



**Федеральное государственное образовательное  
учреждение высшего профессионального образования  
«ФИНАНСОВЫЙ УНИВЕРСИТЕТ  
ПРИ ПРАВИТЕЛЬСТВЕ РОССИЙСКОЙ ФЕДЕРАЦИИ»  
(Финансовый университет)**

Кафедра «Информационные технологии»

А.В.Золотарюк

**ПРОФЕССИОНАЛЬНЫЕ КОМПЬЮТЕРНЫЕ  
ПРОГРАММЫ:  
ПРИМЕНЕНИЕ ПАКЕТА SPSS В СОЦИОЛОГИИ**

Комплексные задания для самостоятельной работы

Для студентов, обучающихся по направлению 040100.62 «Социология»

Москва, 2013

**Федеральное государственное образовательное учреждение  
высшего профессионального образования  
«ФИНАНСОВЫЙ УНИВЕРСИТЕТ  
ПРИ ПРАВИТЕЛЬСТВЕ РОССИЙСКОЙ ФЕДЕРАЦИИ»  
(Финансовый университет)**

**Кафедра «Информационные технологии»**

**А.В.Золотарюк**

**ПРОФЕССИОНАЛЬНЫЕ КОМПЬЮТЕРНЫЕ  
ПРОГРАММЫ:  
ПРИМЕНЕНИЕ ПАКЕТА SPSS В СОЦИОЛОГИИ**

**Комплексные задания для самостоятельной работы**

**Для студентов, обучающихся по направлению 040100.62 «Социология»**

*Одобрено рабочей группой кафедры  
«Информационные технологии»*

**Москва, 2013**

## ОГЛАВЛЕНИЕ

<b>ВВЕДЕНИЕ.....</b>	<b>3</b>
<b>1. ПАКЕТ СТАТИСТИЧЕСКОЙ ОБРАБОТКИ ДАННЫХ SPSS.....</b>	<b>5</b>
1.1.    Общая информация о пакете SPSS и его структуре .....	5
1.2.    Задачи и ключевые возможности поставок пакета SPSS .....	8
1.3.    Схема обработки информации в SPSS .....	15
1.4.    Назначение и общие возможности модулей IBM SPSS.....	17
1.5.    Направления практического использования SPSS.....	21
1.6.    Работа в SPSS.....	25
<b>2. ПРИМЕР РЕШЕНИЯ ЗАДАЧИ В SPSS МЕТОДОМ КЛАСТЕРНОГО АНАЛИЗА.....</b>	<b>40</b>
<b>3. КОМПЛЕКСНОЕ ЗАДАНИЕ НА ОСВОЕНИЕ ТЕХНОЛОГИИ ИСПОЛЬЗОВАНИЯ</b>	
<b>ПАКЕТА SPSS .....</b>	<b>46</b>
3.1.    Общие сведения .....	46
3.2.    Исходные данные для многомерного статистического анализа .....	47
2.3.    Методы исследования и защита работы.....	49
<b>ЛИТЕРАТУРА .....</b>	<b>50</b>

## ВВЕДЕНИЕ

Знание тренда общественных настроений, вызванных различными социально-экономическими процессами, позволяет за счет компетентных решений руководителей нивелировать негативные явления, поддерживать и развивать положительное, что способствует развитию государства, укреплению его экономической и военно-политической мощи, обеспечивает рост народного благосостояния, в первую очередь особо нуждающихся.

Наличие информации, собираемой в ходе исследований общественного мнения, недостаточно для выработки позитивных управленческих воздействий. Необходима ее научно-обоснованная статистическая обработка с целью получения обобщенных результатов, извлечения из массивов исходных данных нетривиальных, явно не просматривающихся знаний, представления итоговых показателей в наглядном графическом виде.

Социологи для обработки накопленных данных применяют пакет SPSS (IBM SPSS Statistics) либо пакет Statistica. Первый пакет является более предпочтительным, так как он, по своей сути, ориентирован, прежде всего, на социологию.

Рабочая программа учебной дисциплины «Профессиональные компьютерные программы», входящей в вариативную часть обязательных дисциплин математического цикла Федерального государственного образовательного стандарта высшего профессионального образования по направлению 040100.62 «Социология», предусматривает изучение и того, и другого пакетов как профессионально-ориентированных компьютерных систем, способствующих решению прикладных задач социологической сферы деятельности.

К сожалению, в ходе практических занятий в вузе, осваивать технологию работы в SPSS часто не представляется возможным из-за отсутствия пакета в перечне установленных лицензионных программ корпоративной сети образовательного учреждения. Причинами такой ситуации являются, прежде всего, очень высокая стоимость лицензий на использование SPSS, а также затруднения в установке бесплатной, ограниченной по срокам использования, ознакомительной версии пакета в нужное время, в соответствии с расписанием учебных занятий, на часть компьютеров разветвленной сети вуза.

Решением проблемы было бы наличие облачной версии пакета, но, к сожалению, фирма IBM, как разработчик SPSS, пока не планирует переход на облачные технологии.

Таким образом, выход из создавшейся ситуации может быть только один: перенести практическое освоение пакета SPSS на самостоятельную работу студентов, предоставив им все методические и технологические материалы по скачиванию и установке ознакомительной версии пакета на личный компьютер и его использованию в ходе решения внеаудиторной лабораторной работы по тематике предметной области социолога.

В настоящих комплексных заданиях описывается общая характеристика пакета SPSS и его применение. Приводится тематика исследовательской внеаудиторной лабораторной работы, требования к ее выполнению, оформлению и публичной защите. Даны массивы исходных данных для каждого из вариантов.

Выполнение заданий требует от студентов проявления профессионально-ориентированных компетенций и знаний, приобретенных в процессе обучения в вузе, а также принятия самостоятельного решения на каждом из технологических этапов работы, выбора наиболее приемлемого пути продолжения аналитического исследования.

# 1. ПАКЕТ СТАТИСТИЧЕСКОЙ ОБРАБОТКИ ДАННЫХ SPSS

## 1.1. Общая информация о пакете SPSS и его структуре

Пакет SPSS (Statistical Package for the Social Sciences — статистический пакет для социальных наук) изначально<sup>1</sup> был разработан как компьютерная программа для статистической обработки данных, предназначенная для проведения прикладных исследований в социальных науках.

В 2009 г. пакет стал называться PASW Statistics (Predictive Analytics SoftWare – интеллектуальное аналитическое программное обеспечение).

С июля 2009 г. пакет сопровождается фирмой IBM (International Business Machines) под именем IBM SPSS Statistics.

В 2013 г. планируется выход очередной версии пакета - IBM SPSS Statistics 22, работающей под управлением различных операционных систем - Windows, MacOSX, Linux.

Имея модульную структуру, пакет SPSS обеспечивает комплексную статистическую обработку – от планирования до управления данными, выполнения анализа и представления результатов. Мощные средства анализа и обработки данных с развитым графическим интерфейсом, удобные меню и простые диалоговые окна существенно упрощают работу пользователя.

Основные блоки SPSS:

- *редактор данных* - гибкая система, внешне похожая на электронную таблицу, служит для определения, ввода, редактирования и просмотра данных;
- *визуализатор* - средство просмотра, упрощающее отображение результатов, позволяющее показывать и скрывать отдельные элементы вывода, изменять порядок вывода результатов, перемещать готовые к презентации таблицы и диаграммы в другие приложения и получать их из других программных приложений;
- *многомерные мобильные таблицы* – служат для отображения результатов анализа, позволяя исследовать таблицы, перемещая строки, столбцы и слои и, таким образом, выявлять важные моменты, которые могут потеряться в стандартных отчетах. Обеспечивают также сравнение групп, расщепляя таблицы таким

---

<sup>1</sup> Первоначально пакет SPSS был разработан в 1968 г. в Чикагском университете; разработчики пакета: Норман Най, Хедли Халл, Дейл Бент

образом, чтобы каждый раз на экран выводилась только одна группа;

- *высококачественная графика* – средство формирования полноцветных диаграмм с высоким разрешением: круговых и линейчатых диаграмм, гистограмм, диаграмм рассеяния, объемных диаграмм и многих-многих других;
- *доступ к базам данных* - конструктор чтения баз данных, позволяющий легко, несколькими нажатиями кнопки мыши, загружать данные из любых источников;
- *преобразование данных* - средство преобразования данных, помогает готовить данные к анализу: можно с легкостью выделять подмножества в данных, объединять категории, добавлять, агрегировать, сливать, расщеплять и транспонировать файлы, а также проводить другие преобразования;
- *Справочная система* – комплекс средств поддержки начинающего пользователя, включающий:
  - *электронный учебник*, содержащий детальный обзор средств и возможностей пакета;
  - *контекстную справку*, помогающую разобраться в конкретных задачах при работе с диалоговыми окнами;
  - *всплывающие определения*, объясняющие статистические термины в мобильных таблицах;
  - *репетитор по статистике*, помогающий в поиске необходимой процедуры обработки;
  - *примеры анализа*, облегчающие интерпретировать результаты в аналогичных типовых задачах;
- *командный язык*, обеспечивающий доступ к дополнительным функциональным возможностям пакета, недоступным через меню и диалоговые окна, и позволяющий сохранять и автоматизировать различные повторяющиеся процессы и задачи. Полная документация по командному языку интегрирована в справочную систему и доступна в виде отдельного PDF-документа «Руководство по синтаксису», который можно вызвать в меню команды «Справка».

Предусмотрены различные варианты поставки продуктов IBM SPSS Statistics:

- IBM SPSS Statistics Standart – включает основные аналитические возможности для решения широкого спектра хозяйственных и исследовательских задач;
- IBM SPSS Statistics Professional – дополнительно содержит средства, связанные с обеспечением качества данных и их полноты, а также автоматизации функций статистики и прогнозирования;
- IBM SPSS Statistics Premium – включает полный набор аналитических методик, систему моделирования на основе структурных уравнений (SEM), средства подробной оценки и проверки выборочных данных, процедуры прямого маркетинга.
- IBM SPSS Statistics for Education (только на английском языке) – включает основные модули, популярные в учебном процессе (для студентов, сотрудников и преподавателей).

В настоящее время пакет представляет собой полнофункциональную исследовательскую систему для статистической обработки и анализа разноплановой информации - бизнес-структурами, представителями государственных органов, научными подразделениями, академическими организациями и вузами. Результаты анализа можно наглядно представлять в виде таблиц и диаграмм различных типов, распространять сетевым пользователям, внедрять полученные результаты в другие программные системы.

Последние версии пакета обладают мощными аналитическими инструментами и средствами ускорения работы, позволяют проще и быстрее выявлять ценную информацию в данных, делать более точные прогнозы и обеспечивать большую отдачу от аналитики при решении прикладных задач. Они содержат все современные средства импорта-экспорта данных, предобработки, реструктуризации и визуальной категоризации данных. Обеспечиваются широкие возможности имитационного моделирования данных с заданными характеристиками методом Монте-Карло (в т.ч., с характеристиками фактических данных, формой распределения и связями между переменными). Тем самым предоставляется возможность специалистам, связанным с планированием и моделированием затрат и рисков, надежнее оценивать диапазон возможных значений прогнозируемого показателя и оценивать его отклик при воздействии на управляемые переменные.



## 1.2. Задачи и ключевые возможности поставок пакета SPSS

### Профессиональная поставка пакета SPSS Statistics Professional

обеспечивает решение всех задач *полного жизненного цикла углубленного, нестандартного анализа* – качества данных, их сложности, автоматизации и прогнозирования.

Ключевые возможности профессиональной поставки реализуются множеством инструментальных средств и процедур.

1. **Линейные модели** включают разнообразные процедуры регрессионного и расширенного статистического анализа, спроектированные с учетом присущих характеристик составных отношений описания данных. Они предполагают применение:
  - обобщенных линейных смешанных моделей (GLMM) для работы с иерархическими данными;
  - процедур общих линейных моделей (GLM) и смешанных моделей;
  - обобщенных линейных моделей (GENLIN):
    - линейной регрессии для ответов с нормальным распределением;
    - логарифмических моделей для двоичных данных;
    - логарифмически-линейных моделей для данных подсчета;
    - реализации других полезных статистических моделей с использованием самых общих формулировок;
  - процедур обобщенных оценочных уравнений (GEE), расширяющих обобщенные линейные модели для возможности обработки связанных осевых и сгруппированных данных.
2. **Нелинейные модели** предоставляют возможность применять к данным более сложные модели. Они обеспечивают использование:
  - полиномиальной логарифмической регрессии (MLR) для прогнозирования категориальных исходов с числом категорий больше двух;
  - двоичной логарифмической регрессии для разбивки данных на две группы;
  - нелинейной регрессии (NLR) и ограниченной нелинейной регрессии (CNLR) для оценки параметров нелинейных моделей;
  - пробит-анализа для оценки значения стимулов.

3. **Настраиваемые таблицы** позволяют пользователям легко определять суть данных и быстро выводить обзор результатов в разных стилях для разных потребителей:

- при наличии дедуктивной статистики можно сравнивать средние значения и соотношения для демографических групп, сегментов потребителей или других категорийных переменных;
- обеспечивается формирование итоговой статистики (от простых подсчетов категорийных переменных до величин дисперсии) и сортировка категорий по любому из используемых итогов статистики;
- предусмотрено использование трех тестов значимости:
  - проверка независимости по критерию хи-квадрат;
  - сравнение средних значений столбцов (проверка по критерию Стьюдента);
  - сравнение соотношений столбцов (проверка по z-критерию);
- интерактивный редактор таблиц позволяет:
  - создавать сводные таблицы с помощью мыши;
  - исключать определенные категории;
  - показывать ячейки с отсутствующими значениями;
  - добавлять в таблицы промежуточные итоги;
- таблицы можно просматривать в реальном режиме времени и изменять по мере создания;
- таблицы можно экспортировать в документы Microsoft Word, Excel, PowerPoint или HTML для использования в отчетах.

4. **Подготовка данных** рационализирует этап подготовки данных аналитического процесса:

- позволяет определить подозрительные или недопустимые варианты, переменные и значения данных;
- позволяет просматривать шаблоны отсутствующих данных и создавать обзоры распределений переменных;
- функция *Optimal Binning* позволяет найти наилучший исход из возможных для алгоритмов, предназначенных для номинальных атрибутов;

- инструмент *Автоматической подготовки данных* (ADP) выявляет и устраняет ошибки качества и условно подставляет отсутствующие значения за один эффективно выполняемый этап;
  - рекомендации и визуализации помогают определить, какие данные следует использовать.
5. ***Значения, показывающие правильность или отсутствие данных*** повышают вероятность получения статистически значимых результатов:
- данные проверяются с нескольких разных точек зрения с помощью одного из шести отчетов диагностики, затем оценивается итоговая статистика и подставляются отсутствующие значения;
  - выполняется диагностика серьезных неполадок подстановки отсутствующих данных;
  - заменяются отсутствующие значения приблизительными;
  - отображается моментальная копия для каждого типа отсутствующего значения и крайние значения для каждого варианта;
  - скрытое отклонение убирается путем замены отсутствующих значений приблизительными; это делается для того, чтобы включить все группы, даже с низким уровнем реагирования.
6. ***Деревья решений*** упрощают определение групп, выявление отношений между группами и прогнозирование будущих событий:
- пакет визуально определяет, как выполняется поток модели, чтобы можно было найти отдельные подгруппы и отношения;
  - обеспечивается создание деревьев классификации непосредственно в пакете, чтобы можно было использовать результаты для сегментации и группировки вариантов непосредственно в данных;
  - правила выбора или классификации/прогнозирования создаются в синтаксисе IBM SPSS Statistics, в операторах SQL или простом тексте (прямой синтаксис).
  - Обеспечивается выбор одного из четырех алгоритмов формирования дерева:
    - CHAID — быстрого статистического алгоритма, который просматривает данные быстро и эффективно, создавая сегменты и профайлы с учетом предпочитаемого исхода;

- исчерпывающего CHAID — модификации CHAID, проверяющей все возможные расщепления для каждой независимой переменной;
- деревьев классификации и регрессии (C&RT) — полноценного двоичного алгоритма деревьев, разделяющего данные и создающего правильные однородные подмножества;
- QUEST — статистического алгоритма, позволяющего выбрать переменные без отклонений и быстро и эффективно построить правильные двоичные деревья.

7. **Функции прогнозирования** позволяют быстрее анализировать данные хронологии и выявлять тенденции:

- позволяют доставлять информацию способом, более подходящим для ответственных руководителей организации, работающих с этой информацией;
- автоматически определяют наиболее подходящий алгоритм (модель) экспоненциального сглаживания для анализа данных хронологии;
- вместо создания переменных по одной позволяют одновременно моделировать сотни рядов с различным временем;
- модели сохраняются в центральном файле: при изменении данных прогнозы обновляются без необходимости повторного задания параметров или повторной оценки моделей;
- можно создавать сценарии, позволяющие автоматически обновлять модели путем добавления новых данных.

**Стандартная поставка пакета SPSS Statistics Standard** содержит основные аналитические возможности для широкого спектра хозяйственных и исследовательских задач и позволяет просматривать данные, формулировать гипотезы для дополнительного тестирования и выполнять процедуры для выявления отношений между переменными, создания кластеров, определения тенденций и составления прогнозов.

Ключевые возможности стандартной поставки включают часть инструментальных средств и процедур профессиональной поставки.

1. **Линейные модели.**
2. **Нелинейные модели.**
3. **Настраиваемые таблицы.**

**Расширенная поставка пакета SPSS Statistics Premium** позволяет использовать дополнительные средства, будучи ориентированной на аналитиков, специалистов по планированию, прогнозированию, анализу обзоров, оценке программ, маркетингу баз данных и пр. Включает большое число полностью интегрируемых функций статистики и связанных продуктов для специализированных аналитических задач, выполняемых разными подразделениями предприятия. Позволяет значительно повысить производительность и добиться высоких результатов в ходе реализации конкретных проектов и бизнес-целей.

Ключевые возможности расширенной поставки включают часть инструментальных средств и процедур профессиональной поставки и новые возможности.

***1. Линейные модели.***

***2. Нелинейные модели.***

***3. Настраиваемые таблицы.***

***4. Подготовка данных.***

***5. Достоверность данных и отсутствующие значения.***

***6. Категориальные и числовые данные:***

- выявляются базовые отношения с помощью карт позиционирования, точечных диаграмм биплот (biplot) и трилинейных точечных диаграмм;
- используя общепринятые процедуры обработки, облегчается работа с номинальными (напр., зарплата) или порядковыми (напр., уровень образования) данными, делая их более понятными;
- визуально интерпретируются наборы данных указанием на взаимосвязь между строками и столбцами в больших таблицах баллов, подсчетов, рейтингов, приоритетов или подобий;
- обрабатываются отличные от нормальных остатки в числовых данных или нелинейные отношения между независимыми переменными (напр., атрибуты потребитель или продукт) и переменной исхода (напр., покупка/без покупки).
- реализуются приемы гребневой регрессии, лассо, эластической сети, выбора переменной и выбора модели как для числовых, так и для категориальных данных.

***7. Древоидные схемы принятия решения***

***8. Функции прогнозирования***

### **9. Моделирование на основе структурных уравнений:**

- проверяются гипотезы и подтверждаются отношения между наблюдаемыми или ненаблюдаемыми переменными, не останавливаясь на регрессии при выявлении дополнительной статистики;
- обеспечивается создание моделей, более реалистично отражающих отношения, поскольку числовые переменные, как наблюдаемые (напр., данные опроса, полученные не экспериментальным путем), так и ненаблюдаемые (напр., удовлетворение и приверженность), могут использоваться для прогнозирования любых других числовых переменных;
- поддерживается визуальное сравнение, подтверждение и уточнение модели;
- многомерный анализ реализуется на расширении стандартных методов исследования - *регрессии, факторного анализа, корреляции и дисперсионного анализа*;
- используется три метода вычисления данных для подстановки: *регрессия, стохастическая регрессия и байесовский подход*.

### **10. Применение начальной загрузки:**

- определяется оценка выборочного распределения оценочной функции путем повторного взятия выборки с заменой из исходной выборки;
- вычисляется оценка стандартных ошибок и доверительных интервалов параметра генеральной совокупности (таких как среднее, медиана, соотношение, отношение шансов, коэффициент корреляции, коэффициент регрессии и др.);
- создаются многовариантные наборы данных для более точного анализа;
- упрощается тестирование стабильности и надежности моделей.

### **11. Дополнительная оценка выборки и тестирование:**

- предусмотрены специализированные инструменты планирования, статистики, необходимые для работы с планами сложной выборки (при многослойной, кластерной или многоэтапной выборке);
- обеспечивается получение улучшенных результатов за счет входа в состав обзорного анализа выборочного плана;

- реализована более точная работа с числовыми и категориальными исходами в планах сложной выборки с использованием алгоритмов анализа и прогнозирования, в т.ч. времени события;
- предусмотрено использование Мастеров при создании планов, анализе данных и интерпретации результатов.

#### **12. Инструменты прямого маркетинга и принятия решений по продукту:**

- сегментация потребителей или контактов путем создания групп;
- профилирование потребителей или контактов с общими характеристиками (для улучшения выбора целевой аудитории при маркетинговых исследованиях);
- выявление уровней предрасположенности (при определении склонных к покупке потребителей);
- определение ответов по почтовому индексу;
- интеграция ответов потребителей.

#### **13. Высококачественные диаграммы и графики:**

- десятки встроенных шаблонов визуализации для отображения результатов анализа;
- создание графиков не требует навыков программирования;
- настройка таблиц стилей и шаблонов графиков;
- отображение графиков в различных ОС;
- поддержка разнообразных источников данных (специальных и общих).

Как видно, поставка Premium имеет существенно более широкие аналитические возможности по сравнению с профессиональной версией пакета.

**Учебная поставка пакета SPSS Statistics for Education** - специальная версия для учебных заведений. Поставляется в ВУЗы, прежде всего США, при заключении соответствующих договоров со значительной скидкой.

Версия выходит только на английском языке.

Возможности версии ограничены.

### 1.3. Схема обработки информации в SPSS

Использование пакета SPSS предусматривает выполнение следующих функциональных задач:

- оперативное получение аналитической информации – ключевых факторов, взаимосвязей и тенденций в исследуемых данных;
- наглядное представление результатов в виде таблиц и диаграмм;
- распространение полученных результатов сетевым пользователям – заказчикам проведенных исследований;
- поддержка процессов оперативного формирования научно-обоснованных оптимальных управленческих решений.

При этом пакет SPSS реализует семь этапов аналитического процесса.

1. Планирование.
2. Сбор данных.
3. Обеспечение доступа к данным.
4. Подготовка данных к анализу.
5. Выполнение анализа.
6. Формирование отчетов.
7. Представление и распространение результатов.

При необходимости некоторые этапы можно пропускать (если они ранее были выполнены) или повторять несколько раз (для получения дополнительных данных).

Графическое представление реализации этапов аналитического исследования представлено на рис. 1.





Рис. 1. Роль функциональных модулей пакета в процессе реализации этапов аналитического исследования

## 1.4. Назначение и общие возможности модулей IBM SPSS

Семейство программных продуктов IBM SPSS – это многомодульный, полностью интегрированный комплекс программ, охватывающий все этапы аналитического процесса – от планирования и сбора данных до анализа, создания отчетов и распространения результатов.

Все программы комплекса имеют интуитивно понятный интерфейс и включают в себя многочисленные функции и процедуры управления данными, реализующие современные статистические методы и алгоритмы, а также средства создания отчетов, что обеспечивает проведение анализа любой степени сложности.

Использование комплекса обеспечивает решение бизнес-проблем и исследовательских задач при помощи статистического программного обеспечения.

Достигается возможность сбора и оперативной обработки массивов информации и получение аналитических сведений, в т.ч. с представлением результатов в виде наглядных таблиц и диаграмм.

Обеспечивается выявление ключевых факторов, взаимосвязей, тенденций, скрытых закономерностей в исходных данных.

Как результат: руководителям предоставляется научно-обоснованная информация, что позволяет своевременно принимать оптимальные управляющие воздействия.

Семейство SPSS представляет собой комплекс четырех взаимосвязанных инструментальных блоков (модулей, программных комплексов, систем, платформ) для прогностической аналитики и анализа данных (рис. 2), использование которых, например, позволяет:

- Собрать всю необходимую информацию о вкусах и мнениях потребителей.
- Спрогнозировать реакцию клиентов на специальные предложения до того, как они осуществляются.
- Влиять на результаты деятельности компании за счет внедрения прогностической аналитики в бизнес-процессы.

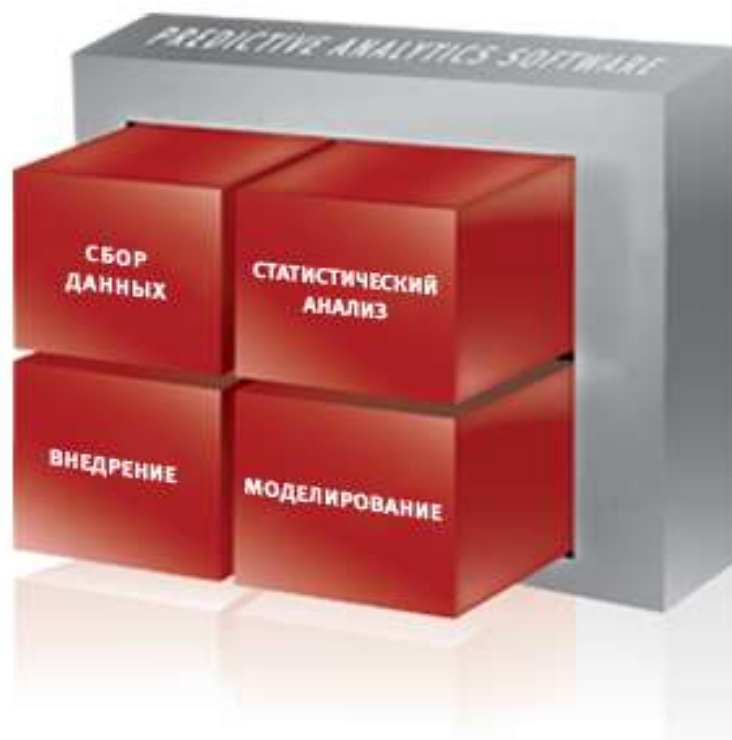


Рис. 2

**IBM SPSS Data Collection - Сбор данных** является программной платформой, позволяющей проводить опросы и маркетинговые исследования наиболее эффективными способами:

- по телефону,
- с помощью портативных компьютеров,
- через сеть Интернет.

**IBM SPSS Statistics - Статистический анализ** - полнофункциональная система, предназначенная для решения бизнес- и исследовательских задач при помощи анализа данных, позволяющая наглядно представлять результаты в виде таблиц и диаграмм, а также распространять и внедрять полученные результаты.

**IBM SPSS Modeler – Моделирование** - программный комплекс, позволяющий принимать обоснованные бизнес-решения, опираясь на надежные модели. Является признанным инструментом Data Mining – извлечения и добычи знаний, позволяет обнаруживать скрытые закономерности, предсказывать и оценивать возможные результаты альтернативных вариантов действий.

**IBM SPSS Collaboration and Deployment Services - Интеграция и внедрение в бизнес-процессы** - семейство продуктов, предназначенное для быстрого и надежного внедрения прогностической аналитики во все

процессы принятия решений, позволяющее соединить прогностическую аналитику и бизнес.

#### **Основные возможности комплекса программ**

- Моделирование структурными уравнениями для поиска скрытых взаимосвязей в данных.
- Определение корректного объема выборки на этапе планирования исследования (размера выборки, соответствующего бюджету и решаемым задачам).
- Распространение отчетов, созданных в SPSS для Windows.
- Распространение в Интернете интерактивных отчетов, созданных в SPSS для Windows .
- Анализ сложных взаимосвязей в данных (методы одномерного и многомерного анализа).
- SPSS Statistics Base — доступ к данным, управление данными, подготовка данных к анализу, анализ данных и создание отчетов; дополнительные модули встраиваются в SPSS Statistics Base и позволяют расширять аналитические возможности ПО.
- Прогнозирование категориальных данных (откликов) и их исследование при помощи карт восприятия (анализ громоздких двух- и многоходовых таблиц, визуализация и исследование категориальных данных и т.п.).
- Построение деревьев классификаций и решений, идентифицирование групп, нахождение взаимосвязи в данных, предсказывание будущих событий.
- Разработка сложных выборок и учет сложности выборки в процессе анализа (детализация приводится далее).
- Определение предпочтений клиентов.
- Инструменты для организации ввода и чистки данных.
- Централизация сбора и ввода данных.
- Рабочее место оператора ввода и чистки данных.
- Точный анализ малых выборок и разрезов.
- Упрощение процесса проверки данных, устранение интенсивной ручной проверки, получение более точных результатов.
- Корректная импутация пропущенных значений (заполнение пропущенных значений для повышения информативности данных и построения адекватных моделей).

- Улучшение качества предсказаний при помощи мощных регрессионных процедур (моделей прогнозирования).
- Анализ централизованно хранящихся данных большого объема.
- Эффективное представление результатов.
- Прогнозирование временных рядов – мощного инструмента для анализа хронологической информации, построения моделей и выявления тенденций.
- Быстрый и полный анализ ответов на открытые вопросы опросов и обследований.

Работа в среде пакета SPSS основывается на применении Конструкторов. Использование Конструктора для обеспечения корректного анализа данных обследования приведено на рис. 3.



Рис. 3

Предусмотрены следующие виды планов отбора:

- *Расслоенный отбор* – повышает точность результатов, полученных по выборочным данным, гарантирует репрезентативность выборки по ключевым группам за счет выборки из подгрупп обследуемой совокупности.

Подгруппами, например, могут быть мужчины и женщины или люди определенных профессий, возрастных групп и т.п.

- *Кластерный отбор* – извлекаются кластеры, представляющие собой группы отбираемых для обследования единиц.

Кластерами могут быть школы (вузы), больницы или географические районы, при этом отбираемыми единицами являются студенты, пациенты или проживающие на данной территории граждане. Кластерный отбор часто помогает снизить стоимость проведения обследования.

- *Многоэтапный отбор* – вначале, на первом этапе, извлекается выборка групп элементов генеральной совокупности; затем формируется выборка второго этапа путем извлечения подвыборки в каждой единице, отобранной на первом этапе.

Используя такую возможность многократно, можно сформировать выборку с большим числом этапов отбора. Например, в обследованиях населения можно формировать выборку членов домашних хозяйств, которые в свою очередь отбираются в районах города.

## **1.5. Направления практического использования SPSS**

**Анализ информации о клиентах (CRM, управление взаимоотношениями с клиентами):**

- Обеспечение лучшего понимания поведения клиентов при помощи построения прогностических профилей
- Прогнозирование различных вариантов поведения клиентов
- Максимизация отдачи от клиентов за счет прогнозирования их интересов при помощи прогностической аналитики (рис. 4).

Например, какие товары и услуги будут необходимы клиентам и, таким образом, максимизировать отдачу от клиентов.

## Архитектура аналитической системы максимизации отдачи от клиентов

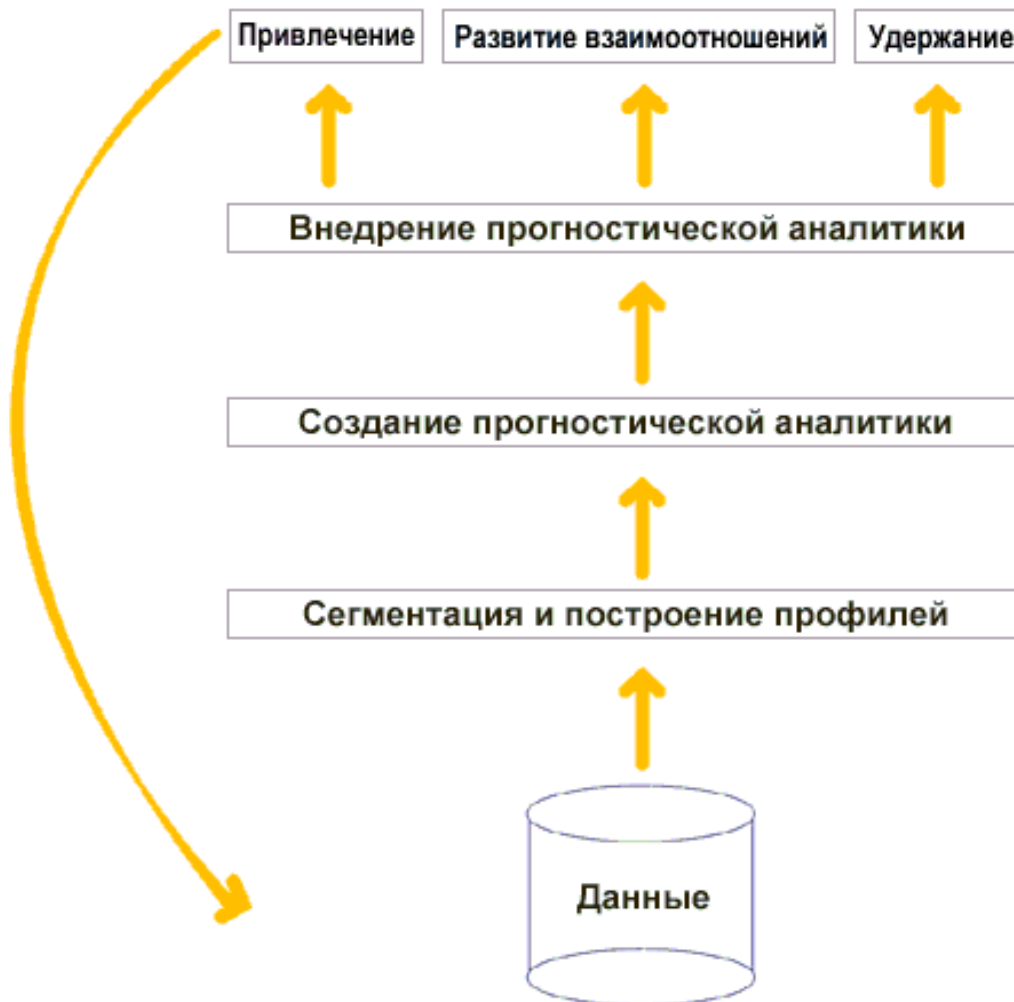


Рис. 4

**Проведение опросов и обследований дает оптимальные результаты на каждом из семи этапов проведения обследования и опросов (рис. 5).**

- Возможность принятия более нацеленных решений, основанных на надежной и достоверной информации, полученной благодаря опросам и обследованиям
- Маркетинговые исследования
- Анализ данных маркетинговых исследований



Рис. 5

### **Data mining (добыча знаний)**

- Решение ключевых бизнес-проблем при помощи современных аналитических технологий
- Обеспечение максимальной отдачи от полученных аналитических результатов за счет их внедрения



Data mining (добыча знаний) – это процесс выявления скрытых фактов и взаимосвязей в больших массивах данных.

Data mining разрешает парадокс: чем больше накапливается данных о клиентах, тем сложнее и дольше эффективно анализировать эти данные и получать значимые выводы. То, что, по идее, должно быть золотоносной жилой, обычно остается неисследованным из-за отсутствия подходящего персонала, недостатка времени и навыков.

**Использование пакета:**

**1) в государственном управлении**

- Аналитическая поддержка принятия управленческих решений
- Повышение эффективности здравоохранения и инфраструктуры
- Обнаружение и предотвращение случаев мошенничества, перерасхода и нецелевого использования средств
- Проведение опросов общественного мнения

**2) в здравоохранении**

- Анализ данных о пациентах
- Поддержка планирования курсов лечения
- Проведение медицинских обследований
- Оптимизация работы медицинских учреждений

**3) в образовательных учреждениях:**

- Обучение студентов прикладной статистике и практическим навыкам анализа данных
- Управление процессом приема/отбора студентов
- Тестирование студентов

**4) в финансовой сфере**

- Анализ финансовых рисков
- Выявление и предотвращение случаев мошенничества в финансовой сфере
- Прогнозирование финансовых показателей

## 1.6. Работа в SPSS

После запуска пакета SPSS следует:

Открыть существующий источник данных  
(например, файл Excel с расширением .xls):

*Переключатель Открыть существующий источник данных (рис.6)*

→ OK →...

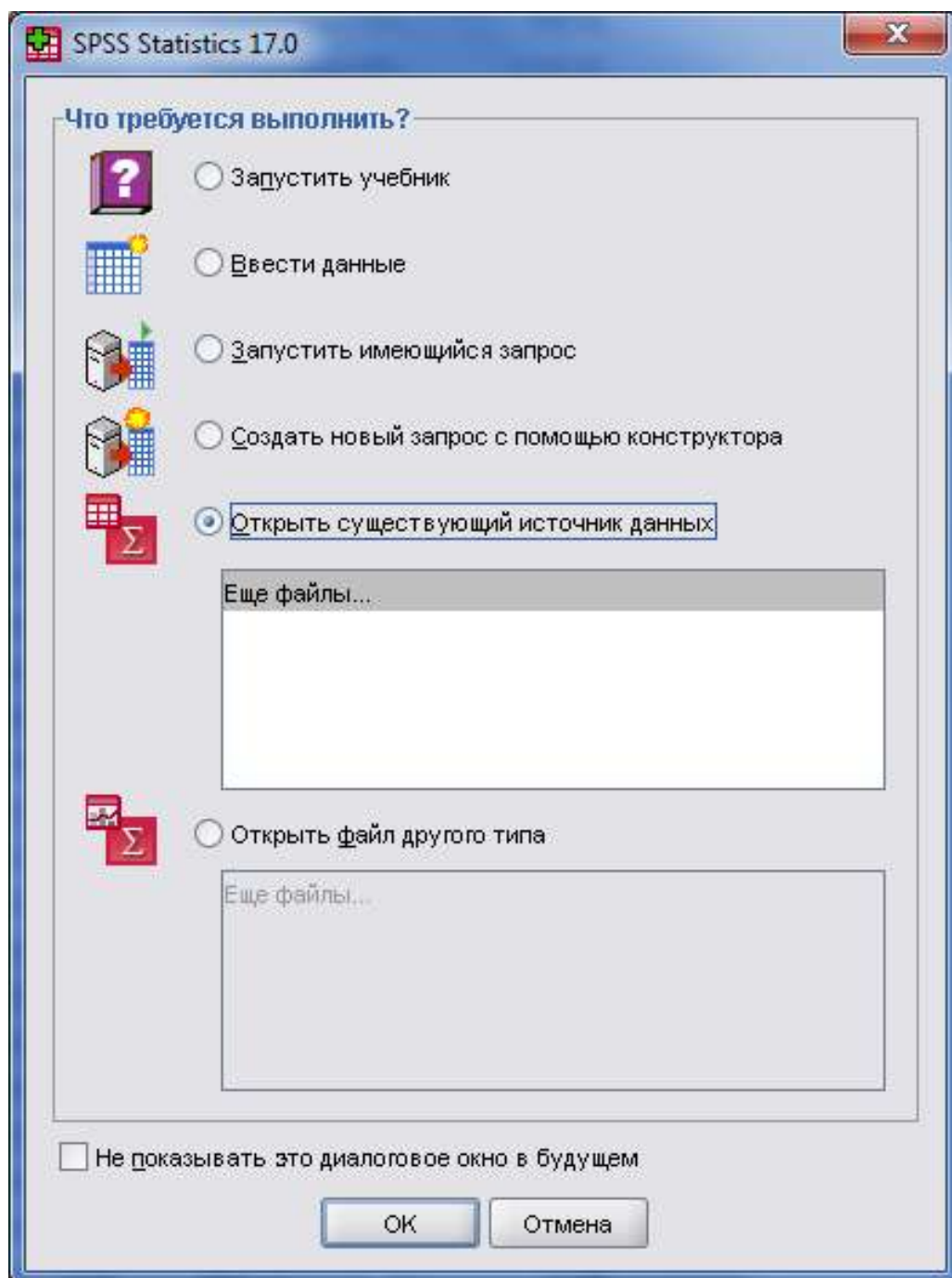
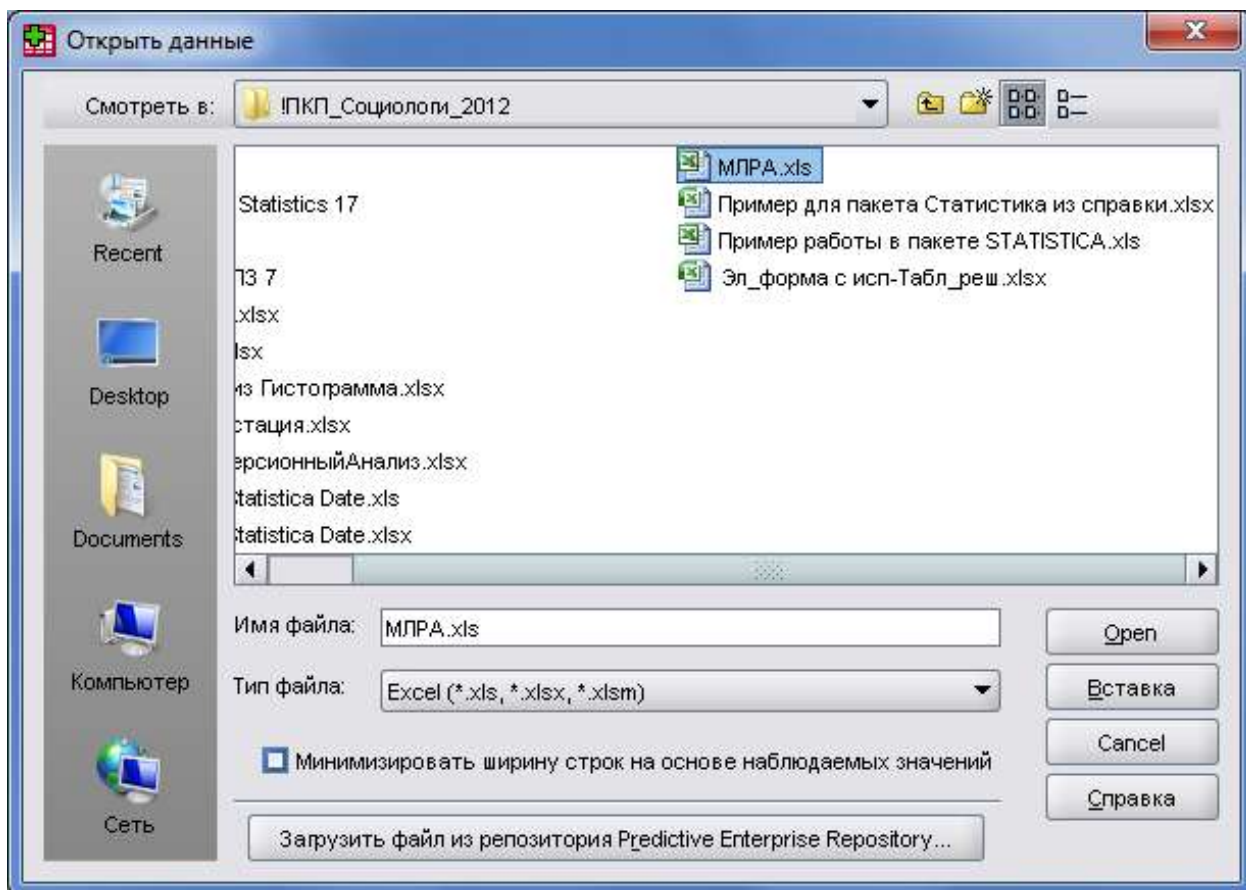
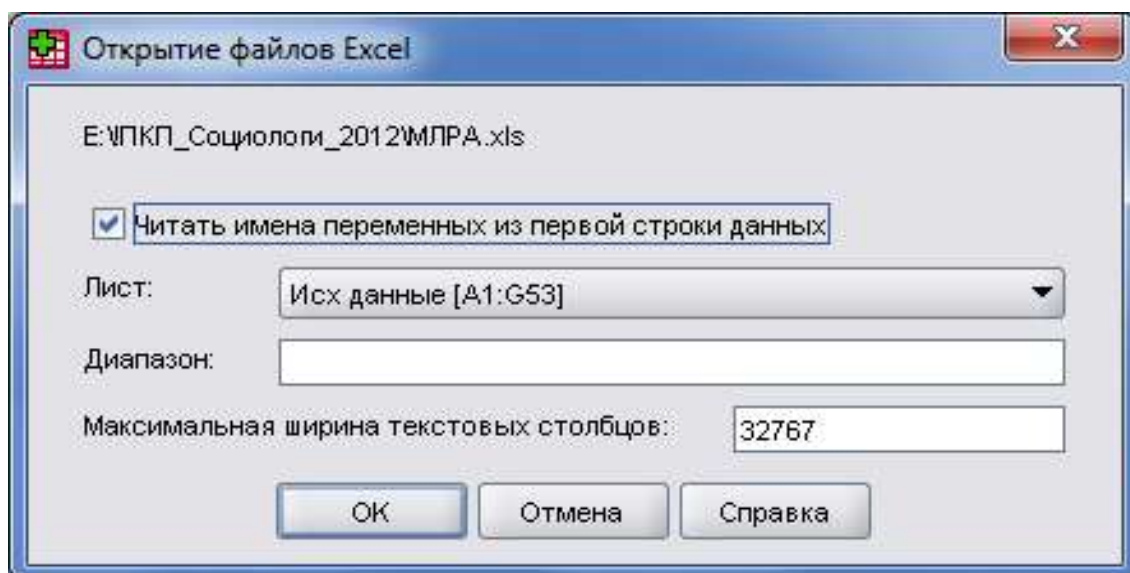


Рис. 6



**Рис. 7**

*Задать тип файла (рис. 7) → Найти и выделить файл →  
Open (Открыть) → ...*



**Рис. 8**

*Флажок Читать имена переменных из первой строки данных (рис. 8)  
→ OK → ...*

SPSS Statistics Data Editor window showing data entry for 21 countries. The 'Data' tab is active, displaying a table with 9 columns: Страна, Y, x1, x2, x3, x4, x5, and two unlabeled columns. The first row is highlighted, showing data for Australia (Австралия).

	Страна	Y	x1	x2	x3	x4	x5	
1	Австралия	74	1,9	16848	2,3	100	15	
2	Австрия	73	1,5	18396	94,0	99	12	
3	Аргентина	68	2,8	3408	12,0	95	20	
4	Бангладеш	53	4,7	202	800,0	35	35	
5	Беларусь	66	1,9	6500	50,0	99	13	
6	Бельгия	73	1,7	17912	329,0	99	12	
7	Бразилия	57	2,7	2354	18,0	81	21	
8	Буркина-Фасо	47	6,9	357	36,0	18	47	
9	Великобритания	74	1,8	15974	237,0	99	13	
10	Вьетнам	63	3,3	230	218,0	88	27	
11	Гаити	43	5,9	383	231,0	53	40	
12	Германия	73	1,5	17539	227,0	99	11	
13	Гондурас	65	4,9	1030	46,0	73	35	
14	Гонконг	75	1,4	14641	5494,0	77	13	
15	Египет	60	3,8	748	57,0	48	29	
16	Замбия	44	6,7	573	11,0	73	46	
17	Индия	58	4,5	275	283,0	52	29	
18	Ирландия	73	2,0	12170	51,0	98	14	
19	Испания	74	1,4	13047	77,0	95	11	
20	Италия	74	1,3	17500	188,0	97	11	
21	Канада	74	1,8	19904	2,8	97	14	

Рис. 9. Данные в окне SPSS после ввода

SPSS Statistics Data Editor window showing the 'Variables' tab. It displays a table with 11 columns: Имя, Тип, Ширина, Десятич..., Метка, Значения, Пропуски, Столбцы, Выравнивание, and Шкала. The first row is highlighted, showing information for the 'Страна' variable.

	Имя	Тип	Ширина	Десятич...	Метка	Значения	Пропуски	Столбцы	Выравнивание	Шкала
1	Страна	Текстовая	17	0		Нет	Нет	17	По левом...	Номинальн...
2	Y	Числовая	11	0		Нет	Нет	11	По право...	Количество
3	x1	Числовая	11	1	x(1)	Нет	Нет	11	По право...	Количество
4	x2	Числовая	11	0	x(2)	Нет	Нет	11	По право...	Количество
5	x3	Числовая	11	1	x(3)	Нет	Нет	11	По право...	Количество
6	x4	Числовая	11	0	x(4)	Нет	Нет	11	По право...	Количество
7	x5	Числовая	11	0	x(5)	Нет	Нет	11	По право...	Количество
8										
9										
10										
11										
12										
13										
14										
15										
16										
17										
18										
19										
20										
21										
22										
23										

Рис. 10. Информация о введенных переменных

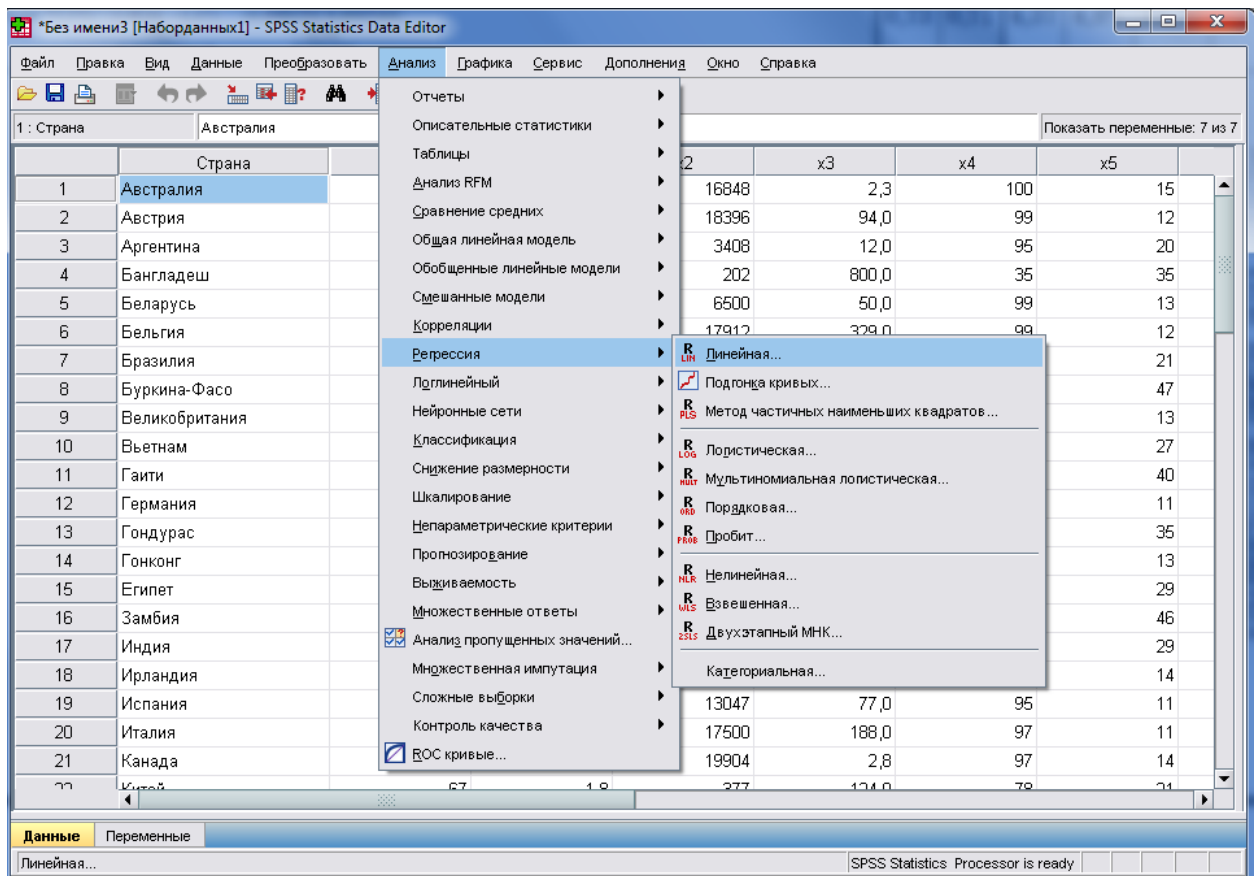


Рис. 11. Меню Анализ > Регрессия

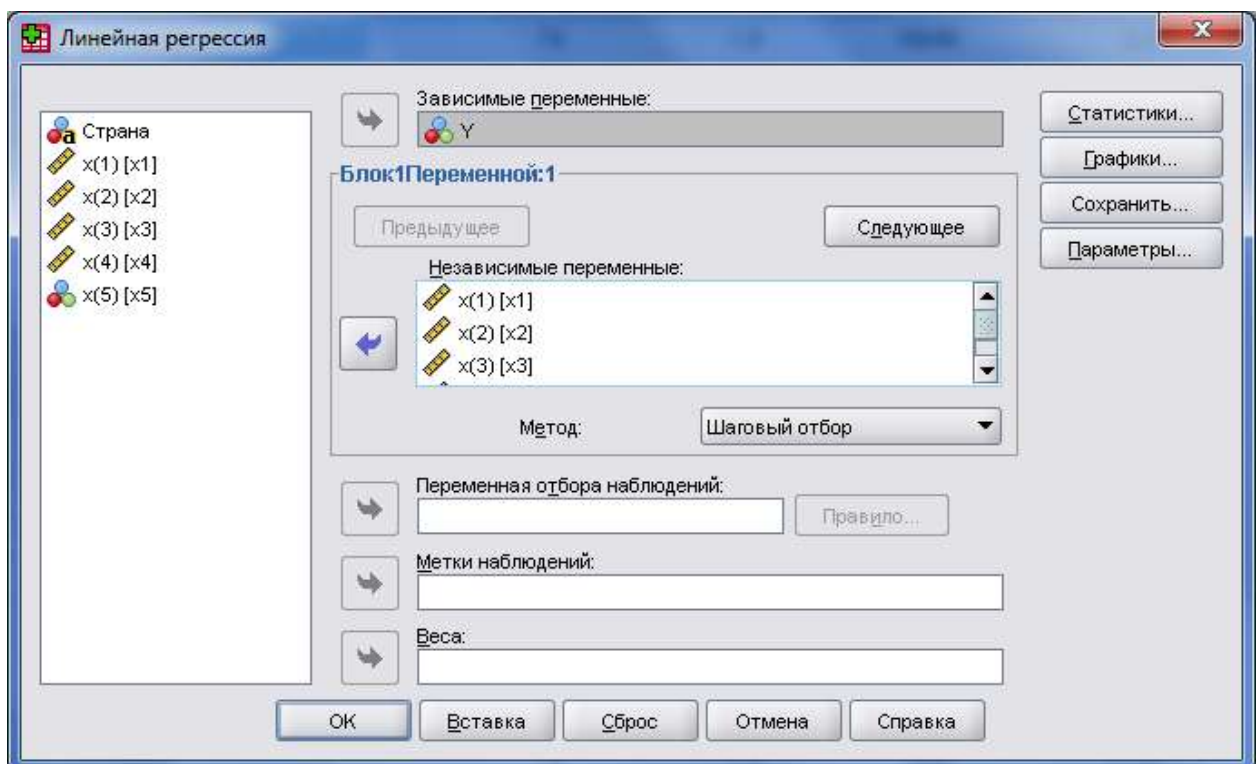


Рис. 12. Окно задания переменных линейной регрессии

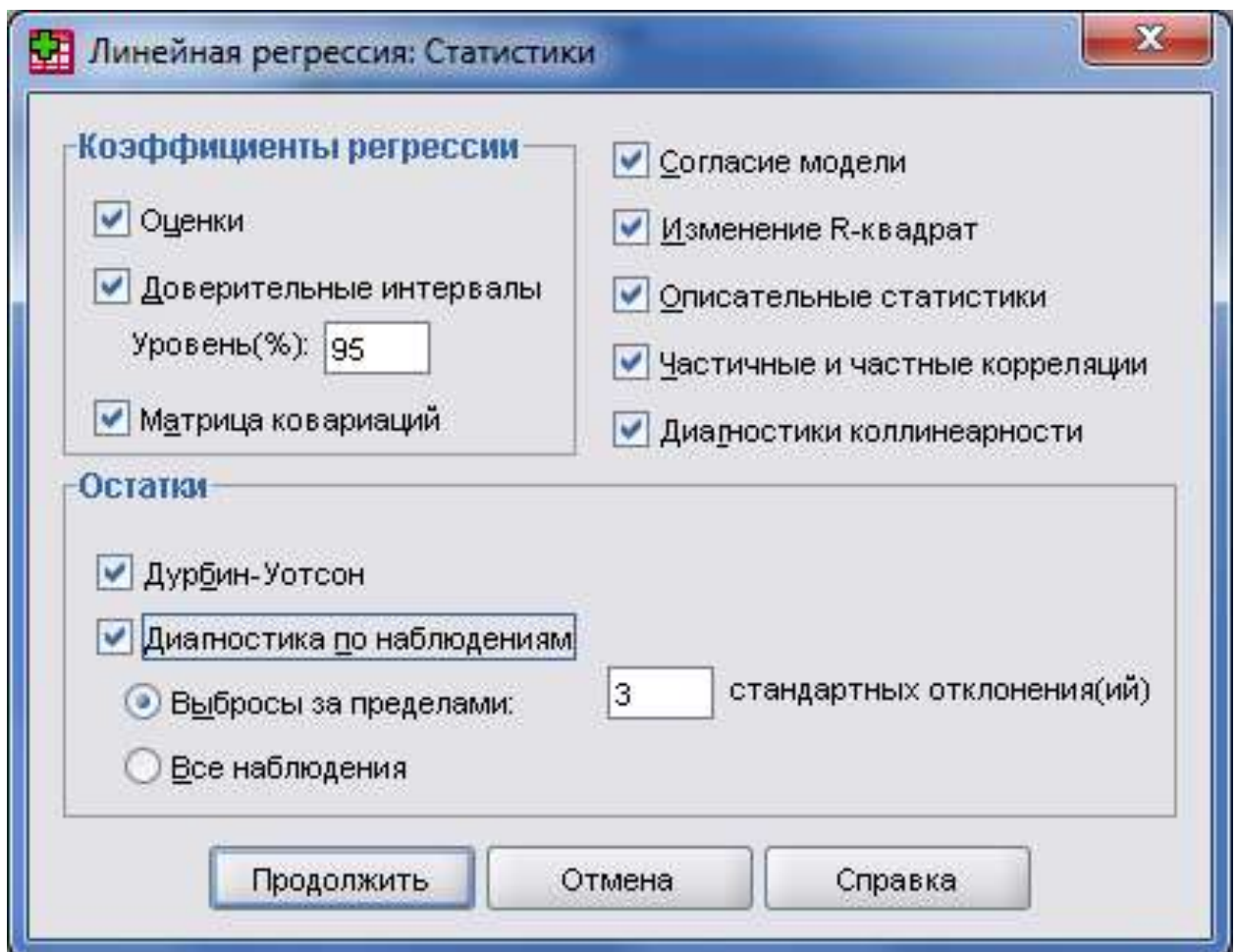


Рис. 13. Окно задания параметров статистики

Последующие окна (рис. 14 – 19) представляют результаты статистики.

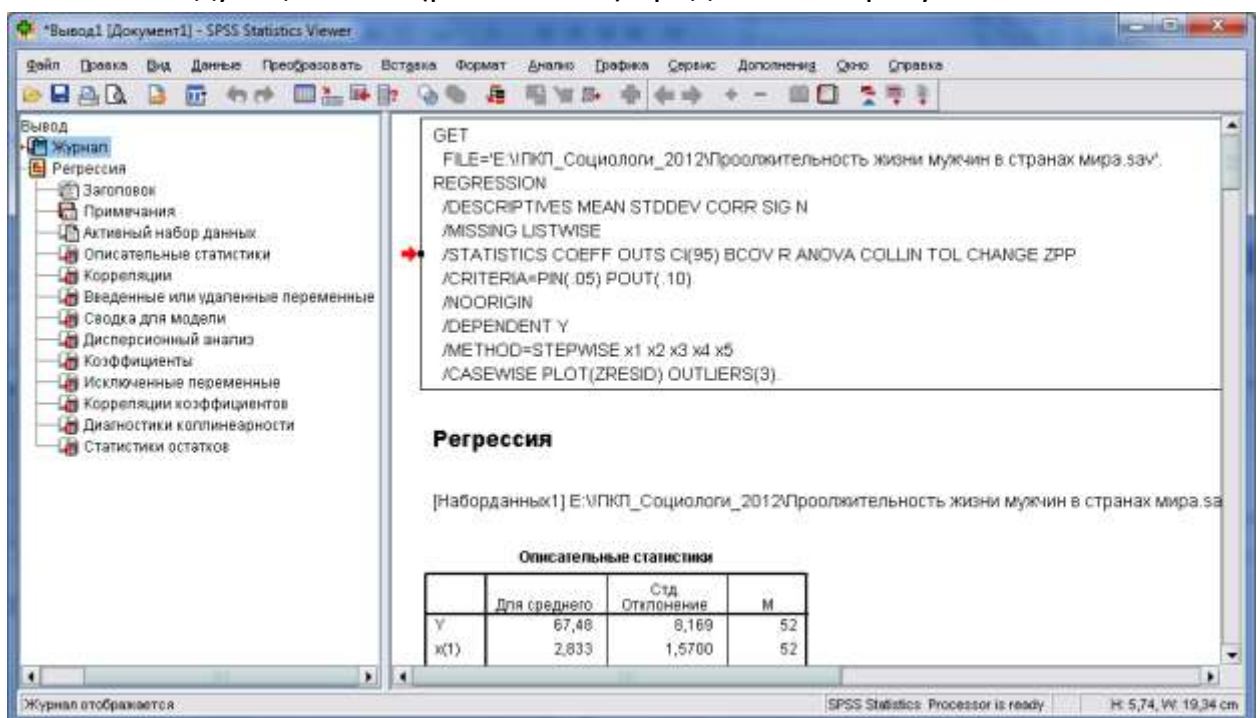


Рис. 14.

### Описательные статистики

	Для среднего	Стд. Отклонение	М
Y	67,48	8,169	52
x(1)	2,833	1,5700	52
x(2)	8421,90	7681,157	52
x(3)	307,723	958,7955	52
x(4)	84,27	20,556	52
x(5)	21,33	10,157	52

**Рис. 15**

### Корреляции

		Y	x(1)	x(2)	x(3)	x(4)	x(5)
Корреляция Пирсона	Y	1,000	-,808	,684	,145	,754	-,817
	x(1)	-,808	1,000	-,576	-,163	-,833	,966
	x(2)	,684	-,576	1,000	,164	,544	-,692
	x(3)	,145	-,163	,164	1,000	-,041	-,142
	x(4)	,754	-,833	,544	-,041	1,000	-,826
	x(5)	-,817	,966	-,692	-,142	-,826	1,000
Знч. (1-сторонняя)	Y	.	,000	,000	,152	,000	,000
	x(1)	,000	.	,000	,124	,000	,000
	x(2)	,000	,000	.	,123	,000	,000
	x(3)	,152	,124	,123	.	,385	,158
	x(4)	,000	,000	,000	,385	.	,000
	x(5)	,000	,000	,000	,158	,000	.
М	Y	52	52	52	52	52	52
	x(1)	52	52	52	52	52	52
	x(2)	52	52	52	52	52	52
	x(3)	52	52	52	52	52	52
	x(4)	52	52	52	52	52	52
	x(5)	52	52	52	52	52	52

**Рис. 16**



**Введенные или удаленные переменные<sup>а</sup>**

Модель	Включенные переменные	Исключенные переменные	Метод
1	x(5)	.	Шаговый (критерий: вероятность F-включения $\leq ,050$ , F-исключения $\geq ,100$ ).
2	x(2)	.	Шаговый (критерий: вероятность F-включения $\leq ,050$ , F-исключения $\geq ,100$ ).
3	x(1)	.	Шаговый (критерий: вероятность F-включения $\leq ,050$ , F-исключения $\geq ,100$ ).
4	.	x(5)	Шаговый (критерий: вероятность F-включения $\leq ,050$ , F-исключения $\geq ,100$ ).

а. Зависимая переменная: Y

**Рис. 17**

**Сводка для модели<sup>е</sup>**

Модель	N	R-квадрат	Скорректированный R-квадрат	Стд. ошибка оценки	Изменения статистик				
					Изменение R квадрат	изменения F	ст.св.1	ст.св.2	Знч. изменения F
1	,817 <sup>а</sup>	,668	,661	4,754	,668	100,569	1	50	,000
2	,834 <sup>б</sup>	,695	,683	4,601	,027	4,380	1	49	,042
3	,853 <sup>с</sup>	,727	,710	4,399	,032	5,617	1	48	,022
4	,852 <sup>д</sup>	,725	,714	4,367	-,002	,310	1	48	,580

- а. Предикторы: (конст) x(5)  
 б. Предикторы: (конст) x(5), x(2)  
 в. Предикторы: (конст) x(5), x(2), x(1)  
 г. Предикторы: (конст) x(2), x(1)  
 е. Зависимая переменная: Y

**Рис. 18**



### Дисперсионный анализ<sup>е</sup>

Модель		Сумма квадратов	ст.св.	Средний квадрат	Щ	Знч.
1	Регрессия	2272,943	1	2272,943	100,569	,000 <sup>а</sup>
	Остаток	1130,037	50	22,601		
	Всего	3402,981	51			
2	Регрессия	2365,662	2	1182,831	55,874	,000 <sup>б</sup>
	Остаток	1037,319	49	21,170		
	Всего	3402,981	51			
3	Регрессия	2474,332	3	824,777	42,631	,000 <sup>с</sup>
	Остаток	928,649	48	19,347		
	Всего	3402,981	51			
4	Регрессия	2468,328	2	1234,164	64,702	,000 <sup>д</sup>
	Остаток	934,653	49	19,075		
	Всего	3402,981	51			

- а. Предикторы: (конст) x(5)  
б. Предикторы: (конст) x(5), x(2)  
с. Предикторы: (конст) x(5), x(2), x(1)  
д. Предикторы: (конст) x(2), x(1)  
е. Зависимая переменная: Y

Рис. 19

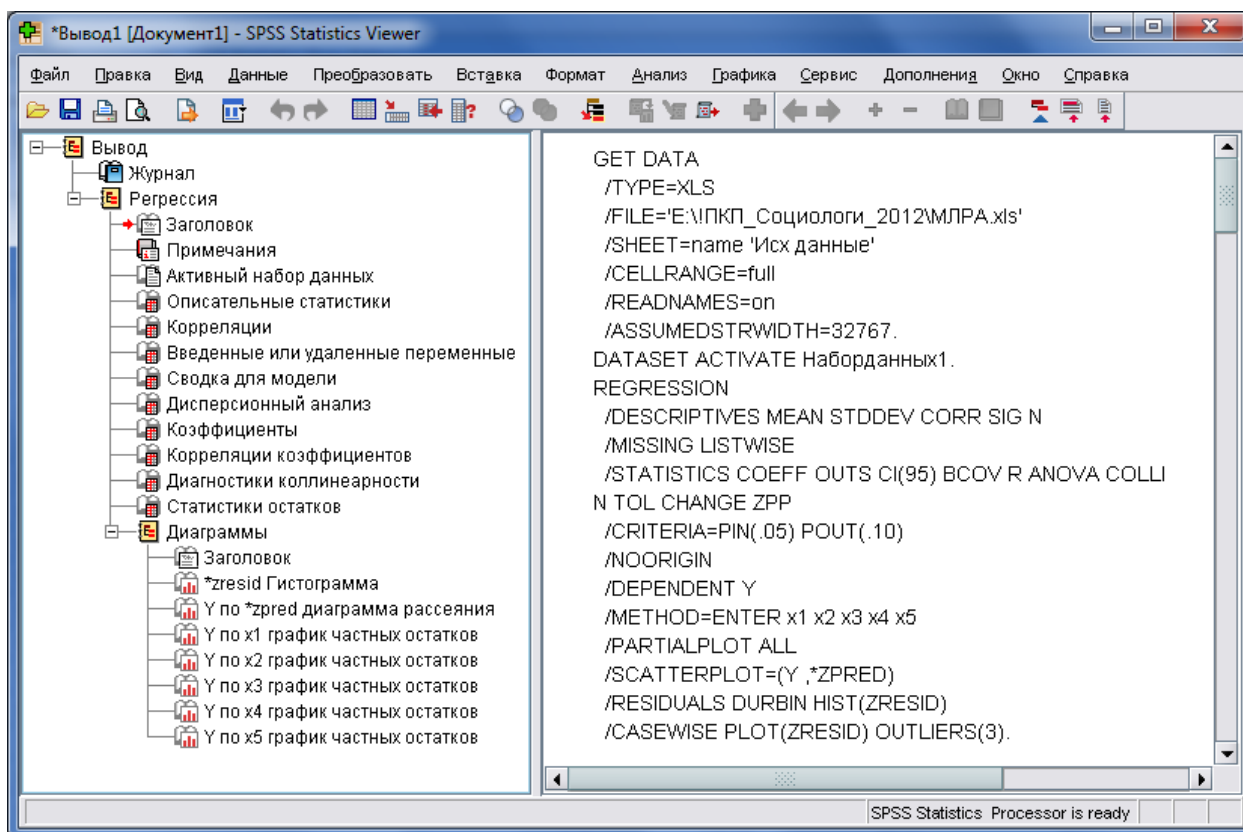


Рис. 20. Примерный вид окна общих результатов обработки

Далее приведены еще некоторые окна с результатами обработки (рис. 21 – 33).

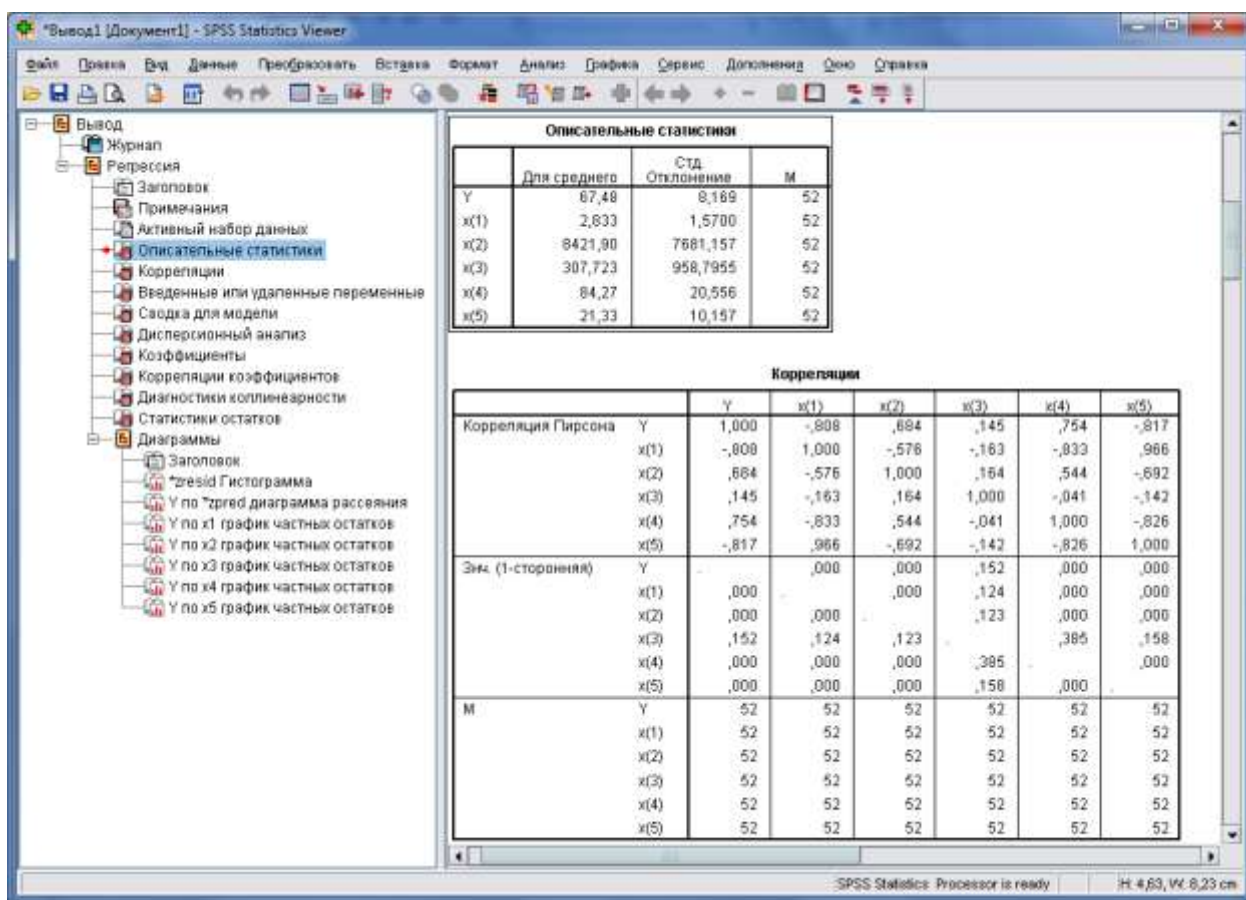


Рис. 21

**Коэффициенты<sup>a</sup>**

Модель	Нестандартизованные коэффициенты		Стандартизованные коэффициенты	t	Знач.	95,0% доверительный интервал для В		Корреляции			Статистики коллинеарности	
	В	Стд. Ошибка	Бета			Нижняя граница	Верхняя граница	Нулевой порядок	Частные	Часть	Толерантность	КРД
1 (Константа)	81,498	1,545		52,734	,000	78,394	84,602					
x(5)	-,657	,066	-,817	-10,028	,000	-,789	-,526	-,817	-,817	-,817	1,000	1,000
2 (Константа)	76,740	2,721		28,198	,000	71,271	82,209					
x(5)	-,530	,088	-,659	-6,036	,000	-,707	-,354	-,817	-,653	-,476	,522	1,917
x(2)	,000	,000	,229	2,093	,042	,000	,000	,684	,286	,165	,522	1,917
3 (Константа)	72,306	3,205		22,564	,000	65,863	78,749					
x(5)	,171	,308	,213	,557	,580	-,447	,790	-,817	,080	,042	,039	25,747
x(2)	,000	,000	,371	3,080	,003	,000	,001	,684	,406	,232	,392	2,552
x(1)	-4,167	1,758	-,801	-2,370	,022	-7,702	-,632	-,808	-,324	-,179	,050	20,082
4 (Константа)	73,679	2,033		36,239	,000	69,594	77,765					
x(2)	,000	,000	,328	3,581	,001	,000	,001	,684	,455	,268	,669	1,495
x(1)	-3,224	,476	-,620	-6,769	,000	-4,182	-2,267	-,808	-,695	-,507	,669	1,495

a. Зависимая переменная: Y

Рис. 22

Исключенные переменные\*

Модель		Бета включения	t	Знч.	Частная корреляция	Статистики коллинеарности		
						Толерантность	КРД	Минимальная толерантность
1	x(1)	-,282 <sup>a</sup>	-,888	,379	-,126	,066	15,083	,066
	x(2)	,229 <sup>a</sup>	2,093	,042	,286	,522	1,917	,522
	x(3)	,030 <sup>a</sup>	,364	,718	,052	,980	1,020	,980
	x(4)	,249 <sup>a</sup>	1,760	,085	,244	,318	3,148	,318
2	x(1)	-,801 <sup>b</sup>	-2,370	,022	-,324	,050	20,082	,039
	x(3)	,015 <sup>b</sup>	,186	,853	,027	,972	1,029	,517
	x(4)	,270 <sup>b</sup>	1,985	,053	,275	,316	3,162	,234
3	x(3)	-,016 <sup>c</sup>	-,210	,834	-,031	,943	1,060	,038
	x(4)	,206 <sup>c</sup>	1,515	,136	,216	,298	3,353	,039
4	x(3)	-,010 <sup>d</sup>	-,124	,902	-,018	,966	1,035	,664
	x(4)	,199 <sup>d</sup>	1,477	,146	,208	,300	3,334	,285
	x(5)	,213 <sup>d</sup>	,557	,580	,080	,039	25,747	,039

a. Предикторы в модели: (конст) x(5)

b. Предикторы в модели: (конст) x(5), x(2)

c. Предикторы в модели: (конст) x(5), x(2), x(1)

d. Предикторы в модели: (конст) x(2), x(1)

e. Зависимая переменная: Y

Рис. 23

Корреляции коэффициентов<sup>a</sup>

Модель			x(5)	x(2)	x(1)
1	Корреляции	x(5)	1,000		
	Ковариации	x(5)	,004		
2	Корреляции	x(5)	1,000	,692	
		x(2)	,692	1,000	
	Ковариации	x(5)	,008	7,055E-6	
		x(2)	7,055E-6	1,349E-8	
3	Корреляции	x(5)	1,000	,644	-,962
		x(2)	,644	1,000	-,499
		x(1)	-,962	-,499	1,000
	Ковариации	x(5)	,095	2,537E-5	-,520
		x(2)	2,537E-5	1,641E-8	,000
		x(1)	-,520	,000	3,091
4	Корреляции	x(2)		1,000	,576
		x(1)		,576	1,000
	Ковариации	x(2)		9,479E-9	2,669E-5
		x(1)		2,669E-5	,227

a. Зависимая переменная: Y

Рис. 24

**Диагностики коллинеарности<sup>а</sup>**

Модель	Размерность	Собственное значение	Показатель обусловленности	Доли дисперсии			
				(Константа)	x(5)	x(2)	x(1)
1	1	1,904	1,000	,05	,05		
	2	,096	4,464	,95	,95		
2	1	2,428	1,000	,01	,01	,03	
	2	,540	2,121	,00	,06	,27	
	3	,033	8,637	,99	,92	,70	
3	1	3,265	1,000	,00	,00	,01	,00
	2	,678	2,195	,00	,00	,19	,00
	3	,054	7,806	,43	,00	,39	,05
	4	,004	28,054	,57	1,00	,41	,94
4	1	2,402	1,000	,01		,03	,02
	2	,544	2,101	,00		,34	,11
	3	,053	6,721	,98		,63	,87

а. Зависимая переменная: Y

**Рис. 25**

**Статистики остатков<sup>а</sup>**

	Минимум	Максимум	Для среднего	Стд. Отклонение	М
Предсказанное значение	51,43	76,32	67,48	6,957	52
Остаток	-11,660	11,608	,000	4,281	52
Стд. Предсказанное значение	-2,308	1,271	,000	1,000	52
Стд. Остаток	-2,670	2,658	,000	,980	52

а. Зависимая переменная: Y

**Рис. 26**

Введенные или удаленные переменные			
Модель	Включенные переменные	Исключенные переменные	Метод
1	x(5), x(3), x(2), x(4), x(1) <sup>a</sup>	.	Принудительное включение
a. Включены все запрошенные переменные			

Сводка для модели<sup>b</sup>

Модель	N	R-квадрат	Скорректированный R-квадрат	Стд. ошибка оценки	Изменения статистик					Дурбин-Уотсон
					Изменение R квадрат	изменения F	ст.св. 1	ст.св. 2	Знач. изменения F	
1	,860 <sup>a</sup>	,740	,712	4,383	,740	26,231	5	46	,000	1,765

a. Предикторы: (конст) x(5), x(3), x(2), x(4), x(1)

b. Зависимая переменная: Y

Дисперсионный анализ<sup>b</sup>

Модель		Сумма квадратов	ст.св.	Средний квадрат	Щ	Знач.
1	Регрессия	2519,376	5	503,875	26,231	,000 <sup>a</sup>
	Остаток	883,605	46	19,209		
	Всего	3402,981	51			

a. Предикторы: (конст) x(5), x(3), x(2), x(4), x(1)

b. Зависимая переменная: Y

Рис. 27

Коэффициенты<sup>a</sup>

Модель		Нестандартизованные коэффициенты		Стандартизованные коэффициенты	t	Знач.	95,0% доверительный интервал для B		Корреляции			Статистики коллинеарности	
		B	Стд. ошибка	Бета			Нижняя граница	Верхняя граница	Нулевой порядок	Частичная	Часть	Толерантность	КРД
1	(Константа)	62,292	7,330		8,510	,000	47,559	77,025					
	x(1)	-3,375	1,866	-,649	-1,809	,077	-7,130	,390	-,808	-,256	-,136	,044	22,777
	x(2)	,000	,000	,353	2,075	,000	,000	,000	,684	,390	,216	,375	2,664
	x(3)	,000	,001	,025	,307	,780	-,001	,002	,145	,045	,023	,837	1,194
	x(4)	,088	,059	,221	1,516	,136	-,029	,205	,754	,218	,114	,265	3,777
	x(5)	,193	,310	,240	,621	,538	-,432	,818	-,817	,081	,047	,038	26,389

a. Зависимая переменная: Y

Корреляции коэффициентов<sup>a</sup>

Модель		x(5)	x(3)	x(2)	x(4)	x(1)
1	Корреляции					
	x(5)	1,000	-,137	,645	,024	-,910
	x(3)	-,137	1,000	-,196	,335	,255
	x(2)	,645	-,196	1,000	-,122	-,522
	x(4)	,024	,335	-,122	1,000	,303
	x(1)	-,910	,255	-,522	,303	1,000
	Ковариации					
	x(5)	,096	-2,981E-5	2,609E-5	,000	-,527
	x(3)	-2,981E-5	4,893E-7	-1,785E-8	1,359E-6	,600
	x(2)	2,609E-5	-1,785E-8	1,701E-8	-9,266E-7	,600
	x(4)	,000	1,359E-6	-9,266E-7	,003	,033
	x(1)	-,527	,600	,600	,033	-3,481

a. Зависимая переменная: Y

Рис. 28

Диагностика коллинеарности <sup>а</sup>									
Модель	Размерность	Собственное значение	Показатель обусловленности	Доли дисперсии					
				(Константа)	x(1)	x(2)	x(3)	x(4)	x(5)
1	1	4,255	1,000	,00	,00	,01	,01	,00	,00
	2	,970	2,094	,00	,00	,02	,57	,00	,00
	3	,650	2,558	,00	,00	,11	,28	,00	,00
	4	,115	6,077	,01	,01	,46	,00	,05	,00
	5	,005	29,103	,60	,41	,04	,15	,83	,10
	6	,004	32,854	,40	,57	,36	,00	,11	,90

а. Зависимая переменная: Y

Статистики остатков <sup>а</sup>					
	Минимум	Максимум	Для среднего	Стд. Отклонение	М
Предсказанное значение	49,66	76,34	67,48	7,028	52
Остаток	-11,814	10,942	,000	4,162	52
Стд. Предсказанное значение	-2,536	1,261	,000	1,000	52
Стд. Остаток	-2,696	2,497	,000	,950	52

а. Зависимая переменная: Y

Рис. 29

## Гистограмма

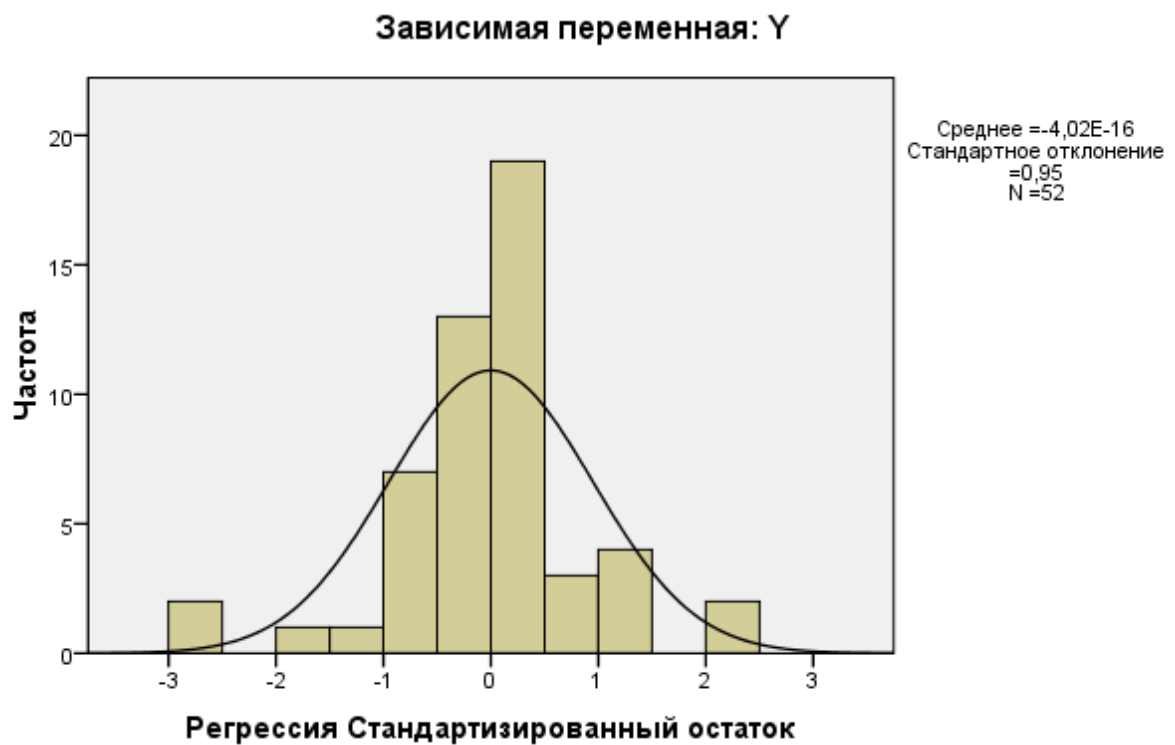


Рис. 30

### Диаграмма рассеяния



Рис. 31

### График частной регрессии

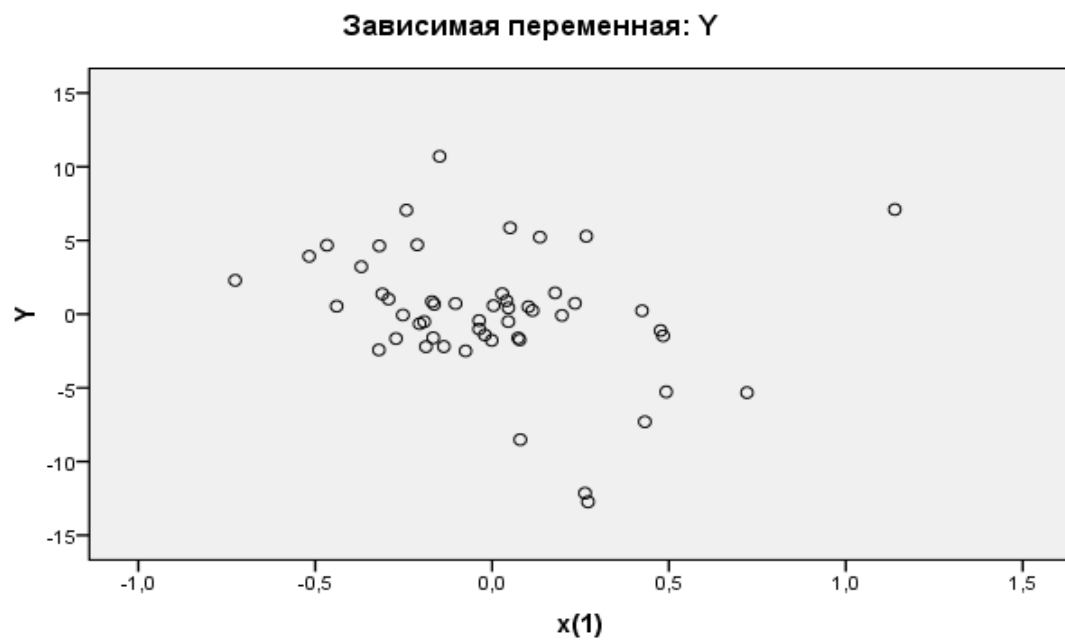


Рис. 32

Корреляции							
		Y	x(1)	x(2)	x(3)	x(4)	x(5)
Y	Корреляция Пирсона	1	-,808	,684	,145	,754	-,817
	Знч.(1-сторон)		,000	,000	,152	,000	,000
	N	52	52	52	52	52	52
x(1)	Корреляция Пирсона	-,808	1	-,576	-,163	-,833	,966
	Знч.(1-сторон)	,000		,000	,124	,000	,000
	N	52	52	52	52	52	52
x(2)	Корреляция Пирсона	,684	-,576	1	,164	,544	-,692
	Знч.(1-сторон)	,000	,000		,123	,000	,000
	N	52	52	52	52	52	52
x(3)	Корреляция Пирсона	,145	-,163	,164	1	-,041	-,142
	Знч.(1-сторон)	,152	,124	,123		,385	,158
	N	52	52	52	52	52	52
x(4)	Корреляция Пирсона	,754	-,833	,544	-,041	1	-,826
	Знч.(1-сторон)	,000	,000	,000	,385		,000
	N	52	52	52	52	52	52
x(5)	Корреляция Пирсона	-,817	,966	-,692	-,142	-,826	1
	Знч.(1-сторон)	,000	,000	,000	,158	,000	
	N	52	52	52	52	52	52

Y — ожидаемая продолжительность жизни мужчины (лет от пяти факторных признаков - регрессоров :

- x(1) — среднего числа детей в семье,
- x(2) — ВВП на душу населения (в долл. США по покупательной способности валют),
- x(3) — плотности населения (в чел. на кв. км),
- x(4) — процента грамотных
- x(5) — рождаемости на 1000 чел.

**Рис. 33**

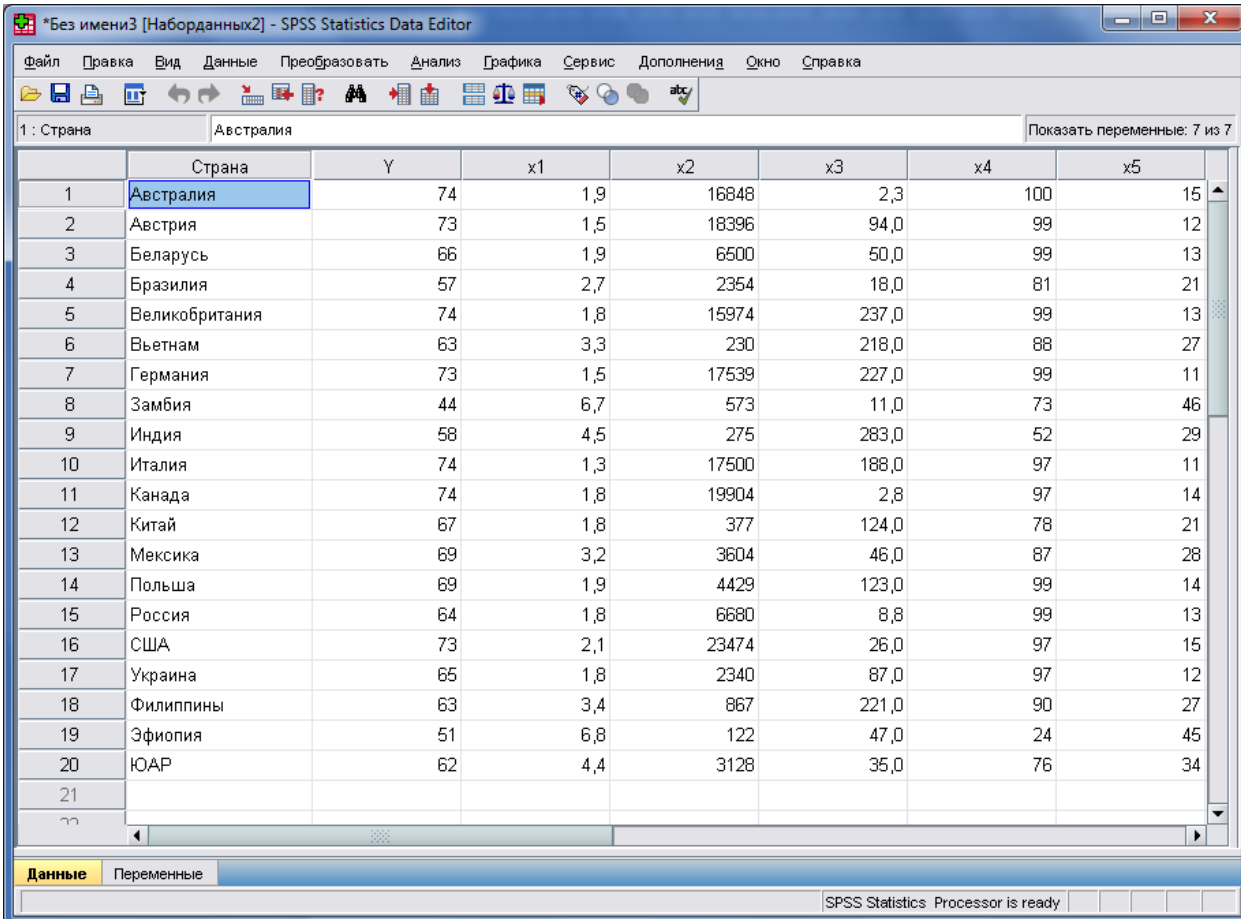


## 2. ПРИМЕР РЕШЕНИЯ ЗАДАЧИ В SPSS МЕТОДОМ КЛАСТЕРНОГО АНАЛИЗА

1. Установить на личный компьютер ознакомительную версию пакета SPSS.
2. Подготовить исходные данные для анализа в файле Excel; сохранить файл (файл **МЛРА-для кластерного анализа.xls** имеется, находится на сетевом диске).
3. Запустить программу SPSS.
4. Загрузить файл с исходными данными в среду SPSS:

*Команда Файл → Открыть → ...*

Появится окно, представленное на рис. 34.



The screenshot shows the SPSS Statistics Data Editor window. The title bar reads '\*Без имени3 [Наборданных2] - SPSS Statistics Data Editor'. The menu bar includes 'Файл', 'Правка', 'Вид', 'Данные', 'Преобразовать', 'Анализ', 'Графика', 'Сервис', 'Дополнения', 'Окно', and 'Справка'. The toolbar contains various icons for file operations, editing, and analysis. The data grid has 21 rows and 7 columns. The first column is labeled 'Страна' and the others are 'Y', 'x1', 'x2', 'x3', 'x4', and 'x5'. The data is as follows:

	Страна	Y	x1	x2	x3	x4	x5
1	Австралия	74	1,9	16848	2,3	100	15
2	Австрия	73	1,5	18396	94,0	99	12
3	Беларусь	66	1,9	6500	50,0	99	13
4	Бразилия	57	2,7	2354	18,0	81	21
5	Великобритания	74	1,8	15974	237,0	99	13
6	Вьетнам	63	3,3	230	218,0	88	27
7	Германия	73	1,5	17539	227,0	99	11
8	Замбия	44	6,7	573	11,0	73	46
9	Индия	58	4,5	275	283,0	52	29
10	Италия	74	1,3	17500	188,0	97	11
11	Канада	74	1,8	19904	2,8	97	14
12	Китай	67	1,8	377	124,0	78	21
13	Мексика	69	3,2	3604	46,0	87	28
14	Польша	69	1,9	4429	123,0	99	14
15	Россия	64	1,8	6680	8,8	99	13
16	США	73	2,1	23474	26,0	97	15
17	Украина	65	1,8	2340	87,0	97	12
18	Филиппины	63	3,4	867	221,0	90	27
19	Эфиопия	51	6,8	122	47,0	24	45
20	ЮАР	62	4,4	3128	35,0	76	34
21							

**Рис. 34. Рабочая область окна SPSS с исходными данными**

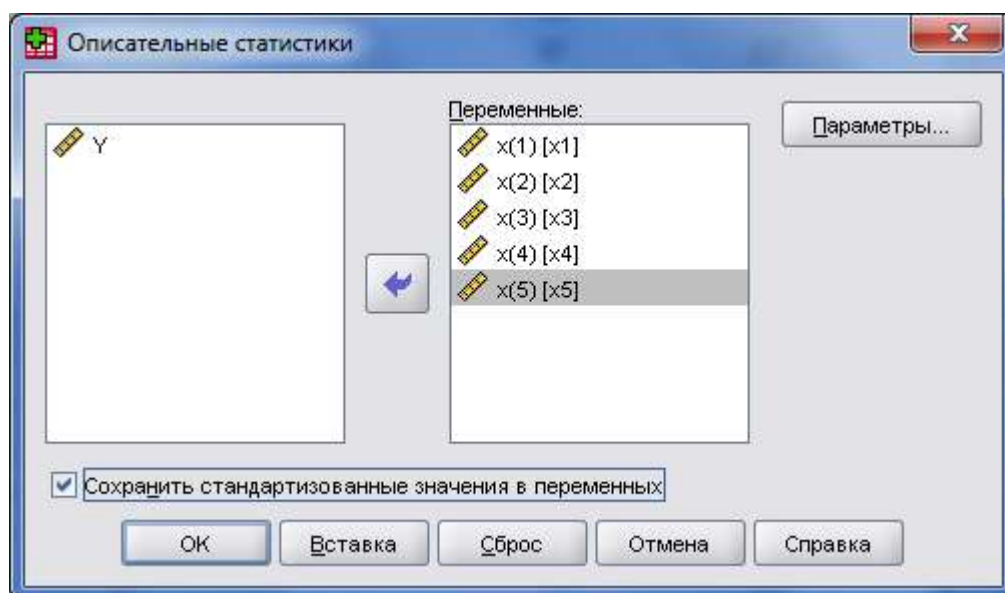
5. Поскольку исходные переменные  $x(1)$ ,  $x(2)$ ,  $x(3)$ ,  $x(4)$ ,  $x(5)$  имеют различную природу и единицы измерения, перед выполнением процедур

кластерного анализа следует стандартизовать исходные данные, перейдя к новым переменным  $z(1)$ ,  $z(2)$ ,  $z(3)$ ,  $z(4)$ ,  $z(5)$  по правилу:

$$z^{(j)} = \frac{x^{(j)} - \overline{x^{(j)}}}{s_{x^{(j)}}}, \quad j = 1, 2, 3, 4, 5.$$

Для этого следует воспользоваться инструментарием «Описательные статистики»:

Команда Анализ → Описательные статистики → Описательные ... → В окне (рис. 35): выбрать независимые переменные  $x(1)$ ,  $x(2)$ ,  $x(3)$ ,  $x(4)$ ,  $x(5)$  → Установить флажок Сохранить стандартизованные значения в переменных → Кнопка ОК



**Рис. 35. Задание параметров обработки исходных данных в окне Описательная статистика**

Стандартизованные значения признаков  $x(1)$ ,  $x(2)$ ,  $x(3)$ ,  $x(4)$ ,  $x(5)$  на 20 объектах исходных данных появятся в рабочей области окна SPSS под именами:  $Zx1$ ,  $Zx2$ ,  $Zx3$ ,  $Zx4$ ,  $Zx5$  (рис. 36).

	Страна	Y	x1	x2	x3	x4	x5	Zx1	Zx2	Zx3	Zx4	Zx5
1	Австралия	74	1,9	16848	2,3	100	15	-0,55430	1,07027	-1,07215	0,69740	-0,55241
2	Австрия	73	1,5	18396	94,0	99	12	-0,79944	1,25870	-0,09041	0,64555	-0,82634
3	Беларусь	66	1,9	6500	50,0	99	13	-0,56656	-0,18937	-0,56148	0,64555	-0,73603
4	Бразилия	57	2,7	2354	18,0	81	21	-0,06404	-0,69405	-0,90407	-0,26778	-0,00457
5	Великобритан...	74	1,8	15974	237,0	99	13	-0,59720	0,96388	1,44054	0,64555	-0,73603
6	Вьетнам	63	3,3	230	218,0	88	27	0,32204	-0,95260	1,23713	0,07518	0,54328
7	Германия	73	1,5	17539	227,0	99	11	-0,81782	1,15438	1,33348	0,64555	-0,91765
8	Замбия	44	6,7	573	11,0	73	46	2,37502	-0,91085	-0,97901	-0,70259	2,27814
9	Индия	58	4,5	275	283,0	52	29	1,02680	-0,94713	1,93302	-1,79147	0,72590
10	Италия	74	1,3	17500	188,0	97	11	-0,92200	1,14963	0,91595	0,54185	-0,91765
11	Канада	74	1,8	19904	2,8	97	14	-0,61559	1,44226	-1,06680	0,54185	-0,64372
12	Китай	67	1,8	377	124,0	78	21	-0,59107	-0,93471	0,23077	-0,44333	-0,00457
13	Мексика	69	3,2	3604	46,0	87	28	0,24237	-0,54189	-0,60430	0,02333	0,63459
14	Польша	69	1,9	4429	123,0	99	14	-0,52979	-0,44147	0,22006	0,64555	-0,64372
15	Россия	64	1,8	6880	8,8	99	13	-0,59720	-0,16746	-1,00256	0,64555	-0,73603
16	США	73	2,1	23474	26,0	97	15	-0,45625	1,87683	-0,81842	0,54185	-0,55241
17	Украина	65	1,8	2340	87,0	97	12	-0,60333	-0,69576	-0,16635	0,54185	-0,82634
18	Филиппины	63	3,4	867	221,0	90	27	0,33430	-0,87506	1,26925	0,17889	0,54328
19	Эфиопия	51	6,8	122	47,0	24	45	2,45469	-0,96575	-0,59359	-3,24331	2,18683
20	ЮАР	62	4,4	3128	35,0	76	34	0,95938	-0,59984	-0,72206	-0,54703	1,16244
21												

**Рис. 36. Стандартизованные значения признаков – исходных данных**

6. Инициировать инструментарий «Иерархический кластерный анализ»:

Команда Анализ → Классификация → Иерархическая кластеризация... →

В окне (рис. 37): выбрать переменные Z-знач:

x(1), z-знач:x(2), z-знач:x(3), z-знач:x(4), z-знач:x(5) →

Метить значениями: Страна →

Переключатель Кластеризовать: Наблюдения →

Флажок Вывести на дисплей: Статистики →

Флажок Вывести на дисплей: Графики → Кнопка Метод...→

В окне рис. 38: → Метод: Ближайший сосед →

Мера Интервальная: Расстояние Евклида→ Кнопка Продолжить →

В окне рис. 37: Кнопка Графики... →

В окне рис. 39: Установить флажок Дендограмма → Кнопка Продолжить →

В окне рис. 37 Кнопка Статистики...→

В окне рис. 40: Установить флажок Порядок агломерации →

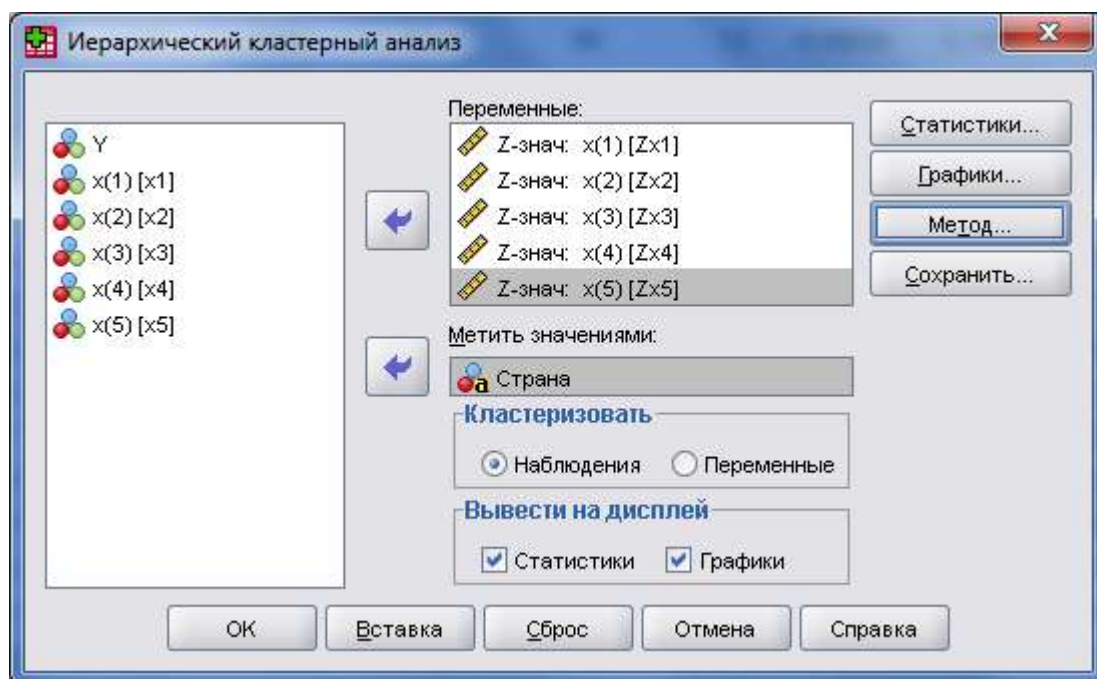
Установить флажок Матрица близостей → Кнопка Продолжить →

В окне рис. 37: Кнопка Сохранить... →

В окне рис. 41: Переключатель принадлежности кластерам: Одно решение

→ Количество кластеров: 2 → Кнопка Продолжить →

В окне рис. 37: Кнопка ОК



**Рис. 37. Окно иерархический кластерный анализ с выбранными параметрами**

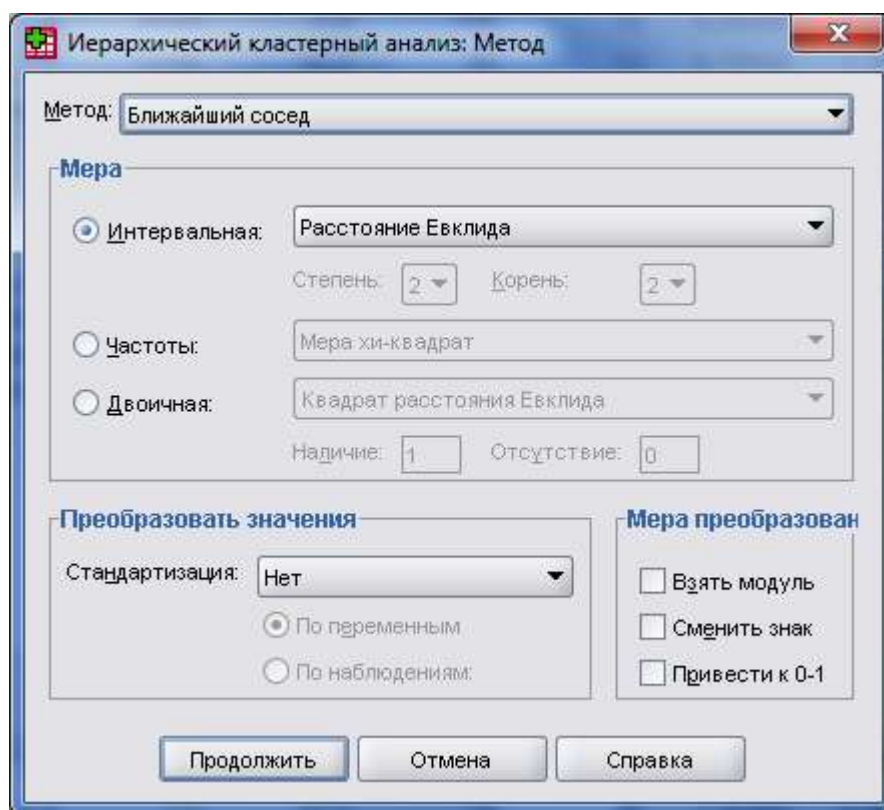


Рис. 38. Окно Иерархический кластерный анализ: Метод

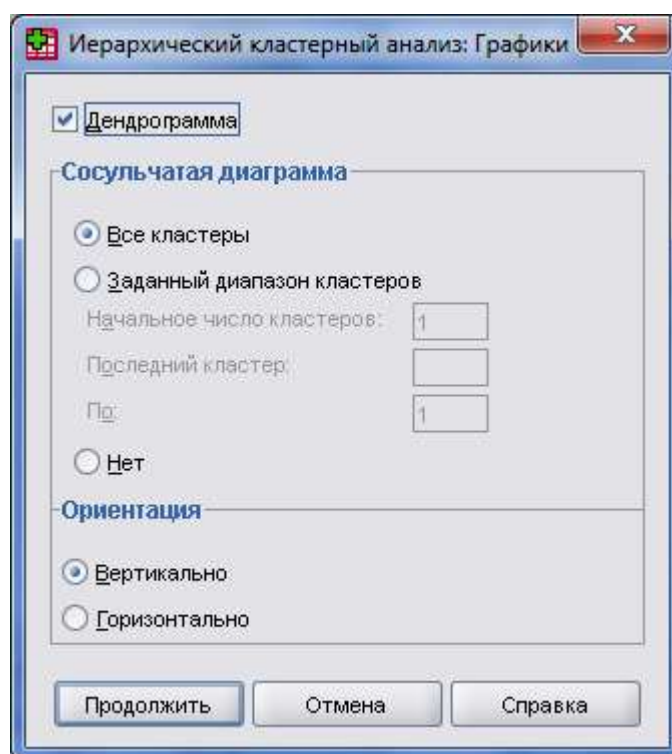
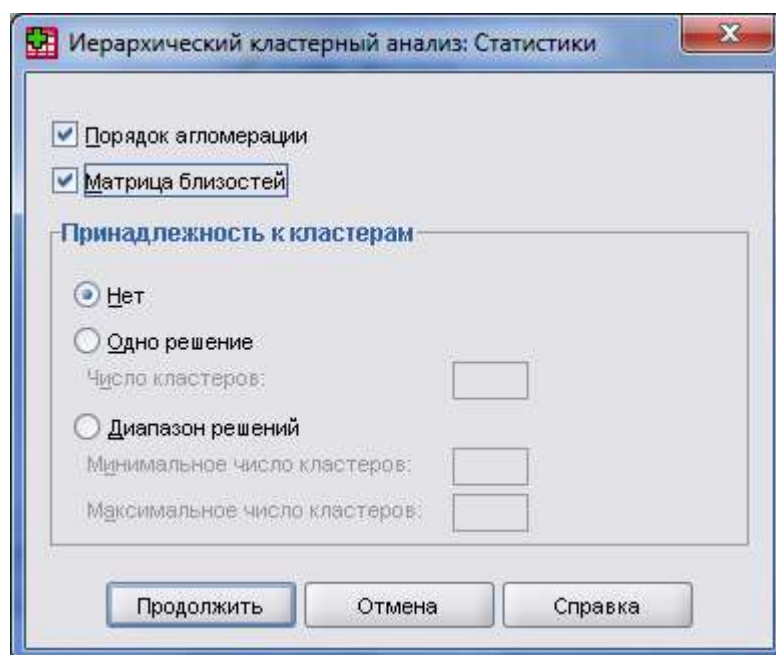
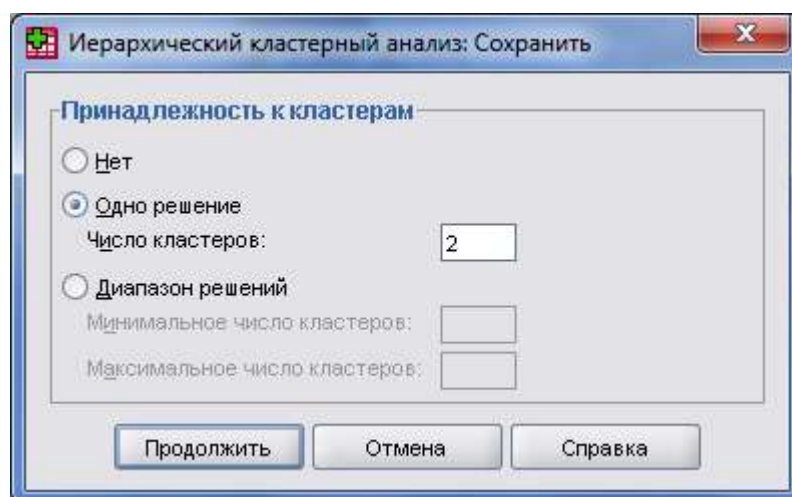


Рис. 39. Окно Иерархический кластерный анализ: Графики



**Рис. 40. Окно Иерархический кластерный анализ: Статистики**



**Рис. 41. Окно Иерархический кластерный анализ: Сохранить**

#### 7. Выполнить анализ результатов. Сделать выводы.

В ходе выполнения анализа целесообразно воспользоваться рекомендациями, приведенными в учебном пособии [2, с. 77 - 91].

### **3. КОМПЛЕКСНОЕ ЗАДАНИЕ НА ОСВОЕНИЕ ТЕХНОЛОГИИ ИСПОЛЬЗОВАНИЯ ПАКЕТА SPSS**

#### **3.1. Общие сведения**

Комплексное задание на освоение технологии использования пакета SPSS представляет собой индивидуальное научно-аналитическое исследование по обработке и анализу социологической информации – внеаудиторную лабораторную работу, выполняемую студентом на личном компьютере в часы самостоятельной работы с применением средств и возможностей пакета SPSS.

Цели исследования:

- освоение возможностей и технологии использования пакета SPSS при решения исследовательских социологических задач;
- выработка у студентов навыков самостоятельной работы с имеющейся исходной информацией;
- приобретение студентами навыков грамотного использования методики анализа социологической информации;
- практическая реализация студентами теоретических знаний с использованием инструментальных средств SPSS;
- формирование и развитие у студентов научно-исследовательских навыков поиска, выборки, анализа и обработки социологической информации;
- комплексное применение компетенций, теоретических знаний, практических навыков и умений, приобретенных студентами при изучении профессиональных компьютерных программ и других профессионально-ориентированных дисциплин.

По результатам выполнения комплексного самостоятельного задания студенты представляют отчет о выполненной работе в виде электронного документа. Он должен содержать информацию о студенте, справочные и исходные данные, результаты решения, необходимые иллюстрации, схемы, графики, другие пояснения. Обязательным в отчете является наличие анализа полученных результатов, обоснованных выводов и рекомендаций.

Результаты выполнения работы подлежат публичному обсуждению (с использованием возможностей многофункционального кафедрального образовательного портала [www.fa-kit.ru](http://www.fa-kit.ru)) и защите.



## 3.2. Исходные данные для многомерного статистического анализа

<sup>1</sup> п / п	Страна	$x^{(1)}$	$x^{(2)}$	$x^{(3)}$	$x^{(4)}$	$x^{(5)}$	$x^{(6)}$	$x^{(7)}$	$x^{(8)}$	$x^{(9)}$	$x^{(10)}$	$x^{(11)}$	$x^{(12)}$
1	Австралия	17800	15	8	7,3	1,9	74	80	16848	2,3	85	100	1,38
2	Австрия	8000	12	11	6,7	1,5	73	79	18396	94	58	99	0,2
3	Аргентина	33900	20	9	25,6	2,8	68	75	3408	12	86	95	1,3
4	Бангладеш	125000	35	11	106	4,7	53	53	202	800	16	35	2,4
5	Беларусь	10300	13	11	19	1,88	66	76	6500	50	65	99	0,32
6	Бельгия	10100	12	11	7,2	1,7	73	79	17912	329	96	99	0,2
7	Бразилия	156600	21	9	66	2,7	57	67	2354	18	75	81	1,28
8	Буркина-Фасо	10000	47	18	118	6,94	47	50	357	36	15	18	2,81
9	Великобритания	58400	13	11	7,2	1,83	74	80	15974	237	89	99	0,2
10	Вьетнам	73100	27	8	46	3,33	63	68	230	218	20	88	1,78
11	Гаити	6500	40	19	109	5,94	43	47	383	231	29	53	1,63
12	Германия	81200	11	11	6,5	1,47	73	79	17539	227	85	99	0,36
13	Гондурас	5600	35	6	45	4,9	65	70	1030	46	44	73	2,73
14	Гонконг	5800	13	6	5,8	1,4	75	80	14641	5494	94	77	-0,09
15	Египет	60000	29	9	76,4	3,77	60	63	748	57	44	48	1,95
16	Замбия	9100	46	18	85	6,68	44	45	573	11	42	73	2,8
17	Индия	911600	29	10	79	4,48	58	59	275	283	26	52	1,9
18	Ирландия	3600	14	9	7,4	1,99	73	78	12170	51	57	98	0,3
19	Испания	39200	11	9	6,9	1,4	74	81	13047	77	78	95	0,25
20	Италия	58100	11	10	7,6	1,3	74	81	17500	188	69	97	0,21
21	Канада	29100	14	8	6,8	1,8	74	81	19904	2,8	77	97	0,7
22	Китай	1205200	21	7	52	1,84	67	69	377	124	26	78	1,1
23	Колумбия	35600	24	6	28	2,47	69	75	1538	31	70	87	2
24	Коста-Рика	3300	26	4	11	3,1	76	79	2031	64	47	93	2,3
25	Куба	11100	17	7	10,2	1,9	74	78	1382	99	74	94	0,95
26	Малайзия	19500	29	5	25,6	3,51	66	72	2995	58	43	78	2,3
27	Марокко	28600	29	6	50	3,83	66	70	1062	63	46	50	2,12
28	Мексика	91800	28	5	35	3,2	69	77	3604	46	73	87	1,9
29	Нидерланды	15400	13	9	6,3	1,58	75	81	17245	366	89	99	0,58
30	Новая Зеландия	3524	16	8	8,9	2,03	73	80	14381	13	84	99	0,57
31	Норвегия	4300	13	10	6,3	2	74	81	17755	11	75	99	0,4
32	ОАЭ	2800	28	3	22	4,5	70	74	14193	32	81	68	4,8
33	Польша	38600	14	10	13,8	1,94	69	77	4429	123	62	99	0,3
34	Португалия	10500	12	10	9,2	1,5	71	78	9000	108	34	85	0,36
35	Россия	149200	13	11	27	1,83	64	74	6680	8,8	74	99	0,2
36	Саудовская Аравия	18000	38	6	52	6,67	66	70	6651	7,7	77	62	3,2
37	Северная Корея	23100	24	6	27,7	2,4	67	73	1000	189	60	99	1,83
38	Сингапур	2900	16	6	5,7	1,88	73	79	14990	4456	100	88	1,2
39	США	260800	15	9	8,11	2,06	73	79	23474	26	75	97	0,99
40	Таиланд	59400	19	6	37	2,1	65	72	1800	115	22	93	1,4
41	Турция	62200	26	6	49	3,21	69	73	3721	79	61	81	2,02



<sup>1</sup> п / п	Страна	$x^{(1)}$	$x^{(2)}$	$x^{(3)}$	$x^{(4)}$	$x^{(5)}$	$x^{(6)}$	$x^{(7)}$	$x^{(8)}$	$x^{(9)}$	$x^{(10)}$	$x^{(11)}$	$x^{(12)}$
42	Украина	51800	12	13	20,7	1,82	65	75	2340	87	67	97	0,05
43	Филиппины	69800	27	7	51	3,35	63	68	867	221	43	90	1,92
44	Финляндия	5100	13	10	5,3	1,8	72	80	15877	39	60	100	0,3
45	Франция	58000	13	9	6,7	1,8	74	82	18944	105	73	99	0,47
46	Чили	14000	23	6	14,6	2,5	71	78	2591	18	85	93	1,7
47	Швейцария	7000	12	9	6,2	1,6	75	82	22384	170	62	99	0,7
48	Швеция	8800	14	11	5,7	2,1	75	81	16900	19	84	99	0,52
49	Эфиопия	55200	45	14	110	6,81	51	54	122	47	12	24	3,1
50	ЮАР	43900	34	8	47,1	4,37	62	68	3128	35	49	76	2,6
51	Южная Корея	45000	16	6	21,7	1,65	68	74	6627	447	72	96	1
52	Япония	125500	11	7	4,4	1,55	76	82	19860	330	77	99	0,3

### Обозначения признаков

- $x^{(1)}$  — численность населения (в тыс. чел.);  
 $x^{(2)}$  — рождаемость (на 1000 чел.);  
 $x^{(3)}$  — смертность (на 1000 чел.);  
 $x^{(4)}$  — младенческая смертность — число детей, умерших в возрасте до 1 г. (на 1000 чел.);  
 $x^{(5)}$  — среднее число детей в семье;  
 $x^{(6)}$  — ожидаемая продолжительность жизни мужчины (в годах);  
 $x^{(7)}$  — ожидаемая продолжительность жизни женщины (в годах);  
 $x^{(8)}$  — ВВП на душу населения (в долл. США по покупательной способности валют);  
 $x^{(9)}$  — плотность населения (количество человек на кв. км);  
 $x^{(10)}$  — процент городского населения;  
 $x^{(11)}$  — процент грамотных;  
 $x^{(12)}$  — прирост населения (% в год).

### Номера признаков для регрессионного анализа

Вариант	Номер результативного признака	Номера факторных признаков (регрессоров)	Вариант	Номер результативного признака	Номера факторных признаков (регрессоров)
1	6	1, 2, 3, 4, 8	19	8	1, 2, 3, 4, 5
2	6	1, 2, 3, 8, 9	20	8	1, 2, 3, 5, 9
3	6	1, 2, 3, 8, 10	21	8	1, 2, 3, 5, 10
4	6	1, 2, 3, 5, 10	22	8	1, 2, 3, 4, 11
5	6	1, 2, 3, 8, 11	23	8	1, 2, 3, 4, 12
6	6	1, 2, 5, 8, 11	24	8	1, 2, 3, 5, 11

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

25	8	1, 2, 3, 9, 11
26	8	1, 2, 3, 5, 12
27	8	1, 2, 3, 4, 11
28	12	1, 2, 3, 4, 5
29	12	1, 2, 3, 5, 9
30	12	1, 2, 3, 5, 10
31	12	1, 2, 3, 4, 11
32	12	1, 2, 3, 4, 8
33	12	1, 2, 3, 5, 11
34	12	1, 2, 3, 9, 11
35	12	1, 2, 3, 5, 8

#### Номера объектов для классификации

Вариант	Номера объектов для классификации
1	1 — 20
2	2 — 21
3	3 — 22
4	4 — 23
5	5 — 24
6	6 — 25
7	7 — 26
8	8 — 27
9	9 — 28
10	10 — 29
11	11 — 30
12	12 — 31

Вариант	Номера объектов для классификации
13	13 — 32
14	14 — 33
15	15 — 34
16	16 — 35
17	17 — 36
18	18 — 37
19	19 — 38
20	20 — 39
21	21 — 40
22	22 — 41
23	23 — 42
24	24 — 43

Вариант	Номера объектов для классификации
25	25 — 44
26	26 — 45
27	27 — 46
28	28 — 47
29	29 — 48
30	30 — 49
31	31 — 50
32	32 — 51
33	33 — 52
34	1 — 7, 16 — 28
35	2 — 8, 17 — 29

## 2.3. Методы исследования и защита работы

Студент проводит исследование в соответствии с заданным преподавателем вариантом работы. В ходе исследования применяются методы анализа - *линейный регрессионный анализ и классификация методом иерархической кластеризации*. Приветствуется инициативный выбор дополнительных методов анализа.

Исходные данные, модели исследования, созданные в пакете SPSS, обоснованные выводы по полученным результатам представляются преподавателю - в электронном виде, включая полученные отчеты.

Краткий отчет о выполненной работе, размещается на сайте кафедры fa-kit.ru с целью его предварительного публичного обсуждения, предшествующей индивидуальной защите.

## **ЛИТЕРАТУРА**

1. Золотарюк А.В. «Профессиональные компьютерные программы». Рабочая учебная программа для студентов, обучающихся по направлению 040100.62 «Социология». –М.: Финуниверситет, кафедра «Информационные технологии», 2012. – 25 с.
2. Компьютерный практикум по прикладной статистике и основам эконометрики: учебное пособие для студентов вузов /В. Н. Калинина, В. И. Соловьев. — М.: Вега-Инфо, 2010. — 140 с.

# **ПРОФЕССИОНАЛЬНЫЕ КОМПЬЮТЕРНЫЕ ПРОГРАММЫ: ПРИМЕНЕНИЕ ПАКЕТА SPSS В СОЦИОЛОГИИ**

**Комплексные задания для самостоятельной работы**

Для студентов, обучающихся по направлению 040100.62 «Социология»

**Автор:**

**Золотарюк Анатолий Васильевич**, кандидат технических наук, доцент, профессор кафедры  
«Информационные технологии»

Компьютерный набор, верстка

А.В. Золотарюк

Формат 60x90/16. Гарнитура Times New Roman

3,2 п.л. 2013 г. Электронное издание

Вычитка и корректура выполнены автором

---

© ФГОБУ ВПО «Финансовый университет при Правительстве Российской Федерации», 2013.

© Кафедра «Информационные технологии», 2013.

© Золотарюк А.В., 2013.