



Создание базы данных в SPSS

Артем Гиль, к.м.н.

ВШУЗ

Первый МГМУ им.И.М.Сеченова

Доступные статистические пакеты

**Необходимо
приобретать
лицензию**

- Excel
- SPSS
- SAS
- STATA

**В бесплатном
доступе**

- PSPP
- EpiInfo
- R

Наиболее часто используются в университетах

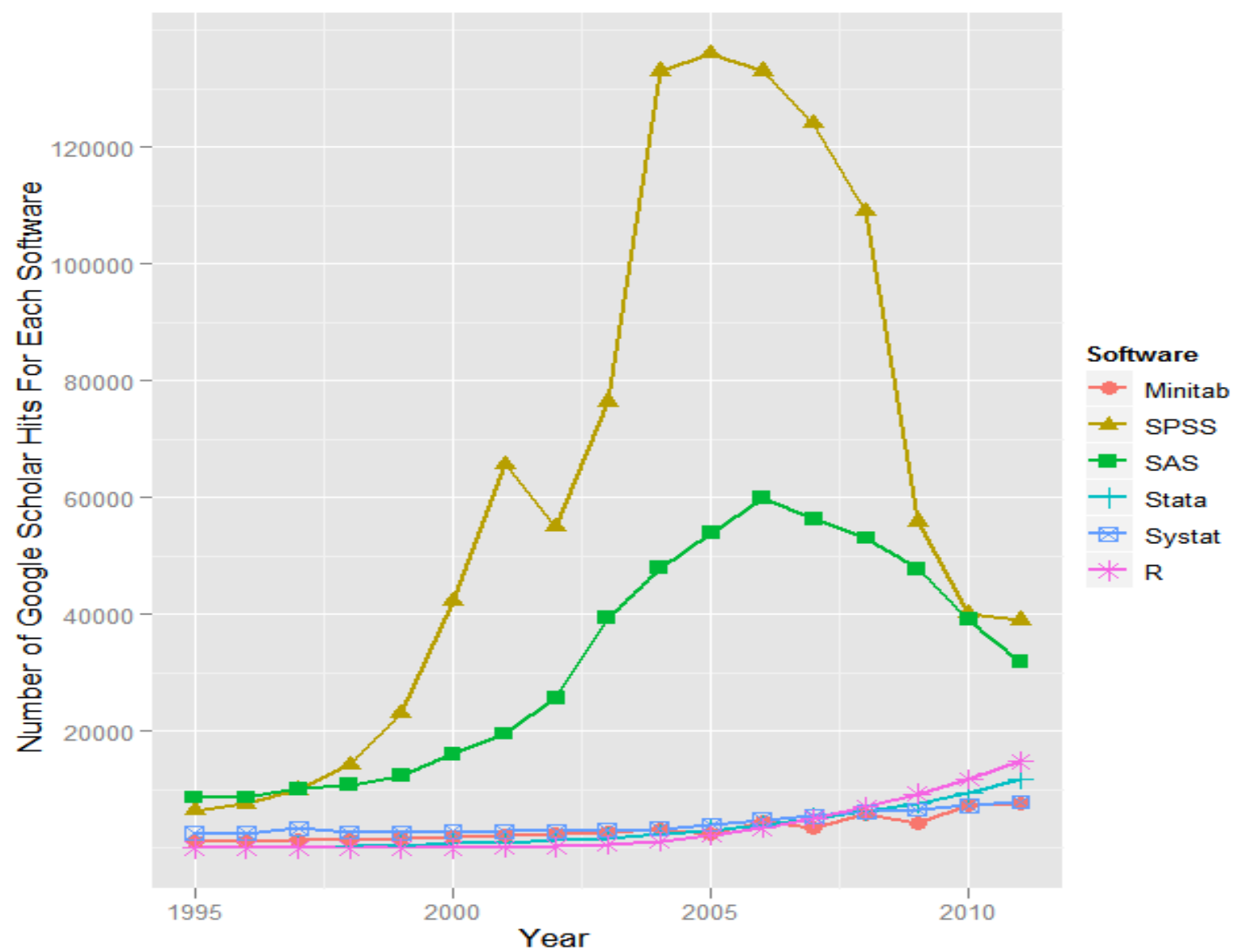


Figure 7a. Use of data analysis software in academic publications as measured by hits on Google Scholar.

SPSS for Windows

SPSS Statistics (аббревиатура англ. «*Statistical Package for the Social Sciences*» — «статистический пакет для социальных наук») — компьютерная программа для статистической обработки данных, один из лидеров рынка в области коммерческих статистических продуктов, предназначенных для проведения прикладных исследований в социальных науках.

IBM SPSS Statistics 24

Что нового:

1. Улучшенное чтение и запись Excel файлов (.xls .xlsx)
 - a. Можно задать % значений в импортируемой переменной, в формате которых будут импортироваться данные из файла Excel
 - b. Игнорирование чтения скрытых в файле Excel строк и столбцов
 - c. Удаление начальных и хвостовых пробелов в строковых значениях
 - d. Появилась возможность задавать имя листа при экспорте данных из SPSS в Excel
 - e. Появилась возможность экспортировать данные на именованный лист в уже существующий файл Excel
 - f. Запись меток переменных вместо имён
2. Чтение данных из текстовых файлов
3. Простой мастер чтения файлов CSV (файл с запятыми, точками с запятой или табуляцией в качестве разделителей полей)

IBM SPSS Statistics 24

Что нового:

4. Чтение и распознавание различных форматов даты и времени
 - a. при чтении текстовых данных программа распознает, что переменная содержит значения даты или времени
 - b. при отображении значений даты и времени в редакторе данных они позволяют представить значения в нужном пользователю виде, например, как «20.01.2016» или как «2016/01/20»
 - c. при записи значений даты и времени в текстовый файл можно предписать, в каком именно виде дата и время будут записаны в текстовый файл

IBM SPSS Statistics 24

Что нового:

5. Меньшая строгость записи некоторых форматов дат и времени
6. Отображение значимости различий долей и средних
 - а. Ранее данные о значимости запрошенных критериев выводились только во вспомогательной таблице (не в той, что содержит оценки долей и средних). Теперь этим можно управлять.
7. Пополнился перечень статистик, для которых возможен расчет стандартных ошибок и доверительных интервалов (напр. можно рассчитать ДИ для суммы значений или для медианы)

SPSS

ЦЕНА

- От \$1000 до \$12000 в зависимости от типа лицензии

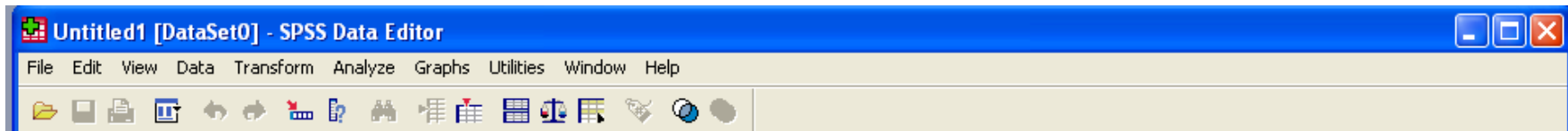
НЕДОСТАТКИ

- Дорогой пакет
- Не подходит для сложных статистических методов (математическое моделирование и т.д.)

ДОСТОИНСТВА

- Простой в использовании статистический пакет
- Очень распространён в университетах и на предприятиях
- Интерфейс основан на окнах меню, но есть и командный язык
- Один из наиболее мощных статистических пакетов, которым легко пользоваться

SPSS Меню & Инструменты



Опции в меню: File, Edit, View, Window, Help - точно также как и в других приложениях Windows.

File - стандартные команды по чтению (Open), сохранению (Save), печати (Print), выходу из программы (Exit)

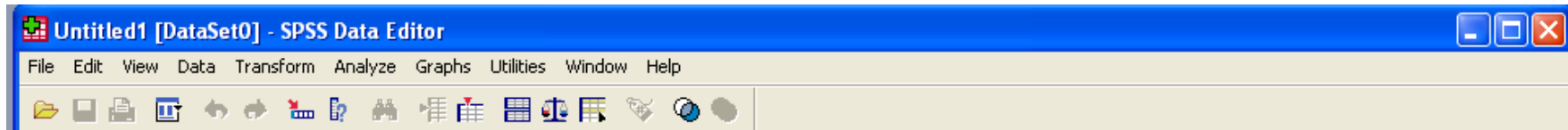
Edit – стандартные коанды: шаг назад (Undo), шаг вперед (Redo), удалить (Cut), скопировать (Copy), вставить данные (Paste)

View – опции по отображению/скрытию инструментов (Toolbars), отображению значений данных (Values) или их меток (Labels)

Window – опции по переключению между разными окнами SPSS

Help – справочник по SPSS

Инструменты (Toolbars), продолжение



Data – опции для манипуляций с данными: сортировка, слияние, разделение ... и др.

Transform – создание новых переменных, перекодировка переменных

Analyze – опции по статистическому анализу данных. Все опции имеют несколько меню для выбора статистик, разновидностей анализа и т.д.

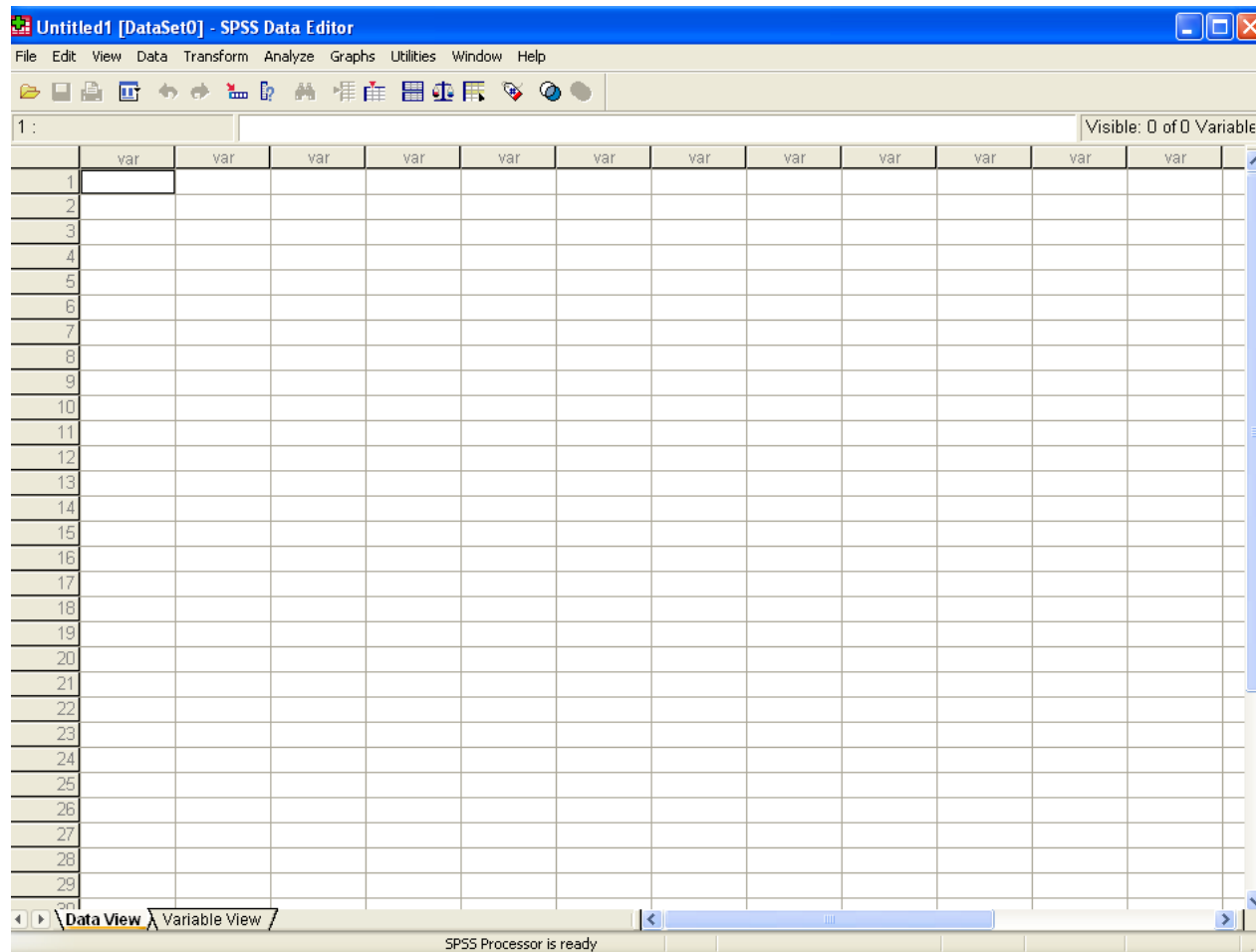
Graphs – опции для построения высококачественных графиков и диаграмм

Utilities – опции по отображению информации по отдельным переменным

Вкладки в редакторе данных (Data Editor)

Две вкладки в редакторе данных:

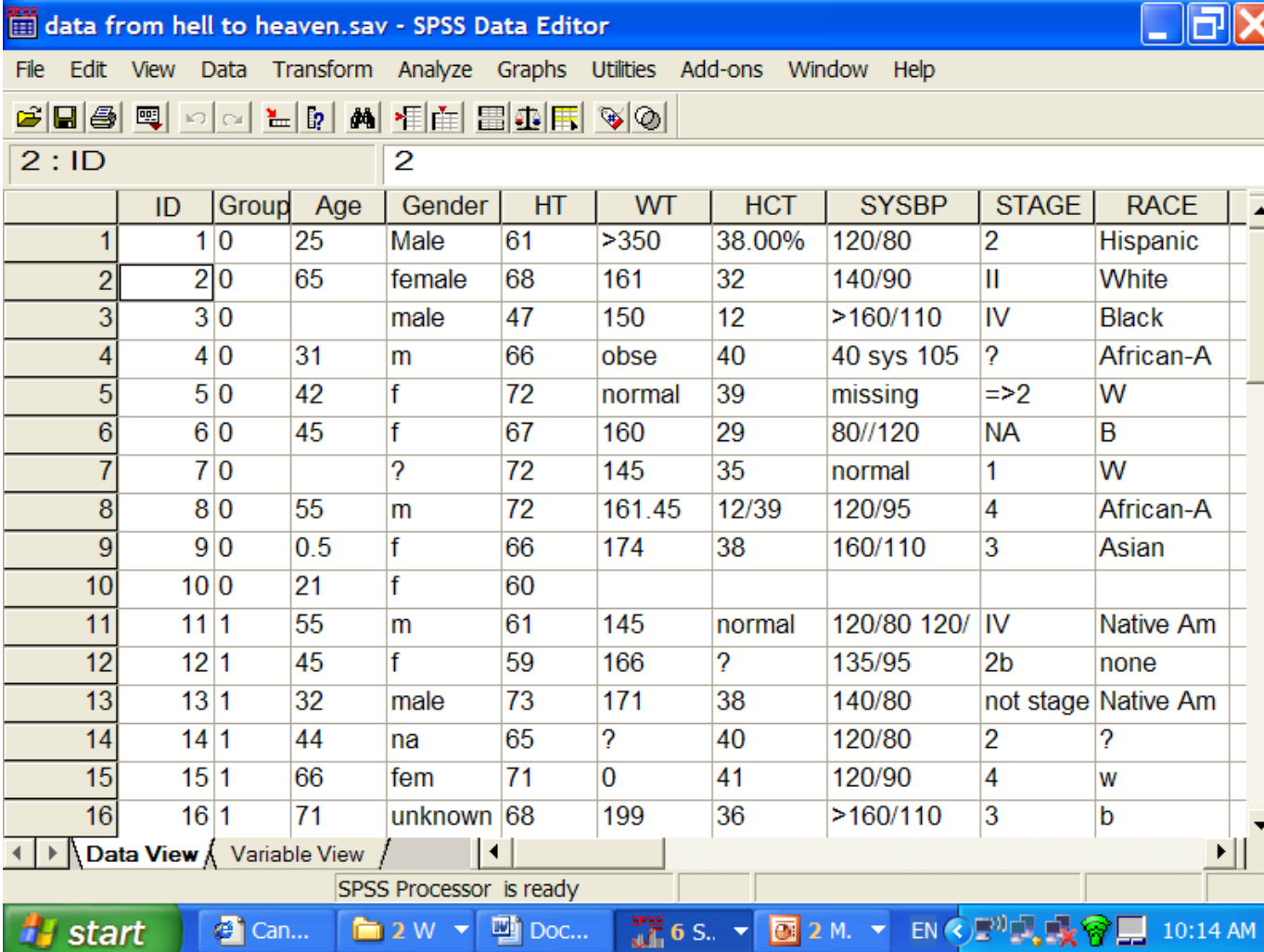
- **Data View** – вкладка для ввода и редактирования данных
- Столбцы – переменные
- Строки – случаи наблюдения (пациенты, респонденты)



Вкладки в редакторе данных (Data Editor)

Две вкладки в редакторе данных:

- **Data View** – вкладка для ввода и редактирования данных с данными



The screenshot shows the SPSS Data Editor window for the file 'data from hell to heaven.sav'. The 'Data View' tab is active, displaying a table with 16 rows of patient data. The columns are labeled: ID, Group, Age, Gender, HT, WT, HCT, SYSBP, STAGE, and RACE. The data is as follows:

	ID	Group	Age	Gender	HT	WT	HCT	SYSBP	STAGE	RACE
1	1	0	25	Male	61	>350	38.00%	120/80	2	Hispanic
2	2	0	65	female	68	161	32	140/90	II	White
3	3	0		male	47	150	12	>160/110	IV	Black
4	4	0	31	m	66	obse	40	40 sys 105	?	African-A
5	5	0	42	f	72	normal	39	missing	=>2	W
6	6	0	45	f	67	160	29	80//120	NA	B
7	7	0		?	72	145	35	normal	1	W
8	8	0	55	m	72	161.45	12/39	120/95	4	African-A
9	9	0	0.5	f	66	174	38	160/110	3	Asian
10	10	0	21	f	60					
11	11	1	55	m	61	145	normal	120/80 120/	IV	Native Am
12	12	1	45	f	59	166	?	135/95	2b	none
13	13	1	32	male	73	171	38	140/80	not stage	Native Am
14	14	1	44	na	65	?	40	120/80	2	?
15	15	1	66	fem	71	0	41	120/90	4	w
16	16	1	71	unknown	68	199	36	>160/110	3	b

Ввод данных (Entering Data)

Перед вводом данных необходимо разработать документ со списком кодов переменных – **Codebook**

Например:

Num = Порядковый номер анкеты (xxx)

ID = Уникальный идентификационный номер пациента/респондента (xxx)

Q1. “Age” = Возраст, полных лет (Сколько Вам полных лет? (xx))

Q2. “Weight” = Вес, кг (согласно измерению на весах (xxx))

Q10. “Educ” = Уровень образования (Каков уровень Вашего образования?
Неполное среднее или ниже = 1, Полное среднее = 2, Начальное профессиональное (техническое училище) = 3, Среднее специальное = 4, Незаконченное высшее = 5, Высшее = 6, Затрудняюсь ответить = 97, Отказ от ответа = 98)

Q12. “Smok” = Курение сигарет (Курите ли Вы в настоящее время? Да = 1, Нет = 2)

Q15. “Rstpulse” = Частота сердечных сокращений в покое (ударов в мин.)

Ввод данных (Entering Data)

File | Open – открыть существующую базу данных SPSS (.sav)

Или

Вручную ввести данные (Enter Data):

1. Составление списка переменных (Define Variables) во вкладке Variable View
2. Ввести данные во вкладке Data view

Или

Прочитать данные из файла Excel (Read Data from Excel file (.xls, .xlsx))

Составление списка переменных при ручном вводе данных (Defining Variables)

- Во вкладке Variable View присвойте имя для переменной в столбце «Name»
- В столбце «Type» задайте подходящий тип переменной

The screenshot shows the SPSS Data Editor window titled '*Employee data.sav [DataSet1] - SPSS Data Editor'. The Variable View tab is active, displaying a table with columns: Name, Type, Width, Decimals, Label, Values, Missing, Columns, Align, and Measure. The first row shows 'id' as a Numeric variable with a width of 4 and 0 decimal places. A 'Variable Type' dialog box is open over the 'id' row, with 'Numeric' selected. The dialog also shows 'Width: 4' and 'Decimal Places: 0'. The background table contains the following data:

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	id	Numeric	4	0	Employee Code	None	None	8	Right	Scale
2	gender	String	1	0	Gender	{f, Female}...	None	1	Left	Nominal
3	education	String	13	0	Education	None	None	13	Right	Scale
4	experience	String	8	0	Level (years)	{0, 0 (Missing)}	0	8	Right	Ordinal
5	jobcategory	String	8	0	Category	{0, 0 (Missing)}	0	8	Right	Ordinal
6	salary	String	8	0	Salary	{\$0, missing}...	\$0	8	Right	Scale
7	hiredate	String	8	0	Hire Date	{\$0, missing}...	\$0	8	Right	Scale
8	hiredate	String	8	0	Hire Date	{0, missing}...	0	8	Right	Scale
9	experience	String	8	0	Experience (months)	{0, missing}...	None	8	Right	Scale
10	classification	String	8	0	Classification	{0, No}...	9	8	Right	Ordinal

Загрузка (чтение) данных из файла Excel

Две опции:

1 – Скопировать данные из файла Excel (copy) и вставить данные (paste) во вкладке Data View в базу данных SPSS

2 – Прочитать (загрузить, открыть) (Read) данные из файла Excel (.xls)

Загрузка (чтение) данных из файла Excel

- Перед чтением в SPSS файла Excel необходимо убедиться, что все текстовые данные (например в категориальных переменных – легкой/средней/тяжелой степени тяжести) переведены в числовой вид
- При переводе данных в числовой вид, которые были введены в файл Excel как текст, необходимо в Microsoft Word создать список кодов переменных (например: «Нет»=0, «Да»=1, «Отказ о ответа»=99 и т.п.)

Загрузка (чтение) данных из файла Excel

Опция 1: Скопируйте (Copy) и вставьте (Paste) данные из базы Excel (spreadsheet) напрямую в редактор данных (Data Editor) во вкладке (Data View)

Опция 2:

- 1) Запустите SPSS.
- 2) Выберите файл: File → Open → Data
- 3) Выберите тип файла в поле (Files of type) Excel (*.xls)
- 4) Выберите необходимый файл Excel и нажмите Open
- 5) Проверьте, что в поле “Read variable names from the first row of data” поставлена галочка OK, что укажет для SPSS что первая строка в файле Excel содержит имена переменных (столбцов данных в Excel)
- 6) Нажмите «Open» и данные из файла Excel появятся в редакторе данных (Data Editor) SPSS

Правила присвоения имен для переменных (Variable Names)

- После ввода данных путем чтения файла Excel или перед ручным вводом данных во вкладке Data View необходимо задать список имен переменных во вкладке Variable View согласно правилам:
 - Имя переменной должно начинаться с буквы
 - Без пробелов
 - Имя должно быть уникальным
 - Допускаются символы: @ # \$
 - Символ «.» точка допускается, но не в качестве последнего символа в имени переменной
 - Символы * ! ? ' не допускаются
 - Символы не являются регистр-специфичными – то есть, например имена переменной ID, id, Id или iD будут расцениваться программой как одинаковые

Правила присвоения имен для переменных (Variable Names)

Некоторые ключевые слова,
используемые в синтаксисе SPSS, не
допускаются как имена переменных:

ALL TO WITH BY AND
OR NOT

Сохранение базы данных (Saving Data Worksheet)

Две опции:

1. Кнопка **Save**



2. **File|Save As**

Рекомендуется в названии файла указывать дату сохранения и версию базы данных, например:

C:\Database\HEALTH_10122016_V1.SAV

C:\USERS\HEALTH_04012017_V2.SAV

C:\USERS\HEALTH_15022017_V3.SAV

Окно просмотра результатов (Output Viewer)

Окно открывается автоматически в ходе работы с данными:
сохранение данных, перекодировка и трансформация
переменных, статистический анализ данных

The screenshot shows the SPSS Output Viewer window with the following content:

Output [Document1] - SPSS Viewer
File Edit View Data Transform Insert Format Analyze Graphs Utilities Window Help

Output
Log
Frequencies
Title
Notes
Active Dataset
Statistics
Educational Level (years)
Log
Frequencies
Title
Notes
Active Dataset
Statistics
Employment Category

Frequencies
[DataSet1] C:\Program Files\SPSS\Employee_data.sav

Statistics
Educational Level (years)

N	Valid	474
	Missing	0

Educational Level (years)

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 8	53	11.2	11.2	11.2
12	190	40.1	40.1	51.3
14	8	1.3	1.3	52.5
15	116	24.5	24.5	77.0
16	99	12.4	12.4	89.5
17	11	2.3	2.3	91.9
18	9	1.9	1.9	93.7
19	27	5.7	5.7	99.4
20	2	.4	.4	99.8
21	1	.2	.2	100.0
Total	474	100.0	100.0	

FREQUENCIES
VARIABLES=jobcat
/ORDER= ANALYSIS .

Frequencies
[DataSet1] C:\Program Files\SPSS\Employee_data.sav

Statistics
Employment Category

N	Valid	474
	Missing	0

SPSS Procedure is ready

Проверка данных после ввода данных (Clean Data)

- Проверьте введены ли метки значений переменных для категориальных, интервальных и ординальных (порядковых) переменных (Values)
- Проверьте указаны ли метки (полные названия) переменных для каждой переменной
- Используя команду “Frequencies” и окно отображения результатов работы с данными для каждой переменной проверьте имеются ли ошибки ввода данных («неправильные значения»)
- Проверьте пропущенные данные (Missing data): если имеются пропуски – поднимите анкеты и проверьте наличие пропущенных данных
- Проверьте можно ли перекодировать текстовые переменные (String variable) в числовую переменную

Проверка данных после ввода данных (Clean Data)

Q1 How old are you

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 12 years old or younger	19	.1	.1	.1
	2 13 years old	17	.1	.1	.2
	3 14 years old	1851	11.3	11.3	11.5
	4 15 years old	4045	24.7	24.7	36.3
	5 16 years old	4234	25.8	25.9	62.2
	6 17 years old	3963	24.2	24.2	86.4
	7 18 years old or older	2215	13.5	13.6	100.0
	8	1	.0	.0	100.0
	9	1	.0	.0	100.0
	Total	16344	99.6	100.0	
Missing	System	66	.4		
Total		16410	100.0		

Сохранение результатов работы с данными (сохранение Output)

При работе с окном просмотра результатов (Output Viewer window) его можно сохранить:

- **File|Save As**
- **формат .spv**

Завершение работы с SPSS

При работе с окном Data Editor выбрать **File|Exit.**

Перед выходом из программы рекомендуется сохранить базу данных (.sav) и файл результатов работы с данными (.spv)

Спасибо за внимание