

**Обзор прикладных
статистических программ для
решения задач
социологического
исследования**

Учебный курс

***«Основы применения прикладных
статистических программ
в социологических исследованиях»
(лекция 1- 2)***

Преподаватель: **Цихончик Надежда
Васильевна**, старший преподаватель
кафедры философии и социологии
САФУ

План лекции

1. Пакеты прикладных программ.
2. Пакеты прикладных программ для задач статистической обработки данных:
 - *Microsoft Excel*
 - *Statistica*
 - *SPSS (PSPP)*
 - *Vortex*
 - *другие пакеты*

1 вопрос лекции.

Пакеты прикладных программ

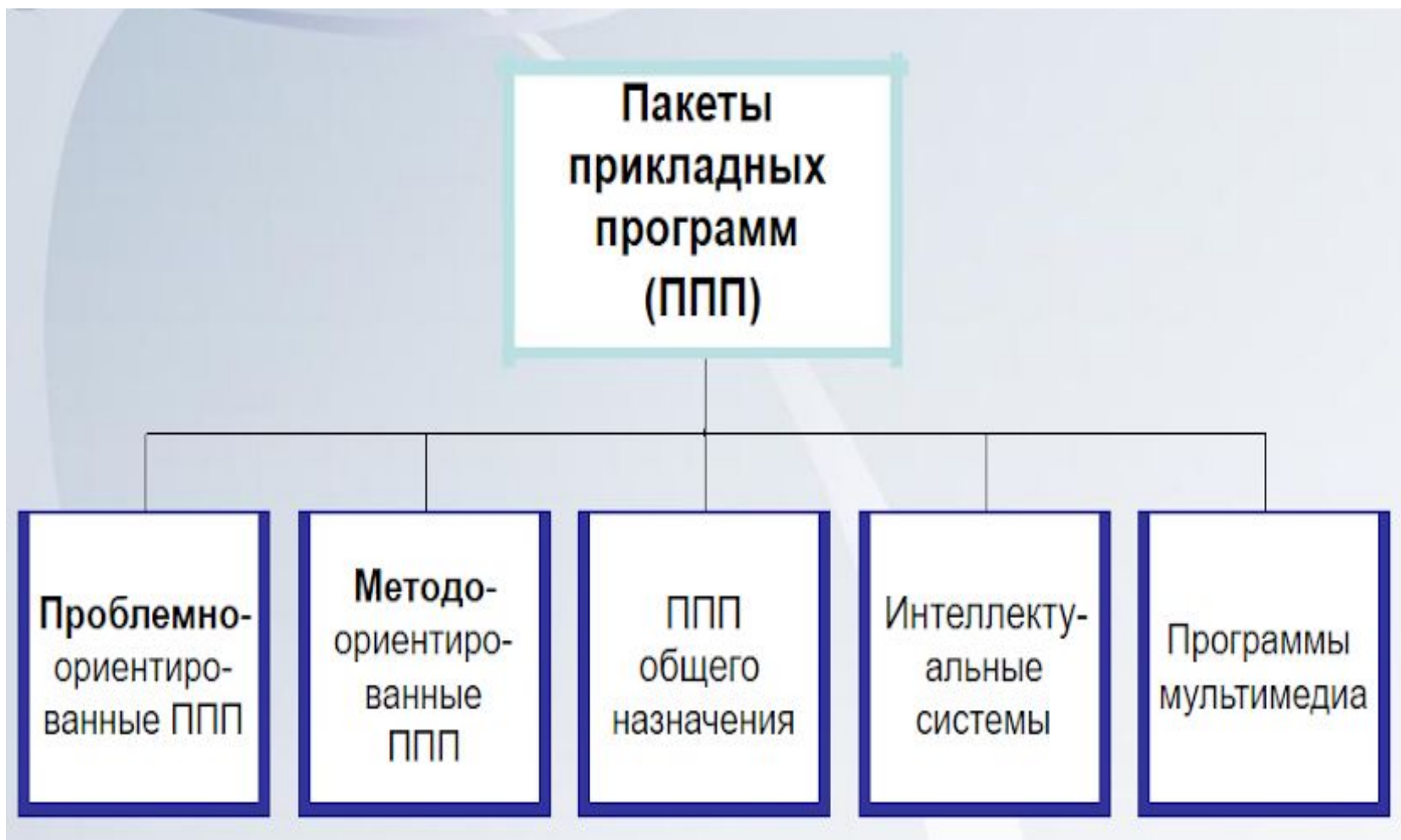
Пакеты прикладных программ

- **Пакеты прикладных программ (ППП)** (application program package) – комплекс взаимосвязанных программ для решения задач определенного класса конкретной предметной области.

Особенности:

- *Являются самым многочисленным классом программных продуктов.*
- *Предназначены для решения конкретных прикладных задач.*
- *Установка программных продуктов на компьютер выполняется квалифицированными пользователями.*
- *Эксплуатацию осуществляют конечные пользователи.*
- *Программные продукты могут быть весьма специфичным для отдельных предметных областей*

Пакеты прикладных программ



Общая характеристика статистических программ

- 2 группа - Методо-ориентированные ППП
- Пакеты статистических программ - компьютерные системы для анализа данных
- ориентированы на решение общих проблем обработки данных, например, на задачи обработки таблиц или на задачи статистического анализа

Виды статистических программ

- Все программы статистической обработки данных можно разделить на профессиональные, полупрофессиональные (популярные) и специализированные
- Большую часть статистических пакетов можно разбить на **две группы** — это статистические пакеты общего назначения и специализированные программные продукты



Виды статистических программ

- ***Универсальные пакеты*** – отсутствие прямой ориентации на специфическую предметную область, предлагают широкий диапазон статистических методов. Из зарубежных универсальных пакетов наиболее распространены BAS, SPSS, Systat, Minilab, Statgraphics, STATISTICA
- ***Специализированные пакеты*** - как правило, реализуют несколько статистических методов или методы, применяемые в конкретной предметной области. Из российских пакетов более известны STADIA, Олимп, Класс-Мастер, КВАЗАР, Статистик-Консультант; американские пакеты – ODA, WinSTAT, Statit и т.д.

Требования к статистическим пакетам общего назначения

1. содержит достаточно полный набор стандартных статистических методов;
2. достаточно прост для быстрого освоения и использования;
3. отвечает высоким требованиям к вводу, преобразованиям и организации хранения данных как в самом пакете, так и обмену с широко распространенными базами данных;
4. имеет широкий набор средств графического представления данных и результатов;
5. имеет подробную документацию, хорошо продуманную с учетом интересов как начинающего пользователя, так и специалиста-статистика.

Требования к СПП

- модульность;
- ассистирование при выборе способа обработки данных;
- использование простого проблемно-ориентированного языка для формулировки задания пользователя;
- автоматическая организация процесса обработки данных и связей с модулями пакета;
- ведение банка данных пользователя и составление отчета о результатах проделанного анализа;
- диалоговый режим работы пользователя с пакетом;
- совместимость с другим программным обеспечением

Литература



А. Наследов

SPSS 19

профессиональный
статистический
анализ данных

ПИТЕР®



2 вопрос лекции.

**Пакеты прикладных программ
для задач статистической
обработки данных**

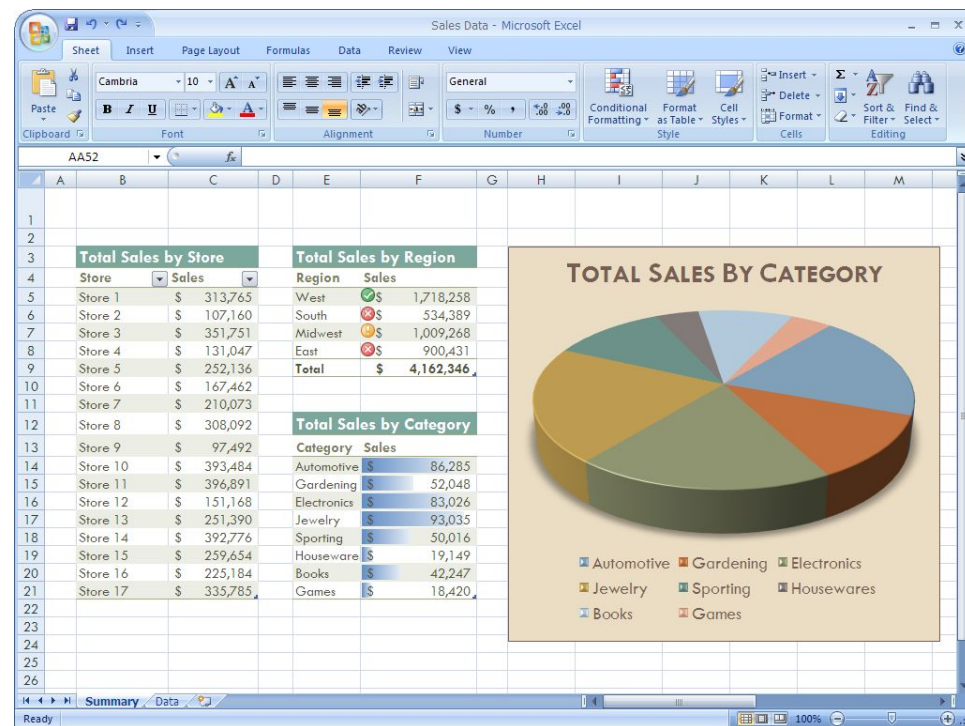


2.1. Microsoft Excel *(Microsoft Office Excel)*

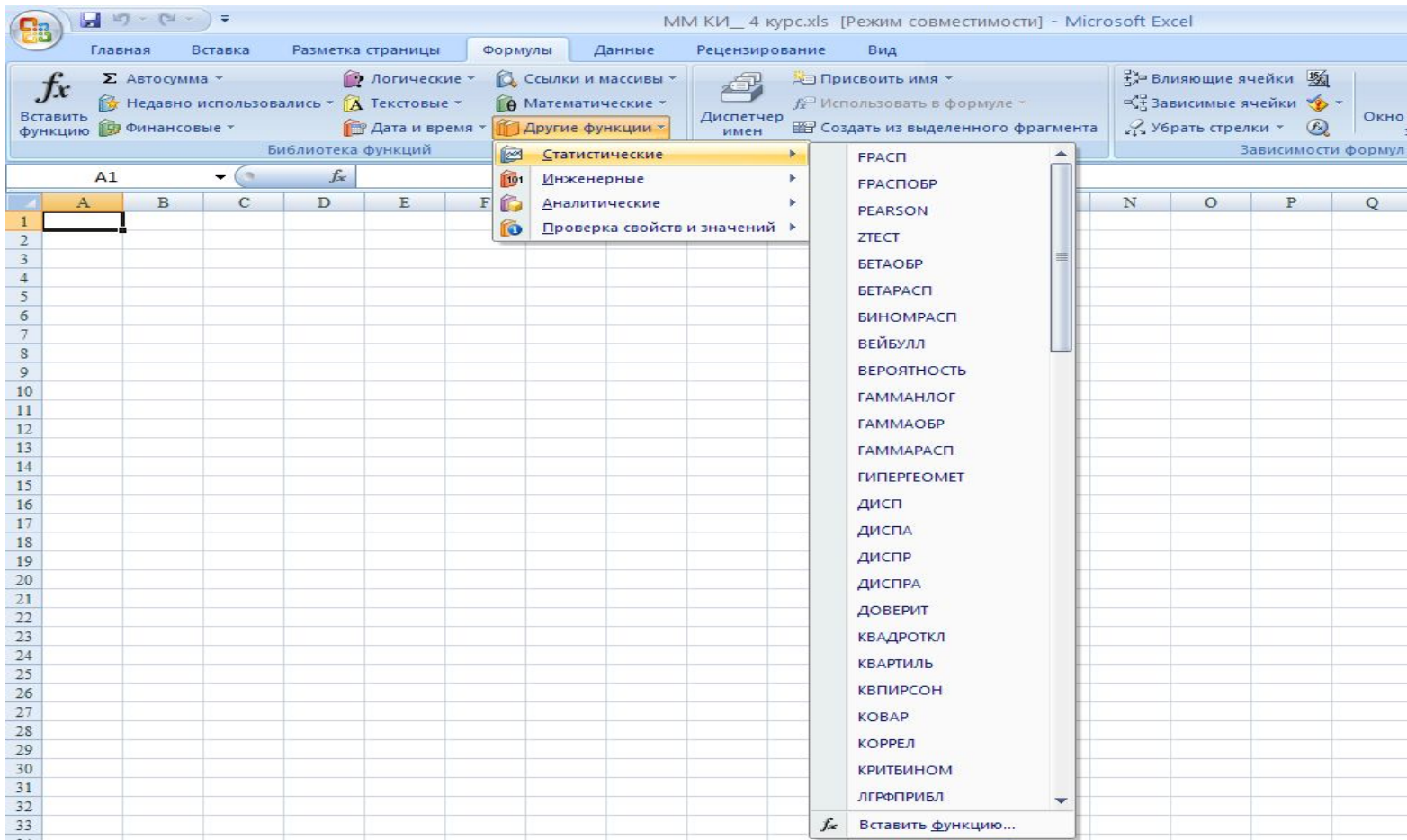
Электронные таблицы (на примере Excel)

Microsoft Excel (Microsoft Office Excel) — программа для работы с электронными таблицами, созданная корпорацией Microsoft.

Она предоставляет возможности экономико-статистических расчетов



Статистические функции Excel



В мастере функций Excel

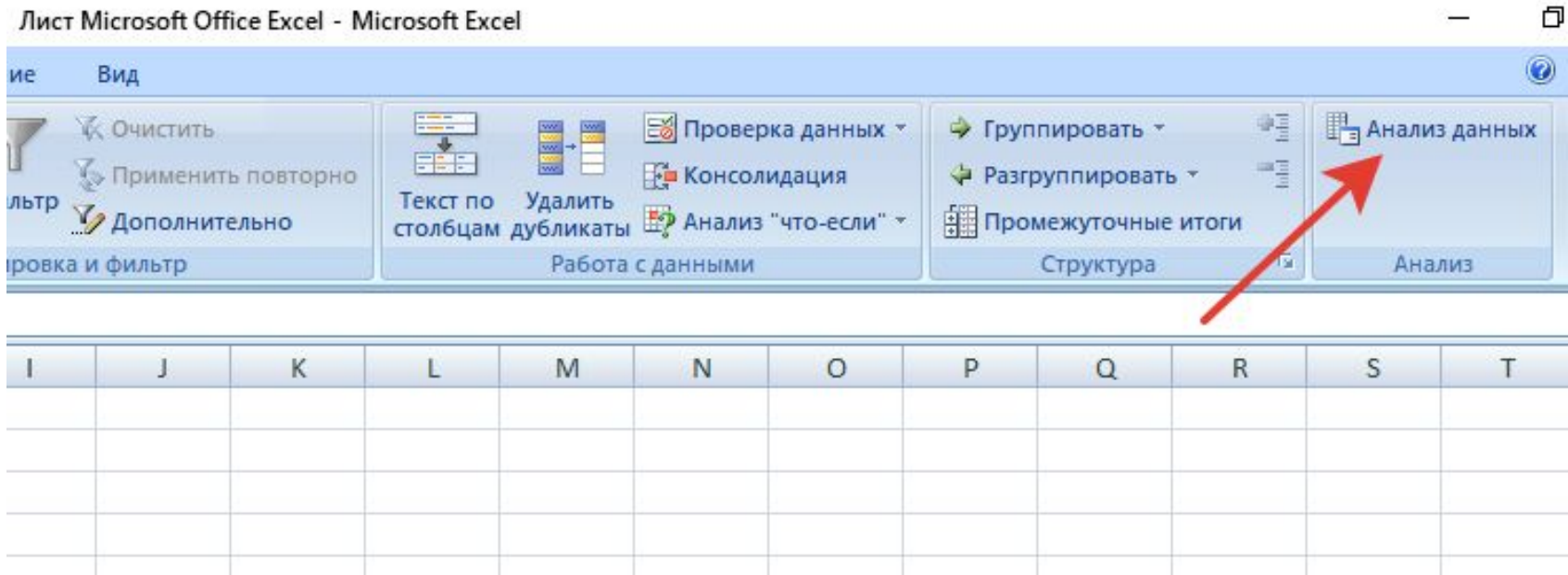
Функция **СРЗНАЧ** .
Функция **МЕДИАНА**
Функция **МОДА**
Функция **ДИСП**
Функция **СТАНДОТКЛОН**
Функция **ЭКСЦЕСС**
Функция **СКОС**
Функция **КВАРТИЛЬ**

1	49		
1	51		
1	49		
1	51		
1	49		
1	51		
1	49		
1	51		
1	49		
491	51		
500	500	сумма	
50	50	среднее	
24010	1,11	дисперсия	
154,95	1,05	станд. отклонение	
1	49	квартили	
1	51	квартили	
1	50	медиана	
1	49	мода	
10	-2,57	эксцесс	
3,16	0	скос(асимметрия)	

Задания для самостоятельной работы

1. Наблюдение посещаемости четырех внеклассных мероприятий в экспериментальном (20 человек) и контрольном (30 человек) классах дали значения (соответственно): 18, 20, 20, 18 и 15, 23, 10, 28. Требуется найти среднее значение, стандартное отклонение, медиану и квартили этих данных.
2. Найти среднее значение, медиану, стандартное отклонение и квартили результатов бега на дистанцию 100 м у группы студентов (с): 12,8; 13,2; 13,0; 12,9; 13,5; 13,1.
3. Определите верхнюю и нижнюю квартиль, выборочную асимметрию и эксцесс для данных измерений роста групп студенток: 164, 160, 157, 166, 162, 160, 161, 159, 160, 163, 170, 171.
4. Найти наиболее популярный туристический маршрут из четырех реализуемых фирмой, если за неделю последовательно были реализованы следующие маршруты: 1, 3, 3, 2, 1, 1, 4, 4, 2, 4, 1, 3, 2, 4, 1, 4, 4, 3, 1, 2, 3, 4, 1, 1, 3

Инструмент «Пакет анализа»



В пакете Excel помимо мастера функций имеется набор более мощных инструментов для работы с несколькими выборками и углубленного анализа данных, называемый **Пакет анализа**, который может быть использован для решения задач статистической обработки выборочных данных



2.2 Statistica



Statistica

- **Statistica** — программный пакет для статистического анализа, разработанный компанией *StatSoft*, реализующий функции анализа данных, управления данными, добычи данных, визуализации данных с привлечением статистических методов.

STATISTICA

- На сегодняшний день разработана 10-я версия пакета (STATISTICA 10).
Русифицирована только 6-я версия (STATISTICA 6).
- Стандартную комплектацию пакета STATISTICA составляют три модуля, которые могут приобретаться как единым пакетом, так и отдельно.
- Базовый пакет STATISTICA Base предоставляет обширные возможности выбора основных типов статистического анализа.

STATISTICA

Пакет предоставляет пользователям следующие возможности статистического анализа данных:

- исследование корреляций между переменными;
- диаграмма рассеяния, матричная диаграмма рассеяния;
- быстрые основные статистики и блоковые статистики (интерактивные средства, позволяющие одним щелчком мыши вычислять основные статистики и строить графики в любой момент в течение сеанса работы);
- интерактивный калькулятор вероятностных распределений (позволяет интерактивно исследовать структуру распределений, например, зависимость вероятности от параметров);
- анализ многомерных откликов, многомерное шкалирование;
- анализ при помощи временных рядов и прогнозирование временных зависимостей, в том числе анализ сезонных колебаний.

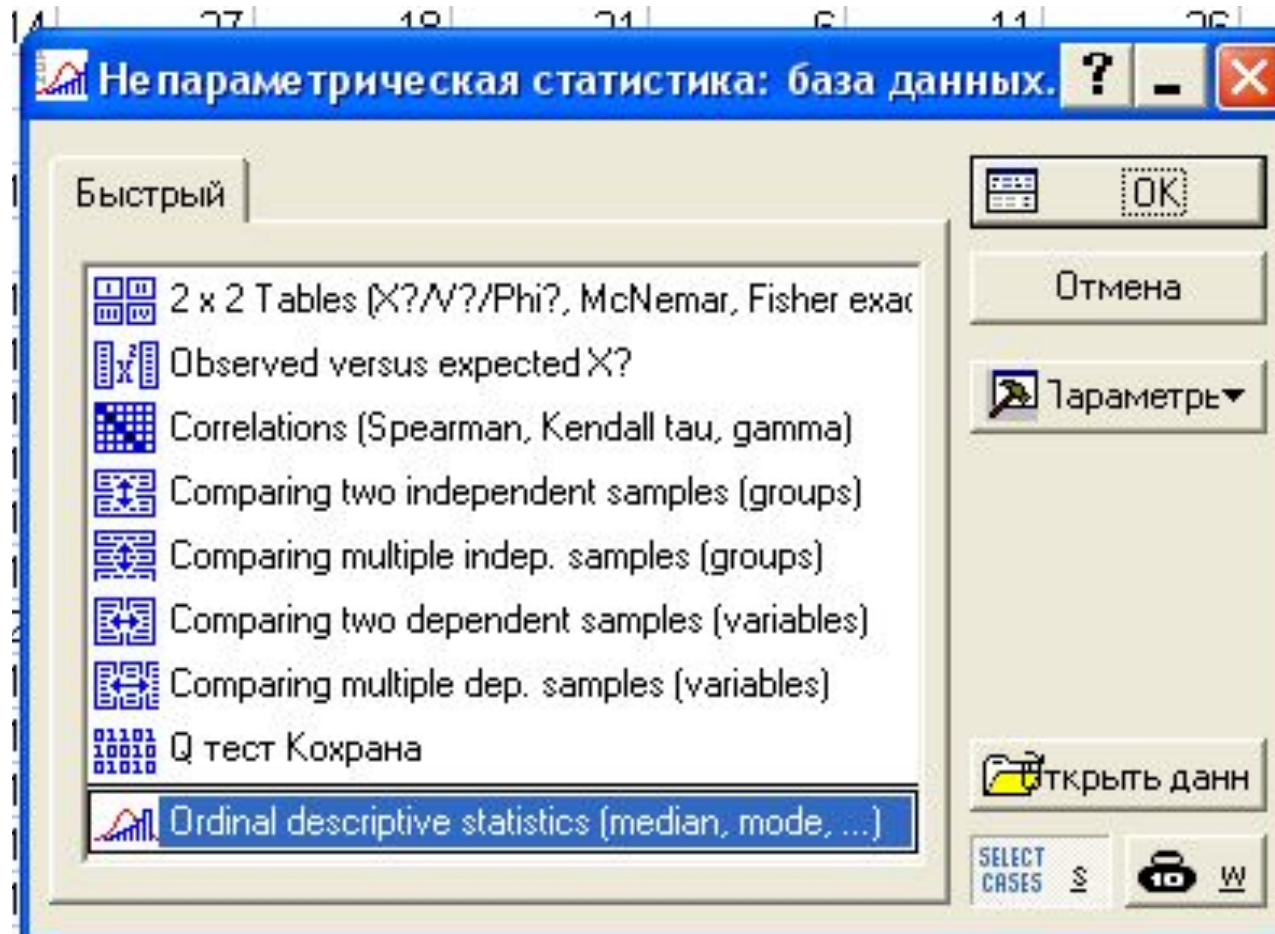
Достоинства STATISTICA

- реализован обмен данными между STATISTICA и Windows-приложениями;
- результаты анализа в виде графиков, таблиц и текста могут быть сохранены в файле с форматом RTF, который открывается и редактируется в MS Word;
- возможность расширения системы при помощи создания программ на встроенном в STATISTICA языке программирования;
- исходные данные из MS Excel можно легко импортировать в STATISTICA;
- возможность записи макросов для автоматизации выполнения однотипных задач;
- программа способна обрабатывать большие массивы данных – базы данных с числом переменных до 32 000 и практически неограниченным числом наблюдений.

STATISTICA



STATISTICA



STATISTICA

The screenshot displays the STATISTICA software interface. On the left, a vertical menu lists various statistical analysis options. The 'Многомерные исследующие методы' (Multivariate Research Methods) option is highlighted in blue. To the right, a secondary menu is open, listing specific multivariate techniques. The background shows a data table with columns labeled '10 пом', '11 манип', '12 позн', '13 извест', '14 работа', and '15 о.ощущ'.

10 пом	11 манип	12 позн	13 извест	14 работа	15 о.ощущ
14	21	15	13	19	26

Menu Options:

- Основная статистика/таблицы
- Множественная регрессия
- ANOVA
- Не параметрический
- Подбор распространения
- Дополнительные линейные/нелинейные модели
- Многомерные исследующие методы**
- Производственная статистика и Six Sigma
- Анализ производительности
- Разработка данных
- Статистика блочных данных
- STATISTICA Visual Basic
- Счетчик вероятности

Multivariate Research Methods Sub-menu:

- Групповой анализ
- Анализ особенности
- Основные компоненты и классификационный анализ
- Канонический анализ
- Надежность/элементный анализ
- Классификационные деревья
- Анализ соотношения
- Многомерное масштабирование
- Дискриминантный анализ
- Общие модели дискриминантного анализа

<http://statsoft.ru/>

ищено | statsoft.ru/products/

[пресс-центр](#) [мероприятия](#) [кон](#)

[Главная](#)

[О компании](#)

[Продукты](#)

[Решения](#)

[Обучение](#)

[Консалтинг](#)

[Ресурсы](#)

[Портал](#)

[ТВ](#)

Поиск

С

Продукты *STATISTICA*

Продукты линейки *STATISTICA* – это мощнейшие инструменты для анализа данных, визуализации, прогнозирования, нейросетевых вычислений, data mining, контроля качества.

Продукты *STATISTICA* полностью переведены на русский язык, включая справочное руководство и документацию!

Преимущества *STATISTICA*

Читать подробнее о [средствах визуализации данных](#), [интеграции *STATISTICA* с Excel и другими приложениями](#), [возможностях автоматизации работы](#), [быстродействию *STATISTICA*](#), [точности вычислений](#).

Есть вопросы?

Специалисты StatSoft всегда на связи.



Линейка

Категории

Пользователи

Приобретение

Продукты серии *STATISTICA* различной комплектации Desktop, Analyst, Modeler, Data Scientist, Visualization Designer включают широкий спектр мощных аналитических инструментов, состоящий из следующих блоков (см. раздел [Новая комплектация](#)):

[Запрос прайс-листа](#)



STATISTICA Base

Широкий набор основных статистик и графических инструментов в понятном интерфейсе со всеми преимуществами, простотой и мощностью технологий *STATISTICA*.



STATISTICA Advanced

Включает все возможности *STATISTICA Base*, а также мощные инструменты для построения линейных/нелинейных моделей – *STATISTICA Advanced Linear/Nonlinear Models*, многомерные технологии анализа данных – *STATISTICA Multivariate Exploratory Techniques*, инструменты для анализа мощности и вычисления объема выборки – *STATISTICA Power Analysis and Interval Estimation*.

Продукты

[Общий обзор](#)

[Новая комплектация](#)

[STATISTICA Base](#)

[STATISTICA Advance](#)

[Промышленная](#)

[STATISTICA](#)

[Нейронные сети](#)

[Data Mining, Text Mining](#)

[Enterprise Solutions](#)

[Управление данными](#) ^{new}

[Специализированные продукты](#)

[Интеграция](#)

[Галерея](#)

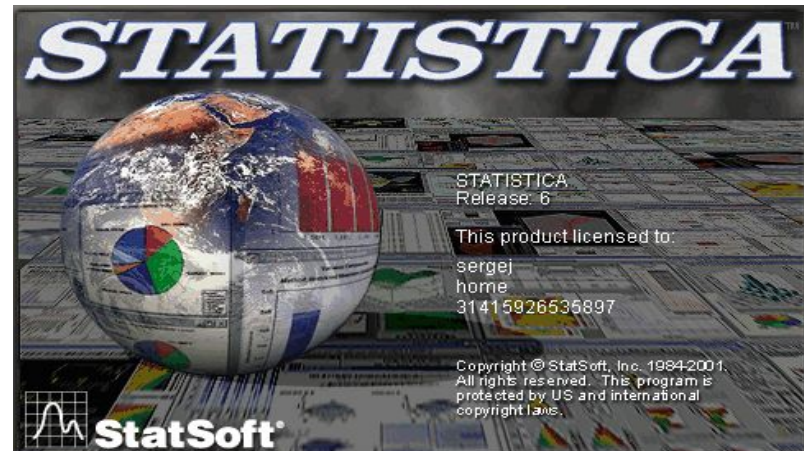
[Trial-версии](#)

[Каталог продуктов](#)

STATISTICA

- Самоучитель по пакету STATISTICA

<https://ieeetusur.ru/books/statistica.pdf>





2.3 SPSS (PSPP)



SPSS

- **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

SPSS

- SPSS является модульной программой. Ее основу составляет базовый *модуль (SPSS Base)*, позволяющий осуществлять *управление данными* и содержащий наиболее распространенные методы статистического анализа данных:
 - проведение *описательной статистики*;
 - построение *линейных и нелинейных моделей*;
 - осуществление преобразования данных;
 - проведение факторного, кластерного, дисперсионного анализов; *вычисление* корреляций;
 - построение графиков; подготовка отчетов и пр.
- Для проведения расширенного и углубленного анализа данных могут быть установлены дополнительные модули пакета. Для пакета *IBM SPSS Statistics 23* разработаны 16 различных модулей

Достоинства SPSS:

- развитый аппарат статистического анализа;
- универсальность (может быть использован для решения широкого круга вопросов из различных предметных областей, требующих проведения статистического анализа данных);
- широкий набор статистических и графических процедур (более 50 типов диаграмм) анализа данных, а также процедур создания отчетов;
- высокая скорость вычислений, простой и удобный интерфейс;
- детальная контекстно-ориентированная справочная система, позволяющая неопытному пользователю с большей легкостью ориентироваться в программе;
- возможность свободного скачивания демонстрационной версии продукта на официальном сайте компании, наличие версий продукта на различных языках;
- совместимость с операционными системами Windows, Mac, Linux;
- наличие значительного количества литературы по работе с пакетом.

Недостатки SPSS:

- высокие требования к системе компьютера (требуется 1GB оперативной памяти, 800MB памяти на жестком диске и процессор с частотой 1GHz и выше);
- высокая цена по сравнению со статистическими пакетами аналогичного уровня (стоимость покупки для индивидуального пользования сроком на год составляет около 1000 долл.⁶).

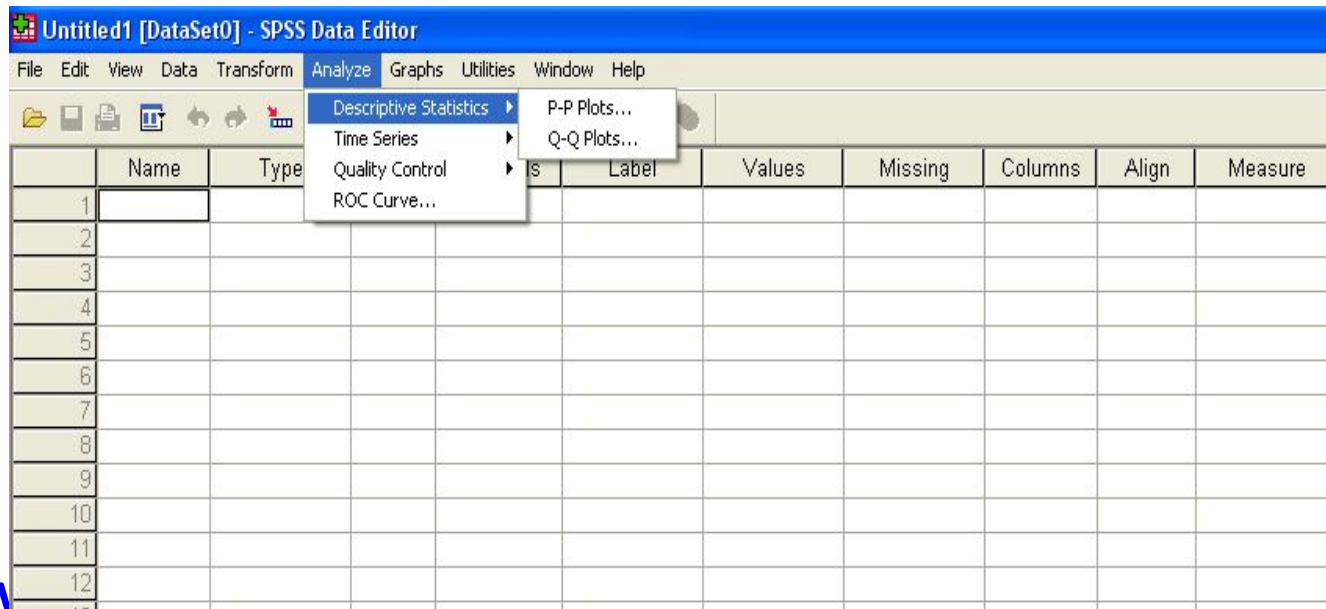
Последняя версия SPSS включает в себя следующие новые возможности

- импорт данных из Excel и SAS;
- экспорт результатов в MS Office, PDF; сохранение результатов в формате HTML;
- одновременная работа с несколькими наборами данных;
- построение диаграммы для переменных с множественными ответами;
- построение диаграммы с двумя осями Y;
- улучшенный редактор синтаксиса с поддержкой *автозавершения* и *цветового кодирования команд*;
- быстрая подготовка данных к анализу посредством Автоматизированной подготовки данных (IBM SPSS Data Preparation), позволяющей облегчить процесс *интеллектуального анализа данных*, выявляя и исправляя ошибки в данных и объясняя пропущенные значения. Также посредством этой функции можно подготовить отчет с рекомендациями о возможности использования данных для анализа.



SPSS

- [Учебное пособие по работе в программе SPSS](#)



- <http://www.rpstat.com.ua/courses/spss/>

[https://www.ibm.com/ru-ru/analytics/ spss-trials](https://www.ibm.com/ru-ru/analytics/spss-trials)



ibm.com/ru-ru/analytics/spss-trials



Аналитика

Список продуктов

Решения

Дополнительно ▾

Пробные версии IBM SPSS

Начните работу с бесплатной пробной версией IBM SPSS уже сегодня. ПО IBM SPSS предлагает простой в использовании комплекс передовых методов, которые помогут вам найти новые возможности, повысить эффективность и снизить риски.

[Попробовать SPSS Statistics бесплатно](#)

[Попробовать SPSS Modeler бесплатно](#)

→ [Есть версии для учебных заведений](#)

PSPP

- **PSPP** — это свободное программное обеспечение для статистического анализа данных. У PSPP имеются как графический пользовательский интерфейс, так и традиционный интерфейс командной строки. Пакет может быть свободной заменой SPSS от IBM.



2.4 *Vortex*

Vortex

- Программа Vortex разрабатывается с 1992 года и применяется для обработки данных социологических, маркетинговых, социально-политических, социально-психологических, социально-медицинских и других исследований.

Программа «VORTEX» предназначена для:

- разработки инструментария сбора данных (анкеты, бланка интервью, теста и т.п.) с возможностью публикации в форматах Microsoft Word, Html и др.;
- ввода первичной информации, собранной в ходе прикладного исследования;
- обработки и анализа этой информации;
- представления полученных результатов анализа в виде таблиц, текстов, графиков и диаграмм с возможностью их переноса в форматы Html, Word, Excel и др.

Применение в социологии

СОЦИОЛОГИЧЕСКИЕ ИССЛЕДОВАНИЯ

Комплекс наиболее востребованных процедур обработки и анализа количественных социологических данных, в том числе:

- расчет объема и определение ошибки репрезентативности для различных видов выборки;
- экстраполяция данных на генеральную совокупность;
- обработка вопросов с множественным выбором;
- обработка открытых и полужакрытых вопросов;
- обработка табличных вопросов;
- обработка ранговых методик, полярных профилей и вопросов с контрольной суммой;
- множество методик разработки вторичных показателей;
- одномерный, двухмерный и многомерный частотный анализ;
- регрессионный, кластерный и другие виды многомерного анализа;
- деревья классификации;
- контент-анализ текстовых данных (в разработке).

www.vortex10.ru

Версия:	Модуль ввода	Базовая	Профессиональная	Полная	Студенческая	Академическая	Для учебных заведений
Ограничения по использованию программы	нет	нет	нет	нет	только для студентов	частным лицам - использование исключительно в учебных и научных целях	использование исключительно в учебных и научных целях
Число установок в комплекте	∞	1	1	1	1	1	30
Документов в обработке	∞	∞	∞	∞	до 300	∞	∞
Формирование структуры переменных	-	✓	✓	✓	✓	✓	✓
Ввод данных	✓	✓	✓	✓	✓	✓	✓
Базовый модуль	-	✓	✓	✓	✓	✓	✓
Профессиональный модуль	-	-	✓	✓	✓	✓	✓
Модуль «Маркетинг»	-	-	-	✓	✓	✓	✓
Модуль «Мониторинг»	-	-	-	✓	✓	✓	✓
Стоимость 1го комплекта, руб.	7 000	14 000	18 500	24 000	800	7 000	40 000
Стоимость 2го и последующих комплектов, руб.	---	14 000	13 000	17 000	800	7 000	40 000
Срок действия лицензии	бессрочная	5 лет	5 лет	5 лет	бессрочная	5 лет	5 лет
Стоимость продления лицензии на 5 лет*	---	14 000	9 250	12 000	---	7 000	20 000

Обучение

<https://www.vortex10.ru/study>

Вас приветствует программа Vortex



Демонстрационная версия 10.23.2.

Выпуск: Апрель 2018 D

Лицензия: Демо-версия

[Обновить программу](#)

[Активировать программу](#)

[Задать вопрос](#)

[Обучение](#)

С чего бы Вы хотели начать работу?

Выбрать ранее загруженное исследование:

C:\Users\Public\Documents\Vortex10\Demo\Выборы в Думу.vit
C:\Users\Public\Documents\Vortex10\Demo\Интересы и ценности.vit
C:\Users\Public\Documents\Vortex10\Demo\Мониторинг УрФО.vit
C:\Users\Public\Documents\Vortex10\Demo\Пиво.vit
C:\Users\Public\Documents\Vortex10\Demo\Полярные профили.vit
C:\Users\Public\Documents\Vortex10\Demo\Пример видео-урока Организация системы переменных.vit
C:\Users\Public\Documents\Vortex10\Demo\Студенты о социальном неравенстве 2016.vit
C:\Users\Public\Documents\Vortex10\Demo\Шаблон CATI.vit

Описание исследования:

В преддверии V Всероссийского социологического Конгресса Российское общество социологов посвящен теме социального неравенства - как Вы его себе представляете, и насколько Российское общество.

Опрошено 3964 студента из 73 вузов 26 городов России. Охвачены все федеральные округ

Таблица №1

Валидные	Значения	Частота	%от опрошенных	Город
1	Екатеринбург	639	16,12	
2	Волгоград	462	11,65	
3	Уфа	368	9,28	
4	Москва	340	8,58	
5	Астрахань	300	7,57	
6	Воронеж	242	6,10	
7	Сургут	242	6,10	
8	Ставрополь	229	5,78	
9	Владивосток	215	5,42	
10	Нижний Новгород	166	4,19	

Переменных: 140

Объектов: 3964

[Выбрать \(искать\) другие исследования](#)

[Создать новое исследование](#)

2.5. Другие пакеты

Другие пакеты статистических программ

- [Matlab](#)
- [Mathematica](#)
- [Mathcad](#)
- [Maple](#)
- **STADIA**
- **STATGRAPHICS PLUS**
- **NCSS**
- **STATA**

Stata

- Пакет Stata – универсальный статистический пакет компании StataCorp⁷. Первая версия пакета была выпущена в 1985 г. На сегодняшний день разработана 11-я версия пакета (Stata 11).
- Пакет Stata позиционируется как инструмент анализа, предназначенный для специалистов, которые занимаются научными исследованиями. По мнению разработчиков, благодаря гибкой модульной структуре пакет применим для анализа данных из различных областей знаний: общественные науки (экономика, политология и пр.), медицина (биостатистика, эпидемиология и пр.) и т. д.

Stata

The screenshot shows the Stata software interface with the following components:

- Menu:** File, Edit, Data, Graphics, Statsbcs, User, Window, Help.
- Review:** Command window showing:

```
1 use "c:\program files\stata11\auto"  
2 tab rep78 foreign, chi2
```
- Обзор (Overview):** A grey arrow pointing to the right.
- Results:** Command window output:

```
2. (/v# option or -set maxvar-) 5000 maximum variables  
running C:\Program Files\stata11\profile.do ...  
. use "c:\program files\stata11\auto"  
(1978 Automobile Data)  
. tab rep78 foreign, chi2
```

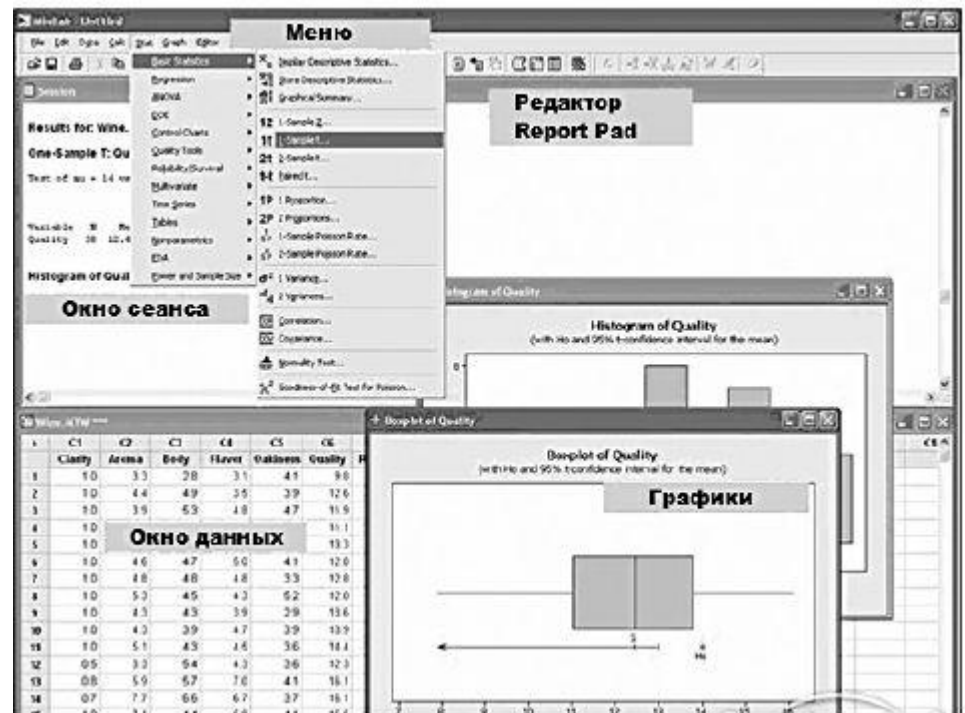
Repair Record 1978	Car type		Total
	Domestic	Foreign	
1	2	0	2
2	8	0	8
3	27	3	30
4	9	9	18
5	2	9	11
Total	48	21	69

pearson chi2(4) = 27.2640 Pr = 0.000
- Переменные (Variables):** A grey arrow pointing to the right.
- Variables List:**

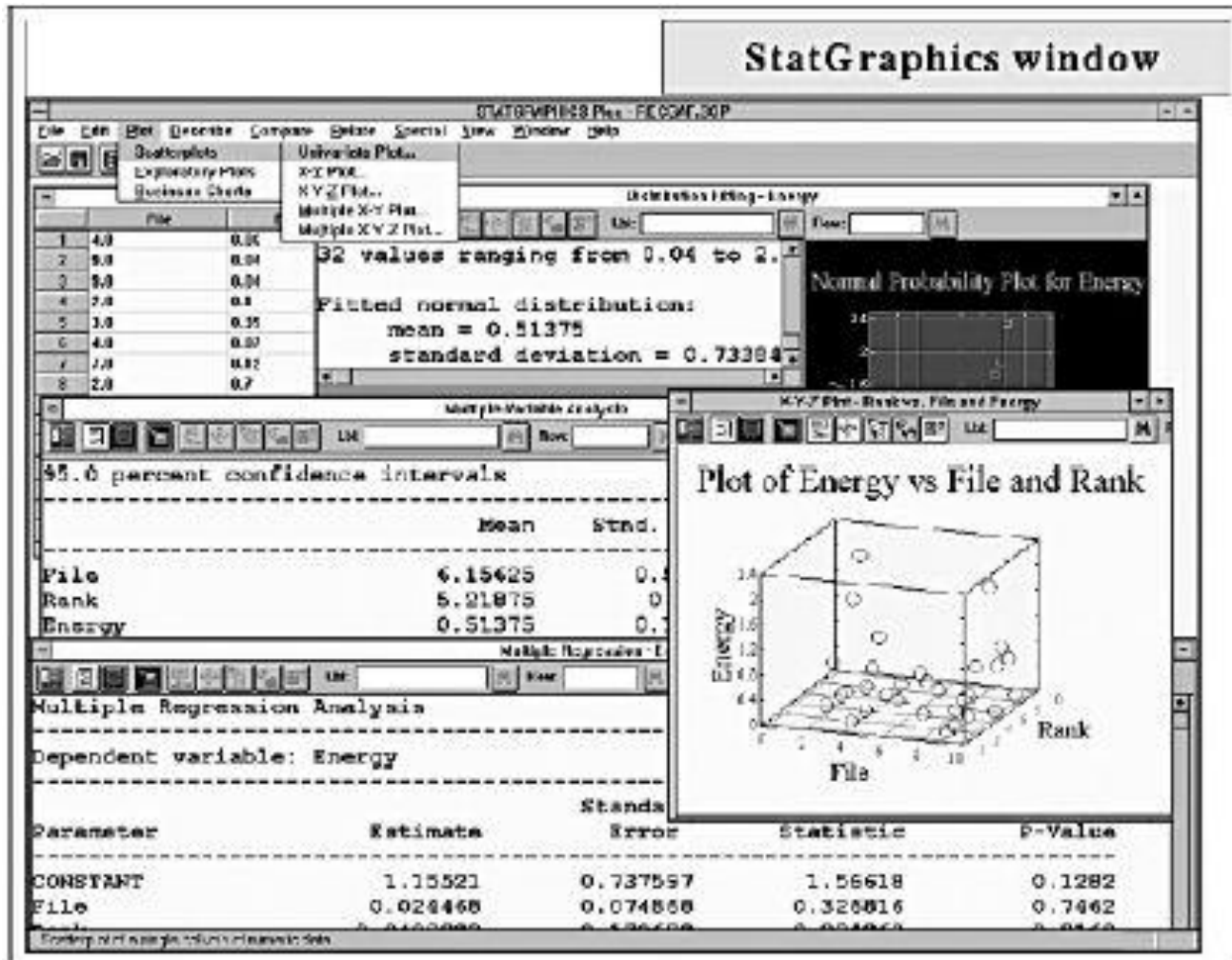
Name	Label	Type
make	Make and Model	str 18
price	Price	int
mpg	Mileage (mpg)	int
rep78	Repair Record 1978	int
headroom	Headroom (in.)	float
trunk	Trunk space (cu. ft.)	int
weight	Weight (lbs.)	int
length	Length (in.)	int
- Команды (Commands):** A grey arrow pointing to the right.
- Command Window:** Shows the command `by foreign: sum mpg`.
- Status Bar:** z:\probs\stata CAP NUM OVR

Minitab

- Последняя версия Minitab 16 позиционируется разработчиками как статистический пакет для аналитической работы на современных предприятиях.



STATGRAPHICS



STADIA

The screenshot displays the STADIA-6.0 software interface. At the top, the title bar reads "STADIA-6.0" and the menu bar includes "Файл", "График=F6", "Вычисл=F7", "Преобр=F8", "Статист=F9", "Окна", and "Помощь=F1". Below the menu bar is a toolbar with icons for file operations and a search field. The main window title is "Таблица данных: tst.std".

The data table contains the following information:

	sample	lc1	lc2	f1	f2	PolyrX	PolirY	ExprY
1	1.9	20.1	7.2	20	22.1	1	5	9
2	2.5	23						7
3	3.4	26						5
4	4.1	19						4
5	5.8	16						3
6	6.6	23						5
7	7.2	31						2
8	8	33						5
9	9.199	28						1
10	0.1							
11	2.4	29						
12	3.5	30						
13	4.8							
14	5.6							
15	6.3							
16	7.3							
17	8.1							
18	1.4							
19	3							
20	1.2							

A dialog box titled "Графики данных" (Data Charts) is open, showing various chart options:

- 1=функциональный
- 2=с отклонениями
- 3=диагр. рассеяния
- 4=распределение
- 5=ящики-с-усами
- 6=столб. диаграмма
- 7=диаграмм-башня
- 8=100%-я башня
- 9=круг. диаграмма
- столбики 3D-эффект
- A=диагр. рассеяния
- B=поверхность
- C=сглаживание
- D=картирование
- E=интерполяция
- G=сглаживание

Additional settings in the dialog include "Шагов= 50" and "Коеф= 0.5". A button at the bottom right reads "X <Esc>=Отменить".

At the bottom of the window, a status bar contains the text "Одна или несколько экспериментальных зависимостей Y=f(X)" and a keyboard shortcut "D&/Rez".

Домашнее задание

1. Скачать установочные программы (PSPP) или демо-версии (SPSS / Vortex)
2. Найти в собственных архивах 1 – 2 базы числовых данных.

Благодарю за внимание!