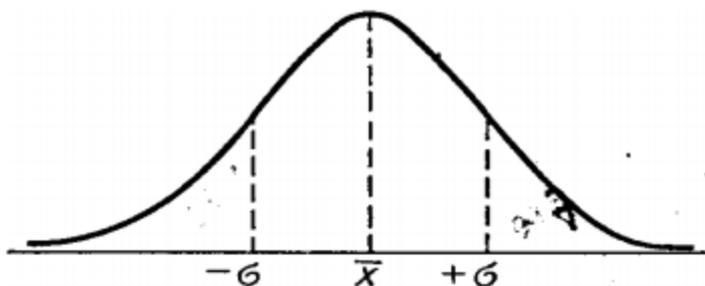
3. МЕРЫ РАСПРЕДЕЛЕНИЯ

- Асимметрия
- Эксцесс

1.1. Обзор основных понятий статистики

НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

Кривая Гаусса:



Формула:

$$f_{отн} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i - M)^2}{2\sigma^2}}$$

- Характеризуется тем, что крайние значения признака в нем встречаются редко, а значения, близкие к средней величине – достаточно часто.
- Некоторая величина отклоняется от среднего под воздействием слабых, независимых друг от друга случайных факторов.
- Имеет место, когда интересующее нас явление подвержено влиянию бесконечного числа случайных факторов, уравновешивающих друг друга.
- Если действует какой-либо однонаправленный фактор, распределение может отличаться от нормального.

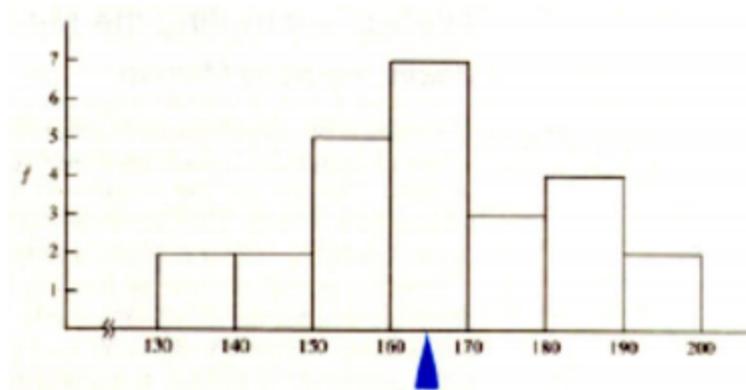
1.1. Обзор основных понятий статистики

СРЕДНЕЕ (Mean)

- Сумма всех значений, отнесенная к общему числу наблюдений (очень чувствительна к выбросам).
- Предполагает, что результат измерения задан в метрической (интервальной) шкале.
- Важнейшее свойство средней величины заключается в том, что она выражает то общее, что присуще всем единицам исследуемой совокупности.
- Типичность средней зависит от степени однородности совокупности. Сумма отклонений от среднего равна 0.

Среднее для **выборки**

$$\bar{X} = \frac{\sum x_i}{n}$$

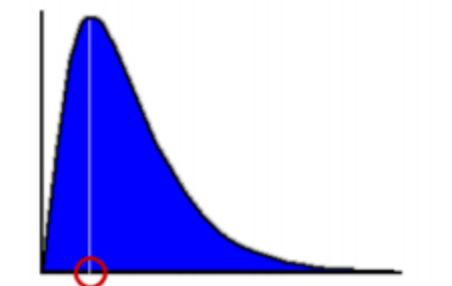


1.1. Обзор основных понятий статистики

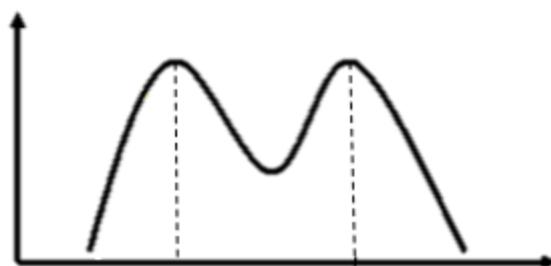
МОДА (Mode)

- Наиболее часто встречающееся значение переменной.
- Обычно используется, когда набор значений ограничен и имеется их частое повторение.
- Если в выборке встречаются одинаково часто два значения, распределение называют бимодальным, если присутствуют несколько часто встречающихся значений – мультимодальным.
- Если все значения в распределении встречаются одинаково часто, то такая выборка не имеет моды.

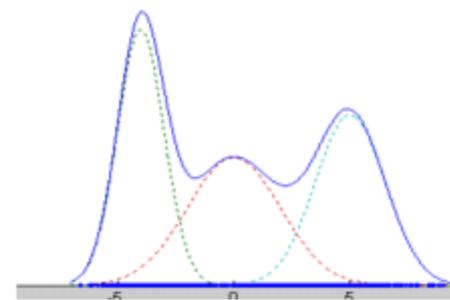
Унимодальное



Бимодальное



Мультимодальное



обычно возникают,
если выборка имеет естественные обособленные подгруппы

1.1. Обзор основных понятий статистики

МЕДИАНА (Median)

- Значение, которое делит распределение пополам: половина значений больше медианы, половина – меньше. «Середина» распределения.
- Когда есть сильные выбросы, лучше использовать медиану, а не среднее.
- Имеет смысл для ранговых и количественных переменных, но не для качественных.

Количество чисел (значений) в ряду

Нечетное

Пример:

Возраст **5** опрошенных (AGE):

18 22 **27** 31 44

Четное

Пример:

Возраст **6** опрошенных (AGE):

18 22 **27** **31** 44 55

$$(27+31):2 = \mathbf{29}$$

1.1. Обзор основных понятий статистики

ДИСПЕРСИЯ (VARIANCE)

- Дисперсия – это среднее арифметическое квадратов разностей полученных значений переменной и ее средним значением.
- Измеряется в единицах переменной, возведённых в квадрат (не всегда удобно).
- Показывает разброс значений признака относительно своего среднего арифметического значения, то есть насколько плотно значения признака группируются вокруг.
- Чем больше разброс, тем сильнее варьируются ответы респондентов в данной группе, тем больше индивидуальные различия между респондентами.
- Используется скорее в различных статистических тестах, а не в описательной статистике.

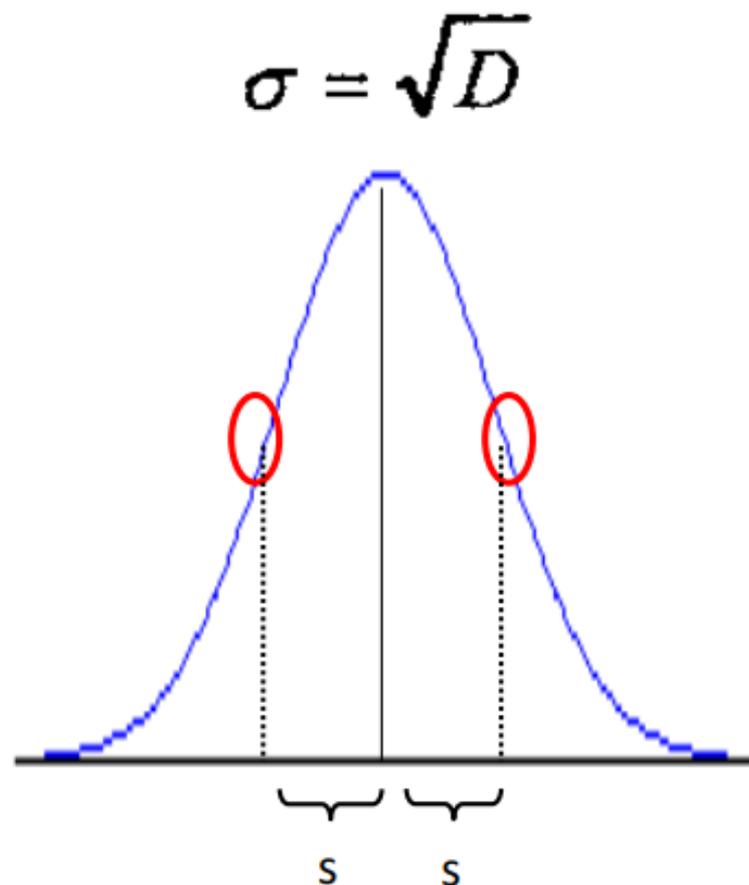
$$D = \frac{(a_1 - S)^2 + (a_2 - S)^2 + \dots + (a_n - S)^2}{n}$$

$a_1, a_2, a_3 \dots a_n$ – данные,
 S – среднее арифметическое
 n – количество чисел в ряду

1.1. Обзор основных понятий статистики

СТАНДАРТНОЕ ОТКЛОНЕНИЕ (Standard Deviation)

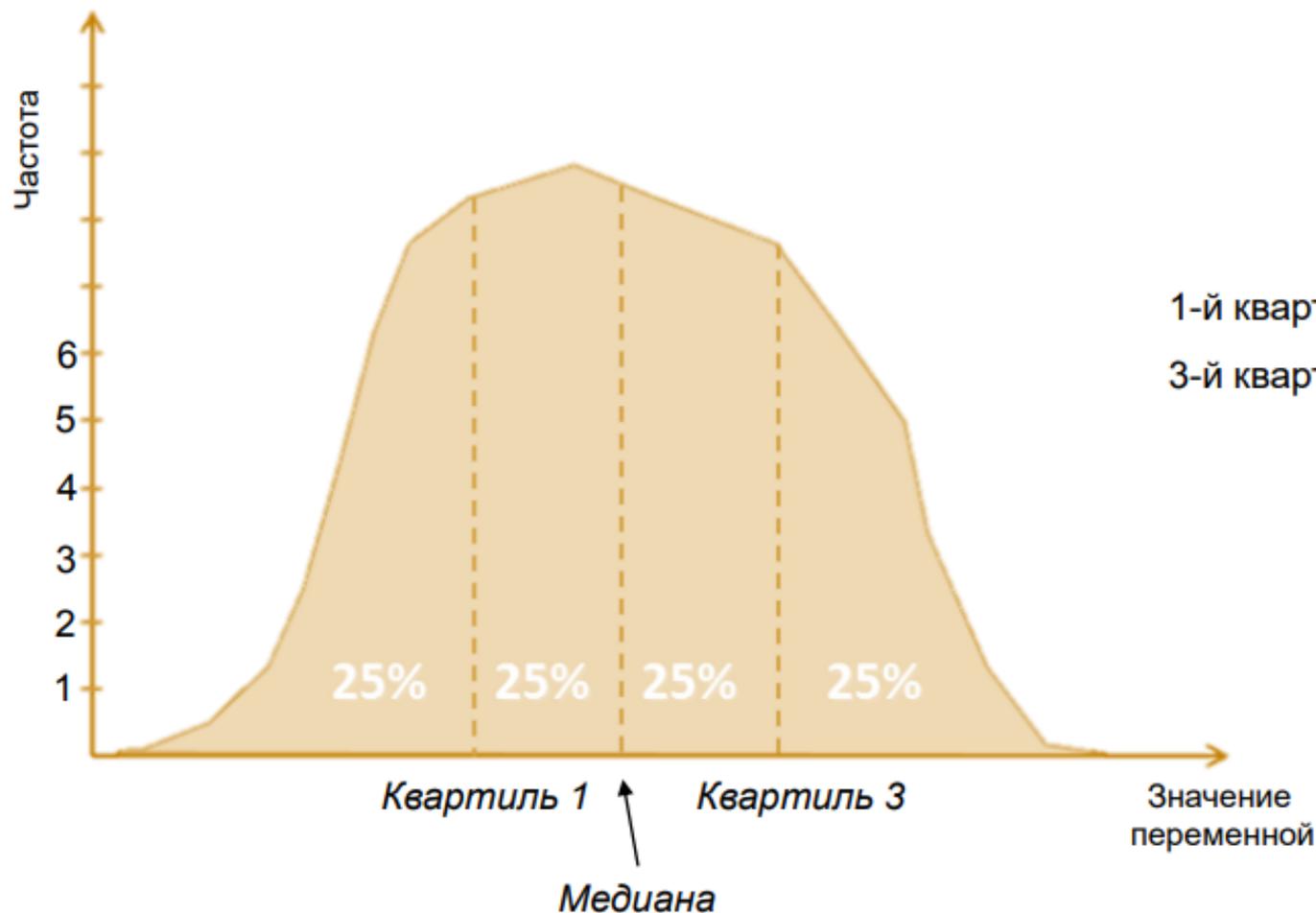
- Среднеквадратическое или стандартное отклонение – мера разброса значений признака около среднего арифметического значения.
- На практике вместо оценки дисперсии чаще используют производную от нее – стандартное отклонение (корень из дисперсии).
- Более наглядно, т.к. его размерность соответствует размерности измеряемой величины (*измеряется в тех же единицах, что и переменная!*)



1.1. Обзор основных понятий статистики

ПРОЦЕНТИЛИ И КВАРТИЛИ

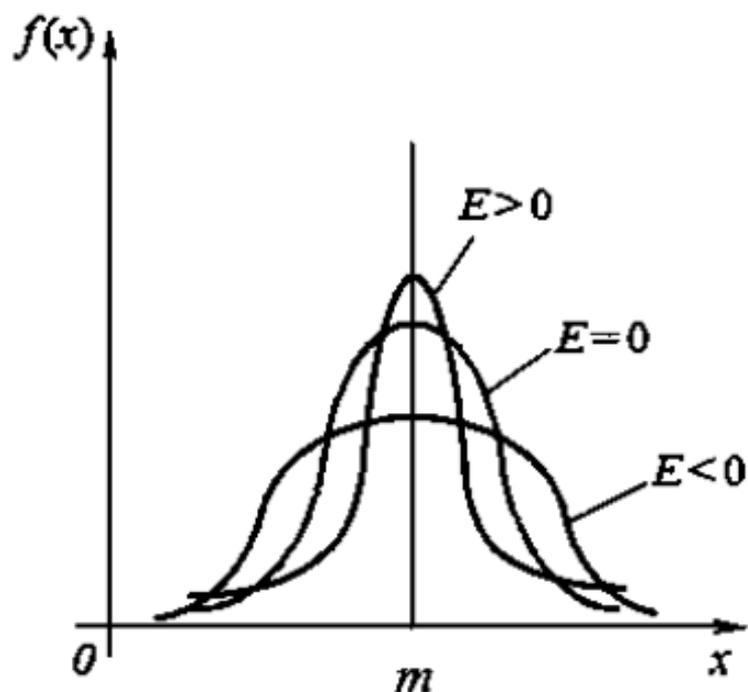
Квартили (quartiles) делят распределение на четыре части так, что в каждой из них оказывается поровну значений (2-й квартиль = медиана)



1.1. Обзор основных понятий статистики

ЭКЦЕСС

Коэффициент эксцесса E (kurtosis) - характеризует степень островершинности распределения. Коэффициент указывает, является ли распределение пологим (при большом значении коэффициента) или крутым.



Для нормального распределения $E=0$
Для островершинного $E>0$
Для плосковершинного $E<0$

$$E_x = \frac{\sum (x_i - M)^4}{n\sigma^4} - 3$$

1.2 ТИПЫ ШКАЛ



1.2. Типы шкал

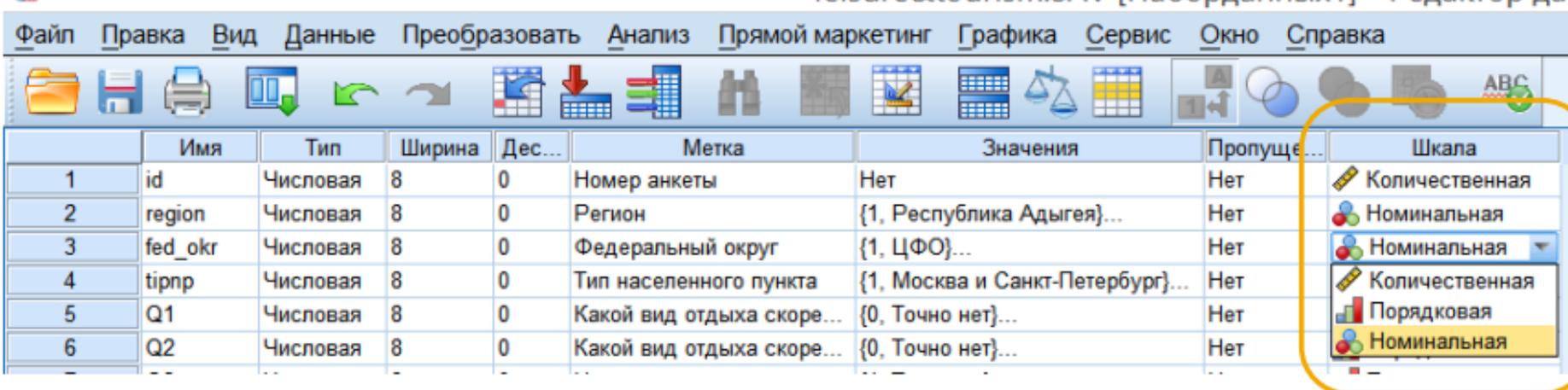
ОСНОВНЫЕ ПОНЯТИЯ

Шкала (Scale) – правило, определяющее, каким образом в процессе измерения каждому изучаемому объекту ставится в соответствие некоторое число или символы.

Шкалирование (Scaling) – процесс создания континуума (последовательного ряда), на котором размещаются измеряемые объекты.

*leisure&tourism.SAV [Наборданных1] - Редактор да

Файл П_равка В_ид Д_анные П_реобразовать А_нализ П_рямой маркетинг Г_рафика С_ервис О_кно С_правка



	Имя	Тип	Ширина	Дес...	Метка	Значения	Пропуще...	Шкала
1	id	Числовая	8	0	Номер анкеты	Нет	Нет	Количественная
2	region	Числовая	8	0	Регион	{1, Республика Адыгея}...	Нет	Номинальная
3	fed_okr	Числовая	8	0	Федеральный округ	{1, ЦФО}...	Нет	Номинальная
4	tipnr	Числовая	8	0	Тип населенного пункта	{1, Москва и Санкт-Петербург}...	Нет	Количественная
5	Q1	Числовая	8	0	Какой вид отдыха скорее...	{0, Точно нет}...	Нет	Порядковая
6	Q2	Числовая	8	0	Какой вид отдыха скорее...	{0, Точно нет}...	Нет	Номинальная

1.2. Типы шкал



Назовите, пожалуйста, форму вашей работы?

1. Работаю по найму
2. Являюсь владельцем предприятия
3. Самостоятельная занятость
4. Не работаю

Как бы вы оценили материальное положение своей семьи?

1. Высокое
2. Выше среднего
3. Среднее
4. Ниже среднего
5. Низкое
6. Затрудняюсь ответить

Оцените, насколько вы согласны с каждым из следующих суждений относительно работы органов власти страны?

(от 1 до 5, где 5 – максимально хорошо)

- Суждение 1
Суждение 2
Суждение 3
И т.п.

Итого: сумма ответов

Сколько вам лет?

_____ (запишите число)

Сколько человек в вашей семье, включая вас?

_____ (запишите число)

1.2. Типы шкал

Номинальная шкала (Nominal) — шкала наименований, которая состоит из значений признаков, не упорядоченных по степени возрастания или убывания.

Пример: национальность, профессия, семейное положение, пол и т.д.

Порядковая шкала (Ordinal) — градации располагаются в определенном порядке относительно возрастания либо убывания интенсивности свойства.

Пример: переменная «Курение» со значениями (1 = некурящий; 2 = изредка курящий; 3 = интенсивно курящий; 4 = очень интенсивно курящий). Переменная сортирована в порядке значимости снизу вверх: умеренный курильщик курит больше, нежели некурящий, а сильно курящий — больше, чем умеренный курильщик и т.д., поэтому порядковая шкала.

Интервальные шкалы (Interval) — основаны на процедурах, обеспечивающих равные или примерно равные расстояния между градациями переменной. В данном случае сравниваются не значения переменных, а расстояния между значениями.

Пример: температура, измеренная в градусах Цельсия. Можно не только сказать, что температура 30 градусов выше, чем 20 градусов, но и то, что увеличение температуры с 10 до 30 градусов вдвое больше увеличения температуры от 20 до 30 градусов.

Шкалы отношений (Метрические) — соответствуют всем требованиям, предъявляемым к шкалам более низких классов.

Пример: возраст. Если Макс 30 лет, а Сергею 60, можно сказать, что Сергей вдвое старше Макса.

1.2. Типы шкал

ЗНАНИЕ ТИПОВ ШКАЛ ПОЗВОЛИТ ВЫБРАТЬ ОПТИМАЛЬНЫЕ МЕТОДЫ АНАЛИЗА ДЛЯ РАЗНЫХ ТИПОВ ДАННЫХ

Статистическая шкала	Математическая значимость
Номинальная	Нет
Порядковая	Порядок чисел
Интервальная	Разность между числами
Шкала отношений	Отношение чисел

В SPSS объединены в одну метрическую шкалу

Шкала, по которой измеряется переменная, накладывает ограничения на выбор меры центральной тенденции



Типическое значение	Номинальные данные	Порядковые данные	Интервальные данные
Мода	●	●	●
Медиана		●	●
Среднее			●

2. ПРОГРАММНЫЙ ПАКЕТ SPSS И ЕГО ВОЗМОЖНОСТИ



Программный пакет **SPSS** и его возможности

- **SPSS** — Statistical Package for the Social Science (статистический пакет для социальных наук)
- Наряду с другими статистическими пакетами (Statistica, STATA, SAS) широко используется специалистами в сфере исследований (социология, психология, маркетинг, медицина и пр.) для обработки и анализа количественных данных.

2.1. Запуск программы, интерфейс, принципы работы

ПО УМОЛЧАНИЮ ПРИ ЗАПУСКЕ SPSS ЗАПУСКАЕТСЯ ОКНО «РЕДАКТОР ДАННЫХ»

Запуск SPSS



Start → All Programs → SPSS Inc → IBM SPSS Statistics

Без имени1 [Наборданных0] - Редактор данных IBM SPSS Statistics

Файл Правка Вид Данные Преобразовать Анализ Прямой маркетинг Графика Сервис Окно Справка

	Имя	Тип	Ширина	Десятич...	Метка	Значения	Пропущен...	Ширина ...	Выравнивание	Шкала	Роль
1											
2											
3											
4											
5											
6											
7											
8											
9											
10											
11											
12											
13											
14											
15											
16											
17											
18											
19											
20											
21											
22											
23											

Данные Переменные

Процессор IBM SPSS Statistics roroa

2.1. Запуск программы, интерфейс, принципы работы

ДВА ИНТЕРФЕЙСА SPSS

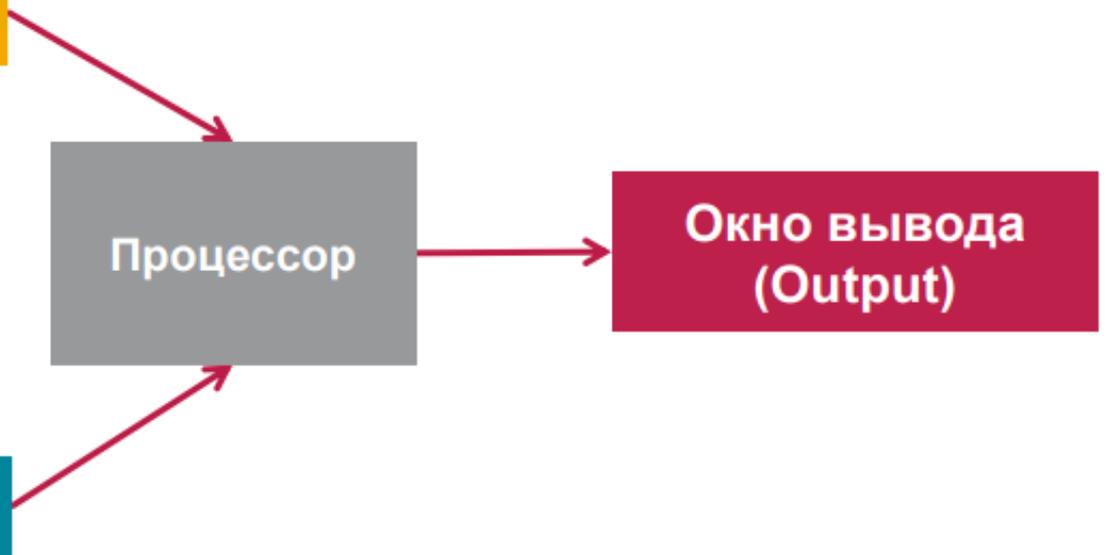
В SPSS реализовано два основных интерфейса работы с данными. Кнопочный – интуитивно более понятный. Синтаксис – язык команд, больше подходит для выполнения более сложных или повторяющихся вычислений.

1. Кнопочный интерфейс

Процессор

Окно вывода
(Output)

2. Синтаксис (Syntax)



2.1. Запуск программы, интерфейс, принципы работы

1. КНОПОЧНЫЙ ИНТЕРФЕЙС

При запуске SPSS пользователю открывается окно для ввода, редактирования и просмотра данных исследования. Данные сохраняются в файле с расширением *.sav

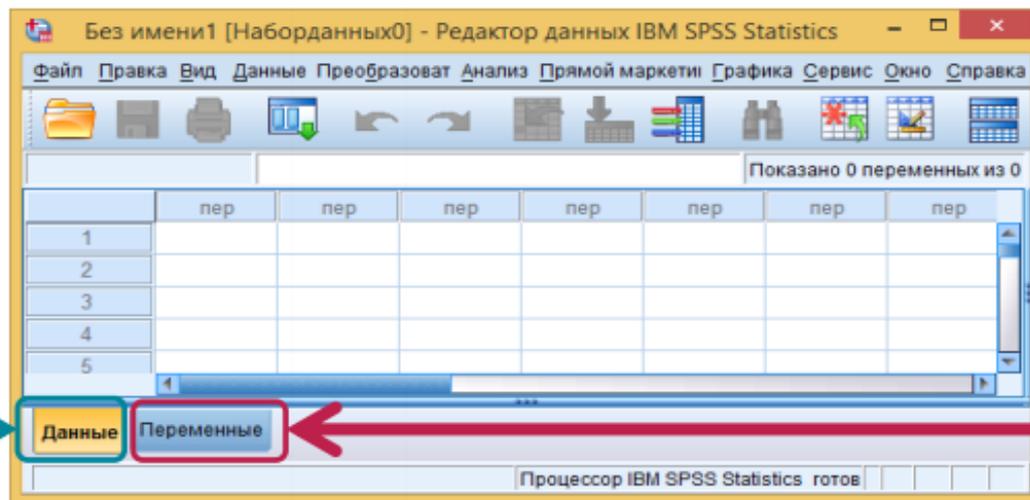
Кнопочный интерфейс состоит из двух вкладок:

- **Вкладка «Data view»**
(Окно данных)

- Окно ввода данных
- Columns: variables (переменные)
- Rows: cases (наблюдения)

- **Вкладка «Variable view»**
(Окно переменных)

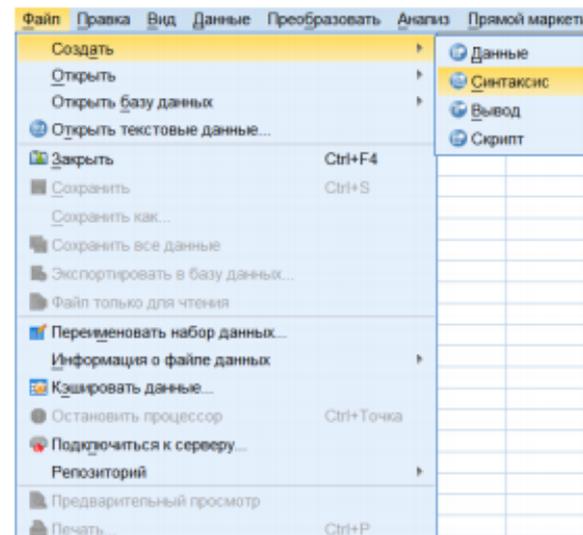
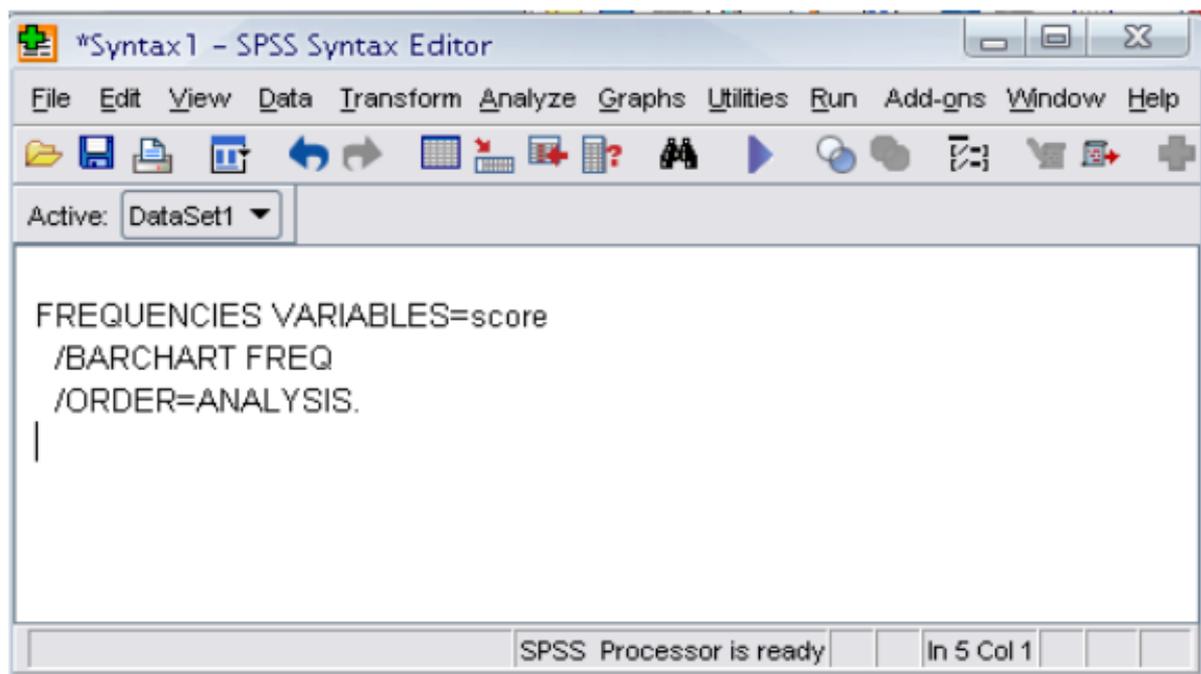
- Окно создания и настройки переменных
- Variables (список переменных)
- Параметры каждой из переменных



2.1. Запуск программы, интерфейс, принципы работы

2. ОКНО РЕДАКТОРА КОМАНДНОГО ЯЗЫКА SYNTAX

Текстовый редактор для ввода синтаксиса – команд обработки данных. **Используется для оптимизации и детальной настройки вычислений, недоступной в кнопочном интерфейсе.**



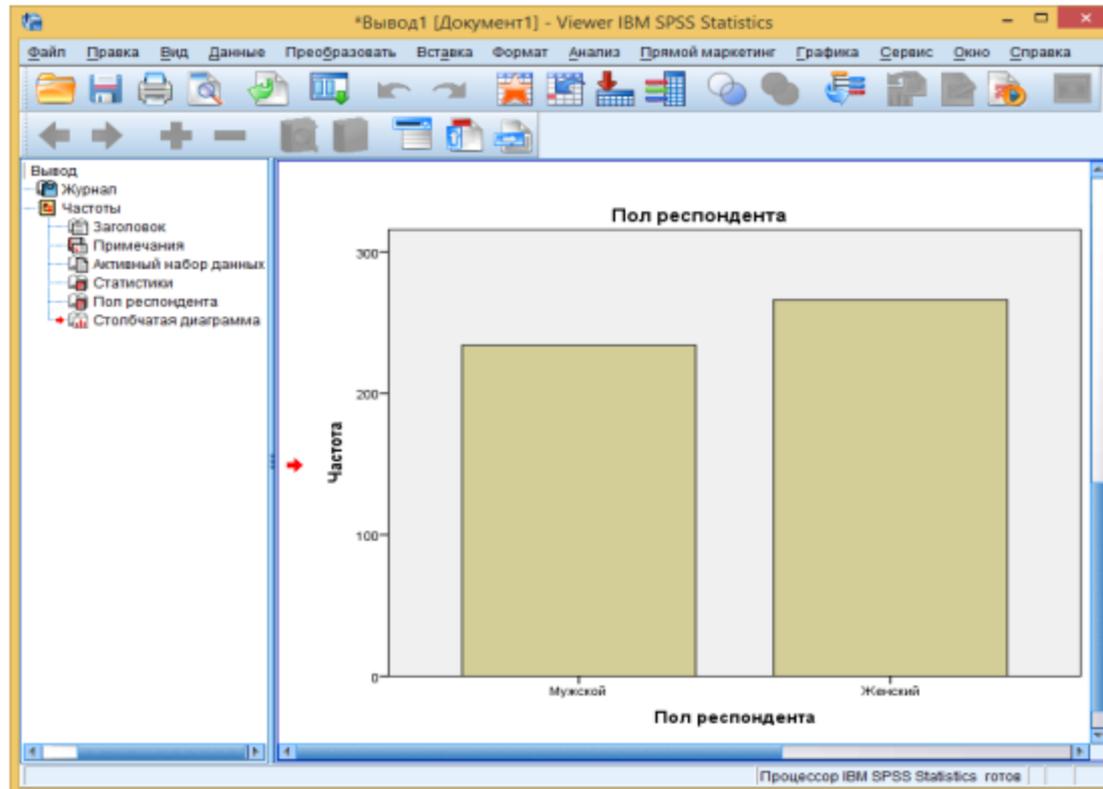
Расширение файла: *.sps

2.1. Запуск программы, интерфейс, принципы работы

3. ОКНО ВЫВОДА

В отличие, например, от Excel, результаты вычислений, анализа данных и их представление в визуальном виде производится **в отдельном окне**.

Окно отображает историю команд (слева), вывод результатов расчетов и возникающие ошибки расчетов (справа). Сохранение осуществляется в файл *.spv или *.spo





2.2 СОЗДАНИЕ ФАЙЛОВ ДАННЫХ. НАСТРОЙКА ПЕРЕМЕННЫХ

2.2. Создание файлов данных. Настройка переменных

3 ЭТАПА СОЗДАНИЯ БАЗЫ ДАННЫХ ДЛЯ АНАЛИЗА

До начала ввода данных в SPSS необходимо создать макет (структуру) переменных (на основе, например, анкеты).

В этом случае вопросы ложатся в основу переменных. У каждой переменной – свои настройки (имя, допустимые значения, тип шкалы и др.)

Структуру базы данных правильнее определить на этапе планирования исследования и разработки инструментария (например, анкеты) в соответствии с гипотезой и задачами исследования.

Шаг 1. Задание имён переменных

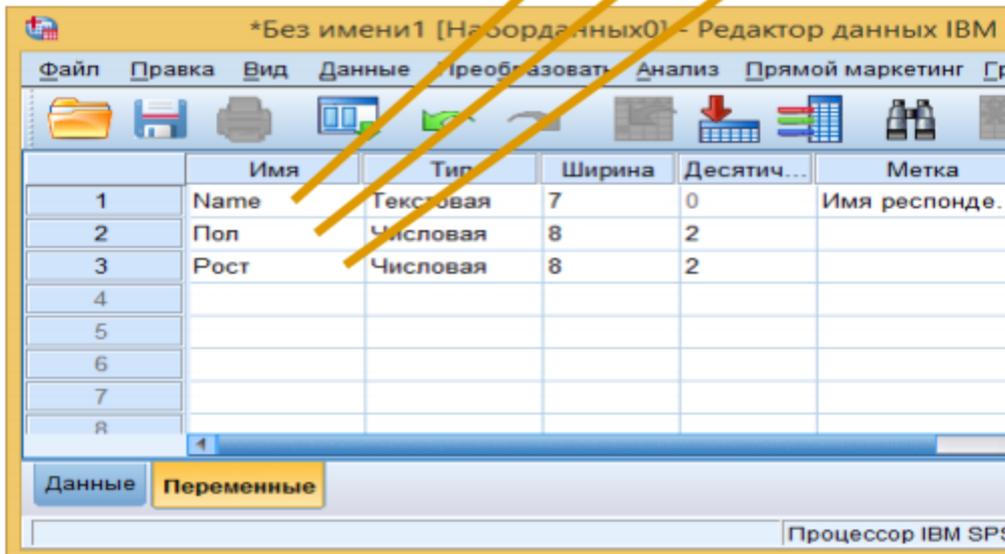
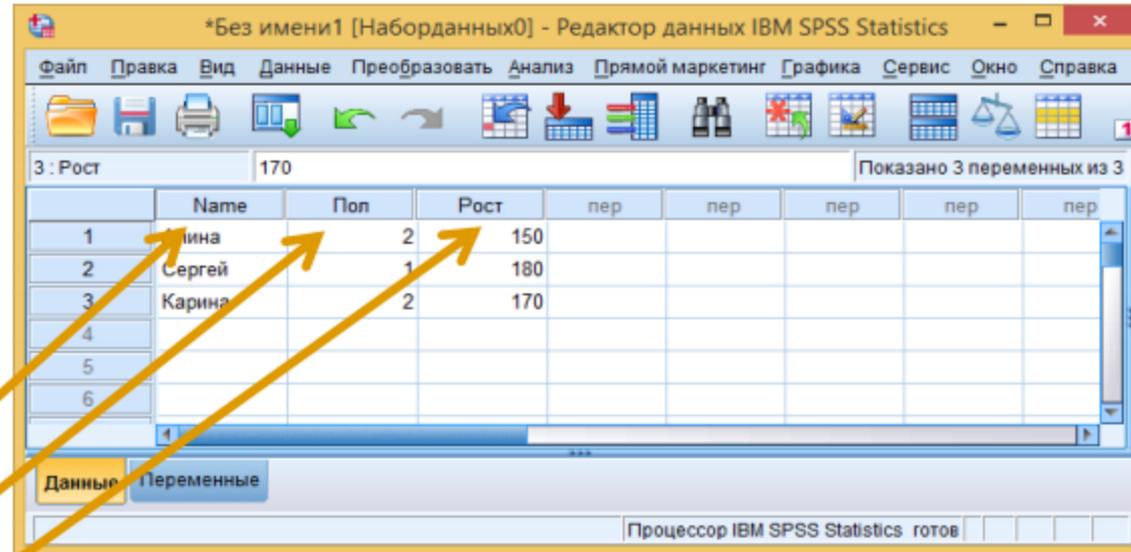
Шаг 2. Определение их параметров

Шаг 3. Ввод данных

2.2. Создание файлов данных. Настройка переменных

ПАРАМЕТРЫ ПЕРЕМЕННЫХ

Перейдя на вкладку Данные после установки параметров переменных, их имена отобразятся в столбцах.



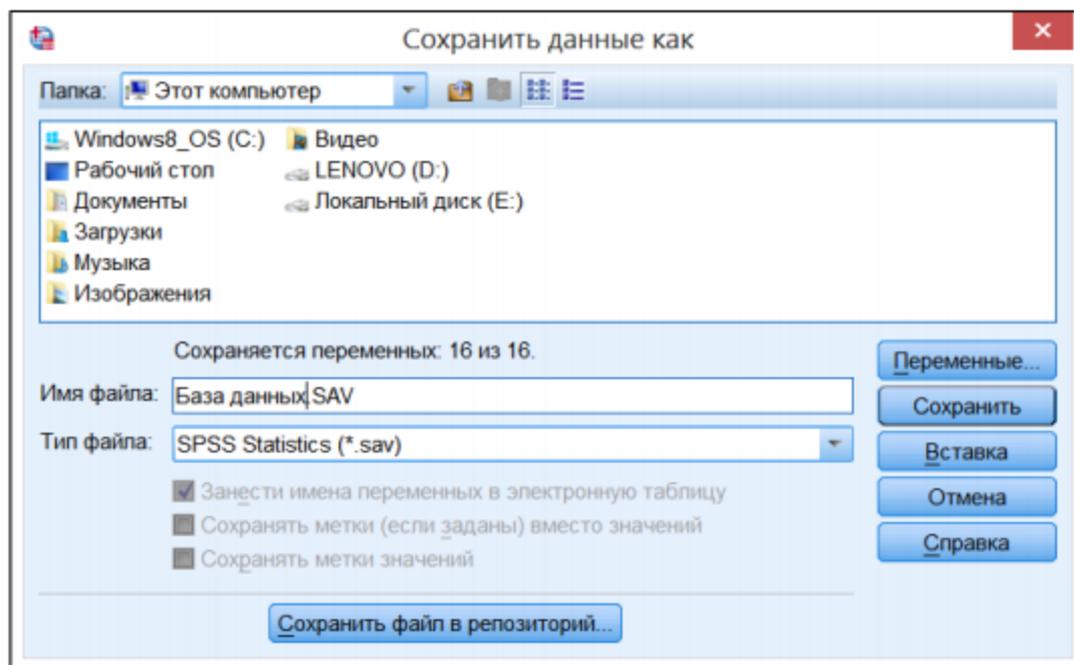
Ввод данных осуществляется **вручную** или **автоматически** (например, через сканер с распознаванием, через онлайн ввод).

2.2. Создание файлов данных. Настройка переменных

СОХРАНЕНИЕ ФАЙЛА ДАННЫХ

После завершения ввода данных и работы с массивом, рабочий файл с данными нужно сохранить.

Массив (базу данных) можно сохранить как в стандартном формате **SPSS (*.sav)**, так и в других форматах, в том числе Excel.





Спасибо за внимание!