**РОССИЙСКАЯ АКАДЕМИЯ НАУК**
**ИНСТИТУТ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ**
**ПРОБЛЕМ НАРОДОНАСЕЛЕНИЯ**

**ПАЦИОРКОВСКИЙ В.В.,**

**ПАЦИОРКОВСКАЯ В.В.**

# SPSS ДЛЯ СОЦИОЛОГОВ

*Учебное пособие*

Москва

2005 г.

ББК.С5в63я73-1
П 21
УДК 314.18 (470) “2005

**Пациорковский В.В., Пациорковская В.В.**
П 21 SPSS для социологов. Учебное пособие. - М.: ИСЭПН РАН. 2005.- 433 с.

Редакторы:
А.В.Пациорковский – научный редактор;
Г.С. Сизова – литературный редактор.

Рецензенты:
И.И.Корчагина - к.э.н., ст.н.с. ИСЭПН РАН.
В.С.Жаромский – к.э.н., начальник отдела Межведомственного Центра социально-экономических измерений РАН и Росстата.
О.В. Крухмалева - к.с.н., н.с. Центра социологических исследований МГУ им. М.В.Ломоносова.

В работе обобщается опыт использования при обработке данных одного из наиболее продвинутых программных продуктов - статистического пакета для социальных наук - SPSS for Windows. Пособие написано хорошим языком и снабжено методическими материалами. Оно предназначено для студентов, изучающих социологию (раздел: методы сбора, обработки и анализа данных), равно как и для широкого круга научных работников, практикующих социологов и других специалистов.

Табл.: 34, Рис.: 84, Графиков: 9, Обрамлений: 49, Библиография: 43 назв.

Печатается по решению Ученого Совета ИСЭПН РАН (протокол №1 от 15.02.05 г.)

Рекомендована Ученым Советом Института социологии Государственного университета гуманитарных наук (протокол №7 от 04.03.05 г.)

ISBN 5-89997-029-4

***Нашим детям и***
***внукам посвящается***

# ВВЕДЕНИЕ

Предлагаемая книга - итог огромного комплекса работ, выполнявшихся авторами в течение последних 10 лет. В рамках этой темы в 1998 - 2002 гг. опубликовано учебное пособие «Использование SPSS в социологии». Первая часть этого пособия вышла в 1998 г. (1) и переиздана в 1999 г. (2), вторая часть - в 2000 г. (3), а третья часть увидела свет в 2002 г. (4).

Это издание нашло своих читателей и получило положительный отклик в частных письмах к авторам и в профессиональных печатных изданиях, а именно в журналах «Социологические исследования» (5, С. 155-156) и «Народонаселение» (6, С. 166-168). Отдельные отзывы на это издание приведены в приложении 1. Авторы выражают признательность всем, кто откликнулся на их многолетний труд.

Опыт реализации первых выпусков пособия показывает, что спрос на такие работы достаточно устойчив и постоянно растет. Сегодня ситуация уже далека от той, которая была несколько лет назад, когда перевод англоязычного руководства по применению SPSS (7) выполнял функцию основного русскоязычного пособия (8) в области обработки и анализа данных в SPSS. Тем не менее, предложение учебных материалов остается ограниченным. В нем преобладают переводные работы с примерами и данными далекими от нашей социальной реальности.

В то же время за рубежом все более широкое распространение получают издания, способствующие освоению специфики использования SPSS в различных областях социального знания (9-12). Как правило, указанные издания подготовлены предметниками. В таких работах на конкретных примерах рассматриваются различные приемы и методы анализа данных, а также дается интерпретация получаемых с их помощью результатов в той или иной предметной области: криминологии, маркетинге, медицине, социологии и др. Это как раз то, к чему мы стремились при подготовке настоящего пособия.

Освоение навыков использования аналитических процедур, предлагаемых в SPSS, тем эффективнее, чем четче и адекватнее понимание

пользователем существа и специфики исследовательского процесса в той или иной области социальных наук. Поэтому в пособии описание аналитических процедур SPSS дается в постоянной увязке с общим контекстом целей, задач и особенно гипотез, в выполняемых в нашей стране социологических, а точнее социально-экономических исследованиях.

**Цель предлагаемого учебного пособия - формирование у начинающих пользователей** (исследователей, управленцев, студентов) **основных навыков работы в SPSS**. Указанная цель достигается посредством передачи опыта в использовании данного пакета, а также обобщения уже имеющихся методических и учебных материалов. Соотнесенность SPSS с социологией сужает рассмотрение многообразия аналитических возможностей самого пакета прикладных программ. Вместе с тем такой подход отражает состояние знаний данной предметной области и позволяет лучше понять перспективы ее развития. В то же время навыки работы в SPSS, которые формируются с помощью настоящего пособия, с успехом могут использоваться в самых различных областях знания и народного хозяйства - от маркетинговых исследований до мониторинга здоровья населения.

Данное пособие в основном наследует и сохраняет уже сложившуюся структуру предыдущего издания. Тем не менее, как по объему, так и содержательно оно является самостоятельной работой и состоит из введения, заключения, развитого технического аппарата и 17 глав, которые сгруппированы в три раздела.

**Первый** из них - «Ввод и контроль данных». В этом разделе дано краткое описание рассматриваемого пакета, а также рассмотрены особенности ввода и контроля данных в SPSS. Последний момент практически еще не нашел отражения в русскоязычной литературе. Например, в одной работе (13, С. 204) даются упражнения на ввод данных, но не обсуждаются проблемы их контроля. Этот этап работ имеет технический характер, поэтому ему и уделяется мало внимания.

В то же время любой прикладник знает, что именно при формировании базы данных с первичной информацией возникает ряд трудностей и ошибок, которые впоследствии накладывают глубокий отпечаток и массу ограничений на результаты анализа и обобщений. В социологии проблема машинного контроля достоверности информации практически всегда остается актуальной (14).

В рассматриваемой предметной области ввод и контроль данных, равно как и последующие этапы анализа и экспериментальных расче-

тов, - все это звенья одной цепи - **процесса социологического исследования**. Использование современных информационных технологий медленно, но неуклонно ведет к ревизии многих представлений, укоренившихся в этой области.

**Второй** раздел - «Анализ данных». Он связан с основными возможностями и правилами статистического анализа, обработки данных и подготовки отчетов с помощью рассматриваемого пакета прикладных программ. По этому вопросу уже имеются отдельные публикации (13, 15-16). Вместе с тем указанные работы обобщают опыт обучения математической статистике в техническом вузе. Вполне естественно, что они не ставят своей целью учитывать особенности и специфику обработки данных социологических исследований.

Основное внимание в этом разделе уделено рассмотрению частот, таблиц сопряженности, мер сравнения и связи.

**Третий** раздел - «Моделирование и Syntax». Он предполагает освоение навыков построения моделей на основе использования регрессионного, факторного, кластерного, дискриминантного анализа и работы в среде «Amos» (Analysis of moment structures - анализ моментных структур). Модуль Amos появился с 7-й версии пакета. По своей идеологии и имеющимся в нем возможностям он намного превосходит LISREL - соответствующий аналог приложения к ранним версиям пакета.

Мы считаем, что использование Amos и сходных с ним продуктов визуального (наглядного) моделирования неизбежно станет весьма мощным инструментом изучения социальных процессов. Такие программные продукты начнут получать быстрое распространение по мере освоения широкими слоями прикладников-социологов теоретических принципов и методологии компьютерного моделирования.

С освоением методов моделирования социальная наука, возможно впервые в своей истории, получает исключительно благоприятные возможности формирования научно-обоснованного, а главное всегда проверяемого знания. Проектно-конструкторские решения и визуальное восприятие модели в Amos могут служить гигантским стимулом к пониманию ее формально-логических оснований и связанных с ними количественных соотношений. Используя для построения моделей такую интеллектуальную среду, легче и проще понять различные аспекты моделирования. В русскоязычной литературе какая-либо информация по данному вопросу пока еще отсутствует.

В пособии все три раздела тесно увязаны между собой. Оно насыщено перекрестными ссылками, позволяющими формировать целостное представление о характере использования SPSS в социологии, а также учебно-методическими материалами (советами, правилами, заданиями для самостоятельной работы, которые завершают каждую главу) и объемной справочной информацией (приложения 2-10).

Помочь преодолеть трудности, возникающие у пользователей на начальных этапах работы с пакетом SPSS, равно как и последующей работы с ним, - задача и «мечта-идея» авторов, набивших много «шишек» на личном опыте освоения тонкостей работы в SPSS. Еще раз подчеркиваем, что это пособие не для искушенных пользователей интеллектуальных продуктов, которые, не глядя, используют любое программное обеспечение.

Авторы пособия не могли использовать получивший сегодня широкое распространение замечательный жанр работ «для чайников», к которым они, по большому счету, себя и относят. Их восприятие пакета SPSS – это восприятие социологов-прикладников, а не разработчиков программного обеспечения или специалистов, практикующих математические методы в социальных науках. Поэтому в пособии **передача опыта и выработка навыков использования пакета основаны на принципах работы по образцу:** «**делай вместе с нами - делай как мы**».

Можно соглашаться или отвергать такой подход. Тем не менее, как мы убедились на собственном опыте и по благожелательным отзывам наших читателей, он довольно хорошо работает в современных условиях. Известно, что многие пользователи, начинающие работу в SPSS, еще осваивают основы текстового процессора Word для Windows и находятся далеко не на «ты» с самой операционной системой Windows. Это реальная проблема многих тысяч людей, и именно им мы бы хотели передать наш опыт.

Вряд ли имеются какие-либо веские основания полагать, что указанное положение дел резко изменится в обозримом будущем. Напротив, разворачивающийся процесс массовой компьютеризации позволяет сделать предположение, что в ближайшее время потребность в таких учебных пособиях и спрос на них будет нарастать.

Материалы исследований, выполненных авторами и их коллегами в 1991-2003 гг., служат основой всех приведенных в пособии примеров и расчетов. Они связаны с российской действительностью и отражают процессы, происходящие в нашей стране. Это обстоятельство весьма

существенно в преподавании. Список файлов, используемых в пособии, приведен в приложении 9.

Среди этих материалов особо следует выделить: базу данных (файл в формате SPSS), содержащую первичную информацию (по трем волнам – 1995, 1996 и 1997 г. панельного исследования, связанного с изменениями условий жизни сельского населения России), а также файл с ее текстовым описанием. Эти два файла (http://www.icpsr.umich.edu/cgi-bin/archive.prl?path=ICPSR&num=2816 и http://www.icpsr.umich.edu/cgi/ab.prl?file=2816) доступны в Интернете (для зарегистрированных пользователей) по сетевому адресу Мичиганского межуниверситетского консорциума политических и социальных исследований (Michigan Inter-University Consortium for Political and Social Research - ICPSR - Университет штата Мичиган, Анн Арбор, США).

Ссылку на сервер ICPSR можно найти и по сетевому адресу Университетской информационной системы «Россия» (http://www.cir.ru).

При подготовке данного издания мы пользовались информационным обеспечением замечательного статистического портала «Statsoft», содержащего богатую коллекцию учебных пособий и аналитических материалов о современном анализе данных. Сетевой адрес этого портала в Рунете: http://www.statsoft.ru/home.

В интернете информация об авторах предлагаемого учебного пособия и выполняемых ими исследованиях может быть получена:

- на сайте Лаборатории социальной инфраструктуры ИСЭПН РАН, в которой работают авторы. Сетевой адрес этого сайта: http://www.isesp-ras.ru/labinfra.htm - (языки русск. и англ.);

- на персональном сайте проф. В.В. Пациорковского – руководителя совместного российско-американского проекта с российской стороны и В.В.Пациорковской. (http://host.iatp.ru/~patsiorkovsky);

- на сайте проф. Давида О'Брайна (университет Миссури-Колумбия, США) – руководителя ранее упомянутого проекта с американской стороны (http://www.ssu.missouri.edu/Faculty/DObrien/Russian VillageProject.htm);

***- на сайте доктора Ларри Дершема - ведущего исполнителя проекта с американской стороны в университете Миссури-Колумбия, США (http://www.ssu.missouri.edu/Rural_Transitions/vilo.htm).***

Считаем уместным сказать слова признательности своим коллегам по ИСЭПН РАН, постоянно проявлявшим живой интерес к этому проекту, нашим первым учителям по SPSS - докторам Давиду О'Брайну и

Ларри Дершему, а также соавтору по первой книге А.И. Петровой.

В подготовке данного издания важную роль сыграла поддержка, оказанная авторам Программой Фулбрайта (The Fulbright Scholar Program). Благодаря гранту, представленному этой Программой, мы получили возможность выполнить в Университете Миссури-Колумбия осенью 2000 – весной 2001 гг. большой объем работ, связанных с предлагаемым учебным пособием. Этому благословенному месту также хочется сказать самые теплые слова признательности.

Выражаем искреннюю благодарность нашим детям. Они консультировали нас по всем вопросам, возникавшим при работе над пособием. Для молодежи использование различных программных продуктов уже рассматривается как норма повседневной жизни. Это заметно отличает ее от многих представителей средних и особенно старших возрастных групп.

Все предложения и замечания, связанные с предлагаемым изданием, в том числе и по вопросам приобретения файлов с данными, на которые имеются ссылки в пособии (приложение 9), просьба посылать по следующим электронным адресам: **patsv@mail.ru**, **patsv@elnet.msk.ru**, или по почте:

Пациорковскому В.В
117218, Москва, Нахимовский проспект, д. 32, ИСЭПН РАН.
Телефон для связи в институте: (095) 332-45-34,
факс: 129-0801,
мобильная связь 8 910 402 1311.

**Дружеский совет**

***К освоению любого программного продукта лучше всего подходть с раскрепощенным сознанием и открытой душой, как в медитации или молитве. Залог успеха здесь состоит в принятии предлагаемой логики и интеллекта, а не в их оценке или критике.***

# Раздел 1

# ВВОД И КОНТРОЛЬ ДАННЫХ

## Глава 1. Подготовка к работе в SPSS

### 1.1. Основные сведения о программе

В начале 1970-х годов Норман Най (Norman Nie), Дейл Бент (Dale Bent) и Хэдлай Халл (Hadlai Hull) зарегистрировали торговый знак SPSS® statistical software. Компания с одноименным названием была создана ими в 1968 г. В 1975 г. компания была преобразована в корпорацию с главным офисом в Чикаго (Chicago, IL USA). За годы существования корпорацией было разработано множество программных продуктов, в том числе и SPSS/PC+™, первая версия которого появилась в 1984 г.

Сегодня **SPSS** – это программный продукт и одновременно защищенная торговая марка всемирно известной американской фирмы SPSS Inc., правление которой так и остается в Чикаго. Этот пакет занимает ведущее положение среди программ, предназначенных для статистической обработки информации. Вместе со всем программным обеспечением указанного профиля он прошел большой путь эволюции: сначала от первых версий SPSS для больших ЭВМ, до версий, ориентированных на PC-DOS/MS-DOS, а затем до версий, работающих в среде Windows. Эта эволюция уже описана в ряде других работ (7-8, 15-21), обратившись к которым можно удовлетворить свое любопытство.

Полезную информацию о самом пакете SPSS можно найти на фирменном портале (http://www.spss.com). Этот портал открывает доступ к большому числу корпоративных и тематических домашних страниц SPSS на многих языках мира. Увы и ах, пока еще не на русском.

На этом портале имеются тематические обобщения использования программных продуктов семейства SPSS в биологии, медицине, инженерии и других областях. Мы надеемся, что наша работа поможет обобщить подобный опыт использования SPSS в социологии.

Использование в интернете поисковой системы Google (http://www.google.com) сразу же позволяет выйти на эту домашнюю страницу, тогда как поисковые системы Alta Vista (http://www.altavista.com) и Yahoo (http://www.yahoo.com) открывают списки всего семейства домашних страниц SPSS, соответственно, на 18000 и 17400 наименований.

В Рунете наиболее полная информация об SPSS может быть найдена на портале российского представительства SPSS (http://www.spss.ru). Здесь можно найти много полезного и интересного, но только за деньги. Причем, все сделано на солидной коммерческой основе. Точно так, как в популярном сюжете: «утром деньги - вечером стулья».

В целом вряд ли будет большим преувеличением сказать, что SPSS, наряду, скажем, с такими пакетами как SAS и STATISTICA (22), - это настоящее и будущее статистической обработки социологической информации. Особенно в этом отношении продуктивны его последние версии, преодолевшие ограничения на использование русского языка и одновременно открывшие широкие возможности моделирования в Amos (23).

SPSS - универсальная система статистического анализа и управления данными. Эта аббревиатура первоначально означала Statistical Package for the Social Science (статистический пакет для социальных наук). Затем исходной аббревиатуре было дано новое толкование: Superior Performance Software System (система программного обеспечения высшей производительности).

Первая версия SPSS для Windows имела порядковый номер 5.0. Затем последовали версии 6.0, 6.1, 7.0, 7.5, 8.0, 9.0 и, наконец, 10.0 и 11.5 и выше. Начиная с SPSS версии 7.0, оболочкой служит минимум Windows95 (NT).

Пакет, наряду с использованием своего собственного типа данных, может считывать данные практически из любых типов файлов и использовать их для создания отчетов в форме таблиц, графиков и

диаграмм, а также вычислять описательные статистики, производить сложный статистический анализ и моделирование.

В начале наших работ (1991-1993 гг.) по совместному российско-американскому проекту (27, С. 6) анализ и обработка собранной первичной информации выполнялись с помощью пакета интегрированных программных средств SPSS/PC для DOS. Занятие это было довольно сложное и трудоемкое. Отсутствие таблицы, соответственно, строчный ввод данных и необходимость использования командного языка - все это делало отношения с пакетом весьма напряженными.

Начиная с 1995 г. мы перешли к использованию пакета SPSS для Windows. В то время основной была версия SPSS 6.1, а в 1997 г. получила распространение версия SPSS 7.5.

Базовая версия SPSS 7.0 была разработана для IBM PC и совместимых с ними компьютеров, имеющих 16 Мбайт оперативной памяти (RAM) и от 30 Мбайт свободного пространства на жестком диске (HDD). Более поздние версии SPSS стали предъявлять довольно жесткие требования к техническим характеристикам PC. К примеру, основными системными требованиями для инсталляции SPSS 10.0 являются:

- Windows 95, Windows 98, Windows NT 4.0 или Windows 2000.
- Процессор Pentium 90 МГц (или более).
- Минимум 16 Мбайт оперативной памяти (рекомендуется 64 Мбайт).
- Минимум 80 Мбайт свободного места на жестком диске (для базовой системы) и еще 80 Мбайт для работы SPSS.
- Привод CD-ROM.
- Видеокарта с минимальным разрешением 800*600 (SVGA).

Еще для инсталляции требуются серийный номер и лицензионный код. Лицензионный код дает возможность инсталлировать базовую систему и модули расширения SPSS, приобретаемые дополнительно.

Порядок инсталляции рассматриваемого пакета такой же, как и для всех других приложений, работающих под управлением операционной системы Windows 95 и ее более поздних версий. Поэтому в данной работе он не рассматривается. Иными словами, мы считаем, что инсталляции программных средств, работающих под Windows, надо учиться несколько ранее, осваивая основы компьютерной грамоты на PC.

SPSS представляет дружественный пользовательский интерфейс, который делает процесс ввода и статистического анализа доступным для начинающего и удобным для опытного пользователя. Редактор

данных пакета позволяет удобно (табличным способом) вводить и корректировать входные данные. SPSS дает возможность получать множество высококачественных графиков и различных диаграмм. С помощью пакета, используя таблицы, простые меню и диалоговые окна, можно выполнять, во-первых, анализ огромных файлов данных с тысячами переменных, и, во-вторых, делать все это без строчной записи команд в языке программирования.

Тем не менее, такой язык под именем «синтаксис» (Syntax) имеется в SPSS. Потребность в его использовании появляется в случаях возникновения у пользователя специальных задач, выходящих за рамки предлагаемого стандартного сервиса, а также при работе по моделированию в Amos. К этому вопросу мы еще вернемся в третьем разделе пособия.

Возможные области применения SPSS: хранение и анализ данных опросов, маркетинговых исследований и продаж, финансовый анализ и др. В социологии пакет позволяет автоматизировать процесс создания баз данных социологической информации, их хранение и обработку. Использование пакета оказывает и обратное влияние на полевую документацию и методы социологического исследования. Например, возможность автоматического построения новых переменных позволяет отказаться от включения в полевую документацию различного рода обобщающих и итоговых индикаторов, а механизация процесса кодировки открывает широкие горизонты для использования открытых вопросов взамен закрытых там, где это более эффективно.

Решение экономических и статистических задач содержит много рутинной работы по сбору и обработке информации. Такие задачи тяготеют к использованию вычислительной техники, хотя многие сложившиеся специалисты и преподаватели вузов продолжают по старинке читать курсы, писать статьи и книги по обработке статистической информации в отрыве от возможностей, предоставляемых вычислительной техникой и программным обеспечением.

Сегодня в решении задач, связанных со сбором и обработкой экономической и социологической информации на персональном компьютере (PC), используются различные отечественные и зарубежные программные продукты. Наиболее известны среди них «Да-система», Mathematica, MathCAD 2000 PRO, Minitab, S-Plus, SAS, SPSS, STATISTICA, Systat.

Любой из названных выше пакетов имеет свои достоинства и недостатки. Вместе с тем пользователь, не имеющий возможности создать

свой программный продукт, который отвечал бы всем требованиям решаемых им задач, вынужден выбирать из уже имеющегося программного обеспечения. Если такая возможность имеется, то это комфортно, правильно и хорошо.

Правда, теперь уже создается ситуация, когда для самого пользователя оказывается необходимым соответствовать определенным требованиям. Без этого ему будет трудно правильно и по средствам сделать свой выбор. Вряд ли нужно доказывать, что это совсем другая ситуация и иные условия работы, по сравнению с теми, которые были ранее. Они предполагают иной тип и подготовки, и переподготовки исследовательского персонала. Настоящее учебное пособие отражает стремление тех, кто его подготовил, найти ответы на эти новые вызовы времени.

Прежде чем перейти к изложению опыта использования пакета SPSS при вводе, контроле и анализе данных эмпирического исследования уместно остановиться на основных моментах запуска пакета, управления в нем и выхода из него. Овладение ими не представляет особых трудностей и обеспечивает надежное использование пакета.

## 1.2. Запуск пакета

При работе пакета под операционной системой Windows 95 и ее более поздними версиями запуск пакета SPSS возможен посредством двойного щелчка левой клавиши мыши на его фирменном значке, который обычно устанавливается непосредственно на рабочем столе Windows или путем выполнения следующей последовательности команд:

**Start** (Пуск)
**Programs** (Программы)
**SPSS**
**SPSS** (фирменный значок).

Щелчок мышью на значке SPSS опять же ведет к запуску пакета.[1]

1. При использовании пакета под устаревшими версиями Windows 3.1 и 3.11 (для рабочих групп) он, как правило, запускается путем двойного щелчка левой клавиши мыши на фирменный значок пакета (SPSS), находящийся непосредственно в Диспетчере программ или в одной из его программных групп, которая была выбрана при установке пакета. Следует также помнить, что в этом случае могут использоваться только ранние версии пакета SPSS, такие как SPSS 6.0.

**Внимание!**

***Приведенная выше форма записи выполнения команд принята в документации SPSS. Она и использована в предлагаемом пособии.***

Следствием начала работы системы служит появление на экране небольшого окна со списком файлов (SPSS for Windows), которое располагается на фоне редактора данных (Data Editor), открывшегося пока еще с пустой таблицей (Untitled – SPSS Data Editor), готовой для ввода данных (рис. 1).

В ранних версиях SPSS эта таблица имела имя «Newdata». Оно указывалось в специальном подзаголовке окна редактора данных (1,С.12).

В последних доступных нам версиях SPSS 10.0 и 11.5 исходная картинка, которая появляется на экране монитора сразу после открытия программного продукта – **редактор данных**, содержит следующие основные элементы (рис. 1):

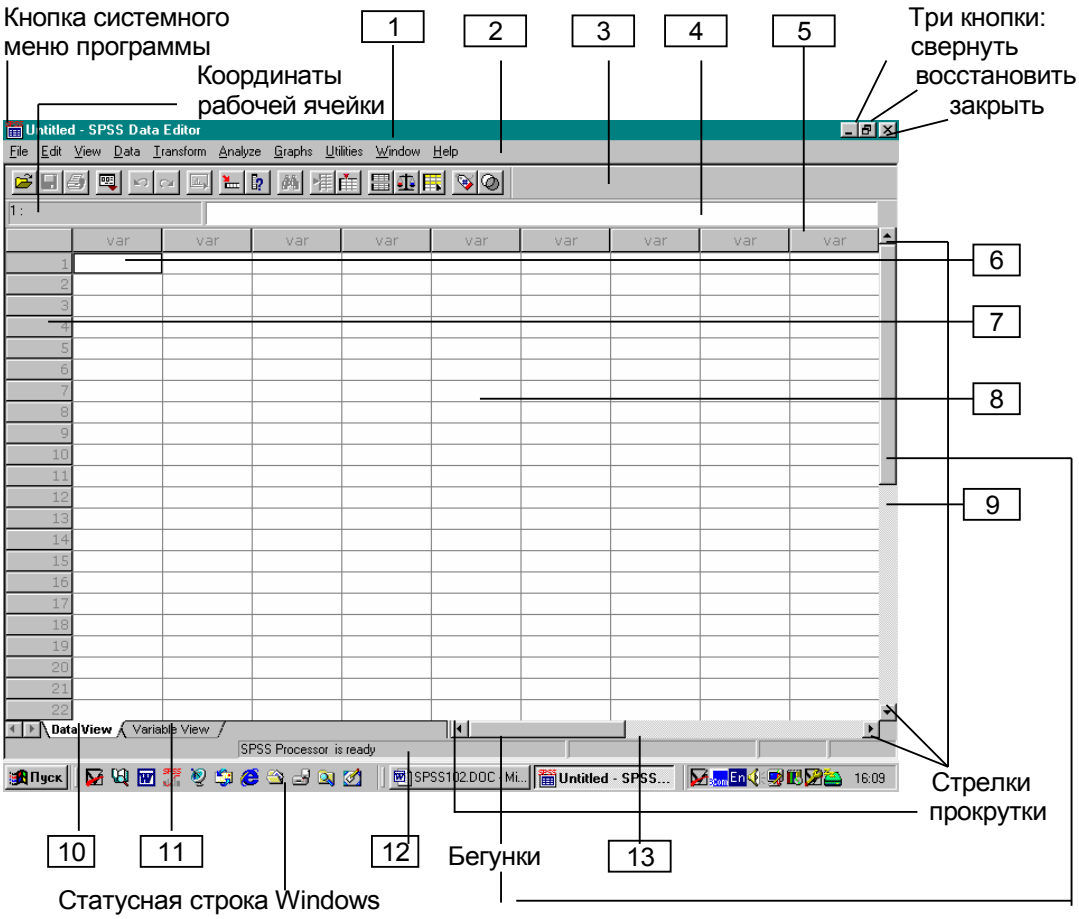



***Рис. 1. Редактор данных***

**1.** ***Заголовок редактора данных*** - это верхняя строка (панель) экрана, в левой части которой находятся: кнопка системного меню программы, имя открытого файла, далее после тире следует имя текущего рабочего окна. В правой части этой строки расположены три стандартные кнопки, характерные для основной массы приложений, работающих под операционной системой Windows.

**2.** ***Главное меню*** **–** следующая после заголовка (вторая сверху) строка экрана. Это очень важный и довольно глубоко структурированный компонент редактора данных. Поэтому его структура описана отдельно в § 1.3.

**3.** ***Стандартная панель инструментов*** - следующая строка после главного меню. Это еще один очень важный и довольно глубоко структурированный компонент редактора данных. Поэтому состав панели инструментов также описан отдельно в § 1.4.

**4.** ***Редактор ячеек*** - следующая после панели инструментов (четвертая сверху) строка, в которой находится горизонтально вытянутое окно. Оно используется при вводе и редактировании данных. Редактор ячеек - составной элемент окна редактора данных. Поэтому порядок работы с ним описан при представлении ввода данных в § 2.7.

**5.** ***Имена колонок в таблице*** – кнопки, которые зрительно составляют пятую сверху строку экрана. Непосредственно под ними находятся ячейки окна редактора. Как видно на рис.1 по умолчанию они имеют имя «var», т.е. сокращенно от variable - переменная. Порядок работы с переменными описан в следующей главе в § 2.3.

**6.** ***Контур рабочей ячейки*** - составной элемент окна редактора данных. Координаты рабочей ячейки всегда выведены на экране в левой части строки редактора ячеек. На рис. 1 в этом месте можно ви-деть 1. Это значит, что рабочей ячейкой в данный момент является первая ячейка по строке. Ее координаты по столбцу на рис. 1 отсутствуют. Связано это с тем, что ячейка пуста. Как только ячейку будет введено любое значение переменной, ее координаты приобретут следующий вид: 1: var00001, который говорит о том, что на данный момент имя этой переменной принято по умолчанию. Естественно, если переменной будет дано имя, то оно и появится в описании координат этой ячейки и всех других ячеек этой колонки (столбца). Порядок работы с рабочей ячейкой описан при представлении ввода данных в главе 2,§ 2.7.

**7.** ***Порядковая нумерация строк в таблице*** – пронумерованные, начиная с единицы, кнопки, которые вертикально расположены в левой части экрана. Непосредственно справа от них находятся ячейки

окна редактора. Нумерация этих кнопок устанавливается системой по умолчанию. Она не может быть изменена пользователем. Поэтому пользователю полезно всегда иметь в первой колонке свой собственный идентификационный номер для каждого вводимого в строку случая. Более полно порядок работы со строками описан в главе 2, § 2.3.

**8.** ***Окно редактора*** – основная, таблично оформленная (с ячейками) часть экрана. При этом одна из ячеек всегда имеет контур, т.е. является рабочей. Визуально окно редактора воспринимается пользователем как электронная таблица, но по существу оно таковой не является. Всякий, кто имеет хотя бы небольшой опыт работы, а значит и возможность сравнения таблиц, например, в Excel и SPSS, быстро увидит заметные различия как ввода, так и обработки данных в этих двух программных продуктах. Порядок работы в окне редактора описан в § 1.5 этой главы и во второй главе § 2.1-2.2.

**9.** ***Прокрутка массива данных по вертикали*** - типовой инструмент просмотра данных по вертикали, характерный практически для всех приложений, работающих под Windows (например, MS Word). Основная особенность устройства этого инструмента состоит в том, что размер его подвижной части (кнопка спуска-подъема или бегунок) всегда обратно пропорционален размеру массива введенного в редактор данных. Иными словами, она тем меньше, чем больше массив. Приход подвижной части прокрутки в нижнее положение обязательно дает возможность видеть последнюю запись случая, введенного в таблицу.

**10.** ***Ярлык просмотра данных окна редактора*** – это закладка Data View (просмотр данных) в левой нижней части экрана. Закладка не находится в рабочем состоянии, всегда имеет заливку (фон). Напротив, находясь в рабочем состоянии, закладка сливается с белым фоном окна редактора. Порядок работы в этом режиме описан в главе 2, § 2.2.

**11.** ***Ярлык просмотра переменных окна редактора*** - это закладка Variable View (просмотр переменных) в левой нижней части экрана. Она находится непосредственно после закладки просмотра данных и, по сути, представляет ее альтернативу. Это значит, что нельзя иметь одновременно открытыми обе эти закладки. Одна из них всегда сменяет другую. Порядок работы в этом режиме описан в главе 2, § 2.2. Наличие закладок Data View (10) и Variable View (11), благодаря которым облегчается как переход между полем ввода данных и описания переменных, так и само описание переменных, одно из

существенных качественных отличий версий SPSS 10.0 и выше от предшествующих версий.

**12.** ***Статусная строка.*** Она расположена в нижней части окна редактора. При вводе сюда курсора появляется надпись «Information area». В этой строке находится информация о текущем состоянии процессора SPSS. При входе в SPSS строка содержит текст: «Starting SPSS Processor» (запуск процессора SPSS). Индикатором готовности системы к работе служит появление надписи «SPSS Processor is ready» (процессор SPSS готов). В следующих за указанной надписью маленьких окнах в ходе расчетов можно наблюдать просмотр системой базы данных по числу случаев. Здесь все, однако, зависит от сложности задачи и быстродействия машины. При ограниченном по числу случаев и переменных массиве, а также хорошем (по быстродействию и объему памяти) компьютере в этих окошках можно заметить лишь мелькание цифр.

**13.** ***Прокрутка массива данных по строке*** - типовой инструмент просмотра данных по горизонтали, характерный практически для всех приложений, работающих под Windows (например, MS Word).

В последних версиях, как видно на рис. 1, самая верхняя строка – заголовок редактора данных, содержит в левом углу условное (данное системой по умолчанию) имя открытого файла «Untitled – SPSS Data Editor», которое указывает на то, что открытый файл еще не получил своего имени. Это имя должен присвоить сам пользователь по результатам работы с файлом. У него, правда, есть возможность открыть (загрузить) один из уже имеющихся файлов. Естественно, он раскроется со своим именем, которое и появится в заголовке редактора.

**Общая информация**

***В отличие от других программных продуктов SPSS пока еще русифицирован лишь частично. Поэтому в нем приходится иметь дело с командами и указаниями только на английском языке.***

Прежде чем перейти непосредственно к описанию основных составляющих редактора данных полезно обратить внимание на тот факт, что он является исходным, но отнюдь не единственным средством работы в SPSS. Другими, не менее важными средствами работы в этом программном продукте, являются:

- окно просмотра (Viewer),
- окно просмотра текста (Text Viewer),
- редактор сводных таблиц (Pivot Table Editor),
- редактор диаграмм (Diagram Editor),
- редактор текстового вывода (Text Output Editor),
- редактор синтаксиса (Syntax Editor),
- редактор сценариев (Script Editor).

Все эти средства открываются в особых окнах. У каждого из них есть собственный интерфейс (меню, панель инструментов и т.п.), который, имея сходный внешний вид, тем не менее, заметно отличается от интерфейса редактора данных и друг от друга характерным только для данного специализированного средства набором команд главного меню и панели инструментов.

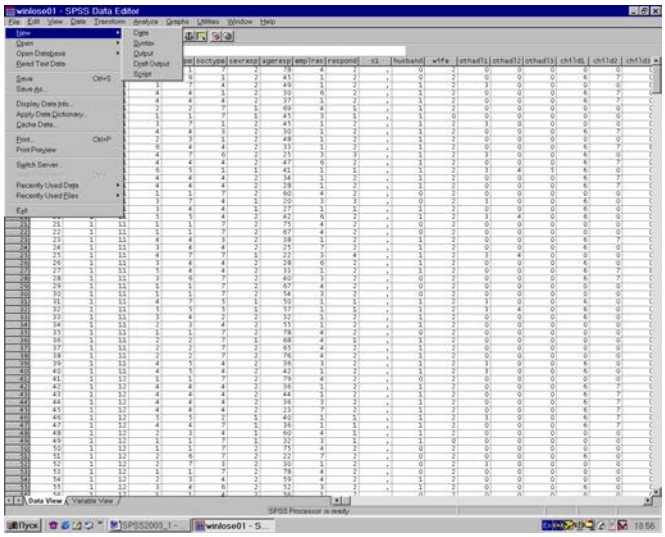


***Рис. 2.***

***Выпадающее окно File – New***

На рис. 2 хорошо видно, как после использования последовательности команд главного меню File – New открывается выпадающее окно. Оно и позволяет получить доступ к некоторым из перечисленных выше программных средств и встроенных модулей SPSS.

Одни из этих функциональных средств (например, окно просмотра и редактор синтаксиса) будут рассмотрены в этом пособии, в то время как другие оставлены для самостоятельного освоения пользователем, уже имеющим первичную подготовку.

## 1.3. Главное меню

Управление в SPSS, как и во многих других приложениях, работающих под операционной системой Windows, осуществляется через главное меню и панели инструментов. В версиях SPSS 10.0 и 11.5 главное меню содержит 10 основных командных кнопок со словесными заголовками (File, Edit и др.). Все они хорошо видны на рис. 1 (вторая строка сверху вниз). В ранних версиях главное меню имело 9 основных командных кнопок (1, С. 16).

Команды главного меню бывают двух видов: прямого действия (вызов такой команды ведет к ее выполнению) и опосредованного (вызов такой команды ведет к открытию дополнительного выпадающего меню с командами прямого действия). Характерным признаком такой команды служит наличие справа от ее имени стрелки, которая так же развернута вправо и свидетельствует о присутствии в данном случае выпадающего меню. Ниже приведено общее описание команд главного меню (слева направо в порядке их размещения).

**File** (файл). Этот пункт меню используется для создания новых файлов или открытия существующих, а также для чтения таблиц и баз данных, созданных другими приложениями.

Так, если открыть этот пункт меню, нажав на File, появляется окно со следующим меню: New (создать новый файл), Open (открыть файл), Open Database (открыть базу данных) – все эти три команды имеют собственное выпадающее меню. Далее следуют: Read Text Data (читать текстовые данные), Display Data Info (выдать информацию о данных), Save (сохранить), Save As (сохранить как), Apply Data Dictionary (директории прикладных данных), Cache Data (временная копия открытого на данный момент файла), Print (печать), Print Preview (просмотр перед печатью), Switch Server (переключатель сервера). Все они - команды прямого действия.

Наконец, три последние команды: Recently Used Data (текущие используемые данные), Recently Used Files (текущие используемые файлы). Обе они имеют выпадающие меню. И завершает список команд пункта меню File - Exit (выход).

После запуска SPSS и появления в окне редактора данных пустой таблицы с именем Untitled, доступны далеко не все из перечисленных выше команд. Так, например, в этом случае еще нечего сохранять и нечего печатать. Соответственно, в пункте меню File все связанные с этими действиями команды не работают. Напротив, при создании но-

вого файла или загрузке уже существующего, в рассматриваемом пункте меню открывается доступ к командам: Save, Save As, Print и др. Отсюда можно сформулировать следующее общее правило работы в системе SPSS:

**Правило 1.**

***Только при создании нового файла или загрузке уже существующего файла становятся доступными все команды меню.***

Если пользователь не знает этих особенностей SPSS, то он очень быстро оказывается в ситуации, которая рождает у него в голове примерно следующие мысли: «либо что-то испортилось в машине или программном продукте, либо его создатели сделали что-то не совсем так, как надо». Появление таких мыслей - надежное свидетельство отсутствия у пользователя основных навыков работы с программными средствами, а отнюдь не обнаружение ошибок или недоработок в программном продукте.

**Edit** (правка). Этот пункт меню используется для корректировки содержимого окна редактора данных. Здесь предоставлена возможность вырезать (Cut), копировать (Copy), вставлять (Paste) и очищать (Clear) строки, колонки и их фрагменты, а также осуществлять поиск данных (Find), устанавливать параметры (Options) и т.д. Мы еще не раз вернемся к рассмотрению порядка использования команд этого меню в разделах, описывающих работу с переменными.

**Veiw** (вид). Указанный пункт меню имеет множество опций, с помощью которых можно производить индивидуальную настройку редактора данных. В частности он позволяет показать или скрыть строку состояния (Status Bar).

Через меню «вид» открывается доступ к панелям инструментов (Toolbars): самого редактора данных, окна просмотра, редактора диаграмм и др. Здесь же имеется возможность изменять число значков на панели инструментов всех редакторов SPSS, в том числе и непосредственно редактора данных.

Кроме того, с помощью пункта меню «вид» можно выбрать другой тип начертания и размер шрифта (Fonts), включить или отключить отображение линий сетки (Grid Lines), указать метки значений (Value Labels) и переменные (Variables).

**Data** (данные). Этот пункт меню используется для осуществления изменений данных. Среди таких изменений в первую очередь можно

отметить: объединение данных, находящихся в различных файлах, создание новых независимых выборок, поиск данных в массиве первичной информации и др. К этому пункту меню приходится обращаться довольно часто:

- Для создания независимой выборки (отбора подмножества наблюдений), которые будут участвовать в дальнейшем анализе, полезно воспользоваться специальной командой Select Cases (отбор случаев).
- Для упорядочивания данных по возрастанию или убыванию используется команда Sort Cases (сортировка случаев).
- Если нужно найти в массиве данных наблюдение, то быстрее всего это можно сделать с помощью команды Go to Case (идти к случаю).
- При постановке задачи, требующей объединения данных, находящихся в разных файлах, используется команда Merge Files (объединить файлы).
- Для разъединения файла на два самостоятельных используется специальная команда Split File (разъединить файл).
- Для объединения случаев в группы путем создания нового файла используется команда Aggregate (агрегировать).

Из достаточно большого выбора возможностей этого пункта меню при анализе данных, по нашему опыту, наиболее часто и продуктивно используются процедуры Sort Cases и Select Cases, которые позволяют упорядочивать данные по нарастанию или убыванию, а также формировать независимые выборки. Более полно названные процедуры преобразования будут описаны ниже в главе 3, § 3.4-3.8.

**Transform** (преобразование). Этот пункт меню используется для изменения (перерасчета) выбранных переменных или для создания (вычисления) новых переменных на основе значений существующих первичных данных. Для преобразования переменных наиболее часто используются команды: Compute (вычислить), Count (подсчет), Recode (перекодировка). Более полно и последовательно названные процедуры преобразования будут описаны ниже в главе 4 (§ 4.1-4.3) и главе 5, § 5.1-5.2.

**Analyze** (анализ)[1] . Этот пункт меню используется для выбора различных статистических процедур, например, получения частотных таблиц, таблиц сопряженности, вычисления корреляции, линейной регрессии. Он содержит большой список различных статистических методов. Каждая из них заканчивается стрелкой, указывающей, что

1. В ранних версиях SPSS этот пункт главного меню назывался статистика – Statistics (1, С.15-16).

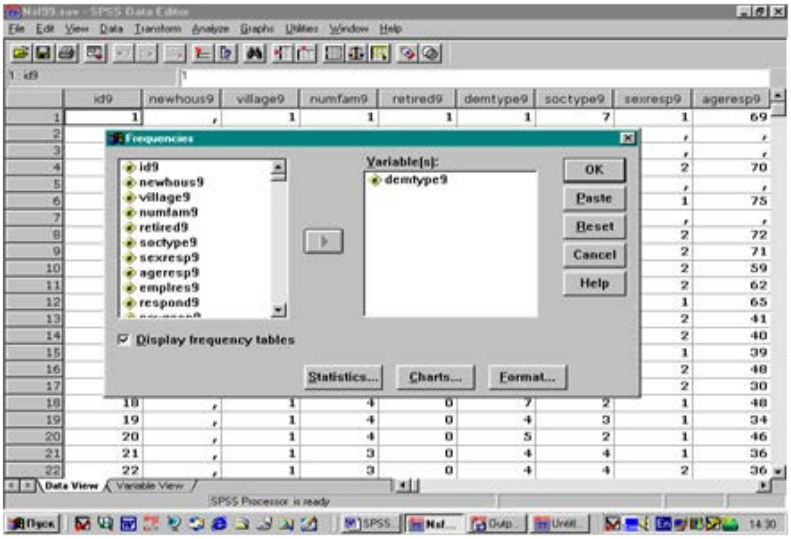
существует еще один уровень - подменю, в котором и перечислены конкретные статистические процедуры.

Например, раскрывая подменю Descriptive statistics (описательные статистики)[1] , получаем окно, содержащее следующие процедуры: Frequencies (частоты или частотное распределение), Descriptives (описания), Explore (исследование), Crosstabs (таблицы сопряженности).

Если выбрать статистическую процедуру и щелкнуть на ней мышью, то на экране появится главное диалоговое окно соответствующей процедуры. Поскольку все главные окна процедур выглядят практически одинаково, опишем главное диалоговое окно процедуры Frequencies (рис. 3).

***Рис. 3.***

***Главное диалоговое окно процедуры Frequencies***




Как видно на рис.3, это окно состоит из списка исходных переменных текущего файла данных (левая часть окна), списка выбранных для анализа переменных (Variables) и двух групп командных кнопок.

Правая часть окна имеет (по вертикали) пять кнопок. При нажатии на любую из них выполняется определенное действие: OK (О'кей) - начать выполнение процедуры, Paste (вставить), Reset (переустановить), Cancel (отменить), Help (помощь). Кроме того, в нижней части окна (по горизонтали) расположены еще три кнопки Statistics (статистики), Charts (диаграммы), Format (форматирование), открывающие доступ к дополнительным диалоговым окнам.

---

1. В ранних версиях SPSS это подменю называлось Summarize (1, С.15).

Маленький квадрат-box, в нижней (левой) части окна, имеет важное сервисное значение. Он установлен по умолчанию, что позволяет видеть в окне просмотра, которое описано далее в § 1.6, таблицу частотного распределения (Display frequency tables). При попытке убрать маркер из этого окошка система выдаст предупреждение о том, что в окне вывода не будет таблицы частотного распределения.

В этом случае в окне просмотра будут указаны только число значимых - валидных (Valid) случаев и число пропущенных (Missing) значений. Более полно и последовательно порядок работы с различными статистическими процедурами будет описан в главах 7-11 и 13-14.

**Graphs** (графики). Этот пункт меню используется для создания графиков и диаграмм. Их самостоятельно генерируют и отдельные процедуры, такие как Crosstabs, Frequencies, Regression и др. Диаграммы могут быть изменены при помощи редактора диаграмм (Chart Editor). Порядок построения графиков и диаграмм дан в главе 12.

**Utilities** (утилиты). Это в основном сервисный пункт меню. Он используется для отображения информации о переменных текущего файла (Variables), о текущем файле данных в целом (File Info), о меню редактора (Menu Editor), о выполнении сценария (Run Script) и др. Его полезно осваивать по мере накопления опыта и навыков работы в SPSS.

**Window** (окно). Этот пункт меню используется для сворачивания всех окон SPSS, открытых в текущей рабочей сессии (Minimize All Windows), а также работы с окнами путем перехода из текущего окна (Data Editor) в текущее окно просмотра (Viewer). Необходимо подчеркнуть, что в текущем окне SPSS может быть показан либо рабочий файл с данными, либо SPSS Draft Viewer с результатами анализа и расчетов, которые выполнены в эту сессию работы в SPSS.

Это значит, что в SPSS отсутствует возможность работы с несколькими окнами одновременно, которая имеется, например, в MS Excel и MS Word. Иными словами, открытие любого нового файла с данными автоматически ведет к сворачиванию предшествующего. Одновременно, указанное обстоятельство позволяет сформулировать еще одно правило, которое очень полезно помнить, приступая к работе в SPSS.

**Правило 2.**

***Любые открывавшиеся ранее (причем не только в текущую сессию) последние девять файлов с данными могут быть легко и быстро возращены в окно редактора путем комбинации команд: File — Recently Used Data — и далее клик мышью на имени интересующего пользователя файла с данными.***

Использование приведенной в правиле 2 комбинации команд делает работу в SPSS более комфортной, позволяя сохранить время и преемственность в анализе данных.

**Help** (справка)[1]. Этот пункт меню используется для получения справочной информации и запуска учебника по работе с SPSS. В 10-й и последующих версиях пакета справка включает следующие сервисные подменю: Topics (темы), Tutorial (обучающая программа), SPSS Home Page (домашняя страница SPSS), Syntax Guide (руководство по синтаксису), которое имеет дополнительное выпадающее меню, Statistics Coach (инструктор по статистике), About (О программе), Register Product (регистрационная карточка пользователя).

Справку в SPSS можно вызвать и другими способами. Например, нажать в любой момент работы функциональную клавишу F1 или, находясь в любом диалоговом окне, нажать переключатель с названием Help (справка) для получения помощи по текущей теме. Использование справки, а тем более обучающих программ, открывает большие возможности в освоении пакета.

Структурно этот пункт меню претерпевает самые заметные изменения от одной базовой версии к другой. В то же время его содержание в целом остается стабильным, а требуемые изменения часто запаздывают.

Вместе с тем для многих пользователей работа в этом пункте меню даже на родном языке (например, в пан-европейской версии Word) представляет известную сложность. Это тем более справедливо относительно доступности англоязычной справки. С целью облегчения доступа к справке мы даем ее содержание в приложении 7.

---

1. В версии пакета 6.0 справка включает следующие сервисные подменю: Contents (содержание), Search for Help on (поиск в справке), SPSS Tutorial (обучающая программа), SPSS Glossary (глоссарий), Help for SPSS/PC+ Users (справка для пользователей SPSS/PC+), Technical Support (техническая поддержка), What's New (что нового), About SPSS (об SPSS).

## 1.4. Панель инструментов

Панель инструментов (Toolbar) – один из важнейших элементов редактора данных. Как видно на рис. 1, она находится сразу под главным меню (третья строка сверху). Назначение панели инструментов - обеспечить быстрый доступ к наиболее часто используемым функциям. Содержание панели инструментов можно изменять. В стандартном варианте она состоит из 17 квадратных кнопок:

На каждую из этих кнопок нанесены символы соответствующей ей функции. Видимо, это и дает основание иногда называть панель инструментов – панелью символов.

В панели можно выделить основные инструменты (слева) и оконно-зависимые инструменты (в правой части). Оконно-зависимые инструменты доступны, когда активными являются либо окно данных, либо окна просмотра (вывода) и синтаксиса.

Первый (слева направо) из основных инструментов в 10-й и более поздних версиях SPSS имеет вид  . Он предназначен для открытия файла (Open File). Благодаря этой кнопке оказывается возможным сразу открыть требуемый файл, как бы отказавшись от комбинации соответствующих команд главного меню (File – Open – выпадающее окно со списком директорий с файлами).

Следующий инструмент, который имеет вид:  -сохранение файла из текущего окна (Save File). Далее следуют:

 - Печать рабочего файла (Print). Дает возможность печатать как весь документ, так и выделенную в нем область. Это особенно важно помнить при печати из текущего окна просмотра, в котором могут быть накоплены результаты большого числа промежуточных расчетов. Не выделив предварительно соответствующую область, пользователь отправляет в печать массу страниц и удивляется происходящему.

 - Вызов диалоговых окон (Dialog Recall). Выводит список последних 12 открывавшихся ранее диалоговых окон из таких команд

главного меню как Analize, Data, Transform. Благодаря этому открывается возможность прямого перехода к наиболее часто используемым статистическим процедурам.

 - Отменить ввод (Undo). Позволяет щелчком мыши убирать из редактора введенные в него ранее данные или их описание. Команда имеет большую глубину действия. Эта ее особенность позволяет легко и быстро вернуться к исходному пункту, требующему корректировки.

 - Вернуть ввод (Redo). Эта команда - обратная предшествующей. Она позволяет щелчком мыши возвращать в редактор ранее убранные данные или их описание. Обе команды (Undo и Redo) становятся доступными только при вводе и редактировании данных.

 - Переход к диаграмме (Go to Chart).

 - Переход к случаю с соответствующим номером (Go to Case Number). Позволяет перейти к определенному случаю, номер которого надо набрать в выпадающем диалоговом окне.

 - Информация о переменных (Variables). Открывает диалоговое окно с описанием выделенных переменных.

 - Найти (Find). Инструмент позволяет найти требуемое значение, которое записывается в специальном диалоговом окне. Оно открывается сразу же после щелчка по этой кнопке.

 - Вставить случай (Insert Case). Щелчок по этой кнопке автоматически ведет к вставке случая (дополнительной строки) над активной ячейкой. При использовании этого инструмента следует быть внимательным и осторожным. Введение нового случая без последующей набивки его значений ведет к росту числа отсутствующих значений и, следовательно, к искажению данных.

 - Вставить переменную (Insert Variable). Щелчок по этой кнопке автоматически ведет к вставке переменной (дополнительного столбца) слева от активной переменной.

 - Разделить файл (Split File). Инструмент позволяет разделить рабочий файл на два самостоятельных файла.

 - Взвесить случаи (Weight Cases).

 - Отобрать случаи (Select Cases). Открывает диалоговое окно, с помощью которого можно отобрать все случаи, удовлетворяющие заданному условию.

 - Метки значений (Value Labels). Этот инструмент позволяет переходить от отображения значений к их меткам.

 - Использование включателей (Use Sets). Очень сильная команда. Она контролирует доступность списка переменных во всех диалоговых окнах статистических процедур. При щелчке мышью по данной кнопке открывается диалоговое окно, состоящее из двух дополнительных окон. Левое из этих окон - пустое, а правое имеет две надписи.

Выделение и перенос надписи «ALLVARIABLES» из правого окна в левое (с последующим OK) ведет к довольно тяжелым последствиям. Оно очищает поля со списками переменных во всех статистических процедурах. При последующей попытке выполнения расчетов система информирует пользователя о недоступности списка переменных. Здесь надо быть очень внимательным и осторожным.

**Дружеский совет**

***Овладев панелью инструментов, можно реже обращаться к главному меню. Это сохраняет время и позволяет эффективно использовать команды прямого действия, заложенные в панели инструментов.***

Все основные команды SPSS могут быть выполнены не только с использованием главного меню и панели инструментов, но и посредством комбинации клавиш. Подобное выполнение команд идет в обход работы с мышью. Поэтому оно особенно эффективно, если мышь работает не совсем хорошо, что, как известно, случается довольно часто у рядовых пользователей. А в случаях коллективного доступа к компьютеру (в учебных классах) подобное положение вещей является скорее нормой, чем отклонением.

Вместе с тем использование комбинации клавиш на клавиатуре предполагает не только наличие хорошей памяти, но и знание основ операционной системы Windows. В качестве примера в обрамлении 1 приведены наиболее важные соответствия команд главного меню,

панели инструментов и комбинации клавиш.

Здесь опять же очень полезна предельная внимательность и осторожность. Так, например, команда главного меню View – Variables и соответствующая ей комбинация клавиш (Ctrl + T) имеют совсем другой характер, чем команда панели инструментов Variables. Последовательность команд View – Variables меняет вид редактора с представления переменных на представление данных, а выполнение команды панели инструментов Variables открывает окно со списком переменных и полем, в котором описывается содержание выделенной переменной.

**Обрамление 1. Основные соответствия команд главного меню, панели инструментов и комбинации клавиш**

| Главное меню | Панель инструментов | Комбинация клавиш |
|---|---|---|
| File - Open | Open File | Ctrl + O |
| File - Save | Save File | Ctrl + S |
| File - Print | Print | Ctrl + P |
| Edit - Undo | Undo | Ctrl + Z |
| Edit - Redo | Redo | Ctrl + R |
| Edit - Find | Find | Ctrl + F |
| View – Data /Variables | Data View, Variable View (переключате-ли в нижнем левом углу редактора данных) | Ctrl + T |

И, конечно, в SPSS, как и во всех приложениях к Windows, действуют такие коронные комбинации клавиш как Ctrl + X (Cut - вырезать), Ctrl + C (Copy -копировать), Ctrl + V (Paste - вставить), которые дублируют соответствующие команды главного меню Edit, но отсутствуют в стандартном варианте панели инструментов.

## 1.5. Редактор данных

Редактор данных обеспечивает удобный способ создания и редактирования файлов данных в форме таблицы. Таблица является прямоугольной. Ее размеры определяются числом наблюдений (случаев) и переменных. Для целей выборочных обследований размеры таблицы практически, как отмечалось выше, не имеют ограничений.

Данные можно вводить в любую ячейку. Если данные вводятся в ячейку за пределами границ таблицы, система автоматически расширит прямоугольник данных, чтобы включить все строки и/или колонки между ячейкой и границами таблицы. Связано это с тем, что в пределах границ таблицы не может быть пустых ячеек.

Отмеченный выше момент имеет существенное значение, поскольку для числовых переменных пустые ячейки конвертируются в системные пропущенные значения, а для строковых переменных пустые ячейки не являются системными пропущенными значениями и имеют тот же статус, что и заполненные ячейки. Более полно последовательность действий, выполняемых при создании и редактировании файлов данных, рассмотрена в главе 2, § 2.3 - 2.5. Здесь же важно усвоить следующее правило:

**Правило 3**

***Отсутствие пустых ячеек в пределах границ таблицы безусловное требование для всех видов работ от оформления таблиц до построения графиков***

## 1.6. Окно просмотра

Окно просмотра (Viewer) появляется сразу после выполнения статистической процедуры. В ранних версиях оно называлось Output или окно вывода (1, С. 19-21). Это окно предназначено для просмотра результатов производимых расчетов. После выполнения в текущую сессию первого комплекса расчетов открывшееся окно имеет заголовок Output1 – SPSS Viewer (рис. 4).

Если далее в текущую сессию при выполнении последующих расчетов окно только сворачивается, но не закрывается, то каждый новый результат вычислений помещается по очереди в конце окна просмотра. В том случае, если после просмотра окно будет закрыто, то результаты следующих вычислений будут открыты в новом окне с заголовком Output2 – SPSS Viewer.

Ранее уже отмечалось, что «самым слабым местом в системе было и остается окно вывода» (4, С. 121). Опыт показывает, что наши оценки совпадали с направлением поиска разработчиков системы. Окно просмотра (Viewer) как бы явилось их ответом на отмеченную выше «зло-

бу дня» в последних версиях системы. В этих версиях окно просмотра (Viewer) выполняет функцию оболочки для существовавшего ранее независимо окна вывода (Output). Мы считаем, что это нужный шаг в правильном направлении.

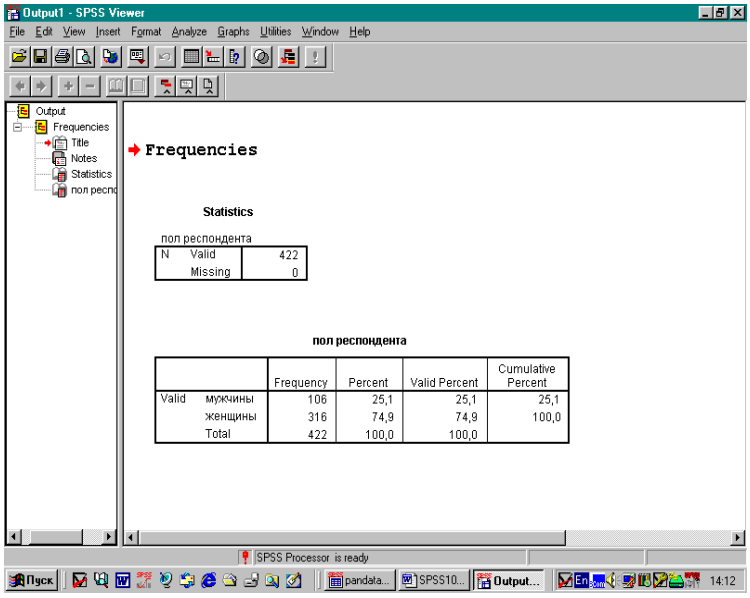



***Рис. 4.***

***Окно просмотра***

Окно просмотра состоит из главного меню, панели инструментов и двух подокон. Главное меню и панель инструментов окна просмотра заметно отличаются от идентичных элементов окна редактора. Указанные функциональные различия могут быть легко освоены в ходе самостоятельной работы.

В правом подокне помещаются таблицы с результатами расчетов и построенные графики. В левом подокне находится обзор содержания результатов, которые отображаются в виде отдельных блоков. Ширину подокон можно изменять перетаскиванием разделительной границы при помощи мыши.

Каждый блок левого подокна имеет имя выполненной процедуры и свой значок. Этому значку предшествует небольшой прямоугольник со знаком (-). Внутри каждого блока имеется заголовок (Title) и примечания (Notes). Далее идет перечисление элементов блока с соответствующими значками.

Благодаря такой иерархической конструкции блоков, можно производить поиск необходимых элементов, переставлять их местами, копировать, удалять и т.д. Здесь также есть весьма интересная возможность свернуть, не уничтожая, любой блок.

Для этого надо щелкнуть мышью на прямоугольнике со знаком (-). Знак поменяется на (+), а все расчеты, таблицы этого блока, выведенные в правой части окна просмотра, исчезнут. Чтобы их вернуть, надо повторно щелкнуть мышью на прямоугольнике, тогда. (+) превратится в (-), а вся информация этого блока расчетов вновь появится в окне просмотра.

В целом конструкция левого подокна идентична конструкции проводника Windows Explorer в Windows. Полученные в окне просмотра данные статистических расчетов можно редактировать и сохранять в текстовом файле для последующего использования (просмотра, печати и др.).

Сохранение текстового файла аналогично сохранению файла данных (глава 3, § 3.3). По умолчанию при сохранении (Save Data) текстового файла он получает расширение **.spo** (в ранних версиях .lst), но при использовании команды Save as (сохранить как...) можно задать любое из предлагаемых системой расширение.

Открытие сохраненного текстового файла аналогично открытию файла данных, но с той лишь разницей, что в меню необходимо выбрать:

**File**

**Open**

**Output.**

После выбора требуемого файла и нажатия кнопки OK (выполнить команду), выбранный файл появится на экране в окне просмотра, и имя окна вывода станет именем файла. Чаще всего окно просмотра используется только для получения результатов статистических процедур и сохранения их в файле. Последующая работа с файлом осуществляется в других текстовых редакторах (например, Word), которые удобнее, многофункциональнее и более приспособлены для работы с текстами, вставками, таблицами, графиками и т.п.

В ранних версиях окно вывода открывалось автоматически, с самого начала сеанса работы в SPSS, и располагалось сразу за окном редактора данных. В первом издании настоящего учебного пособия окно вывода хорошо видно на рис. 1 (1, С. 12).[1]

---

[1] После выполнения статистической процедуры в версии 6.0 результат работы в виде таблиц или последовательности чисел помещается в окно вывода. При этом рассматриваемое окно становится активным, т.е. перемещается на передний план, закрыв собой окно редактора данных (саму таблицу с первичной информацией).

При печати из текущего окна вывода, в котором могут быть накоплены результаты большого числа промежуточных расчетов, как уже отмечалось ранее (глава 1, § 1.4), следует руководствоваться следующим правилом:

**Правило 4.**

***Печати из текущего окна просмотра всегда должно предшествовать выделение блока (блоков), подлежащих печати.***

Одной из важных особенностей окна просмотра в последних версиях SPSS является возможность преобразования выводимых данных (таблиц и графиков) в форматы, поддерживаемые браузерами (соответственно, с расширением **.html** для таблиц и **.jpg** для графиков-рисунков), что позволяет публиковать их в сети интернет с минимальными дополнительными затратами труда и времени. Такого рода трансформация осуществляется с помощью команды главного меню окна просмотра **File-Export**, а также кнопки **Export,** находящейся на панели инструментов окна просмотра. Более того, если ваш компьютер имеет сетевое подключение, то теперь непосредственно из окна просмотра можно отправить письмо, используя команды: **File-Send Mail**.

## 1.7. Завершение сеанса работы в SPSS

Для окончания работы в SPSS необходимо сначала закрыть все рабочие окна (сохранив или уничтожив находящиеся в них данные), а затем выбрать в главном меню следующую последовательность команд:

**File**

**Exit.**

При работе под Windows 95 и последующими версиями этой операционной системы для закрытия пакета можно использовать кнопку (со значком «X») команды «закрыть» управляющего меню, находящуюся в строке заголовка (правый верхний угол экрана). После нажатия этой кнопки система обязательно задаст вопрос относительно необходимости сохранения данных текущего рабочего файла. Автома-

тическая реакция или пренебрежение этим вопросом может дорого стоить начинающему или забывчивому пользователю.

В случае необходимости временного выхода из SPSS, например, с целью копирования данных в Word (при подготовке отчета), очень удобно пользоваться кнопкой команды «свернуть» управляющего меню, также находящейся в строке заголовка в правом верхнем углу экрана. Но если файл с другим приложением уже открыт, то тогда можно и не трогать открытый редактор данных, окно просмотра или что-то другое из самой системы SPSS, а просто кликнуть мышью на имени нужного файла в нижней строке экрана. Точно таким же образом можно опять вернуться к работе в SPSS.

**Дружеский совет**

***Если Вы первый раз вошли в SPSS, то единственно полезный совет, который можно дать в таком случае — выйти из него и опять войти.Это поможет вам избежать неприятного для начинающего пользователя чувства, когда входишь в программу, а выйти из нее не можешь.***

## Задание для самостоятельной работы

1. Как можно открыть SPSS? Откройте и закройте программу.
2. Назовите основные элементы окна редактора данных.
3. Откройте файл с данными.
4. Создайте и сохраните файл с данными.
5. Как выглядит строка заголовка редактора данных после входа в SPSS?
6. Как выглядит та же строка редактора после открытия файла с данными?
7. Как выглядит строка заголовка окна просмотра после выполнения статистической процедуры?
8. Какие команды имеет главное меню окна редактора данных?
9. Сравните команды главного меню редактора данных и окна просмотра.
10. Опишите символы кнопок панели инструментов.
11. Куда ведет комбинация команд главного меню: Edit – Options?
12. Назовите основные расширения файлов, используемые в SPSS.
13. О чем следует помнить, начиная печать в окне просмотра?

14. В чем различие окна просмотра (Viewer) и окна вывода (Output)?

15. В чем состоит функциональное назначение закладок: просмотр данных (Data View) и просмотр переменных (Variable View) в SPSS?

16. Выполните команду Export окна просмотра.

17. Если можно, отправьте электронную почту из окна просмотра.

18. Какие дублирующие команды главного меню, панели инструментов и комбинации клавиш вы знаете?

19. В чем различие команд Variables в главном меню View и Utilities?

20. В чем различие команд подменю Data и Tpansform?

21. В чем различие команд подменю Edit и View?

22. В чем различие команд подменю Analyze и Graphs?

23. Какая кнопка панели инструментов соответствует комбинации команд главного меню File - Open?

24. Какая кнопка панели инструментов соответствует комбинации команд главного меню File - Print?

25. Какая кнопка панели инструментов соответствует комбинации команд главного меню Data - Select Cases?

26. Где в главном меню находится команда, которую дублирует кнопка Split File панели инструментов?

27. Где в главном меню находится команда, которую дублирует кнопка Value Labels панели инструментов?

28. Как в таблицу редактора данных SPSS можно вставить переменную?

29. Как в таблицу редактора данных SPSS можно вставить случай?

# Глава 2. Ввод данных

## 2.1. Общие замечания

Хорошо известно, что социологические исследования имеют довольно жестко регламентированную технологию выполнения работ. Без программы исследования вряд ли можно сделать полноценный инструментарий. В свою очередь, без инструментария нечего делать в поле. И только собрав первичную информацию, можно ставить задачи ее ввода, контроля, анализа и обработки.

Эта технология сохраняется и при обработке первичной информации с использованием пакета SPSS. Последовательность шагов, требуемая для решения задач социологического исследования, продолжает оставаться жестко заданной. Каждый шаг по-своему важен, его практически нельзя исключить или выполнить в другом порядке.

Например, нельзя вводить информацию, предварительно не закодировав ее, или пытаться выполнить статистический анализ, не проведя контроля введенных данных. Обработку собранных в поле данных лучше всего выполнять в приведенной ниже последовательности, поэтапно:

- подготовительный этап;
- ввод и корректировка данных;
- контроль данных;
- получение результатов статистических процедур;
- анализ данных и подготовка отчета.

В этой части пособия мы подробнее остановимся на нашем опыте использования SPSS для обработки социологической информации: при ее подготовке к вводу, а также на особенностях ввода и контроля данных с помощью рассматриваемого пакета.

Здесь и далее конкретным примером служат данные панельного исследования, выполненного под руководством и с участием авторов пособия в 1995-2003 гг. в трех российских селах: Латоново (Ростовская обл.), Венгеровка (Белгородская обл.) и Святцово (Тверская обл.).

Единица наблюдения в обследованиях - сельское домохозяйство. Объем выборки - 508 домохозяйств в 1995 г., 508 - в 1996 г., 500 - в

1997 г., 463 - в 1999 г. и 422 - в 2003 г.

Специфика панельного обследования предполагает наличие информации об изменениях во времени для каждого конкретного наблюдения. Другими словами, в панели должны быть одни и те же единицы наблюдения, равно как и одни и те же индикаторы. А вот с переменными в базе данных дела обстоят несколько по иному.

В случае записи панели в виде отдельных файлов (в нашем случае по каждому из трех лет самостоятельно) нумерация единиц наблюдения, а также имена переменных остаются неизменными. В то же время в случае записи панели в одном файле нумерация единиц наблюдения сохраняется как бы по умолчанию. Однако имена переменных в этом случае должны иметь отличительные (в нашем случае по годам наблюдения) метки. В принципе запись панели в одном файле является избыточной. Тем не менее, она оказалась очень полезной для целей контроля ввода данных и их анализа.

С учетом сказанного, конкретно в окончательный вариант базы данных, созданной по результатам проведения четвертой волны панели в 1999 г. вошло только 422 домохозяйства, в которых, каждый раз, начиная с 1995 г., в 1996-97 гг. и 1999 г. опрашивался один и тот же респондент. Другие 86 домохозяйств, в которых фиксировалось изменение единицы наблюдения, вошли в базовый файл, но оказались за пределами файла панели.

В качестве инструментария использовался опросный лист (приложение 2), состоящий из шести разделов и содержащий 70 вопросов. При этом ряд вопросов состоял из набора подвопросов, обладающих общим содержанием и объединенных в целях компактного представления. Ежегодно каждая анкета содержала около 550 переменных.

Небольшая разница в числе переменных первого и последующих этапов связана с необходимостью фиксации на втором и третьем этапах панели изменений, происходящих в массиве опрашиваемых (выбытии, смерти респондента, ликвидации домохозяйства и т.п.). Методика адаптации инструментария к таким изменениям хорошо видна в приложении 3.

В анкете использовались главным образом закрытые вопросы (например, вопросы 6-10, приложение 2). Открытые вопросы (например, вопрос 63, приложение 2) шифровались и кодировались на этапе ручного контроля полевой документации.

## 2.2. Подготовительный этап

*Основной смысл рассматриваемого этапа состоит в выполнении работ, обеспечивающих адаптацию анкеты к виду, позволяющему использовать средства автоматизации при ее обработке и выполнении расчетов.*

Еще на этапе разработки инструментария в бланке формализованного интервью во всех закрытых вопросах было выполнено кодирование ответов опрашиваемых числами (приложение 2). Эти числа и использовались интервьюерами при фиксации ответов респондентов.

В качестве примера рассмотрим один из разделов нашей анкеты «Социально - демографические характеристики семьи» (приложение 2). Этот раздел включает 10 вопросов: отношение к главе семьи, число исполнившихся лет, пол, образование всех членов семьи, национальность, состояние в браке, занятость, место работы, предприятие, должность взрослых ее членов.

Всем возможным ответам на перечисленные вопросы присваивались свои коды: **отношение к главе семьи:** муж-1, жена-2, другие взрослые члены семьи 3-5, дети до 18 лет -6-8; **пол:** мужской-1, женский-2; **национальность:** русский-1, украинец-2, другая-3; **состояние в браке**: женат (замужем)-1, холост-2, вдовец (вдова)-3, разведен-4; **занятость**: полный день-1, часть дня-2, безработный-3, пенсионер-4, нетрудоспособный (инвалид)-5, домохозяйка-6, по уходу за ребенком-7, учащийся-9; **место работы**: свое село-1, другое село-2, райцентр-3, город-4; **предприятие**: ТОО/АО-1, колхоз-2, общественное обслуживание-3, фермерское хозяйство-4, другой агробизнес-5, другой бизнес-6; **должность**: руководитель-1, специалист-2, служащий-3, рабочий/колхозник-4, фермер-5, другие-6 (приложение 2).

Открытые вопросы кодировались после полевых работ в ходе подготовки массива к вводу. Например, вопрос: **«Если домохозяйство новое, то почему?»** (приложения 2 и 3) принимал, исходя из характера полученных ответов, три значения (0 - прежнее домохозяйство, 1 - старая усадьба, но новые хозяева, 2 - новое домохозяйство, взятое в связи с недоступностью старого).

Следующий важный шаг - присвоение каждому индикатору анкеты восьмисимвольного смыслового имени с использованием букв латинского алфавита (приложение 4). Эти имена и становятся именами **пе-**

**ременных** с момента их введения в систему.

Например, переменная под именем «sexresp7» - производное от индикатора «пол респондента» в массиве 1997 г.; переменная «demtype5» - производная от индикатора «демографический тип семьи» в массиве 1995 г. Легко заметить, что последняя цифра смыслового имени указывает на год проведения интервью.

При обработке данных **панельного исследования** этот нюанс имеет существенное значение, приносящее массу удобств и большую экономию времени. Эти удобства возникают в связи с тем, что смысловые имена одних и тех же переменных отличаются только последней цифрой: demtype5, demtype6, demtype7.

Подобная организация смысловых имен ведет к тому, что в алфавитном перечне одни и те же переменные, индикаторы которых предметно - хронологически находятся в разных частях массива, стоят вместе (друг за другом). Опыт показывает, что в случаях достаточно большого перечня переменных указанный подход позволяет в дальнейшем экономить массу сил и времени.

В результате выполнения рассматриваемого шага к массиву анкет с первичной информацией добавляется еще один бланк со всеми индикаторами, расписанными по смысловым именам - переменным, со всеми возможными в данном исследовании кодами индикаторов, а также с указанием размера ячейки (ширины переменной). Формат такого бланка, названного нами «**Макет ввода данных в ЭВМ**», дан в приложении 4. В других работах сходный по назначению документ называется «кодировочной таблицей» (16, С. 26-27).

Конечно, можно дискутировать: нужно ли создавать такого рода макет в качестве отдельного документа. Можно обойтись и расписыванием всех имен переменных, их значений (кодов) и ширины этих значений в пустом бланке анкеты (фактически используя его в дальнейшем как макет), или, не тратя сил и времени, сразу ввести всю необходимую информацию непосредственно в систему (используя режим Variable View). В принятии одного из этих решений многое зависит от объема выполняемых работ (числа переменных) и от того, кто и как в дальнейшем собирается работать с массивом.

В том случае, если число переменных ограниченно (менее 100) и разработчик сам предполагает пользоваться формируемой им базой данных, то в принципе все возможно. Но, если переменных много и если анализом будут заниматься два и более пользователей, то без макета ввода данных в той или иной форме не обойтись.

В противном случае каждый пользователь, при каждом обращении к данным, вынужден будет сначала выполнить функции дешифровщика, а затем уже при благоприятном решении этой задачи, что в принципе представляется весьма и весьма проблематичным, он может заняться анализом данных. Все это позволяет нам сформулировать следующее довольно категоричное утверждение:

**Правило 5**

***Разработка макета ввода данных в соответствии с треблваниями SPSS является непременным предварительным условием ввода самих данных.***

Перед вводом данных выполняется визуальный контроль правильности и полноты заполнения анкеты и кодировки. Этот контроль позволяет выявить ошибки в заполнении анкет, которые возникают в результате неправильных записей, произведенных в анкете интервьюером, найти логические несоответствия (перепутан принятый в анкете порядок записи членов семьи, что в дальнейшем при панельном обследовании делало некорректным проводимый анализ), обнаружить ошибки в расчетах, выполняемых внутри анкеты.

Результатом указанных работ оказывается массив полевой документации, который теперь уже подготовлен к вводу данных. Конкретно, в нашем случае на последнем этапе работ (в 1999 г.) - это массив, состоящий из 463 заполненных и проверенных анкет. Их ввод и послужил основанием для последующего формирования панели, состоящей ежегодно из 422 случаев в 1995-1997 гг. Все следующие виды работ связаны непосредственно со средством автоматизации, а именно с пакетом SPSS.

## 2.3. Формирование таблицы для ввода данных

В этом параграфе описан общий порядок подготовки базовой таблицы SPSS к вводу данных. В 10-й и последующих версиях системы непосредственно работа по вводу данных выполняется в режиме **Data View** редактора данных. Сам ввод данных остался сходным с тем, что был в предшествующих версиях системы (1, С. 32). В то же время порядок их описания претерпел заметные изменения (1, С. 29). Основ

ной смысл подготовки базовой таблицы к вводу данных как раз и состоит в выполнении предварительных работ по созданию электронной версии макета ввода данных. Формирование электронного макета ввода данных выполняется в специальном режиме Variable View редактора данных.

Именно для этой цели, как отмечалось в предшествующем параграфе, на стадии подготовки инструментария выполняется работа по построению макета, т.е. присвоению уникального имени каждой переменной и заданию ее ширины. Выполнение последовательности действий по формированию таблицы – вводу имен переменных и их описания, предполагает знание следующих важных особенностей структуры окна редактора данных.

**Каждая строка** таблицы представляет собой место для записи случая или наблюдения. Любая анкета вводимого массива данных в полевых условиях называется «наблюдением», а в электронном формате наблюдение принято именовать «случаем». При вводе данных число наблюдений (случаев) равно числу анкет. Например, при проведении полевых работ в панельном исследовании 1995-2003 гг. число обследованных домохозяйств по годам составляло: в 1995 г. – 508, в 1996 г. – 508, в 1997 г. – 500, в 1999 г. – 463 и в 2003 г. – 382 наблюдения. Соответственно, в редакторе данных SPSS массив для каждого года состоит из 508, 500, 463 и 382 случаев. А панель, скажем, 1995-2003 гг. состоит из 382 случаев.

**Каждая колонка** представляет собой место для записи одной переменной. Любой вопрос анкеты имеет как минимум один индикатор и, следовательно, должен характеризоваться как минимум одной переменной. Например, вопрос 27 нашего обследования «**К какой группе обеспеченности вы можете отнести свою семью**?» (приложение 2) имеет один индикатор и, соответственно, одну переменную - group7 (приложение 4).

Уместно обратить внимание на тот факт, что вопрос анкеты нельзя отождествлять с индикатором. Один вопрос может иметь и несколько индикаторов. Например, вопрос 30 «**Какая социальная и другая помощь необходима вашей семье**?» Этот вопрос имеет 9 индикаторов: увеличение пенсий, увеличение пособий на детей, по ремонту дома и построек, при обработке приусадебного участка, льготное обеспечение лекарствами, в воспитании детей, по уходу за престарелыми и помощи им, в борьбе с алкоголизмом, душевное участие (приложение 2).

Соответственно, столько же колонок и переменных должно быть выделено в таблице окна редактора данных. В нашем примере для 1997 г. - это переменные pension7, helpch7, overhau7, cultiv7, helpmed7, educch7, helpold7, helpalc7, mental7 (приложение 4). Вопросы, не имеющие индикаторов с числовыми значениями, не могут быть введены в таблицу и обработаны ею.

Колонки и строки состоят из ячеек. Каждая ячейка представляют собой пересечение случая и переменной. Значение одной переменной записывается в одну ячейку. При этом в пределах границ сформированной таблицы всегда выполняется **правило отсутствия пустых ячеек** (правило 3), равно как и еще три следующих правила:

**Правило 6**

***Вопросы анкеты, не имеющие индикаторов с числовыми значениями, не могут быть ввкдены в таблицу.***

**Правило 7**

***В SPSS каждой строке присваивается порядковый номер, а именем каждой колонки является префикс var с последующей порядковой нумерацией.***

**Правило 8**

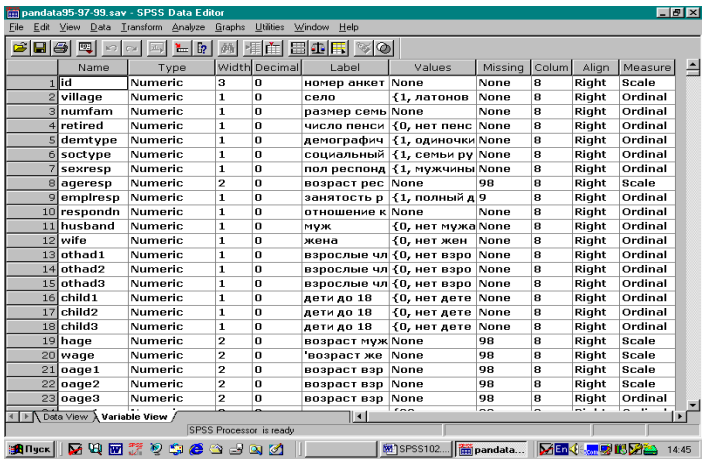
***Порядковый номер в таблице и вводимый разработчиками порядковый номер анкеты (id) не всегда совпадают. Этот момент имеет принципиальное значение для ввода, автоматизированного контроля, поиска ошибок кодировки и анализа данных.***

## 2.4. Имена переменных и их типы

Для создания переменной необходимо: дважды щелкнуть мышью на пустой колонке с надписью var или на ярлыке **Variable View** (просмотр переменных) в редакторе данных (рис.1). Главное меню также дает возможность перехода от просмотра данных к просмотру переменных (**View - Variables**). Эта цель может быть достигнута и путем комбинации клавиш: **Ctrl+T**. Во всех указанных случаях произойдет переход от режима просмотра данных к режиму просмотра переменных (рис. 5).

**Рис. 5.**

**Редактор данных в режиме Variable view-просмотр переменных**




В режиме просмотра переменные с их описанием записываются по строкам. Строки по умолчанию, как видно на рис. 5, имеют сплошную нумерацию. Это обстоятельство не только позволяет контролировать количество и последовательность записи переменных, но и создает ряд удобств, когда возникают задачи их переноса или удаления.

Одна строка содержит полное описание одной переменной. Полное описание переменной включает в себя десять характеристик: **Name** (имя переменной), **Type** (тип), **Width** (ширина), **Decimals** (число знаков после запятой), **Label** (метка), **Values** (метки значений), **Missing** (пропущенные значения), **Columns** (колонки), **Align** (выравнивание), **Measure** (шкала), названия которых, в указанном порядке (слева направо), хорошо видны в заголовках колонок режима просмотра переменных (рис.5).

Работая в этом режиме, можно последовательно, строка за строкой вводить переменные, создавая электронное описание (макет) формируемой базы данных. Более того, в режиме просмотра переменных

можно сэкономить время. Эта экономия достигается путем переноса параметров описания одной переменной на другие подобные переменные путем их копирования. К вопросу копирования переменных мы вернемся немного позже.

В 10-й и последующих версиях SPSS технология работ, связанных с записью и описанием переменных, претерпела заметные изменения. Теперь уже отсутствует необходимость, как это было в ранних версиях системы, для описания одной переменной пользоваться разными диалоговыми окнами: Define Variable, Define Labels (1, С. 25-31).

**Имена переменных.** Для допустимых имен переменных действуют следующие правила:

· длина имени не может превышать 8 знаков;

· имя должно начинаться с буквы. Остальные символы могут быть любыми буквами, цифрами, точкой или знаками @, #, _, $. Применяются только буквы латинского алфавита;

· имена переменных не могут оканчиваться точкой. Необходимо избегать имен переменных, оканчивающихся символом подчеркивания (чтобы избежать конфликта с переменными, которые автоматически создаются некоторыми процедурами);

· в именах не могут быть использованы пробелы и специальные символы (например, !, ?, “, и *);

· каждое имя переменной должно быть единственным. Например, в нашем пятиволновом панельном исследовании ответ на вопрос анкеты - возраст респондента обозначен ageresp5 (1995 г.), ageresp6 (1996 г.), ageresp7 (1997 г.), ageresp9 (1999 г.), ageresp3 (2003 г.).

Кроме системных требований, перечисленных выше, имя переменной обязательно должно удовлетворять и одному из основных пользовательских требований, а именно **оно должно быть узнаваемо**. В целом можно сказать, что при написании имен переменных имеются три основные возможности:

· имя переменной создается на основе корня слова соответствующего английского эквивалента (как это сделано нами в приведенном выше примере);

· имя переменной создается на основе транслитерации, т.е. путем написания русского слова или его основы латинскими буквами;

· имя переменной создается как комбинация словесного выражения и числа, отражающего порядковый номер переменной.

Каждая из указанных возможностей имеет свои преимущества и недостатки. Поэтому любой пользователь как индивидуальный, так и

коллективный неизбежно должны выработать свой стиль написания имен переменных.

**Тип переменных**. По умолчанию пакет SPSS предполагает, что все вводимые переменные являются числовыми **(Numeric)** с максимальной длиной 8 знаков, причем дробная часть состоит из двух знаков. Это основной тип переменных, т.к. статистический анализ ведется именно с такими переменными.

**Числовые переменные** имеют следующие особенности (характеристики):

· допустимые значения - цифры, знаки плюс или минус перед цифрами и десятичный показатель;

· ширина определяется следующим образом: **Width** - общая ширина числа (включая число разрядов перед десятичной точкой, десятичная точка (1 знак) и число разрядов после десятичной точки);

· **Decimal Places** - число отображаемых разрядов после десятичной точки.

Чтобы изменить ширину переменной, надо в поле Type соответствующей переменной щелкнуть мышью в ячейке на кнопке с тремя точками. Открывшееся диалоговое окно Variable Type (определение типа переменных) в средней части содержит информацию о ширине переменной Width и количестве десятичных знаков Decimal Places (рис.6).

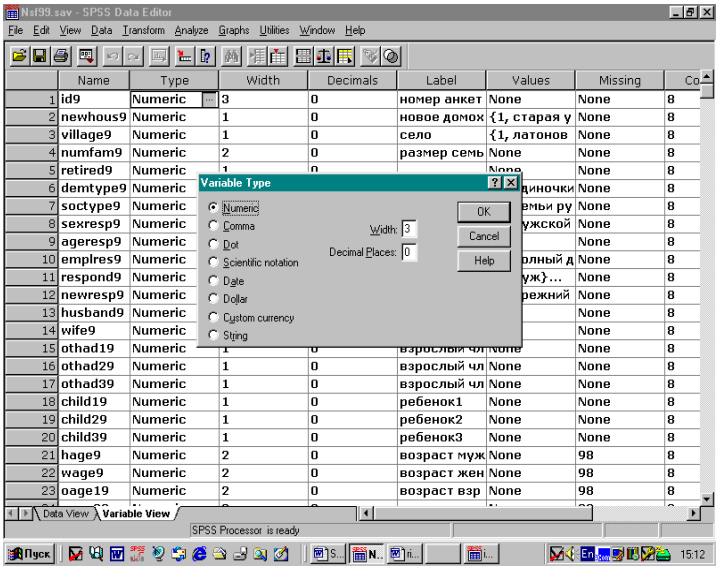



***Рис. 6.***

***Диалоговое окно Define variable type -определение типа переменной***

Можно здесь внести требуемые изменения. Изменения можно вносить также и в следующих за типом переменной столбцах (Width и Decimals) в режиме просмотра переменных (рис. 5). Для этого в поле Width (формат столбца) необходимо щелкнуть на кнопке лифта (вверх-вниз) и установить необходимое число позиций. Аналогично устанавливается количество десятичных разрядов в соседнем столбце под названием Decimals (десятичные разряды).

Например, для переменной id7 - номер анкеты (1997 г.) общая ширина числа равна 3, т.к. количество анкет в нашем обследовании не превышает трехзначного числа. Для переменной demtype7 - демографический тип семьи, общая ширина числа равна 1, т.к. эта переменная может принимать значения только от 1 до 7, т.е. однозначное число. При этом количество десятичных разрядов и в том и в другом случае равно 0 (приложение 4).

Кроме числовых в SPSS существуют и другие типы переменных: **Comma** (запятая), **Dot** (точка), **Scientific notation** (экспоненциальное представление), **Date** (дата), **Dollar** (доллар), **Custom currency** (специальная валюта), **String** (строка). Этот перечень типов переменных представлен в диалоговом окне **Define variable type** (рис. 6).

Выбор или изменение типа переменной можно осуществить, пометив в предложенном списке необходимый тип. Как уже отмечалось, по умолчанию заданным являются числовые (Numeric) переменные. Указанное обстоятельство существенно и может рассматриваться в качестве особого правила:

**Правило 9**

***По умолчанию пакет SPSS предполагает, что все вводимые переменные являются числовыми (Numeric).***

## 2.5. Метки

**Метка** (Label) используется для раскрытия (более подробного описания) имени переменной и закодированной числовой информации при получении результатов анализа. Метка переменной может содержать до 256 символов. В последних версиях системы метка без всяких проблем может быть написана и на русском языке, что авторы пособия и призывают пользователей делать. Подробное описание имени пере-

менной осуществляется в поле Label. Например, метка переменной sexresp7 - пол респондента; метка переменной group7 - группа обеспеченности семьи и т.п. Наличие хорошей метки на родном языке значительно облегчает как последующее чтение и понимание статистических расчетов, так и подготовку отчетов.

**Метки значений** (Values) **-** это название, позволяющее более подробно описать возможные значения переменной. Для присвоения метки значения необходимо выполнить следующие действия**:**

· Щелкнуть в поле **Values** на кнопке с тремя точками. Откроется диалоговое окно **Value Labels** - определение меток значений (рис.7).

· Ввести в поле **Value** значение переменной.

· Нажать клавишу «**Tab**» для перехода в поле **Value Label**. Здесь следует ввести метку значения.

· Щелкнуть на кнопке **Add** (добавить).

Например, для переменной sexresp7 метки ее значений: 1 - мужской, 2 - женский; для переменной group7 метки ее значений: 1 - очень бедная, 2 - бедная, 3 - среднего достатка, 4 - выше средней обеспеченности, 5 - с высокими доходами, 9 - затрудняюсь ответить (приложение 4).

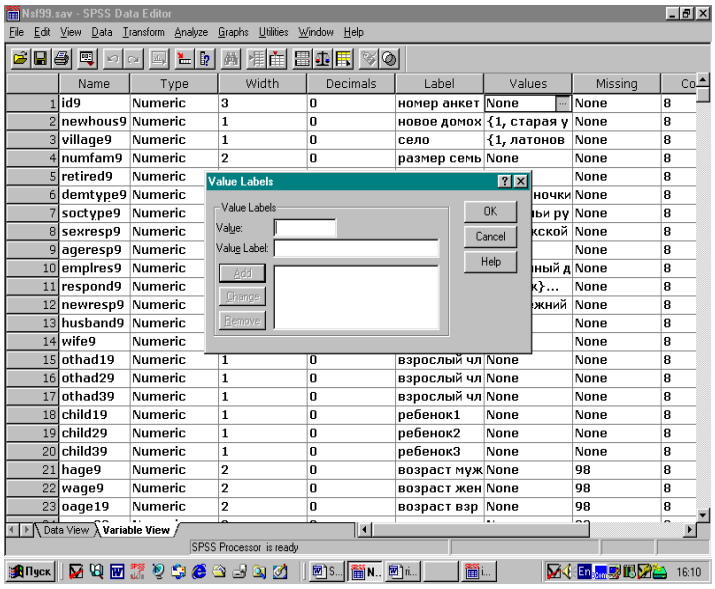



***Рис. 7.***

***Диалоговое окно Define Value Labels -определение меток значений***

Для изменения метки значения переменной необходимо выполнить последовательность действий:

· Выделить значение и метку из списка.
· Ввести новую метку.
· Щелкнуть на кнопке Change (изменить).

Для удаления метки значения переменной необходимо последовательно выполнить действия:

· Выделить метку и значение из списка.
· Щелкнуть на кнопке Remove (удалить).

**Дружеский совет**

***Описание меток и меток значений очень важно для наглядности и читаемости результатов последующих статистических расчетов. Здесь существенны оба момента: полнота описания и его язык.***

Как уже отмечалось ранее в § 2.4, имена переменных могут быть записаны только латинскими буквами. В то же время описание меток и меток значений вполне допустимо и на русском языке. От этого и от полноты самого описания, собственно, и зависит читаемость результатов статистических расчетов.

## 2.6. Пропущенные значения

В SPSS существует два типа пропущенных или отсутствующих значений:

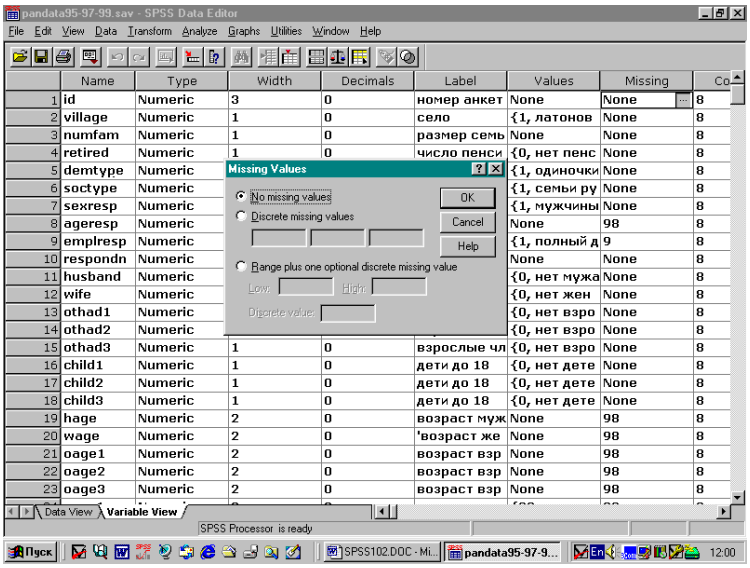
· **Системные (System missing)**. Это пустые значения, обозначаемые запятой. Такие запятые появляются в таблице, если в какую-нибудь ячейку не ввести значение.

· **Пользовательские (User missing)**. Это значения, которые пользователь определяет как пустые при выполнении статистических процедур.

При определении пользовательских пропущенных значений необходимо нажать кнопку с тремя точками в поле **Missing (**пропущенные значения). Откроется диалоговое окно **Missing Values** (рис. 8), в котором выбирается один из следующих альтернативных вариантов:

***Рис. 8.***

***Диалоговое окно Define Missing Values -определение пропущенных значений***




**No missing values** (нет пропущенных значений). Все значения - значимые. Это установка - по умолчанию.

**Discrete missing values** (дискретные пропущенные значения). Можно ввести до 3 таких значений.

**Range plus one optional discrete missing value** (диапазон плюс одно пропущенное значение). Все значения в диапазоне между верхним (High) и нижним (Low) пределами и еще одно дополнительное значение вне диапазона - пропущенные.

Образцом установки пользовательского пропущенного значения может служить следующий пример. Если на вопрос 11 нашей анкеты – «**Имеете ли вы или кто-нибудь из членов вашей семьи свое дело**?» (приложение 2), респондент отвечает - нет, то следующий 12 вопрос – «**Если имеете, то, как оказалось возможным его начать**?» - не имеет смысла. В этом случае в эту ячейку проставляется пропущенное значение, закодированное как 8.

Другой случай, если в составе семьи нет детей в возрасте до 18 лет (вопрос 1, приложение 2), т.е. значения переменных child17, child27, child37 (первый, второй и третий ребенок) равны 0, то значения переменных csex17, csex27, csex37 - пол детей (приложение 4) являются пропущенными. Их опять же можно принять равными 8, тогда пол детей автоматически уходит из расчетов.

Если теперь вернуться к переменной, фиксирующей число детей в семье, то в ней 0 - отсутствие детей является значимой величиной.

Среднее число детей в семье в этом случае будет относиться ко всем семьям массива, в том числе к одиночкам, брачным парам престарелых, т.е. к таким семьям, в которых заведомо не может быть детей до 18 лет. В то же время, если в данном случае объявить 0 как пропущенное значение, то базой расчета среднего числа детей в семье окажутся только те семьи, в которых есть несовершеннолетние дети.

Сходная ситуация возникает и в случае определения среднего размера пенсии или заработной платы. Размер средней заработной платы в массиве будет довольно заметно различаться в зависимости от того, относим ли мы ее ко всем опрошенным или только к работающим. Более того, статистически рассчитывать среднюю пенсию, как и среднюю заработную плату, можно только среди пенсионеров и работающих. Пропущенные значения как раз и позволяют это делать. В этом и состоит их содержательный смысл.

**Дружеский совет**

***Пропущенные значения очень важны при выполнении статистических расчетов. Именно они позволяют изменить их базу. Поэтому при выполнении расчетов этот параметр всегда полезно контролировать.***

## 2.7. Другие характеристики переменных

В режиме просмотра переменных имеются еще три позиции: **Columns**, **Align**, **Measure.** Они описывают различные свойства переменных (рис. 5).

**Columns** (столбцы). Это поле определяет ширину, которую будет иметь столбец в таблице данных при отображении значений. По умолчанию она равна 8. Ширину столбца также можно изменить непосредственно в окне редактора данных, поместив указатель мыши на разделитель между двумя заголовками столбцов с именами переменных. Затем путем перетаскивания можно расширить или сузить соответствующий столбец.

**Align –** сокращение от Alignment (выравнивание). Здесь можно задать вид выравнивания значений, т.е. определить, как они будут отображаться в таблице. По умолчанию системой предлагается вырав-

нивание по правому краю (Right). Можно задать выравнивание по левому краю (Left) или по центру (Center). Для внесения изменений выравнивания необходимо щелкнуть на кнопке (со стрелкой вниз) данного поля и выбрать один из трех предлагаемых вариантов.

**Measure** (измерение). С помощью этой характеристики имеется возможность задать шкалу переменной. Система допускает, что шкала может быть **Nominal** - номинальной (шкала наименований), **Scale** - порядковой или **Ordinal** - метрической. По умолчанию системой принимается метрическая шкала измерения. К вопросу о шкалах измерения мы еще вернемся ниже во второй части пособия.

## 2.8. Общие настройки

С помощью общих настроек можно установить некоторые полезные параметры, которые будут действовать во всех командах системы и при всех последующих обращениях к ней.

Для задания интересующих общих настроек необходимо выполнить в главном меню следующую последовательность команд:

**Edit**

**Options.**

На экране появится диалоговое окно **Options -**параметры (рис. 9).

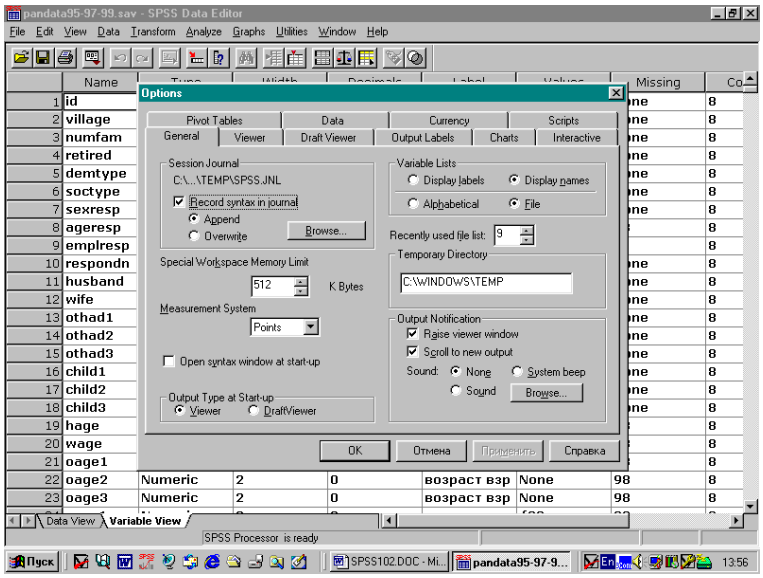



***Рис. 9.***

***Диалоговое окно Options -параметры***

В этом диалоговом окне находится 10 регистрационных карт: General (общие), Viewer (окно просмотра), Draft Viewer (окно текстового режима), Output Labels (обозначение выводимых значений), Charts (диаграммы), Interactive (интерактивный режим), Pivot Tables (сводные таблицы), Data (данные), Currency (денежная единица), Scripts (сценарии).

Каждая из этих карт выполняет определенные функции. Например, установка порядка вывода для списка переменных (в порядке ввода переменных в таблицу или в алфавитном порядке) производится в регистрационной карте General. Следует иметь ввиду, что после установки SPSS всегда по умолчанию действует опция **File** - в порядке ввода переменных в таблицу. Сортировка списка переменных в алфавитном порядке произойдет после установки опции **Alphabetical.** При этом машина выдает сообщение, что данная установка будет действовать только при последующем обращении к рабочему файлу. Это означает, что при желании работать со списком переменных, записанных в алфавитном порядке, в текущую сессию необходимо перезагрузить систему.

Установить тип и размер шрифта заголовков (Title Font) и текста (Text Output Font) в окне просмотра можно, используя карту Viewer. Здесь же задаются размеры страницы в блоке Text Output Page Size.

Используя карту Draft Viewer (окно текстового режима), можно установить внешний вид таблиц и текста.

В карте Output Labels (обозначение выводимых значений) можно выбрать, будут ли для обозначения переменных указываться их имена или соответствующие метки. При этом метки установлены по умолчанию. Можно выбрать и установить режим, при котором и имена переменных, и их метки будут выводиться одновременно. Для обозначения категорий переменной можно выбрать значение переменной или метку значения, или оба варианта одновременно.

Карта Charts (диаграммы) служит для установки в графиках и диаграммах шрифта, цвета, разных штриховок, типов линий, компоновки рамки и отображения координатной сетки.

Interactive (интерактивный режим). В этой карте имеются настройки для выбора параметров интерактивных графиков.

Карта Pivot Tables (сводные таблицы) служит для выбора внешнего вида сводных таблиц.

Изменения формата данных происходят в карте Data. Установка по умолчанию: ширина (Width) равна 8, количество знаков после запятой

(Decimal places) - 2 знака.

В карте Currency (денежная единица) можно указать денежный формат.

Карта Scripts (сценарии) позволяет активировать автоматические сценарии.

Для сравнения можно вспомнить, как устанавливались отдельные настройки в предыдущих версиях пакета (1, С. 31-32). Так, для задания требуемых общих настроек в версии SPSS 6.0 необходимо выбрать в меню последовательность команд: Edit – Preferences. Далее, например, чтобы определить порядок вывода переменных в списках, необходимо в диалоговом окне общих настроек в поле Display Order for Variable Lists выбрать Alphabetical (в алфавитном порядке) или File (в порядке ввода переменных в таблицу).

## 2.9. Ввод и корректировка данных в таблице

Ввод данных - это заполнение ячеек таблицы числами, содержащимися в первичной информации. Следует иметь в виду, что в SPSS имеется специальное приложение для ввода данных - SPSS Data Entry (глава 5, § 5.4). Сам редактор данных позволяет осуществлять их ввод в любом порядке - по наблюдениям или по переменным. Эта технология ввода данных описана ниже.

Для введения данных в ячейку необходимо выполнить следующие действия:

· щелкнуть мышью на нужной ячейке (первая ячейка находится в верхнем левом углу). Вокруг ячейки появится рамка, означающая, что ячейка активна;

· ввести требуемое значение. Это значение появляется в строке ввода и в ячейке. Чтобы перейти к следующей ячейке, надо либо нажать клавишу **Enter** (и тогда активируется ячейка под заполненной), либо с помощью стрелок на клавиатуре перемещаться вверх, вниз, вправо, влево. Как мы уже отмечали, ввод данных можно производить и по строкам, и по столбцам.

Операция ввода данных повторяется вплоть до заполнения всех ячеек, находящихся на пересечении вводимых случаев (наблюдений) и переменных таблицы. При этом полезно помнить, что:

· Если значение вводится в колонку за пределами границ таблицы, то автоматически создается новая переменная. Параметры этой пере-

менной устанавливаются по умолчанию. При этом все случаи (кроме текущего) получают системные пропущенные значения.

· Если значение вводится в строку за пределами границ таблицы, автоматически создается новое наблюдение. При этом все переменные (кроме той, что введена последней) получают системные пропущенные значения.

· При вводе числовых переменных целые значения, превышающие заданную ширину, вводятся, но редактор данных показывает их на экране символами (число перед E + число, обозначающее количество знаков, превышающее заданную ширину). Если далее производятся какие-то расчеты, то в окне просмотра отражаются значения переменных с превышением заданной ширины.

Необходимость выработки навыков корректировки данных в таблице является неотъемлемым компонентом их ввода. Ввести ошибочно другую цифру, перепутать столбцы и строки – все это неизбежные человеческие издержки ввода данных.

Корректировка данных в таблице предполагает: их изменение в ячейке, удаление всей строки, содержащей ошибочно введенный случай, удаление столбца, содержащего ненужную переменную, вставку дополнительных строк и столбцов для ввода новых случаев и переменных. Все эти операции требуется выполнять постоянно. Поэтому порядок их выполнения должен быть хорошо отработан и усвоен.

**Для того**, **чтобы изменить данные в ячейке**, необходимо выполнить следующую последовательность действий:

· встать на нужную ячейку с помощью мыши или стрелок на клавиатуре;

· изменить требуемое значение;

· перейти в другую ячейку с помощью кнопки **Enter** или стрелок на клавиатуре.

В результате выполнения указанной последовательности действий новое значение закрепится в нужной ячейке.

**Для удаления случая** (строки) необходимо выполнить следующую последовательность действий:

· Щелкнуть на номере наблюдения (слева от строки) или выбрать любую ячейку в строке и нажать **Shift + пробел**. Выделится вся строка.

При выделении нескольких наблюдений используется метод щелчка и протягивание мышью или набор клавиш **Shift + стрелки.**

· Далее в главном меню следует выполнить последовательность команд: Edit – Clear.

Выбранная область (из одного или нескольких случаев) будет удалена. Все наблюдения ниже нее будут подтянуты вверх. Сходным образом случай может быть удален и с помощью клавиши **Del**.

**Для удаления переменной** (столбца) необходимо выполнить следующую последовательность действий:

· Щелкнуть на имени переменной (в верхней части столбца) или выбрать любую ячейку в столбце и нажать **Ctrl + пробел**. Выделится весь столбец. Для выделения нескольких переменных используется метод щелчка и протягивания мышью или набор клавиш **Shift + стрелки**.

· Далее в главном меню следует выполнить последовательность команд: Edit – Clear.

Выбранные переменные (переменная) будут удалены. Все переменные, которые находятся справа от удаленных, будут сдвинуты влево. Сходным образом переменная (ые) может быть удалена и с помощью клавиши **Del**.

Последние версии пакета позволяют производить удаление переменных из файла, находясь в режиме Variable View. Для этого достаточно щелкнуть мышью слева от имени переменной. Это действие ведет к ее выделению. Далее можно использовать клавишу Del.

**Для вставки случая** (строки) необходимо выполнить следующую последовательность действий:

· Находясь в режиме просмотра данных (Data View), выделить ячейку, над которой надо вставить случай (строку).

· Нажать **правую** клавишу мыши и выбрать в появившемся меню **Insert Case**.

· Нажать левую клавишу мыши.

Вставка новой переменной может быть выполнена и посредством использования кнопки **Insert Case** на панели инструментов (глава 1, § 1.4).

**Для вставки новой переменной между имеющимися переменными** необходимо выполнить следующую последовательность действий:

· Находясь в режиме просмотра данных (Data View), выделить столбец, перед которым надо вставить переменную.

· Нажать правую клавишу мыши и выбрать в появившемся меню Insert Variables.

· Нажать левую клавишу мыши.

Вставка новой переменной может быть выполнена и посредством использования кнопки Insert Variables на панели инструментов (глава 1, § 1.4).

В результате выполнения указанной выше последовательности действий в таблице появится столбец с новой переменной. То же самое можно сделать в режиме просмотра переменных (Variables View). Изюминка здесь состоит в том, что в этом случае надо выделять ту строку с переменной, перед которой будет вставлена новая строка. При этом автоматически происходит изменение (сдвиг) в нумерации переменных.

**Правило 10**

***Ввод данных следует делать только после полного описания переменных в таблице окна редактора.***

### Задание для самостоятельной работы

1. Что такое макет ввода данных?
2. Как правильно: сначала вводить данные, а затем описывать переменные или сначала описать переменные, а затем вводить данные?
3. Назовите основные характеристики переменной.
4. Какая комбинация клавиш позволяет менять режим Data View на режим Variable View и наоборот?
5. Опишите последовательность действий при удалении строки.
6. В чем различие метки (Lable) и метки значения (Values)?
7. Что такое пропущенное значение?
8. Какие пропущенные значения есть в SPSS?
9. Какие действия выполняются при изменении данных в ячейке?
10. Какие возможности имеются в написании имен переменных?
11. Назовите основные действия при вводе данных.
12. Опишите последовательность действий при удалении переменной.
13. Какие закладки имеются в диалоговом окне Options?
14. Опишите последовательность действий при вставке нового столбца.

15. Введите в новый файл пять переменных, опишите их и набейте условные данные для пяти случаев.

16. Опишите последовательность действий при вставке новой переменной между уже имеющимися переменными.

17. Как можно выделить переменную и наблюдение?

18. Как можно выделить группу переменных и наблюдений?

19. В чем различие представления списка переменных в опции File и опции Alphabetical в диалоговом окне Options?

20. Опишите устройство диалогового окна Value Labels (определение меток значений) и порядок работы в нем.

21. Опишите устройство диалогового окна Missing Values (определение пропущенных значений) и порядок работы в нем.

22. Какой порядок подготовки базовой таблицы SPSS к вводу данных?

23. В каком режиме редактора данных выполняется их ввод?

24. В каком режиме редактора данных выполняется описание переменных?

25. В чем смысл правила отсутствия пустых ячеек?

26. Как устанавливается порядковый номер строки в базовой таблице SPSS?

27. В чем отличие порядкового номера строки в базовой таблице SPSS и идентификационного номера случая (id)?

28. Какие основные требования предъявляются к написанию имен переменных в SPSS?

29. Что такое ширина (Width) переменной?

30. Что такое число разрядов после десятичной точки (Decimal Places)?

31. Какие типы переменных фиксируются в режиме их просмотра (Variable View) в редакторе данных SPSS?

# Глава 3. Работа с файлами данных

## 3.1. Создание нового файла

При создании нового файла его в первую очередь необходимо сохранить. Обычно многие пользователи сначала, если файл создается для ввода данных, сразу же и начинают их вводить, а затем уже сохраняют файл с целью обеспечения возможности его дальнейшего (текущего и последующего) использования. В действительности такой порядок действий всегда связан с риском потери уже введенных данных, описания переменных или текста. Более правильно сначала сохранить новый файл под подходящим для него именем, а затем уже работать в нем.

Таким образом, рекомендуемая последовательность действий при создании нового файла для ввода данных имеет следующий вид:

**File**

**New**

**Data.**

Результат выполнения этих действий – появление в окне редактора новой таблицы с готовыми для ввода (пустыми) ячейками. В строке заголовка при этом, как уже отмечалось ранее (глава 1, § 1.4), появляется надпись «**Untitled – SPSS Data Editor**». Лучше всего, именно сейчас, следующим шагом, еще не приступая к работе с новым файлом, сохранить его.

Сохранение нового файла предполагает выполнение следующей последовательности действий в главном меню:

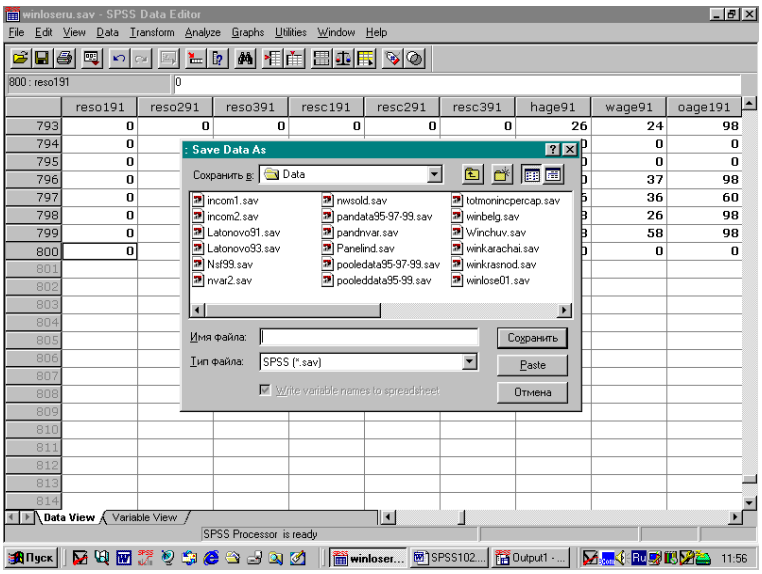
**File**

**Save As.**

Две другие возможности, позволяющие достигнуть того же результата: использование панели инструментов, а именно кнопки -  сохранение файла из текущего окна - **Save File** и комбинации клавиш **Ctrl + S** (глава 1, § 1.4).

*Рис. 10.*

***Диалоговое окно Save Data As - сохранение нового файла***




Если файл новый (без имени), то, независимо от того, какая из трех указанных возможностей будет реализована, откроется диалоговое окно **Save Data As** (сохранить файл данных как). Это окно и показано на рис. 10.

В принципе оно сделано и работает также, как и все окна подобного плана во всех приложениях к Windows. Поэтому, если пользователь имеет опыт работы, скажем, в MS Word, у него вряд ли могут возникнуть трудности, связанные с созданием и сохранением нового файла.

На практике это означает, что пользователь должен сначала выбрать директорию (папку), в которой будет сохранен файл (строка **Save in** -сохранить в). По умолчанию системой там будет указываться имя папки, в которой ранее уже создавался новый файл. В данном случае, как это видно на рис. 10, имя папки «Data». Следующим шагом необходимо вписать в окно строки **File name** - имя файла, то уникальное имя, которое наиболее полно отразит содержание сохраняемого файла. При этом система уже сама готова дать соответствующее расширение. Об этом говорит надпись у строки **Save as type** – сохранить как тип. В нашем примере выставлено расширение для сохранения всех файлов с данными (*.sav).

Специфика SPSS как раз и проявляется в названиях расширений, используемых им файлов. В SPSS имена файлов данных имеют расширение **.sav.** Файл будет сохранен под введенным именем. Повто

ряем, хорошо, если само имя файла информативно, т.е. позволяет визуально идентифицировать содержащиеся в нем данные. В нашем случае файлы назывались соответственно по годам: nsf95.sav, nsf96.sav, nsf97.sav, nsf99.sav. Все пользователи знали: NSF - это аббревиатура фонда, финансировавшего наш проект, а именно National Science Foundation. Объединенный файл за три года получил имя pandata_95-96-97.sav, а за пять лет – pandata_95-97-99.sav, что указывало на хранение в нем данных панельного исследования, причем цифрами фиксировался год проведения каждой волны.

После введения имени файла, как и всегда в таких случаях, для выполнения команды остается нажать кнопку **Save** - сохранить. Результат выполнения команды - появление уникального имени нового файла в строке заголовка редактора данных.

**Правило 11**

***Сначала надо сохранить новый файл под подходящим для него именем, а затем уже работать в нем.***

**Правило 12**

***Имена файлов с данными имеют в SPSS расширение .sav. Файлам окон просмотра присваивается расширени .spo. Файлы командного языка синтаксис имеют расширение .sps.***

## 3.2. Открытие рабочего файла

Всякий раз, приступая к сеансу работы в SPSS, необходимо открыть рабочий файл. Для этого следует выполнить следующие команды главного меню:

**File**

**Open**

**Data.**

Две другие возможности, позволяющие достигнуть того же результата: использование панели инструментов, а именно кнопки  которая предназначена для открытия файла (Open File) и комбинации

клавиш Ctrl + O (глава 1, § 1.4). В результате выполнения любой из перечисленных выше команд откроется диалоговое окно Open File (открыть файл).

В этом окне выбирается имя требуемого файла и нажимается кнопка **Open** - открыть окно или делается двойной щелчок левой клавишей мыши на имени открываемого файла, что, как известно, дает тот же результат. Спустя некоторое время, в зависимости от конфигурации (быстродействия) компьютера и размера открываемого файла, требуемая информация появится в редакторе данных.

Появление имени вводимого файла в строке заголовка окна редактора (глава 1, § 1.4) свидетельствует о загрузке файла. Так, при загрузке нашего файла с данными четырех волн панельного исследования в строке заголовка появляется запись «c:\spsswin\nsf\pandata_95-97-99.sav». Затем в таблице отображаются имена переменных, порядковые номера наблюдений (опросных листов) и, в самую последнюю очередь, появляются заполненные данными ячейки таблицы. За пределами таблицы обозначены затененные номера строк (слева) и затененные заголовки колонок (сверху). В любой момент пользователь может добавить в них новые переменные и число наблюдений.

Дополнительными индикаторами выполнения команды открытия файла служат сигналы работы процессора в статусной строке, а также отсчет выставляемых машиной случаев (анкет), фиксируемый в одном из окошек статусной строки. При этом сама таблица какое-то время остается неизменной.

## 3.3. Сохранение файла

Работая с таблицей, можно вносить определенные изменения и дополнения в нее, будь то внесение новых наблюдений или создание новых переменных, а также корректировка данных. Чтобы сохранить текущую таблицу в файл данных со всеми изменениями для последующей работы, необходимо выбрать в меню:

**File**

**Save.**

Две другие возможности, позволяющие достигнуть того же результата: использование панели инструментов, а именно кнопки  -сохранение файла из текущего окна - **Save File** и комбинации клавиш

Ctrl+S (глава 1, § 1.4). Если файл старый (уже имеет имя), то, независимо от того, какая из трех указанных возможностей будет реализована, измененный файл данных автоматически сохраняется и записывается поверх предыдущей версии.

**Правило 13**

***Перед окончанием сеанса работы, если рабочий файл не сохранен, то система всегда напомнит о необходимости «сохранить изменения в файле». При этом откроется окно с вопросами и возможными вариантами ответов «да» или «нет».***

## 3.4. Обединение и разделение файлов

Иногда требуется объединить два или несколько файлов данных. **Объединение файлов - это добавление данных из одного файла к данным другого файла, загруженного в настоящий момент в редактор данных**. В зависимости от характера данных, содержащихся в различных файлах, можно выделить два основных типа объединения.

**Объединение файлов, содержащих одинаковые переменные, но различные наблюдения**. Например, при анкетировании нескольких сел, удобнее вначале (при вводе и корректировке данных) разместить их отдельно по каждому селу в своем файле (по принципу: одно село - один файл – один оператор ввода). При этом размер файла будет значительно меньше, и работа с ним будет идти быстрее. Это особенно удобно для экономии времени и контроля работы операторов в случае ввода данных несколькими операторами одновременно. Все это имеет и другие известные преимущества, но на этапе общего контроля и анализа данных возникает задача слияния нескольких файлов в один файл.

**Объединение файлов, содержавших одинаковые наблюдения и различные переменные**. Например, при анкетировании за несколько лет можно при вводе расположить данные за каждый год в отдельном файле, а при анализе слить их в один файл.

В нашей работе использовалось объединение файлов, содержащих одинаковые переменные, но различные наблюдения. Объединя

лись 3 файла, которые содержали данные по трем селам (соответственно, Латоново, Венгеровка, Святцово) за каждый год обследования.

Для объединения таких файлов необходимо загрузить первый файл в редактор данных. С учетом, что в ходе полевых работ наблюдения уже имеют сплошную нумерацию, то в качестве исходного файла удобнее брать файл, содержащий случаи, начиная с первого. В обследовании 1999 г. это был файл с данными, собранными в селе Латоново (latonovo99.sav), который в автоматической нумерации, задаваемой по умолчанию системой, включал наблюдения с 1-го по 144-й номер. Сохраняем этот файл в директории Data под новым именем – nsf99.sav. Если этого не сделать сразу же, то несколько позже он может оказаться сохраненным (по ошибке, забывчивости или невнимательности пользователя) с именем исходного рабочего файла.

Далее в 10-й и последующих версиях SPSS[1] следует выбрать в главном меню последовательность команд:

**Data**

**Merge Files**

**Add Cases.**

Выполнение этой последовательности команд ведет к открытию окна **Add Cases: Read File** – Добавить случаи: читать файл (рис. 11).

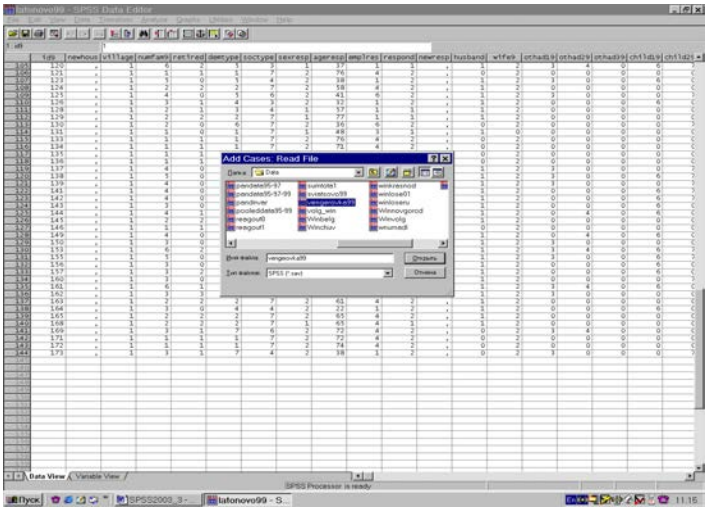



***Рис. 11.***

***Первое дополнительное окно - Add Cases: Read File команды объединения файлов***

[1] В предыдущих версиях пакета процедура объединения файлов осуществлялась с использованием пункта главного меню Edit.

В этом первом дополнительном диалоговом окне необходимо выбрать файл, который требуется добавить и выделить его. При этом автоматически имя файла прописывается в строке File name - имя файла. Этот момент и зафиксирован на рис. 11. Рисунок сделан так, чтобы хорошо было видно число случаев в рабочем файле, а так же та нижняя (с пустыми ячейками) часть таблицы, в которой по результатам выполнения команды окажутся данные присоединяемого файла.

В нашем случае это будет файл с данными по селу Венгеровка (vengerovka99.sav), который включал наблюдения с 145-го по 289-й номер. Последующее нажатие кнопки **Open** (открыть) ведет к открытию в редакторе данных второго дополнительного окна команды объединения файлов (рис. 12).

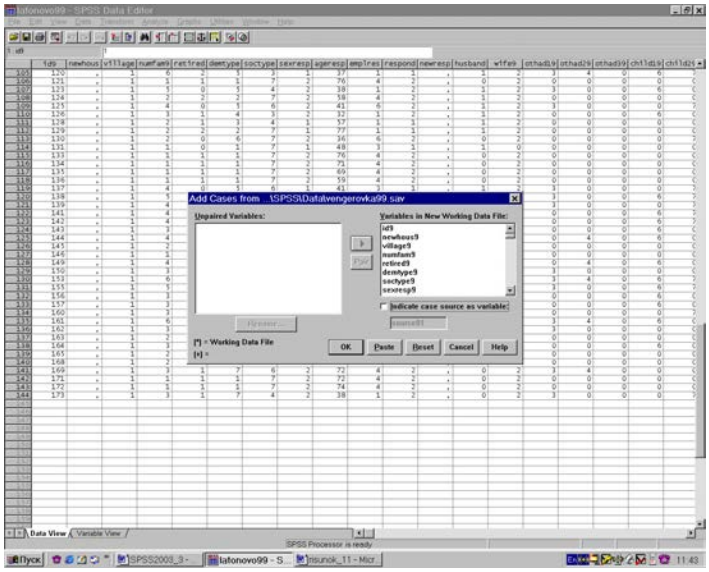



***Рис. 12.***

***Второе дополнительное окно – Add Cases from команды объединения файлов***

При условии идентичности переменных в объединяемых файлах их список появляется в правом поле – **Variables in New Working Data File**, а левое поле - **Unpaired Variables** – остается пустым. Далее следует нажать кнопку ОК. В редакторе окажутся данные двух исходных файлов, объединенные в один файл. Далее следует контрольное сохранение.

В нашем примере этот файл теперь уже будет включать с 1-го по 289-й случаи. Если объединяемые файлы имеют различные переменные, то система в правом поле (Variables in New Working Data File)

выставит список общих переменных для обоих объединяемых файлов, а в левом поле - Unpaired Variables (со знаками + или *) окажутся переменные, которых нет в одном из объединяемых файлов. При этом знак плюс указывает на тот факт, что непарные переменные принадлежат присоединяемому файлу, а звездочка указывает на то, что они принадлежат рабочему файлу.

Если далее идти по пути выполнения команды объединения файлов, нажав кнопку ОК, то в объединенный файл войдут только идентичные (парные для обоих файлов) переменные. Если же пользователь желает включить все или отдельные непарные переменные в объединенный файл, каждую из них необходимо выделить и с помощью кнопки со стрелкой перенести из левого в правое поле. Выделение переменной в левом поле активирует не только кнопку переноса, но и кнопку **Rename -** переименование, позволяющую редактировать имена переменных, находящихся в этом поле.

Здесь возможна и обратная операция, а именно: убрать из будущего объединенного файла одну или несколько переменных, которые имеются в обоих исходных файлах. Для этого, как можно догадаться, необходимо выделить такую переменную в правом поле и с помощью все той же, но уже поменявшей свое направление стрелки, перенести ее в левое поле. Система укажет, что убираемая переменная принадлежит и тому, и другому файлу, но выполнит последующую команду на их объединение. Новый файл окажется уже без удаленной переменной(ых).

Тем же путем присоединяется следующий файл. В нашем примере это будет файл с данными по селу Святцово (sviatsovo99.sav), который включал наблюдения с 291-го по 422-й номер. В результате в одном файле оказываются объединенными данные всего массива обследования (номера случаев с 1-го по 422-й) за 1999 г., а новый файл, как уже отмечалось ранее, получает и новое хорошо узнаваемое разработчиками имя - nsf99.sav.

**В случае объединения файлов, содержащих одинаковые наблюдения и различные переменные**, необходимо выбрать в главном меню последовательность команд:

**Data**

**Merge Files**

**Add Variables.**

Выполнение этой последовательности команд ведет к открытию диалогового окна Add Variables: Read File (Добавить переменные: читать файл). Все дальнейшие действия в этом случае аналогичны действиям, выполнявшимся при открытии первого дополнительного окна, и весьма сходные с действиями, выполнявшимися при открытии второго дополнительного окна, которые были описаны ранее в этом параграфе при рассмотрении порядка добавления случаев.

В нашей практике было объединение файлов, которые содержали данные по годам обследования [nsf95.sav (1995 г.), nsf97.sav (1997 г.) и nsf99.sav (1999 г.)], в файл с данными трех волн панели за 1995 -1999 гг. - pandata_95-97-99.sav. В любом случае, как при объединении файлов с различными наблюдениями, так и особенно при объединении файлов с различными переменными, всегда требуется выполнение определенных условий.

**Правило 14**

***Система без проблем объединит файлы:***
***- если при добавлении случаев в объединяемых файлах все переменные идентичны;***
***- если при добавлении переменных объединяемые файлы содержат одинаковое число случаев.***

При нарушении этих условий возникает необходимость либо корректировки данных в процессе объединения файлов, либо последующей «ручной» чистки в объединенном файле переменных (при добавлении случаев) и случаев (при добавлении переменных), содержащих системные пропуски. Это весьма и весьма существенный момент, который может быть усвоен и преодолен только в повседневном опыте, связанном с эксплуатацией системы обработки данных в SPSS.

Как случаи, так и переменные могут переноситься из одного файла в другой файл «ручным» путем. Этот эффект достигается посредством **выделения** (клик мышью) и **копирования** (Ctrl + C) предполагаемых к переносу данных в одном файле, а затем последующего открытия другого файла, установки в нем курсора в место предполагаемого переноса и **вставки** (Ctrl + P) в него переносимых данных. Если при этом будет выделяться вся строка (с номером случая) и вся колонка (с именем переменной), то в случае их идентичности система сделает

все так же, как и при выполнении описанной выше последовательности команд по объединению файлов.

Опыт показывает, что использование и той, и другой возможности объединения файлов и переноса данных из одного файла в другой оправдано и зависит от определенных обстоятельств. Если число дополняемых случаев (переменных) ограниченно, то быстрое использование комбинации клавиш копирования и вставки вполне оправдано.

Выполнение комбинации команд по объединению файлов более эффективно и безопасно, когда объединяются большие файлы, содержащие значительное число случаев и переменных. Если полученный результат всегда будет одним и тем же для пользователя, то для системы перенос данных из одного файла в другой и объединение файлов – совершенно разные операции.

**Разделение файла**. В SPSS отсутствует последовательность команд главного меню, которая имеет своим прямым результатом разделение файла на два или более файла. Вместе с тем эта задача может быть легко решена как путем ее задания в языке системы Syntax, так и «ручным» путем.

Логически задача по разделению файлов является обратной относительно задачи их объединения. Поэтому, имея файл nsf99.sav, содержащий данные обследования трех упоминавшихся ранее сел, вполне разумно в случае необходимости разделить его на три исходные файла.

Самый короткий путь решения такой задачи предполагает:

· **сохранение** исходного рабочего файла nsf99.sav под новым именем, скажем, latonovo99.sav,

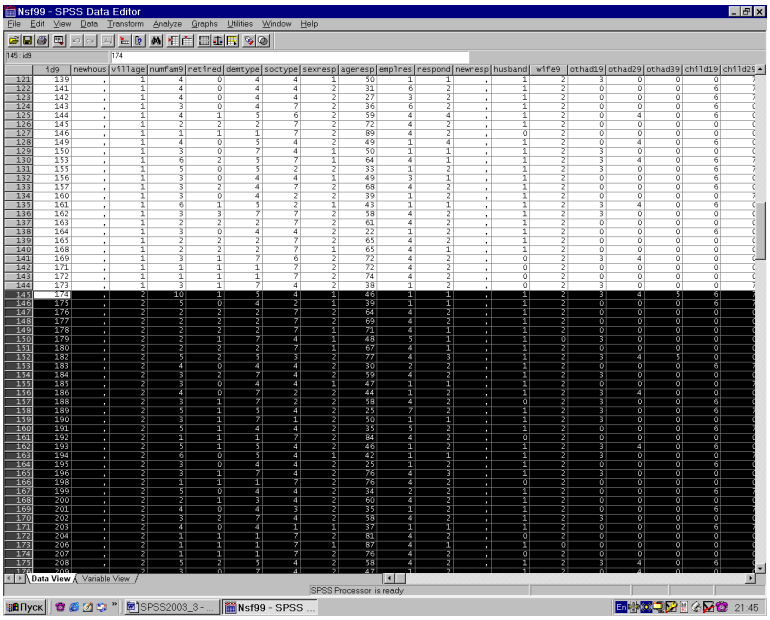
· **выделение** (путем протягивания с левой стороны у клавиш сплошной нумерации) мышью всех случаев, относящихся к двум другим селам, как это показано на рис. 13,

· **удаление** выделенных случаев с помощью клавиши Del или последовательности команд главного меню Edit - Cut, или комбинации клавиш Ctrl + X.

Как видно на рис. 13 черная линия выделения проходит по случаю, имеющему 145-й номер в сплошной нумерации системы. Контролем здесь служит смена кодов сел в третьей слева переменной village9 с единицы (Латоново) на двойку (Венгеровка). Сходным образом в случае необходимости могут быть созданы отдельные файлы с данными и по двум другим селам нашего массива (Венгеровке и Святцово).

*Рис. 13.*

*Разделение файла*




Описанная процедура **фактического** разделения одного файла на два или несколько самостоятельных файлов по своей сути является, если так можно выразиться, насильственной. Она существенно отличается от предлагаемого системой изящного и очень эффективного при контроле и анализе данных **условного** (виртуального) разделения файла. К рассмотрению этой возможности мы сейчас и переходим.

## 3.5. Разделение случаев на группы - Split File

Использование процедуры **Split File**, имя которой можно перевести как «разделить файл», позволяет выполнять анализ данных рабочего файла раздельно по группам. Под группой при этом понимается определенное количество случаев с одинаковыми значениями признаков. Для того, чтобы можно было производить обработку по группам, файл необходимо отсортировать по группирующим переменным. В принципе в качестве группирующей может быть взята любая переменная с первичной информацией (глава 2, § 2.4).

Таким образом, оказывается, что для анализа массива данных в разрезе села вовсе не обязательно фактическое разделение файла, которое может быть полезным для достижения каких-то других целей.

На этом этапе работ вполне уместно воспользоваться возможностями, которые предлагает рассматриваемая процедура.

Для ее использования необходимо выбрать в главном меню последовательность команд:

**Data**

**Split File.**

При выполнении указанной последовательности команд откроется диалоговое окно **Split File** (рис. 14).

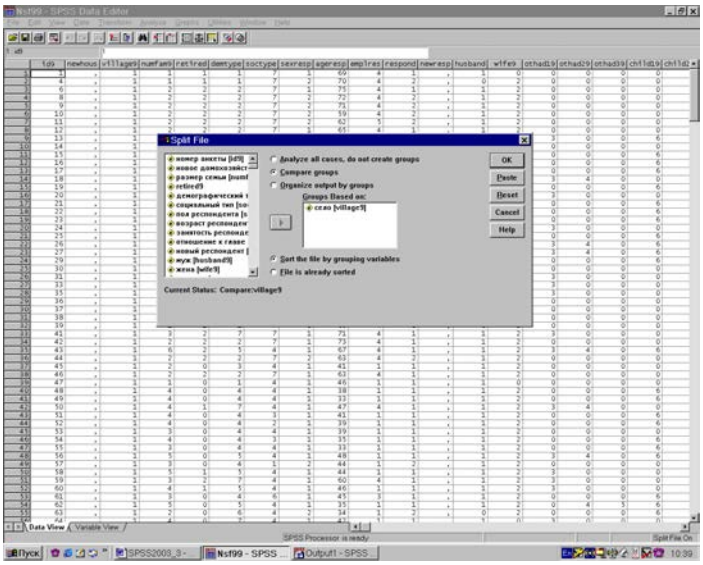



***Рис. 14.***

***Диалоговое окно Split File***

Системой здесь допускаются три основные возможности (опции):

· исходно, по умолчанию, разделение на группы отсутствует. Об этом свидетельствует метка в первой (верхней) строке меню: **Analyze all cases, do not create groups** (анализировать все случаи, не создавая групп);

· следующая опция - **Compare groups** (сравнивать группы). Эта опция и установлена в окне на рис. 14;

· последняя опция - **Organize output by groups** (разделить вывод на группы).

При выборе одной из двух последних опций группировочный признак задается путем переноса соответствующей переменной из левого большого подокна со списком переменных в правое подокно - Groups Based on (группы, образованные на основе). Как видно на рис.14, в качестве группировочного признака нами задана все та же переменная village9 (село в выборке 1999 г.).

Завершение выполнения команды посредством нажатия кнопки ОК не ведет к внешнему изменению файла. Связано это с тем, что, как уже отмечалось ранее, массив организован и исходно введен по селам.

Вместе с тем, если бы в качестве группировочного признака был задан пол респондента, то даже зрительно массив в таблице претерпел бы заметные изменения. Случайное чередование респондентов по полу в массиве первичной информации приобрело бы упорядоченный характер: сначала пошли бы все мужчины, а затем женщины. Это обусловлено не отсутствием галантности у мужчин, а тем, что они имеют код равный 1, тогда как у женщин код равен 2.

В результате выполнения рассматриваемой команды в текущем сеансе, вплоть до ее отмены, все расчеты будут выполняться в разрезе группировочного признака. Выполнение этой команды может быть отменено посредством возврата в диалоговое окно Split File, установки в нем первой опции Analyze all cases, do not create groups и последующей команды ОК.

## 3.6. Отбор случаев - Select Cases

Для некоторых операций над переменными в системе предусмотрена возможность отбора случаев, с которыми будет выполняться данная операция, т.е. формирование **направленной** или **случайной выборки**. В отличие от предшествующей рассматриваемая процедуры ведет к зрительно фиксируемым преобразованиям рабочего файла.

Например, необходимо подсчитать в массиве число женщин и мужчин среди респондентов старше 50 лет или число семей в массиве с доходом ниже прожиточного минимума. Такого рода задачи вполне осмыслены и актуальны. Но они почти всегда предполагают задание предварительного условия. Скажем, отбора респондентов в возрасте старше 50 лет с помощью переменной (возраст респондента).

В SPSS предусмотрена возможность задания такого рода условия. Она реализуется с помощью команды **Select Cases** (отбор случаев)**.** Если стоит задача отбора какой-либо части случаев (подмножества) из всей имеющейся в данном файле совокупности наблюдений, то в главном меню необходимо выбрать следующую последовательность команд:

**Data**

**Select Cases.**

При этом откроется диалоговое окно Select Cases, которое дает возможность выбора ряда вариантов отбора. Это окно состоит из подокна со списком переменных, которое располагается слева, и большой рамки Select, содержащей различные опции отбора случаев, находящейся в центральной части диалогового окна.

По умолчанию здесь всегда установлена опция - **All Cases** (все наблюдения), которая находится в рамке Select первой сверху. Для задания условия отбора подмножества наблюдений следует выбрать опцию **If condition is satisfied** (если удовлетворяет условию) и нажать кнопку **If**. После чего откроется дополнительное окно для ввода условия, которое аналогично всем окнам, содержащим кнопку If. Этот момент и зафиксирован на рис. 15.

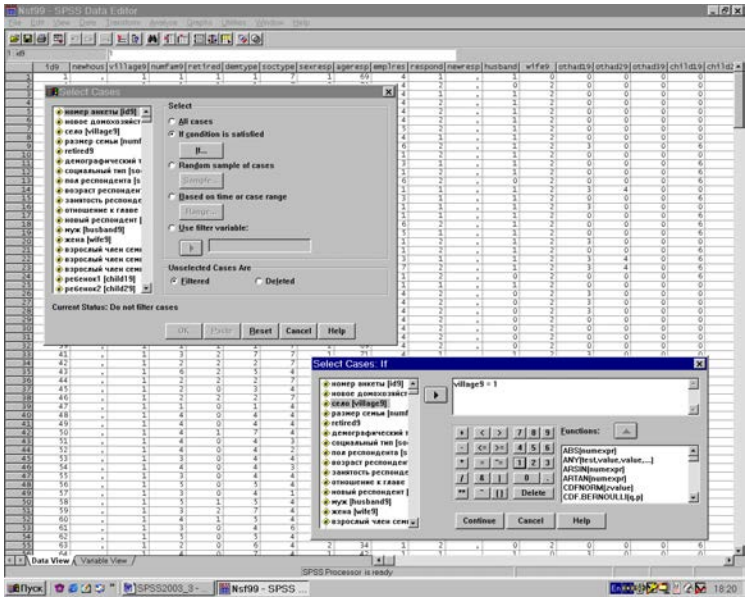



***Рис. 15.***

***Диалоговые окна процедуры Select Cases***

Для обеспечения возможности задания условий отбора, разработчики системы предусмотрели в этом подокне набор сервисных инструментов. Среди них в первую очередь следует отметить:

· типовой сервис, обеспечивающий доступ к списку переменных рабочего файла и их перенос в подокно, в котором задаются условия отбора;

· небольшую клавиатуру, позволяющую задание чисел и основных математических (сложения, деления, умножения, равенства и др.) и логических действий (конъюнкции, дизъюнкции);

· поле со списком предлагаемых системой функций (которые будут рассмотрены в отдельном параграфе следующей главы) и кнопкой их переноса в поле задания условий отбора;

· уже упоминавшуюся ранее кнопку **Continue**, позволяющую после окончания задания условий отбора, возвращаться в главное диалоговое окно рассматриваемой команды.

При отборе случайной выборки наблюдений необходимо использовать опцию **Random sample of cases** (случайная выборка наблюдений) и нажать кнопку **Sample** (выборка). Далее откроется диалоговое окно **Random Sample** (случайная выборка).

Для указания объема выборки следует остановиться на одном из двух предлагаемых системой вариантов:

· **Approximately** (приблизительно) задает процент от общего числа наблюдений. Система порождает случайную выборку из приблизительно определенного процента наблюдений.

· **Exactly** (точно). Число наблюдений определяется пользователем.

После задания условия необходимо указать способ отображения в редакторе данных неотобранных наблюдений. Для этого в окне **Unselected Cases Are** (невыбранные наблюдения) нужно выбрать одну из двух возможностей:

· Первая - **Deleted** (уничтожить). При этом невыбранные наблюдения физически удаляются из таблицы. Следует помнить, что наблюдения могут быть восстановлены, если закрыть файл данных, не сохраняя какие-либо изменения, а затем открыть его. Они, однако, будут потеряны навсегда в случае сохранения изменений в файле данных.

· Вторая - **Filtered** (отфильтровать). Это менее опасная опция. Невыбранные наблюдения отсутствуют в анализе, но остаются в таблице. В этом случае создается новая переменная с именем **filter_$** со значением 1 для выбранных наблюдений и со значением 0 - для невыбранных. Как всякая новая переменная filter**_$** появляется в последнем правом столбце таблицы.

Для удобства просмотра невыбранные наблюдения автоматически отмечаются диагональной линией, проведенной через номер строки. Если во время работы отменить фильтрацию (в команде Select Cases выбрать опцию All Cases), то можно использовать и ранее невыбранные наблюдения. Фильтр может быть уничтожен и стандартным путем, а именно в результате выделения переменной filter_$ и последующего использования клавиши Del.

Если произвести отбор наблюдений с другим условием, то переменная filter_$ пересчитывается с учетом заданного условия. Отбор наблюдений действует только на один сеанс работы, т.е. если сохранить файл с фильтрацией и завершить сеанс работы, то при последующем открытии файла переменная filter_$ остается, но невыбранные наблюдения восстанавливаются.

Следующий пример иллюстрирует возможность использования команды Select Cases. В файле данных находится информация по трем селам, необходимо провести анализ по одному из них. В файле есть переменная village9, которая однозначно идентифицирует село (Латоново - 1, Венгеровка - 2, Святцово – 3. См.: приложения 2 и 4).

Предположим, пользователя интересует информация только по с. Латоново. Поставив в команде Select Cases условие отбора: village9 = 1, весь последующий статистический анализ будет проводиться только для этого села. Установка этого условия в дополнительном диалоговом окне Select Cases: If и зафиксирована на рис.15.

Далее необходимо нажать в этом окне кнопку **Continue**, окно закроется и произойдет возврат в окно Select Cases, в котором для выполнения команды необходимо нажать кнопку ОК. Как уже отмечалось ранее, в результате выполнения команды будет создана переменная filter_$, которая активизирует (сделает доступными для анализа) только случаи, заданные по условию: Select Cases: If village9=1 (если село=1).

Другой пример иллюстрирует использование команды Select Cases в сочетании с командой Case Summaries. Потребность их совместного использования возникает в случае необходимости проверки значения какой-нибудь переменной по конкретному набору анкет, скажем, пол респондента (sexresp9) по трем анкетам: 200, 340 и 422.

С этой целью в команде Select Cases надо задать условие отбора: id9=200 | id9=340 | id9=422. В результате выполнения этой команды в таблице в порядковой нумерации наблюдений останутся неперечеркнутыми только три заданных случая.

Далее, используя команду Case Summaries, выбираем в списке переменных sexresp9. Выполнение этой команды ведет к появлению в окне вывода информации о том, что в данном случае во всех трех наблюдениях респондентами являются женщины.

## 3.7. Сортировка случаев - Sort Cases

Система позволяет сортировать данные рабочего файла в соответствии со значениями одной или нескольких переменных. Решение задач такого рода связано с использованием процедуры **Sort Cases** (сортировка случаев).

Для ее использования необходимо выбрать в главном меню последовательность команд:

**Data**

**Sort Cases.**

При выполнении указанной последовательности команд откроется диалоговое окно Sort Cases (рис. 16).

Далее следует перенос в этом окне имени переменной из левого поля со списком переменных в правое поле **Sort by** (сортировать по) с выбором одной из опций сортировки по нарастанию (**Ascending**) или по убыванию (**Descending**) значения признака. Последующее выполнение команды путем нажатия кнопки ОК ведет к сортировке массива данных по нарастанию (убыванию) значений заданного признака.

В данном примере на рис. 16, установлена переменная «совокупный месячный доход семьи» (sumtota9) в массиве 1999 г. При выборе опции, предполагающей сортировку по нарастанию значений признака, в результате сортировки домохозяйства распределятся в массиве данных в порядке роста месячного совокупного дохода от минимального к максимальному. В результате массив приобретет очень выигрышную конфигурацию для изучения бедности (27, С. 296-314).

Сортировка может выполняться по нескольким признакам. Более того, для каждого из них может быть установлена своя опция по возрастанию или убыванию значения сортирующего признака. При этом исходным основанием сортировки всегда будет первая перенесенная переменная.

Сортировки очень эффективное и мощное средство анализа особенно для количественных признаков, таких как: возраст, месячный доход, размер семьи, площадь обрабатываемого земельного участка, объем выпускаемой продукции и др. Несколько ниже, в том числе и в следующем параграфе, мы еще вернемся к примерам использования рассматриваемой процедуры.

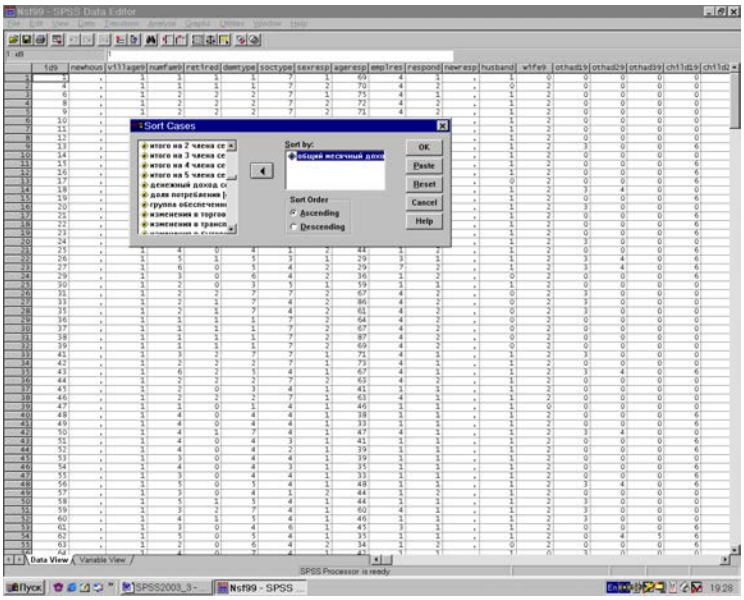



***Рис. 16.***

***Диалоговое окно процедуры Sort Cases***

## 3.8. Создание индивидуального файла

При работе с нашими данными постоянно возникала необходимость в информации о семье в целом (число людей в массиве, число мужчин и женщин в массиве, средний возраст, среднее число членов в семье, число детей в семье и др.). Между тем, как видно из приложения 2 (раздел 1), исходно собиралась информация по каждому члену семьи отдельно. Это обстоятельство требовало преобразования собранной в поле информации.

Такого рода задача, как и многие другие, может решаться несколькими способами. Один из них состоит в том, чтобы создать новый файл, записав в нем последовательно (в таблице сверху вниз) всех членов семьи с основными социально-демографическими характеристиками. Базой для преобразования при этом должен служить существующий файл с данными обследования, в котором информация по каждому индикатору для каждого члена семьи записана в качестве самостоятельной переменной (приложение 4, раздел 1).

В отличие от исходного файла, фиксирующего **структуру семьи** и содержащего 500 случаев (1997 г.), новый файл, который мы называли **индивидуальным**, насчитывает 1400 случаев (число всех людей во

всех семьях, попавших в выборку того же года). Порядок создания этого файла и описан ниже.

В основном рабочем файле выделяем последовательно, начиная с переменной «husband7» (муж), каждого члена семьи, встав на название переменной и щелкнув мышью один раз. Выделится нужная колонка. Далее идем в главное меню, выбираем последовательно:

**Edit**

**Copy.**

Отмеченный фрагмент забирается в карман (Clipboard) и вставляется командой Paste в предварительно созданный индивидуальный файл. Причем для вставки первой переменной (муж) необходимо установить курсор в новом файле на название первой колонки, пока еще не имеющей своего имени. После щелчка мышью появляются все значения соответствующей переменной (муж) с числом наблюдений равным 500. Переменной присваивается имя (famstr7), которое фиксирует структуру семей в массиве. Далее необходимо перенести из основного файла всех остальных членов семьи.

Начиная с 10-й версии пакета SPSS, нельзя, скопировав столбец с переменной, перенести ее вслед за первой (как это делалось в предыдущих версиях). Копировать необходимо не столбец, а наблюдения (встав курсором на первый случай, протянуть выделение до последнего случая). Все следующие члены семьи переносятся из основного файла и записываются в эту же колонку: с 501 случая - жена, с 1001-первый взрослый член семьи и т.д.

Таким путем создаются переменные: «возраст», «пол», «образование» и др. Для удобства работы в индивидуальный файл можно предварительно блоком перенести все переменные, из которых и будет осуществляться последующее построение требуемых новых переменных.

В процессе работы с данными панельного исследования нам приходилось создавать специальный файл, все переменные в котором выстраивались по годам, т.е. каждому наблюдению соответствовало 4 (четырехлетнее панельное обследование) значения переменной. Здесь, кроме описанного выше способа, еще использовалась процедура Sort Cases по Id. Представляется целесообразным описать построение и такого файла.

В панельном исследовании (четырехлетнее) опрашивались одни и те же респонденты, которым каждый год (т.е. четыре раза) задавались одни и те же вопросы. Соответственно имена переменных в

итоговом файле отличаются только указанием года (demtype5, demtype6, demtype7, demtype9). В результате итоговый файл panel-data_95-97-99 представляет собой таблицу с 422 наблюдениями и 2168 (542*4) переменными по 4 годам.

Для анализа динамики, т.е. изменения значений различных переменных по годам, требовалось иметь файл, в котором каждая переменная расписывалась бы последовательно по каждой волне (каждому году) наблюдения. В 1997 г. по итогам трех волн панели построение подобного файла выполнялось с помощью Syntax (глава 17, § 17.3). В 1999 г. создание этого файла было выполнено «вручную» следующим образом:

· В рабочем файле paneldata_95-97-99 переменные выделялись путем щелчка мышью на их названии и копировались (Edit - Copy).

· В предварительно созданном новом файле (pooleddata_95-97-99) выделялось столько столбцов, сколько переносилось переменных из файла paneldata_95-97-99. Далее использовалась последовательность команд Edit - Paste.

· В paneldata_95-97-99 выделялись и копировались такие же переменные за следующий год.

· Курсор устанавливался в нижней части столбца, куда и вставлялись данные. Эта операция была повторена четыре раза.

· В новом файле каждой переменной присваивалось новое имя. Например, переменная для демографического типа семьи, составляющей которой по годам стали переменные: demtype5, demtype6, demtype7, demtype9 - получила название demtype.

По результатам этого этапа в файле pooleddata_95-97-99 все переменные получились выстроенными по годам вниз по очереди: 422 анкеты за 1995 г., 422 - за 1996 г. и т.д. - всего 1688 случаев. Между тем, как уже отмечалось выше, для нас было важно организовать данные в файле таким образом, чтобы в нем по каждому случаю числовые значения каждой переменной шли последовательно год за годом, как бы повторяясь четыре раза, от 1-го случая до 422-го случая.

Для достижения этой цели на втором этапе преобразований использовалась процедура **Sort Cases**. При этом в качестве сортирующего признака брался идентификационный номер анкеты (переменная id). В результате выполнения сортировки и был получен файл, в котором все случаи представлены по годам: от 1-го случая с 4-мя значениями каждой переменной до последнего 422-го случая с

теми же 4-мя значениями каждой переменной (всего, все те же 1688 случаев).

Смеем надеяться, внимательный читатель понимает, что для возврата к организации данных в файле, которая была до момента сортировки, надо опять отсортировать файл, но уже по переменной «год проведения интервью» (year). Этим замечанием мы стремимся привлечь внимание всех пользователей к огромной аналитической значимости включения сначала в первичную информацию, а затем и вводимые в систему SPSS данные, казалось бы таких технических характеристик массива, как номер анкеты и год проведения исследования.

Кстати сказать, в первичной документации и, соответственно, при формировании файлов с данными, мы ставим номер анкеты на первое место в качестве кодировочного признака, а год проведения полевых работ - на последнее. Такая организация массива данных в SPSS, кроме всего прочего, позволяет визуально контролировать совпадение автоматической нумерации задаваемой системой и нумерации случаев в массиве данных. Последний момент приобретает особенное значение в панельных исследованиях, в которых во времени неизбежно выпадение определенной части случаев.

В целом необходимо отметить, что описанные в этой главе различные аспекты работы с файлами данных, от их создания и сохранения до отбора случаев и сортировки, связаны с двумя пунктами главного меню, а именно **File** и **Data**. Каждый из этих пунктов включает в себя и другие команды для работы с файлами данных. Одни из них описаны в следующих главах пособия, тогда как другие, так и остались за пределами нашего внимания. Это относится к отдельным командам пункта главного меню Data, таким как Transpose, Orthogonal Design и некоторым другим. Пока еще в нашей практике не встречалось задач, предполагающих использование перечисленных процедур.

**Дружеский совет**

***Работая с файлами данных, всегда полезно помнить о командах главного пункта меню Data, позволяющих решать многие задачи контроля и анализа данных.***

## Задание для самостоятельной работы

1. Назовите последовательность команд при создании нового файла.
2. Какое расширение в SPSS имеет файл с данными?
3. Какая последовательность команд при сохранении нового файла?
4. В чем специфика объединения файлов в SPSS?
5. Для решения каких задач служит пункт главного меню Data?
6. Приведите несколько примеров команд пункта главного меню Data.
7. Опишите процедуру Merge Files.
8. Как изменяется рабочий файл при выполнении команды Sort Cases?
9. Опишите процедуру Select Cases.
10. В чем специфика разделения файла в SPSS?
11. Как изменяется рабочий файл при выполнении команды Select Cases?
12. Опишите процедуру Split File.
13. Каким образом можно вернуть файл в исходное состояние после выполнения команд Sort Cases и Select Cases?
14. Как файл можно дополнить новыми случаями?
15. Опишите процедуру Sort Cases.
16. Как файл можно дополнить новыми переменными?
17. Для чего в массиве данных нужна переменная с номером случая?
18. В чем особенность копирования переменных в последних версиях системы?
19. В каких командах используются и для чего нужны опции по нарастанию (Ascending) - по убыванию (Descending)?

# Глава 4. Преобразование данных

Наряду с возможностями выполнения различных операций с файлами данных, которые были рассмотрены в предшествующей главе, и статистических расчетов, которые будут рассмотрены далее во второй части пособия, в SPSS существует еще одна замечательная особенность. эта особенность связана с возможностью преобразования исходных (первичных) данных и их дополнения путем агрегирования, вычисления весов случаев, ранговых преобразований, создания новых переменных и др.

Среди перечисленных операций по преобразованию данных особенно важными, открывающими действительно новые возможности анализа в социологических и социально-экономических исследованиях, на наш взгляд, являются различные преобразования, связанные с созданием новых переменных. Поэтому именно им в настоящей главе и уделено основное внимание.

Команды, позволяющие преобразовать данные, находятся в пункте главного меню **Transform**, имя которого говорит само за себя и вряд ли нуждается в переводе.

## 4.1. Создание новой переменной с использованием процедуры Compute

Потребность в преобразовании данных существует практически всегда как на этапе контроля, так и при их анализе. Здесь важно обратить внимание на два момента: во-первых, четкое понимание возможностей создания новых переменных неизбежно должно оказывать обратное влияние на структуру полевой документации. Учет указанного обстоятельства в первичной документации позволяет как экономить время и силы при сборе данных, так и избегать ошибок расчета, допускаемых интервьюерами в вопросах, требующих вычислений (например, вопрос 24, приложение 2).

Во-вторых, задачи создания новых переменных могут возникать и решаться не только на этапе анализа, но и при вводе и контроле данных. Более полно этому вопросу уделено внимание в главе 5, § 5.2.

Новая переменная может создаваться путем вычисления по уже имеющимся переменным. при этом она автоматически размещается в последнем справа столбце таблицы. следующим шагом ее можно перенести на любое удобное для пользователя место в массиве данных.

Решение задач такого рода оказывается возможным благодаря команде **Compute** (вычислить). Например, требуется вычислить денежный доход на одного члена семьи. В файле данных имеются переменные numfam7 - размер семьи и total7 - денежный доход семьи (приложение 4). Логически новая переменная с уникальным именем midtot7 (средний доход в 1997 г.) должна быть рассчитана по следующей формуле: midtot7=total7/numfam7. Это выражение означает, что средний денежный доход на одного члена семьи равен денежному доходу семьи, разделенному на число ее членов.

Другой пример, для вычисления новой переменной child7 (число детей в семье в 1997 г.) необходимо провести следующие преобразования: child7=child17/6+child27/7+child37/8. Логика таких преобразований основана на том, что в переменных child17 (первый ребенок), child27 (второй ребенок), child37 (третий ребенок) при наличии ребенка проставляется код значения его места в структуре семьи (6, 7 или 8 соответственно), а если ребенка нет, то код данной позиции равен 0. Поэтому в нашем случае мы и должны сначала разделить сам на себя код каждого ребенка, чтобы получить единицу (если ребенок есть) и ноль (если его нет), а затем уже выполнить сложение. Если бы мы сразу кодировали наличие ребенка – 1, а его отсутствие – 0, то выражение, позволяющее получить число детей в семье, имело бы вид: child7=child17+child27+child37.

Возникает вопрос, **где и как можно записать приведенные выше выражения с тем, чтобы вычислить значения требуемых признаков для каждого случая?** Ответ на этот вопрос как раз и можно получить с помощью главного диалогового окна команды **Compute**.

Для использования этой команды необходимо в главном меню выполнить следующую последовательность действий:

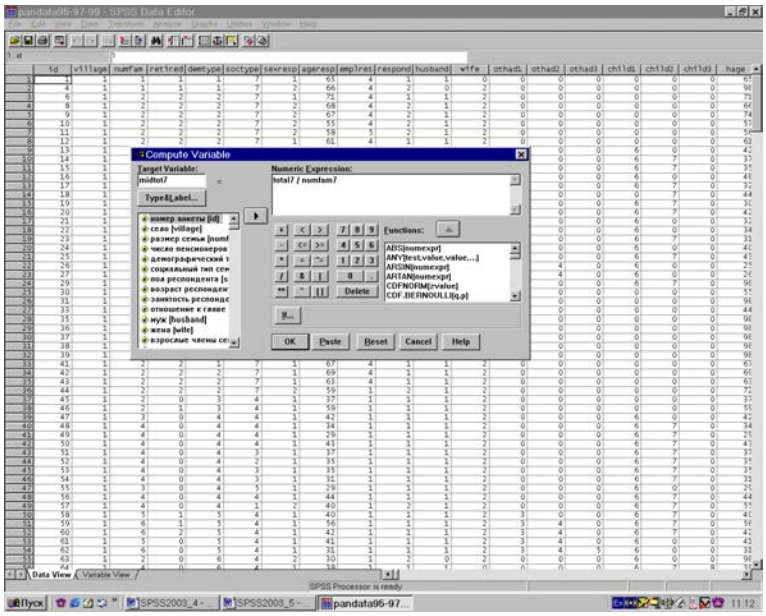
**Transform**

**Compute.**

Выполнение указанной последовательности команд ведет к открытию главного диалогового окна Compute Variable (рис. 17).

*Рис. 17.*

*Главное диалоговое окно Compute Variable*




В левом верхнем углу этого окна имеется небольшое поле **Target Variable** (целевая переменная). Сюда надо вписать имя новой переменной. В данном случае видно записанное нами имя из первого примера midtot7 (средний доход). Клавиша **Type & Label**, находящаяся под этим поле, позволяет сделать полное описание новой переменной, в которое мы, исходя из собственного опыта, рекомендуем включать и логическую форму вычисления переменной. Не сделав этого сразу, потом (в первую очередь при написании текста отчета или статьи) её придется вспоминать мучительно долго.

В центральной части окна находится поле **Numeric Expression** (числовое выражение), в которое нами в данном случае внесено выражение: total7/numfam7. Между полями Target Variable и Numeric Expression в поле диалогового окна стоит знак равенства. Это означает, что система всегда готова произвести вычисление по заданной формуле. Важно только, чтобы сама формула была задана корректно.

Для обеспечения благоприятных условий задания числовых выражений система предлагает большой набор сервисных инструментов. Среди них в первую очередь следует отметить:

· типовой сервис, обеспечивающий доступ к списку переменных рабочего файла и их перенос в поле числового выражения;

· небольшую клавиатуру, обеспечивающую задание чисел и основных математических действий (сложения, деления, умножения, равенства и др.);

· поле со списком предлагаемых системой функций и кнопкой их переноса в поле числового выражения, которое, как мы и обещали ранее (в главе 3, § 3.6), будет рассмотрено ниже;

· уже упоминавшуюся ранее кнопку **If**, позволяющую задать особые условия, при которых записанное числовое выражение может быть вычислено.

Числовое выражение может включать существующие имена переменных, константы, арифметические операторы и функции. Для вставки элементов можно использовать панель вычислений, список переменных и список функций или набирать их вручную.

Панель вычислений содержит числа, арифметические операторы, операторы сравнения и логические операторы. Ею можно пользоваться как калькулятором, используя мышь, или просто как эталоном для уточнения специальных символов.

Для того, чтобы применить преобразования к подмножеству наблюдений, необходимо использовать условные выражения. Если результат такого выражения является истинным, то к этому наблюдению применяется преобразование.

Для задания условного выражения необходимо: щелкнуть мышью на кнопке **If** (если). Откроется диалоговое окно **If Cases**. Здесь можно выбрать одну из двух возможностей:

· **Include all cases** (включить все наблюдения). Эта опция устанавливается по умолчанию. С ее помощью преобразования вычисляются для всех наблюдений. Например, в переменной размер семьи - numfam7 необходимо выполнить анализ для семей, состоящих из пяти человек. Для решения этой задачи в поле Target Variable вписывается имя новой переменной -nnumfam7, а в поле Numeric Expression вписываем выражение: numfam7=5. Далее следует выполнение команды – OK. Результат – новая переменная со значением 1 для семей, состоящих из пяти человек, и 0 для всех остальных случаев. При этом кнопка **If** не используется.

· **Include if case satisfies condition** (включить отбор наблюдений, удовлетворяющих условию).

В этом случае преобразования вычисляются для наблюдений, удовлетворяющих условию, введенному в текстовое поле. Условие может включать имена существующих переменных, константы, арифметические операторы, числовые и другие функции, операторы сравнения.

Например, для анализа особенностей поведения семей, состоящих из тех же 5-и человек, в одном из трех сел, скажем, Латоново, которое в рабочем файле имеет код равный 1. Для решения этой задачи в поле Target Variable вписывается имя новой переменной - numfam71, а в поле Numeric Expression вписываем выражение: numfam7=5. Затем идет нажатие кнопки **If**. В открывшемся окне устанавливается опция **Include if case satisfies condition,** которая активизирует находящееся ниже текстовое поле. В этом поле делается следующая запись: village7=1. Наконец, следует выполнение команд Continue и OK. Результат – новая переменная со значением 1 для семей, состоящих из пяти человек, 0 для всех остальных семей в селе Латоново и пропущенными системными значениями для остального массива.

Процедура Compute, будучи универсальной, является наиболее мощной процедурой по созданию новых переменных. По этой же причине она является и наиболее сложной для использования. Другие процедуры, часто используемые для создания новых переменных, - **Recode** и **Count**.

## 4.2. Создание новой переменной с использованием процедуры Recode

В ходе анализа данных довольно часто оказывается необходимым изменить значения переменных путем их перекодировки. Это особенно полезно делать с целью сжатия информации, скажем, при переходе от непрерывных значений к интервалам.

Хорошим примером здесь может служить переменная возраст респондента (ageresp7), фиксируемая в первичных данных в виде «числа исполнившихся лет» (приложение 2). А так как разброс возраста очень велик (могут быть любые значения от 18 до 99), то удобнее сжать информацию по возрастным интервалам, которые выбираются в зависимости от типа решаемых задач. Благодаря этому абсолютное значение возраста респондента можно преобразовать, например, в следующую интервальную шкалу:

- возраст от 18 до 30 закодировать как - 1;
- от 31 до 45 - 2;
- от 46 до 60 - 3;
- от 61 и старше - 4.

С этой целью уместно создать новую переменную agegr7. Во всех наблюдениях, где ageresp7 принимает значение от 18 до 30, она будет равна 1; от 31 до 45 - 2 и т.д. Для того, чтобы создать новую переменную на основе перекодирования значений, необходимо выбрать в главном меню:

**Transform**

**Recode**

**Into Different Variables**.

Откроется диалоговое окно **Recode Into Different Variables** (перекодировать в другие переменные). Процедура Recode имеет две опции, причем на первом месте в дополнительном меню стоит опция **Into Same Variable** (перекодировка внутри исходной переменной), а на втором - **Into Different Variables** (перекодировать в другие переменные).

Использование первой опции ведет к модификации первичной информации, так как будет перекодирована сама исходная переменная. Естественно, что в этом случае всегда надо «семь раз подумать», и как показывает опыт, лучшее решение здесь вообще «не резать». Практически всегда в данном случае правильнее создать новую переменную, т.е. работать с опцией Into Different Variables. Основное диалоговое окно этой опции и показано в верхней части на рис. 18.

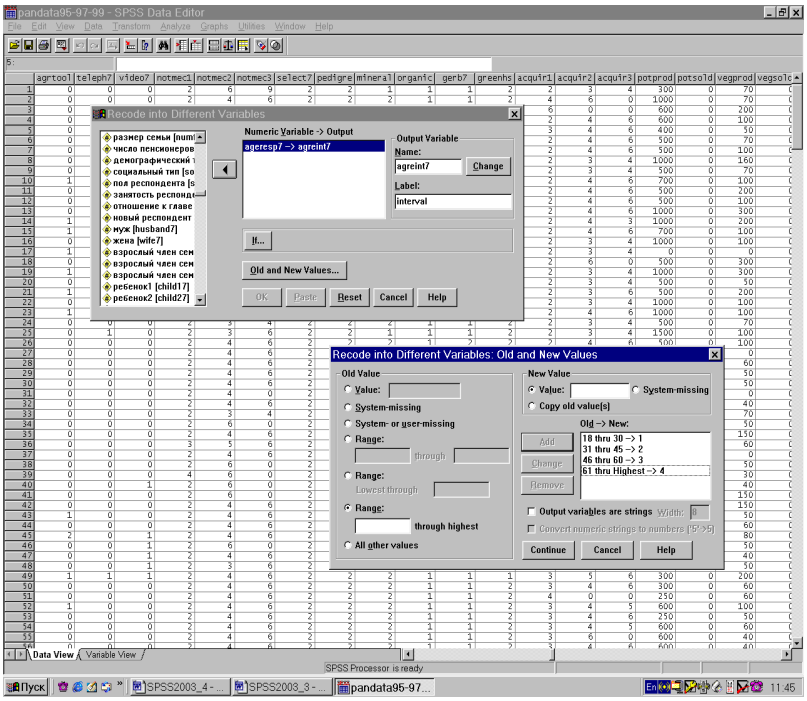



***Рис. 18.***

***Диалоговое окно Recode Into Different Variables - перекодировать в другие переменные***

В этом окне слева на экране - список переменных, которые могут быть исходными для перекодирования (**Input Variable**). Из него путем выделения выбирается нужная переменная и нажимается кнопка переноса - «стрелка вправо». Выбранная переменная переносится в центральное поле - **Numeric Variable - > Output Variable.**

Справа на экране находится поле **Output Variable**, которое используется для ввода новой переменной. Далее полезно, но необязательно заполнить поле **Lable**, в котором можно дать описание новой переменной. Наконец, следующим шагом необходимо нажимать кнопку **Change** (изменить).

В результате выполнения описанной выше последовательности команд в центре экрана - в поле, которое уже описано выше (Input Variable - > Output Variable), к введенному ранее имени исходной переменной после значка **- >** добавится имя новой переменной, которая будет построена на выходе. В данном случае здесь можно видеть следующее выражение: ageresp7 **- >** agreint7, которое словесно описывается как «возраст респондентов, в числе исполнившихся лет, преобразуется в интервальную шкалу».

Кнопка **If** задает условие отбора наблюдений, над которыми будет выполнена операция. Она работает аналогично той же кнопки, описанной ранее в командах Select Cases (глава 3, § 3.6) и Compute (глава 4, § 4.1).

Для ввода конкретных значений перекодировки необходимо нажать кнопку **Old and New Values** (старые и новые значения). При этом откроется дополнительное диалоговое окно с тем же именем: **Old and New Values**, которое показано в нижней части экрана на рис. 18.

Для каждого значения (или интервала), которое требуется перекодировать, в находящихся в левой части окна полях задается старое значение исходной переменной. Здесь можно использовать пять основных возможностей перекодировки или задания интервалов:

- **Old Value** - старое значение;
- **Range through** - от и до;
- **Range lowest through** - меньше чем;
- **Range through highest** - больше чем;
- **All other values** - все другие значения.

Открытие доступа к любому из указанных полей достигается путем пометки мышью, стоящего слева от него круглого чек-бокса. В качестве дополнительных возможностей система предлагает учет или иск-

лючение из преобразований пропущенных значений (System-missing, System-or user-missing).

Работа здесь выполняется по шагам. Сначала в одно из полей вводятся значения исходной переменной. Затем, в находящиеся в верхнем правом контуре **New Value** поля для новых значений (**Value** или **Copy old value(s)** -копировать старые значения), вводятся новые значения. Опыт показывает, что чаще всего используется поле Value, для работы в котором требуется пометка стоящего слева от него чек-бокса.

После ввода в поле Value нового значения следует щелкнуть мышью на кнопке **Add** (добавить). Задав все интервалы, необходимо нажать кнопку **Continue**, которая позволяет вернуться в главное диалоговое окно рассматриваемой команды. Далее, если все условия, которые заданы для выполнения команды, признаются пользователем корректными, необходимо нажать кнопку OK. Как и во всех подобных случаях, система либо создаст новую переменную (если условия были действительно заданы корректно), либо откроет окно и информацией о невозможности выполнения команды в связи с ошибочностью заданных условий.

Например, в преобразованиях, зафиксированных на рис. 18, в поле Old Value требовалось начать с задания исходного значения возраста с интервалом: от и до. Связано это с тем, что по условию респонденты не могли быть моложе 18 лет. Для этого мышью выделялся чек-бокс поля Range through, что и открывало к нему доступ посредством его активизации (из матового оно сразу же становится белым).

Сюда и вводились исходные значения: Range 18 through 30, а в поле Value - новое значение: 1, и далее следовало нажатие кнопки Add. В результате в поле Old -> New появлялась запись: 18 thru 30 - > 1. Затем в поле Old Value задавалось выражение: Range 31 through 45, а в поле Value – новое значение - 2 и т.д. Для задания последнего интервала (61 год и старше) сначала активизировалось поле Range through highest (от и выше), а затем в него вводилось старое значение: Range 61 through highest. Соответственно, в поле Value набивалось порядковое значение - 4. При необходимости каких-либо корректировок новой шкалы интервалов, любая из вновь созданных в поле Old -> New записей (после ее выделения мышью) может изменяться с помощью кнопки **Change** и убираться – кнопка **Remove**.

Далее следовало нажатие кнопок: Continue и OK. В результате выполненных преобразований в новой переменной  абсолютное значе

ние возраста каждого респондента заменялось его принадлежностью к одной из четырех возрастных групп.

## 4.3. Создание новой переменной с использованием процедуры Count

Процедура **Count** (счет) предназначена для подсчета повторений одного или нескольких одинаковых значений в списке переменных, в том числе и для подсчета числа повторений в заданном интервале.

Например, требуется подсчитать в массиве данных 1997 г. число детей в возрасте от 1 до 7 лет. Новая переменная - cage7, содержащая такую информацию по каждому наблюдаемому случаю, может быть получена на основе преобразований, выполняемых с тремя исходными переменными: cage17, cage27, cage37 (число лет каждого из трех, фиксировавшихся в обследовании детей в возрасте до 18 лет).

Для того, чтобы создать новую переменную с помощью процедуры Count, необходимо выбрать в главном меню:

**Transform**

**Count.**

Откроется диалоговое окно **Count Occurrences of Values within Cases** - подсчет повторений, которое и показано первым сверху на рис.19.

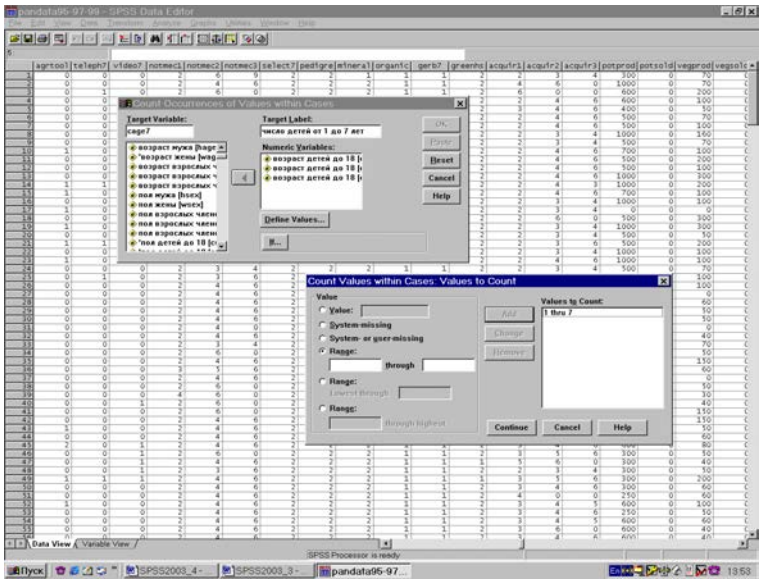



***Рис. 19.***

***Основное и дополнительное диалоговые окна процедуры Count***

Поле, находящееся в левой верхней части окна, называется **Target Variable** (целевая переменная). В него вводится имя новой переменной. Справа от него находится поле **Target Lable**, в которое полезно ввести описание новой переменной. Ниже расположен список переменных, из которого следует выбрать необходимые и перенести в правую часть - **Variables** (переменные).

При переносе из исходного списка переменных в окно Variables меняется его название на Numeric Variables (числовые переменные) и открывается доступ к дополнительным окнам **Define Values** (определение значений) и **If** (если).

В случае использования кнопки If открывается диалоговое окно **Count Occurrences: if Cases** (подсчет повторений: если), в котором по ранее описанным правилам работы с заданием условий (глава 3, § 3.6 и глава 4, § 4.1), задается условие для выполнения необходимой операции. Далее для выполнения команды опять же следуют кнопки Continue в дополнительном и OK в главном диалоговом окне.

В приведенном нами примере на рис. 19 в поле Target Variable введено имя новой переменной cage7 (число детей в массиве 1997 г. в возрасте от 1 до 7 лет), в поле Target Label введена метка: «число детей от 1 до 7 лет», в список Variables перенесены переменные: cage17, cage27, cage37.

Далее при нажатии на кнопку Define Values открывается дополнительное диалоговое окно: **Count Values within Cases: Values to Count** - значения для подсчета. На рис. 19 это окно показано нижним.

В левой части этого окна указываются необходимые интервалы (аналогично процедуре Recode). В нашем случае в левой части окна необходимо выбрать опцию Range through и заполнить ее поля значениями: 1 through 7.

Путем нажатия кнопки Add (добавить) заданный интервал значений переносится в правую часть экрана Values to Count (значения для подсчета). Затем следует нажатие кнопки Continue в дополнительном диалоговом окне (в результате чего оно закрывается) и кнопки OK в главном диалоговом окне рассматриваемой процедуры. Результат выполнения команды - новая переменная с указанием числа детей в возрасте от 1 до 7 лет в каждой наблюдаемой семье (случае).

## 4.4. Логические выражения и функции

Во многих задачах, связанных с преобразованием файлов с данными (Select Cases), равно как и с преобразованием самих данных (Compute, Count, Recode), постоянно возникает потребность в написании логических выражений и использовании различных функций, предлагаемых системой.

Как при использовании функций, так и при написании логических выражений, необходимо иметь хотя бы самые общие представления о работе с логическими операторами, которые служат одним из основных предметов рассмотрения различных направлений математического анализа, а именно: логики высказываний (24), булевой алгебры (25) и теории множеств (26).

У нас нет какой-либо возможности и морального права давать основы логики высказываний. Тот, кто испытывает такую потребность, должен обратиться к указанным выше первоисточникам.

На наш взгляд, в данном контексте полезно отметить **четыре** наиболее важных момента:

· Любое числовое значение переменной открывает возможности его соотношения (сравнения) с другими числовыми значениями.

· Отношения числовых значений формулируются в логических выражениях посредством различных операторов отношений.

· Функциональное выражение в SPSS представляет собой записанное с помощью специального командного языка системы - синтаксиса логическое выражение. В таких выражениях операторы отношений и логические операторы интегрированы в функциональные операторы самой системы SPSS.

Любое логическое или функциональное выражение предполагает вербальное (словесное) описание отношения сравниваемых переменных, а также их словесных и числовых значений.

**Правило 15**

***Словесная формулировка условия является непременным требованием задания логического или функционального выражения.***

Например, переменная «**пол**» (sex7) имеет два словесных значения (индикатора): мужской и женский и два числовых значения «мужской-1», «женский-2». Благодаря такой организации, оказывается возмож

ным с помощью процедуры Select Cases отобрать для анализа только мужчин или только женщин. При этом всегда задача предварительно формулируется сначала в виде словесного выражения: «отобрать мужчин» или «отобрать женщин», а затем и в виде логического выражения, соответственно: «**если пол равен единице** (sex7=1)» или «**если пол равен двойке** (sex7=2)».

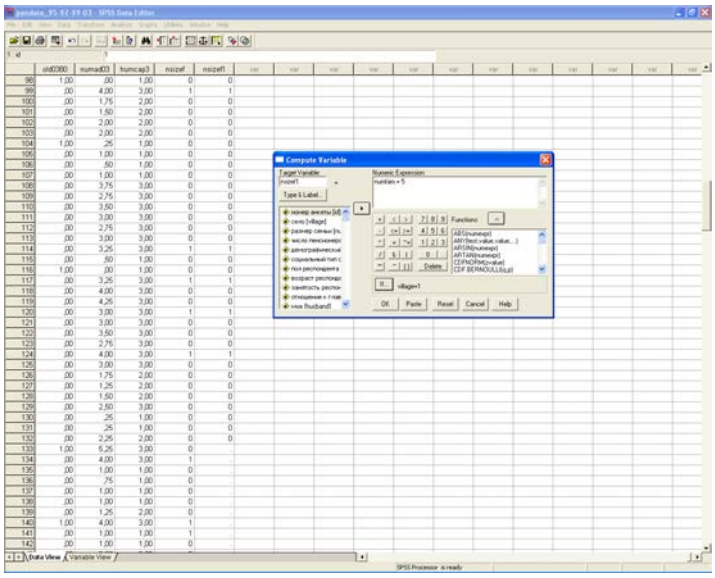
В диалоговых окнах соответствующих процедур средства записи логических выражений размещены в центральной части в виде небольшой клавиатуры, обеспечивающей возможность задания чисел, отношений (сложения, деления, умножения, равенства и др.), а также логических действий. Конъюнкция (&), дизъюнкция (|) и отрицание (-) все это логические действия (обрамление 3).

Средства записи функциональных выражений находятся в правой части диалогового окна. Здесь размещаются поле, содержащее список предлагаемых системой функций, и кнопка их переноса в поле записи числового выражения. Все это хорошо видно на рис. 20.

Условно говоря, любая задача может быть сформулирована в виде логического выражения, но сделать это, не имея опыта, довольно сложно. Для целей смягчения и обхода указанных трудностей и служат функциональные выражения.

***Рис.20.***

***Диалоговое окно записи логических выражений и функций***




Клавиатура содержит два блока кнопок. В первом слева блоке имеется 15 кнопок со знаками арифметических действий, операторов отношений и логических операторов. Во втором блоке имеется 12 кнопок (включая две нижних кнопки-клавиши). Его основу

составляют цифры от 0 до 9.

Арифметические операторы. Первая слева колонка кнопок в порядке движения сверху вниз включает кнопки: плюс, минус, умножение, деление, возведение в степень. Они вводятся путем нажатия (клик мышью) кнопок.

Операторы отношений, которые могут вводиться как путем нажатия соответствующей кнопки, так и с помощью буквенного сочетания (альтернативного текста). Всего таких операторов шесть (по три верхних кнопки во второй и третьей колонках слева). Для целей наиболее удобного восприятия они даны ниже в табличной форме (обрамление 2).

**Обрамление 2. Операторы отношений**

| Значение | Знак на кнопке | Буквенная запись |
|---|---|---|
| Меньше – Less then | < | LT |
| Больше – Greater then | > | GT |
| Меньше или равно – Less then or equal to | <= | LE |
| Больше или равно – Greater then or equal to | >= | GE |
| Равно – Equal to | = | EQ |
| Не равно – Not equal to | ~= | NE или <> |

Из приведенного обрамления видно, что буквенная запись логических операторов имеет своей основой их значение, записанное на английском языке.

Логические операторы, как и операторы отношений, могут вводиться и путем нажатия соответствующей кнопки, и с помощью буквенной записи (обрамление 3).

**Обрамление 3. Логические операторы**

| Значение | Приоритеты | Знак на кнопке | Буквенная запись |
|---|---|---|---|
| Логическое Не | 1 | ~ | NOT |
| Логическое И | 2 | & | AND |
| Логическое ИЛИ | 3 | \| | OR |

Логические операторы занимают две нижние кнопки в среднем ряду первого блока и предпоследнюю кнопку в правом ряду. Последняя кнопка в правом ряду первого блока имеет знак скобок ( ). Наличие

этого знака необходимо потому, что арифметические действия и логические операторы имеют определенную последовательность выполнения (приоритеты). Последовательность выполнения арифметических действий хорошо известна, что же касается логических операторов, то в приведенной выше таблице они даны в порядке приоритетности их выполнения.

**Правило 16**

***Формулируя логические выражения важно помнить, что та часть выражения, которая заключена в скобках, всегда представляет собой высший приоритет.***

Цифры. Они вводятся путем нажатия (клик мышью) соответствующих кнопок. При записи выражений цифры могут вводиться и непосредственно с основной клавиатуры компьютера. В этом блоке имеется кнопка с точкой для ввода десятичных значений, а также клавиша для удаления (Delete) выделенных фрагментов логических выражений.

Функции. Дополнительное диалоговое окно **If** предлагает большое число различных функций. Для переноса функции в редактор условий ее следует выделить, а затем щелкнуть мышью по кнопке с треугольником, которая находится между полем со списком функций и полем редактора записи условий логических выражений. Другая возможность переноса – двойной клик мышью по имени функции. В редакторе условий функция будет записана там, где предварительно находился курсор. Как правило, это начало записи условия.

Функция, вставленная в выражение, всегда требует дальнейшего редактирования. Вопросительные знаки, стоящие в скобках после имени функции, требуется заменить именами переменных, которые можно написать или перенеси из их списка. Количество вопросительных знаков в исходном выражении функции указывает на минимальное количество вставляемых аргументов.

Предлагаемые системой SPSS функции (functions) могут быть сгруппированы следующим образом: **логические (***logical***), строковые (***string***), арифметические (arithmetic), статистические (***statistical***), даты и времени (***date and time***), обработки отсутствующих значений (***treatment of missing-value***), поиск (извлечение) значений случаев (***search functions***), статистического распределения и генерации случайных чисел (***random variable and distribution functions***).**

Более полную информацию по данному вопросу можно найти во всех руководствах по использованию системы SPSS и языка «Синтаксис», сопровождающих каждую ее новую версию (например, 21, С. 45-75).

Краткое описание средств записи логических выражений с указанием вида функции приведено в следующей табличной подборке (обрамление 4), которая опирается на рисунок, отображающий диалоговое окно записи логических выражений и функций (рис. 20).

Формируя обрамление 4, мы старались на конкретных примерах показать процесс перехода от принятых в среде разработчиков качественных, словесных выражений к логическим выражениям и далее к функциям.

**Обрамление 4. Примеры записи отдельных выражений**

| **Словесное описание выражения** | **Логическое выражение** | **Функциональное (численное) выражение** | **Вид функции** |
|---|---|---|---|
| Возраст респондента больше или равен 30 лет и меньше или равен 50 лет в массиве 1999 г. | ageresp9 >= 30 & ageresp9 <= 50 | RANGE(ageresp9, 30,50) | Логическая |
| Отобрать значения переменной «село» равные единице или двойке. | village9 = 1 \| village9=2 | ANY(village9,1,2) | Логическая |
| Проверить истинность значений переменной село, которые находятся в интервале от 1 до 3. | village9=1 \| village9= 2 \| village9= 3 | ANY(village9,1,2,3) | Логическая |
| Вывести значения для переменной размер семьи в 1997 г. | namfam7 | VALUE(namfam7) | Строковая |
| Вывести абсолютные значения для переменной размер семьи в 1997 г. | namfam7 | ABS(namfam7) | Арифметическая |
| Число детей в семье в 1997 г. (при кодировке есть –1, нет – 0). | child7=child17 + child27 + child37 | child7=SUM (child17, child27, child37). | Статистическая |
| Средний денежный доход на члена семьи (руб. в месяц) в 1997 г. | midtot7=total7 / numfam7 | midtot7=MEAN(total 7 / numfam7). | Статистическая |

Разумеется, здесь очень важно как сочетание знаний и навыков логического мышления, так адекватное понимание анализируемых социальных явлений. Например, исходя из нашего опыта, для многих начинающих разработчиков демографический тип семьи представляет собой довольно сложную интеллектуальную конструкцию даже на уровне корректного словесного описания. Разумеется, что задать его логическое выражение в этом случае не представляется возможным.

Как видно из приведенных выше в табличной форме примеров, отдельные функции имеют сходный результат преобразования (например, строковая - VALUE и арифметическая - ABS). Другие, как, например, та же арифметическая – ABS, имеют своим результатом фактически абсолютные значения исходной переменной для каждого случая, третьи (ANY) позволяют проверять истинность значений и отбирать заданные значения, а четвертые (RANGE, SUM, MEAN) ведут к глубокому преобразованию исходных данных, позволяя создавать новые, отсутствовавшие при сборе данных переменные.

Система предъявляет определенные требования к порядку записи логических выражений и функций. Так, в первом приведенном нами примере: «Возраст респондента больше или равен 30 лет и меньше или равен 50 лет в массиве 1999 г.», логически корректной является запись 30 <= ageresp9 <= 50, но система заявит о ее некорректности, так как по условию выражение должно начинаться с имени переменной. Поэтому в данном случае корректной является запись, приведенная в примере, «ageresp9 >= 30 & ageresp9 <= 50». Указанное правило действует и для функций, в которых переменная и ее аргументы всегда заключены в скобки, указывающие на высший приоритет при выполнении расчетов.

В целом система предлагает огромное число функций различного вида. При современном состоянии социальных знаний далеко не все из них могут быть использованы в анализе и преобразованиях данных социологических исследований. Вместе с тем без использования огромных возможностей по преобразованию данных, предлагаемых системой, социологический анализ будет беднее и ограниченнее.

## 4.5. Пример использования вычислительных операций при создании новой переменной

Примером построения новой переменной с использованием вышеуказанных операций может служить созданная нами новая переменная - человеческий капитал - humcap97 (27, С. 190-234). Особенность этой переменной состоит в том, что для ее расчета требовалось создать ряд новых промежуточных переменных.

При определении переменной humcap97 за основу брался возраст всех членов семьи. Было выделено 9 возрастных групп и создано 9 новых переменных: возраст 1-7 лет (7old97), 8-11 лет (11old97), 12-14 лет (14old97), 15-16 лет (16old97), 17-65 лет (65old97), 66-70 лет (70old97), 71-74 года (74old97), 75-79 лет (79old97), 80-97 лет (80old97).

Каждая из перечисленных переменных создавалась следующим путем:

**Transform**

**Count**.

Откроется диалоговое окно Count Occurrences of Values within Cases. В левую часть окна Target Variable вводится имя новой переменной. Из списка переменных выбирается поэтапно, в зависимости от вышеуказанных интервалов, возраст всех членов семьи (сначала только детей, затем и детей, и взрослых для интервала 17-65, потом только взрослых членов семьи) и переносим в правую часть -Variables.

При нажатии на кнопку Define Values открывается диалоговое окно Count Values within Cases: Values to Count. В левой части указываются необходимые интервалы. Нажатием кнопки Add они переносятся в правую часть экрана Values to Count. Далее для завершения операции нажимаются кнопки Continue и OK.

В результате получаются 9 новых переменных, отражающих число людей в заданных возрастных интервалах. Экспертным путем была установлена степень участия членов семьи в хозяйственной деятельности крестьянского двора. Для наиболее активного возраста от 17 до 65 лет коэффициент участия – 1. Далее все коэффициенты участия нормируются по 1. В возрастных группах от 1 до 7 лет и для лиц старше 80 лет они равны 0. В возрастных группах от 8 до 11 лет и от 75 до 79 лет они составляют 0,25. В возрастных группах от 12 до 14 лет и от 71 до 74 лет они равны 0,50. Наконец, в возрастных группах от 15 до 16 лет и от 66 до 70 лет они составляют 0,75. Посредством

использования команды Compute (с учетом указанных коэффициентов) создается новая промежуточная переменная:

**Transform**

**Compute.**

Эта промежуточная переменная numadl97 - взвешенное число работников в домохозяйствах определяется по формуле:

**numadl97 = sum((7old97 * 0),(11old97 * 0,25),(14old97 * 0,50),(16old97 * 0,75),(65old97*1),(70old97*0,75),(74old97* 0,50),(79old97 * 0,25), (80old97 * 0))**

Следующий шаг - переход к созданию новой переменной humcap97 (человеческий капитал). Все значения numadl97 группируем: от 0 до 1 - семьи с низким человеческим капиталом; от 1,1 до 2,25 - семьи со средним человеческим капиталом и от 2,26 и выше - с высоким человеческим капиталом. В этом случае используем команду Recode. Выбираем в меню:

**Transform**

**Recode**

**Into Different Variables**.

Откроется диалоговое окно Recode Into Different variables. Из списка переменных выбирается numadl97 для перекодирования и нажимается кнопка «стрелка вправо». Справа на экране вводится имя новой переменной humcap97 и нажимается Change. В центре экрана - в области, которая называется: Input Variable -> Output Variable, - появятся введенные значения: numadl97-> humcap97.

Для ввода конкретных значений перекодирования необходимо нажать кнопку Old and New Values. Откроется окно Old and New Values. Каждому интервалу, который требуется перекодировать, задается старое значение из входной переменной: **Old Value - Range through** от 0 до 1. И новое значение для выходной переменной: **New Value - Value** - конкретное значение – 1.

Далее следует щелчок мышью по кнопке Add. Эта операция повторяется для интервалов 1,1 - 2,25 (Value =2) и 2,26 и выше (Value =3). Для выполнения команды нажимается кнопка OK. Результат, теперь уже всех преобразований, – новая переменная «человеческий капи

тал». Как видно из приведенного примера, ее расчет требует решения большого комплекса задач с использованием нескольких процедур. С использованием командного языка синтаксис вся эта работа может быть выполнена в одном блоке команд (глава 17, § 17.2).

Создание новых переменных в принципе относится к этапу формирования базы данных. Проблема состоит в том, что задача построения новых переменных возникает, как правило, именно в ходе статистических расчетов и анализа. И на их основе делаются расчеты и проверяются гипотезы.

Более того, отдельные процедуры такие, как регрессионный и факторный анализ, сами генерируют новые переменные (главы 13-15). Поэтому в действительности ситуация здесь довольно сложная, так как при построении новых переменных база данных первичной информации растет за счет дополнения ее фактически вторичной (расчетной) информацией. Постепенно разделить эти две части без специальных меток и границ становится все более трудным.

**Дружеский совет**

***Преобразование данных и связанное с ним построение новых переменных - мощнейшее средство анализа. Поэтому умение использовать процедуры, которые описаны в данной главе, - непременное условие успешного освоения системы SPSS.***

## Задание для самостоятельной работы

1. Что такое преобразование данных в системе SPSS?
2. В чем различие пунктов главного меню Transform и Data?
3. Какие процедуры преобразования данных вы знаете?
4. Перечислите последовательность команд при создании новой переменной с помощью процедур Compute, Count, Recode.
5. Какова роль кнопки If в процедурах Compute, Count и Recode?
6. Опишите особенности работы с диалоговым окном Recode Into Different Variables (перекодировать в другие переменные) в процедуре Recode.
7. Как работать с окном Compute Variable в процедуре Compute?

8. Опишите особенности работы с основным и дополнительным диалоговыми окнами процедуры Count.

9. Какие операторы вы знаете?

10. Перечислите виды функций, предлагаемых системой SPSS.

11. Опишите структуру и особенности работы с диалоговым окном записи логических выражение и функций.

12. Как можно перенести функции в логическое выражение в SPSS?

13. Чем отличаются вербальное, логическое и функциональное выражения?

14. Как редактировать функции в логических выражениях?

15. В чем отличие новой и исходной (исходных) переменных, принимавших участие в преобразованиях данных?

16. Напишите логическое выражение от вербального выражения: «брачная пара», при условии, что в переменной демографический тип семьи (demtype) одиночки = 1, брачная пара пенсионеров = 2, брачная пара с минимум одним работником = 3, нуклеарная семья = 4 и т.д.

17. Что означают два следующие выражения: midtot7=total7/ numfam7 и midtot7=MEAN(total7/numfam7)?

18. Что произойдет при выполнении последовательности команд: Transform - Recode - Into Same Variable?

19. Что произойдет при выполнении последовательности команд: Transform - Recode - Into Different Variables?

20. Для каких целей можно использовать возможность преобразования данных в SPSS?

21. Что означает создание новой переменной?

22. Где размещаются новые переменные?

23. Какая последовательность команд открывает главное диалоговое окно Compute Variable?

24. Что такое целевая переменная (Target Variable)?

25. Вспомните, пожалуйста, правила и советы, приведенные в этой главе.

# Глава 5. Контроль правильности ввода данных

## 5.1. Особенности этапа контроля

После ввода данных, непосредственно перед выполнением различных статистических расчетов аналитического характера, возникает необходимость контроля и исправления ошибок, которые были сделаны при вводе данных. Опыт показывает, что предварительный (визуальный) контроль заполнения анкет перед вводом в ЭВМ не исключает сохранения определенного числа ошибок интервьюера и кодировщика, хотя и ведет к их сокращению.

Сам ввод данных добавляет вероятность появления новых ошибок, связанных с работой операторов. Ошибок набивки, как правило, тем больше, чем больше операторов занято вводом. Последний момент особенно важен, так как сегодня только очень мощные структуры могут использовать труд профессиональных операторов. В обычной же ситуации российских исследований и разработок практически все виды работ выполняются одними руками. Подобная ситуация ведет к мультипликации ошибок кодировки и ввода.

Практически всегда кодировка, ввод и контроль данных тем качественнее, чем качественнее сделан документ, который мы называем «Макет ввода данных» (приложение 4). В работе немецких авторов, сходный по назначению документ называется «Кодировочная таблица» (16, С. 26-27).

Как бы не называли этот документ, его цель – установить соответствие между вопросами и индикаторами опросного листа и переменными, используемыми системой SPSS при машинной обработке данных. При обработке больших массивов, когда работа неизбежно должна выполняться несколькими кодировщиками и операторами ввода, отсутствие такого руководства (инструкции) кодировки и ввода фактически исключает возможность выполнения указанных работ.

В целом при решении задач поиска и исправления ошибок ввода, полезно помнить, что они бывают двух основных видов: **случайные и систематические.** Оба вида этих ошибок появляются как на этапе

сбора информации, так и при ее вводе.

Источниками появления ошибок могут служить и ограниченность используемой методики, и условия проведения опроса, и психофизиологические особенности самих опрашиваемых и опросчиков. Поэтому с ошибками необходимо работать как при визуальном контроле результатов опроса и кодировке первичной информации, так и по итогам ее ввода при подготовке данных к анализу и обработке с помощью системы SPSS.

Для целей поиска и исправления ошибок ввода в SPSS весьма эффективно использовать следующие процедуры: Sort Cases (сортировка случаев), Freguencies (частотные таблицы), Find (поиск), Compute (вычисление), Case Summaries (просмотр наблюдений) и Select Cases (отбор наблюдений), которые будут рассмотрены ниже как самостоятельно, так и в сочетании с другими процедурами.

**Дружеский совет**

***Ошибка — неизбежный спутник опроса и вводв данных. Их не надо бояться, но нужно тщательно искать и стремиться исправить еще по итогам ввода данных. В противном случае наличие ошибок создаст массу проблем в ходе анализа.***

## 5.2. Использование различных процедур для целей контроля данных

Процедура **Sort Cases** (сортировка случаев) позволяет сортировать данные по возрастанию или убыванию значения признака. Сразу же после ввода данных сортировка массива по возрастанию **Id –** идентификационного номера позволяет получить массив в порядке от первого до последнего случая независимо от порядка ввода их в массив и числа операторов, выполнявших его формирование.

Для выполнения процедуры Sort Cases необходимо выбрать в главном меню последовательность команд:

**Data**

**Sort Cases.**

Перенос Id в подокно Sort by, проверка установленной по умолчанию пометки Ascending в подокне Sort Order и последующее OK для завершения выполнения команды, в конечном счете, и позволяют окончательно подготовить массив к машинному контролю. С выполнения рассмотренной выше последовательности команд мы и рекомендуем начинать этап контроля данных.

Процедура **Frequencies** (частоты) создает таблицы, содержащие частоту встречаемости каждого значения переменной. Для выполнения процедуры Frequencies необходимо выбрать в меню:

**Analyze**

**Descriptive Statistics**

**Frequencies**.

Откроется диалоговое окно Frequencies - частоты (рис. 3, глава 1, § 1.3). В этом окне слева появится список переменных, которые можно выбрать, нажав кнопку «стрелка вправо». Выбранные переменные попадут в правый список. После нажатия кнопки OK, процедура начнет выполняться, и результаты будут выдаваться в окне просмотра.

Частотные таблицы используются в следующих случаях:

**При суммировании данных.** Например, требуется подсчитать число опрошенных мужчин и женщин. Для этого достаточно построить частотную таблицу по переменной «пол респондента». В нашем случае имя этой переменной для массива 1999 г. (sexresp9).

**При обнаружении ошибочных значений.** Например, значение 4 в частотной таблице по признаку села village9 указывает на ошибки при вводе данных, поскольку каждое из трех сел имеет значения, соответственно, 1, 2 или 3. Такого рода ошибки, возникающие при вводе данных, легко обнаружить и устранить. Делается это как раз с помощью построения частотных таблиц, позволяющих обнаружить указанные ошибки, и их последующей идентификации с использованием команд поиска или отбора случаев, которые рассмотрены ниже.

**При фиксации наблюдений с необычными значениями.** Например, в анкете может быть указано, что в семье имеется 10 тракторов. В частотной таблице, построенной по этой переменной (tractor9), такой случай сразу же получает отражение. В принципе подобное возможно, но, зная сельскую реальность, цифра вызывает подозрение и должна быть перепроверена путем идентификации данного случая. Это делается опять же с использованием команд поиска или отбора случаев.

**При проверке соответствия значений отдельных переменных.** При фиксации в исследовании структуры семьи в случае проверки соответствующего раздела «Социально-демографические характеристики семьи» удобно принимать за основу исходные частотные распределения по каждому члену семьи (первому, второму, третьему и другим членам семьи). При таком подходе итоговые значения частотных характеристик по каждому члену семьи (пол, образование, возраст, национальность, специальность, занятость) довольно легко и удобно контролировать посредством постоянного сравнения с соответствующими значениями характеристик, которые приняты за основу. Например, в массиве 2003 г. первый член семьи – муж есть в 288 из 382 случаев. Следовательно, каждая из всех других его характеристик (пол, образование, национальность и т.д.) должна встречаться в 288 случаях. Любое отклонение здесь свидетельствует о наличии ошибки.

Процедура **Find** (поиск). Ошибочные данные и данные, вызывающие сомнение, можно найти в редакторе данных с помощью функции поиска.

Для того, чтобы найти какие-то конкретные значения одной переменной, следует сначала выделить искомую переменную, щелкнув мышью на ее названии. При этом курсор автоматически переходит в первую ячейку выделенной переменной, которая сразу же окажется оконтуренной (в рамке). Указанное положение курсора необходимо для того, чтобы поиск был произведен по всему массиву от первого до последнего случая.

Далее выбираем в главном меню последовательность команд:

**Edit**

**Find**.

Откроется диалоговое окно Find Data in Variable (название переменной). Ввести в поле Find what то значение, которое требуется найти. Далее следует команда Find Next (кнопка в нижнем левом углу диалогового окна). Курсор автоматически установится на ячейке с указанным значением выбранной переменной. Если такое значение не найдено, об этом система также выдает сообщение. Поиск следующего случая предполагает повторное выполнение команды Find Next.

Процедура **Select Cases** (отбор случаев). Ошибочные значения какой-то переменной, равно как и ее значения, вызывающие сомнение, можно найти в редакторе данных и с помощью команды отбора случаев.

Для того, чтобы найти такие значения необходимо использовать следующий путь:

**Data**

**Select Cases.**

Отбор случаев тем более эффективен, чем больше искомых значений необходимо отобрать. В то же время использование функции поиска эффективно при необходимости идентификации одного, двух наблюдений.

Рассматриваемая процедура особенно продуктивна при последующем использовании процедуры просмотра отобранных наблюдений.

Процедура **Case Summaries** (просмотр наблюдений) предназначена для вывода на экран переменных и наблюдений в табличном виде. Нами эта команда использовалась не только для сравнения изменений значений идентичных переменных панельного исследования за три года, но и в качестве средства контроля правильности ввода данных.

Для выполнения этой команды необходимо выбрать в меню:

**Analyze**

**Reports**

**Case Summaries**.

На экране появится диалоговое окно Summarize Cases (список наблюдений). В правой части окна - список всех переменных. После выбора переменной и нажатия кнопки «стрелка вправо», переменная появится в левой части экрана, в списке переменных, которые будут выдаваться в окно просмотра. После нажатия кнопки OK отобранные переменные в табличном виде отобразятся в окне просмотра**.**

Например, логично предположить, что в контролируемом массиве данных каждому члену семьи соответствуют идентичные числовые значения по всему набору характерных для рассматриваемых членов семьи признаков (вопросы 1 – 7, приложение 2).

Выполнив рассматриваемую процедуру для переменных: муж, возраст мужа, пол мужа, образование мужа и других его социальных характеристик, в первой из двух таблиц окна просмотра (Case Processing Summary, Included, N) можно сразу увидеть, что для всех этих характеристик свойственно одно и то же число. Отличие рассматриваемого числа для какой-то переменной - свидетельство ошибки ввода, которую можно найти путем просмотра значений соответствующих характеристик для каждого отдельного случая во второй таблице окна

просмотра (Case Summaries). Фиксация существенного разброса чисел - знак большой тревоги, связанной с наличием огромного числа ошибок ввода.

Другая интересная возможность использования данной процедуры описана нами в одной из опубликованных работ (27, С. 9, 231). Она связана с выполнением следующей последовательности команд: **Data – Sort cases**, и далее **Analyze – Reports – Case Summaries** – перенос из общего списка переменных в дополнительное окно списка имен переменных, подлежащих проверке или анализу, в том числе и идентификационного номера (id), без которого найти необходимый случай просто невозможно. Выполнение приведенной выше последовательности команд при необходимости открывает возможность полного описания любого конкретного случая. Указанное обстоятельство имеет исключительно важное познавательное значение.

Процедура **Compute** (вычислить) предназначена для создания новых переменных или изменения уже существующих. Она может быть полезна в следующих случаях:

- при создании новых переменных на базе первичной информации при подготовке к анализу данных. Этот случай в общем виде описан в главе 4, § 4.1, в то же время конкретный пример его использования дан в настоящей главе (§ 5.5);
- при создании новых переменных на этапе контроля данных;
- при изменении переменных на этапе исправления ошибок.

Два последних случая имеют непосредственное отношение к данному разделу и будут описаны ниже.

На этапе контроля новые переменные уместно создавать для проверки правильности ручных расчетов, выполнение которых в отдельных случаях (например, при фиксации доходов из различных источников для каждого члена семьи) неизбежно в ходе полевых работ. Конкретно, в нашем случае суммарный доход первого члена семьи (переменная total17) был рассчитан вручную. Для его проверки создается новая переменная – новый суммарный доход первого члена семьи (newtot17): newtot17 = psalw17 + salwag17 + pensio17 + alimon17 + ch17 + incplo17 + divid17 + income17 + othben17.

Далее сравниваются значения исходной переменной (total17) со значениями вновь созданной переменной (newtot17). Но самое главное, что указанные расчеты необходимо сначала выполнить для каждого взрослого члена семьи, а затем и для семьи в целом. Эта процедура более подробно описана в данной главе (§ 5.5).

**При контроле данных может возникнуть необходимость корректировки пропущенных значений переменных,** которые описаны в главе 2, § 2.6**.** Выполнение этой операции предполагает изменение значения переменной. Например, при вводе данных в массив 1997 г. по продаже молока (milksol7) пропущенное значение этой переменной задавалось равным 8 (приложение 4).

Между тем, в некоторых случаях оператором ошибочно набивалось число 88. Содержательно это вело к тому, что в хозяйствах, не реализующих молоко, указанные значения воспринимались как продажа 88 литров молока. Другими словами, данная ошибка вела к завышению объемов продаж молока.

Исправление такого рода ошибок осуществляется с помощью процедуры Compute следующим образом. Выбрать в меню:

**Transform**

**Compute**.

Откроется диалоговое окно Compute Variable (вычислить переменную). Далее есть два пути исправления ошибки.

*Первый путь* связан с использованием логического выражения. В этом случае в левой части окна Target Variable (целевая переменная) вводится имя переменной (в данном случае milksol7), которая получает в правом окне Numeric Expression (числовое выражение) вычисленное значение, равное 8.

Для задания условного выражения используется кнопка If. Она открывает дополнительное окно, в которое записывается условие «milksol7=88». Вербально сформулированное условие можно записать так: **«Если значение переменной равно 88, то его следует заменить значением 8»**. А логическое выражение указанного условия в целом имеет вид: «milksol7=8 if milksol7=88».

*Второй путь* связан с использованием функции «MISSING». В этом случае, как и в первом, в левой части окна Target Variable вводится имя переменной (milksol7), а в правое окно переносится функциональное выражение «MISSING(?)», в скобки которого вставляется имя исходной переменной. При этом запись приобретает вид: «MISSING (milksol7)». Далее опять используется кнопка If и задается условие: «milksol7=88». Последующее выполнение команд Continue и OK ведет к исправлению соответствующих ошибок в исходной переменной. При этом в случае использования функции (независимо от того, какие уста-

новки использовались при описании пропущенного значения для данной переменной) в качестве нового пропущенного значения устанавливается 0.

Перед выполнением команды ОК в основном диалоговом окне данной процедуры система всегда просит подтверждения на изменение существующей переменной «Change existing variable?». Последнее обстоятельство связано с тем, что в окно Target Vatiable было вписано имя исходной, а не новой переменной.

Преобразование типа: «**Если переменная имеет значение равное 88, то его необходимо заменить на 8**» можно легко и просто выполнить с помощью процедуры Recode (перекодировка). Для этого необходимо использовать следующий путь:

**Transform**
**Recode**
**Into Same Variable**.

Далее в диалоговом окне **Recode into Same Variables** из левого поля со списком переменных переносим необходимую переменную в правое поле Numeric Variables. Нажимаем клавишу **Old and New Values.** Она открывает дополнительное диалоговое окно **Recode into Same Variables: Old and New Values** (глава 4, § 4.2). В этом окне в обрамлении **Old Value** устанавливается опция **Value**, и в ее поле вносится исходное значение переменной - 88.

Затем в правой части окна в обрамлении **New Value**, в поле **Value** вносим новое значение – 8 и нажимаем кнопку **Add.** В результате в поле **Old->New** появляется выражение: 88 ->8. Нажатие кнопки Continue ведет к возврату в главное диалоговое окно процедуры. Далее выполняем команду ОК. В результате в переменной происходит смена всех ошибочных значений 88 на новые значения равные 8.

**Дружеский совет**

***В этом, как и во многих других случаях, разработчики каждый раз сами должны принимать решение, что для них нужно в данном случае: преобразовать исходную или создать новую переменную.***

## 5.3. Практика решения задач контроля

Решая задачи поиска и исправления ошибок ввода, как уже отмечалось ранее, мы исходили из предпосылки, что они бывают двух основных видов: случайные и систематические. Оба вида этих ошибок могут быть допущены и допускаются как на этапе сбора информации, так и при ее вводе.

Первый вид ошибок характерен для квалифицированных интервьюеров, кодировщиков и операторов. Систематические ошибки связаны с квалификацией и психофизическими особенностями персонала. Вместе с тем они могут быть заложены и в саму полевую документацию как результат ошибок разработки программы и методики исследования. Поэтому задача машинного контроля с использованием различных вычислительных операций и процедур, позволяющих выявить и исключить как технические, так и логические ошибки, практически всегда остается актуальной (14 , С. 145-146). Использование в качестве интервьюеров, кодировщиков и операторов неквалифицированного технического персонала, взятого со стороны, заведомо ведет к появлению ошибок и того, и другого вида.

При проверке данных в первую очередь полезно использовать статистическую процедуру **Frequencies** (частоты). Эта процедура, как и описанная в предшествующем параграфе 5.2 процедура Case Summaries, позволяет выявить многие ошибочные значения переменных.

Выполнение рассматриваемой процедуры, как уже отмечалось ранее, откроет диалоговое окно **Frequencies** (рис. 3, глава 1, § 1.3). В этом окне слева появится список переменных, которые можно выбрать, нажав кнопку «стрелка вправо». Выбранные переменные попадут в правый список. После нажатия кнопки OK, процедура начнет выполняться, и результаты будут выдаваться в окно вывода. Пример такого вывода показан в табл. 1.

Из табл. 1, которая представляет собой слегка модернизированную в текстовом процессоре MS Word частотную таблицу окна вывода SPSS, видно, что в переменной demtype7 в одном случае присутствует ошибочное значение 8. Его необходимо найти и исправить. Быстрее всего это можно сделать с помощью команды Find. Порядок ее выполнения уже описан ранее в § 5.2.

Сейчас же мы только восстановим этот путь. Сначала выделяем переменную. Далее следует: Edit – Find – 8 – Find next – рамка курсора оконтуривает искомый случай, который требует исправления. Если бы

у нас был еще такой же случай, то следует повторить команду Find Next и рамка курсора встанет в искомую ячейку. Понятно, что искать десять или двадцать таких случаев, хотя и не смертельно, но, условно говоря, скучно. Для этой цели и существует отбор случаев, который сразу решит всю эту задачу.

**Таблица 1. Демографический тип семьи (DEMTYPE7)**

| Value Label | Value | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Одиночки | 1 | 101 | 21.8 | 21.8 | 21.8 |
| Супружеские пары пенсионеров | 2 | 52 | 11.2 | 11.2 | 33.0 |
| Супружеские пары работников | 3 | 32 | 6.9 | 6.9 | 40.0 |
| Супружеские пары с детьми | 4 | 140 | 30.2 | 30.2 | 70.2 |
| Супружеские пары с детьми и др.родственниками | 5 | 59 | 12.7 | 12.7 | 82.9 |
| Неполные семьи | 6 | 15 | 3.2 | 3.2 | 86.2 |
| Смешанные (прочие) | 7 | 63 | 13.6 | 13.6 | 99.8 |
| | 8 | 1 | 0.2 | 0.2 | 100.0 |
| Total | | 463 | 100.0 | 100.0 | |

В случаях, когда надо проверить однозначность записи, например, пол мужа/жены, полезно использовать следующий путь:

**Data**

**Select Cases**.

Выбираем опцию If condition is satisfied и нажимаем кнопку If, открывается другое окно, там пишем имя проверяемой переменной с заданным условием: hsex7 (пол мужа) не равен 1.

Далее Continue и OK. Создается фильтр. При этом на экране слева в порядковой нумерации, задаваемой системой, появляются зачеркнутые и не зачеркнутые номера анкет. Не зачеркнутые номера анкет отвечают заданному нами условию, т.е. пол мужа не равен 1. Можно найти эти анкеты, визуально просматривая весь массив. Эта работа

облегчается тем, что после выполнения команды курсор устанавливается на первом по порядку случае с ошибкой, которую сразу же можно и исправить. Тем более, что повторное выполнение процедуры установит курсор на следующей по порядку ошибке. Все же этот путь достаточно трудоемкий, особенно при наличии большого числа случаев в массиве.

Минимизация затрат сил и времени здесь может быть выполнена с использованием процедур **Frequencies** или **Case Summaries**. При наличии фильтра **з**адание в них идентификационного номера анкеты (Id) выводит в окне просмотра номера случаев с ошибками. Естественно, все это не избавляет от ручной работы, связанной с внесением исправлений в массив.

Для целей контроля может быть использована и процедура **Compute**, которая в отличие от рассмотренных выше, позволяет вносить исправления машинным путем. Примером ее использования для таких целей может служить описанное в предшествующем параграфе 5.1 исправление пропущенных значений в переменной, связанной с продажами молока.

Отдельные переменные можно проверить путем контроля их взаимосвязи. Например, «занятость» с «занимаемой должностью», «местом работы», «предприятием» взрослых членов семьи (приложение 2). Типичная ошибка - при отсутствии занятости указываются занимаемая должность, место работы и тип предприятия.

Для обнаружения таких ошибок эффективно использовать процедуру **Crosstabs** (таблица сопряженности), которая позволяет пересечь две переменные. Например, hempl7 (занятость мужа) и hpos7 (должность мужа) или wempl7 (занятость жены) и wpos7 (должность жены). Аналогично указанная задача решается для всех других взрослых членов семьи.

Для выполнения этой процедуры в меню выбирается:

**Analyze**
**Descriptive Statistics**
**Crosstabs**.

В открывшемся диалоговом окне Crosstabs из поля со списком переменных переносятся искомые переменные, соответственно, занятость респондента в поле Row(s), а должность в поле Column(s). При контроле расчеты лучше всего выполнять для абсолютных

значений, поэтому, не задавая каких-либо дополнительных установок, сразу же выполняем команду OK. В результате получаем в окне просмотра следующую информацию (табл. 2).

**Таблица 2. Контроль характеристик занятости и должности опрошенного**

| Занятость | Должность | | | | | | |
|---|---|---|---|---|---|---|---|
| | Руковод. | Специал. | Служ. | Рабоч. | Фермер | Другое | Всего |
| Полная | 18 | 20 | 7 | 155 | 6 | 4 | 210 |
| Частичная | - | 1 | 1 | 6 | - | 1 | 9 |
| Безработн. | - | - | - | - | - | 4 | 4 |
| Пенсионер | - | - | - | 1 | - | - | 1 |
| Total | 18 | 21 | 8 | 162 | 6 | 9 | 224 |

Из данных табл. 2 видно, что общее число занятых и занимающих определенную должность опрошенных совпадает (224 чел.), и это хорошо. Но среди пенсионеров имеется один рабочий. По условиям опроса, если пенсионер работает, то он должен идти по строке полная или неполная занятость, а если он не работает, то его вообще не должно быть среди работников. Отсюда вывод: надо искать соответствующий случай и делать в нем исправления.

Поиск ошибочного наблюдения в этом случае, как и ранее, может быть выполнен с помощью описанных выше команд. При этом используется следующий путь: Select Cases – If hempl7=4 & hpos7=4 – Continue – OK. Далее, выполняются поиск (Find) или просмотр наблюдений (Case Summaries), или частоты (Frequencies) для идентификации и исправления случаев с ошибками.

Особого контроля требуют переменные, полученные в результате расчетов интервьюеров. Например, расчет суммарного дохода каждого члена семьи и семьи в целом (вопросы 24-26, приложение 2). Здесь чаще всего встречаются арифметические ошибки.

В принципе такие расчеты, казалось бы, должны делаться только машинным путем. Вместе с тем, с точки зрения технологии опроса, все выглядит далеко не так однозначно. Во-первых, наличие суммарной оценки дохода способствует созданию более благоприятного климата отношений интервьюера и респондента, а, следовательно, и повышению достоверности собираемой информации.

Во-вторых, для целей нашего интервью необходим был расчет доли натурального потребления (вопрос 25) и общего месячного дохода (вопрос 26). Получить эту информацию без расчета месячного дохода (последняя строка в вопросе 24) практически было невозможно. Другими словами, наличие ошибок подобного рода в полевой документации как бы предполагается и закладывается технологией самих работ по сбору первичной информации.

Включение в анкету таких обобщенных характеристик семьи, как демографический и социальный типы (титул опросного листа, приложение 2), которые не спрашивались у респондента, а определялись по определенному алгоритму интервьюерами, в принципе является избыточным с точки зрения технологии обработки данных в SPSS.

Нами эти характеристики были включены, несмотря на критическое отношение американских коллег, чисто по причинам консерватизма и недоверия к возможностям SPSS. Ведь наш опыт контактов с техникой формировался во времена, когда еще отсутствовал персональный компьютер. Обработка данных в те годы занимала больше времени, чем полевые работы. Сейчас же, выполнив полевые работы в российских селах в июне-августе, мы уже в сентябре имели введенные и отконтролированные данные.

Возвращаясь к проблеме контроля, уместно отметить, что его смысл в данном случае состоит в сравнении исходных сумм в первичной документации с машинными расчетами. Используется следующий путь:

**Transform**

**Compute**.

Далее вводим имя новой переменной (например, суммарный доход первого члена семьи newtot1). В правом окне ищем знак суммы (SUM) и переносим стрелочкой вверх в рабочее окно, а из левого окна (список переменных) переносим все виды доходов первого члена семьи. Получаем формулу: **SUM** (psalw17, salwag17, pensio17, alimon17, ch17, incplo17, divid17, income17, othben17). Нажимаем на OK.

Новая переменная newtot1 появляется в таблице и является расчетной суммой доходов первого члена семьи. Затем ищем разницу между полученной суммой и имеющейся ранее total17 (суммарный доход первого члена семьи).

Используем последовательность команд:

**Data**

**Select Cases.**

Кнопка If открывает другое окно, где задается условие newtot1-total17 **не равно 0.** Далее следуют Continue и OK. Создается фильтр, в котором есть все случаи расхождений (не перечеркнутый номер по порядку). Для поиска ошибок используется путь:

**Analyze**

**Reports**

**Case Summaries.**

Из списка переменных выбираются Id7, newtot1, total17 и нажимается OK. В таблице окна просмотра выводятся номера анкет для всех случаев с расхождением данных. Эта процедура повторяется для каждого взрослого члена семьи.

Сходная последовательность процедур выполняется и при проверке расчетов по переменной total7 («суммарный доход семьи», приложение 4). В результате этих расчетов создается новая переменная «суммарный доход семьи» - newtota7, полученная уже машинным путем.

В случае необходимости пересчета общего месячного дохода семьи (sumtota7) с учетом доли потребления продуктов питания из личного подсобного хозяйства к месячному денежному доходу семьи (quota7) используется путь:

**Transform**

**Compute.**

Вводится выражение: **sumtot7** (имя новой переменной) = **newtota7** (новый суммарный доход)*** (100+quota7):100**. Полученный результат сравнивается с исходным доходом, фиксировавшимся в первичной информации (sumtota7). Вполне вероятно, что такие ошибки можно и не контролировать, и даже исходно не вводить эти данные. В таком случае необходимо сразу делать расчет новых переменных по всей структуре суммарных доходов, начиная от итогового денежного дохода каждого взрослого члена семьи (предпоследняя строка в вопросе 24, приложение 2) и кончая общим месячным доходом (вопрос 26, приложение 2). Порядок создания новых переменных описан ранее в главе 4, § 4.1-4.3.

В целом контроль ввода данных – это довольно трудоемкая и мало приятная задача. Видимо, поэтому в различных учебниках ей уделяется явно недостаточное внимание. Тем не менее, эту работу необходимо делать. Следует подчеркнуть, что она имеет не только техническое, но и содержательное значение.

Наблюдательному исследователю выполнение работ по контролю данных позволяет, во-первых, полнее освоить командный комплекс системы SPSS. Многие операции, используемые при контроле данных, в последующем используются и при их анализе. Во-вторых, выполнение указанных работ дает возможность быстрее и полнее освоить подготавливаемый к анализу массив данных. Наконец, в-третьих, хорошее знание массива позволяет уже на начальных этапах работы с ним увереннее и корректнее формулировать самые разнообразные предварительные гипотезы и обобщения.

**Общая информация**

***Наш опыт показывает, что выполнение всех шагов, связанных с вводом и контролем данных в таблице SPSS, позволяет получить базу данных, хорошо подготовленную к статистическим расчетам. Краткая итоговая инструкция выполняемых с этой целью шагов приведена в приложении 5.***

## 5.4. Модуль ввода данных

Описанная в предшествующих параграфах этой главы практика ввода и контроля данных связана с выработкой у пользователя соответствующих навыков работы. Между тем, примерно в середине 90-х годов в SPSS для указанных целей был создан специализированный модуль, который помогает обеспечить быстрый и качественный ввод данных. Имя этого программного средства «SPSS Data Entry». Сейчас уже реализуются его 3-я и 4-я версии.

Эта программа разработана для ввода данных с форм и вопросников. Опираясь на эти формы, система автоматически создает соответствующие переменные. Такие формы можно дополнять правилами проверки, как отдельных переменных (Validation rules), так и логического соответствия нескольких переменных (Checking rules).

Иными словам, в отличие от мягкого призыва к созданию макетов ввода данных, который был сформулирован ранее, такая организация ввода связанна с жестким требованием создания системного варианта подобной документации, который реализован в формах и вопросниках.

При организации работ с помощью формы данные вводятся путем проставления отметок в элементах управления формой или путем ввода значений в текстовые окна. Форма запоминает данные и сохраняет их в файле данных. Вместе с тем, как и в таблице основного пакета, для организации скоростного ввода в Data Entry имеется возможность использования табличного формата ввода данных.

Согласно информации производителя, программное обеспечение рассматриваемого приложения имеет несколько составляющих элементов:

- **DataEntry Builder** - создание форм/вопросников и правил чистки;
- **DataEntry Station** - ввод данных на локальных компьютерах;
- **DataEntry Enterprise Server** - организация сбора/ввода данных в Интернете/Интранете или в локальной сети (http://www.spss.ru/products/dew/). Вполне естественно, что за все это надо нести дополнительные расходы ресурсов машины и финансовых средств пользователя.

Если исходить из условий работы на «заводах» по производству социально-экономической информации, связанной с формированием общественного мнения или спроса потребителей, то использование систем типа Data Entry дает несомненные преимущества в экономии времени и сил.

Для таких структур характерна примерно следующая технологическая цепочка действий: **получение заказа -** разработка **стандартизированного опросного листа –** формирование **большого массива выборки** (в тысячах наблюдений) – использование **для опроса сети интервьюеров (**которые по условию не будут иметь доступа к данным) – использование услуг **операторов ввода данных** (которые мыслят в терминах скорости ввода сотен знаков в минуту) – использование услуг **служб, выдающих расчеты** на любой вкус аналитиков – продукция самих **аналитиков**.

Основная особенность такой организации работ – глубокое разделение труда, ведущее к полному отчуждению аналитиков от этапов сбора и обработки данных. Многим из них, уже имеющим заготовки макро выводов и обобщений, данные по большому счету и не нужны.

Это хорошо видно по публикациям и комментариям аналитиков в СМИ. В таких случаях текст и данные очень часто находятся в контрастах и противоречиях. Здесь взгляды и интересы управляют данными.

Совсем иная технологическая цепочка организации работ по сбору и анализу данных наблюдается в науке, учебном процессе**,** управлении и корпоративном секторе. Она может быть представлена следующим образом: **формулировка гипотезы - разработка индивидуализированного опросного листа – формирование ограниченного массива выборки** (в сотнях наблюдений) – **использование для опроса, ввода и анализа данных собственных сил** (личных и групповых ресурсов очень ограниченного круга людей).

При такой организации работ ввод и контроль данных, равно как и их сбор, нельзя рассматривать в качестве технической задачи. Для человека, работавшего в поле, ввод наблюдения внутренне воспроизводится как формализованное интервью, а систематическая повторяемость или разброс отдельных значений уже на этапе ввода служат источником формулировки различных гипотез и предположений. Это мир эмпиризма, в котором данные управляют выводами и обобщениями. На широкий круг таких пользователей и ориентировано настоящее учебное пособие. Поэтому мы и говорим о своем консерватизме в отношении использования модуля Data Entry. На родине рассматриваемого продукта наши коллеги в подобных случаях говорят: «Too much technology!» (переизбыток технологии).

## Задание для самостоятельной работы

1. Что такое контроль данных?
2. Назовите виды контроля, которые вы знаете.
3. Какие команды SPSS используются при контроле в первую очередь?
4. В чем специфика использования команды Select Cases при контроле данных?
5. Какие виды ошибок ввода вы знаете?
6. Приведите несколько примеров ошибок ввода.
7. Как можно изменить пропущенные значения?
8. В чем специфика использования команды Find при контроле данных?

9. В чем отличие случайных ошибок от систематических ошибок?

10. В чем специфика использования команды Case Summaries при контроле данных?

11. В чем отличие использования информации, содержащейся в двух таблицах окна вывода команды Case Summaries, при контроле данных?

12. В чем специфика использования команды Frequencies при контроле?

13. Что делают с открытыми вопросами при подготовке массива данных к кодировке и вводу?

14. Как можно использовать процедуру Crosstabs для контроля данных?

15. В чем специфика использования команды Compute при контроле данных?

16. Чем следует руководствоваться при вводе данных?

17. В чем специфика использования команды Recode при контроле?

18. Для каких целей необходим идентификационный номер анкеты?

19. В чем отличие выполнения работ в окнах Recode into Same Variables и Recode into Different Variables процедуре Recode?

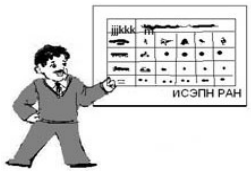
20. Для каких целей необходим «Макет ввода данных»?

21. Какие источники ошибок вы знаете?

22. Приведите пример словесного описания логического выражения, используемого в процедуре Compute при контроле данных.

23. Как выполняется процедура Find?

24. Перечислите основные шаги работ, выполняемых при вводе и контроле данных в SPSS.




# Раздел 2.

# АНАЛИЗ ДАННЫХ: ОБЩИЕ ПРИНЦИПЫ, СУММАРНЫЕ СТАТИСТИКИ И ГРАФИКИ

## Глава 6. На пути к анализу данных

### 6.1. Цели и задачи социологического анализа

Сбор, ввод и контроль данных, которые завершаются созданием базы данных или файла с расширением .sav (формат SPSS), открывают дорогу к последующей обработке и анализу первичной информации (глава 3, § 3.1). В SPSS анализ данных может иметь разную глубину и выполняться с помощью широкого набора статистических методов.

Возможности использования предлагаемых в SPSS процедур и методов расчета обуславливаются не только навыками работы с программным продуктом. В первую очередь они связаны с четкой фиксацией и пониманием целей, задач и основных гипотез исследования. Эта работа должна выполняться еще на этапе разработки программы исследования как описательно, так и путем операционализации понятий и индикаторов в полевой документации: анкетах, опросных листах и бланках формализованных интервью.

Конечно, все это общие правила, которые каждым исследователем (группой) выполняется по своему усмотрению. А это значит, что, если бы в социологии был такой же жесткий контроль за выполнением этих правил, как на автодорогах, то исследовательские происшествия,

связанные с их нарушением и ведущие к снижению достоверности и надежности выводов, фиксировались бы непрерывным гулом сирен соответствующей службы. Справедливости ради уместно отметить, что при таком контроле и общественные затраты на исследования и разработки в рассматриваемой области были бы гораздо выше. В любом случае, сформировав файл данных, перед тем, как приступить к их анализу, всегда полезно иметь в виду следующие три весьма важные момента.

Первый из них связан с выборочным характером основной массы социологических исследований. **Выборочная совокупность (выборка) -** единицы наблюдения, тем или иным способом и в том или ином количестве, отобранные из конкретной генеральной совокупности для их изучения в процессе социологического исследования.

Отсюда следует, что любой анализ данных выборочного обследования предполагает актуализацию знаний о г**енеральной совокупности,** как наборе всех объектов (единиц наблюдения), относительно которых исследователь собирается делать выводы, анализируя данные своей выборочной совокупности (29, С. 21). Актуализация необходима для сравнения значений известных (доступных) характеристик генеральной совокупности и соответствующих значений характеристик наблюдаемого объекта изучения (выборочной совокупности).

Например, изучая структуру и размер сельской семьи, равно как и средний размер оплаты труда и пенсий, мы были просто обязаны знать значения этих характеристик на страновом и региональном уровнях. Учет данного факта важен как в плане контроля надежности данных, так и для собственно аналитических целей, на которые ниже будет обращено более пристальное внимание.

**Напоминание**

***Анализ на уровне первичных распределений (частот), описательной статистики и сравнения средних возможен и имеет смысл только с учетом сделанного выше замечания.***

Второй момент связан с необходимостью актуализации знаний и понимания целей, задач и основных гипотез исследования к моменту начала этапа обработки и анализа данных. Построение любой таблицы распределения, равно как и анализ любых связей, по существу предс

тавляют собой процесс проверки различных гипотез. При этом в случае корректной постановки задачи фиксация тесной связи работает на подтверждение гипотезы, а ее отсутствие позволяет делать вывод о том, что проверяемая гипотеза не подтвердилась. Как известно, в науке проблема состоит совсем не в том, чтобы каждая гипотеза подтверждалась.

Куда более важно, чтобы в процессе исследования разработчики не действовали по принципу пересечения всего со всем, в основе которого лежит известный метод проб и ошибок. В социологии, на этапе анализа данных, последовательная реализация этого принципа ведет к расчету и распечатке огромной массы таблиц, основная часть которых по условию никогда не будет просмотрена и осмыслена.

В этом плане SPSS дает широкую возможность предварительного просмотра и корректировки выполняемых расчетов. Вместе с тем указанное обстоятельство открывает скорее перспективы экономии бумаги, чем непосредственного повышения эффективности и качества анализа данных. Последняя задача может быть решена только в том случае, если разработчик, задавая ту или иную таблицу, как бы предсказывает характер предполагаемого распределения и строит для самого себя его вербальное описание.

Содержание такого рода суждения в скрытом или явном виде неизбежно содержит гипотезу. Для целей ее подтверждения или опровержения в режиме реального времени рассматриваемый программный продукт и является незаменимым помощником.

**Напоминание**

***Анализ на уровне построения таблиц распределения, корреляционных связей и регрессии возможен и имеет смысл только с учетом сделанного выше замечания.***

Более того, при регулярном обращении и работе с SPSS он, как и любой тренажер, закрепляет и повышает навыки рефлексии и аналитического (научного) мышления, выполняемого в терминах причинно-следственных связей.

Наконец, последний третий момент связан с корректным пониманием функционального назначения и характера **переменных,** используемых в анализе. Этот момент имеет исключительно важное значение и будет предметом специального рассмотрения в следующем параграфе.

## 6.2. Переменные и их роль в анализе данных

В главе 2, § 2.3-2.5 уже отмечалось, что все вводимые в исходную таблицу SPSS данные в разрезе колонок представляют собой переменные. Любая переменная имеет три основных свойства: **уникальное имя, функциональное назначение** и **числовое значение** (т.е. принадлежит определенному типу чисел).

С точки зрения использования аналитических процедур (технически) уникальное имя переменной служит гарантом корректности и однозначности выполняемых расчетов. Система в этом случае использует довольно жесткие механизмы защиты.

**Правило 17**

***В SPSS каждая переменная имеет уникальное имя. Разработчику полезно знать эти имена, понимпть их условные записи и уметь ими пользоваться.***

Она в принципе не допускает повторения в одном файле имен переменных, заставляя забывчивых пользователей вернуться и исправить допущенную ошибку.

Переменные могут быть интервьюируемые и расчетные.

**Интервьюируемые переменные** всегда имеют своим источником первичную информацию, получаемую непосредственно в ходе опроса.

**Расчетные переменные**, напротив, всегда есть результат использования какой-то процедуры обобщения или расчленения первичной информации. Подобного рода процедура может быть реализована непосредственно в ходе выполнения расчетов в SPSS, и тогда ее результатом будет новая переменная (глава 4, § 4.1-4.3).

Расчетная переменная может быть получена и в ходе визуального контроля и кодировки собранной в поле первичной информации еще до этапа ввода ее в ЭВМ. Хорошим примером расчетной переменной такого рода в наших исследованиях является демографический тип семьи (приложение 4, demtype7). По сложившейся еще с домашинных времен традиции, мы, в целях контроля, делаем это непосредственно в полевой документации (приложение 2, код определения: демографического типа семьи).

При этом в действительности, во-первых, как бы напрасно тратится труд и время в поле, а во-вторых, совершается отступление от исходных принципов ввода и обработки данных. Это отступление выра

жается в том, что мы вводим в массив в качестве кодировочной (первичной) переменной фактически расчетную (вторичную) переменную. Как результат - мы повышаем вероятность ошибок кодировки и увеличиваем объем машинного контроля.

Справедливости ради, следует отметить, что машинный расчет практикуемой нами демографической типологии семьи не может быть выполнен для отдельных типов без фиксации дополнительной первичной информации. Это касается, в первую очередь, способности различить нуклеарную семью и относительно молодых бабушку и дедушку с внуком (внуками), равно как и неполную семью и моложавую бабушку с внуком (внуками). И если взвешивать, какое из двух зол большее: после полевое кодирование или сбор дополнительной информации в поле, которая, как заведомо известно, касается лишь около 10% массива наблюдаемых случаев, то решение вряд ли будет принято в пользу увеличения объемов полевых работ.

С точки зрения существа выполняемых расчетов функциональное назначение переменных играет решающую роль в проверке закладываемых в различные расчеты гипотез, а числовые значения переменных служат основным критерием для определения возможности использования в анализе тех или иных статистических процедур.

По своему функциональному назначению переменные бывают зависимые (dependent) и независимые (independent). В расчетах зависимые переменные, в соответствии с логикой причинно- следственных связей, всегда выступают в виде функции (следствия), а независимые в виде аргумента (причины).

При построении таблиц распределения зависимая переменная всегда должна быть в сказуемом (по колонке, столбцу), а независимая в подлежащем (по строке). В SPSS требуется неукоснительное следование этому простому правилу. Связано это с тем, что при выполнении одних расчетов, например, в случае последовательности команд: **Analyze – Descriptive Statistics - Crosstabs** в соответствующих окнах процедуры Crosstabs необходимо задать переменные по строкам (rows) и столбцам (columns). В то же время в других случаях, например, при выполнении последовательности команд **Analyze - Regression - Linear** в соответствующих окнах процедуры Linear следует задать зависимую (dependent) и одну или несколько (по усмотрению разработчика) независимых (independents) переменных. А при выполнении последовательности команд **Analyze – Compare Means – One-Way ANOVA** появляется задача введения в

соответствующие окна зависимых переменных (Dependent list) и фактора (Factor), т.е. фактически независимой переменной.

Опыт показывает, что выполнение этих в сущности простых требований довольно часто служит камнем преткновения при выполнении расчетов. Здесь предполагается определенный уровень понимания основ статистики, а как раз с этим и связаны главные трудности многих разработчиков в социальных науках.

В свою очередь независимые переменные подразделяются на несколько основных групп: контрольные, промежуточные, предсказывающие. Все эти группировки в известном смысле условны. Тем не менее, эта условность весьма существенна при анализе данных. Другими словами, как бы по умолчанию предполагается, что разработчик знает назначение каждой переменной.

В действительности этому вопросу практически не уделяется внимание ни при формулировке гипотез и операционализации понятий, ни при разработке инструментария исследования. А так как макеты ввода и обработки данных, как правило, не разрабатываются в виде отдельного документа (приложение 4), то на этапе анализа и начинают возникать трудности.

**Правило 18**

***Приступая к расчетам в SPSS, разработчику следует помнить функциональное назначение переменных и быть готовым к принятию решения по этому вопросу в каждом конкретном случае.***

Указанные трудности обусловлены тем, что во многих случаях у разработчика при постановке задачи на выполнение тех или иных расчетов имеется довольно смутное представление о том, что от чего зависит или, образно говоря, где лошадь, а где телега. Единственная возможность преодоления подобных трудностей - предварительная группировка переменных на зависимые, независимые факторы и независимые контрольные.

Наконец, еще одно важное свойство любой переменной - ее числовое значение или тип. По этому основанию все переменные разбиваются на две основные группы: числовые и номинальные (рис. 21).

Примером числовых переменных, т.е. переменных, образованных по количественному признаку, могут служить возраст и доход (непрерывные или интервальные), размер семьи и число детей в семье (пре

рывные - дискретные), оценка здоровья и удовлетворенность работой в баллах (прерывные - порядковые или ранжированные).

Хорошим примером номинальных (атрибутивных) переменных, т.е. переменных, образованных по качественному признаку, является постоянно осуществляемое в социальных исследованиях приписывание числового значения полу (мужской - 1, женский - 2).


**Рис. 21. Типы переменных**

Понимание типа переменной важно для правильного выбора статистических процедур, допускаемых при использовании различных чисел. Связано это с тем, что основная масса статистических показателей может быть использована только в расчетах с количественными переменными, к которым относятся числовые непрерывные и дискретные величины.

В самой системе SPSS от ранних версий к поздним идет эволюция типов выделяемых переменных. В ранних версиях SPSS 6.0 и 7.5 выделялись следующие типы переменных: неупорядоченные категории (unordered categories) или номинальные числа; упорядоченные или порядковые категории (ordered categories); число наблюдений (counts) и непрерывные (measurements), значение которых измеряется в определенных единицах. (8, С. 7). При этом термины количественные переменные (quantitative variables) и натуральные числа (numeric) соотносятся только с двумя последними категориями.

В версии SPSS 11.5 переменные уже имеют несколько иную конфигурацию. В глоссарии этой версии, куда можно попасть, используя

путь **Help-Topics-Glossary**, выделены следующие типы переменных:

**Категориальная (categorical)** – порядковая или номинальная переменная (A variable with a discrete number of values; an ordinal or nominal variable.).

**Дихотомическая (dichotomous)** – переменная. Она имеет два значения. (A term for a variable that has two possible values).

**Фактор (factor)** – это категориальная переменная, которая служит дополнительной переменной в линейной модели. (A categorical variable that has been added to a general linear model).

**Интервальная (interval)** переменная – это количественная переменная с числовыми значениями. (Quantitative variables measured on a numeric scale in which distances between the points on the scale can be compared meaningfully. Interval variables have numeric values, rather than coded values).

**Номинальная (nominal)** переменная. Переменная с приписанными числами, например, пол или занятость. (A variable can be treated as nominal when its values represent categories with no intrinsic ranking).

**Числовая (numeric)** переменная. Переменная, в которой число имеет свой формат - доходы, заработная плата, цены. (A variable whose values are numbers. Numeric variables can be displayed in many different formats).

**Порядковая (ordinal)** переменная. Переменная с упорядоченными значениями, например, оценки здоровья, успеваемости от 1 до 5 или удовлетворенность различными сторонами жизни от 1 до 7 (A variable can be treated as ordinal when its values represent categories with some intrinsic ranking).

**Шкальная (scale)** переменная. Переменная является шкальной, если можно вычислить ее средние значения. Примерами таких переменных могут быть возраст в годах, доход в рублях (A variable can be treated as scale when its values represent ordered categories with a meaningful metric, so that distance comparisons between values are appropriate. Examples of scale variable include age in years and income in thousands of dollars).

В приведенном выше тексте определения типа переменной даны вместе с их описанием в документации SPSS. На это обстоятельство важно обратить внимание, поскольку под такими названиями различные типы переменных даются в окнах всех выполняемых статистических процедур SPSS.

Надо сказать, что эволюция системы еще далека от своего заверше-

ния. Так в версии SPSS 11.5 при расчете одной из кластерных моделей появилась отсылка на продолжающиеся переменные – continuous variables (глава 15, параграф 15.2), определения которых еще нет даже в ее справочнике (Help – Topics - Glossary). Полное описание типов переменных в справочной системе 11.5 версии SPSS можно получить, используя путь: Help-Data Editor-Variable View-Variable Type, а методов их измерения - Help-Data Editor-Variable view-Variable Measurement Level.

**Правило 19**

***Приступая к расчетам в SPSS, зазработчику полезно знать и помнить типы переменных и характер числа каждой переменной.***

В последних версиях SPSS, в некоторых процедурах при выполнении расчетов требуется задание типа переменной. Движение в данном направлении свидетельствует о полноте и последовательности учета разработчиками системы технологии и специфики исследовательской деятельности.

В обрамлении 5 в табличной форме приведены, взятые из наших исследований, примеры уникальных имен, функционального назначения и числовых значений ряда переменных. Эта информация приведена с целью демонстрации и наглядного прояснения технологического единства различных этапов социологического исследования, а именно разработки программы, инструментария, макета ввода данных и их обработки.

Показатели (колонка 1) представляют собой операциональные понятия, пришедшие из программы в инструментарий и полевую документацию исследования (приложение 2).

Уникальные имена переменных (колонка 2) фиксируют представление различных показателей сначала в макете ввода данных, пример его разработки приведен нами в приложении 4, а затем и в файле самого пакета SPSS, в котором будут храниться введенные данные.

Функциональное назначение переменных (колонка 3) позволяет определить роль каждой переменной при выполнении статистических расчетов с помощью процедур, предлагаемых в SPSS.

Для целей анализа важным, как уже отмечалось ранее, является разбиение переменных на две основные группы (зависимые и независи-

мые) и две дополнительные группы (контрольные и промежуточные).

**Обрамление 5. Основные переменные панельного исследования и их роль в анализе данных (опросный лист 1997 г.)**

| Показатели | Имя переменной | Назначение переменной | Характер числа/ шкалы | Возможные вычислительные процедуры в SPSS |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 |
| Село | village7 | Контрольная | Номинальное | Frequencies, Crosstabs, Explore |
| Размер семьи | numfam7 | Независимая | Интервальное | Frequencies, Crosstabs, Explore, Compare Means, Correlate, Regression |
| Дем. тип семьи | demtype7 | Независимая | Номинальное | Frequencies, Crosstabs, Explore |
| Пол | sexresp7 | Независимая | Номинальное | Frequencies, Crosstabs, Explore |
| Возраст | ageresp7 | Независимая | Интервальное | Frequencies, Crosstabs, Explore, Compare Means, Correlate, General Linear Model, Regression |
| Наличие собственного дела | busines7 | Зависимая | Номинальное | Frequencies, Crosstabs, Explore |
| Наличие птицы | poultry7 | Зависимая | Интервальное | Frequencies, Crosstabs, Explore, Compare Means, Correlate, General Linear Model, |
| Денежный доход семьи | total7 | Зависимая | Числовое | Frequencies, Crosstabs, Explore, Compare Means, Correlate, General Linear Model, Regression |
| Удовл. здоровьем | healsat7 | Зависимая | Порядковое | Frequencies, Crosstabs, Explore, Compare Means, Correlate |
| Чувство депрессии | depress7 | Промежуточная | Порядковое | Frequencies, Crosstabs, Explore, Compare Means, Correlate |

Характер числа (колонка 4) задает тип переменной. По этому основанию переменные делятся на:

- **номинальные** (nominal);
- **порядковые** (ordinal);
- **интервальные** (interval);
- **числовые** (numeric).

Наконец, последняя (пятая) колонка содержит перечень возможных вычислительных процедур в SPSS для каждой переменой. Из клеток этой колонки видно, что только дискретные и непрерывные переменные позволяют выполнять широкий набор различных статистических процедур, предлагаемых SPSS. Два других типа переменных (порядковые и номинальные) имеют в этом плане весьма жесткие ограничения.

В сформулированных выше соображениях по умолчанию принимается положение о том, что в основе рядов распределений, которые стоят за приведенными в таблице переменными, лежит нормальное распределение. В противном случае все рассуждения подобного рода не имеют смысла. Связано это с тем, что возможности использования тех или иных процедур регламентируются не только типом переменной, но и характером ее распределения. Последний момент требует специального внимания и рассмотрения. В той или иной степени он освещен практически во всех пособиях по статистике и использованию математических методов в социологии.

Визуально нормальность распределения лучше всего видна на графиках, в которых задаются средние значения и стандартное отклонение. В нормальном распределении средние значения (средняя, мода и медиана) совпадают, а плотность распределения симметрична относительно среднего (условно говоря, разброс случаев близок к однородности, без больших выбросов).

**Правило 20**

***Тип переменной (характер ее числа) выполняет роль ограничителя при выборе для расчетов различных статистических процедур, равно как и задания соответствующих команд доя их выполнения.***

## 6.3. SPSS и методы математической статистики и социологии

В нашей стране имеется богатый опыт разработки проблемы использования методов математической статистики в социологии. Эта традиция, идущая со времен ручных расчетов, использования перфорационных машин и мощных ЭВМ, ставила своей задачей способствовать повышению глубины и достоверности социологического анализа путем распространения основ математической статистики.

В многочисленных работах математиков и статистиков от социологии, выполнявших миссию повышения математической культуры социологов-гуманитариев, в качестве основного инструмента культурного регентства практически всегда выступала математическая формула расчета того или иного показателя и ее краткое вербальное описание (28). При этом действия социолога-аналитика направлялись следующим образом: измеряемый признак - шкала измерения - допустимая статистика.

Указанная последовательность событий регламентировала весь ход анализа. В обрамлении 6 в табличной форме приведен, получивший широкое распространение, пример такого подхода.

Вполне возможно, что такой подход был уместен и дидактически выверен. А так как сама система расчетов была исключительно трудоемкой, требующей много сил и времени, то до их выполнения дело доходило довольно редко. Как правило, анализ ограничивался сопоставлением абсолютных и относительных величин, а также сравнением средних значений различных признаков. Здесь вряд ли необходимо делать ссылки, так как потребуется приводить обширную библиографию работ такого плана, опубликованных в 1970-1999 гг.

К сожалению, по мере распространения компьютеризации указанный подход претерпел очень слабые изменения. А между тем ситуация радикально изменилась. Сегодня, используя такие специализированные пакеты прикладных программ как SPSS и другие близкие ему по замыслу интеллектуальные продукты, куда легче выполнить комплекс расчетов, чем написать ту или иную формулу, лежащую в их основе или описать шкалу измерения. При этом работа в SPSS задает и требует несколько иного хода рассуждений. Можно сказать, что в случае его использования, действия социолога-аналитика теперь

уже регламентируются следующим образом: переменная - допустимая статистика - выполнение процедуры.

**Обрамление 6. Уровни измерения и их характеристики***

| Шкала | Описание шкалы | Отношения, задаваемые на шкале | Примеры допустимой статистики |
|---|---|---|---|
| Наименований | Использование чисел или символов только для классификации объектов | 1.Эквивалентность | Частота (частость); мода; энтропия H; меры взаимозависимости: Q, Ф, r-Пирсона, T-Чупрова, K-Крамера |
| Порядковая | Иерархическая соподчиненность объектов одного класса с объектами других классов | 1.Эквивалентность<br>2.”Больше, чем” | Медиана; квантили; меры взаимозависимости: r -Спирмена; т-Кендалла |
| Интервальная | Знание расстояния между двумя любыми числами на шкале (в дополнение к порядковой шкале) | 1.Эквивалентность<br>2. ”Больше, чем”<br>3.Знание отношений между любыми двумя интервалами | Средние арифметические; дисперсии; меры взаимозависимости: r-Пирсона; R-множественный коэффициент корреляции; все известные операции с натуральными числами |
| Отношений | Независимость отношения любых двух точек шкалы от единицы измерения (интервальная шкала плюс истинная нулевая точка) | 1.Эквивалентность<br>2. ”Больше, чем”<br>3.Знание отношений между любыми двумя интервалами<br>4. Знание отношений между любыми двумя шкальными значениями | То же |

*Источник:* Статистические методы анализа информации в социологических исследованиях. Отв. ред. Г.В.Осипов. -М.: Наука, 1979, С.22; Рабочая книга социолога.- М.: Наука, 1977, С. 171.

Сопоставление содержания табличных форм 5-6 со всей очевидностью свидетельствует о характере и глубине различий двух описанных выше подходов, равно как и о направлении происходящих перемен. Сугубо логические мыслительные действия, связанные с признаками и шкалами измерения, существенно отличаются от реальных операций с переменными, которые осуществляются путем выполнения процедур (команд) статистических расчетов, т.е. фактически по прототипам.

Можно сказать, что новые интеллектуальные продукты позволяют освободить социологов-аналитиков от необходимости освоения огромного пласта знаний, связанных с особенностями и спецификой статистических расчетов. Разработчики подобных продуктов как бы предлагают взамен воспользоваться их знаниями в этой области и не тратить зря силы и время. Проблема состоит в том, что этим, в полном смысле слова благим даром, надо уметь воспользоваться. И здесь социологи-математики вполне могут продолжить выполнение своей миссии. Но делать это уже нужно на несколько иной основе.

В новых условиях основная часть работ, связанных с использованием математических методов в социологии, во-первых, должна быть привязана к процедурам интеллектуальных продуктов, предназначенных для выполнения статистических расчетов, а во-вторых, стремиться к качественному (вербальному и рефлексивному) описанию особенностей выполняемых расчетов.

Имеются веские основания утверждать, что при сохранении существующего положения дел эта мощная традиция потеряет свою культуррегентскую функцию и начнет работать, как это обычно бывает в подобных случаях, в первую очередь на саму себя. Тенденции движения в данном направлении можно наблюдать повсеместно, но лучше всего они проявляются в ограниченности разработок по использованию пакетов прикладных программ (таких как SPSS) в социальных исследованиях.

В подтверждение данного тезиса уместно привести одно из писем, полученных нами после публикации первой части предыдущей работы:

«Уважаемый Валерий Валентинович.

Вас беспокоит аспирант Московского НИИ глазных болезней им. Гельмгольца. С большим интересом прочитал первую часть книги «Использование SPSS в социологии». Сам я, конечно же, не социолог, а врач. Хочу использовать (и использую) SPSS в своей диссертации

для обработки данных. Я располагаю 9-ой версией пакета. Если первой частью ввода и контроля данных я достаточно хорошо овладел, то непосредственно со статистической обработкой возникают существенные проблемы, заключающиеся в том, что я практически не знаю, для чего необходим тот или иной вид анализа данных, например: T-тест, тест ANOVA, различные виды корреляций (бивариантная, частичная, дистантная), регрессионный анализ, LOG-линейный анализ, кластерный, дискриминантный анализ, шкалирование, непараметрические тесты. То есть, если о том, для чего они нужны, хотя бы возможно где-то прочитать, то интерпретировать появляющиеся на экране данные просто невозможно.

Кроме того, мне необходимо проводить корреляционный анализ между не только количественными или между качественными признаками, но и между качественными и количественными признаками. В пакете раздел корреляций не позволяет мне отслеживать такого рода связи. Может быть, для анализа взаимосвязи необходимо использовать другой вид анализа?

Уважаемый Валерий Валентинович! Не могли бы Вы по возможности рекомендовать какую-либо литературу по данным методам анализа, применительно к пакету SPSS, ориентированную на достаточно неподготовленного пользователя. Может быть, во второй части вашей книги вы даете какие-либо объяснения по расшифровке получающихся данных. Возможно ли где-нибудь ознакомиться со второй частью вашей книги?

Заранее благодарю,

С уважением, Ю. В. Мой e-mail: ...» (3, С. 22-23).

**Лозунг дня**

***Математики-социолги, обратите внимаие на потребности социологов-аналитиков в освоении пакетов прикладных программ типа SPSS!***

## Задание для самостоятельной работы

1. Как цели и задачи выборочного исследования связаны со сбором данных и их анализом?
2. В чем различие показателя, индикатора и переменной?
3. Какие типы переменных вы знаете?

4. Чем расчетная переменная отличается от интервьюируемой?
5. Как соотносятся расчетные данные и первичная информация?
6. В чем различие зависимых и независимых переменных?
7. Приведите примеры различного типа переменных.
8. В чем различие числовых и номинальных переменных?
9. Какие статистические процедуры допускается использовать при работе с переменными, содержащими номинальные числа?
10. Какие статистические процедуры допускается использовать при работе с переменными, содержащими натуральные числа?
11. Какие статистические процедуры допускается использовать при работе с переменными, содержащими порядковые числа?
12. С точки зрения работы в SPSS, в чем отличие количественной переменной и натурального числа?
13. Сформулируйте различие двух следующих подходов к анализу данных:

а) переменная - допустимая статистика - выполнение процедуры;

б) измеряемый признак - шкала измерения - допустимая статистика.

14. Как в SPSS можно найти определения типов переменных?
15. Как использование компьютерных систем обработки данных связано программой социологического исследования?
16. Как использование компьютерных систем обработки данных связано инструментарием (полевой документацией) социологического исследования?
17. Как использование компьютерных систем обработки данных связано с гипотезами, формулируемыми в программе социологического исследования и последующем анализе первичной информации?
18. Что такое «операционализация понятий» в социологическом исследовании?
19. Какие особенности файлов с данными социологических исследований полезно помнить, приступая к их анализу с помощью пакетов прикладных программ?
20. Какие основные свойства переменных необходимо принимать во внимание при анализе данных в SPSS?
21. Зачем разработчику нужно знать имена переменных?
22. Что такое функциональное назначение переменной?
23. Приведите пример категориальной переменной.
24. Приведите пример дихотомической переменной.
25. Какие правила сформулированы в этой главе?

# Глава 7. Аналитические возможности SPSS

В SPSS анализ и изучение данных, содержащихся в открытом на каждый данный момент файле, можно вести, прежде всего, посредством использования двух основных сервисных возможностей главного меню:

· статистических процедур (раздел главного меню **Analyze**, который в ранних версиях назывался - Statistics);

· графиков (раздел главного меню **Graphs**).

Кроме того, каждый из указанных типов анализа может быть выполнен с помощью использования специального командного языка SPSS - синтаксиса (раздел главного меню **File - New - Syntax,** а в ранних версиях -SPSS Syntax).

При этом результаты статистического анализа в любом случае будут представлены в окне просмотра данных **Output – SPSS Viewer** (глава 1, § 1.6). Ниже, в текущей и последующих главах 8-11, описаны различные возможности использования статистических процедур. Порядок выполнения работ в разделе главного меню Graphs изложен в главе 12, а описание командного языка Syntax дано в главе 17.

## 7.1. Основные сведения о статистических процедурах в SPSS

SPSS представляет возможности использования ряда статистических процедур для анализа социологической информации. В различных версиях SPSS, от базовой 6.1 до 8.0, раздел меню «Statistics» (статистика) содержит список категорий статистических методов. Каждая из них заканчивается знаком стрелки, указывающей, что существует еще один уровень - подменю, в котором и перечислены конкретные статистические процедуры.

Начиная с версии SPSS 9.0, раздел главного меню «Statistics» преобразован в «**Analyze**». При этом в нем существенно расширен выбор базовых статистических процедур, но порядок и принципы их

использования остались неизменными, как и во всех предшествующих версиях.

Ниже приведен перечень процедур меню Statistics с кратким описанием их содержания в получившей наиболее широкое распространение в нашей стране базовой версии SPSS 6.1. Более полно все процедуры будут описаны в отдельных параграфах.

Процедура **Frequencies** (частоты) является средством детального описания данных. С этой процедуры начинается первичный анализ социологической информации. Полученные первичные распределения дают представление о частоте встречаемости (в абсолютном и относительном выражении) анализируемых переменных. Таблицы частот пригодны для суммирования и отражения данных.

Процедура **Crosstabs** (таблицы сопряженности). Она позволяет вскрыть сопряженность переменных. Полученные двумерные таблицы показывают частоту встречаемости одной переменной в зависимости от другой.

Процедура **Descriptives** (дескриптивные или описательные статистики). дает описание средних, квадратичного отклонения, дисперсии и др. статистик для нормального распределения, а также минимальное значение, размах и сумму для ассиметричного распределения с количественной переменной.

Процедура **Explore** (исследовать). Дает возможность описания подмножеств наблюдений с помощью разнообразных статистик (подсчет частот и процентов, средних и др.) и графиков.

Процедура **Means** (средние). Позволяет делать расчет большого числа статистик, таких как средняя, стандартное отклонение, дисперсия и др.

Процедуры **One-Sample T Test** (одновыборочный t-критерий), **Independent-Samples T Test** (t - критерий для независимых выборок), **Paired-Samples T Test** (t -критерий для парных выборок), **One-Way ANOVA** (однофакторный дисперсионный анализ) - все они предназначены для проверки различных гипотез о средних значениях количественных переменных.

Процедура **Correlate** (корреляция). Она позволяет установить меру линейной связи между двумя количественными переменными.

Процедура **Regression** (регрессия). Показывает зависимость среднего значения результативного признака (зависимой переменной) от одного или нескольких факторов (независимых переменных).

Приведенное выше описание меню Statistics отражает его содержа-

ние для версии SPSS 6.1. Последующие версии SPSS 7.5 и 11.5 предлагают значительно больший выбор процедур. Среди них, теперь уже в меню **Analyze**: **Classify** (операции с порядковыми переменными), **Cluster** (кластерный анализ), **Discriminant** (дискриминантный анализ), **Data Reduction** (редукция данных), **Factor** (факторный анализ), **Correspondence Analysis** (анализ соответствий), **Time Series** (динамические ряды) и др.

Одни из этих процедур используются в социологии (например, регрессионный и факторный анализ), тогда как другие еще ждут своего освоения. В любом случае, как мы уже отмечали ранее, в SPSS, в отличие от известного текстового процессора MS Word, все версии сохраняют преемственность и сопоставимость в обе стороны как от верхней к нижней, так и от нижней к верхней.

Различия идут в основном по линии затрат и наращивания сервисных возможностей. Поэтому здесь главное овладеть основными принципами, которые едины для всех версий, но более доступны для освоения в процессе работы с базовой версией.

**Напоминание**

***В SPSS каждая новая версия требует больше затрат памяти машины и других ресурсов, чем открывает новых возможностей.***

## 7.2. Порядок выполнения статистических процедур

В SPSS статистический анализ начинается с открытия файла данных. Порядок выбора и открытия файла данных для целей анализа описан подробно в главе 3 (§ 3.2). Кратко, он состоит из выполнения нескольких команд. В главном меню выбирается последовательность команд: File - Open - Data. В результате раскрывается диалоговое окно Open Data File (открыть файл данных). Далее выделяется имя нужного файла в списке имен файлов и делается щелчок мышью на OK или двойной щелчок мышью на имени открываемого файла.

Содержимое файла появится в редакторе данных. По умолчанию предполагается, что первичная информация, представленная в редакторе данных, корректно введена и отконтролирована (глава 5, § 5.1-5.3). В противном случае массив начинает рассыпаться при первой же

попытке анализа с использованием статистических процедур.

Доступ к статистическим процедурам закрыт, если в редакторе (рабочем файле) отсутствуют данные или в окнах статистических процедур не выполнены операции, связанные с формированием списка переменных, подлежащих расчету. Логика здесь простая, но железная: «**Не ходи в систему расчетов, не имея данных**». Она, кстати сказать, радикально отличается от господствовавшей продолжительное время логики многих разработок по теории измерения, социальному и экономико-математическому моделированию, которые весьма широко практиковали свои формально-логические построения в отсутствии исходных данных.

Следующий шаг - выбор статистической процедуры. Для выбора определенной статистической процедуры необходимо обратиться к разделу главного меню - статистики Analyze, который содержит список общих категорий статистических методов.

Каждая из категорий этого меню, как уже отмечалось ранее, заканчивается значком стрелки, указывающей на существование следующего уровня, в котором перечислены статистические процедуры. Так, для того, чтобы в версии SPSS 9.0 и выше получить частотные таблицы (статистическая процедура Frequencies), следует выбрать в главном меню Analyze, а внутри этого меню открыть подменю Descriptive Statistics (сводка). И только в последнем выпадающем меню появляется доступ непосредственно к процедуре Frequencies. Таким образом, последовательность выполнения команд имеет следующий вид: **Analyze - Descriptive Statistics – Frequencies**.

В ранних версиях указанная последовательность действий имела следующий вид: **Statistics - Summarize - Frequencies.**

Для получения таблиц сопряженности (статистическая процедура Crosstabs) в последних версиях необходимо выполнить последовательность команд: **Analyze - Descriptive Statistics – Crosstabs,** а в ранних - **Statistics - Summarize - Crosstabs.**

После появления на экране главного диалогового окна конкретной процедуры, предстоит выбрать и установить в соответствующие поля переменные, которые предполагается включить в анализ, а также, в случае необходимости, ввести требуемые параметры (опции) в дополнительные диалоговые окна.

После нажатия кнопки OK в главном диалоговом окне выполняемой статистической процедуры, результаты расчетов появятся в окне просмотра. В последних версиях системы в верхней

строке этого окна будет написано: «Output1 SPSS Viewer». Скорость открытия окна просмотра зависит от технических параметров используемого компьютера, объема памяти, которая съедается файлом, открытом в редакторе данных, а также комплексом заданных расчетов.

Наличие редактора данных, в котором содержится первичная информация, и специального окна просмотра, в котором даются результаты расчетов, выполняемых с помощью различных статистических процедур, - характерная особенность SPSS. Она заметно отличает его как от электронных таблиц типа Excel, так и других программных продуктов, ориентированных на ввод, хранение и обработку статистической информации.

**Напоминания**

***Выбор переменных обусловлен целями, задачами и проверяемыми гипотезами, а также ранее сформулированным в предыдущей главе правилом 18.***

***Выбор дополнительных параметров из диалоговых окон обычно обусловлен ранее сформулированным в предыдущей главе правилом 19.***

***Выбор статистической процедуры обусловлен целями, задачасми и проверяемыми гипотезами, а также ранее сформулированным в предыдущей главе правилом 20.***

**Обобщение**

***Статистический анализ в SPSS начинается с открытия файла данных. Следующий шаг предполагает выбор в меню конкретной статистической процедуры, открытие ее главного диалогового окна, установку в соответствующие подокна переменных, которые будут включены в анализ, а также задание дополнительных параметров в специальных диалоговых окнах выбранной процедуры.***

Более полно выполнение конкретных статистических процедур (с фиксацией последовательности шагов оперирования в дополнительных диалоговых окнах каждой из них, а также конкретными примерами расчетов) будет описано ниже.

## 7.3. Главные диалоговые окна

После выполнения последовательности команд, открывающих доступ к той или иной статистической процедуре, на экране появляется ее **главное диалоговое окно**. Почти все главные диалоговые окна процедур выглядят одинаково, т.е. примерно так, как показано ранее в главе 1 на рис 3 (главное диалоговое окно процедуры Frequencies), или, как показано ниже, на примере главного диалогового окна процедуры **Means** (рис. 22).

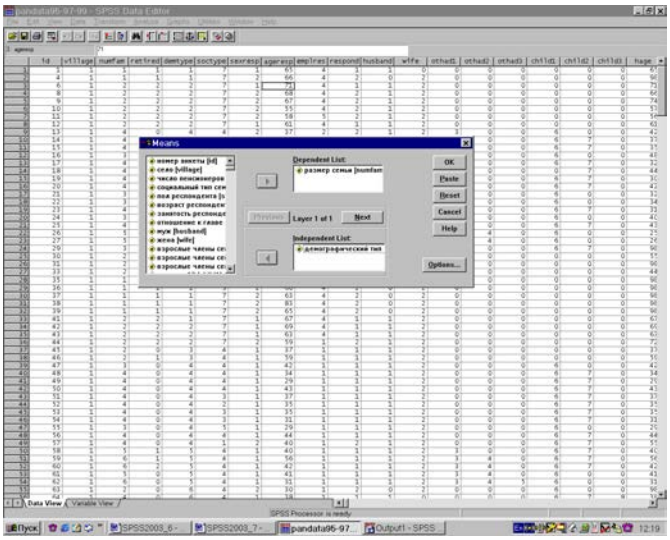



***Рис. 22.***

***Главное диалоговое окно процедуры Means***

Главное диалоговое окно каждой статистической процедуры имеет три основные элемента: **список исходных переменных, список (списки) выбранных переменных и кнопки (выключатели) основных и дополнительных команд.** Ниже детально рассмотрен каждый из этих элементов.

**Список исходных переменных (Source variable list).** Он включает в себя переменные текущего (открытого в редакторе) файла данных. Этот список вытянут по вертикали и расположен в левой части главного диалогового окна статистической процедуры в виде отдельного поля.

В списке переменные могут быть даны в алфавитном порядке или в порядке их ввода в редактор данных. Переключение выполняется с помощью комбинаций команд главного меню **Edit-Options** или панели

инструментов редактора данных. Порядок выполняемых при этом действий описан ранее в главе 2, § 2.8.

Несколько повторяясь, напомним, что при выполнении указанной выше последовательности команд на экране появится диалоговое окно общих настроек с открытой закладкой - **General**. Для определения порядка вывода переменных в списках главных диалоговых окон статистических процедур необходимо в части окна общих настроек**,** которая имеет имя: «Variable Lists» (в ранних версиях ее название было полнее**:** «**Display Order for Variable Lists**») выбрать одну из двух возможностей:

- **Alphabetical** - в алфавитном порядке,
- **File** - в порядке ввода переменных в таблицу.

В результате установки нужной опции и выполнения команды список переменных в главном диалоговом окне статистической процедуры приобретет тот вид, который более удобен для пользователя. После установки SPSS всегда (по умолчанию) действует опция **File**, т.е. в списке переменных последние даются в порядке их ввода в таблицу.

В другой, предложенной там же альтернативе, необходимо выбрать еще одну из двух возможностей:

**Display labels** – показывать описание переменных

**Display names** – показывать только имена переменных.

Наличие описания переменных создает более комфортные условия работы. Как уже отмечалось ранее (глава 2, § 2.4), одна из основных проблем здесь - это ввод данных с порядковым именем переменных, заданных по умолчанию. В этом случае подокно списка переменных будет содержать колонку имен переменных var00001, var00002, var00003 и т.д. Работать с таким окном, в котором несколько десятков, а то и сотен переменных, не только постороннему пользователю, но и самим разработчикам довольно сложно.

Здесь экономия времени при вводе данных на описании их значений и меток начинает выходить «боком», т.е. оборачивается многократным ростом затрат времени, сил и дискомфорта при обработке и анализе данных. В случае отсутствия описания значений и меток, все результаты расчетов в окне просмотра будут даваться с именами переменных, которые заданы по умолчанию, т.е. var00001, var00002, var00003 и др.

Эта проблема становится особенно актуальной, если к анализу подключаются пользователи «со стороны». Иными словами, те, кто не принимал непосредственного участия в разработке инструментария.

Например, новые работники, аспиранты, заказчики, покупатели массива. В таких случаях наличие инструментария (приложение 2) и электронного макета (приложение 4) становится просто необходимым условием расчетов. Иначе массив данных может представлять интерес скорее для криптографа-дешифровщика, чем социолога-аналитика.

**Дружеский совет**

***Если вы не описали значения и метки данных, сделайте это немедленно, а затем приступайте к выполнению статистических процедур.***

**Список (списки) выбранных переменных [Variable(s)].** Это один или более списков переменных, выбранных для анализа. Они вводятся в специальные поля, расположенные, как правило, в центральной части главного диалогового окна статистической процедуры. Например, в главном диалоговом окне процедуры Frequencies (рис. 3) такое поле всего одно, но в него можно внести несколько переменных. А в главном диалоговом окне процедуры Means (рис. 22) таких видимых полей два: верхнее поле - **Dependent List** и нижнее - **Independent List.** В то же время, если в нижнем поле использовать функцию «**Layer**», то выяснится, что можно вводить (использовать) сразу несколько независимых списков. Благодаря этому открывается возможность многомерного анализа. Иными словами, можно получить среднюю заработную плату не только по всему массиву в целом, но и в разрезе мужчин и женщин специалистов, служащих или рабочих и т.п.

Как уже многократно отмечалось ранее, для выбора переменной необходимо выделить ее из списка в левом поле (щелкнув на ней мышью) и нажать на кнопку со стрелкой, указывающей вправо. Она появится в списке выбранных переменных.

Если при этом возникнет задача вернуть ее на место в список исходных переменных, то надо выделить требуемую переменную и опять нажать на стрелку, которая сразу же после выделения переменной примет обратное направление.

Список исходных переменных, поле для переноса анализируемых переменных, а также кнопки со стрелкой для перемещения переменной из одного списка в другой для главного диалогового окна статистических процедур Friquencies и Means хорошо видны на уже многократно упоминавшихся рис. 3 и рис. 22.

**Напоминание**

***Каждому окну списка выбранных переменных соответствует своя стрелка. Два окна — две стрелки.***

В отдельных случаях для целей анализа можно выбрать сразу несколько переменных. Для выбора группы переменных, расположенных в списке последовательно друг за другом, можно использовать технику «держать и тащить» (click-and-drag), которая довольно часто используется при работ с текстовыми файлами в пакете прикладных программ Word.

Другой способ: щелкнуть мышью на первой переменной группы, затем на последней переменной, удерживая при этом нажатой клавишу **Shift**. Наконец, третий способ: держать кнопку **Ctrl** и идти по выбираемым переменным. Сходным образом можно вернуть переменные назад в их исходный список. Только в данном случае работа идет в списке ранее перенесенных переменных и со стрелкой, указывающей влево.

**Кнопки – выключатели основных и дополнительных команд.** Это кнопки, при нажатии на которые выполняется какое-либо действие. Например, с их помощью осуществляется запуск команды, получение справки или переход к дополнительным диалоговым окнам с целью задания требуемых для выполнения необходимых расчетов параметров.

В правой части окна вертикально расположен ряд кнопок, указывающих системе на необходимость что-то сделать, например, запустить процедуру. В главном диалоговом окне каждой статистической процедуры обычно пять таких кнопок. При движении сверху вниз они стоят в следующем порядке:

- Кнопка, задающая начало выполнения процедуры. Традиционно на ней стоит надпись «**OK**».

- Кнопка **Paste** (вставить) - сгенерировать заданную команду в виде командной строки и поместить этот текст в окно **Syntax** (синтаксиса). После этого можно изменить содержимое окна в редакторе главного окна синтаксиса (если необходимо) и запустить процедуру на выполнение расчетов уже из этого окна с помощью команды Run (глава 17, § 17.2).

- Кнопка **Reset** (переустановить) - дает возможность сбросить в главном диалоговом окне все, ранее установленные в ходе выполнения расчетов параметры статистической процедуры, установить значения, принятые по умолчанию, и очистить списки ранее выбранных переменных.

- Кнопка **Cancel** (отменить) - позволяет отменить все изменения, сделанные в главном диалоговом окне данной статистической процедуры при последнем обращении к нему, а также закрыть это окно.

- Кнопка **Help** (справка) - стандартная подсказка по текущему главному диалоговому окну статистической процедуры.

На рис. 3 и рис. 22 все перечисленные выше пять кнопок основных команд главных диалоговых окон статистических процедур Friquencies и Means можно видеть в крайнем правом столбце.

Кнопки (выключатели) дополнительных команд связаны, главным образом, с особенностями выполнения тех или иных статистических процедур. Например, для того чтобы получить частотные статистики, построить частотные диаграммы, модифицировать формат вывода частотных таблиц и др.

На рис. 3 эти три кнопки дополнительных команд для главного диалогового окна процедуры Frequencies видны внизу с надписями **Statistics, Charts, Format**. А вот у процедуры Means (рис. 22) всего лишь одна такая кнопка - **Options**. При нажатии этих кнопок открываются соответствующие дополнительные диалоговые окна, которые и описаны в следующем параграфе.

**Напоминания**

***В течение всего сеанса работы в SPSS все диалоговые окна сохраняют свое содержимое. В ходе одного сеанса работы при каждом новом открытии диалогового окна оно будет содержать те же параметры, что и при предшествующем обращении к нему.***

***Главные диалоговые окна статистических процедур отражают общий подход, который заложен в SPSS, к выполнению расчетов. Поэтому их устройство имеет общие черты для всех статистических процедур.***

## 7.4. Дополнительные диалоговые окна

Главное диалоговое окно статистической процедуры обычно содержит минимум информации, необходимой для формулировки общего задания на выполнения расчетов. Другие параметры статистических расчетов задаются в дополнительных диалоговых окнах и опциях.

Все кнопки главного диалогового окна, которые имеют многоточие после названия, служат для вызова дополнительных окон. Например, в окне **Frequencies** (рис. 3), как отмечалось ранее, три таких кнопки: **Statistics...** (открывает дополнительное диалоговое окно расчета частотных статистик), **Charts...** (открывает дополнительное диалоговое окно построения частотных диаграмм) и **Format...** (открывает дополнительное диалоговое окно форматирования частотной таблицы).

При нажатии кнопки **Statistics...** (статистики) откроется дополнительное диалоговое окно «частотные статистики», которое можно увидеть на рис. 23.

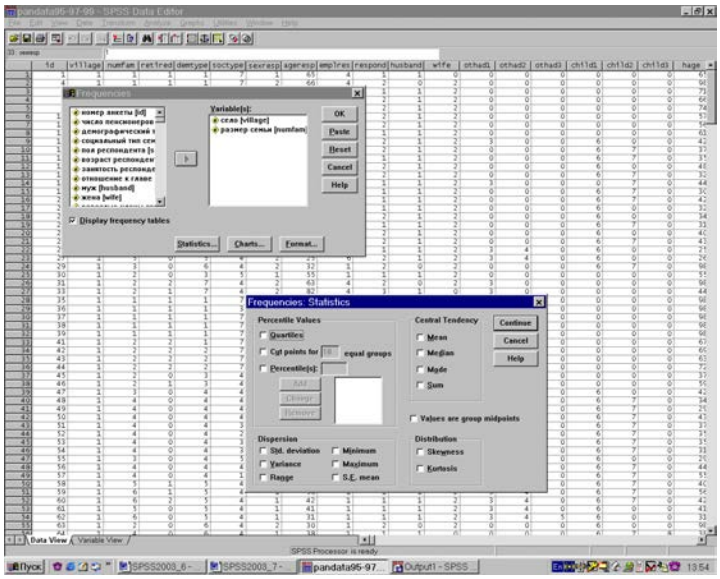



***Рис. 23.***

***Дополнительное диалоговое окно процедуры Frequencies: Statistics -частотные статистики***

Благодаря дополнительным диалоговым окнам открываются большие возможности выполнения статистических расчетов с учетом **типа переменной**. Так, например, для получения статистик, вычисляемых для числовых переменных, можно выбрать в левой верхней части

дополнительного диалогового окна **Percentile Values** (значения процентилей) одно из следующих заданий: **Quartiles** (квартили) - представляют 25, 50 и 75-й процентили; **Cut points for n equal groups** (пороговые значения для n равных групп) - представляют значения процентилей, разделяющих выборку на равные по количеству группы наблюдений (по умолчанию число групп равно 10). При этом можно задать любое число групп от 2 до 100; **Percentile(s)** - значения процентилей, задаваемые пользователем.

Левый нижний блок окна **Dispersion** (дисперсия) позволяет выбрать одну или несколько из следующих возможностей: **Std. deviation** (стандартное отклонение) - показатель отличия наблюдения от среднего значения; **Variance** (характеристики разброса) - показатель отличия наблюдения от среднего, равный квадрату стандартного отклонения; **Range** (размах) - разность между наибольшим и наименьшим значениями наблюдения; **Minimum** (минимум) - наименьшее значение наблюдения; **Maximum** (максимум) - наибольшее значение наблюдения**; S.E. mean** (стандартная ошибка среднего) - показатель изменчивости среднего значения выборки.

В верхней части дополнительного диалогового окна имеется блок **Central Tendency** (центральная тенденция - числа, задающие центральные показатели частотного распределения). В ней можно дань задание вычислить: **Mean** арифметическое среднее; **Median** (медиана) - значение, выше/ниже которого попадает половина наблюдений; **Mode (**мода) - наиболее часто встречающееся значение; **Sum (**сумма) - сумма всех значений.

В нижней правой части дополнительного диалогового окна находится блок **Distribution** (характеристики распределения). В нем можно выбрать одно из двух заданий **Skewness** (асимметрия) - показатель степени несимметричности распределения и **Kurtosis** (эксцесс) - показатель степени концентрации наблюдений вокруг центральной точки. Более полное формальное и содержательное описание указанных выше статистических показателей имеется во множестве публикаций, среди которых, как одну из наиболее доступных, можно выделить (30, С. 31-44).

В правой части дополнительного диалогового окна находятся три кнопки: **Continue** (продолжить), **Cancel** (отменить), **Help** (справка). При этом две последние имеют здесь те же функции, что и в главном диалоговом окне статистической процедуры. Нажатие верхней кнопки **Continue** означает завершение работы в дополнительном диалоговом

окне и ведет к возврату в главное диалоговое окно статистической процедуры, в котором, для выполнения расчетов по всему комплексу заданных параметров, необходимо нажать кнопку OK.

При необходимости внесения каких-то дополнений в расчеты перед нажатием кнопки OK еще есть возможность вернуться в то или иное дополнительное окно и внести соответствующие изменения. В случае же нажатия кнопки OK к корректировке расчетов также можно вернуться, но только после выполнения расчетов, а еще лучше после осмысления их результатов, которые, как отмечалось ранее, даются в специальном окне просмотра**.** Сделать это нужно будет уже в режиме повторения задания путем нового обращения к разделу главного меню **Analyze** и открытию главного диалогового окна интересующей статистической процедуры.

Выше был рассмотрен общий случай возможности получения частотных статистик в одном из трех дополнительных диалоговых окон **Statistics...** (статистики) главного диалогового окна статистической процедуры **Frequencies**.

Нажав вторую из трех кнопок **Charts...** (диаграммы), мы как бы заявляем о своем желании построения столбиковых диаграмм и гистограмм для ранее выбранной переменной (переменных). Это значит, что при нажатии кнопки **Charts** откроется диалоговое окно «частотные диаграммы», которое можно увидеть на рис. 24.

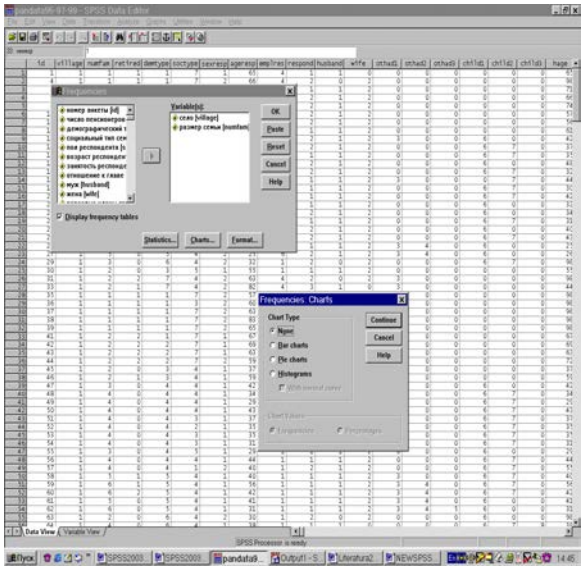



***Рис. 24.***

***Дополнительное диалоговое окно Charts - частотные диаграммы процедуры Frequencies***

Для выбора типа диаграммы в блоке **Chart Type** необходимо установить одну из опций: **None** (никаких) – эта опция установлена системой по умолчанию, **Bar charts** (столбиковые диаграммы), **Histograms** (гистограммы) – это опция доступна только при использовании числовых переменных.

Доступ к работе с опцией **With normal curve** (график с нормальной кривой) открывается лишь при построении гистограммы. Указанная опция накладывает на гистограмму кривую нормального распределения. Пример такого наложения можно увидеть на графике 4(глава 12).

Для столбиковых диаграмм имеется возможность установить разметку по вертикальным осям. Здесь по умолчанию задается опция **Frequencies** (частоты), а опция **Percentages** (проценты) требует дополнительной установки в поле **Axis Label Display** (разметка осей).

При нажатии кнопки **Format...** в окне **Frequencies** откроется дополнительное диалоговое окно «Формат частотной таблицы», которое можно увидеть на рис. 25.

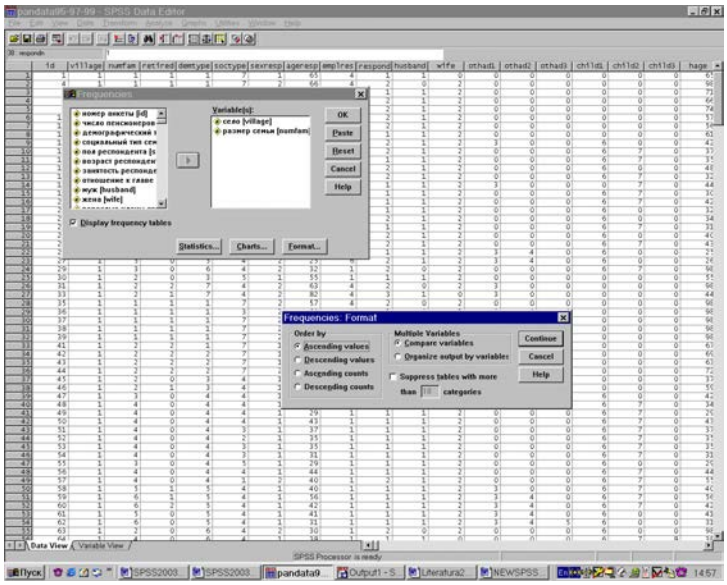



***Рис. 25.***

***Дополнительное диалоговое окно Format - формат частотной таблицы процедуры Frequencies***

Используя опции этого окна можно выбрать порядок, в котором данные будут представлены в частотной таблице (по возрастанию значений **Ascending values**, по убыванию значений **Descending values**, по возрастанию количества **Ascending counts**, по убыванию количества **Descending counts**), а также можно выбрать формат страницы **Page Format** (стандартный, сжатый и т.п.).

В ходе анализа довольно часто возникает необходимость расположить данные в порядке возрастания или убывания их значений. Ее решение возможно как с использованием процедуры **Frequencies**, так и другим, несомненно, более эффективным путем с использованием процедуры Sort Cases (глава 3, § 3.7).

**Напоминание**

***Дополнительные диалоговые окна статистических процедур фиксируют специфику их выполнения, которая заложена в логику статистических расчетов. Поэтому они имеют существенные особенности и для каждой статистической процедуры.***

## 7.5. Особенности работы с окном просмотра

Окно просмотра-вывода **(Output1 - SPSS Viewer)** открывается после выполнения статистической процедуры (глава 1, § 1.6).

В окне просмотра результаты могут быть даны в виде текста, таблиц сопряженности, корреляционных матриц, графиков и т.п. В SPSS 11.5 формат представления результатов в окне просмотра может быть задан путем выполнения в главном меню последовательности команд: **Edit - Options –** закладка **Data.** Далее, требуется установить необходимые опции в оконтуренных полях (обрамлениях): **Text Output Page Size** и **Text Output Font.** Для описания меток используется путь: **Edit - Options –** закладка **Output Labels**, в которой устанавливаются соответствующие опции.

В ранних версиях системы, формат представления результатов в окне вывода устанавливался путем выполнения в главном меню последовательности команд: Edit – Preferences – закладка Output.

Содержимое окна просмотра можно редактировать и сохранять в файле для последующего использования. Вместе с тем опыт показывает, что форма подачи результатов в окне просмотра, возможно, является одним из наиболее слабых мест (с точки зрения дружественного отношения к пользователю) в рассматриваемом пакете прикладных программ.

Здесь можно указать на целый букет проблем: от заголовков таблиц и имен переменных до их описания. Справедливости ради, следует сказать, что подобное положение дел в большей мере свойственно

ранним версиям SPSS. В них возможности изменения характеристик окна вывода весьма ограничены. Идя по пути **Edit - Preferences - Output,** можно изменить, прежде всего, установки страницы (ширину и высоту), а также границы таблиц.

В версии 7.5 и выше, идя по пути **Edit - Options - Paviot Tables**, можно изменить формат вывода таблицы. По умолчанию здесь, как и в нижних версиях, дается формат «System Defoult». Вместе с тем при желании теперь уже можно выбрать и другие, более изящные форматы таблиц, например, «Academic.tlo», «Hoddog.tlo», «Modern.tlo» и др. Необходимо только понять, что вы с этим собираетесь делать.

В последних версиях SPSS имеется возможность редактирования отдельных составляющих окна вывода (заголовков таблиц, легенд и т.д.). Для этого полезно фиксировать связь между красной стрелочкой слева в основной части окна просмотра и выделенным текстом в левом поле этого же окна. Красная стрелка стоит всегда у выделенной части окна просмотра. Двойной клик мышью обрамляет ее, свидетельствуя об ее готовности к редактированию. Для открытия доступа к инструментам редактирования в главном меню следует выполнить следующую последовательность команд:

**View**

**Toolbar**.

При этом в окне просмотра появляется небольшая панель инструментов, позволяющая менять шрифты, их размеры, выравнивать или центрировать текст и т.п.

Здесь есть два важных момента. Первый из них связан с тем, что получение в окне просмотра более изящной таблицы требует дополнительных ресурсов системы, которые, как правило, ограничены (особенно в части оперативной памяти). Второй - связан с тем, что в социологии результаты расчетов, таблица или график редко бывают конечным продуктом. Скорее всего, они часть текста (научного доклада, отчета или публикации). А вот для целей подготовки таких документов рассматриваемый специализированный программный продукт и не предназначен. И это хорошо, так как в противном случае у многих из нас не хватило бы на наших ПК ресурсов для его установки.

Кстати сказать, программные продукты, интегрирующие электронные таблицы и текстовые процессоры - это отработанное направление компьютеризации. Его примерами могут служить такие еще совсем недавно широко известные системы, как Lotus, Framework и др.

Поэтому лучшая форма работы с результатами расчетов, которые даются в окне просмотра SPSS, - их перенос в текстовый процессор Word, в котором, собственно, и готовятся все отчеты и выходные тексты, таблицы и графики.

В SPSS можно также открывать новые и ранее уже сохраненные дополнительные окна просмотра. Для того, чтобы открыть окно просмотра нового текстового файла, следует выполнить в главном меню следующую хорошо известную и постоянно повторяемую последовательность команд: File - New - Output.

Выполнение этих команд откроет новое окно просмотра. Для того, чтобы открыть окно просмотра с ранее сохраненным текстовым файлом, следует выполнить в главном меню уже описанную в ранее (глава 1, § 1.6) последовательность команд: File – Open - Output.

Откроется диалоговое окно «открыть файл вывода» (**Open Output**). Оно и показано на рис. 26.

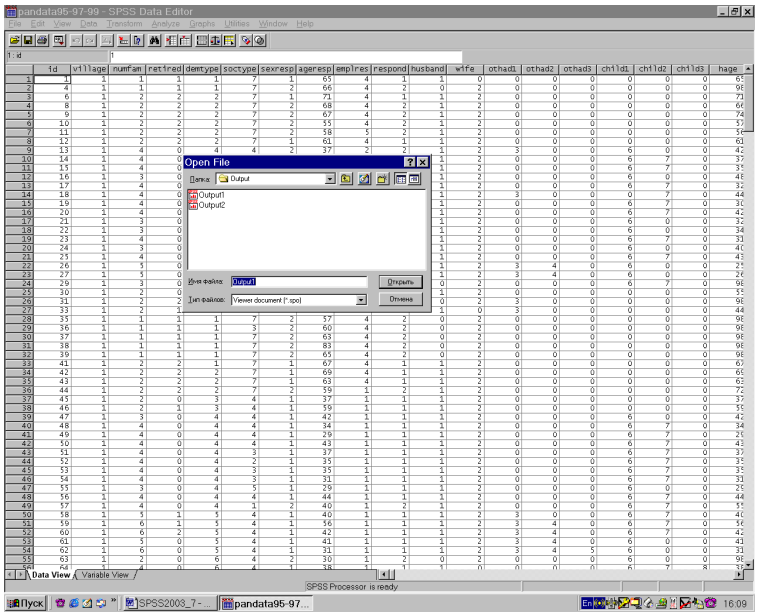



***Рис. 26.***

***Диалоговое окно Open Output - открыть файл окна вывода***

В открывшемся окне со списком файлов следует выделить требуемый файл и нажать на OK. Откроется окно просмотра, содержащее текстовый файл. Файл может быть файлом вывода или файлом синтаксиса, сохраненным ранее, либо файлом, созданным в другом приложении и сохраненным в текстовом формате. Окно открытия файла содержит стандартные элементы такого рода окна.

По умолчанию SPSS производит поиск по всем файлам в текущей директории с расширением **.spo** (**Viewer document).** Это термин, которым сегодня принято обозначать файлы, содержащие текстовое описание результатов выполнения расчетов и сеансов работы в SPSS. В ранних версиях для этой цели использовались файлы с расширением **.lst** (**Listing file -** файл листинга).

Если открыто несколько окон просмотра, SPSS, как мы уже отмечали ранее, по умолчанию посылает результаты в то окно, которое открывается автоматически в начале сеанса, т.е. в **Output1,** и именно оно получило название: **окно назначения вывода**. Иными словами, несмотря на то, что можно открывать множество окон просмотра, в течение любого сеанса работы допустимо существование лишь одного окна назначения вывода.

В то же время окно назначения вывода можно переназначить. Для этого перед началом расчетов следует открыть (сделать активным) одно из ранее сохраненных окон вывода. Именно в нем, вплоть до специальных указаний, система и будет выдавать все результаты расчетов, выполняемые в ходе текущего сеанса работы. Все новые результаты расчетов будут добавляться в конец текстового файла, находящегося в назначенном окне вывода. Окно остается назначенным окном вывода до тех пор, пока не выберется другое.

Изменения в файлах окна просмотра не сохраняются до тех пор, пока не дать команду их сохранить. Для сохранения изменений в существующем текстовом файле следует сделать окно просмотра, содержащее файл, активным, и выбрать в меню: **File - Save.** Модифицированный файл будет сохранен с прежним именем. Причем, как это принято повсеместно в таких случаях, он запишется поверх прежней версии файла и сделает ее уже недоступной для последующего использования.

Для того, чтобы сохранить текущие результаты из окна просмотра в новом текстовом файле или в файле другого формата, следует сделать это окно просмотра активным и выполнить последовательность команд: **File - Save As.** В результате выполнения этих команд откроется диалоговое окно **Save As ...** (сохранить просмотр как). Для сохранения текущих результатов необходимо, в соответствии с общими правилами записи файлов, ввести имя вновь сохраняемого файла и щелкнуть на кнопку OK.

**Дружеский совет**

***С окном просмотра лучше не экспериментировать. Его полезно принять таким, каким оно есть, и использовать для перекачивания результатов расчетов в Word, который открывает гораздо более привлекательные возможности построения таблиц, графиков и подготовки текстов к публикации.***

## Задание для самостоятельной работы

1. Какие вы можете назвать статистические процедуры в SPSS?
2. Перечислите общие принципы выполнения статистических процедур в SPSS?
3. Пропишите путь к статистической процедуре Crosstabs.
4. В чем различие между главным и дополнительным диалоговым окном статистической процедуры?
5. Опишите основные составляющие главного диалогового окна статистической процедуры.
6. Опишите основные составляющие дополнительного диалогового окна статистической процедуры.
7. Как устроено окно просмотра (вывода) в SPSS?
8. Какое расширение имеют файлы окна просмотра в последних версиях системы?
9. Где хранятся файлы окна просмотра?
10. Как можно сменить окно просмотра?
11. Какую последовательность команд необходимо выполнить для сохранения файла?
12. Для каких целей служит диалоговое окно Open Output?
13. Как можно изменить установки в окне просмотра?
14. В чем различие окна просмотра в последних версиях SPSS и окна вывода в ранних его версиях?
15. В каком случае в SPSS закрыт доступ к выполнению статистических расчетов?
16. Чем редактор данных отличается от окна просмотра?
17. Чем обусловлен выбор статистической процедуры при выполнении расчетов?
18. Для каких целей нужен список переменных, который имеется в главном диалоговом окне любой статистической процедуры?

19. Как можно изменить порядок вывода переменных в списке переменных главного диалогового окна статистической процедуры?

20. Для каких целей необходимы имена и описание переменных в SPSS?

21. Как в окне просмотра будут выдаваться результаты расчетов в случае отсутствия описания и значений меток переменных?

22. Назовите кнопки-выключатели основных и дополнительных команд.

23. Для каких целей служит дополнительное диалоговое окно Statistics?

24. Для каких целей служат дополнительные диалоговые окна Charts и Plots?

25. Для каких целей служит дополнительное диалоговое окно Format?

26. Куда ведет кнопка Paste главного диалогового окна любой процедуры в SPSS?

27. Использование какой кнопки в дополнительном диалоговом окне позволяет вернуться в главное диалоговое окно?

28. Какая кнопка главного диалогового окна ведет к выполнению расчетов и их выдачи в окне просмотра?

29. Куда ведет кнопка Help главного диалогового окна любой процедуры в SPSS?

30. Как можно изменить формат выдачи результатов расчетов в окне просмотра?

31. Как можно редактировать результаты расчетов в окне просмотра?

32. Как можно сохранить результаты расчетов, полученные в окне просмотра?

# Глава 8. Описательные статистики и отчеты

В отчетах и описательных статистиках (Descriptive Statistics) имеется возможность выполнения восьми различных статистических процедур: OLAP (Online Analytical Processing) Cubes, Case Summaries, Report Summaries in Rows, Report Summaries in Columns, Frequencies, Descriptives, Explore, Crosstabs. В этой главе рассмотрен порядок выполнения расчетов в основной массе перечисленных выше процедур, а именно: Frequencies, Descriptives, Explore, Case Summaries, Report Summaries in Rows и in Columns. В то же время выполнение расчетов с помощью процедуры Crosstabs, в связи с их важностью и распространенностью в социологических исследованиях, рассматривается отдельно в следующей главе. А выполнение расчетов в OLAP Cubes, в связи с их простотой и изящностью, оставлено для самостоятельного освоения.

## 8.1. Базовая процедура расчета частот - Frequencies

Процедура **Frequencies** (частоты) позволяет строить статистические ряды распределения (31, С. 27-28). Ряды распределения помогают изучать структуру анализируемой совокупности. Они строятся на основе разделения всей совокупности наблюдений на качественно однородные группы по определенному признаку, который выбирается в зависимости от целей и задач исследования.

Другими словами, каждый ряд распределения характеризует состав изучаемых явлений всегда только по одному признаку. В социологии такого рода ряды величин принято называть «первичными распределениями». Собственно, с их построения и начинается анализ уже прошедших контроль в системе SPSS данных.

Например, требуется подсчитать число мужчин и женщин среди респондентов. Сходная задача описывалась нами ранее в главе 5 (§ 5.3). Но там она выполнялась с целью контроля правильности ввода

данных. В настоящий момент она рассматривается в качестве аналитической задачи. Ее решение связано с построением ряда распределения по полу респондента (переменная sexresp9). Для выполнения процедуры Frequencies необходимо выбрать в меню:

**Analyze**

**Descriptive Statistics**

**Frequencies**.

В ранних версиях SPSS для достижения указанной цели требовалось выполнение последовательности команд: Statistics-Summarize-Frequencies. В любом случае выполнение указанной последовательности команд ведет к открытию диалогового окна Frequencies. В этом окне слева появится список переменных. В нем, как уже многократно говорилось ранее, можно выделить интересующие переменные и переместить их в соседнее окно, нажав кнопку «стрелка вправо» (глава 1, § 1.3, рис. 3). Выбранные переменные попадут в правый список. После нажатия кнопки OK процедура начнет выполняться, и результаты будут выдаваться в окне просмотра.

Пример таблицы, построенной по переменной sexresp9 (пол респондента) и появившейся в окне просмотра, представлен ниже в обрамлениях 7 и 8.

**Обрамление 7. Окно просмотра результатов расчета частот по полу респондента в выборке 1999 г. (версия SPSS 11.5)**

пол респондента

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | мужчины | 106 | 25,1 | 25,1 | 25,1 |
| | женщины | 316 | 74,9 | 74,9 | 100,0 |
| | Total | 422 | 100,0 | 100,0 | |

Сопоставление структуры окон просмотра одной из последних и ранней версии системы хорошо показывает его эволюцию. Сейчас - это уже фактически таблица, которую, в принципе, можно целиком перенести в Word, используя, после ее предварительного выделения, последовательность команд главного меню окна вывода: **Edit-Copy objects** (Ctrl+K). Это означает, что вы скопировали в текстовый процессор объект. Любая попытка его редактирования в качестве таблицы Word связана с трудностями, которые вряд ли удастся преодолеть.

Попытка переноса таблицы путем использования стандартного **Edit-Copy** (Ctrl+C) технически вполне допустима, но форматировать

и редактировать такую таблицу в текстовом процессоре Word будет довольно сложно.

**Обрамление 8. Окно вывода результатов расчета частот по полу респондента в выборке 1999 г. (версия SPSS 6.1)**

SEXRESP9 пол респондента

| Value Label | Value | Frequency | Percent | Valid Percent | Cum Percent |
|---|---|---|---|---|---|
| мужской | 1 | 106 | 25,1 | 25,1 | 25,1 |
| женский | 2 | 316 | 74,9 | 74,9 | 100,0 |
| Total | | 422 | 100,0 | 100,0 | |
| Valid cases | 422 | | | | |

В ранней версии системы (обрамление 8) в верхней строке таблицы окна вывода дается имя переменной и ее описание на русском языке (конечно, если оно вводилось при формировании файла с данными). В поздней версии (обрамление 7) выведено только описание переменной (пол респондента). Это значит, что в установках стоит опция «Display labels» (глава 7, § 7.3.).

Value Label - имя переменной (в данном случае, «мужской» и «женский» пол).

Value - это варианты или определенные числовые значения варьирующего признака (в данном случае, «пол респондента»: мужской-1, женский -2).

Frequency - частоты или абсолютные числа, показывающие, сколько раз встречается тот или иной вариант (31, С. 28). В нашем случае - это 106 мужчин и 316 женщин.

Percent - процентное выражение числовых значений переменной с учетом пропущенных значений.

Valid Percent - валидное процентное выражение числовых значений переменной без учета пропущенных значений.

Cum Percent - накопленный (кумулятивный) процент.

Total (всего) показывает общее число наблюдений (в данном случае, их 422) и 100% по колонкам Percent и Valid Percent.

Последняя строка, выходящая за пределы ряда распределения, показывает общее число валидных случаев (422).

Если в итоговом документе необходимо иметь описанные выше данные в хорошей табличной форме, то лучший выход скопировать полу-

ченный в окне вывода ряд распределения и перенести его в предварительно построенную в текстовом процессоре, в котором пишется отчет об исследовании (текст), «родную» таблицу. Сегодня для этих целей обычно используется MS Word. А это значит, что необходимо научиться работать с таблицами указанного текстового процессора. К примеру, в нем рассматриваемая таблица может быть преобразована следующим образом (таб. 3).

**Таблица 3.Распределение опрошенных по полу в выборке 1999 г.**

| Пол респондента | Абс. (чел.) | Относ. (%) |
|---|---|---|
| Мужской | 107 | 25,4 |
| Женский | 315 | 74,6 |
| Итого | 422 | 100,0 |

В приведенном примере показано, что анализ осуществляется на основе 422 домохозяйств, входящих в базу данных панели 1995 - 1997 - 1999 гг. Следует иметь в виду, что переменная (пол респондента) номинальная. Это значит, что рассматриваемый ряд распределения построен по качественному признаку (пол). В статистике такой ряд распределения называется **атрибутивным.** Проведение различных статистических расчетов в таком ряду распределения не имеет смысла. Для этой цели в качестве примера лучше использовать ряд распределения, образованный по количественному признаку. Такой ряд распределения называется **вариационным.**

«Вариационные ряды бывают: а) прерывные, которые носят название дискретных или ранжированных, т.е. расположенных в порядке возрастания от наименьшего значения к наибольшему; и б) непрерывные, называемые интервальными» (31, С. 28).

В качестве примера построим ранжированный вариационный ряд. Основой расчетов будет порядковая переменная gamb9 (оценка качества медицинского обслуживания). Статистические расчеты выполняются с использованием дополнительных диалоговых окон рассматриваемой процедуры. Все эти окна уже описаны в главе 7 (§ 7.4).

Находясь в главном диалоговом окне процедуры Frequencies (рис. 3), нажимаем на первую слева кнопку в нижнем ряду **Statistics...** и открываем дополнительное диалоговое окно расчета частотных статистик (рис. 23).

В этом окне задаем все требуемые в анализе статистики. Затем при желании открываем следующие дополнительные диалоговые окна

**Charts... -** построение частотных диаграмм (рис. 24) и **Format... -** для выбора порядка представления переменных в частотной таблице (по возрастанию или убыванию), а также выбора формата страницы, который может быть стандартный, сжатый и т.п. (рис. 25).

**Обрамление 9. Окно просмотра частотного распределения**

GAMB9 обслуживание в поликлинике

| Value Label | Value | Frequency | Percent | Valid Percent | Cum Percent |
|---|---|---|---|---|---|
| очень плохо | 1 | 2 | .5 | .5 | .5 |
| | 2 | 40 | 9.5 | 9.5 | 10.0 |
| | 3 | 168 | 39.8 | 39.8 | 49.8 |
| | 4 | 146 | 34.6 | 34.6 | 84.4 |
| отлично | 5 | 50 | 11.8 | 11.8 | 96.2 |
| затр. ответить | 9 | 16 | 3.8 | 3.8 | 100.0 |
| Total | | 422 | 100.0 | 100.0 | |

Hi-Res Chart # 3:Bar chart of обслуживание в поликлинике

| | | | | | |
|---|---|---|---|---|---|
| Mean | 3.706 | Std err | .065 | Median | 4.000 |
| Mode | 3.000 | Std dev | 1.343 | Variance | 1.804 |
| Kurtosis | 6.746 | S E Kurt | .237 | Skewness | 2.155 |
| S E Skew | .119 | Range | 8.000 | Minimum | 1.000 |
| Maximum | 9.000 | Sum | 1564.000 | | |

| Percentile | Value | Percentile | Value | Percentile | Value |
|---|---|---|---|---|---|
| 25.00 | 3.000 | 50.00 | 4.000 | 75.00 | 4.000 |

Valid cases 422

В обрамлении 9 показано, как выглядит ряд распределения оценок обслуживания в поликлинике (переменная gamb9) в окне вывода одной из ранних версий системы. В результате расчетов в окне просмотра выведены: частотная таблица, ссылка на построенный график, а также все требовавшиеся статистики (средняя арифметическая, медиана, мода, стандартное отклонение и др.). Все указанные статистики будут рассмотрены в последующих параграфах. Сравнение данных обрамлений 7-8 и 9 позволяет наглядно видеть различия построения рядов распределения для переменных, образованных по качественному и количественному признакам.

**Дружеский совет**

***Анализ данных лучше всего начинать с построения частотных таблиц и диаграмм, т.е. с выполнения статистической процедуры Frequencies***

## 8.2. Описательные статистики - Descriptives

Процедура **Descriptives** (дескриптивные или описательные статистики) вычисляет: среднюю, максимальное и минимальное значения, стандартное отклонение и др. В принципе Descriptives вычисляет статистики, которые доступны и в процедуре Frequencies (частоты). Но для непрерывных переменных она делает это более эффективно (сохраняя много времени и места), т.к. не сортирует переменные в частотную таблицу.

Для использования этой процедуры необходимо выполнить последовательность команд:

**Analyze**

**Descriptive Statistics**

**Descriptives.**

В ранних версиях SPSS для достижения указанной цели требовалось выполнение последовательности команд: Statistics-Summarize-Descriptives. Выполнение указанной последовательности команд ведет к открытию главного диалогового окна рассматриваемой статистической процедуры. Это окно показано на рис 27.

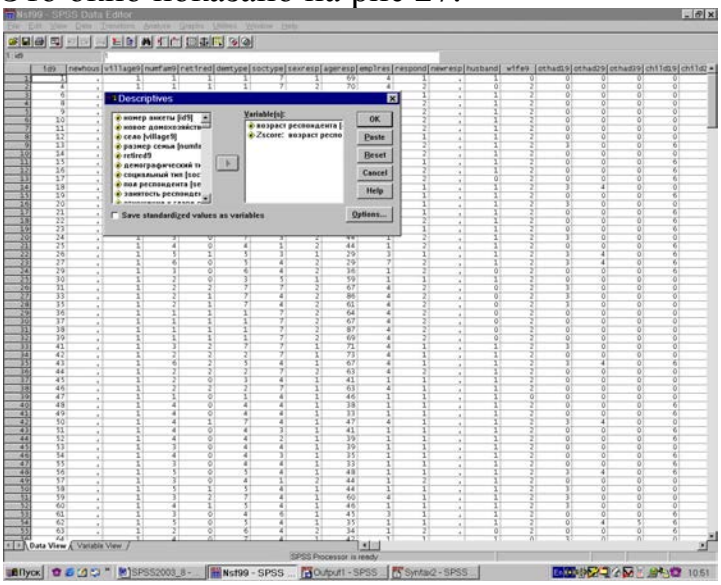



***Рис. 27.***

***Главное диалоговое окно процедуры Descriptives***

Главное диалоговое окно процедуры Descriptives имеет стандартный вид и состоит из двух полей. Одно из них (левое) содержит список исходных переменных. Другое поле служит для переноса и последующего анализа интересующих разработчиков переменных.

Кроме того, в левом нижнем углу оно содержит очень важный чекбокс, позволяющий, в случае установки данной опции, сохранять нормированные значения в качестве новой переменной (Save standardized values as variables). Более полно этот вопрос освещен в текущей главе и в главе 13, § 13.2.

В обрамлениях 10 и 11 приведен пример использования, соответственно, Frequencies и Descriptives с целью получения суммарных статистик для одной непрерывной переменной - возраста респондента (ageresp9).

При использовании Frequencies частотная таблица заняла более двух страниц (в обрамлении 10 показаны только начало и конец частотной таблицы), и только потом были получены суммарные статистики: средняя, медиана, мода, максимальное и минимальное значение, стандартное отклонение.

**Обрамление 10.Суммарные статистики возраста респондента в окне просмотра процедуры Frequencies**

Statistics

возраст респондента

| | | |
|---|---|---|
| N | Valid | 422 |
| | Missing | 0 |
| Mean | | 54,75 |
| Median | | 58,00 |
| Mode | | 39 |
| Std. Deviation | | 17,19 |
| Minimum | | 22 |

возраст респондента

| | | Frequency | Percent | Valid Percent | Cum Percent |
|---|---|---|---|---|---|
| Valid | 22 | 1 | ,2 | ,2 | ,2 |
| | 24 | 1 | ,2 | ,2 | ,5 |
| | 25 | 3 | ,7 | ,7 | 1,2 |
| | … | … | … | … | |
| | 95 | 1 | ,2 | ,2 | 100,0 |
| Total | | 422 | 100,0 | 100,0 | |

При использовании Descriptives суммарные статистики были получены сразу. Правда, следует обратить внимание, что в процедуре Descriptives недоступны вычисления медианы и моды (обрамление 11).

**Обрамление 11. Суммарные статистики возраста респондента в окне просмотра процедуры Descriptives**

Descriptive Statistics

| | N | Minimum | Maximum | Mean | Std. Deviation |
|---|---|---|---|---|---|
| возраст респондента | 422 | 22 | 95 | 54,75 | 17,19 |
| Valid N (listwise) | 422 | | | | |

Как уже отмечалось ранее, процедура Descriptives дает описание средних, стандартного отклонения, дисперсии и др. статистик для нормального распределения, а также минимальное значение, размах и сумму для асимметричного распределения с количественной переменной. Все эти статистики можно получить, открыв дополнительное диалоговое окно **Options** главного диалогового окна процедуры Descriptives (рис. 28).

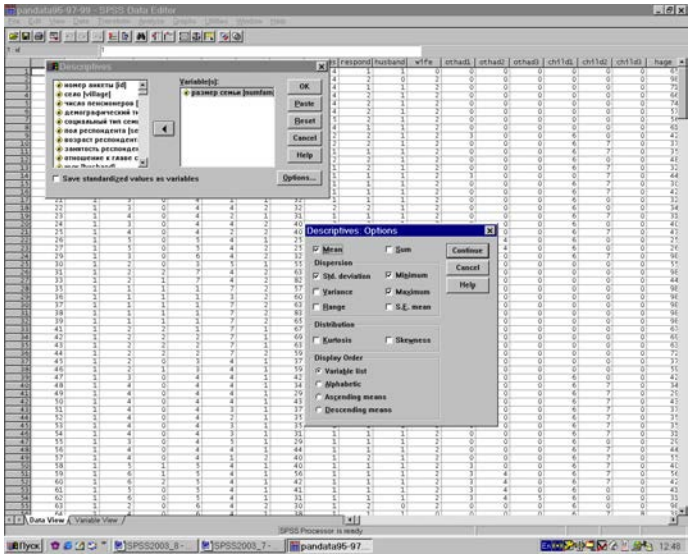



***Рис. 28.***

***Дополнительное диалоговое окно Options - опции процедуры Descriptives***

Кроме того, в дополнительном диалоговом окне Options предусмотрен выбор порядка представления переменных (Display Order) при получении результатов в окне вывода. Иными словами, если анализи

руется одновременно несколько переменных, то результаты могут быть представлены в следующем виде:

- по возрастанию (Ascending means) средних значений переменных;
- по убыванию (Descending means) средних значений переменных;
- в алфавитном порядке (Alphabetic) выбранных переменных;
- в порядке занесения переменных в список диалогового окна Descriptives (Variable list).

Выбор того или иного порядка представления переменных может быть сделан лишь в каждом конкретном случае. Тогда как по умолчанию в окне просмотра результаты всегда выдаются в порядке занесения переменных в список для анализа, т.е. в поле Variable(s).

Процедура Descriptives имеет еще одну важную особенность. Она позволяет нормировать переменные и сохранять их в качестве новых переменных в рабочем файле. Для этой цели требуется пометить флажком квадратный бокс в нижней левой части главного диалогового окна процедуры (рис. 27). Эта опция называется: Save standardized values as variables (сохранять стандартизированные значения в качестве переменных). Новая переменная будет автоматически записываться в конец рабочего файла с дополнительной буквой «z», которая встанет на первое место в имени стандартизируемой переменной. Если в имени исходной переменной уже было 8 знаков, то последний знак стирается. В нашем случае, например, переменная возраст респондента (ageresp9) преобразована в новую переменную – zageresp. Результаты преобразования приведены в таб. 4.

**Таблица 4. Сопоставление исходных и стандартизированных значений переменной возраст респондента**

| Перемен-ные | Описательные статистики | | | | | | |
|---|---|---|---|---|---|---|---|
| | N | Range | Min | Max | Sum | Mean | Std. Deviation |
| Возраст респондента | 422 | 73 | 22 | 95 | 23105 | 54,75 | 17,19 |
| Zscore: возраст респондента | 422 | 4,24677 | -1,90530 | 2,34147 | ,00000 | 3,7339923E-16 | 1,0000000 |

В качестве другого примера здесь можно привести нормирование рядов распределения при сопоставлении тенденций механического движения населения в стране и одном из субъектов федерации, скажем, в Карачаево-Черкессии в 1991-2003 гг. Огромная разница в масштабах абсолютных цифр (миллионы и тысячи) не позволяет здесь делать какие-либо прямые сопоставления. Нормирование этих переменных и последующее создание графика дает возможность увидеть общность происходящих процессов. Повсеместно люди вынуждены уходить из родных мест в поисках лучшей жизни.

Нормирование значений переменных при стандартизации выполняется по формуле: **Z = X** (исходное значение) **– M** (среднее значение переменной) **/ S** (стандартное отклонение). Необходимость такого рода операции появляется тогда, когда значения нескольких переменных требуется привести к соизмеримым числам или общему показателю. Этот вопрос будет рассматриваться еще не один раз в главах 13-16.

**Дружеский совет**

***Процедуру Descriptives лучше всего использовать для быстрого получения суммарных статистик количественных переменных.***

## 8.3. Исследовательские статистики - Explore

После ввода данных и проверки их на корректность довольно часто возникает потребность предварительного (экспресс) анализа. Такая потребность вполне разумна как с точки зрения быстрого получения необходимой информации, так и с точки зрения проверки массива с помощью простой исследовательской техники.

Всегда полезно найти возможные объяснения в случае обнаружения малооправданной изменчивости данных. Например, если в распределении значений данных существует пропуск, или некоторые значения являются экстремальными - сильно отличающимися от остальных, либо форма распределения, создаваемая числовыми значениями, кажется странной. Для всех этих целей и полезно использовать процедуру **Explore** (исследование).

Для выполнения рассматриваемой процедуры необходимо реализовать последовательность команд:

**Analyze**

**Descriptive Statistics**

**Explore**.

В ранних версиях SPSS для достижения указанной цели требовалось выполнение последовательности команд: Statistics-Summarize-Explore. В результате выполнения этой последовательности команд откроется главное диалоговое окно процедуры Explore (рис. 29).

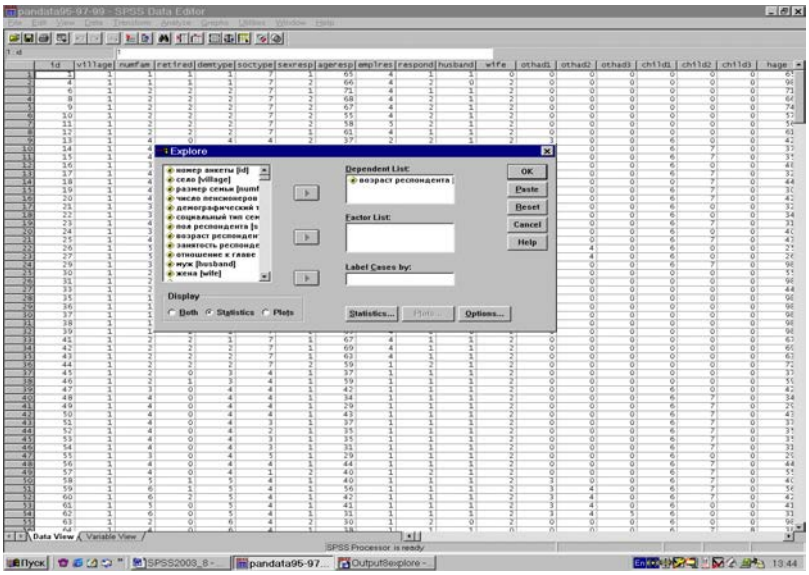



***Рис. 29. Главное диалоговое окно процедуры Explore***

Как видно на рис 29, это окно сходно с главными диалоговыми окнами других статистических процедур. Слева находится список переменных, из которого они выбираются для выполнения процедуры. Выбор переменной опять же осуществляется путем выделения в списке имени переменной.

Выбранная переменная с помощью стрелок перемещается в одно из трех полей, находящихся в средней части окна. Эти поля в порядке от верхнего к нижнему называются: список зависимых переменных (Dependent List), список независимых переменных (Factor List) и метки случаев (Label Cases by). Различие между зависимыми переменными и факторами в рассматриваемом окне означает, что может быть выполнен углубленный анализ по группам случаев (например, анализ возрастной структуры или доходов домохозяйств отдельно по каждому селу, попавшему в выборку).

Анализ может быть выполнен и без группирующей переменной (т.е. по возрастной структуре или доходам домохозяйств в массиве в целом). Для этого необходимо перенести исследуемую переменную(ые) в поле Dependent List и начать выполнение процедуры.

Далее, в левой нижней части окна в поле Display полезно пометить флажком формат вывода информации. Ее можно вывести в виде описательной статистики (Statistics), графика (Plots), или использовать обе возможности (Both). По умолчанию метка как раз и задает выполнение статистики и графиков, т.е. функцию Both, стоящую на первом месте.

Кроме того, в главном диалоговом окне процедуры Explore имеются три кнопки: Statistics..., Plots..., Options..., указывающие на наличие трех дополнительных диалоговых окон (рис. 30-32). Благодаря своим сервисным возможностям, рассматриваемая процедура позволяет визуально изучить распределение значений для различных групп, проверить нормальность распределения и однородность дисперсии и т.п. С этой целью и следует использовать дополнительное диалоговое окно Statistics... (рис. 30).

На рис. 30 видно, что дополнительное диалоговое окно Statistics процедуры Explore содержит несколько маленьких полей. Одно из них - Descriptives. Сделав в нем пометку, можно получить в окне просмотра описание диаграммы ствол-лист (Stem & Leaf ).

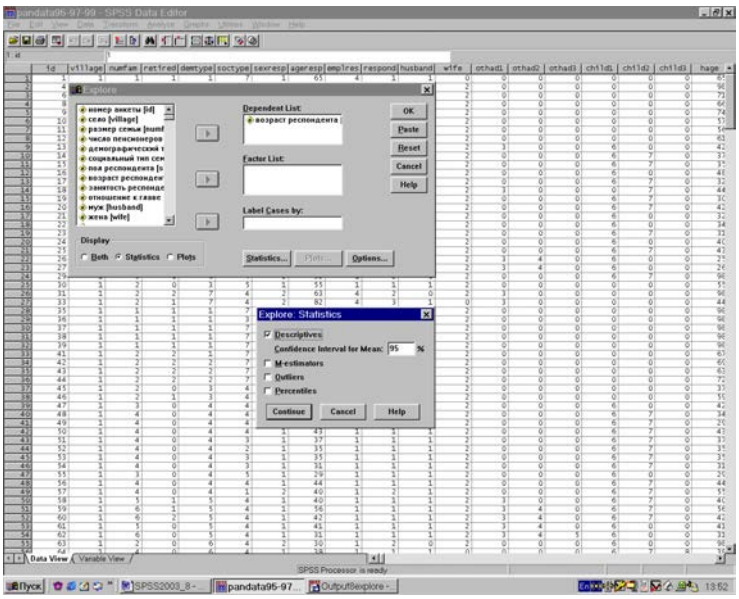



***Рис. 30.***

***Дополнительное диалоговое окно Statistics процедуры Explore***

Каждая строка такой описательной диаграммы (обрамления 12 и 13) содержит определенный интервал значений. Числа слева (Frequency) показывают точное число наблюдений, имеющих значения, указанные в строке.

Следующий ряд - ствол (Stem). Это цифра старшего числа -значения, представленного в строке. Далее идет разряд - лист (Leaf). Сколько разрядов игнорируется - указывается в ширине ствола (Stem width).

Если ширина ствола, как в нашем случае, равна 10, то указываются все разряды, если 100 - то игнорируются единицы, если 1000 - то игнорируются единицы и десятки и т.д. Игнорирование происходит, когда много разнообразных значений в большом интервале.

Каждый лист (Each leaf) показывает, сколько значений он содержит. Если это значение не равно 1, то оно может теряться. Так, в случае, когда Each leaf равен 2, если значение присутствует только один раз, оно вообще не показывается, а если присутствует три раза, то показывается всего 1 раз. Это говорит о том, что система не выводит точное число самих значений.

**Обрамление 12. Диаграмма ствол-лист для возраста респондента в массиве 1999 г.**

**Case Processing Summary**

| | Cases | | | | | |
|---|---|---|---|---|---|---|
| | Valid | | Missing | | Total | |
| | N | Percent | N | Percent | N | Percent |
| возраст респондента | 422 | 100,0% | 0 | ,0% | 422 | 100,0% |

**Descriptives**

| | | | Statistic | Std. Error |
|---|---|---|---|---|
| возраст респондента | Mean | | 54,75 | ,84 |
| | 95% Confidence Interval for Mean | Lower Bound | 53,11 | |
| | | Upper Bound | 56,40 | |
| | 5% Trimmed Mean | | 54,44 | |
| | Median | | 58,00 | |
| | Variance | | 295,480 | |
| | Std. Deviation | | 17,19 | |
| | Minimum | | 22 | |
| | Maximum | | 95 | |
| | Range | | 73 | |
| | Interquartile Range | | 30,00 | |
| | Skewness | | ,103 | ,119 |
| | Kurtosis | | -1,135 | ,237 |

```
возраст респондента Stem-and-Leaf Plot


 Frequency    Stem &  Leaf

     2,00        2 .  &
    14,00        2 .  58999&
    42,00        3 .  00000111112233334444
    58,00        3 .  55555566667777778889999999999
    39,00        4 .  0001111112233334444
    30,00        4 .  55566677888889
    18,00        5 .  00012234
    29,00        5 .  57788888999999
    45,00        6 .  000011111222222333444
    45,00        6 .  555667777788888999999
    45,00        7 .  00011111222223334444
    30,00        7 .  5556666777889
     5,00        8 .  14
    16,00        8 .  556779&
     3,00        9 .  0&
     1,00        9 .  &

 Stem width:  10

 Each leaf:       2 case(s)

 & denotes fractional leaves.
```

Если в строке есть значения, которые вообще не показаны, то такая строка метится символом &. После слова Extremes - идут значения, которые называются экстремальными, т.е. значения сильно отличающиеся от остальных. В норме такие значения должны встречаться довольно редко.

В нашем примере ширина ствола (Stem width) равна 10, поэтому в значениях не теряются разряды. Характеристика Each leaf равна 2, поэтому каждый лист - это 2 значения, и они могут теряться. В первой строке 2 значения, которые попадают в интервал от 20 до 24. Какие это значения точно сказать нельзя, т.к. система их не выдала. Но можно предположить, что такие значения встречаются по одному разу.

Во второй строке 14 значений с возрастом - 25, 28, 29. Здесь возраста 25 может быть либо 2, либо 3 случая, возраста 28 - то же самое, возраста 29 - 6 или 7 случаев, а какое-то из значений 26-27 или оба могут быть потеряны, т.к. строка помечена &.

В тринадцатой строке значения от 80 до 84. Указаны только 81 и 84. Других значений нет, т.к. строка не помечена &. Всего значений должно быть 5. Поэтому можно сказать, что какое-то из этих двух значений присутствует 3 раза.

## Обрамление 13. Диаграмма ствол-лист для доходов семьи

**Case Processing Summary**

| | Cases | | | | | |
|---|---|---|---|---|---|---|
| | Valid | | Missing | | Total | |
| | N | Percent | N | Percent | N | Percent |
| общий месячный доход | 422 | 100,0% | 0 | ,0% | 422 | 100,0% |

**Descriptives**

| | | | Statistic | Std. Error |
|---|---|---|---|---|
| общий месячный доход | Mean | | 2881,84 | 106,28 |
| | 95% Confidence Interval for Mean | Lower Bound | 2672,93 | |
| | | Upper Bound | 3090,75 | |
| | 5% Trimmed Mean | | 2676,87 | |
| | Median | | 2591,00 | |
| | Variance | | 4766722,9 | |
| | Std. Deviation | | 2183,28 | |
| | Minimum | | 304 | |
| | Maximum | | 27675 | |
| | Range | | 27371 | |
| | Interquartile Range | | 2391,75 | |
| | Skewness | | 4,387 | ,119 |
| | Kurtosis | | 41,200 | ,237 |

```
общий месячный доход Stem-and-Leaf Plot

Frequency    Stem &  Leaf
       4,00        0 .  34
      61,00        0 .  666666777777777888888888999999
      50,00        1 .  00000001111222223333344
      37,00        1 .  555666777888889999
      49,00        2 .  000000111111122233334444
      51,00        2 .  55555566666677777778889999
      45,00        3 .  000001111122233333444
      34,00        3 .  5556667777788889
      29,00        4 .  00011122333444
      18,00        4 .  5567889
      13,00        5 .  012444
       9,00        5 .  56&
       6,00        6 .  034
       5,00        6 .  9&
       1,00        7 .  &
      10,00 Extremes    (>=7828)

Stem width:    1000
Each leaf:       2 case(s)

  & denotes fractional leaves.
```

Диаграмма ствол-лист для доходов семьи (обрамление 13) приведена для демонстрации и большей наглядности различий отдельных значений. В этом случае ширина ствола (Stem width) равна 1000, т.е. в значениях переменной игнорируются десятки и единицы.

Например, третья строка диаграммы будет содержать значения: 1000, 1100, 1200, 1300, 1400 (независимо от того, какие в значениях десятки и единицы: число 1156 будет показано как 1100, 1489 - как 1400). В отличие от предыдущего примера, в этом примере появились экстремумы.

Как видно из обрамлениий 12 - 13, вместе с диаграммой ствол-лист выдаются различные статистики: минимальное значение, максимальное значение, среднее, медиана, стандартное отклонение и т.д. Все эти статистики были описаны ранее, при рассмотрении процедур частоты - Frequencies (§ 8.1) и описательные статистики - Descriptives (§ 8.2).

**Обрамление 14. Экстремальные значения в диаграмме ствол-лист для возраста респондента**

Extreme Values

| 5 Highest | Case # | 5 Lowest | Case # |
|---|---|---|---|
| 95 | Case: 387 | 22 | Case: 164 |
| 91 | Case: 486 | 24 | Case: 397 |
| 90 | Case: 252 | 25 | Case: 421 |
| 90 | Case: 357 | 25 | Case: 189 |
| 89 | Case: 500 | 25 | Case: 195 |

**Обрамление 15. Экстремальные значения в диаграмме ствол-лист для доходов семьи**

Extreme Values

| 5 Highest | Case # | 5 Lowest | Case # |
|---|---|---|---|
| 26357 | Case: 317 | 300 | Case: 426 |
| 23342 | Case: 377 | 304 | Case: 386 |
| 12146 | Case: 80 | 329 | Case: 327 |
| 8110 | Case: 87 | 375 | Case: 207 |
| 7524 | Case: 408 | 375 | Case: 204 |

Дополнительно можно выдать номера строк (анкет) с 5 максимальными и 5 минимальными значениями (обрамления 14 и 15). Для получения дополнительной таблицы с экстремальными значениями надо выбрать опцию выбросы (Outliers) еще при формировании задания в подокне Statistics.

В обрамлениях 14 и 15 четко видны по 5 максимальных (Highest) и 5 минимальных (Lowest) значений переменных, соответственно, возраста респондента и дохода семьи, с конкретным указанием номера каждого из этих случаев (Case) в анализируемом массиве данных. Для построения рассмотренных выше диаграмм чуть раньше в главном диалоговом окне процедуры Explore необходимо было, как и во всех других сходных случаях, сначала выделить из списка переменных требуемые переменные и перенести их в центральное верхнее поле, которое называется список зависимых переменных (Dependent List).

Можно получить диаграммы ствол-лист для той же самой переменной (переменных) в зависимости от другой переменной (переменных), то есть выполнить анализ по группирующей переменной, если таковую задать в находящемся чуть ниже подокне список факторов (Factor List). В этом случае, на каждое значение заданной переменной (переменных) - фактора (Factor List) будет строиться своя диаграмма для зависимой переменной (переменных) Dependent List.

В обрамлениях 16 и 17 показан пример, когда для переменной «возраст респондента» построена диаграмма ствол-лист в зависимости от пола респондента. В случае, если в списке переменных (Dependent List) задано несколько исходных переменных, то диаграммы будут строиться последовательно для каждой из них. При этом для начинающего пользователя понять что-нибудь будет довольно сложно. Поэтому здесь лучше не торопиться, а совершать пошаговые движения. В любом случае, использование рассматриваемой процедуры предполагает определенный уровень статистической подготовки.

**Обрамление 16. Диаграмма ствол-лист возраста респондента в зависимости от его пола (мужчины) (фрагмент окна просмотра)**

возраст респондента Stem-and-Leaf Plot for
SEXRESP9= мужской

| Frequency | Stem & | Leaf |
|---|---|---|
| 2,00 | 2 | . 89 |
| 9,00 | 3 | . 112233344 |
| 16,00 | 3 | . 5555677789999999 |
| 13,00 | 4 | . 0111222333344 |
| 15,00 | 4 | . 566667788888889 |
| 3,00 | 5 | . 002 |
| 6,00 | 5 | . 778999 |
| 15,00 | 6 | . 000111122233444 |
| 15,00 | 6 | . 555677777888899 |
| 9,00 | 7 | . 001112223 |
| 2,00 | 7 | . 57 |
| ,00 | 8 | . |
| 1,00 | 8 | . 7 |
| 1,00 | 9 | . 0 |

Stem width: 10
Each leaf: 1 case(s)

При выполнении функции Explore диаграмма ствол-лист выдается по умолчанию (Display-Both) или при задании опции Display-Statistics. Одновременно и также по умолчанию выдается график типа «ящичковая диаграмма». Этот график описан в главе 12. Дополнительно можно получить график типа гистограмма, который также описан в главе 12.

**Обрамление 17. Диаграмма ствол-лист возраста респондента в зависимости от его пола (женщины) (фрагмент окна просмотра)**

возраст респондента Stem-and-Leaf Plot for
SEXRESP9= женский

| Frequency | Stem & | Leaf |
|---|---|---|
| 2,00 | 2 | . 24 |
| 12,00 | 2 | . 555678899999 |
| 33,00 | 3 | . 000000000011111111222333333444444 |
| 42,00 | 3 | . 555555556666666777777777788888999999999999 |
| 26,00 | 4 | . 00000111111111233334444444 |
| 15,00 | 4 | . 555556667788899 |

```
      15,00    5    . 000001112223344
      23,00    5    . 5577888888889999999999
      30,00    6    . 000000111111222222222233334444
      30,00    6    . 555666777777888888999999999999
      36,00    7    . 000001111111122222222333333444444444
      28,00    7    . 55555566666666677777778888999
       5,00    8    . 11444
      15,00    8    . 555556667778999
       2,00    9    . 01
       1,00    9    . 5
Stem width:  10
  Each leaf:      1 case(s)
```

Процедура Explore дает возможность получать самостоятельно (непосредственно) разнообразные графики. Например, такие как: гистограммы, диаграммы «ствол-лист», ящичковые диаграммы, нормальную вероятностную бумагу, диаграммы типа «разброс против среднего», а также тесты на однородность дисперсии (тест Левена), на нормальность распределения (тесты Шапиро-Уилкса и Лильефорса) и оценки максимального правдоподобия.

С этой целью и следует использовать дополнительное диалоговое окно графики - Plots... (рис. 31). Для получения графиков с помощью дополнительного диалогового окна Plots рассматриваемой процедуры в качестве первого шага следует выполнить типовую последовательность команд:

**Analyze**

**Descriptive Statistics**

**Explore**

**Plots**.

В ранних версиях SPSS для достижения указанной цели требовалось выполнение последовательности команд: **Statistics-Summarize-Explore-Plots**. Открывшееся окно Plots предлагает выбор из четырех групп графиков: Boxplots, Descriptive, Normally plots with Tests и Spread vs. Level with Levene Test.

Для построения того или иного типа графиков, равно как и нескольких графиков, одновременно в окошках, стоящих слева от имени каждого графика, следует установить необходимую опцию. После нажатия клавиши Continue окно графиков закрывается и идет возврат в главное диалоговое окно процедуры Explore, Для выполнения команды здесь, как всегда в таких случаях, необходимо нажать клавишу OK.

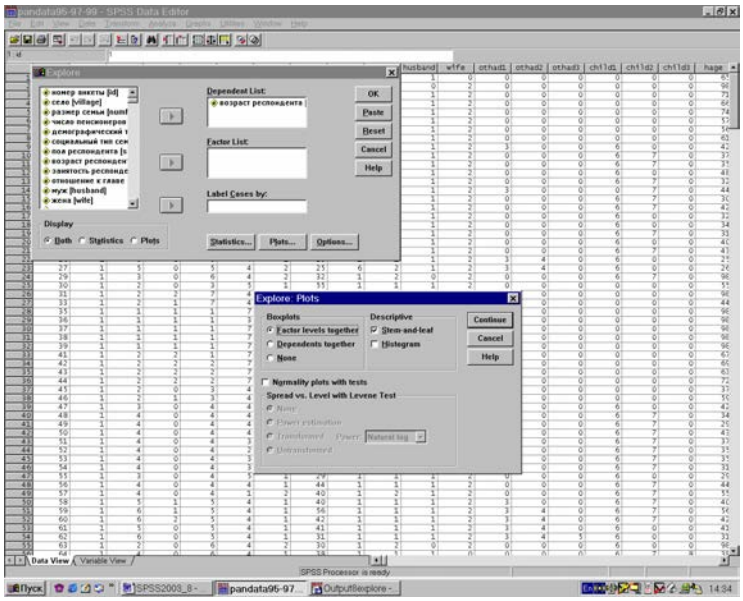



***Рис. 31.***

***Дополнительное диалоговое окно Plots процедуры Explore***

К примеру, на рис. 31 видно, что контур Descriptive дополнительного окна Plots открывает возможность получения гистограмм (Histogram). Порядок их построения рассмотрен в главе 12 § 12.2.

Для того, чтобы выдать нужный в данный момент график, необходимо в дополнительном окне Plots установить опцию на его имени, а у имени ненужного на данный момент графика снять отметку.

Дополнительное диалоговое окно Options (опции) процедуры Explore (рис. 32) позволяет включать или исключать пропущенные значения (Missing Values) при выполнении исследований статистических значений различных переменных.

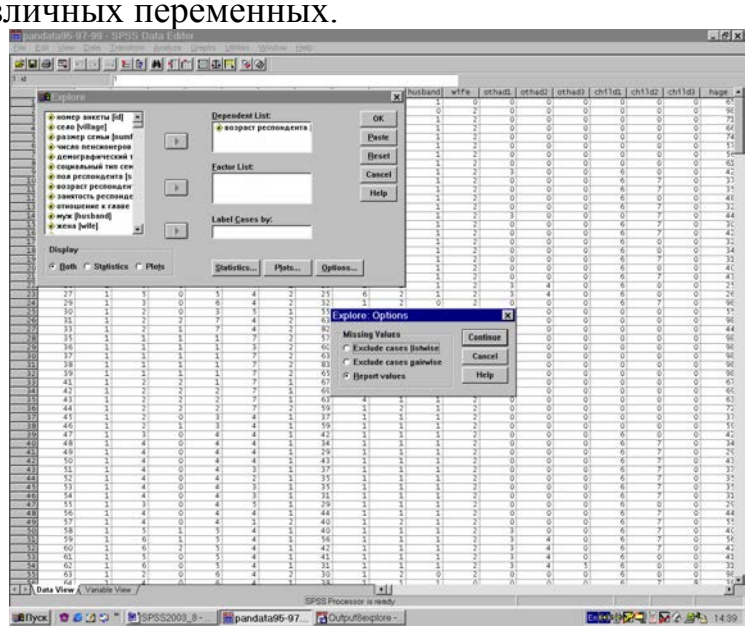



***Рис. 32.***

***Дополнительное диалоговое окно Options процедуры Explore***

Одновременно использование рассматриваемой кнопки открывает возможность регулирования вывода на экран (Display) различных составляющих выполняемых расчетов (диаграмм и таблиц).

**Правило 21**

***В SPSS использование дополнительных диалоговых окон Statistics и Plots процедуры Explore всегда ведет к получению описания диаграммы в окне просмотра и публикации графика.***

## 8.4. Возможности процедуры Case Summaries

Эта процедура, название которой можно перевести как «сводка случаев», доступна, начиная с версии SPSS 7.5 и всех последующих. Для работы в ней, как уже отмечалось ранее, необходимо выполнить последовательность команд:

**Analyze**

**Reports**

**Case Summaries**.

В предыдущих версиях SPSS аналогом процедуры Case Summaries является процедура List Cases, которая описана ранее в первом издании пособия (1, С. 45). Процедура List Cases позволяет просматривать отобранные переменные и визуально сравнивать их значения. Например, изменение дохода домохозяйства по годам в многолетнем панельном исследовании для каждого наблюдения отдельно.

Процедура Case Summaries также позволяет просматривать все эти изменения, но уже по подгруппам наблюдений с набором различных статистик. Одновременно эта процедура дает возможность вывести значения переменной для первого и последнего наблюдения в файле данных, что может оказаться весьма полезным для формирования более полного и законченного представления о характере распределения данных в наблюдаемом массиве.

В нашей практике обе эти процедуры использовались как на этапе контроля данных (глава 5, § 5.2-5.3), так и в аналитических целях. Фактически процедура Case Summaries представляет собой мощный инструмент не только количественного, но и качественного анализа. С ее помощью любой случай может быть представлен и изучен как формализованное интервью (глава 5, § 5.2 и 27, С. 9, 231).

**Дружеский совет**

***Использование процедуры Case Summaries весьма полезно в эвристическом плане, т.е. При осмыслении значений переменных и их изменения во времени.***

## 8.5. Отчет по итогам по строкам - Report Summaries in Rows

Процедура **Report Summaries in Rows** (отчет об итогах по строкам) позволяет выводить в окно просмотра различные характеристики переменных: минимальное и максимальное значения, средняя, сумма и др.

Для выполнения рассматриваемой процедуры в качестве первого шага реализуется типовая последовательность команд:

**Analyze**

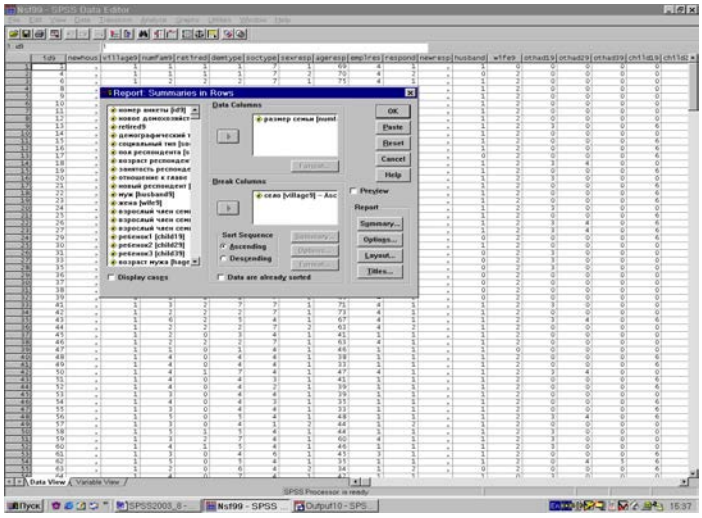
**Reports**

**Report Summaries in Rows.**

В ранних версиях SPSS это была последовательность следующих команд: **Statistics-Summarize-Report Summaries in Rows**. При этом открывается главное диалоговое окно данной процедуры, которое можно видеть на рис. 33.

***Рис. 33.***

***Главное диалоговое окно процедуры Report Summaries in Rows – отчет об итогах по строкам***



Специфика рассматриваемой процедуры состоит в том, что она позволяет одновременно выводить на экран различные статистические значения исследуемой переменной (переменных). Для этого их надо перенести в верхнее центральное поле Data Columns, а требуемый для расчета набор значений различных статистических показателей необходимо задать в дополнительном диалоговом окне Summary, которое находится в секции Report (нижняя часть главного диалогового окна).

Этот момент имеет принципиальное значение, так как в центральной нижней секции Break Columns имеется еще одно дополнительное диалоговое окно Summary. Оба дополнительных диалоговых окна Summary имеют идентичную структуру (рис 34), но совершенно разные целевые функции.

В то время как первое из них предназначено для расчета статистических значений переменной(ых), находящейся в поле Data Columns, второе служит целям описания указанных значений для распределения исходной переменной(ых) по переменной, которая введена в поле Break Columns.

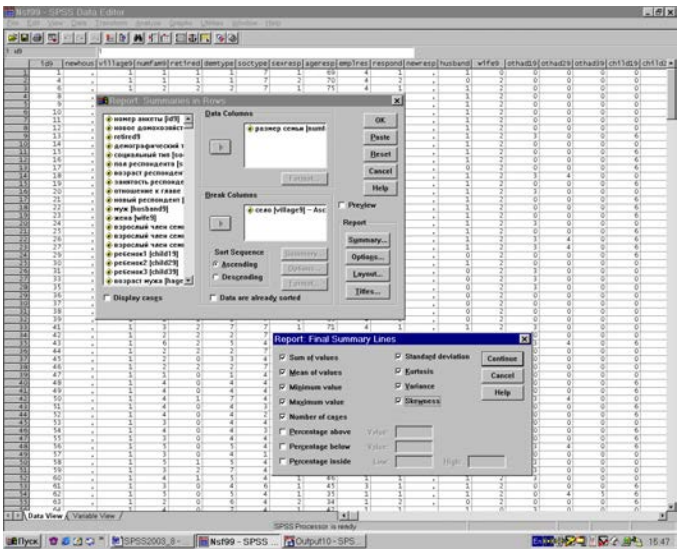



***Рис. 34.***

***Дополнительное диалоговое окно Summary процедуры Report Summaries in Rows***

В последних версиях системы SPSS такого рода досадных совпадений названий команд, окон и опций уже значительно меньше, чем в ранних версиях. Но о ни есть, а поэтому надо быть внимательным пользователем и не пугаться различных трудностей, возникновение которых неизбежно на начальных этапах работы.

На рис. 34 хорошо виден большой набор статистических показателей, которые могут быть получены с помощью дополнительного

диалогового окна Summary. Выбор того или иного показателя предполагает установку опции в поле, находящемся слева от его названия.

**Обрамление 18. Окно просмотра процедуры Report Summaries in Rows**

| Page 1 | | | |
|---|---|---|---|
| село | демографический тип | коровы | свиньи |
| латоново | одиночки | | |
| | Minimum | 0 | 0 |
| | Maximum | 2 | 3 |
| | супружеские пары, пенсионеры | | |
| | Minimum | 0 | 0 |
| | Maximum | 3 | 3 |
| | супружеские пары,работники | | |
| | Minimum | 0 | 1 |
| | Maximum | 4 | 3 |
| | супружеские пары с детьми | | |
| | Minimum | 0 | 0 |
| | Maximum | 5 | 10 |
| | супруж. пары с детьми и родств. | | |
| | Minimum | 0 | 0 |
| | Maximum | 6 | 26 |
| | неполные семьи | | |
| | Minimum | 0 | 0 |
| | Maximum | 2 | 22 |
| | прочие | | |
| | Minimum | 0 | 0 |
| | Maximum | 4 | 10 |
| Minimum | | 0 | 0 |
| Maximum | | 6 | 26 |
| ..... | | | |
| Grand Total | | | |
| Mean | | 1 | 2 |
| Minimum | | 0 | 0 |
| Maximum | | 8 | 26 |

В обрамлении 18, взятом в качестве примера, выведены результаты расчета среднего (Mean), минимального (Minimum) и максимального (Maximum) значений для переменных число коров и свиньей в домохозяйстве. Эти значения получены в разделе итогов (Grand Total) в нижней части окна просмотра.

Обе эти переменные задавались в поле Data Columns, а расчет статистик для них задавался в подокне Report - Summary. Напротив, для переменных село и демографический тип, которые вводились в поле Break Columns, к сожалению, с ущербом для наглядности, были заданы одинаковые статистики минимума и максимума в том же поле Break Columns - Summary. Тем не менее, в обрамлении 18 хорошо видно различие расчетов по каждой переменной. Окно просмотра содержало еще две страницы с расчетами для двух других сел нашего массива, которые убраны из примера в целях экономии места.

Рассматриваемая процедура легко и просто позволяет получать данные, которые практически недоступны для получения с помощью других процедур. В этом плане особенно эффективным является использоваие функций суммирования (Sum), минимального (Minimum) и максимального (Maximum) значений, которые отсутствуют в довольно близкой по духу, но существенно отличающейся содержательно процедуре таблицы сопряженности (Crosstab).

**Правило 22**

***Использование процедуры Report Summaries in Rows позволяет оперативно, в доступном виде и, главное, каждый раз одновременно получать Minimum, Maximum, Mean, Sum и многие другие значения исследуемых переменных.***

## 8.6. Отчет об итогах по столбцам - Report Summaries in Columns

Процедура **Report Summaries in Columns** (отчет об итогах по столбцам), подобно предшествующей процедуре, также позволяет выводить на экран различные характеристики переменных, такие как: минимальное и максимальное значения, средняя, сумма и др. Для выполнения рассматриваемой процедуры в качестве первого шага следует реализовать стандартную последовательность команд:

**Analyze**

**Reports**

**Report Summaries in Columns**.

В ранних версиях здесь использовалась последовательность команд: **Statistics-Summarize-Report Summaries in Columns**. При этом открывается главное диалоговое окно данной процедуры, которое можно видеть на рис. 35.

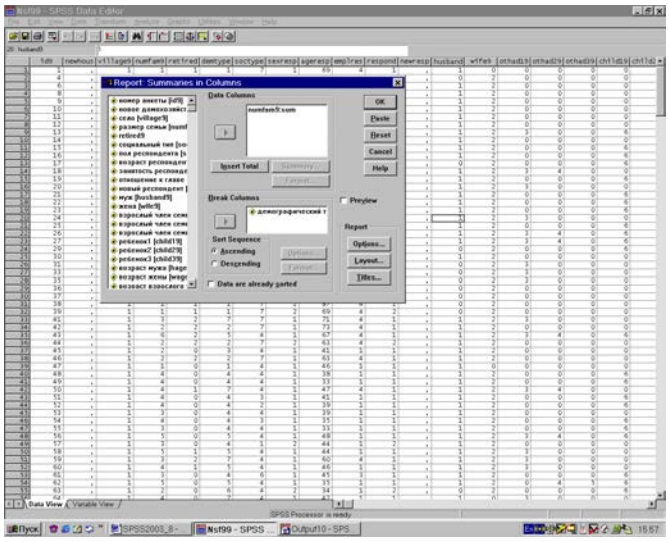



***Рис. 35.***

***Главное диалоговое окно процедуры Report Summaries in Columns - отчет об итогах по столбцам***

Специфика рассматриваемой процедуры состоит в том, что она позволяет одновременно выводить на экран только одно статистическое значение исследуемой переменной (переменных). Для выполнения процедуры числовые переменные надо установить в верхнем центральном поле Data Columns, а требуемый для расчета статистический показатель задать в дополнительном диалоговом окне Summary отдельно для каждой из выбранных переменных (рис. 36).

Как видно на рис. 36, дополнительное диалоговое окно Summary рассматриваемой процедуры практически не отличается от дополнительного диалогового окна предшествующей процедуры (рис. 34). Если на рис. 34 можно видеть уже заданные указания для расчета нескольких статистических показателей, то на рис. 36 возможен и виден только один вариант задания.

В отличие от главного диалогового окна предшествующей процедуры (рис. 33), главное диалоговое окно рассматриваемой процедуры (рис. 35) содержит только одно дополнительное диалоговое окно Summary. Причем доступ к нему открывается только после переноса минимум двух переменных в поле Data Columns.

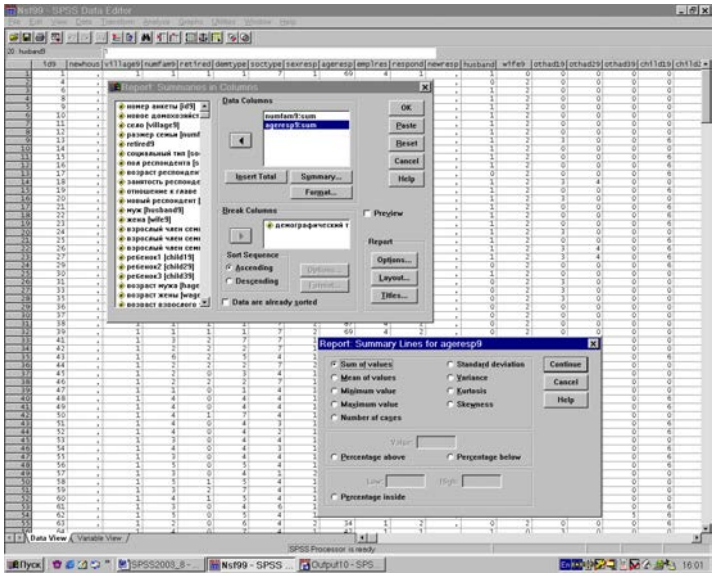


***Рис. 36.***

***Дополнительное диалоговое окно Summary процедуры Report Summaries in Columns***

Последний момент имеет существенное значение, так как в случае помещения переменной(ых) в центральное нижнее поле секции Break Columns именно по ней будут выполняться расчеты для переменной(ых), заданной в соответствующем поле секции Data Columns. Результат таких расчетов и приведен ниже в обрамлении 19.

Особенность вывода данных, приведенных в обрамлении 19, состоит в том, что в нем отсутствует как общее число каждого вида животных в наблюдаемой выборке (по колонке), так и общее число животных в целом в каждом селе (по строке).

**Обрамление 19. Окно просмотра расчета числа свиней и коров в обследуемых селах**

| свиньи | коровы |
|---|---|
| село | Sum | Sum |
| латоново | 347 | 162 |
| венгеровка | 317 | 217 |
| **святцово** | **88** | **196** |

Общее число каждого вида животных в наблюдаемой выборке (по столбцу) будет рассчитано в том случае, если переменные вводятся только в поле секции Data Columns, а соответствующее поле секции Break Columns остается пустым. Результаты расчетов приведены в обрамлении 20.

**Обрамление 20. Окно просмотра расчета числа каждого вида животных в выборке**

| свиньи | коровы |
|---|---|
| Sum | Sum |
| Grand Total | |
| 752 | 575 |

Для получения суммарной численности животных в выборке необходимо использовать команду Insert Total, которая находится в секции Data Columns. При этом ее использование имеет свои весьма специфические особенности. Они состоят в том, что, во-первых, эта команда допустима только в случае, если, как уже отмечалось ранее, в поле Data Columns предварительно введены минимум две переменные.

Во-вторых, при выполнении этой команды в подокне Data Columns под списком введенных ранее переменных появляется только ее имя «Total».

В-третьих, эта команда может быть выполнена лишь в том случае, если в качестве следующего шага опять будет открыто дополнительное диалоговое окно Summary данной процедуры и в нем будет задан расчет соответствующего статистического показателя. Она задается путем переноса переменных, подлежащих расчету, из левого поля Data Columns в правое поле Summary Column.

Наконец, в-четвертых, в поле Summary function дополнительного диалогового окна Summary, которое открывается при появлении в поле Data Columns под списком ранее введенных переменных командного слова Total, необходимо выбрать (рис. 37) одну из предлагаемых возможностей расчета итогов (Total).

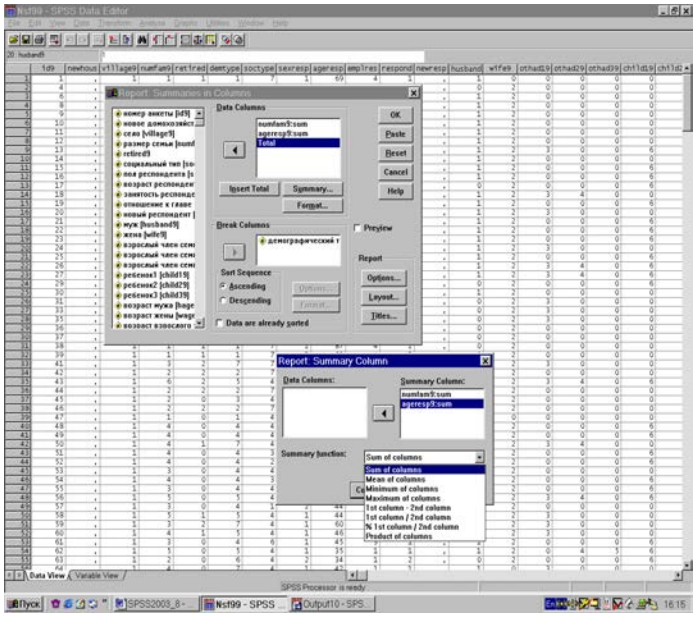
**Обрамление 21. Окно просмотра расчета общего числа животных в выборке**

| свиньи | коровы | |
|---|---|---|
| Sum | Sum | Total |
| Grand Total | | |
| **752** | **575** | **1327** |

По умолчанию в поле Summary function стоит команда Sum of columns, а дополнительно можно выбрать и установить среднюю, минимум и максимум по столбцу и другие статистические показатели. Пример получения в результате выполнения команды Insert Total таких относительно новых данных, которые довольно сложно получить с использованием других процедур, приведен в обрамлении 21.

***Рис. 37.***

***Подокно Summary function дополнительного диалогового окна Summary процедуры Report Summaries in Columns***




В этом случае число 1327 животных в обрамлении 21 и есть итог выполнения команды Insert Total. Не только общее число каждого вида животных в наблюдаемой выборке, равно как и их число в целом по выборке, но и их сумма в каждом селе могут быть рассчитаны, если команда Insert Total будет выполнена при введении в подокно Break Columns соответствующей переменной, обозначающей село (обрамление 22).

**Обрамление 22. Окно просмотра расчета общего числа животных в каждом из обследуемых сел**

| свиньи | коровы | |
|---|---|---|
| село | Sum | Sum | Total |
| латоново | 347 | 162 | 509 |
| венгеровка | 317 | 217 | 534 |
| святцово | 88 | 196 | 284 |

Чисто техническое объединение данных двух последних обрамлений в текстовом процессоре Word позволяет получить таблицу, которая, в полном смысле этого слова, просится в доклад (табл. 5).

**Таблица 5. Общее число животных в каждом селе и в целом по выборке 1999 гг.**

| Село | Свиньи | Коровы | Всего |
|---|---|---|---|
| Латоново | 347 | 162 | 509 |
| Венгеровка | 317 | 217 | 534 |
| Святцово | 88 | 196 | 284 |
| Всего | 752 | 575 | 1327 |

Получить такую таблицу путем использования процедуры Crosstab, выполнив всего две итерации, практически невозможно. А использование для этих целей процедуры Frequencies потребует предварительного трехразового выполнения расчетов с помощью процедуры Select Cases, связанной с отбором случаев, входящих в каждое из трех сел, в качестве независимой выборки (глава 5, § 5.2).

Вполне естественным здесь является вопрос: «В чем конкретно различие процедур Report Summaries in Columns и Report Summaries in Rows»? В обрамлениях 23 и 24 приведены конкретные примеры идентичных расчетов для каждой из двух рассматриваемых процедур.

Для наглядности в обрамлении 24 показатели, которые соответствуют данным обрамления 23 (доля домохозяйств в каждом селе с доходом более 1800 руб. в месяц), выделены жирным шрифтом.

Сравнение данных в обрамлениях 23-24 позволяет сделать вывод о том, что использование процедуры Report Summaries in Columns дает возможность выполнить расчеты, которые для процедуры Report Summaries in Rows в принципе могут рассматриваться как частный случай.

В целом процедуры, объединенные в подмению дескриптивные статистики, важны постоянно: как при подготовке данных к анализу, так и непосредственно при проведении их анализа, подготовке различных аналитических записок и отчетов. Овладение этими процедурами открывает путь ко многим другим видам статистического анализа данных и выполнения процедур непосредственно в системе SPSS.

**Обрамление 23. Окно просмотра процедуры Report Summaries in Columns**

| общий месячный доход | |
|---|---|
| село | > 1800 |
| латоново | 76,4% |
| венгеровка | 74,5% |
| святцово | 52,6% |

**Обрамление 24. Окно просмотра процедуры Report Summaries in Rows**

| село | общий месячный доход |
|---|---|
| латоново | |
| Mean | 3132 |
| **> 1800** | **76,4%** |
| < 900 | 8,3% |
| венгеровка | |
| Mean | 3215 |
| **> 1800** | **74,5%** |
| < 900 | 10,3% |
| святцово | |
| Mean | 2247 |
| **> 1800** | **52,6%** |
| < 900 | 18,8% |
| Grand Total | |
| Mean | 2882 |
| > **1800** | **68,2%** |
| < 900 | 12,3% |

**Правило 23**

***Использование процедуры Report Summaries in Columns позволяет оперативно, в доступном виде, но каждый раз отдельно, получать различные значения (Minimum, Maximum, Mean, Sum и др.) исследуемых переменных.***

## Задание для самостоятельной работы

1. Назовите, пожалуйста, описательные статистики, доступные для расчетов в SPSS.
2. Какие статистические процедуры из меню Reports и Descriptive Statistics рассмотрены в данной главе?

3. Что такое частотный анализ?
4. В чем отличие процедур Frequencies и Descriptives?
5. Как можно построить график в процедуре Frequencies?
6. В чем отличие расчета частот в ранних и поздних версиях SPSS?
7. Какие вы знаете описательные статистики?
8. Какие описательные статистики можно получить в процедуре Descriptives?
9. Как выполняются расчеты в Descriptives?
10. Какие особенности процедуры Explore вы можите назвать?
11. Какие описательные статистики можно получить в процедуре Explore?
12. В чем состоит специфика процедуры Case Summaries?
13. Какая процедура соответствует Case Summaries в ранних версиях SPSS?
14. Чем процедура Report Summaries in Rows отличается от процедуры Report Summaries in Columns?
15. Можно ли построить графики в процедурах меню Reports?
16. В каких процедурах меню Descriptive Statistics можно построить графики, а в каких этого нельзя сделать?
17. В чем специфика процедуры корни OLAP (OLAP Cubes)?
18. Какие дополнительные диалоговые окна имеются в процедуре Frequencies?
19. Что такое нормирование переменных?
20. В чем отличие окон просмотра процедур Frequencies и Descriptives?
21. Что такое описательная диаграмма?
22. Какая процедура SPSS позволяет получить описательную диаграмму?
23. Опишите структуру диаграммы «ствол-лист».
24. Какая процедура позволяет нормировать переменные?
25. Почему в социологии пока еще так редко используются описательные статистики?
26. В чем состоят особенности процедур Report Summaries in Columns и Report Summaries in Rows?

# Глава 9. Таблицы сопряженности

Традиционно в социологии построение и описание таблиц распределения (сопряженности) является одним из основных методов анализа данных выборочных исследований (интервью, опросов и наблюдений). Статистическая таблица - это систематизация обработки исходных данных в особой форме. Текст в таблице сведен к минимуму, а числовые данные объясняются заголовками, подлежащим и сказуемым (31, С. 34, 39).

Любая таблица включает в себя название (общие заголовки) и внутренние заголовки, отражающие содержание строк (подлежащее) и столбцов (сказуемое). Подлежащее - это то, о чем идет речь в таблице. Сказуемое - это совокупность различных показателей, выраженных соответствующими цифровыми данными, которыми характеризуется подлежащее статистической таблицы (31, С.40). Метод группировок и построение статистических таблиц позволяет установить наличие или отсутствие связей между факторами (подлежащее) и результативными признаками (сказуемое), описать обнаруженные связи, а также определить некоторые количественные характеристики (средние и отклонения от них).

В последних версиях SPSS есть два основных пути построения таблиц. Один из них – базовая процедура Crosstabs, другой – подменю Tables.

## 9.1. Построение двухмерных таблиц: процедура Crosstabs

Процедура **Crosstabs** создает таблицы, содержащие частоту встречаемости каждого значения первой переменной (подлежащее) в разрезе второй переменной (сказуемое). Для выполнения процедуры Crosstabs необходимо выбрать в меню:

**Analyze**

**Descriptive Statistics**

**Crosstabs**.

Как и при работе с другими предшествующими процедурами, в ранних версиях SPSS для достижения указанной цели требовалось выполнение последовательности команд: Statistics-Summarize-Crosstabs. После выполнения указанных последовательностей команд откроется диалоговое окно Crosstabs (рис. 38).

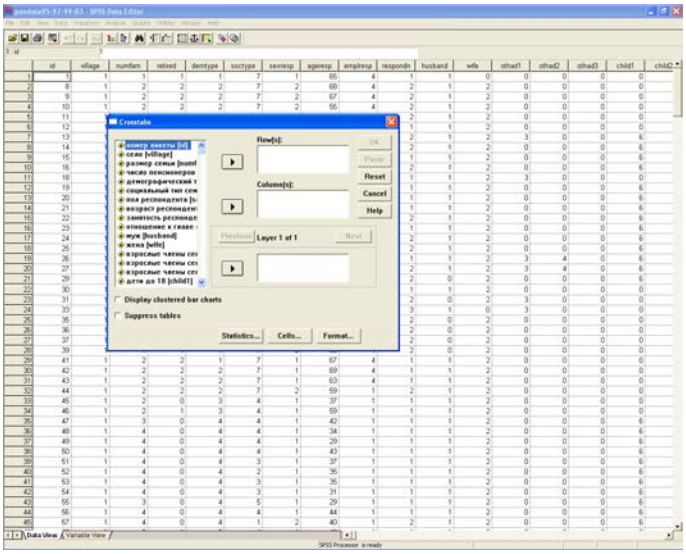


***Рис. 38.***

***Главное диалоговое окно процедуры Crosstabs***

В этом окне слева стоит список переменных, из которого и выбираются переменные, необходимые для построения таблицы. Выбор переменной осуществляется путем выделения в списке ее имени (установкой мышью курсора на имени и нажатием ее левой клавиши - клик).

На следующем шаге выделенная переменная, путем нажатия одной из двух кнопок «стрелка вправо», переносится в поле **Rows** (строки) или **Columns** (столбцы). Здесь и возникает содержательная исследовательская задача: какая переменная в таблице будет записана в качестве подлежащего, а какая в качестве сказуемого. Ее решение предполагает наличие предварительных гипотез и знание характеристик первичных распределений анализируемого массива данных.

Возврат переменных из полей Rows и Columns назад в список переменных выполняется в обратном порядке. Переменная, находящаяся в любом поле, выделяется. При этом соответствующая кнопка - стрелка вправо уже изменила ориентацию и приобрела вид «стрелка влево». Выделенная переменная из поля для анализа возвращается в список переменных путем нажатия данной кнопки. Без выполнения

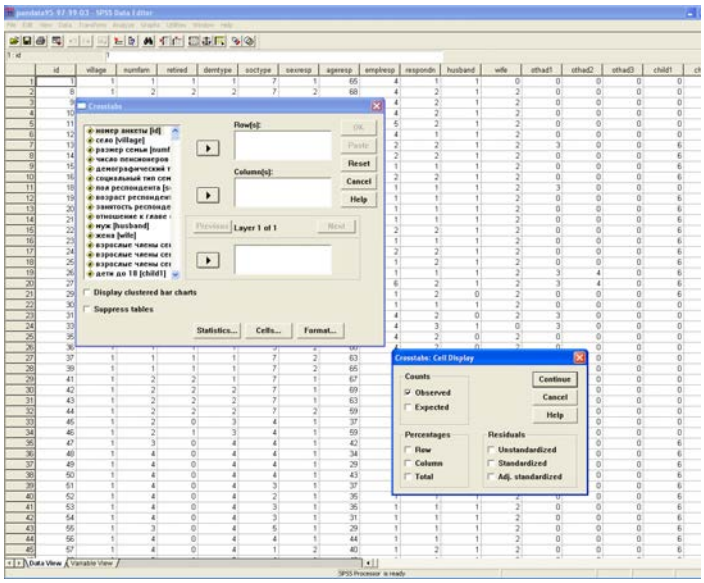
описанной выше команды, переменная, введенная в поле анализа, будет там сохраняться до конца текущего сеанса работы в SPSS.

Как видно на рис. 38, главное диалоговое окно рассматриваемой процедуры содержит в средней части еще один блок управления (Previous Layer 1 of 1 Next), который используется при построении многомерных таблиц и будет рассмотрен в следующем параграфе.

Для задания статистик, например, вычисления процентов по строкам и столбцам таблицы, используется кнопка **Cells** (проценты), рис. 39, открывающая дополнительное диалоговое окно рассматриваемой процедуры. Работа в любом из дополнительных окон завершается нажатием кнопки **Continue**. Выполнение этой команды ведет к возврату в главное окно процедуры, в котором после нажатия кнопки OK процедура начнет выполняться, а результаты появятся в окне просмотра.

***Рис. 39.***

***Дополнительное диалоговое окно Cells процедуры Crosstabs***




Пример двухмерной таблицы из окна просмотра, которая построена по переменным «село» и «пол респондента» с использованием дополнительного окна Cells (итоговые проценты), представлен в обрамлении 25. Таблица, полученная в окне просмотра, содержит много информации, в которой сразу разобраться довольно сложно.

Первая верхняя строка - это заголовок таблицы. Совершенно ясно, что в таком виде он может быть использован только для аналитических (внутренних) целей, но не для передачи заказчику или публикации.

Его более или менее приемлемая интерпретация на русском языке дана в заголовках табл. 6-7.

**Обрамление 25. Окно просмотра процедуры Crosstabs**

**село * пол респондента Crosstabulation**

| | | | пол респондента | | |
|---|---|---|---|---|---|
| | | | мужчины | женщины | Total |
| село | латоново | Count | 56 | 76 | 132 |
| | | % within село | 42,4% | 57,6% | 100,0% |
| | | % within пол респондента | 57,1% | 26,8% | 34,6% |
| | | % of Total | 14,7% | 19,9% | 34,6% |
| | венгеровка | Count | 21 | 110 | 131 |
| | | % within село | 16,0% | 84,0% | 100,0% |
| | | % within пол респондента | 21,4% | 38,7% | 34,3% |
| | | % of Total | 5,5% | 28,8% | 34,3% |
| | святцево | Count | 21 | 98 | 119 |
| | | % within село | 17,6% | 82,4% | 100,0% |
| | | % within пол респондента | 21,4% | 34,5% | 31,2% |
| | | % of Total | 5,5% | 25,7% | 31,2% |
| Total | | Count | 98 | 284 | 382 |
| | | % within село | 25,7% | 74,3% | 100,0% |
| | | % within пол респондента | 100,0% | 100,0% | 100,0% |
| | | % of Total | 25,7% | 74,3% | 100,0% |

Исходя из записи заголовка таблицы в окне просмотра (обрамление 25), можно сформулировать общее представление о формате записи заголовков таблицы в SPSS. Такое общее представление кратко описано в правиле 24:

**Правило 24**

***Формат записи заголовка таблицы в SPSS выглядит следующим образом: «Сопряженность подлежащего и сказуемого».***

Таким образом, в нашем примере может быть записан следующий заголовок таблицы: «Распределение пола опрошенных по селу».

Фактически рассмотренная нами двумерная таблица позволяет вести анализ по разным основаниям. Используя проценты по колонке, можно, например, построить табл. 6.

**Таблица 6. Распределение пола опрошенных по селу**

| Пол | Мужской | | Женский | |
|---|---|---|---|---|
| | Абс. чис. | % | Абс. чис. | % |
| Латоново | 56 | 57,2 | 76 | 26,8 |
| Венгеровка | 21 | 21,4 | 110 | 38,7 |
| Святцово | 21 | 21,4 | 98 | 34,5 |
| Итого | 98 | 100 | 284 | 100 |

Табл. 6 характеризует абсолютное и относительное соотношение респондентов по полу в выборке в 2003 г. Анализ табл. 6 показывает, что на село Латоново приходится больше всего опрошенных мужчин (56 из 98 чел. или 57,2% общего числа респондентов - мужчин). В то время как представительность женщин - респондентов в трех деревнях распределена равномерно.

Напротив, используя проценты по строке, можно построить табл.7.

**Таблица 7. Распределение опрошенных мужчин и женщин по селу**

| Пол | Мужской | | Женский | | Всего | |
|---|---|---|---|---|---|---|
| | Абс. чис | % | Абс. чис. | % | Абс. чис. | % |
| Латоново | 56 | 42,4 | 76 | 57,6 | 132 | 100 |
| Венгеровка | 21 | 16,0 | 110 | 84,0 | 131 | 100 |
| Святцово | 21 | 17,6 | 98 | 82,4 | 119 | 100 |

Табл. 7 отличается от табл. 6 тем, что в ней анализ идет по строке, а не по столбцу, как в предыдущем случае. Здесь уже важны различия по полу в каждом селе. Для представленных в обрамлении 25 расчетов построенные таблицы не являются исчерпывающими.

Хотя табл. 6 и табл. 7 выглядят весьма сходным образом, в действительности они имеют в качестве объектов анализа и рассмотрения совершенно разные признаки. В первом случае - это сам пол (мужской или женский), а во втором – соотношение мужчин и женщин в каждом селе. Строя, например, «распределение пола опрошенных по селу»,

пользователь находится в сфере анализа гендерных отношений. В то же время, строя «распределение сел по полу», он изучает особенности сельского расселения.

Из приведенных примеров видно, что SPSS позволяет строить таблицы, рассчитывая удельные веса по строке и столбцу одновременно. Система предлагает самим пользователям принять решение о том, что им необходимо получить в результате расчетов.

**Правило 25**

***Использование процедуры Crosstabs предполагает повышенное внимание пользователя при установке задания вычисляемых процентов. Поведение по принципу «задай все, потом разберемся» связано с пустой тратой времени и сил самого пользователя и ресурсов PC.***

Кроме кнопки Cells диалоговое окно Crosstabs содержит еще две кнопки, открывающие дополнительные диалоговые окна: Statistics с набором возможностей расчета различных статистик и Format для выбора порядка вывода данных.

Расчет различных статистик сопряженности, а именно: коэффициентов связи и корреляции признаков, составляющих таблицу, - возможен с помощью дополнительного диалогового окна Statistics. Оно и показано на рис. 40.

Это окно имеет сходную структуру с одноименным окном в описанной ранее процедуре Frequencies (глава 1, § 1.3 и глава 8, § 8.1). Но если окно Statistics в процедуре Frequencies предназначено для расчета частотных статистик (дисперсии, средних, асимметрии), то окно Statistics в процедуре Crosstabs предназначено для описания мер связи ($\chi^2$, корреляции).

Рассматриваемое окно имеет три основные поля. Их использование обусловлено типом переменных, составляющих таблицу. В этом плане базовая версия SPSS 6.1 предлагает следующие возможности расчета различных коэффициентов: для номинальных чисел (Nominal Data: Contingency coefficient, Phi and Gramer's V, Lambda, Uncertainly coefficient) - центральное левое поле; для порядковых чисел (Ordinal Data: Gamma, Somers'd, Kendall's tau-b, Kendall's tau-c) - центральное правое поле. И для номинальных чисел, сопряженных в таблице с интервальными числами, - нижнее левое поле (Nominal by Interval: Eta).

В качестве отдельных опций для выполнения специальных расчетов предлагаются $\chi^2$ (Chi-square) - верхнее левое поле, корреляция (Correlation) - верхнее правое поле, а также каппа (Kappa) и риск (Risk) - два поля, расположенные внизу справа. Последние версии системы имеют тенденцию к расширению набора предлагаемых для расчетов статистик мер связи, но порядок работы при этом всегда остается неизменным. Например, 10-я версия SPSS позволяет дополнительно вычислять: McNemar и др.

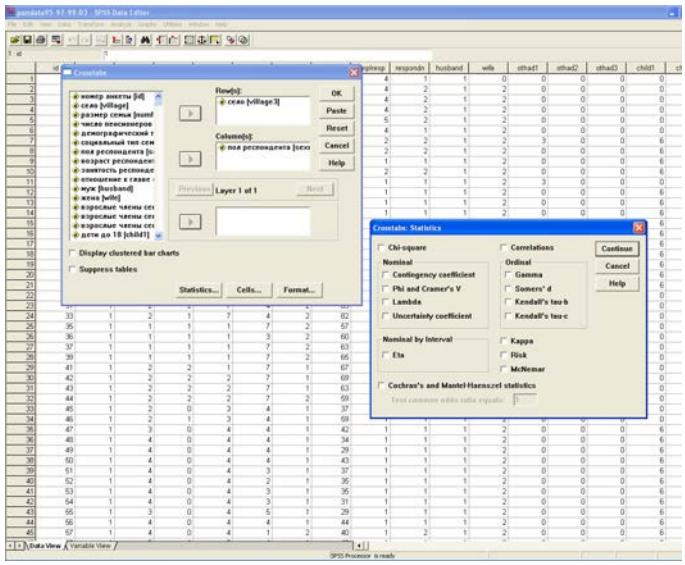


***Рис. 40.***

***Дополнительное диалоговое окно Statistics процедуры Crosstabs***

Выбор для расчета того или иного коэффициента осуществляется путем указания курсором в маленьком квадрате (box), стоящем слева от имени каждого коэффициента. При выполнении процедуры Crosstabs все статистики, получаемые с помощью дополнительного диалогового окна процедуры Statistics, выводятся сразу и непосредственно внизу окна просмотра под выведенной таблицей.

Необходимость анализа связей в таблице сопряженности показана ниже на конкретном примере. В табл. 8 приведено распределение числа выкармливаемых в домохозяйстве свиней по размеру семьи.

Эта таблица содержит абсолютные данные, которые весьма сложно поддаются интерпретации и, как правило, не приводятся в отчетах. Наиболее распространенным в социологии является представление данных в таблицах в относительных величинах.

**Таблица 8. Число свиней в домохозяйстве по размеру семьи в выборке 1999 г.**

| Число членов семьи | Число свиней в домохозяйстве | | | | | Всего |
|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 и > | |
| 1 | 73 | 9 | 5 | 1 | 0 | 88 |
| 2 | 42 | 28 | 16 | 8 | 10 | 100 |
| 3 | 20 | 14 | 22 | 11 | 19 | 86 |
| 4 | 9 | 22 | 25 | 11 | 13 | 80 |
| 5 и > | 11 | 8 | 14 | 14 | 17 | 68 |
| Всего | 155 | 81 | 82 | 45 | 59 | 422 |

В таком случае сопряженность признаков, включенных в таблицу, просматривается вполне определенно и нагляднее (табл. 9). В примере, приведенном в табл. 9, видно, что почти половина (47,1%) домохозяйств, в которых нет поросят, приходятся на одиночек.

**Таблица 9. Структура распределения выкармливаемых в домохозяйстве свиней по размеру семьи (в %)**

| Число членов семьи | Число свиней | | | | | Всего |
|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 и > | |
| 1 | 47.1 | 11.1 | 6.1 | 2.2 | 0.0 | 20.9 |
| 2 | 27.1 | 34.6 | 19.5 | 17.8 | 10.2 | 23.7 |
| 3 | 12.9 | 17.3 | 26.8 | 24.4 | 32.2 | 20.4 |
| 4 | 5.8 | 27.2 | 30.5 | 24.4 | 22.0 | 19.0 |
| 5 и > | 7.1 | 9.9 | 17.1 | 31.1 | 35.6 | 16.1 |
| Всего | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |

Напротив, наличие четырех и более свиней (почти в 90% домохозяйств) отмечается в семьях, состоящих из трех и более человек. Можно сказать, что и в том, и другом случае связь просматривается наглядно, но и при расчете относительных величин значения мер связи остаются недоступными.

Обрамления 26 и 27 позволяют увидеть как различия мер связи, которые рассчитываются для номинальных («распределение пола опрошенных по селу») и порядковых переменных («распределение числа свиней в домохозяйстве по размеру семьи»), так и их значения. Расчет выполнен по выборке 1999 г.

## Обрамление 26. Окно просмотра процедуры Crosstabs

**село * пол респондента Crosstabulation**

Count

| | | пол респондента | | |
|---|---|---|---|---|
| | | мужчины | женщины | Total |
| село | латоново | 58 | 86 | 144 |
| | венгеровка | 25 | 120 | 145 |
| | святцево | 23 | 110 | 133 |
| Total | | 106 | 316 | 422 |

**Directional Measures**

| | | | Value | Asymp. Std. Error[a] | Approx. T[b] | Approx. Sig. |
|---|---|---|---|---|---|---|
| Nominal by Nominal | Lambda | Symmetric | ,086 | ,022 | 3,680 | ,000 |
| | | село Dependent | ,119 | ,031 | 3,680 | ,000 |
| | | пол респондента Dependent | ,000 | ,000 | ,[c] | ,[c] |
| | Goodman and Kruskal tau | село Dependent | ,032 | ,013 | | ,000[d] |
| | | пол респондента Dependent | ,063 | ,025 | | ,000[d] |
| | Uncertainty Coefficient | Symmetric | ,037 | ,014 | 2,527 | ,000[e] |
| | | село Dependent | ,028 | ,011 | 2,527 | ,000[e] |
| | | пол респондента Dependent | ,054 | ,021 | 2,527 | ,000[e] |

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

c. Cannot be computed because the asymptotic standard error equals zero.

d. Based on chi-square approximation

e. Likelihood ratio chi-square probability.

**Symmetric Measures**

| | | Value | Approx. Sig. |
|---|---|---|---|
| Nominal by Nominal | Phi | ,252 | ,000 |
| | Cramer's V | ,252 | ,000 |
| | Contingency Coefficient | ,244 | ,000 |
| N of Valid Cases | | 422 | |

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

Например, практически все меры связи в обрамлении 26 указывают на значимость связи. Для всех мер связи она (Approx.Sig.) составляет , 000.

# Обрамление 27. Окно просмотра процедуры Crosstabs

**размер семьи * 'сколько свиней' Crosstabulation**

Count

| | | 'сколько свиней' | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | нет | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Total |
| размер семьи | 1 | 66 | 11 | 6 | | | | | | 83 |
| | 2 | 39 | 45 | 19 | 4 | 1 | | | | 108 |
| | 3 | 16 | 14 | 23 | 6 | 8 | 1 | | | 68 |
| | 4 | 8 | 29 | 34 | 6 | 10 | 7 | 1 | 1 | 96 |
| | 5 | 9 | 5 | 12 | 7 | 5 | 2 | | 1 | 41 |
| | 6 | 2 | 5 | 6 | 1 | 5 | | | | 19 |
| | 7 | | | 4 | 2 | | | | | 6 |
| | 9 | 1 | | | | | | | | 1 |
| Total | | 141 | 109 | 104 | 26 | 29 | 10 | 1 | 2 | 422 |

**Chi-Square Tests**

| | Value | df | Asymp. Sig. (2-sided) |
|---|---|---|---|
| Pearson Chi-Square | 202,482[a] | 49 | ,000 |
| Likelihood Ratio | 206,287 | 49 | ,000 |
| Linear-by-Linear Association | 91,928 | 1 | ,000 |
| N of Valid Cases | 422 | | |

a. 42 cells (65,6%) have expected count less than 5. The minimum expected count is ,00.

**Directional Measures**

| | | | Value | Asymp. Std. Error[a] | Approx. T[b] | Approx. Sig. |
|---|---|---|---|---|---|---|
| Ordinal by Ordinal | Somers' d | Symmetric | ,444 | ,032 | 13,808 | ,000 |
| | | размер семьи Dependent | ,460 | ,033 | 13,808 | ,000 |
| | | 'сколько свиней' Dependent | ,429 | ,031 | 13,808 | ,000 |

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

**Symmetric Measures**

| | | Value | Asymp. Std. Error[a] | Approx. T[b] | Approx. Sig. |
|---|---|---|---|---|---|
| Ordinal by Ordinal | Kendall's tau-b | ,444 | ,032 | 13,808 | ,000 |
| | Kendall's tau-c | ,395 | ,029 | 13,808 | ,000 |
| | Gamma | ,551 | ,038 | 13,808 | ,000 |
| | Spearman Correlation | ,532 | ,037 | 12,865 | ,000[c] |
| Interval by Interval | Pearson's R | ,467 | ,037 | 10,832 | ,000[c] |
| N of Valid Cases | | 422 | | | |

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

c. Based on normal approximation.

Вместе с тем сопряженность рассматриваемых признаков очень слабая. На это указывают значения показателя Value, которые не поднимаются выше 0,119. В то же время при высокой сопряженности признаков они должны были бы стремиться к значению 0,999.

Напротив, практически все меры связи в обрамлении 27 показывают высокую степень сопряженности между размером сельской семьи (число членов в семье) и числом выкармливаемых в домохозяйстве свиней в 1999 г.

В статистике $\chi^2$ (Chi-Square) значение (Value), равное 202,482 для числа степеней свободы (df) 49, соответствует наблюдаемому уровню значимости (Sig.) меньшему, чем ,000.

Это значит, что в данном случае нулевая гипотеза для критерия $\chi^2$ Пирсона отвергается. И напротив, можно утверждать, что связь между размером семьи и числом выкармливаемых в домохозяйстве свиней существует.

Значения симметричных мер связи, а именно коэффициентов (Somers'd Symmetric) и Kendall's tau-b совпадают и составляют 0,444. В то же время значение коэффициента корреляции Спирмена (Spearman Correlation) еще выше и составляет 0,532. Все это позволяет говорить о достаточно высокой степени связи. Такой вывод оказывается возможным на основании того, что указанные меры связи принимают значения от 0 до 1. В приведенном нами примере, это фактически означает, что в 5 случаях из 10 число свиней в домохозяйстве сопряжено с числом членов в сельской семье.

**Правило 26**

***Использование процедуры Crosstabs предполагает понимание содержательного смысла мер связи и ограничительной роли типа переменной при выполнении статистических расчетов.***

## 9.2. Таблицы большей размерности

В SPSS логика построения и представления многомерных таблиц имеет свою специфику. Она несколько отлична от традиционно устоявшегося в социологии порядка разработки таблиц, когда наращивание размерности таблицы достигается путем наглядного

разбиения подлежащего и сказуемого на дополнительные признаки. А сама таблица по мере наращивания размерности приобретает все более сложный и трудно читаемый характер.

Разработчики SPSS пошли в этом случае несколько иным путем. Они создали алгоритм, позволяющий преобразовать двухмерную таблицу в таблицу любой размерности.

**Правило 27**

***Особенность SPSS состоит в том, что в нем таблица любой размерности первично представляется как двухмерная таблица.***

Многомерная таблица исходно начинает задаваться как двухмерная, порядок построения которой описан в предшествующем параграфе. Вместе с тем, как отмечалось ранее, главное диалоговое окно процедуры Crosstabs (§ 9.1, рис. 38) содержит в средней части специальный блок (Previous Layer 1 of 1 Next), который и используется при построении многомерных таблиц.

Порядок построения трехмерной таблицы выглядит следующим образом: сначала из списка переменных задаются подлежащее (строка) и сказуемое (столбец) требуемой таблицы (рис. 38), затем в блок (Previous Layer 1 of 1 Next) вводится управляющая переменная (Layer). Для этого из списка переменных с помощью нижней кнопки («стрелка вправо») в поле данного блока переносится выделенная переменная. При этом открывается доступ к выключателю Next.

В случае необходимости повышения размерности таблицы, каждая новая управляющая переменная вводится путем предварительного нажатия выключателя Next. После введения управляющих переменных, при выполнении команды ОК, в окне просмотра будут получены двухмерные таблицы для каждого значения управляющей переменной.

В качестве примера посмотрим, как выглядит в окне просмотра SPSS трехмерная таблица, в которой в обследуемых селах (по строке) фиксируются задержки в выплате заработной платы и пенсии (управляющая переменная) по полу опрашиваемых (столбец) в выборке 1999 г. (обрамление 28).

## Обрамление 28. Окно просмотра трехмерной таблицы сопряженности

**село * пол респондента * 'задержки в выплате зарплаты,пенсии' Crosstabulation**

| 'задержки в выплате зарплаты,пенсии' | | | | пол респондента | | Total |
|---|---|---|---|---|---|---|
| | | | | мужчины | женщины | |
| да | село | латоново | Count | 42 | 54 | 96 |
| | | | % within село | 43,8% | 56,3% | 100,0% |
| | | | % within пол респондента | 53,2% | 21,7% | 29,3% |
| | | | % of Total | 12,8% | 16,5% | 29,3% |
| | | венгеровка | Count | 18 | 99 | 117 |
| | | | % within село | 15,4% | 84,6% | 100,0% |
| | | | % within пол респондента | 22,8% | 39,8% | 35,7% |
| | | | % of Total | 5,5% | 30,2% | 35,7% |
| | | святцево | Count | 19 | 96 | 115 |
| | | | % within село | 16,5% | 83,5% | 100,0% |
| | | | % within пол респондента | 24,1% | 38,6% | 35,1% |
| | | | % of Total | 5,8% | 29,3% | 35,1% |
| | Total | | Count | 79 | 249 | 328 |
| | | | % within село | 24,1% | 75,9% | 100,0% |
| | | | % within пол респондента | 100,0% | 100,0% | 100,0% |
| | | | % of Total | 24,1% | 75,9% | 100,0% |
| нет | село | латоново | Count | 16 | 32 | 48 |
| | | | % within село | 33,3% | 66,7% | 100,0% |
| | | | % within пол респондента | 59,3% | 47,8% | 51,1% |
| | | | % of Total | 17,0% | 34,0% | 51,1% |
| | | венгеровка | Count | 7 | 21 | 28 |
| | | | % within село | 25,0% | 75,0% | 100,0% |
| | | | % within пол респондента | 25,9% | 31,3% | 29,8% |
| | | | % of Total | 7,4% | 22,3% | 29,8% |
| | | святцево | Count | 4 | 14 | 18 |
| | | | % within село | 22,2% | 77,8% | 100,0% |
| | | | % within пол респондента | 14,8% | 20,9% | 19,1% |
| | | | % of Total | 4,3% | 14,9% | 19,1% |
| | Total | | Count | 27 | 67 | 94 |
| | | | % within село | 28,7% | 71,3% | 100,0% |
| | | | % within пол респондента | 100,0% | 100,0% | 100,0% |
| | | | % of Total | 28,7% | 71,3% | 100,0% |

В этом обрамлении видно, что система объединила в окне просмотра две отдельные двухмерные таблицы.Эти таблицы построены по значениям «да» и «нет» управляющей переменной «задержки в выплате заработной платы/пенсии».

В одной из них (верхняя часть таблицы окна просмотра) выполнены расчеты для случаев, связанных с задержкой выплат (ответ «да»), в другой (нижняя часть таблицы окна просмотра) - для случаев, связанных с их отсутствием (ответ «нет»).

В табл. 10 приведены данные обрамления 28, преобразованные в текстовом процессоре Word. Данные, приведенные в обрамлении 28 и табл.10, имеют различный вид. Во-первых, данные окна просмотра содержат проценты по строке, столбцу, итогу по селу и двум значениям управляющей переменной. В то же время в табл. 10 представлены проценты задержки в выплате заработной платы/пенсии и полу (столбцы) по селу. Можно сказать, что данные табл. 10 представляют собой частный случай данных, содержащихся в обрамлении 28. Во-вторых, как было сказано выше, таблица, выведенная в окне просмотра (обрамление 28), состоит из двух двухмерных таблиц. Одна из них - село по полу для тех, у кого есть задержки с выплатой заработной платы или пенсии, а другая - село по полу для тех, у кого таких задержек нет.

**Таблица 10. Задержки в выплате зарплаты и пенсии в обследуемых селах по полу опрошенных**

| Село | Задержки в выплате заработной платы/пенсии | | | |
|---|---|---|---|---|
| | Есть | | Нет | |
| | Муж. | Жен. | Муж. | Жен. |
| Латоново | 53,2 | 21,7 | 59,3 | 47,8 |
| Венгеровка | 22,7 | 39,8 | 25,9 | 31,3 |
| Святцово | 24,1 | 38,5 | 14,8 | 20,9 |
| Итого | 100,0 | 100,0 | 100,0 | 100,0 |

Примечание: В таблице отдельные десятые доли процента округлены.

При построении таблиц большей размерности описанный выше порядок работы сохраняется. Для этого, после ввода расчетных переменных в строку и столбец, следует задать дополнительно две управляющих переменных (для четырехмерной таблицы), три (для пятимерной таблицы) и т.д. Двухмерные таблицы в этом случае выдаются для каждой комбинации значений управляющей переменной в поряд

ке, обратном тому, как они задавались. Понятно, что непосредственно в таком виде таблицы можно использовать в анализе, но они не украсят отчет или публикацию.

Для задания двух и более управляющих переменных, как отмечалось ранее, используется выключатель Next. При этом в рабочее состояние приходит выключатель Previous**,** позволяющий возвращаться к ранее введенным управляющим переменным.

В последних версиях SPSS идеология построения многомерных таблиц, используемая в процедуре Crosstabs, начала ревизоваться в процедурах команды Tables.

**Правило 28**

***В процедуре Crosstabs при построении многомерных таблиц используется специальный блок задания управляющих переменных - Layer.***

## 9.3. Порядок построения таблиц в меню Tables

В последних версиях SPSS в меню Analyze появилась команда **Tables**, позволяющая выполнять расчеты для большой группы таблиц: **Basic Tables** (простые таблицы), **Custom Tables** (пользовательские таблицы), **General Tables** (общие таблицы), **Multiply Response Tables** (таблицы множественных ответов, **Tables of Frequencies** (таблицы частот) и другие. Одна из основных особенностей выполнения расчетов с использованием команды Tables – хорошее качество вывода таблиц в окне просмотра. В обрамлениях 29-30 показаны результаты расчетов, которые выполнены для тех же переменных, что и в обрамлении 25, с помощью двух процедур: простых и общих таблиц.

В обрамлении 31 показана многомерная пользовательская таблица. Как видно из обрамлений 29-31 в результате выполненных расчетов в окне просмотра по всем трем процедурам выведены сходные таблицы. И все они заметно отличаются от такой же таблицы в обрамлении 25. Процедуры расчета пользовательских и простых таблиц теперь уже сделаны для работы в интерактивном режиме. Это и позволило построить в одной из них трехмерную таблицу, которая выведена в обрамлении 31. Не только внешний вид, но и логика построения таких таблиц заметно отличаются от тех, что используются в процедуре Crosstabs.

Порядок работы с интерактивными процедурами, за которыми, несомненно, будущее, описан в главе 12, § 12.4 на примере построения интерактивных графиков.

**Обрамление 29. Окно просмотра Basic Tables: село (Down), пол (Across).** Использованы опции: Titles, Totals – Table-margin totals.

**Село по полу**

| | | пол респондента | | |
|---|---|---|---|---|
| | | мужчины | женщины | Table Total |
| село | латоново | 56 | 76 | 132 |
| | венгеровка | 21 | 110 | 131 |
| | святцево | 21 | 98 | 119 |
| Table Total | | 98 | 284 | 382 |

**Обрамление 30.Окно просмотра General Tables: село (Rows), пол (Columns).** Использованы опции: Insert Total, Titles

**Село по полу**

| | | пол респондента | | |
|---|---|---|---|---|
| | | мужчины | женщины | Total |
| село | латоново | 56 | 76 | 132 |
| | венгеровка | 21 | 110 | 131 |
| | святцево | 21 | 98 | 119 |
| Total | | 98 | 284 | 382 |

**Обрамление 31. Окно просмотра Custom Tables: Демографический тип (Rows), село и пол (Columns).** Приняты все установки умолчания. Дополнительно использована опция Hide

| | | село | | | | | |
|---|---|---|---|---|---|---|---|
| | | латоново | | венгеровка | | святцево | |
| | | пол респондента | | пол респондента | | пол респондента | |
| | | мужчины | женщины | мужчины | женщины | мужчины | женщины |
| демографический тип семьи | одиночки | 3 | 15 | 1 | 19 | 3 | 22 |
| | супр.пары пенсионеров | 6 | 7 | 3 | 8 | 3 | 8 |
| | супр. пары работников | 6 | 2 | 3 | 5 | 4 | 12 |
| | супр.пары с детьми | 27 | 28 | 8 | 35 | 5 | 29 |
| | супр.пары сдетьми и др.родственниками | 10 | 7 | 2 | 19 | 0 | 6 |
| | неполные семьи | 1 | 5 | 0 | 2 | 1 | 1 |
| | прочие | 3 | 12 | 4 | 22 | 5 | 20 |

Еще более радикально в Tables изменяется порядок расчета частотных таблиц (Tables of Frequencies). При первом соприкосновении использование поля Frequencies for повышает качество выводимой в окне просмотра частотной таблицы, но не ведет к изменению содержания. Ситуация резко меняется, когда используются поля обрамления Subgroups (In each Table и Separate Tables). Пример расчета частот для демографического типа семьи (поле Frequencies for) по селам (поле In each Table) приведен на табличной вставке 1.

**Вставка 1. Частотная таблица (Tables of Frequencies)**

| | село | | | | | |
|---|---|---|---|---|---|---|
| | латоново | | венгеровка | | святцово | |
| | демографический тип семьи | | демографический тип семьи | | демографический тип семьи | |
| | Count | % | Count | % | Count | % |
| одиночки | 18 | 13,6% | 20 | 15,3% | 25 | 21,0% |
| супр.пары пенсионеров | 13 | 9,8% | 11 | 8,4% | 11 | 9,2% |
| супр. пары работников | 8 | 6,1% | 8 | 6,1% | 16 | 13,4% |
| супр.пары с детьми | 55 | 41,7% | 43 | 32,8% | 34 | 28,6% |
| супр.пары сдетьми и др.родственниками | 17 | 12,9% | 21 | 16,0% | 6 | 5,0% |
| неполные семьи | 6 | 4,5% | 2 | 1,5% | 2 | 1,7% |
| прочие | 15 | 11,4% | 26 | 19,8% | 25 | 21,0% |

Из приведенного примера видно. что в результате изменения технологии и порядка расчетов в Tables теперь уже частотная таблица внешне мало чем отличается от таблицы сопряженности, порядок построения которой описан ранее в настоящей главе. Напротив, полученная таким образом частотная таблица имеет существенные различия по сравнению с таблицей частот, расчет которой выполнен с использованием процедуры Frequencies (глава 8, § 8.1).

Разумеется, каждая из процедур Tables имеет свои особенности, позволяя строить самые разнообразные таблицы. Освоение некоторых из них связано с известными трудностями. Их преодоление обязательно для каждого, кто хочет работать регулярно и самостоятельно в SPSS.

## Задание для самостоятельной работы

1. Что такое таблица сопряженности?
2. Какая команда главного меню SPSS позволяет работать с таблицами?
3. Как строится двухмерная таблица?
4. Как строится многомерная таблица?
5. Опишите главное диалоговое окно процедуры Crosstabs.
6. Какие статистики могут быть рассчитаны в процедуре Crosstabs?
7. В чем особенность записи заголовка таблицы в окне вывода SPSS?
8. Что такое Layer?
9. Зачем при построении таблиц нужны управляющие переменные?
10. Какие правила, связанные с расчетом таблиц сопряженности, вы знаете?
11. В чем состоит особенность SPSS при выводе таблиц?
12. Зачем нужны кнопки Next и Previous в процедуре Crosstabs?
13. Какие статистики допустимы при расчете связи номинальных чисел в таблице сопряженности?
14. Какие таблицы строятся с использованием процедур команды Tables?
15. Постройте общую, пользовательскую и простую таблицы.
16. В чем состоит особенность построения простой таблицы (Basic Tables) в подменю Tables?
17. В чем состоит особенность построения общей таблицы (General Tables) в подменю Tables?
18. В чем состоит особенность построения пользовательской таблицы (Custom Tables) в подменю Tables?
19. В чем состоит особенность построения частотной таблицы (Tables of Frequencies) в подменю Tables?
20. Как интерактивные процедуры отличаются от традиционных процедур SPSS?

# Глава 10. Меры сравнения

## 10.1. Характеристика мер сравнения

В практике научных исследований и разработок обычно фиксируются две основные черты зависимости между переменными. Одна из них – величина зависимости, а другая – надежность зависимости (32). Измерение величины зависимости нашло широкое распространение в социологии, в то же время вопрос о надежности каждой конкретной зависимости либо не обсуждается, либо как бы снимается при обсуждении репрезентативности выборочного исследования в целом.

Такой исследовательский опыт часто ведет к далеким от корректности выводам и обобщениям. Тем не менее, его стабильность и воспроизводимость во времени имеет под собой весьма сложную констелляцию обстоятельств, из которых можно выделить минимум три наиболее веских.

Первое обстоятельство связано с постоянно обсуждаемым в социологии вопросом об ограниченности измерения. Указанная ограниченность обусловлена преобладанием в социологических исследованиях номинальных чисел (33). Признавая обоснованность этого обстоятельства, следует отметить, что оно не абсолютно. Числовые переменные – непрерывные, дискретные, порядковые, также как и номинальные, имеют широкое распространение в описании социальных характеристик (3, С.14). Возраст, доход, размер семьи, жилища и сбережений – все это лишь отдельные примеры таких характеристик.

Второе обстоятельство связано с известной новизной социологических данных и, как следствие, ограниченностью их использования в повседневной практике управления социальными процессами. Отсюда первичность спроса на сами данные и вторичность спроса на их надежность и обоснованность. В этом плане российская социология еще весьма далека, скажем, от прикладной экономики и эконометрики. Например, приведенная ниже табл. 11 с фиксацией характеристик надежности измерения – уже норма для американской социологии и абсолютная экзотика для социологических публикаций в нашей стране.

**Таблица 11. Среднее значение стресса (CES-D Score*) по селу и году****

| Село | 1995 | 1996 | 1997 |
|---|---|---|---|
| Латоново (n=157) | 20.27 | 21.01 | 20.08 |
| Венгеровка (n=156) | 24.20 | 22.37 | 22.35 |
| Святцово (n=150) | 23.67 | 24.22 | 23.84 |
| Всего (n=463) | 22.70 | 22.51 | 21.92 |

* CES-D по селу и году: 1995 F(2)=6.43, p<.01; Латоново по Венгеровке, p<.01, Латоново по Святцово, p<.05. 1996 F(2)=3.87, p<.05; Латоново по Святцово, p<.05.
1997 F(2)=4.10, p<.05; Латоново по Святцово, p<.05.
** Источник: David J.O'Brien, Valeri V.Patsiorkovski, Larry D.Dershem. Household Capital and the Agrarian Problem in Russia. Aldershot (UK): Ashgate, 2000, p.198

В приведенном примере данные самой таблицы имеют смысл только вместе с описанием статистических характеристик надежности их связи, а именно: статистического уровня значимости (p-уровень) и F-критерия, проверяющего равенство дисперсий в двух группах.

Третье обстоятельство связано с относительно меньшей наглядностью и довольно сложными расчетами, которые необходимо выполнить для проверки надежности связи.

SPSS, как и другие близкие к нему программные продукты, в части мер сравнения (Compare Means) открывает широкие возможности для таких расчетов, позволяя сделать их нормой исследовательской практики. В профессиональном сообществе такое развитие принимается с сомнениями и критикой. Например, М. Косолапов считает, что легкость освоения SPSS и других сходных продуктов может иметь тяжелые последствия, связанные со снижением уровня статистической грамотности. Дословно эта мысль выражена так: «Существует пропасть между видимой легкостью, доступностью компьютерных программ, реализующих довольно сложные методы анализа эмпирической информации, и сложностью понимания сути этих методов, их реальных возможностей и ограничений в их использовании. Причем, указанная коллизия большинством социологов не осознается» (34). Такие же опасения, но сформулированные по другому случаю, приведены и в главе 14, § 14.1.

Здесь уместно сделать несколько замечаний. Первое – мысль о том, что хорошо бы совместить освоение прикладных программ анализа данных с углубленным изучением статистики – бесспорна. Но,

что же делать социологам-прикладникам и студентам, если более чем за 10 лет использования таких пакетов математики-социологи все еще не нашли время подготовить учебные пособия по данному предмету, написанные хорошим русским языком и на материалах нашей, а не чужой жизни.

Второе - представляется, что провозглашаемый тезис о легкости освоения SPSS находится в остром противоречии с реальным положением вещей. Наш собственный опыт, равно как и популярность первых выпусков нашего учебного пособия, а также содержание получаемых на них отзывов, свидетельствуют в пользу прямо противоположной тенденции. Более справедливо и правильно сказать, что разработчикам от социологии освоение прикладных программ статистической обработки данных дается большим потом и кровью.

Третье - указанная коллизия большинством социологов осознается и переживается очень остро. Поэтому и получили распространение различные учебные курсы по SPSS как традиционных, так и новых дистанционных форм обучения. Причем, повсеместно на таких курсах слушателям даются примеры и материалы либо из переводных фирменных руководств, либо из других не менее далеких от нашей социальной реальности источников.

Наконец, четвертое и последнее замечание. Мы считаем, что в современных условиях только посредством освоения специализированных программных продуктов может быть повышена как общая культура использования статистических методов в социологии в целом, так и освоена сфера измерения надежности зависимости в частности. И лишь таким путем социология и социологи могут реально укрепить и повысить свой научный статус.

Более того, на сайте, к которому, как мы полагаем, автор высказанных выше соображений имеет непосредственное отношение, содержится следующая сентенция: «SPSS широко применяется для обучения студентов и аспирантов прикладному статистическому анализу. Практически все университеты США, Канады и Европы используют SPSS как для проведения научных исследований, так и для преподавания прикладной статистики и смежных дисциплин» (35). Вряд ли следует доказывать, что и у нас в этом плане нет иного пути, разве что освоить еще и другие близкие пакеты, например, SAS или STATISTICA.

Возвращаясь к предмету рассмотрения данной главы, полезно отметить, что в SPSS для целей сравнения и проверки средних

величин используются три основные процедуры: Means, T Test и One Way ANOVA. Общее направление поиска процедур сравнения средних: Analyze - Compare Means. Далее, в открывшемся подменю, в зависимости от поставленной задачи, следует сделать выбор между процедурами Means, T Test и One Way ANOVA. Ниже это движение будет совершаться в порядке, который предложен самой системой.

## 10.2. Средние

Ранее, при рассмотрении процедур Explore, Descriptives, Frequencies, уже упоминались описательные (суммарные) статистики: среднее, максимальное и минимальное значения, стандартное отклонение, равно как и различные способы их вычисления (глава 8, § 8.2). Среднее значение – очень информативная мера «центральной тенденции» наблюдаемой переменной. Она дает возможность делать выводы относительно распределения в целом.

В SPSS процедура Means (средние) вычисляет, а также позволяет получить для сравнения в табличном виде обобщенные статистики сразу для нескольких переменных, распределенных по какому-нибудь признаку. Например, средние значения дохода семьи, заработной платы и пенсий в разных селах. Для выполнения рассматриваемой процедуры необходимо в главном меню реализовать последовательность команд:

**Analyze**

**Compare Means**

**Means**.

При ее выполнении система открывает главное диалоговое окно этой процедуры (рис. 41).

В этом окне слева находится список переменных, из которого следует выбрать необходимые переменные. Перенося переменные в верхнее или нижнее поле, надо быть внимательным и четко представлять: для какой (их) зависимой (верхнее поле) переменной будет выполняться расчет средних, равно как и какая (ие) переменная будет рассматриваться в качестве независимой (нижнее поле).

В качестве примера можно привести расчет среднего дохода домохозяйств по селам. Для этой цели из списка переменных выбирается переменная sumtota9 (совокупный доход) и переносится в верхнее поле Dependent List (список зависимых переменных). Следующей выбирается переменная village9 (село). Она переносится в нижнее

поле Independent List (список независимых переменных). Здесь, как и ранее, имена переменных из используемого нами рабочего файла приводятся с целью наглядности оперирования с ними. Далее, следует известная команда на выполнение расчетов.

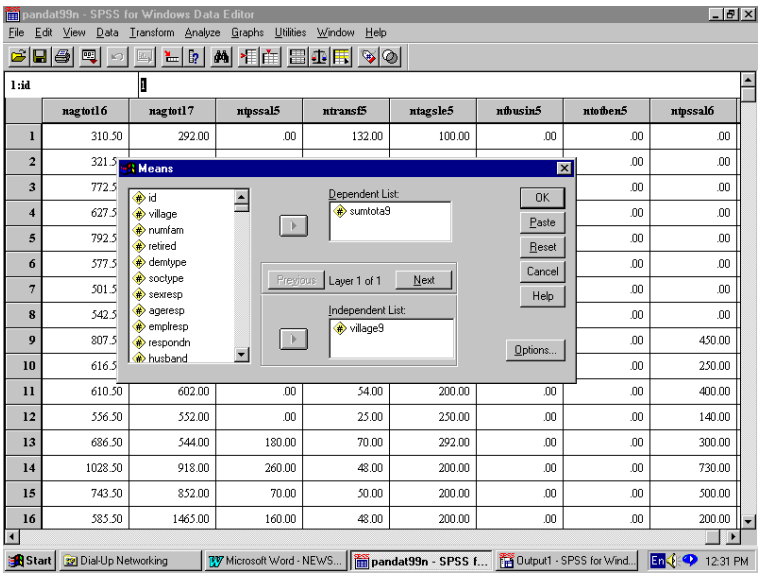



***Рис. 41.***

***Главное диалоговое окно процедуры Means***

Результаты выполнения расчетов из окна просмотра даны в табл. 12-13. В окне просмотра выводятся две таблицы. Первая из них, Case Processing Summary, содержит общие характеристики (число значимых случаев и их доля, число пропущенных значений и их доля, а также итог). В настоящем тексте эта таблица отсутствует. Вторая таблица окна просмотра содержит все заданные для расчета статистики. Она и приведена ниже для двух видов расчетов.

В табл.12 и 13 даны результаты расчетов среднего совокупного дохода домохозяйств по трем селам. В целях наглядности, кодовые значения сел изменены на их названия. В табл.12 статистики рассчитаны по умолчанию, т.е. без использования дополнительного диалогового окна Options. Поэтому системой посчитаны только среднее (Mean), число случаев (N) и стандартное отклонение (Std. Deviation). Эти базовые статистики среднего в принципе позволяют рассчитать многие дополнительные показатели, такие как сумма, дисперсия и др. Например, зная, что сумма есть произведение среднего на число случаев, ее можно вычислить и с помощью калькулятора. Другой пример, возведя в квадрат стандартное отклонение, можно получить дисперсию.

**Таблица 12. Средний доход домохозяйств по селам в 1999 г. (вторая таблица окна просмотра с статистиками, принятыми по умолчанию)**

| VILLAGE9 | Mean | N | Std. Deviation |
|---|---|---|---|
| Латоново | 3132.06 | 144 | 2075.27 |
| Венгеровка | 3215.50 | 145 | 2672.72 |
| Святцово | 2247.17 | 133 | 1465.61 |
| Всего | 2881.84 | 422 | 2183.28 |

Сама система предлагает расчет всех этих и многих других статистик среднего. В версиях SPSS 9.0 и выше в дополнительном диалоговом окне Options рассматриваемой процедуры для расчета предлагается 21 статистическая характеристика. Кроме того, в нижней части слева в этом окне имеется еще два маленьких, но весьма важных чек-бокса, которые обеспечивают возможность выполнения дополнительных расчетов (таблицы ANOVA и др.).

Для получения доступа к таким расчетам в дополнительном диалоговом окне Options процедуры Means следует задать все требуемые статистики. Это можно сделать путем их выделения в левом поле и последующего переноса в правое поле (рис. 42).

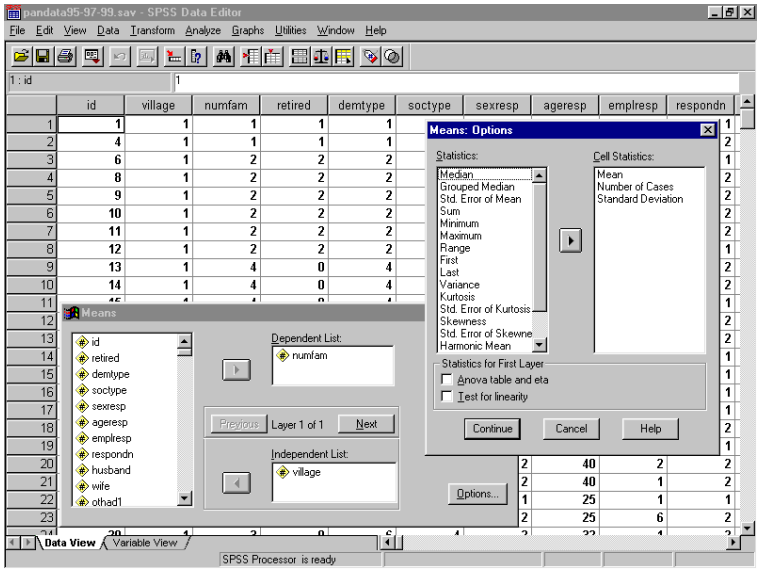



***Рис. 42.***

***Дополнительное диалоговое окно Options - опции процедуры Means***

Результаты таких расчетов и видны в табл. 13. В приведенном ниже примере дополнительно заданы: сумма (Sum), дисперсия (Variance), минимальное (Minimum) и максимальное значения (Maximum). Полный перечень статистик среднего, взятый по версии SPSS 11.5, приведен, с кратким описанием и переводом, в приложении 8.

**Таблица 13. Средний доход домохозяйств по трем селам в 1999 г.** (вторая таблица окна просмотра с дополнительно установленными статистиками)

| VILLAGE9 | Mean | N | Std. Deviation | Sum | Variance | Mini-mum | Maxi-mum |
|---|---|---|---|---|---|---|---|
| Латоново | 3132.06 | 144 | 2075.27 | 451016 | 9501606.7 | 499 | 15183 |
| Венгеровка | 3215.50 | 145 | 2672.72 | 466247 | 7016460.9 | 622 | 27675 |
| Святцово | 2247.17 | 133 | 1465.61 | 298873 | 2014348.6 | 304 | 8276 |
| Всего | 2881.84 | 422 | 2183.28 | 1216136 | 6485929.3 | 304 | 27675 |

Сравнение данных табл. 12 и табл. 13 позволяет лучше понять порядок статистических расчетов. Из табл. 12, в которой представлены результаты расчетов, выполненных системой по умолчанию, довольно затруднительно видеть их содержательный смысл.

В то же время из данных табл. 13, зная формулу расчета средней и стандартного отклонения (приложение 8), порядок расчетов становится абсолютно прозрачным. Так, любое из четырех значений средней может быть получено путем деления соответствующей суммы на соответствующее число случаев (например, для с. Латоново расчет среднего дохода имеет вид: 451016 / 144 = 3132.06).

В свою очередь, любое из четырех значений стандартного отклонения, как результата расчета корня квадратного из выборочной дисперсии (приложение 8), может быть получено путем выполнения указанной операции с соответствующим значением Variance (например, для с. Венгеровка расчет стандартного отклонения имеет следующий вид: $\sqrt{7016460.9} = 2672.72$).

Как уже отмечалось выше, можно одновременно получить характеристики среднего сразу по нескольким переменным. В табл.14. показаны результаты расчета характеристик среднего для четырех переменных: NSRDOH (душевой доход с учетом натуральных поступле-

ний), NSRDOH1 (душевой доход без учета натуральных поступлений), TPSSALA9 (зарплата) и PENSUM9 (пенсия).

**Таблица 14. Средний душевой доход, зарплата и пенсия по трем селам** (вторая таблица окна просмотра)

| VILLAGE9 | Статистики | NSRDOH | NSRDOH1 | TPSSALA9 | PENSUM9 |
|---|---|---|---|---|---|
| Латоново | Mean | 1068.32 | 733.62 | 793.09 | 346.65 |
| | N | 144 | 144 | 144 | 144 |
| | Std. Deviation | 529.77 | 448.38 | 1832.97 | 377.96 |
| Венгеровка | Mean | 1050.20 | 713.69 | 579.43 | 358.29 |
| | N | 145 | 145 | 145 | 145 |
| | Std. Deviation | 772.21 | 738.84 | 638.32 | 322.28 |
| Святцово | Mean | 915.81 | 681.15 | 388.28 | 399.36 |
| | N | 133 | 133 | 133 | 133 |
| | Std. Deviation | 462.39 | 590.05 | 631.76 | 334.21 |
| Всего | Mean | 1014.03 | 710.24 | 592.10 | 367.26 |
| | N | 422 | 422 | 422 | 422 |
| | Std. Deviation | 608.96 | 603.84 | 1196.96 | 345.79 |

Из данных табл. 14 хорошо виден характер и особенности средних значений и разброса по отдельным видам доходов в различных селах. По вполне понятным причинам, различия значений средних минимальны для размера пенсий (Std. Deviation = 345.79). Они, напротив, максимальны для заработной платы (Std. Deviation = 1196.96).

При этом средняя заработная плата в с.Латоново более чем в 2 раза превышает соответствующий показатель для с. Святцово. А значение Std. Deviation для самого с. Латоново почти наполовину выше соответствующего значения по всему массиву. Возникает вопрос: «Как подобные различия в заработной плате возможны?» Анализ показывает, что структура занятости в с. Латоново существенно отличается от структуры занятости в других селах. Здесь многие работают вне села и вне сферы сельскохозяйственного производства. Однако это уже совершенно другая содержательная проблема.

Забегая несколько вперед, заметим, что именно подобного рода случаи, в первую очередь, требуют не только измерения самой зависимости, но и измерения надежности этой зависимости. Содержательно такого рода явления связаны с тем, что любая независимая выборка (например, зарплата в с. Латоново) как статистическое распределение

может обладать совсем иными свойствами, чем сходное распределение в генеральной совокупности (заработная плата в обследованном массиве).

Система позволяет получать характеристики среднего и по нескольким основаниям одновременно. При этом используется принцип построения многомерных таблиц с использованием управляющих переменных, который ранее был описан в главе 9, § 9.2.

Там, где это возможно, анализ средних имеет очень важное информативное значение. Распространенные суждения скептического характера «о средней температуре по больнице» привлекательны своей парадоксальностью, но далеки от истины. На учете свойств средних величин держится многое в науке и вся инженерия.

**Правило 29**

***Общее правило при анализе средних следующее:***
***- чем больше размер выборки и меньше разброс, тем выше надежность оценок среднего;***
***- чем меньше выборка и больше разброс, тем оценка среднего менее надежна.***

## 10.3. T тест - T Test

Процедура **T Test** дает возможность проверять средние значения отдельных переменных, а также сравнивать средние значения двух или нескольких переменных между собой по одному или нескольким основаниям. Эта задача корректна только для вариационных рядов распределения. Вариационный ряд – это распределение, образованное по количественному признаку. По характеру составляющих их чисел такие ряды, условно говоря, можно разбить на три группы. Они могут состоять:

- из случайных чисел;
- из систематически повторяющихся чисел;
- из чисел, плотность распределения которых симметрична относительно среднего.

Последнюю группу рядов в статистике принято называть нормальным распределением.Два первых вида распределения представляют собой условные крайности. В то же время нормальное распределение встречается в статистике и теории вероятностей наиболее часто. В известном смысле, оно выполняет роль образца. В SPSS проверить нормальность распределения позволяет процедура Explore (глава 8, §

8.3).

Все другие ряды распределения находятся в описанном выше пространстве распределений, тяготея или удаляясь от нормального распределения, к двум крайностям. В качестве примеров таких видов распределения можно указать на распределение Стьюдента (38,С.325-326), $\chi^2$ (38,С.327-329), F-распределение (38,С.330-331) и др.

При тестировании рядов распределения, состоящих из случайных чисел, само среднее такого ряда и все его характеристики (дисперсия, стандартное отклонение, стандартная ошибка и др.) оказываются случайными числами. Напротив, при попытке тестирования ряда распределения, состоящего из систематически повторяющихся чисел, сразу становится видно, что его средняя равна самому числу, а все ее характеристики равны нулю. Поэтому такой ряд распределения фактически не подлежит тестированию, о чем система и сообщает в окне просмотра.

Тестирование нормального распределения показывает, что его среднее значение совпадает с медианой и модой, а пик плотности находится в точке с абсциссой, равной среднему значению. При увеличении дисперсии плотность нормального распределения рассеивается относительно среднего значения, а при уменьшении дисперсии она сжимается, имея тенденцию к концентрации вдоль оси симметрии. Все эти и другие свойства нормального распределения делают это распределение столь популярным и надежным инструментом познания.

К этому следует добавить, что огромная часть социальных процессов тяготеет к нормальному распределению. В то же время можно встретить как социальные явления, имеющие характер систематически повторяющихся близких по своему значению чисел (например, нивелируемый в течение последних лет размер пенсии в нашей стране), так и социальные явления, имеющие случайный характер, на чем, например, держится вся система страхования жизни и имущества, лотерей и игорных домов.

Еще одно важное замечание. Возраст (в числе исполнившихся лет) представляет собой хороший пример социального явления, имеющего характер нормального распределения. Но такой пример имеет смысл, когда речь идет о возрасте населения страны или большого города. В то же время возраст учеников в классе (взятый в тех же единицах измерения) будет тяготеть к распределению с постоянно повторяющимися числами.

Обратный пример – распределение по заработной плате в нашем массиве. Как уже отмечалось ранее, в с. Латоново велика доля занятых не только вне села, но и в отраслях близких по оплате труда к полярным условиям (таможня, нефтебаза, газстрой). Отсюда в первичном распределении и появляется большой разброс, а его средняя заметно превышает средний народнохозяйственный уровень оплаты труда в сельском хозяйстве. В то же время проверка этого распределения по независимой переменной «село» позволяет получить практически идеальную картину для с. Венгеровка и Святцово, в которых очень низок разброс именно в связи с тем, что почти все работают в сельском хозяйстве.

Другими словами и еще раз - **ряд распределения, который сформирован на основе независимой выборки, может обладать совершенно иными характеристиками, чем основная выборка.** Эту особенность рядов распределения полезно помнить как при рассмотрении процедур тестирования средних, так и при анализе, и использовании данных социологических исследований.

Обобщения, сделанные на основе анализа представительной выборки, могут быть как вполне корректны, так и далеки от корректности. Такое же утверждение справедливо и для обобщений, сделанных на основе анализа независимых выборок. Любая таблица распределения фактически и строится на основе использования принципа независимых выборок. Вряд ли будет большим отступлением от истины сказать, что именно анализ таблиц распределения служит основой различных выводов и обобщений в социологических исследованиях и разработках.

Для целей тестирования средних в SPSS предлагается три вида процедуры T Test:

- одновыборочный t-критерий **(One Sample T Test),**
- t-критерий для независимых выборок **(Independent-Sample T Test),**
- t-критерий для парных выборок **(Paired-Sample T Test).**

Порядок выполнения всех этих процедур, равно как и их содержательные особенности, и будут рассмотрены ниже. Процедуры T Test описаны в последовательности их представления в соответствующем меню системы.

**One Sample T Test** - процедура тестирования статистических характеристик среднего одной или нескольких переменных. Например, сельхозпродукция домохозяйства, средний душевой доход, средняя

заработная плата, средняя пенсия и т.п. Процедура проверяет среднее значение каждой отдельно взятой переменной в отличие от константного значения, соответствующего определенным условиям (уровню значимости и степени свободы). Таблица константных значений приведена в 38,С.327-329. Путь, с помощью которого открывается главное диалоговое окно рассматриваемой процедуры, следующий:

**Analyze**

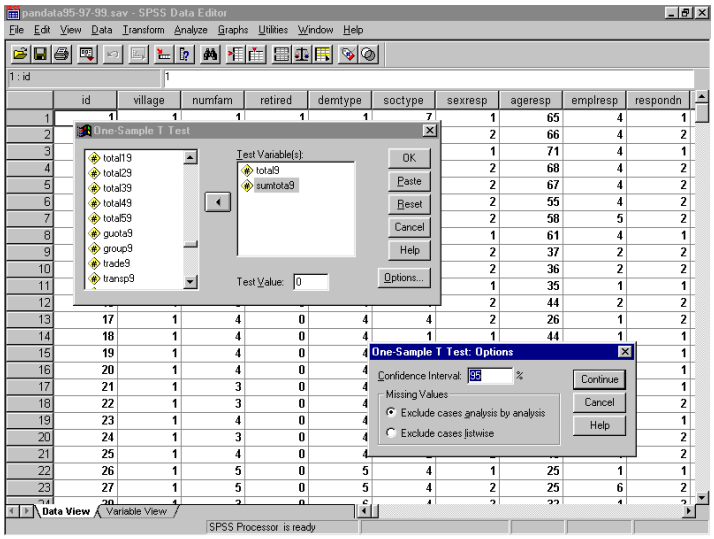
**Compare Means**

**One Sample T Test**.

При выполнении указанной последовательности команд система открывает требуемое главное диалоговое окно (рис. 43).

***Рис. 43.***

***Главное и дополнительное диалоговые окна процедуры One Sample T Test***




Как видно на рис. 43, задание тестируемых переменных выполняется достаточно легко и удобно. Делается это с помощью их выделения из списка переменных рабочего файла (как всегда левый список) и последующего переноса в правый список. В тесте может участвовать минимум одна переменная.

На рис. 43 конфигурация главного диалогового окна рассматриваемой процедуры указывает, что одновременно может тестироваться и несколько переменных. Правое поле Test Variable(s) устроено так, что в него можно перенести несколько переменных. Система выполнит процедуру тестирования, как всегда, при получении команды ОК.

Для процедур тестирования, включая и описываемую, в системе всегда имеется лимит на максимально допустимое число одновременно тестируемых переменных. В случае превышения установленного по умолчанию лимита система выдаст (в специальном окне) предупреждение об этом. Это предупреждение должно быть принято во внимание.

Особенность рассматриваемой процедуры состоит в том, что она выполняется отдельно для каждой из тестируемых переменных. Установка порогового значения в нижнем чек-боксе (Test Value) или смена параметров доверительного интервала, который дан по умолчанию в дополнительном диалоговом окне Options (рис. 43), не меняют существа дела.

То же самое следует сказать и о боксах пропущенных значений (Missing Values) в дополнительном диалоговом окне Options рассматриваемой процедуры. T Test использует только значимые данные в тестируемой переменной.

Результаты тестирования переменных «производство продукции» (картофеля, овощей, мяса и молока) приведены в табл. 15-16. В окне просмотра результатов выполнения этой процедуры, как и при расчете среднего значения, даются две таблицы: One-Sample Statistics (табл. 15) и One-Sample Test (табл. 16). В данном случае обе таблицы содержат различные статистики. Поэтому обе они и приведены в тексте.

**Таблица 15**. **Результаты выполнения процедуры One-Sample Test для переменных, связанных с производством отдельных видов продукции** (первая таблица окна просмотра)

**One-Sample Statistics**

| | N | Mean | Std. Deviation | Std. Error Mean |
|---|---|---|---|---|
| POTPROD9 | 422 | 1940.13 | 2258.69 | 109.95 |
| VEGPROD9 | 422 | 383.51 | 681.25 | 33.16 |
| MEATPRO9 | 422 | 335.42 | 324.14 | 15.78 |
| MILKPRO9 | 422 | 2889.22 | 2939.75 | 143.10 |

Табл. 15 содержит основные характеристики среднего (первая таблица окна просмотра), а в табл. 16 представлены результаты тестирования, которые взяты из второй таблицы окна просмотра. Окно просмотра этой процедуры претерпело в последних версиях системы существенные изменения. На наш взгляд, в версии SPSS 6.1 оно

полнее отражало специфику тестирования. Результаты тестирования выводились в отдельной таблице для каждой переменной (4, С. 128), тогда как позже они стали выводиться отдельной строкой в общей таблице.

**Таблица 16. One-Sample Test - Тестирование средних для переменных, связанных с производством отдельных видов продукции** (вторая таблица окна просмотра)

**One-Sample Test**

| | Test Value = 0 | | | | | |
|---|---|---|---|---|---|---|
| | t | df | **Sig. (2-tailed)** | Mean Difference | 95% Confidence Interval of the Difference | |
| | | | | | Lower | Upper |
| POTPROD9 | 17.645 | 421 | .000 | 1940.13 | 1724.00 | 2156.25 |
| VEGPROD9 | 11.564 | 421 | .000 | 383.51 | 318.32 | 448.69 |
| MEATPRO9 | 21.257 | 421 | .000 | 335.42 | 304.41 | 366.44 |
| MILKPRO9 | 20.190 | 421 | .000 | 2889.22 | 2607.93 | 3170.51 |

Из табл. 15-16 видно, что процедура One Sample T Test выполняет расчет различных статистик для двух выводимых в окно просмотра таблиц. Эти статистики и описаны ниже. Здесь и далее, в тех случаях, если статистика упоминается в первый раз, она выделяется жирным шрифтом. В тех случаях, когда статистика уже упоминалась ранее, рядом с ней без повторного описания дается только ее название.

Для первой таблицы окна просмотра характерно наличие следующих статистик среднего:

- N – число случаев.
- Mean – среднее**.**
- Std. Deviation – стандартное отклонение.
- **Std. Error Mean** – стандартная ошибка среднего. Ее определение дано в приложении 8. Существенным здесь является знание о благоприятных допустимых значениях этой статистической характеристики: они должны быть вне интервала от -2 до +2. В этом случае нулевая гипотеза о равенстве средних значений тестируемой переменной отвергается.

Для второй таблицы окна просмотра характерно наличие

следующих статистик среднего:

- **Test Value**. По умолчанию он равняется 0. Изменения Test Value в интервале от 1 до 99 требует одновременного изменения уровня доверительного интервала (Confidence Interval).

**- t** - t-распределение Стьюдента. Эта статистика среднего важна в тех случаях, когда неизвестна дисперсия выборки (22, С. 65).

**- df** – число степеней свободы (Degrees of Freedom). Это параметр распределения $\chi^2$. При допущении о равенстве дисперсий этот параметр всегда представляет собой целое положительное число. Условно говоря, чем это число больше, тем лучше. При больших степенях свободы (более 30) t-распределение практически совпадает со стандартным нормальным распределением. В статистике связь числа степеней свободы с распределением $\chi^2$ имеет табличное представление, которое приведено в 38, С. 327-329.

**- Sig. (2-tailed)** – это p - уровень статистической значимости (двухсторонний) t–критерия. В отличие от одностороннего уровня статистической значимости, который будет рассмотрен далее, он предполагает анализ различия средних в двух направлениях. Как и уровень односторонней значимости, этот показатель зависит в основном от объема выборки. В очень больших выборках даже слабые зависимости между переменными будут значимыми.

Обычно, если p меньше или равно 0.05, то это рассматривается как приемлемая граница статистической значимости. Она говорит о 5-ти процентной вероятности ошибки. Поэтому только результаты на уровне p=0.01 рассматриваются как статистически значимые. А результаты с уровнем p=0.005 или p=0.001 - как высоко значимые.

Если по результатам теста p уровень значимости имеет в окне просмотра значение .000, то это свидетельствует о фиксации самого высокого уровня значимости, который можно получить в стандартных расчетах (по умолчанию). **В этом случае нулевая гипотеза о равенстве средних значений тестируемой переменной отвергается**.

**- Mean Difference** – различие среднего. В данном случае оно равно среднему значению тестируемого признака (см.: значения Mean в табл. 15 и Mean Difference в табл. 16).

**- Confidence Interval of the Difference** - доверительный интервал, который по умолчанию равен 95% различий между средним и гипотетическим значением. Доверительный интервал для среднего представляет интервал значений вокруг оценки, где с данным уровнем доверия находится «истинное» (неизвестное) среднее тестируемого распреде-

ления. Ширина доверительного интервала зависит от объема или размера выборки, а также от разброса (изменчивости) данных. Увеличение размера выборки делает оценку среднего более надежной. Увеличение разброса наблюдаемых значений уменьшает надежность оценки. Доверительный интервал имеет нижнее (Lower) и верхнее (Upper) значения. Как видно из табл. 16, средняя всегда находится внутри этого интервала.

Из выполненного в табл. 16 тестирования ряда переменных видно, что наблюдаемый уровень значимости практически идеален для всех тестируемых переменных. Нулевая гипотеза о независимости значений отдельных случаев в каждой из этих переменных отвергается, а распределение каждой из них фактически совпадает со стандартным нормальным распределением.

**Independent-Sample T Test -** процедура проверки характеристик среднего одной или нескольких переменных по определенному основанию. Например, производство отдельных видов продукции (картофеля, овощей, мяса и молока), средний душевой доход, средняя заработная плата и средняя пенсия в двух из трех обследованных сел. Процедура t-критерий для независимых выборок может быть выполнена только для двух групп одновременно (в нашем случае, скажем, для двух, а не трех сел). Путь, с помощью которого можно открыть главное диалоговое окно рассматриваемой процедуры:

**Analyze**

**Compare Means**

**Independent-Sample T Test**.

При выполнении указанной последовательности команд система открывает главное диалоговое окно этой процедуры (рис. 44).

Это окно имеет одновременно сходство и существенные различия по сравнению с таким же окном предшествующей процедуры. Как видно из рис. 44, сходство состоит в наличии правого верхнего списка Test Variable(s), в который и переносятся тестируемые переменные.

Различие связано здесь с наличием третьего поля Grouping Variable. В него переносится переменная, два значения которой будут использоваться в качестве основания группировки. После переноса переменной для задания требуемых значений необходимо нажать кнопку Define Groups. В открывшемся одноименном окне имеется возможность выбора двух различных опций.

Одна из них предполагает использование опции Use specified values

(рис. 44), которая дается по умолчанию. В поля записей Group 1 и Group 2 вносятся коды двух из нескольких значений переменной, которая служит основанием группировки. Далее следуют известные команды: продолжить (Continue) и выполнить (OK).

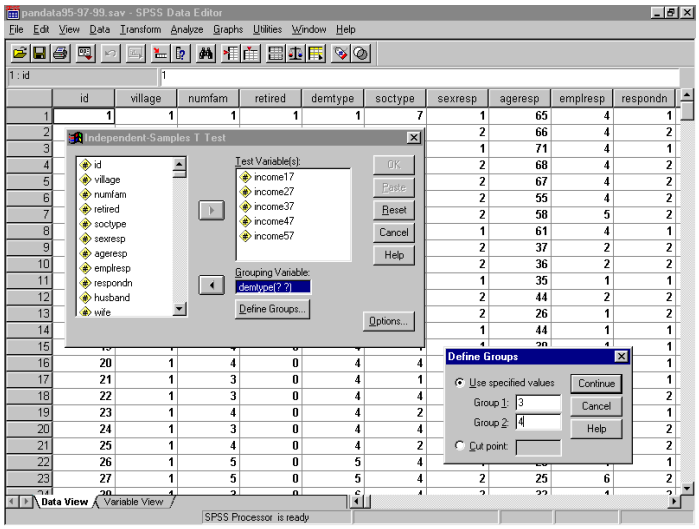



***Рис. 44.***

***Главное диалоговое окно и окно Define Groups процедуры Independent-Sample T Test***

Другая опция – Cut Point позволяет разделить множество значений группирующей переменной на две подгруппы в любом интересующем пользователя разрезе. Например, возьмем в качестве группирующей переменной демографический тип семьи. У этой переменной имеется семь различных значений: одиночки, брачные пары, нуклеарные семьи и др. (приложение 4). Тогда при пометке Cut Point и задании числа 7 (прочие семьи) в открывшейся строке записей, анализ будет выполнен в разрезе двух независимых выборок: =>7 и <=7.

Таким образом, «прочие семьи» составят первую независимую выборку, а шесть остальных демографических типов семьи образуют вторую. Эта опция открывает огромные возможности при решении задач, которые связаны с выбором различных вариантов анализа (по семьям с детьми, по пенсионерам и др.).

В табл.17 приведен фрагмент окна просмотра, полученный при выполнении процедуры Independent-Sample T Test, для тестируемых переменных: производства сельхозпродукции в домохозяйствах по группировочной переменной «село». Для этой переменной в качестве оснований группировок взяты два из трех сел, представленных в рабочем файле. Указание об этом содержится в заголовке табл. 17.

В окне просмотра этой процедуры, точно так же как и предшествующей, даются две таблицы Group Statistics и Independent

Samples Test. Первая из них соответствует окну просмотра, описанному ранее в предшествующей процедуре, а вторая характерна для данного теста. Она и приведена в табл. 17.

**Таблица 17. Independent Samples Test для переменных, связанных с производством отдельных видов продукции** (картофеля, овощей и мяса)**, и группировочной переменной «село»** (группа 1 – Латоново, группа 2 – Венгеровка) (вторая таблица окна просмотра)

**Independent Samples Test**

| | | Levene's Test for Equality of Variances | | t-test for Equality of Means | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | F | Sig | t | df | Sig. 2-tailed | Mean Difference | Std. Error Difference | 95% Confidence Interval of the Difference | |
| | | | | | | | | | Lower | Upper |
| POTPROD9 | * | 146.4 | .000 | -17.62 | 287 | .000 | -2868.1 | 162.77 | -3188.4 | -2547.73 |
| | ** | | | -17.67 | 171.7 | .000 | -2868.1 | 162.31 | -3188.4 | -2547.74 |
| VEGPROD9 | * | 10.5 | .001 | -4.41 | 287 | .000 | -293.73 | 66.54 | -424.69 | -162.77 |
| | ** | | | -4.42 | 153.5 | .000 | -293.73 | 66.32 | -424.75 | -162.71 |
| MEATPRO9 | * | 1.0 | .306 | -.22 | 287 | .826 | -8.75 | 39.73 | -86.95 | 69.44 |
| | ** | | | -.22 | 275.9 | .826 | -8.75 | 39.75 | -87.01 | 69.51 |

** Equal variances assumed, ** Equal variances not assumed*

Из сравнения табл. 16 и 17 видно, что в отличие от предшествующей, рассматриваемая процедура рассчитывает несколько дополнительных статистик. Ее основная особенность состоит в том, что она дает две независимые группы статистик: Levene's Test for Equality of Variances (критерий Левена) и t-test for Equality of Means (тест на проверку равенства средних).

Первая из этих групп - критерий Левена для проверки гипотезы о равенстве дисперсий (Levene's Test for Equality of Variances) имеет два показателя:

**- F** – критерий, проверяющий действительно ли отношение дисперсий значимо больше 1. Если это различие значимо, т.е значение F больше 1, то нулевая гипотеза отвергается и принимается альтернативная гипотеза о существовании различия между средними. F-распределение сосредоточено на положительной полуоси (38, С. 330-331). Оно,

в отличие от нормального распределения, несимметрично (22, С. 69).

**- Sig** - p - наблюдаемый уровень статистической значимости. Его другое распространенное название p-value. Как и описанный ранее Sig. (2-tailed) двухсторонний, это показатель, находящийся в убывающей зависимости от надежности результата. Более высокий p-уровень соответствует более низкому уровню доверия к найденной зависимости между переменными.

**Этот показатель служит основой принятия или опровержения нулевой гипотезы.** При этом все зависит от его значения. Ранее уже отмечалось, что, если наблюдаемый уровень значимости достаточно мал (равен или меньше 0.05 или 0.01), то нулевая гипотеза отвергается.

Как видно из табл. 17, оба показателя F и Sig рассчитываются только при допущении предположения о равенстве значений дисперсии [Equal variances assumed (Equal)]. Критерий Левена для проверки гипотезы о равенстве дисперсий свободен от допущения предположения о нормальности тестируемого распределения.

Вторая группа статистик - t-тест на равенство средних (t-test for Equality of Means). Здесь рассчитываются следующие статистики:

- t - t-распределение Стьюдента.

- df - число степеней свободы. При допущении предположения о неравенстве дисперсий (Equal variances not assumed) это число, как видно из примера в табл. 17, может быть дробным.

- Sig. (2-tailed) 2-Tail Sig – уровень значимости (двухсторонний).

- Mean Difference – различие среднего. В данном случае оно равно разнице (остатку) средних значений тестируемого признака в двух выделенных группах (независимых выборках).

**- Std. Error Difference (SE of Diff)** – стандартная ошибка различия среднего. Она характеризует стандартное отклонение выборочного среднего, рассчитанное по выборке размера n из генеральной совокупности. Это зависит от дисперсии генеральной совокупности (сигма) и объема выборки (n).

- 95% CI for Diff – доверительный интервал.

Все эти статистики рассчитываются в рассматриваемой процедуре как для случая равных значений дисперсии Equal variances assumed (Equal), так и для случая, допускающего предположение о различии (неравных) значений дисперсии Equal variances not assumed (Unequal). Пять из этих параметров рассчитывались и описаны в предыдущей процедуре: t-распределение Стьюдента (t**),** число степеней свободы

(df), двухсторонний уровень значимости (2-Tail Sig), различие среднего (Mean Difference) и доверительный интервал (95% CI for Diff). Лишь одна статистика - стандартная ошибка различия среднего (SE of Diff) - характерна именно для данного теста.

Возвращаясь к результатам тестирования по независимой выборке (табл. 17), можно сказать, что первые две переменные «производство картофеля» и «производство овощей» имеют хорошие характеристики. Для этих переменных нулевая гипотеза о равенстве дисперсий и средних в каждой из двух рассматриваемых выборок отвергается.

Подобный вывод не может быть сделан в отношении последней переменной «производство мяса». Нулевая гипотеза о равенстве дисперсий и средних в каждой из двух рассматриваемых выборок фактически подтверждается.

**Paired-Sample T Test** (t – критерий для парных выборок) - процедура сравнения значений средних величин двух переменных. Например, сравнение душевого дохода с учетом натуральных поступлений и без учета натуральных поступлений.

Система предлагает следующий путь, с помощью которого можно открыть главное диалоговое окно рассматриваемой процедуры «t - критерий для парных выборок»:

**Analyze**

**Compare Means**

**Paired-Sample T Test**.

При выполнении указанной последовательности команд система открывает главное диалоговое окно этой процедуры (рис. 45).

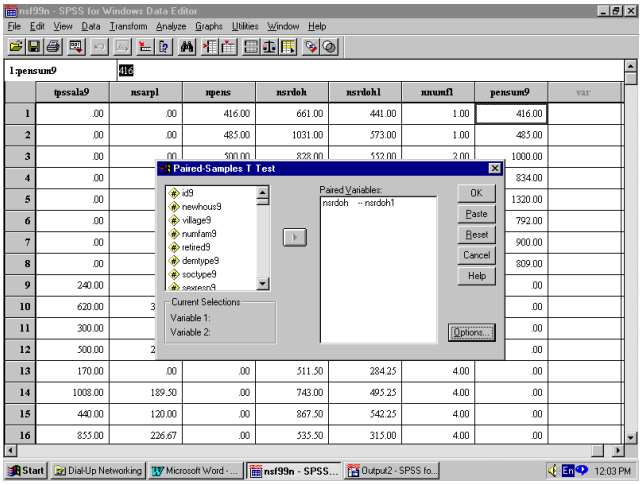



***Рис. 45.***

***Главное диалоговое окно процедуры Paired-Sample T Test***

В этой процедуре перенос переменных из левого в правое поле возможен только парой. Вначале надо выделить одну переменную, и в нижней части окна сразу же высвечивается ее имя в первой строке Variable1. При выделении второй переменной, она появляется во второй строке Variable2. Затем обе переменные, в виде пары NSRDOH - NSRDOH1, переносятся в правое поле (Paired Variables list) путем нажатия стандартной для таких случаев кнопки со стрелкой.

Далее можно взять следующую пару переменных, и так можно выбрать несколько пар. Возможна и обратная операция изъятия переменных из предполагаемых расчетов, которая выполняется путем выделения пары и возврата ее в левое поле со списком переменных.

В табл. 18 и 19 видно, как результаты тестирования представлены в окне просмотра в версии SPSS 11.5. При выполнении этой процедуры в окне просмотра появляются уже не две таблицы, как ранее, а целых три: Paired Samples Statistics, Paired Samples Correlations и Paired Samples Test.

В данном случае, первая и третья таблицы идентичны своим предшественницам в ранее рассмотренных процедурах тестирования. А вторая таблица окна просмотра этой процедуры новая. Она и приведена в табл. 18.

**Таблица 18. Paired Samples Correlations (t – критерий для парных выборок) месячного душевого дохода домохозяйств с учетом** (NSRDOH) **и без учета натуральных поступлений** (NSRDOH1) (вторая таблица окна просмотра)

**Paired Samples Correlations**

| | | N | Correlation | Sig. |
|---|---|---|---|---|
| Pair 1 | NSRDOH - NSRDOH1 | 422 | .889 | .000 |

Табл. 18 содержит две важные характеристики.

Первая – Sig. Эта характеристика уже рассматривалась в предшествующем тесте.

Вторая - **Correlation** (корреляция) - совершенно новая характеристика, дающая название всей таблице (Paired Samples Correlations).

Корреляция – мера зависимости двух величин. Наиболее известна корреляция Пирсона. Коэффициент корреляции является безразмерной величиной. Его значение лежит в интервале между –1 и +1.

Положительная зависимость двух величин тем теснее, чем она ближе к +1. В этом случае при возрастании одной величины растет и другая величина.

Отрицательная зависимость двух величин тем теснее, чем она ближе к –1. Здесь при возрастании одной величины другая величина уменьшается.

Значение .000 свидетельствует об отсутствии корреляции. Корреляция высокая, когда на графике зависимость можно представить прямой линией (с положительным или отрицательным углом наклона). Важно, что значение коэффициента корреляции не зависит от масштаба измерения. Полнее корреляция рассматривается в следующей главе.

В табл. 19 показаны результаты расчетов, выведенные в третьей таблице окна просмотра. Данные этой таблицы позволяют утверждать, что статистики, рассчитываемые в этой части процедуры Paired Samples Test, не имеют характерной специфики, по сравнению с ранее рассмотренными тестами.

Все они уже описаны в предшествующих параграфах и вряд ли нуждаются в повторном рассмотрении. Содержательно из табл. 18 и 19 видно, что нулевая гипотеза о равенстве средних рассматриваемой пары переменных отвергается, а сами эти переменные имеют высокую степень положительной взаимосвязи (коэффициент корреляции равен 0,889).

**Таблица 19. Paired Samples Test (t – критерий для парных выборок) месячного душевого дохода домохозяйств с учетом и без учета натуральных поступлений** (третья таблица окна просмотра)

**Paired Samples Test**

| | | Paired Differences | | | | | t | df | Sig. (2-tailed) |
|---|---|---|---|---|---|---|---|---|---|
| | | Mean | Std. Deviation | Std. Error Mean | 95% Confidence Interval of the Difference | | | | |
| | | | | | Lower | Upper | | | |
| Pair 1 | NSRDOH-NSRDOH1 | 303.79 | 285.98 | 13.92 | 276.42 | 331.15 | 21.822 | 421 | .000 |

**Правило 30**

***Общий принцип Т-тестов — оценка наблюдаемой зависимости путем сравнения ее с максимально мыслимой зависимостью.***

## 10.4. Однофакторный дисперсионный анализ

Однофакторный дисперсионный анализ (One Way ANOVA) - процедура сравнения разброса (вариации) средних величин нескольких зависимых переменных по отношению к одной независимой переменной (фактору). Например, производство продукции (картофеля, овощей, мяса и т.п.) в домохозяйствах разных сел, или в разные годы, или в различных демографических типах семьи.

В основе этой процедуры лежит гипотеза о равенстве средних. В отличие от одновыборочного t критерия (one -sample t test), описанного ранее, однофакторный дисперсионный анализ может рассматриваться как двухвыборочный t критерий **(**two-sample t test**)**. Эта процедура позволяет повысить чувствительность анализа путем сравнения выборочных дисперсий, а не самих средних, как это делается в одновыборочном t тесте.

Путь, с помощью которого можно открыть главное диалоговое окно рассматриваемой процедуры:

**Analyze**

**Compare Means**

**One Way ANOVA.**

При выполнении указанной последовательности команд система открывает главное диалоговое окно этой процедуры (рис. 46).

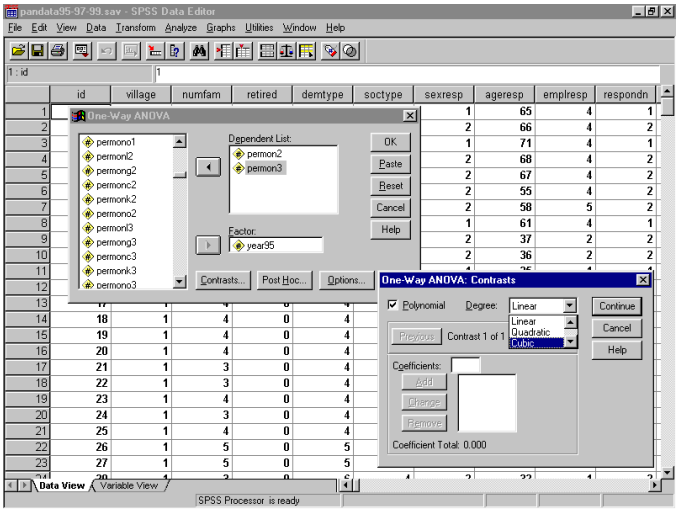



***Рис. 46.***

***Главное и дополнительное -Contrasts диалоговые окна процедуры One Way ANOVA***

В открывшемся диалоговом окне, из левого списка в правый верхний список «Dependent List», переносятся зависимые переменные, а в правое нижнее поле «Factor» переносится одна - независимая переменная. В разрезе этой переменной и проводится анализ.

Рассматриваемая процедура имеет два вида сравнения средних: априорные различия (a priori contrasts) и различия, выявленные по итогам тестирования (post hoc tests) – апостериорные. Поэтому в нижней части главного диалогового окна и находятся три кнопки, открывающие доступ к дополнительным диалоговым окнам: Contrasts…, Post Hoc…, Options… (рис. 46).

В дополнительном диалоговом окне **Contrasts** (рис.46) задается расчет априорных различий (a priori contrasts). При пометке опции **Polynomial** открывается доступ к выпадающему списку **Degree**, в котором можно выбрать: Linear, Quadratic, Cubic, $4^{th}$, $5^{th}$.

Дополнительное диалоговое окно **Post HocMultiple Comparisons** (рис. 47) дает возможность тестирования в режиме post hoc tests, т.е. апостериорно.

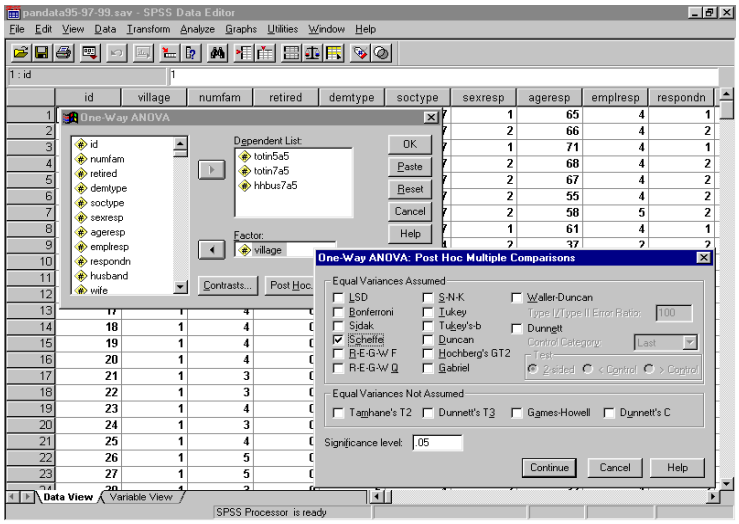



***Рис. 47.***

***Дополнительное диалоговое окно Post HocMultiple Comparisons процедуры One Way ANOVA***

Это окно имеет большое число опций-статистик (чек-боксов: LSD, Duncan, Scheffe и др.), которые необходимо установить для целей выполнения теста. Все указанные опции объединены в два блока:

- статистики, используемые в случае допущения равенства дисперсий (Equal Variances Assumed);

- статистики, используемые при отсутствии такого допущения (Equal Variances NotAssumed).

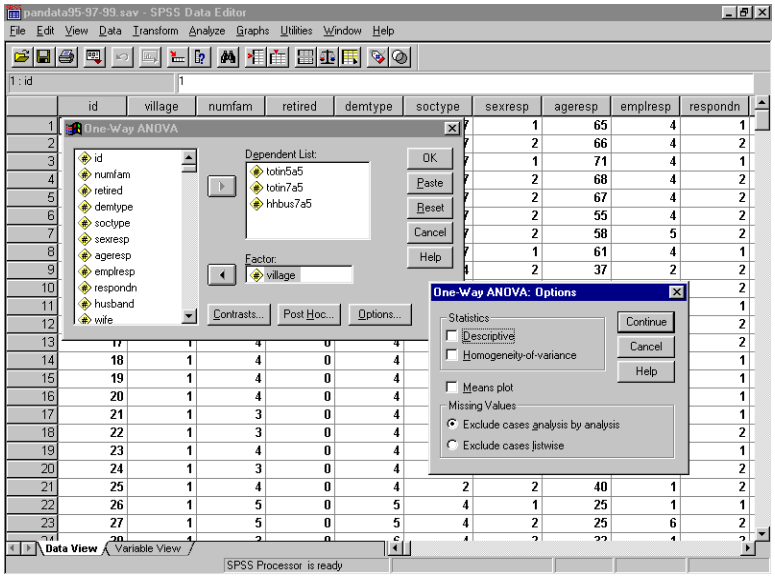
В нижней части этого дополнительного диалогового окна указан уровень статистической значимости (significance level), который по умолчанию равен 0,05.

Обычно после получения статистически значимого результата в дисперсионном анализе полезно знать, какие средние вызвали наблюдаемый эффект. Это делается с помощью процедур апостериорного сравнения, содержащегося в дополнительном диалоговом окне Post HocMultiple Comparisons. Указанные процедуры специально рассчитаны так, чтобы учитывать более двух выборок. Одна из таких процедур, а именно тест Scheffe и использована в приведенном ниже примере.

И, наконец, в дополнительном диалоговом окне **Options** (рис.48) имеются опции для задания дополнительных статистик, построения графика средних, а также расчетов с включением и исключением пропущенных значений.

***Рис. 48.***

***Дополнительное диалоговое окно Options процедуры One Way ANOVA***




В качестве примера ниже приведен анализ средних переменных «производство картофеля» в 1995, 1997, 1999 гг. в обследованных селах и надежности зависимости между анализируемыми средними с помощью процедуры One Way ANOVA (табл. 20-21).

**Таблица 20. ANOVA – однофакторный дисперсионный анализ для производства картофеля по годам в разных селах (первая таблица окна просмотра)**

**ANOVA**

| | | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| POTPROD5 | Between Groups | 176034524.980 | 2 | 88017262.490 | 24.535 | .000 |
| | Within Groups | 1503123358.669 | 419 | 3587406.584 | | |
| | Total | 1679157883.649 | 421 | | | |
| POTPROD7 | Between Groups | 561887160.940 | 2 | 280943580.470 | 63.800 | .000 |
| | Within Groups | 1845054924.368 | 419 | 4403472.373 | | |
| | Total | 2406942085.308 | 421 | | | |
| POTPROD9 | Between Groups | 632185430.540 | 2 | 316092715.270 | 87.385 | .000 |
| | Within Groups | 1515620931.803 | 419 | 3617233.727 | | |
| | Total | 2147806362.344 | 421 | | | |

**Таблица 21. ANOVA – однофакторный дисперсионный анализ для производства картофеля по годам в разных селах (вторая таблица окна просмотра)**

**Multiple Comparisons Scheffe**

| Dependent Variable | (I) VILLAGE | (J) VILLAGE | Mean Difference (I-J) | Std. Error | Sig. | 95% Confidence Interval | |
|---|---|---|---|---|---|---|---|
| | | | | | | Lower Bound | Upper Bound |
| POTPROD5 | 1 | 2 | -1,546.12 | 222.83 | .000 | -2,093.51 | -998.73 |
| | | 3 | -966.78 | 227.78 | .000 | -1,526.33 | -407.22 |
| | 2 | 1 | 1,546.12 | 222.83 | .000 | 998.73 | 2,093.51 |
| | | 3 | 579.34 | 227.41 | .040 | 20.71 | 1,137.97 |
| | 3 | 1 | 966.78 | 227.78 | .000 | 407.22 | 1,526.33 |
| | | 2 | -579.34 | 227.41 | .040 | -1,137.97 | -20.71 |
| POTPROD7 | 1 | 2 | -2776.60 | 246.88 | .000 | -3383.06 | -2170.14 |
| | | 3 | -1624.64 | 252.37 | .000 | -2244.58 | -1004.70 |
| | 2 | 1 | 2776.60 | 246.88 | .000 | 2170.14 | 3383.06 |
| | | 3 | 1151.95 | 251.95 | .000 | 533.04 | 1770.87 |
| | 3 | 1 | 1624.64 | 252.37 | .000 | 1004.70 | 2244.58 |
| | | 2 | -1151.95 | 251.95 | .000 | -1770.87 | -533.04 |
| POTPROD9 | 1 | 2 | -2868.11 | 223.75 | .000 | -3417.77 | -2318.45 |
| | | 3 | -794.29 | 228.73 | .003 | -1356.17 | -232.41 |
| | 2 | 1 | 2868.11 | 223.75 | .000 | 2318.45 | 3417.77 |
| | | 3 | 2073.82 | 228.35 | .000 | 1512.87 | 2634.77 |
| | 3 | 1 | 794.29 | 228.73 | .003 | 232.41 | 1356.17 |
| | | 2 | -2073.82 | 228.35 | .000 | -2634.77 | -1512.87 |

- The mean difference is significant at the .05 level.

Используя описанный выше путь и открыв главное диалоговое окно рассматриваемой процедуры, выбираем из списка переменных необходимые для анализа переменные. В нашем примере это три переменные POTPROD5, POTPROD7, POTPROD9 (производство картофеля соответственно в 1995, в 1997 и 1999 гг.). В качестве фактора используется переменная VILLAGE (село). Результаты расчетов из окна просмотра системы представлены в табл. 20-21.

В табл. 20 наиболее существенную роль имеют две последние колонки F и Sig. Все числа в последней колонке имеют значение 0.000, что свидетельствует о высоком уровне значимости зависимости между анализируемыми переменными. В колонке с характеристикой F все числа существенно больше 1, что также свидетельствует в пользу значимости и надежности рассматриваемой зависимости.

В табл. 21 наиболее важной является колонка Sig. Она свидетельствует о значимости зависимости как по годам, так и между производством картофеля в каждом из трех наблюдаемых сел.

В целом данные окна вывода процедуры ANOVA для производства картофеля по годам в разных селах, приведенные в табл. 20-21, позволяют отразить характеристики надежности фиксируемой связи при сопоставлении среднего производства картофеля в 1995-1999 гг. в домохозяйствах трех рассматриваемых сел (табл. 22).

**Таблица 22. Среднее производство картофеля в 1995-1999 гг. в домохозяйствах трех сел (кг)**

| Село | Производство картофеля в домохозяйствах | | |
|---|---|---|---|
| | **1995 (n=422)** | 1997 **(n=422)** | **1999 (n=422)** |
| Латоново (n=144) | 812.85 | 543.40 | 704.31 |
| Венгеровка (n=145) | 2,358.97 | 3320.00 | 3572.41 |
| Святцово(n=133) | 1,779.62 | 2168.05 | 1498.59 |
| Всего (N=422) | 1,648.79 | 2009.48 | 1940.13 |

Примечания: 1995 г. F=24.53, p<.001; Scheffe, Латоново х Венгеровка и Святцово p<.001; Венгеровка х Святцово p<.05;
1997 г. F=63.80, p<.001; Scheffe, Латоново х Венгеровка и Святцово p<.001;
1999 г. F=87.38, p<.001; Scheffe, Латоново х Венгеровка p<.001, Латоново х Святцово p<.005, Венгеровка х Святцово p<.001.

Последняя таблица содержит не только сами данные, но и информацию об уровне их надежности и объеме всей и каждой из независимых

выборок. Она приведена в том виде, в котором многократно использована в различных докладах и публикациях (36).

**Правило 31**

***Общий принцип процедуры ANOVA - сравнение разброса (вариации) средних величин нескольких зависимых переменных по отношению к одной независимой переменной (фактору).***

**Задание для самостоятельной работы**

1. Что такое меры сравнения?
2. Какие меры сравнения используются в SPSS?
3. Как из главного меню можно прийти к мерам сравнения?
4. Что означает величина зависимости между переменными?
5. Что значит надежность зависимости между переменными?
6. В чем смысл средних значений?
7. Что позволяет делать процедура Means?
8. Какие переменные могут использоваться в мерах сравнения?
9. Какие статистики среднего позволяет рассчитывать процедура Means?
10. Какие дополнительные диалоговые окна имеются в процедуре Means?
11. Какие статистики среднего позволяет рассчитывать окно Options?
12. Что такое Т- тест?
13. Какие виды Т – тестов предлагаются в SPSS?
14. Что такое одновыборочный t-критерий (One Sample T Test)?
15. Что такое t-критерий для независимых выборок?
16. Что такое t-критерий для парных выборок (Paired-Sample T Test)?
17. В чем состоит специфика каждого из трех видов Т - тестов?
18. В чем состоит специфика однофакторного дисперсионного анализа?
19. Какие дополнительные окна имеются в процедуре One Way ANOVA?

# Глава 11. Анализ связей

## 11.1. Описание корреляционной зависимости

В предыдущей главе (§ 10.3), двигаясь в последовательности представления процедур в системе SPSS, мы уже дали некоторые сведения о корреляции. Повторимся **- корреляция** - мера зависимости двух или нескольких величин. Это такое отношение между признаками, в котором в зависимости от изменения одного признака изменяется значение и другого признака. Область математической статистики, занятая изучением подобного рода отношений, называется корреляционным анализом. В основе корреляционного анализа лежит представление о типе, форме и тесноте (плотности) как важнейших свойствах связи.

**По типу** корреляционные связи делятся на положительные (прямые) и отрицательные (обратные). При положительной связи с увеличением (уменьшением) значений одного признака значения другого также возрастают (убывают). При отрицательной связи увеличение (уменьшение) одного признака вызывает уменьшение (увеличение) другого.

**По форме** корреляционные связи делятся на линейные (прямолинейные) и нелинейные (криволинейные). Отсюда появляется потребность в разнообразии коэффициентов корреляции. Одни из них описывают линейные функции, другие – нелинейные. Проверить характер функции можно с помощью диаграмм рассеяния.

**Теснота** (плотность) связи - степень сопряженности между двумя явлениями, признаками, величинами. Связь считается более тесной в том случае, когда каждому значению одного признака соответствуют близкие друг другу, тесно расположенные около своей средней величины, значения другого признака. Связь менее тесна, если эти значения заметно отклоняются (сильно варьируют) от среднего значения.

Корреляция бывает парная и множественная. Парная корреляция выявляет тип, форму и плотность связи между двумя признаками. Множественная корреляция выявляет взаимосвязь между несколькими признаками.

Перечисленные выше основные характеристики корреляционной связи: тип, форма и плотность - измеряются с помощью таких статистических характеристик, как коэффициент корреляции, корреляционное отношение, коэффициент регрессии и др. Первые два из них рассмотрены ниже в этой главе, а коэффициент регрессии рассмотрен в главе 13, § 13.1.

Коэффициент корреляции – безразмерная величина, значение которой не зависит от масштаба измерения. Как принято об этом писать в статистике, корреляция между ростом и весом будет одной и той же, независимо от того, выполнены ли измерения в сантиметрах и килограммах или в дюймах и фунтах.

Значения коэффициента корреляции (§ 10.3) находятся в интервале от –1 до +1. При этом –1 свидетельствует о полной отрицательной зависимости, +1 – о полной положительной зависимости, а 0 об отсутствии таковой. Более полную информацию по данному вопросу можно найти в любом учебнике по статистике (31, С. 119-126, или 32).

Еще один важный момент в анализе корреляций обусловлен различием истинных и ложных корреляций. Здесь и возникает огромная проблема, которая пока еще не обсуждается в социологии. В связи с тем, что ложная корреляция не очевидна, она может быть фиксирована либо путем продолжительных наблюдений и жизненного опыта, либо экспериментально. По вполне понятным причинам случай здесь плохой помощник.

Эта проблема вполне естественна, так как в социологии фактически отсутствует опыт экспериментальных расчетов, позволяющих формулировать гипотезу и, проверяя ее непосредственно в эксперименте, фиксировать ложную корреляцию, а также, что не менее важно для науки, накапливать информацию о таких корреляциях. Поэтому за рамками учебников по статистике, где в качестве примеров ложных корреляций можно встретить связь длины волос и роста, а также связь размеров ущерба от пожаров и числа пожарных, принявших участие в их тушении, социологам в этом плане сказать пока еще практически нечего.

Для социолога в первичном анализе материалов и файлов с данными полевых исследований фиксировать ложную корреляцию достаточно трудно, а главное, в этом пока еще нет никакой внутренней необходимости и заинтересованности. Исходного материала и данных всегда значительно больше, чем требуется для подготовки отчета или публикации. Со временем это достоинство первоначального этапа

развития социологической науки становится фундаментальным тормозом ее дальнейшего движения.

Социологическое сообщество, комфортно чувствующее себя в рамках первичного анализа, не стремится создавать благоприятные условия для распространения вторичного анализа, а, следовательно, и экспериментальных разработок. Как результат, каждое социологическое исследование продолжает рассматриваться в качестве новаторского, а база данных остается доступной только разработчикам.

Установка на необходимость открытия доступа к базам данных с первичной информацией вызывает негативную реакцию и отторжение. В их основе лежит известный эффект «собаки на сене», даже у лучших представителей социологического сообщества. Методология вторичного анализа обсуждается в узком кругу профессионалов (37), а вопросы развития экспериментальной социологии все еще ждут своей постановки. Отсюда вывод – эгоизм и ограниченность сознания социологов представляет собой основную угрозу развития социологической науки.

Прикрываясь соображениями о специфике социального знания, многие социологи фактически заняты разработкой и распространением идеологических догм. Такие специалисты далеки от поддержки стандартов научного знания. Более того, многие из них делают все, что могут, для дискредитации этих стандартов в общественном сознании. Учитывая важность проблематики, которая связана с экспериментальными расчетами в социологии, в том числе с поиском как действительных, так и ложных корреляций, мы еще вернемся к ней в этой главе.

В SPSS для целей анализа корреляционных связей используются несколько процедур. Направление их поиска: **Analyze - Correlate**. Далее, в выпадающем меню в зависимости от поставленной задачи, требуется сделать выбор между тремя процедурами: **Bivariate, Partial, Distances**.

## 11.2. Парная корреляций — Bivariate

**Bivariate** - основная процедура измерения корреляционной связи, используемая в социологии. С ее помощью измеряются парные корреляции. Путь, открывающий возможность выполнения этой процедуры, следующий:

**Analyze**

**Correlate**

**Bivariate.**

При выполнении указанной последовательности команд система открывает главное диалоговое окно процедуры Bivariate (рис. 49).

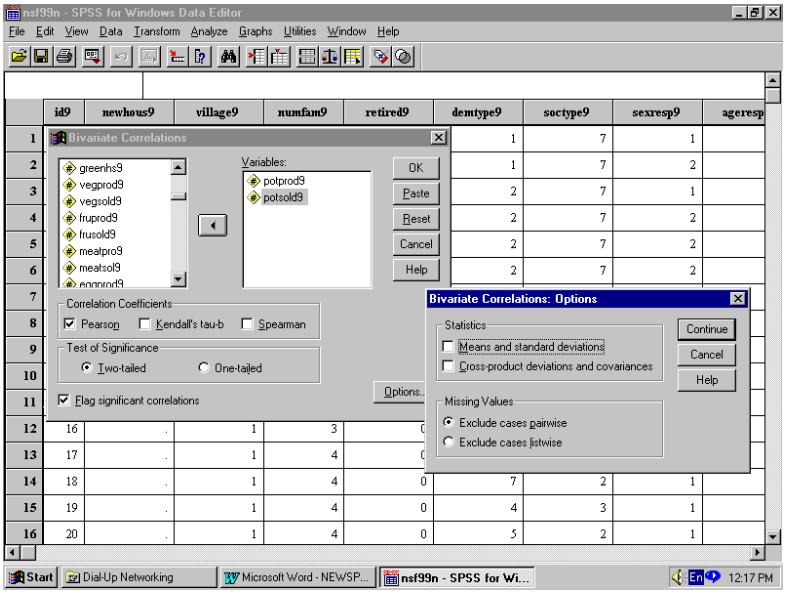



***Рис. 49.***

***Главное и дополнительное диалоговое окно Options процедуры Bivariate***

В открывшемся главном диалоговом окне данной процедуры, ранее уже многократно описанным способом, выбираем из левого поля интересующие переменные и переносим их для анализа в правое поле, которое называется Variables. Здесь, правда, есть один нюанс. Для выполнения этой процедуры система исходно требует переноса минимум двух переменных.

При первоначальном выделении одной переменной и ее переносе команда выполнения процедуры (OK) остается недоступной, что и свидетельствует о незавершенности действий пользователя по заданию условий для выполнения расчетов. К тому же, в данном случае система позволяет переносить и несколько переменных. При этом корреляция будет рассчитываться отдельно для каждой пары переменных, а в окне вывода результаты расчетов появятся в общей квадратной матрице.

Следующий важный шаг – выбор коэффициента корреляции. Он выполняется в первом нижнем контуре, который так и называется «Correlation Coefficients». Система дает возможность расчета трех коэффициентов корреляции: Пирсона (Pearson), Кендалла (Kendall's

tau-b) и Спирмена (Spearman).

**Напоминание**

***При выборе коэффициента корреляции важно помнить, что коэффициент Пирсона приспособлен только для числовых переменных и линейных связей, коэффициент Кендалла применим к переменным с упорядоченными и ранжированными числами и нелинейным связям, коэффициент Спирмена – к переменным с упорядоченными числами и нелинейным связям.***

Расчет одного, а при большом желании и всех трех коэффициентов можно задать путем пометки курсором стоящих напротив каждого из них маленьких чек-боксов. Затем в следующем нижнем контуре, который называется «Test of Significance», необходимо выбрать тип теста на значимость связи. Система предлагает два таких типа: двухсторонний - Two-tailed и односторонний- One-tailed. Оба они описаны ранее в § 10.3.

По умолчанию система самостоятельно задает расчет коэффициента Пирсона и двусторонний тест значимости, а также вводит пометку в окне просмотра значимой связи – «Flag significant correlations».

В дополнительном диалоговом окне Options можно задать такие статистики, как среднее, стандартное отклонение, ковариации. Последним шагом к выполнению команды, как и в других подобных случаях, следует нажатие кнопки ОК.

В табл. 23 приведены результаты расчета корреляционной зависимости для двух числовых переменных «выращено картофеля» и «продано картофеля» в выборке 1999 г. В данном случае экспериментаторами заданы только переменные, а все другие параметры расчетов использованы системой по умолчанию.

Из данных табл. 23 видно, что по умолчанию система рассчитывает и выводит три основные характеристики: значение коэффициента корреляции Пирсона, двухстороннюю значимость связи, а также число случаев в выборке (N=422 и 420, соответственно).

В этом эксперименте значение коэффициента корреляции Пирсона равно .900, т.е. оно очень высокое. А двухсторонняя значимость связи - Sig равна .000, что свидетельствует в пользу значимости связи, а, следовательно, и надежности корреляции.

**Таблица 23. Корреляционная зависимость (Correlations) для переменных производство** (POTPROD9) **и продажа** (POTSOLD9) **картофеля**

| | | **POTPROD9** | **POTSOLD9** |
|---|---|---|---|
| POTPROD9 | Pearson Correlation | 1.000 | .900** |
| | Sig. (2-tailed) | . | .000 |
| | N | 422 | 420 |
| POTSOLD9 | Pearson Correlation | .900** | 1.000 |
| | Sig. (2-tailed) | .000 | . |
| | N | 420 | 420 |

** Correlation is significant at the 0.01 level (2-tailed).

**Таблица 24. Корреляционная зависимость для переменных производство** (POTPROD9) **и продажа** (POTSOLD9) **картофеля**

**Nonparametric Correlations**

| | | | **POTPROD9** | **POTSOLD9** |
|---|---|---|---|---|
| Kendall's tau_b | POTPROD9 | Correlation Coefficient | 1.000 | .619 |
| | | Sig. (2-tailed) | . | .000 |
| | | N | 422 | 420 |
| | POTSOLD9 | Correlation Coefficient | .619 | 1.000 |
| | | Sig. (2-tailed) | .000 | . |
| | | N | 420 | 420 |
| Spearman's rho | POTPROD9 | Correlation Coefficient | 1.000 | .728 |
| | | Sig. (2-tailed) | . | .000 |
| | | N | 422 | 420 |
| | POTSOLD9 | Correlation Coefficient | .728 | 1.000 |
| | | Sig. (2-tailed) | .000 | . |
| | | N | 420 | 420 |

** Correlation is significant at the .01 level (2-tailed).

Если при задании параметров расчетов пометить для расчета еще и коэффициенты Кендалла и Спирмена, то в окне вывода можно увидеть данные, приведенные в табл. 24.

Обе табл. 23 и 24, представляющие окно вывода рассматриваемой процедуры, близки как по структуре, так и содержательно. Основное различие связано с несколько меньшим значением коэффициентов корреляции в табл. 24. Это различие обусловлено особенностями самих коэффициентов. Оно не меняет существа дела в данном эксперименте. В данном случае, который характеризуется использованием числовой переменной и линейной связью, коэффициент корреляции Пирсона адекватнее других показателей.

В основе расчета коэффициента корреляции лежит квадратная матрица. В этой матрице каждая переменная пересечена дважды – один раз сама с собой, а второй раз с другой переменной, поэтому в квадрантах по диагонали значения коэффициента корреляции и уровня значимости оказываются идентичными.

Наконец, еще один очень важный момент. Он связан с основанием выбора того или иного коэффициента корреляции. Такое основание может быть получено путем предварительного построения и анализа диаграммы рассеяния для коррелируемых переменных (график 1) (порядок работы с графиками рассмотрен в следующей главе).

Построение диаграммы рассеяния для коррелируемых переменных позволяет контролировать два основных параметра, важных для принятия решения, связанного с выбором того или иного коэффициента корреляции, а именно:

- однородность выборки;
- характер функции.

На диаграмме рассеяния (график 1), которая получена путем выполнения последовательности команд: Graphs – Scatter – Simple, отображены отношения между двумя рассматриваемыми переменными. Она дает возможность видеть, что в двух выбранных для эксперимента переменных наблюдается высокая однородность выборки и линейность функции.

Наличие однородности в выборке подтверждается тем фактом, что на графике рассеяния оба распределения представляют собой как бы единое целое. Наблюдается только один выброс.

Линейность функции подтверждается тем фактом, что в выборке фактически все случаи, включая и выброс, тяготеют к невидимой прямой регрессии, идущей от пересечения осей координат снизу вверх. Отсюда и вывод, который был использован ранее - коэффициент корреляции Пирсона наиболее адекватен для данного эксперимента.

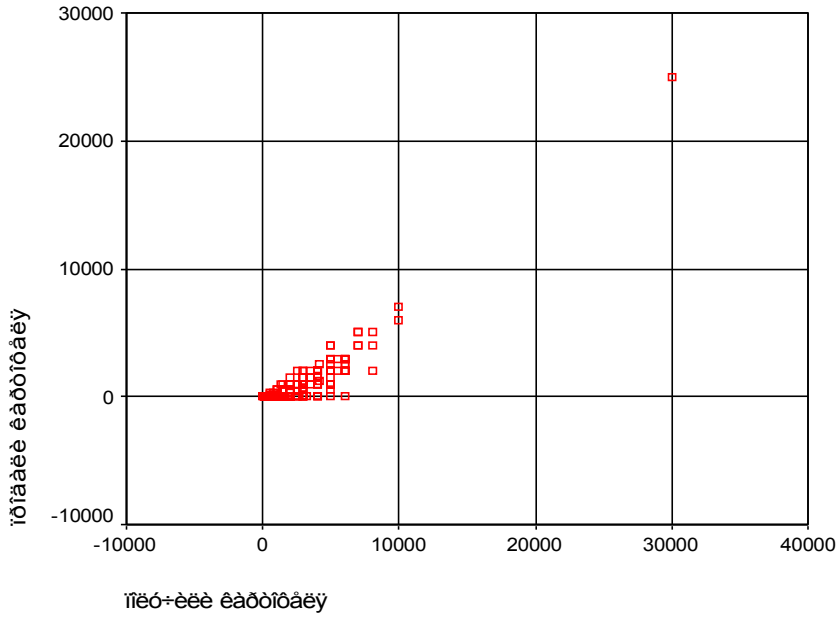



**График 1. Диаграмма рассеяния для переменных производство и продажа картофеля**

**Правило 32**

***Уровень значимости, вычисленный для каждой корреляции, представляет собой главный источник информации о ее надежности. Поэтому ссылка в тексте или таблице на значение коэффициента корреляции должна делаться вместе с указанием значимости связи.***

## 11.3. Частная корреляция - Partial

Частная корреляция (**Partial**) – процедура множественной корреляции, позволяющая использовать при расчетах коэффициента корреляции дополнительную, контрольную переменную. Использование этой процедуры дает возможность определения **ложной корреляции**. Ложная корреляция связана с влиянием на зависимость между парой переменных третьей переменной, которая в случае измере-

ния парной корреляции выпадает из рассмотрения.

В сравнении с парной, частная корреляция, благодаря вводу контрольной переменной, расширяет условия эксперимента. В этом случае из измерения изымается общая, т.е. пересекающаяся с контрольной переменной, часть корреляционного отношения. Отсюда, как следствие, уменьшение значения коэффициента корреляции.

В рассматриваемой процедуре используется специальный коэффициент - partial correlation coefficients, с помощью которого фиксируется только **линейная связь** между парой переменных, в то время как они испытывают влияние одной или нескольких добавочных переменных, используемых в качестве контрольных. Partial correlation coefficients – безразмерная величина, которая имеет значение в интервале от –1 до +1.

Путь, открывающий возможность выполнения этой процедуры:

**Analyze**

**Correlate**

**Partial.**

При выполнении указанной последовательности команд система открывает главное диалоговое окно этой процедуры (рис. 50).

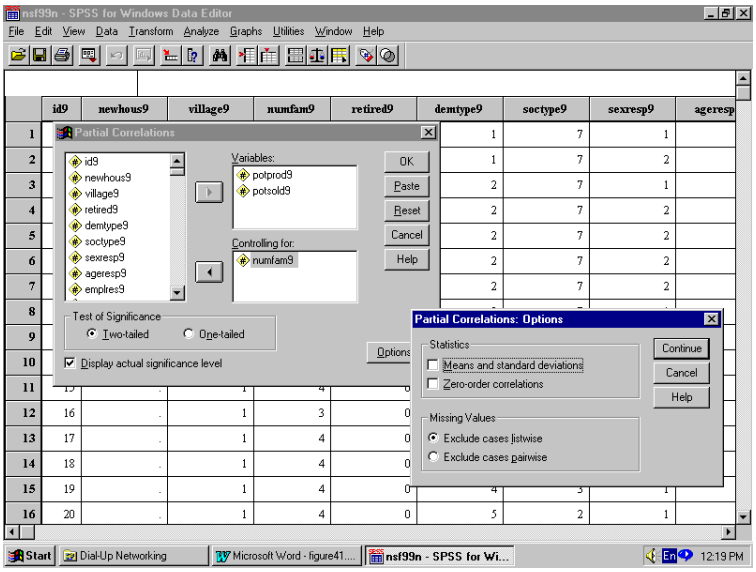



***Рис. 50.***

***Главное и дополнительное диалоговое окно Options процедуры Partial***

На рис. 50 можно видеть как главное, так и дополнительное диалоговое окно процедуры Partial, которые в целях экономии места показаны вместе. В отличие от главного диалогового окна предшествующей

процедуры в данном случае имеется третий список – «Controlling for», который и необходимо использовать для переноса в него контрольной(ых) переменной(ых).

Другое важное отличие – отсутствие в данном случае возможности выбора коэффициента корреляции. Как уже отмечено выше, «partial correlation coefficients» – один на все случаи жизни, а потому и задан системой по умолчанию.

Дополнительное диалоговое окно Options отличается от аналогичного в процедуре парной корреляции только предложением статистики Zero-order correlations вместо Cross-product deviations and covariances. Эта статистика позволяет получить матрицу простых корреляций между всеми переменными, включая контрольную. В отличие от других процедур, где, как уже отмечалось ранее, в последних версиях системы окно вывода имеет табличный формат, в этой процедуре даже в версии SPSS 10.0 окно просмотра сохранило формат близкий к версии 6.0. Поэтому оно и дается в обрамлении 32.

В этом обрамлении показаны результаты определения корреляционной зависимости двух переменных: «производство картофеля» и «продажа картофеля» с использованием контрольной переменной «совокупный доход домохозяйства» в выборке 1999 г. Все эти три переменные – числовые. В дополнительном диалоговом окне установлены опции Mean and standard deviations и Zero Order Partials.

Из данных, приведенных в этом обрамлении, видно, что система выполнила три вида запрашиваемых расчетов. В первом блоке даются среднее и стандартное отклонение, а также число случаев для всех переменных. Во втором блоке представлена матрица Zero Order Partials. И, наконец, в третьем блоке показаны непосредственно коэффициенты корреляции.

В предыдущем параграфе в качестве примера для расчета корреляций были использованы те же переменные. Если сравнить полученные коэффициенты корреляции, то можно более наглядно представить себе смысл частной корреляции. В парной корреляции была получена очень высокая зависимость (.900) между производством и продажей картофеля. А что произойдет, если попытаться выявить влияние какой-то переменной на эту связь? В качестве такой контрольной переменной в нашем примере выбрана переменная «суммарный доход». Как видно из обрамления 32, корреляционная зависимость осталась по-прежнему очень высокой, хотя и несколько ослабла (.892).

**Обрамление 32. Корреляция переменных «производство картофеля» (POTPROD9) и «продажи картофеля» (POTSOLD9) по контрольной переменной «совокупный доход домохозяйства» (SUMTOTA9)**

| Variable | Mean | Standard Dev | Cases |
|---|---|---|---|
| POTPROD9 | 1949.3643 | 2260.0834 | 420 |
| POTSOLD9 | 562.7381 | 1559.6871 | 420 |
| SUMTOTA9 | 2893.6500 | 2181.7252 | 420 |

**- - - P A R T I A L  C O R R E L A T I O N  C O E F F I C I E N T S - - -**

Zero Order Partials

| | POTPROD9 | POTSOLD9 | SUMTOTA9 |
|---|---|---|---|
| POTPROD9 | 1.0000 | .9005 | .3709 |
| | ( 0) | ( 418) | ( 418) |
| | P= . | P= .000 | P= .000 |
| POTSOLD9 | .9005 | 1.0000 | .2948 |
| | ( 418) | ( 0) | ( 418) |
| | P= .000 | P= . | P= .000 |
| SUMTOTA9 | .3709 | .2948 | 1.0000 |
| | ( 418) | ( 418) | ( 0) |
| | P= .000 | P= .000 | P= . |

(Coefficient / (D.F.) / 2-tailed Significance)
" . " is printed if a coefficient cannot be computed

**- - - P A R T I A L  C O R R E L A T I O N  C O E F F I C I E N T S - - -**

Controlling for.. SUMTOTA9

| | POTPROD9 | POTSOLD9 |
|---|---|---|
| POTPROD9 | 1.0000 | .8915 |
| | ( 0) | ( 417) |
| | P= . | P= .000 |
| POTSOLD9 | .8915 | 1.0000 |
| | ( 417) | ( 0) |
| | P= .000 | P= . |

(Coefficient / (D.F.) / 2-tailed Significance)
**" . "** is printed if a coefficient cannot be computed

В обрамлении 33 представлены расчеты корреляции для трех переменных «производство картофеля», «продажа картофеля», «производство мяса» по контрольной переменной «суммарный доход» без использования окна Options. Из этого обрамления видно, что в данном случае рассчитываются только коэффициенты корреляции и уровень значимости связи переменных.

**Обрамление 33. Корреляция переменных «производство картофеля» (POTPROD9), «продажа картофеля» (POTSOLD9), «производство мяса» (MEATPRO9) по переменной «совокупный месячный доход домохозяйства» (SUMTOTA9)**

```
- - - P A R T I A L   C O R R E L A T I O N   C O E F F I C I E N T S - - -

Controlling for..  SUMTOTA9

                POTPROD9   POTSOLD9   MEATPRO9
POTPROD9         1.0000      .8915      .1472
                 (    0)    ( 417)     ( 417)
                 P= .       P= .000    P= .003
POTSOLD9          .8915     1.0000      .1293
                 ( 417)     (    0)    ( 417)
                 P= .000    P= .       P= .008
MEATPRO9          .1472      .1293     1.0000
                 ( 417)     ( 417)     (    0)
                 P= .003    P= .008    P= .

(Coefficient / (D.F.) / 2-tailed Significance)
" . " is printed if a coefficient cannot be computed
```

Понимая важность постановки вопроса о фиксации ложных корреляций в социологии, мы, как и обещали в предыдущем параграфе (§ 11.1), проделали огромную экспериментальную работу в поисках примера ложной корреляции непосредственно в своих массивах данных.

Действительно, «кто ищет, тот всегда найдет». В табл. 25 приведен приер расчета коэффициента парной корреляции Пирсона (процедура Bivariate) для переменных, фиксирующих наличие свиней (число) и производства мяса (кг) в выборке сельских домохозяйств в 1999 г.

**Таблица 25. Корреляционная зависимость для переменных «наличие свиней» (PIG9) и «производство мяса» (MEATPRO9) в сельских домохозяйствах в 1999 г.**

**Correlations**

| | | PIG9 | MEATPRO9 |
|---|---|---|---|
| PIG9 | Pearson Correlation | 1.000 | .710 |
| | Sig. (2-tailed) | | .000 |
| | N | 422 | 422 |
| MEATPRO9 | Pearson Correlation | .710 | 1.000 |
| | Sig. (2-tailed) | .000 | . |
| | N | 422 | 422 |

** Correlation is significant at the 0.01 level (2-tailed).

А в обрамлении 34 приведен пример расчета коэффициента частной корреляции (процедура Partial) для тех же переменных, фиксирующих наличие свиней и производство мяса. При этом в качестве контрольной переменной (и в целях сохранения однородности выборки) здесь взята «продажа мяса» теми же сельскими домохозяйствами и в том же 1999 г.

**Обрамление 34. Частная корреляционная зависимость для переменных «наличие свиней»** (PIG9) **и «производство мяса»** (MEATPRO9) **по контрольной переменной «продажа мяса»** (MEATSOL9)

```
- - - P A R T I A L   C O R R E L A T I O N   C O E F F I C I E N T S - - -

Controlling for..  MEATSOL9

                    PIG9            MEATPRO9
PIG9                1.0000           .2093
                    (    0)          ( 381)
                    P= .             P= .000
MEATPRO9           .2093            1.0000
                   ( 381)            (    0)
                    P= .000          P= .

(Coefficient / (D.F.) / 2-tailed Significance)

" . " is printed if a coefficient cannot be computed
```

В соответствии с расчетами, представленными в табл. 25, видно, что для переменных «наличие свиней» (PIG9) и «производство мяса» (MEATPRO9) фиксируется положительная корреляционная зависимость на уровне значения коэффициента Пирсона равного .710. Это достаточно хорошая зависимость. На первый взгляд, она кажется тривиальной. Действительно, вряд ли кому в сельской местности придет в голову держать в это трудное время свиней в качестве хобби или домашнего зоопарка.

Расчет частной корреляции этих же переменных по контрольной переменной «продажа мяса» показывает, что между исходными переменными существует положительная зависимость на уровне значения коэффициента частной корреляции равного .209. Эта зависимость намного слабее по сравнению с предыдущей зависимостью. Вместе с тем, и в том, и в другом случае уровень значимости связи очень высок (p=.000).

Возникающая здесь коллизия может быть описана следующим образом: хотя для переменных «наличие свиней» и «производство мяса» наблюдается высокая корреляционная зависимость, все же свиней в

домохозяйствах держат не для потребления, а, главным образом, для продажи свинины. Изымая из корреляционного отношения значительную часть случаев, отвечающих заданному критерию: наличие свиней – производство - продажа мяса, мы и фиксируем низкую связь для случаев, имеющих чистый вид: наличие свиней - производство мяса.

Понимание этого факта позволяет совершенно по-другому и в более обобщенном виде воспринять следующие высказывания некоторых опрашиваемых селян: «Выращиваешь поросенка целый год, а нам потом достается хвост и ливер. Все остальное идет на продажу. Деньги нужны, а взять их негде: зарплату-то совсем не дают».

К тому же и статистически связь между наличием свиней и производством мяса при внимательном рассмотрении оказывается достаточно сложной. Во-первых, график рассеяния для переменных «наличие свиней» и «производство мяса» (график 2) выглядит совершенно по-другому, чем аналогичный график для переменных «производство картофеля» и «продажа картофеля» (график 1).

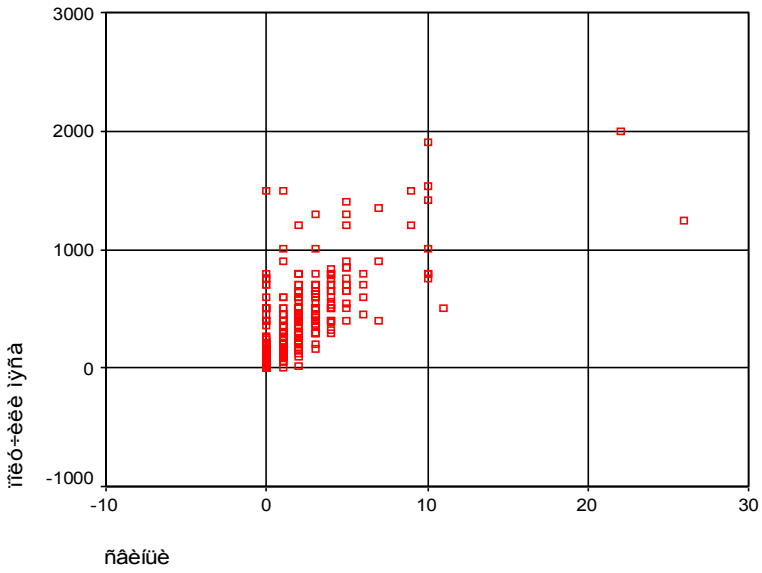



**График 2. Диаграмма рассеяния для переменных наличие свиней и производство мяса**

Слойный характер рассеяния случаев как по вертикали, так и по горизонтали показывает, что имеется много случаев, когда нет свиней, а есть производство мяса. Об этом говорит нижняя горизонталь рас

сеяния случаев. То же самое повторяется и при наличии одного, двух и более выкармливаемых поросят (слои по вертикали). И это естественно, так как селяне, кроме свиней, держат на мясо еще и телят, и птицу (гусей, уток, кур и др.). Во-вторых, что не менее важно, есть случаи, когда свиней много, а произведено мяса меньше (т.е. в домохозяйстве есть хряк, свиноматка и масса поросят на продажу). Наличие таких случаев еще сильнее искривляет связь, делая ее криволинейной (в этом плане очень полезно еще раз сравнить диаграммы рассеяния на графиках 1 и 2), что и требует отказа от использования коэффициента корреляции Пирсона.

**Правило 33**

***Общий принцип процедуры Partial – измерение корреляционной зависимости пары переменных посредством использования дополнительной (третьей) контрольной переменной.***
***Частная корреляция – основной инструмент поиска ложных корреляций.***

## 11.4. Мера сходства или различия - Distances

Мера сходства или различия (**Distances**) - процедура, выполняющая расчет для статистик, измеряющих сходства (similarities) и различия (dissimilarities) как между парой переменных, так и между случаями, составляющими ту или иную переменную. Обычно она используется в качестве вспомогательной совместно с другими процедурами, предлагаемыми системой в разделе главного меню Analyze, такими как факторный анализ, кластерный анализ и др. Эта процедура используется только для числовых (numeric) переменных.

В системе SPSS эта процедура появилась только в поздних версиях, начиная с версии SPSS 7.5. Интересно, что в российском издании руководства для пользователей этой версии отмечается, что в подменю «Correlate» есть три пункта, но рассматриваются только два: парная и частная корреляции (8, С. 161). Мы не знаем случаев самостоятельного использования этой процедуры в социологии. Тем не менее, ниже дано ее краткое описание. Для открытия этой процедуры используется следующий путь:

**Analyze**
**Correlate**
**Distances**.

При выполнении указанной последовательности команд система открывает главное диалоговое окно этой процедуры (рис. 51). В открывшемся главном диалоговом окне процедуры, ранее уже многократно описанным способом, выбираем из левого поля интересующие переменные и переносим их для анализа в правое поле.

Это поле называется «Variables». Здесь есть один не совсем маленький нюанс. А именно, необходимо заранее принять решение, что будет рассчитываться: мера различия между случаями или переменными. Отсюда открываются две возможные последовательности действий.

Первая последовательность действий связана с расчетом меры различия между случаями. Для этого необходимо выбрать в левом списке одну (или более) переменную и считать distances между случаями (эта возможность задана по умолчанию в контуре Compute Distances).

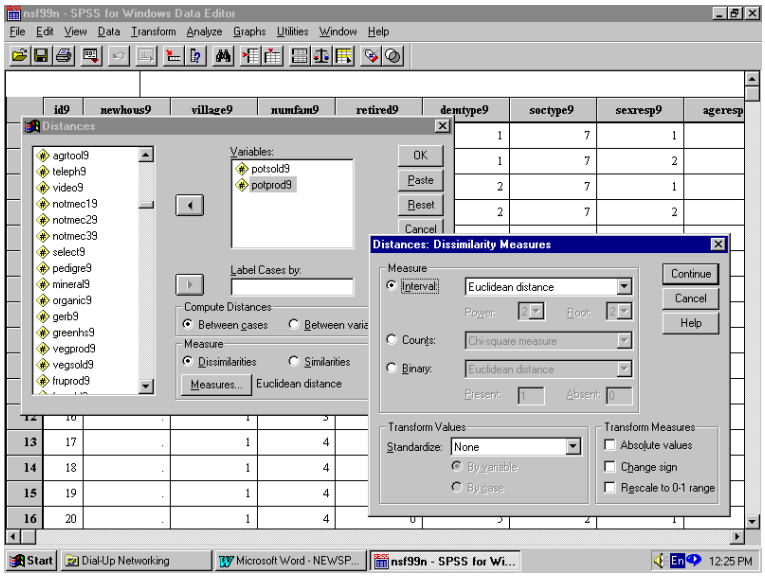



***Рис. 51.***

***Главное и дополнительное диалоговое окно процедуры Distances***

Вторая последовательность действий связана с расчетом меры различия между переменными. Для этого необходимо выбрать в левом поле две или более переменных и считать distances между переменными, пометив предварительно в контуре Compute Distances опцию «между переменными» - Between variables.

По умолчанию задан также и расчет различий (dissimilarities) в контуре Measure. Если в этом контуре пометить альтернативную опцию

Similarities (сходство), то анализ будет проводиться уже в направлении поиска сходства.

В дополнительном диалоговом окне, которое открывается после нажатия на кнопку Measures (в контуре с тем же названием), можно задать расчет различных статистик расстояния, изменить шкалу, выделить группы и т.п. При этом каждая статистика должна выбираться в зависимости от установленной ранее опции - сходство или различие.

Если выбираются сходства (similarities), то рядом с кнопкой Measures высвечивается надпись Pearson correlation. И тогда при нажатии на указанную кнопку открывается дополнительное диалоговое окно (similarities Measures), в котором в контуре Measure по умолчанию помечена опция Interval (есть еще опция Binary), а в прямоугольном боксе задан расчет Pearson correlation.

В случае выбора статистики различия (dissimilarities) рядом с кнопкой Measures высвечивается надпись «Euclidean distance». При использовании этого выключателя открывается дополнительное диалоговое окно (dissimilarities Measures), а в контуре Measure по умолчанию устанавливается опция Interval, которая открывает возможность расчета «Euclidean distance».

Для наглядности показа работы этой процедуры выбрано небольшое число случаев. Если попытаться использовать весь массив (в 422 случая) и показать различия между случаями, то в приводимом примере машина средней мощности затратит на выполнение подобных расчетов много времени. В результате она построит матрицу 422 X 422 случая, которая по условию не может быть сразу открыта в окне просмотра.

Сначала будет выдан первый фрагмент (100 по вертикали и 80 по горизонтали), который надо расширять, «кликая» мышью специальные красные стрелки на концах матрицы. Попытка распечатать такое окно просмотра займет не один десяток страниц.

В стремлении обойти эти трудности, в качестве примера в базе данных в выборке 1999 г берутся только случаи «месячного денежного дохода семьи» (total9), меньшие или равные 377 руб. Для этого используется процедура сортировки данных (Sort Cases) по возрастанию переменной total9 (глава 3, § 3.7 и глава 5, § 5.2). Все случаи переменной total9, которые меньше или равны 377, отбираются с помощью процедуры Select Cases (глава 5, § 5.2). Таких случаев в выборке имеется 7, а их значения распределяются следующим образом:

300 руб., 304 руб., 329 руб., 375 руб., 375 руб., 377 руб, 377 руб.

Затем в контуре «Compute Distances» главного диалогового окна процедуры Distances принимается установленная по умолчанию опция расчета distances между случаями (between cases). А в контуре Measure устанавливается опция различия (Dissimilarities). Последней, как всегда, следует команда OK.

Результаты выполнения команды, полученные в окне просмотра в виде двух таблиц, приведены ниже. Первая таблица окна просмотра имеет информативный характер (табл. 26). В ней содержатся данные о числе случаев и пропущенных значениях.

Вторая таблица окна просмотра представляет собой матрицу пересечений всех заданных случаев (табл. 27). В ней рассчитываются расстояния, в данном примере - это различия, между каждой парой случаев: первый и второй отличаются на 4 руб. (304-300=4), первый и третий – на 29 руб. (329-300=29), а шестой и седьмой случаи не имеют различий (377-377=0) и т.д.

**Таблица 26. Различия (dissimilarities) между случаями переменной «месячный денежный доход семьи» (total9) (первая таблица окна просмотра)**

**Case Processing Summary**

| Cases | | | | | |
|---|---|---|---|---|---|
| Valid | | Missing | | Total | |
| N | Percent | N | Percent | N | Percent |
| 7 | 100.0% | 0 | .0% | 7 | 100.0% |

**Таблица 27. Различия (dissimilarities) между случаями переменной «месячный денежный доход семьи» (total9) (вторая таблица окна просмотра)**

**Proximity Matrix**

| | Euclidean Distance | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 1 | | 4.000 | 29.000 | 75.000 | 75.000 | 77.000 | 77.000 |
| 2 | 4.000 | | 25.000 | 71.000 | 71.000 | 73.000 | 73.000 |
| 3 | 29.000 | 25.000 | | 46.000 | 46.000 | 48.000 | 48.000 |
| 4 | 75.000 | 71.000 | 46.000 | | .000 | 2.000 | 2.000 |
| 5 | 75.000 | 71.000 | 46.000 | .000 | | 2.000 | 2.000 |
| 6 | 77.000 | 73.000 | 48.000 | 2.000 | 2.000 | | .000 |
| 7 | 77.000 | 73.000 | 48.000 | 2.000 | 2.000 | .000 | |

This is a dissimilarity matrix

Если при установке параметров расчетов опции различия (dissimilarities), между случаями задать более одной переменной, то во второй таблице окна просмотра будут показаны различия в паре переменных между каждой парой случаев (табл. 28).

При установке опции сходства (Similarities) для одной переменной вторая таблица окна вывода показывает корреляцию между векторами значений. В этом случае вторая таблица окна просмотра имеет вид, который показан на табл. 29.

**Таблица 28. Различия (dissimilarities) между случаями пары переменных «месячный денежный доход семьи» и «совокупный месячный доход семьи»**

**Proximaty Matrix**

| | Euclidean Distance | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 1 | | 86.093 | 269.564 | 294.703 | 294.703 | 372.055 | 279.803 |
| 2 | 86.093 | | 354.882 | 377.733 | 377.733 | 455.883 | 362.428 |
| 3 | 269.564 | 354.882 | | 49.041 | 49.041 | 107.331 | 48.010 |
| 4 | 294.703 | 377.733 | 49.041 | | .000 | 79.025 | 16.125 |
| 5 | 294.703 | 377.733 | 49.041 | .000 | | 79.025 | 16.125 |
| 6 | 372.055 | 455.883 | 107.331 | 79.025 | 79.025 | | 95.000 |
| 7 | 279.803 | 362.428 | 48.010 | 16.125 | 16.125 | 95.000 | |

This is a dissimilarity matrix

**Таблица 29. Сходства (similarities) между случаями переменной total9**

**Proximity Matrix**

| | Correlation between Vectors of Values | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 1 | | 9999.999 | 9999.999 | 9999.999 | 9999.999 | 9999.999 | 9999.999 |
| 2 | 9999.999 | | 9999.999 | 9999.999 | 9999.999 | 9999.999 | 9999.999 |
| 3 | 9999.999 | 9999.999 | | 9999.999 | 9999.999 | 9999.999 | 9999.999 |
| 4 | 9999.999 | 9999.999 | 9999.999 | | 9999.999 | 9999.999 | 9999.999 |
| 5 | 9999.999 | 9999.999 | 9999.999 | 9999.999 | | 9999.999 | 9999.999 |
| 6 | 9999.999 | 9999.999 | 9999.999 | 9999.999 | 9999.999 | | 9999.999 |
| 7 | 9999.999 | 9999.999 | 9999.999 | 9999.999 | 9999.999 | 9999.999 | |

This is a similarity matrix

Если выбрано более одной переменной, то вторая таблица окна просмотра по результатам расчетов имеет вид сходный с табл. 30.

**Таблица 30. Сходства (similarities) между случаями пары переменных «месячный денежный доход семьи» и «совокупный месячный доход семьи»**

**Proximaty Matrix**

| | Correlation between Vectors of Values | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 1 | | .000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 2 | .000 | | .000 | .000 | .000 | .000 | .000 |
| 3 | 1.000 | .000 | | 1.000 | 1.000 | 1.000 | 1.000 |
| 4 | 1.000 | .000 | 1.000 | | 1.000 | 1.000 | 1.000 |
| 5 | 1.000 | .000 | 1.000 | 1.000 | | 1.000 | 1.000 |
| 6 | 1.000 | .000 | 1.000 | 1.000 | 1.000 | | 1.000 |
| 7 | 1.000 | .000 | 1.000 | 1.000 | 1.000 | 1.000 | |

This is a similarity matrix

При выборе в контуре «Compute Distances» главного диалогового окна рассматриваемой процедуры опции Distances between variables определяются различия между каждой парой переменных (табл.31).

**Таблица 31. Различия (dissimilarities) между переменными «месячный денежный доход семьи» (TOTAL9) и «совокупный месячный доход семьи» (SUMTOTA9)**

**Proximity Matrix**

| | Euclidean Distance | |
|---|---|---|
| | TOTAL9 | SUMTOTA9 |
| TOTAL9 | | 719.718 |
| SUMTOTA9 | 719.718 | |

This is a dissimilarity matrix

А при выборе в контуре Measure опции Similarities определяются сходства между каждой парой переменных (табл. 32).

**Таблица 32. Сходства (similarities) между переменными TOTAL9 и SUMTOTA9**

**Proximity Matrix**

| | Correlation between Vectors of Values | |
|---|---|---|
| | TOTAL9 | SUMTOTA9 |
| TOTAL9 | | .882 |
| SUMTOTA9 | .882 | |

This is a similarity matrix

Для примера, в наших расчетах использованы переменные «месячный денежный доход семьи» (TOTAL9) и «совокупный месячный доход семьи» (SUMTOTA9). Результаты выполнения расчетов различия и сходства показаны в табл. 31-32. В данном случае, как мы уже отмечали ранее, для выполнения расчетов требуются минимум две переменные.

Внимательный пользователь, посмотрев на главное диалоговое окно рассматриваемой процедуры (рис 51), может заметить, что в тексте выше отсутствует описание поля «Label Cases by». Оно имеет кнопку управления со стрелкой вправо. Наши попытки использовать его для каких-либо целей не имели успеха.

**Правило 34**

***Общий принцип процедуры «Мера сходства или различия» (Distances) - выполнение расчетов, позволяющих измерять сходства (similarities) и различия (dissimilarities) между случаями, составляющими переменную, и между парой переменных.***

**Дружеский совет**

***При описании конкретной корреляционной связи всегда полезно помнить и давать в тексте информацию о ее важнейших свойствах:***

***- типе (положительная или отрицательная);***
***- форме (линейная или нелинейная);***
***- характере (парная или множественная);***
***- используемом при ее измерении коэффициенте корреляции;***
***- тесноте (плотности связи) – числовое значение используемого коэффициента корреляции;***
***- значимости связи (Sig).***

## Задание для самостоятельной работы

1. Что такое корреляция?
2. Назовите основные характеристики корреляционной связи.
3. Какие типы корреляционной связи вы знаете?
4. Какие формы имеют корреляционные связи?

5. Что такое теснота (плотность) корреляционной связи?
6. Что такое парная корреляция?
7. Что такое множественная корреляция?
8. Что такое коэффициент корреляции?
9. Какие коэффициенты корреляции вы знаете?
10. Какая размерность у различных коэффициентов корреляции?
11. Что такое ложная корреляция?
12. Какие коэффициенты корреляции можно рассчитать в SPSS?
13. Как в SPSS из главного меню можно пройти к подменю Correlate?
14. Какие виды корреляции рассчитываются с помощью подменю Correlate?
15. Что такое процедура Bivariate?
16. В чем особенность диалоговых окон процедуры Bivariate?
17. Для каких чисел используется коэффициент корреляции Пирсона?
18. В чем состоит специфика процедуры Partial?
19. В чем особенность диалоговых окон процедуры Partial?
20. Для каких целей полезно использовать процедуру Distances?
21. В чем особенность диалоговых окон процедуры Distances?
22. Назовите основные шаги, необходимые для расчета корреляции.

# Глава 12. Графическое представление данных

## 12.1. Подменю Graphs

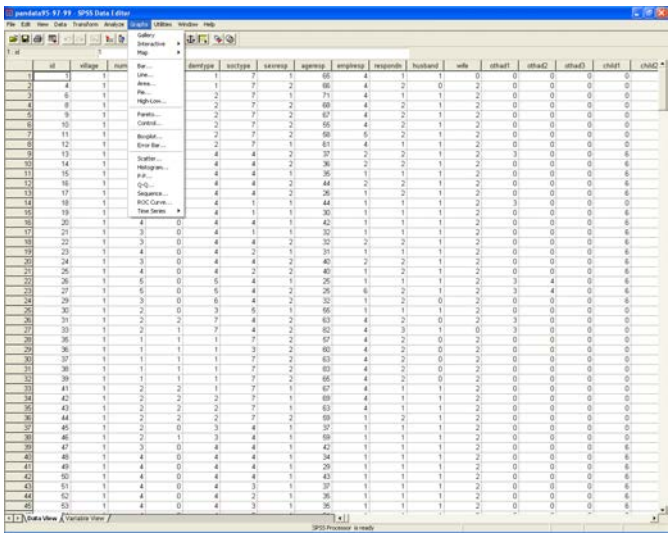
Графики - это возможность визуального представления и изучения данных. В SPSS графики можно получать двумя основными способами. Первый способ связан с использованием подменю **Graphs** (графики) в главном меню системы**.**

Второй способ связан с использованием дополнительных диалоговых окон в отдельных статистических процедурах, таких как, например, частоты (Frequencies), об этом упоминалось в главе 7, § 7.4 и главе 8, § 8.1, или исследовательские статистики (Explore) – глава 8, § 8.3.

SPSS представляет широкие возможности получения графиков для анализа социологических данных. В различных версиях SPSS от базовой 6.1 до 11.5 раздел главного меню Graphs содержит обширный перечень графиков, доступных для построения (рис. 52)

***Рис. 52.***

***Выпадающий список графиков раздела главного меню Graphs***



На рис. 52 можно видеть полный перечень графических процедур меню Graphs в получившей сегодня наиболее широкое распространение версии SPSS 10.0. Ниже дано краткое описание их содержания.

Более полно отдельные категории графиков будут описаны далее как в этом, так и в других параграфах. В порядке представления команд и графиков в подменю Graphs (рис. 52) следует назвать:

**Gallery –** сервисная команда. Она позволяет, во-первых, видеть образцы различных графиков с их названиями и, во-вторых, переходить к их описанию в справочной системе (Help). Этот переход осуществляется посредством двойного клика мышью по иконке того или иного графика.

**Interactive** – интерактивная графика. Она дает возможность взаимодействия пользователя с графиками. Это мощное средство, с которым имеет смысл начинать работать только после освоения стандартных графиков, представленных ниже.

**Map –** устанавливает графики и диаграммы на географической карте. Как и интерактивные графики, этот сервис не для начинающих пользователей. Он предполагает наличие особого типа данных, привязанных к местности, и понимание географических электронных систем.

Далее следуют команды, позволяющие строить различные графики и диаграммы:

**Bar** - столбиковая диаграмма;

**Line** - линейный график;

**Area -** площадной график;

**Pie -** круговая диаграмма;

**High-Low -** уровневый график;

**Pareto -** график Парето;

**Control -** график контрольная;

**Boxplot -** ящичковая диаграмма;

**Error Bar -** график ошибки в столбиковой диаграмме;

**Scatter -** график рассеяния;

**Histogram -** гистограмма;

**P-P -** кумулятивный (comulative) график относительно нормального распределения;

**Q-Q -** график квантилей (quantiles) относительно нормального распределения;

**Sequence -** график последовательности распределения случаев в массиве;

**Time Series** - графики временных серий (Autocorrelations, Cross-Correlations).

Эти графики присутствуют практически во всех версиях SPSS, поэ-

тому их часто называют стандартными. Стандартные графики можно разбить на две неравные группы. Одну из них представляют графики и диаграммы, использующиеся в первую очередь в учебных и презентационных целях. Сюда относятся столбиковые и круговые диаграммы, линейные графики. Другая большая группа графиков, в первую очередь, служит инструментом познания в научно-исследовательской и инженерной деятельности. Сюда относятся гистограммы, ящичковые диаграммы, графики рассеяния и практически все другие графики из приведенного выше списка.

Разумеется, не все эти графики могут найти широкое распространения в социологии и социально-экономических исследованиях. Тем не менее, использование графиков в научно-исследовательских расчетах и представлении их результатов может рассматриваться в качестве надежного индикатора состояния дел и уровня подготовки кадров в той или иной науке. Тот факт, что в нашей стране в социологических публикациях графики пока еще нашли ограниченное распространение, свидетельствует далеко не в пользу социологического сообщества.

## 12.2. Построение графиков

В SPSS при построении графиков почти постоянно используется одна и та же последовательность команд и действий. Поэтому главное в работе с графиками понимание целевого назначения конкретного графика и освоение прототипа их построения. Последнему моменту и уделено основное внимание в предлагаемом параграфе, тогда как содержательное понимание графиков требует предварительной подготовки.

Для графического представления данных обычно используется гистограмма. Гистограмма - это график, в котором по оси X располагаются значения анализируемой переменной, разделенные на равные интервалы (ширина столбика), а по оси Y - количество значений переменной, попадающих в каждый интервал (высота столбика). Значения, расположенные по горизонтальной оси, являются средними точками диапазона значений. Каждый столбец в гистограмме представляет число наблюдений, значения которых попадает в соответствующий интервал.

При построении гистограммы используется следующая последовательность команд:

**Graphs**

**Histogram.**

Открывшееся в результате выполнения этих команд главное диалоговое окно процедуры Histogram можно видеть на рис. 53.

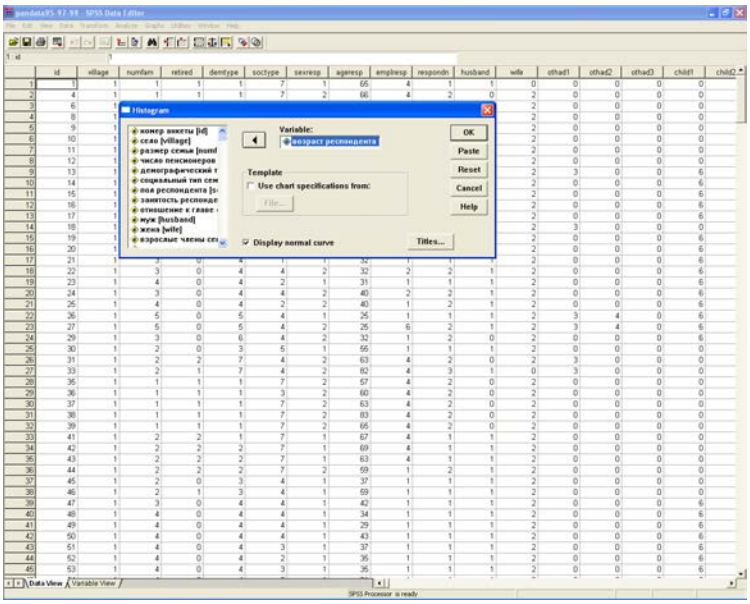



***Рис. 53.***

***Главное диалоговое окно процедуры Histogram -гистограмма***

В этом окне из находящегося слева списка переменных следует выбрать нужную переменную, выделить ее и перенести с помощью стрелки вправо, в центральное верхнее поле Variable, и нажать кнопку OK.

Гистограмма возраста респондента, например, будет содержать по оси X возраст, разбитый на интервалы по 5 лет (график 3). А по оси Y - число случаев встречаемости данного значения признака в массиве. Это значит, что каждый столбец содержит возраст: плюс - минус 2 года от значения, отмеченного на оси. Так, первый столбец содержит возраст респондента, колеблющийся в интервале от 18 до 22 лет. По вполне понятным причинам таких респондентов в нашем массиве меньше 10. Напротив, респондентов в возрасте около 35 лет и около 65 лет больше всего в выборке – 55 и 52, соответственно. И это имеет смысл, так как выборка строилась по главам сельских домохозяйств.

Другое дело, что распределение по возрасту респондента оказалось двугорбым (бимодальным). Об этом свидетельствует провал в возрастной группе около 50 лет. Указанное обстоятельство говорит о

неоднородности выборки, которая должна учитываться исследователями в работе с нею. Оно особенно четко фиксируется при наложении на график кривой нормального распределения (график 4)

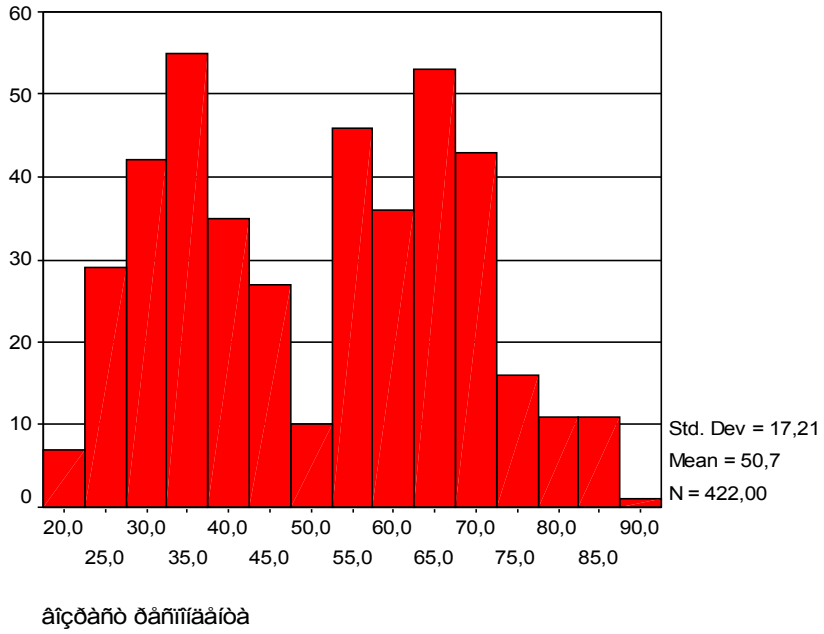



**График 3. Гистограмма возраста респондентов, попавших в выборку в опросе 1999 г.**

На графике 3 видно, что при построении гистограмм выводятся такие статистические показатели, как стандартное отклонение (Std. dev.), среднее (Means) и общее число наблюдений (N). Они даются в обрамлении в правом нижнем углу. Если в нижней части главного диалогового окна установить опцию Display normal curve (показать нормальную кривую), то гистограмма приобретет иной вид. Он показан на графике 4.

В правом нижнем углу главного диалогового окна рассматриваемой процедуры имеется кнопка Titles (заголовки), позволяющая делать описание графика, в том числе и на русском языке. Установки графика могут быть взяты и из какого-нибудь файла, хранящегося в самой системе или, скажем, в текстовом процессоре. С это целью необходимо использовать находящееся в центральной части главного диалогового окна рассматриваемой процедуры обрамление «Template». Активизация кнопки «File», находящейся в этом обрамлении, и доступ

к ней появляются после пометки мышью чек-бокса «использовать спецификацию графика из» (Use chart specification from).

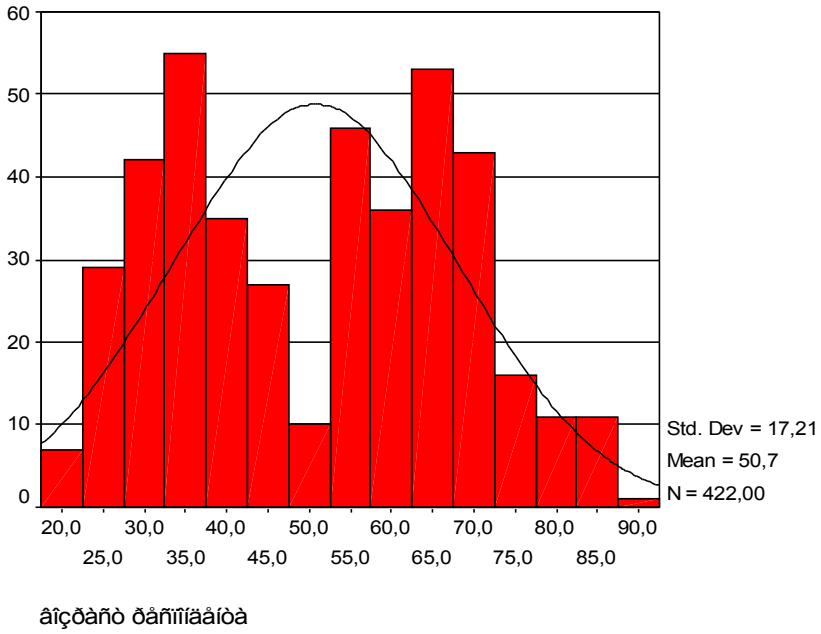



**График 4. Гистограмма возраста респондентов с кривой нормального распределения**

Более обобщенное представление о данных дает **ящичковая диаграмма**. Она показывает не действительные значения, а суммарную статистику для всего распределения. При построении ящичковой диаграммы используется следующая комбинация команд:

**Graphs**

**Boxplot**.

В результате выполнения приведенной выше последовательности команд открывается окно, позволяющее выбрать один из двух типов ящичковой диаграммы: Simple (простая) и Clustered (кластерная). Это окно и показано на рис. 54. По умолчанию в нем установлены опции: «простая ящичковая диаграмма» (она находится в черном жирном контуре) и опция «Summaries for groups of cases» (суммирование одинаковых случаев внутри переменной).

Фиксация этих исходных установок и последующее нажатие кнопки Difine открывает, соответственно, одно из двух главных диалоговых

окон процедуры Boxplot. Такой путь построения характерен для многих графиков (Bar, Line, Area, Pie и др.).

Далее, как и всегда в главных диалоговых окнах, из списка переменных следует выбрать нужную переменную (если был выбран тип независимой переменной) и нажать OK или из списка переменных выбрать две переменные и нажать кнопку OK.

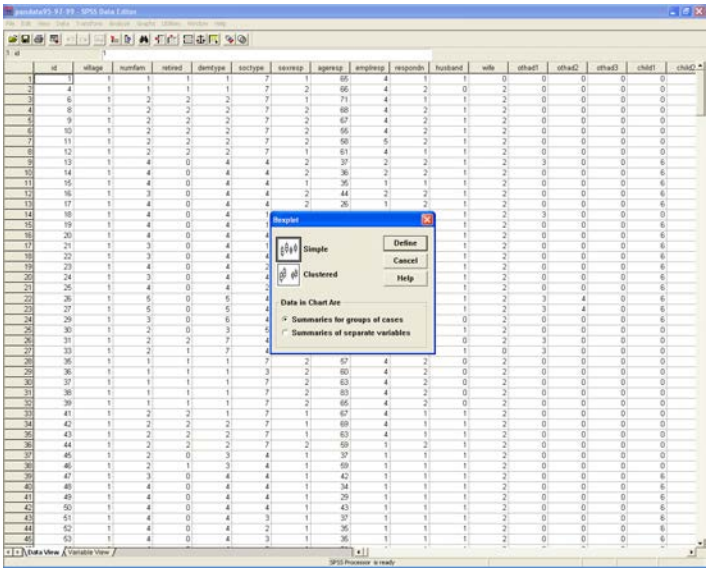



***Рис. 54.***

***Вводное диалоговое окно процедуры Boxplot -ящичковая диаграмма***

В одном случае график будет представлять собой один ящичек (график 5). В другом случае - несколько ящичков для каждого значения второй переменной (график 6).

Ящичковые диаграммы, как гистограммы и диаграммы ствол-лист (Stem & Leaf), описанные в главе 8, § 8.3, дают информацию о распределении наблюдаемых значений. Эти диаграммы имеют медиану – 50-й процентиль (горизонтальная линия внутри ящичка), нижнюю границу - 25-й процентиль, верхнюю границу - 75-й процентиль. Длина ящичка соответствует межквартильному диапазону, который является разницей между 25 и 75 процентилями.

Ящичковая диаграмма включает две категории наблюдений с их значениями. Наблюдения со значениями, превышающими три длины ящичка от верхнего или нижнего его края, называются экстремальными значениями. На ящичковой диаграмме они обозначаются звездочками. Наблюдения со значениями, лежащими в диапазоне от 1,5 до 3 длин ящичка от верхнего или нижнего его края, называются выбросами и обозначаются кружочками.

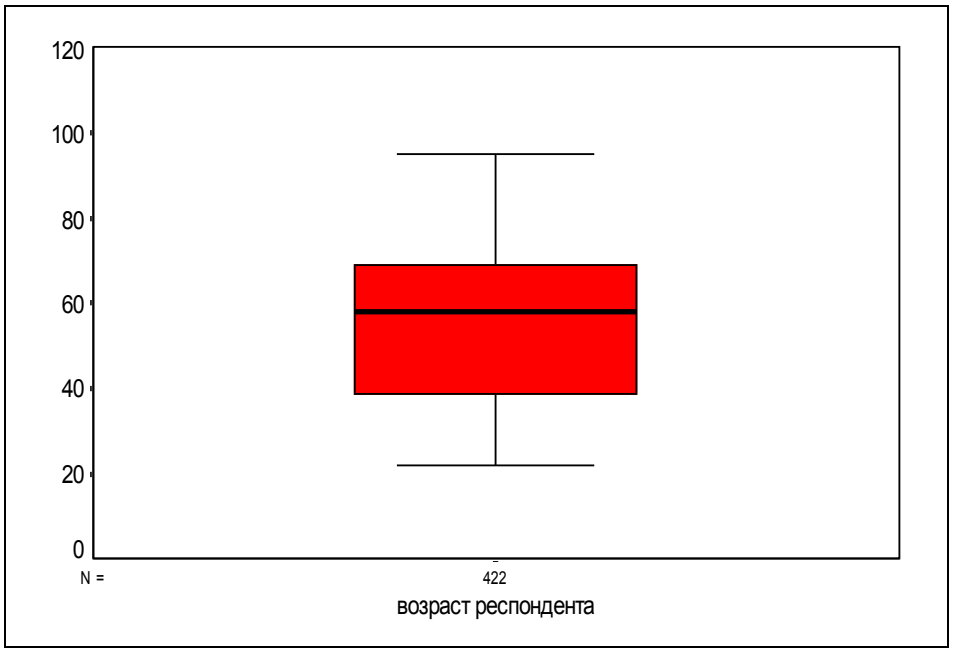



**График 5. Ящичковая диаграмма «Simple» (вторая опция)**

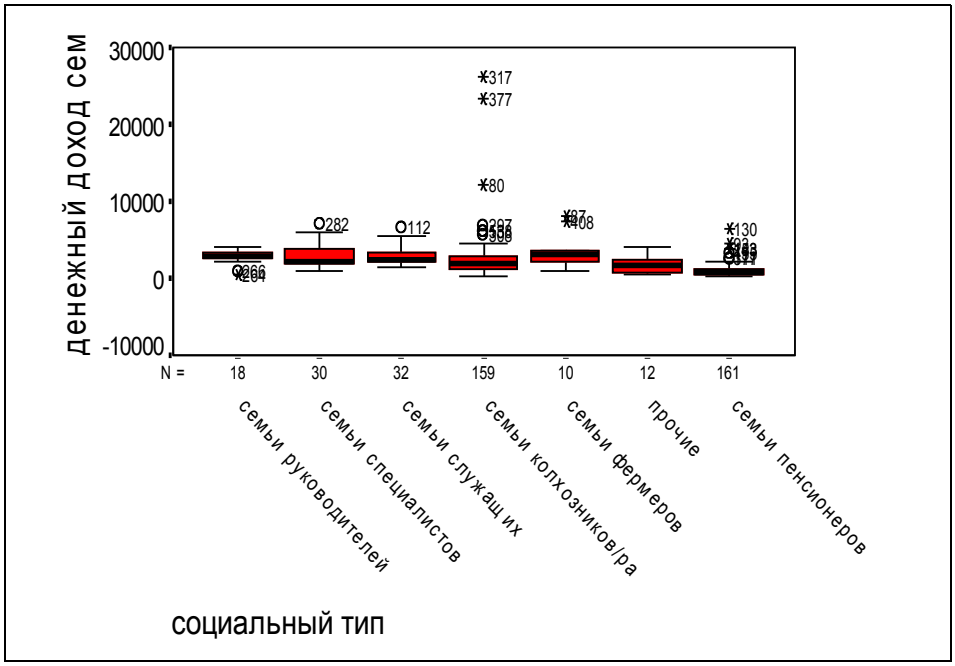



**График 6. Ящичковая диаграмма «Simple» (первая опция)**

Наибольшее и наименьшее из наблюдаемых значений, которые не являются выбросами, также показаны на графике. К этим значениям проведены линии от концов ящичка. Эти линии иногда называются усиками. Ящичковая диаграмма особенно полезна для сравнения распределений значений в нескольких группах.

Некоторые графики могут служить иллюстрацией к статистическим функциям. Например, круговая диаграмма прекрасно проиллюстрирует функцию частоты (Frequencies). В обрамлении 35 представлено частотное распределение массива по признаку «демографический тип семьи».

**Обрамление 35. Распределение демографического типа семьи в сельской местности в выборке 1999 г.**

**демографический тип**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | одиночки | 88 | 20,9 | 20,9 | 20,9 |
| | супружеские пары, пенсионеры | 48 | 11,4 | 11,4 | 32,2 |
| | супружеские пары,работники | 20 | 4,7 | 4,7 | 37,0 |
| | супружеские пары с детьми | 107 | 25,4 | 25,4 | 62,3 |
| | супруж. пары с детьми и родств. | 65 | 15,4 | 15,4 | 77,7 |
| | неполные семьи | 11 | 2,6 | 2,6 | 80,3 |
| | прочие | 83 | 19,7 | 19,7 | 100,0 |
| | Total | 422 | 100,0 | 100,0 | |

Это же распределение показано на графиках 7-8, но уже в виде круговой диаграммы. Оно является результатом выполнения последовательности команд: Graphs – Pie – подокно Pie Chart, в котором по умолчанию стоит опция «Summaries for groups of cases» (суммирование случаев с одинаковыми значениями внутри переменной) – Define – подокно Define Pie: Summaries for groups of cases, в котором надо выделить и установить в поле Define slices by (определить сектора с помощью) искомую переменную, а также обратить внимание на установки в центральном верхнем обрамлении Slices Represent (сектора представлены).

По умолчанию здесь стоит опция - N of cases (число случаев). Это значит, что сектора созданы из расчета суммирования числа случаев с одинаковыми значениями, но здесь можно установить и другие опции, в том числе проценты. Далее можно, используя кнопку «Titles», ввести описание графика, а можно и сразу выполнить его расчет, нажав кнопку OK.

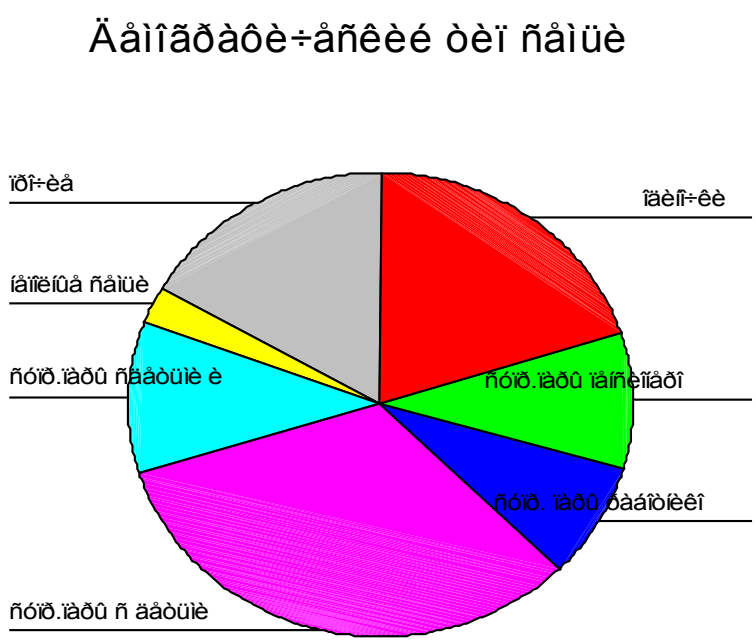



**График 7. Круговая диаграмма (Pie), выполненная в цвете**


**График 8. Круговая диаграмма, выполненная в штрихах**

Графики 7-8 отличаются свойствами (цветовая гамма и черно-белая штриховка). Эти свойства изменяются либо непосредственно через главное меню системы, либо в окне просмотра. Сначала система по умолчанию всегда выдает цветной график.

Пользователь может изменить эти установки посредством выполнения следующей последовательности команд: Edit–Options – закладка Chart – нижнее левое обрамление Fill Patterns and Line Stiles–Cycle through colors, then patterns (выделение посредством цвета) или Cycle through patterns (выделение посредством штриховки). Установка одной из этих двух опций и ведет к выведению графиков в цвете или штриховке.

Свойства могут быть изменены и из окна просмотра. В этом случае пользователь должен выполнить ту же последовательность команд, но уже не через главное меню системы, а через меню окна просмотра и вновь построить график. При таком подходе порядок изменения свойств более правильно было бы описать в следующем параграфе. Полезно помнить, что раз измененные свойства сохраняются не только до конца текущего сеанса, но и для последующей работы, то с изменением свойств не следует торопиться.

Общая последовательность шагов по построению графиков может быть резюмирована следующим образом:

- После открытия раздела главного меню Graphs далее выбирается тип графика: столбиковая диаграмма (bar), круговая диаграмма (pie), гистограмма (histogram), ящичковая диаграмма (boxplot) и т.д.
- На следующем шаге выбирается уточненный вид выбранного графика, который можно строить по независимой переменной или рассматривать в взаимосвязи нескольких переменных.
- И, наконец, открывается главное диалоговое окно соответствующей процедуры для выбора переменной, которая будут участвовать при построении графика. В этом же окне выбираются и другие параметры графика (в каждом типе графика свои). Например, во всех типах графиков можно задать параметр показа абсолютных значений (N of cases) или процентов (% of cases), а с помощью использования кнопки «Title» ввести описание графика.

## 12.3. Окно просмотра и редактирование графиков

После выполнения команды OK графики появляются в специальном окне. В ранних версиях SPSS оно называлось «Chart Carousel», а в последних версиях стало называться. «Output – SPSS Viewer». В этом окне просмотра (глава 7, § 7.1 и 7.5) последовательно накапливаются все выданные в одну сессию графики.

В случае необходимости редактирования графика двойной клик мышью по нему открывает специальное окно редактора «Chart – SPSS Chart Editor» с требуемым графиком. Но в отличие от окна Output, где информация текстовая и результат новой функции добавляется в конец окна, в окне Chart одновременно можно видеть на экране только один график.

Окно редактора графиков имеет главное меню и панель инструментов. С их помощью можно изменять цвет графиков, размер и шрифт надписей, разворачивать графики вдоль осей и производить другие специфические (зависящие от типа графика и интереса пользователя) действия.

Так, в круговой диаграмме можно сдвинуть один или несколько из сегментов круга. Это имеет смысл делать в том случае, если по тексту надо сделать акцент на соответствующем сегменте.

Для выполнения указанной операции следует поместить стрелку курсора на требуемый сектор и выделить его, кликнув мышью. При этом внешний контур сектора, дуга окружности получит ожерелье черных бусинок. Далее клик по четвертой справа кнопке Explode Slice на панели инструментов окна редактора и сдвиг сегмента выполнен. В нашем примере мы выделили брачные пары с детьми до 18 лет в качестве базовых, нуклеарных семей (график 9).

Как видно на графике 9, в нем не только выделен сектор, но и установлены процентные значения секторов. Последняя операция выполняется путем установки курсора в окне редактора на надпись одного из секторов и последующего клика. При этом все надписи обрамляются в маленькие квадратики. Последующий клик мышью в выделенной надписи открывает подокно Pie Options. В этом подокне в обрамлении нами была установлена опция Percent. Далее OK, и результат выполнения команды представлен на графике 9.

В окне редактора могут быть изменены цвета сегментов или штриховок. С этой целью необходимо выделить требуемый сегмент и в панели инструментов нажать третью слева кнопку Color. При этом открывается окно с аналогичным названием и набором различных цветов. Выделение одного из цветов и последующее нажатие кнопки Apply ведет к изменению цвета выделенного сегмента. Таким путем на графике 8, для повышения качества печати, черный фон штриховки изменен на светло-серый фон. Это окно имеет и другие возможности работы с цветами, которые при желании могут быть освоены самостоятельно

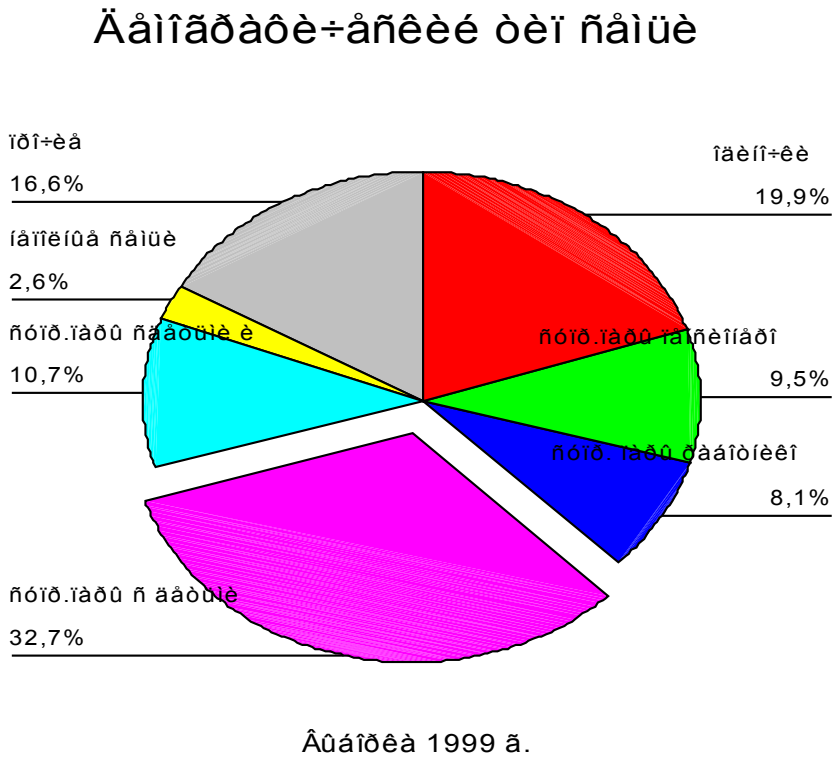



**График 9. Круговая диаграмма с отодвинутым сектором нуклеарных семей и установкой процентных значений секторов**

В панели инструментов окна редактора графиков такие команды-кнопки, как Color, Explode Slice или Fill Pattern (тип штриховки), имеют общее назначение. Но есть и команды, которые специализированы для тех или иных видов графиков, например, Line Stile, Bar Stile (пятая и шестая кнопки слева). Освоение панели инструментов окна редактора пойдет гораздо быстрее и легче, если

обратить внимание на тот факт, что установка курсора (стрелки мыши) на ту или иную кнопку высвечивает ее название в левой нижней части окна редактора.

Окно просмотра графиков дает возможность перейти, во-первых, к графикам с расширением .cht, строившимся в ранних версиях SPSS (3, С. 54-55). С этой целью используется путь: команда меню «вставка» Insert – Old Graph. Далее открывается окно «Data», позволяющее перейти к окну «Output», в котором графики сохраняются системой по умолчанию.

Во-вторых, окно просмотра графиков позволяет перейти от стандартных к интерактивным графикам. Для этого используется путь: Insert – Interactive 2-D Graph или Interactive 3-D Graph. Особенности построения таких графиков рассмотрены в следующем параграфе.

Команда «вставка» имеет еще ряд важных особенностей. С ее помощью можно описывать и редактировать все текстовое содержание графика: заголовок (Insert - New Title), текст (Insert -New Text), сопрягать график с объектом (Insert – Object) и др.

Окно просмотра дает возможность сохранять графики в виде файлов, чтобы в дальнейшем их использовать (особенно это важно, если график подвергся кропотливому редактированию). Для этого, находясь в окне просмотра, необходимо выбрать в меню стандартную последовательность команд: File - Save As.

Далее следует задать имя файла и нажать кнопку OK. В отличие от файлов с данными, для которых используется расширение **.sav** (глава 3, § 3.1), в SPSS для графиков имеется специальное расширение. В ранних версиях таковым было расширение .cht, а в последних версиях используется расширение **.spo.** При дальнейшей работе сохраненный ранее график можно вызвать в окно просмотра с помощью последовательности команд: File – Open – Output. Далее следует выбрать файл с заданным ранее именем.

Копирование графика также можно выполнить непосредственно из окна просмотра (Output – SPSS Viewer). Для этого надо либо использовать команды Edit-Copy меню этого окна, либо поставить стрелку мыши на график и кликнуть правой кнопкой мыши. Эта операция открывает окно, позволяющее копировать график (Copy) и переносить его как объект (Copy objects). Делая вставку копии (Paste) в текстовый процессор, надо быть внимательным и отслеживать положение курсора на экране текстового процессора. Связано это с тем, что именно на месте курсора в открытом текстовом файле или в новом окне появится

переносимый график.

В окне просмотра есть и другие функции. Они позволяют: выводить графики на печать (File – Print), экспортировать их (File - Export) в Web Browser (для создания и сохранения сетевых файлов с расширением **.html**), переносить графики (Edit - Cut), вставлять их (Edit - Paste After), выделять весь график (Edit – Select All) или его фрагменты (Edit – Select – далее выбор фрагмента), а также переходить в окно редактора графиков (Edit - SPSS Chart Object - Open).

Однажды, оказавшись в окне просмотра, в дальнейшем можно непосредственно из этого окна выполнять все последующие расчеты, используя команды меню «Analyze» и «Graphs». В то же время из окна просмотра нельзя делать преобразования данных. Для этого его следует свернуть или закрыть и вернуться в редактор данных.

## 12.4. Интерактивные графики и карты

Интерактивные графики дают пользователю много новых возможностей. Правда, их выбор ограничен наиболее часто используемыми графиками, среди которых гистограммы и диаграммы (круговые, линейные, ошибок, рассеяния, столбиковые и ящичковые). Более полное представление о наборе графиков можно получить, используя последовательность команд главного меню: Graphs – Interactive. Для сопоставления стандартных и интерактивных графиков ниже будет кратко описан порядок построения круговой диаграммы.

Началом в построении интерактивной круговой диаграммы служит следующий путь:

**Graphs**

**Interactive**

**Pie**

**Simple**.

Выполнение указанной последовательности команд приводит к открытию главного диалогового окна искомой процедуры (рис. 55).

Это окно, и по устройству, и по характеру работы с ним, имеет заметные отличия по сравнению с ранее описанными главными диалоговыми окнами других процедур. Здесь обращают на себя внимание закладки (регистрационные карты). В зависимости от особенностей построения графика такие окна содержат четыре или пять закладок. Как видно на рис. 55, в данном случае их четыре.

Первая из них, в порядке движения слева направо, «**Assign Variables»** (перетаскивание переменных) имеется в главных диалоговых окнах всех типов графиков. Вторая карта, как правило, отражает специфику создаваемого графика и носит его имя. В данном случае – это карта «**Pies»** или круговые диаграммы. Две последние карты: «**Titles»** (заголовки) и «**Options»** (свойства).

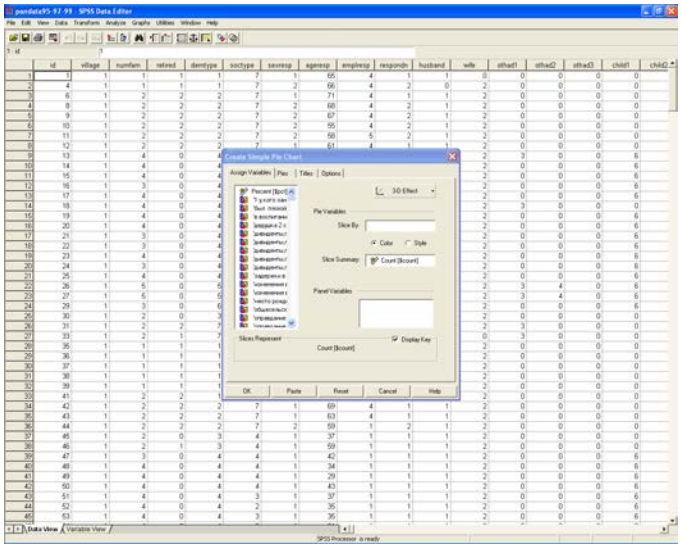


***Рис. 55.***

***Главное диалоговое окно процедуры Interactive Pie***

Карта Assign Variables открывается по умолчанию и содержит: поле со списком переменных, поле для перетаскивания переменных (Slice By), кнопку 3-D Effect, позволяющую делать выбор между трехмерными и двухмерными (2-D Coordinate) графиками, поле Slice Summary с одной из управляющих переменных Count или Percent, которые задают основание для построения графика (он может строится по абсолютным или относительным значениям), и поле Panel Variables (панельные переменные), позволяющее строить диаграммы с разбиением рассматриваемого признака по контрольной (панельной) переменной.

Основные особенности построения интерактивных графиков состоят в следующем. Во-первых, исходный список переменных в них доступен к преобразованию. Установка указателя мыши, который здесь имеет вид руки, на имени любой переменной и последующий клик правой кнопкой мыши, открывает поле, позволяющее сортировать переменные в алфавитном порядке (Sort by Name), в порядке их ввода в редактор данных (Sort by File Order), а также по типу (Sort by Type).

При этом система сама разбивает все переменные на два типа: категориальные и метрические (количественные). Об отнесении переменной к тому или иному типу свидетельствует стоящая в списке переменных перед ее именем пиктограмма. Если переменные рассортировать по типу, то сначала будут идти метрические, а затем категориальные переменные. Вместе с тем условные переменные «счет» и «процент» всегда будут оставаться на первом месте независимо от типа сортировки. Они имеют и свою пиктограмму.

Во-вторых, в интерактивных графиках, в отличие от стандартных, переменные не переносятся, а перетаскиваются из списка в поля Slice By, Slice Summary и Panel Variables. Эта операция требует некоторого навыка и выполняется путем установки указателя мыши на имени переменной, фиксации левой кнопки мыши в нажатом состоянии и последующего перетаскивания переменной в нужное поле. Результат перетаскивания переменной «демографический тип» в поле Slice By с последующим нажатием кнопки OK и приведен на рис. 56.

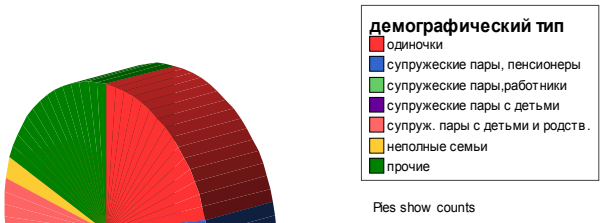



***Рис. 56.***

***Окно просмотра после построения Interactive Pie - круговой диаграммы***

На диаграмме под описанием легенды с указанием цвета и названия секторов содержится надпись: «Pies show count» (круг показывает частоты), свидетельствующая о том, что при установке опции Slice Summary была принята опция, заданная по умолчанию.

Если перед выполнением команды построения диаграммы в карту Pies ввести опции, задающие описание диаграммы, а в карту Titles

вписать заголовок диаграммы, то после выполнения команды OK в окне просмотра появится диаграмма, приведенная на рис. 57.

При этом использовалась последовательность команд:

· Pies - обрамление Slice Labels- помечены боксы - Category, Count, Percent, а в открывшейся после пометки бокса Category строке Location выбрана опция All Outside (все снаружи),

· Titles (поле Chart Title) вписан заголовок диаграммы: Диаграмма 1. Распределение демографического типа семьи.

***Рис. 57.***

***Окно просмотра после построения круговой диаграммы - Interactive Pie с заполнением карт Pies и Titles***


Как видно на рис. 57, приведенная исходно проекция диаграммы не совсем удобна для ее чтения. Этот недостаток может быть исправлен. Для этого следует перейти в окно редактирования диаграммы. Оно открывается двойным кликом мыши по телу диаграммы. При этом она сначала выделяется (получает обрамление), а затем на второй клик открывается окно редактора.

Использование в окне редактора инструментов вертикального и горизонтального вращения графика (два колеса в подокне 3-D), изменяющих его проекцию, а также кнопки Lighting (освещение) позволяют привести диаграмму в более удобный для чтения вид. Новая редакция диаграммы, выполненная в окне редактора графиков, приведена на рис. 58.

Окно редактора позволяет делать массу других преобразований интерактивных графиков и диаграмм. Например, в закладке Titles кроме описания заголовка (верхнее подокно Chart Title) можно сделать

подзаголовок (среднее подокно Chart Subtitle) и комментарий (нижнее подокно Caption).

Закладка Options позволяет сортировать сектора диаграммы по возрастанию и убыванию (обрамление Sort – опции Ascending и Descending), а также задавать различные форматы вывода графиков в цвете и полутонах (контур Chart Look –строки Classic, Education, Grayscale-серые полутона и др.). Эти и другие особенности интерактивных графиков очень ценны в учебных и презентационных целях.

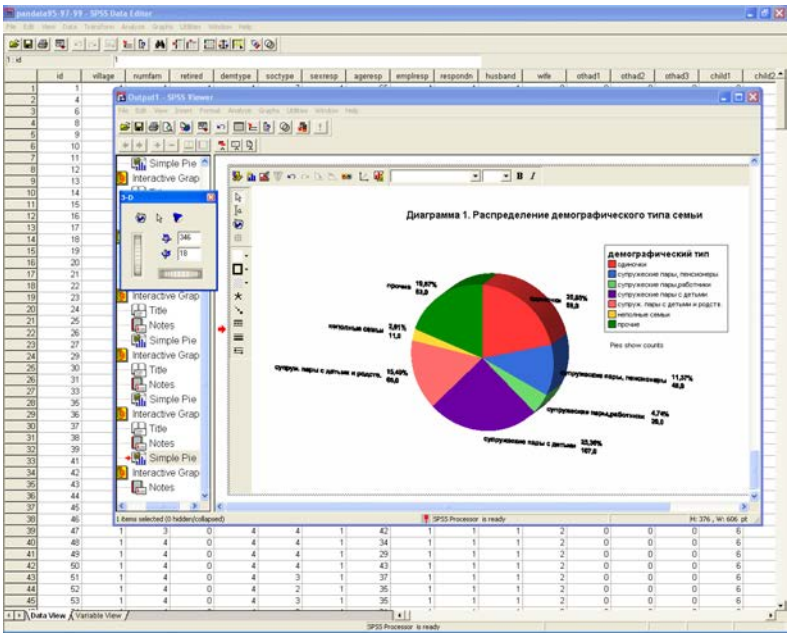



***Рис. 58.***

***Круговая диаграмма в окне редактора Interactive Pie***

Сходные соображения можно высказать и относительно использования картографических сервисных возможностей SPSS. Исходная комбинация команд, позволяющая начинать работу с картами:

**Graphs**

**Map.**

В результате ее выполнения открывается окно с тремя группами дополнительных команд. Первая из них (Range of Values, Graduated Symbol, Dot Density, Individual Values) дает возможность наносить на карту масштабированные символы и точки, а также индивидуальные и ранжированные значения признаков; вторая (Bar Chart, Pie Chart) - строить графики и диаграммы; третья (Multiple Themes) как бы объединяет все ранее названные возможности. Главное диалоговое окно, открывающееся после выбора команды Multiple Themes, представлено на рис. 59.

Это окно имеет 8 закладок. Первая из них - Map. Она и открывается по умолчанию. Далее следуют закладки конкретного типа расчетов (Values, Range, Dot, Bar, Pie, Symbols) и замыкает семейство закладок – Advanced. Для нанесения на карту тех или иных данных требуется сначала задать общие параметры в первой закладке Map, а затем идти в одну из следующих закладок и задать там оставшиеся параметры.

Уже здесь на первом шаге выясняется, что рассматриваемая процедура требует особого типа данных, привязанных к географической местности. Это должны быть или переменные с названием стран, столиц, больших городов, или географические координаты. Разумеется, в массивах социологических данных такая информация пока еще отсутствует.

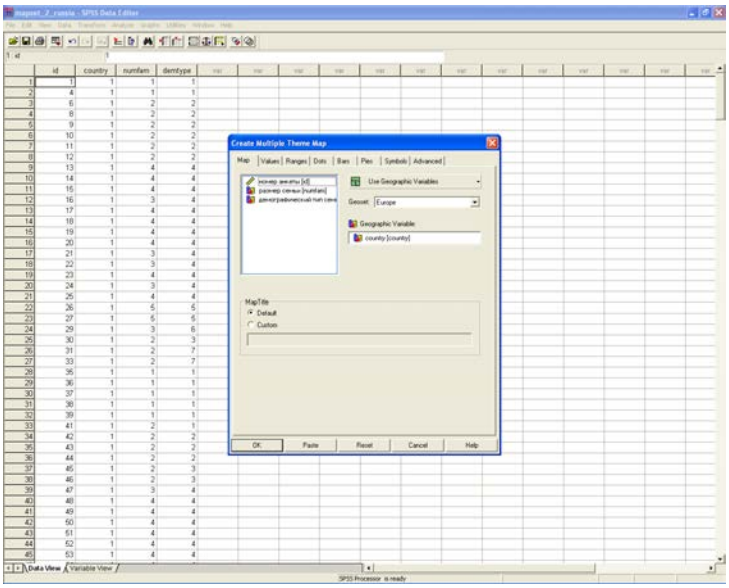



***Рис. 59.***

***Главное диалоговое окно процедуры Create Multiple Theme***

Использование географической переменной предполагает перетаскивание ее в поле Geographic Variable. Далее в поле Geoset необходимо установить опцию географической карты, соответствующей вводимым данным.

Так как в нашем примере в качестве географической переменной взята переменная «страна» со значениями «Беларусь», «Россия» и «Украина», то в поле Geoset установлена карта Европы (Europe).

Другой, более близкой и лучше масштабированной для поставленных целей карты, система не предлагает.

Следующим шагом открывается закладка «круговая диаграмма» (Pie). В этой закладке в поле Slice By перетаскивается переменная «демографический тип семьи». Далее, можно делать различные дополнительные установки, а можно и дать OK на выполнение команды. Результаты этой работы представлены на рис. 60.

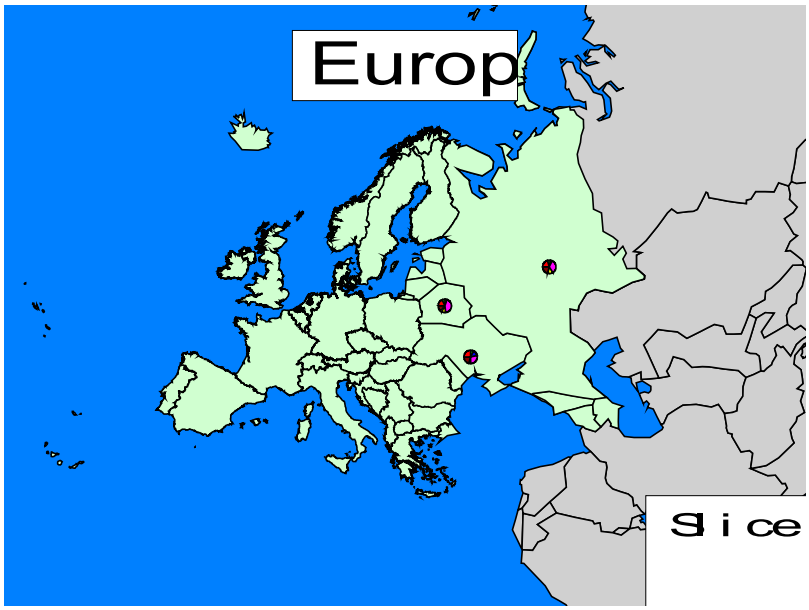



***Рис. 60.***

***Окно просмотра команды Multiple Themes процедуры Map***

Система, действительно, привязала все к требуемой местности. Но для нас полученные эффекты трудно сопоставить с затратами, не говоря уже о том, что приведенный пример условен, так как у нас нет данных, которые отвечали бы всем требованиям географических электронных системам. А в сервисе, предлагаемом SPSS, нет карт, хорошо отражающих пространственное положение СНГ.

**Дружеский совет**

***Многие стандартные графики такие, как ящичковые диаграммы и гистограммы, очень полезно использовать для целей предварительного анализа и изучения массивов данных выборочных исследований.***

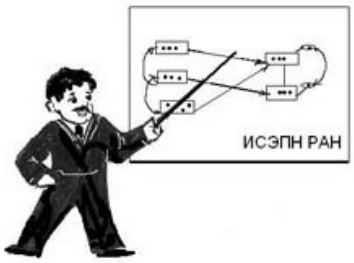
**Дружеский совет**

***Интерактивные графики более удобны для решения учебных задач и представительских целей.***

***Карты – процедура, которую полезно освоить. Возможно, такой шаг будет стимулировать учет пространственно-территориальных особенностей изучаемых объектов при проведении полевых работ.***

## Задание для самостоятельной работы

1. Что такое графики?
2. Назовите основные виды графиков.
3. Как в SPSS из главного меню можно пройти к подменю Graphs?
4. В чем смысл команды Gallery в подменю Graphs?
5. Что такое столбиковая диаграмма?
6. Что такое круговая диаграмма?
7. Что такое линейный график?
8. Что такое график рассеяния?
9. Что такое гистограмма?
10. Что такое ящичковая диаграмма?
11. Перечислите последовательность шагов при построении графиков.
12. Постройте столбиковую диаграмму.
13. Постройте гистограмму.
14. Чем гистограмма отличается от столбиковой диаграммы?
15. Постройте круговую диаграмму.
16. Как соотносятся таблица частот и круговая диаграмма?
17. В чем особенности главного диалогового окна процедур, позволяющих строить различные графики?
18. Чем суммирование случаев в переменной отличается от суммирования отдельных переменных?
19. Как можно задать описание графика?
20. Используя кнопку заголовки, попробуйте выполнить описание графика.
21. В чем особенности окна вывода графиков?
22. Назовите команды панели инструментов окна редактора графиков, которые вы знаете.




# Раздел 3

# МОДЕЛИРОВАНИЕ И SYNTAX

## Глава 13. Регрессионный анализ

### 13.1. Основные понятия

В качестве инструмента моделирования регрессионный анализ позволяет прогнозировать значение одной (зависимой) переменной на основе значения другой (независимой) переменной.

Разумеется, если последовательно исходить из позиции ограниченных возможностей использования числа в социологии (33), то все математическое моделирование исключается из арсенала средств познания социальных явлений. Между тем, в социологии имеется ряд стандартных характеристик, по отношению к которым использование методов математического моделирования может быть весьма эффективным средством познания.

Среди таких характеристик следует обратить внимание на возраст, доходы, размер семьи, число лет обучения. За этими характеристиками стоят настоящие числа, и социологам было бы весьма опрометчиво не использовать эту возможность. Различные оценки, ранги и даже номинальные числа, широко используемые в социологии, имеют сравнительно ограниченные возможности применения в моделировании, но такие возможности есть и их также нужно реализовать.

В математике задача прогнозирования значения зависимой переменной на основе значения независимой переменной решается путем описания корреляционной зависимости между этими двумя призна-

ками в уравнении вида:

$$y = b \cdot x + a$$

Это математическое выражение корреляционной зависимости получило название «уравнение регрессии». Параметры данного уравнения: b - коэффициент регрессии и a – смещение по оси ординат. В отличие от корреляционного анализа, позволяющего установить наличие статистически значимых связей между переменными и оценить степень их тесноты, регрессионный анализ дает возможность описания конкретного вида зависимостей между переменными (38, С. 91).

«Уравнение регрессии тем лучше описывает корреляционную зависимость, чем ближе она к линейной и чем больше ее достоверность. В случае нелинейной зависимости математически запись может выражаться в виде более сложных уравнений различных кривых линий (экспоненциальной кривой, параболы, гиперболы и т.д.). При наличии достоверной криволинейной корреляционной зависимости можно подобрать уравнение, хорошо ее описывающее» (39).

Как правильно отмечают авторы приведенного утверждения, эта возможность становится особенно близкой к реализации при наличии электронно-вычислительной техники. Пакет SPSS и представляет собой одно из средств, открывающих возможность выполнять такого рода расчеты на PC для пользователей, далеких от программирования и специальных знаний математической статистики.

## 13.2. Порядок построения линейной регрессионной модели

Использование регрессионного анализа предполагает предварительное установление статистически значимых связей между переменными и оценки степени их тесноты, т.е. всего того, что связано с корреляцией и рассматривалось ранее в главе 11.

При постановке задачи, связанной с прогнозированием значения зависимой переменной по значению независимой переменной, используется последовательность команд:

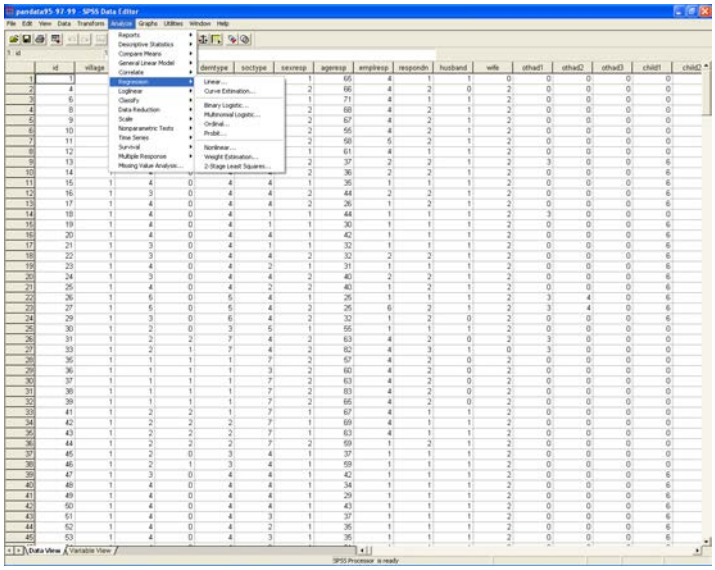
**Analyze**

**Regression.**

Открывшееся в результате выполнения этих команд выпадающее окно (подменю) процедуры **Regression** можно видеть на рис. 61.

*Рис. 61.*

***Выпадающее окно процедуры Regression***



Как видно на этом рисунке, выпадающее окно рассматриваемой процедуры содержит ряд команд, позволяющих рассчитывать различные виды регрессии: линейную (Linear), бинарную логистическую (Binary Logistic), порядковую (Ordinal), пробит (Probit), нелинейную (Nonlinear) и др.

При выборе вида регрессии следует принимать во внимание минимум два обстоятельства: во-первых, тип предполагаемых к использованию в анализе переменных. Во-вторых, естественную природу связи рассматриваемых явлений, позволяющую корректное использование переменных в качестве функции и аргумента. Например, нельзя получать пенсию, не будучи пенсионером или инвалидом, выпивка всегда предшествует опьянению, страховой случай имеет место только при условии предварительного страхования и т.д.

В целях экономии места и времени в предлагаемом тексте приведен пример расчета лишь одного - наиболее простого вида регрессии, а именно линейной (Linear) регрессии. Для выполнения рассматриваемой процедуры необходимо в главном меню выбрать:

**Analyze**

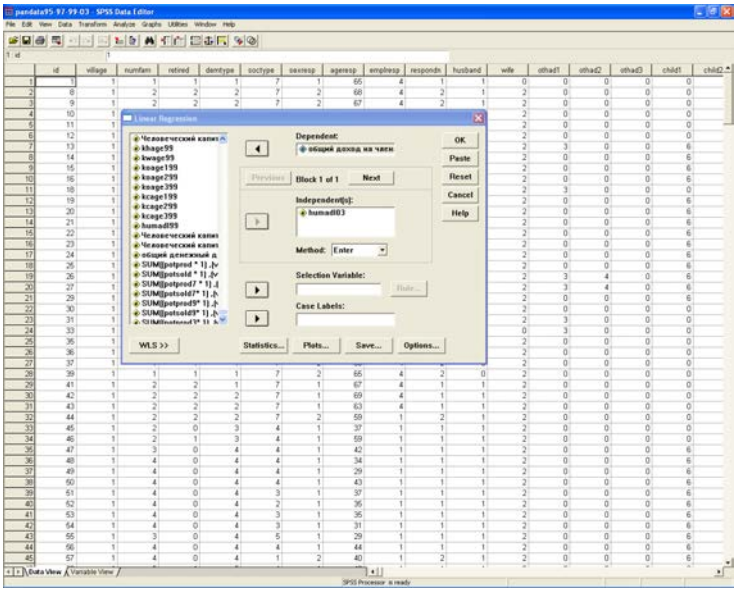
**Regression**

**Linear.**

При выполнении указанной последовательности команд система открывает главное диалоговое окно этой процедуры (рис. 62).

***Рис. 62.***

***Главное диалоговое окно Linear Regression***




Главное диалоговое окно процедуры линейной регрессии имеет семь основных составляющих элементов:

· поле со списком исходных переменных рабочего файла;

· строка для переноса зависимой переменной (Dependent);

· поле для переноса независимых переменных (Independents). Как видно из устройства самого поля, можно переносить несколько переменных;

· строка переноса переменной (Selection Variables) с выключателем (Rule), для одного из значений которой можно построить модель;

· строка (Case Labels) для вывода информации по наиболее отклоняющимся от средней тенденции случаям;

· кнопка WLS, открывающая возможность введения весов (левая нижняя часть окна);

· обрамление (Block), позволяющее вводить независимые переменные блоками. Этот сервис дает возможность одновременного расчета нескольких моделей. По умолчанию здесь всегда выставлен первый блок (Block 1 of 1) для расчета 1-й модели.

Кроме того, в нижней части окна, как и во всех других главных диалоговых окнах, имеются сервисные кнопки-выключатели, позволяющие задавать необходимые статистики (Statistics), строить графики (Plots) и сохранять (Save) в редакторе данных новые расчетные переменные.

Ниже приведен пример расчета линейной регрессии для двух переменных: продаж мяса (зависимая переменная) и объемов его произ-

водства (независимая переменная). Обе переменные взяты из выполненной нами панели обследования сельских домохозяйств в 1995-2003 гг. В этом примере приведены данные 2003 г.

При выборе переменных во внимание принималось два обстоятельства: во-первых, обе они количественные и выражены в одних единицах (кг), во-вторых, продажа мяса предполагает его предварительное производство. Конечно, такого рода последовательность производства и продаж продукции сельского подворья связана с допущением, что селяне знают разницу между продажей продукции производителем и ее перепродажей посредником. Опыт показывает, что они повседневно сталкиваются с этой проблемой.

Предварительный корреляционный анализ двух названных выше переменных показал, что между ними имеется статистически значимая корреляционная связь. Коэффициент Пирсона равняется ,961. Для выполнения расчетов указанные переменные перенесены в соответствующую строку для зависимой переменной (Dependent) и поле для независимых переменных (Independents). Это и зафиксировано на рис. 62.

В дополнительном диалоговом окне Statistics установки, стоящие по умолчанию (Estimates и Model fit), сохранены и новых установок не добавлено. В дополнительном диалоговом окне графиков (Plots) установлены параметры расчета диаграммы рассеяния (Scatter), показывающей связь удаленных остатков и нормированных предсказанных значений: Y=SDRESID, X=ZPRED, а также флажки для вывода гистограммы (Histogram) и диаграммы нормальной вероятности линии регрессии (Normal probability plot).

В контуре графиков рассеяния из самого наличия поля с аббревиатурами различных диаграмм рассеяния видно: во-первых, что можно задать несколько видов диаграмм, во-вторых, при этом пользователь приходит в прямое соприкосновение с командами языка синтаксиса. Например, ZPRED - Standardized predicted (нормированные предсказываемые значения), ZRESID - Standardized residuals (нормированные остатки) RESID - Unstandardized residuals (ненормированные остатки), SDRESID - Standardized deleted residuals (нормированные удаленные остатки) и др.

Здесь трудно что-нибудь советовать, за исключением использования справочника (Help), справочной литературы (21, p. 641) и наблюдательности при использовании команд системы. Исходные сведения о командном языке «Syntax» даны в последней главе учебного посо-

бия. Больше никаких дополнений не требуется. Поэтому можно дать известную команду OK на выполнение.

## 13.3. Окно просмотра и интерпретация и модели

Результаты расчетов приведены в обрамлениях 36-38. В целях удобства описания, таблицы и графики пронумерованы и выделены в отдельные обрамления. При этом сохранен порядок их вывода в окне просмотра. Вся дополнительная информации, полученная при повторных расчетах модели с другими дополнительными установками, имеет по тексту специальные оговорки. Рассмотрим шаг за шагом результаты расчетов, полученные в окне просмотра и приведенные в обрамлении 36.

**Обрамление 36. Окно просмотра регрессионной модели зависимости продаж мяса от объемов его производства в сельском домохозяйстве в выборке 2003 г.**

**Regression**

**Таблица 1.**

**Variables Entered/Removed[b]**

| Model | Variables Entered | Variables Removed | Method |
|---|---|---|---|
| 1 | Мясо получили (кг)[a] | , | Enter |

a. All requested variables entered.

b. Dependent Variable: Мясо продали (кг)

**Таблица 2**.

**Model Summary[b]**

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|---|---|---|---|---|
| 1 | ,961[a] | ,923 | ,923 | 69,11 |

a. Predictors: (Constant), Мясо получили (кг)

b. Dependent Variable: Мясо продали (кг)

## Таблица 3.

**ANOVA[b]**

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| 1 | Regression | 18678020 | 1 | 18678020,42 | 3911,058 | ,000[a] |
| | Residual | 1552101 | 325 | 4775,695 | | |
| | Total | 20230121 | 326 | | | |

a. Predictors: (Constant), Мясо получили (кг)

b. Dependent Variable: Мясо продали (кг)

## Таблица 4.

**Coefficients[a]**

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | | |
| 1 | (Constant) | -64,398 | 6,206 | | -10,377 | ,000 |
| | Мясо получили (кг) | ,776 | ,012 | ,961 | 62,538 | ,000 |

a. Dependent Variable: Мясо продали (кг)

## Таблица 5.

**Casewise Diagnostics[a]**

| Case Number | Std. Residual | Мясо продали (кг) |
|---|---|---|
| 123 | -3,387 | 400 |
| 143 | 3,235 | 1300 |
| 191 | 3,546 | 1500 |
| 227 | -3,560 | 0 |
| 234 | -3,050 | 400 |
| 263 | -4,510 | 400 |
| 344 | -4,683 | 0 |

a. Dependent Variable: Мясо продали (кг)

## Таблица 6.

**Residuals Statistics[a]**

| | Minimum | Maximum | Mean | Std. Deviation | N |
|---|---|---|---|---|---|
| Predicted Value | -56,64 | 1254,94 | 241,39 | 239,36 | 327 |
| Residual | -323,64 | 245,06 | -3,72E-15 | 69,00 | 327 |
| Std. Predicted Value | -1,245 | 4,234 | ,000 | 1,000 | 327 |
| Std. Residual | -4,683 | 3,546 | ,000 | ,998 | 327 |

a. Dependent Variable: Мясо продали (кг)

*Первая таблица* (введенные/перемещенные переменные) говорит о том, что у нас была одна модель (левый столбец Model 1). Введена одна независимая переменная (второй столбец слева).

В таблице говорится о том, что не было каких-либо перемещений переменных (третий столбец слева с системным пропущенным значением – запятой), а при расчете модели использовался метод Enter, который был установлен системой по умолчанию и принят нами. Информация о зависимой переменной дана в примечании к рассматриваемой таблице.

Метод Enter - норма расчета линейной регрессии. Он задает последовательный перебор всех данных, вовлеченных в анализ. Для других видов регрессии, например, множественной линейной регрессии, больше подходит пошаговый метод (Stepwise), который может быть установлен с помощью команды «Method».

*Вторая таблица* (сводная таблица по модели) в столбцах слева направо подтверждает, что расчет сделан по одной модели (первый столбец), квадратный корень из меры определенности - R равен ,961 (второй столбец). В нашем случае и во всех случаях линейной регрессии значение R совпадает со значением коэффициента корреляции Пирсона.

Мера определенности (R-квадрат), как характеристика качества регрессионной прямой, а также степени соответствия между моделью и исходными данными, имеет значение ,923 (третий столбец). Значение меры определенности всегда находится в интервале от 0 до 1, и одновременно оно всегда несколько меньше значения R. Мера определенности показывает, какая часть изменчивости зависимой переменной может быть объяснена независимой переменной. В нашем примере эта величина составляет 92%, что с полным основанием можно рассматривать как хороший результат.

Далее в двух столбцах рассматриваемой таблицы идут дополнительные статистики и их числовые значения: смещенного R-квадрат (третий столбец) и стандартной ошибки оценки (четвертый столбец).

*Третья таблица* дает описание дисперсии. Как и предшествующие таблицы, она содержит в первом (левом) столбце номер модели. Затем идет столбец со значениями суммы квадратов: в уравнении регрессии (Regression = 18678020), остатков (Residual = 1552101) и их итоговой суммы (Total = 20230121). Этот столбец очень информативен, поскольку здесь фактически описано, как вычисляется мера определенности (Regression/Total = R Square). В нашем случае она, как уже говорилось

ранее, равняется ,923.

Следующий столбец содержит значения степеней свободы (глава 10, § 10.3). В нашем примере степень свободы равна 1. Такая степень свободы характерна для линейной регрессии. Для множественной регрессии она определяется числом независимых переменных. Условно говоря, максимально возможное число степеней свободы равно числу случаев в зависимой переменной минус единица. В нашем случае оно составляет 326. Это значит, что продажа мяса наблюдается в 327 хозяйствах из 382, попавших в выборку.

Далее, следует столбец со средним значением квадрата. Вполне естественно, что для регрессии оно такое же, как и в столбце с суммой квадратов, а для остатков значительно меньше. В нашем примере они равны, соответственно, 18678020,42 и 4775,695.

Еще один столбец содержит результаты вычисления контрольной величины F для регрессии. Она проверяет существование ненулевых коэффициентов регрессии. «Значение F велико, когда независимые переменные помогают объяснить вариацию зависимой переменной» (8, С. 183). В нашем случае она равна 3911,058. Характерная особенность F состоит в том, что эта величина всегда должна быть положительной или, в крайнем случае, равняться 0. Отрицательное F – свидетельство отсутствия соответствия между моделью и используемыми в ней данными. Об этом неизбежно будет еще более твердое указание в следующем столбце.

Последний столбец рассматриваемой таблицы содержит очень важную информацию, фиксирующую уровень статистической значимости связи (глава 10, § 10.3). Линейная регрессионная модель зависимости является надежной, если уровень значимости не превышает 0,05 (5%). В нашем случае Sig = ,000, что свидетельствует как в пользу высокого уровня статистической значимости линейной связи двух переменных модели, так и в пользу надежности зависимости продаж мяса от его производства, зафиксированной в модели.

*Четвертая таблица* окна просмотра содержит описание коэффициентов уравнения регрессии. Именно в этой таблице выводятся такие расчетные показатели уравнения регрессии, как коэффициент регрессии – b и константа a, о которых говорилось в самом начале главы при описании уравнения регрессии. Вся эта информация дана в столбце «не стандартизированные коэффициенты» (колонка B). В нашем случае смещение по оси ординат (константа) равно - 64,398, а коэффициент регрессии равен ,776.

Содержательно это означает, что, если в 2003 г. на сельском подворье произведено 1000 кг мяса, то объем его продаж подворьем составит примерно 710 кг. (объем продаж = ,776 · 1000 – 64,398). Отрицательное значение константы свидетельствует о том, что домохозяйства не продают все произведенное мясо. Это вполне естественно, так как кроме продаж есть еще и натуральное потребление.

Еще одна колонка в рассматриваемом столбце выводит стандартные ошибки для константы и коэффициента регрессии. Далее, идет столбец с стандартизированным (нормированным) коэффициентом. В уравнении линейной регрессии значение этого коэффициента то же, что и значение R в табл. 2 (Model Summary). В нашем примере оно равно ,961.

В следующем столбце - t представляет собой контрольную величину, которая получена путем деления значений константы и коэффициента регрессии на их стандартные ошибки. В нашем случае значения t равны, соответственно -10,377 и 62,538.

Наконец, крайний справа столбец таблицы коэффициентов опять, как и в предшествующей таблице, выводит значение уровня статистической значимости (Sig), но теперь он представляет уровень значимости для каждого коэффициента регрессии. При 5%-ном уровне значимости можно считать неравными нулю только те коэффициенты, для которых значение Sig. не превышает 0,05.

Две последние таблицы окна просмотра (диагностика случаев и статистика остатков) имеют важное контрольное значение. Данные *пятой таблицы* (диагностика случаев) позволяют контролировать непосредственно в массиве случаи, в которых остатки продаж мяса отклоняются от среднего значения. Система при этом выводит номера случаев, указывая их в порядке, заданном редактором данных. Использование процедуры Case Summaries (глава 5, § 5.2) позволяет вывести идентификационный номер, значения производства и продаж мяса для каждого случая, равно как и другие важные для контроля данные.

Например, системой в табл.5 выведен 191-й случай. Это подворье произвело 1700 кг мяса, а продало 1500 кг или 88,2% от произведенного мяса. В то же время в среднем по массиву доля продаж мяса, как отмечалось ранее, составляет 71,4% от его производства. Это максимальное отклонение от среднего по продажам мяса, но, как показывает анализ, в нем нет ничего экстраординарного.

Сходные соображения могут быть высказаны и по случаю 344. Он представляет собой пример, который дает минимальный остаток

(здесь произведено 500 кг мяса и нет продаж). Анализ показывает, что в данном случае мать, живя на селе, обеспечивает (кормит) мясом семью своих городских детей. Иными словами, и в том, и в другом крайних случаях отклонения продаж мяса нет ошибки сбора или ввода данных. Указанные отклонения имеют содержательную, а не технико-методическую природу.

*Шестая таблица* (статистика остатков) важна для контроля адекватности самой модели. В ней приводятся минимальные, максимальные и средние значения: предсказанные моделью (Predicted Value), остатков (Residual) и нормированных остатков (Std. Residual).

Для каждого наблюдения остаток – это разница между наблюдаемым значением зависимой переменной в выборке и значением, которое предсказано моделью. «Остатки позволяют оценить ошибки (e) модели, и если модель адекватна, то остатки подчиняются нормальному распределению» (8, С. 185). С целью проверки указанного положения и задавалось построение гистограммы, которая будет рассмотрена ниже.

В последнем столбце табл. 6 выводится число случаев в зависимой переменной. В нашем примере, как уже отмечалось ранее, оно равняется 327. О том, что эта информация относится к зависимой переменной, говорится в примечании к данной таблице.

Если при формулировке задания на расчет модели в дополнительном диалоговом окне установить другие опции, например, описательные статистики (Descriptives), матрица ковариаций (Covariance matrix), то в окне просмотра на первом месте окажется таблица с описательными статистиками. А на втором месте – корреляционная таблица (обрамление 37). Только после этих двух таблиц будут следовать таблицы, расчет которых задан по умолчанию и которые приведены в обрамлении 36.

В этом случае появляется возможность сразу увидеть число наблюдений (N=327) и значение коэффициента корреляции Пирсона, а также проверить формулу расчета регрессии по средним (Mean) производства и продаж мяса (,776 · 394,01 – 64,398 = 241,39). Эта операция позволяет лучше понять первичный смысл регрессионного анализа. Его создатель, Ф.Гальтон, рассматривал математическое описание конкретного вида зависимостей, как «регрессию к среднему состоянию» (38, С. 91).

**Обрамление 37. Дополнительные таблицы окна просмотра**

**Descriptive Statistics**

| | Mean | Std. Deviation | N |
|---|---|---|---|
| Мясо продали (кг) | 241,39 | 249,11 | 327 |
| Мясо получили (кг) | 394,01 | 308,42 | 327 |

**Correlations**

| | | Мясо продали (кг) | Мясо получили (кг) |
|---|---|---|---|
| Pearson Correlation | Мясо продали (кг) | 1,000 | ,961 |
| | Мясо получили (кг) | ,961 | 1,000 |
| Sig. (1-tailed) | Мясо продали (кг) | , | ,000 |
| | Мясо получили (кг) | ,000 | , |
| N | Мясо продали (кг) | 327 | 327 |
| | Мясо получили (кг) | 327 | 327 |

На графике 1 (обрамление 38) приведена гистограмма нормированных остатков зависимой переменной «продажа мяса» и кривой нормального распределения. «Нормированные остатки – это обычные остатки, деленные на стандартное отклонение остатков выборки, в результате чего (если модель адекватна) их среднее равно 0, а стандартное отклонение равно 1» (8, С. 185).

Таким образом, подтверждение адекватности модели можно увидеть как на графике, так и еще раньше в табл. 6 окна просмотра. Уже из данных этой таблицы видно, что среднее (Mean) нормированных остатков (Std. Residual) равно 0, а их стандартное отклонение (Std. Deviation) очень близко к 1. Указанная информация выводится и непосредственно с самим графиком (легенда в его нижнем правом углу).

В отличие от табличных данных график наглядно показывает, что остатки в целом подчиняются закону нормального распределения, но лишь в целом, что и видно на графике по его сдвигу вправо и вверх. Это обстоятельство, правда, не свидетельствует против адекватности модели.

**Обрамление 38. Окно просмотра регрессионной модели зависимости продаж мяса от объемов его производства (графики)**

## Charts

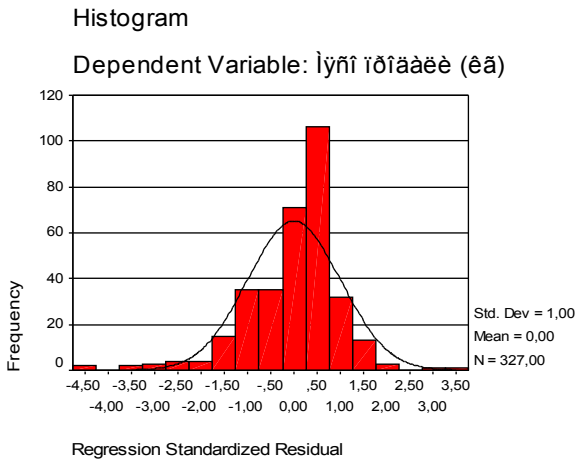



**График 1. Histogram**

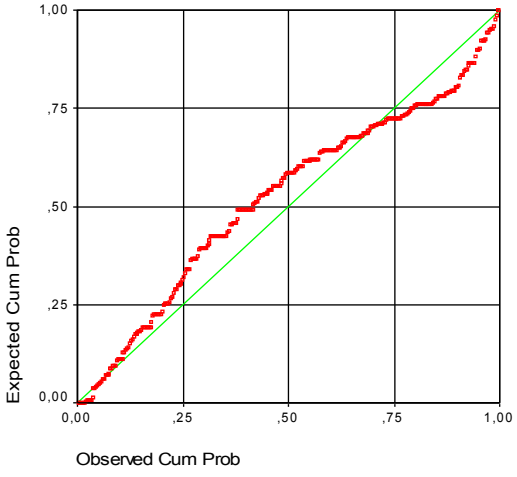



**График 2. Normal P-P Plot of Regression Standardized Residual Dependent Variable: Мясо продали (кг)**

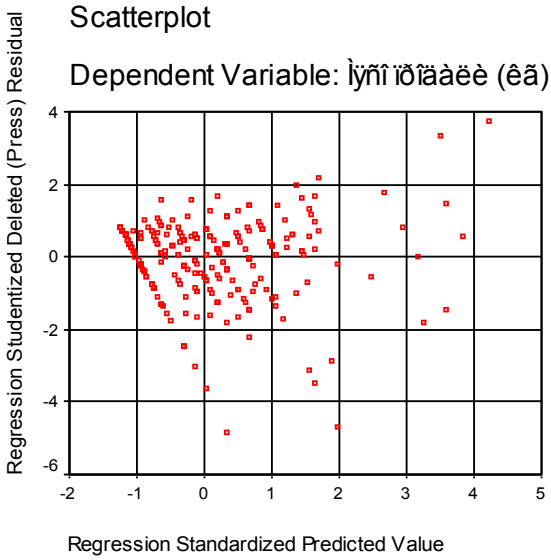



**График 3.**

График 2 (диаграмма нормальной вероятности нормированных остатков зависимой переменной «продажа мяса») показывает, как нормированные остатки зависимой переменной обрамляют прямую наилучшей аппроксимации (линию регрессии). Предел мечтаний здесь - точное попадание остатков на прямую регрессии. Но для этого они все должны быть взяты из нормального распределения. Как видно на графике, в нашем случае это условие выполняется лишь частично. Близость этих двух линий и общность их направленности - свидетельство линейной связи.

График 3 (диаграмма рассеяния) показывает связь удаленных остатков и нормированных предсказанных значений. «Многие исследователи предпочитают рассматривать в диаграммах стьюдентизированные остатки, поскольку, когда модель для вычисления остатков адекватна, их среднее равно 0, а дисперсия равна 1. Если ошибки распределены нормально и модель адекватна, около 95% остатков попадут в отрезок между –2 и +2 (только 1 из тысячи попадет вне отрезка плюс-минус 3)» (8, С. 189). На графике 3 видно, что в нашем случае это условие выполняется не совсем полно.

В заключение полезно обратить внимание еще на один важный момент. Он связан с сохранением и использованием различных вспомогательных значений (расстояния, остатки, интервалы), получаемых в ходе расчетов регрессионной модели. Эти значения можно сохранять в виде новых переменных и использовать в дальнейших расчетах

при контроле данных и в качестве независимых переменных. Для сохранения вспомогательных значений в качестве новых переменных используется команда Save (сохранить) главного диалогового окна рассматриваемой процедуры. Эта команда позволяет сохранять расчетные характеристики в нормированном (Standardized) и ненормированном (Unstandardized) виде.

При этом переменные с нормированными прогнозируемыми значениями получают имена zpr_1, zpr_2, а с ненормированными значениями – pre_1, pre_2, с нормируемыми остатками - zre_1, zre_2, а ненормируемыми остатками – res_1, res_2, с ковариациями - cov_1, cov_2 и т.д. Новые переменные сохраняются непосредственно в редакторе данных. Они устанавливаются в конце списка переменных рабочего файла. Команда Save позволяет сохранять новые переменные и в специальном файле. Для этой цели в ее подокне имеется опция XML – файл. Такой сервис дает возможность использовать сохраненные данные в приложениях SPSS.

## 13.4. Другие опции и методы регрессии

В главном диалоговом окне линейной регрессии возможность ввода нескольких переменных в поле Independents связана с расчетом различных видов множественной регрессии. Большое число переменных не гарантирует хорошей модели, но оно всегда усложняет ее построение и интерпретацию. От пользователя зависит, как будет строиться модель множественной регрессии. Это можно делать путем пошагового ввода переменных с просчетом каждого шага, а также путем ввода всей совокупности переменных и последующего пошагового исключения все тех же переменных с минимальной важностью. В любом случае при расчете множественной регрессии требуется переход на пошаговый метод (Stepwise), который устанавливается с помощью команды-выключателя Method.

При построении модели множественной регрессии минимально важными считаются те переменные, которые вносят наименьший вклад в аппроксимацию. Это видно по значению приведенного $R^2$ (Adjusted R Square). Если указанный коэффициент растет, то переменная улучшает модель. Если он снижается, то она ухудшает модель. Индикатором важности конкретной переменной в сопоставлении с важностью других переменных служит значение коэффициента t в

таблице коэффициентов – Coefficients (обрамление 36, табл. 4). В случае множественной регрессии переменная с минимальным значением статистики t служит первым кандидатом на удаление.

Обрамление Block главного диалогового окна  тех видов регрессии, где оно имеется, позволяет вводить независимые переменные блоками. Благодаря этому, открывается возможность одновременного расчета нескольких моделей. Иными словами, все то, что делалось по шагам, может быть получено сразу за одну итерацию в одном окне просмотра. По умолчанию, как отмечалось ранее, здесь всегда выставлен первый блок (Block 1 of 1) для расчета 1-й модели.

В приведенном нами примере модель строилась для всего массива. Поэтому здесь довольно сложно избежать выбросов, которые и дали о себе знать. Строка переноса переменной (Selection Variables) позволяет строить модель для одного из указанных значений переменной. Это повышает точность расчетов, выполняемых в более однородной выборке. Скажем, в массиве есть переменная с высокой объяснительной способностью - демографический тип семьи. Ее перенос в строку Selection Variables ведет к появлению в ней выражения: demtype 3=? Одновременно открывается доступ к выключателю Rule. Его использование позволяет вставить в окошко открывшейся строки одно из семи значений введенной переменной, а именно: значение 5 – семьи с брачной парой, детьми до 18 лет и другими родственниками (приложение 4). Последующее Continue и в строке Selection Variables выражение приобретает вид demtype3=5. Модель строится только для сложных семей, о чем будет указано в примечании к каждой таблице. Таким путем модели могут быть построены для каждого типа семьи.

Установка идентификационного номера в строку (Case Labels) позволяет сразу получить диагностику случаев, привязанную к каждому наблюдению (обрамление 36, табл. 5). Без этого сервиса, для идентификации вывода информации по наиболее отклоняющимся от средней тенденции случаям, требуется дополнительно использовать процедуру Case Summaries.

Кнопка WLS открывает возможность введения весов. Введение сюда идентификационного номера или того же демографического типа не даст ничего, кроме соответствующего примечания в каждой таблице окна просмотра. А введение в эту клетку таких характеристик, как взвешенное число единиц техники на подворье или человеческий капитал (взвешенная способность подворья к труду), изменяет параметры модели.

Расчет каждого вида регрессии имеет свою специфику, но она не выходит за рамки общих принципов выполнения статистических расчетов в SPSS. Все ограничения этого метода связаны с особыми требованиями к типу (глава 6, § 6.2) используемых переменных. Нелинейная регрессия предъявляет наиболее жесткие требования к числовому значению зависимой переменной, которое в этом случае должно быть интервальным. Линейная регрессия используется в случае, если зависимая переменная имеет интервальное или порядковое числовое значение. Бинарная логистическая регрессия допускает использование в качестве зависимой дихотомическую переменную.

Сходные требования регрессионный анализ предъявляет и к числовому значению независимой переменной. В линейной регрессии она должна быть интервальной, а в множественной линейной регрессии, в дополнение к интервальной независимой переменной, можно использовать в качестве ковариант категориальные переменные. Бинарная логистическая регрессия - самая амбивалентная к типу независимой переменной, а множественная логистическая регрессия позволяет использовать только категориальные независимые переменные.

**Правило 35**

***Критерии оценки качества регрессионной модели:***
***- значение меры определенности (R-квадрат) тяготет к 1;***
***- уровень значимости ( Sig.) не должен превышать 0,05 (5%);***
***- средние нормированных предсказанных значений (Std. Predicted Value) и нормированных остатков (Std. Residual) равняются 0, а их нормированные отклонения равняются 1.***

***Основания для включения-исключения переменных в модель:***
***- если коэффициент приведенного $R^2$ (Adjusted R Square) растет, то переменная улучшает модель, если же он снижается, то переменная ухудшает модель;***
***- индикатором важности переменной служит значение коэффициента t в таблице коэффициентов– Coefficients.***

**Правило 36**

***Уравнение регрессии может быть построено по данным таблицы коэффициентов (Coefficients).***

## Задание для самостоятельной работы

1. Что такое регрессионный анализ?
2. Сформулируйте цели применения регрессионного анализа.
3. Какой вид имеет уравнение регрессии?
4. Какие виды регрессионного анализа предлагает SPSS?
5. Какие требования к переменным предъявляет регрессионный анализ?
6. В чем состоят основные свойства графического представления уравнения прямолинейной регрессии?
7. Что такое мера определенности?
8. Какой уровень значимости связи говорит о надежности модели?
9. Что такое коэффициенты уравнения регрессии?
10. Что специфично для задания установок на расчет линейной регрессии?
11. Какие таблицы окна просмотра регрессионной модели вы знаете?
12. Как проверить надежность регрессионной модели?
13. Как рассчитываются значения величин F и T в модели регрессии?
14. Какие имена получают новые переменные, генерируемые регрессией?
15. О чем говорит легенда в правом нижнем углу гистограммы в окне просмотра регрессионной модели?
16. Какие опции имеются в главном диалоговом окне бинарной логистической регрессии?
17. В чем функция строки Selection Variables главного диалогового окна линейной регрессии?
18. В чем специфика расчета бинарной логистической регрессии?
19. Какие опции имеются в главном диалоговом окне линейной регрессии?
20. В чем специфика расчета нелинейной регрессии?
21. Как связаны регрессионный и корреляционный анализы?
22. Какие опции имеются в главном диалоговом окне нелинейной регрессии?
23. По каким данным окна просмотра при расчете регрессии можно построить уравнение регрессии?
24. О чем говорят правила, сформулированные в этой главе?

# Глава 14. Факторный анализ

## 14.1. Основные понятия

Факторный анализ в еще большей мере, чем регрессионный, требует одновременно как понимания математических принципов решения подобного рода задач, так и высокой исследовательской квалификации разработчиков, позволяющей интерпретировать далеко не очевидные результаты моделирования. В связи с этим обращает на себя внимание его очень ограниченное использование в публикациях результатов социологических исследований.

В любом случае обсуждаемые в данной главе методы моделирования вряд ли можно отнести к переходящим из текста в текст социологическим предупреждениям следующего типа: «Подчас полагают, что для получения эффективного результата достаточно лишь прагматично и крайне неразборчиво использовать известные методики, якобы автоматически приносящие функциональный результат» (40, С. 4). Конечно, предупреждающие знаки весьма полезное средство организации дорожного движения, но и они, как известно, не избавляют от происшествий на дорогах. Это тем более справедливо в отношении процесса познания, который существенно отличается от организации дорожного движения.

Между тем, моделирование социально-экономических процессов с использованием факторного анализа требует таких специальных знаний и навыков, равно как и хорошо согласующихся исходных данных, что представить себе крайне неразборчивое использование этой методики довольно проблематично. Скорее, напротив, в социологии пока еще очень редко можно встретить такого рода эксперименты. Вполне естественно, что указанное положение вещей тормозит познание социальных явлений.

В самом общем виде факторный анализ может быть определен как метод сведения нескольких наблюдаемых переменных в одну новую переменную. Эта новая переменная (фактор) исходно присутствует в массиве как бы в скрытом (латентном) состоянии. Методология

факторного анализа позволяет, во-первых, установить связь фактора с наблюдаемыми переменными, и, во-вторых, ввести в массив фактор в качестве новой переменой, которую в дальнейшем можно использовать как независимую.

Модели с латентными переменными применяются при решении следующих задач:

· понижение размерности признакового пространства в данных типа «объект-признак»,

· классификация объектов на основе сжатого признакового пространства,

· косвенные оценки признаков, не поддающихся непосредственному измерению,

· преобразование исходных переменных к более удобному для интерпретации виду,

· создание структурной теории исследования объектов (39).

Использование в факторном анализе некоторого множества наблюдаемых переменных ведет к их организации в несколько факторов, которые требуют содержательной интерпретации. Это обстоятельство даже с использованием таких программных продуктов, как SPSS, продолжает оставаться довольно сложной задачей.

Можно сказать, что, если для моделирования взято 5 переменных и получено 3 фактора или взято 14 переменных и получено 7 факторов, то такой результат подозрителен в своей основе, равно как и осмыслить его будет довольно сложно. Но если в анализе использовалось, скажем, 12 переменных, а получено 4 или еще лучше 3 фактора, то такое решение может привести к хорошим результатам. С учетом сказанного легче понять, почему в SPSS факторный анализ находится в меню Data Reduction (редукция данных).

## 14.2. Построение факторной модели

В 10-й версии SPSS в Data Reduction выпадающее меню содержит только одну команду - Factor. Из этого совершенно не следует, что система не позволяет рассчитывать различные виды факторного анализа. Напротив, их множество, но они задаются через дополнительные диалоговые окна, а не через выпадающее окно, как это происходит в предшествующей процедуре Regression.

Видимо, в данном случае разработчикам системы было важно обратить внимание на оба момента: на редукцию данных и непосредственно на факторный анализ. Можно предположить, что, как и все предшествующие, это была переходная версия в эволюции системы. Поэтому уже в следующей версии SPSS 11.5 выпадающее окно Data Reduction содержит три возможности расчетов: Factor, Correspondence Analyses и Optimal Scaling.

При постановке задачи, связанной с факторным анализом, используется следующая последовательность команд:

**Analyze**

**Data Reduction**

**Factor**.

Открывшееся в результате выполнения этих команд главное диалоговое окно рассматриваемой процедуры можно видеть на рис. 63.

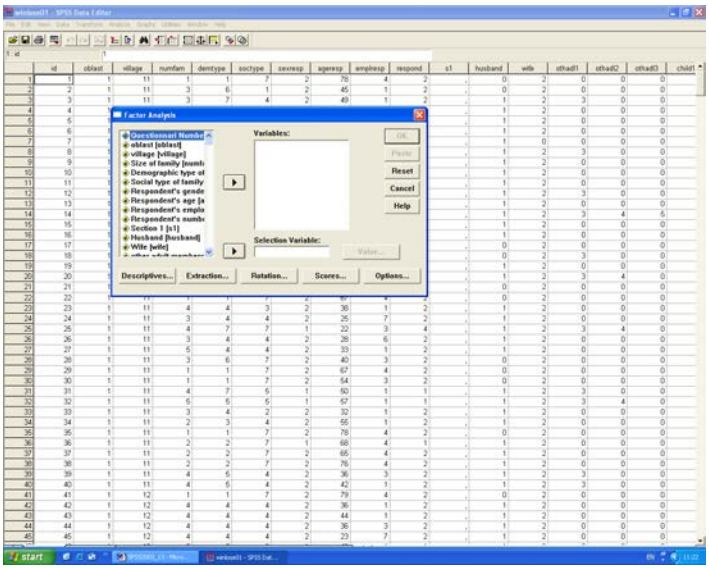



***Рис. 63.***

***Главное диалоговое окно процедуры Factor***

Работа в главном диалоговом окне начинается с традиционного переноса переменных из их списка (левое поле) в центральное поле Variables. Трудно сказать, сколько переменных можно одновременно использовать в анализе.

Следует помнить, что по умолчанию в установках двух дополнительных диалоговых окон стоит максимум в 25 итераций при расчете матрицы компонентов (Extraction –Maximum Iterations for Convergence) и при вращении матрицы компонентов (Rotation - Maximum Iterations for Convergence). А это значит, что при увеличении

числа переменных система неизбежно попросит переустановить число итераций.

Строка Selection Variables позволяет вносить контрольную переменную. При этом после стандартного переноса переменной система потребует установить одно из ее значений, которое и будет использовано в анализе. Установка значения переменной происходит в специальном окне, которое открывается с помощью кнопки Value.

Эта кнопка активируется сразу же после переноса переменной в строку Selection Variables. Например, установка в этой строке переменной respsex (пол респондента) =1 ведет к тому, что модель будет строиться только в разрезе мужчин респондентов. Таким образом, модель получает мощное организующее начало, которое играет решающую роль в ее интерпретации.

В случае, когда строка Selection Variables оставлена пустой, модель строится по всему массиву. В действительности такой подход продуктивен лишь при условии, что анализ выполняется по совокупности однородных переменных, описывающих какое-то цельное явление. Например, в анализе используются переменные, фиксирующие различное отношение к реформам, бизнесу или семейной жизни.

Во всех других случаях отказ от использования строки Selection Variables может быть либо связан с редукцией большого числа переменных в факторы с целью их последующего использования в анализе, либо он еще больше усложняет последующую интерпретацию модели, сводя эффективность проделанной работы к нулю. Видимо, последний случай и имеют в виду авторы цитируемой ранее работы (40), предупреждая о возможности автоматического получения функционального результата в условиях распространения современных информационных технологий.

Нам, однако, представляется, что это как раз тот случай, когда с водой выплескивают ребенка. В экспериментальной науке сделать что-нибудь без ошибок и новых проб практически невозможно. Наивно полагать, что те, кто стоят у истоков профессиональной карьеры, могут вести себя, как и те, кто ее уже завершает. Здесь можно надеяться только на сокращение числа проб и ошибок за счет высокого уровня профессиональной подготовки, а также уровня развития и состояния самой научной дисциплины.

Отличительная черта главного диалогового окна Factor Analysis – пять дополнительных диалоговых окон. Их выключатели находятся в нижней части и имеют слева направо следующие имена: Descriptives

(описательные статистики), Extraction (отбор), Rotation (вращение), Scores (значения), Options (опции). Для выполнения расчетов в разрезе минимума использования необходимых для построения модели опций во всех дополнительных окнах там, где это требуется, по умолчанию уже расставлены соответствующие флажки.

При этом не выполняется вращение матрицы, а факторы, полученные в результате расчетов, не будут сохранены. Для целей вращения матрицы необходимо выполнить последовательность команд: Rotation - обрамление Method - установка Varimax. По умолчанию здесь стоит опция None (нет вращения).

Новые переменные сохраняются в редакторе данных с использованием последовательности команд: Scores - Save as variables (установка флажка) – далее открывается доступ к обрамлению Method – Regression (по умолчанию). При этом новым переменным (факторам) присваиваются имена: fac1_1, fac2_1, fac3_1 и т.д.

Отбор переменных для факторного анализа – один из основных его этапов. В приведенном ниже примере для целей расчета в качестве контрольной переменной взята переменная из массива 2001 г., которая фиксирует оценку респондента (по пятибалльной шкале) к факту выигрыша-проигрыша его семьи в ходе реформ 1991-2001 гг. Социальный портрет выигравших – проигравших по большому числу различных признаков описан нами ранее (27, С.321). При подготовке к анализу шкала оценок (полностью проиграли -1, проиграли - 2, сохранили свои позиции - 3, выиграли - 4, полностью выиграли -5) была преобразована с использованием команд Transform – Recode в номинальную шкалу (проиграли -1, сохранили позиции – 2, выиграли –3).

Объем выборки 800 семей. Первичная и новая шкалы имеют следующие частотные характеристики: шкала оценок (полностью проиграли – 38,9%, проиграли – 22,6%, сохранили свои позиции – 27,3%, выиграли – 7,9%, полностью выиграли –2,4); новая шкала (проиграли –61,5%, сохранили позиции – 27,3%, выиграли –10,3%).

Исходно в строку Selection Variables было введено значение – 3 новой переменной «семья выиграла - проиграла в ходе реформ» (FAMWL3SC=3). Указание имени переменной в данном случае методически уместно в связи с тем, что ниже оно будет выводиться в примечаниях ко всем таблицам окна просмотра (обрамление 39). Еще раз подчеркиваем, что в приведенном ниже примере факторный анализ выполняется только для семей опрошенных, которые считают, что их семьи выиграли от проведения реформ.

В поле Variables перенесено 11 переменных, а именно:

1. Число взрослых членов семьи (чел.).
2. Общая оценка здоровья взрослых членов семьи (в баллах).
3. Общее число лет учебы взрослых членов семьи (лет).
4. Общее число лет взрослых членов семьи (в годах).
5. Денежный доход на одного члена семьи (руб./месяц).
6. Совокупный доход на одного члена семьи (руб./месяц).
7. Взвешенное число единиц техники на подворье (ед. авт. машин).
8. Число детей (чел.).
9. Размер семьи (чел.).
10. Удовлетворенность респондента положением в стране (в баллах).
11. Удовлетворенность респондента жизнью в селе (в баллах).

В приведенном списке переменных они указаны в порядке, который дан в матрице вращения – Rotated Component Matrix (обрамление 39). Следующий этап связан с контролем установок различных опций в дополнительных диалоговых окнах. Эта работа выполнена и описана ниже в порядке размещения выключателей дополнительных диалоговых окон (слева направо).

· Установки в Descriptives (описательные статистики) приняты по умолчанию (флажок стоит только на Initial solution- начальные решения).

· В Extraction (отбор) приняты все установки умолчания, в том числе метод главных компонент (Principal Component), с помощью которого, как утверждают специалисты, решается 90% задач факторного анализа (38, С.180). Дополнительно установлен флажок на Scree plot (точечная диаграмма).

· В установках Rotation (вращение) опция None (нет) заменена на опцию Varimax (варимах).

· В установках Scores (значения) сохранена опция Regression и дополнительно офлажкованы опции Save as variable, а также Display factor score coefficient matrix.

· В Options (опции) сохранена исходная установка Exclude cases listwise и дополнительно установлены флажки в опциях Sorted by size и Suppress absolute values less than. В последней опции после установки флажка открылся доступ в квадрат справа. В нем стоящее по умолчанию минимально выводимое в таблицу значение 0.1 заменено на большее значение - 0.4.

Все эти шаги связаны с логической последовательностью решения задач факторного анализа, среди которых, как отмечается в специальной литературе, первой следует считать проблему робастности (устойчивости), второй – общности, третьей – факторов, четвертой - вращения, пятой – оценки значений факторов, а шестой – проблему динамических моделей (38, С. 178).

С решением пяти первых из указанных задач в SPSS связаны описанные выше установки в дополнительных диалоговых окнах. Справедливости ради следует отметить, что, расставляя установки и опции, мы преследовали еще две дополнительные цели. Они связаны, во-первых, с минимизацией расчетов и, соответственно, данных в окне просмотра, во-вторых, с максимумом простоты и удобства последующей работы с выводимыми в окне просмотра данными.

Теперь можно дать указание на выполнение команды. О том, что все установки и переменные введены корректно, система подтверждает выполнением расчетов и их выведением в окне просмотра. В противном случае в окне просмотра будут указаны причины, по которым система не может выполнить факторный анализ.

## 14.3. Окно просмотра и интерпретация факторной модели

При строгом соответствии расчетов, которые заданы по умолчанию, система выводит в окне просмотра только три таблицы: Communalities (общности), Total Variance Explained (объясненная суммарная дисперсия) и Component Matrix (матрица компонентов). Вместе с тем, поскольку неповернутое факторное решение представляет малозначимую информацию, в нашем примере введены, как отмечалось ранее, дополнительные опции, а это сразу ведет за собой рост числа таблиц и диаграмм в окне просмотра. Результаты расчета модели приведены в обрамлении 39.

В связи с тем, что даже при реализации минимума опций факторного анализа в окне просмотра выдается огромный набор таблиц, в обрамлении 39 оставлены только наиболее значимые для понимания данные. В то же время сама последовательность таблиц и графиков в окне просмотра, полученная на основе описанной ранее последовательности команд, сохранена и приведена ниже.

В окне просмотра непосредственно под названием выполненной процедуры система дала информацию о том, что она не может вывести корреляционную матрицу. И это вполне естественно, поскольку для этого необходимо было в подокне Descriptives установить флажок в контуре Correlation Matrix – Coefficients. Далее, в окне просмотра (обрамление 39) идет несколько таблиц.

**Обрамление 39. Основные составляющие окна просмотра процедуры Factor**

**Factor Analysis**

**Correlation Matrix[a]**

a. This matrix is not positive definite.

**Communalities[a]**

| | Initial | Extraction |
|---|---|---|
| Size of family | 1,000 | ,988 |
| Respondent-satisfied with village life | 1,000 | ,800 |
| Respondent-satisfied with situation in country | 1,000 | ,784 |
| Money total household income per person per month | 1,000 | ,950 |
| Sum total household income per person per month | 1,000 | ,951 |
| ADULTAGE | 1,000 | ,754 |
| NUMADULT | 1,000 | ,979 |
| NUMCHILD | 1,000 | ,987 |
| EDUYEARA | 1,000 | ,866 |
| TOTADHEA | 1,000 | ,895 |
| Взвешенная техника (машина=1, гр.машина= 1,25, трактор=1,25, мотоц.=0,75, др.=0,5 | 1,000 | ,719 |

Extraction Method: Principal Component Analysis.

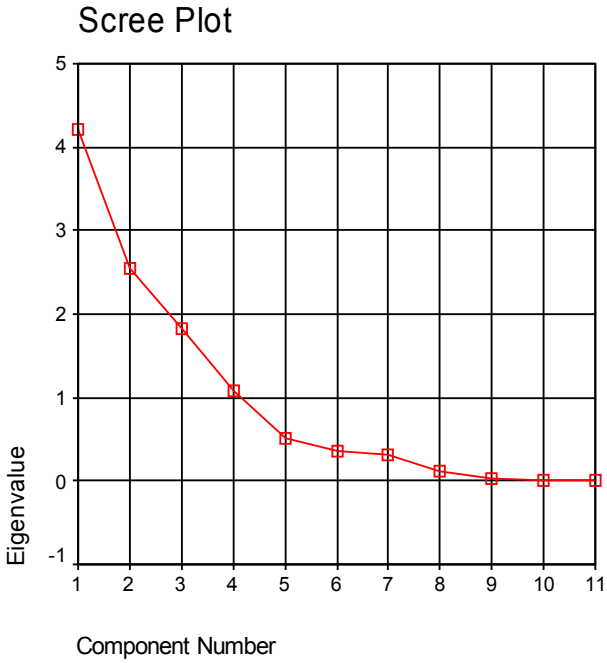
a. Only cases for which FAMWL3SC = 3 are used in the analysis phase.

**Total Variance Explained[a]**

| Component | Initial Eigenvalues | | | Extraction Sums of Squared Loadings | | | Rotation Sums of Squared Loadings | | |
|---|---|---|---|---|---|---|---|---|---|
| | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % |
| 1 | 4,209 | 38,264 | 38,264 | 4,209 | 38,264 | 38,264 | 3,945 | 35,866 | 35,866 |
| 2 | 2,543 | 23,120 | 61,385 | 2,543 | 23,120 | 61,385 | 2,516 | 22,871 | 58,738 |
| 3 | 1,834 | 16,674 | 78,059 | 1,834 | 16,674 | 78,059 | 1,625 | 14,776 | 73,513 |
| 4 | 1,086 | 9,874 | 87,933 | 1,086 | 9,874 | 87,933 | 1,586 | 14,420 | 87,933 |
| 5 | ,512 | 4,652 | 92,585 | | | | | | |
| 6 | ,367 | 3,340 | 95,925 | | | | | | |
| 7 | ,307 | 2,791 | 98,716 | | | | | | |
| 8 | ,108 | ,983 | 99,699 | | | | | | |
| 9 | ,032 | ,292 | 99,991 | | | | | | |
| 10 | ,001 | ,009 | 100,000 | | | | | | |
| 11 | -6,12E-16 | -5,564E-15 | 100,000 | | | | | | |

Extraction Method: Principal Component Analysis.

a. Only cases for which FAMWL3SC = 3 are used in the analysis phase.




**Component Matrix[a,b]**

| | Component | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| NUMADULT | ,970 | | | |
| TOTADHEA | ,914 | | | |
| EDUYEARA | ,910 | | | |
| Size of family | ,866 | | | |
| ADULTAGE | ,783 | | | |
| Money total household income per person per month | | ,943 | | |
| Sum total household income per person per month | | ,941 | | |
| Взвешенная техника (машина=1, гр.машина= 1,25, трактор=1,25, мотоц.=0,75, др.=0,5 | | ,797 | | |
| Respondent-satisfied with situation in country | | | ,753 | ,443 |
| Respondent-satisfied with village life | | | ,687 | ,465 |
| NUMCHILD | | | ,674 | -,661 |

Extraction Method: Principal Component Analysis.

a. 4 components extracted.

b. Only cases for which FAMWL3SC = 3 are used in the analysis phase.

**Rotated Component Matrix[a,b]**

| | Component | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| NUMADULT | ,982 | | | |
| TOTADHEA | ,935 | | | |
| EDUYEARA | ,917 | | | |
| ADULTAGE | ,858 | | | |
| Money total household income per person per month | | ,964 | | |
| Sum total household income per person per month | | ,963 | | |
| Взвешенная техника (машина=1, гр.машина= 1,25, трактор=1,25, мотоц.=0,75, др.=0,5 | | ,769 | | |
| NUMCHILD | | | ,984 | |
| Size of family | ,670 | | ,727 | |
| Respondent-satisfied with village life | | | | ,869 |
| Respondent-satisfied with situation in country | | | | ,867 |

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 5 iterations.

b. Only cases for which FAMWL3SC = 3 are used in the analysis phase.

**Component Transformation Matrix** (преобразованная матрица компонентов)- убрана.

**Component Plot in Rotated Space** (диаграмма компонент в повернутом пространстве) - убрана.

**Component Score Coefficient Matrix** (коэффициенты матрицы нагрузок после вращения) - убрана.

**Component Score Covariance Matrix** (ковариационная матрица нагрузок) – убрана.

*Первая таблица* окна просмотра – Communalities (общности). Она содержит список компонент, а также две колонки: Initial (начальные общности) и Extraction. При использовании метода главных компонент их число равно числу переменных. В списке компоненты приведены в порядке ввода переменных в поле Variables.

В колонке Initial начальные значения общности при использовании метода главных компонент должны быть равны 1. Таковыми они и являются в рассматриваемой таблице.

В колонке Extraction дается расчетная оценка общностей. Общности могут иметь значение в интервале от 0 до 1. При этом близость значения общности к единице говорит о большой объяснительной способности общих факторов по отношению к дисперсии данной переменной. А близость к нулю свидетельствует о том, что общие факторы очень слабо объясняют дисперсию переменной. В нашем примере все значения общностей больше 0,7, что свидетельствует в пользу высокой объяснительной способности общих факторов.

В примечании к этой и всем другим таблицам в окне просмотра дается ссылка на используемый метод («Метод оценки общностей: анализ главных компонент») и на те случаи в массиве, для которых построена модель (в рассматриваемом примере это FAMWL3SC=3 или, как указывалось ранее, семьи, выигравшие в ходе реформ).

*Вторая таблица* - Total Variance Explained (объясненная суммарная дисперсия). Эта таблица имеет важное информационное значение. В ней компоненты упорядочены по их весу в факторной модели. При этом те из компонентов, значение которых больше 1, оказываются вверху списка. Такие компоненты и называются факторами, и только их система SPSS учитывает при построении результирующей модели. Указание об этом содержится в дополнительном диалоговом окне Extraction – обрамление Extract – опция Eigenvalues over (по умолчанию здесь установлена 1).

Из этой таблицы (обрамление 39) видно, что в нашем примере сформировано четыре фактора. В последовательности, указанной в таблице, первый из них объясняет 38,3%, второй – 23,2%, третий – 16,7%, а четвертый – 9,9% дисперсии, а все вместе они объясняют 87,9% дисперсии. Приведенные данные говорят об очень хорошей объяснительной способности модели.

*График* - Scree Plot (точечная диаграмма). Он представляет собой дополнительный инструмент, в котором графически представлены данные, описанные выше в предшествующей таблице. Крутая часть графика пересекает значение 1 по вертикальной оси при значении равном 4 по горизонтальной оси. Все остальные значения меньше 1, но больше или равны 0. Это значит, как уже отмечалось ранее, в нашем примере выделено четыре фактора.

От генерируемого системой числа факторов можно отказаться в

пользу их сокращения. Это делается в дополнительном диалоговом окне Extraction – контур Extract - устанавливается опция Number of factors. При этом открывается доступ в стоящее справа от него окошко, в которое следует ввести цифру с требуемым числом факторов. Одновременно автоматически закрывается доступ к опции Eigenvalues over. В нашем примере число факторов было сокращено с 4-х до 3-х. Основные результаты этих расчетов можно видеть в обрамлении 40.

*Третья таблица* - Component Matrix (матрица компонентов). В этой таблице приведены коэффициенты (или нагрузки), которые соответствуют переменным четырех неповернутых факторов (компонент). Поэтому в литературе можно встретить и другое название этой таблицы – матрица нагрузок (8, С 327).

Как и в предшествующей таблице, компоненты в рассматриваемой таблице упорядочены по весу в факторной модели. Но теперь уже факторы идентифицируются. В нашем случае наибольшую нагрузку имеет число взрослых в семье (NUMADULT) – 0,970, а наименьшую положительную нагрузку имеет удовлетворенность респондента жизнью в селе (0,465). Число детей в семье (NUMCHILD) имеет отрицательное значение нагрузки (-0,661).

В связи с тем, что нами была установлена опция Suppress absolute values less than (не выводить абсолютные значения меньше, чем 0,4), в матрице отсутствуют многие значения нагрузок. Благодаря этому, она стала удобна для просмотра и интерпретации. В противном случае это была бы таблица с полностью заполненными ячейками. Поэтому выделить в ней зрительно переменные с высокими нагрузками было бы довольно трудно.

С учетом того, что нами была установлена опция вращения матрицы, вполне уместно было бы отказаться от вывода рассматриваемой таблицы. Для этой цели в дополнительном диалоговом окне Extraction – контур Display – следует убрать установленное по умолчанию выполнение опции Unrotated factor solution. Мы оставили эту таблицу с целью более наглядной фиксации результатов вращения матрицы.

*Четвертая таблица* - Rotated Component Matrix (повернутая матрица компонентов). Как и предшествующие нагрузки, «нагрузки, преобразованные с помощью ортогонального вращения, - это корреляции переменных с фактором» (8, С. 327). Поэтому, на первый взгляд, четвертая и третья таблицы довольно близки. Вместе с тем вращение матрицы значительно повышает точность модели. Это становится хорошо видно при более внимательном рассмотрении и

сопоставлении двух матриц.

Во-первых, в рассматриваемой матрице значения почти всех нагрузок выросли. Уменьшились нагрузки только переменных «размер семьи» и «взвешенная техника». Во-вторых, такие трансформации привели к заметным изменениям в самой модели.

В простой матрице компонент размер семьи был четвертым фактором. В повернутой матрице им стал возраст взрослых членов семьи (ADULTAGE). Еще более радикально, а главное содержательно, изменились нагрузки числа детей в семье, что привело к образованию ими вместе с размером семьи самостоятельного (третьего) фактора. Благодаря этому, четвертый фактор (удовлетворенность опрошенных) получил большую определенность.

Первый и второй фактор композиционно не претерпели существенных изменений. В этом и проявляется уточняющая роль вращения при очень высокой объяснительной способности модели. Из данных рассматриваемой таблицы видно, что второй фактор связан с доходами и свидетельствует о материальном положении семьи.

В то же время первый – общий и наиболее значимый фактор – число взрослых членов семьи (реальных кормильцев), их возраст, здоровье и образование – связан с человеческим и социальным капиталом семьи. Очень интересно, что в результате вращения снизились нагрузки размера семьи в этом факторе. И это вполне естественно, так как сам по себе размер семьи ничего не говорит о человеческом или социальном капитале. Тем не менее, размер семьи единственная из одиннадцати переменных, имеющая высокую нагрузку по двум факторам (первому и третьему). Совершенно не случайно, как уже отмечалось нами ранее в других работах (27, С. 152-171), что под напором происходящих перемен структура семьи претерпевает заметные изменения. К сожалению, дальнейшая интерпретация модели выходит далеко за рамки предмета рассмотрения в данной главе.

*Пятая таблица* - Component Transformation Matrix (преобразованная матрица компонентов). Это промежуточная таблица, фиксирующая корреляции между факторами после вращения. Она необходима для расчетов, представленных в следующей шестой таблице. В целях экономии места рассматриваемая таблица убрана из обрамления 39.

График - Component Plot in Rotated Space (диаграмма компонент в повернутом пространстве). Этот график выводит пространство координат, число которых равно числу факторов в модели. Такой график легко понять в случае двухфакторной модели, но уже при трех факторах он становиться весьма сложным для восприятия. В нашем случае число факторов равно четырем, а потому с целью упрощения этот график убран из обрамления 39.

*Шестая таблица* - Component Score Coefficient Matrix (коэффициенты матрицы нагрузок после вращения). В этой таблице выводятся корреляции между переменными и факторами. Ее значения являются результатом умножения матрицы нагрузок после вращения (табл.4) на матрицу корреляций между факторами (табл. 5). Эта таблица убрана из обрамления 39.

*Седьмая таблица* - Component Score Covariance Matrix (ковариационная матрица нагрузок). Это квадратная матрица, которая построена по числу факторов (в нашем примере 4x4). В ней значения по диагонали (фактора по фактору) равно 1, а все остальные значения равны 0. Эта таблица, как и две предшествующие, убрана из обрамления 39.

Заканчивая описание окна просмотра факторной модели и последовательности ее построения, уместно обратить внимание на тот факт, что система предлагает и критерии оценки качества модели. С этой целью можно использовать статистику $\chi^2$, которая выдается при выделении факторов обобщенным методом наименьших квадратов (Extraction – обрамление Method – установить опцию Generalized least squares) или методом максимального правдоподобия (Extraction – обрамление Method - установить опцию Maximum likelhood).

Сопоставление данных, приведенных в обрамлениях 39 и 40, показывает, что сжатие модели с четырех до трех факторов практически не коснулось первого и второго факторов. Оно было реализовано за счет объединения третьего (размер семьи и число детей в семье) и четвертого фактора (удовлетворенность респондента жизнью в селе и стране) в один фактор.

За сжатие модели приходится платить немалую цену, которая связана со снижением ее объяснительной способности. В данном случае она снизилась практически на 10%. С учетом высокой объяснительной способности исходной модели (88%) такое снижение не затрагивает ее основ. Понятно, что так бывает далеко не во всех случаях.

**Обрамление 40. Основные данные окна просмотра при сокращении числа факторов модели**

Total Variance Explained[a]

| Component | Initial Eigenvalues | | | Extraction Sums of Squared Loadings | | | Rotation Sums of Squared Loadings | | |
|---|---|---|---|---|---|---|---|---|---|
| | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % |
| 1 | 4,209 | 38,264 | 38,264 | 4,209 | 38,264 | 38,264 | 4,112 | 37,385 | 37,385 |
| 2 | 2,543 | 23,120 | 61,385 | 2,543 | 23,120 | 61,385 | 2,542 | 23,108 | 60,493 |
| 3 | 1,834 | 16,674 | 78,059 | 1,834 | 16,674 | 78,059 | 1,932 | 17,566 | 78,059 |
| 4 | 1,086 | 9,874 | 87,933 | | | | | | |
| 5 | ,512 | 4,652 | 92,585 | | | | | | |
| 6 | ,367 | 3,340 | 95,925 | | | | | | |
| 7 | ,307 | 2,791 | 98,716 | | | | | | |
| 8 | ,108 | ,983 | 99,699 | | | | | | |
| 9 | ,032 | ,292 | 99,991 | | | | | | |
| 10 | ,001 | ,009 | 100,000 | | | | | | |
| 11 | -6,12E-16 | -5,564E-15 | 100,000 | | | | | | |

Extraction Method: Principal Component Analysis.

a. Only cases for which FAMWL3SC = 3 are used in the analysis phase.

**Rotated Component Matrix[a,b]**

| | Component | | |
|---|---|---|---|
| | 1 | 2 | 3 |
| NUMADULT | ,982 | | |
| TOTADHEA | ,944 | | |
| EDUYEARA | ,920 | | |
| ADULTAGE | ,819 | | |
| Size of family | ,781 | | ,494 |
| Money total household income per person per month | | ,962 | |
| Sum total household income per person per month | | ,961 | |
| Взвешенная техника (машина=1, гр.машина= 1,25, трактор=1,25, мотоц.=0,75, др.=0,5 | | ,767 | |
| Respondent-satisfied with situation in country | | | ,765 |
| NUMCHILD | | | ,713 |
| Respondent-satisfied with village life | | | ,706 |

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 4 iterations.

b. Only cases for which FAMWL3SC = 3 are used in the analysis phase.

В обрамлениях 41 и 42 приведены основные данные двух факторных моделей, которые построены для двух других значений переменной «семья опрошенного выиграла или проиграла в ходе реформ». А именно в обрамлении 41 приведены данные модели по проигравшим семьям, а в обрамлении 42 по семьям, сохранившим свои позиции.

Сопоставление данных обрамлений 39 и 41-42 позволяет избежать многих вопросов, которые неизбежно должны возникнуть при построении и интерпретации только одной модели - для выигравших. При построении всех трех моделей использовался один и тот же набор переменных.

**Обрамление 41. Основные данные модели проигравших семей**

Total Variance Explained

| Component | Initial Eigenvalues | | | Extraction Sums of Squared Loadings | | | Rotation Sums of Squared Loadings | | |
|---|---|---|---|---|---|---|---|---|---|
| | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % |
| 1 | 4,216 | 38,323 | 38,323 | 4,216 | 38,323 | 38,323 | 3,649 | 33,173 | 33,173 |
| 2 | 2,346 | 21,324 | 59,647 | 2,346 | 21,324 | 59,647 | 2,196 | 19,962 | 53,135 |
| 3 | 1,261 | 11,468 | 71,115 | 1,261 | 11,468 | 71,115 | 1,823 | 16,577 | 69,712 |
| 4 | 1,067 | 9,702 | 80,817 | 1,067 | 9,702 | 80,817 | 1,222 | 11,105 | 80,817 |
| 5 | ,792 | 7,202 | 88,020 | | | | | | |
| 6 | ,665 | 6,046 | 94,065 | | | | | | |
| 7 | ,457 | 4,157 | 98,223 | | | | | | |
| 8 | ,130 | 1,178 | 99,401 | | | | | | |
| 9 | ,039 | ,356 | 99,757 | | | | | | |
| 10 | ,027 | ,243 | 100,000 | | | | | | |
| 11 | 1,605E-15 | 1,459E-14 | 100,000 | | | | | | |

Extraction Method: Principal Component Analysis.

a. Only cases for which FAMWL3SC = 1 are used in the analysis phase.

**Rotated Component Matrix[a,b]**

| | Component | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| NUMADULT | ,975 | | | |
| TOTADHEA | ,882 | | | |
| ADULTAGE | ,835 | | | |
| EDUYEARA | ,831 | | | |
| Money total household income per person per month | | ,941 | | |
| Sum total household income per person per month | | ,937 | | |
| Взвешенная техника (машина=1, гр.машина= 1,25, трактор=1,25, мотоц.=0,75, др.=0,5 | | ,477 | | |
| NUMCHILD | | | ,891 | |
| Size of family | ,658 | | ,689 | |
| Respondent-satisfied with situation in country | | | | ,794 |
| Respondent-satisfied with village life | | | | ,746 |

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 5 iterations.

b. Only cases for which FAMWL3SC = 1 are used in the analysis phase.

Для всех трех групп построенные модели показывают высокую объяснительную способность (соответственно, для выигравших - 87,9%, проигравших - 80,8% и сохранивших свои позиции - 83,4%) и очень близкую, но далеко не тождественную конфигурацию факторов. Вполне резонно предположить, что за этими мелкими различиями кроются особенности латентных факторов, которые специфичны для каждой из рассматриваемых групп и которые ускользнули при моделировании по общему набору компонент.

**Обрамление 42. Основные данные модели устоявших семей**

Total Variance Explained[a]

| Component | Initial Eigenvalues | | | Extraction Sums of Squared Loadings | | | Rotation Sums of Squared Loadings | | |
|---|---|---|---|---|---|---|---|---|---|
| | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % |
| 1 | 4,550 | 41,363 | 41,363 | 4,550 | 41,363 | 41,363 | 3,846 | 34,965 | 34,965 |
| 2 | 2,242 | 20,385 | 61,748 | 2,242 | 20,385 | 61,748 | 2,281 | 20,734 | 55,699 |
| 3 | 1,329 | 12,084 | 73,832 | 1,329 | 12,084 | 73,832 | 1,708 | 15,523 | 71,223 |
| 4 | 1,053 | 9,568 | 83,400 | 1,053 | 9,568 | 83,400 | 1,340 | 12,178 | 83,400 |
| 5 | ,668 | 6,074 | 89,474 | | | | | | |
| 6 | ,584 | 5,305 | 94,779 | | | | | | |
| 7 | ,373 | 3,389 | 98,169 | | | | | | |
| 8 | ,130 | 1,181 | 99,350 | | | | | | |
| 9 | ,038 | ,343 | 99,693 | | | | | | |
| 10 | ,034 | ,307 | 100,000 | | | | | | |
| 11 | 3,406E-16 | 3,096E-15 | 100,000 | | | | | | |

Extraction Method: Principal Component Analysis.

a. Only cases for which FAMWL3SC = 2 are used in the analysis phase.

**Продолжение Обрамления 42.**

**Rotated Component Matrix[a,b]**

| | Component | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| NUMADULT | ,970 | | | |
| ADULTAGE | ,890 | | | |
| TOTADHEA | ,885 | | | |
| EDUYEARA | ,821 | | | |
| Size of family | ,721 | | ,580 | |
| Money total household income per person per month | | ,966 | | |
| Sum total household income per person per month | | ,953 | | |
| NUMCHILD | | -,441 | ,807 | |
| Взвешенная техника (машина=1, гр.машина= 1,25, трактор=1,25, мотоц.=0,75, др.=0,5 | | | ,653 | |
| Respondent-satisfied with situation in country | | | | ,828 |
| Respondent-satisfied with village life | | | | ,783 |

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 5 iterations.

b. Only cases for which FAMWL3SC = 2 are used in the analysis phase.

Обращает на себя внимание сама конфигурация факторов. Все они связаны с взрослыми членами семьи (их число, оценка здоровья, возраст и образование). Общий фактор «число взрослых членов семьи» имеет большой вес (соответственно, 0,982, 0,975 и 0,970). То, что в числе факторов не оказалось характеристик материального положения, свидетельствует в пользу огромного значения для семьи концентрации в ней «добытчиков». Все остальное как бы является результатом их активности.

Действительно, анализ средних показывает, что композиция структурных характеристик семьи имеет весьма заметные различия для выигравших и проигравших (табл. 33).

**Таблица 33. Средние значения числа взрослых членов семьи, их возраста, числа детей и размера всей семьи для выигравших-проигравших в ходе реформ**

**Report**

| FAMWL3SC | | NUMADULT | ADULTAGE | NUMCHILD | Size of family |
|---|---|---|---|---|---|
| 1,00 | Mean | 2,22 | 109,1159 | ,71 | 2,92 |
| | N | 492 | 492 | 492 | 492 |
| | Std. Deviation | ,922 | 43,72350 | 1,005 | 1,522 |
| 2,00 | Mean | 2,43 | 118,4404 | ,76 | 3,19 |
| | N | 218 | 218 | 218 | 218 |
| | Std. Deviation | 1,028 | 51,89549 | ,854 | 1,511 |
| 3,00 | Mean | 2,71 | 116,6341 | 1,11 | 3,82 |
| | N | 82 | 82 | 82 | 82 |
| | Std. Deviation | 1,000 | 44,10132 | ,956 | 1,458 |
| Total | Mean | 2,33 | 112,4609 | ,76 | 3,09 |
| | N | 792 | 792 | 792 | 792 |
| | Std. Deviation | ,972 | 46,29571 | ,967 | 1,536 |

Из данных табл. 33 видно, что среднее число взрослых среди выигравших семей составляет 2,7, а среди проигравших – 2,2 чел. И видимо, именно это обстоятельство, сочетая в себе элементы высокой способности к труду и взаимного страхования, имеет решающее значение. Иными словами, возраст, дети, размер семьи, равно как и доходы, только дополняют его.

В пользу указанного вывода говорят и три других важных обстоятельства, которые наблюдаются в последние годы. Во-первых, наметившаяся тенденция роста в сельской местности доли сложных (многопоколенных) семей. По нашим данным, в 1991-2003 гг. их доля выросла с 10,6% до 16,8%. Эта тенденция может реализоваться главным образом за счет возврата в родительскую семью взрослых детей или за счет вхождения в нуклеарную семью родителей. И в том и другом случае в семье идет увеличение числа ее взрослых членов. А это, как видно из выполненного анализа, повышает шансы семьи сохранить или даже улучшить свои позиции в трудных условиях перехода к рыночным отношениям.

Во-вторых, идет сокращение удельного веса нуклеарных семей и брачных пар. По нашим данным, в тот же период доля брачных пар сократилась с 30,7% до 17%. В-третьих, наблюдения показывают, что в отношении происходящих перемен в самых неблагоприятных условиях оказываются нуклеарные и неполные семьи. В этих семьях

концентрация бедности идет на фоне ограниченных возможностей их взрослых членов.

Уместно отметить, что такие интересные результаты моделирования были получены далеко не простым путем. Они представляют собой плод огромного объема экспериментальных работ и расчетов, выполненных авторами.

Расчеты начинались с введения 34 переменных, которые при моделировании объединились в 12 факторов, суммарно объяснявших 72% дисперсии. Сокращение числа компонент до 19 вело к образованию 7 факторов и понижало объяснительную способность модели до 49%.

А завершились расчеты, как уже описано выше, построением трех моделей, включающих 11 компонент, 4-е фактора и имеющих объяснительную способность в интервале от 80 до 88%. Как и во многих других случаях, здесь уместно известное выражение: «дорогу осилит идущий». Заметим, что идти по этой дороге, хотя и сложно, но интереснее и продуктивнее, чем расставлять на ней предупреждающие знаки.

**Правило 37**

***Критерии оценки качества факторной модели:***
***- высокие значения отобранных общностей;***
***- высокая объяснительная способность модели;***
***- значения коэффициента $\chi^2$.***

**Правило 38**

***Основания включения-исключения компонент факторной модели:***
***- низкие значения компонент в объясненной суммарной дисперсии;***
***- близость компонент к 0 на точечной диаграмме - Scree Plot;***
***- повышение объяснительной способности модели при исключении компоненты.***

## Задание для самостоятельной работы

1. Что такое факторный анализ?
2. Для решения каких задач используется факторный анализ?
3. В чем различие переменной, компоненты и фактора?
4. Где находится факторный анализ в SPSS?
5. Опишите устройство главного диалогового окна Factor Analysis процедуры Factor.
6. В чем назначение строки Selection Variables в факторном анализе?
7. Перечислите дополнительные диалоговые окна процедуры факторного анализа в SPSS?
8. Какие опции по умолчанию стоят в дополнительных диалоговых окнах процедуры факторного анализа в SPSS?
9. Укажите последовательность команд, которую надо выполнить для вращения матрицы.
10. Что такое общности (Communalities) в факторном анализе?
11. В чем смысл информации, содержащейся в таблице Total Variance Explained (объясненная суммарная дисперсия) окна просмотра?
12. Какова цель вывода в окне просмотра точечной диаграммы - Scree Plot при построении факторной модели?
13. Как можно сократить число факторов?
14. Что происходит с объяснительной способностью модели при сокращении числа факторов?
15. В чем смысл вращения матрицы компонент?
16. Как матрица компонент отличается от повернутой матрицы?
17. Какие таблицы по умолчанию выводятся в окне просмотра факторной модели?
18. Что такое вес фактора в модели?
19. Чем факторы отличаются от компонент?
20. Что объясняют факторы?
21. Какие значения могут принимать общности в факторном анализе?
22. В чем смысл информации, содержащейся в повернутой матрице компонентов - Rotated Component Matrix?
23. Какие правила сформулированы в этой главе?

# Глава 15. Методы многомерной классификации

## 15.1. Кластерный и дискриминантный анализ

Методы многомерной классификации связаны с разбиением случаев или переменных на группы с целью преобразования множества данных в компактную форму. В SPSS для расчетов предлагается два таких метода: кластерный и дискриминантный анализ.

При постановке задачи, связанной с многомерной классификацией, используется путь: Analyze - Classify. Открывшееся в результате выполнения этих команд выпадающее меню Classify (классифицировать) представлено на рис. 64.

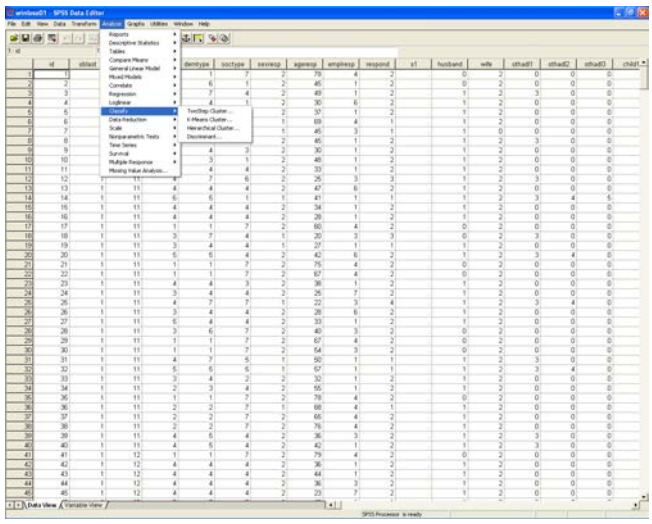


***Рис. 64.***

***Выпадающее меню Classify***

Как видно на этом рисунке, выпадающее меню Classify открывает доступ к процедурам, позволяющим рассчитывать различные виды многомерной классификации. Среди таких процедур версия SPSS 11.5 предлагает: TowStep Cluster (двухшаговый кластер), K-Means Cluster

(кластеризацию методом k-средних), Hierarchical Cluster (иерархический кластерный анализ) и Discriminant (дискриминантный анализ). Заметим, что еще в предшествующей версии системы двухшаговый кластер отсутствовал.

При выборе процедуры многомерной классификации требуется принять решения, которые связаны с типом переменной, с выбором метода нормировки, способом компоновки кластеров и др. Например, метод k-средних лучше подходит для задач с большим числом случаев (200 и более). В то же время иерархическая кластеризация громоздка при работе с большими выборками. Это связано с тем, что она вычисляет матрицу расстояний с элементами для каждой пары наблюдений (8, С. 290).

Дискриминантный анализ требователен к данным. Используемые в нем в качестве предикторов (независимых переменных) данные должны быть количественными и нормально распределенными. При наличии таких независимых переменных для целей повышения качества классифицирующей функции можно использовать и номинальные переменные, но только как дополнительные, а не основные переменные (8, С. 237).

Кластерный анализ можно применять к интервальным данным, частотам, бинарным переменным. Тем не менее, он предполагает, что все переменные, вовлеченные в анализ, независимы. «Продолжающиеся переменные» должны быть количественными и нормально распределенными, а категориальные переменные должны иметь несколько значений (multinomial distribution).

Как для кластерного, так и дискриминантного анализа важно, чтобы используемые в нем переменные измерялись в сравнимых числах (8, С. 237 и 290). Для основной массы практикующих социологов– это фактически еще не актуализированная постановка задачи. Поэтому в социологии, в связи с отсутствием традиции предварительной (при разработке методики и инструментария) сопоставимости шкал, использование методов многомерной классификации часто, если не всегда, предполагает предварительное нормирование (z-преобразование) исходных данных (глава 8, § 8.2).

Сами разработчики SPSS, например, при построении модели двухшагового кластера рекомендуют использовать: парную корреляцию (Bivariate Correlations) - для тестирования независимости двух продолжающихся переменных; таблицы сопряженности (Crosstabs) - для тестирования независимости двух категориальных переменных; ана-

лиз средних (Means) - для тестирования независимости между продолжающимися и категориальными переменными; исследовательские статистики (Explore) - для тестирования нормальности продолжающихся переменных и тест $\chi^2$ (Chi-Square Test) - для проверки множественного распределения категориальной переменной (более подробно по данному вопросу см.: Help- Topics-Base System-TwoStep Cluster Analysis -TwoStep Cluster Analysis Data Considerations). Все это предполагает не только наличие специальных знаний, умений и навыков, но и соответствующих данных, используемых для построения модели.

## 15.2. Построение и описание кластерной модели

Многомерная классификация позволяет получить естественные группировки (кластеры) близких по своей природе случаев, которые не видны с первого взгляда. Используемые при этом алгоритмы имеют существенные отличия от принципов традиционной классификации, которые формулируются, исходя из видимых свойств объектов.

Например, в своей работе мы широко используем демографический тип семьи. Эта многомерная типология строится на основании первичных данных, причем не все из них формализуются в расчетах. Иными словами, демографический тип – результат знаний и многолетнего опыта.

Разумеется, что такого рода знания существуют далеко не для всех возможных конфигураций в распределении различных наблюдений. Вряд ли кто из практикующих социологов или демографов сможет дать ответ о распределении характеристик образования в семейных парах по занятости или месту работы супругов. Ответ на этот вопрос не может быть получен из анализа частот или таблиц сопряженности.

В то же время кластерный анализ вполне может помочь в решении указанной или сходной задачи. С его помощью можно выделить «сгущения точек» и разбить их на однородные совокупности объектов. Не случайно один из синонимов термина «кластерный анализ» - «таксономия».

Основу кластерного анализа составляют меры близости и расстояний между объектами.

Рассмотрим порядок построения модели многомерной классификации на примере использования двухшагового кластерного анализа.

Для решения указанной задачи в главном меню выполняются команды:

**Analyze**

**Classify**

**TowStep Cluster.**

При выполнении указанной последовательности команд система открывает главное диалоговое окно этой процедуры (рис. 65).

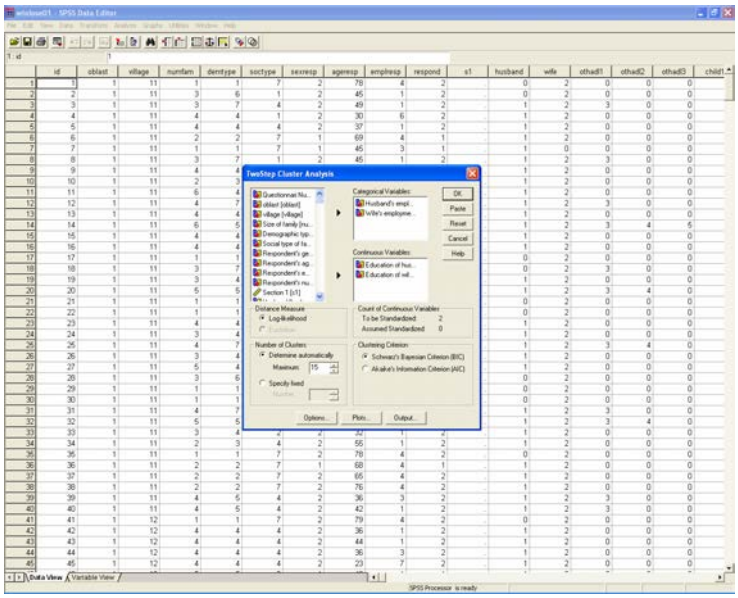


***Рис. 65.***

***Главное диалоговое окно процедуры TowStep Cluster***

Открывшееся главное диалоговое окно рассматриваемой процедуры имеет свои особенности. Оно содержит несколько контуров и всего три дополнительных диалоговых окна с выключателями: Options, Plots и Output.

Кроме того, в нем имеются и стандартные элементы: поле со списком переменных, поле для переноса категориальных переменных (Categorical Variables) и поле для переноса «продолжающихся переменных» (Continuous Variables). Как уже отмечено ранее в главе 6 (§ 6.2), эта новация по типам переменных еще отсутствует в глоссарии самого SPSS (Help-Topics-Glossary).

Конструкция главного диалогового окна рассматриваемой процедуры такова, что на контуры приходится основная часть установок, вводимых системой по умолчанию. Причем, каждый контур имеет две опции.

Контур Distance Measure (измерение расстояния) имеет опции Log-likelihood (лог-правдоподобия) и Euclidian (евклидово расстояние). Системой по умолчанию установлена опция «лог-правдоподобия».

Контур Count of Continuous Variables (расчет продолжающихся переменных) имеет опции To be Standardized (подлежит нормированию) и Assumed Standardized. Системой по умолчанию установлена опция «подлежит нормированию».

Контур Number of Clusters имеет опции Determine automatically (обусловлено автоматически) и Specify fixed Number (задать число кластеров), установив которую необходимо и задать число кластеров. Системой по умолчанию установлена опция «обусловлено автоматически». Заданное в ней максимальное число кластеров (Maximum) равно 15.

Последний контур Clustering Criterion (критерии кластеризации) также имеет две опции BIC и AIC, из которых первая задана по умолчанию.

Дополнительное диалоговое окно Options (опции) сопряжено с контуром Count of Continuous Variables посредством своего собственного контура Standardization of Continuous Variables (нормирование продолжающихся переменных). Названия двух его полей Assumed Standardized и To be Standardized соответствуют именам опций в контуре Count of Continuous Variables. В этом окне нет установок, заданных по умолчанию.

Дополнительное диалоговое окно Plots (графики) содержит ряд опций, позволяющих выводить круговую диаграмму кластера (Cluster pie chart), доверительный интервал (Confidence level) и другую информацию. Здесь нет установок, заданных по умолчанию.

В дополнительном диалоговом окне Output (окно вывода) в контуре Statistics по умолчанию установлены две опции: Descriptives by cluster (описание по кластерам) и Cluster frequencies (частоты кластеров). Два других контура Working Data File (рабочий файл данных) и XML File (XML файл) позволяют, соответственно, сохранить в рабочем файле кластер в качестве переменной с именем tsc_8589 (комбинация цифр здесь случайная и взята из нашего примера), а также экспортировать файл с кластерной моделью.

Исходя из перечисленных ранее требований, для построения кластерной модели взяты две пары переменных. Они фиксируют: распределение занятости и места работы одной пары взрослых членов семьи и число лет учебы другой пары взрослых членов семьи.

В качестве категориальных взяты две перекодированные (Transform – Recode) переменные: «место работы мужа» (hbusnew) и «занятость четвертого взрослого члена семьи» (oempl2n). В выборке из 800 семей (2001г.) мужья есть в 612, из них работают 386, а четвертый взрослый член семьи есть в 127 случаях. Характеристики типа предприятия, на котором работают мужья, имеют следующие частотные распределения: ТОО/АО – 22,0%, СПК – 26,9%, колхоз - 7,8%; общественное обслуживание –13,7%, фермерство и др. агробизнес – 9,6%, другой бизнес – 12,7%, самозанятость – 7,3%.

Занятость четвертых взрослых членов семьи имеет следующие частотные распределения: заняты полный рабочий день – 42,5%, безработные – 11,0%; пенсионеры –37,8%, инвалиды – 3,1%, по домохозяйству и уходу за ребенком – 5,6%. Результаты проверки множественного распределения категориальных переменных с использованием непараметрического теста $\chi^2$ (Chi-Square) представлены в обрамлении 43.

Из обрамления 43 видно, что для той и другой переменной при высокой значимости критерия $\chi^2$ (p=,000) наблюдаются значимые отклонения от ожидаемых частот. Это обстоятельство опровергает нулевую гипотезу о равномерном распределении частот тестируемых переменных.

В качестве продолжающихся взяты переменные «число лет обучения жены» (weduc) и «число лет обучения первого взрослого члена семьи» (adult1). В выборке они представлены, соответственно, в 769 и 325 случаях. Первый взрослый – это фактически третий член семьи (после мужа и жены). Предварительное тестирование показало, что эта пара переменных соответствует требованиям, предъявляемым кластерным анализом к продолжающимся переменным.

Во-первых, обе они числовые. Во-вторых, у них сопоставимые шкалы. Минимум обучения у жены и первого взрослого члена семьи 0 лет. Максимум – у них двоих одинаков –18 лет. Среднее число лет обучения у жены – 9,96 лет, а другого члена семьи– 8,88 лет.

В-третьих, тестируемые переменные независимы. Предварительное тестирование показало, что переменные «число лет обучения мужа» и «число лет обучения жены» тесно связаны (коэффициент корреляции Пирсона равен ,655 при значимости связи ,000). В то же время в выбранной паре переменные независимы между собой (коэффициент корреляции Пирсона равен ,163 при значимости связи ,004).

**Обрамление 43. Проверка множественного распределения категориальных переменных**

**NPar Tests**
**Chi-Square Test**
**Frequencies**

**OEMPL2N**

| | Observed N | Expected N | Residual |
|---|---|---|---|
| 1 | 54 | 25,4 | 28,6 |
| 3 | 14 | 25,4 | -11,4 |
| 4 | 48 | 25,4 | 22,6 |
| 5 | 4 | 25,4 | -21,4 |
| 6 | 7 | 25,4 | -18,4 |
| Total | 127 | | |

**HBUSNEW**

| | Observed N | Expected N | Residual |
|---|---|---|---|
| 1 | 85 | 55,1 | 29,9 |
| 2 | 104 | 55,1 | 48,9 |
| 3 | 30 | 55,1 | -25,1 |
| 4 | 53 | 55,1 | -2,1 |
| 5 | 37 | 55,1 | -18,1 |
| 6 | 49 | 55,1 | -6,1 |
| 7 | 28 | 55,1 | -27,1 |
| Total | 386 | | |

**Test Statistics**

| | OEMPL2N | HBUSNEW |
|---|---|---|
| Chi-Square[a,b] | 88,787 | 91,016 |
| df | 4 | 6 |
| Asymp. Sig. | ,000 | ,000 |

a. 0 cells (,0%) have expected frequencies less than 5. The minimum expected cell frequency is 25,4.

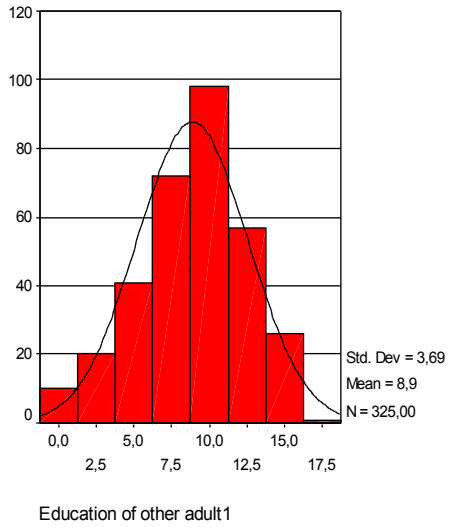
b. 0 cells (,0%) have expected frequencies less than 5. The minimum expected cell frequency is 55,1.

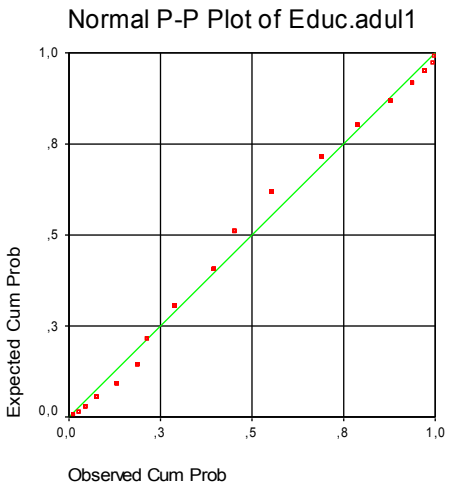
В-четвертых, они соответствуют закону нормального распределения. Для проверки последнего обстоятельства построены графики и диаграммы нормального распределения для переменных «число лет обучения жены» и «число лет обучения третьего члена семьи». Результаты тестирования приведены в обрамлении 44.

**Обрамление 44. Гистограммы числа лет учебы мужа и жены с кривой нормального распределения и их диаграммы нормальной вероятности**

Обучение 1-го взрослого члена семьи

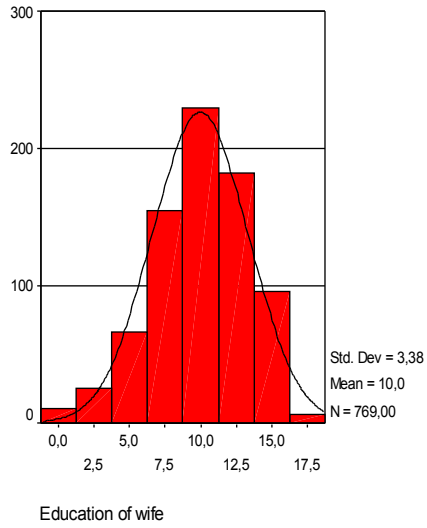
**Graph**

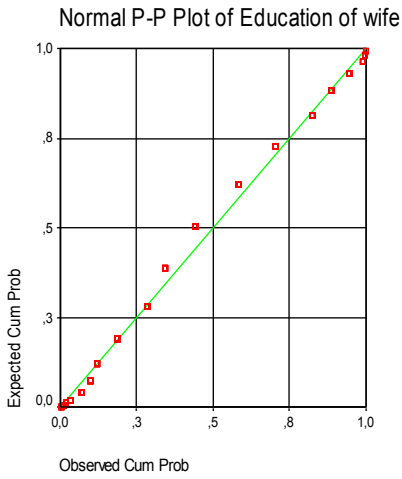



Обучение жены

**Graph**



Как показывают данные обрамления 44, используемые в рассматриваемом примере продолжающиеся переменные, вполне соответствуют требованиям, предъявляемым условиями кластерного анализа. Они – числовые, независимые и подчиняются закону нормального распределения. Можно сделать следующий шаг, а именно: ввести выбранные для построения кластерной модели переменные в соответствующие поля, оставить все установки, которые система SPSS расставила по умолчанию, и дать команду OK на выполнение расчетов.

Результаты выполняемых расчетов представлены в обрамлении 45. Из данных этого обрамления видно, что при расчете двухшаговой кластерной модели с установками, которые заданы по умолчанию, в окно просмотра выводится четыре таблицы. Таким образом, подготовительная работа к построению модели, связанная с поиском и тестированием исходных переменных, заняла значительно больше сил, времени и места, чем построение самой модели. Отсюда, правда, вряд ли справедливо делать далеко идущие выводы. Рассмотрим таблицы окна просмотра полученной кластерной модели.

**Обрамление 45. Окно просмотра данных двухшаговой кластерной модели**

## TwoStep Cluster

**Cluster Distribution**

| | | N | % of Combined | % of Total |
|---|---|---|---|---|
| Cluster | 1 | 49 | 53,8% | 6,1% |
| | 2 | 42 | 46,2% | 5,3% |
| | Combined | 91 | 100,0% | 11,4% |
| Excluded Cases | | 709 | | 88,6% |
| Total | | 800 | | 100,0% |

## Cluster Profiles

**Centroids**

| | | Education of wife | | Education of other adult1 | |
|---|---|---|---|---|---|
| | | Mean | Std. Deviation | Mean | Std. Deviation |
| Cluster | 1 | 10,82 | 2,028 | 8,57 | 3,360 |
| | 2 | 12,48 | 1,954 | 9,95 | 3,527 |
| | Combined | 11,58 | 2,150 | 9,21 | 3,488 |

## Frequencies

**HBUSNEW**

| | | 1 | | 2 | | 3 | | 4 | | 5 | | 6 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Frequency | Percent | Frequency | Percent | Frequency | Percent | Frequency | Percent | Frequency | Percent | Frequency | Percent |
| Cluster | 1 | 18 | 100,0% | 28 | 100,0% | 0 | ,0% | 0 | ,0% | 0 | ,0% | 3 | 21,4% |
| | 2 | 0 | ,0% | 0 | ,0% | 9 | 100,0% | 14 | 100,0% | 8 | 100,0% | 11 | 78,6% |
| | Combined | 18 | 100,0% | 28 | 100,0% | 9 | 100,0% | 14 | 100,0% | 8 | 100,0% | 14 | 100,0% |

**OEMPL2N**

| | | 1 | | 3 | | 4 | | 5 | | 6 | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Frequency | Percent | Frequency | Percent | Frequency | Percent | Frequency | Percent | Frequency | Percent |
| Cluster | 1 | 17 | 42,5% | 6 | 66,7% | 18 | 52,9% | 2 | 100,0% | 6 | 100,0% |
| | 2 | 23 | 57,5% | 3 | 33,3% | 16 | 47,1% | 0 | ,0% | 0 | ,0% |
| | Combined | 40 | 100,0% | 9 | 100,0% | 34 | 100,0% | 2 | 100,0% | 6 | 100,0% |

*Первая таблица* – распределение объектов по кластерам (Cluster Distribution). Она показывает, что в результате расчетов сформировалось два кластера. В их формировании приняли участие только 91 из 800 случаев в выборке или 11,4% выборочной совокупности. Основная масса случаев оказалась исключенной из модели, о чем свидетельствует информация, выведенная в рассматриваемой таблице (Excluded Cases– 709 или 88,6%). Случаи, вошедшие в два кластера, сгруппированы следующим образом: 1-й кластер включает 49 или 53,8%, а 2-й – 42 или 46,2% случаев, составивших модель.

*Вторая таблица* – центроиды (Centroids). Она имеет общий подзаголовок – профили кластера (Cluster Profiles). Эта таблица содержит информацию об основаниях выделения кластеров, подлежащую интерпретации. В данном случае она прозрачна. Кластеры выделены по отклонению значений продолжающихся переменных - число лет обучения жены (Education of wife) и первого взрослого члена семьи (Education of other adult1) от их средних значений в совокупности рассматриваемых случаев.

Средние значения числа лет обучения для жены и первого взрослого члена семьи равны, соответственно, 11,58 и 9,21 лет. Один кластер составили значения ниже среднего, соответственно, 10,82 и 8,57 лет. Другой кластер составили значения выше среднего, соответственно, 12,48 и 9,95 лет обучения.

*Третья и четвертая таблицы* имеют общий подзаголовок – частоты (Freqencies). Сюда включены таблицы: «место работы мужа» (hbusnew) и «занятость четвертого взрослого члена семьи» (oempl2n), показывающие частотное распределения значений указанных двух переменных по двум выделенным кластерам.

Таблица «место работы мужа» показывает, как строго распределились мужья в выделенных кластерах по месту работы. В 1-й кластер попали все те, кто работает в ТОО/АО или СПК, а во 2-й кластер - в основном те, кто работает в общественном обслуживании или частном бизнесе. Последняя частотная таблица, характеризующая занятость второго взрослого члена семьи, менее выразительна, но и она говорит о социальных преимуществах, сгруппированных во втором кластере. В нем больший удельный вес занятых полный рабочий день, меньший удельный вес безработных, пенсионеров, а также совсем отсутствуют инвалиды и занятые дома по хозяйству.

Кластеризация, при всей трудоемкости и жесткости требований к переменным, несет большой эвристический заряд. Традиция социоло

гического мышления такова, что мы работаем в основном с взаимосвязанными и взаимозависимыми переменными. Благодаря этому, при построении таблиц сопряженности, как правило, строятся не все возможные таблицы, а только уже «обкатанные» причинно-следственные связи, по которым и идут анализ и объяснение.

Такой подход к анализу данных имеет как достоинства, так и недостатки. Достоинства связаны с сокращением трудоемкости и включенностью в общий контекст поиска (всегда можно сослаться на сходные выводы или найти объект для критики), а недостатки - с ограниченным объемом первичных данных, вовлекаемых в научный поиск. И как следствие, постоянное обсуждение довольно узкого круга социально-экономических проблем.

Кластеризация как раз и показывает, что группировки по независимым переменным таят в себе много нового и интересного. Являясь описательной, а не объяснительной процедурой, кластерный анализ исключает какие-либо статистические выводы о причинно-следственных связях в самих кластерах и между ними. Вместе с тем он позволяет лучше понять структуру и структурные отношения в наблюдаемой совокупности объектов.

## 15.3. Построение и интерпретация дискриминантной модели

Дискриминантный анализ, как и кластерный, позволяет классифицировать наблюдения путем их разбиения на группы. Его особенность состоит в объяснительной способности предсказывать разбиение по признакам группировок на основе значений независимых переменных. Поэтому в отличие от кластерного анализа, который при определенных условиях проявляет сходство с факторным анализом, дискриминантный анализ несет в себе элементы регрессионного анализа.

Дискриминантный анализ предполагает использование одной группирующей (grouping variable) и нескольких независимых переменных (predictor variables). Группирующая переменная может иметь два и больше значений. В любом случае система требует задание максимального и минимального значений. Случаи со значениями, не попавшими в заданный интервал, исключаются из анализа, о чем система добросовестно проинформирует в окне просмотра. При этом довольно

широко и часто используется группирующая переменная со значениями 0 и 1 (нет-есть), которая позволяет получать хорошо «дискриминируемую» классификацию.

Независимые переменные, как уже отмечалось ранее, должны быть количественными и нормально распределенными, и только в качестве дополнительных к ним можно использовать переменные с номинальным числом. Последние при этом должны быть перекодированы в фиктивные (dummy) или контрастные (contrast) переменные. Все перечисленные требования направлены на создание условий, исключающих ситуацию попадания в модели одного случая в две группы. Об этих и других особенностях дискриминантного анализа можно узнать из справочника системы, используя путь Help-Topic-Base System-Discriminant Analysis, или посредством кнопки Help непосредственно в главном диалоговом окне процедуры Discriminant.

Рассмотрим порядок построения модели в дискриминантном анализе. Для решения указанной задачи необходимо в главном меню выбрать:

**Analyze**
**Classify**
**Discriminant**.

При выполнении указанной последовательности команд система открывает главное диалоговое окно процедуры дискриминантного анализа (рис. 66). Это окно имеет свои особенности. В отличие от предшествующей процедуры оно не имеет контуров. Зато в нем есть пять дополнительных диалоговых команд. Кроме того, в нем имеются и стандартные элементы: поле со списком переменных, поле для переноса группирующей переменной (Grouping Variable), а также поле для переноса независимых (Independents) переменных, которые в справочнике системы описываются, как предсказывающие переменные (predictor variables).

В центральной части окна (над кнопкой Method) имеются две альтернативные опции. Одна из них – Enter independents together (одновременный учет всех независимых переменных). Она установлена по умолчанию. Другая – Use stepwise method (использовать пошаговый метод) при необходимости может быть установлена самим пользователем. Одновременно при этом открывается доступ к находящейся под указанными опциями кнопке «метод» (Method).

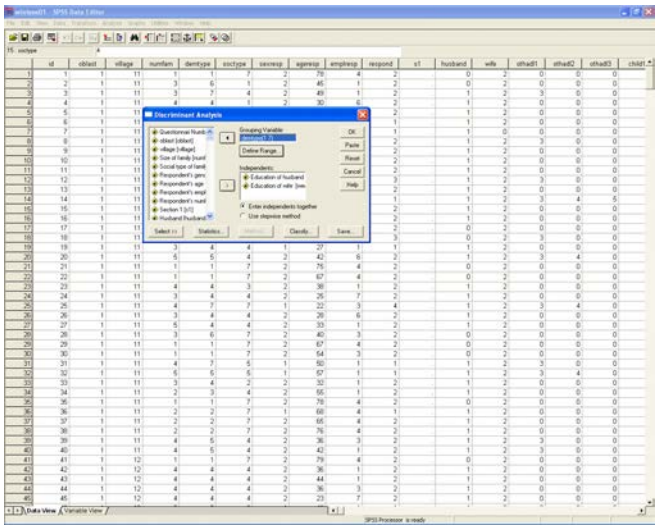


***Рис. 66.***

***Главное диалоговое окно процедуры Discriminant***

Дополнительные диалоговые окна слева направо имеют следующие выключатели: Select, Statistics, Method, Classify, Save. Во всех дополнительных диалоговых окнах минимум опций, необходимых для расчета модели, установлен по умолчанию. При этом к услугам пользователя имеется огромный выбор дополнительных опций, использование которых связано со спецификой решаемых им задач классификации. Ниже приведено описание всех дополнительных диалоговых окон.

Кнопка **Select** (отбор) открывает доступ к строке переноса переменной (Selection Variables) с выключателем (Value). При необходимости можно построить модель для одного из значений переменной, которая перенесена в рассматриваемую строку. Эта функция идентична таким же функциям отбора переменных, которые были рассмотрены в предшествующих главах (глава 13, § 13.2 и глава 14, § 14.2). По умолчанию она не задействована, поэтому модель строится по общим данным для группирующей и независимых переменных.

Кнопка **Statistics** (статистики) открывает дополнительное диалоговое окно Discriminant Analysis: Statistics. Оно имеет три контура: Descriptives (описательные статистики), Matrices (матрицы) и Function Coefficients (коэффициенты функции) с различными опциями, которые уже описаны ранее. В этом диалоговом окне отсутствуют опции, установленные по умолчанию.

Кнопка **Method** (метод), по умолчанию, как отмечалось выше, она недоступна. Доступ к ней открывается посредством установки опции главного диалогового окна - Use stepwise method (использовать пошаговый метод). Это позволяет открыть дополнительное диалоговое окно Discriminant Analysis: Stepwise Method. Оно имеет три контура: Method (метод), в котором по умолчанию стоит опция Wilk's Lambda (Лямбда Уилкса), Criteria (критерий) и Display (вывод на экран), в котором по умолчанию установлена опция Summary of steps (суммирование шагов).

Кнопка **Classify** (классифицировать) открывает дополнительное диалоговое окно Discriminant Analysis: Clsssification. Оно содержит четыре контура и один бокс. По умолчанию здесь установлены две опции: в контуре Prior Probabilities - опция All groups equal и в контуре Use Covariance Matrix - опция Within-groups. Все другие опции, в том числе и для вывода графиков, оставлены на усмотрение пользователя.

Наконец, кнопка **Save** (сохранить) открывает дополнительное диалоговое окно Discriminant Analysis: Save, в котором имеется возможность установки трех опций, но нет опций, установленных по умолчанию. Установка здесь опции Predicted group membership позволяет сохранять в рабочем файле новые переменные с именами dis_1, dis_2 и т.д.

Установка опции Discriminant Scores (значения дискриминанта) позволяет сохранять значения дискриминанта в виде новых переменных с именами dis1_1, dis1_2. А установка опции Probabilities of group membership сохраняет в рабочем файле новые переменные с именами dis1_3, dis2_3. Эти переменные фиксируют вероятность отнесения каждого случая к одной из выделяемых при моделировании групп. Использование всех этих возможностей система оставляет на усмотрение пользователя.

В рассмотренном ниже примере в качестве группирующей переменной взят демографический тип семьи, а в качестве независимых переменных - число лет учебы основной брачной пары (мужа и жены) и пол респондента. При построении модели использованы только опции, которые установлены системой по умолчанию. Результаты расчетов приведены в обрамлении 46.

Как видно из обрамления 46, окно просмотра модели, построенной на опциях, которые установлены системой по умолчанию, имеет семь таблиц.

## Обрамление 46. Окно просмотра процедуры Discriminant

## Discriminant

**Analysis Case Processing Summary**

| Unweighted Cases | | N | Percent |
|---|---|---|---|
| Valid | | 583 | 72,9 |
| Excluded | Missing or out-of-range group codes | 0 | ,0 |
| | At least one missing discriminating variable | 217 | 27,1 |
| | Both missing or out-of-range group codes and at least one missing discriminating variable | 0 | ,0 |
| | Total | 217 | 27,1 |
| Total | | 800 | 100,0 |

**Group Statistics**

| Demographic type of family | | Valid N (listwise) | |
|---|---|---|---|
| | | Unweighted | Weighted |
| Retired couple | Education of husband | 113 | 113,000 |
| | Education of wife | 113 | 113,000 |
| | Respondent's gender | 113 | 113,000 |
| Employed couple | Education of husband | 47 | 47,000 |
| | Education of wife | 47 | 47,000 |
| | Respondent's gender | 47 | 47,000 |
| Employed couple with children | Education of husband | 171 | 171,000 |
| | Education of wife | 171 | 171,000 |
| | Respondent's gender | 171 | 171,000 |
| Employed couple with children and other adults | Education of husband | 157 | 157,000 |
| | Education of wife | 157 | 157,000 |
| | Respondent's gender | 157 | 157,000 |
| Other | Education of husband | 95 | 95,000 |
| | Education of wife | 95 | 95,000 |
| | Respondent's gender | 95 | 95,000 |
| Total | Education of husband | 583 | 583,000 |
| | Education of wife | 583 | 583,000 |
| | Respondent's gender | 583 | 583,000 |

## Analysis 1
## Summary of Canonical Discriminant Functions

**Eigenvalues**

| Function | Eigenvalue | % of Variance | Cumulative % | Canonical Correlation |
|---|---|---|---|---|
| 1 | ,344[a] | 97,4 | 97,4 | ,506 |
| 2 | ,007[a] | 2,0 | 99,4 | ,084 |
| 3 | ,002[a] | ,6 | 100,0 | ,044 |

a. First 3 canonical discriminant functions were used in the analysis.

**Wilks' Lambda**

| Test of Function(s) | Wilks' Lambda | Chi-square | df | Sig. |
|---|---|---|---|---|
| 1 through 3 | ,737 | 176,331 | 12 | ,000 |
| 2 through 3 | ,991 | 5,270 | 6 | ,510 |
| 3 | ,998 | 1,139 | 2 | ,566 |

**Standardized Canonical Discriminant Function Coefficients**

| | Function | | |
|---|---|---|---|
| | 1 | 2 | 3 |
| Education of husband | ,497 | -,467 | 1,009 |
| Education of wife | ,609 | ,238 | -1,025 |
| Respondent's gender | ,222 | ,900 | ,380 |

**Structure Matrix**

| | Function | | |
|---|---|---|---|
| | 1 | 2 | 3 |
| Education of wife | ,891* | -,028 | -,452 |
| Education of husband | ,832* | -,377 | ,407 |
| Respondent's gender | ,196 | ,923* | ,331 |

Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions
Variables ordered by absolute size of correlation within function.

*. Largest absolute correlation between each variable and any discriminant function

**Functions at Group Centroids**

| Demographic type of family | Function | | |
|---|---|---|---|
| | 1 | 2 | 3 |
| Retired couple | -1,127 | -,007 | ,023 |
| Employed couple | ,243 | -,254 | -,058 |
| Employed couple with children | ,412 | -,020 | ,052 |
| Employed couple with children and other adults | ,373 | ,070 | -,015 |
| Other | -,138 | ,053 | -,067 |

Unstandardized canonical discriminant functions evaluated at group means

*Первая таблица* –Analysis Case Processing Summary (общий анализ случаев). Она выводит общий обзор действительных и отсутствующих значений. Из данных этой таблицы видно, что группирующая переменная представлена 583 случаями (72,9%) из 800, имеющихся в выборке. Два из семи демографических типов семьи (одиночки и неполные семьи), которые составляют 27,1% массива, выпали из анализа. В них нет брачной пары, т.е. они не соответствуют условию, которое сформулировано введенными независимыми переменными.

*Вторая таблица* – Group Statistics (статистики групп). Эта таблица дает статистику случаев для каждой из пяти семейных групп, вошедших в анализ, и для всего анализируемого массива в целом. В каждой группе число случаев соответствует числу случаев, имеющихся в трех независимых переменных. Если бы были заданы опции на вывод описательных статистик (среднее, стандартное отклонение), то они были бы выведены в дополнительных столбцах рассматриваемой таблицы.

*Третья таблица* – Eigenvalues (собственные значения). Она открывает собой группу собственно аналитических таблиц, о чем система информирует в соответствующих подзаголовках. В этой таблице представлены основные характеристики трех канонических дискриминантных функций, вошедших в анализ. В первом слева столбце представлены сами дискриминантные функции, их число равно числу независимых переменных, т.е. три.

Большие собственные значения (второй слева столбец) указывают на удачно подобранные дискриминантные функции. Дисперсия (третий слева столбец) говорит о том, что в нашем случае основная объяснительная нагрузка (97%) лежит на первой дискриминантной функции. Далее, следуют два столбца с простыми и кумулятивным процентами. Последний столбец содержит значение коэффициента канонической корреляции. Это значение тем лучше, чем оно ближе к 1.

*Четвертая таблица* – Wilk's Lambda (Лямбда Уилкса). Эта таблица дает информацию о значимости отличия в выделенных дискриминантных группах средних значений дискриминантной функции. В приводимом примере они значимы только для первой из трех групп.

*Пятая таблица* – Standardized Canonical Discriminant Function Coefficient (нормированные канонические коэффициенты дискриминантной функции). Она позволяет увидеть корреляцию каждой переменной с нормированными значениями самой функции.

*Шестая таблица* – Structure Matrix (структурная матрица). В этой таблице, как видно из примечания к ней, даются объединенные корреляции внутри групп между дискриминантными переменными и нормированными каноническими значениями дискриминантной функции. При этом переменные здесь располагаются не в порядке их ввода, как это было в предшествующей таблице, а в соответствии с абсолютными значениями корреляции в первой дискриминантной функции. Поэтому наибольшие корреляционные значения ,891 и ,832 помечены звездочкой и выведены на первое и второе место.

*Седьмая таблица* – Functions at Group Centroids (функции групповых центроидов). Из примечания к этой таблице видно, что она содержит ненормированные канонические дискриминантные функции, оцениваемые по групповым средним значениям. В нашем примере наиболее высокие средние значения в первой дискриминантной функции имеют нуклеарные и сложные семьи.

В принципе рассматриваемая процедура позволяет не только формировать дискриминантные группы, но и прогнозировать принадлежность к исходной группе по дискриминантной функции. Для этого в дополнительном диалоговом окне Classify необходимо установить две опции: Casewise results (результаты отдельных случаев) и Summary table (сводная таблица).

**Правило 39**

***Критерии оценки качества кластерной модели:***
- ***выполнение требований к данным, предъявляемым анализом;***
- ***разбиение массива минимум на два кластера;***
- ***возможность интерпретации кластеров.***

**Правило 40**

***Критерии оценки качества дискриминантной модели:***
- ***большие собственные значения;***
- ***значение коэффициента канонической корреляции;***
- ***значимости отличия в выделенных дискриминантных группах средних значений дискриминантной функции;***
- ***объединенные корреляции внутри групп между дискриминантными переменными и нормированными каноническими значениями дискриминантной функции.***

## Задание для самостоятельной работы

1. Что такое многомерная классификация?
2. Какие задачи решает многомерная классификация?
3. Какие виды многомерной классификации имеются в SPSS?
4. Что такое кластерный анализ?
5. Что такое дискриминантный анализ?
6. В чем различие фактора, кластера и дискриминантной функции?
7. Опишите устройство главного диалогового окна Cluster Analysis.
8. В чем назначение строки Selection Variables?
9. Перечислите контуры и дополнительные диалоговые окна процедуры TowStep Cluster в SPSS?
10. Какие опции по умолчанию стоят в дополнительных диалоговых окнах процедуры TowStep Cluster в SPSS?
11. Какие опции по умолчанию стоят в процедуре Discriminant в SPSS?
12. Назовите имена новых переменных, создаваемых кластерным и дискриминантным анализом?
13. Какие критерии качества кластерной и дискриминантной моделей?

# Глава 16. Моделирование в среде «AMOS»

## 16.1. Описание модуля

В качестве средства моделирования данных Amos (Analysis of moment structures - анализ моментных структур) впервые появился в виде самостоятельного приложения к 7-й версии SPSS. Уже в 9-й версии пакета, оставаясь самостоятельным модулем, он, тем не менее, был включен в качестве составной части в главное меню. Предварительно установив это приложение[1], его было можно открыть, используя путь: Analyze – Amos.

Начав интеграцию процедур визуального моделирования в основной пакет программы, разработчики SPSS открыли для социальной науки новые и весьма благоприятные возможности освоения методов моделирования. К сожалению, видимо, коммерческие интересы возобладали, поэтому в последующих версиях системы движение в этом направлении было прекращено, а прямой путь к самому приложению был изъят из главного меню.

Между тем, будучи интегрированным в SPSS, Amos или сходный с ним продукт неизбежно станет весьма мощным инструментом изучения социальных процессов. Такие программные продукты начнут получать быстрое распространение по мере освоения широкими слоями прикладников-социологов теоретических принципов и методологии компьютерного моделирования.

С освоением методов моделирования социальная наука, возможно впервые в своей истории, получает исключительно благоприятные возможности формирования научно-обоснованного, а главное всегда проверяемого знания. Проектно-конструкторские решения и визуальное восприятие модели в Amos могут служить гигантским стимулом к пониманию ее формально-логических оснований и связанных с ними

---

1. В распространяемых на нашем рынке версиях, например, CD: SPSS 10.0.5 (2001, Delta-MM Corp.) и SPSS 11.5 - CD: Вся статистика от SPSS 2003 (2003, Prism Digital company LTD), Amos все еще отсутствует.

количественных соотношений. Используя для построения моделей такую интеллектуальную среду, легче и проще понять различные аспекты моделирования.

Еще одним важным достоинством Amos является его совместимость с такими известными программными продуктами, как MS Excel, FoxPro, Access и др. Имея базовое расширение файла .amw, совместимое с родовыми файлами SPSS с расширением .sav, Amos позволяет выполнять расчеты с файлами, имеющими расширение .xls, .dbf, .mdb, .wk1, .wk3, .wk4, .txt, .csv и др. Это весьма существенно расширяет возможности его использования. Резонно предположить, что у нас в стране пользователей Excel или Access гораздо больше, чем пользователей SPSS.

В целом Amos 4.0 предполагает выполнение двух основных типов работ. Первый из них связан с выполнением работ по построению и расчетам различных характеристик моделей. Эти работы выполняются в среде и с помощью **Amos Graphics**.

Второй тип связан с выполнением работ по построению моделей программным путем в особой среде, которая называется Amos Basic. Эта среда совместима и поддерживает большую группу различных языков программирования от MS Visual Basic до MS C++ (41, С. 45). В настоящем пособии последний аспект вообще не рассматривается. А основное внимание уделено выработке навыков работы в Amos Graphics.

С интеграцией Amos в SPSS в нем стали появляться не характерные ранее для пакета черты, связанные с проблемами совместимости последующих версий с предыдущими. Так, модели, построенные в Amos 3.0, открываются в Amos 4.0, но при сохранении модифицируются и становятся недоступными в Amos 3.0. С целью преодоления этого противоречия разработчиками введена дополнительная (хорошо знакомая по работе в MS Word) команда, рекомендующая сохранять модели верхней версии под новыми именами.

**Напоминание**

***Освоение среды моделирования Amos предполагает:***
- ***знание основ статистики и статистического анализа;***
- ***наличие навыков работы в SPSS;***
- ***умение работать с объектами (таблицы, графики, рисунки или OLE-objects в MS Word).***

## 16.2. Рабочий стол и инструменты моделирования Amos Graphics

После установки Amos доступен тремя способами:

- Один из них связан с использованием Amos 3.0 или 3.61 в качестве приложения к версиям SPSS 7.0 и 8.0.

- Другой способ связан с использованием Amos 4.0 и выше в версиях SPSS 9.0 и 10.0. В этом случае для того, чтобы при уже открытом рабочем файле войти в Amos, необходимо выполнить следующую последовательность команд главного меню в версиях SPSS 9.0 и 10.0: Analyze – Amos. При выполнении указанной последовательности команд на фоне ваших данных сначала появляется фирменная табличка, информирующая о том, что выполняется команда открытия Amos 4, а затем на экране появляется рабочий стол с инструментами проектирования (рис. 67).

- Третий способ – использование самой операционной системы Windows для открытия Amos 4.0 и выше после его установки. В этом случае предполагается выполнение последовательности команд:

**Start**

**Programs**

**Amos 4**

**Amos Graphics.**

После выполнения указанной последовательности команд, как и в случае использования пути Analyze - Amos, сначала, но уже не в самой системе SPSS, а на фоне основного окна Windows появляется та же самая фирменная табличка, информирующая о том, что выполняется команда открытия Amos 4. Если все идет хорошо и правильно, то в результате выполнения описанной выше последовательности команд на экране появляется несколько необычный на первый взгляд рабочий стол с инструментами проектирования (рис. 67).

В SPSS при отсутствии открытого рабочего файла доступ к Amos закрыт, а при наличии любого открытого рабочего файла Amos может сразу выставить на своем рабочем столе последний, выполненный ранее опыт моделирования. Любая модель, построенная в Amos и сохраненная с расширением (*.amw), может быть открыта с помощью двойного клика на ее имени в правом окне Windows Explorer. При этом, даже не открывая рабочий файл SPSS, пользователь получает возможность работать с этой моделью. Только в случае отсутствия

файла с характерным для Amos расширением .amw, гарантировано открытие чистого рабочего стола.

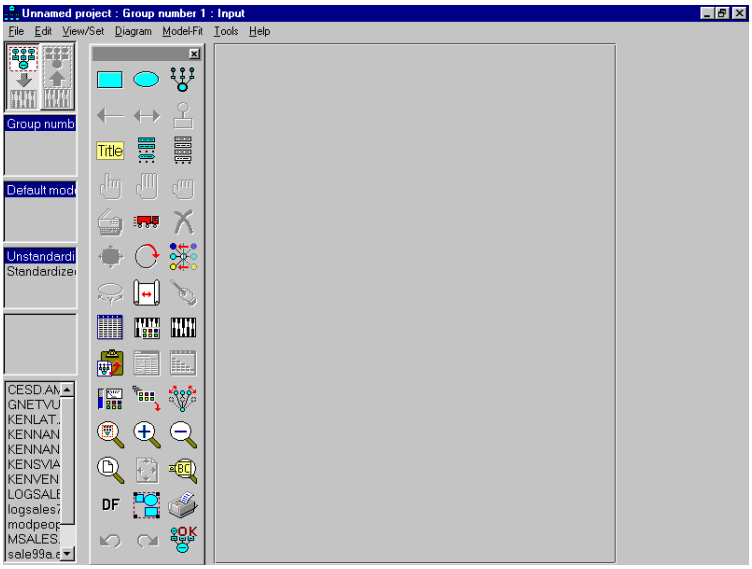



***Рис. 67.***

***Рабочий стол и инструменты моделирования Amos Graphics***

Для построения новой модели следует выполнить стандартную последовательность команд главного меню самого Amos, а именно File-New. При желании открыть другую модель, надо использовать список файлов с моделями, который имеется на экране в панели заголовков (нижняя часть левой колонки) или идти уже по известному пути, т.е. непосредственно в Amos (а не в SPSS) и выполнить последовательность команд: File-Open-Data - имя файла с расширением .amw. С этой целью необходимо в списке «тип файла» основного окна Data установить команду «все файлы». Двойной клик на имени модели открывает окно с названием имени модели и приписки – «SPSS for Windows Syntax Editor». Уже из названия этого окна видно, что в нем могут быть выполнены все работы, связанные с изменением и корректировкой модели. Но для этого надо быть хорошо подготовленным пользователем.

**Внимание**

***Через главное меню SPSS путем выполнения команд File – Open – Data - имя файла .amw модель не может быть открыта. В этом случае открывается описание модели в языке Syntax.***

Как видно на рис 67, экран Amos 4.0 включает в себя следующие пять основных составляющих: рабочий стол, главное меню, строка состояния в верхней части экрана, панель заголовков и панель инструментов. Все они обстоятельно описаны ниже.

1. **Рабочий стол** (large rectangle) - основная, взятая в обрамление, серая часть экрана. Условно говоря, она представляет собой лист бумаги. Именно в этой части выполняется вся черновая работа по построению модели.

2. **Главное меню** (menu) находится там же и выглядит так же, как и главное меню SPSS (глава 1, § 1.2). Оно включает в себя следующие разделы: File, Edit, View/Set, Diagram, Model-Fit, Tools, Help.

3. **Строка состояния** - в верхней части экрана. Она информирует об имени файла, находящегося на рабочем столе и его состоянии (Input или Output). На краях этой строки, как и в SPSS, находится слева от имени файла одна кнопка (пиктограмма модели), а справа - три стандартные кнопки управления всем экраном (свернуть, расшиить, закрыть).

4. **Панель заголовков** (several menu titles), находящаяся в левой части экрана, состоит из шести блоков, расположенных друг под другом.

*Первый блок сверху* содержит две большие кнопки со стрелками. Они предназначены для смены вида рабочего стола (Input - Output) и действуют только в альтернативном режиме.

Кнопка входа (Input) доступна по умолчанию (направление стрелки вниз). Она обеспечивает доступ к файлам, панели инструментов и рабочему столу. Только в этом режиме открываются, строятся и просчитываются модели.

Кнопка выхода (Output) становится доступной путем нажатия после построения модели. Она позволяет видеть на экране и распечатать такие характеристики модели, как числовые значения связей (ковариаций и корреляций), а также дисперсию для каждой переменной модели. Смена режимов в этом блоке может быть осуществлена непосредственно после выполнения расчетов модели.

*Второй блок* – номер группы (Group number 1). По умолчанию здесь стоит указание на выполнение расчетов в группе № 1.

Переход к следующей группе осуществляется путем клика мышью по заголовку окна. Группа аналогична одной из версий (итераций) модели. Их может быть несколько, и надо быть внимательным к тому, где вы находитесь в данный момент. В противном случае может возник

нуть ситуация, когда вроде бы ничего не произошло, а на рабочем столе появилась какая-то старая, уже отвергнутая ранее, версия модели. При наличии нескольких групп в окне рассматриваемого блока появляется их список. В результате этого открывается возможность обращения к ним и работы с ними.

*Третий блок* – модель по умолчанию (Default model).

*Четвертый блок* – ненормированных и нормированных оценок (Unstandardized estimates или Standardized estimates). Как и в случае первого блока здесь действует режим альтернативы. По умолчанию задаются ненормированные оценки.

*Пятый блок* – описание результатов (Writing output). Заголовок и текст в этом блоке появляются лишь по результатам расчета модели. В момент расчета модели здесь можно видеть, как прокручиваются данные. Это может служить хорошим признаком приближения к благоприятному (в смысле построения модели, а не ее характеристик) исходу. В случае такого исхода на правой стенке окна появляется прокрутка для просмотра основных характеристик модели.

*Шестой (последний) блок* - список файлов с моделями. Он не имеет отдельного заголовка. Как и в предшествующем блоке, здесь есть прокрутка для просмотра списка файлов с моделями (в случае их наличия).

5. **Панель инструментов** (toolbar) с 42 кнопками (иконками) команд, необходимых для построения, расчета и редактирования модели. При открытии Amos 4.0 панель инструментов по умолчанию находится между рабочим столом и панелью заголовков. Будучи втиснутой между ними, она часто имеет вид двух колонок иконок, уходящих в нижнюю часть экрана. Приведение ее в нормальное рабочее состояние предполагает раздвижку стенок до трех колонок. Делается это точно таким же путем, как в MS Word, т.е. с помощью мыши изменяются параметры строк и колонок таблицы. В результате такого рода операции на экране появляется 14 рядов иконок по 3 в каждом ряду (рис 67). Они описаны ниже в порядке представления сверху вниз и слева направо.

*1-й ряд.* 1.1. Прямоугольник (rectangle) – элемент блок-схемы модели, в который записывается имя переменной. Он может быть задан как непосредственно кликом мыши с последующим переносом на рабочий стол, так и через главное меню путем: Diagram – Draw Observed или клавишей F3 клавиатуры. Использование любой из этих возможностей открывает доступ к иконке. Она вдавливается и

начинает светиться, а на нижнем конце стрелки курсора появляется белый прямоугольник, свидетельствующий о готовности системы выполнить данную команду, связанную с рисованием оболочки для переменной. Каждая переменная имеет свою оболочку. Выход из режима использования этой и других подобных команд панели инструментов выполняется либо кликом мышью по той же иконке, что ведет к выходу из панели инструментов, либо путем нажатия другой кнопки на панели инструментов, что ведет к смене инструмента моделирования.

1.2. Эллипс (ellipse) – элемент блок-схемы модели, в который записывается имя латентной переменной (unobserved variable) или ошибки (error). Он может быть задан как непосредственно кликом мыши с последующим переносом на рабочий стол, так и через главное меню путем Diagram – Draw Unobserved или клавишей F4 клавиатуры. Эллипс - оболочка латентной переменной.

1.3. Индикаторы латентной переменной (draw a latent variable or draw indicator variable) - элемент блок-схемы модели. Они одновременно задают дополнительный рисунок и комплекс параметров латентной переменной. Этот элемент не имеет самостоятельного значения. Он как бы одевается на уже построенные эллипсы-оболочки путем его выделения и клика мышью в момент ее нахождения внутри оболочки латентной переменной (эллипса). При каждом дополнительном клике элементы этой иконки будут как бы размножаться на оболочке латентной переменной. Элемент дублирует команду главного меню Diagram-Draw Indicator Variable.

*2-й ряд.* 2.1. Односторонняя стрелка (draw path) предназначена для обозначения зависимостей между переменными (regression weights). Она может быть задана как непосредственно кликом мыши с последующим переносом на рабочий стол, так и через главное меню путем Diagram-Draw Path или клавишей F5 клавиатуры. В любом случае после активирования[1] этого элемента установка стрелки может быть выполнена только путем помещения курсора в оболочку независимой переменной (при этом ее контур окрашивается в красный свет) и последующего протягивания курсора со стрелкой к зависимой переменной (при этом ее контур становится белым). Стрелка устанавливается в направлении от независимой к зависимой переменной при выделении их контуров.

[1] «Активация - возбуждение или усиление активности, переход в деятельное состояние». «Активировать - производить, вызывать активацию чего-л.» (44, С. 28).

2.2. Двухсторонняя стрелка (draw covariance) предназначена для обозначения взаимосвязей (covariances) и взаимозависимостей (correlations) между переменными. Она может быть задана как непосредственно кликом мыши с последующим переносом на рабочий стол, так и через главное меню путем: Diagram Draw Covariance или клавишей F6 клавиатуры.

Этот элемент устанавливается таким же образом, как и односторонняя стрелка (2.1). Двухсторонняя стрелка может устанавливаться только между независимыми переменными и между ошибками.

2.3. Установка неизвестной переменной (add a unique variable) - элемент блок-схемы модели (небольшой рисунок, представляющий комбинацию прямоугольника и эллипса), которым одновременно задается оболочка и параметры ошибки (error). Этот элемент, как и элемент 1.3, не имеет самостоятельного значения. Его отличие состоит в том, что он может устанавливаться на оболочке любых переменных, включенных в модель.

Установка этого элемента на переменную автоматически делает ее зависимой переменной. Это значит, что на нее уже нельзя устанавливать элементы взаимосвязи (2.2). При повторных кликах по оболочке этот элемент не размножается, а как бы двигается справа налево по контуру оболочки, совершая за четыре клика оборот в $360^0$. Этот элемент дублирует команду главного меню Diagram-Draw Unique Variable.

*3-ряд.* 3.1. Заголовок модели (title) – иконка, с помощью которой непосредственно на рабочем столе можно написать название и другие интересующие экспериментатора характеристики модели. Особенность ее использования состоит в том, что, во-первых, текст будет помещен на рабочем столе там, где в этот момент стоит курсор.

Во-вторых, клик мышью в месте написания текста при активированной кнопке заголовка ведет к открытию дополнительного диалогового окна Figure Caption, в котором текст и пишется. Это окно имеет команды изменения размера, выделения и центрирования текста. Выйти из режима написания текста можно, нажав комбинацию клавиш Ctrl-Enter.

При первичном написании текста и отсутствии опыта такого рода работы довольно трудно получить хороший заголовок. Но клик по черновому тексту на рабочем столе опять возвращает в режим его написания, который теперь уже можно использовать с целью редакти

рования и доводки до нужной кондиции. Этот элемент дублирует команду главного меню Diagram-Figure Caption.

3.2. Список переменных модели (list variables in model). Клик мышью по этой иконке (три маленьких зеленых прямоугольника) открывает дополнительное диалоговое окно Variables in model, которое позволяет видеть названия переменных, вписанных в уже построенные базовые элементы блок-схемы модели - прямоугольники (1.1) и эллипсы (1.2). Этот список можно видеть только в случае, если подобная работа уже проделана. Дублирует команду главного меню View/Set-Variable in Model.

3.3. Список переменных файла с данными, используемого для построения модели (list variables in data set). Клик мышью по этой иконке позволяет видеть в открывающемся при этом диалоговом окне Variables in dataset названия всех переменных, переносить или переписывать те из них, которые необходимо вносить в уже построенные оболочки - прямоугольники (1.1). Перенос переменной осуществляется путем ее выделения (клик мышью по переменной) и перетаскивания ее названия (с помощью появившейся на конце стрелки курсора пиктограммы в виде стопки страниц) в оболочку переменной. Как правило, эта операция требует последующего редактирования, так как с названием переменной перетаскивается и ее описание. Дублирует команду главного меню View/Set-Variable in Dataset.

*4-й ряд.* 4.1. Выделение одного объекта (select one object) – иконка (кисть руки с указательным пальцем), позволяющая выделять отдельные элементы блок-схемы модели.

Эта команда может выполняться и через главное меню Edit-Select и с помощью клавиши F2. Активирование этой кнопки и последующая установка курсора на любой элемент модели делает его контур сначала красным, а в случае клика (выделения) - голубым. С помощью этой команды постепенно, один за другим можно выделить по очереди все элементы модели.

4.2. Выделение всех объектов (select all objects) – иконка (раскрытая кисть руки), позволяющая выделить все объекты модели одновременно. Эта команда аналогична команде главного меню Edit-Select all. Ее выполнение выделяет все элементы рабочего стола (окрашиваются в голубой цвет).

4.3. Снятие выделения всех объектов (deselect all objects) – иконка (сжатая кисть руки), позволяющая снять выделение. Эта команда аналогична команде главного меню Edit-Deselect all и кнопке F11. Выпол

нение этой команды отменяет выделение всех элементов рабочего стола и возвращает им заданный по умолчанию, черный цвет.

*5-й ряд.* 5.1. Дублирование объектов (duplicate objects) – инструмент, позволяющий дублировать базовые элементы (оболочки 1.1 и 1.2) блок-схемы модели. Эта команда выполняется путем активации иконки и последующего помещения курсора с индикатором команды в предполагаемую к тиражированию оболочку. При этом контур оболочки становится красным. Клик мышью в данном положении позволяет тащить автоматически появляющуюся ее копию в нужное место рабочего стола. Команда позволяет сделать все оболочки одного размера, что естественно улучшает вид модели. Дублирует команду главного меню Edit-Duplicate.

5.2. Перемещение объектов (move objects) – инструмент, позволяющий изменять положение базовых элементов (оболочек - 1.1 и 1.2 и стрелок – 2.1 и 2.2) блок-схемы модели. Команда выполняется путем активирования иконки и последующего помещения курсора с индикатором команды на предполагаемый к перемещению объект. При этом контур объекта становится красным. Клик мышью в данном положении позволяет тащить объект в любое место на рабочем столе. Команда дает возможность перестраивать блок-схему модели. Ее использование может способствовать улучшению вида модели. Дублирует команду главного меню Edit-Move и комбинацию клавиш Ctrl+M.

5.3. Удаление объектов (erase objects) – инструмент, дающий возможность удалять различные элементы модели. Команда выполняется путем активации иконки и последующего помещения курсора с индикатором команды на удаляемый объект. При этом контур объекта становится красным. Клик мышью в данном положении ведет к удалению выделенного объекта. Этот элемент дублирует команду главного меню Edit-Erase и клавишу Del.

*6-й ряд.* 6.1. Изменение размера объектов (change the shape of objects) - инструмент, позволяющий изменять размеры оболочек (1.1-1.3 и 2.3) и двухсторонних стрелок (2.2). Команда выполняется путем активирования иконки и последующего помещения курсора с индикатором команды на объект, размеры которого предполагается изменить. При этом контур объекта становится красным. Клик мышью в данном положении и последующее протягивание ведут к изменению размеров выделенного объекта. Дублирует команду главного меню Edit-Shap of Object.

6.2. Поворот индикаторов латентных переменных (rotate the indicators of a latent variables) - инструмент, позволяющий поворачивать все индикаторы латентной переменной (1.2) справа налево вокруг контура переменной. Команда выполняется путем активирования иконки и последующего помещения курсора с индикатором команды в контур эллипса латентной переменной. При этом контур становится красным. Каждый клик мышью в данном положении ведет к изменению положения индикаторов латентной переменной (1.3) на $90^0$. Дублирует команду главного меню Edit-Rotate.

6.3. Зеркальное отражение индикаторов латентной переменной (reflect the indicators of a latent variable) - инструмент, позволяющий перебрасывать все индикаторы латентной переменной (1.2) справа налево или сверху вниз вокруг контура переменной. Команда выполняется путем активирования иконки и последующего помещения курсора с индикатором команды в контур эллипса латентной переменной. При этом контур становится красным. В зависимости от исходного размещения индикаторов каждый клик мышью в данном положении ведет к изменению положения индикаторов латентной переменной (1.3) на $180^0$, т.е. туда и обратно, но всегда в одной плоскости. Дублирует команду главного меню Edit-Reflect.

*7-й ряд.* 7.1. Перемещение параметров оценок (move parameter value) - инструмент, позволяющий перемещать параметры оценок различных объектов (1.1-1.3 и 2.1-2.3). Команда выполняется путем активирования иконки и последующего помещения курсора с индикатором команды (белый квадратик на конце курсора) в контур объекта. При этом контур становится красным. Независимо от исходного размещения параметров оценок каждый клик мышью в данном положении ведет к их перемещению в позицию, указанную индикатором этой команды. Дублирует команду главного меню Edit-Move Parameter.

7.2. Изменение положения рабочего стола на экране (reposition the path diagram on the screen) - инструмент, позволяющий двигать рабочий стол по экрану. Команда выполняется путем активирования иконки и последующего помещения курсора с индикатором команды (маленький прямоугольник со словом scroll) в любое место рабочего стола. Клик мышью с его фиксацией и последующей протяжкой сверху вниз и справа налево ведет к изменению положения рабочего стола. При этом полезно обращать внимание на то, что происходит с обрамлением стола. Дублирует команду главного меню Diagram-Scroll.

7.3. Указка (touch up a variable) - инструмент презентации, позволяющий указывать на отдельные объекты модели путем их временного выделения. Команда выполняется путем активирования иконки и последующего помещения курсора с индикатором команды (указка) на интересующий объект, контур которого при этом становится красным. Дублирует команду главного меню Edit-Touch Up и комбинацию клавиш Ctrl+H.

8-й ряд. 8.1. Выделение файла с данными (select data files) - команда, позволяющая получать доступ к файлу с данными и его различным характеристикам. Команда выполняется путем активирования иконки, что ведет к открытию окна Data Files. Дублирует команду главного меню File-Data Files и комбинацию клавиш Ctrl+D.

8.2. Анализ свойств модели (analysis properties) - исключительно важная команда, с помощью которой задаются расчетные характеристики модели. Команда выполняется путем активирования иконки, что ведет к открытию окна Analysis Properties. Это окно имеет 9 подокон, из которых два: Estimation и Output - обязательны к заполнению. Дублирует команду главного меню View/Set-Analysis Properties и комбинацию клавиш Ctrl+A.

8.3. Расчет модели (calculate estimate) - основная команда, позволяющая выполнить расчет модели. Команда выполняется путем активирования иконки, что ведет либо к выполнению расчетов (о чем свидетельствует начало движения текста и чисел в пятом блоке панели заголовков), либо к появлению окна, в котором система информирует, почему она не может выполнить расчет. Дублирует команду главного меню Model-Fit-Calculate Estimate и комбинацию клавиш Ctrl+F9.

*9-й ряд.* 9.1. Копирование (copy the path diagram to the clipboard)-сервисная команда, позволяющая копировать рабочий стол. Эта иконка дублирует команду главного меню Edit-Copy (to clipboard) и комбинацию клавиш Ctrl+C.

9.2. Просмотр текстового описания модели (view text) - одна из важнейших команд, позволяющая видеть и читать текстовый файл в окне вывода. Команда может быть выполнена только после расчета модели (8.3). Она выполняется путем нажатия на иконку. В результате на экране открывается окно, имеющее текстовый файл с именем модели и расширением .amo. Эта иконка дублирует команду главного меню View/Set-Text Output .

9.3. Просмотр электронной таблицы модели (view spreadsheet) - команда, позволяющая получать результаты расчета модели в виде

электронной таблицы. Команда выполняется путем нажатия на иконку. В результате открывается окно, имеющее самостоятельное меню с несколькими разделами (Title, Variable Summary, Parameter Summary, Notes for Group, Notes for Model, Minimization History). Выделение любого раздела кликом мышью в верхнем левом списке ведет к открытию его содержания в отдельном поле, справа от списка. Этот элемент дублирует команду главного меню View/Set-Table Output.

*10-й ряд.* 10.1. Характеристики объекта (object properties) - сервисная команда, позволяющая получать доступ к свойствам объекта (тексту описания, параметрам, формату и др.). Эта иконка дублирует команду главного меню View/Set-Object Property и комбинацию клавиш Ctrl+O. Окно с характеристиками объекта открывается и более удобным способом: двойным кликом мышью по оболочке объекта.

10.2. Перемещение характеристик объекта (drag properties) - иконка, открывающая окно Drag Properties. Пометка в нем отдельных характеристик объекта позволяет получать их для всех объектов. Дублирует команду главного меню Edit-Drag Properties и комбинацию клавиш Ctrl+G.

10.3. Сохранение симметрии (preserve symmetries) - сервисная команда, позволяющая перемещать на рабочем столе ранее выделенную группу объектов. При этом сохраняется весь рисунок этой группы объектов.

*11-й ряд.* 11.1. Увеличение выделенной части (zoom in on area that you select) - сервисная команда, позволяющая увеличивать выделенный объект или группу объектов. Эта команда, как и некоторые другие, например, две следующие, нечто специальное, рассчитанное на большого любителя. Дублирует команду главного меню Diagram-Zoom.

11.2. Просмотр отдельной части (view a smaller area of the path diagram) - сервисная команда. Ее выполнение ведет к увеличению размеров рабочего стола. Это возможность просмотра его отдельных частей. Дублирует команду главного меню Diagram-Zoom In и клавишу F7.

11.3. Просмотр большей части (view a bigger area of the path diagram) - сервисная команда, выполнение которой ведет к уменьшению размеров рабочего стола и соответственно возможности просмотра его в целом, как это и задано по умолчанию. Дублирует команду главного меню Diagram-Zoom Out и клавишу F8.

*12-й ряд.* 12.1. Показ текущей страницы рабочего стола на экране

(show the entire page on the screen) - совершенно замечательная сервисная команда. Эта команда спасет «жизнь» и настроение пользователя, использовавшего три предшествующие иконки. Она возвращает рабочий стол и все, что на нем было на момент последнего сохранения файла, в исходное состояние. Дублирует команду главного меню Diagram-Zoom Page и клавишу F9.

12.2. Компоновка положения и размеров модели на рабочем столе (resize the path diagram to fit on a page) - очень нужная и полезная команда. Она позволяет делать последний штрих в рисунке модели перед печатью. В результате ее выполнения модель автоматически размещается на странице наилучшим образом. Команда выполняется путем клика мышью по иконке. Этот элемент дублирует команду главного меню Edit-Fit to Page и комбинацию клавиш Ctrl+F.

12.3. Лупа (examine the path diagram with a loupe) - сервисная команда, позволяющая смотреть через хорошую лупу на любое место на рабочем столе. Команда выполняется путем клика мышью по иконке. Дублирует команду главного меню Diagram-Loupe и клавишу F12.

*13-й ряд.* 13.1. Степень свободы (DF) - сервисная команда, позволяющая видеть в специальном окне - Degrees of freedom различные характеристики степени свободы. Команда выполняется путем клика мышью по иконке. Дублирует команду главного меню Model-Fit-Degrees of freedom.

13.2. Связь выделенного объекта (link the selected object) - сервисная команда, позволяющая выделить все имеющееся на рабочем столе и снять это выделение. Эта команда как бы объединяет в себе возможности двух команд 4.2-4.3. Дублирует команду главного меню Edit-Link.

13.3. Печать (printer). Эта команда выводит модель на печать. Дублирует команду главного меню File-Print и комбинацию клавиш Ctrl+P.

*14-й ряд.* 14.1. Возврат к положению, которое предшествует совершенному действию (undo or undo the previous change), - сервисная команда, позволяющая отменить неосторожное действие на рабочем столе. Дублирует команду главного меню Edit-Undo и комбинацию клавиш Ctrl+Z.

14.2. Вперед (redo or undo the previous undo) - возвращает к положению, которое было до выполнения предшествующей команды. Дублирует команду главного меню Edit-Redu  и комбинацию клавиш Ctrl+Y.

14.3. Перерисовка модели на рабочем столе (redraw the path diagram on the screen) - сервисная команда, фиксирующая уже имеющийся на

рабочем столе рисунок и текст. Она не обладает замечательными свойствами команды 12.1. Выполняется сразу же при нажатии на иконку. Дублирует команду главного меню Diagram-Redraw diagram.

**Обобщение**

***Функционально команды панели инструментов могут быть разбиты на четыре группы:***
***Первая - элементы блок-схемы модели: оболочки, стрелки, индикаторы (1.1-2.3).***
***Вторая - инструменты построения и редактирования модели (3.1-7.1, 12.2).***
***Третья - инструменты расчета и вывода текстового описания модели (8.1--8.3, 9.2-9.3).***
***Четвертая - сервисные инструменты (7.2-7.3, 9.1, 10.1-12.1, 12.3-14.3).***

Структурирование экрана и интерфейса в Amos 4.0 имеет весьма заметные отличия по сравнению с предшествующими версиями 3.60 и 3.61. Эти различия хорошо заметны, но они не имеют принципиального характера. Просто панель заголовков в ранних версиях еще оставалась составной частью панели инструментов, которая при этом оказывалась весьма перегруженной и имела два варианта предоставления доступа к выполняемым командам.

**Правило 41**

***Практически все команды панели инструментов и панели заголовков дублируются командами главного меню.***
***Раздел главного меню, в котором находится та или иная дублирующая команда, связан с ее функциональным назначением.***

## 16.3. Построение блок-схемы модели — Input

Знакомство с рабочим столом и инструментами моделирования в Amos Graphics, а также умение оперировать ими (при выполнении прочих необходимых условий, связанных с наличием соответствующих данных) позволяет приступить к построению модели. Для этого неплохо знать теорию моделирования, но так как этому нигде не учат, то использование метода «проб и ошибок», равно как и «учебы на марше», просто неизбежны для основной массы пользователей.

Всего этого нельзя бояться, но надо проявить наблюдательность и определенную способность делать позитивные выводы из первых почти неизбежно негативных результатов. Здесь, как и во многих других случаях, сначала полезно выяснить для самого себя, какую модель нужно построить, и что должно получиться в результате ее построения.

В руководстве по использованию Amos 4.0, например, приведено более 20 конкретных расчетов в качестве образцов для разработки различных моделей (19, pp. 61-386). Оценка дисперсии нескольких переменных и их взаимосвязей (estimating variance and covariance), оценка и тестирование гипотез о средних (estimating and testing hypotheses about means), линейная регрессия (conventional linear regression), факторный анализ (factor analysis) – все эти и многие другие расчеты подобного рода представляют собой различные способы моделирования.

Общий вид одной из наиболее простых моделей приведен на рис. 68. С помощью этой модели устанавливаются и измеряются корреляционные связи между четырьмя переменными. В контурах переменных (прямоугольниках) видны их имена, а сами переменные соединены элементами взаимосвязи (двухсторонние стрелки).

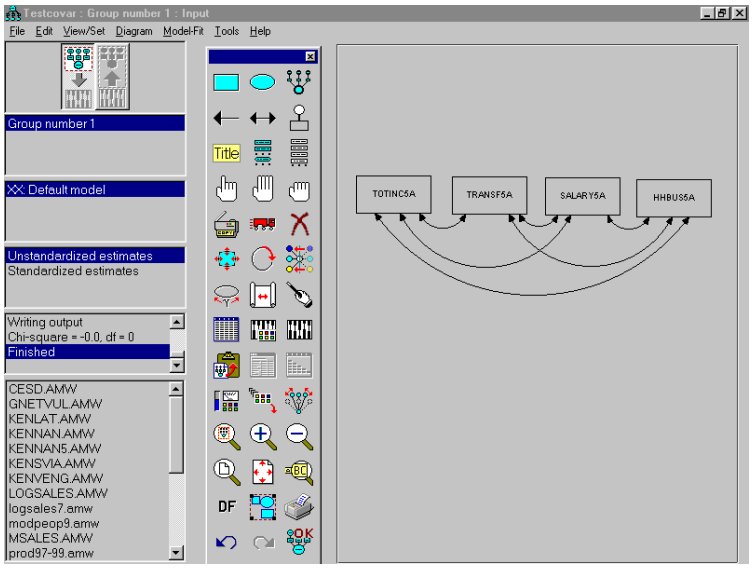



***Рис. 68.***

***Общий вид корреляционной модели в Amos Graphics***

В качестве еще одного примера, ниже на рис. 69, приведен общий вид факторной модели. В этой модели уже шесть переменных (прямоугольники) объединены в два фактора (окружности). Каждый фактор связан с тремя переменными. У каждой переменной есть

ошибка (эллипс). Факторы связаны между собой ковариацией, а с переменными - связями зависимости.

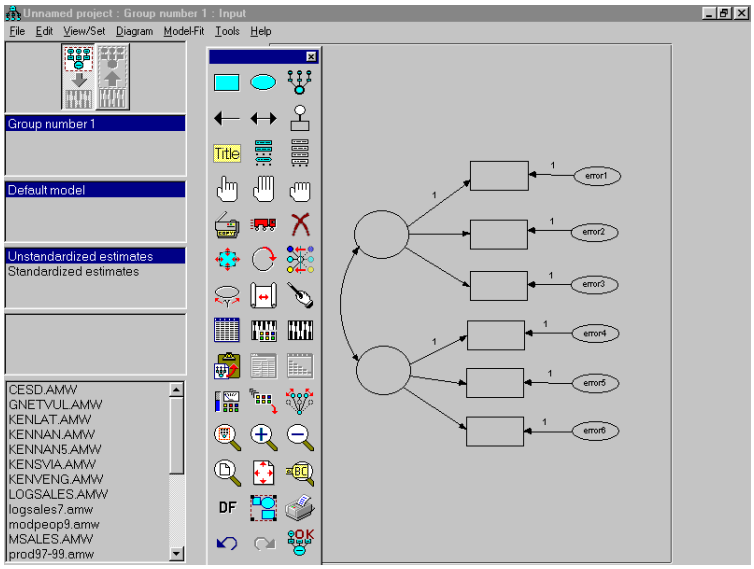



***Рис. 69.***

***Общий вид факторной модели в Amos Graphics***

Чем и как модели, построенные в Amos, отличаются, с одной стороны, от расчетов, выполненных непосредственно путем использования различных статистических процедур (сравнения средних, регрессии, факторного анализа и др.), а с другой – от традиционно довольно широко используемых в социальных науках блок-схем, именуемых многими гуманитариями моделями?

В первом случае основное отличие связано с визуализацией взаимосвязей и зависимостей рассматриваемых переменных. Во втором случае это отличие много существеннее. Дело в том, что в модели, в отличие от простой блок-схемы, все элементы (переменные и их связи) просчитаны и имеют количественные характеристики, а сама модель также имеет набор характеристик, свидетельствующих о ее значимости и надежности.

Проще и лучше начинать строить модели, ставя и решая простые задачи в той последовательности, как они приведены в уже упомянутом руководстве по моделированию в рассматриваемой системе. Не имея такой возможности, мы рассмотрим ниже пример из нашего собственного опыта построения моделей.

В наших работах мы постоянно обращаемся к проверке различных гипотез, связанных с изучением зависимости производства и реализации сельхозпродукции в сельских домохозяйствах от различных факторов. В качестве таких факторов могут выступать: наличие рабочих рук в домохозяйстве, наличие земли и скота, доступность сельскохозяйственной техники, посадочного материала и многие другие. Если мы желаем проверить подобного рода гипотезу, то у нас должна быть база данных с соответствующими переменными-индикаторами.

Начиная такой анализ, мы заранее принимаем решение, какие переменные будут зависимыми, какие независимыми, а какие промежуточными. Рабочий стол рассматриваемой системы моделирования одинаково интерактивен по всей поверхности. Он реагирует не на место переменной, а на конфигурацию ее оболочки и связей. Поэтому модель может иметь различную ориентацию на поверхности экрана (сверху вниз, снизу вверх, справа налево и слева направо). В своих опытах мы используем ориентацию модели слева направо, что по известным причинам культурно-языкового характера наиболее приемлемо в российской традиции (рис. 70).

В связи с этим, приступая к построению модели, мы закладываем слева на рабочем столе независимые, а справа - зависимые переменные. На рис. 70 видно, что в данном случае в модели имеется 5 независимых переменных и три зависимых. При этом результирующей является правая нижняя переменная (продажа сельхозпродукции).

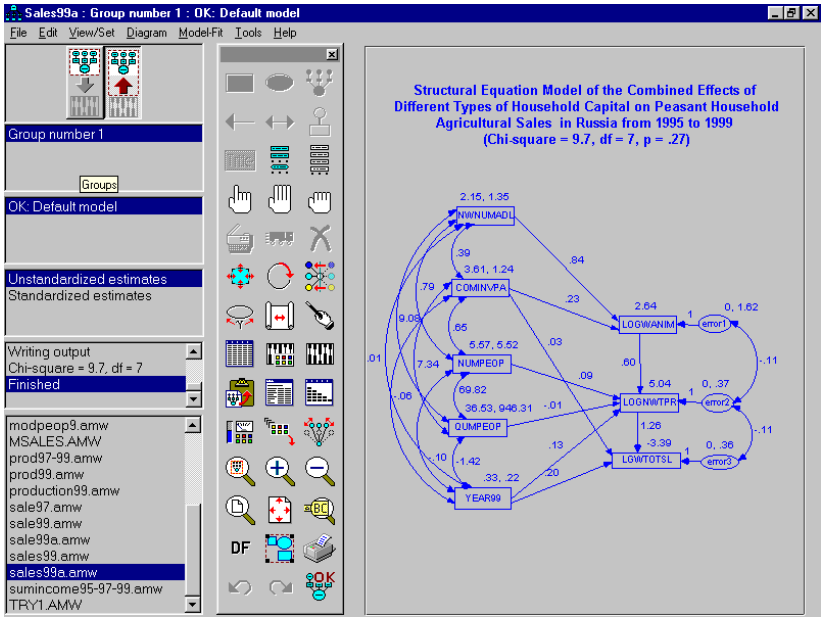



***Рис. 70.***

***Регрессионная модель в Amos Graphics***

Две другие зависимые переменные, а именно: производство сельхозпродукции и наличие скота (средняя и верхняя переменные справа на рис. 70), - фактически являются промежуточными. Они необходимы потому, что, как видно на рис. 70, только две независимые переменные прямо связаны с продажами (вторая слева сверху и последняя слева снизу).

Именно от этих переменных идут напрямую стрелки к продажам. Все остальные независимые переменные оказывают опосредованное влияние на продажу посредством сильного влияния на наличие скота и производство продукции. Производство продукции содержательно должно рассматриваться одновременно как независимая переменная в отношении продаж и как зависимая переменная по отношению к наличию скота, социальным связям, году и др. Все это справедливо и в отношении переменной «наличие скота».

Как строится такая модель? Обычно это делается в три этапа. Два из них: подготовка различных элементов блок-схемы (оболочек и стрелок), а также внесение в них имен переменных - описаны ниже в данном параграфе. Третий - расчет модели, выделен в отдельный параграф в связи с его результирующей значимостью.

*Первый этап построения модели.* Этот этап состоит из нескольких шагов. В качестве исходного шага на рабочий стол из панели инструментов (в соответствии с описанием, приведенным выше в 1.1) переносится контур для переменной.

Следующий (второй) шаг связан с многократным дублированием этого контура (5.2) и зрительно удобным размещением (5.3) на рабочем столе контуров предполагаемых независимых и зависимых переменных. На рис. 70 видно, что как для независимых, так и для зависимых переменных используется один и тот же контур.

Далее (третий шаг), независимые переменные связываются между собой двухсторонними ковариационными стрелками (2.2).

Четвертый шаг - установление связей между независимыми и зависимыми переменными. Эти связи устанавливаются посредством переноса и введения односторонних стрелок в направлении от независимой к зависимой переменной (2.1).

Пятый шаг - установка ошибки. По условиям расчетов каждая зависимая переменная обязательно должна иметь ошибку. Ошибка устанавливается либо с помощью эллипса (1.2) и идущей от него к зависимой переменной односторонней стрелки (2.1), либо, что более просто и правильно, с помощью установки индикатора (2.3).

В результате выполнения описанной выше последовательности действий на рабочем столе появляется общий вид блок-схемы модели, который можно видеть на рис. 71.

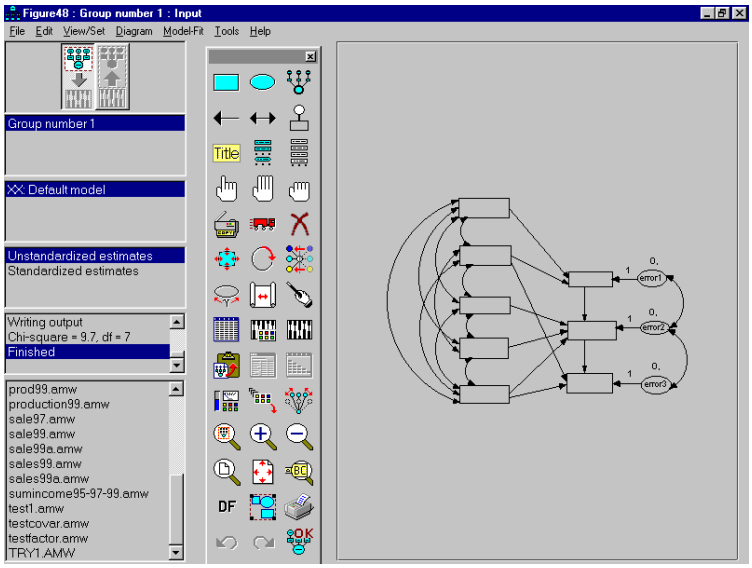



***Рис. 71.***

***Общий вид блок-схемы регрессионной модели на рабочем столе Amos Graphics***

*Второй этап построения модели.* Этот этап связан с внесением имен переменных в заготовленные оболочки блок-схемы модели. Как и предшествующий этап, он предполагает выполнение последовательности нескольких шагов. Основная особенность его состоит в том, что здесь требуется индивидуальная работа с каждой заготовленной оболочкой.

Начиная работу по заполнению контуров переменных их именами, следует сделать двойной клик мышью по оболочке. В результате этого заготовленный для переменной контур выделяется (становится красным) и берется в пунктирное обрамление, а на экране появляется окно для описания объекта - Object Properties (рис. 72).

Окно Object Properties имеет пять закладок (Color, Text, Parameters, Format, Visbility). По умолчанию открывается вкладка Text, которая наиболее необходима в данном случае. Как видно на рис 72, эта закладка имеет четыре поля (Font size, Font stile, Variable name, Variable label). Названия этих полей говорят сами за себя. Центральным здесь и визуально, и по существу является поле Variable name (имя переменной). В данном случае только оно и подлежит обязательному заполнению. В него вписывается имя переменной, как это и показано на рис. 72. Проблема состоит в том, что это имя  нужно либо знать заранее на

память, либо откуда-то списать, например, из того же макета или кодировочной таблицы (приложение 4).

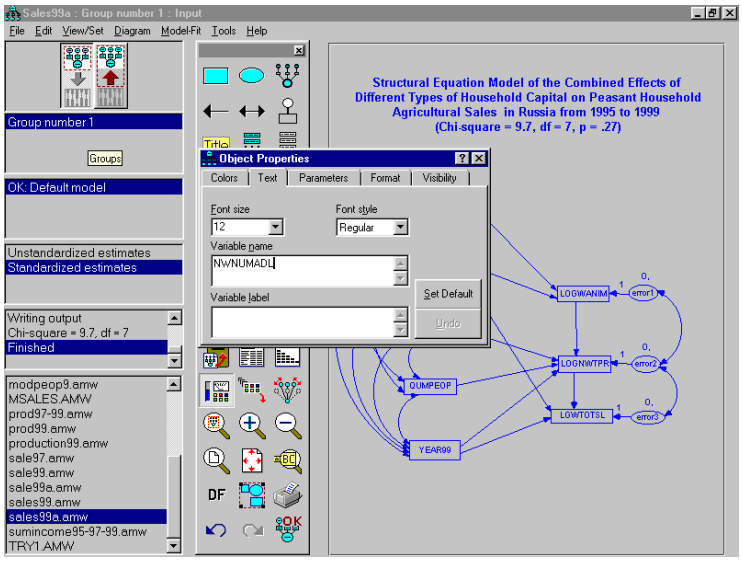



*Рис 72.*

***Окно Object Properties в Amos Graphics***

Вся эта работа может быть куда более приятной и выполняться почти автоматически. Но для этого требуется сразу воспользоваться услугами списка переменных рабочего файла с данными (3.3). При этом имя переменной непосредственно (после его выделения в списке) может сразу перетаскиваться в оболочку (рис. 73).

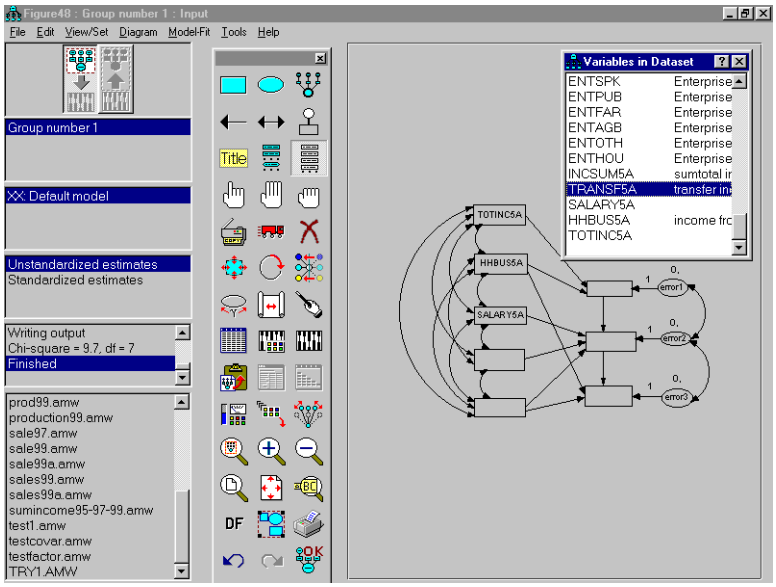



*Рис 73.*

***Окно List variables in data set в Amos Graphics***

Здесь, правда, есть один нюанс. Если файл с данными сделан хорошо и правильно, то в нем наряду с именем переменной имеется и ее описание (label). Оно неизбежно переместится вместе с именем. Между тем, описанию нет и не может быть места в контуре. Поэтому оно займет поверхность рабочего стола за пределами контура переменной. Решение этого вопроса предполагает повторное открытие окна Object Properties и уничтожение описания переменной (но не ее имени, разумеется) во вкладке Text списка Variable label (метка переменной).

Выполняя такую работу, мы как бы уже из контура переменной автоматически вписываем имя переменной и в окно Object Properties. Еще одно достоинство использования иконки 3.3 связано с тем, что она позволяет избежать ошибок, почти неизбежно появляющихся при написании имен переменных вручную. Эти ошибки будут отслежены системой и предъявлены к исправлению, но только на следующем этапе, при попытке выполнить расчет модели, поскольку заполнение контуров именами переменных автоматически формирует список используемых в модели переменных (3.2).

Особого внимания требует формирование описания ошибок. Дело в том, что такая переменная как «error» (ошибка) отсутствует в любом исходном файле с данными. Соответственно, там нет и ее количественных характеристик. В то же время система устроена так, что она не может просчитать модель без числовых характеристик ошибки. Отсюда вывод: их надо вводить.

Если ошибка вводится с помощью иконки 2.3, то все эти вопросы решаются в один момент и автоматически. Если же ошибка вводится стандартным путем с помощью эллипса (1.2) и стрелки (2.1), то после двойного клика по контуру эллипса (рис. 74) возникает необходимость использовать тройной путь:

- окно Object Properties - вкладка Text - список Variable name. Сюда вписывается имя переменной, например, error1. Две ошибки, как и любые другие переменные, не могут иметь одно имя.

- окно Object Properties - вкладка Parameters - поле Mean, в которое следует ввести число 0 (рис. 74).

Далее, после закрытия окна Object Properties с именем переменной, следует двойной клик по стрелке, соединяющей ошибку с зависимой переменной. Эта команда ведет к выделению стрелки и открытию нового окна Object Properties. Вкладка Parameters этого окна содержит поле Regression weight, в которое необходимо вписать число 1 (рис. 75).

*Рис. 74.*

***Вкладка Parameters окна Object Properties для переменной error1 в Amos Graphics***

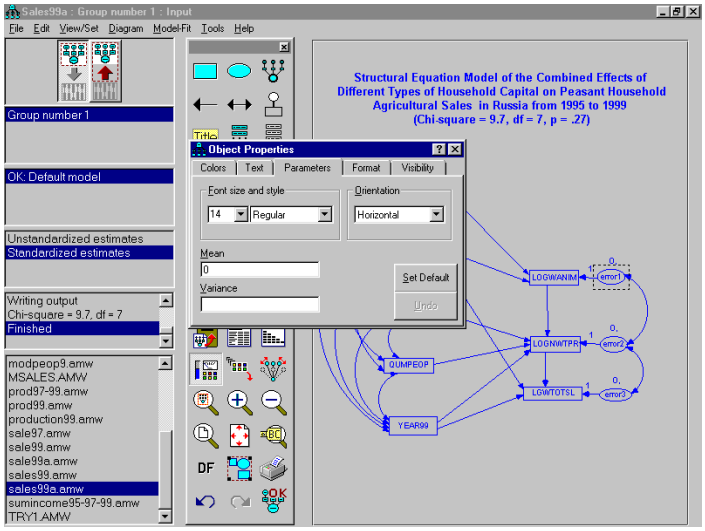
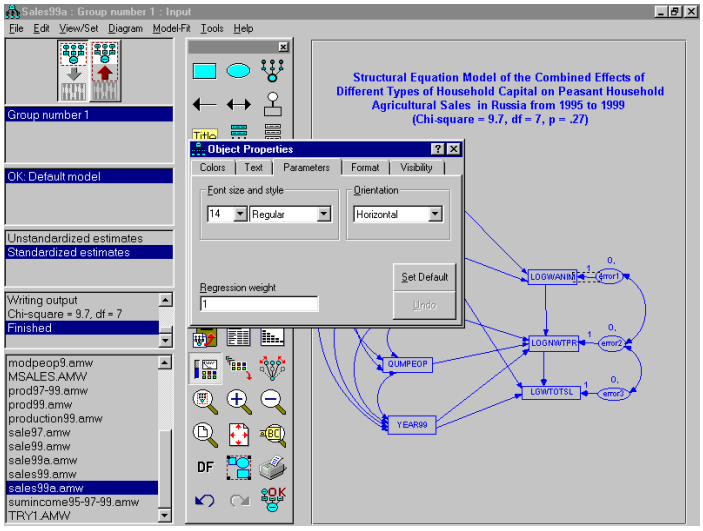



*Рис. 75.*

***Вкладка Parameters окна Object Properties для стрелки между ошибкой (error1) и зависимой переменной в Amos Graphics***




На рис. 74-75 видно, что параметры оценок (числовые характеристики ошибок) устанавливаются еще до выполнения расчетов. Их положение на рабочем столе может корректироваться с помощью иконки 7.1. Еще один важный момент связан с установкой двухсторонних стрелок ковариации на ошибки. Как правило, этого не следует делать сразу при построении блок-схемы, но затем в процессе расчетов и доводки модели до кондиции такой шаг очень часто имеет решающее значение.

В качестве заключительного шага подготовки модели к расчетам следует дать ей заголовок. Это можно сделать, используя соответствующую функциональную команду (иконка 3.1). Полезно и подкорректировать положение модели на рабочем столе, пользуясь функцией компоновки положения и размеров модели на рабочем столе (иконка 12.2). При этом в случае необходимости можно использовать и команды редакционного плана: дублирование, перемещение, удаление объектов и др. (иконки 5.1-7.1). Все это будет уже штрихи и лоск, свидетельствующие о завершении второго этапа подготовки блок-схемы модели к расчетам.

**Напоминание**

***Рабочий стол позволяет строить модели любой ориентации и конфигурации. Но это не значит, что все они будут в равной степени приемлемы для восприятия и публикации.***

## 16.4. Расчет модели - Output

Расчет модели составляет последний - третий этап ее построения. В качестве двух необходимых подготовительных шагов к расчету модели следует задать параметры, которые связаны с анализом ее свойств. И это вполне разумно. Обладая гигантскими возможностями, система должна знать, какие расчеты экспериментатор хотел бы выполнить, и что конкретно он хотел бы получить по итогам расчетов данной модели.

Эта часть работы выполняется с помощью иконки «анализ свойств модели» - analysis properties (8.2), или с помощью команды главного меню View/Set - Analysis Properties, или комбинации клавиш Ctrl+A. Выполнение этой команды одним из трех указанных способов ведет к открытию окна Analysis Properties (анализ свойств).

Окно Analysis Properties состоит из 9 самостоятельных закладок: Permutations, Random #, Title, Output (Formatting), Output, Bootstrap, Estimation, Numerical, Bias. Все эти вкладки имеют определенный

смысл и предназначены для выполнения различных задач, возникающих в ходе построения и расчета моделей.

Тем не менее, две из перечисленных выше девяти закладок - Estimation и Output - имеют решающее значение для выполнения расчетов. Они обязательны к заполнению, а поэтому и приведены на рис. 76-77. Система установлена таким образом, что, открывая окно Analysis Properties, она по умолчанию первой окрывает закладку Title. При желании иметь заголовок не только на рабочем столе, но и в описании модели ее можно заполнить, но это уже сервис, а не необходимость.

Необходимость же выполнения расчетов предполагает в качестве первого шага открытие и заполнение закладки Estimation. Эта закладка имеет массу функций, о чем свидетельствуют ее 11 чек-боксов для установки опций (рис. 76). Для выполнения минимума расчетов *необходимо установить три опции*, как это и показано на рис. 76.

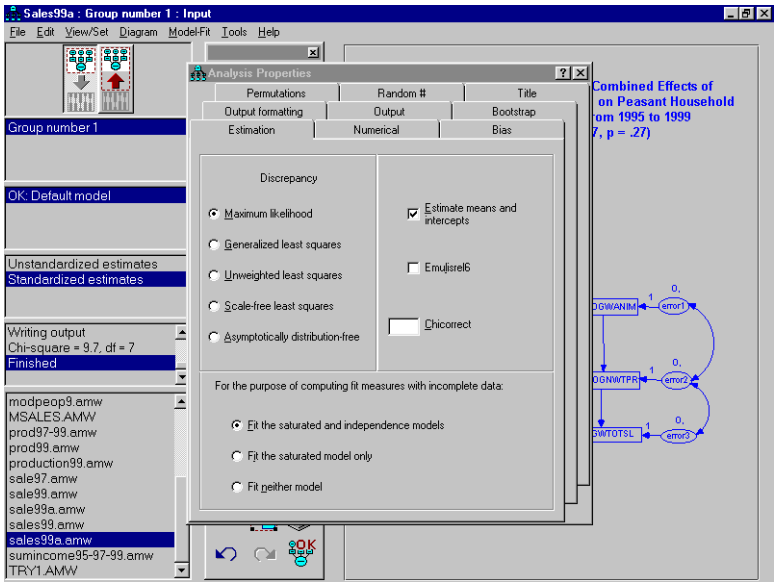



***Рис. 76.***

***Закладка Estimation окна Analysis Properties в Amos Graphics***

*Следующий - необходимый и уже достаточный шаг* - открытие закладки Output и задания в ней минимума параметров, которые будут описаны в окне вывода системы (рис. 77).

Конечно, можно задать и все параметры закладок Estimation и Output. Первая из них, как уже отмечалось выше, имеет 11, а вторая 16 различных опций. Здесь, правда, полезно помнить, что все это не только избыточные расчеты как для системы, так и для самого экспериментатора, но и избыточная информация окна вывода, которое и

без того имеет минимум 5-6 страниц довольно сложного текста. Выполнение двух описанных выше шагов делает блок-схему модели практически готовой к расчетам.

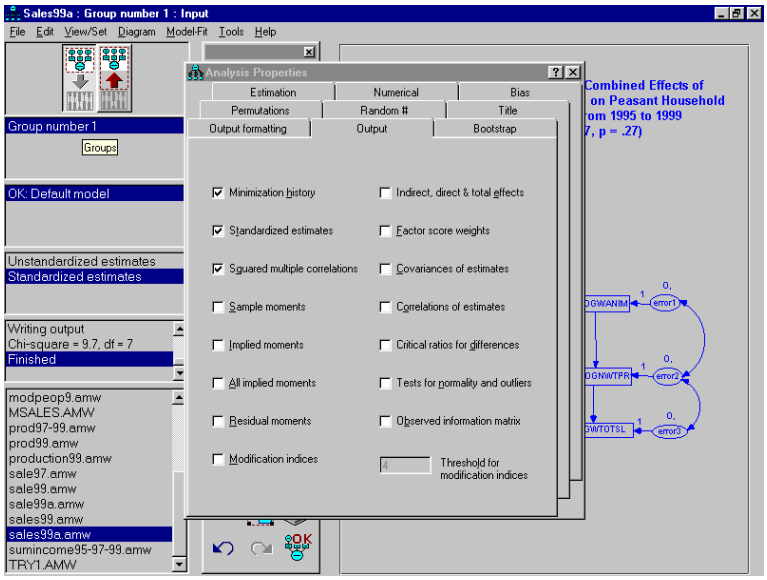



***Рис. 77.***

***Закладка Output окна Analysis Properties в Amos Graphics***

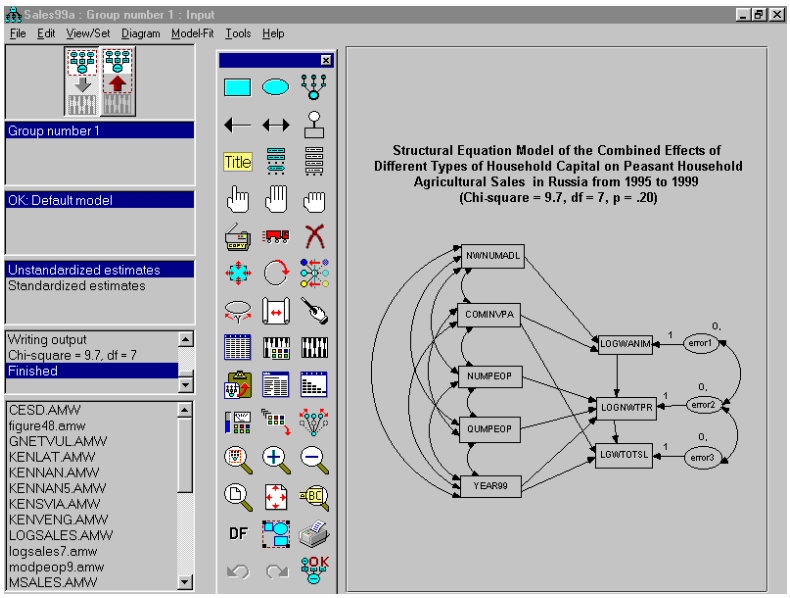
Указание на расчет параметров модели дается с помощью иконки «расчет модели» - calculate estimate (8.3), или с помощью команды главного меню Model-Fit - Calculate Estimate, или посредством комбинации клавиш Ctrl+F9. Выполнение этой команды одним из трех указанных способов ведет, если все сделано правильно, к выполнению расчетов.

О выполнении расчетов свидетельствует движение текста и чисел в пятом блоке панели заголовков. При благополучном завершении расчетов модели в этом блоке появляются ее основные свойства, такие как $\chi^2$ и степень свободы. Если при установке параметров расчетов или еще ранее при построении блок-схемы была допущена ошибка, система прекращает выполнять команду расчета модели и открывает окно, в котором дается информация о причинах такого решения.

При исправлении ошибки и повторном задании команды «расчет модели» система будет выдавать окна на исправление ошибок (если они имеются) в порядке их важности для выполнения расчетов модели. В случае же благополучного завершения расчета модели в пятом блоке панели заголовков устанавливаются, как отмечалось ранее, такие свойства модели как $\chi^2$ и степень свободы, и появляется надпись «Finished» (рис.78).

***Рис. 78.***

***Регрессионная рекурсивная модель в Amos Graphics***




Появление указанных характеристик свидетельствует о завершении расчетов модели и возможности просмотра окна вывода с ее текстовым описанием или электронной таблицы. Сама блок-схема модели при этом внешне не претерпевает каких-либо заметных изменений. Однако теперь на рабочем столе уже находится модель (рис. 78). Эта модель, правда, весьма заметно отличается от модели, которая представлена на рис. 70. Последняя, в отличие от модели на рис. 78, имеет количественные характеристики как для переменных, так и для связей.

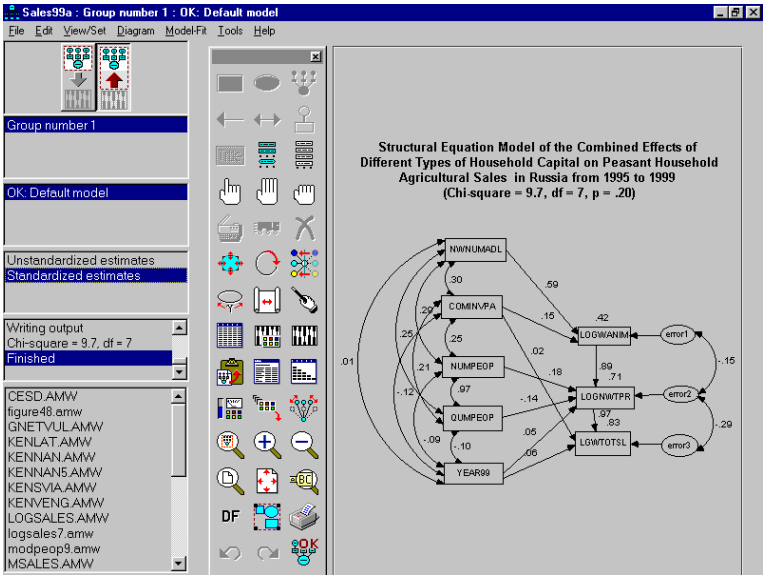
Для того, чтобы модель приобрела такой вид, необходимо изменить режим рабочего стола и перейти из режима Input в режим Output. Делается это с помощью смены кнопок первого верхнего блока панели заголовков (several menu titles).

Такой вид модели, возможно, несколько перегружен числовой информацией. Разработчики системы, понимая это, предлагают пользователю «золотую середину». Модель такого плана и представлена на рис. 79.

Этот вид модели получается с помощью все той же панели заголовков. Только теперь уже, находясь в режиме Output, необходимо перейти от ненормированной (Unstandardized estimate) к нормированной оценке (Standardized estimates). На рис. 79 видно, как в четвертом блоке панели заголовков произошла смена оценок.

***Рис. 79.***

***Регрессионная рекурсивная модель - Standardized estimates в Amos Graphics***




Этим практически и исчерпывается все, что можно видеть на рабочем столе. Если пользователь желает узнать несколько больше о своей модели, то ему необходимо идти в режим просмотра окна вывода.

Окно вывода (output) открывается с помощью иконки «текстовое описание модели» - view text (9.2) или команды главного меню View/Set - text Output.

Электронная таблица (view spreadsheet), в которой также содержится описание модели, открывается с помощью иконки view spreadsheet (9.3) или команды главного меню View/Set - Table Output.

В целях сохранения наглядности и одновременно экономии места, распечатка модели (рис. 78-79) дана в приложении 6 («Окно вывода Amos Graphics»). В этом окне содержится систематическое описание модели. Оно имеет большой объем информации, которую можно структурировать на группы следующим образом:

- имя файла рассчитанной модели;
- дату и время выдачи информации;
- список и число переменных модели с разбивкой на зависимые и независимые наблюдаемые переменные, а также на ненаблюдаемые независимые переменные (ошибки);
- замечание о типе модели. Рассматриваемая модель рекурсивна (recursive). Это значит, что в ней нет обратных зависимостей. Если в модели имеются обратные зависимости, то такая модель нерекурсивна (non-recursive). Такой пример дан на рис. 80;

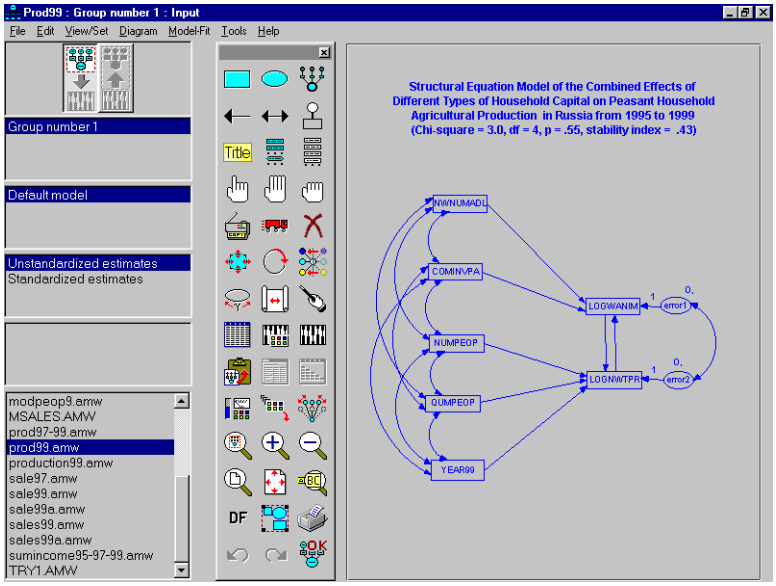



***Рис. 80.***

***Вид регрессионной нерекурсивной модели в Amos Graphics***

- размер выборки - sample size (1266 случаев);
- основные характеристики модели: Chi-square, Degrees of freedom, Probobility level;
- регрессионные веса оцененных переменных (regression weidhts);
- стандартизированные регрессионные веса (SRW);
- оценку среднего (means);
- оценку ковариации для независимых переменных (covariances);
- оценку корреляции для независимых переменных (correlations);
- оценку дисперсии (variances);
- суммарное описание моделей (summary of models) и некоторые другие характеристики модели.

Из приведенных выше различных характеристик модели в первую очередь полезно контролировать:

- *Значения и соотношение трех основных характеристик*: $\chi^2$ (Chi-square), степень свободы (Degrees of freedom) и вероятность (Probobility level). Значения и соотношение этих характеристик проверяются по таблице распределения $\chi^2$ (38, С. 327-329). Идея расчета модели состоит в том, что рассчитываемая модель должна соответствовать некоторой теоретической модели. Это значит, что чем выше уровень вероятности, тем расчетная модель ближе к теоретической. Здесь все выглядит как бы совсем наоборот по сравнению с уровнем значимости в анализе средних. В нашем случае p=.20. Это довольно высокое значение вероятности. Степень свободы в нашей модели df=7.

В таблице распределения $\chi^2$ (38, С. 327-329) соотношению df=7/p=.20 соответствует значение $\chi^2$ = 9.803. В случае нашей модели значение $\chi^2$= 9.746. Другими словами, налицо имеется практически полное совпадение теоретического и расчетного соотношения. Отсюда вывод об устойчивости и надежности рассчитанной модели.

- *Оценки регрессии, ковариации и корреляции.* Здесь есть три колонки: оценка (estimate), стандартная ошибка (S.E.) и критический уровень (C.R.). Оценка тем лучше, чем она ближе к 1. Критический уровень равен отношению оценка/стандартная ошибка. Он значим, если равен или превышает число 1.96 (41, p.74). В принципе, если C.R. не имеет значения, то связь между переменными может быть убрана. Переменные, не имеющие связей или со связями ниже C.R., должны быть убраны из модели. Это основной путь улучшения характеристик модели.

- *Индекс стабильности* (stability index) для нерекурсивных моделей. Он обычно находится в конце окна вывода. Его значения должны находиться в интервале от -1 до +1 (41, p. 183). В этом случае модель стабильна. Если значение индекса стабильности более 1, то либо мало случаев в выборке, либо данные некорректны.

На улучшении соотношения основных характеристик модели и оценок строится вся стратегия расчета модели. Она вряд ли может быть построена за одну или несколько итераций. Расчет любой модели требует массу времени и сил, а главное, проверки множества гипотез, совпадающих или очень удаленных от нулевой гипотезы. Нам, например, так и остается непонятным, каким образом оказалось так, что в панели 1995-1997 гг. аренда земли имела весьма высокое значение (36а, р. 156), а в панели 1995-1999 гг. при реструктурировании той же модели, она оказалась малозначимой и выпала из расчетов (рис. 78-79).

Справедливости ради следует сказать, что в российской социологии проблемы моделирования теоретически обсуждаются уже довольно продолжительное время (42). К сожалению, число прикладников-социологов, которые могли бы использовать предлагаемые в названных работах методы моделирования, было и устойчиво остается меньшим, чем число авторов этих книг. И связано это, конечно, не с тем, что работы подобного рода обладают какими-то фундаментальными недостатками. Напротив, они имеют массу достоинств, которые доступны в основном математикам-социологам. Работы этого плана позволяют им углублять и развивать весьма и весьма увлекательную тематику моделирования.

В противовес такому подходу моделирование в среде Amos открывает большие возможности именно перед социологами-прикладниками. В пользу его распространения работают минимум четыре обстоятельства. Во-первых, распространение компьютерных технологий повсеместно ведет к повышению роли моделирования в познании и проектно-конструкторской деятельности. Во-вторых, визуализация и наглядность компьютерного моделирования делают его доступным для широкого круга пользователей. В-третьих, моделирование позволяет получить новое знание, которое вряд ли может быть получено с помощью других методов. В-четвертых, если в построении модели участвуют несколько экспериментаторов, то дух тотализатора на гонках и одновременно комбинационного расчета в шахматах может вполне овладеть их умами и эмоциями. Победа (хорошие характеристики модели), как правило, приходит неожиданно, а именно путем тестирования очень эффектных и отнюдь не лежащих на поверхности связей.

Против распространения методов моделирования в социальных науках в целом и в том числе в социологии в настоящее время также работают несколько важных обстоятельств. Среди них в первую очередь следует отметить ограниченность владения социологами современными интеллектуальными технологиями. Ситуация здесь усугубляется тем, что, как правило, носители традиционного научного авторитета оказываются менее подготовленными к восприятию нового, чем начинающие исследователи.

Свою долю в консервацию сложившегося положения дел вносят и коммерческие распространители новых технологий. В отсутствии спроса на такие разработки они продолжают заниматься прямым маркетингом, фактически отказываясь от участия в формировании самого спроса. Подтверждением этого печального факта непосредственно в наших условиях служит как отсутствие таких продуктов на открытом рынке, так и слабая заинтересованность дистрибьюторов в организации обучения потенциальных пользователей, которая опять же мотивируется отсутствием спроса. Именно такой ответ был получен одним из авторов этих строк в российском офисе SPSS на вопрос об отсутствии у них учебных курсов по модулю «Amos».

**Правило 42**

***Прежде чем выполнять команду «расчет модели», система всегда требует сохранения блок-схемы, открыв окно с соответствующим запросом-указанием.***

## Задание для самостоятельной работы

1. Что такое модель в Amos Graphics?
2. В чем состоит специфика моделирования в Amos Graphics?
3. Назовите основные составляющие экрана в Amos Graphics.
4. Что такое рабочий стол в Amos Graphics?
5. В чем состоят особенности главного меню в Amos Graphics?
6. Что такое строка состояния в Amos Graphics?
7. Как устроена панель заголовков в Amos Graphics?
8. В чем состоит специфика панели инструментов в Amos Graphics?
9. Какие элементы панели инструментов вы знаете?
10. Как строится модель в Amos Graphics?
11. Какие виды моделей можно строить в в Amos Graphics?
12. Постройте, используя свои данные, корреляционную модель.
13. Опишите окно вывода модели.
14. Какие виды связей фиксирует Amos Graphics?
15. В чем отличие рекурсивной и нерекурсивной моделей?
16. Для чего необходима «ошибка» при построении модели?
17. Что такое блок-схема модели в Amos Graphics?
18. Какие оценки регрессии, корреляции и ковариации выводятся в Amos Graphics?
19. Для каких целей полезно использовать моделирование в социологии?
20. В чем различие моделирования в Amos и Classify, Reduction, Regression?
21. Какие три основные характеристики регрессионной модели в Amos?
22. Как проверяется соотношение трех основных характеристик регрессионной модели: $\chi^2$ (Chi-square), степень свободы (Degrees of freedom) и вероятность (Probobility level) в Amos?
23. Что такое и как проверяется критический уровень (C.R.) в регрессионной модели в Amos?

# Глава 17. Командный язые «СИНТАКСИС»

## 17.1. О синтаксисе

В предыдущих главах описан порядок выполнения расчетов в SPSS с помощью диалоговых окон. Благодаря такой щадящей организации работ, система открывает большие возможности выполнения статистических расчетов и управления данными для каждого начинающего пользователя.

В основе интерфейса диалоговых окон лежит специальный командный язык - «синтаксис» **(Syntax)**. В SPSS этот язык можно использовать в качестве языка программирования команд. С его помощью имеется возможность выполнять как стандартные статистические процедуры, так и создавать новые, которые еще отсутствуют в предлагаемом на данный момент стандартном наборе команд SPSS. Справедливости ради следует отметить, что выбор этих команд растет в каждой последующей версии системы.

Освоение командного языка SPSS требует определенных усилий. Надо быть довольно продвинутым пользователем, чтобы отказаться от весьма удобного выполнения расчетов с помощью диалоговых окон и встать на тернистую тропу задания тех же команд путем их описания и выполнения с помощью командного языка синтаксиса. Тем не менее, избежать полностью соприкосновения с синтаксисом постоянному пользователю довольно сложно.

Необходимость прямых контактов с синтаксисом возникает уже при постановке задач по преобразованию данных (глава 4, § 4.4), а своего максимума она достигает при моделировании (главы 13-16). Поэтому знание основных команд синтаксиса полезно не только для понимания общих принципов работы SPSS, но и для повседневного комфорта при контактах с системой.

Для работы с командным языком в SPSS есть специальное окно. Оно предназначено для ввода и выполнения команд с помощью синтаксиса. Это окно имеет текстовый формат записи команд. Его можно открыть двояким образом.

1. При постановке задачи, связанной с подготовкой нового текстового файла, окно синтаксиса открывается с помощью главного меню. С этой целью необходимо выбрать в меню последовательность команд:

**File**

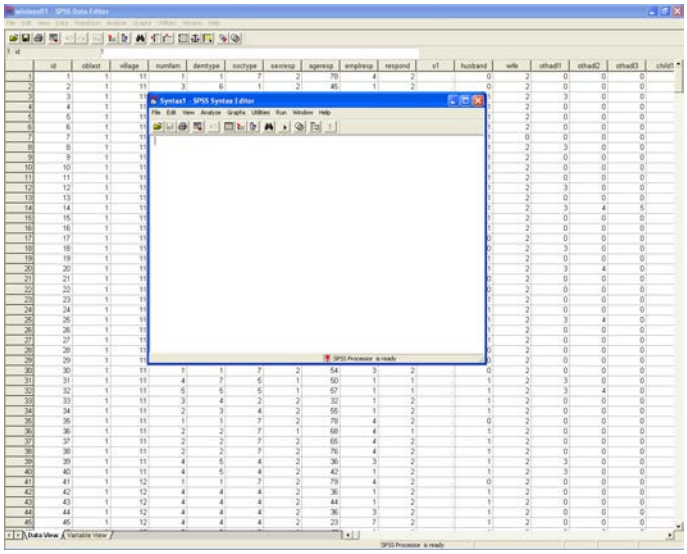
**New**

**Syntax.**

В результате выполнения указанной последовательности команд на фоне окна Newdata или окна с данными рабочего файла появится окно Syntax1 (рис. 81). В этом окне, как в окне обычного текстового редактора, можно набирать любую информацию, но, чтобы использовать его по назначению, необходимо давать описание задач в синтаксисе.

***Рис. 81.***

***Диалоговое окно синтаксиса New-Syntax1***



2. Для того, чтобы открыть уже существующий и сохраненный ранее в SPSS текстовый файл, в главном меню следует выбрать:

**File**

**Open**

**Syntax.**

Откроется диалоговое окно **Open File** со списком ранее сохраненных файлов с расширением **.sps**, которое используется для файлов синтаксиса.

Эти файлы по умолчанию хранятся в директории с данными – Data. Но их нельзя открывать, используя путь: File-Open-Data. Система сразу даст информацию о том, что этот файл не может быть открыт как файл с данными. Поэтому здесь надо быть внимательным и использовать путь: File - Open - Syntax - окно Open File - выделение и открытие необходимого файла.

В результате выполнения этих команд откроется окно Syntax с требуемым файлом (рис. 82).

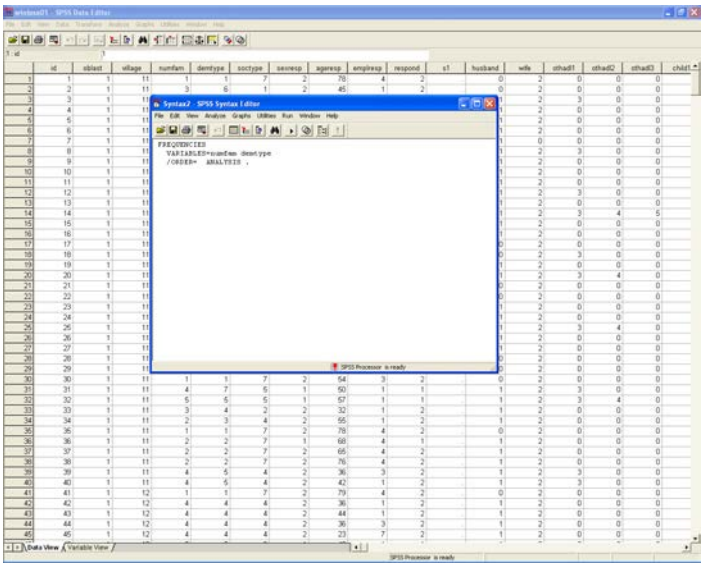


***Рис. 82.***

***Диалоговое окно синтаксиса Open Syntax***

Возникает вопрос: «Откуда берутся такие файлы?» Имеется несколько возможностей их формирования. Такой файл мог быть написан ранее и сохранен непосредственно в SPSS. Его можно создать в другом приложении (например, в Word) и сохранить в текстовом формате. В последующем этот файл можно скопировать и перенести в окно Syntax. В окне синтаксиса можно писать любые тексты (отчеты, доклады), но лучше это делать в Word (45).

Файл, который будет написан на командном языке Syntax, можно получить из главного диалогового окна любой статистической процедуры SPSS. Для этого необходимо в главном и дополнительных окнах статистической процедуры задать все требуемые параметры для ее выполнения (глава 7, § 7.2), а затем, вместо нажатия кнопки OK нажать находящуюся под ней кнопку **Paste** (глава 7, § 7.3).

В результате выполнения этой команды откроется окно Syntax, в котором будет дано описание всех ранее заданных для выполнения процедур расчета. Если в этом окне посмотреть справку посредством команды меню Help, то система откроет содержание справки с отсылкой к руководству по синтаксису (Syntax Guide). В то же время, если в окне синтаксиса версии SPSS 11.5 на панели инструментов нажать кнопку Syntax Help, то система выдаст описание командного синтаксиса для текущего содержания окна.

Например, в случае расчета частот – это будет командный синтаксис частот (Frequencies Command Syntax), а в случае расчета таблиц сопряженности – Crosstabs Command Syntax. При этом курсор должен находиться на теле команды, а не на свободной строчке окна. В противном случае система выдаст сообщение: «Command Not Found. Sorry. Unable to locate this command. Try using the Index to find the command you want».

Здесь следует обратить внимание еще на одно важное обстоятельство. При инсталляции системы SPSS на компьютер Syntax Reference Guide отсутствует среди установок, инсталлируемых по умолчанию. При обращении к нему, в случае его отсутствия, система выдаст рекомендацию, связанную с необходимостью его дополнительной установки.

Устанавливая это приложение, следует использовать путь: Setup program - Custom – Components – установить опцию Syntax Guide. Кстати сказать, это руководство на 1500 страниц в формате **.pdf** файла. Оно требует наличия «Adobe Acrobat» и, в случае отсутствия этой программы, автоматически устанавливает ее.

При работе с окном Syntax можно использовать диалоговое окно Variables (переменные) для копирования имен переменных и их последующей вставки в окно синтаксиса. Для этого необходимо использовать комбинацию команд меню окна Syntax **Utilities** (утилиты) - **Variables** или специальную кнопку Variables, которая находится на его панели инструментов.

Открывающееся подокно позволяет, путем выделения имени переменной и последующего нажатия кнопок Paste-Close, перенести переменю (или их группу, используя клавишу Ctrl) непосредственно на нужное место в окне синтаксиса. Переменная встанет туда, где стоит курсор. Поэтому предварительно следует поставить курсор на требуемое место.

В окне синтаксиса команды набираются и запускаются по одной или группами. Сначала выбирается и выделяется команда, которую требуется запустить. Это можно сделать одним из двух способов: с использованием мыши (методом щелчка и протягивания) и с использованием клавиатуры (Shift + стрелки).

Если нужно запустить только одну команду, то можно поместить курсор в любом месте командной строки. Если же требуется запустить все команды в окне синтаксиса, то лучше использовать команду Select

All (выбрать все) в меню Edit (редактор). Далее, следует набрать комбинацию клавиш Ctrl + R.

В последних версиях SPSS в окне синтаксиса имеется специальная кнопка Run (выполнить). Эта кнопка дублирует одноименную команду Run в меню окна Syntax. Используя путь **Run-All**, можно сразу выделить и выполнить все расчеты.

При этом курсор должен находиться перед началом текста первой команды или после точки в конце команды. Только в этом случае система выделит весь требуемый фрагмент. Здесь надо быть внимательным и осторожным. Система не может выполнить команду, если курсор стоит в «чистом поле» окна синтаксиса.

В ранних версиях системы список команд и их формат можно найти, используя раздел Utilities - Command Index окна синтаксиса. В версии SPSS 11.5 это можно делать, либо используя комбинацию команд Help (справка) - Syntax Guide, либо посредством кнопки Syntax Help на панели инструментов.

**Дружеский совет**

***Лучший способ овладеть командным языком Syntax - регулярное использование команд Paste-Run при выполнении статистических процедур и кнопки Syntax Help на панели инструментов окна синтаксиса.***

**Правило 43**

***Комбинация команд Paste-Run (в окне синтаксиса) эквивалентна команде OK в главном диалоговом окне любой процедуры.***

## 17.2. Формат записи в синтаксисе

Любой пользователь согласится с тем, что хорошо было бы сесть и написать в окне синтаксиса команды для выполнения каких-либо расчетов самостоятельно. Но каждый разумный пользователь понимает, что для этого необходимо знать как сами команды, так и порядок их написания. С учетом отсутствия у пользователя знаний командного языка SPSS, скопируем несколько записей разных команд из окон

синтаксиса в файл, который находится в текстовом процессоре Word. С этой целью ниже и сформировано обрамление 47.

В терминах выполнения статистических расчетов с помощью диалоговых окон в этом обрамлении объединены четыре отдельных процедуры. А именно: расчет частот, построение таблицы сопряженности, расчет средних и коэффициента корреляции Пирсона. Разумеется, для выполнения комплекса таких расчетов требуется время и навигация по всем четырем процедурам, пошаговый доступ к которым открывается посредством использования команды главного меню Analyze.

Как видно из обрамления 47, каждая процедура начинается с новой строки командой (FREQUENCIES, CROSSTABS, MEANS, CORRELATIONS), занимающей первую строку задаваемого выражения, и заканчивается точкой.

**Обрамление 47. Формат команд в Syntax**

```
FREQUENCIES
  VARIABLES=hage9 wage9 oage19 oage29 oage39 cage19 cage29 cage39
  /STATISTICS=VARIANCE MINIMUM MAXIMUM MEAN MEDIAN MODE
  /HISTOGRAM  NORMAL
  /ORDER=  ANALYSIS .
CROSSTABS
  /TABLES=demtype  BY village
  /FORMAT= AVALUE TABLES
  /CELLS= COUNT COLUMN TOTAL .
MEANS
  TABLES=sumtotal sumtota7 sumtota9 sumtota3  BY village3
  /CELLS MEAN COUNT STDDEV  .
CORRELATIONS
  /VARIABLES=sumtota3 meatpro3 meatsol3 milkpro3 milksol3 eggprod3 eggsold3 potprod3
potsold3 vegprod3 vegsold3
  /PRINT=TWOTAIL NOSIG
   /MISSING=PAIRWISE .
```

Эта команда (command) может быть названа **основной** или собственно **командой.** Она управляет всем ходом расчетов в SPSS. Поэтому в одном выражении может быть только одна основная команда. Синтаксис имеет огромное число таких команд.

По характеру их использования все основные команды делятся на несколько групп: сервисные - Utility commands, определители файлов - File definition commands, ввода программ - Input program commands, преобразований - Transformation commands, ограничения преобразований -Restricted commands и процедуры - Procedures (21, pp. 840-842). В табл. 34 дан перечень части таких команд.

**Таблица 34. Группировка некоторых основных команд синтаксиса**

| Группа | Команды |
|---|---|
| Сервисные | COMMENT, DISPLAY, DOCUMENT, EDIT, END, DATA, ERASE, FINISH, HELP, INCLUDE, N OF CASES, SUBTITLE, TITLE |
| Определители файлов | ADD FILES, DATA LIST, FILE TYPE, GET, HOST, IMPORT, INPUT PROGRAM, UPDATE |
| Ввода программ | END CASE, END FILE, POINT, RECORD TYPE |
| Преобразований | ADD VALUE LABELS, COMPUTE, COUNT, IF, MISSING VALUES, STRING, WEIGHT, WRITE |
| Ограничения преобразований | FILTER, SAMPLE, SELECT IF, TEMPORARY |
| Процедуры | BEGIN DATA, EXECUTE, EXPORT, GRAPH, SAVE |

*Источник:* 21, pp. 840-842.

Из приведенного в табл. 34 перечня команд языка программирования в SPSS видно, что они отражают структуру естественного (английского) языка. Основу этих команд составляют существительные (комментарий, документ, данные) и глаголы (удалить, включить, выполнить), хотя встречаются и союзы (если). Поэтому элементарные знания английского языка могут оказать содействие при освоении синтаксиса.

Следующая строка во всех выражениях начинается со **вспомогательной команды** или **подкоманды** (subcommand). В нашем примере в двух случаях использована вспомогательная команда «VARIABLES» и в двух случаях - «TABLES». В каждом выражении может быть несколько вспомогательных команд. В первом выражении в обрамлении 47 четыре вспомогательные команды: – «VARIABLES», «STATISTICS», «HISTOGRAM» и «ORDER». Они отделяются друг от друга косой чертой, которая может отсутствовать только перед первой вспомогательной командой, поскольку система «хорошо знает» различие между основной и вспомогательной командой.

В каждой строке выражения все, что находится между вспомогательной командой и следующей косой чертой, представляет собой описание содержания или **спецификацию** (specification) вспомогательной команды. Как видно из выражений, записанных в обрамлении 47, в качестве спецификаций могут выступать: знаки равенства (арифметическая операция); имена переменных [возраст мужа. (hage9), воз

раст жены (wage9)]; ключевые слова (keywords) – VARIANCE, MINIMUM, MAXIMUM; функции (function) – BY, IF, WITH и др. Интересно, что такое замечательное выражение как Count (счет), в зависимости от ситуации может быть и командой, и ключевым словом, и функцией; то же самое следует сказать о выражениях File (файл), If (если) и др.

В результате использования перечисленных выше элементов командного языка и установленного порядка их записи, в обрамлении 47 записано четыре ранее упомянутые команды. Все они имеют и словесное содержание. Например, команда, записанная в первом выражении, имеет следующее вербальное описание (в этом описании - абзац и строка команды эквивалентны):

«Рассчитать частоты для переменных:

возраст мужа, возраст жены, возраст первого взрослого члена семьи, возраст второго взрослого члена семьи, возраст третьего взрослого члена семьи, возраст первого ребенка, возраст второго ребенка, возраст третьего ребенка.

Для каждой переменной рассчитать описательные статистики: разброс, минимум, максимум, среднее значение, медиану, моду.

Для каждой переменной построить гистограмму с нормальной кривой.

Выполнить это задание».

Вербальное описание трех других выражений в обрамлении 47, каждый желающий может выполнить самостоятельно. Если теперь скопировать содержание обрамления 47 и перенести его в окно синтаксиса, то система, после его выделения и нажатия кнопки Run, выполнит все расчеты в один прием. Такой подход может оказаться эффективным, если пользователь регулярно выполняет один и тот же или очень близкий комплекс расчетов. В этом случае формирование блоков команд и использование их в качестве шаблонов может даже для начинающего пользователя сохранить массу времени и сил. Напомним, что последовательность команд, набранная в окне Syntax, может быть сохранена в файле с расширением .sps и, благодаря этому, стать доступной для многократного использования.

Эффективность такого подхода различна и для отдельных процедур. Скажем, делать шаблон для частотного анализа или описательных статистик имеет смысл только в учебных целях. Связано это с тем, что диалоговые окна этих процедур позволяют в один «присест» выполнить огромный объем расчетов для большого числа переменных.

В то же время однофакторный дисперсионный анализ (глава 10, § 10.4) допускает одновременное исследование нескольких переменных только по одному фактору. Это ограничение можно обойти, используя в окне синтаксиса шаблон, приведенный в обрамлении 48.

**Обрамление 48. Запись команд в окне синтаксиса для одновременного расчета в ANOVA по разным группам зависимых переменных и разным факторам**

```
ONEWAY
  ageresp numfam BY village
  /MISSING ANALYSIS .
ONEWAY
  sumtotal  BY demtype
  /MISSING ANALYSIS .
ONEWAY
  hage wage oage1 oage2 oage3 cage1 cage2 cage3  BY sexresp
  /MISSING ANALYSIS .
```

Такая запись команд в окне синтаксиса позволяет выполнить расчет в ANOVA одновременно для возраста респондента и размера семьи по селу, для совокупного месячного дохода семьи по ее демографическому типу и для возраста всех членов семьи по полу респондента. Используя диалоговое окно соответствующей процедуры, пришлось бы выполнить минимум в три раза больше расчетов. Такая ситуация характерна для линейной регрессии, факторного анализ и ряда других процедур. Во всех этих случаях при необходимости выполнения большого числа расчетов могут использоваться сходные решения.

Еще одно важное направление использования окна синтаксиса связано с созданием новых переменных. В главе 4 (§ 4.5) приведен пример расчета новой переменной, фиксирующей три группы человеческого капитала в семье (низкий, средний, высокий) в выборке 1997 г. При этом использовались три процедуры: Count - для создания девяти промежуточных переменных, Compute - для создания одной суммарной переменной и Recode – для создания итоговой переменной, которая делит человеческий капитал в семье на три группы.

Сама по себе задача, связанная с использованием нескольких процедур, должна наводить на мысль о необходимости использования окна синтаксиса. Пример ее решения для определения уровня человеческого капитала в семье в выборке 2003 г. с помощью командного языка дан в обрамлении 49.

**Обрамление 49. Расчет трех групп человеческого капитала**

```
COUNT
  old037 = cage13 cage23 cage33  (1 thru 7)
  /old0311 = cage13 cage23 cage33  (8 thru 11)
  /old0314 = cage13 cage23 cage33  (12 thru 14)
  /old0316 = cage13 cage23 cage33  (15 thru 16)
  /old0365 = hage3 wage3 oage13 oage23 oage33 cage13 cage23 cage33  (17 thru 65)
  /old0370 = hage3 wage3 oage13 oage23 oage33 (66 thru 70)
  /old0374 = hage3 wage3 oage13 oage23 oage33 (71 thru 74)
  /old0379 = hage3 wage3 oage13 oage23 oage33 (75 thru 79)
  /old0380 = hage3 wage3 oage13 oage23 oage33 (80 thru 97)  .
EXECUTE .
COMPUTE numad03 = SUM((old037 * 0),(old0311 * 0.25),(old0314 * 0.5),(old0316
 * 0.75),(old0365 * 1),(old0370 * 0.75),(old0374 * 0.5),(old0379 * 0.25)
 ,(old0380 * 0)) .
EXECUTE .
RECODE
  numad03
  (0 thru 1=1)  (1.1 thru 2.25=2)  (2.26 thru Highest=3)  INTO  humcap3 .
    EXECUTE .
```

Если учесть, что в случае нашего пятиволнового панельного исследования человеческий капитал в семье надо было рассчитать для каждого года, то наличие такого шаблона можно рассматривать как весьма эффективное средство экономии сил и времени.

В специальной литературе при написании и редактировании командного синтаксиса рекомендуют выполнять следующие правила:

· Каждая команда должна начинаться с новой строки и заканчиваться точкой.

· Вспомогательные команды отделяются друг от друга при помощи косой черты - обратного слежа (/).

· Текст, взятый в одинарные кавычки (используемый для идентификации меток), должен находиться в одной строке.

· Строка с программным синтаксисом не должна превышать 80 знаков.

· Для многих команд в синтаксисе допускается использование сокращенных форм и аббревиатур. Например, Aggregate=AGG, COMPUTE=COMP, ADD VALUE LABLES=ADD VAL LAB, EXAMINE VARIABLES=EXA VAR и др. При первом знакомстве это затрудняет восприятие командного языка.

· В качестве десятичного разделительного знака в спецификациях должна применяться точка (.), независимо от установок операционной системы Windows.

· Спецификации (ключевые слова и переменные) отделяются друг от друга пробелом.

· Ввод пробела или переход на новую строку разрешается в точке применения одиночного пробела, то есть перед и после косой черты, скобок, арифметических операторов или между именами переменных (21, pp. 11-19; 16, С. 559; 43, pp. 4-10).

Повторяясь, отметим, что работа с командами искусственного языка программирования - синтаксиса, во-первых, предполагает знание корней и словосочетаний соответствующих выражений естественного английского языка, а во-вторых, владения навыками записи и общими принципами использования команд искусственных языков программирования. Конечно, все это затруднительно для начинающих пользователей.

Вместе с тем, всегда полезно помнить, что, работая с главным меню, панелью инструментов SPSS и диалоговыми окнами различных процедур, вы постоянно находитесь в среде командного языка. COMPUTE, COUNT, EDIT, FILTER, HELP, RECODE и др. - все это команды синтаксиса. И если, выполняя эти команды, хотя бы изредка (нажав, как было описано ранее, кнопку Paste) смотреть их написание в окне Syntax, то можно весьма быстро перестать бояться формы их записи и начать фиксировать определенные повторения и последовательности их написания и структурной организации.

При регулярной работе в системе эта наблюдательность начнет окупаться очень быстро. В то же время для случайного пользователя вся премудрость командного языка может оставаться тайной за семью печатями, которую он сам для себя создает.

**Дружеский совет**

***Окно Syntax полезно использовать в следующих случаях:***
***- когда изменяется первичная информация в файле данных и одновременно требуется выполнение того же набора команд;***
***- когда необходимо выполнять одни и те же команды с разными переменными (заменить имя в команде проще, чем выполнить расчет посредством диалоговых окон);***
***- если невозможно выполнить команду через меню из-за большого размера файла.***

## 17.3. Преобразование файлов с помощью синтаксиса

Ниже приведен пример преобразования файла с использованием окна Syntax. У нас эта задача впервые возникла в 1997г. После трех лет панельного исследования (приложения 2-3) был создан файл «paneldata», в котором 463 наблюдения были записаны по строкам, а в столбцах записывались 542 * 3=1626 переменных.

Для решения задач, связанных с анализом изменения значений каждой переменной по годам, исходный файл данных требовалось преобразовать в файл, в котором каждая переменная расписана по трем годам последовательно. Другими словами, новый файл должен иметь вид, в котором по строкам необходимо расписать 463 * 3=1389 наблюдений, а по столбцам 542 переменные.

Выбор возможностей для решения такой задачи весьма ограничен. Можно заново набить данные в требуемом формате, что, естественно, требует огромных затрат труда и времени.

В принципе есть возможность переформатировать файл вручную, т.е. путем выделения и переноса в новый файл переменных (из колонки старого файла в строку нового файла) и записи имен 542 новых переменных. Такой путь возможен, но довольно сложен в реализации, т.к. при переносе возникает масса проблем.

И, наконец, последняя возможность - воспользоваться окном Syntax с целью преобразования исходного файла данных «paneldata» в новый файл «pooleddata». Последовательность выполнения различных команд синтаксиса по преобразованию файла в SPSS 11.5 несколько отличается от предшествующих версий (3, С. 67-73) и описана ниже.

В качестве примера приведены очень удобные для контроля и наглядности преобразования с возрастом опрашиваемого. Это связано с тем, что каждый последующий год он должен быть на единицу больше.

Прежде всего, открываем исходный файл paneldata_95-96-97. Далее, выполняется последовательность команд:

**File**
**New**
**Syntax.**

В левом верхнем углу открывшегося окна Syntax делается следующая запись:

**write outfile = nageresp.dat**
**/ageresp**
**/ageresp6**
**/ageresp7.**

Откуда взялось это магическое выражение «write outfile» и что оно означает? Write outfile (написать файл, доступный при работе с различными приложениями SPSS) - основная команда синтаксиса. Она относится к группе преобразующих команд (Transformation commands).

Эти команды надо либо знать, либо брать их, используя путь Help - Syntax Guide. Еще одна дополнительная возможность - взять их из специального руководства SPSS по Syntax, что мы и сделали в данном случае (21, p. 826; 43, p.1319). На худой конец можно заглянуть в приведенную ранее табл. 34, которая, безусловно, не даст ответы на все вопросы, но, возможно, будет способствовать их правильному решению.

Имя новой переменной «**nageresp**». Первая буква (n) и означает, что она новая (new). Имя новой переменной, хотя и сходно со старым (ageresp), но оно действительно новое, уникальное и содержит, как и полагается, не более 8 знаков (глава 2, параграф 2.4). Словесно оно может быть описано как «новая переменная для возраста опрашиваемого». Имя этой новой переменной записывается со специальным расширением **.dat**, запись которого обязательна. Указанное расширение означает, что преобразованию подлежат данные рабочего файла.

Между основной командой и именем новой переменной стоит арифметический знак равенства, позволяющий системе идентифицировать вновь создаваемый файл.

Таким образом, в нашем примере первая строка записи в окне синтаксиса «write outfile = nageresp.dat» означает, что ниже будет описан порядок формирования файла с именем nageresp и расширением .dat.

Далее идет задание на формирование необходимого файла. В столбец, каждая строка которого начинается с обратного слежа, записаны три имени переменных, взятых по годам из базового файла:

**/ageresp –** возраст респондента в 1995 г.

**/ageresp6 –** возраст респондента в 1996 г.

**/ageresp7. –** возраст респондента в 1997 г. Имя последней переменной можно найти в приложении 4.

Наличие косой черты перед именем каждой переменной означает, что далее идет запись вспомогательной команды. В данном случае

имя самой вспомогательной команды система воспринимает по умолчанию. Но это правило действует далеко не для всех вспомогательных команд. Запись в столбце заканчивается точкой.

В качестве следующего шага в меню синтаксиса выполняется команда Run-All, которая может быть продублирована нажатием кнопки Run current command на панели инструментов окна синтаксиса или комбинацией клавиш Ctrl + R. Если в записи была небольшая ошибка, то команда не выполнится. Если в записи допущена серьезная ошибка, то откроется окно Output с ее описанием. Если в конце будет отсутствовать точка, то откроется маленькое окно с указанием на необходимость поставить таковую.

**Внимание**

***Запись в столбце обязательно заканчивается точкой, а курсор стоит после нее. Более того, для решения нашей задачи запись должна вестись только по столбцу. Запись по строке ведет к формированию файла с другой структурой данных.***

В случае же, когда все правильно, команда начнет выполняться, а в нижней строке окна редактора данных справа от записи SPSS Processor is ready появляется дополнительная запись **Transformations pending**. В этом случае надо идти в главное меню SPSS и выполнять последовательность команд:

**Transform**

**Run Pending Transforms.**

В предшествующих версиях системы эта команда называлась - Run transform planted. О выполнении последней команды свидетельствует прокрутка всего массива в нижней строке главного окна с файлом paneldata. Далее, опять выполняются команды главного меню:

**File**

**Read Text Data.**

В предшествующих версиях системы эта команда называлась - Read ASCII. Сам ход преобразования данных в версии SPSS 11.5 уже существенно отличается от хода преобразования в предшествующих версиях, описанного нами ранее (3, С. 67-73).

В результате выполнения команды Read Text Data открывается окно Open File с готовой к открытию директорией Data. В нижней строке этого окна - Files of type стоит заготовка на открытие текстового файла

(Text .txt), которая должна быть заменена на All files или, что еще лучше в данном случае, на файлы типа Data (.dat). Тогда в списке файлов, хранящихся в директории Data, в окне Open File окажутся только файлы с требуемым расширением .dat. Среди них обязательно будет вновь созданный файл «nageresp.dat».

Следующим шагом выделяем имя этого файла и открываем его (кнопка Open). При этом открывается диалоговое окно – **Text Import Wizard - Step 1 of 6** (рис. 83). Это окно предполагает выполнение последовательности команд, которая состоит из 6-и шагов. Указанная последовательность команд реализуется посредством открытия доступа к выключателю «Next», приглашающему совершить следующий шаг. Само это окно имеет, как видно на рис. 83, довольно сложную структуру.

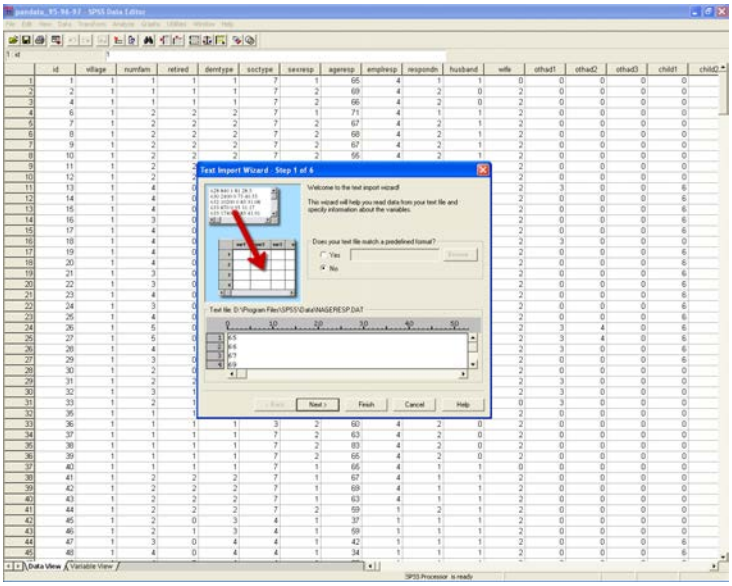


***Рис. 83.***

***Диалоговое окно Text Import Wizard - Step 1 of 6***

Тем не менее, уже в нижней части этого окна можно видеть, как будет преобразован файл с данными. С этим и связана основная функция 1-го шага – дать возможность предварительного просмотра преобразуемых данных. Используя выключатель Next, последовательно выполняем все следующие шаги. При этом каждый новый шаг открывает новое диалоговое окно, в котором все требуемые установки уже заданы.

Окно 2-го шага содержит информацию о переменных. Окно 3-го шага содержит информацию о наблюдениях (случаях), вошедших в

преобразуемый файл. Окно 4-го шага содержит информацию о том, как система будет считывать данные, преобразованные в новый файл. Окно 5-го шага содержит информацию об имени каждой переменной, вошедшей в новый файл, и формате ее данных. Окно 6-го шага позволяет сохранить всю информацию о преобразовании в отдельном файле и затем при необходимости воспользоваться ее.

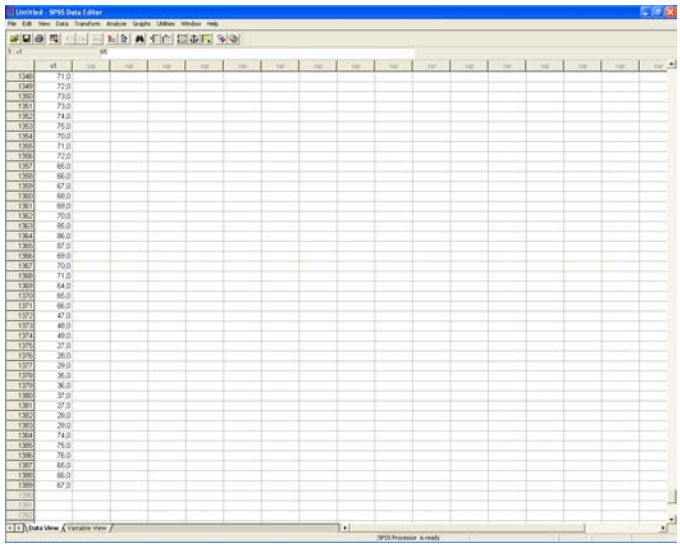
Так сохраняется описание преобразований в командном языке, но не сам результат преобразований – файл с новым форматом данных. Для вывода этого файла необходимо после прохождения всех шести шагов выполнить команду «Finish». Кнопка с таким именем находится справа от выключателя Next.

Если, еще находясь на 1-ом шаге, выполнить команду Finish, то система сразу перейдет с 1-го на 6-й шаг. Повторное нажатие выключателя Finish выбросит сначала указание на необходимость сохранить данные в текущем рабочем файле (в нашем случае – это paneldata_95-96-97), а затем откроет новый рабочий файл с одной переменой v1 (рис. 84).

В отличие от исходного файла, в котором каждая из трех преобразованных переменных содержала 463 случая, в новом файле имеется одна переменная, содержащая 1389 случаев (рис. 84). При этом значение возраста в каждой тройке случаев отличается на единицу (первое - минимальное, а последнее – максимальное).

***Рис. 84.***

***Фрагмент окна рабочего файла с новой переменной v1***



В качестве следующего промежуточного шага полезно в Variable View переименовать безымянную переменную «v1» в более подходящую для данного случая переменную с именем «nageresp». Если после этого, используя путь: File – Save As, сохранить этот, пока еще не названный файл, под именем «pooleddata_95-96-97», то тем самым будет положено начало формированию искомого файла с преобразованными данными.

Причем этот файл, в отличие от всех переходных файлов, будет иметь расширение **.sav**, которое принято в SPSS для файлов с данными. Такой результат означает завершение поставленной ранее задачи по преобразованию данных, связанных с возрастом респондента и созданию нового файла pooleddata_95-96-97.

Конечно, чтобы этот файл стал жизнеспособным, необходимо выполнить массу описанных выше преобразований переменных. Здесь уместно подчеркнуть, что каждая новая переменная по результатам преобразований будет появляться в новом безымянном файле с именем v1. Ее перенос в формируемый файл pooleddata_95-96-97 предполагает установку курсора на имени переменной с последующим выделением всего столбца.

Следующий шаг - переход в главное меню и выполнение последовательности команд: Edit –Copy. Затем открываем новый файл с именем pooleddata_95-96-97, идем в конец файла, помечаем новую колонку в 1389 случаев и выполняем последовательность команд главного меню: Edit – Paste. Далее, выполняем команду Save и можем считать завершенным еще один важный этап по преобразованию данных и формированию основ нового файла.

В отличие от исходного файла paneldata_95-96-97, новый файл pooldata_95-96-97 обладает рядом совершенно удивительных особенностей, позволяющих фиксировать динамику изменений различных признаков и строить замечательные модели.

По большому счету результатом выполнения всей последовательности преобразований должен стать новый файл данных, в котором 1389 случаев и 542 новые переменные. Постановка такой задачи уже сама по себе рассчитана на большого любителя преобразований данных. Ее последовательная реализация весьма близка к поведению начинающего исследователя при построении таблиц сопряженности. Такой пользователь, как уже отмечалось ранее, старается пересечь все со всем, а уже потом разбираться с тем, что получилось (глава 9, § 9.1).

Основная проблема здесь состоит в том, что далеко не все переменные имеют шанс быть значимыми при построении модели. Хотя на данный момент это обстоятельство и не очевидно по отношению к каждой конкретной переменной. Нельзя забывать, что опыт моделирования социологической информации чрезвычайно мал и ограничен.

В этих условиях делать какие-либо категорические утверждения довольно проблематично. В пользу такого вывода свидетельствуют и, приведенные в предыдущих главах этого раздела, примеры моделирования данных. Например, расчет факторной (глава 14, § 14,3) и еще в большей мере кластерной (глава 15, § 15.2), и дискриминантной (глава 15, § 15.3) моделей. Фиксируемые в этих моделях отношения между переменными далеко не очевидны. В то же время примеры регрессионных моделей, которые приведены в главах 13, § 13.3 и 16, §16.3-16.4, видимо, могут рассматриваться как весьма показательные и наглядные образцы моделирования данных социологических исследований.

В любом случае для целей нашего анализа мы трансформировали примерно 20% исходного файла, создав, таким образом, новый файл, содержащий 1389 случаев и около 100 новых переменных. В значительной мере на основе анализа данных этого файла и была написана книга: «Household Capital and the Agrarian Problem in Russia» (36a).

Сегодня уже не очень трудно представить себе какая огромная часть полезной и нужной информации просто не была востребована. Это обстоятельство в еще большей мере типично для всей массы социологических исследований. Оно существенно обедняет как научный поиск, так и использование данных социологических исследований в практике управления и повышении уровня информированности о социальных процессах, происходящих в обществе.

**Лозунг дня**

***Социологи, повышайте КПД своих исследований и разработок!***

В заключение уместно еще раз отметить, что владение основами командного языка Syntax открывает перед пользователем огромные возможности и делает более осмысленной саму работу в SPSS.

**Напоминание**

***В практике стандартных расчетов окно Syntax фактически не используется, но периодически посматривать в него весьма и весьма полезно для понимания особенностей и специфики выполнения команд в SPSS.***

*Задание для самостоятельной работы*

1. Как называется командный язык в SPSS?
2. Как в SPSS можно получить информацию о командном языке?
3. Что такое окно Syntax?
4. Как можно открыть окно Syntax?
5. Как можно увидеть текущую запись командного языка в главном диалоговом окне статистической процедуры?
5. Какие основные элементы имеет Syntax?
6. Что такое основная команда в синтаксисе?
7. Приведите примеры основных команд.
8. Что такое вспомогательная команда в синтаксисе?
9. Приведите примеры вспомогательных команд.
10. Что такое спецификации в синтаксисе?
11. Приведите примеры спецификаций.
12. В каких случаях полезно использовать окно синтаксиса?
13. Чем в SPSS файл с расширением .dat отличается от файла с расширением .sav?
14. Опишите диалоговое окно Text Import Wizard.
15. Как выделяется вспомогательная команда при записи в окне синтаксиса?
16. Чем заканчивается запись команды в синтаксисе?
17. С чего начинается запись команды в синтаксисе?
18. Какие правила командного языка вы знаете?
19. Где должен находиться курсор при выполнении команды в синтаксисе?
20. Как открываются ранее сохраненные файлы командного языка?
21. Какое расширение имеют сохраненные файлы с записью команд синтаксиса?
22. Как устанавливается в SPSS Syntax Reference Guide?

# ЗАКЛЮЧЕНИЕ

Работая над этой и другими книгами (1-4, 27, 36), мы, как бы вместе с пакетом SPSS, прошли весьма значительный путь его эволюции от версии 6.1 до версии 11.5 для Windows в 1994-2005 гг. Конечно, история SPSS куда более продолжительна (23, С. 385), но ее изучение выходит за рамки нашей темы. Как и всякое масштабное явление, рассматриваемый программный продукт развивается противоречиво, проявляя заметные достоинства и недостатки.

К числу достоинств можно отнести:

- Дружественный интерфейс, равно как и глубокое понимание разработчиками характера и особенностей восприятия, а также потребностей реальных и потенциальных пользователей.

- Максимальную технологическую соотнесенность с процессом ввода и обработки первичной социально-экономической информации.

- Высокую культуру и доступность выполнения требований обоснования достоверности статистической информации (вся область сравнения средних), которая, несомненно, окажет влияние на повышение общего научного уровня результатов эмпирических исследований в социологии и всей социальной науке в целом.

- Выход на прямое и непосредственное моделирование социальных процессов (Amos), позволяющий существенно повысить познавательный потенциал и практическую значимость эмпирических исследований и разработок.

- Командный язык, который открывает перед пользователем возможность самостоятельно формирования команд, выходящих за рамки стандартных решений.

Среди недостатков уместно отметить:

- Эволюция системы идет в направлении формирования неоправданно опережающих требований к обновлению технического парка пользователями, что, естественно, весьма затруднительно для образовательных и научных учреждений, а также исследовательских коллективов. Начиная освоение системы, мы поставили SPSS 6.1 на 386 компьютер и были счастливы в течение нескольких лет. А все последующие версии с нарастающим эффектом стали предъявлять требования к техническим характеристикам компьютеров (прежде всего, размеру жесткого диска и оперативной памяти). Более того, версии SPSS 10.0 и выше стали формулировать и повышенные требования к

монитору (видимо, как прямую плату за цветную панель инструментов в Amos 4.0 и выше).

- Системное введение годового срока лицензионного доступа к пакету, а также платные условия его последующего обновления ставят потребителей в исключительную и несимметричную зависимость как от производителей, так и посредников-распространителей системы.

- В интерфейсе многозначное использование отдельных терминов иногда ставит начинающих пользователей в довольно сложное положение. Конкретно, в версиях SPSS 10.0 и выше к таким терминам следует отнести «Data» и «View». В самом деле, открытие рабочего файла выполняется путем File – Open – Data. В главном меню на экране две из нескольких основных команд: View и Data. В левом нижнем углу под таблицей с данными можно видеть две небольшие кнопки: View и Data. И, наконец, в разделе главного меню View последняя строка высвечивается в обратной зависимости от того, что выставлено на экране в редакторе (данные или их описание) всегда стоит Data или View. Сходная картина наблюдалась в предшествующих версиях до SPSS 9.0 для термина «Statistics». Выполнение этой команды главного меню, в конечном счете, всегда вело к выполнению последовательности команд: Statistics – Statistics (в главных диалоговых окнах различных процедур). Поэтому появление в главном меню раздела Analyze представляется вполне оправданным шагом.

- Отказ от введения модуля Amos непосредственно в сам пакет тормозит, но не остановит развитие моделирования в социально-экономических исследованиях.

- Самым слабым местом в системе было и остается окно просмотра (SPSS Viewer). Хотя справедливости ради следует сказать, что в рассматриваемый период в этом плане была проделана огромная работа. В последних версиях системы окно просмотра существенно (в лучшую сторону) отличается от окна вывода (Output) ранних версий. Более того, появление в системе специального приложения Smart Viewer дает основание полагать, что разработчики системы сами озабочены и стремятся к разработке более качественных способов представления данных. При этом возникают почти философские проблемы эволюции системы от специализации к универсализации, которые уже стали камнем преткновения для многих других программных продуктов. Понятно, что обсуждение этого вопроса выходит за рамки данного пособия.

- Пользователям без специальной подготовки трудно понять основа-

ния, по которым «Мера сходства или различия» включена в состав корреляционного анализа (глава 11, § 11.4). Это процедура анализа независимых переменных, которые по определению не связаны между собой. Соответственно, здесь нет причинно-следственных связей, а есть отношения подобия, и это уже совсем другая логика.

В целом освоение программных продуктов, подобных SPSS, вряд ли справедливо рассматривать как довольно легкое занятие или техническую задачу повышения квалификации кадров. Скорее, это похоже на инвестирование в человеческий капитал. Выработка специфических навыков и умений позволяет его носителям использовать современные информационные технологии. В свою очередь, это ведет к тому, что человеческий капитал приобретает совершенно новые свойства.

Видимо, освоение широкими кругами социологической общественности компьютерных возможностей доступности, анализа и моделирования данных будет способствовать сначала распространению вторичного анализа, а затем приведет к формированию нового направления, которое может быть названо «экспериментальной социологией».

Понятие «экспериментальная социология» в приводимом контексте включает в себя два основных аспекта. Один из них - связан с распространением моделирования в социологии. А другой – с формированием культуры систематического вторичного анализа, контроля и воспроизводимости результатов социологических исследований.

Оба этих обстоятельства могут иметь решающее значение для формирования и укрепления статуса социологии как науки. При таком развитии событий социология неизбежно начнет изживать свое идеологическое прошлое, а социологи в большей мере станут походить на ученых и меньше, чем сегодня, будут заниматься псевдонаучным теоретизированием. Это не значит, что в социологической среде и в обществе в целом исчезнут концептуальные схемы и разработки, направленные на критику и совершенствование существующих общественных отношений, но это значит, что такая критика, равно как и такие разработки будут более реалистическими и эмпирически обоснованными.

В конечном счете, развитие событий в указанном направлении будет вести к тому, что само существование социологов, не имеющих опыта работы с эмпирическими данными, станет проблематичным. В этом плане овладение широкими массами исследователей и управленцев программными продуктами, позволяющими сохранять социологичес-

кие данные и выполнять их анализ, может иметь заметные последствия.

Оно, во-первых, должно способствовать улучшению качества и результативности самих социологических исследований и разработок. Во-вторых, такое развитие событий неизбежно приведет к повышению в профессиональном сообществе **иммунитета** к бесплодному теоретизированию и эмпирически необоснованным выводам, которые пока еще являются скорее нормой, чем исключением.

Возвращаясь к целевому назначению и задачам представленного выше учебного пособия, считаем необходимым, хотя бы в общем виде, сформулировать критерии его эффективности. Например, можно попросить нашего читателя зрительно представить себе и описать словесно, каким будет результат выполнения последовательности команд: Analyze - Descriptive Statistics - Frequencies – выделение и перенос переменной – OK? Если читатель или слушатель сможет дать правильный ответ, то все мы не напрасно провели время за книгой и в аудиториях.

Если же наш читатель сможет дать правильный ответ и на более сложный вопрос: «Что будет записано в окне синтаксиса после выполнения команд: Analyze - Descriptive Statistics - Frequencies – выделение и перенос переменной – Paste?», то можно с уверенностью сказать, что все мы тратили свое время, силы и деньги с большой пользой.

Любой преподаватель, владеющий основами SPSS, используя приведенные тесты в качестве образцов, может сформулировать огромное число таких вопросов и тестов как для текущей семинарской работы, так и для зачетной и экзаменационной сессии. Будет очень хорошо, если кто-то из читателей или пользователей этой книги обменяется с нами своим опытом.

# ЛИТЕРАТУРА


1. Пациорковский В.В., Петрова А.И., Пациорковская В.В. Использование SPSS в социологии. Часть 1. Ввод и контроль данных: Учебное пособие.- М.: ИСЭПН РАН, 1998.- 116 с.

2. Пациорковский В.В., Петрова А.И., Пациорковская В.В. Социология. Ввод и контроль данных в SPSS: Учебное пособие.- М.: МИРЭА, 1999.- 112 с.

3. Пациорковский В.В., Дершем Л.Д., Петрова А.И., Пациорковская В.В. Использование SPSS в социологии. Часть 2. Анализ данных: общие принципы, суммарные статистики и графики: Учебное пособие.- М.: ИСЭПН РАН, 2000.- 150 с.

4. Пациорковский В.В., Петрова А.И., Пациорковская В.В. Использование SPSS в социологии. Часть 3. Анализ данных: меры сравнения, прогнозирование и моделирование: Учебное пособие.- М.: ИСЭПН РАН, 2002.- 152 с.

5. Крухмалева О.В., Савина Н.Е. Использование SPSS в социологии. Учебное пособие в 3-х частях: Ч. I. Ввод и контроль данных; Ч.II. Анализ данных: общие принципы, суммарные статистики и графики; Ч. III. Анализ данных: меры сравнения, прогнозирование и моделирование (Пациорковский В.В., Петрова А.И., Пациорковская В.В.). М.: ИСЭПН РАН, 1998, 2000, 2002. // Социологические исследования: 2003, № 5.- С. 155-156.

6. Лылова О., Рыжкина Е. Использование SPSS в социологии. // Народонаселение: 2002, № 3.- С. 136-138.

7. SPSS 7.5. Applications Guide.- Chicago: SPSS Inc., 1997.

8. SPSS Base 7.5 для Windows. Руководство по применению.- М.: Центр Общечеловеческих Ценностей, 1997.

9. Bablie E. The Practice of Social Research. 9th Edition.- Belmont, CA: Wadsworth/Thomson Learning, 2001.

10. Healey J., Bablie E., Halley F. Exploring Social Issues. Using SPSS for Windows.- Thousand Oaks, California: Pine Forge Press, 1997.

11. Bablie E., Halley F. Adventures in Social Research: Data Analysis Using SPSS for Windows.- Thousand Oaks, California: Pine Forge Press, 1997.

12. Dowdall G., Bablie E., Halley F. Adventures in Criminal Justice Research: Data - Analysis Using SPSS for Windows.- Thousand Oaks, California: Pine Forge Press, 1996.

13. Афанасьев В., Афанасьев Д. SPSS в студенческой аудитории. // Компьютер пресс: 1998, № 3,.- С.202-205.

14. Бызов Л.Г., Пациорковский В.В., Шкрабкина И.А. Контроль достоверности социологической информации с помощью ЭВМ. В сб.: Организационно-методические проблемы социологического опроса.- М.: ИСИ АН СССР, 1986.- С. 145-165.

15. Афанасьев В.И. Методические указания по курсу математической статистики с применением пакета SPSS.- М.: МЭИ, 1996.

16. Бююль А., Цефель П. SPSS: Искусство обработки информации. Анализ статистических данных и востстановление скрытых закономерностей.- М.: Диасофт, 2001.- 486 с.

17. Тюрин Ю.Н., Макаров А.А. Анализ данных на компьютере.- М.: Финансы и статистика, 1995- 384 с.

18. Математические методы в социально-экономических и археологических исследованиях.- М.: Наука, 1981.

19. SPSS для Windows. Руководство пользователя. Книга 1. Базовая система версии 6.1. Интерфейс. Разведочный анализ данных.- М.: АО "Статистические Системы и Сервис", 1995.

20. SPSS/PC+ 4.0 Base Manual.- Chicago: Marija J. Norusis/SPSS Inc., 1990.

21. SPSS 6.1 Syntax Reference Guide.- Chicago: SPSS Inc., 1994. 941 p.

22. Боровиков В.П. Популярное введение в программу STATISTICA.- М.: КомпьютерПресс, 1998.

23. Amos Users' Guide Version 3.6.- Chicago: SmallWaters Corporation, 1997.

24. Сетевой адрес: http://www.edu.nsu.ru/noos/metod/logint/21.htm

25. Сетевой адрес: http://www.icsti.su/ibd/Sart2.asp?T1=BBB

26. Сетевой адрес: http://www.krugosvet.ru/articles/15/1001551/1001551 a1.htm

27. Пациорковский В.В. Сельская Россия: 1991-2001 гг.- М.: Финансы и статистика, 2003.- 368 с.

28. Маслов П.П. Статистика в социологии.- М.: Статистика, 1971; Статистические методы анализа информации в социологических исследованиях. Отв. ред. Г.В.Осипов.- М.: Наука, 1979; Цыба В.Г. Математико-статистические основы социологических исследований.- М.: Финансы и статистика, 1981; Анализ нечисловой информации в социологических исследованиях.- М.: Наука, 1985 и др.

29. Методика и техника статистической обработки первичной социологической информации (отв. ред. Г.В. Осипов).- М.: Наука, 1968.

30. Кюн Ю. Описательная и индуктивная статистика.- М.: Финансы и статистика, 1981.

31. Кожухарь Л.И. Основы общей теории статистики.- М.: Финансы и статистика, 1999.

32. Электронный учебник Statsoft.- М,: Statsoft Inc., 2001.
Web: http://www.statsoft.ru/home/textbook/default.html

33. Чесноков С. Песни в жизни персонажа.- М.: Socio.ru, 2001.
Web: http://www.socio.ru/public/chesnokov/Text.zip

34. Косолапов М. Искушение математикой.- М.: ФОМ, 2001.
Web: http://socium.fom-discurs.ru/?act=thread&forum=1

35. SPSS - Российское отделение.- М.: SPSS, 2001.
Web: http://www.spss.ru/atwork.htm

36 а) O'Brien, D. J., V. V. Patsiorkovski and L. D. Dershem. Household Capital and the Agrarian Problem in Russia.- Aldershot:Ashgate. 2000; б) O'Brien, D. J., V. V. Patsiorkovski and L. D. Dershem. Rural responses to land reform in Russia: an analysis of household land use in Belgorod, Rostov and Tver' Oblasts from 1991 to 1996. / Land Reform in the Former Soviet Union and Eastern Europe. S. K. Wegren (ed.).- London: Routledge. 1998. Pp. 35-61.

37. Сазонов Б.В. Роль вторичного анализа в методологизации социального

познания и в развитии профессии «социальный исследователь». // Системные исследования. Ежегодник 1997.- М.: УРСС, 1998.

38. Дубров А.М., Михтарян В.С., Трошин Л.И. Многомерные статистические методы.- М.: Финансы и статистика, 2000. 352 с.

39. Сетевой адрес: http://www.kgafk.ru/kgufk/html/uchmetrologia9.html

40. Константиновский Д.Л., Овсянников А.А.,.Покровский Н.Е. Итоговый аналитический отчет по результатам реализации проектов по социологии вузов-участников Инновационного проекта развития образования. Сетевой адрес: http://www.sociolog.net/nfpk.doc

41. Arbuckle J., Wothke W. Amos 4.0 User's Guide.- Chicago, IL: SmallWaters Corporation, 1999.

42. Моделирование в социологических исследованиях.- М.: Наука, 1978; Многомерный анализ социологических данных.- М.: ИС АН СССР, 1981; Математические методы и модели в социологии.- М.: ИС АН СССР, 1991 и др.

43. SPSS 11.5 Syntax Reference Guide.- Chicago: SPSS Inc.2002. 1490 p.

44. Современный словарь иностранных слов.- М.: Русский язык, 2000.

45. Стариков В.Н., Абросимова М.А. Как подготовить научную публикацию, учебное пособие, научный отчет, диссертацию с помощью пакета Microsoft Word 7.0 для Windows.- Уфа: БГУ, 1997.- 126 с.

# ПРИЛОЖЕНИЯ

Приложение 1.

Отдельные отзывы

*Письмо 1.*

Уважаемый Валерий Валентинович!

Меня зовут Екатерина. Я руковожу социологической службой Центра политологических исследований при ДГУ. В прошлом году во время стажировки в Институте социологии РАН я приобрела ваше учебное пособие «Использование SPSS в социологии. Часть 1». Буду очень признательна, если Вы сообщите мне, вышла ли вторая часть этого издания, и где его можно приобрести.

С уважением, Е.И.

*Письмо 2.*

От кого: «G. S/» <sg…@spss.com>
Кому: «"Valeri Patsiorkovski"» <patsv@mail.ru>
Дата: Thu, 19 Apr 2001 11:47:03 -0500
Тема: RE: Re[4]: SPSS

Dear Valeri:

I received your SPSS books. Thanks so much! They were a hit at SPSS Corporate headquarters in Chicago, where I am today!

S. J. G.
Ph.D.Vice President for Strategic Planning,
Public SectorSPSS, Inc. 2000 N. 14th Street, Suite 320 Arlington, VA.

*Письмо 3.*

От: «K.I.» <ki…@ceu.edu.pl>
Дата: Fri, 25 Apr 2003

Uvazhaemyj Valerij Valentinovich!

Kazhdyj raz, obraschajas' k knige po SPSS, khochu vas poblagodarit'. Ochen' poleznoje rukovodstvo. Nadejus', chto ne poslednee. Hope to hear from you.

Yours, Katya.

*Письмо 4.*

From: «metod» <metod@library.tversu.ru>
To: «Valeri Patsiorkovski» <patsv@mail.ru>
Date: Wed, 16 Jul 2003 13:49:17 +0400
Subject: Re: SPSS text book

Еще раз, добрый день, глубокоуважаемый Валерий Валентинович! Только что получила книги и благодарю Вас бесконечно. Думаю, что эти книги будут полезны и нам библиотечным работникам и самое главное, студентам факультета управления и социологии. Всего доброго. Б.Е.И.

*Письмо 5.*

From: «М.Б.» <mb...@mail.ru >
To: patsv@mail.ru
Date: Tue, 27 Apr 2004 05:16:40
Subject: Приобретение книги по SPSS

Добрый день!!!

Меня зовут М.Б. Я работаю в Управлении социального мониторинга аппарата администрации Сахалинской области. Нам нужно приобрести вашу книгу по SPSS. Прошу Вас сообщите об условиях приобретения книги.

С уважением, М.Б.

Приложение 2.

ИНСТИТУТ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ПРОБЛЕМ
НАРОДОНАСЕЛЕНИЯ РАН
УНИВЕРСИТЕТ МИССУРИ-КОЛУМБИЯ
МП «СОЦИАЛЬНАЯ НАУКА»

**ОПРОСНЫЙ ЛИСТ - 3**
(Фрагмент, полностью этот документ опубликован в 1, С. 65-81)

«СОЦИАЛЬНЫЕ СВЯЗИ, СОЦИАЛЬНЫЙ КАПИТАЛ И АДАПТАЦИЯ К СОЦИАЛЬНЫМ ИЗМЕНЕНИЯМ В СЕЛЬСКОЙ МЕСТНОСТИ РОССИИ»
(третий этап трехлетнего панельного исследования)

| | |
|---|---|
| Номер анкеты (домохозяйства) | _____ |
| Если домохозяйство новое, то почему? _____ | _____ |
| Село | _____ |
| Размер семьи | _____ |
| Число пенсионеров | _____ |
| Демографический тип семьи | _____ |
| Социальный тип семьи | _____ |
| Пол респондента | _____ |
| Возраст респондента | _____ |
| Занятость респондента | _____ |
| Отношение к главе семьи | _____ |
| Если респондент новый, то почему? _____ | _____ |

Полевые работы по данному проекту финансируются
Национальным научным фондом (NSF) США

Москва-Колумбия,
1997 г.

Раздел 1. СОЦИАЛЬНО-ДЕМОГРАФИЧЕСКИЕ ХАРАКТЕРИСТИКИ СЕМЬИ

| Состав семьи | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| 1. **Отношение к главе семьи:** муж - 1, жена - 2, другие взрослые - 3-5, дети до 18 лет - 6-8 | | | | | | | | |
| 2. **Число исполнившихся лет** (детям до года - 1) | | | | | | | | |
| 3. **Пол**: мужской-1, женский-2 | | | | | | | | |
| 4. **Национальность**: русский-1 украинец- 2, другая- 3 | | | | | | | | |
| 5**. Образование**: число полных лет учебы | | | | | | | | |
| 6. **Состояние в браке**: женат (замужем)-1, холост-2, вдовец (вдова)-3, разведен-4 | | | | | | | | |
| 7. **Занятость**: полный день- 1 часть дня-2, безработный-3, пенсионер-4, нетрудоспособный (инвалид)-5, домохозяйка-6, по уходу эа ребенком-7, учащийся-8 | | | | | | | | |
| 8. **Место работы:** свое село-1, др. село-2, райцентр-3, город-4 | | | | | | | | |
| 9. **Предприятие:** ТОО/АО-1, колхоз- 2, обществ. бсл.- 3, ферм.хоз-во- 4, другой агробизнес- 5, другой бизнес- 6 | | | | | | | | |
| 10. **Должность:** руководитель-1, специалист- 2, служащий- 3, рабочий/колхозник- 4, фермер- 5 | | | | | | | | |

Раздел П. ТРУД, ДОХОДЫ, ЛИЧНОЕ ПОДСОБНОЕ ХОЗЯЙСТВО

11. **Имеете ли вы или кто-нибудь из членов вашей семьи свое дело (назовите)?** (в агробизнесе- 1, в другом бизнесе- 2, нет- 3, затр. ответить- 9) _____
12. **Если имеете, то как оказалось возможным его начать?** (источники средств: личные сбережения- 1, _____
от реализации продукции- 2, кредит- 3, заем- 4, _____
что-то другое- 5, затр. отв.- 9) _____
13. **Есть ли у вас партнеры по бизнесу?** (да- 1, нет- 2, затр. отв.- 9) _____
14. **Если есть, то кто они?** (член семьи- 1, друг- 2, _____
сосед- 3, сослуживец- 4, другие- 5, затр. отв.- 9) _____

15. **Какая площадь обрабатываемого вами участка земли** (в том числе: приусадебного участка, арендуемой и находящейся в вашем пользовании земли)? Затр. Ответить- 999.

Приусадебный участок (га) _____
Аренда (га) _____
Другие виды землепользования (га) _____

16. **Брали ли вы лично или кто-то другой из членов вашей семьи кредит (заем) в 1991-1997 гг.** (год взятия кредита - от 1 до 7, не брали - 8, затр. Ответить - 9) _____

17. **Если кредит брался, то**:

на какой срок ( лет) _____
на какую сумму (тыс. рубл.) _____
под какой процент (%) _____
для какой цели (строительство - 1, покупка техники - 2, коммерция - 3, другая - 4) _____
кредитор (родные - 1, близкие - 2, банк - 3, специальная программа - 4, другой - 5, затрудняюсь ответить - 9) _____

18. **Есть ли в вашем хозяйстве** (и сколько)?

Коровы (телята) _____
Лошади _____
Свиньи _____
Другой скот (козы, овцы) _____
Птица _____
Автомашина _____
Мотоцикл/мотороллер _____
Трактор _____
Другие с/х машины (мотоблок и пр.) _____
Другое механическое оборудовние _____
Телефон _____
Видеотехника _____

19. **Если у вас нет сельскохозяйственной техники**, **как вы обрабатываете свою землю?** (договариваетесь с колхозом- 1, ТОО/АО- 2, частными лицами- 3, родными-4, знакомыми-5, делаете сами-6, затр.отв.- 9)? _____

20. **Используете ли вы в своем хозяйстве современные технологии?** (да- 1, нет- 2, затр. отв.- 9)

Селекционные семена и саженцы _____
Племенной скот _____
Минеральные удобрения _____
Органические удобрения _____
Гербициды и пестициды _____
Теплицы _____
Что еще? _____ _____

21. **Если используете, то где приобретаете?** _____
(в колхозе-1, ТОО/АО- 2, на рынке- 3, в частном _____
секторе-4, магазине-5, у знакомых и родственни- _____
ков-6, через земельный комитет-7, затр. отв.- 9)

22. **Какую продукцию вы получили в своем личном хозяйстве за последний год (в кг)?** Затр.отв.- 99999

| | Получили | Продали |
|---|---|---|
| Картофель | _____ | _____ |
| Овощи | _____ | _____ |
| Фрукты | _____ | _____ |
| Мясо(включая птицу) | _____ | _____ |
| Молоко | _____ | _____ |
| Сено | _____ | _____ |
| Другое | _____ | _____ |

23. **Собираетесь ли вы при благоприятных условиях повысить продуктивность своего хозяйства в 1997-1998 гг.?** _____
(значительно-1,немного-2, нет-3, сокращу- 4, затр. отв.-9)

24. **Укажите, пожалуйста, сумму денежных доходов каждого взрослого члена вашей семьи за последний месяц (тыс. руб.) по следующим источникам** (затр. отв.- 9):

| | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Зарплата, премия (основные) | | | | | |
| Зарплата, премия (дополн.раб) | | | | | |
| Пенсия, стипендия | | | | | |
| Алименты | | | | | |
| Пособие на детей | | | | | |
| Доходы от продажи сельхозпр | | | | | |
| Дивиденды, паевые, проценты | | | | | |
| Доходы от предприн. Деятельн. | | | | | |
| Прочие денежные доходы | | | | | |
| Итого для каждого члена семьи | | | | | |

Всего денежный доход семьи: _____

Приложение 3.

**Методика фиксации изменений в массиве опрашиваемых трех волнового панельного исследования (1995 - 1997 гг.)**

| ОПРОСНЫЙ ЛИСТ- 1 | ОПРОСНЫЕ ЛИСТЫ 2-3 |
|---|---|
| Номер анкеты) (домохозяйства) _____ | Номер анкеты (домохозяйства) _____ |
| Если домохозяйство новое, то почему?_____ | |
| Село | Село |
| Размер семьи | Размер семьи |
| Число пенсионеров | Число пенсионеров |
| Демографический тип семьи | Демографический тип семьи |
| Социальный тип семьи | Социальный тип семьи |
| Пол респондента | Пол респондента |
| Возраст респондента | Возраст респондента |
| Занятость респондента | Занятость респондента |
| Отношение к главе семьи | Отношение к главе семьи |
| **Если респондент новый, то почему?_____** | |

Приложение 4.

**Фрагмент макета ввода данных в ЭВМ**
(Полностью документ опубликован в 1, С.. 83-107)
ОПРОСНЫЙ ЛИСТ – 3

СОЦИАЛЬНЫЕ СВЯЗИ, СОЦИАЛЬНЫЙ КАПИТАЛ И АДАПТАЦИЯ К СОЦИАЛЬНЫМ ИЗМЕНЕНИЯМ В СЕЛЬСКОЙ МЕСТНОСТИ РОССИИ
(третий этап трехлетнего панельного исследования)

| Индикаторы и их коды | Имя переменной | Ширина переменной |
|---|---|---|
| **Номер анкеты (домохозяйства)** | Id7 | 000 |
| **Если домохозяйство новое,**<br>прежнее - 0,<br>старая усадьба-новые хозяева-1,<br>новое д/х - недоступность старого-2 | newhous7 | 0 |
| **Село**<br>Латоново - 1,<br>Венгеровка - 2,<br>Святцово - 3 | village7 | 0 |

| Индикаторы и их коды | Имя переменной | Ширина переменной |
|---|---|---|
| **Размер семьи** | numfam7 | 0 |
| **Число пенсионеров** | retired7 | 0 |
| **Демографический тип семьи** | demtype7 | 0 |
| одиночки - 1, супружеские пары пенсионеров-2, супружеские пары работников-3, супружеские пары с детьми -4, супружеские пары с детьми и родственниками-5, неполные семьи-6, прочие-7 | | |
| **Социальный тип семьи** | soctype7 | 0 |
| семьи руководителей 1, семьи специалистов -2, семьи служащих-3, семьи колхозников и рабочих -4, семьи фермеров -5, прочие -6, семьи пенсионеров -7 | | |
| **Пол респондента** | sexresp7 | 0 |
| **Возраст респондента** | ageresp7 | 00 |
| **Занятость респондента** | emplres7 | 0 |
| **Отношение к главе семьи** | respond7 | 0 |
| **Если респондент новый** | newresp7 | 0 |
| прежний - 0, прежний респондент умер - 1, прежний респондент недоступен-2, отказ - 3 | | |

Раздел 1. СОЦИАЛЬНО-ДЕМОГРАФИЧЕСКИЕ ХАРАКТЕРИСТИКИ СЕМЬИ

Состав семьи:

| | | |
|---|---|---|
| 1. **Отношение к главе семьи**: | | |
| муж -1 | husband7 | 0 |
| жена -2 | wife7 | 0 |
| другие взрослые члены семьи -3-5 | othad17 | 0 |
| | othad27 | 0 |
| | othad37 | 0 |
| дети до 18 лет - 6- 8 | child17 | 0 |
| | child27 | 0 |
| | child37 | 0 |
| 2. **Число исполнившихся лет** | | |
| муж | hage7 | 00 |
| жена | wage7 | 00 |

| Индикаторы и их коды | Имя переменной | Ширина переменной |
|---|---|---|
| другие взрослые члены семьи | oage17 | 00 |
| | oage27 | 00 |
| | oage37 | 00 |
| дети до 18 лет | cage17 | 00 |
| | cage27 | 00 |
| | cage37 | 00 |
| 3. **Пол**: мужской-1, женский-2 | | |
| муж | hsex7 | 0 |
| жена | wsex7 | 0 |
| другие взрослые члены семьи | osex17 | 0 |
| | osex27 | 0 |
| | osex37 | 0 |
| дети до 18 лет | csex17 | 0 |
| | csex27 | 0 |
| | csex37 | 0 |
| 4. **Национальность** | | |
| русский-1,украинец-2, другая-3 | | |
| муж | hnat7 | 0 |
| жена | wnat7 | 0 |
| другие взрослые члены семьи | onat17 | 0 |
| | onat27 | 0 |
| | onat37 | 0 |
| 5. **Образование** (число полных лет учебы) | | |
| муж | heduc7 | 00 |
| жена | weduc7 | 00 |
| другие взрослые члены семьи | oeduc17 | 00 |
| | oeduc27 | 00 |
| | oeduc37 | 00 |
| 6. **Состояние в браке**: женат (замужем)-1, холост-2, вдовец(вдова)-3, разведен-4 | | |
| муж | husm7 | 0 |
| жена | wifem7 | 0 |
| другие взрослые члены семьи | omarst17 | 0 |
| | omarst27 | 0 |
| | omarst37 | 0 |
| 7. **Занятость**: полный день-1, часть дня-2, безработный-3, пенсионер-4, нетрудо-способный (инвалид)-5, домохозяйка-6, по уходу эа ребенком-7, учащийся-8 | | |
| муж | hempl7 | 0 |
| жена | wempl7 | 0 |
| другие взрослые члены семьи | oempl17 | 0 |

| Индикаторы и их коды | Имя переменной | Ширина переменной |
|---|---|---|
| | oempl27 | 0 |
| | oempl37 | 0 |
| 8. **Место работы**: свое село-1, другое село-2, райцентр-3, город-4 | | |
| муж | hplace7 | 0 |
| жена | wplace7 | 0 |
| другие взрослые члены семьи | oplace17 | 0 |
| | oplace27 | 0 |
| | oplace37 | 0 |
| 9. **Предприятие**: ТОО/АО-1, колхоз- 2, обществ. обслуживание- 3,ферм.хоз-во- 4, другой агробизнес- 5, другой бизнес- 6 | | |
| муж | hbus7 | 0 |
| жена | wbus7 | 0 |
| другие взрослые члены семьи | obus17 | 0 |
| | obus27 | 0 |
| | obus37 | 0 |
| 10. **Должность**: руководитель-1, специалист- 2, служащий- 3, рабочий/колхозник- 4, фермер- 5 | | |
| муж | hpos7 | 0 |
| жена | wpos7 | 0 |
| другие взрослые члены семьи | opos17 | 0 |
| | opos27 | 0 |
| | opos37 | 0 |
| Раздел П. ТРУД, ДОХОДЫ, ЛИЧНОЕ ПОДСОБНОЕ ХОЗЯЙСТВО | | |
| 11. **Имеете ли вы или кто-нибудь из членов вашей семьи свое дело**? (в агробизнесе- 1, в другом бизнесе- 2, нет- 3, затр. ответить- 9) | busines7 | 0 |
| 12. **Если имеете, то как оказалось возможным его начать**? | busine17 | 0 |
| | busine27 | 0 |
| три варианта ответа (источники средств: личные сбережения-1, от реализации продукции-2, кредит-3, заем-4, что-то другое-5, затр.отв.- 9) | busine37 | 0 |
| . **Есть ли у вас партнеры по бизнесу**? (да- 1, нет- 2, затр. отв.- 9) | partner7 | 0 |
| 14. **Если есть, то кто они**? | part17 | 0 |
| три варианта ответа | part27 | 0 |
| член семьи- 1, друг- 2, сосед- 3, сослуживец- 4 другие- 5, затр. отв.- 9 | part37 | 0 |

| Индикаторы и их коды | Имя переменной | Ширина переменной |
|---|---|---|
| 15. **Какая площадь обрабатываемого вами участка земли** (га) (в том числе: приусадебного участка, арендуедуемой и находящейся в вашем пользовании земли)? Затр. ответить- 999. | | |
| Приусадебный участок | plot7 | 000 |
| Аренда | tenantr7 | 000 |
| Другие виды землепользования | othland7 | 000 |
| 16. **Брали ли вы лично или кто-то другой из членов вашей семьи кредит (заем) в 1991-1997 гг.** ? (год взятия кредита - от 1 до 7, не брали - 8, затр. ответить - 9 | credit7 | 0 |
| 17. **Если кредит брался, то**: | | |
| на какой срок ( лет) | yr7 | 00 |
| на какую сумму (тыс.руб.) | sum7 | 000000 |
| под какой процент (%) | per7 | 000 |
| для какой цели (строительство - 1, покупка техники-2, коммерция-3, другая-4) | aim7 | 0 |
| кредитор (родные - 1, близкие - 2, банк - 3, специальная програм-ма - 4, другой - 5, затр.отв.- 9) | cred7 | 0 |
| 18. **Есть ли в вашем хозяйстве**: | | |
| Коровы (телята) | cow7 | 0 |
| Лошади | horses7 | 0 |
| Свиньи | pig7 | 00 |
| Другой скот (козы, овцы) | sheep7 | 00 |
| Птица | poultry7 | 000 |
| Автомашина | car7 | 0 |
| Мотоцикл/мотороллер | motcycl7 | 0 |
| Трактор | tractor7 | 0 |
| Другие с/х машины | agrmoto7 | 0 |
| Другое механическое оборудовние | agrtool7 | 0 |
| Телефон | teleph7 | 0 |
| Видеотехника | video7 | 0 |
| 19. **Если у вас нет сельскохозяйственной техники, как вы обрабатываете свою землю**? (три варианта ответа: | | |

| Индикаторы и их коды | Имя переменной | Ширина переменной |
|---|---|---|
| договариваетесь с колхозом-1, ТОО/АО- 2, частными лицами- 3, родными-4, знакомыми-5, делаете сами-6, затр.отв.- 9) | notmec17<br>notmec27<br>notmec37 | 0<br>0<br>0 |
| 20. **Используете ли вы в своем хозяйстве современные технологии**? (да- 1, нет- 2, затр.отв.- 9) | | |
| Селекционные семена и саженцы | select7 | 0 |
| Племенной скот | pedigre7 | 0 |
| Минеральные удобрения | mineral7 | 0 |
| Органические удобрения | organic7 | 0 |
| Гербициды и пестициды | gerb7 | 0 |
| Теплицы | greenhs7 | 0 |
| 21.**Если используете, то где приобретаете**? (три варианта ответа: в колхозе-1, ТОО/АО- 2, на рынке- 3, в частном секторе-4, магазине-5, у знакомых и родственников-6, через земельный комитет-7, затр.отв.- 9) | acquir17<br>acquir27<br>acquir37 | 0<br>0<br>0 |
| 22. **Какую продукцию вы получили в своем личном хозяйстве за последний год** (в кг)? Затр.отв.- 9999 | | |
| Картофель | | |
| получили | potprod7 | 00000 |
| продали | potsold7 | 0000 |
| Овощи | | |
| получили | vegprod7 | 0000 |
| продали | vegsold7 | 0000 |
| Фрукты | | |
| получили | fruprod7 | 0000 |
| продали | frusold7 | 0000 |
| Мясо(включая птицу) | | |
| получили | meatpro7 | 0000 |
| продали | meatsol7 | 0000 |
| Молоко | | |
| получили | milkpro7 | 00000 |
| продали | milksol7 | 00000 |
| Сено | | |
| получили | hayprod7 | 00000 |
| продали | haysold7 | 00000 |
| **23. Собираетесь ли вы при благо-** | | |

| Индикаторы и их коды | Имя переменной | Ширина переменной |
|---|---|---|
| **приятных условиях повысить продуктивность своего хозяйства в 1997-1998 гг.**? (значительно-1, немного-2, нет-3, сокращу- 4, затр. ответить.-9) | promote7 | 0 |
| 24. **Укажите, пожалуйста, сумму денежных доходов каждого взрослого члена вашей семьи за последний месяц (в тыс. руб.) по следующим источникам** (затр. Отв.- 9): | | |
| **Зарплата, премия** (основные) | | |
| муж | psalw17 | 0000000 |
| жена | psalw27 | 0000000 |
| другие взрослые члены семьи | psalw37 | 0000000 |
| | psalw47 | 0000000 |
| | psalw57 | 0000000 |
| **Зарплата, премия** (дополн.раб) | | |
| муж | salwag17 | 0000000 |
| жена | salwag27 | 0000000 |
| другие взрослые члены семьи | salwag37 | 0000000 |
| | salwag47 | 0000000 |
| | salwag57 | 0000000 |
| **Пенсия, стипендия** | | |
| муж | pensio17 | 000000 |
| жена | pensio27 | 000000 |
| другие взрослые члены семьи | pensio37 | 000000 |
| | pensio47 | 000000 |
| | pensio57 | 000000 |
| **Алименты** | | |
| муж | alimon17 | 000000 |
| жена | alimon27 | 000000 |
| другие взрослые члены семьи | alimon37 | 000000 |
| | alimon47 | 000000 |
| | alimon57 | 000000 |
| **Пособие на детей** | | |
| муж | ch17 | 000000 |
| жена | ch27 | 000000 |
| другие взрослые члены семьи | ch37 | 000000 |
| | ch47 | 000000 |
| | ch57 | 000000 |
| **Доходы от продажи сельхозпр**. | | |
| муж | incplo17 | 0000000 |
| жена | incplo27 | 0000000 |

| Индикаторы их коды | Имя переменной | Ширина переменной |
|---|---|---|
| другие взрослые члены семьи | incplo37 | 0000000 |
| | incplo47 | 0000000 |
| | incplo57 | 0000000 |
| **Дивиденды, паевые, проценты** | | |
| муж | divid17 | 000000 |
| жена | divid27 | 000000 |
| другие взрослые члены семьи | divid37 | 000000 |
| | divid47 | 000000 |
| | divid57 | 000000 |
| **Доходы от предприн. деятельн**. | | |
| муж | income17 | 0000000 |
| жена | income27 | 0000000 |
| другие взрослые члены семьи | income37 | 0000000 |
| | income47 | 0000000 |
| | income57 | 0000000 |
| **Прочие денежные доходы** | | |
| муж | othben17 | 000000 |
| жена | othben27 | 000000 |
| другие взрослые члены семьи | othben37 | 000000 |
| | othben47 | 000000 |
| | othben57 | 000000 |
| **Итого для каждого члена семьи** | | |
| муж | total17 | 0000000 |
| жена | total27 | 0000000 |
| другие взрослые члены семьи | total37 | 0000000 |
| | total47 | 0000000 |
| | total57 | 0000000 |
| **Всего денежный доход семьи** | total7 | 00000000 |

Приложение 5.

# Инструкция пошагового выполнения работ по вводу и контролю данных социологических исследований

**Шаг 1.**

Подготовка макета опросного листа: присвоение всем переменным уникальных имен с использованием латинских букв.(не более 8 символов) и определение ширины переменной (количество знаков).

**Шаг 2.**

Создание рабочего файла:( File - Save as. - Имя файла - Ok).

**Шаг 3.**

Создание переменных в таблице. *Внесение имен переменных* путем двойного щелчка мышью на затененном названии столбца (var), и использовании открывшегося окна Define Variable (определить переменную), *определение их типа* (кнопка Type в окне), *определение меток* (кнопка Lables), *пропущенных значений* (если таковые имеются) (кнопка Missing).

**Шаг 4.**

Заполнение ячеек таблицы данными путем установки курсора на нужную ячейку: в строке редактора ячейки появляется набранное число и после нажатия Enter переносится в ячейку.

**Шаг 5.**

Полностью заполненная таблица по строкам и столбцам представляет собой базу данных, перед анализом которой следует провести контроль. Контроль данных осуществляется с помощью статистических процедур, таких как Frequencies (частотные таблицы позволяют определить некорректные значения), Crosstabs (таблицы сопряженности позволяют проверить взаимосвязанные переменные), Compute (создание новых переменных), List Cases (просмотр наблюдений) и Select Cases (отбора наблюдений).

Приложение 6.

## Окно вывода Amos Graphics

sales99a
Sunday, May 27, 2001 09:15:59

Amos
by James L. Arbuckle
Version 4.01


1507 E. 53rd Street - #452
Chicago, IL 60615 USA
773-667-8635
Fax: 773-955-6252
http://www.smallwaters.com

**********

Title
sales99a: Sunday, May 27, 2001 09:15 PM

Your model contains the following variables

| | | |
|---|---|---|
| LOGNWTPR | observed | endogenous |
| LOGWANIM | observed | endogenous |
| LGWTOTSL | observed | endogenous |
| QUMPEOP | observed | exogenous |
| COMINVPA | observed | exogenous |
| NUMPEOP | observed | exogenous |
| YEAR99 | observed | exogenous |
| NWNUMADL | observed | exogenous |
| error2 | unobserved | exogenous |
| error1 | unobserved | exogenous |
| error3 | unobserved | exogenous |

| | |
|---|---|
| Number of variables in your model: | 11 |
| Number of observed variables: | 8 |
| Number of unobserved variables: | 3 |
| Number of exogenous variables: | 8 |
| Number of endogenous variables: | 3 |

Summary of Parameters

| | Weights | Covariances | Variances | Means | Intercepts | Total |
|---|---|---|---|---|---|---|
| Fixed: | 3 | 0 | 0 | 0 | 0 | 3 |
| Labeled: | 0 | 0 | 0 | 0 | 0 | 0 |
| Unlabeled: | 9 | 12 | 8 | 5 | 3 | 37 |
| Total: | 12 | 12 | 8 | 5 | 3 | 40 |

NOTE:

The model is recursive.
Sample size: 1266
Model: Default model
Computation of degrees of freedom

| | |
|---|---|
| Number of distinct sample moments: | 44 |
| Number of distinct parameters to be estimated: | 37 |
| --- | --- |
| Degrees of freedom: | 7 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 0e | 5 | 0.0e+000 | -6.4191e-001 | 1.00e+004 | 8.62590592560e+003 | 0 | 1.00e+004 |
| 1e* | 5 | 0.0e+000 | -5.3583e+000 | 8.17e-001 | 4.20450804666e+003 | 17 | 8.67e-001 |
| 2e* | 4 | 0.0e+000 | -1.0952e+001 | 2.27e-001 | 2.21412659305e+003 | 6 | 1.16e+000 |
| 3e | 3 | 0.0e+000 | -5.0718e-001 | 3.27e-002 | 1.90010906197e+003 | 7 | 7.65e-001 |
| 4e | 2 | 0.0e+000 | -6.9893e-002 | 4.47e-001 | 1.20030764825e+003 | 13 | 7.66e-001 |
| 5e | 0 | 2.4e+004 | 0.0000e+000 | 7.34e-001 | 5.66186024530e+002 | 6 | 9.77e-001 |
| 6e | 0 | 5.6e+003 | 0.0000e+000 | 8.15e-001 | 3.27164807181e+002 | 4 | 0.00e+000 |
| 7e | 0 | 5.0e+003 | 0.0000e+000 | 5.16e-001 | 8.88006574998e+001 | 1 | 9.83e-001 |
| 8e | 0 | 5.7e+003 | 0.0000e+000 | 1.68e-001 | 1.54340359597e+001 | 1 | 1.15e+000 |
| 9e | 0 | 6.0e+003 | 0.0000e+000 | 4.34e-002 | 9.82690039918e+000 | 1 | 1.07e+000 |
| 10e | 0 | 6.0e+003 | 0.0000e+000 | 6.02e-003 | 9.74602611211e+000 | 1 | 1.01e+000 |
| 11e | 0 | 5.9e+003 | 0.0000e+000 | 1.69e-004 | 9.74598305867e+000 | 1 | 1.00e+000 |

Minimum was achieved
Chi-square = 9.746
Degrees of freedom = 7
Probability level = 0.203

Maximum Likelihood Estimates
Regression Weights:

| Label | Estimate | S.E. | C.R. |
|---|---|---|---|
| LOGWANIM <---- COMINVPA | 0.226 | 0.034 | 6.698 |
| LOGWANIM <---- NWNUMADL | 0.843 | 0.033 | 25.806 |
| LOGNWTPR <---- LOGWANIM | 0.603 | 0.017 | 35.475 |
| LOGNWTPR <----- QUMPEOP | -0.005 | 0.002 | -2.478 |
| LOGNWTPR <------ YEAR99 | 0.127 | 0.037 | 3.437 |
| LOGNWTPR <----- NUMPEOP | 0.089 | 0.028 | 3.136 |
| LGWTOTSL <---- LOGNWTPR | 1.256 | 0.022 | 56.412 |
| LGWTOTSL <------ YEAR99 | 0.196 | 0.043 | 4.587 |
| LGWTOTSL <---- COMINVPA | 0.031 | 0.019 | 1.679 |

| Standardized Regression Weights: | Estimate |
|---|---|
| LOGWANIM <---- COMINVPA | 0.151 |
| LOGWANIM <---- NWNUMADL | 0.586 |
| LOGNWTPR <---- LOGWANIM | 0.888 |
| LOGNWTPR <----- QUMPEOP | -0.143 |
| LOGNWTPR <------ YEAR99 | 0.053 |

| | | |
|---|---|---|
| LOGNWTPR <----- NUMPEOP | 0.184 | |
| LGWTOTSL <---- LOGNWTPR | 0.967 | |
| LGWTOTSL <------ YEAR99 | 0.063 | |
| LGWTOTSL <---- COMINVPA | 0.024 | |

| Means: | Label | Estimate | S.E. | C.R. |
|---|---|---|---|---|
| | NWNUMADL | 2.153 | 0.033 | 65.879 |
| | COMINVPA | 3.614 | 0.031 | 115.266 |
| | NUMPEOP | 5.569 | 0.066 | 84.299 |
| | QUMPEOP | 36.531 | 0.865 | 42.237 |
| | YEAR99 | 0.333 | 0.013 | 25.150 |

| Intercepts: | Label | Estimate | S.E. | C.R. |
|---|---|---|---|---|
| | LOGWANIM | 2.638 | 0.127 | 20.802 |
| | LOGNWTPR | 5.037 | 0.109 | 46.278 |
| | LGWTOTSL | -3.388 | 0.179 | -18.918 |

| Covariances: | Label | Estimate | S.E. | C.R. |
|---|---|---|---|---|
| COMINVPA <---> NWNUMADL | | 0.390 | 0.038 | 10.245 |
| NUMPEOP <----> COMINVPA | | 0.647 | 0.076 | 8.527 |
| QUMPEOP <-----> NUMPEOP | | 69.821 | 2.825 | 24.711 |
| QUMPEOP <------> YEAR99 | | -1.424 | 0.410 | -3.476 |
| YEAR99 <-----> NWNUMADL | | 0.006 | 0.015 | 0.406 |
| QUMPEOP <----> NWNUMADL | | 9.079 | 1.037 | 8.753 |
| QUMPEOP <----> COMINVPA | | 7.342 | 0.986 | 7.445 |
| YEAR99 <-----> COMINVPA | | -0.062 | 0.015 | -4.140 |
| NUMPEOP <----> NWNUMADL | | 0.795 | 0.080 | 9.937 |
| NUMPEOP <------> YEAR99 | | -0.103 | 0.031 | -3.308 |
| error2 <-------> error1 | | -0.114 | 0.035 | -3.246 |
| error3 <-------> error2 | | -0.107 | 0.015 | -7.181 |

| Correlations: | Estimate |
|---|---|
| COMINVPA <---> NWNUMADL | 0.301 |
| NUMPEOP <----> COMINVPA | 0.247 |
| QUMPEOP <-----> NUMPEOP | 0.966 |
| QUMPEOP <------> YEAR99 | -0.098 |
| YEAR99 <-----> NWNUMADL | 0.011 |
| QUMPEOP <----> NWNUMADL | 0.254 |
| QUMPEOP <----> COMINVPA | 0.214 |
| YEAR99 <-----> COMINVPA | -0.117 |
| NUMPEOP <----> NWNUMADL | 0.291 |
| NUMPEOP <------> YEAR99 | -0.093 |
| error2 <-------> error1 | -0.146 |
| error3 <-------> error2 | -0.291 |

| Variances: | Label | Estimate | S.E. | C.R. |
|---|---|---|---|---|
| | COMINVPA | 1.243 | 0.049 | 25.150 |
| | NWNUMADL | 1.351 | 0.054 | 25.150 |
| | error1 | 1.618 | 0.066 | 24.572 |
| | QUMPEOP | 946.306 | 37.627 | 25.150 |
| | NUMPEOP | 5.520 | 0.219 | 25.150 |
| | YEAR99 | 0.222 | 0.009 | 25.150 |
| | error2 | 0.374 | 0.016 | 23.711 |
| | error3 | 0.362 | 0.018 | 20.376 |

| Squared Multiple Correlations: | Estimate |
|---|---|
| LOGWANIM | 0.420 |
| LOGNWTPR | 0.709 |
| LGWTOTSL | 0.833 |

Summary of models

| Model | NPAR | CMIN | DF | P | CMIN/DF |
|---|---|---|---|---|---|
| Default model | 37 | 9.746 | 7 | 0.203 | 1.392 |
| Saturated model | 44 | 0.000 | 0 | | |
| Independence model | 8 | 27659.887 | 36 | 0.000 | 768.330 |

| Model | DELTA1 NFI | RHO1 RFI | DELTA2 IFI | RHO2 TLI | CFI |
|---|---|---|---|---|---|
| Default model | 1.000 | 0.998 | 1.000 | 0.999 | 1.000 |
| Saturated model | 1.000 | 1.000 | | 1.000 | |
| Independence model | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |

| Model | PRATIO | PNFI | PCFI |
|---|---|---|---|
| Default model | 0.194 | 0.194 | 0.194 |
| Saturated model | 0.000 | 0.000 | 0.000 |
| Independence model | 1.000 | 0.000 | 0.000 |

| Model | NCP | LO 90 | HI 90 |
|---|---|---|---|
| Default model | 2.746 | 0.000 | 15.182 |
| Saturated model | 0.000 | 0.000 | 0.000 |
| Independence model | 27623.887 | 27080.089 | 28173.962 |

| Model | FMIN | F0 | LO 90 | HI 90 |
|---|---|---|---|---|
| Default model | 0.008 | 0.002 | 0.000 | 0.012 |
| Saturated model | 0.000 | 0.000 | 0.000 | 0.000 |
| Independence model | 21.866 | 21.837 | 21.407 | 22.272 |

| Model | RMSEA | LO 90 | HI 90 | PCLOSE |
|---|---|---|---|---|

| | | | | |
|---|---|---|---|---|
| Default model | 0.018 | 0.000 | 0.041 | 0.991 |
| Independence model | 0.779 | 0.771 | 0.787 | 0.000 |

| Model | AIC | BCC | BIC | CAIC |
|---|---|---|---|---|
| Default model | 83.746 | 84.276 | | |
| Saturated model | 88.000 | 88.631 | | |
| Independence model | 27675.887 | 27676.002 | | |

| Model | ECVI | LO 90 | HI 90 | MECVI |
|---|---|---|---|---|
| Default model | 0.066 | 0.064 | 0.076 | 0.067 |
| Saturated model | 0.070 | 0.070 | 0.070 | 0.070 |
| Independence model | 21.878 | 21.448 | 22.313 | 21.878 |

| Model | HOELTER .05 | HOELTER .01 |
|---|---|---|
| Default model | 1826 | 2399 |
| Independence model | 3 | 3 |

Execution time summary:

| | |
|---|---|
| Minimization: | 0.220 |
| Miscellaneous: | 0.940 |
| Bootstrap: | 0.000 |
| Total: | 1.160 |

Приложение 7.

# Содержание окна справки (Help)

| | |
|---|---|
| Help Contents | Содержание справки |
| Select a topic below.<br>Arrows indicate common tasks | Выбор нижеследующих тем(подтем).<br>Стрелки, указывающие общие задачи |
| General Information | Общая информация |
| Help on SPSS menu commands | Справка по командам меню SPSS |
| Getting Help in SPSS<br>Searching Help for a topic | Получение помощи в SPSS<br>Поисковая помощь для темы |
| Problems | Проблемы |
| Something isn't working<br>Frequently asked questions | Что-то не работает<br>Часто задаваемые вопросы |
| SPSS windows | Окна SPSS |
| The Data Editor<br>Output windows<br>The Chart Carousel<br>Chart windows<br>Syntax windows | Редактор данных<br>Окна вывода<br>Карусель диаграмм<br>Окна диаграмм<br>Окна синтакса |
| Operations | Действия |
| The Toolbar<br>Using the menus<br>Using the dialog boxes<br>The SPSS command language<br>OLE,<br>DDE,<br>SPSS API and<br>Drag & Drop | Панель инструментов<br>Использование меню<br>Использование диалоговых окон<br>Командный язык SPSS<br>Связь и внедрение объектов<br>Динамический обмен данными<br>Программный интерфейс при-<br>ложений SPSS<br>Способ переноса “тащить и<br>бросить” |
| Data and files | Данные и файлы |
| The Data Editor | Редактор данных |
| Bringing data into SPSS<br>Defining variables<br>Entering data<br>Editing data values | Внесение данных в SPSS<br>Определение переменных<br>Ввод данных<br>Редактирование значений данных |

Приложение 8.

# Статистики среднего в SPSS 10.0

(дополнительное диалоговое окно Options в процедуре Means)

**Статистики среднего, заданные по умолчанию**

1. Название русское: **Арифметическое среднее**
   Название английское: **Mean**
   Краткое определение: Частное от деления суммы всех наблюдений ($\sum x$) на их число n
   Математическая формула: $\overline{N} = \sum x/n$

2. Название русское: **Число случаев**
   Название английское: **Number of cases**
   Краткое определение: Число случаев ряда распределения
   Обозначение:: n

3. Название русское: **Стандартное отклонение**
   Название английское: **Standard Deviation**
   Краткое определение: Показатель величины отличия наблюдений от среднего, выраженный в тех же единицах, что и измеряемая величина
   Математическая формула: $sd_x=\sqrt{\Sigma x^2/n}$

**Статистики среднего, которые можно задать дополнительно**

4. Название русское: **Медиана**
   Название английское: **Median**
   Краткое определение: Значение признака, находящегося в середине ряда распределения.
   Обозначение: Me

5. Название русское: **Медиана интервального ряда**
   Название английское: **Grouped median**
   Краткое определение: Значение медианы в медианном интервале ряда распределения

6. Название русское: **Стандартная ошибка среднего**
   Название английское: **Standard error of the mean**
   Краткое определение: Стандартная мера оценки того, насколько результат выборки отличается от действительных фактов генеральной совокупности вследствие

случайных колебаний выборочных характеристик.

Математическая формула: $S_{Mx} = sd_x/\sqrt{n}$

7. Название русское: **Сумма**
   Название английское: **Sum**
   Краткое определение: Сумма значений переменной

   Обозначение: $\sum x$

8. Название русское: **Минимум**
   Название английское: **Minimum**
   Краткое определение: Минимальное значение переменной

   Обозначение: Min

9. Название русское: **Максимум**
   Название английское: **Maximum**
   Краткое определение: Максимальное значение переменной

   Обозначение: Max

10. Название русское: **Первое значение переменной**
    Название английское: **First**

11. Название русское: **Последнее значение переменной**

    Название английское: **Last**

12. Название русское: **Дисперсия** (разброс признака)
    Название английское: **Variance**
    Краткое определение: Показатель величины отличия наблюдений от среднего. Значение дисперсии равно квадрату стандартного отклонения

    Математическая формула: $\sigma = sd_x^2$

13. Название русское: **Выборочный коэффициент эксцесса**

    Название английское: **Kurtosis**
    Краткое определение: Показатель степени концентрации наблюдений вокруг центральной точки

    Математическая формула: $\gamma_i=\eta_i/\eta_j^2$

14. Название русское: **Стандартная ошибка**

| | | |
|---|---|---|
| | | **эксцесса** |
| | Название английское: | **Standard error of kurtosis** |
| 15. | Название русское: | **Выборочный коэффициент асимметрии** |
| | Название английское: | **Skewness** |
| | Краткое определение: | Показатель степени несимметричности распределения |
| | Математическая формула: | $\gamma_i=\eta_i/\eta_j^{3/2}$ |
| 16. | Название русское: | **Стандартная ошибка коэффициента асимметрии** |
| | Название английское: | **Standard error of skewness** |
| 17. | Название русское: | **Средняя гармоническая** |
| | Название английское: | **Harmonic mean** |
| | Краткое определение: | Величина, обратная средней арифметической из обратных значений признака |
| | Обозначение: | H |
| 18. | Название русское: | **Процент от общей суммы** |
| | Название английское | **Percentage of total sum** |
| 19. | Название русское: | **Процент от общего числа случаев** |
| | Название английское: | **Percentage of total N** |

Приложение 9.

## Список файлов с данными (.sav), используемыми в пособии

1. Три волны панели 1995-1996-1997 гг. Объем выборки 463 домохозяйства. Основной файл: pandata_95-96-97.
2. Три волны панели 1995-1996-1997 гг. Индивидуальный файл для структуры семьи. Число случаев 3704. Файл создан дополнительно к основному: individual_data_panel_95-96-97.
3. Три волны панели 1995-1996-1997 гг. Дополнительный файл для динамического и структурного анализа: pooleddata_95-96-97.
4. Три волны панели 1995-1997-1999 гг. Объем выборки 422 домохозяйства. Основной файл: pandata_95-97-99.
5. Четыре волны панели 1995-1997-1999-2003 гг. Объем выборки 382 домохозяйства. Основной файл: pandata_95-97-99-03.
6. Разовое обследование 2001 г. Объем выборки 800 домохозяйств. Основной файл: winlose_01.

Приложение 10.

## Полезные сетевые адреса

Англо-русский словарь терминов и функций программы SPSS. http://www.marketresearch.ru/spssdoc/voc.doc

База данных панели 1995-1996-1997 гг. http://www.icpsr.umich.edu/cgi-in/archive.prl?path=ICPSR&num=286

Булева алгебра: http://www.icsti.su/ibd/Sart2.asp?T1=BBB

Домашняя страница рабочего места авторов в ИСЭПН РАН. http://www.isesp-ras.ru/labinfra.htm

Домашняя страница авторов. http://host.iatp.ru/~patsiorkovsky

Использование SPSS для анализа социологической информации. http://socionet.narod.ru/CHART3-1.zip

Источники информации по статистическому анализу данных. http://www.unn.ru/rus/f14/k2/courses/borisova/sources.htm

Ноздряков Р. Количественные методы анализа (дистанционное обучение). http://vle3.projectharmony.ru/dlscripts/dlmanage.exe

Регрессионный анализ. http://www.kgafk.ru/kgufk/html/uchmetrologia9.html

Соколова М.И, Гречков В.Ю. Маркетинговые исследования. Учебник.- М.: МГИМО (У) МИД РФ. http://www.marketing.spb.ru/lib-research/sokol

Статистический портал. http://www.statsoft.ru/home

Текстовый файл описания базы данных панели 1995-1996-1997 гг. http://www.icpsr.umich.edu/cgi/ab.prl?file=2816 (англ. язык)

Теория множеств. http://www.krugosvet.ru/articles/15/1001551/1001551a1.htm

Теория факторного анализа. http://intecs.ur.ru/page.php?handle=service.faktoranalizteory

# ГЛОССАРИЙ

Анализ (Analyze) – раздел главного меню, позволяющий выполнять статистические расчеты.
База данных - файлы с данными.
Вид (View) – раздел главного меню, позволяющий изменять вид редактора данных.
Главное диалоговое окно - устройство, позволяющее выполнять расчеты без написания формул.
Главное меню – устройство, открывающее доступ выполнения различных команд и процедур.
Графики (Graphs) - раздел главного меню, позволяющий строить графики.
Данные (Data) - раздел главного меню, позволяющий агрегировать, сортировать данные, делить и объединять файлы с данными.
Диаграмма (Chart) - одно из дополнительных диалоговых окон.
Дополнительное диалоговое окно - устройство, позволяющее учитывать специфику выполняемых расчетов.
Команда (Command) – инструкция, на основе которой системой выполняются все требуемые действия. Базовый элемент командного языка – «синтаксиса» и интерфейса SPSS
Кнопки (выключатели) – устройства, позволяющие открывать дополнительные диалоговые окна и выполнять команды.
Окно (Window) - раздел главного меню, позволяющий сворачивать открытое окно рабочего файла.
Окно просмотра (Viewer) - окно, в котором выводятся результаты выполняемых расчетов. В ранних версиях его эквивалент – окно вывода (Output).
Панель инструментов (Toolbar) – набор кнопок, дублирующих выполнение наиболее часто используемых команд.
Переменная (Variable) - столбец таблицы редактора данных (индикатор, характеристика наблюдения-случая).
Преобразование (Transform) - раздел главного меню, позволяющий создавать новые переменные.
Пропущенные значения (Missing Values) – пустые значения, возникшие в результате отсутствия данных или назначенные пользователем в качестве таковых.

Процедура – совокупность команд.

Рабочий файл – файл с открытыми на каждый данный момент данными.
Редактор данных (Data Editor) – основное устройство в формате таблицы, позволяющее работать с данными.
Редакторование (Edit) - раздел главного меню, позволяющий вставлять, копировать, искать данные и изменять настройки редактора данных.
Синтаксис (Syntax) – командный язык SPSS.
Случай (Case) – строка таблицы редактора данных (формализованная запись наблюдения).
Справка (Help) – раздел главного меню, позволяющий получить справку по SPSS.
Статистики (Statistics) – одно из дополнительных диалоговых окон.
Утилиты (Utilities) - раздел главного меню, позволяющий получать различную сервисную информацию о рабочем файле.
Файл (File) - раздел главного меню, позволяющий создавать, открывать и сохранять файлы с данными.
Формат (Format) – одно из дополнительных диалоговых окон.
Ячейка таблицы – пересечение случая и переменной в таблице редактора данных SPSS.

# ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

# INDEX

# СОДЕРЖАНИЕ

ББК.С5в63я73-1

Учебное пособие
**SPSS ДЛЯ СОЦИОЛОГОВ**

Авторы:
**Пациорковский Валерий Валентинович,**
**Пациорковская Валентина Викторовна**

Редакторы:
**А.В.Пациорковский** – научный редактор
**Г.С.Сизова** – литературный редактор

Компьютерная верстка и оригинал-макет **В.В.Вдовенко**
Оформление художника **В.В.Коробановой**



РИЦ ИСЭПН РАН
117218, Москва. Нахимовский пр-т, д. 32.
Тел.: (095) 125-73-02, Факс: (095) 129-08-01

Подписано в печать 1.04.2005 г.
Бумага офсетная. Формат 60 x 90 / 16
Печ. л. 26. Тираж 1000 экз.

Отпечатано в ООО «Доминант»
Зак. № 22 Тир. 1000 экз.

Цена договорная

Об авторах:

**Пациорковская Валентина Викторовна** - старший научный сотрудник ИСЭПН РАН, руководитель полевых работ и обработки данных российско-американского исследовательского проекта: «Изменение условий жизни сельского населения России в 1991-2005 гг.». Инженер-экономист, выпускница Государственного университета управления им. С.Орджоникидзе (1972 г.). Автор ряда научных работ. Сфера интересов: методы сбора и обработки социологической информации.

*«Работа в SPSS помогла мне лучше понять, как неукоснительно в социальных отношениях выполняется закон больших чисел».*

*В.В.Пациорковская*

**Пациорковский Валерий Валентинович** - д.э.н., профессор, зав.лаб. ИСЭПН РАН, один из создателей и руководитель с российской стороны упомянутого выше российско-американского исследовательского проекта. Выпускник философского факультета МГУ им. М.В.Ломоносова (1969 г.). Кандидатскую диссертацию защитил в Институте социологии РАН в 1975 г., докторскую - в ИСЭПН РАН в 1994 г. Автор более 150 публикаций. Сфера интересов: социальная инфраструктура, условия и качество жизни населения, методология социологических исследований.

*«Знакомство с SPSS помогло мне реализовать логические способности, которые до этого выражались лишь в любви к шахматам».*

*В.В.Пациорковский*

Электронный адрес: patsv@mail.ru

Сетевой адрес:
http://www.isesp-ras.ru/labinfra.htm