

УДК 311
ББК 65.051
В 64

Авторы-составители: Н. В. Лацкевич, канд. экон. наук, доцент;
Д. А. Степаненко, ст. преподаватель

Рецензенты: М. А. Грибовская, канд. физ.-мат. наук, доцент
Белорусского торгово-экономического университета
потребительской кооперации;
С. А. Дешеня, канд. экон. наук, доцент Белорусского
торгово-экономического университета потребительской
кооперации

Рекомендовано к изданию научно-методическим советом учреждения образования «Белорусский торгово-экономический университет потребительской кооперации». Протокол № 5 от 11 июня 2013 г.

Возможности системы Statistica 6.0 : пособие для реализации со-
В 64 держания образовательных программ высшего образования I и II ступеней / авт.-сост. : Н. В. Лацкевич, Д. А. Степаненко. – Гомель : учреждение образования «Белорусский торгово-экономический университет потребительской кооперации», 2014. – 52 с.
ISBN 978-985-540-141-5

В пособии описаны возможности и структура системы Statistica 6.0, особенности ее интерфейса. Приведены документы и краткий обзор методов системы, представлены пошаговые примеры проведения кластеризации, факторного и канонического анализа. Пособие является методическим руководством при научных исследованиях. Предназначено для студентов и магистрантов экономических специальностей.

УДК 311
ББК 65.051

ISBN 978-985-540-141-5

© Учреждение образования «Белорусский торгово-экономический университет потребительской кооперации», 2014

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

Прикладные статистические методы широко используются в практической деятельности работников самых различных сфер. Знание основ статистических методов необходимо специалистам, работающим в естественнонаучных и технических областях, а также представителям гуманитарных профессий: экономистам, социологам, психологам, лингвистам.

Возможности компьютеров в обработке громадных объемов информации сделали доступными для широкого пользователя самые современные методы статистического анализа. В настоящее время разработано большое количество статистических пакетов программ, представляющих удобную современную форму программного обеспечения.

Применение статистических пакетов упрощает использование статистических методов, однако при этом необходимо не только собрать и правильно ввести данные, выбрать тот или иной способ их обработки, но и понимать основные идеи статистических методов и, что особенно важно, теоретическое обоснование этих методов. При осмысленном применении статистических методов невозможны курьезы и абсурдные выводы, которыми так богата история статистики.

Предлагаемое пособие по системе Statistica 6.0 написано на основе данных фирмы Statsoft Russia, а также различной учебно-методической литературы по данному программному продукту.

Основная цель его состоит в том, чтобы на простых примерах показать, как можно использовать статистический пакет системы Statistica 6.0 для выполнения статистических расчетов, проведения анализа результатов деятельности субъекта хозяйствования и формулирования выводов.

Данное пособие является дополнительным методическим материалом в рамках изучения дисциплин «АРМ экономиста» и «Экономика организации (предприятия)». Оно может представлять интерес для студентов и магистрантов экономических специальностей при выполнении научных, курсовых и дипломных проектов.

1. ВОЗМОЖНОСТИ СИСТЕМЫ STATISTICA 6.0 И ОСОБЕННОСТИ ЕЕ ИНТЕРФЕЙСА

Универсальная статистическая система Statistica разработана и производится фирмой StatSoft, Inc. (США). Последние версии системы полностью совместимы со средой Windows. В Беларуси программный продукт распространяется через представителей фирмы StatSoft Russia. Подробную информацию о пакете можно получить на официальном сайте компании. Работа с системой подробно описана в ряде книг и электронных источниках [1–3].

Система Statistica – это универсальная интегрированная система, предназначенная для статистического анализа и обработки данных. Программа поддерживает большое количество методов обмена данными из других приложений, которые позволяют импортировать их практически из любой базы.

Одной из широко распространенных версий является система Statistica 6.0. Сфера применения системы достаточно обширна (рисунок 1).

Интерфейс Statistica 6.0 является интуитивно понятным, легко настраивается в соответствии с пользовательскими задачами и аналогичен интерфейсу стандартных приложений Windows, поэтому легко осваивается пользователем.

Анализ данных проводится интерактивно, в режиме последовательно открывающихся диалоговых окон. Любое окно анализа сконструировано таким образом, что на первой вкладке содержатся только самые необходимые кнопки, а на последующих вкладках – углубленные методы и специальные опции.

Система Statistica 6.0 включает графический модуль, содержащий удобные инструменты для эффективной и наглядной визуализации данных, проведения графического анализа; более 10 000 различных типов графиков, имеющих богатые возможности редактирования, интерактивной настройки (вращение, масштабирование, прозрачность и др.).

Одним из важных свойств программных продуктов Statistica 6.0 является их быстрое действие при работе с большим объемом данных и вычислительная мощность приложений, требующих регулярного построения запросов к базам данных, комплексного управления данными.

Язык программирования системы Statistica 6.0 *Visual Basic* позволяет создавать дополнительные приложения, реализующие как простые преобразования данных, так и сложные вычислительные процедуры. В системе Statistica 6.0 можно записать макросы, которые автоматизируют повторяющиеся шаги, также используется подмножество языка запросов *SQL (Structured Query Language – структурированный язык запросов)* для задания критериев импорта записей из баз данных.



Рисунок 1 – Сферы применения системы Statistica 6.0

Система Statistica 6.0 полностью удовлетворяет основным стандартам Windows, включая OLE DB и DDE. В системе реализованы мощные инструменты построения запросов, с помощью которых может быть осуществлен быстрый доступ к любым базам данным (Oracle, MS SQL Server, Informix, Access и др.).

При входе в систему Statistica 6.0 пользователя встречает окно приглашения (рисунок 2).

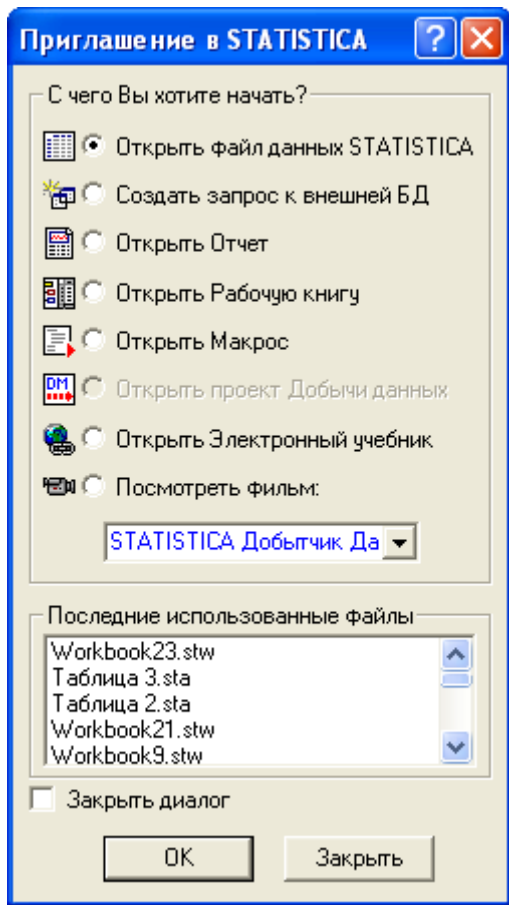


Рисунок 2 – Окно *Приглашение в Statistica 6.0*

В системе Statistica 6.0 предлагается несколько вариантов индивидуального режима работы (пользовательского интерфейса):

- интерактивный интерфейс;
- язык *SVB*;
- Web-интерфейс.

При этом нужно иметь в виду следующее:

- во многих отношениях эти варианты пользовательского интерфейса не исключают друг друга, поэтому их можно использовать в различных комбинациях в зависимости от конкретной ситуации и индивидуальных потребностей;

- настраиваемые меню и панели инструментов могут использоваться для совмещения разных вариантов интерфейса, например, для быстрого доступа к макросам или наиболее часто используемым файлам;

- почти все характеристики этих интерфейсов определяются соответствующими настройками по умолчанию, которые можно изменять по желанию пользователя (преобразовывая тем самым внешний вид и поведение программы); как правило, рекомендуется настроить систему так, чтобы полностью использовать возможности пакета Statistica с учетом своих предпочтений и специфики решаемых задач.

Даже без предварительной настройки конфигурация системы Statistica 6.0, предлагаемая по умолчанию, дает пользователю возможность получать нужные ему результаты в рамках различных вариантов интерфейса. Благодаря принципу альтернативного доступа, последовательно и всесторонне реализованному в пользовательском интерфейсе пакета, система Statistica 6.0 обладает возможностью поддерживать различные стили работы. Так, например, доступ к наиболее часто используемым средствам осуществляется любым из следующих способов:

- через обычные выпадающие меню;
- через последовательности «горячих» клавиш;
- с помощью панелей инструментов («линеек») и активных полей строки состояния;

- через пользовательские панели инструментов (всплывающие или зафиксированные панели инструментов с кнопками, которые определяет сам пользователь и которые могут соответствовать пользовательским процедурам или последовательностям клавиш);

- через контекстные меню, связанные с конкретными объектами (ячейками, значками рабочих книг, графическими объектами), которые вызываются нажатием правой кнопки мыши на данном объекте.

Прежде чем выбрать какой-то один определенный стиль, имеет смысл попробовать различные варианты пользовательского интерфейса системы Statistica 6.0, а затем выбрать наиболее оптимальный.

Интерактивный пользовательский интерфейс в системе Statistica 6.0 используется чаще всего (рисунок 3).

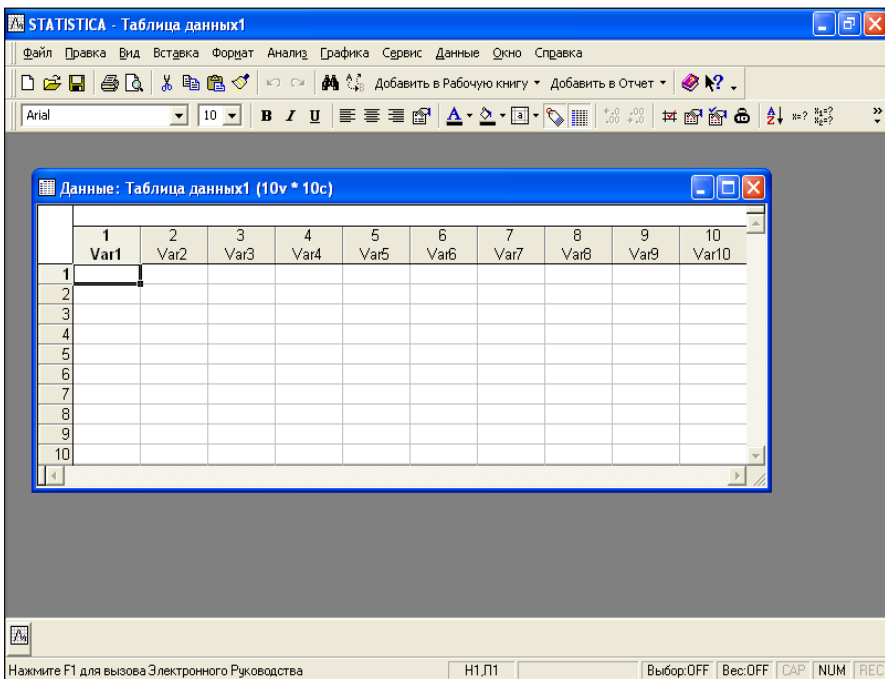


Рисунок 3 – Рабочее окно интерактивного пользовательского интерфейса программы Statistica 6.0

Характеристиками данного интерфейса являются следующие:

- как и во многих популярных программах, строки меню и различные панели инструментов отображаются в верхней части экрана (они могут быть настроены под пользователя);
- в нижней части экрана слева располагаются панель *Анализ* (содержащая свернутые диалоговые окна анализов и графиков) и строка состояния (также имеются контекстные меню, доступные с помощью правой кнопки мыши);
- файлы могут быть отображены в электронных таблицах, рабочих книгах, отчетах или отдельных окнах.

Система Statistica 6.0 предоставляет широкие возможности настройки пользовательского интерфейса, позволяя приспособивать стандартный пользовательский интерфейс для решения предметных задач. Например, можно добавлять обычные опции либо их изменять.

2. КРАТКАЯ ХАРАКТЕРИСТИКА СТРУКТУРЫ ПРОГРАММЫ STATISTICA 6.0

Система Statistica 6.0 состоит из более чем 10 000 вычисляемых процедур, сгруппированных в отдельные модули по основным направлениям анализа. В каждом модуле можно применять определенный способ обработки данных, не обращаясь к процедурам из других модулей. При этом можно легко переключаться из одного модуля в другой при выполнении сложных исследований.

Архитектура системы Statistica 6.0 и реализованный в ней аппарат многозадачности позволяют не только без каких-либо усилий сравнивать результаты анализа, полученные разными процедурами, но и одновременно выполнять их в разных окнах приложения.

Быстро переключаться с одного модуля на другой можно:

- щелкая кнопкой мыши по их значкам на рабочем столе;
- активируя соответствующее окно приложения (если оно уже было открыто);
- выбирая окна в диалоговом окне *Переключатель модулей* (эту операцию можно настроить так, чтобы было удобно обращаться к модулям, которые используются чаще всего).

В целом систему Statistica 6.0 с позиции анализа можно представить в следующем виде:

1. *Базовые модули*, которые обеспечивают обширный вывод основных статистик в сочетании с мощностью, производительностью и простотой использования технологии Statistica 6.0 и позволяют осуществить первичный анализ данных и подготовить аналитический отчет. Раздел включает все графические инструменты системы Statistica 6.0, а также следующее:

- описательные статистики;
- парные и частные корреляции;
- t -критерии для зависимых и независимых выборок, независимых переменных;
- одновыборочный t -критерий;
- группировку и однофакторный дисперсионный анализ;
- таблицы частот, сопряженности, флагов и заголовков, многомерных откликов;
- вероятностный калькулятор;
- множественную регрессию;
- дисперсионный анализ;
- непараметрические статистики;
- подгонку распределений;

- блоковые статистики.

2. Модуль *Углубленные методы анализа* системы Statistica 6.0, представляющий собой большой набор самых современных инструментов для моделирования и прогнозирования и включающий возможность автоматического выбора оптимальных моделей и расширенные интерактивные средства визуализации. Позволяет строить сложные объяснительные модели зависимостей и прогноза с помощью классических моделей временных рядов.

Все анализы снабжены разнообразными интерактивными графиками и встроенным редактором Visual Basic. Данный модуль включает следующее:

- общие линейные модели;
- обобщенные линейные и нелинейные модели;
- общие регрессионные модели;
- общие модели частных наименьших квадратов;
- компоненты дисперсии;
- анализ выживаемости;
- нелинейное оценивание;
- множественную нелинейную регрессию;
- логлинейный анализ таблиц частот;
- временные ряды и прогнозирование;
- моделирование структурными уравнениями.

3. Модуль *Многомерный разведочный анализ* системы Statistica 6.0, предоставляющий широкий выбор разведочных технологий, начиная с кластерного анализа до расширенных методов классификационных деревьев, в сочетании с огромным набором средств интерактивной визуализации для построения моделей. В состав модуля входят:

- кластерный анализ;
- факторный анализ;
- анализ главных компонент и классификация;
- канонический анализ;
- надежность и позиционный анализ;
- деревья классификации;
- анализ соответствий;
- многомерное шкалирование;
- дискриминантный анализ;
- общие модели дискриминантного анализа.

4. Модуль *Промышленная статистика и Шесть сигма* системы Statistica 6.0, который предусматривает следующее:

- карты контроля – полностью настраиваемые (вызываемые из других приложений), простые, доступные и гибкие карты с набором

опций автоматизации и возможностью создания пользовательских ярлыков для упрощения ежедневных задач;

- анализ процессов – обширный пакет для анализа пригодности процесса, анализа повторяемости и воспроизводимости измерений, и запуска других процедур контроля и улучшения качества;

- планирование экспериментов (ПЭ) – самый большой выбор методов планирования экспериментов и соответствующих технологий визуализации, включая интерактивные профили желательности и широкий спектр статистик остатков и др.

При этом процедуры данного модуля также позволяют осуществлять процедуры базового модуля.

5. Модуль *Анализ мощности, оценка объема выборки* системы Statistica 6.0, являющийся чрезвычайно точным и удобным в использовании инструментом исследования, предназначенным для анализа всех аспектов статистической мощности и расчета объема выборки.

6. Модуль *Нейронные сети* системы Statistica 6.0 и др.

3. ОТДЕЛЬНЫЕ ДОКУМЕНТЫ ПРОГРАММЫ STATISTICA 6.0

3.1. Рабочие книги

Рабочие книги являются стандартным средством управления выводом. В рабочей книге каждый документ представлен в виде вкладки. Говоря техническим языком, рабочие книги Statistica 6.0 являются оптимизированными *ActiveX*-контейнерами, которые позволяют эффективно обрабатывать большое количество документов.






Рабочая книга – это документ системы Statistica 6.0 (файл с расширением, который объединяет разного вида информацию, относящуюся к одним и тем же данным). Так, в рабочей книге могут размещаться таблица данных, таблицы результатов анализов, графики, отчеты, программы пользователя, которые были созданы или использованы во время анализа данных. Управлять содержимым рабочей книги и размещением результатов анализа можно на вкладке *Диспетчер вывода* окна параметров системы, которое вызывается командой *Сервис/Параметры*.

Рабочую книгу можно создать в отдельном окне с помощью команды *Файл/Создать*, активизировав вкладку *Рабочая книга* (если установить флажок *Поместить пустую таблицу*, то рабочая книга

будет содержать таблицу из 10 столбцов и 10 строк по умолчанию). Новая рабочая книга имеет порядковый номер, который зависит от номера создаваемой книги при одном сеансе работы с системой Statistica 6.0 (например, *Workbook 1*). При сохранении рабочей книги с помощью команды *Файл/Сохранить как* можно изменить ее имя.

Каждая рабочая книга содержит два поля: навигационное дерево в стиле Explorer в левой части и поле просмотра файлов в правой части. В навигационном дереве может находиться несколько различных узлов, которые используются для создания логической структуры. Вкладки в нижней части поля просмотра документов используются для удобства переходов между документами текущего узла. Преимуществом вертикального расположения вкладок является то, что при этом они отображаются в виде нескольких строк, а не одной длинной строки, что облегчает выбор нужной вкладки. Эти вкладки также можно скрыть, увеличив тем самым поле просмотра рабочей книги.

Документы можно организовать в виде иерархии папок, или узлов, документов (по умолчанию для каждого нового анализа создается своя папка). С использованием дерева просмотра можно легко управлять структурой, изменяя положение отдельных документов, папок или целых ветвей дерева, используя значки «+» и «-» возле каждой папки, которые позволяют развернуть или свернуть ее содержание, либо буфер обмена. При этом при настройке иерархии папок можно различными способами перенести в другие рабочие книги целые ветви дерева.

Каждому элементу дерева рабочей книги соответствует определенный значок. Значок  обозначает папку, которая может содержать ряд документов и мелких папок. Значок  представляет собой электронную таблицу,  – отчет,  – макрос,  – график.

Документы, не являющиеся документами системы Statistica 6.0, обозначаются соответствующими по типу значками.

3.2. Электронные таблицы

Таблица данных – это документ системы Statistica 6.0 (файл с расширением *.sta), в котором содержатся в табличной форме исходные данные. Таблицу можно создать с помощью команды *Файл/Создать* (вкладка *Таблица*).

Электронная таблица состоит из строк и столбцов. Столбцы таблицы называются *Переменные*, а строки – *Наблюдения* (рисунок 4).

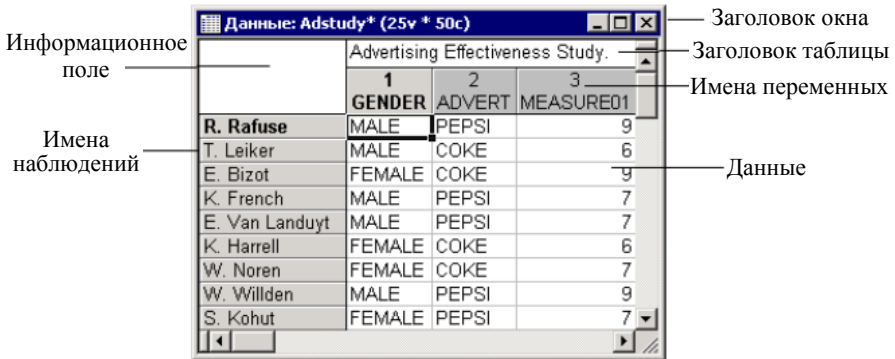


Рисунок 4 – Окно электронной таблицы системы Statistica 6.0

Например, в медицине наблюдения – это пациенты, переменные – пол, возраст, дата поступления в больницу, дата диагноза, даты операции, перевода в другую больницу, выписки и т. д. Можно также считать, что столбцы (переменные) соответствуют различным случайным величинам, а строки (наблюдения) – выборочным значениям этих величин.

При создании таблицы также указывается префикс переменных как универсальная часть названия всех переменных (по умолчанию *Var*). Имя переменной можно затем изменить в созданной таблице на более информативное, вызвав окно спецификации переменной (например, двойным щелчком по имени переменной).

Таким образом, стандартная таблица – это двумерная таблица, которая может содержать практически неограниченное число наблюдений (строк) и переменных (столбцов). При этом каждая ячейка может включать практически неограниченное число символов.

Формат отображения и тип данных можно оставить по умолчанию, так как их можно будет уточнить в процессе редактирования свойств каждой переменной.


Так как электронные таблицы системы Statistica могут содержать макросы и поддерживают настраиваемый пользовательский интерфейс, то эти мультимедийные таблицы можно использовать в качестве оболочки различных приложений, презентаций, анимаций и т. д. Например, можно добавить окно списка опций или ряд кнопок в верхнем левом углу рабочего окна через команды *Настройка/Панели инструментов (Команды/Макросы)*.

Окно электронной таблицы содержит несколько основных частей.


Данные в системе Statistica 6.0 представлены в виде набора наблюдений и переменных. Каждое наблюдение состоит из набора значений переменных. Обычно в первом столбце файла содержатся имена наблюдений.



В заголовке окна отображается название электронной таблицы. Если электронная таблица является исходной таблицей данных, то заголовков окна также содержит число переменных и наблюдений.


Для того, чтобы выделить всю электронную таблицу, необходимо щелкнуть кнопкой мыши в углу *Информационного поля*, которое располагается в верхнем левом углу окна электронной таблицы (при этом указатель должен иметь вид обычной стрелки).


Для выделения *Информационного поля* (например, с целью форматирования) один раз нужно щелкнуть кнопкой мыши в верхнем левом углу *Информационного поля* (указатель мыши должен быть в виде знака )

Поле *Заголовков таблицы* расположено непосредственно над именами переменных в верхней части окна. Чтобы выделить это поле, необходимо щелкнуть по нему один раз. Чтобы ввести или отредактировать текстовую информацию, следует дважды щелкнуть по этому полю и нажать клавиши *Ctrl-Enter* или *Alt-Enter* для ввода новой строки.

Для ввода или редактирования текста в ячейке *Имя наблюдения* необходимо дважды щелкнуть по любой позиции. Для того, чтобы выделить всю строку наблюдения (для редактирования), нужно один раз щелкнуть по средней или правой части ячейки *Имя наблюдения* (указатель мыши в этом случае имеет вид )

Для просмотра спецификаций отдельной переменной нужно дважды щелкнуть по полю *Имя переменной*. Для выделения поля *Имя переменной* (например, форматирования) один раз нужно щелкнуть по верхней части этого поля (указатель мыши должен при этом иметь вид  вместо )

Для выделения всего столбца переменных (например, с целью редактирования) нужно один раз щелкнуть по нижней части поля *Имя переменной* (указатель мыши в этом случае имеет вид )

Чтобы автоматически установить подходящую ширину строки или столбца, дважды необходимо щелкнуть по правой границе (указатель мыши в этом случае принимает вид )

Остальная область электронной таблицы содержит данные. В ячейки могут быть также вставлены звук, видео, графики, анимация, отчеты с внедренными объектами или любые документы, использующие технологию *ActiveX*. При работе с данными часто приходится иметь

дело с ситуациями, когда часть данных не была по каким-либо причинам измерена. В этом случае ячейка остается пустой и ей присписывается некоторый специальный код – код пропущенных данных (*Код ПД*), значение которого по умолчанию равно –9999 (обычно в настройках системы Statistica 6.0 длина ограничена 1000 символами для предотвращения случайной вставки большого объема нежелательной информации). При этом имеются широкие возможности форматирования текста в ячейках.

3.3. Отчеты

Отчеты в системе Statistica 6.0 предлагают более традиционный способ представления результатов, чем рабочие книги. Каждый объект отображается последовательно в документе текстового редактора. Тем не менее, технология создания отчетов обладает широкими функциональными возможностями. Например, как и рабочая книга, каждый отчет системы Statistica 6.0 является *ActiveX*-контейнером. Это означает, что они могут содержать в себе ряд совместимых документов. Каждый объект в нем является настраиваемым и допускает редактирование непосредственно в окне *Отчет*.

Таким образом, отчет – это документ системы Statistica 6.0 (файл с расширением **.str* или **.rtf*), в котором сохраняются результаты статистического анализа. Отчет может содержать любую текстовую, числовую и графическую информацию. Отчет можно создать с помощью команды *Файл/Создать*, активизировав вкладку *Отчет*. Кроме того, отчет может создаваться автоматически при проведении процедур статистического анализа. Документы с расширением **.rtf* информацию о дереве объектов не сохраняют.

Созданные таблица данных и отчет могут быть размещены в отдельных файлах или включены в одну рабочую книгу.

Очевидными преимуществами такой организации вывода отчета являются возможности добавления примечаний и комментариев к объектам, а также быстрой прокрутки и просмотра результатов вывода в таком виде, к какому привыкли пользователи.

Кроме того, отчеты содержат и сохраняют дополнительную информацию, включающую в себя подробный журнал заданных для анализа опций. Объем дополнительной информации указывается во вкладке *Диспетчер вывода* диалогового окна *Параметры*.

Очевидным недостатком традиционных отчетов является их одноуровневая структура, хотя для некоторых пользователей или приложений такая структура может оказаться предпочтительной.

В контекстном меню дерева отчета (вызываемом правой кнопкой мыши в любом месте дерева) доступны команды вставки, извлечения, переименования и удаления элементов дерева отчета.

3.4. Графики

Система Statistica 6.0 предоставляет широкий спектр возможностей по созданию графических документов, а также их настройке. Как и другие документы системы Statistica 6.0, графики являются *ActiveX*-контейнерами. Существуют два основных типа графиков:

- графики исходных данных;
- графики блоковых данных.

Они могут быть вызваны из меню *Графика* (рисунок 5), панели инструментов *Графика*, контекстных меню и меню стартовой кнопки системы Statistica 6.0. Основные типы графиков используют одинаковые опции для настроек и одинаковые виды графиков.

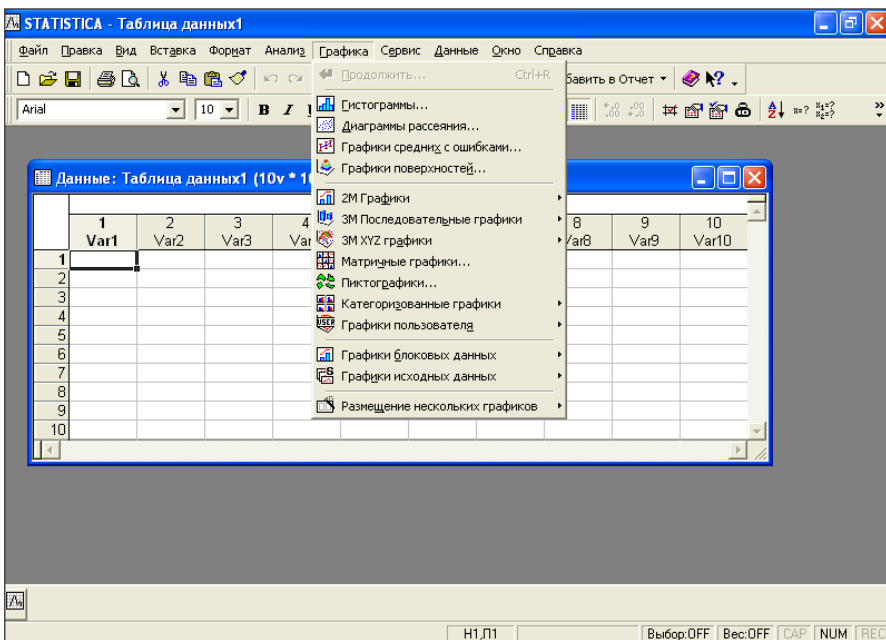


Рисунок 5 – Меню *Графика* в системе Statistica 6.0

Графики исходных данных включают первичные данные из текущей электронной таблицы данных. Обычно на них представляются переменные или подмножества их значений, если используются условия выбора наблюдений.

Графики блоковых данных совершенно не зависят от первичных данных или файла данных. Они предоставляют возможность визуализации числовых значений выделенного в данный момент блока электронной таблицы. Выделенный блок может содержать значения как определенных пользователей подмножеств численных результатов, так и значений произвольно выбранных подмножеств исходных данных.

4. КРАТКИЙ ОБЗОР МЕТОДОВ СИСТЕМЫ STATISTICA 6.0

4.1. Базовые модули

4.1.1. Основные статистики и таблицы

Базовый модуль системы Statistica 6.0 предлагает широкий набор методов для статистического анализа.

Процедура *Основные статистики и таблицы* представлена на рисунке 6.

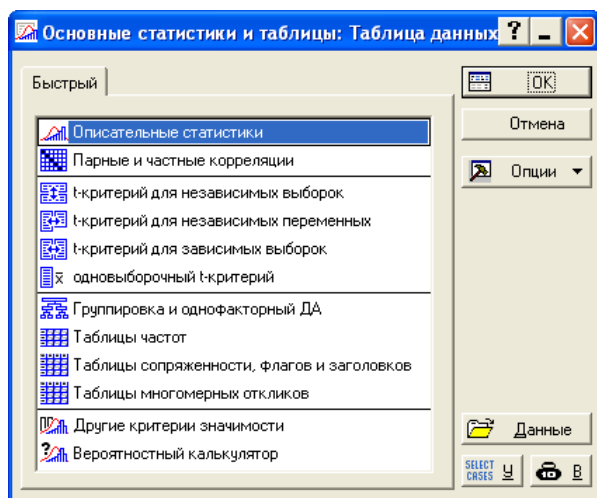


Рисунок 6 – Окно *Таблица данных* процедуры *Основные статистики и таблицы* системы Statistica 6.0

Система вычисляет практически все используемые описательные статистики общего характера: медиану, моду, квартили, заданные пользователем процентиля, среднее значение и стандартное отклонение, квартильный размах, доверительные интервалы для среднего значения, асимметрию и эксцесс (и их стандартные ошибки), гармоническое и геометрическое среднее, а также многие специальные описательные статистики.

Как и во всех других модулях системы Statistica 6.0, в анализе данных доступны разнообразные графики и диаграммы, в том числе различные виды диаграмм размаха и гистограмм, гистограммы двумерных распределений (трехмерные и категоризованные), двух- и трехмерные диаграммы рассеяния с помеченными подмножествами данных, нормальные и полунормальные вероятностные графики и графики с исключенным трендом, графики «квантиль-квантиль», «вероятность-вероятность» и т. д.

Имеется набор критериев для подгонки нормального распределения к данным (критерии Колмогорова – Смирнова, Лилиефорса и Шапиро – Уилкса). Процедуры для подгонки многих других типов распределений можно найти в разделах *Анализ процессов Statistica 6.0* и *Графики*.

Группировка. Практически все описательные статистики и графики могут быть построены для данных, категоризованных (сгруппированных) по значениям одной или нескольких группирующих переменных.

Например, с помощью нескольких щелчков мыши можно сгруппировать имеющиеся данные о людях по полу и возрасту, а затем просмотреть категоризованные гистограммы, диаграммы размаха, нормальные вероятностные графики, диаграммы рассеяния и т. д. В случае, если было выбрано более двух категоризованных переменных, автоматически будет построен каскад соответствующих графиков.

Имеется возможность производить категоризацию по числовым (непрерывным) переменным, например, можно потребовать, чтобы значения переменной были разбиты на заданное число интервалов; с помощью инструмента перекодировки можно в реальном времени задать специальный способ перекодировки переменной (возможна категоризация практически любой сложности, заданная через соотношения между любыми переменными файла данных).

В дополнение к этому, в системе имеется специализированная процедура иерархической группировки, позволяющая осуществлять категоризацию данных по многим (до шести) переменным и строить различные категоризованные графики, описательные статистики и кор-

реляционные матрицы для подгрупп (пользователь может в интерактивном режиме игнорировать некоторые факторы в полной таблице группировок и исследовать статистики по маргинальным таблицам).

Многочисленные возможности форматирования и расстановки меток позволяют получать таблицы и отчеты презентационного качества, содержащие длинные метки и описания переменных. При этом важно отметить, что процедура группировки выполняется для чрезвычайно больших объемов данных (например, по одной категоризирующей переменной можно построить до 300 групп), а ее результаты содержат все соответствующие статистики дисперсионного анализа (включая полные таблицы ANOVA, критерии проверки гипотез типа критерия Левена однородности дисперсии, семь различных апостериорных (post-hoc) критериев и т. д.).

Как и во всех других модулях системы Statistica 6.0, для достижения высокой, не имеющей аналогов в сравнении с другими пакетами точности результатов здесь можно производить вычисления с повышенной точностью (если нужно – с четырехкратной).

Благодаря интерактивному характеру системы изучение данных становится очень простым делом. Например, графики для разведочных статистик можно построить непосредственно по любой таблице результатов, просто указав мышью на отдельные ячейки или группы ячеек. Одним щелчком мыши можно получать каскады графиков (в том числе сложных, например, со множественными категориями).

В дополнение к большому числу готовых статистических графиков пользователь может самостоятельно задавать различные типы визуализации исходных данных, описательных статистик, взаимосвязей между статистиками, группировок и категоризаций с помощью средств прямого доступа, что позволяет существенно упростить задачу. Средства графического разведочного анализа объединены с собственно статистическими процедурами, что существенно облегчает визуальный анализ данных (например, в интерактивном режиме можно удалять выбросы, выделять подмножества данных, осуществлять сглаживание и подгонку функций, а богатые средства работы с кистью позволяют легко выявлять и (или) выделять нужные данные).

Корреляции. В системе имеется большой набор методов для исследования корреляций между переменными. Прежде всего, могут быть вычислены все основные характеристики связи между переменными, в том числе коэффициенты корреляции Пирсона, ранговой корреляции Спирмена, Кендалла, Гамма, тетрагорический, Фи, Крамера, сопряженности Соммера, неопределенности, частные и частичные корреляции, автокорреляции и т. д.

Нелинейные корреляции, регрессия для цензурированных данных и другие более специализированные меры корреляции реализованы в процедурах *Нелинейное оценивание* и *Анализ выживаемости*.

Корреляционные матрицы могут быть вычислены с построчным, попарным удалением пропусков или с подстановкой среднего вместо пропущенных значений. Как и во всех других модулях системы Statistica 6.0, для достижения высокой, не имеющей аналогов среди других пакетов, точности результатов здесь можно производить вычисления с повышенной точностью (если нужно – с четырехкратной). Как и все численные результаты, корреляционные матрицы в системе Statistica 6.0 выводятся в виде таблиц, имеющих богатые возможности форматирования данных и визуализации численных результатов. Так, можно «указать» на конкретный корреляционный коэффициент и вызвать для него контекстное меню всевозможных «описательных диаграмм» (диаграммы рассеяния с доверительными интервалами, различные объемные гистограммы двумерных распределений, вероятностные графики и т. д.).

Богатые средства закрашивания позволяют выделять (или, наоборот, затенять) отдельные точки на диаграмме рассеяния и таким образом оценивать их влияние на положение линии регрессии (и других подогнанных кривых). Таким образом исследуются, например, выбросы или резко выделяющиеся наблюдения. Поддерживаются разнообразные форматы глобального вывода корреляций. Значимые коэффициенты корреляции могут автоматически выделяться цветом, каждую ячейку таблицы результатов можно расширить и посмотреть число наблюдений (n), по которым вычислен коэффициент и уровень значимости (p), можно запросить подробные результаты, включающие все описательные статистики (попарные средние и стандартные отклонения, В-веса, пересечения и т. д.).

Как и все численные результаты, корреляционные матрицы выводятся в виде таблиц, поддерживающих операцию масштабирования и интерактивно управляемый формат вывода значащих цифр (например, от +0,4 до +0,41358927645193). Таким образом, матрицы больших размеров можно сжимать с помощью операции уменьшения либо изменением формата вывода, что делается перетягиванием границ столбцов с помощью мыши. Это облегчает зрительное восприятие и, в частности, помогает быстро находить коэффициенты, превосходящие заданную величину или имеющие определенный уровень значимости (соответствующие ячейки в таблице результатов будут помечены красным цветом).

Диаграмма рассеяния, матричная диаграмма рассеяния, анализ по группам. Как и во всех других диалоговых окнах вывода, здесь доступны различные общие параметры графического вывода, позволяющие проводить дальнейшее изучение закономерностей и взаимосвязей между переменными; например, двух- и трехмерные диаграммы рассеяния (с метками наблюдений или без них) служат для выявления зависимостей по подмножествам наблюдений или последовательностям переменных.

Корреляционные матрицы могут быть категоризованы группирующими переменными и представляться графически в виде категоризованных диаграмм рассеяния. Могут быть также выведены последовательности таблиц результатов «группировки корреляционных матриц» (по одной матрице на каждое подмножество наблюдений), которые возможно использовать в дальнейшем при работе в процедуре *Моделирование структурными уравнениями* в модуле *Углубленные методы анализа*.

Вся корреляционная матрица может быть представлена на одном графике (со сколь угодно большим разрешением) в виде матрицы диаграмм рассеяния. Такие матрицы можно интерактивно просматривать, увеличивая нужные участки графика или прокручивая график в режиме увеличения. Имеется также возможность строить категоризованные матричные диаграммы рассеяния (одна матричная диаграмма на каждое подмножество данных). Можно поступить иначе и построить матричные диаграммы рассеяния для нескольких подмножеств (например, задаваемых уровнями группирующей переменной или сколь угодно сложными условиями выбора наблюдений), где отдельные подмножества данных изображаются различными маркерами.

Для визуализации корреляционных матриц и поиска в них глобальных закономерностей имеется еще много других графических методов (контурные графики, несглаженные поверхности, пиктограммы и т. д.). Все эти операции требуют лишь нескольких щелчков кнопкой мыши, а для выбора параметров предоставляются различные средства быстрого доступа. Одновременно на экране можно открыть любое число таблиц результатов и графиков, что принципиально облегчает интерактивный разведочный и сравнительный анализ.

Быстрые основные статистики. В дополнение к описательным статистикам и корреляциям, таблицам частот в системе Statistica 6.0 имеются интерактивные средства, позволяющие одним щелчком кнопкой мыши вычислять основные статистики и строить графики в любой момент в течение сеанса работы.

Быстрые основные статистики доступны из всех панелей инструментов и контекстных меню и выдают результат немедленно без каких-либо промежуточных диалоговых окон, если только пользователь не выберет другой режим. Более того, по умолчанию эти быстрые процедуры даже не требуют, чтобы были выбраны переменные для анализа – они выбираются автоматически по выделенному блоку ячеек или диапазону столбцов (или строк) в открытой в данный момент таблице исходных данных или таблице результатов.

Быстрые основные статистики – очень удобный способ быстро получать общие сведения об исследуемых переменных. Сюда входит полный набор описательных статистик (включая порядковые статистики), корреляции и таблицы частот вместе с ожидаемыми нормальными частотами и другие числовые характеристики, а также большой набор встроенных статистических графиков. Выходные результаты для всех статистик и графиков могут быть категоризованы (сгруппированы по значениям некоторой переменной).

Блочные статистики. Кроме того, все входные и выходные таблицы системы Statistica 6.0 поддерживают инструмент *Блочные статистики*, выдающий статистики для произвольного выделенного блока ячеек в открытой в данный момент таблице исходных данных или таблице результатов.

Результаты можно выводить по столбцам или по строкам выделенного блока в зависимости от смысла содержащихся в нем данных (исходные данные, нагрузки факторов, средние значения, частоты и т. д.).

Вероятностный калькулятор. Из любой панели инструментов системы Statistica 6.0 доступен удобный интерактивный *Калькулятор вероятностных распределений*. Он поддерживает множество типов стандартных распределений (бета, Коши, хи-квадрат, экспоненциальное, экстремальное (Гумбеля), Фишера, гамма, Лапласа, логнормальное, логистическое, Парето, Релея, t (Стьюдента), Вейбулла и Z (нормальное)). Графики (плотности вероятности и функции распределения) отображаются в интерактивном режиме, что позволяет наглядно представить себе то или иное распределение, изменяя его с помощью кнопок микропрокрутки (при нажатии левой кнопки мыши изменится последняя, а при нажатии правой – предпоследняя значащая цифра числа).

Имеется возможность строить составные графики распределений в нужном диапазоне. Таким образом, калькулятор дает возможность интерактивно исследовать структуру распределений (например, зависимость вероятности от параметров).

T-критерии. В системе Statistica 6.0 могут быть вычислены *t*-критерии для зависимых и независимых выборок, а также для независимых переменных и одномерочный *t*-критерий.

Как и в других процедурах системы Statistica 6.0, здесь в каждом меню результатов доступны функции подробной диагностики и графические инструменты. Так, например, в случае использования *t*-критериев для независимых выборок имеется возможность вычислить *t*-критерий с отдельными оценками дисперсий и критерий Левена однородности дисперсии, построить различные диаграммы размаха, категоризованные гистограммы, вероятностные графики, категоризованные диаграммы рассеяния и т. д. Специализированные критерии групповых различий реализованы в некоторых других модулях (например, *Непараметрическая статистика*, *Анализ выживаемости*, *Надежность и позиционный анализ*).

Таблицы частот, таблицы сопряженности, таблицы флагов и заголовков, многомерных откликов. В системе Statistica 6.0 имеются разнообразные средства для табуляции непрерывных и категориальных переменных, переменных множественного отклика и многомерных дихотомий. Вид таблиц и формат данных регулируются многочисленными функциями. Так, например, в таблицах для переменных множественного отклика или многомерных дихотомий маргинальные частоты и процентные показатели могут вычисляться по отношению к общему числу респондентов либо числу ответов, переменные множественного отклика можно обрабатывать парами, имеются различные варианты обработки пропущенных данных.

Таблицы частот могут быть вычислены для заданных пользователем логических условий выбора наблюдений (условия могут быть любой степени сложности и связывать любые переменные в текущем файле данных), относящих наблюдения к той или иной категории в таблице. Все таблицы можно настроить для вывода результатов в форме конечного отчета (для печати). Например, не имеющую аналогов в других пакетах «многоходовую итоговую таблицу» можно вывести с группировками и иерархическим упорядочением факторов; в ячейках таблицы сопряженности могут выводиться проценты по столбцу, по строке или проценты от общего числа наблюдений в таблице; для описания категорий в таблице можно использовать длинные метки значений; частоты, превышающие заданный пользователем уровень, можно выделить цветом и т. д.

Система может выводить кумулятивные (суммарные, или накопленные) частоты и относительные частоты, логит- и пробит-преобразованные частоты, ожидаемые нормальные частоты (а также значения

статистик Колмогорова – Смирнова, Лилиефорса и Шапиро – Уилкса), ожидаемые и остаточные частоты в таблицах сопряженности и др. Среди критериев значимости в таблицах сопряженности реализованы критерии: хи-квадрат Пирсона, максимума правдоподобия и с поправкой Йетса, хи-квадрат МакНемара, точный критерий Фишера (одно- и двухсторонний), Фи, тетрафорическая корреляция; доступны также следующие статистики: Кендалла, Гамма, Спирмена, Соммера и др.

Графические средства включают простые, категоризованные (множественные) и трехмерные гистограммы, гистограммы кросс-сечений (для любых «срезов» одно-, двух- и многовходовых таблиц) и много других типов графиков, в том числе уникальный в своем роде частотный график взаимодействий, суммирующий частоты для сложных таблиц сопряженности (подобно диаграмме средних в модели дисперсионного анализа). В интерактивном режиме можно просматривать целые каскады графиков (в том числе сложных, например, множественные категоризованные графики и графики взаимодействий).

4.1.2. Множественная регрессия

В модуле *Множественная регрессия* реализован полный набор методов линейной и нелинейной регрессии. Эти методы включают простую, множественную, пошаговую (с пошаговым (блоковым) включением или исключением предикторов), иерархическую, нелинейную (полиномиальную, экспоненциальную, логарифмическую и т. д.) и гребневую (ридж) регрессию, с включением или без включения константы (свободного члена), модели взвешенных наименьших квадратов (рисунок 7).

Система вычисляет полный набор статистик и дополнительных анализов, включая полную таблицу регрессии (со значениями стандартной ошибки для β и свободного члена, коэффициента детерминации (R^2) и скорректированного R^2 для моделей с константой и без нее и таблицей дисперсионного анализа), матрицы частных и получастных корреляций, корреляции и ковариации коэффициентов регрессии, матрицу выметания (обратную матрицу), статистику Дарбина – Уотсона, расстояния Махаланобиса и Кука, удаленные остатки, доверительные интервалы для предсказанных значений и многие другие статистики.

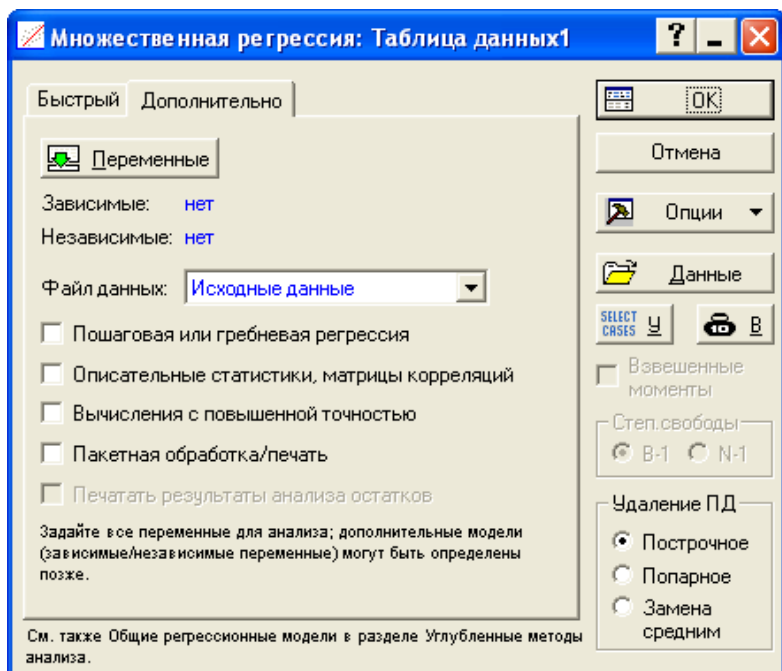


Рисунок 7 – Окно модуля *Множественная регрессия* системы Statistica 6.0

Предсказанные значения и анализ остатков. Широкие возможности анализа остатков и выбросов предоставляют многочисленные графики: диаграммы рассеяния, гистограммы, нормальные и полунормальные вероятностные графики, графики с удаленным трендом, графики частных корреляций, различные построчные графики и диаграммы остатков и выбросов и др.

Значения зависимой переменной для отдельных наблюдений можно просмотреть визуально с помощью разведочных пиктографиков и других многомерных графиков, доступных непосредственно из меню таблицы результатов. Остаточные и предсказанные значения можно автоматически добавлять к текущему файлу данных. Процедура прогнозирования позволяет проводить анализ типа «что если» и интерактивно вычислять предсказанные значения по задаваемым с клавиатуры значениям предикторов.

4.2. Модуль «Углубленные методы анализа»

4.2.1. Общие линейные модели

Процедура *Общие линейные модели* системы Statistica 6.0 предназначена для анализа откликов одной или нескольких непрерывных переменных как функции одной или нескольких категориальных или непрерывных независимых переменных. Процедура является не только вычислительным средством, но также и наиболее удобным и полным из доступных приложений, предоставляющим широкий выбор опций, графиков, сопутствующих статистик и расширенных диагностических функций (рисунок 8).

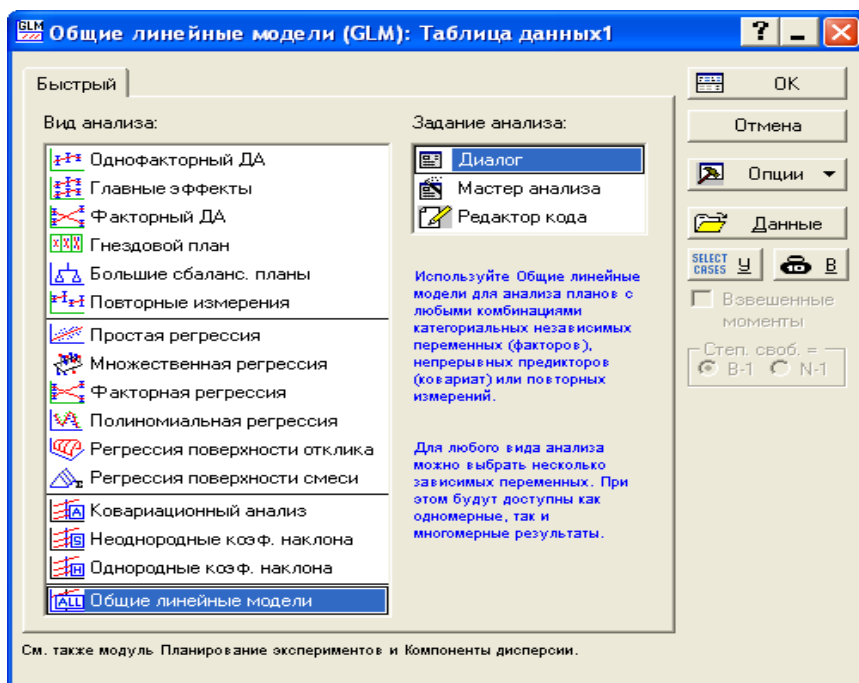


Рисунок 8 – Окно процедуры *Общие Линейные модели* системы Statistica 6.0

Система Statistica 6.0 предлагает наиболее богатые опции для поддержки задач, для которых не существует стандартных решений. Процедура *Общие линейные модели* вычисляет все стандартные результаты, включая таблицы с одномерными и многомерными теста-

ми, описательными статистиками и т. п. Также процедура предлагает широкий выбор результатов и графических опций, которые обычно не доступны в других программах, простые способы тестирования линейных комбинаций параметров оценивания, спецификации пользовательских членов ошибки и эффектов, расширенные апостериорные методы сравнения межгрупповых эффектов, повторные эффекты измерений и взаимодействий между повторными измерениями.

4.2.2. Общие регрессионные модели

Общие регрессионные модели системы Statistica 6.0 предлагают пользователю уникальную высокоэффективную реализацию стандартных функций в общей линейной модели, а также включают широкий набор пошаговой регрессии и технологии лучшего подмножества построения моделей, поддерживающие непрерывные и категориальные переменные. Пошаговые методы и методы лучшего подмножества построения моделей составных планов можно использовать в данной процедуре, включая планы с эффектами для категориальных предсказанных переменных (рисунок 9).

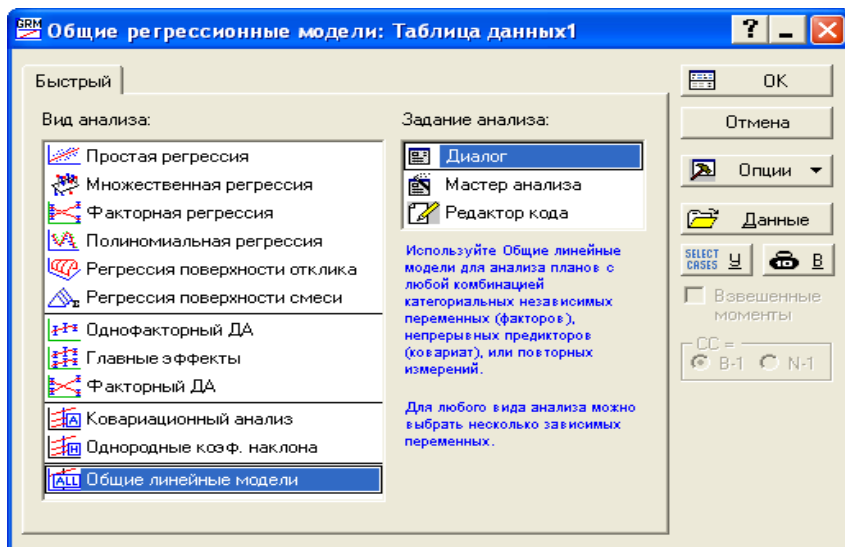


Рисунок 9 – Окно процедуры *Общие регрессионные модели* системы Statistica 6.0

Процедура *Общие регрессионные модели* не ограничивает анализы в плане содержания непрерывных предсказанных переменных. Кроме этого, исключительные опции регрессионных результатов включают карты Парето параметров оценивания, целые модельные критерии с различными методами оценивающих моделей, частные и получастные корреляции и т. п.

4.2.3. Анализ временных рядов

В процедуре *Временные ряды и прогнозирование* реализован широкий набор методов описания, построения моделей, декомпозиции и прогнозирования временных рядов, как во временной, так и в частотной области. Все процедуры полностью совместимы, и результаты анализа одной модели (например, остатки, вычисленные для модели *АРПСС*) можно использовать для дальнейшего анализа (например, вычисления автокорреляции остатков) (рисунок 10).

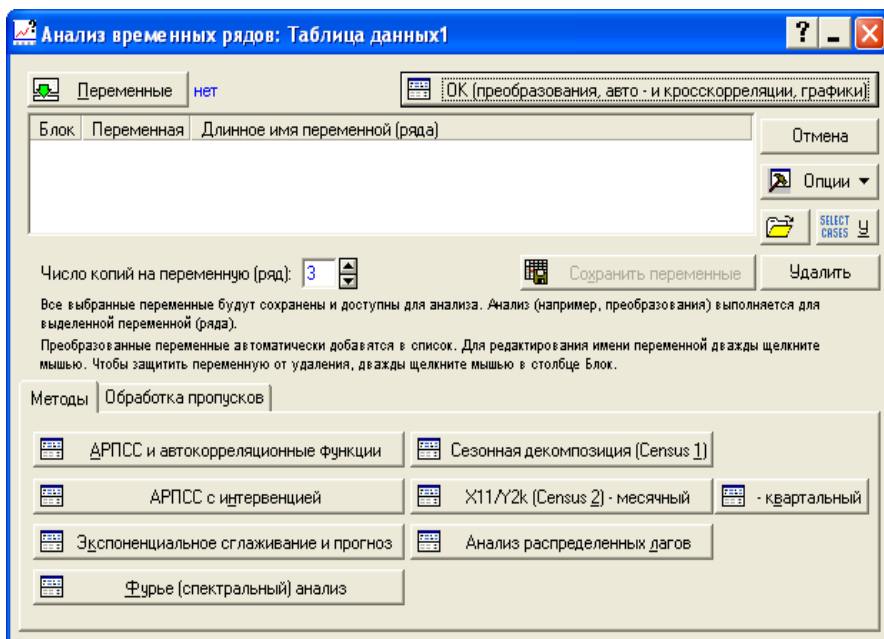


Рисунок 10 – Окно *Анализ временных рядов* процедуры *Временные ряды и прогнозирование* системы Statistica 6.0

Также имеются самые разнообразные возможности для просмотра и графического представления одномерных и многомерных рядов. Можно анализировать очень длинные ряды (более 100 тыс. наблюдений). С многомерными рядами (в случае многомерных исходных данных или с рядами, полученными на различных этапах анализа) можно работать в активной рабочей области, где их можно просматривать и сопоставлять друг с другом.

Система Statistica 6.0 автоматически отмечает все этапы анализа временного ряда и сохраняет полную историю преобразований и полученные результаты (остатки модели *АРПСС*, сезонную составляющую и т. д.). Поэтому пользователь всегда имеет возможность вернуться к более раннему этапу анализа или отобразить на графике исходный ряд и его преобразования. Информация о последовательных преобразованиях хранится в виде длинных меток переменных, поэтому при сохранении вновь полученных рядов в файле данных автоматически сохраняется вся «история» каждого из рядов.

С помощью различных преобразований исходного временного ряда можно понять его структуру и имеющиеся в нем закономерности. В модуле реализованы такие часто используемые преобразования, как удаление тренда, удаление автокорреляций, сглаживание скользящими средними (невзвешенными или взвешенными – с весами, заданными пользователем или вычисленными по методам Даниеля, Тьюки, Хэмминга, Парзена и Бартлета), медианное сглаживание (среднее значение заменено медианой), простое экспоненциальное сглаживание, взятие разностей, суммирование, вычисление остатков, сдвиг, косинус-сглаживание, преобразование Фурье, а также обратное преобразование Фурье и др. Можно выполнить анализ автокорреляций, частных автокорреляций и кросскорреляций. Далее рассмотрим конкретные процедуры анализа временных рядов.

АРПСС и автокорреляционные функции, АРПСС с интервенцией. Процедура *Анализ временных рядов* включает полную реализацию модели авторегрессии и проинтегрированного скользящего среднего (*АРПСС*). Модель может включать константу. Перед построением модели ряд может быть подвергнут преобразованию, которое автоматически будет отменено после построения прогноза по *АРПСС*, при этом предсказанные значения и их стандартные ошибки будут выражены через значения исходного (а не преобразованного) ряда.

Могут быть вычислены приближенные и точные суммы квадратов из условия максимума правдоподобия. Уникальной особенностью модели *АРПСС* процедуры *Анализ временных рядов* является способность анализировать модели с длинными периодами сезонности (с лагом

до 30). Стандартный набор результатов содержит оценки параметров, стандартные ошибки и корреляции. Предсказанные значения могут быть представлены в числовой и графической форме и добавлены к исходному ряду. Имеются многочисленные дополнительные функции для исследования остатков модели *АРПСС*, в том числе большой набор графических средств.

Реализация модели *АРПСС* в модуле *Временные ряды* позволяет проводить анализ прерванных временных рядов (рядов с интервенциями). Имеется возможность использовать одновременно несколько различных интервенций (до 6). Доступны следующие виды интервенций: однопараметрические скачкообразные, двухпараметрические постепенные, временные (характер воздействия можно просмотреть на графике). Для всех прерванных моделей могут быть построены прогнозы, которые можно вывести на график (вместе с исходным рядом) и, если требуется, добавить к исходному ряду.

Сезонное и несезонное экспоненциальное сглаживание. В процедуре *Анализ временных рядов* полностью реализованы все 12 классических моделей экспоненциального сглаживания. Задание модели может включать аддитивную или мультипликативную сезонную составляющую и (или) линейный, экспоненциальный или демпфированный тренд; в частности, доступны популярные модели с линейным трендом Холта – Винтера.

Пользователь может задавать начальное значение параметров сглаживания, начальное значение тренда и (если требуется) сезонные факторы. Для тренда и сезонной составляющей могут быть заданы независимые параметры сглаживания. Для определения лучшей комбинации параметров используется метод поиска на сетке; в таблицах результатов для всех комбинаций значений параметров сглаживания вычисляется средняя ошибка, средняя абсолютная ошибка, сумма квадратов ошибок, среднеквадратическая ошибка, средняя относительная ошибка и средняя абсолютная ошибка. Наименьшие значения этих ошибок выделяются цветом.

Имеется возможность автоматического поиска лучшего набора параметров в смысле среднеквадратической, средней абсолютной или средней относительной ошибки (для этого используется общая процедура минимизации). Все результаты преобразования экспоненциальным сглаживанием, остатки и прогноз на требуемое число шагов можно в дальнейшем проанализировать и изобразить на графике. Для оценки адекватности модели используются графики, на которых вместе с исходным рядом в подходящем масштабе по оси *Y* изображаются его сглаженный вариант, прогноз и ряд остатков.

Классическая сезонная декомпозиция (метод Census I). Имеется возможность задать произвольный сезонный лаг и выбрать либо аддитивную, либо мультипликативную сезонную модель. Программа вычисляет скользящие средние, отношения или разности, сезонные компоненты, ряд с сезонной поправкой, сглаженную тренд-циклическую и нерегулярную компоненты. Все эти составляющие ряда доступны для дальнейшего анализа, например, для проверки адекватности можно построить гистограммы, нормальные вероятностные графики и т. д.

Месячная сезонная корректировка (метод Census II). Процедура *Анализ временных рядов* включает полную реализацию метода X-11 сезонной корректировки, принятого Статистическим управлением США (US Bureau of the Census). Структура всех функций и диалоговых окон соответствует требованиям и соглашениям, описанным в документации Bureau of the Census. Можно выбрать либо аддитивные, либо мультипликативные модели.

Пользователь может дополнительно вычислить априорные поправки на число рабочих дней и сезонные поправки. Колебания числа рабочих дней оцениваются регрессионными методами (с правильной обработкой крайних членов ряда) и затем (по желанию) используются для корректировки ряда. Реализованы стандартные средства для градуировки выбросов, вычисления сезонных факторов и вычисления тренд-циклической компоненты (имеется возможность выбирать несколько типов взвешенного скользящего среднего; кроме того, программа может сама находить оптимальную длину и тип скользящего среднего).

Итоговые компоненты ряда (сезонная, тренд-циклическая, нерегулярная) и ряд с внесенной сезонной поправкой всегда доступны для дальнейшего анализа и вывода на график. Кроме того, все они могут быть сохранены для дальнейшего исследования другими методами и (или) в других программах. Все компоненты выводятся на графики в различной форме, включая категоризованные графики по месяцам (кварталам).

Анализ распределенных лагов. С помощью реализованных в процедуре *Анализ временных рядов* методов анализа полиномиальных распределенных лагов можно выполнять оценку моделей с обычными лагами и лагами Алмона. Для анализа распределений переменных модели имеется ряд графических средств.

Спектральный анализ (Фурье) и кросс-спектральный анализ. Процедура *Анализ временных рядов* включает полную реализацию методов спектрального анализа (Фурье) одного ряда и кросс-спектральный анализ двух рядов. Преимущества реализации спектрального анализа в системе Statistica 6.0 особенно отчетливо проявляются при

анализе очень длинных временных рядов (с более чем 250 тыс. наблюдений) и не предполагают каких-либо ограничений по длине ряда (в частности, длина исходного ряда не обязательно должна быть четной). Вместе с тем, иногда бывает разумно предварительно увеличить или уменьшить длину ряда.

Стандартные методы предварительной обработки ряда включают косинус-сглаживание, вычитание среднего и удаление тренда. Результаты обычного спектрального анализа содержат коэффициенты частоты, периода, коэффициенты при синусах и косинусах, периодограммы и оценку спектральной плотности. Оценка плотности может быть вычислена с помощью весов Даниеля, Хэмминга, Бартлетта, Тьюки, Парзена или с весами и шириной, заданными пользователем.

Очень полезно, особенно при работе с длинными рядами, иметь возможность выводить в убывающем порядке заранее заданное число точек периодограммы или спектральной плотности; таким образом можно легко обнаружить резкие пики периодограммы и спектральной плотности для длинных рядов. Имеется возможность вычислить d -критерий Колмогорова – Смирнова для значений периодограммы, чтобы проверить, подчиняются ли они экспоненциальному распределению (является ряд белым шумом или нет).

Для представления результатов анализа имеются различные типы графиков, можно отобразить коэффициенты при синусах и косинусах, периодограмму, лог-периодограмму, спектральную и лог-спектральную плотности по отношению к частотам, периодам и лог-периодам. В случае длинного исходного ряда имеется возможность выбрать конкретный сегмент (период), для которого будут изображаться соответствующие периодограмма и график спектральной плотности, тем самым будет улучшено их «разрешение».

При кросс-спектральном анализе, в дополнение к результатам обычного спектрального анализа каждого отдельного ряда, вычисляется кросс-периодограмма (вещественная и мнимая части), ко-спектральная плотность, квадратурный спектр, кросс-амплитуда, значения когерентности, усиления и фазовый спектр. Все эти величины могут быть выведены на график, где по горизонтальной оси будут откладываться частота, период или лог-период для всего интервала периодов (соответственно, частот) либо для выбранного пользователем диапазона. Указанное пользователем количество наибольших значений кросс-периодограммы (вещественных или мнимых) может быть выведено в убывающем порядке в виде таблицы результатов, что позволяет легко выявлять на ней резкие пики для длинных исходных рядов.

Как и во всех других процедурах модуля *Временные ряды*, все полученные ряды могут быть добавлены в активную рабочую область и затем подвергнуты дальнейшему исследованию с помощью других методов анализа временных рядов или средствами других модулей системы Statistica.

Прогнозирование на основе регрессионных методов. Наконец, в системе Statistica 6.0 реализованы регрессионные методы анализа временных рядов для переменных с запаздыванием (лагом) или без него, в том числе регрессия, проходящая через начало координат, нелинейная регрессия и интерактивное прогнозирование по методу «что если».

4.3. Модуль «Многомерный разведочный анализ»

4.3.1. Методы кластеризации

В модуле *Кластерный анализ* системы Statistica 6.0 реализован полный набор методов кластерного анализа данных, включая методы *k*-средних, иерархической кластеризации и двухвходового объединения. Данные могут поступать как в исходном виде, так и в виде матрицы расстояний между объектами. Наблюдения, переменные или (и) наблюдения и переменные можно кластеризовать, используя различные меры расстояния (евклидово, квадрат евклидова, городских кварталов (манхэттеновское), Чебышева, степенное, процент несогласия и другие) и различные правила объединения (связывания) кластеров (одиночная, полная связь, невзвешенное и взвешенное попарное среднее по группам, невзвешенное, взвешенное расстояние между центрами, метод Варда и другие) (рисунок 11).

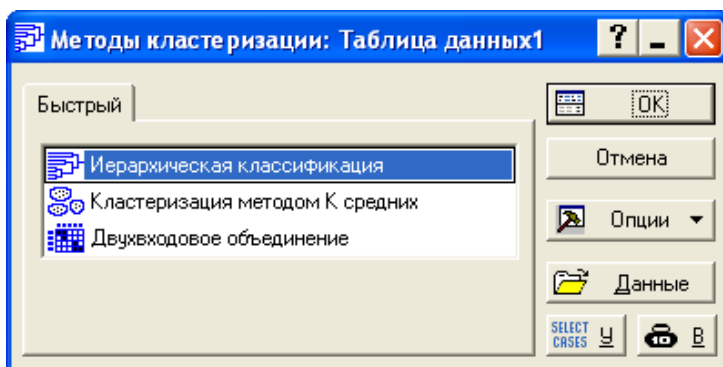


Рисунок 11 – Окно процедуры *Методы кластеризации* системы Statistica 6.0

Матрицы расстояний можно сохранять для дальнейшего анализа в других модулях системы Statistica 6.0. При проведении кластерного анализа методом k -средних пользователь имеет полный контроль над начальным расположением центров кластеров. Могут быть выполнены чрезвычайно большие планы анализа: так, например, при иерархическом (древовидном) связывании можно работать с матрицей из 90 тыс. расстояний. Помимо стандартных результатов кластерного анализа, в модуле доступен также разнообразный набор описательных статистик и расширенных диагностических методов (полная схема объединения с пороговыми уровнями при иерархической кластеризации, таблица дисперсионного анализа при кластеризации методом k -средних). Информация о принадлежности объектов к кластерам может быть добавлена к файлу данных и использоваться в дальнейшем анализе. Графические возможности модуля *Кластерный анализ* включают настраиваемые дендрограммы, двухвходовые диаграммы объединений, графическое представление схемы объединения, диаграмму средних при кластеризации по методу k -средних и многое другое.

4.3.2. Факторный анализ и анализ главных компонент

Модуль *Факторный анализ* системы Statistica 6.0 содержит широкий набор статистик и методов факторного анализа (а также иерархического факторного анализа) с расширенной диагностикой и большим многообразием исследовательских и разведочных графиков (рисунок 12).

В модуле *Анализ главных компонент и классификация* системы Statistica 6.0 можно выполнять общий и иерархический косоугольный анализ главных компонент и главных факторов для наборов данных, содержащих до 300 переменных. Выходные результаты системы – собственные значения (обычные, кумулятивные и относительные), нагрузки факторов и коэффициенты факторных баллов (которые можно добавить к файлу входных данных, просмотреть на пиктограмме и в интерактивном режиме перекодировать), а также некоторые более специальные статистики и диагностики. В распоряжении пользователя имеются следующие методы вращения факторов: варимакс, биквартимакс, квартимакс и эквимакс (по нормализованным либо первоначальным нагрузкам), а также косоугольные вращения (рисунок 13).

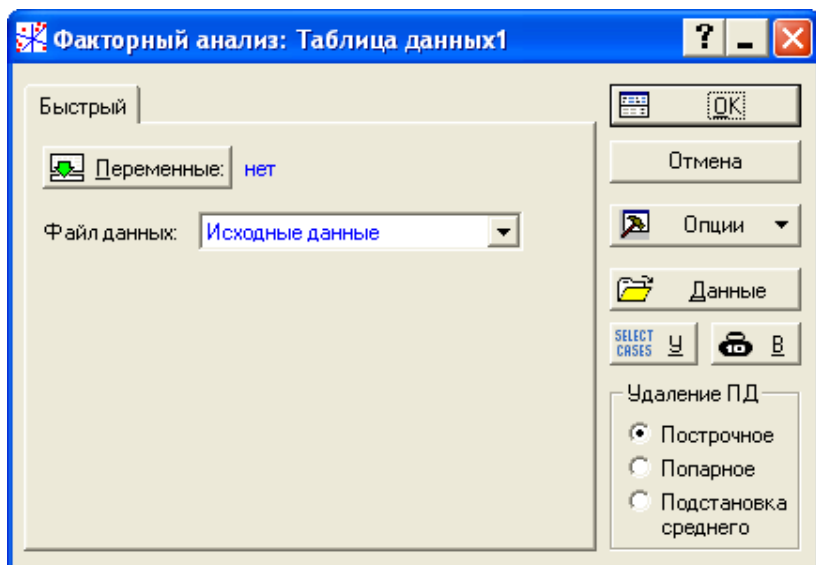


Рисунок 12 – Окно модуля *Факторный анализ* системы Statistica 6.0

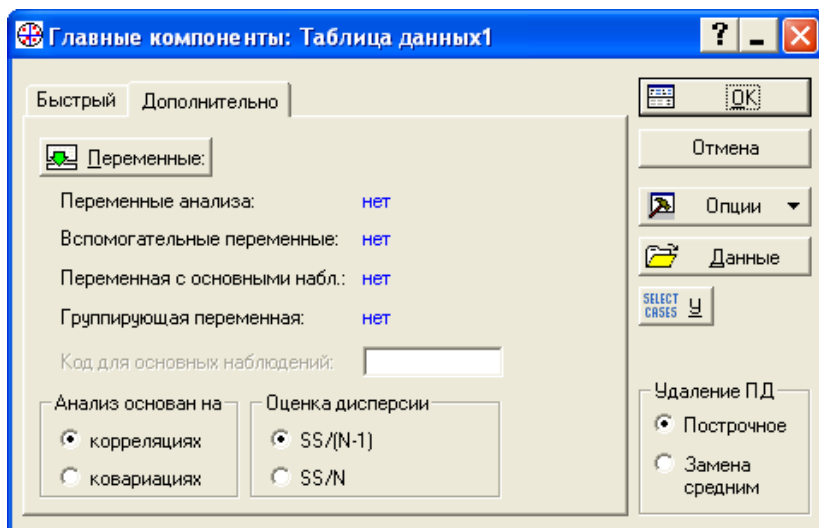


Рисунок 13 – Окно процедуры *Главные компоненты* системы Statistica 6.0

Пространство факторов можно визуально просматривать «срез за срезом» на двух- или трехмерных диаграммах рассеяния с отмеченными точками данных; среди других графических средств – графики «каменистой осыпи», различные типы диаграмм рассеяния, гистограммы, линейные графики и др. После того как факторное решение определено, пользователь может вычислить (воспроизвести) корреляционную матрицу и оценить согласованность факторной модели путем анализа остаточной корреляционной матрицы (или остаточной дисперсионной (ковариационной) матрицы). На входе можно использовать как исходные данные, так и матрицы корреляций. Подтверждающий факторный анализ и другие, связанные с ним виды анализа, могут быть выполнены средствами процедуры *Моделирование структурными уравнениями* из модуля *Многомерный разведочный анализ* системы Statistica 6.0, где специальный *Мастер подтверждающего факторного анализа* проведет пользователя через все этапы построения модели.

4.3.3. Канонический анализ

В модуле *Канонический анализ* системы Statistica 6.0 реализован полный набор методов канонического анализа (дополняющий методы канонического анализа, встроенные в другие модули). Работать можно как с файлами исходных данных, так и с корреляционными матрицами; вычисляются все стандартные статистики канонической корреляции (собственные векторы и собственные значения, коэффициенты избыточности, канонический вес, нагрузки, дисперсии, критерии значимости для каждого из корней и другие), а также некоторые расширенные диагностики. Для каждого наблюдения могут быть вычислены значения канонических переменных, которые затем можно просмотреть на встроенных пиктографиках, а также добавить к файлу данных (рисунки 14).

В модуле *Каноническая корреляция* имеется большой набор графических средств (включая диаграммы собственных значений и канонических корреляций, диаграммы значений канонических переменных и много других). Подтверждающий анализ структурных связей между латентными переменными может быть также выполнен средствами процедуры *Моделирование структурными уравнениями* из модуля *Многомерный разведочный анализ* системы Statistica 6.0.

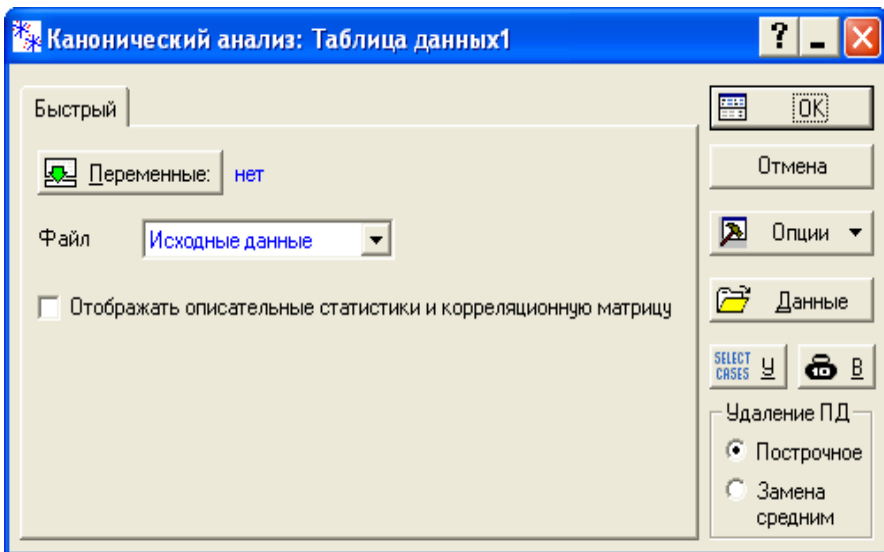



Рисунок 14 – Окно модуля *Канонический анализ* системы Statistica 6.0

5. ПОШАГОВЫЕ ПРИМЕРЫ В ПРОГРАММЕ STATISTICA 6.0

Пример 1. Корреляционный анализ

Постановка задачи. Используя данные о работе 10 торговых организаций, представленные в электронной книге на рисунке 15, с применением корреляционного метода оценим степень тесноты и направление взаимосвязи между хозяйственно-финансовыми результатами деятельности: товарооборотом, прибылью, численностью работников и среднегодовой стоимостью основных средств.

Техническое выполнение задачи в системе Statistica 6.0. Вычислим матрицу корреляций из файла данных (рисунок 15). Сначала вызовем стартовую панель *Основные статистики и таблицы* в меню *Анализ* (рисунок 16) или нажмем кнопку  в нижнем левом углу экрана (рисунок 17).

№ организации	1 товарооборот, млн р.	2 прибыль, млн р.	3 численность работников, чел.	4 среднегодовая стоимость основных средств, млн р.
1	1495	39	75	2054
2	2425	14	97	1813
3	399	3	79	3301
4	3040	97	83	984
5	4195	58	95	1186
6	1732	10	64	1427
7	5040	27	115	1614
8	3475	94	125	2654
9	2351	58	93	1372
10	1920	14	90	2750

Рисунок 15 – Данные о работе 10 торговых организаций (исходные данные)

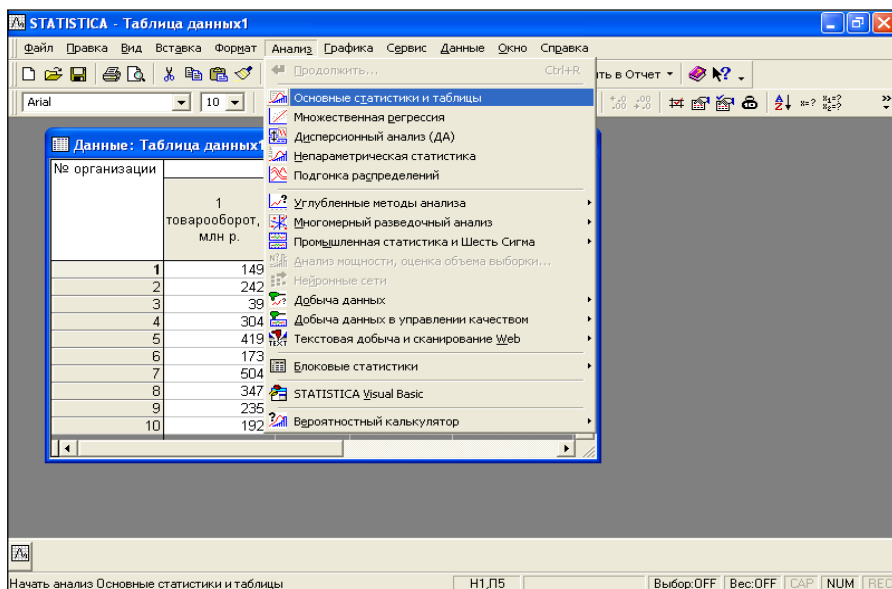


Рисунок 16 – Вызов панели *Основные статистики и таблицы* (1-й вариант)

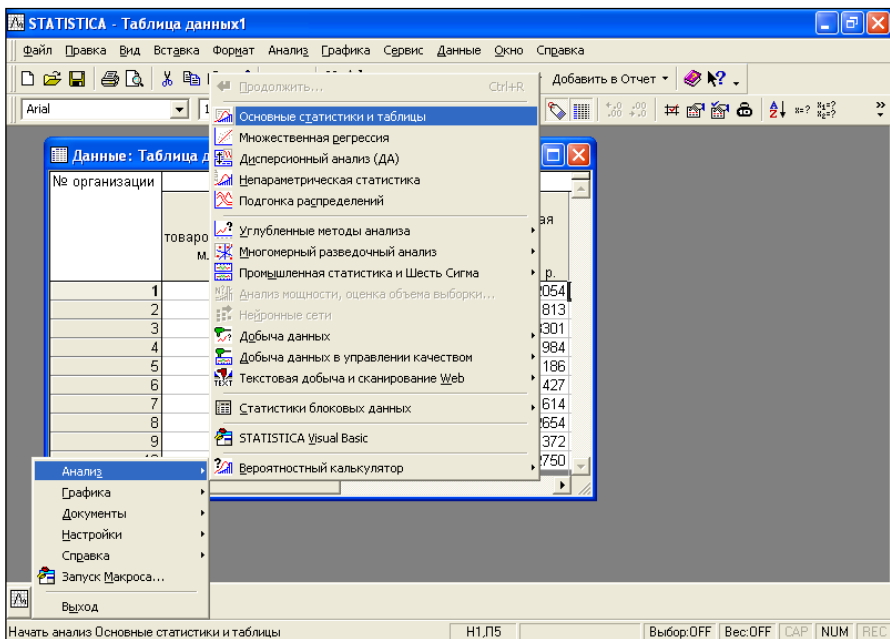


Рисунок 17 – Вызов панели *Основные статистики и таблицы* (2-й вариант)

При вызове стартовой панели убедимся, что в электронной таблице не выделен блок, т. е. группа ячеек (чтобы отменить выделение блока, щелчком кнопкой мыши по любой другой ячейке в электронной таблице). Если блок выделен, система Statistica будет проводить анализ именно этого блока.

На панели *Основные статистики и таблицы* выберем пункт *Парные и частные корреляции* (см. рисунок 6). После нажатия кнопки *OK* появится диалоговое окно *Парные и частные корреляции* (рисунок 18).

Как и большинство диалоговых окон *Анализа*, окно *Парные и частные корреляции* содержит набор опций. Как правило, предлагается не менее двух типов анализа.

Вкладка *Парные* этого окна содержит опции, позволяющие вычислить парные корреляции нескольких переменных и отобразить результаты в различных формах, например, вывести уровень значимости и т. д.

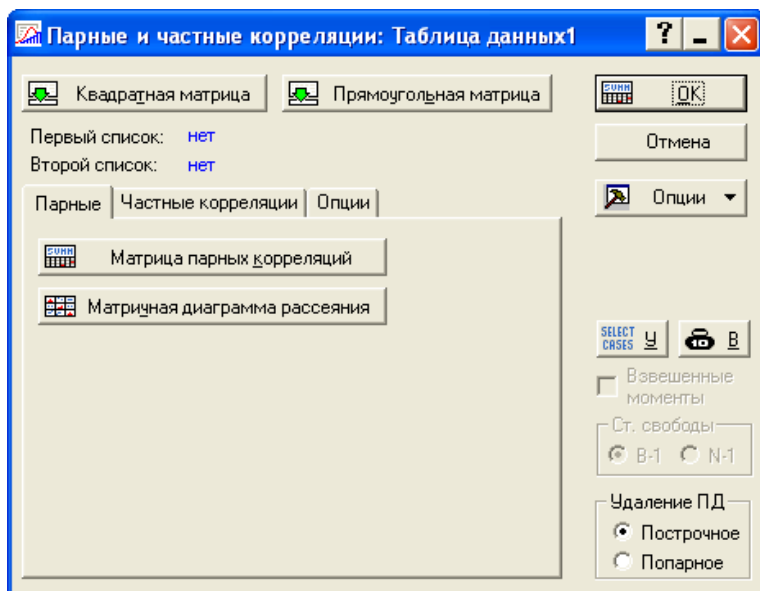


Рисунок 18 – Диалоговое окно *Парные и частные корреляции* системы Statistica 6.0

Вкладка *Частные корреляции* содержит опции для вычисления как парных, так и частных корреляций, а также опции сохранения матриц, получения некоторых данных статистики или графиков. В зависимости от выбранного анализа могут быть доступны также дополнительные вкладки.

Все диалоговые окна программы Statistica 6.0 поддерживают «контекстную подсказку». Если пользователь не знает, что делать дальше, просто необходимо нажать кнопку *OK* или *Итоги*, и программа сама сделает следующий логический шаг. При этом, если какой-либо этап был упущен, программа попросит ввести недостающую информацию (например переменные для анализа).

После нажатия кнопки *Квадратная матрица* (или кнопки *OK*) появится диалоговое окно выбора переменной (рисунок 19).

Диалоговое окно выбора переменной поддерживает разные способы выделения переменных (включая стандартные соглашения Windows о выделении при нажатых клавишах *Shift* и *Ctrl* произвольных или непрерывных списков переменных соответственно). Нажав кнопку *Инфо*, пользователь имеет возможность просмотреть отсортированный список всех значений и описательных статистик переменной.

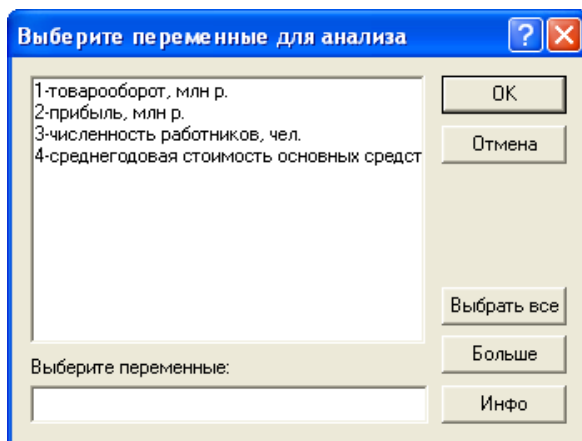


Рисунок 19 – Диалоговое окно для ввода переменных

В нашем примере нажмем кнопку *Выбрать все*, а затем *OK*. После этого мы вернемся в диалоговое окно *Парные и частные корреляции*. Далее отметим *Удаление ПД Парное* (справа в нижней части окна) и щелкнем по кнопке *OK* для получения матрицы корреляций между товарооборотом, прибылью, численностью работников, среднегодовой стоимостью основных средств. Заметим, что вместо кнопки *OK* можно было выбрать строку *Матрица парных корреляций*.

Результаты корреляции представлены на рисунке 20.

Корреляции (Таблица данных1)				
Отмеченные корреляции значимы на уровне $p < ,05000$				
N=10 (Построчное удаление ПД)				
Переменная	товарооборот, млн р.	прибыль, млн р.	численность работников, чел.	среднегодовая стоимость основных средств, млн р.
товарооборот, млн р.	1,00	0,46	0,70	-0,52
прибыль, млн р.	0,46	1,00	0,41	-0,37
численность работников, чел.	0,70	0,41	1,00	0,11
среднегодовая стоимость основных средств, млн р.	-0,52	-0,37	0,11	1,00

Рисунок 20 – Результаты корреляции

Выводы по задаче. В примере товарооборот тесно связан с численностью работников (0,70), в средней степени с прибылью (0,46). При этом наблюдается обратная по направлению и средняя по силе взаимосвязь между среднегодовой стоимостью основных средств и това-

рооборотом $(-0,52)$, что негативно характеризует использование основных средств.

Пример 2. Регрессионный анализ

Постановка задачи. На основании данных о работе 10 торговых организаций, представленных в электронной книге на рисунке 15, выполним регрессионный анализ по двум показателям: товарообороту (зависимый признак) и среднесписочной численности работников (влияющий признак) с целью определения уравнения регрессии.

Техническое выполнение задачи в системе Statistica 6.0. Для запуска метода *Простая регрессия* в модуле *Углубленные методы анализа* выберем процедуру *Общие регрессионные модели* через меню *Анализ* (см. рисунок 16) или через кнопку в нижнем левом углу (см. рисунок 17). Появится стартовая панель (рисунок 21).

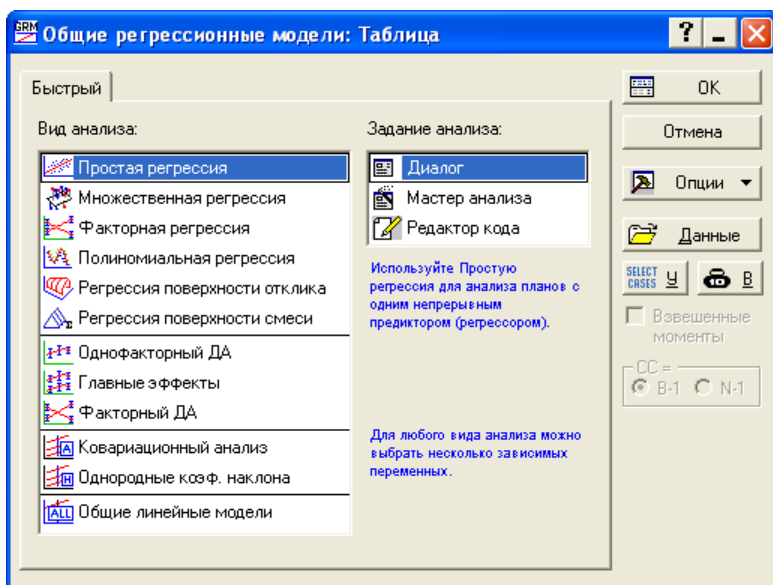


Рисунок 21 – Диалоговое окно *Общие регрессионные модели* системы *Statistica 6.0*.

Это диалоговое окно используется для задания различных видов регрессионного анализа. В нашем примере выберем метод *Простая регрессия* в поле *Вид анализа*, а в поле *Задание анализа* выберем ре-

жим *Диалог*, затем нажмем кнопку *OK* на стартовой панели для отображения диалогового окна *Однофакторный ДА* (рисунок 22).

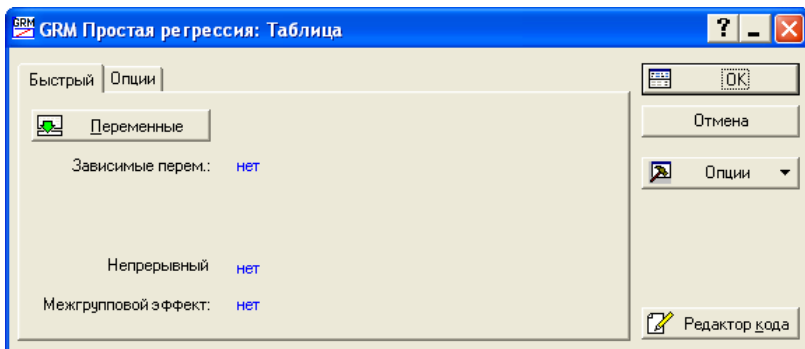


Рисунок 22 – Диалоговое окно *GRM Простая регрессия*

Нажав на кнопку *Переменные*, в качестве зависимой переменной отметим товарооборот, а в качестве предиктора – численность работников (нажмем кнопку *OK*, чтобы вернуться в предыдущее окно). После повторного нажатия кнопки *OK* в диалоговом окне *GRM Простая регрессия* система произведет анализ и выведет полученные результаты в окне *GRM Результаты* (рисунок 23).

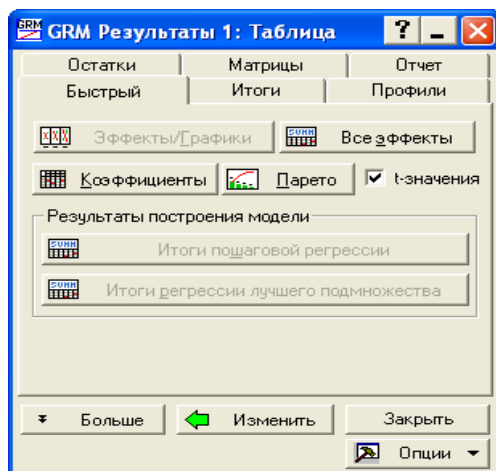


Рисунок 23 – Диалоговое окно *GRM Результаты*

С помощью этого окна можно обратиться к различным описательным графикам и таблицам результатов. Окно содержит вкладки,

позволяющие быстро выбирать требуемые результаты. В нашем примере мы будем пользоваться результатами, доступными во вкладке *Быстрый*. Для вывода необходимого уравнения (его параметров) нажмем кнопку *Все эффекты* (рисунки 24–26).

		SS модели и SS остатков (Таблица)				
Зависим. Перемен.	Множест. R	Множест. R2	Скоррект R2	SS Модель	сс Модель	
товарооборот, млн р.	0,704749	0,496671	0,433755	8359077	1	

Рисунок 24 – Результаты регрессионного анализа – *SS модели и SS остатков*

		Одномерные критерии значимости для товарооборот, млн р. (Таблица)				
		Сигма-ограниченная параметризация				
		Декомпозиция гипотезы				
Эффект	SS	Степени Свободы	MS	F	p	
Св.член	1728494	1	1728494	1,632362	0,237203	
численность работников, чел.	8359077	1	8359077	7,894178	0,022855	
Ошиб.	8471131	8	1058891			

Рисунок 25 – Результаты регрессионного анализа – *Одномерные критерии значимости*

		Оценки параметров (Таблица)			
		Сигма-ограниченная параметризация			
Эффект	товарооборот, млн р. Парам.	товарооборот, млн р. Ст. Ош.	товарооборот, млн р. t	tc	
Св.член	-2245,50	1757,535	-1,27764		
численность работников, чел.	52,98	18,855	2,80966		

Рисунок 26 – Результаты регрессионного анализа – *Оценки параметров*

Выводы по задаче. В нашем примере между товарооборотом и численностью работников установлена следующая форма уравнения взаимосвязи:

$$Y = -2245,5 + 52,98 \cdot X,$$

где Y – товарооборот, млн р.;

X – численность работников, чел.

Значимость свободного члена составляет 76%, коэффициента регрессии – 98%, множественный R равен 0,7 (как в результате корреляционного анализа). В целом значимость уравнения составляет 98%, что позволяет его использовать для прогнозирования товарооборота в зависимости от численности работников.

Пример 3. Кластерный анализ

Постановка задачи. Проведем кластерный анализ по трем показателям: товарообороту, прибыли и среднесписочной численности работников (рисунок 15) с целью группировки 10 торговых организаций.

Техническое выполнение задачи в системе Statistica 6.0. Для запуска модуля *Кластерный анализ* выберем команду *Многомерный разведочный анализ / Кластерный анализ* (рисунок 27). Появится стартовая панель *Методы кластеризации* (см. рисунок 11), которая предлагает выбрать три метода: иерархическую классификацию (древовидную кластеризацию), метод k -средних и двухвходовое объединение.

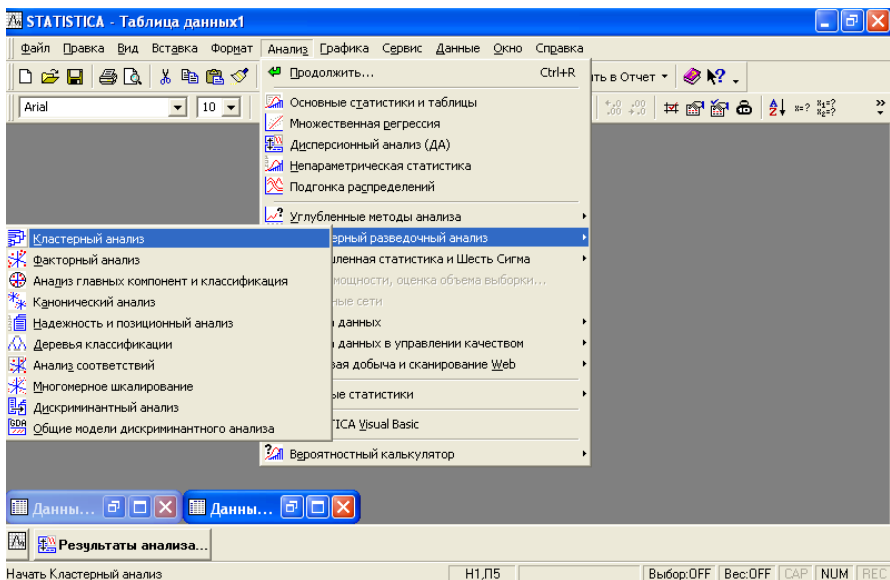


Рисунок 27 – Активация модуля *Кластерный анализ* Statistica 6.0

На первом этапе кластеризации воспользуемся процедурой *Иерархическая классификация* (рисунок 28).

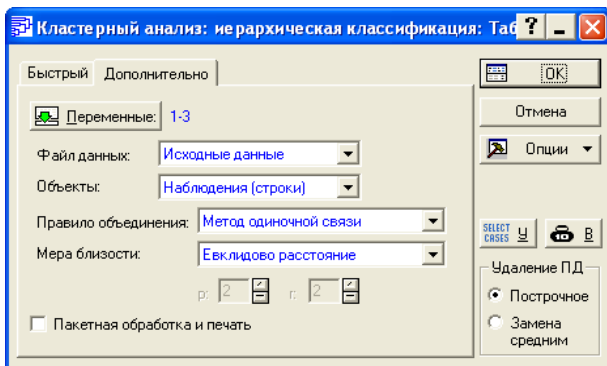


Рисунок 28 – Диалоговое окно *Кластерный анализ* процедуры *Иерархическая классификация*

Последовательно вводим переменные 1–3 (товарооборот, прибыль и численность работников), затем отмечаем, что объектами наблюдения в нашем примере являются строки и нажимаем кнопку *OK*. Возникает окно *Результаты иерархической классификации* (рисунок 29), в котором предлагается во вкладке *Быстрый* выбрать тип дендрограммы. Выбираем вертикальную дендрограмму. Результат кластеризации – график на рисунке 30.

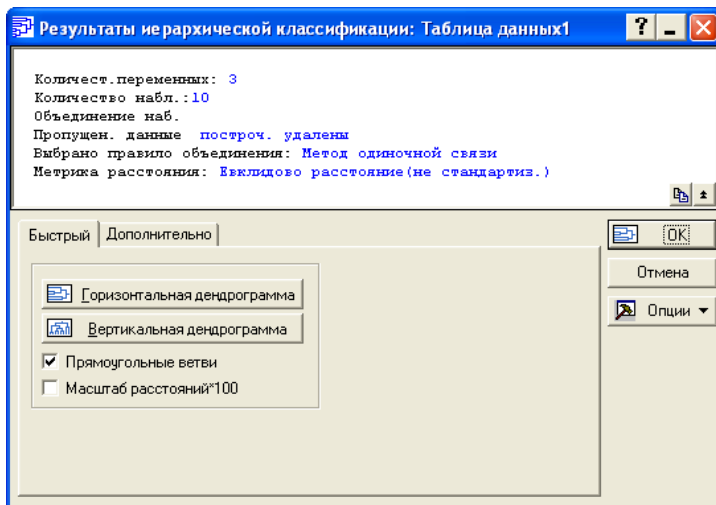


Рисунок 29 – Диалоговое окно *Результаты иерархической классификации*

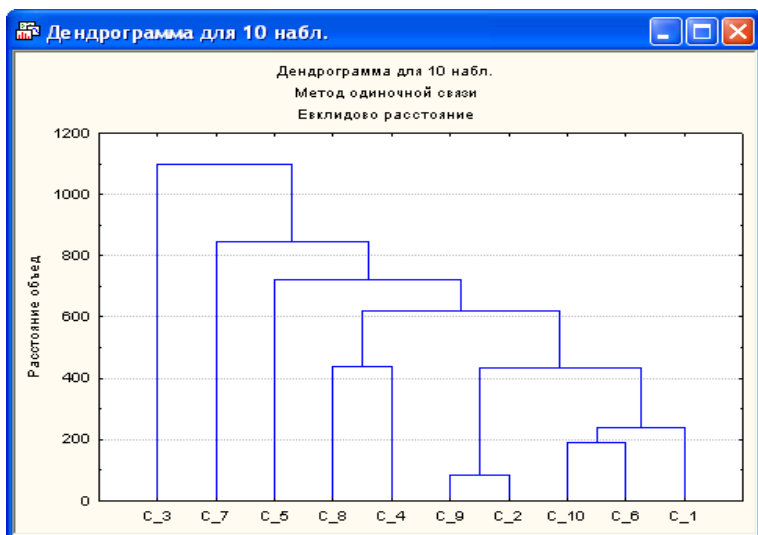


Рисунок 30 – Результаты древовидной кластеризации

Осуществим группировку методом k -средних, выполнив принудительную группировку на 3 кластера. Для этого активируем диалоговое окно *Кластеризация методом k -средних* (рисунок 31).

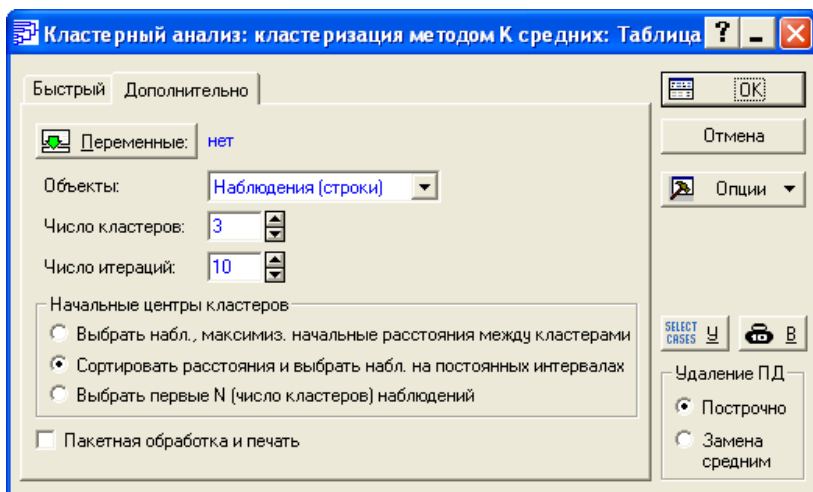


Рисунок 31 – Диалоговое окно *Кластеризация методом k -средних* процедуры *Кластерный анализ*

Вновь вводим переменные 1–3 (товарооборот, прибыль и численность работников), затем отмечаем, что кластеров в нашем примере должно быть 3, и нажимаем кнопку *OK*.

Как и в предыдущем случае, возникает окно *Результаты метода k-средних* (рисунок 32), в котором предлагается выбрать форму получения результатов кластеризации.

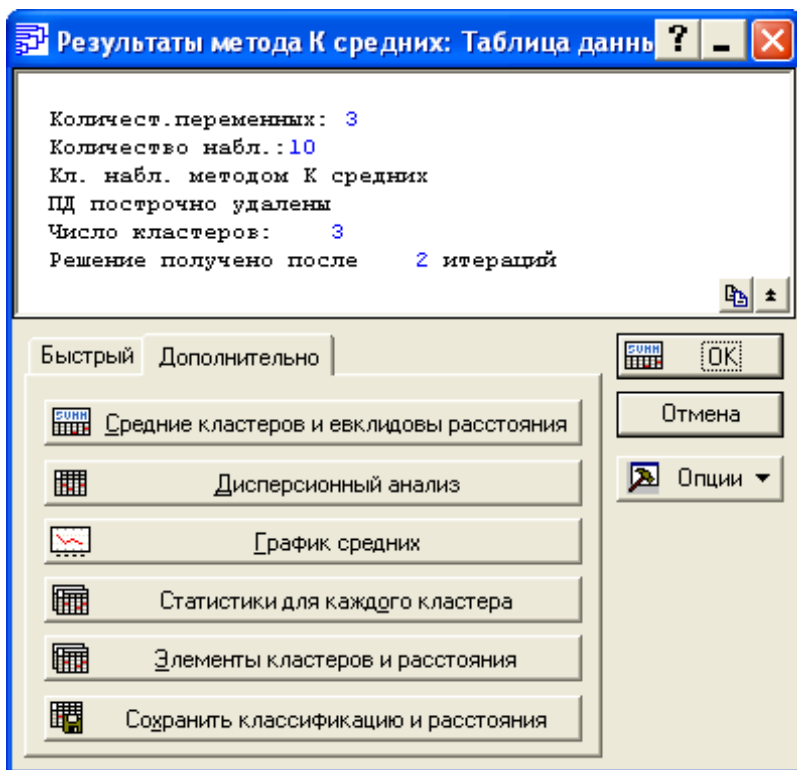


Рисунок 32 – Диалоговое окно *Результаты метода k-средних*

В нашем примере во вкладке *Дополнительно* выбираем *Элементы кластеров и расстояния*. Программа Statistica 6.0 отдельно по каждому из 3 кластеров выдает электронные таблицы «Данные: Элементы кластера...» (рисунок 33).

Данные: Элементы кластера номер 1 (Таблица данных1)*

Элементы кластера номер 1 (Таблица данных1) и расстояния до центра кластера. Кластер содержит 3 набл.						
Набл.Но.	Набл.Но.	Набл.Но.				
С 5	С 7	С 8				
Расст.	25,92725	464,1920	440,2623			

Данные: Элементы кластера номер 2 (Таблица данных1)*

Элементы кластера номер 2 (Таблица данных1) и расстояния до центра кластера. Кластер содержит 6 набл.						
Набл.Но.	Набл.Но.	Набл.Но.	Набл.Но.	Набл.Но.	Набл.Но.	
С 1	С 2	С 4	С 6	С 9	С 10	
Расст.	384,2592	153,5648	508,8954	248,2074	110,6814	139,6290

Данные: Элементы кластера номер 3 (Таблица данных1)*

Элементы кластера номер 3 (Таблица данных1) и расстояния до центра кластера. Кластер содержит 1 набл.						
Набл.Но.						
С 3						
Расст.	0,00					

Рисунок 33 – Результаты кластеризации методом k -средних (электронные таблицы)

Выводы по задаче. Данные рисунка 30 показывают, что 10 торговых предприятий можно сгруппировать в следующие кластеры: кластер 1 (предприятия 8 и 4), кластер 2 (предприятия 9 и 2), кластер 3 (предприятия 10, 6 и 1) и кластер 4 (предприятия 3, 7 и 5). Однако визуально это только условное деление.

Данные рисунка 33 показывают, что 10 торговых предприятий целесообразно сгруппировать в следующие кластеры: кластер 1 (предприятия 5, 7 и 8), кластер 2 (предприятия 1, 2, 4, 6, 9 и 10), кластер 3 (предприятие 3), для которых средние характеристики являются более значимыми.

СПИСОК РЕКОМЕНДУЕМОЙ ЛИТЕРАТУРЫ

1. **Вуколов, Э. А.** Основы статистического анализа. Практикум по статистическим методам и исследованию операций с использованием пакетов Statistica и Excel : учеб. пособие / Э. А. Вуколов. – 2-е изд., испр. и доп. – М. : ФОРУМ, 2011. – 464 с.

2. **Основы** статистической обработки информации : пособие / авт.-сост. : О. И. Еськова, Л. П. Авдашкова, М. А. Грибовская. – Минск : Беларусь, 2011. – 175 с.

3. **Продукты** // StatSoft Russia. Официальный сайт [Электронный ресурс]. – Режим доступа : http://statsoft.ru/products/STATISTICA_Base/. – Дата доступа : 11.05.2013.

СОДЕРЖАНИЕ

Пояснительная записка.....	3
1. Возможности системы Statistica 6.0 и особенности ее интерфейса.....	4
2. Краткая характеристика структуры программы Statistica 6.0.....	9
3. Отдельные документы программы Statistica 6.0.....	11
3.1. Рабочие книги	11
3.2. Электронные таблицы	12
3.3. Отчеты	15
3.4. Графики	16
4. Краткий обзор методов системы Statistica 6.0.....	17
4.1. Базовые модули.....	17
4.1.1. Основные статистики и таблицы	17
4.1.2. Множественная регрессия	24
4.2. Модуль «Углубленные методы анализа»	26
4.2.1. Общие линейные модели	26
4.2.2. Общие регрессионные модели	27
4.2.3. Анализ временных рядов	28
4.3. Модуль «Многомерный разведочный анализ»	33
4.3.1. Методы кластеризации.....	33
4.3.2. Факторный анализ и анализ главных компонент	34
4.3.3. Канонический анализ	36
5. Пошаговые примеры в программе Statistica 6.0.....	37
Пример 1. Корреляционный анализ.....	37
Пример 2. Регрессионный анализ.....	42
Пример 3. Кластерный анализ.....	45
Список рекомендуемой литературы	50

Учебное издание

**ВОЗМОЖНОСТИ СИСТЕМЫ
STATISTICA 6.0**

Пособие

**для реализации содержания образовательных программ
высшего образования I и II ступеней**

Авторы-составители:

Лацкевич Наталья Васильевна

Степаненко Дмитрий Александрович

Редактор **Е. Г. Привалова**

Технический редактор **И. А. Козлова**

Компьютерная верстка **Л. Г. Макарова**

Подписано в печать 22.05.14. Формат 60 × 84 ¹/₁₆.
Бумага типографская № 1. Гарнитура Таймс. Ризография.
Усл. печ. л. 3,02. Уч.-изд. л. 3,20. Тираж 130 экз.
Заказ № 30-05-14.

Издатель и полиграфическое исполнение:
учреждение образования «Белорусский торгово-экономический
университет потребительской кооперации».

Свидетельство о государственной регистрации издателя,
изготовителя, распространителя печатных изданий
№ 1/138 от 08.01.2014.

Просп. Октября, 50, 246029, Гомель.

**БЕЛКООПСОЮЗ
УЧРЕЖДЕНИЕ ОБРАЗОВАНИЯ
«БЕЛОРУССКИЙ ТОРГОВО-ЭКОНОМИЧЕСКИЙ
УНИВЕРСИТЕТ ПОТРЕБИТЕЛЬСКОЙ КООПЕРАЦИИ»**

Кафедра сферы услуг и статистики

**ВОЗМОЖНОСТИ СИСТЕМЫ
STATISTICA 6.0**

**Пособие
для реализации содержания образовательных программ
высшего образования I и II ступеней**

Гомель 2014