

Выполнила:
учитель информатики и ИКТ
Полюхина И. Ю.

ПРОГРАММА SPSS

2014 г.

ИСТОРИЯ



SPSS является самой распространённой программой для обработки статистической информации.

Два студента Норман Най (Norman Nie) и Дейл Вент (Dale Bent), специализировавшиеся в области политологии в 1965 году пытались отыскать в Стенфордском университете Сан-Франциско компьютерную программу, подходящую для анализа статистической информации.



Вскоре они разочаровались в своих попытках, так как имеющиеся программы оказывались более или менее непригодными, неудачно построенными или не обеспечивали наглядность представления обработанной информации. К тому же принципы пользования менялись от программы к программе.

The image shows the SPSS logo, which consists of the letters "SPSS" in a bold, red, sans-serif font. The logo is centered within a white rectangular box that has a thin blue border. The background of the slide is a light yellow with a subtle diagonal line pattern.

Так, не долго думая, они решили разработать собственную программу, со своей концепцией и единым синтаксисом. В их распоряжении тогда был язык программирования FORTRAN и вычислительная машина типа IBM 7090.

Уже через год была разработана первая версия программы, которая, еще через год, в 1967, могла работать на IBM 360. К этому времени к группе разработчиков присоединился Хэдлай Халл (Hadlai Hull).



IBM 7090



IBM 360

Figure 100. Console — IBM System/360 Model 30

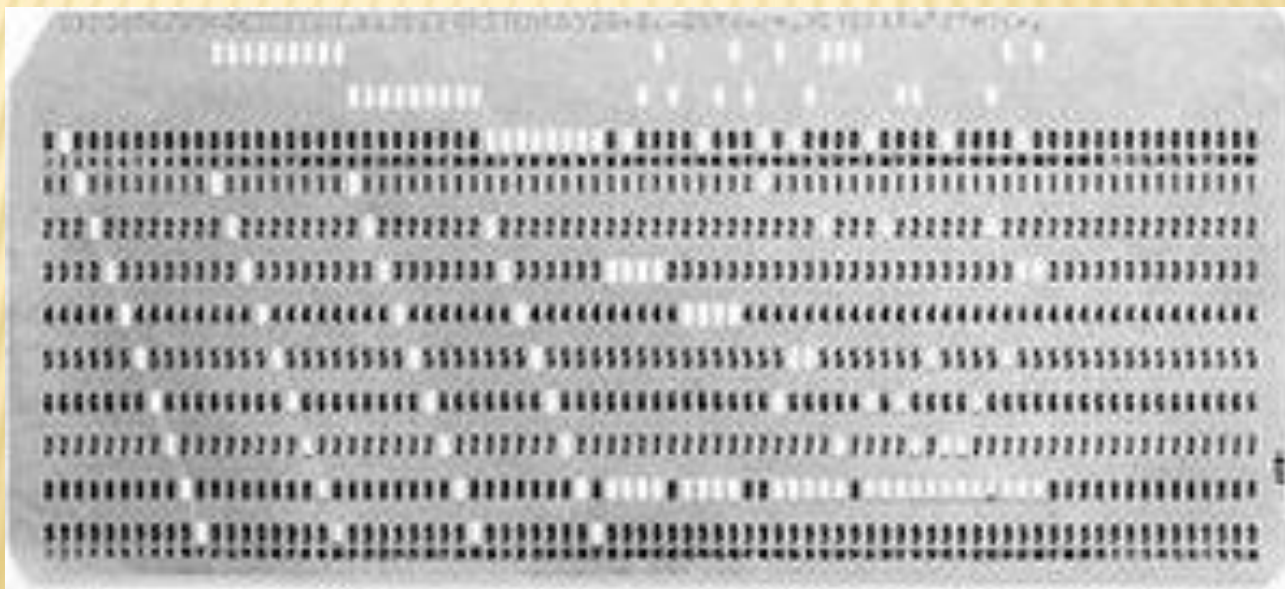
Как известно из истории развития информатики, программы тогда представляли собой пакеты перфокарт. Как раз на это указывает и исходное название программы, которое авторы дали своему продукту: SPSS — это аббревиатура от Statistical Package for the Social Science.

В 1970 году работа над программой была продолжена в Чикагском университете, а Норман Най основал соответствующую фирму — к тому моменту уже было произведено шестьдесят инсталляций. Первое руководство для пользователей описывало одиннадцать различных процедур.



Спустя пять лет SPSS была уже инсталлирована шестьсот раз, причём под разными операционными системами. С самого начала версиям программы присваивали соответствующие порядковые номера. В 1975 была разработана уже шестая версия (SPSS6). До 1981 последовали версии 7, 8 и 9.

Командный язык (синтаксис) SPSS в то время был ещё не так хорошо развит, как сейчас, и естественно ориентирован на перфокарты. Поэтому так называемые управляющие карты SPSS состояли из идентификационного поля (столбцы 1-15) и из поля параметров (столбцы 16-80).



В 1983 году командный язык SPSS был полностью переработан, синтаксис стал значительно удобней. Что бы отметить этот факт, программа была переименована в SPSSX, где буква X должна была служить как номером версии в римскими числами, так и сокращением для extended (расширенный).



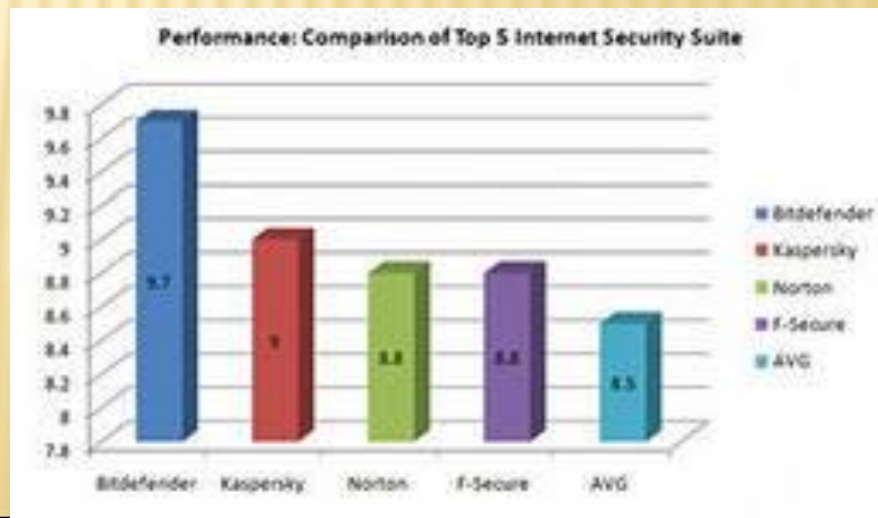
Так как применение перфокарт к этому моменту уже стало историей, то программа SPSS и информация, подлежащая обработке, сохранялись в отдельных файлах на винчестерах больших ЭВМ, которые тогда использовались повсеместно. Год от года постоянно увеличивалось и количество процедур



С появлением персональных компьютеров была разработана также и PC-версия SPSS, с 1983 года появилась PC-версия SPSS\PC+. рассчитанная на MS-DOS. Позже, с момента основания в 1984 году европейского торгового представительства в Горинхеме в Нидерландах, SPSS стал широко применяться и в Европе. В настоящее время это самое распространённое программное обеспечение для статистического анализа во всём мире.



Для того, чтобы отразить возможность использования программы во всех областях, имеющих отношение к статистическому анализу, буква X вновь была удалена из названия марки, а исходной аббревиатуре присвоено новое значение: Superior Performance Software System (система программного обеспечения высшей производительности).



Если PC версия SPSS/PC+ была чуть усовершенствованной версией для больших ЭВМ, то SPSS для операционной системой Windows (SPSS for Windows) стала большим шагом вперёд.

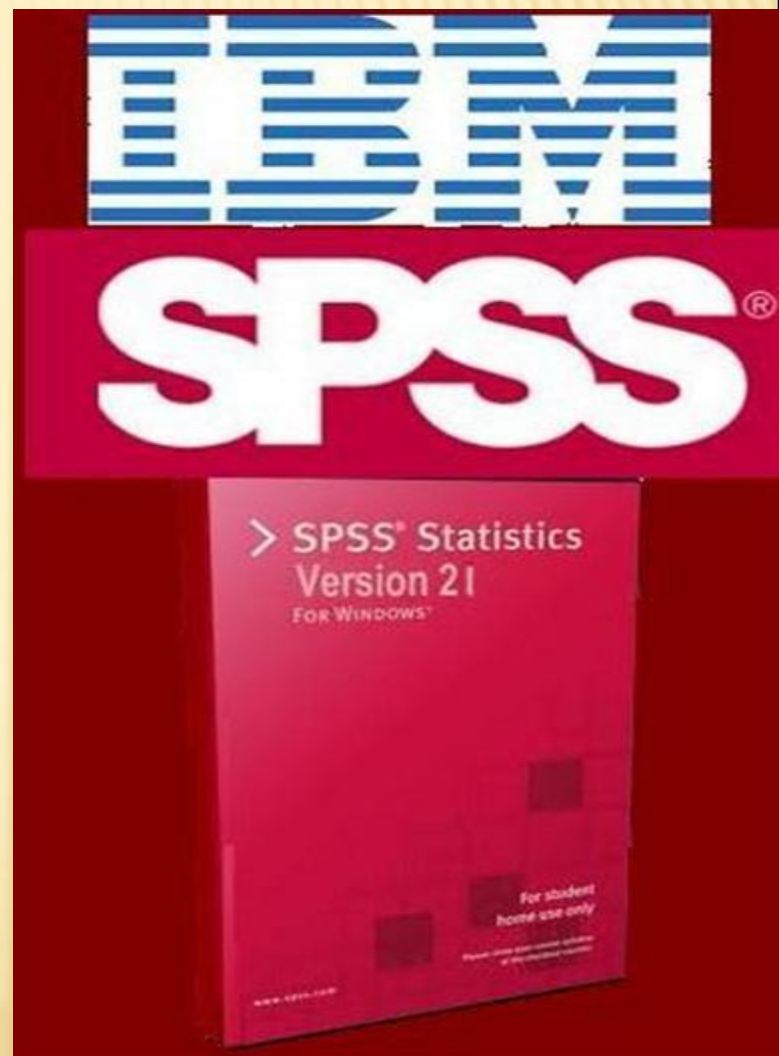


Эта версия SPSS обладает всеми возможностями версии для больших ЭВМ. За некоторыми немногочисленными исключениями, программой можно пользоваться без особых знаний в области прикладного программирования. Вызов необходимых процедур статистического анализа происходит при помощи стандартной техники, применяемой в Windows, то есть с помощью мыши и соответствующих диалоговых окон.



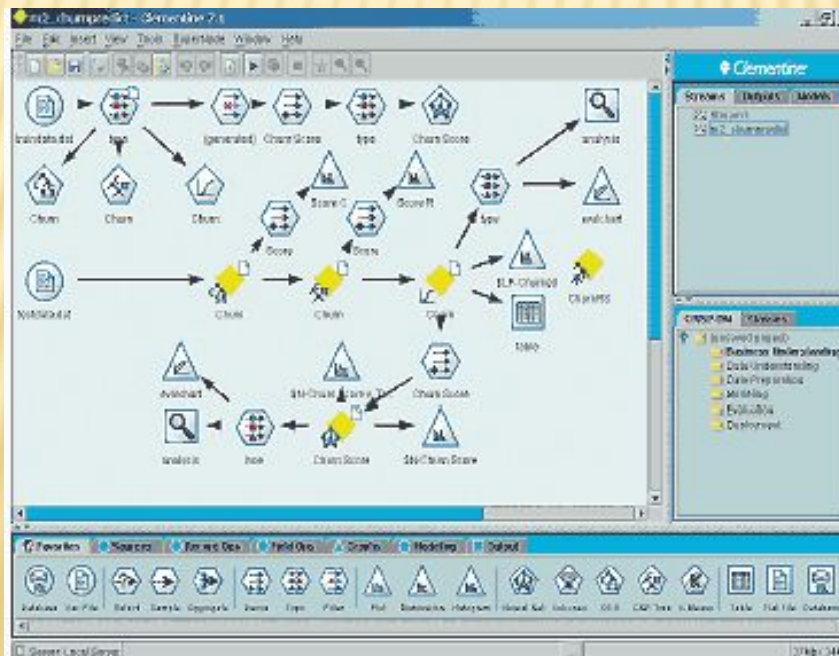
В сентябре 2012 г. вышла очередная версия пакета - IBM SPSS Statistics 21, работающая под управлением ОС Windows, MacOSX и Linux.

Но на данном этапе во многих учебных заведениях пользуются версией SPSS 17.0.



МОДУЛИ SPSS

Основу программы SPSS составляет SPSS Base (базовый модуль), предоставляющий разнообразные возможности доступа к данным и управления данными. Он содержит методы анализа, которые применяются чаще всего.



Традиционно вместе с SPSS Base (базовым модулем) поставляются ещё два модуля: Advanced Models (продвинутые модели) и Regression Models (регрессионные модели). Эти три модуля охватывают тот спектр методов анализа, который входил в раннюю версию программы для больших ЭВМ.



SPSS BASE (БАЗОВЫЙ МОДУЛЬ)

SPSS Base входит в базовую поставку. Он включает все процедуры ввода, отбора и корректировки данных, а также большинство предлагаемых в SPSS статистических методов. Наряду с простыми методиками статистического анализа, такими как частотный анализ, расчет статистических характеристик, таблиц сопряженности, корреляций, построения графиков, этот модуль включает t-тесты и большое количество других непараметрических тестов, а также усложненные методы, такие как многомерный линейный регрессионный анализ, дискриминантный анализ, факторный анализ, кластерный анализ, дисперсионный анализ, анализ пригодности (анализ надежности) и многомерное шкалирование.

*stroke_survival.sav [DataSet2] - IBM SPSS Statistics Data Editor

File Edit View Data Transform **Analyze** Direct Missing Values Graphs Utilities Add-ons Window Help

Reports
Descriptive Statistics
Tables
Compare Means
General Linear Model
Generalized Linear Models
Mixed Models
Correlate
Regression
Loglinear
Neural Networks
Classify
Dimension Reduction
Scale
Nonparametric Tests
Forecasting
Survival
Multiple Response
Missing Value Analysis...
Multiple Imputation
Complex Samples
Quality Control
ROC Curve...

Automatic Linear Modeling...
Linear...
Curve Estimation...
Partial Least Squares...
Binary Logistic...
Multinomial Logistic...
Ordinal...
Probit...
Nonlinear...
Weight Estimation...
2-Stage Least Squares...
Optimal Scaling (CATREG)...

Value: 42 of 42 Variables

	gender	active	obesity	diabetes	bp
54	Female	Yes	No	No	Hypotension
74	Male	Yes	Yes	No	Hypertension
74	Female	Yes	Yes	Yes	Hypertension
74	Male	Yes	No	No	Normal
			No	No	Hypertension
			Yes	No	Normal
			No	Yes	Normal
			No	No	Normal
			No	No	Normal
			No	No	Hypertension
			Yes	No	Normal
			No	No	Normal
			No	No	Hypotension
			Yes	No	Normal
			Yes	Yes	Normal
			Yes	No	Hypertension
			No	No	Hypertension
64	Male	Yes	No	No	Hypertension
55-64	Female	No	No	No	Normal
83	Male	Yes	No	No	Hypotension
67	Female	Yes	Yes	No	Normal

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21

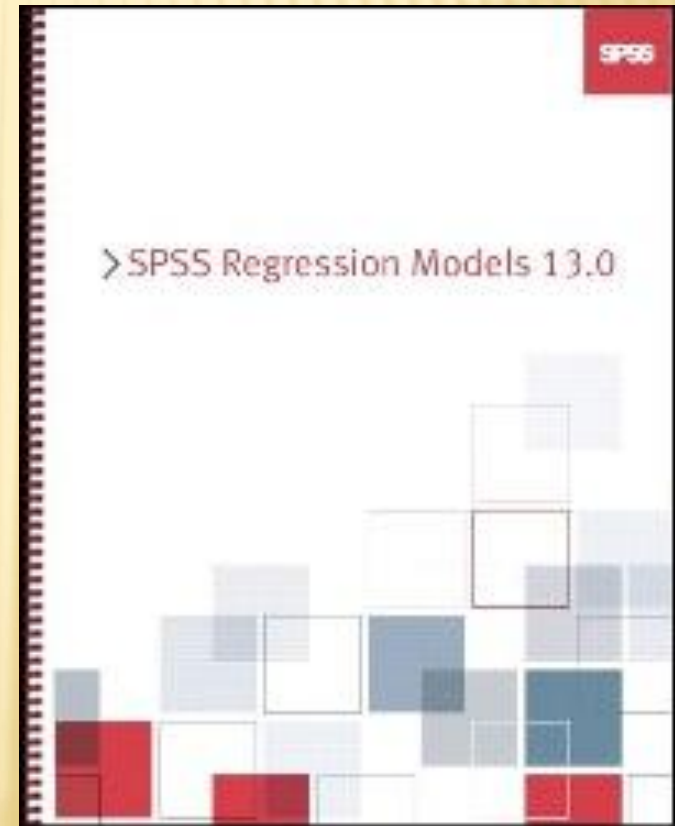
Data View Variable View

Linear...

IBM SPSS Statistics Processor is ready

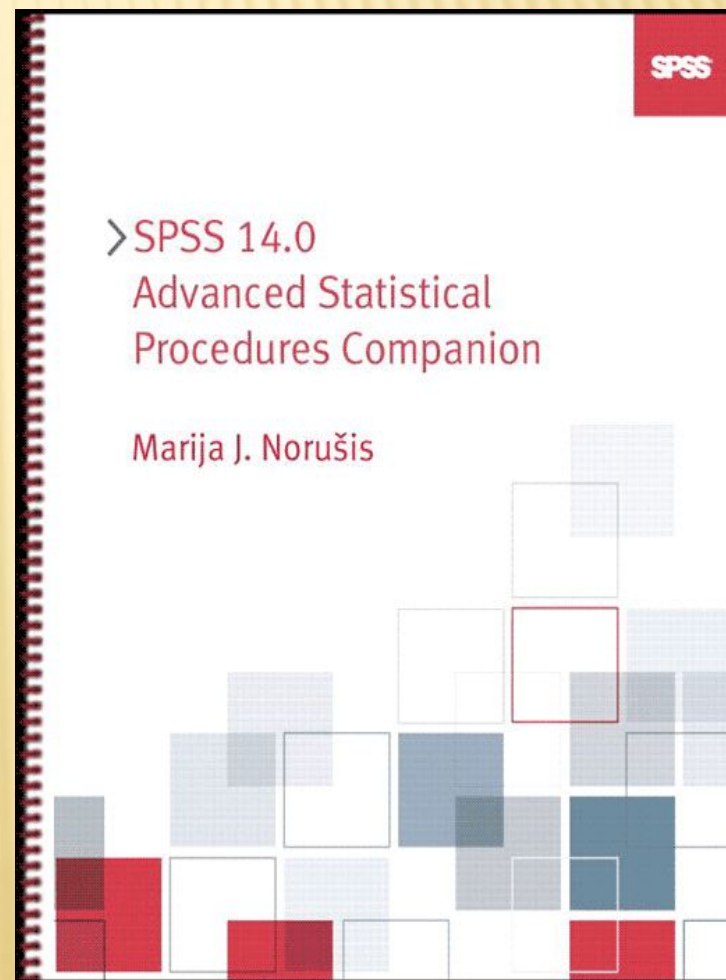
REGRESSION MODELS

Данный модуль включает в себя различные методы регрессионного анализа, такие как: бинарная и мультиномиальная логистическая регрессия, нелинейная регрессия и пробит-анализ.



ADVANCED MODELS

В этот модуль входят различные методы дисперсионного анализа (многомерный, с учетом повторных измерений), общая линейная модель, анализ выживания, включая метод Каплана-Майера и регрессию Кокса, логлинейные, а также логитлоглинсйные модели.



TABLES

Модуль Tables служит для создания презентационных таблиц. Здесь предоставляются более широкие возможности по сравнению со упрощенными частотными таблицами и таблицами сопряженности, которые строятся в SPSS Base (базовом модуле).

The screenshot displays the SPSS interface with three main windows open:

- Custom Tables Dialog Box:** Shows the configuration for creating a custom table. The 'Variables' list includes 'Respondent's Sex', 'Race of Respondent', 'Region of the U.S.', 'General Happiness', 'Is Life Exciting or Dull', 'Number of Births', 'Number of Children', 'Age of Respondent', 'Highest Year of Schooling', and 'Highest Year Spent in School'. The 'Columns' section shows a hierarchy: 'Is Life Exciting or Dull' (with sub-columns 'Exciting' and 'Dull'), 'Routine' (with sub-columns 'Count' and 'Count'), and 'Dull' (with sub-columns 'Count' and 'Count'). The 'Layer Output' section is set to 'Show each category as a layer'.
- Custom Tables Output Window:** Displays the results for 'Table 1' titled 'Region of the United States North East'. It shows a table with columns for 'Is Life Exciting or Dull' (Exciting, Routine, Dull) and 'Count'. The data is broken down by 'General Happiness' (Very Happy, Pretty Happy, Not Too Happy).
- Data View Window:** Shows the raw data for the 'General Social Survey' dataset. The first few rows are visible, showing variables like 'sex', 'race', 'region', 'happy', 'life', and 'sibs'.

AMOS

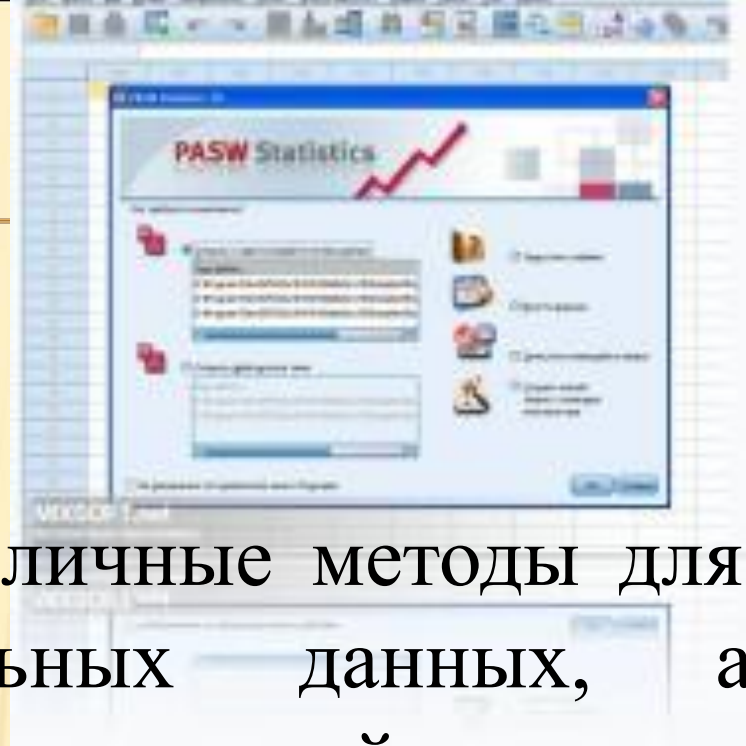
Amos (Analysis of moment structures — анализ моментных структур) включает методы анализа с помощью линейных структурных уравнений. Целью программы является проверка сложных теоретических связей между различными признаками случайного процесса и их описание при помощи подходящих коэффициентов. Проверка проводится в форме причинного анализа и анализа траектории. При этом пользователь в графическом виде должен задать теоретическую модель, в которую вместе с данными непосредственных наблюдений могут быть включены и так называемые скрытые элементы.



ANSWERTREE

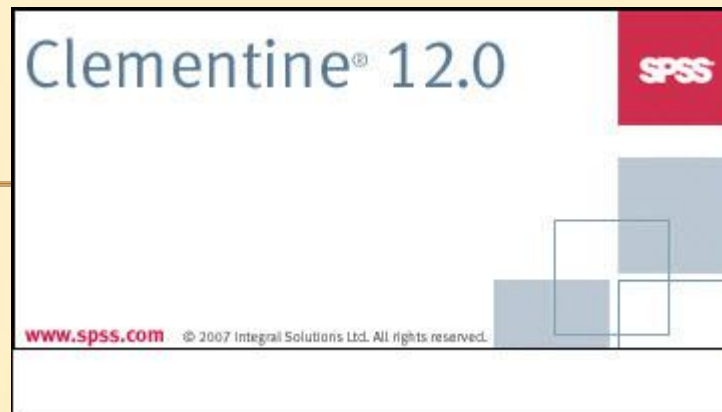
AnswerTree (дерево решений) включает четыре различных метода автоматизированного деления данных на отдельные группы (сегменты). Деление проводится таким образом, что частотные распределения целевой (зависимой) переменной в различных сегментах значительно различаются. Типичным примером применения данного метода является создание характерных профилей покупателей при исследовании потребительского рынка. AnswerTree является преемницей программы СНАШ (Chi squared interaction Detector — детектор взаимодействий на основе хи-квадрата).

CATEGORIES



Модуль содержит различные методы для анализа категориальных данных, а именно: анализ соответствий и три различных метода оптимального шкалирования (анализ однородности, нелинейный анализ главных компонент, нелинейный канонический корреляционный анализ).

CLEMENTINE



Clementine — это программа для data mining (добычи знаний), в которой пользователю предлагаются многочисленные подходы к построению моделей, к примеру, нейронные сети, деревья решений, различные виды регрессионного анализа. Clementine представляет собой "верстак" аналитика, при помощи которого можно визуализировать процесс моделирования, перепроверять модели, сравнивать их между собой. Для удобства пользования программой имеется вспомогательная среда внедрения результатов.

CONJOINT (СОВМЕСТИНЫЙ АНАЛИЗ)

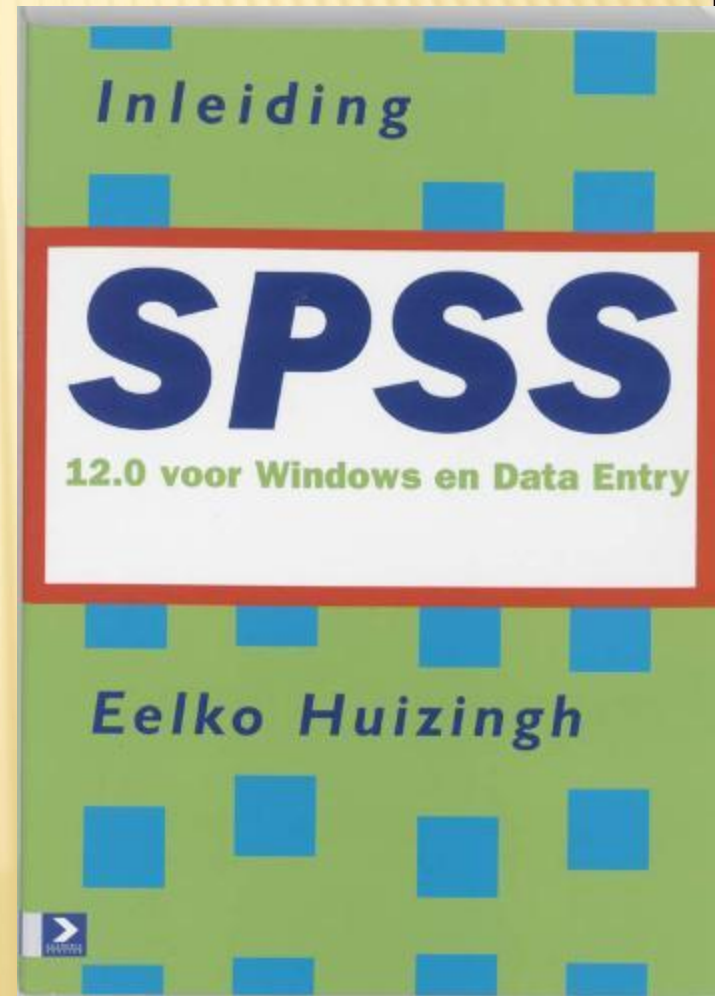
IBM SPSS Conjoint 21

IBM.

Совместный анализ применяется при исследовании рынка для изучения потребительских свойств продуктов на предмет их привлекательности. При этом опрашиваемые респонденты по своему усмотрению должны расположить предлагаемые наборы потребительских свойств продуктов в порядке предпочтения, на основании которого можно затем вывести так называемые детализированные показатели полезности отдельных категорий каждого потребительских свойства.

DATA ENTRY (ВВОД ДАННЫХ)

Программа Data Entry предназначена для быстрого составления вопросников, а также ввода и чистки данных. Заданные на этапе создания вопросника вопросы и категории ответов потом используются в качестве меток переменных и значений.



EXACT TESTS (ТОЧНЫЕ ТЕСТЫ)

Данный модуль служит для вычисления точного значения вероятности ошибки (величины p) в условиях ограниченности данных при проверке по критерию χ^2 (Chi-Quadrat-Test) и при непараметрических тестах. В случае необходимости для этого также может быть применён метод Монте-Карло (Monte-Carlo).

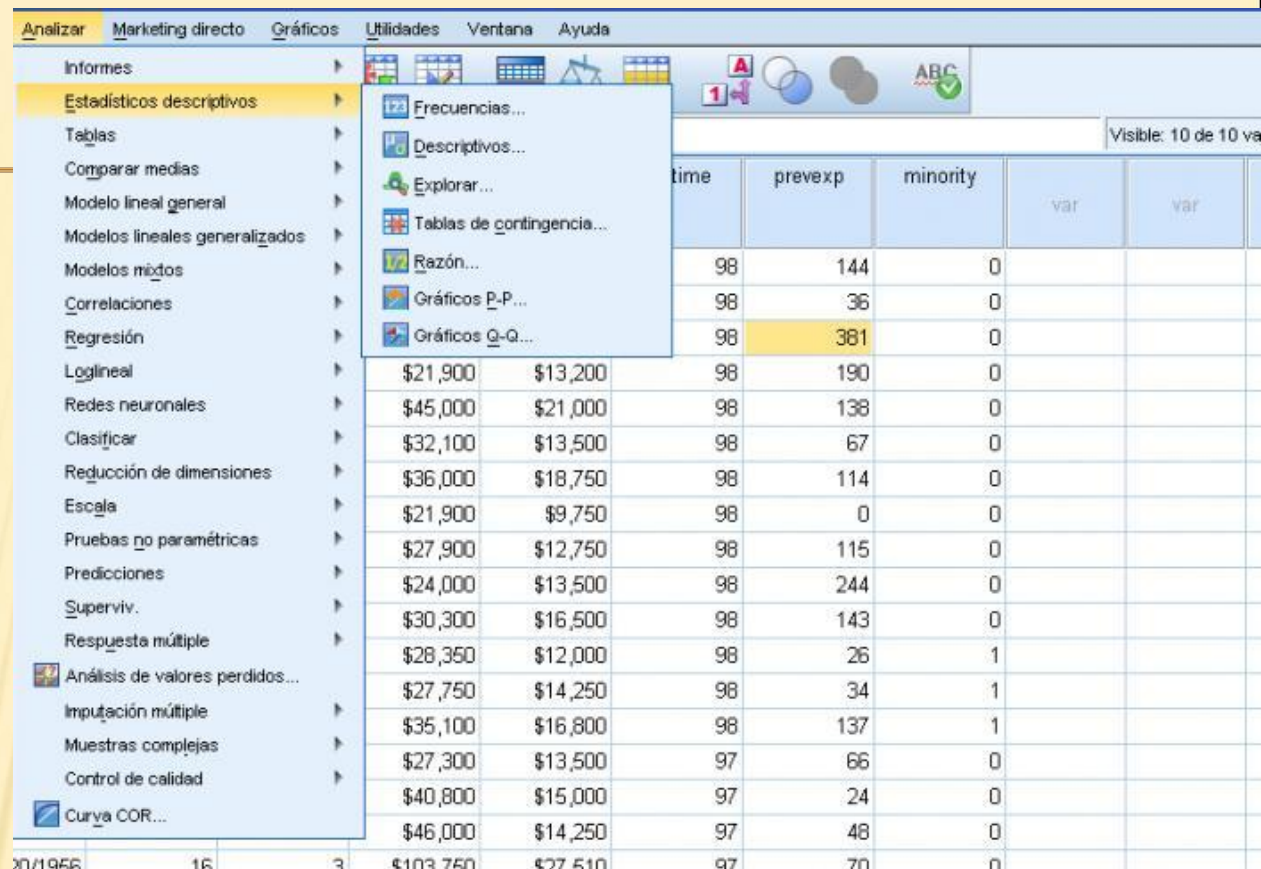


SAMPLEPOWER

При помощи SamplePower может быть определён оптимальный размер выборки для большинства методов статистического анализа, реализованных в SPSS.



GOLDMINER



The screenshot shows the GOLDMINER software interface. The 'Analizar' menu is open, displaying a list of statistical analysis options. The background shows a data table with columns 'time', 'prevexp', 'minority', 'var', and 'var'.

Analizar Menu Options:

- Informes
- Estadísticos descriptivos
- Tablas
- Comparar medias
- Modelo lineal general
- Modelos lineales generalizados
- Modelos mixtos
- Correlaciones
- Regresión
- Loglineal
- Redes neuronales
- Clasificar
- Reducción de dimensiones
- Escala
- Pruebas no paramétricas
- Predicciones
- Superviv.
- Respuesta múltiple
- Análisis de valores perdidos...
- Imputación múltiple
- Muestras complejas
- Control de calidad
- Curva COR...

Sub-menu for 'Estadísticos descriptivos':

- Frecuencias...
- Descriptivos...
- Explorar...
- Tablas de contingencia...
- Razón...
- Gráficos P-P...
- Gráficos Q-Q...

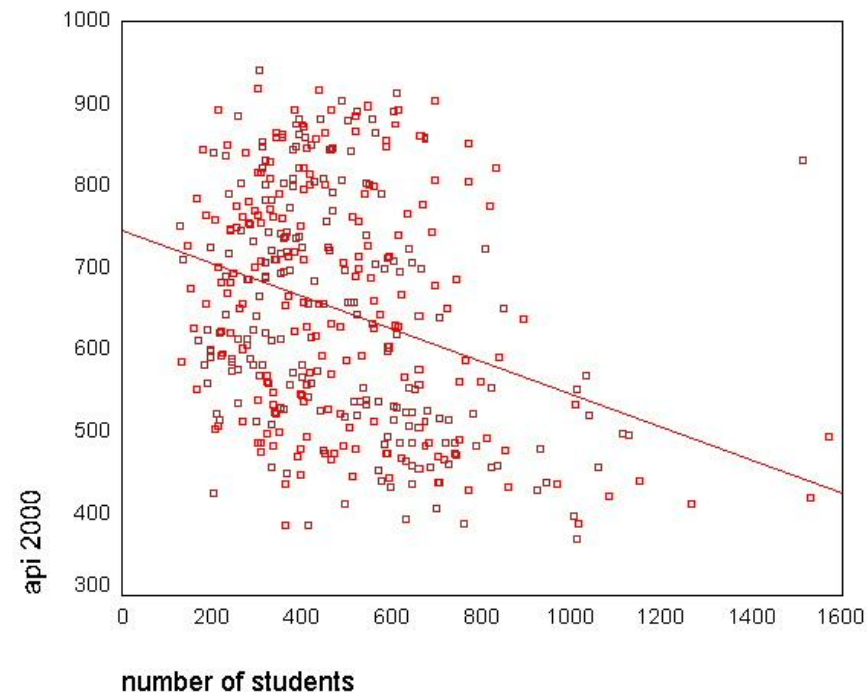
Background Data Table:

time	prevexp	minority	var	var
98	144	0		
98	36	0		
98	381	0		
\$21,900	\$13,200	98	190	0
\$45,000	\$21,000	98	138	0
\$32,100	\$13,500	98	67	0
\$36,000	\$18,750	98	114	0
\$21,900	\$9,750	98	0	0
\$27,900	\$12,750	98	115	0
\$24,000	\$13,500	98	244	0
\$30,300	\$16,500	98	143	0
\$28,350	\$12,000	98	26	1
\$27,750	\$14,250	98	34	1
\$35,100	\$16,800	98	137	1
\$27,300	\$13,500	97	66	0
\$40,800	\$15,000	97	24	0
\$46,000	\$14,250	97	48	0
\$103,750	\$27,510	97	70	0

Программа содержит специальную регрессионную модель для регрессионного анализа упорядоченных зависимых и независимых переменных.

SPSS MISSING VALUE ANALYSIS

Данный модуль служит для анализа и восстановления закономерностей, которым подчиняются пропущенные значения. Он предоставляет различные варианты замены недостающих значений.



TRENDS



Программа содержит специальную регрессионную модель для регрессионного анализа упорядоченных временных рядов. Модуль Trends содержит различные методы для анализа временных рядов, такие как: модели ARIMA, экспоненциальное сглаживание, сезонная декомпозиция и спектральный анализ.

Пример работы в SPSS

ИМПОРТ ДАННЫХ ИЗ ФАЙЛА ФОРМАТА EXCEL
(ACCESS) В SPSS

ПОДГОТОВКА ДАННЫХ ДЛЯ ИСПОЛЬЗОВАНИЯ ПАКЕТА SPSS.

Данные могут быть введены непосредственно в SPSS или могут быть импортированы из текстовых файлов, файлов формата EXCEL, ACCESS и т.д. Пусть нашем случае имеется файл данных в формате EXCEL (рис 1.1). Приступим к импорту данных в файл SPSS.

Эконометрика - задача из методки						
	A	B	C	D	E	F
1	Y	X1	X2	X3	X4	X5
2	126	1	4	15	17	100
3	137	2	4.8	14.8	17.3	98.4
4	148	3	3.8	15.2	16.8	101.2
5	191	4	8.7	15.5	16.2	103.5
6	274	5	8.2	15.5	16	104.1
7	370	6	9.7	16	18	107
8	432	7	14.7	18.1	20.2	107.4
9	445	8	18.7	13	15.8	108.5
10	367	9	19.8	15.8	18.2	108.3
11	367	10	10.6	16.9	16.8	109.2
12	321	11	8.6	16.3	17	110.1
13	307	12	6.5	16.1	18.3	110.7
14	331	13	12.6	15.4	16.4	110.3
15	345	14	6.5	15.7	16.2	111.8
16	364	15	5.8	16	17.7	112.3
17	384	16	5.7	15.1	16.2	112.9
18						

Рис.1.1. Файл исходных данных в формате EXCEL

Для импорта данных из файла формата EXCEL в SPSS выполните следующие действия:

1. После запуска программы SPSS и появления на экране пустого файла SPSS, подобного изображенному на Рис. 1.2, выберите в строке меню ***Файл-Открыть-Данные*** и задайте имя EXCEL файла с вашими данными, его тип, после чего нажмите кнопку OPEN (Рис. 1.3).

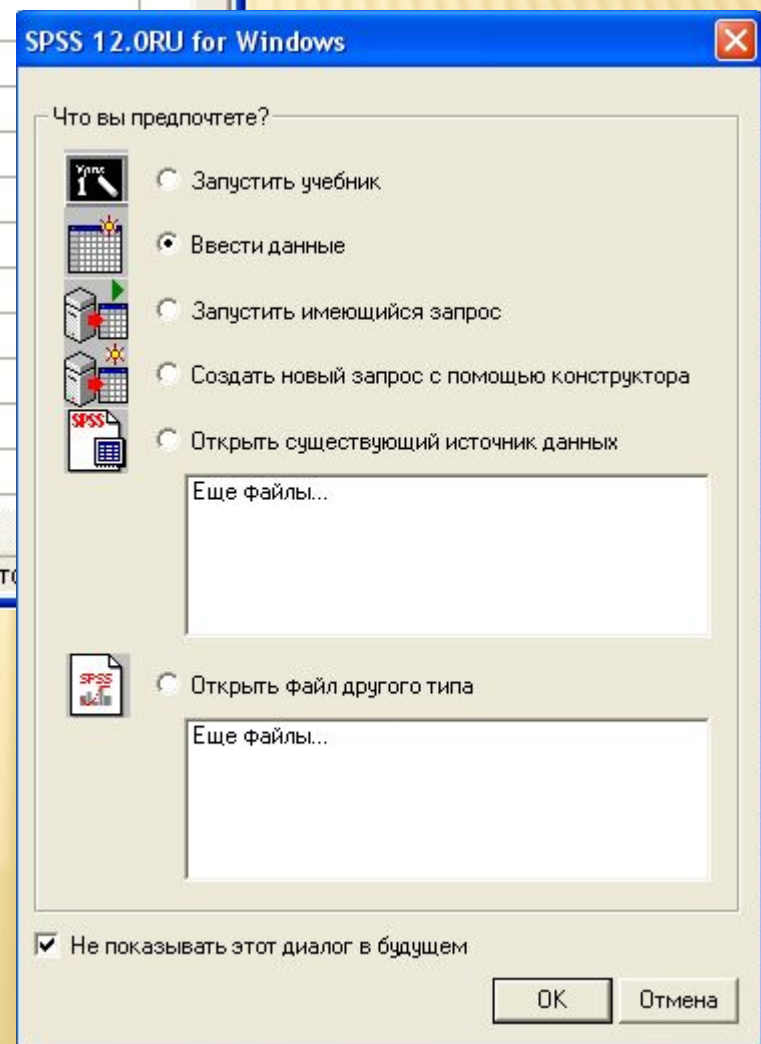
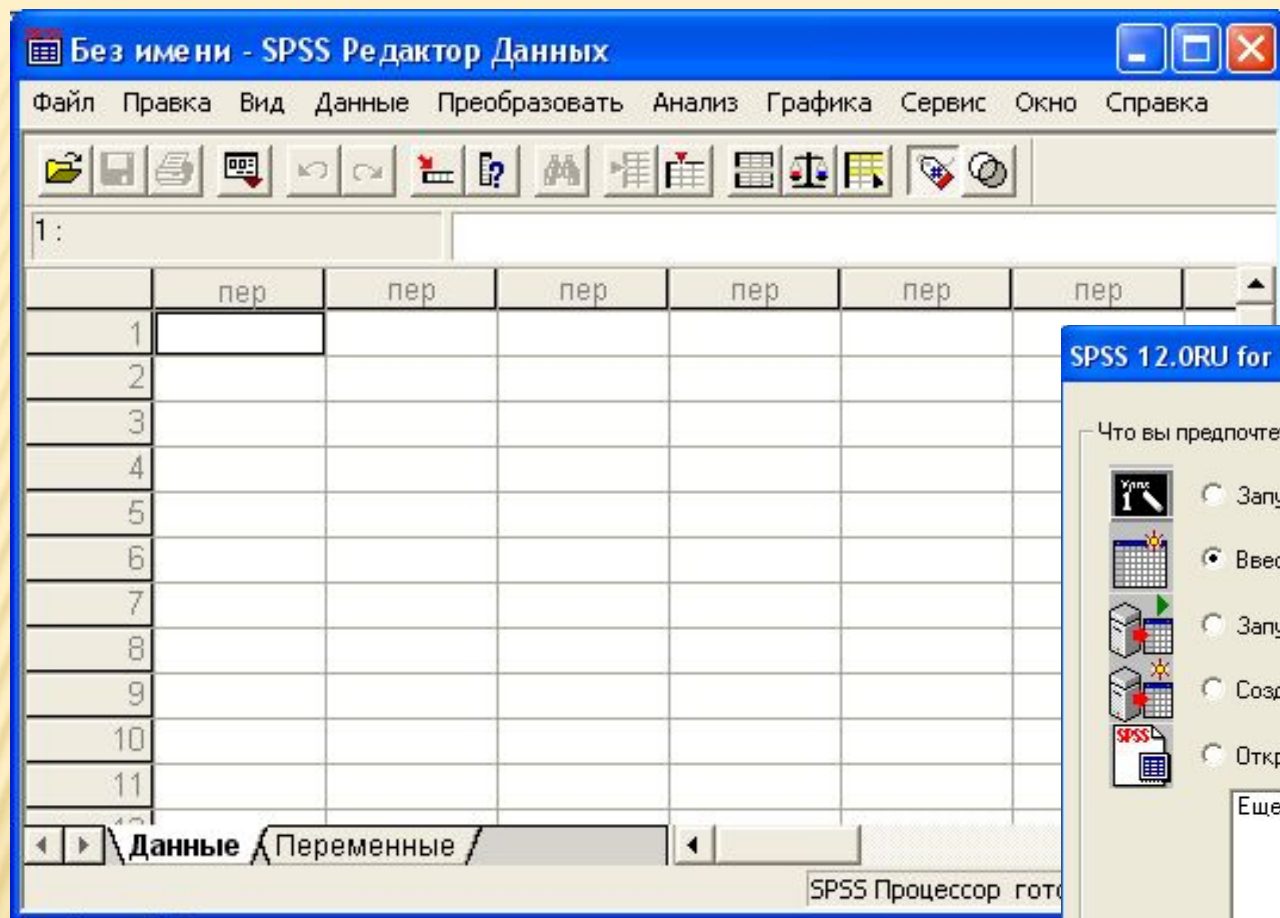


Рис. 1.2. Окно редактора данных в SPSS.

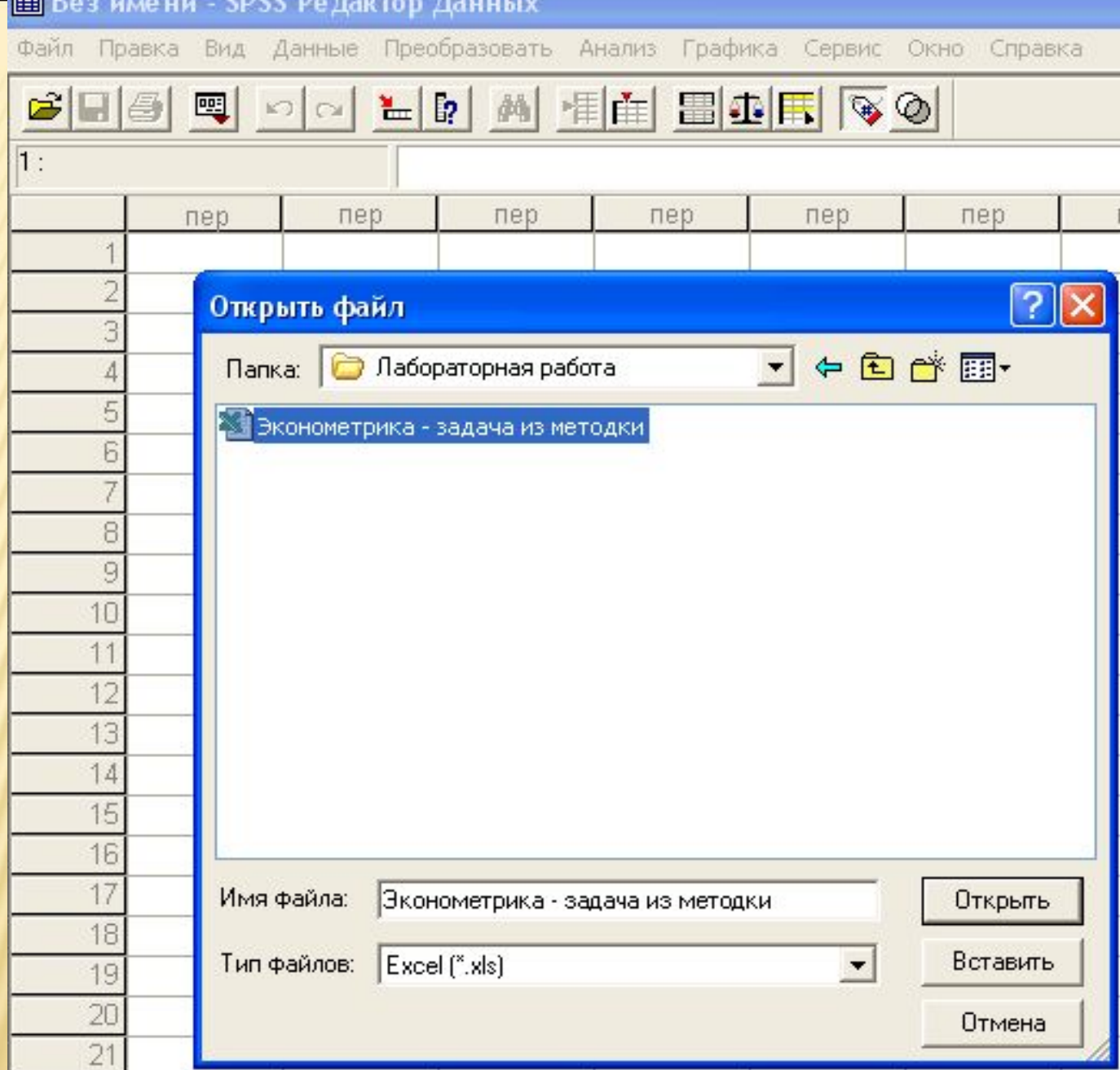


Рис.1.3. Импорт данных в SPSS.

2. В новом диалоговом окне (Рис.1.4) задайте диапазон ячеек (например, A1:F17). Если в первой строке EXCEL файла находятся имена переменных, то установить флажок ***Читать имена переменных из первой строки данных.*** Щелкните на кнопке ОК.

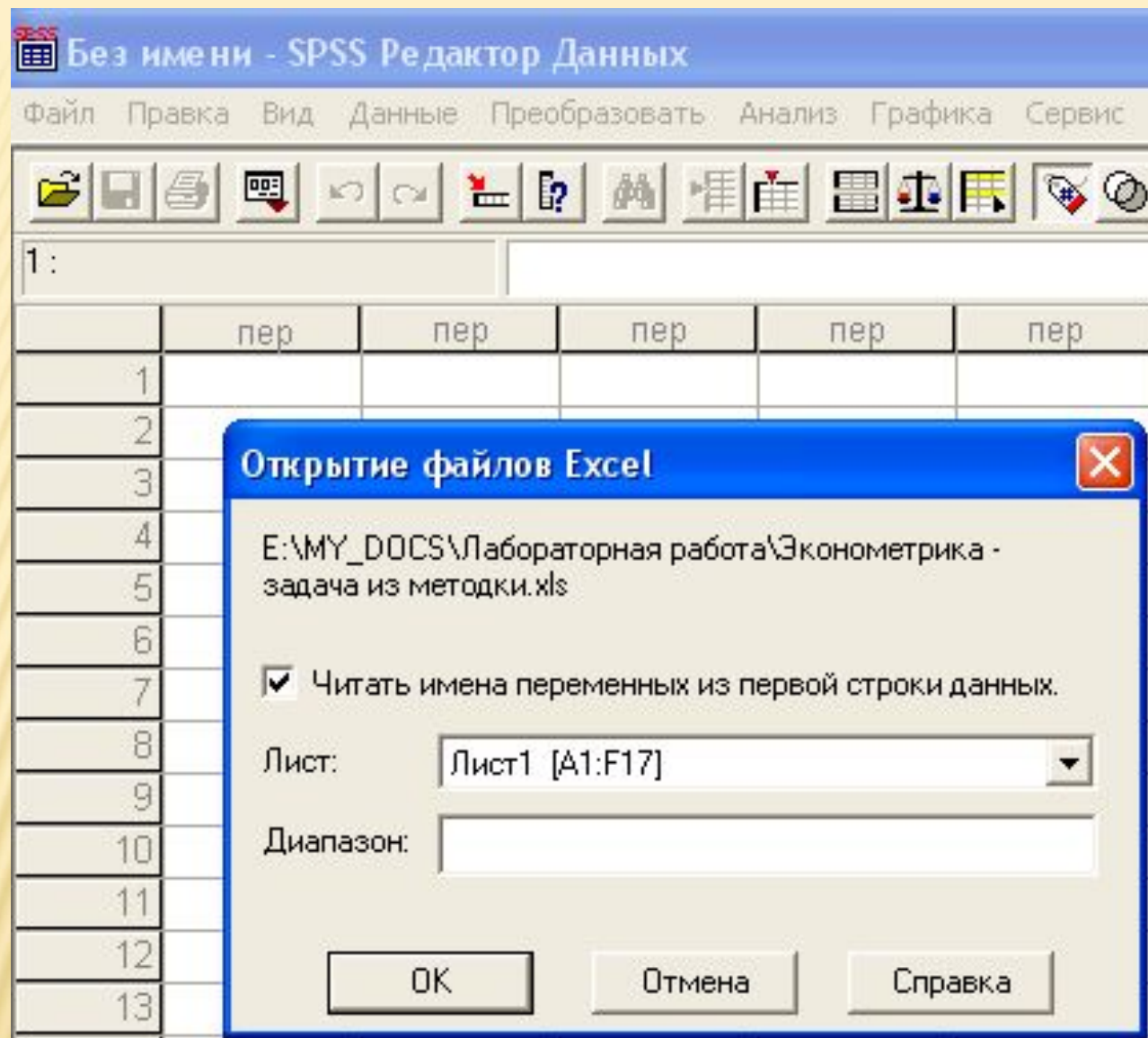


Рис.1.4. Импорт данных в SPSS.

3. На экране вы увидите
импортированные данные в формате
SPSS (Рис. 1.5).

Без имени - SPSS Редактор Данных

Файл Правка Вид Данные Преобразовать Анализ Графика Сервис Окно Справка

1: Y 126

	Y	X1	X2	X3	X4	X5	пер	пер
1	126	1	4.0	15.0	17.0	100.0		
2	137	2	4.8	14.8	17.3	98.4		
3	148	3	3.8	15.2	16.8	101.2		
4	191	4	8.7	15.5	16.2	103.5		
5	274	5	8.2	15.5	16.0	104.1		
6	370	6	9.7	16.0	18.0	107.0		
7	432	7	14.7	18.1	20.2	107.4		
8	445	8	18.7	13.0	15.8	108.5		
9	367	9	19.8	15.8	18.2	108.3		
10	367	10	10.6	16.9	16.8	109.2		
11	321	11	8.6	16.3	17.0	110.1		
12	307	12	6.5	16.1	18.3	110.7		
13	331	13	12.6	15.4	16.4	110.3		
14	345	14	6.5	15.7	16.2	111.8		
15	364	15	5.8	16.0	17.7	112.3		
16	384	16	5.7	15.1	16.2	112.9		

Данные Переменные / SPSS Процессор готов

Рис.1.5. Данные задачи представлены в формате SPSS.