

Министерство образования и науки Российской Федерации  
АМУРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

**ОСНОВЫ ПРИМЕНЕНИЯ  
SPSS  
В СОЦИОЛОГИИ**

Благовещенск  
Издательство АмГУ  
2020

ББК 60.5  
Л 47

*Печатается по решению  
редакционно-издательского совета  
Амурского государственного  
университета*

*Рецензенты:* Н.Ю. Щека, *доцент, канд. социол. наук*  
Н.В. Болтенков, *доцент ДальГАУ, канд. пед. наук*

Леонов, А.К.

Основы применения SPSS в социологии : учеб. пособие. 2-е изд., перераб. и дополн. / А.К. Леонов. – Благовещенск : Изд-во Амур. гос. ун-та, 2020. – 162 с.

Учебное пособие включает изложение основных принципов работы в прикладной статистической программе SPSS, а также комплекс практических заданий и контролирующие материалы, и предназначена для студентов, обучающихся по направлению подготовки 39.03.01 – «Социология».

© Амурский государственный университет, 2020  
© Леонов А.К., 2020

## СОДЕРЖАНИЕ

<b>Введение</b>	4
<b>Глава 1. Подготовительный этап обработки данных</b>	6
1.1. Программный комплекс SPSS. Общий обзор	6
1.2. Подготовка первичных данных. Работа с анкетой	13
1.3. Матрица данных SPSS. Файлы данных	19
1.4. Редактор данных	23
1.5. Оценка и контроль качества первичных данных	34
1.6. Тестирование качества базы данных. Чистка массива данных	37
1.7. Сервисные процедуры с данными	42
1.8. Модификация данных: Перекодирование переменных	53
1.9. Модификация данных: Вычисление переменных	60
1.10. Модификация данных: Дополнительные процедуры	64
<b>Глава 2. Этап статистической обработки данных</b>	73
2.1. Основы статистики и исследование данных	73
2.2. Статистические методы исследования данных	83
2.3. Частотный анализ	87
2.4. Статистические характеристики	96
2.5. Таблицы сопряженности	108
2.6. Обработка множественных ответов	117
2.7. Графики	124
2.8. Редактирование таблиц. Модуль Tables	131
2.9. Многомерные методы анализа данных: обзор	137
<b>Промежуточные проверочные работы и кроссворды к ним</b>	143
<b>Итоговый тест</b>	148
<b>Рекомендуемая литература</b>	155
<b>Приложение. Памятка по работе в SPSS</b>	157

## ВВЕДЕНИЕ

Целями данного учебного пособия является изучение основ наиболее популярной в социологии прикладной программы статистической обработки и анализа эмпирических данных – SPSS (Statistical Package for the Social Science), формирование у студентов практических навыков использования универсальной системы SPSS в конкретных социологических исследованиях. Учебное пособие может использоваться в числе базовой литературы при подготовке лекций и проведении практических занятий по дисциплине «Основы применения прикладных статистических программ в социологических исследованиях».

Данный учебный курс представляет собой последовательное изложение базовых функциональных возможностей системы статистического анализа и управления данными – SPSS, включающее в себя следующие задачи:

- подготовка матрицы базы данных (БД) – создание шаблона для ввода первичной информации;
- формирование БД, импорт, экспорт, агрегирование данных;
- редактирование внесенных данных; выполнение сервисных процедур с данными;
- модифицирование исходной БД, трансформация структуры БД;
- работа с данными: проведение первичных расчетов, получение дескриптивных статистик;
- подготовка и выполнение операций с множественными вопросами;
- построение таблиц сопряженности, расчет коэффициентов статистической связи;
- формирование выдачи: работа с таблицами и графиками.

Рекомендуется, чтобы в ходе семинарских занятий студенты с помощью пакета программ SPSS отрабатывали полученные на лекции знания и тем самым реализовывали полный цикл статистических и сервисных процедур, используемых в реальных эмпирических исследованиях.

В учебном пособии приведены снимки экрана основных процедур работы в статистическом пакете IBM SPSS Statistics 21.0. Однако практически весь материал может быть применен и к более ранним версиям. Основные отличия будут иметь место только в интерфейсе. Каждый параграф пособия имеет следующую структуру: описание работы с процедурами (включая алгоритмы и снимки экрана основных процедур), контрольные вопросы по лекции, практические задания. Также в книге приводятся две промежуточные проверочные работы, включающие списки вопросов и кроссворды, итоговый тест (два варианта), рекомендуемая литература для более глубокого изучения предмета, краткая памятка по работе с SPSS, содержащая обзор основных процедур и алгоритмов.

Отличительными чертами данного учебного пособия являются: наличие указанных дидактических материалов, направленных на проверку теоретических знаний и формирование практических навыков работы студентов в программном комплексе; адаптация к применению SPSS при анализе данных, полученных в ходе социологических исследований; демонстрация возможностей одной из последних версий рассматриваемого программного обеспечения.

## ГЛАВА 1. ПОДГОТОВИТЕЛЬНЫЙ ЭТАП ОБРАБОТКИ ДАННЫХ

### 1.1. Программный комплекс SPSS. Общий обзор

Внедрение компьютеров буквально во все сферы человеческой деятельности является на сегодняшний день, наверное, самым очевидным итогом научного прогресса. И как следовало ожидать, в существенной степени компьютеризация изменила характер самих научных исследований, в том числе в психологии и социальных науках. Компьютер обычно применяется исследователями для выполнения такой работы, которая считается самой скучной и утомительной: учет и организация исходных данных, вычисления различных показателей и пр. Это позволяет исследователю проводить более глубокий анализ данных, больше времени уделять интерпретации и выдвигению новых предположений, то есть заниматься тем, что обычно считается самым приятным и интересным в любом исследовании, но что остается за пределами возможностей компьютера. Данный курс является основой по применению компьютерной программы SPSS – очень мощного и широко распространенного средства компьютерного анализа данных в психологии и социальных науках. SPSS – это аббревиатура от Statistical Package for the Social Science (статистический пакет для социальных наук). Как следует из названия, SPSS представляет собой множество различных программ, предназначенных для анализа данных в социальных науках. Эти программы позволяют организовывать ввод данных, гибко менять их структуру, применять к ним самые современные методы обработки или их последовательность и получать результаты в удобной и наглядной форме. Все это множество программ объединено в единую систему, обеспечивающую простой и дружественный диалог с исследователем и снабженную исчерпывающей справочной поддержкой. Благодаря такой дружественности система SPSS легко

доступна для освоения даже тем, кто имеет минимальные навыки владения компьютером<sup>1</sup>.

### **Немного истории.**

Два студента Норман Най (Norman Nie) и Дейл Вент (Dale Bent), специализировавшиеся в области политологии в 1965 году пытались отыскать в Стенфордском университете Сан-Франциско компьютерную программу, подходящую для анализа статистической информации. Вскоре они разочаровались в своих попытках, так как имеющиеся программы оказывались более или менее непригодными, неудачно построенными или не обеспечивали наглядность представления обработанной информации. К тому же принципы пользования менялись от программы к программе.

Так, не долго думая, они решили разработать собственную программу, со своей концепцией и единым синтаксисом. В их распоряжении тогда был язык программирования FORTRAN и вычислительная машина типа IBM 7090. Уже через год была разработана первая версия программы, которая, еще через год, в 1967, могла работать на IBM 360. К этому времени к группе разработчиков присоединился Хэдлай Халл (Hadlai Hull).

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

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

Спустя пять лет SPSS была уже инсталлирована шестьсот раз, причём под разными операционными системами. С самого начала версиям программы

---

<sup>1</sup> Наследов А.Д. SPSS : Компьютерный анализ в психологии и социальных науках : самоучитель / А.Д. Наследов. – СПб. : Питер, 2005. – С. 12

присваивали соответствующие порядковые номера. В 1975 была разработана уже шестая версия (SPSS6). До 1981 последовали версии 7, 8 и 9.

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

Первая версия SPSS для Windows имела порядковый номер 5. Затем последовали версии 6.0 и 6.1 с некоторыми нововведениями в статистической и графической областях; версия 6.1 была первой статистической программой для Windows, которая использовала 32 битную архитектуру Windows 3.1. Это можно было заметить по более высокой скорости выполнения вычислений. Усовершенствования коснулись также и интерфейса пользователя. В конце концов, была выпущена версия 6.1.3, которая уже могла работать и под Windows 95 и под NT.

В начале 1996 года появилась 7-я версия SPSS, сначала как версия 7.0, а затем 7.5. Наряду с расширением возможностей в сфере статистики, разница между этими двумя версиями заключалась в том, что в версии 7.5 как меню, так и интерфейс программы были выполнены уже не только на английском, но и на других наиболее распространенных языках.

Самым весомым отличием версии 7 по отношению к предыдущим версиям, был абсолютно новый подход к выводу информации на экран. Так, во первых, получил новые очертания так называемый Viewer (Окно просмотра), и, во вторых, более приятный внешний вид приобрели таблицы результатов расчётов (мобильные таблицы). Появившаяся технология мобильных таблиц позволяет перестраивать полученные таблицы различным способами.

За версией 7.5 последовала версия 8.0, прогресс которой заключался в усовершенствовании графической оболочки. Возможность составления интерактивных графиков предоставляет ряд преимуществ по сравнению с



традиционными графиками, которые являются стандартом для многих других пакетов.

Версия 9.0 включала в себя несколько новых статистических методов, в т.ч. многозначную логистическую регрессию, и несколько новых графических возможностей, расширяющих область интерактивных графиков.

Версия, описываемая в этом пособии имеет порядковый номер 21.0.

### **Модули программы.**

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

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

Наряду с тремя упомянутыми, существует еще ряд специальных дополнительных модулей и самостоятельных программ, число которых постоянно растёт, так что пользователям следует постоянно знакомиться с информацией о нововведениях в SPSS.

В этой книге описываются базисный модуль, а также модули Regression Models, Advanced Models и Tables. Назначением последнего модуля является составление презентационных таблиц. В книге не рассматриваются лог-линейные модели, анализ выживания и многомерное шкалирование, а также процедура составления презентаций<sup>2</sup>.

**Анализ данных с применением компьютера включает выполнение ряда необходимых шагов.**

1. Определение структуры данных.

---

<sup>2</sup> Бююль А., Цёфель П. SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей : Пер. с нем. / А. Бююль, П. Цёфель. – СПб.: ООО «ДиаСофтЮП», 2005. – С. 14–16

2. Ввод данных в компьютер в соответствии с их структурой и требованиями программы.

3. Задание метода обработки данных в соответствии с задачами исследования.

4. Получение результата обработки данных.

5. Интерпретация результата обработки.

Шаги 1 (подготовительный) и 5 (заключительный) не способна выполнить ни одна компьютерная программа – их исследователь делает сам. Помощь компьютера (шаги 2-4) заключается, в конечном итоге, в переходе от длинной последовательности чисел к более компактной их последовательности. На «вход» компьютера исследователь подает массив исходных данных, который недоступен осмыслению, но пригоден для компьютерной обработки (шаг 2). Затем исследователь дает программе команду на обработку данных в соответствии с задачей исследования структурой данных (шаг 3). На «выходе» исследователь получает результат обработки (шаг 4) – тоже массив данных, только меньший, доступный осмыслению и содержательной интерпретации. При этом исчерпывающий анализ данных обычно требует многократной их обработки с применением разных методов. По всем параметрам SPSS for Windows является сложным и мощным статистическим пакетом. Однако несмотря на сложность, средства взаимодействия входящих в пакет программ с пользователем являются весьма дружественными. С помощью пакета SPSS можно проводить практически любой анализ данных, известный в социальных науках, а последние версии программы находят применение в самых разных научных областях и в мире бизнеса<sup>3</sup>.

### **Запуск программы**

Пожалуй, основное, что вам необходимо (по крайней мере в рамках этой книги), – знать, как запустить программу SPSS. Для этого на большинстве компьютеров вам следует, щелкнув на кнопке Пуск, последовательно

---

<sup>3</sup> Наследов А.Д. SPSS : Компьютерный анализ в психологии и социальных науках : самоучитель / А.Д. Наследов. – СПб. : Питер, 2005. – С. 13–14

переместить указатель мыши сначала на пункт Программы, а затем на пункт SPSS for Windows. Когда последний окажется выделенным, раскроется программная группа пакета. Щелкните мышью на команде SPSS for Windows. После этого программа SPSS запустится. Если значок SPSS находится на рабочем столе, для запуска программы достаточно дважды щелкнуть на этом значке. Помимо запуска SPSS вы также должны уметь переключаться между окнами программ с помощью панели задач. Для работы с пакетом SPSS это весьма актуально, поскольку он представляет собой набор из нескольких одновременно выполняющихся программ, имеющих собственные окна. При запуске SPSS вы видите на экране единственное окно редактора данных, однако как только вы займетесь статистическим анализом ваших данных, на экране также появится окно вывода (рис. 1). В зависимости от выполняемых вами действий число открытых окон может меняться, однако окна редактора данных и вывода, как правило, присутствуют на экране постоянно. Иногда пакет SPSS переключается между окнами автоматически, в остальных случаях вам потребуется осуществлять переключение вручную. Когда несколько программ выполняются одновременно, каждой из них соответствует кнопка на панели задач.

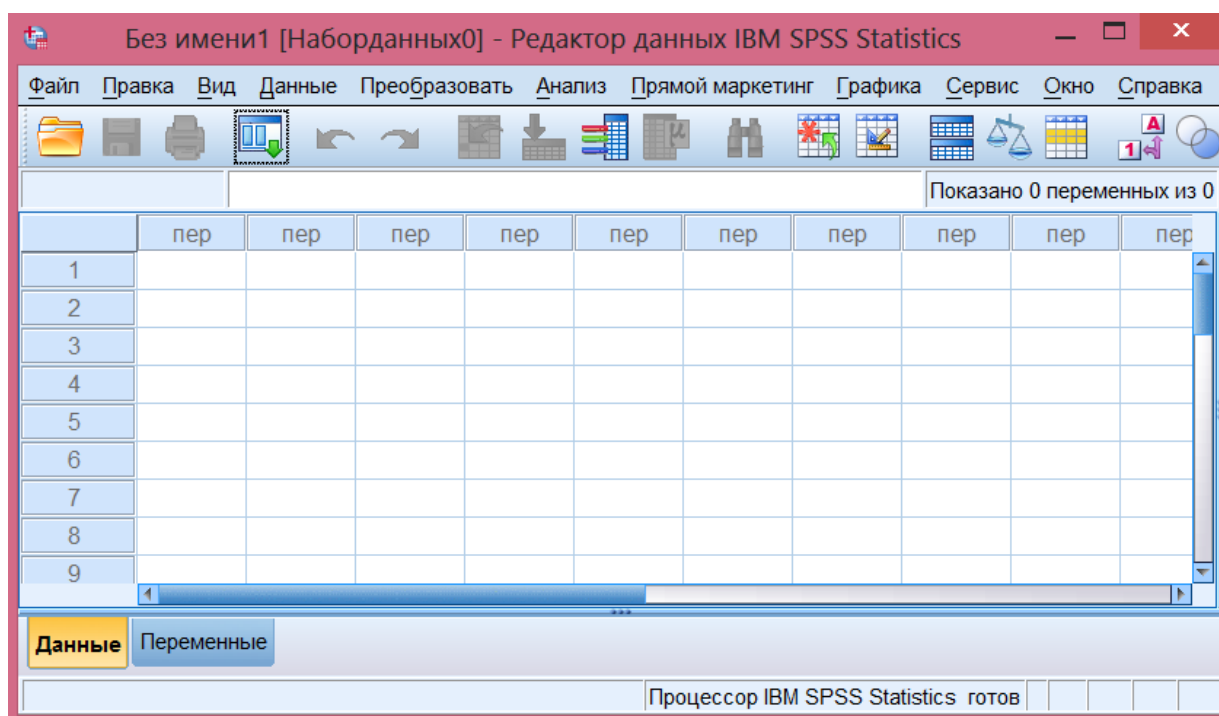


Рис. 1

## **Кнопки.**

Кнопки используются практически в каждом окне. Они выглядят по-разному и могут находиться как внутри окна, так и в его строке заголовка. Есть несколько стандартных кнопок, которые выполняют одну и ту же функцию независимо оттого, окну какой программы они принадлежат.

## **Окна.**

Существует 5 типов окон, которые при работе с пакетом SPSS используются чаще других.

- Окно приветствия, предназначенное для оперативного доступа к имеющимся базам данных, встроенному электронному учебнику, запросам и т.д.

- Главное окно программы появляется при запуске SPSS, с него начинается работа с пакетом.

- Диалоговое окно *Открыть данные* позволяет получить доступ к ранее созданным файлам.

- Базовые диалоговые окна хотя и зависят от конкретной процедуры, имеют схожие элементы интерфейса.

- Помимо этих трех типов окон особое значение имеет окно вывода. Окно вывода появляется каждый раз после окончания обработки данных. Оно содержит результаты обработки, а также краткие пояснения по их интерпретации<sup>4</sup>.

## **Использованная литература**

Бююль А., Цёфель П. SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей : Пер. с нем. / А. Бююль, П. Цёфель. – СПб.: ООО «ДиаСофтЮП», 2005. – 608 с.

Наследов А.Д. SPSS : Компьютерный анализ в психологии и социальных науках : самоучитель / А.Д. Наследов. – СПб. : Питер, 2005. – 416 с.

---

<sup>4</sup> Там же. С. 22–23

## **Контрольные вопросы**

1. Кратко опишите Историю разработки SPSS, его Цель и функции
2. Назовите основные Модули SPSS
3. Опишите процесс Инсталляции SPSS

## **Практические задания**

1. Установите программный комплекс SPSS на свой компьютер.
2. Запустите SPSS любым удобным способом (через меню «Пуск» или воспользовавшись ярлыком SPSS на рабочем столе).
3. Выберите в окне приветствия пункт «Ввести данные» и откройте главное окно программы.
4. В настройках программы установите русский язык интерфейса (Edit ► Options).
5. С помощью окна «Открытие файла» (Open file) откройте файл данных «куратор 2011», расположенный в папке «базы ЦСИ АмГУ».

## **1.2. Подготовка первичных данных. Работа с анкетой**

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 пакета SPSS.

Основной смысл подготовительного 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. Эти числа и использовались интервьюерами при фиксации ответов респондентов.

Следующий важный шаг – присвоение каждому вопросу анкеты восьмисимвольного смыслового имени. Причем первым символом имени должна быть буква. Эти имена и становятся именами переменных с момента их введения в систему.

***Переменная – это вопрос анкеты и набор ответов (их кодов) к нему.***

В результате выполнения рассматриваемого шага к массиву анкет с первичной информацией добавляется еще один бланк со всем индикаторами, расписанными по смысловым именам – переменным – со всеми возможными в данном исследовании кодами индикатора, а также с указанием размера ячейки (ширины переменной). Форма такого бланка называется «Макет ввода данных в ЭВМ». В других работах сходный по назначению документ называется «**кодировочной таблицей**».

Подытожим задачи, которые решаются при составлении кодировочной таблицы:

1. Кодировочная таблица устанавливает соответствие между отдельным вопросам анкеты и переменными.

2. Кодировочная таблица устанавливает соответствие между возможными значениями переменных и кодовыми числами.

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.

Основной 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 *Переменные* 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. 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. Соответственно, столько же колонок и переменных должно быть явлено в таблице окна редактора данных. Колонки и строки состоят из ячеек. Каждая ячейка представляет собой пересечение случая и переменной. Значение одной переменной записывается в одну ячейку<sup>5</sup>.

### **Типы вопросов и особенности их ввода.**

#### **1. Альтернативные и множественные вопросы.**

Альтернативный вопрос предполагает выбор респондентом только одного варианта ответа:

#### **Пол респондента**

- 1 Мужской
- 2 Женский

#### **Слушаете ли вы радио?**

1. Да
2. Нет

#### **Как Вы оцениваете бытовые условия в общежитии?**

1. Хорошо
2. Удовлетворительно
3. Плохо

Множественный вопрос предполагает выбор нескольких вариантов ответа (более одного). Например:

#### **Какие радиостанции Вы слушаете?**

- 1 Авторадио
- 2 Эхо Москвы
- 3 Love-радио
- 4 Европа-плюс
- 5 Радио-Шансон.

---

<sup>5</sup> Бююль А., Цёфель П. SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей : Пер. с нем. / А. Бююль, П. Цёфель. – СПб.: ООО «ДиаСофтЮП», 2005. – С. 26–27



Респондент может выбрать любое количество вариантов ответа, если это неограниченно в формулировке вопроса:

**Какие радиостанции Вы слушаете наиболее часто (выберите не более двух)?**

2. Закрытые, полузакрытые и открытые вопросы.

В закрытом вопросе жестко задан набор вариантов ответа, из которого респондент выбирает соответствующий своему мнению. Свой вариант ответа респондент предложить не может: его выбор строго ограничен набором вариантов ответа, заданным социологом. Например:

**Ваше образование?**

- |   |                         |   |   |
|---|-------------------------|---|---|
| 1 | Начальное или ниже      | 4 | Среднее специальное (техникум)                  |
| 2 | Неполное среднее        | 5 | Незаконченное высшее (не менее 3-х курсов вуза) |
| 3 | Среднее (школа или ПТУ) | 6 | Высшее  |

В полузакрытом вопросе помимо набора вариантов ответа, заданного социологом, у респондента есть возможность дать собственный вариант ответа:

**Как Вы готовились к поступлению в АмГУ?**

- 1 Самостоятельная подготовка к ЕГЭ
- 2 Репетиторство
- 3 Лицей АмГУ
- 4 Другое (напишите, что именно) \_\_\_\_\_

В данном случае респондент может внести свой вариант, например, «курсы при другом вузе». Варианты ответов, данные респондентами, необходимо закодировать вручную в соответствии с общей кодировкой конкретного вопроса.

Открытый вопрос вообще не имеет готовых вариантов ответа в анкете. Респондент самостоятельно и в произвольной форме отвечает на такой тип вопроса. Например:

**Ваши предложения по совершенствованию и повышению эффективности работы Комбината питания АмГУ:**

---

Ответы на открытый вопрос необходимо закодировать вручную. Рекомендуется это сделать до ввода данных в базу, создав таблицу кодификатора.

### 3. Простые и табличные вопросы.

Простой вопрос представляет собой единичную переменную в отличие от табличного, в котором содержится ряд взаимосвязанных переменных. Например:

**Как вы оцениваете созданные в АмГУ условия для получения доступного и качественного образования?**

	Очень плохо	Плохо	Хорошо	Отлично	<i>Затрудняюсь ответить</i>
<b>1. обеспечение вуза квалифицированными преподавателями</b>	1	2	3	4	5
<b>2. обеспечение эффективного взаимодействия между студентами и преподавателями</b>	1	2	3	4	5
<b>3. поддержание благоприятного психологического климата в вузе</b>	1	2	3	4	5
<b>4. развитие материально-технической базы вуза</b>	1	2	3	4	5
<b>5. оснащение вуза техническими средствами (компьютеры, выход в интернет и др.)</b>	1	2	3	4	5
<b>6. обеспечение поддержки и развития талантливой молодежи</b>	1	2	3	4	5
<b>7. условия быта (общежитие)</b>	1	2	3	4	5
<b>8. обеспечение питанием (столовая, кафетерий, буфеты)</b>	1	2	3	4	5
<b>9. обеспечение здоровья студентов</b>	1	2	3	4	5
<b>10. комплексная безопасность студентов</b>	1	2	3	4	5
<b>11. профилактика девиантного (отклоняющегося) поведения студентов</b>	1	2	3	4	5

Иными словами, табличный вопрос – это совокупность подвопросов.

### Использованная литература

Бююль А., Цёфель П. SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей : Пер. с нем. / А. Бююль, П. Цёфель. – СПб.: ООО «ДиаСофтЮП», 2005. – 608 с.

## Контрольные вопросы

1. Что такое «Переменные» и «Значения переменных»?
2. Чем отличаются Альтернативные и множественные вопросы?
3. Чем отличаются Закрытые, полузакрытые и открытые вопросы?
4. Как закодировать открытые вопросы?
5. Чем отличаются Простые и табличные вопросы?

## Практические задания

1. Найдите в анкете, предложенной преподавателем:
  - а) альтернативные и множественные вопросы
  - б) закрытые, полузакрытые и открытые вопросы
  - в) простые и табличные вопросы
2. Придумайте вопрос следующего типа:
  - а) закрытый альтернативный вопрос
  - б) полузакрытый множественный вопрос
  - в) открытый множественный вопрос
  - г) табличный вопрос
3. Составьте кодировочную таблицу для первых 5 вопросов анкеты, предложенной преподавателем, по образцу ниже:

Номер и формулировка вопроса в анкете	Имя переменной	Все возможные значения переменной или их диапазон
1. Пол респондента	В_1	1, 2
2. Ваш возраст	В_2	От 18 до 74

4. Составьте кодификатор для любого открытого вопроса в анкете.

### 1.3. Матрица данных SPSS. Файлы данных

Этот параграф поможет вам сделать первый шаг к статистическому анализу – ввести данные в программу. Как вы уже знаете из предыдущей главы, в SPSS для управления данными используется специальная программа – редактор данных. Окно редактора данных представляет собой электронную таблицу, в которой столбцы отражают переменные, а строки – объекты

(случаи). На пересечении строки и столбца находится ячейка, в которой сохраняется значение переменной (столбца) для данного объекта (строки). Для работы с таблицами, не уместяющимися в пределах окна, предусмотрены вертикальные и горизонтальные полосы прокрутки, а также автоматические функции перехода к заданной переменной или объекту (рис. 2).

Показано 29 переменных из 29

	V1	V2	V2per	V3	V4	V5_1	V5_2	V5
1	2	3	2	4	1	1	.	.
2	1	1	1	2	0	1	.	.
3	1	1	1	3	1	.	.	.
4	1	1	1	3	1	.	.	.
5	1	1	1	2	0	1	.	.
6	2	1	1	5	0	6	7	.
7	2	1	1	5	2	1	3	.
8	2	1	1	4	2	1	4	.
9	1	1	1	3	2	1	.	.
10	1	1	1	4	2	1	.	.

Процессор IBM SPSS Statistics готов

Рис. 2

Перед тем как начать ввод данных, необходимо определить структуру будущего файла. Для этого вы должны ответить себе на вопрос, как будут использоваться в анализах те или иные переменные. К сожалению, многие пользователи начинают задумываться об этом гораздо позже, чем следует, и в этом кроется причина многих неудач при проведении анализа данных: чем сложнее план исследования, тем больше шансов, что из-за необдуманных действий исследователя он завершится неудачей. Итак, первое, что следует сделать, – определить последовательность действий при обработке данных. Кроме того, необходимо четко представлять себе структуру и взаимосвязи переменных в вашем плане исследования. Ниже перечислены наиболее характерные для файлов данных ошибки и недостатки.

- Отсутствуют ключевые переменные (пол, возраст и т. п.), являющиеся основой для анализа.

- Переменная плохо отражает содержание соответствующей реальной величины (например, на сложный вопрос имеются только два варианта ответа: «да» и «нет»).

- При большом количестве независимых переменных отсутствуют зависимые переменные, отражающие цель исследования (или наоборот).

- Недостаточно независимых переменных, влияющих на заданную зависимую переменную.

Эти примеры демонстрируют, что залогом успеха как исследования в целом, так и создания файла данных в частности является тщательно продуманный выбор структуры данных. Следует отметить, что структура файла данных должна соответствовать плану исследования. С другой стороны, план исследования должен быть составлен так, чтобы его исходные данные можно было бы обработать в соответствии с задачами и гипотезами исследования. Самый оптимальный и простой путь обеспечения этих соответствий – определение структуры данных на этапе планирования исследования, еще до сбора данных. Это позволит избежать большинства типичных ошибок, относящихся как к планированию исследования, так и к организации данных. На этапе планирования исследования структура данных может быть задана в виде предварительного списка переменных с указанием их типов и диапазонов возможных значений<sup>6</sup>.

### **Типы переменных.**

Каждая переменная – это имеющее значение для исследователя основание, позволяющее отличать объекты друг от друга. На предварительном этапе следует выделять два типа переменных: количественные и категориальные (номинативные). Количественная переменная позволяет различать объекты по уровню выраженности некоторого свойства, например:

---

<sup>6</sup> Наследов А.Д. SPSS : Компьютерный анализ в психологии и социальных науках : самоучитель / А.Д. Наследов. – СПб. : Питер, 2005. – С. 40–41

средний балл отметки, тестовый показатель и пр. Идентификация количественных переменных на предварительном этапе не составляет труда: обычно они соответствуют тому, что исследователь намеревается измерить. Второй тип – категориальные (номинативные) переменные. Обычно они используются как основания для деления объектов (испытуемых) на группы или категории: пол, класс, возрастная категория, уровень дохода и пр. Типичная ошибка начинающего исследователя – игнорирование возможных оснований для деления объектов на группы в качестве самостоятельных номинативных переменных в структуре данных.

Важным свойством номинативных переменных является возможность их представления в виде набора целых чисел. Например, трем видам внешкольных увлечений (хобби) учащихся (спорт, компьютер, искусство) можно сопоставить числа 1, 2 и 3 соответственно. Числовое представление данных в компьютерных программах всегда предпочтительнее символьного, поскольку обработка чисел происходит быстрее, проще и с меньшей вероятностью ошибок. Кроме того, числовое представление легко модифицировать: вы можете переназначить числа, соответствующие созданным элементам, а также (что часто требуется на практике) без проблем включить в анализ новые элементы. Например, если в группе окажется учащийся, увлечение которого не соответствует перечисленным, будет полезно включить в переменную хобби категорию с названием другие и присвоить ей число 4<sup>7</sup>.

Третьим типом переменных являются порядковые. Значения порядковой переменной поддаются сортировке от меньшего к большему и наоборот. Например, оценка состояния объекта: «отлично», «хорошо», «удовлетворительно», «плохо» и «очень плохо»; оценка частоты какого-либо явления: «часто», «редко», «почти никогда». К вопросам, относящимся к такому типу переменной, можно отнести: «Оцените по 10-балльной шкале сложность обучения в нашем университете» и т.п.

---

<sup>7</sup> Там же. С. 42

**Способы ввода первичных данных в компьютер.** Вводить первичные данные (анкеты) в компьютер можно одним оператором или несколькими. Во втором случае необходимо иметь в виду, что кодировка открытых вопросов должна бы обязательно одинаковой у всех операторов. В дальнейшем частные базы данных от каждого оператора объединяются в одну.

### **Использованная литература**

Наследов А.Д. SPSS : Компьютерный анализ в психологии и социальных науках : самоучитель / А.Д. Наследов. – СПб. : Питер, 2005. – 416 с.

### **Контрольные вопросы**

1. Что подразумевается под строками и столбцами матрицы данных?
2. Зачем необходимо уделять внимание структуре данных?
3. Каковы типы переменных?
4. Каковы способы ввода данных?

### **Практические задания**

1. Запустите SPSS и откройте файл данных «куратор 2011».
2. Покажите вторую, третью и четвертую переменные.
3. Покажите ответы по этим переменным в анкетах (записях) №№ 41-45.
4. В предложенной преподавателем анкете укажите, к какому типу переменной (номинальной, порядковой или количественной) относится каждый вопрос.

## **1.4. Редактор данных**

Начнем с описания переменных во вкладке *Переменные* (рис. 3).

	Имя	Тип	Ширина	Десяти...	Метка	Значения	Пропущен...	Ширина...	Вывравниван...	Шкала	Роль
1	V1	Числовая	8	0	ПОЛ	{1, Юноши и...	Нет	8	☰ По право...	🔴 Номинальн...	👉 Входная
2	V2	Числовая	8	0	НАСЕЛЕННЫ...	{1, Благо...	Нет	8	☰ По право...	🔴 Номинальн...	👉 Входная
3	V2рег	Числовая	8	0	РЕГИОН	{1, Благо...	Нет	8	☰ По право...	🔴 Номинальн...	👉 Входная
4	V3	Числовая	8	0	КЛАСС	{1, 5-8 кла...	Нет	8	☰ По право...	🔴 Номинальн...	👉 Входная
5	V4	Числовая	8	0	ПРОФИЛЬ КЛ...	{0, без про...	Нет	8	☰ По право...	🔴 Номинальн...	👉 Входная
6	V5_1	Числовая	8	0	В КАКИЕ ВУЗ...	{1, АмГУ}...	Нет	8	☰ По право...	🔴 Номинальн...	👉 Входная
7	V5_2	Числовая	8	0	В КАКИЕ ВУЗ...	{1, АмГУ}...	Нет	8	☰ По право...	🔴 Номинальн...	👉 Входная
8	V5_3	Числовая	8	0	В КАКИЕ ВУЗ...	{1, АмГУ}...	Нет	8	☰ По право...	🔴 Номинальн...	👉 Входная
9	V6	Числовая	8	0	В КАКОЙ ИЗ ...	{1, АмГУ}...	Нет	8	☰ По право...	🔴 Номинальн...	👉 Входная
10	V7	Числовая	8	0	КАК БЫ ВЫ О...	{1, АмГУ}...	Нет	8	☰ По право...	🔴 Номинальн...	👉 Входная
11	V8_1	Числовая	8	0	ПО КАКИМ П	{1, Русски	Нет	8	☰ По право...	🔴 Номинальн...	👉 Входная

Рис. 3

### Имя переменной

Параметр *Имя* определяет имя переменной. Чтобы задать имя первой переменной просто введите его с клавиатуры в текущую ячейку. Имя второй переменной окажется в том же столбце под первой, то есть во второй строке; имя третьей переменной – в третьей строке, и т. д. Для перемещения между строками можно воспользоваться клавишами «вверх» и «вниз».

Имя переменной не является произвольным. Существует ряд соглашений, которым оно должно удовлетворять:

- ▶ длина имени не должна превышать 8 символов;
- ▶ в имени могут использоваться любые буквы, цифры, символы @, #, \$, однако имя всегда должно начинаться с буквы, а символ . (точка) не может стоять в конце имени;
- ▶ имена всех переменных должны быть разными;
- ▶ буквы верхнего и нижнего регистров символов различаются, то есть имена ID, id, Id и iD воспринимаются программой как идентичные;
- ▶ имена переменных не должны совпадать с каким-либо из зарезервированных слов (all, ne, eq, to, le, It, by, or, gt, and, not, ge, with).

Следующим шагом является определение типа переменной, что позволит вам максимально эффективно использовать ваши данные в будущем.

### Тип переменной



Параметр *Тип* определяет тип переменной. Как видите, текущим типом переменной является тип Числовой. В подавляющем большинстве случаев вам придется иметь дело именно с числовыми данными. В тех редких случаях, когда значения переменных представляют собой буквы или буквосочетания (слова), необходимо установить переключатель Текстовая. Строчные данные, в отличие от числовых, могут включать буквы и другие символы, то есть нести текстовую информацию. В частном случае строчная переменная может хранить число, однако обработка такого «числа» будет производиться так, как будто оно является текстом. Как правило, строчные переменные не подлежат обработке. Поэтому их следует избегать, за редким исключением – например, когда данная переменная содержит имена людей.

Необходимость в выборе строчного типа возникает, когда данные переносятся из программы Excel в SPSS путем копирования и вставки. Если значения какой-либо из переменных являются буквенными (например, «м» и «ж»), то перед вставкой этой переменной необходимо изменить ее тип. В противном случае вставлены не будут.

Остальные 6 типов переменных, которые можно задать с помощью переключателей, присутствующих в диалоговом окне, практически не используются при обработке данных, и поэтому мы не будем их рассматривать.

Поскольку все переменные имеют по умолчанию тип Числовой, никаких дополнительных действий с ними производить не нужно.

### **Ширина переменной**

Параметр *Ширина* позволяет задать максимальное количество знаков, которое может иметь переменная, включая дробную часть. Например, ширина переменной № (идентификатор) составляет 3 знака, поскольку все числа, которые мы намерены вводить в нее, содержат от 1 до 3 знаков. На практике определить заранее ширину переменной гораздо труднее, поскольку не всегда известно, какие данные нам будет нужно вводить в нее в будущем. Поэтому следует задавать ширину переменной с гарантированным запасом: ее можно ограничить потом, после ввода данных.

В большинстве случаев нет необходимости менять заданную по умолчанию ширину переменной.

### **Дробная часть числа**

Параметр *Десятичные* предназначен для задания числа десятичных знаков после запятой в случае, если тип переменной допускает использование дробных чисел. Для строковых переменных значение в ячейке Десятичные автоматически устанавливается равной нулю, а для цифровых – равным 2. У строковых переменных значение данного параметра недоступно для изменения.

### **Метки переменных**

С помощью параметра *Метка* можно создать метку переменной. Как правило, метка используется в тех случаях, когда содержание переменной недостаточно ясно отражено в ее названии. По сути метка – это пояснение к имени переменной, то есть формулировка вопроса в той форме, в какой он сформулирован в анкете. Метка переменной – это название, позволяющая описать переменную более подробно. Метка переменной может содержать до 256 символов. В метках переменных различаются прописные и строчные буквы. Они отображаются в том виде, в каком были введены.

### **Метки значений переменных**

Параметр *Значения* позволяет управлять наименованиями уровней (категорий) переменной. Под уровнем, или категорией, понимается целочисленное значение переменной, имеющее определенный смысл, то есть формулировки вариантов ответов на данный вопрос анкеты. Например, переменная пол имеет два уровня: 1 – жен (женский) и 2 – муж (мужской). Буквосочетания или слова, поставленные в соответствие уровням переменной, например, жен и муж, называются метками значений и отражают смысл разных значений переменной для исследователя. Иными словами, метки значений – это название, позволяющее более подробно описать возможные значения переменной. Так, например, в случае переменной пол можно задать метку "женский" для значения "1" и метку "мужской" для значения "2".

### **Пропуски**

Параметр **Пропущенные** используется очень редко, поскольку программа и так позволяет учитывать пропуски в данных. Необходимость в этом параметре возникает, когда требуется различать причины пропусков значений. Например, пропуск в данных может быть обусловлен тем, что респондент еще не опрошен, а может быть и так, что он отказался отвечать на данный вопрос. Так, в отношении переменной вуз (предполагаемый для поступления вуз) нам необходимо различать тех учащихся, которых мы не успели опросить, и тех, которые еще не определились. Для еще не опрошенных учащихся мы будем оставлять пустую ячейку, а неопределившихся учащихся будем обозначать цифрой 9. Если ввести значение 9 в столбец Пропущенные, то оно не будет использоваться в дальнейшем при обработке наряду с пустыми ячейками. После щелчка на кнопке с многоточием, появляющейся при перемещении фокуса ввода в ячейку столбца Пропущенные, открывается диалоговое окно Пропущенные значения, в котором можно определить необходимые обозначения пропусков.

### **Столбцы**

Параметр **Ширина столбца**, в отличие от своего соседа слева, требуется для всех переменных. С его помощью можно управлять шириной (в символах) столбцов вкладки просмотра данных (**Данные**). Поле Ширина столбца определяет ширину, которую будет иметь в таблице данный столбец при отображении значений. Ширину столбца также можно изменить непосредственно в окне редактора данных. Для этого поместите указатель мыши на разделитель между двумя заголовками столбцов с именами переменных. Вид указателя изменится. Появившаяся двойная стрелка указывает, что соответствующий столбец можно расширить или сузить путем перетаскивания.

### **Выравнивание**

Параметр **Выравнивание** позволяет управлять расположением данных внутри ячейки. Здесь можно задать вид выравнивания значений, т.е.

определить, как они будут отображаться в таблице. Возможные виды выравнивания – По правому краю, По левому краю и По центру.

### **Шкала измерения**

Как и у предыдущего параметра, значение параметра *Шкала* выбирается в раскрывающемся списке из трех доступных: Количественная, Порядковая и Номинальная. Иными словами, здесь указывается тип переменной. По умолчанию принимается метрическая шкала измерения<sup>8</sup>.

### **Особенности описания переменных отдельных типов вопросов.**

*Альтернативный вопрос* – описывается с помощью одной переменной.

Особенность описания *множественного вопроса* определяется спецификой ввода данных в SPSS вообще. Дело в том, что одна ячейка в базе данных SPSS – это ответ на один вопрос, НО (!) одна ячейка может содержать один и только один код ответа, то есть только один ответ. В случае с множественным вопросом поступают следующим образом: либо 1) вопрос описывают столькими переменными, сколько респондент может дать ответов на данный вопрос, либо 2) создают отдельную переменную для каждого возможного варианта ответа. И в том, и в другом случае множественный вопрос в базе данных представлен несколькими переменными.

*Табличный вопрос* – это совокупность подвопросов, каждый из которых описывается отдельной переменной. То есть сколько в табличном вопросе подвопросов, столько и переменных для его описания.

### **Редактирование описаний переменных**

Чтобы вставить новую переменную между двумя соседними, щелкните сначала на правой из них, а затем – на кнопке Вставить переменную. Будет создан пустой столбец, а все переменные, находящиеся справа, окажутся сдвинутыми на один столбец. Чтобы удалить переменную, необходимо выделить ее, щелкнув на ней правой кнопкой мыши, а затем выбрать в раскрывшемся меню пункт Удалить.

---

<sup>8</sup> Наследов А.Д. SPSS : Компьютерный анализ в психологии и социальных науках : самоучитель / А.Д. Наследов. – СПб. : Питер, 2005. – С. 44–51

## **ВВОД ДАННЫХ.**

Итак, когда все переменные созданы, а их параметры определены, настало время заняться вводом данных. Ввод самих данных гораздо проще, чем задание параметров переменных. Для ввода значений в создаваемый файл данных предназначена вкладка *Данные*.

Существует два метода непосредственного ввода данных в ячейки электронной таблицы: по строкам (переменным) и по столбцам (объектам). В зависимости от конкретных данных тот или другой способ оказывается предпочтительным.

Ввод данных по переменным предполагает последовательное заполнение всех строк одного столбца значениями одной переменной, затем другой переменной и т. д. Вы вводите значения в левую верхнюю ячейку окна, затем нажимаете клавишу «вниз» или Enter, переходя на следующую строку, заполняете еще ячейку и продолжаете эти действия до тех пор, пока не заполните весь первый столбец. Затем заполняете столбец, соответствующий второй переменной, и т. д., пока все переменные не будут введены. Таким образом, с визуальной точки зрения ввод данных по переменным осуществляется «по вертикали».

Ввод данных по объектам заключается в последовательном заполнении каждой строки значениями всех переменных. Сначала вводится значение в левую верхнюю ячейку окна, затем с помощью клавиши «вправо» или Tab осуществляется переход в соседний столбец (переменную), вводится значение переменной и т. д. до тех пор, пока вся строка не будет заполнена. После этого аналогичным способом заполняются вторая, третья и все остальные строки. В отличие от предыдущего способа ввод данных по объектам осуществляется «по горизонтали»<sup>9</sup>.

### **Особенности ввода ответов на отдельные типы вопросов.**

Ввод ответа на *альтернативный вопрос* – это ввод одной цифры, кодирующей данный ответ респондента.

---

<sup>9</sup> Там же. С. 51–52

Особенность ввода ответов на *множественный вопрос* определяется спецификой ввода данных в SPSS вообще. Дело в том, что одна ячейка в базе данных SPSS – это ответ на один вопрос, НО (!) одна ячейка может содержать один и только один код ответа, то есть только один ответ. В случае с множественным вопросом поступают следующим образом:

1) каждый отдельный ответ респондента на такой тип вопроса кодируется в отдельную ячейку (*дихотомный метод*):

Вопрос: "Как можно сделать партию более привлекательной?".

Предлагались следующие варианты ответов:

- больше активности в период между выборами
- повышение эффективности общих собраний
- больше неформальных встреч
- открытые общие собрания
- большая близость к населению на местах
- лучше информировать членов партии об актуальных событиях
- привлечение не членов партии к различным партийным проектам
- больше мероприятий по актуальным политическим темам на местах

В методе множественной дихотомии для каждой из возможностей ответа определяется отдельная переменная. В рассматриваемом примере для этого понадобится восемь переменных. Если член партии отметит ответ "больше активности в период между выборами", соответствующая переменная получит значение "1", если нет – "0", если член партии отметит ответ "повышение эффективности общих собраний", соответствующая переменная получит значение "1", если нет – "0" и т.д. для остальных переменных. Таким образом мы получим восемь переменных с кодовыми значениями 0 и 1. Кодовые значения при этом выбираются произвольно, однако для всех ответов они должны быть одинаковы и введены в компьютер на правильном месте;

2) кодируется каждый возможный выбор варианта ответа респондентом (*категориальный метод*):

Альтернативный способ кодирования множественных ответов предоставляет метод множественных категорий, или категориальный. Для применения этого метода должно быть известно максимальное количество возможных ответов. Это количество можно, например, задать в анкете (указанием типа "Отмечайте не более пяти вариантов") или установить после проверки анкет.

Чтобы узнать, почему члены партии, не имеющие партийного поручения, не хотят его получить или не участвуют в партийной работе иным образом, в анкете задавался вопрос "Что мешает Вашему участию в партийной работе?". После вопроса было помещено указание, что можно отметить не более пяти из приводимых вариантов ответа. Так как количество ответов составляет не более пяти, для того, чтобы закодировать все варианты ответов будет достаточно пяти переменных. Каждая из пяти переменных кодируется одинаковыми категориями, причем вне зависимости от количества данных ответов область этих пяти переменных заполняется слева направо<sup>10</sup>.

И в том, и в другом случае множественный вопрос в базе данных представлен несколькими ячейками.

*Табличный вопрос* – это совокупность подвопросов. Каждый такой подвопрос кодируется в базе данных и вводится отдельно.

## **РЕДАКТИРОВАНИЕ ДАННЫХ**

### **Изменение содержимого ячейки**

Щелкните на ячейке, содержимое которой вы намерены отредактировать, введите новое значение, а затем перейдите в любую соседнюю ячейку с помощью клавиш Enter, Tab или клавиш со стрелками.

### **Вставка нового объекта**

Если вам необходимо вставить новый объект (строку) между двумя соседними строками, щелкните сначала на нижней из них, а затем – на кнопке *Вставить наблюдения* панели инструментов. В результате будет создана пустая строка, а номера строк, находящихся ниже, увеличатся на единицу.

---

<sup>10</sup> Иллюстрированный самоучитель по SPSS [Электронный ресурс]. URL: [www.learnspss.ru](http://www.learnspss.ru)

## **Копирование и вырезание содержимого ячеек**

Для того чтобы скопировать или вырезать содержимое одной или нескольких ячеек, сначала необходимо их выделить. Активная ячейка (то есть ячейка, в которой находится фокус ввода) всегда является выделенной. Чтобы выделить группу ячеек, нажмите левую кнопку мыши на угловой ячейке будущей группы, и удерживая эту кнопку, перетащите указатель к противоположной угловой ячейке группы, после чего кнопку мыши отпустите. Если вам необходимо выделить целиком строку или столбец, щелкните соответственно на нужном объекте или на имени переменной. Когда желаемые ячейки выделены, в меню *Правка* выберите команду Копировать или Вырезать.

## **Вставка ячеек**

Для того чтобы вставить предварительно скопированные в буфер обмена данные, нужно переместить фокус ввода в левую верхнюю ячейку группы, в которую будет осуществляться вставка, и в меню *Правка* выбрать команду Вставка. Следует помнить о двух неприятностях, связанных с операцией вставки.

► Если в области, в которую планируется вставить данные, уже содержатся какие-либо другие данные, то после выполнения вставки они будут утеряны.

► При копировании и вставке данных их смысл не анализируется программой, и вы можете случайно заменить значения одной переменной значениями любой другой переменной. Последствия такой вставки могут быть весьма плачевными, поскольку разрешить возникшую путаницу, как правило, весьма непросто.

Мы приводим несколько рекомендаций, которые могут быть полезными при выполнении вставки.

Перед удалением и вставкой важных фрагментов сохраняйте текущий файл. В случае ошибки вы сможете открыть этот файл снова.

По возможности создавайте новые переменные и строки для вставки данных.



Как правило, вам придется иметь дело с копированием, вырезанием и вставкой столбцов (переменных) или строк (объектов) целиком. Однако особую аккуратность следует проявлять при вставке блоков столбцов или строк.

► Будьте особенно внимательны при вставке значений переменных разных типов, например, не допускайте вставки значений строковых переменных в столбцы, соответствующие числовым переменным. Это может привести к трудно идентифицируемым ошибкам в процессе статистического анализа<sup>11</sup>.

### **Использованная литература**

Иллюстрированный самоучитель по SPSS [Электронный ресурс]. URL: [www.learnspss.ru](http://www.learnspss.ru)

Наследов А.Д. SPSS : Компьютерный анализ в психологии и социальных науках : самоучитель / А.Д. Наследов. – СПб. : Питер, 2005. – 416 с.

### **Контрольные вопросы**

1. Для чего нужна вкладка Variable View (Переменные)?
2. Назовите основные параметры описаний переменных и их назначение
3. Для чего нужна вкладка Data View (Данные)?
4. Опишите процесс ввода и редактирования данных

### **Практические задания**

1. Запустите СПСС и выберите вкладку *Переменные* в окне редактора данных.
2. Во вкладке *Переменные* сделайте полное описание выделенных переменных анкеты, предложенной преподавателем.
3. Между вторым и третьим вопросом вставьте новую переменную.
4. Удалите эту новую переменную.

---

<sup>11</sup> Наследов А.Д. SPSS : Компьютерный анализ в психологии и социальных науках : самоучитель / А.Д. Наследов. – СПб. : Питер, 2005. – С. 53–54

5. Не закрывая файл, перейдите во вкладку *Данные*.
6. Введите в полученную матрицу свои ответы на данную анкету, а затем ответы своего соседа по парте (соседа по комнате, друга, брата, сестры, мамы или папы).
7. Вставьте новую строку между первой и второй строками.
8. Удалите эту новую строку.
9. Скопируйте первую и вторую строки и вставьте их в конце данных.

### **1.5. Оценка и контроль качества первичных данных**

Когда данные введены в компьютер, не следует сразу же приступать к анализу. На первом этапе сами данные следует подвергнуть подробному и всестороннему исследованию. Подобное исследование преследует несколько целей, первичная из которых: Обнаружение ошибок ввода.

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, там 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». В работе немецких авторов, сходный по назначению документ называется «Кодировочная таблица».

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.

Первый вид ошибок характерен для квалифицированных интервьюеров, кодировщиков и прочих. 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. Использование в качестве интервьюеров, кодировщиков и операторов неквалифицированного технического персонала, взятого со стороны, заведомо ведет к появлению ошибок и того, и другого вида.

Источниками появления ошибок могут служить и ограниченность используемой методики, и условия проведения опроса, и психофизиологические особенности самих опрашиваемых и интервьюеров. 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.

**Методы диагностики базы данных.** Существует два основных метода диагностики базы данных SPSS: выборочный и автоматический. Первый заключается в визуальном просмотре 10 % полученных от интервьюера анкет на предмет ошибок, неточностей, неполноты, нелогичности в фиксировании ответов респондента. Второй метод – автоматическая диагностика – осуществляется при помощи самого программного комплекса SPSS и позволяет учитывать не только ошибки интервьюеров, но и кодировщиков, а также оперативно находить их и исправлять. Ниже будет более подробно рассмотрен второй метод.

Для целей поиска и исправления ошибок ввода в SPSS весьма эффективно использовать следующие процедуры: **Сортировать наблюдения, Частоты, Найти, Итоги по наблюдениям и Отобрать наблюдения**, которые будут рассмотрены ниже как самостоятельно, так и в сочетании с другими процедурами<sup>12</sup>.

### **Использованная литература**

Пациорковский В.В., Пациорковская В.В. SPSS для социологов: Учебное пособие. ИСЭПН РАН. – М., 2005. – 433 с.

### **Контрольные вопросы**

1. Что такое «качество данных»? От чего оно зависит?
2. Назовите основные Виды ошибок?
3. Каковы основные Методы диагностики БД?
4. Что такое «Выборочный (10%) контроль правильности ввода данных»?
5. Что такое «Автоматическая диагностика БД»?
6. Что такое «качество первичных данных»? По каким признакам можно судить о качестве данных?

---

<sup>12</sup> Пациорковский В.В., Пациорковская В.В. SPSS для социологов: Учебное пособие. ИСЭПН РАН. – М., 2005. – С. 99–100

7. Зачем необходимо контролировать качество первичных данных после ввода данных?

8. Приведите примеры случайных и систематических ошибок ввода данных и причины их возникновения.

9. Как устранить ошибки ввода данных? Укажите достоинства и недостатки выборочного и автоматического метода диагностики первичных данных.

### 1.6. Тестирование качества базы данных. Чистка массива данных

Процедура **Сортировать наблюдения** позволяет сортировать данные по возрастанию или убыванию значения признака. Сразу 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 **Сортировать наблюдения** необходимо выбрать в главном меню последовательность команд (рис. 4):

**Данные**

**Сортировать наблюдения**

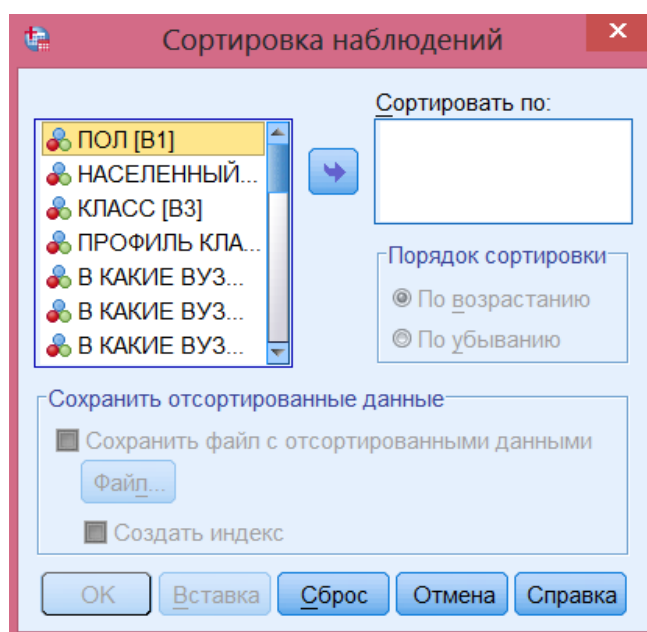


Рис. 4

С выполнения рассмотренной выше последовательности команд мы и рекомендуем начинать этап контроля данных.

Самый точный метод проверки данных (то есть значений всех переменных) на ошибки при вводе состоит в том, чтобы командами меню

## **Анализ**

### **Отчеты**

#### **Итоги по наблюдениям**

вывести их список и сравнить каждое значение с оригиналом (например, анкетой). Однако этот способ требует очень много времени, особенно при большом объеме данных. Поэтому решиться на проведение такой скучной и утомительной работы можно только в редких случаях – как правило, когда объем данных ограничен (рис. 5).

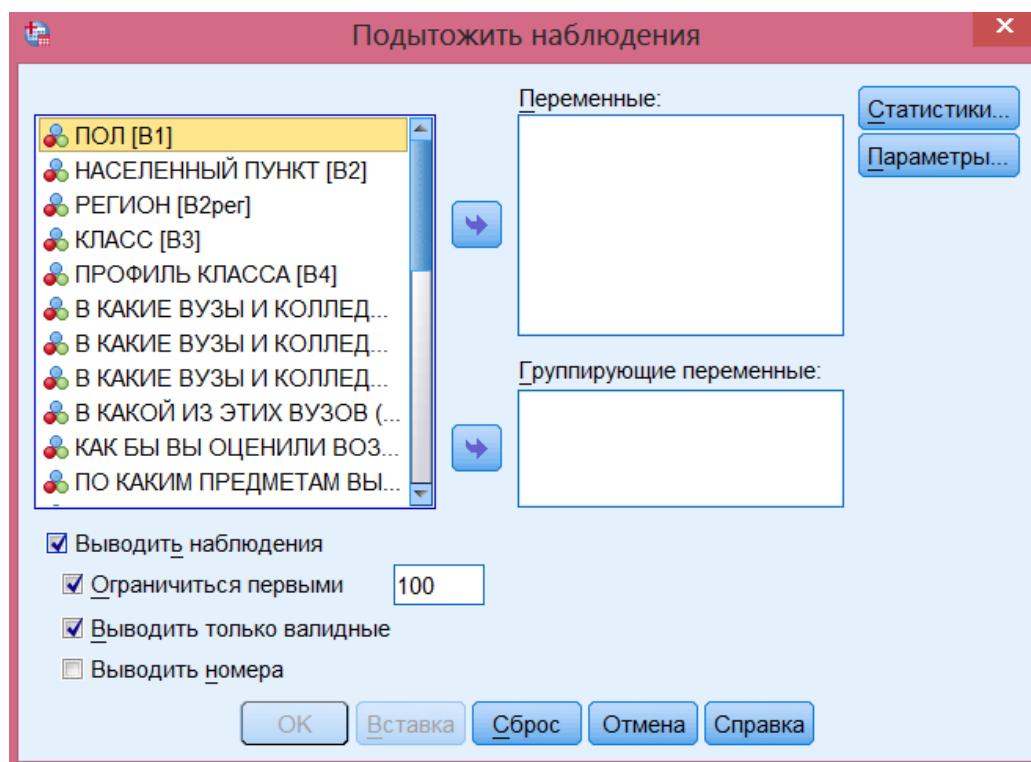


Рис. 5

В общем случае рекомендуется проводить частотный анализ значений переменных; для этого служат команды меню

## **Анализ**

### **Описательные статистики**

#### **Частоты**

Результаты этого анализа при внимательном рассмотрении позволяют выявить недопустимые значения (рис. 6). Например, если переменная содержит данные роста в сантиметрах, то значение 384, обнаруживаемое при частотном анализе, явно свидетельствует о том, что в данных имеется ошибка. После проведения частотного анализа это значение можно отыскать в файле данных и исправить. Следовательно, при изучении частотных таблиц особое внимание надо обращать на максимальное и минимальное значения. Однако если вместо возраста 65 лет было введено, например, значение 56, то при помощи частотной таблицы эту ошибку обнаружить невозможно.

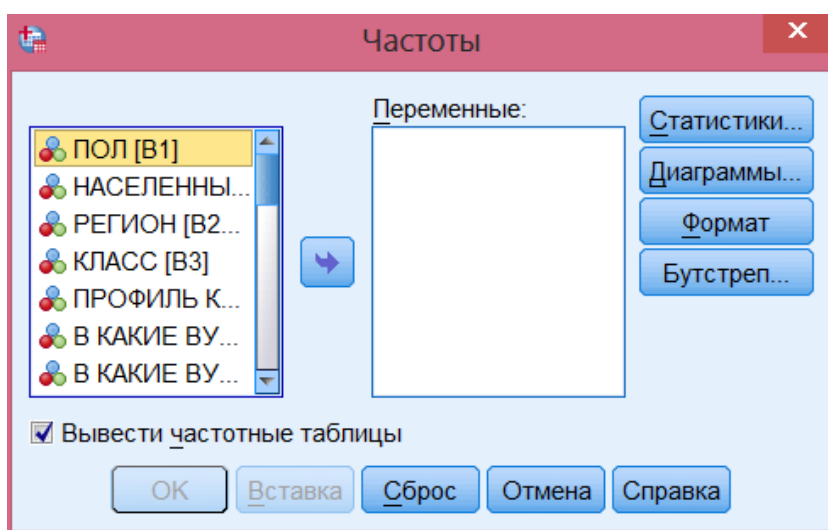


Рис. 6

Обладая некоторыми практическими навыками и фантазией, с помощью описанных и им подобных способов можно выявить большое количество ошибок ввода. Все такие ошибки обязательно должны быть исправлены. Даже если наблюдений несколько тысяч, то даже одно-единственное противоречивое значение наносит вред вашему исследованию: создается впечатление, что работа по сбору и подготовке информации выполнена поверхностно.

Откроется диалоговое окно **Частоты**. В этом окне слева появится список переменных, которые можно выбрать, нажав кнопку «стрелка вправо». Выбранные переменные попадут в правый список. После нажатия кнопки ОК, процедура начнет выполняться, и результаты будут выдаваться в окне просмотра.

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 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

## **Правка**

### **Найти**

Откроется диалоговое окно *Найти и заменить*. Ввести в поле *Поиск* то значение, которое требуется найти. Далее следует команда *Найти далее* (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 (рис. 7). Поиск следующего случая предполагает повторное выполнение команды *Найти далее*<sup>13</sup>.

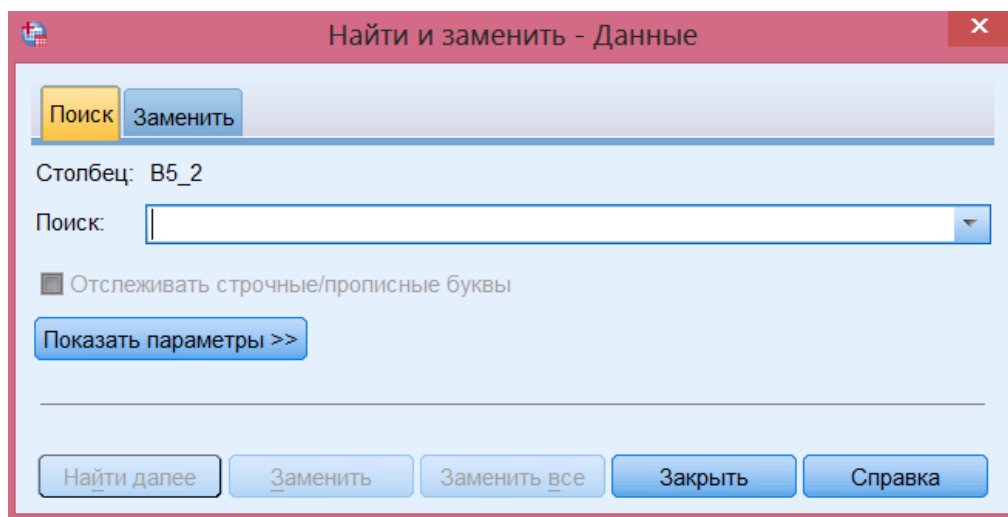


Рис. 7

<sup>13</sup> SPSS: обработка статистической информации [Электронный ресурс]. URL: [www.ispss.ru](http://www.ispss.ru); Пациорковский В.В., Пациорковская В.В. SPSS для социологов: Учебное пособие. ИСЭПН РАН. – М., 2005. – С. 101–103



## Использованная литература

SPSS: обработка статистической информации [Электронный ресурс].  
URL: [www.ispss.ru](http://www.ispss.ru)

Пациорковский В.В., Пациорковская В.В. SPSS для социологов: Учебное пособие. ИСЭПН РАН. – М., 2005. – 433 с.

## Контрольные вопросы

1. Каковы возможности процедуры Итоги по наблюдениям?
2. Каковы возможности процедуры Частоты?
3. Зачем нужна процедура Поиск?

## Практические задания

1. С помощью процедуры Sort Cases (Сортировать наблюдения...) протестируйте в базе данных «первокурсники (с ошибками)» одну переменную на свой выбор.

2. С помощью процедуры Case summaries... (Итоги по наблюдениям...) введите список двух-трех переменных на свой выбор и просмотрите их значения в сводной таблице для первых 30 анкет.

3. С помощью процедуры Frequencies... (Частоты) протестируйте две переменные на свой выбор.

4. С помощью процедур Frequencies... (Частоты) и Find... (Найти...) протестируйте все переменные и найдите 10 ошибок ввода данных. Результаты оформите в таблицу по образцу ниже:

№ №	Имя переменной, в которой обнаружена ошибка	Номер анкеты, в которой обнаружена ошибка по данной переменной	Суть ошибки (указать ошибочное значение, а в скобках – допустимые значения)
	V_1	12	3 (допустимо 1 и 2)
	V_1	137	4 (допустимо 1 и 2)
	V_5	14	5 (допустимо 1-4 и 98)
	V_14	2	7 (допустимо 1-4, 97, 98)
	...	...	...
0	V_24	231	99 (допустимо 1-10)

5. С помощью процедуры Sort Cases (Сортировать наблюдения...) найдите все пропущенные значения по переменной *v\_23* (Вопрос «Ваш пол»).

### 1.7. Сервисные процедуры с данными

Материал этого параграфа посвящен вопросам эффективной работы с исходными данными. Описанные здесь операции весьма полезны в большинстве случаев обработки и анализа данных, так как практически всегда существует необходимость в предварительной подготовке и преобразовании исходных данных. Поэтому изложенные рекомендации по форматированию данных помогут вам работать с программой гораздо свободнее. Иногда желательно упорядочить данные файла по какому-либо признаку, например по результатам выполнения какого-либо задания. Нередко возникает необходимость обработки не всех данных файла, а лишь их подмножества, выделяемого по определенным критериям (например, по полу, классу, успеваемости и пр.). Существует и обратная задача: если данные хранятся в нескольких небольших файлах, может возникнуть потребность в их объединении для последующего анализа.

Перечисленные проблемы указывают на то, что для регулярной аналитической работы недостаточно умения вводить данные и применять к ним статистические процедуры. Возникает задача эффективного управления данными. Способы решения этой задачи бывают весьма нетривиальными. Тем не менее представленного в этой главе материала вполне достаточно, чтобы научиться свободно манипулировать данными даже новичку.

Несмотря на то, что навыки управления данными приходят с опытом и требуют некоторого терпения, обязательно освоите их. Это придаст процессу исследования гибкость, простоту и легкость. Тогда выполнение статистических процедур, казавшихся сложными и громоздкими, станет для вас интуитивно понятным<sup>14</sup>.

---

<sup>14</sup> Наследов А.Д. SPSS : Компьютерный анализ в психологии и социальных науках : самоучитель / А.Д. Наследов. – СПб. : Питер, 2005. – С. 57–58

## 1. Объединение данных. Слияние файлов.

Иногда требуется объединить два или несколько файлов данных. Такая необходимость возникает, например, когда анкеты вводят несколько кодировщиков, или когда анкеты от разных групп респондентов вводятся в разные базы (горожане в один файл, а селяне в другую), или когда ответы одних и тех же респондентов на различные вопросы вводятся в разные базы данных. Объединение файлов – это добавление данных из одного файла к данным другого файла.

Работа с данными нескольких файлов одновременно иногда способна сбить с толку даже опытных пользователей. Нередко файлы данных создаются при помощи разного программного обеспечения и имеют разные форматы. Это порождает различные проблемы их совместного использования, из которых мы рассмотрим наиболее типичную: дополнение рабочего файла SPSS содержимым внешнего файла. При этом возможны две ситуации дополнения данных рабочего файла SPSS:

- из внешнего файла Excel;
- из внешнего файла SPSS.

Мы рассмотрим оба варианта, но сначала сформулируем общие рекомендации:

^ если вы намерены добавлять переменные, убедитесь, что порядок следования объектов в рабочем и внешнем файлах одинаков;

^ если вы намерены добавлять объекты, убедитесь, что порядок следования переменных в рабочем и внешнем файлах одинаков;

^ настройте форматы каждой переменной рабочего файла данных, чтобы они соответствовали данным внешнего файла (если добавляемая переменная содержит буквенные символы, то либо замените буквы числами во внешнем файле, либо поменяйте тип переменной на строчную в рабочем файле);

^ перед добавлением данных создайте резервную копию рабочего файла, чтобы к ней можно было вернуться в случае неудачного переноса данных.

Следует иметь в виду, что чем больше соответствия между структурами рабочего и внешнего файлов, тем меньше вероятность ошибок при слиянии данных.

Если вы выполнили указанные рекомендации, то перенос данных из таблицы Excel не составит труда при помощи команд меню *Правка: Копировать* и *Вставить*.

При помощи той же последовательности действий можно перенести данные в рабочий файл из внешнего файла SPSS.

Указанная последовательность действий требует большого внимания и определенной сноровки. Поэтому вряд ли этот метод может быть рекомендован неочень искушенным пользователям. К счастью, существует более надежный способ объединения данных – *слияние рабочего и внешнего файлов*.

Слияние файлов допустимо, когда и рабочий, и внешний файлы созданы при помощи редактора данных SPSS и имеют одинаковые имена переменных (когда мы добавляем объекты) и одинаковые число и порядок следования объектов (когда мы добавляем переменные). Таким образом, если внешним является файл Excel, то перед слиянием необходимо на его основе создать внешний файл SPSS. Структура создаваемого файла должна быть максимально согласована со структурой рабочего файла. Для этого, дополнительно к общим рекомендациям, изложенным ранее, нужно настроить форматы каждой переменной файлов так, чтобы они были одинаковыми, и проверить идентичность имен переменных файлов.

Хотя эти рекомендации обязательными не являются, все же следует иметь в виду, что чем больше соответствия между структурами файлов, тем меньше вероятность ошибок при слиянии. Особое внимание следует уделять идентичности имен переменных, их форматам и расположению в файле данных. Обратите внимание также на то, что при добавлении переменных одинаковый порядок следования объектов в файлах обязателен.

В зависимости от характера данных, содержащихся в различных файлах, можно выделить два основных типа слияния данных: объединение объектов (респондентов) и объединение переменных (вопросов анкет).

*Добавление объектов.*

С помощью диалогового окна *Добавить наблюдения* можно выбрать внешний файл данных, предназначенный для слияния с открытым (рабочим) файлом. Если нужный внешний файл есть в списке диалогового окна, то для доступа к нему достаточно щелкнуть на его имени (рис. 8):

**Данные**

**Слить файлы**

**Добавить наблюдения**

**Обзор...**

SPSS без проблем объединит файлы, если при добавлении респондентов в объединяемых файлах все вопросы (их переменные) идентичны.

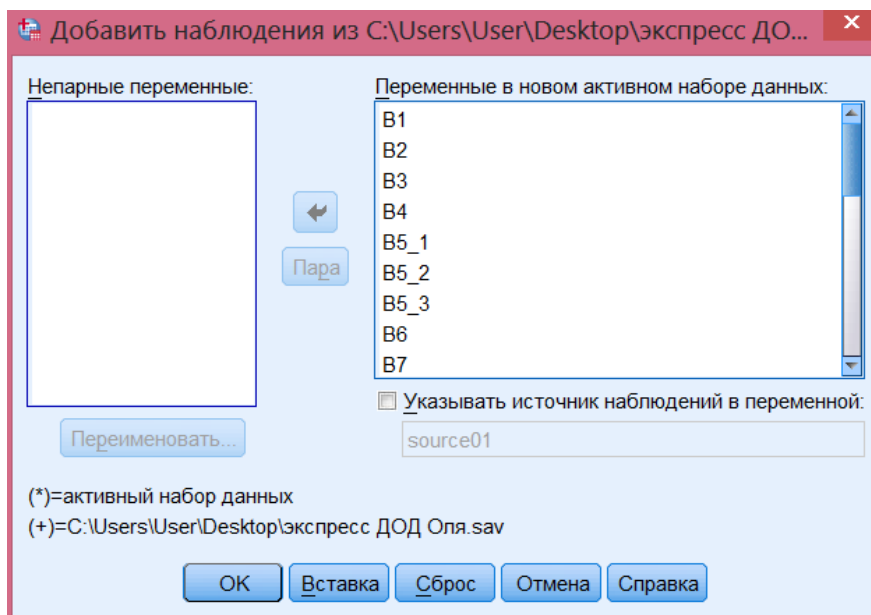


Рис. 8

*Добавление переменных.*

Слияние с добавлением переменных очень похоже на слияние с добавлением объектов. Начало выполнения обеих пошаговых процедур фактически одинаково: на экране появляется диалоговое окно, в котором вы указываете имя внешнего файла и щелкаете на кнопке *Обзор*.

## Данные

### Слить файлы

### Добавить переменные

### Обзор...

Следующим и очень важным действием является задание ключевой переменной (рис. 9). Необходимо, чтобы порядок следования значений в ключевых переменных обоих файлов был одинаков. В нашем примере в роли ключевой переменной рекомендуется использовать переменную № (идентификационный номер), поскольку ее значения уникальны для каждого объекта. Значения переменной № должны быть упорядочены, например по возрастанию. После того как вы убедитесь, что в обоих файлах ключевые переменные отсортированы одинаково, установите флажок *Сопоставить наблюдения на основе ключевых переменных*, выделите ключевую переменную в списке *Исключенные переменные* и щелкните сначала на нижней кнопке со стрелкой, а потом – на кнопке ОК<sup>15</sup>.

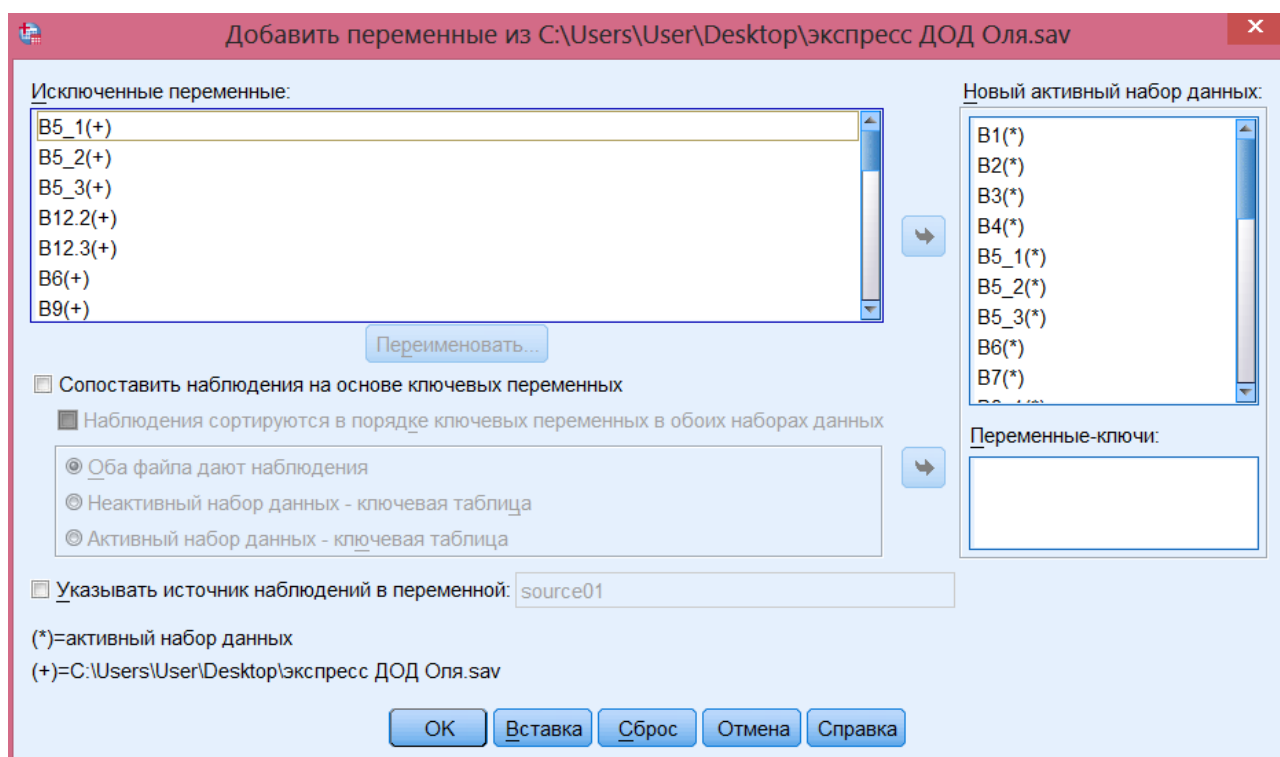


Рис. 9

<sup>15</sup> Там же. С. 78–82

SPSS без проблем объединит файлы, если при добавлении вопросов (их переменных) объединяемые файлы содержат одинаковых респондентов по основным контрольным признакам (пол, возраст, образование и т.п.).

## 2. Сортировка наблюдений.

Команда **Сортировать наблюдения** предназначена для реорганизации данных файла. Эта операция очень распространена, поскольку позволяет расположить информацию в том порядке, в котором это удобно исследователю в текущий момент. Подобные операции над данными можно выполнить с помощью команды **Сортировать наблюдения** (рис. 10).

Данные

### Сортировать наблюдения...

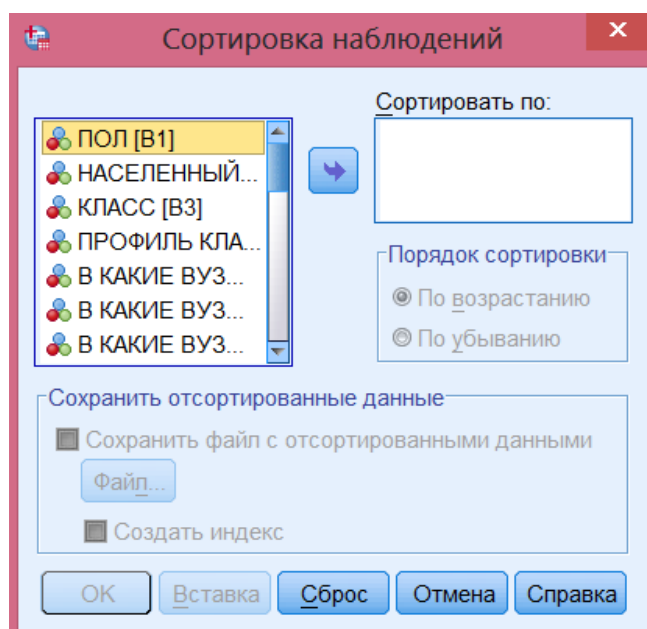


Рис. 10

Параметры команды **Сортировать наблюдения**, в отличие от параметров предыдущих команд, настраиваются в одном небольшом диалоговом окне. Вы выбираете переменные в исходном списке и с помощью переключателей По возрастанию или По убыванию указываете порядок сортировки. Для текстовых переменных сортировка по возрастанию означает сортировку в алфавитном порядке.

При определении параметров сортировки можно задавать сразу несколько переменных. Это означает, что сначала данные будут отсортированы

но значению первой выбранной переменной, затем объекты, имеющие одинаковые значения первой переменной, сортируются по значению второй переменной и т. д. Например, если в список *Сортировать по* ввести переменные класс и пол, то все учащиеся будут перечислены по классам, а внутри каждого класса сначала будут следовать девушки, затем – юноши<sup>16</sup>.

### **3. Разделение наблюдений на группы.**

В SPSS можно выполнять анализ данных отдельно по группам. Группой в этом контексте называется определенное количество наблюдений с одинаковыми значениями признаков<sup>17</sup>. Иными словами SPSS позволяет разделить респондентов в базе данных на различные группы для последующего анализа каждой группы в отдельности. Например, всех респондентов можно разделить на мужчин и женщин, или по студенческим группам, или населенным пунктам и т.п. В первом случае все переменные со значением признака 1 (женский) образуют одну группу, а все переменные со значением признака 2 (мужской) – другую группу. С каждой группой можно проводить определенные операции, например, выполнять частотный анализ (рис. 11).

**Данные**

**Расщепить файл...**

---

<sup>16</sup> Там же. С. 77

<sup>17</sup> Пациорковский В.В., Пациорковская В.В. SPSS для социологов: Учебное пособие. ИСЭПН РАН. – М., 2005. – С. 67



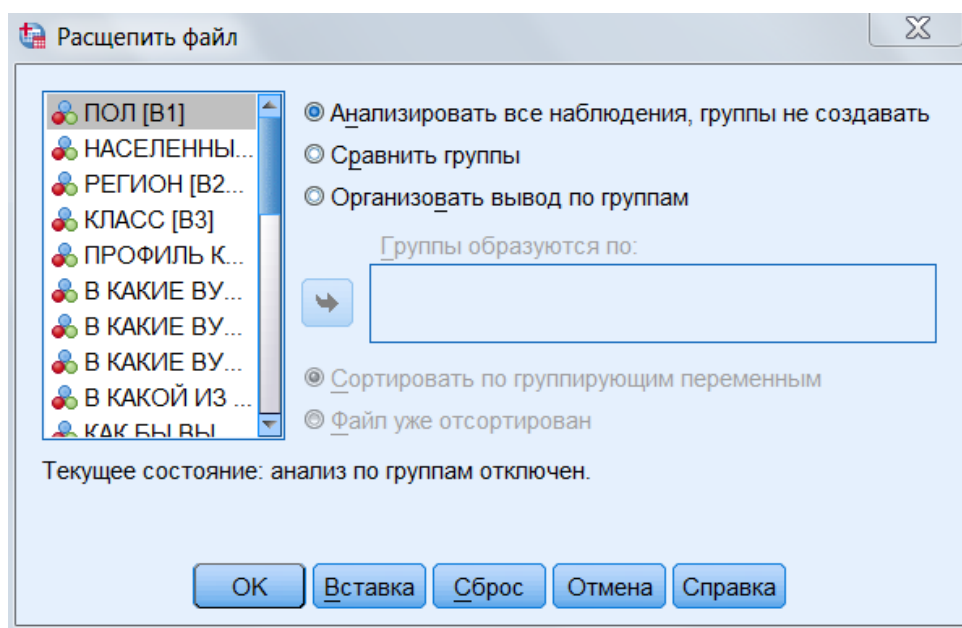


Рис. 11

Системой здесь допускаются три основные возможности (опции):

- исходно, по умолчанию, разделение на группы отсутствует. Об этом свидетельствует метка в первой опции: *Анализировать все случаи, не создавая групп*;
- следующая опция – *Сравнивать группы*;
- последняя опция – *Разделить вывод на группы*.

При выборе одной из двух последних опций группирующий признак задается путем переноса соответствующей переменной из левого подокна со списком переменных в правое подокно – *Группы образуются по*.

В результате выполнения рассматриваемой команды в текущем сеансе, вплоть до ее отмены, все расчеты будут выполняться в разрезе группирующего признака. Выполнение этой команды может быть отменено посредством возврата в диалоговое окно, установки в нем первой опции и последующей команды ОК.

#### 4. Отбор наблюдений.

Отбор данных – это выбор наблюдений по определенным критериям; так, например, при опросе избирателей можно отобрать только мужчин, голосующих за конкретную партию, а при опросе студентов – только

студенток, обучающихся по направлению «Социология». После этого все вычисления будут проводиться только с этими отобранными наблюдениям.

Команда **Отобразить наблюдения** позволяет пользователю выбирать для обработки не все, а часть данных, удовлетворяющих заданным условиям. Поскольку необходимость в этом возникает довольно часто, команда **Отобразить наблюдения** является одной из самых востребованных при проведении исследований. Так, исследователю могут понадобиться статистические сведения, касающиеся учащихся одного из классов или определенного пола, и т. д. Это означает, что ему нужно указать программе, какие данные следует выделить для обработки. Именно для этого предназначена команда **Отобразить наблюдения** (рис. 12).

## Данные

### Отобразить наблюдения

#### Если выполнено условие – Если...

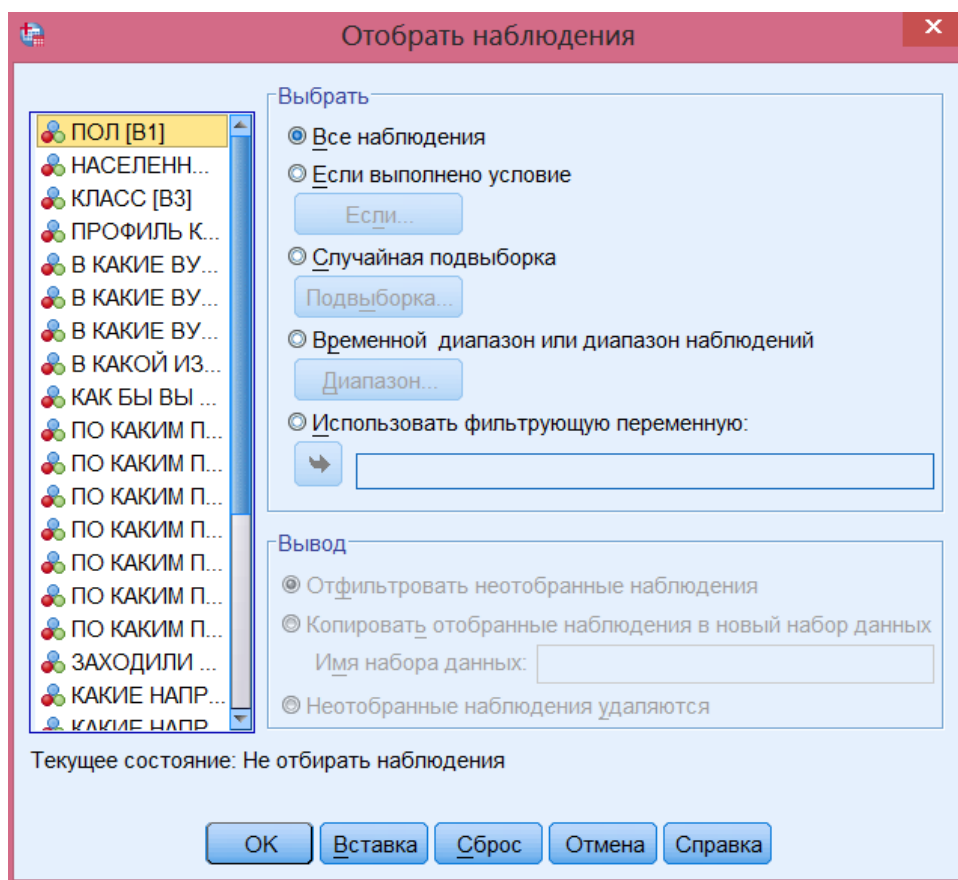


Рис. 12

В диалоговом окне **Отобразить наблюдения** имеются две группы переключателей. В группу *Выбрать* входит 5 переключателей, из которых в данный момент нас интересует только два.

^ Переключатель *Все наблюдения* фактически предназначен для отмены операции, поскольку при его установке выбранными полагаются все объекты файла.

^ Переключатель *Если выполнено условие* позволяет задать условие отбора объектов с помощью кнопки *Если*. При щелчке на этой кнопке появляется диалоговое окно, очень похожее на окно. Чтобы задать условие отбора, вы можете использовать клавиатуру, панель калькулятора, список функций, а также буфер обмена.

Часто используемые операторы (табл. 1):

Таблица 1

=	равно
>	больше
<	меньше
>=	больше или равно
<=	меньше или равно
≠	не равно
&	И
	ИЛИ
~	НЕ

Далее приведены примеры некоторых условий отбора:

$пол = 1$

Отбираются только девушки.

$класс = 3$

Отбираются ученики класса «В».

$класс \leq 3$

Отбираются ученики классов «А» и «Б».

$вуз \geq 2 \ \& \ вуз \leq 3$

$вуз > 1 \ \& \ вуз < 4$

При задании любого из этих условий отбираются ученики, выбирающие экономические и технические вузы.

Чтобы ввести переменную в поле задания условия отбора, достаточно щелкнуть сначала на ней, а затем – на кнопке со стрелкой. Знаки операций можно вводить как с клавиатуры, так и с панели калькулятора.

После проведения анализа выбранной подгруппы вы можете вернуться к полному набору данных, установив в окне **Отобразить наблюдения** переключатель *Все наблюдения*. Номера выбранных объектов, в отличие от исключенных, после выполнения команды **Отобразить наблюдения** не зачеркиваются, что позволяет легко находить их визуально<sup>18</sup>.

### **5. Извлечение случайной выборки.**

При большом количестве наблюдений для экономии времени может быть полезно использовать небольшую случайную выборку при первой предварительной проверке гипотезы. Чтобы извлечь случайную выборку из совокупности всех наблюдений, выполните следующие действия<sup>19</sup>:

**Данные**

**Отобразить наблюдения**

**Случайная подвыборка – Подвыборка...**

### **Использованная литература**

Наследов А.Д. SPSS : Компьютерный анализ в психологии и социальных науках : самоучитель / А.Д. Наследов. – СПб. : Питер, 2005. – 416 с.

Пациорковский В.В., Пациорковская В.В. SPSS для социологов: Учебное пособие. ИСЭПН РАН. – М., 2005. – 433 с.

### **Контрольные вопросы**

1. В каких ситуациях требуется Объединение данных? Слияние файлов данных?
2. Для чего нужна Сортировка наблюдений?

---

<sup>18</sup> Наследов А.Д. SPSS : Компьютерный анализ в психологии и социальных науках : самоучитель / А.Д. Наследов. – СПб. : Питер, 2005. – С. 74–76

<sup>19</sup> Пациорковский В.В., Пациорковская В.В. SPSS для социологов: Учебное пособие. ИСЭПН РАН. – М., 2005. – С. 71

3. Для чего нужно Разделение наблюдений на группы?
4. Для чего нужна операция Выбор наблюдений?
5. Для чего извлекать из БД случайную выборку?

### **Практические задания**

1. Прежде чем, объединить несколько файлов данных в один, создайте копию файла «куратор 2011» и назовите ее «куратор-база2». При помощи команд меню «Копировать» и «Вставить» вставьте пять первых анкет из файла «куратор 2011» в файл «куратор-база2».

2. С помощью процедуры *Добавить наблюдения* добавьте в базу данных «куратор-база2» все анкеты файла «куратор 2011».

3. Откройте базу данных «куратор 2011». С помощью процедуры *Сортировать наблюдения* отсортируйте анкеты по факультетам.

4. С помощью процедуры *Расщепить файл* разделите анкеты на две группы: «юноши» и «девушки».

5. Отмените результат процедуры *Расщепить файл*.

6. С помощью процедуры *Отобразить наблюдения* отберите всех девушек факультета социальных наук, оценивающих деятельность куратора на 6 баллов и выше.

7. Отмените результат процедуры *Отобразить наблюдения*.

8. С помощью процедуры *Случайная выборка* отберите в случайном порядке примерно 50 процентов анкет.

### **1.8. Модификация данных: Перекодирование переменных**

Для проведения анализа часто бывает необходимо выполнить преобразование данных. На основе первоначально собранных данных можно создать новые переменные и изменить кодирование. Подобные преобразования называются модификацией данных.

В ходе анализа данных 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. Перекодирование численных данных необходимо, например, тогда, когда первоначальное разнообразие исходных данных не нужно для последующего анализа. В этом случае перекодирование означает уменьшение объема обрабатываемой информации. Перекодирование данных можно выполнить вручную или автоматически. Мы рассмотрим оба этих метода.

### **1. Ручное перекодирование значений.**

*Перекодировка существующей переменной.*

Иногда у вас может возникнуть необходимость изменить кодирование какой-либо переменной. Это может быть обусловлено двумя причинами.

^ Вы работаете одновременно с несколькими файлами данных, которые создавались разными людьми и содержат одни и те же переменные, но закодированные по-разному.

^ В процессе исследования вам стало ясно, что текущее кодирование какой-либо из переменных можно улучшить.

Перекодировка существующей переменной выполняется с помощью команды:

#### **Преобразовать**

**Перекодировать в те же переменные... – Старые и новые значения...**

Операция перекодирования в те же переменные управляется двумя диалоговыми окнами (рис. 13).

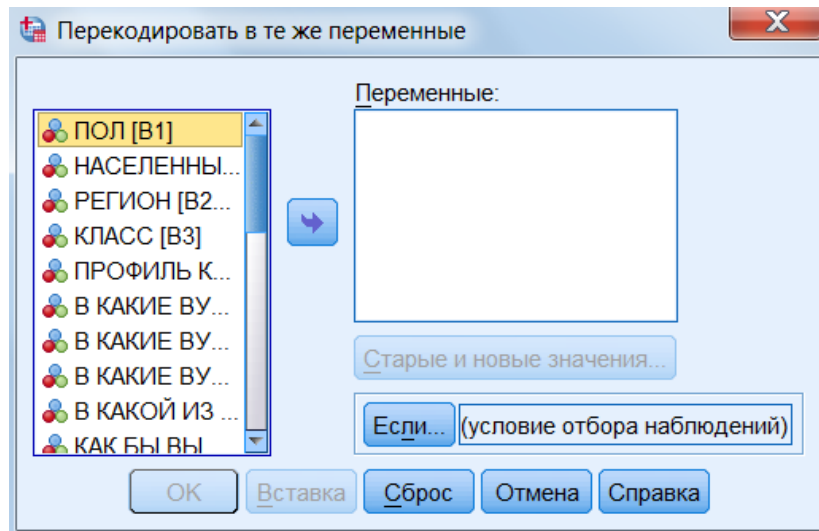


Рис. 13

Первым действием, которое необходимо выполнить в окне **Перекодировать в те же переменные**, является заполнение списка перекодируемых переменных *Переменные*. Имена переменных содержатся в списке слева и выбираются при помощи кнопки со стрелкой.

При щелчке на кнопке *Старые и новые значения* открывается второе диалоговое окно. Далее описаны переключатели группы *Старое значение* (рис. 14).

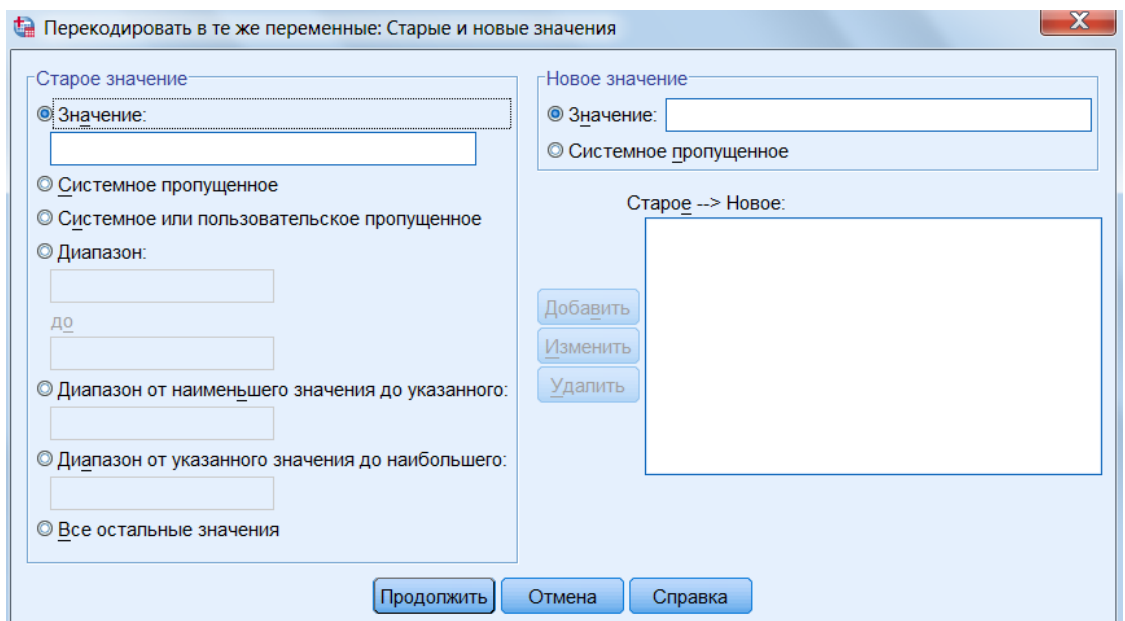


Рис. 14

^ Значение. При установке этого переключателя в поле рядом нужно указать значение.

^ Системное пропущенное значение. Установка этого переключателя означает физическое отсутствие значения.

^ Системное или пользовательское пропущенное значение. Установка этого переключателя означает, что значение либо не задано, либо не может участвовать в анализе.

^ Диапазон. Это имя соответствует сразу трем переключателям, позволяющим тремя способами указать диапазон значений:

> от нижней до верхней границы;

> от наименьшего до заданного значения;

> от заданного до наибольшего значения.

> Все остальные значения. Этот переключатель устанавливается во всех остальных случаях.

Далее, вы задаете одно или несколько старых и новых значений и помещаете их в список Старое-->Новое<sup>20</sup>.

*Перекодировка в новую переменную.*

Перекодировка в новую переменную означает создание новой переменной, значения которой определяются на основе замены множества значений существующей переменной небольшим числом категорий. Как правило, эта процедура применяется, когда необходимо разделить выборку объектов на подгруппы по некоторому количественному признаку.

Команда перекодирования в другие переменные выполняется также с помощью двух диалоговых окон:

### **Преобразовать**

**Перекодировать в другие переменные... – Выходная переменная – Старые и новые значения...**

В первом окне вы можете задать входную и выходную переменные (рис. 15).

---

<sup>20</sup> Наследов А.Д. SPSS : Компьютерный анализ в психологии и социальных науках : самоучитель / А.Д. Наследов. – СПб. : Питер, 2005. – С. 71–73



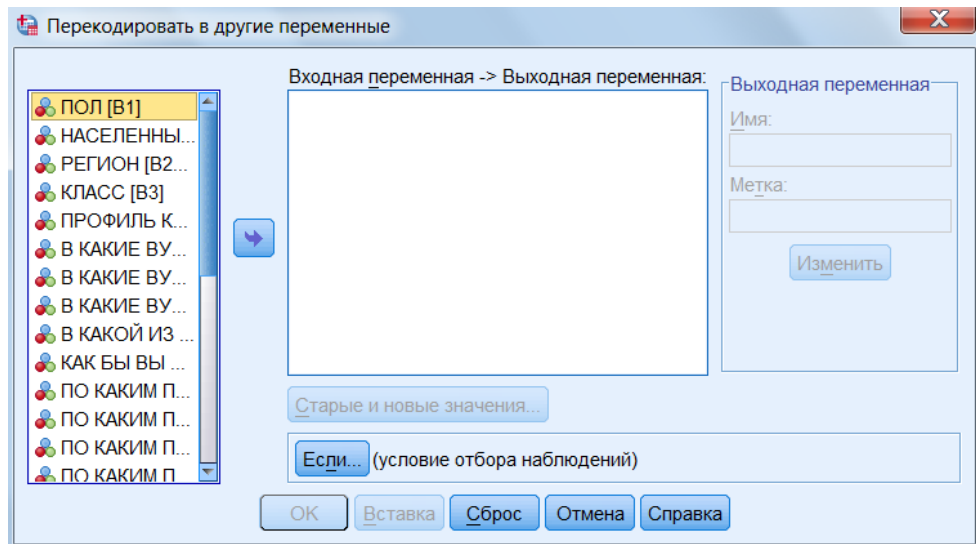


Рис. 15

Окно **Перекодировать в другие переменные** имеет вполне обычный вид: в его левой части находится список доступных переменных. Первое, что необходимо сделать, – выделить переменную, на основе которой создается новая переменная, и переместить ее в список *Входная переменная -> Выходная переменная*. Затем в поле *Имя* области *Выходная переменная* следует ввести имя новой переменной. Щелчок на кнопке *Изменить* приведет к появлению переменной в предыдущем списке.

Щелчок на кнопке *Старые и новые величины* приведет к открытию второго диалогового окна. В нем вы можете задать градации новой переменной, которые будут соответствовать диапазонам уровней старой переменной. Теперь требуется задать соответствие между диапазонами значений исходной переменной и градациями новой переменной.

Новыми производными переменными вы можете управлять так же, как и переменными, созданными «обычным» способом: назначать метки, вручную изменять значения, менять расположение в файле данных и т.п.<sup>21</sup>

## 2. Редактирование меток значений переменной.

Обратите внимание на то, что процедура ручного перекодирования никак не влияет на метки значений. Другими словами, после выполнения ручного

<sup>21</sup> Там же. С. 68–70

перекодирования необходимо заново задать метки всем значениям, изменившим свой смысл<sup>22</sup>.

### 3. Автоматическое перекодирование.

Если категории не были закодированы непрерывно начиная с 1, то это может приводить к негативным последствиям при решении многих задач в SPSS. Поэтому для преобразования значений численных или строковых переменных в непрерывную последовательность целых чисел в SPSS реализована возможность автоматического перекодирования (рис. 16). Часто к автоматическому перекодированию приходится прибегать, когда необходимо текстовую переменную преобразовать в числовую<sup>23</sup>.

#### Преобразовать

#### Автоматическая перекодировка... – Новое имя

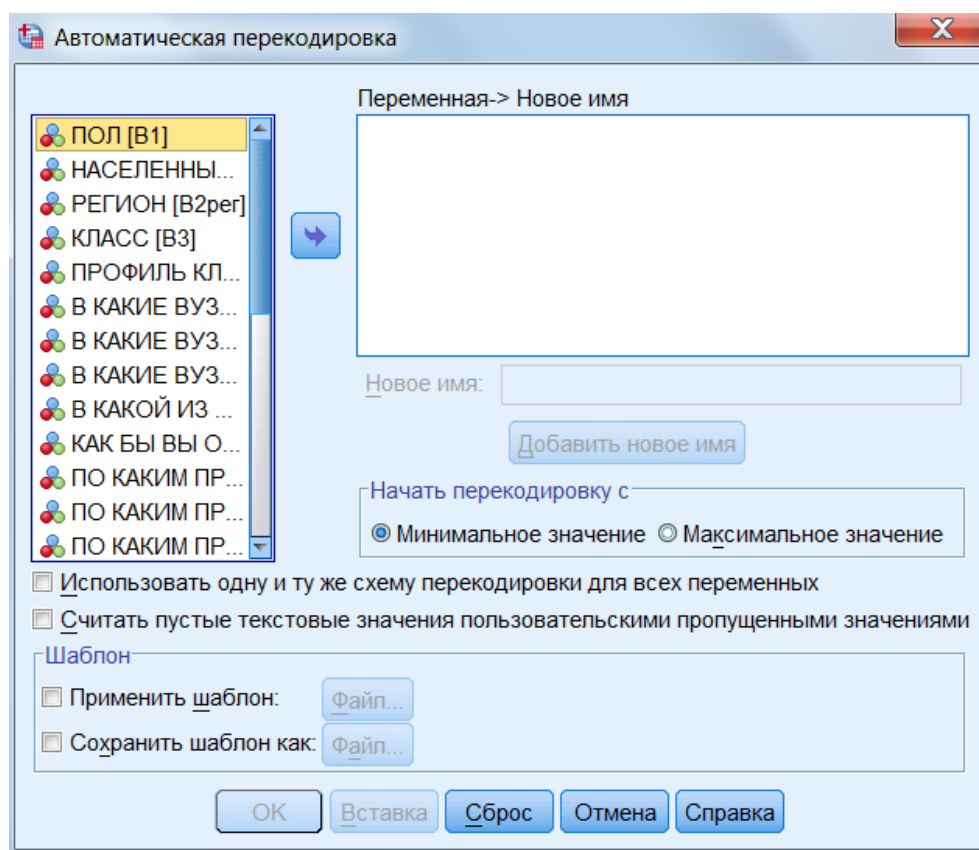


Рис. 16

<sup>22</sup> Там же. С. 73

<sup>23</sup> Бююль А., Цёфель П. SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей : Пер. с нем. / А. Бююль, П. Цёфель. – СПб.: ООО «ДиаСофтЮП», 2005. – С. 135

При решении указанной выше задачи каждой категории строковой переменной будет присвоен свой код, в порядке возрастания или убывания.

### **Использованная литература**

Бююль А., Цёфель П. SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей : Пер. с нем. / А. Бююль, П. Цёфель. – СПб.: ООО «ДиаСофтЮП», 2005. – 608 с.

Наследов А.Д. SPSS : Компьютерный анализ в психологии и социальных науках : самоучитель / А.Д. Наследов. – СПб. : Питер, 2005. – 416 с.

### **Контрольные вопросы**

1. Что такое Перекодирование значений переменной? Когда оно требуется?
2. Опишите алгоритм процедуры Ручное перекодирование?
3. В каких случаях требуется Автоматическое перекодирование?

### **Практические задания**

1. Перекодирование существующей переменной.

А. В базе данных «родители 2010» перекодируйте все переменные табличного вопроса № 10 (v10\_1 ... v10\_12), посвященного оценке созданных в АмГУ условий для получения доступного и качественного образования, следующим образом: ответы «Очень плохо» и «Плохо» преобразовать в один ответ «Плохо»; ответы «Хорошо» и «Отлично» преобразовать в один ответ «Хорошо»; ответ «Затрудняюсь ответить» оставить без изменения.

Б. Во вкладке Переменные внесите соответствующие изменения в метки значений (формулировки ответов) по данным переменным (v10\_1 ... v10\_12).

2. Перекодирование в новую переменную.

А. Перекодируйте переменную «Факультет обучения» (v2) в новую переменную «Направленность обучения» (v17) следующим образом: ИФФ, ФМиИ, ФДиТ, ЭФ объединить в одну категорию «Естественно-техническая»;

ФМО, ФСН, ФФ, ЮФ объединить в одну категорию «Социально-гуманитарная»; ЭкФ отнести к категории «Экономика и управление».

Б. Во вкладке Переменные внесите соответствующие изменения в метки значений (формулировки ответов) новой переменной (v17).

В. Переместите новую переменную (v17) в положение сразу после вопроса о факультете обучения.

### 3. Автоматическое перекодирование.

А. С помощью процедуры *Автоматическая перекодировка* создайте новую переменную (v3n) с порядковыми номерами для переменной «Номер группы» (v3).

Б. Переместите новую переменную (v3n) в положение сразу после вопроса о номере группы.

## 1.9. Модификация данных: Вычисление переменных

Потребность в преобразовании данных может возникнуть практически всегда как на этапе контроля, так и при 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 (программе социологического исследования и инструментарии) позволяет как экономить время и силы при сборе данных, так и избегать, ошибок расчета, допускаемых интервьюерами в вопросах, требующих вычислений (например, средний доход семьи на одного члена семьи в месяц). Во-вторых, задачи создания новых переменных могут возникать и решаться не только на этапе анализа, но и при вводе и контроле данных<sup>24</sup>.

Путем вычислений в SPSS можно образовать новые переменные и добавить их в файл данных (рис. 17):

### Преобразовать

---

<sup>24</sup> Пациорковский В.В., Пациорковская В.В. SPSS для социологов: Учебное пособие. ИСЭПН РАН. – М., 2005. – С. 79

## Вычислить переменную... – Вычисляемая переменная

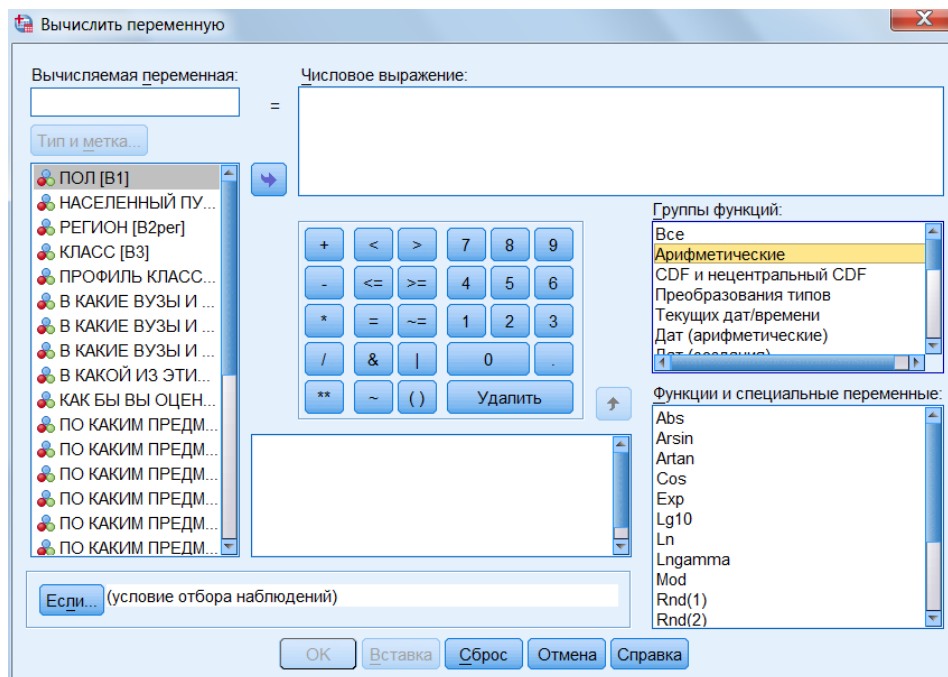


Рис. 17

Диалоговое окно **Вычислить переменную** содержит в левой части уже ставший привычным список доступных переменных. Над списком, в левом верхнем углу окна, имеется поле *Вычисляемая переменная*, в которое необходимо ввести имя создаваемой или вычисляемой переменной. В поле *Числовое выражение* вводится выражение, с помощью которого вычисляется новая переменная. Вы можете вводить в это поле символы с клавиатуры, через буфер обмена, а также пользоваться перечисленными ниже вспомогательными элементами интерфейса окна, позволяющими формировать выражение. В этом выражении могут использоваться имена существующих переменных, константы, арифметические и логические операторы и функции.

### *Арифметические функции*

■ *ABS (numexpr)*: Функция ABS возвращает абсолютное значение (модуль). Если переменная Celsius имеет значение -6,5, ABS (celsius) возвращает 6,5, а ABS (celsius + 3) – значение 3,5.

■ *RND (numexpr)*: Функция RND округляет до ближайшего целого числа. Если переменная celsius имеет значение 3,6, RND (celsius) возвращает 4, а RND (celsius + 6) – значение 10.

■ *TRUNC (numexpr)*: Функция отбрасывает дробную часть значения; округления не происходит. Если переменная *celsius* имеет значение 3,9, *TRUNC (celsius)* возвращает 3, а *TRUNC (celsius + 4)* – значение 7.

■ *MOD (numexpr, modulus)*: Функция *MOD* возвращает остаток от деления первого аргумента (*numexpr*) на второй (*modulus*). Если переменная *jaehr* имеет значение 1994, *MOD (jaehr, 100)* возвращает 94.

■ *SQRT (numexpr)*: Функция *SQRT* возвращает квадратный корень. Если переменная *zahl1* имеет значение 9, *SQRT (zahl1)* возвращает значение 3.

#### *Статистические функции*

Статистические функции могут иметь любое количество параметров.

■ *SUM (numexpr, numexpr,...)*: Функция *SUM* возвращает сумму значений допустимых аргументов. *SUM (zahl1, zahl2, zahl3)* возвращает сумму значений трех переменных.

■ *MEAN (numexpr, numexpr,...)*: Функция *MEAN* возвращает среднее арифметическое допустимых аргументов. *MEAN (42, 19, 29)* возвращает значение 30.

■ *SD (numexpr, numexpr,...)*: Функция *SD* возвращает стандартное отклонение значений допустимых аргументов.

■ *VARIANCE (numexpr, numexpr,...)*: Функция *VARIANCE* возвращает дисперсию значений допустимых аргументов.

■ *MIN (numexpr, numexpr,...)*: Функция *MIN* возвращает наименьшее из значений допустимых аргументов.

■ *MAX (numexpr, numexpr,...)*: Функция *MAX* возвращает наибольшее из значений допустимых аргументов.

Функциям *SUM*, *MEAN*, *MIN* и *MAX* требуется хотя бы один допустимый аргумент. Остальные аргументы могут содержать отсутствующие значения. Если это свойство, принятое по умолчанию, требуется деактивировать, то к имени функции через точку прибавляют количество необходимых аргументов, например, *MEAN. 10*. В этом случае значение функции

вычисляется только тогда, когда существует хотя бы указанное количество аргументов (в данном примере 10).

#### *Функции обработки пропущенных значений*

■ *NMISS (variable,variable,...)*: Функция NMISS возвращает количество пропущенных значений в списке переменных.

■ *NVALID (variable,variable,...)*: Функция NVALID возвращает количество допустимых значений в списке переменных.

В результате выполнения процедуры будет создана новая переменная, которая разместится в крайнем правом столбце файла данных. Если вам удобнее держать данную переменную ближе к началу файла, можете воспользоваться операциями вырезания и вставки в нужную позицию.

После завершения процедуры программа возвращает вас в окно редактора данных, чтобы вы могли просмотреть содержимое новой переменной<sup>25</sup>.

### **Использованная литература**

Наследов А.Д. SPSS : Компьютерный анализ в психологии и социальных науках : самоучитель / А.Д. Наследов. – СПб. : Питер, 2005. – 416 с.

Пациорковский В.В., Пациорковская В.В. SPSS для социологов: Учебное пособие. ИСЭПН РАН. – М., 2005. – 433 с.

### **Контрольные вопросы**

1. Для чего нужны Дополнительные переменные?
2. Как вычислить новые переменные?

### **Практические задания**

1. С помощью процедуры *Вычислить переменную* создайте новую переменную (v9n), в которой бы вычислялось среднее арифметическое оценок

---

<sup>25</sup> Наследов А.Д. SPSS : Компьютерный анализ в психологии и социальных науках : самоучитель / А.Д. Наследов. – СПб. : Питер, 2005. – С. 64–67

эффективности формирования различных качеств студента (переменные в9\_1 ... в9\_12).

2. Какие еще новые переменные можно было вычислить в этой базе данных? Вычислите их.

## 1.10. Модификация данных: Дополнительные процедуры

### 1. Подсчет частоты

В SPSS есть возможность подсчитать количество появления одного и того же значения или значений для определенной переменной<sup>26</sup>. Процедура *Подсчет частоты* предназначена для подсчета повторений одного или нескольких одинаковых значений в списке переменных, в том числе и для подсчета числа повторений в заданном интервале<sup>27</sup>. Например, это позволяет подсчитать количество детей в семье респондента в возрасте до 7 лет (дошкольников), если в анкете имелся вопрос о возрасте всех детей.

#### Преобразовать

**Подсчитать значения в наблюдениях... – Вычисляемая переменная – Задать значения...**

Откроется диалоговое окно **Подсчитать значения в наблюдениях** (рис. 18).

---

<sup>26</sup> Бююль А., Цёфель П. SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей : Пер. с нем. / А. Бююль, П. Цёфель. – СПб.: ООО «ДиаСофтЮП», 2005. – С. 129

<sup>27</sup> Пациорковский В.В., Пациорковская В.В. SPSS для социологов: Учебное пособие. ИСЭПН РАН. – М., 2005. – С. 87



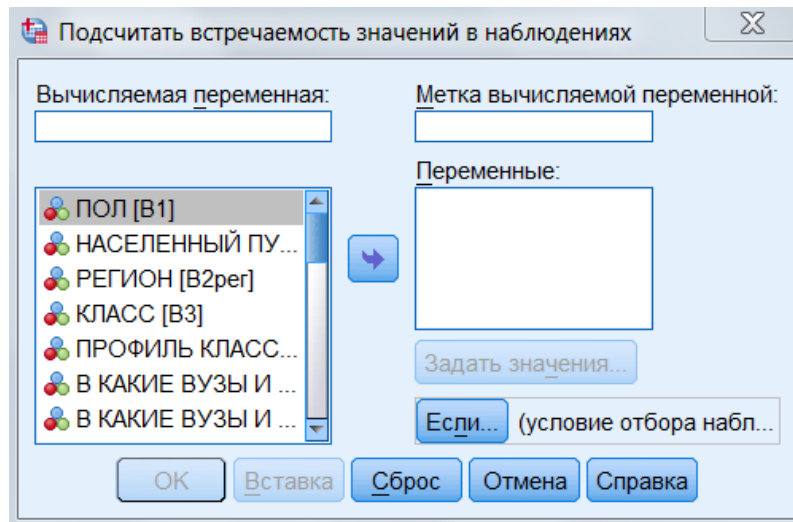


Рис. 18

Это диалоговое окно разделено на следующие части:

- *Вычисляемая переменная:* указывается имя переменной, в которой будут содержаться подсчитанные значения.
- *Метка:* указывается метка для вычисляемой переменной.
- *Переменные:* Этот список содержит переменные, выбранные из списка исходных переменных, хранящихся в файле данных, для которых нужно подсчитать определенные значения. Список не может одновременно содержать численные и строковые переменные.

Щелкните на кнопке *Задать значения* (рис. 19).

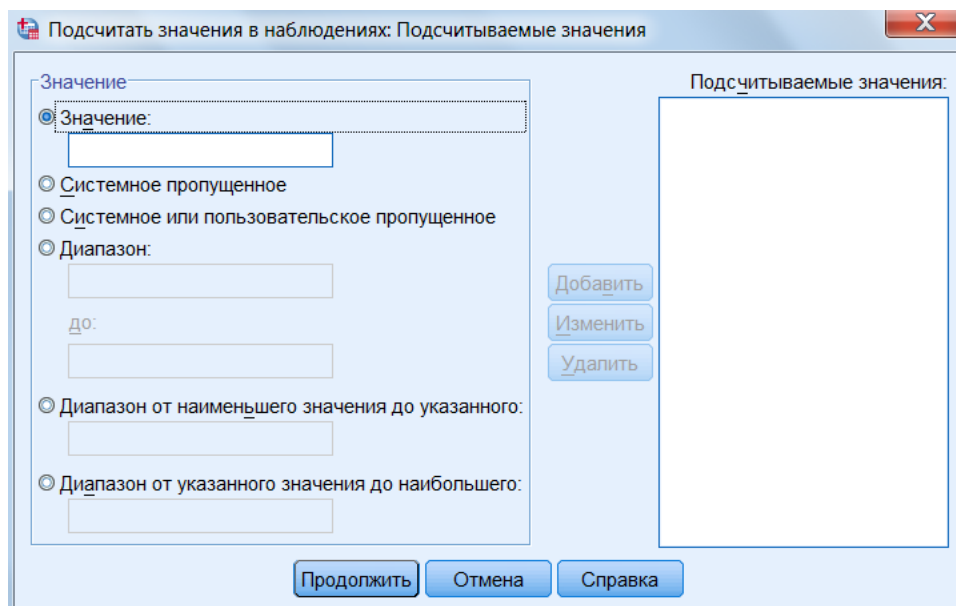


Рис. 19

Откроется диалоговое окно *Подсчитать значения в наблюдениях: подсчитываемые значения*. Это диалоговое окно служит для определения подсчитываемых значений и аналогично окну ручной перекодировки переменных. Можно задать отдельное значение, диапазон или сочетание того и другого<sup>28</sup>. Если требуется подсчитать повторяемость нескольких значений, щелкните после выбора опции на кнопке *Добавить*. В этом случае будет подсчитана частота повторений каждого значения, присутствующего в списке *Переменные*.

## **2. Построение индекса.**

Индексом называют объединение нескольких отдельных вопросов анкеты в едином показателе, который характеризует сложные, многоплановые состояния – например, показатель уровня жизни или уровня интеллекта<sup>29</sup>.

## **3. Агрегирование данных.**

На базе значений одной или нескольких группирующих переменных (переменных разбиения: пол, номер группы, факультет, район...) можно объединить наблюдения в группы (агрегировать) и создать новую переменную или новый файл данных, содержащий по одному значению для каждой группы разбиения. Для этого SPSS предоставляет большое количество функций агрегирования<sup>30</sup>. Агрегирование данных позволяет создавать такие значения переменных, каждое из которых представляет собой результат объединения группы исходных значений, например среднее. В процессе агрегирования задается группирующая переменная (например, пол, номер группы, факультет, район...), каждое значение которой неоднократно встречается в исходных данных. Затем для каждого значения группирующей переменной вычисляются

---

<sup>28</sup> Бююль А., Цёфель П. SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей : Пер. с нем. / А. Бююль, П. Цёфель. – СПб.: ООО «ДиаСофтЮП», 2005. – С. 130–131

<sup>29</sup> Там же. С. 138

<sup>30</sup> Там же. С. 143

новые значения агрегируемых переменных исходя из их исходных значений по заданной функции (например, среднее)<sup>31</sup>.

Иными словами, если вам необходимо присвоить каждому респонденту в определенной группе новое значение, зависящее от ответов всех респондентов, входящих в данную группу, то целесообразно воспользоваться процедурой **Агрегирование данных** (рис. 20):

### Данные

**Агрегировать данные...** – **Группирующие переменные** – **Агрегируемые переменные**

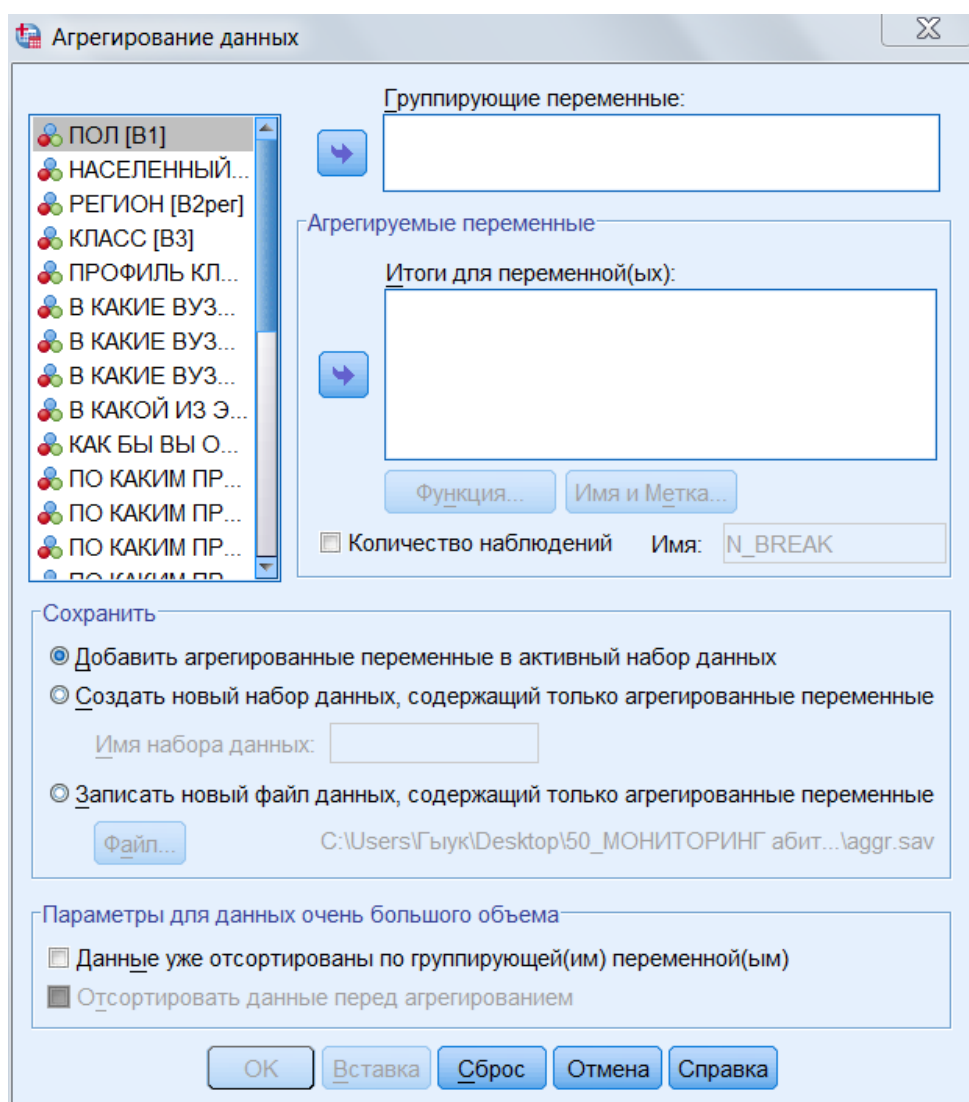


Рис. 20

<sup>31</sup> Наследов А.Д. SPSS 19: профессиональный статистический анализ данных / А.Д. Наследов. – СПб. : Питер, 2011. – С. 83

Например, вы хотите присвоить респонденту, проживающему в конкретном районе, среднее значение дохода всех респондентов, проживающих в данном районе. Тогда, группирующей переменной будет «район», а агрегируемой – «доход респондента». При выполнении данной процедуры появиться новая переменная, в которой каждому респонденту, проживающему в конкретном районе, присвоится значение среднего дохода его земляков.

#### 4. Ранговые преобразования.

В SPSS существует возможность задавать ранги для измеренных значений переменной, вычислять процентные ранги и формировать процентильные группы, добавляя в файл данных соответствующие переменные<sup>32</sup>. В простейшем случае процедура рангового преобразования позволяет проранжировать респондентов по определенному признаку, например доходу.

#### Преобразовать

#### Ранжировать наблюдения... – Типы рангов... – Совпадающие...

Откроется диалоговое окно **Ранжировать наблюдения** (рис. 21). Затем необходимо выбрать переменную, по которой будет происходить ранжирование респондентов, вид ранга и процедуру при совпадении значений.

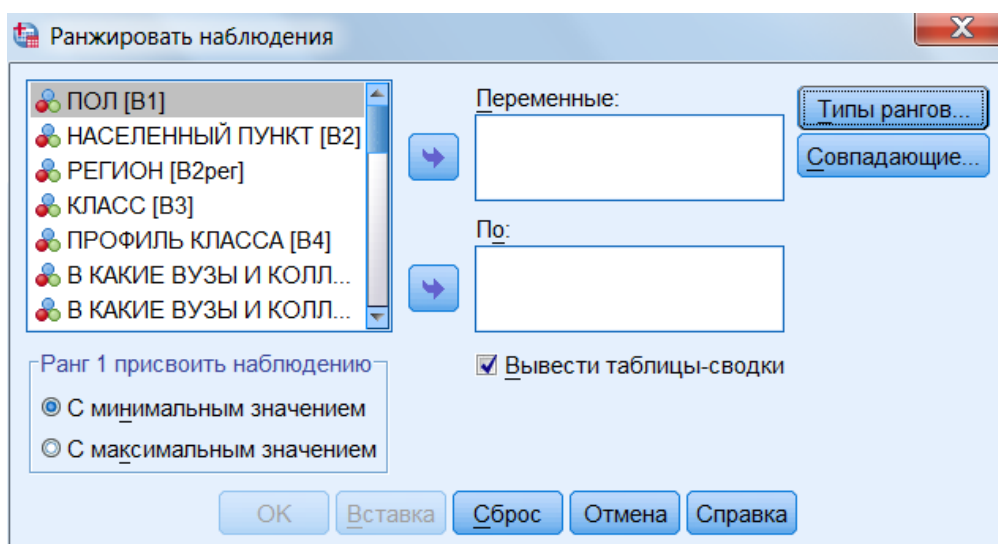


Рис. 21

<sup>32</sup> Наследов А.Д. SPSS : Компьютерный анализ в психологии и социальных науках : самоучитель / А.Д. Наследов. – СПб. : Питер, 2005. – С. 67

Настройки окна *Совпадающие* указывают, как программа будет поступать при появлении одинаковых величин (рис. 22).

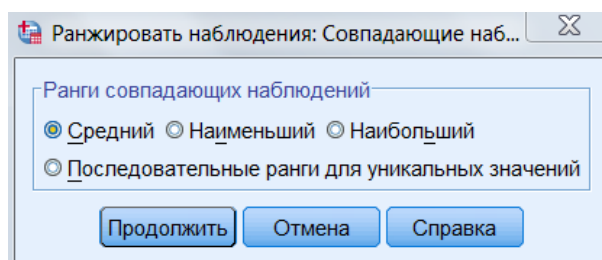


Рис. 22

По умолчанию принято (и, как правило, это наилучший вариант), что присваивается среднее из значений рангов этих величин. При установке *Наименьший* все значения получают наименьший, при установке *Наибольший* – наибольший из этих рангов. При выбранной опции *Последовательные ранги* все связанные наблюдения получают одинаковый ранг; следующему наблюдению присваивается следующее по порядку целое число. Поэтому максимальный присвоенный ранг равен не общему количеству значений, а количеству различных значений.

Перечисленные четыре способа присвоения рангов можно пояснить с помощью простого примера, в котором семь значений расположены по убыванию (табл. 2)<sup>33</sup>:

Таблица 2

<i>Значение</i>	<i>Средний</i>	<i>Наименьший</i>	<i>Наибольший</i>	<i>Последовательный</i>
190	1	1	1	1
187	2,5	2	3	2
187	2,5	2	3	2
185	5	4	6	3
185	5	4	6	3
185	5	4	6	3
184	7	7	7	4

В диалоге **Ранжировать наблюдения** можно, щелкнув на кнопке *Типы рангов*, открыть диалоговое окно *Ранжировать наблюдения: Типы рангов*. В этом окне представлены шесть типов рангов.

<sup>33</sup> Бююль А., Цёфель П. SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей : Пер. с нем. / А. Бююль, П. Цёфель. – СПб.: ООО «ДиаСофтЮП», 2005. – С. 146–147

## 5. Взвешивание наблюдений.

SPSS предоставляет возможность определения веса данных. При этом данным, относящимся к разным наблюдениям, присваиваются различные весовые коэффициенты посредством так называемой переменной взвешивания. Эта процедура может быть полезной в следующих ситуациях:

- Данная выборка не является репрезентативной, то есть частотные характеристики выборки, состоящей из переменных, достаточных для обеспечения репрезентативности, не соответствуют частотным характеристикам генеральной совокупности.

- Анализ данных, которые уже представлены в виде частотных таблиц<sup>34</sup>.

*Коррекция при отсутствии репрезентативности.*

Принцип заключается в том, что для каждого значения переменной вычисляется весовой коэффициент как отношение необходимого значения в генеральной совокупности к существующему в реализованной выборке:

*Весовой коэффициент = необходимое значение / существующее значение.*

Например, вместо 46% мужчин и 54% женщин (статистическое распределение в генеральной совокупности) интервьюеры опросили 49% мужчин и 51% женщин (реализованная выборка). Это происходит в результате некачественной работы интервьюеров. Исправить данную погрешность можно следующим образом.

Сначала рассчитываются весовые коэффициенты для мужчин и женщин по указанной выше формуле:

Вес для мужчин =  $46 / 43 = 1,0698$

Вес для женщин =  $54 / 57 = 0,9474$

Далее вводим новую переменную ВЕС\_пола, в которой мужчинам присваиваем значение веса для мужчин, а женщинам – для женщин.

Далее с помощью команд

**Данные**

**Взвесить наблюдения...**

---

<sup>34</sup> Там же. С. 150–151

выбрать переменную для взвешивания (ВЕС\_пола) и «перевзвесить» респондентов (рис. 23).

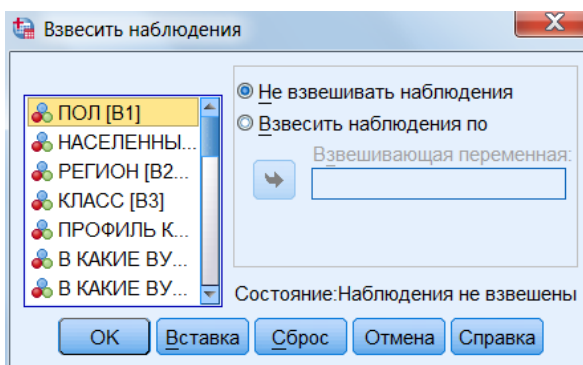


Рис. 23

### Использованная литература

Бююль А., Цёфель П. SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей : Пер. с нем. / А. Бююль, П. Цёфель. – СПб.: ООО «ДиаСофтЮП», 2005. – 608 с.

Наследов А.Д. SPSS: Компьютерный анализ в психологии и социальных науках : самоучитель / А.Д. Наследов. – СПб. : Питер, 2005. – 416 с.

Наследов А.Д. SPSS 19: профессиональный статистический анализ данных / А.Д. Наследов. – СПб. : Питер, 2011. – 400 с.

Пациорковский В.В., Пациорковская В.В. SPSS для социологов: Учебное пособие. ИСЭПН РАН. – М., 2005. – 433 с.

### Контрольные вопросы

1. Для чего нужен Подсчет частоты появлений определенных значений?
2. Для чего нужно Построение индекса?
3. Что такое Агрегирование данных?
4. Назовите основные Типы рангов
5. В каких ситуациях требуется Взвешивание наблюдений?

### Практические задания

1. Откройте базу данных «первокурсники 2010». С помощью процедуры *Подсчет значений в наблюдениях* введите новую переменную, в которой

указывалось бы количество способов подготовки к поступлению в АмГУ, использованных респондентом.

2. С помощью процедуры *Агрегировать данные* введите новую переменную, в которой респонденту присваивалось бы среднее значение сложности сдаваемых экзаменов на его факультете.

3. Проранжируйте следующие данные о балле ЕГЭ по обществознанию, используя различные способы присвоения рангов:

95	48	48	74	62	62	62	100	51	62
----	----	----	----	----	----	----	-----	----	----

Результаты оформите в таблицу по образцу ниже:

Значение	Средний	Наименьший	Наибольший	Последовательный
190	1	1	1	1
187	2,5	2	3	2
187	2,5	2	3	2
185	5	4	6	3
185	5	4	6	3
185	5	4	6	3
184	7	7	7	4

4. Вычислите веса для переменной «возраст», имея следующие данные:

Возрастная группа	Необходимое значение в генеральной совокупности (%)	Существующее значение в реализованной выборке (%)
18-24 года	16	23
25-29 лет	12	8
30-39 лет	19	9
40-49 лет	17	21
50-59 лет	19	22
60 лет и старше	17	17

Результаты оформите в таблицу по образцу ниже (значения весов округлите до сотых):

Возрастная группа	Весовой коэффициент (вес)
18-24 года	1,12
25-29 лет	1,00
30-39 лет	0,25
40-49 лет	0,78
50-59 лет	2,09
60 лет и старше	2,75

5. Придумайте индекс, объединяющий несколько отдельных переменных в базе данных «первокурсники 2010» в едином показателе.



## ГЛАВА 2. ЭТАП СТАТИСТИЧЕСКОЙ ОБРАБОТКИ ДАННЫХ

### 2.1. Основы статистики и исследование данных

Владение приемами работы с такой программой, как SPSS требует предварительных познаний в области статистики. Здесь мы коротко остановимся на некоторых основных понятиях, с которыми непременно должен быть знаком пользователь, если он хочет использовать SPSS. В первую очередь сюда относятся предварительные оценки, которые выполняются перед проведением любого статистического теста: классификация переменных по статистическим шкалам, проверка наличия нормального распределения и выделение независимых и зависимых выборок.

#### **Предварительные условия для проведения статистического теста.**

В большинстве случаев перед применением статистического теста ставится вопрос: каков характер заданных условий? В частности, необходимо выяснить следующие моменты:

- К какой статистической шкале относится данная переменная?
- Если речь идёт о переменных с интервальной шкалой, то подчиняются ли они закону нормального распределения?
- Являются ли сравниваемые выборки зависимыми или независимыми?

#### *Типы статистических шкал*

В эмпирическом исследовании могут встречаться, к примеру, следующие переменные (табл. 3):

Таблица 3

Вопрос	Варианты ответов и их коды
Пол респондента	1 = мужской 2 = женский
Семейное положение респондента	1 = холост / не замужем 2 = женат / замужем 3 = вдовец / вдова 4 = разведен(а)
Курение	1 = некурящий 2 = изредка курящий 3 = интенсивно курящий 4 = очень интенсивно курящий
Личный доход в месяц	1 = до 15000 р. 2 = от 15000 до 45000 р. 3 = от 45000 р.
Коэффициент интеллекта	
Возраст	

Рассмотрим сначала графу Пол. Мы видим, что назначение соответствия цифр 1 и 2 обоим полам абсолютно произвольно, их можно было поменять местами или обозначить другими цифрами

Мы, конечно, не имеем в виду, что женщины стоят на ступеньку ниже мужчин, или что мужчины значат меньше, чем женщины. Следовательно, отдельным числам не соответствует никакого эмпирического значения. В этом случае говорят о переменных, относящихся к номинальной шкале. В нашем примере рассматривается переменная с номинальной шкалой, имеющая две категории. Такая переменная имеет еще одно название – дихотомическая.

Такая же ситуация и с переменной Семейное положение. Здесь также соответствие между числами и категориями семейного положения не имеет никакого эмпирического значения. Но в отличие от Пола, эта переменная не является дихотомической – у нее четыре категории вместо двух.

Возможности обработки переменных, относящихся к номинальной шкале очень ограничены. Собственно говоря, можно провести только частотный анализ таких переменных. К примеру, расчет среднего значения для переменной Семейное положение, совершенно бессмысленен. Переменные, относящиеся к номинальной шкале часто используются для группировки, с помощью которых совокупная выборка разбивается по категориям этих

переменных. В частичных выборках проводятся одинаковые статистические тесты, результаты которых затем сравниваются друг с другом.

В качестве следующего примера рассмотрим переменную Курение. Здесь кодовым цифрам присваивается эмпирическое значение в том порядке, в котором они расположены в списке. Переменная Курение, в итоге, сортирована в порядке значимости снизу вверх: умеренный курильщик курит больше, нежели некурящий, а сильно курящий – больше, чем умеренный курильщик и т.д. Такие переменные, для которых используются численные значения, соответствующие постепенному изменению эмпирической значимости, относятся к порядковой шкале.

Однако эмпирическая значимость этих переменных не зависит от разницы между соседними численными значениями. Так, несмотря на то, что разница между значениями кодовых чисел для некурящего и изредка курящего и изредка курящего и интенсивно курящего в обоих случаях равна единице, нельзя утверждать, что фактическое различие между некурящим и изредка курящим и между изредка курящим и интенсивно курящим одинаково. Для этого данные понятия слишком расплывчаты.

К классическими примерами переменных с порядковой шкалой относятся также переменные, полученные в результате объединения величин в классы, как Месячный доход в нашем примере.

Кроме частотного анализа, переменные с порядковой шкалой допускают также вычисление определенных статистических характеристик, таких как медианы. В некоторых случаях возможно вычисление среднего значения. Если должна быть установлена связь (корреляция) с другими переменными такого рода, для этой цели можно использовать коэффициент ранговой корреляции.

Рассмотрим теперь коэффициент интеллекта (IQ). Не только его абсолютные значения отображают порядковое отношение между респондентами, но и разница между двумя значениями также имеет эмпирическую значимость. Например, если у Ганса IQ равен 80, у Фрица – 120 и у Отто – 160, можно сказать, что Фриц в сравнении с Гансом настолько же

интеллектуальнее насколько Отто в сравнении с Фрицем (а именно – на 40 единиц IQ). Однако, основываясь только на том, что значение IQ у Ганса в два раза меньше, чем у Отто, исходя из определения IQ нельзя сделать вывод, что Отто вдвое умнее Ганса.

Такие переменные, у которых разность (интервал) между двумя значениями имеет эмпирическую значимость, относятся к интервальной шкале. Они могут обрабатываться любыми статистическим методами без ограничений. Так, к примеру, среднее значение является полноценным статистическим показателем для характеристики таких переменных.

Наконец, мы достигли наивысшей статистической шкалы, на которой эмпирическую значимость приобретает и отношение двух значений. Примером переменной, относящейся к такой шкале является возраст: если Макс 30 лет, а Морицу 60, можно сказать, что Мориц вдвое старше Макса. Шкала, к которой относятся данные называется шкалой отношений. К этой шкале относятся все интервальные переменные, которые имеют абсолютную нулевую точку. Поэтому переменные относящиеся к интервальной шкале, как правило, имеют и шкалу отношений.

Подводя итоги, можно сказать, что существует четыре вида статистических шкал, на которых могут сравниваться численные значения (табл. 4):

*Таблица 4*

<i>Статистическая шкала</i>	<i>Эмпирическая значимость</i>
Номинальная	Нет
Порядковая	Порядок чисел
Интервальная	Разность чисел
Шкала отношений	Отношение чисел

На практике, в том числе в SPSS, различие между переменными, относящимися к интервальной шкале и шкале отношений обычно несущественно. То есть в дальнейшем практически всегда речь будет идти о переменных, относящихся к интервальной шкале.

Пользователь SPSS должен четко разбираться в видах статистических шкал и при выборе метода обращать внимание на то, чтобы были определены надлежащие виды шкал<sup>35</sup>.

*Закон распределения. Нормальное распределение.*

Многочисленные методы, с помощью которых обрабатываются переменные, относящиеся к интервальной шкале, исходят из гипотезы, что их значения подчиняются нормальному распределению. При таком распределении большая часть значений группируется около некоторого среднего значения, по обе стороны от которого частота наблюдений равномерно снижается (рис. 24).

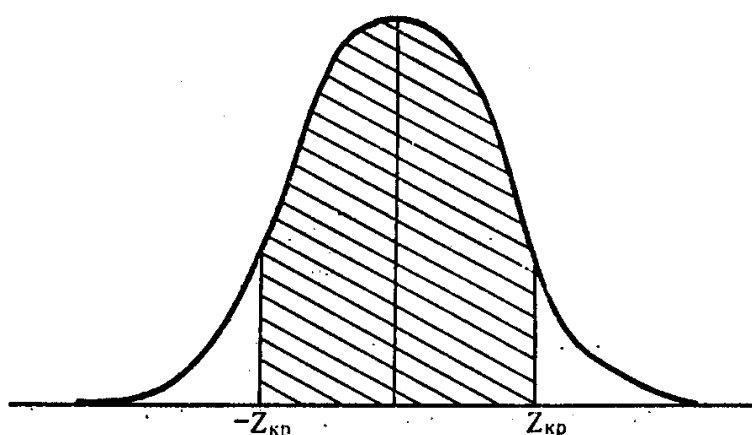


Рис. 24

Перед применением любого метода, который предполагает существование нормального распределения, наличие последнего нужно проверять в первую очередь (рис. 25).

## **Анализ**

### **Описательные статистики**

#### **Частоты**

**Диаграммы... – Гистограммы (Показать на гистограмме нормальную кривую)**

---

<sup>35</sup> Бююль А., Цёфель П. SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей : Пер. с нем. / А. Бююль, П. Цёфель. – СПб.: ООО «ДиаСофтЮП», 2005. – С. 82–84

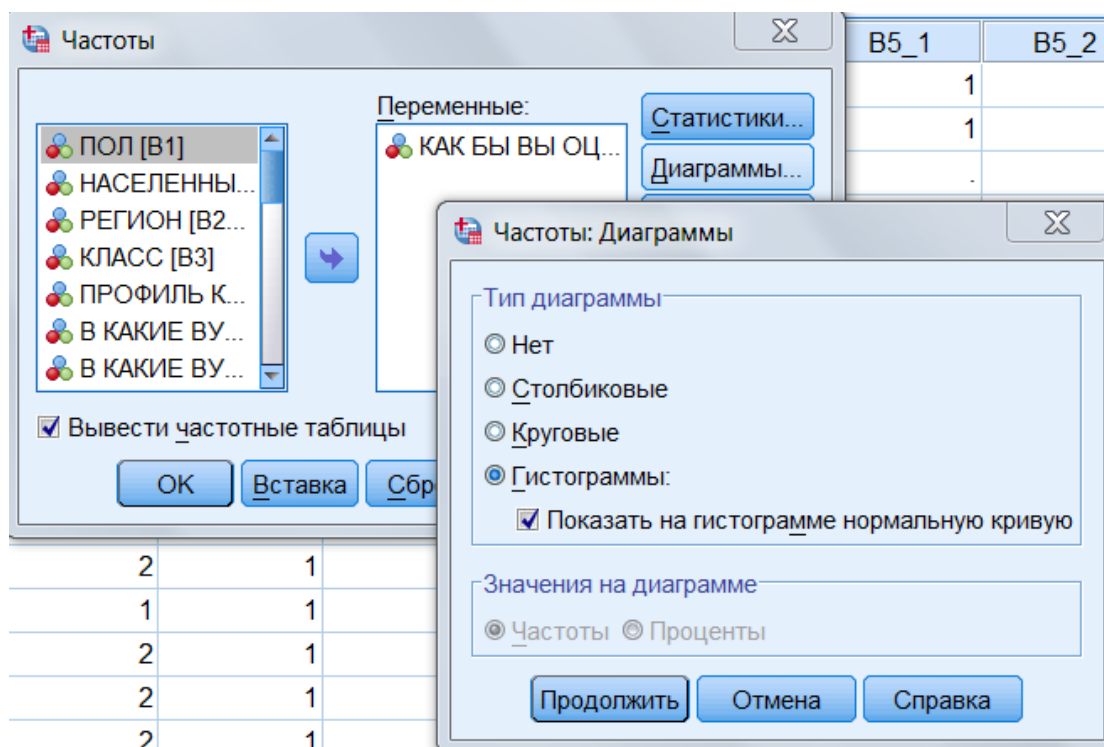


Рис. 25

Если визуальное сравнение реальной гистограммы с кривой нормального распределения кажется недостаточным, можно применить тест Колмогорова-Смирнова, который находится в меню *Анализ* в наборе непараметрических тестов (рис. 26)<sup>36</sup>.

## Анализ

### Непараметрические критерии

#### Устаревшие диалоговые окна

#### Одновыборочный Колмогорова-Смирнова...

<sup>36</sup> Там же. С. 85

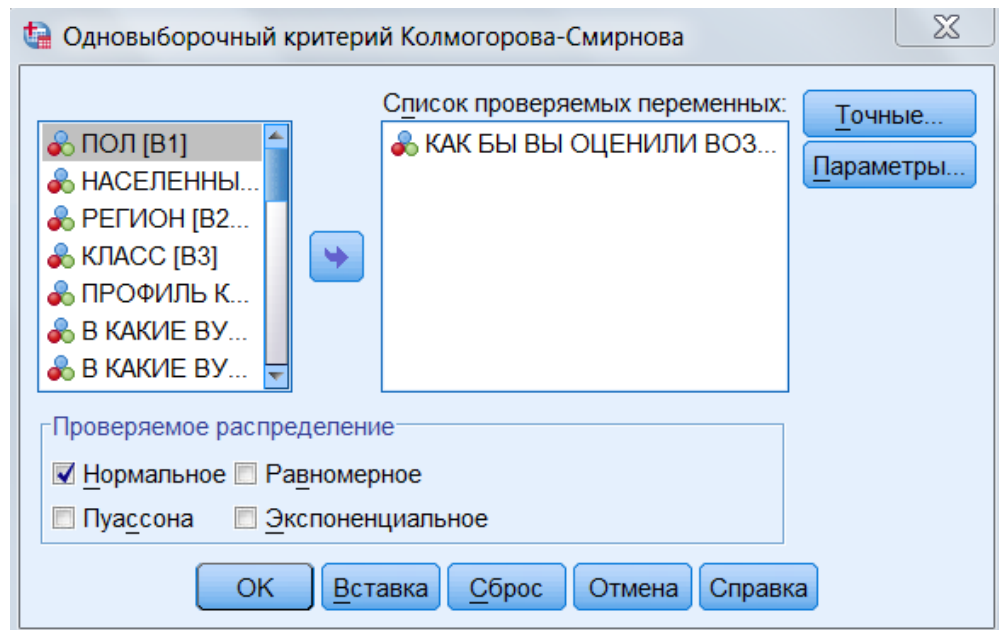


Рис. 26

### Статистические методы.

Прежде чем мы сможем применить статистические методы или строить графики, естественно, следует представить собранные данные в форме, пригодной для обработки. Теперь можно начинать статистическую обработку введенных данных. Учтите, что анализ может быть выполнен только для наблюдений, сгруппированных определенным образом<sup>37</sup>.

#### *Описательный (дескриптивный) анализ.*

Этот вид анализа включает описательное представление отдельных переменных. К нему относятся создание частотной таблицы, вычисление статистических характеристик или графическое представление. Частотные таблицы строятся для переменных, относящихся к номинальной шкале и для порядковых переменных, имеющих не слишком много категорий.

Для переменных относящихся к номинальной шкале нельзя вычислить никаких значимых статистических характеристик. Наиболее часто для порядковых переменных и переменных, относящихся к интервальной шкале, но не подчиняющихся нормальному распределению, вычисляются медианы и оба

<sup>37</sup> Там же. С. 88–89

квартиля; при небольшом числе категорий можно использовать вариант для концентрированных данных.

Для переменных, относящихся к интервальной шкале и подчиняющихся нормальному распределению, чаще всего вычисляется среднее значение и стандартное отклонение или стандартная ошибка. Однако следует выбрать только одну из этих двух характеристик разброса. Для переменных, относящихся ко всем статистическим шкалам, можно построить большое разнообразных графиков, на которых представлены частоты, средние значения или другие характеристики.

#### *Аналитическая статистика.*

Практически любой статистический анализ наряду с чисто описательными операциями включает те или иные аналитические методы (тесты значимости), при применении которых в конечном счете определяется вероятности ошибки  $p$ .

Большая группа тестов служит для выяснения того, различаются ли две или более различных выборки по своим средним значениям или медианам. При этом учитывается разница между независимыми выборками (разные наблюдения) и зависимыми выборками. В зависимости количества выборок (две или более), от того, зависимы ли выборки или нет, относятся ли переменные к интервальной или порядковой шкале, подчиняются ли нормальному распределению – применяются специализированные тесты.

Очень часто встречается ситуация, когда сравниваются различные группы наблюдений или значений переменных, относящихся к номинальной шкале. В этом случае строятся таблицы сопряженности. Другая группа тестов касается исследования связей между двумя переменными, то есть выявления корреляций и восстановления регрессий.

Кроме этих довольно простых статистических методов существуют также более сложные методы многомерного анализа, в которых обычно одновременно используется очень много переменных. К примеру, если требуется свести большое количество переменных к меньшему количеству «пучков



переменных», называемых факторами, то проводится факторный анализ. Если же наша цель, противоположна – объединить заданные наблюдения, образовав из них кластеры, то применяется кластерный анализ.

В определенной группе многомерных тестов вводится различие между зависимой переменной, называемой также целевой и несколькими независимыми переменными (переменными влияния или прогнозирования).

При мультиномиальной логистической регрессии и порядковой регрессии могут также использоваться ковариации, относящиеся к интервальной шкале. Независимые переменные, относящиеся к номинальной шкале, при двоичной логистической регрессии, дискриминантом анализе и многозначном регрессионном анализе должны быть дихотомическими либо раскладываться на набор дихотомических переменных. Логит-логарифмические линейные модели рассматриваются не в этой книге, а во втором томе, посвященном методам исследования рынка и общественного мнения.

Кроме упомянутых здесь, существует еще несколько методов анализа, например, пробит анализ или анализ надежности<sup>38</sup>.

### **Статистические процедуры в SPSS.**

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. При этом действия социолога-аналитика направлялись следующим образом: *измеряемый признак - шкала измерения - допустимая статистика.*

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

---

<sup>38</sup> Там же. С. 89–90

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 регламентируются следующим образом: ***переменная - допустимая статистика - выполнение процедуры.***

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<sup>39</sup>.

### **Использованная литература**

Бююль А., Цёфель П. SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей : Пер. с нем. / А. Бююль, П. Цёфель. – СПб.: ООО «ДиаСофтЮП», 2005. – 608 с.

Пациорковский В.В., Пациорковская В.В. SPSS для социологов: Учебное пособие. ИСЭПН РАН. – М., 2005. – 433 с.

---

<sup>39</sup> Пациорковский В.В., Пациорковская В.В. SPSS для социологов: Учебное пособие. ИСЭПН РАН. – М., 2005. – С. 128–130

### **Контрольные вопросы**

1. Назовите основные группы Статистических методов, реализуемых в SPSS
2. Какие существуют Типы статистических шкал?
3. Что такое Нормальное распределение?

### **Практические задания**

1. Воспользовавшись статистическим сборником (Регионы России. Социально-экономические показатели. 2011 : Стат. сб. / Росстат. – М. : 2011. – 990 с.), скопируйте в пустую базу данных SPSS любой показатель по всем регионам и проверьте при помощи построения гистограммы нормальность распределения данного показателя.
2. Проверьте нормальность распределения данного показателя при помощи теста Колмогорова-Смирнова.

## **2.2. Статистические методы исследования данных**

**Предварительные условия для проведения статистического теста**  
(Проверка закона распределения)

В качестве примера рассмотрим нормальное распределение возраста (рис. 27):

### **Анализ**

#### **Описательные статистики**

##### **Частоты...**

- Статистики (Экссесс и Асимметрия)
- Диаграммы (Гистограммы с нормальной кривой).

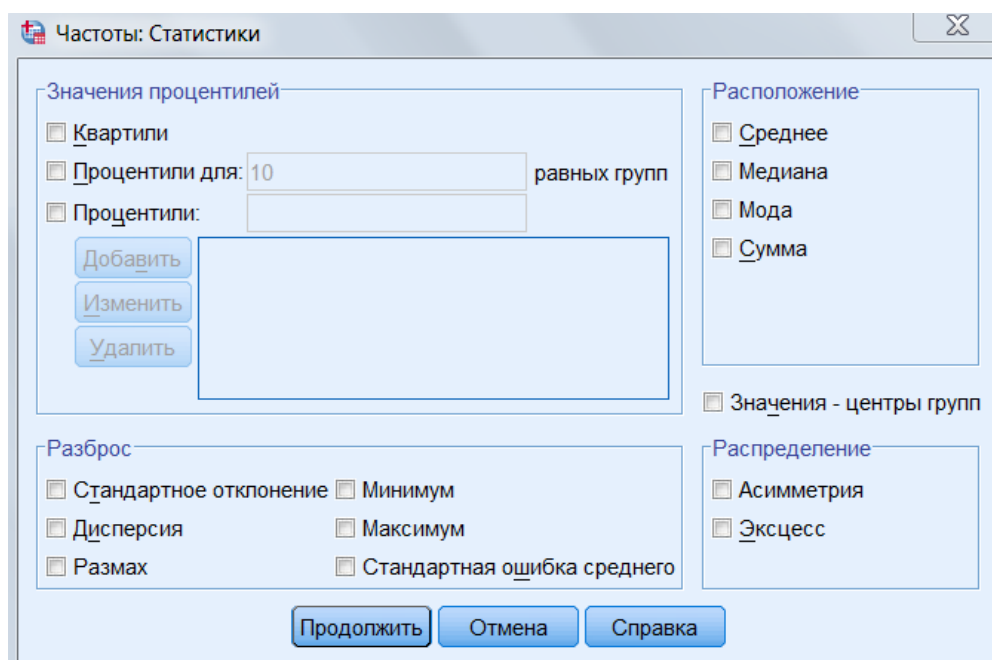


Рис. 27

На диаграмме нанесена кривая нормального распределения (Колокол Гаусса). Реальное распределение в большей или меньшей степени отклоняется от этой идеальной кривой. Выборки, строго подчиняющиеся нормальному распределению, на практике, как правило, не встречаются. Поэтому почти всегда необходимо выяснить, можно ли реальное распределение считать нормальным и насколько значительно заданное распределение отличается от нормального<sup>40</sup>.

*Эксцесс* является мерой «сглаженности» («остро-» или «плосковершинности») распределения. Если значение эксцесса близко к 0, это означает, что форма распределения близка к нормальному виду. Положительный эксцесс указывает на «плосковершинное» распределение, у которого максимум вероятности выражен не столь ярко, как у нормального. Значения эксцесса, превышающие 5,0, говорят о том, что по краям распределения находится больше значений, чем вокруг среднего. Отрицательный эксцесс, напротив, характеризует «островершинное»

<sup>40</sup> Бююль А., Цёфель П. SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей : Пер. с нем. / А. Бююль, П. Цёфель. – СПб.: ООО «ДиаСофтЮП», 2005. – С. 85

распределение, график которого более вытянут по вертикальной оси, чем график нормального распределения. Считается, что распределение с эксцессом в диапазоне от -1 до +1 примерно соответствует нормальному виду. В большинстве случаев вполне допустимо считать нормальным распределение с эксцессом, по модулю не превосходящим 2.

*Асимметрия* показывает, в какую сторону относительно среднего сдвинуто большинство значений распределения. Нулевое значение асимметрии означает симметричность распределения относительно среднего значения, положительная асимметрия указывает на сдвиг распределения в сторону меньших значений, а отрицательная асимметрия – в сторону больших значений. В большинстве случаев за нормальное принимается распределение с асимметрией, лежащей в пределах от -1 до +1. В исследованиях, не требующих высокой точности результатов, нормальным считают распределение с асимметрией, по модулю не превосходящей 2<sup>41</sup>.

### **Статистические процедуры в SPSS.**

SPSS представляет возможности использования ряда статистических процедур для анализа социологической информации. Ниже приведен перечень процедур пункта меню *Анализ* с кратким описанием их содержания.

**Отчеты** – предоставляют наиболее общую информацию о базе данных: метка, тип, шкала измерения переменной и ее допустимые значения, а также введенные данные в первоначальном виде.

**Описательные статистики** – содержат в себе ряд описательных статистических процедур:

**Частоты.** Являются 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

---

<sup>41</sup> Наследов А.Д. SPSS: Компьютерный анализ в психологии и социальных науках : самоучитель / А.Д. Наследов. – СПб. : Питер, 2005. – С. 106–107

переменных. Таблицы частот пригодны для суммирования и отражения данных.

*Описательные.* Дают описание средних, квадратичного отклонения, дисперсии и др. статистик для нормального распределения, а также минимальное значение, размах и сумму для асимметричного распределения количественной переменной.

*Разведочный анализ.* Дает возможность описания подмножеств наблюдений с помощью разнообразных статистик (подсчет частот и процентов, средних и др.) и графиков.

*Таблицы сопряженности.* Позволяет вскрыть сопряженность переменных. Полученные двумерные таблицы показывают частоту встречаемости одной переменной в зависимости от другой.

*Сравнение средних.* Предназначено для проверки различных гипотез о средних значениях количественных переменных.

*Корреляции.* Позволяет установить меру линейной связи между двумя переменными.

*Регрессия.* Показывает зависимость среднего значения результативного признака (зависимой переменной) от одного или нескольких факторов (независимых переменных).

*Классификация.* Позволяет объединить заданные наблюдения, образовав из них кластеры по определенным признакам.

*Снижение размерности.* Позволяет свести большое количество переменных к меньшему количеству "пучков переменных", называемых факторами<sup>42</sup>.

---

<sup>42</sup> Пациорковский В.В., Пациорковская В.В. SPSS для социологов: Учебное пособие. ИСЭПН РАН. – М., 2005. – С. 134–135

## **Использованная литература**

Бююль А., Цёфель П. SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей : Пер. с нем. / А. Бююль, П. Цёфель. – СПб.: ООО «ДиаСофтЮП», 2005. – 608 с.

Наследов А.Д. SPSS: Компьютерный анализ в психологии и социальных науках : самоучитель / А.Д. Наследов. – СПб. : Питер, 2005. – 416 с.

Пациорковский В.В., Пациорковская В.В. SPSS для социологов: Учебное пособие. ИСЭПН РАН. – М., 2005. – 433 с.

## **Контрольные вопросы**

1. Как при помощи гистограммы проверить нормальность распределения значений переменной?
2. Что такое асимметрия частотного распределения?
3. Что такое эксцесс частотного распределения?

## **Практические задания**

1. Воспользовавшись статистическим сборником (Регионы России. Социально-экономические показатели. 2011 : Стат. сб. / Росстат. – М. : 2011. – 990 с.), скопируйте в пустую базу данных SPSS любой показатель по всем регионам и проверьте при помощи построения гистограммы нормальность распределения данного показателя.

2. Проверьте нормальность распределения данного показателя при помощи вычисления *Асимметрии* и *Эксцесса*.

## **2.3. Частотный анализ**

1. Первым этапом статистического анализа данных, как правило, является **частотный анализ**. Анализ частотных распределений результатов количественного социологического исследования – это первый шаг при обработке собранной информации. Во многих случаях этот анализ не является,

строго говоря, анализом данных, а выполняет функции получения общих представлений об изучаемых социальных группах.

Первый шаг одномерного описательного анализа для объяснения какого-то явления – его описание. Результаты любого массового опроса содержат ответы большого числа респондентов на широкий круг анкетных вопросов. Даже в рамках только одного вопроса анкеты объем исходной информации достаточно велик для того, чтобы можно было охватить его одним взглядом и каким-то образом суммировать. Именно задачу сжатия исходной информации, компактного ее представления для дальнейшего осмысления и решают методы одномерного описательного анализа.

Одномерный описательный анализ решает поставленную задачу взаимодополняющими методами:

- построения частотных распределений;
- графического представления поведения анализируемой переменной;
- получения статистических характеристик распределения анализируемой переменной<sup>43</sup>.

Процедура **Частоты** позволяет строить статистические ряды распределения. 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.

---

<sup>43</sup> Крыштановский А.О. Анализ социологических данных с помощью пакета SPSS : учеб. пособие для вузов / А.О. Крыштановский; Гос. ун-т – Высшая школа экономи-ки. – М. : Изд. дом ГУ ВШЭ, 2006. – С. 13–14



Например, требуется подсчитать число мужчин и женщин среди респондентов. Для выполнения процедуры **Частоты** необходимо выбрать в меню (рис. 28):

**Анализ**

**Описательные статистики**

**Частоты**

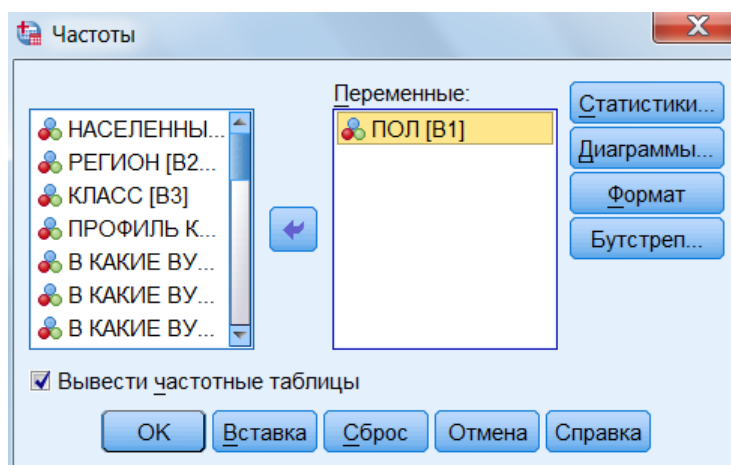


Рис. 28

Выполнение указанной последовательности команд ведет к открытию диалогового окна **Частоты**. 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». Выбранные переменные попадут в правый список. После нажатия кнопки ОК процедура начнет выполняться, и результаты будут выдаваться в окне просмотра<sup>44</sup>.

Результаты появятся в окне просмотра результатов. Перед самой частотной таблицей выводится небольшая таблица с обзором допустимых и отсутствующих значений (рис. 29).

<sup>44</sup> Пациорковский В.В., Пациорковская В.В. SPSS для социологов: Учебное пособие. ИСЭПН РАН. – М., 2005. – С. 153–154

Статистики

ПОЛ		
N	Валидные	113
	Пропущенные	1

		ПОЛ			
		Частота	Процент	Валидный процент	Накопленный процент
Валидные	Юноши	42	36,8	37,2	37,2
	Девушки	71	62,3	62,8	100,0
	Итого	113	99,1	100,0	
Пропущенные	Системные пропущенные	1	,9		
Итого		114	100,0		

Рис. 29

Каждая строка частотной таблицы описывает одно возможное значение.

*Частоты* – частоты или абсолютные числа, показывающие, сколько раз встречается тот или иной вариант.

*Процент* – процентное выражение числовых значений переменной с учетом пропущенных значений.

*Валидный процент* – валидное процентное выражение числовых значений переменной без учета пропущенных значений.

*Накопленный процент* – накопленный (кумулятивный) процент<sup>45</sup>.

Делать выводы о том, много или мало респондентов отметили при опросе ту или иную градацию в вопросе, опираясь на значения в колонке *Частоты*, невозможно, поскольку необходимо постоянно соотносить эти числа с общим количеством опрошенных. Поэтому удобнее использовать колонку *Процент* (процент), которая содержит процентные значения для каждой из частот.

Колонка *Валидный процент* связана с такой важной в социологической практике характеристикой, как «Отсутствие ответа». Мы знаем, что в ходе любого массового опроса какая-то часть опрошиваемых не отвечает на поставленные вопросы. Причины такого рода «неответов» различны. Это и просто нежелание людей давать информацию по тем или иным показателям.

<sup>45</sup> Бююль А., Цёфель П. SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей : Пер. с нем. / А. Бююль, П. Цёфель. – СПб.: ООО «ДиаСофтЮП», 2005. – С. 92

Это и отсутствие собственного мнения по определенным вопросам. Возможности преодоления проблемы «неответов» на этапе сбора социологической информации достаточно подробно рассматриваются у разных авторов, однако очевидно, что эту проблему нельзя решить полностью.

На этапе работы с собранными данными проблема «неответов» может быть сформулирована следующим образом: как анализировать ту информацию, которая может быть квалифицирована как «отсутствие ответа».

Необходимо отметить, что на этот вопрос нет однозначного ответа. В зависимости от характера решаемых задач существуют разные подходы к анализу информации, которая соответствует «неответам». Отметим, что числовые коды, связанные с «неответами», называют *пропущенные данные*.

Вопрос о том, какой из показателей – процент опрошенных, либо процент ответивших необходимо использовать для выявления определенных социологических закономерностей, некорректен. Оба показателя несут определенную информацию и, как правило, используются одновременно, однако их интерпретация существенно различна. Например, если в ходе опроса, за кого собираются голосовать респонденты на предстоящих выборах, мы получим, что за кандидата А собирается голосовать 20% опрошенных и 40% ответивших, то оба этих числа представляют интерес. Действительно, первое число говорит нам, что 20% общего количества взрослого населения собирается поддержать кандидата А на будущих выборах. Поскольку коды пропущенных данных в такого рода опросах получают, как правило, те респонденты, которые говорят, что не будут участвовать в выборах, то число 40% говорит нам о том, сколько процентов может набрать кандидат А в ходе голосования<sup>46</sup>.

**2. Чтобы получить описательную статистику** числовых переменных, можно щелкнуть в диалоге **Частоты** на кнопке **Статистики**. Откроется

---

<sup>46</sup> Крыштановский А.О. Анализ социологических данных с помощью пакета SPSS : учеб. пособие для вузов / А.О. Крыштановский; Гос. ун-т – Высшая школа экономи-ки. – М. : Изд. дом ГУ ВШЭ, 2006. – С. 17–19

диалоговое окно *Частоты: Статистики*. В этом окне представлены такие показатели, как процептили и описательные статистики (рис. 30):

- значения процептилей (в частности квартилей),
- разброс (стандартное отклонение, дисперсия, размах, минимум и максимум, стандартная ошибка среднего),
- расположение (среднее, медиана, мода, сумма),
- распределение (асимметрия и эксцесс).

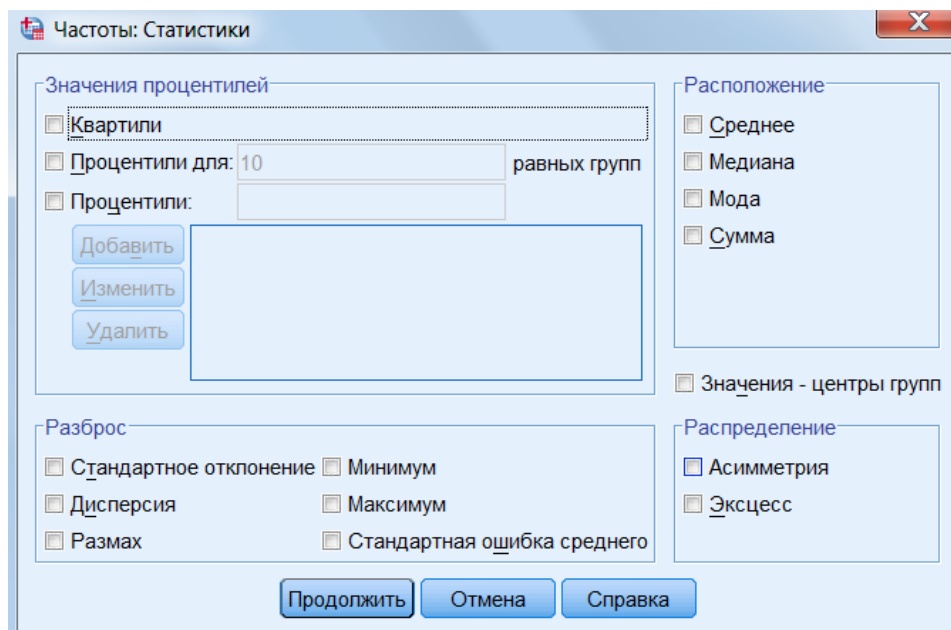


Рис. 30

3. Сейчас мы попробуем вывести частотную таблицу для интересующей переменной, отсортированную по убыванию частоты. Поступите следующим образом:

- Выберите в меню команды

**Анализ**

**Описательные статистики**

**Частоты**

- Перенесите переменную в список выходных переменных.
- Щелкните на кнопке **Формат...** Откроется диалоговое окно *Частоты: Формат* (рис. 31).

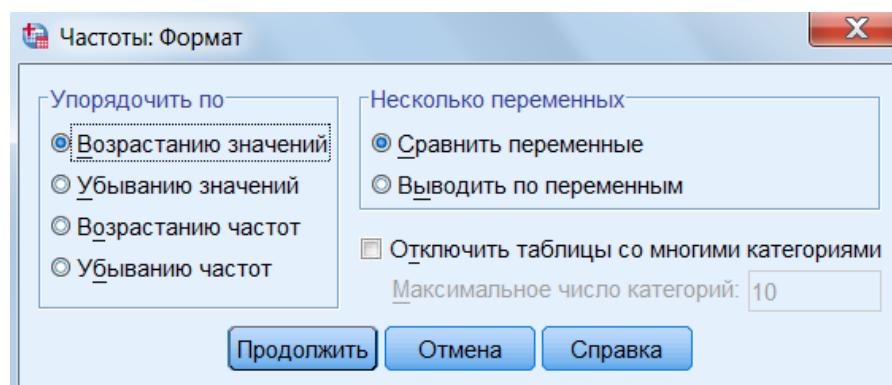


Рис. 31

В группе *Упорядочить по* можно выбрать порядок, в котором будут отображены значения в частотной таблице. Возможны следующие варианты:

- **Возрастанию значений:** Данные сортируются по возрастанию значений. Это настройка по умолчанию.

- **Убыванию значений:** Данные сортируются по убыванию значений.

- **Возрастанию частот:** Данные сортируются по возрастанию частот.

- **Убыванию частот:** Категории сортируются по убыванию частот.

Кроме того, флажок *Отключить таблицы со многими категориями* позволяет избежать вывода длинных частотных таблиц<sup>47</sup>.

4. Результаты частотного распределения можно представить графически. Наиболее популярные формы – это столбиковые и круговые диаграммы. Команды для построения графических диаграмм могут выполняться либо непосредственно из модуля вычисления одномерных частотных распределений (команда **Частоты**), либо из специального блока команд **Графика**, в котором представлены возможности графического анализа пакета программ SPSS. Графические диаграммы в качестве метода построения одномерных частотных распределений повышают наглядность полученных закономерностей и могут

<sup>47</sup> Бююль А., Цёфель П. SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей : Пер. с нем. / А. Бююль, П. Цёфель. – СПб.: ООО «ДиаСофтЮП», 2005. – С. 99

использоваться, прежде всего, для презентации результатов социологических исследований<sup>48</sup>.

Для примера мы создадим столбчатую диаграмму для частотного распределения интересующей переменной. Поступите следующим образом:

- Выберите в меню команды

## **Анализ**

### **Описательные статистики**

#### **Частоты**

- Перенесите переменную в список выходных переменных.
- Щелкните на кнопке *Диаграммы...*. Откроется диалоговое окно *Частоты: Диаграммы* (рис. 32).

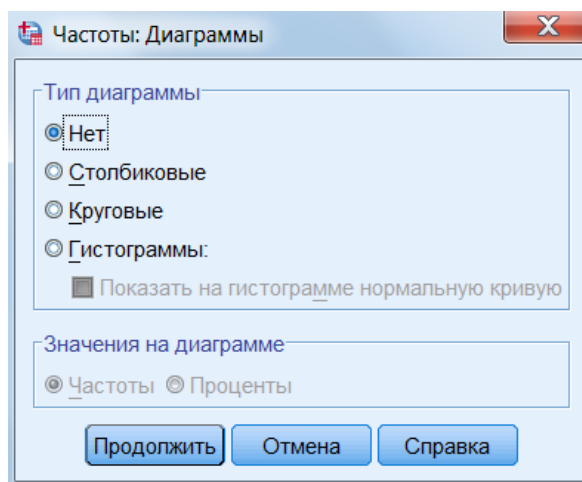


Рис. 32

- Выберите в группе *Тип диаграммы* пункт *Столбиковые*, а в группе *Значения на диаграмме* – пункт *Проценты*.

- Подтвердите выбор кнопкой *Продолжить*. Вы вернетесь в диалог **Частоты**.

#### **Примечания:**

- 1) частотный анализ может проводиться с использованием сервисных процедур: разделение переменных на группы (позволяет вести частотный

---

<sup>48</sup> Крыштановский А.О. Анализ социологических данных с помощью пакета SPSS : учеб. пособие для вузов / А.О. Крыштановский; Гос. ун-т – Высшая школа экономи-ки. – М. : Изд. дом ГУ ВШЭ, 2006. – С. 23

анализ отдельно по каждой группе респондентов) и отбор наблюдений (позволяет вести частотный анализ отдельно по данной группе респондентов);

2) частотный анализ проводится и с модифицированными данными (перекодированными переменными, вычисленными переменными, индексами и т.д.).

### **Использованная литература**

Бююль А., Цёфель П. SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей : Пер. с нем. / А. Бююль, П. Цёфель. – СПб.: ООО «ДиаСофтЮП», 2005. – 608 с.

Крыштановский А.О. Анализ социологических данных с помощью пакета SPSS : учеб. пособие для вузов / А.О. Крыштановский; Гос. ун-т – Высшая школа экономи-ки. – М. : Изд. дом ГУ ВШЭ, 2006. – 281 с.

Пациорковский В.В., Пациорковская В.В. SPSS для социологов: Учебное пособие. ИСЭПН РАН. – М., 2005. – 433 с.

### **Контрольные вопросы**

1. Опишите Частотные таблицы?
2. Какие существуют Статистические характеристики?
3. Что такое Формат частотных таблиц?
4. Каковы Виды ранжирования (сортировки) результатов?

### **Практические задания**

1. Откройте одну из трех баз данных на выбор: «куратор 2011», «первокурсники 2010» или «родители 2010». Выведите частотные таблицы для всех альтернативных вопросов и поясните полученные данные (Что вы видите в каждой таблице?).

2. Постройте частотную таблицу по любому альтернативному вопросу с условием, что ответы на данный вопрос в таблице будут отсортированы от самого частотного (с наибольшим процентом) до наименее частотного (с наименьшим процентом).

3. Постройте круговую диаграмму для любого альтернативного вопроса.
4. Выведите частотную таблицу для любого альтернативного вопроса, предварительно расщепив наблюдения по полу (или студенческой группе, или факультету – на ваш выбор).
5. Выведите частотную таблицу для любого альтернативного вопроса, но только для одной конкретной группы респондентов, например, девушек (или юношей, или студентов определенного факультета – на ваш выбор).

## 2.4. Статистические характеристики

### *1. Вывод описательных статистик.*

Описательные статистики – это различные вычисляемые показатели, характеризующие распределение значений переменной. Эти показатели условно можно разбить на несколько групп. Первая группа – меры центральной тенденции, вокруг которых «группируются» данные: среднее значение, медиана и мода. Вторая группа характеризует изменчивость значений переменной относительно среднего: стандартное отклонение и дисперсия. Диапазон изменчивости характеризуется минимумом, максимумом и размахом. Асимметрия и эксцесс представляют меру отклонения формы распределения от нормального вида. Кроме того, существуют величины, выражающие погрешности некоторых статистик: стандартная ошибка среднего, стандартная ошибка асимметрии и стандартная ошибка эксцесса. Последние два показателя вычисляются программой вместе с асимметрией и эксцессом по умолчанию<sup>49</sup>.

Чтобы получить описательную статистику числовых переменных, можно щелкнуть в диалоге **Частоты** на кнопке **Статистики**. Откроется диалоговое окно *Частоты: Статистики* (рис. 33).

---

<sup>49</sup> Наследов А.Д. SPSS: Компьютерный анализ в психологии и социальных науках : самоучитель / А.Д. Наследов. – СПб. : Питер, 2005. – С. 105



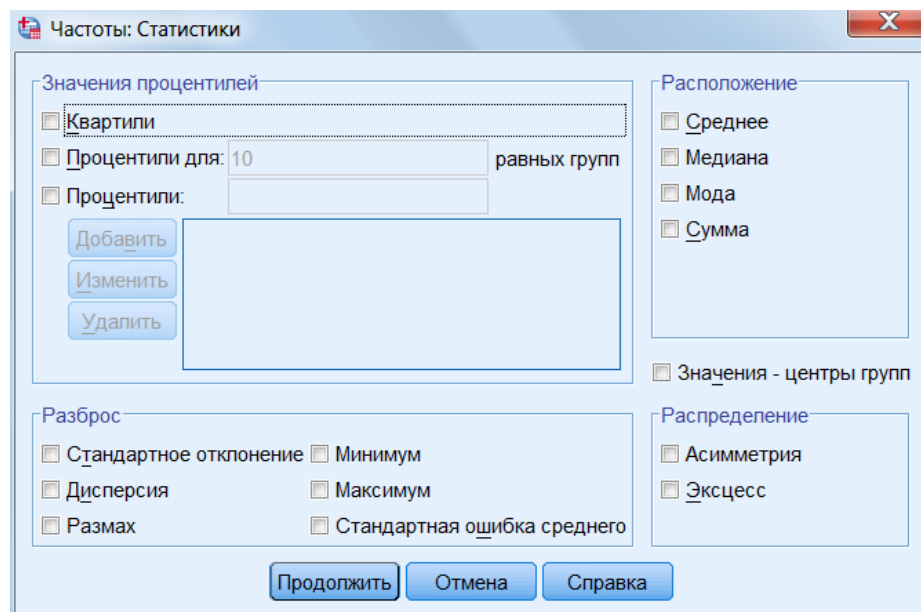


Рис. 33

В группе **Значения процентилей** можно выбрать следующие варианты:

- **Квартили:** Будут показаны первый, второй и третий квартили. Первый квартиль ( $Q_1$ ) – это точка на шкале измеренных значений, ниже (левее) которой располагаются 25 % измеренных значений. Второй квартиль ( $Q_2$ ) – это точка, ниже которой располагаются 50 % измеренных значений. Второй квартиль также называется медианой. Третий квартиль ( $Q_3$ ) – это точка на шкале измеренных значений, ниже которой располагаются 75 % значений. Если данные имеются только в форме порядкового отношения, то качестве меры разброса используется межквартильная широта.

- **Процентили для:** Будут вычислены значения процентилей, разделяющие выборку на группы наблюдений, которые имеют одинаковую ширину, то есть включают одно и то же количество измеренных значений. По умолчанию предлагается количество групп 10. Если задать, к примеру, 4, то будут показаны квартили, то есть квартили соответствуют процентиям 25, 50 и 75. Видно, что число показываемых процентилей на единицу меньше заданного числа групп.

- **Процентили:** Здесь имеются в виду значения процентилей, определяемые пользователем. Введите значение процентия в пределах от 0 до 100 и щелкните на кнопке *Добавить*. Повторите эти действия для всех желаемых

значений процентилей. Значения в порядке возрастания будут показаны в списке. Например, если ввести значения 25, 50 и 75, то мы получим квартили. Можно задавать любые значения процентилей, например, 37 и 83. В первом случае (37) будет показано значение выбранной переменной, ниже которого лежат 37 % значений, а во втором случае (83) – значение, ниже которого располагаются 83 % значений.

В группе **Разброс** можно выбрать следующие меры разброса:

- **Стандартное отклонение:** Стандартное отклонение – это мера разброса измеренных величин; оно равно квадратному корню из дисперсии. В интервале шириной, равной удвоенному стандартному отклонению, который отложен по обе стороны от среднего значения, располагается примерно 67% всех значений выборки, подчиняющейся нормальному распределению.

- **Дисперсия:** Дисперсия – это квадрат стандартного отклонения и, следовательно, эта характеристика также является мерой разброса измеренных величин. Она определяется как сумма квадратов отклонений всех измеренных значений от их среднеарифметического значения, деленная на количество измерений минус 1.

- **Размах:** Размах – это разница между наибольшим значением (максимумом) и наименьшим значением (минимумом).

- **Минимум:** Наименьшее значение.

- **Максимум:** Наибольшее значение.

- **Стандартная ошибка среднего:** Это стандартная ошибка среднего значения. В интервале шириной, равной удвоенной стандартной ошибке, отложенному вокруг среднего значения, располагается среднее значение генеральной совокупности с вероятностью примерно 67 %. Стандартная ошибка определяется как стандартное отклонение, деленное, на квадратный корень из объема выборки.

*Стандартная ошибка среднего* является характеристикой точности, или стабильности, величины, для которой она вычисляется. В контексте программы SPSS стандартная ошибка используется для среднего значения, асимметрии и

эксцесса. Ее смысл заключается в следующем. Вы можете, взяв определенное количество случайно выбранных значений генеральной совокупности, составить выборку и вычислить для нее среднее значение. Повторив эту операцию некоторое количество раз, вы получите набор средних значений выборок, которые также представляют собой некоторое распределение. Стандартное отклонение этого распределения и будет являться стандартной ошибкой для среднего значения генеральной совокупности. Аналогичным способом вычисляются стандартные ошибки для асимметрии и эксцесса. Чем меньше значение стандартной ошибки, тем выше стабильность величины, для которой она вычисляется.

В группе *Расположение* можно выбрать следующие характеристики:

■ *Среднее*: Среднее значение – это арифметическое среднее измеренных значений; оно определяется как сумма значений, деленная на их количество. Например, если имеется 12 измеренных значений и их сумма составляет 600, то среднее значение будет  $x = 600 : 12 = 50$ .

■ *Медиана*: Медиана – это точка на шкале измеренных значений, выше и ниже которой лежит по половине всех измеренных значений. Например, если измеренные значения таковы:

3785463928 4,

то сначала они располагаются в порядке возрастания: 2334456788 9.

В данном случае медианой будет значение 5. Всего у нас 11 измеренных значений, следовательно, медианой является шестое значение. Выше него располагается 5 значений, и ниже – тоже 5. При нечетном количестве значений медиана всегда будет совпадать с одним из измеренных значений. При четном количестве медиана будет средним арифметическим двух соседних значений. Например, если имеются следующие измеренные значения:

3445678899

то медиана в этом случае будет равна:  $(6 + 7) : 2 = 6,5$ .

■ *Мода*: Мода – это значение, которое наиболее часто встречается в выборке. Если одна и та же наибольшая частота встречается у нескольких значений, то выбирается наименьшее из них.

■ *Сумма*: Сумма всех значений.

Одной из важнейших характеристик при описании поведения отдельных переменных является показатель средней тенденции. Возможности использования различных мер средней тенденции для шкал различного типа приведены в таблице ниже (табл. 5):

Таблица 5

№ п/п	Уровень измерения	Допустимые меры средней тенденции
1	Номинальный	Мода
2	Порядковый	Мода, медиана
3	Метрический	Мода, медиана, среднее арифметическое

В группе *Распределение* можно выбрать следующие меры несимметричности распределения:

■ *Асимметрия*: Коэффициент асимметрии – это мера отклонения распределения частоты от симметричного распределения, то есть такого, у которого на одинаковом удалении от среднего значения по обе стороны выборки данных располагается одинаковое количество значений. Если наблюдения подчиняются нормальному распределению, то асимметрия равна нулю. Для проверки на нормальное распределение можно применять следующее правило: Если асимметрия значительно отличается от нуля, то гипотезу о том, что данные взяты из нормально распределенной генеральной совокупности, следует отвергнуть. Если вершина асимметричного распределения сдвинута к меньшим значениям, то говорят о положительной асимметрии, в противоположном случае – об отрицательной.

■ *Экссесс*: Коэффициент вариации (экссесс) указывает, является ли распределение пологим (при большом значении коэффициента) или крутым. Коэффициент вариации равен нулю, если наблюдения подчиняются нормальному распределению. Поэтому для проверки на нормальное распределение можно применять еще одно правило: Если коэффициент вари-

ации значительно отличается от нуля, то гипотезу о том, что данные взяты из нормально распределенной генеральной совокупности, следует отвергнуть.

Как правило, для переменных, относящихся к интервальной шкале и подчиняющихся нормальному распределению, в качестве основной характеристики используют среднее значение, а в качестве меры разброса – стандартное отклонение или стандартную ошибку. Для порядковых или интервальных переменных, не подчиняющихся нормальному распределению, – соответственно медиану или первый и третий квартили. Для переменных относящихся к номинальной шкале, нельзя дать других значимых характеристик кроме моды.

В диалоге есть еще один флажок:

■ Значения – центры групп): Если установить этот флажок, то при вычислении медианы и остальных значений процентилей оценки этих характеристик будут определяться для концентрированных данных<sup>50</sup>.

## ***2. Исследование данных.***

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 **Разведочный анализ** (в ранних версиях – **Исследовать**).

---

<sup>50</sup> Бююль А., Цёфель П. SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей : Пер. с нем. / А. Бююль, П. Цёфель. – СПб.: ООО «ДиаСофтЮП», 2005. – С. 92–95

Для выполнения рассматриваемой процедуры необходимо реализовать последовательность команд:

## Анализ

### Описательные статистики

#### Разведочный анализ

В результате выполнения этой последовательности команд укроется главное диалоговое окно процедуры **Исследовать** (рис. 34). 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.

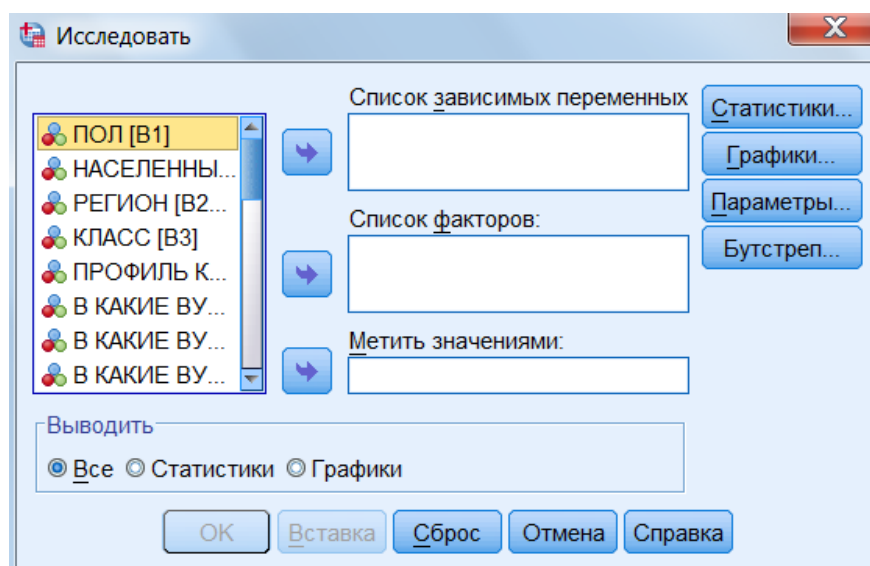


Рис. 34

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, список независимых переменных и метки случаев. Различие между зависимыми переменными и факторами в рассматриваемом окне означает, что может быть проведен углубленный анализ по группам случаев (например, анализ возрастной структуры или доходов домохозяйства отдельно по каждому населенному пункту, 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 *Список зависимых переменных* и начать выполнение процедуры.

Далее, в левой нижней части окна в поле *Выводить* полезно пометить флажком формат вывода информации. Ее можно вывести в виде описательной статистики (*Статистики*), графика (*Графики*), или использовать обе возможности (*Все*). По умолчанию метка как раз и задает выполнение статистики и графиков, т.е. функцию *Все*, стоящую на первом месте.

Кроме того, в главном диалоговом окне процедуры *Explore* имеются три кнопки: *Статистики...*, *Графики...*, *Параметры...*, указывающие на наличие трех дополнительных диалоговых окон. 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 *Статистики...*

Процедура ***Разведочный анализ*** 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 Лиллие-форса) и оценки максимального правдоподобия. С этой целью и следует использовать дополнительное диалоговое окно графики – *Графики...*<sup>51</sup>.

### ***3. Получение сводки для наблюдений.***

Команда ***Итоги по наблюдениям*** предназначена для получения с различной степенью детализации упорядоченного списка всех данных файла или

---

<sup>51</sup> Пациорковский В.В., Пациорковская В.В. SPSS для социологов: Учебное пособие. ИСЭПН РАН. – М., 2005. – С. 162–164

их подмножества. Это обычно необходимо при определении состава и качества данных для их дальнейшего редактирования или анализа (рис. 35).

## Анализ

### Отчеты

#### Итоги по наблюдениям...

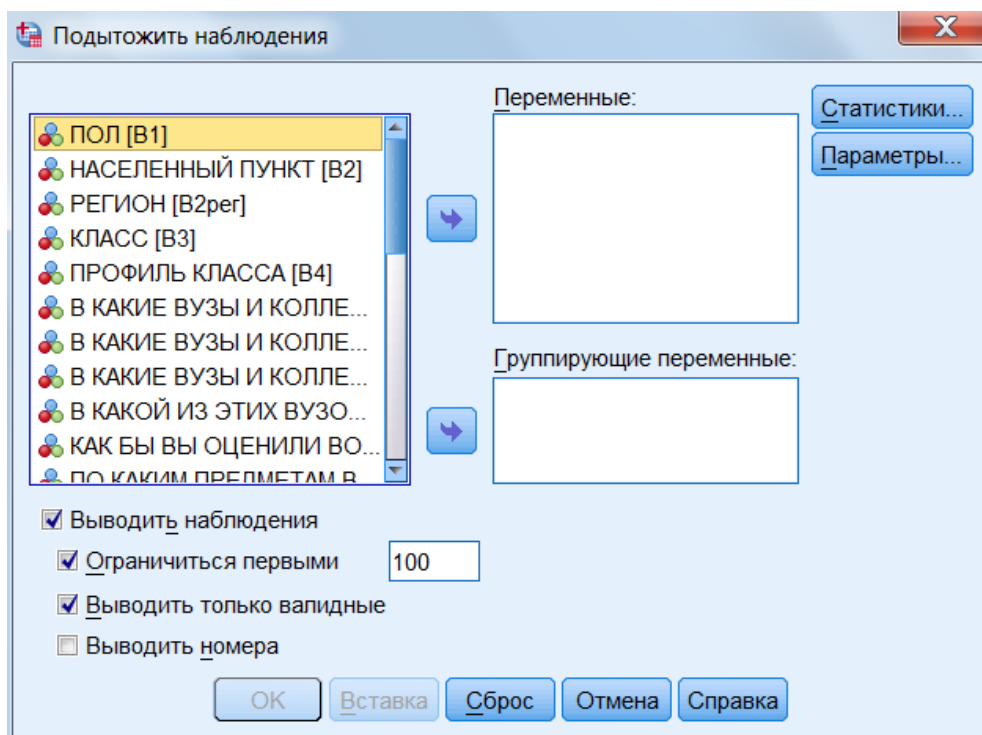


Рис. 35

Список в левой части окна позволяет выбирать переменные для сводки. Назначение остальных элементов окна описано далее.

^ В список *Переменные* вы можете перенести все или некоторые переменные для сводки.

^ Список *Группирующие переменные* определяет порядок перечисления объектов. Если список пуст, объекты будут перечислены в том порядке, в котором они перечислены в файле.

^ Флажок *Выводить наблюдения* управляет отображением значений переменных, и по умолчанию он установлен. Если его сбросить, результат будет содержать только число значений каждой переменной в различных категориях.



^ Флажок *Ограничиться первыми* позволяет ограничить число объектов (строк) для сводки. Если он установлен, то становится доступным поле, расположенное справа от флажка, в которое можно ввести нужное для учета число объектов. По умолчанию это число равно 100. Если флажок сброшен, то учитываются все объекты.

^ Флажок *Выводить только валидные* определяет, нужно ли включать в результат отсутствующие значения переменных. По умолчанию он установлен, однако вы будете часто его сбрасывать, чтобы иметь возможность отследить, какие значения отсутствуют.

^ Флажок *Выводить номер* определяет, нужно ли включать в выводимые данные номера объектов. Несмотря на то что по умолчанию он сброшен, практически всегда его необходимо устанавливать, поскольку для корректировки данных вам понадобится знать номер нужного объекта.

> Кнопка *Статистики...* обеспечивает доступ к диалоговому окну, в котором можно построить список описательных статистик.

> Кнопка *Параметры...* позволяет открыть диалоговое окно, в котором можно задать, нужно ли учитывать объекты, содержащие пропущенные значения, а также отредактировать содержимое заголовков и подзаголовков<sup>52</sup>.

#### **4. Применение t-теста для сравнения средних.**

Различные варианты обработки данных с применением t-критерия позволяют сделать вывод о различии двух средних значений. SPSS позволяет применять 3 варианта t-критерия<sup>53</sup>:

> t-критерий для независимых выборок, предназначен для сравнения средних значений двух выборок. Для сравниваемых выборок должны быть определены значения одной и той же переменной:

---

<sup>52</sup> Бююль А., Цёфель П. SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей : Пер. с нем. / А. Бююль, П. Цёфель. – СПб.: ООО «ДиаСофтЮП», 2005. – С. 167–168

<sup>53</sup> Наследов А.Д. SPSS: Компьютерный анализ в психологии и социальных науках : самоучитель / А.Д. Наследов. – СПб. : Питер, 2005. – С. 145–146

## **Анализ**

### **Сравнение средних**

#### **T-критерий для независимых выборок**

> t-критерий для зависимых (парных) выборок, позволяет сравнить средние значения двух измерений одного признака для одной и той же выборки:

## **Анализ**

### **Сравнение средних**

#### **T-критерий для парных выборок**

> t-критерий для одной выборки, позволяет сравнить среднее значение этой выборки с некоторой эталонной величиной:

## **Анализ**

### **Сравнение средних**

#### **Одновыборочный t-критерий**

Результат сравнения средних значений с применением t-критерия оценивается по уровню значимости ( $p$ ), который не должен превышать 0,05.

## **Использованная литература**

Бююль А., Цёфель П. SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей : Пер. с нем. / А. Бююль, П. Цёфель. – СПб.: ООО «ДиаСофтЮП», 2005. – 608 с.

Наследов А.Д. SPSS: Компьютерный анализ в психологии и социальных науках : самоучитель / А.Д. Наследов. – СПб. : Питер, 2005. – 416 с.

Пациорковский В.В., Пациорковская В.В. SPSS для социологов: Учебное пособие. ИСЭПН РАН. – М., 2005. – 433 с.

## **Контрольные вопросы**

1. Назовите основные Описательные статистики и их виды
2. Какие виды t-теста сравнения средних существуют? Каково их назначение?
3. Что такое Разведочный анализ?

4. Для чего нужна процедура Итоги по наблюдениям?

### **Практические задания**

1. Создайте условную базу данных СПСС, в которой были бы отражены баллы ЕГЭ по обществознанию для двух студенческих групп. Для этого:

1) введите в пустую базу данных СПСС переменную «Балл ЕГЭ по обществознанию», содержащую 30 произвольных значений (в диапазоне от 0 до 100);

2) введите вторую переменную «Студенческая группа», имеющую два значения 1 – «группа № 1» и 2 – «группа № 2», так, чтобы первые пятнадцать наблюдений относились к первой группе, а оставшиеся пятнадцать наблюдений – ко второй.

2. Выведите все статистические характеристики для переменной «Балл ЕГЭ по обществознанию». Поясните значение каждой статистики.

3. Проведите экспресс-анализ (разведочный анализ) переменной «Балл ЕГЭ по обществознанию» с целью сравнить статистические характеристики двух студенческих групп.

4. Выберите подходящую процедуру t-теста и сравните средние. Сделайте вывод.

5. Откройте на выбор одну из двух баз данных: «первокурсники 2010» или «родители 2010». Выведите все средние для табличного вопроса. Поясните получившиеся результаты.

6. Выведите все средние для табличного вопроса, предварительно отобрав конкретную группу респондентов (например, юношей, девушек, студентов конкретного факультета – на ваш выбор).

7. Выведите все средние для табличного вопроса, предварительно расщепив базу данных по определенному признаку (полу, факультету, студенческой группе – на ваш выбор).

## 2.5. Таблицы сопряженности

До сих пор мы рассматривали только отдельные переменные. Мы проводили частотный анализ, а также описывали отдельные переменные статистическими характеристиками, такими как минимум, максимум и среднее значение. Методы анализа такого рода называются одномерными. В текущей главе мы перейдем к двумерному анализу и займемся выяснением вопроса, существует ли взаимосвязь между двумя или более переменными.

В SPSS имеется большое количество разнообразных процедур, при помощи которых можно произвести анализ связи между двумя переменными. Связь между неметрическими переменными, то есть переменными, относящимися к номинальной шкале или к порядковой шкале с не очень большим количеством категорий, лучше всего представить в форме таблиц сопряженности. Для этой цели в SPSS реализован тест при котором проверяется, есть ли значимое различие между наблюдаемыми и ожидаемыми частотами. Кроме того, существует возможность расчета различных мер связанности. До сих пор мы рассматривали только отдельные переменные. Мы проводили частотный анализ, а также описывали отдельные переменные статистическими характеристиками, такими как минимум, максимум и среднее значение. Методы анализа такого рода называются одномерными. В текущем параграфе мы перейдем к двумерному анализу и займемся выяснением вопроса, существует ли взаимосвязь между двумя или более переменными<sup>54</sup>.

1. К наиболее часто используемым инструментам изучения взаимосвязи двух переменных относятся методы анализа таблицы сопряженности. Анализ

---

<sup>54</sup> Бююль А., Цёфель П. SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей : Пер. с нем. / А. Бююль, П. Цёфель. – СПб.: ООО «ДиаСофтЮП», 2005. – С. 180

таблицы является весьма простым и наглядным, и вместе с тем эффективным инструментом изучения одновременно двух переменных<sup>55</sup>.

Для создания таблиц сопряженности и вычисления меры связанности на их основе, выберите в меню команды:

## Анализ

### Описательные статистики

#### Таблицы сопряженности

Откроется диалоговое окно *Таблицы сопряженности* (рис. 36).

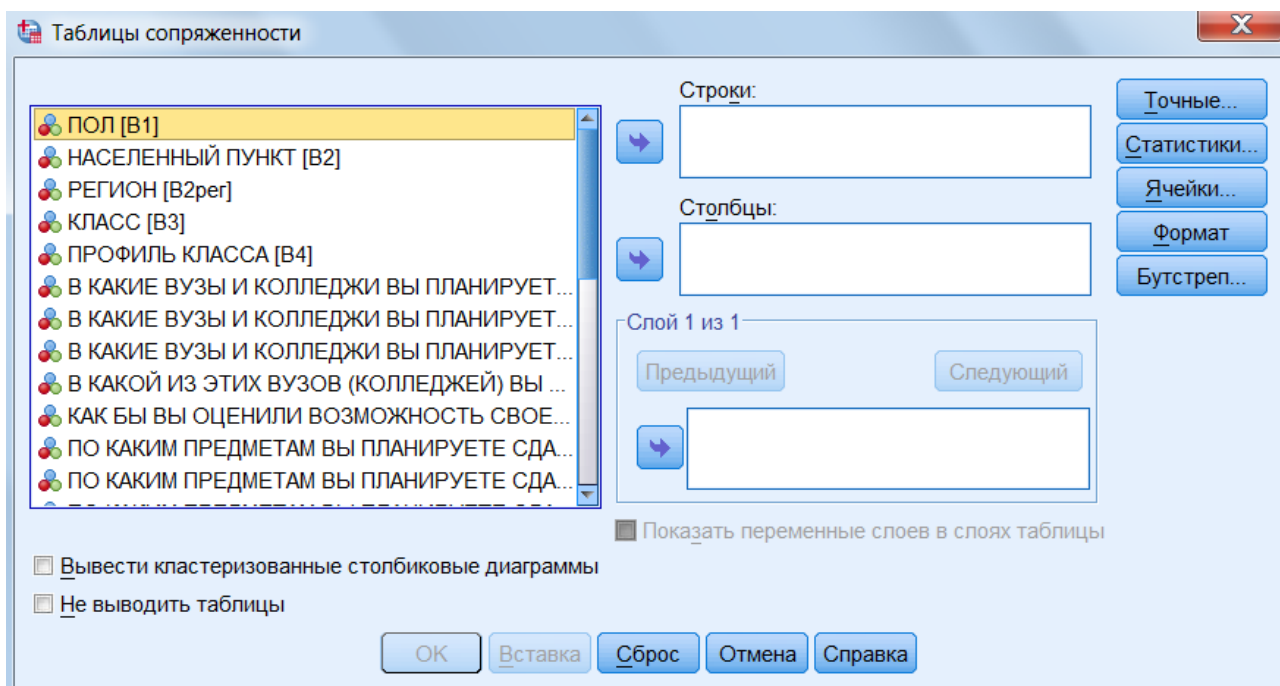


Рис. 36

Список исходных переменных содержит переменные открытого файла данных. Здесь можно выбрать переменные для строк и столбцов таблицы сопряженности. Для каждого сочетания двух переменных будет создана таблица сопряженности. Например, если в списке *Строки* находится три переменных, а в списке *Столбцы* – две, то мы получим  $3 \times 2 = 6$  таблиц

<sup>55</sup> Крыштановский А.О. Анализ социологических данных с помощью пакета SPSS : учеб. пособие для вузов / А.О. Крыштановский; Гос. ун-т – Высшая школа экономи-ки. – М. : Изд. дом ГУ ВШЭ, 2006. – С. 40

сопряженности<sup>56</sup>. 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. На следующем шаге выделенная переменная, путем нажатия одной из двух кнопок «стрелка вправо», переносится в поле *Строки* или *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. При анализе зависимостей двух переменных важнейшим является вопрос о том, какую из переменных считать зависимой, т.е. подверженной влиянию, а какую – независимой, т.е. влияющей.

Для задания статистик, например, вычисления процентов по строкам и столбцам таблицы, используется кнопка **Ячейки**, открывающая дополнительное диалоговое окно рассматриваемой процедуры (рис. 37).

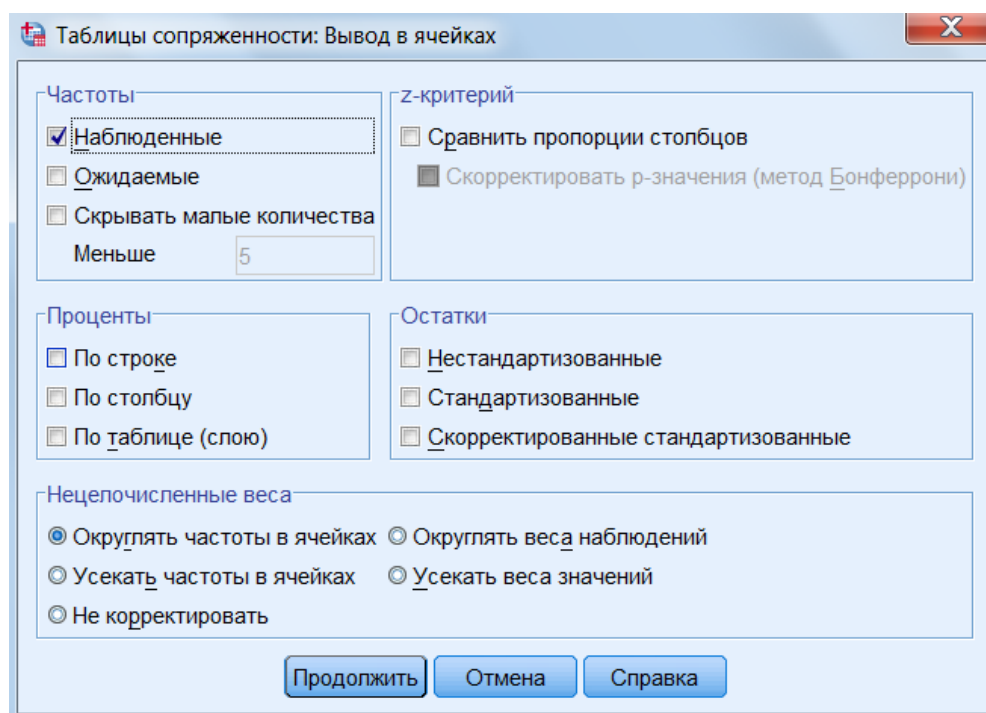


Рис. 37

<sup>56</sup> Бююль А., Цёфель П. SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей : Пер. с нем. / А. Бююль, П. Цёфель. – СПб.: ООО «ДиаСофтЮП», 2005. – С. 180

В группе **Проценты** можно выбрать один или более из нижеследующих вариантов отображения:

■ *По строке*: Вычисляются процентные значения по строкам: количество наблюдений в каждой ячейке, отнесенное к сумме по строке.

■ *По столбцу*: Вычисляются процентные значения по столбцам: количество наблюдений в каждой ячейке в отношении к сумме столбца.

■ *По таблице (слою)*: Вычисляются полные процентные значения: количество наблюдений в каждой ячейке, отнесенное к общей сумме наблюдений.

Работа в любом из дополнительных окон завершается нажатием кнопки **Продолжить**. Выполнение этой команды ведет к возврату в главное окно процедуры, в котором после нажатия кнопки ОК процедура начнет выполняться, а результаты появятся в окне просмотра<sup>57</sup>.

Первая таблица содержит информацию о числе самих наблюдений; два наблюдения содержат пропущенные значения по крайней мере в одной из двух участвующих переменных. Вторая таблица – это собственно таблица сопряженности (рис. 38).

---

<sup>57</sup> Пациорковский В.В., Пациорковская В.В. SPSS для социологов: Учебное пособие. ИСЭПН РАН. – М., 2005. – С. 186–187

Таблица сопряженности ПРОФИЛЬ КЛАССА \* ПОЛ

			ПОЛ		Итого
			Юноши	Девушки	
ПРОФИЛЬ КЛАССА	без профиля / профиль не указан	Частота	10	9	19
		% в ПОЛ	23,8%	12,9%	17,0%
	социально-гуманитарный	Частота	6	14	20
		% в ПОЛ	14,3%	20,0%	17,9%
	социально-экономический	Частота	8	12	20
		% в ПОЛ	19,0%	17,1%	17,9%
	МЧС	Частота	0	4	4
		% в ПОЛ	0,0%	5,7%	3,6%
	юридический	Частота	4	8	12
		% в ПОЛ	9,5%	11,4%	10,7%
	химико-биологический	Частота	2	3	5
		% в ПОЛ	4,8%	4,3%	4,5%
	физико-математический	Частота	10	16	26
		% в ПОЛ	23,8%	22,9%	23,2%
	естественно-научный	Частота	1	0	1
		% в ПОЛ	2,4%	0,0%	0,9%
	медицинский	Частота	0	1	1
		% в ПОЛ	0,0%	1,4%	0,9%
	аэрокосмический	Частота	0	1	1
		% в ПОЛ	0,0%	1,4%	0,9%
	филологический	Частота	1	2	3
		% в ПОЛ	2,4%	2,9%	2,7%
Итого		Частота	42	70	112
		% в ПОЛ	100,0%	100,0%	100,0%

Рис. 38

На пересечении строк и столбцов находятся числа, показывающие, какое количество единиц анализа (в данном случае – респондентов) обладают одновременно данными градациями по выбранным переменным. Внизу таблицы сопряженности располагаются суммарные данные по всем колонкам, а с правого края таблицы – аналогичные суммы по всем строкам. Иными словами, сбоку справа и снизу находятся одномерные частотные распределения для переменных, использованных в таблице<sup>58</sup>.

**Примечание:** с помощью процедуры *Таблицы сопряженности* SPSS позволяет построить таблицы не только по двум признакам, но одновременно по трем и более, то есть – таблицы большей размерности (многомерные).

Порядок построения трехмерной таблицы выглядит следующим образом: сначала из списка переменных задаются подлежащее (строка) и сказуемое (столбец) требуемой таблицы, затем в *Слой* вводится управляющая переменная

<sup>58</sup> Бююль А., Цёфель П. SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей : Пер. с нем. / А. Бююль, П. Цёфель. – СПб.: ООО «ДиаСофтЮП», 2005. – С. 181



Для этого из списка переменных с помощью нижней кнопки («стрелка вправо») в поле данного блока переносится выделенная переменная.

После введения управляющих переменных, при выполнении команды ОК, в окне просмотра будут получены двухмерные таблицы для каждого значения управляющей переменной.

Кроме кнопки *Ячейки* диалоговое окно **Таблицы сопряженности** содержит еще две кнопки, открывающие дополнительные диалоговые окна: *Статистики* с набором возможностей расчета различных статистик и *Формат* для выбора порядка вывода данных.

2. Можно изменить порядок сортировки переменных строк в таблице сопряженности, щелкнув в диалоговом окне **Таблицы сопряженности** на кнопке *Формат...* Откроется диалоговое окно **Таблицы сопряженности: Формат** (рис. 39).

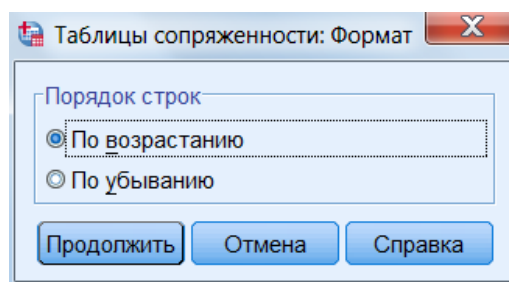


Рис. 39

В группе **Порядок строк** можно выбрать один из следующих вариантов сортировки значений:

- *По возрастанию*: Значения переменных строк отображаются в порядке возрастания от наименьшего к наибольшему. Это настройка по умолчанию.

- *По убыванию*: Значения переменных строк отображаются в порядке убывания от наибольшего к наименьшему<sup>59</sup>.

3. Чтобы сделать более наглядными данные, содержащиеся в таблицах сопряженности, их можно представить визуально. Для этого поступите следующим образом: установите в диалоге **Таблицы сопряженности** флажок

---

<sup>58</sup> Там же. С. 186–187

Вывести кластеризованные столбиковые диаграммы. Тогда на диаграмме будут показаны две группы столбцов для двух переменных строк<sup>60</sup>.

4. Чтобы получить статистические критерии для таблиц сопряженности, щелкните на кнопке *Статистики* в диалоговом окне **Таблицы сопряженности**. Откроется диалоговое окно Таблицы сопряженности: Статистики (рис. 40).

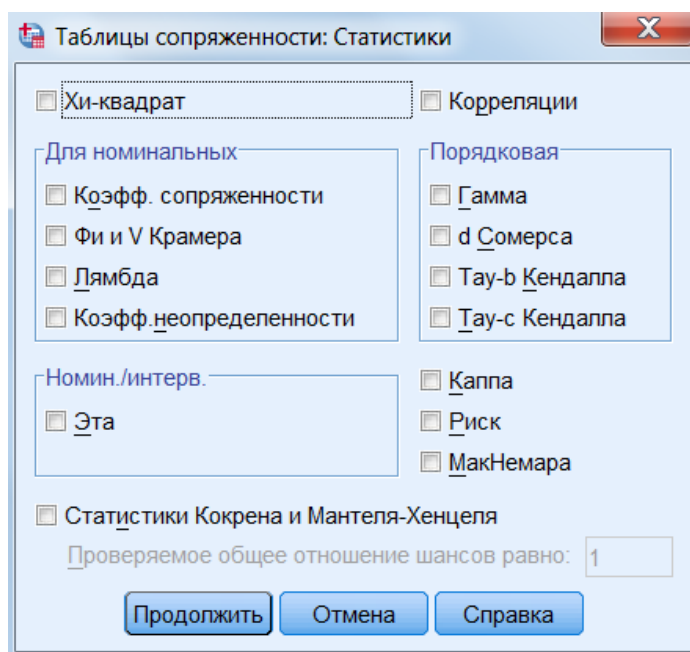


Рис. 40

Флажки в этом диалоговом окне позволяют выбрать один или несколько критериев.

- Тест хи-квадрат ( $\chi^2$ )
- Корреляции
- Меры связанности для переменных, относящихся к номинальной шкале
- Меры связанности для переменных, относящихся к порядковой шкале
- Меры связанности для переменных, относящихся к интервальной шкале
- Коэффициент каппа ( $\kappa$ )
- Мера риска
- Тест Мак-Немара

<sup>60</sup> Там же. С. 188

## ■ Статистики Кохрана и Мантеля-Хэнзеля

Критерий хи-квадрат имеет большое значение в статистических вычислениях, поэтому ему уделим особое внимание<sup>61</sup>.

Помимо частот (или наблюдаемых величин) SPSS может вычислять ожидаемые значения для каждой ячейки таблицы. Ожидаемое значение вычисляется в предположении, что две номинативные переменные независимы друг от друга. Рассмотрим простой пример. Пусть в комнате находится 100 человек, из которых 30 являются мужчинами, а 70 – женщинами. Если известно, что из этих 100 человек 10 увлекаются искусством, то в случае, если увлечение не зависит от пола, мы будем ожидать, что из 10 увлекающихся искусством 3 являются мужчинами, а 7 – женщинами. Сопоставляя эти ожидаемые частоты с наблюдаемыми частотами, мы можем судить о том, действительно ли два номинативных признака не связаны. Чем больше расхождение наблюдаемых и ожидаемых частот, тем, очевидно, два признака сильнее связаны друг с другом. Целью применения критерия независимости  $\chi^2$  и является установление степени соответствия между наблюдаемыми и ожидаемыми значениями ячеек.

В основе критерия независимости лежит вычисление величины  $\chi^2$  определяемой как сумма отношений квадратов отклонений наблюдаемой величины  $f_0$  от ожидаемой величины  $f_e$  к ожидаемой величине каждой ячейки (формула 1).

$$\chi^2 = \sum \frac{(\text{наблюдаемое} - \text{ожидаемое})^2}{\text{ожидаемое}} \quad (1)$$

При больших отклонениях  $f_0$  от  $f_e$  величина  $\chi^2$  также становится большой. Вместе с  $\chi^2$  вычисляется р-уровень значимости. При  $p > 0,05$  считается, что различия между наблюдаемыми и ожидаемыми значениями незначительны. В противном случае предположение о независимости двух номинативных

---

<sup>61</sup> Там же. С. 190–191

переменных отклоняется и делается вывод о том, что две классификации (переменные) зависят друг от друга<sup>62</sup>.

### **Использованная литература**

Бююль А., Цёфель П. SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей : Пер. с нем. / А. Бююль, П. Цёфель. – СПб.: ООО «ДиаСофтЮП», 2005. – 608 с.

Крыштановский А.О. Анализ социологических данных с помощью пакета SPSS : учеб. пособие для вузов / А.О. Крыштановский; Гос. ун-т – Высшая школа экономи-ки. – М. : Изд. дом ГУ ВШЭ, 2006. – 281 с.

Наследов А.Д. SPSS: Компьютерный анализ в психологии и социальных науках : самоучитель / А.Д. Наследов. – СПб. : Питер, 2005. – 416 с.

Пациорковский В.В., Пациорковская В.В. SPSS для социологов: Учебное пособие. ИСЭПН РАН. – М., 2005. – 433 с.

### **Контрольные вопросы**

1. Что такое Двумерные таблицы? Их назначение и роль в социологии?
2. Каковы Форматы таблиц сопряжённости?
3. Что такое критерий хи-квадрат? И для чего он нужен при построении таблиц сопряженности?

### **Практические задания**

1. Откройте базу данных «первокурсники 2010». Отберите для двумерного анализа демографические вопросы (запишите их номера в тетради).
2. Постройте таблицы сопряженности, которые позволили бы:
  - 2.1. сравнить участие в олимпиадах и конференциях девушек и юношей;
  - 2.2. сравнить посещение сайта АмГУ студентами, обучающимися на разных факультетах.

---

<sup>62</sup> Наследов А.Д. SPSS: Компьютерный анализ в психологии и социальных науках : самоучитель / А.Д. Наследов. – СПб. : Питер, 2005. – С. 114–115

2.3. ответить на вопрос: «На каком факультете обучается больше всего студентов, которые поступили именно на тот факультет, на который больше всего хотели?».

2.4. выявить связь между желанием продолжить обучение в магистратуре и образованием отца (или матери – на ваш выбор).

2.5. выявить связь между желанием продолжить обучение в магистратуре и материальным положением семьи респондента.

3. Выведите любую перекрестную таблицу с упорядочением переменной в строке по убыванию.

4. Выведите любую перекрестную таблицу со столбчатой диаграммой. Постройте на основе этой таблицы столбчатую диаграмму в Excel.

5. Перекрестные таблицы повышенной сложности. Постройте таблицы сопряженности, которые позволили бы:

5.1. определить, на какие факультеты подавали студенты документы помимо того, на который поступили.

5.2. определить, студенты какой направленности обучения (Естественно-научная, Социально-гуманитарная или Экономическая) более склонны к академической карьере.

5.3. узнать процент иногородних студентов, которые будут проживать в общежитии.

## **2.6. Обработка множественных ответов**

Рассмотрим особенности кодирования и анализа множественных ответов. Вопросы, на которые можно дать несколько ответов одновременно (это и есть множественные ответы), имеются во многих анкетных исследованиях. Для кодировки и анализа таких множественных ответов SPSS предоставляет два различных метода: метод множественной дихотомии и категориальный метод.

### **1. Дихотомный метод.**

1.1. В первую очередь мы должны сообщить компьютеру, что эти некоторые переменные принадлежат к одному «набору», к одному множественному вопросу (рис. 41).

## Анализ

### Множественные ответы

#### Задать наборы множественных ответов...

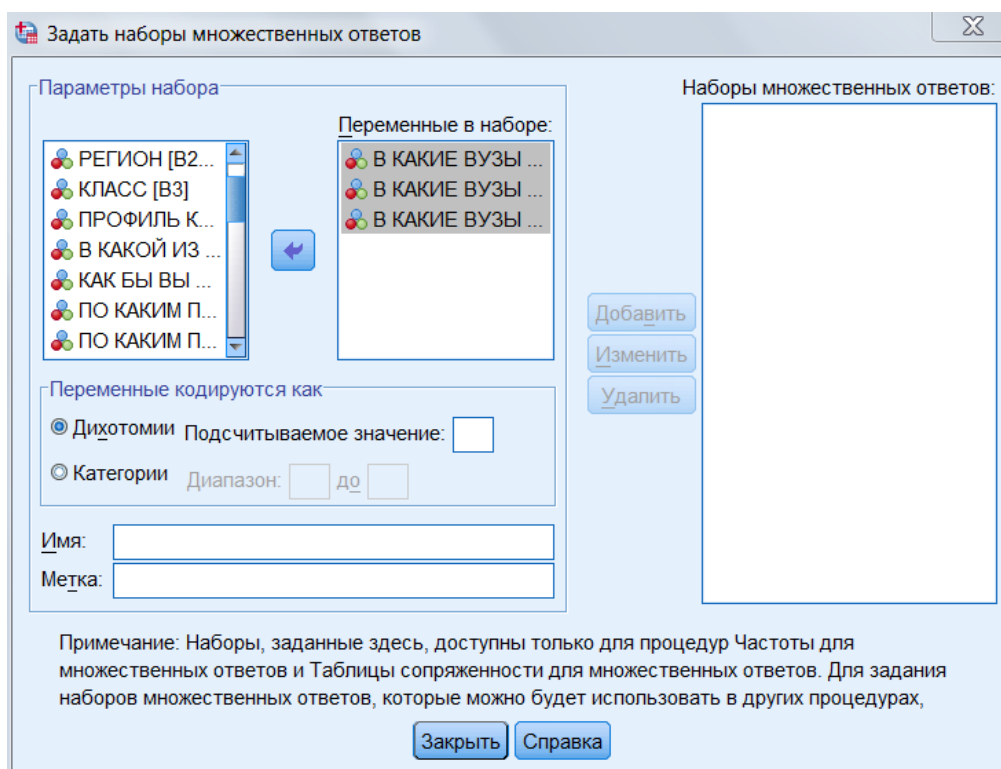


Рис. 41

- Выделите в списке исходных переменных интересующие переменные и перенесите их в список *Переменные в наборе*.
  - Задайте дихотомическую кодировку переменных (опция *Дихотомии* в группе *Переменные кодируются как*). Эта настройка выбирается по умолчанию. В поле *Подсчитываемое значение* введите «1» (ту цифру, которой кодировался выбор респондентом ответа).
    - Присвойте набору имя и метку.
    - Щелкните на кнопке *Добавить*, и созданный набор будет внесен в список наборов множественных ответов *Наборы множественных ответов*.
- SPSS начинает имена наборов переменных со знака доллара.

- Щелкните по кнопке *Заккрыть*, чтобы закончить процесс определения набора.

1.2. Чтобы создать частотную таблицу для дихотомического набора, выберите команды меню (рис. 42):

## Анализ

### Множественные ответы

#### Частоты...

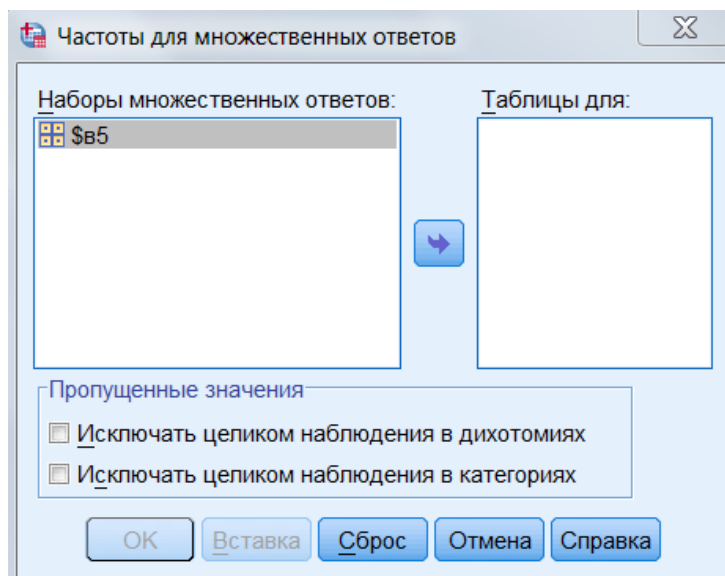


Рис. 42

В списке *Наборы множественных ответов* этого диалога отображаются уже определенные наборы переменных.

- Перенесите набор в список *Таблицы для*.
- Щелкните на кнопке *ОК*.

Для наблюдаемых частот выводятся два разных процентных значения. При определении первого из них наблюдаемая частота отнесена к общему числу ответов (190), а при определении второго – к общему числу ответивших респондентов (102) (рис. 43).

		Ответы		Процент наблюдений
		N	Процент	
Вузы <sup>a</sup>	АМГУ	96	50,5%	94,1%
	АГМА	5	2,6%	4,9%
	БГПУ	26	13,7%	25,5%
	ДальГАУ	7	3,7%	6,9%
	ДВФУ	22	11,6%	21,6%
	МГУ	1	0,5%	1,0%
	СПбГУ	6	3,2%	5,9%
	юридические при МВД (ОрЮИ, ДВЮИ...))	7	3,7%	6,9%
	МГТУ	1	0,5%	1,0%
	ДВВКУ	1	0,5%	1,0%
	ДВГМУ	1	0,5%	1,0%
	СПбГАВМ	1	0,5%	1,0%
	ТОГУ	6	3,2%	5,9%
	ВГУЭС	1	0,5%	1,0%
	СФУ	1	0,5%	1,0%
	СПбГИКИТ	1	0,5%	1,0%
	Академия художеств им. Репина	1	0,5%	1,0%
	Иннополис	1	0,5%	1,0%
	МЭИ	1	0,5%	1,0%
	ХПИ	1	0,5%	1,0%
РАП (РГУП)	1	0,5%	1,0%	
ОГУ	1	0,5%	1,0%	
Академический колледж АМГУ	1	0,5%	1,0%	
Всего		190	100,0%	186,3%

Рис. 43

**1.3.** Таблицы сопряженности можно создавать между двумя наборами переменных, а также между набором и "обычной" переменной. Так, к примеру, нам необходимо в одной таблице сопряженности отобразить соотношение между набором и переменной ПОЛ (рис. 44):

### Анализ

#### Множественные ответы

#### Таблицы сопряженности...



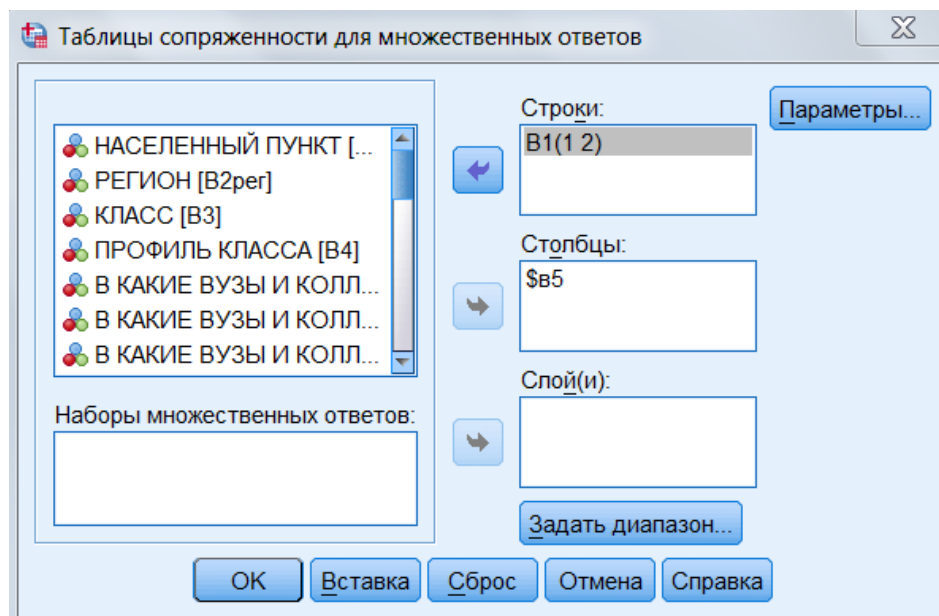


Рис. 44

В списке исходных переменных показаны переменные файла. В списке наборов множественных ответов показан ранее определенный набор.

- Перенесите в список переменных строк (или столбцов) набор, а в список переменных столбцов (или строк) – переменную ПОЛ. Эта переменная появится в списке столбцов (строк) с двумя вопросительными знаками, заключенными в скобки. Если таблица сопряженности строится между элементарными переменными (не являющимися наборами) и наборами, то для первых следует задать диапазон значений.

- Щелкните на кнопке *Задать диапазон*.

Откроется диалоговое окно *Задать диапазон переменной*.

- Задайте минимальное и максимальное значение
- Нажмите *Продолжить*
- Щелкните на кнопке *Параметры*.

Абсолютные частоты в ячейках выводятся всегда.

Дополнительно в группе *Проценты в ячейках* можно выбрать одну или несколько характеристик (рис. 45):

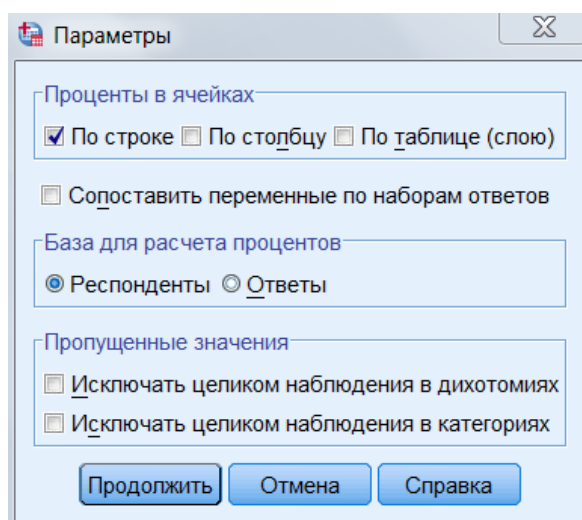


Рис. 45

- По строке: Отображаются проценты для строки.
- По столбцу. Отображаются проценты для столбца.
- По таблице (слою): Отображаются общие проценты для таблицы.

Полученные проценты соответствуют отношению частот к числу ответивших респондентов<sup>63</sup>.

## 2. Категориальный метод.

2.1. Сначала определим набор переменных. Выполните следующие действия:

### Анализ

#### Множественные ответы

##### Задать наборы множественных ответов...

- Выделите интересующие переменные и перенесите их в список *Переменные в наборе*.
- Задайте категориальную кодировку переменных (опция *Категории*). В полях *Диапазон ... до* укажите диапазон от минимального значения до максимального.
- Присвойте набору имя и метку.

<sup>63</sup> Бююль А., Цёфель П. SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей : Пер. с нем. / А. Бююль, П. Цёфель. – СПб.: ООО «ДиаСофтЮП», 2005. – С. 207–213

- Щелкните на кнопке *Добавить*, и созданный набор будет внесен в список наборов множественных ответов.

- Щелкните на кнопке *Закреть*, чтобы завершить определение набора.

2.2. Для того, чтобы создать частотную таблицу, выберите в меню команды, приведенные в пункте 1.2 данного параграфа.

2.3. На основе наборов со множественными категориями также можно строить таблицы сопряженности с другими переменными. Порядок действий аналогичен указанному в пункте 1.3 данного параграфа<sup>64</sup>.

### **Использованная литература**

Бююль А., Цёфель П. SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей : Пер. с нем. / А. Бююль, П. Цёфель. – СПб.: ООО «ДиаСофтЮП», 2005. – 608 с.

### **Контрольные вопросы**

1. Как зафиксировать Наборы множественных ответов?
2. Что такое Дихотомный метод?
3. Что такое Категориальный метод?
4. Как построить Частотные таблицы для множественных ответов?
5. Как построить Таблицы сопряженности для множественных ответов?

### **Практические задания**

1. Откройте базу данных «первокурсники 2010». Отберите для анализа множественные вопросы (запишите их номера в тетради). Определите, с помощью какого метода – дихотомного или категориального – их кодировали.

2. Выберите три множественных вопроса. Задайте для каждого из них наборы переменных.

3. Постройте для этих трех множественных вопросов частотные таблицы и поясните полученные результаты.

---

<sup>64</sup> Там же. С. 213–217

4. Постройте для этих трех множественных вопросов таблицы сопряженности с полом, материальным положением, факультетом обучения респондента (всего должно получиться 9 таблиц сопряженности). Поясните полученные результаты.

## 2.7. Графики

Пакет SPSS for Windows обладает обширным арсеналом мощных и эффективных средств построения диаграмм. Как правило, диаграмма зависит от конкретной статистической процедуры, следовательно, рассматривать все тонкости построения диаграмм имеет смысл лишь в контексте обработки данных. В этом параграфе собраны лишь общие сведения, касающиеся диаграмм вообще, и по большей части эти сведения связаны с редактированием уже созданных диаграмм.

Далее перечислены наиболее часто используемые виды графиков.

^ **Столбиковые диаграммы**, или гистограммы для дискретных переменных. Эти графики применяются для отображения распределения частот встречаемости значений переменной, имеющей небольшое количество градаций. Например, с помощью столбиковой диаграммы удобно представить распределение учащихся по трем классам или по их ориентации на поступление в вузы четырех типов, и т. д. Столбиковые диаграммы делятся на простые, кластерные и стековые (состыкованные).

^ **Гистограммы** внешне напоминают столбиковые диаграммы, однако, как правило, иллюстрируют распределение объектов по диапазонам значений непрерывной переменной (имеющей большое число возможных значений). С помощью гистограммы было бы удобно представить распределение учащихся по диапазонам значений успеваемости или диапазонам тестовых значений.

^ **Линейные графики** применяются для иллюстрации различных зависимостей между данными.

^ **Круговые диаграммы**, как и столбиковые, зачастую применяются для иллюстрации распределений в различных категориях. Используются только при иллюстрации альтернативных вопросов.

^ **Парето-диаграммы** сочетают свойства столбиковых диаграмм и линейных графиков; при этом столбики представляют число элементов данных в различных категориях, а линии показывают накопленные частоты.

^ **Коробчатые диаграммы** основаны на процентилях и являются прекрасным средством отображения распределения данных.

^ **Диаграммы рассеивания** часто используются для отображения корреляций между переменными. Различают простые и оверлейные диаграммы.

^ **Диаграммы столбцов ошибок** включают столбцы, отображающие стандартную погрешность измерения или доверительный интервал для сравниваемых групп<sup>65</sup>.

### **1. Построение диаграмм.**

В SPSS есть два способа построения графиков: 1) связан с использованием дополнительных диалоговых окон в отдельных статистических процедурах, например Частоты и Таблицы сопряженности; 2) связан с использованием подменю Графики.

SPSS предоставляет дополнительную очень удобную возможность построения графиков в диалоговом режиме. В меню **Графика** имеется подменю *Устаревшие диалоговые окна*, в котором представлен широкий спектр команд, позволяющих строить диаграммы в интерактивном режиме: добавлять переменные, изменять категории данных и т.д. Другими словами, работая исключительно с графиками, вы можете добиться такого же результата, как если бы применили статистические процедуры. Иногда такая возможность может оказаться полезной для пользователя, однако до этого необходимо

---

<sup>65</sup> Наследов А.Д. SPSS: Компьютерный анализ в психологии и социальных науках : самоучитель / А.Д. Наследов. – СПб. : Питер, 2005. – С. 84–85

освоить стандартные приемы работы с SPSS как в отношении статистики, так и в отношении графики<sup>66</sup>.

Как правило, с помощью элементов интерфейса верхней части окна вы можете выбрать тип диаграммы, а в нижней части указать, какие данные вы намерены использовать. Типы диаграмм будут меняться в зависимости от вида диалогового окна; для того чтобы указать нужный тип, щелкните на кнопке с миниатюрой диаграммы слева от ее названия. Варианты используемых данных для большинства типов диаграмм одни и те же (рис. 46).

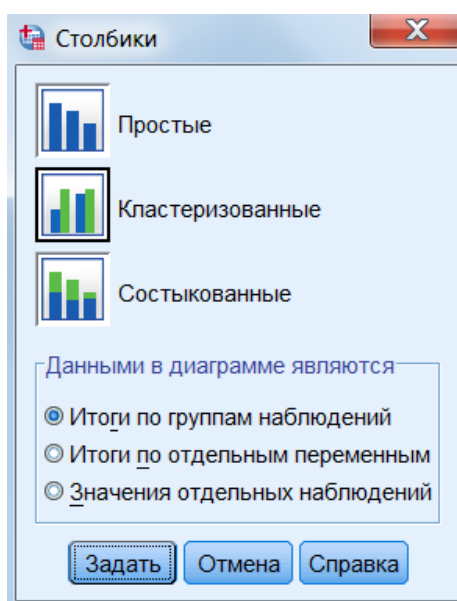


Рис. 46

^ Переключатель *Итоги по группам наблюдений* означает, что задействуется одна переменная, а столбцы диаграммы отразят число объектов каждой градации этой переменной. Например, для переменной *класс* диаграмма будет состоять из трех столбцов, соответствующих трем классам учащихся; для переменной *пол* – из двух столбцов, отображающих число учащихся женского и мужского пола; для переменной *вуз* – из четырех столбцов, представляющих количество учащихся в каждой из четырех групп предпочтительных вузов.

^ Переключатель *Итоги по отдельным переменным* указывает на то, что диаграмма будет содержать несколько столбцов, каждый из которых будет

---

<sup>66</sup> Там же. С. 85

соответствовать среднему значению одной из переменных. Как правило, при построении диаграмм этого типа используют переменные, связанные между собой по смыслу. Такой переменной является, например, переменные **Баллы ЕГЭ**; диаграмма, построенная для нее, отразит средние баллы всех учащихся для каждого из пяти экзаменов.

^ Переключатель *Значения отдельных наблюдений* предназначен для файлов данных с относительно небольшим числом объектов. Разумеется, если число объектов файла велико, то можно выбрать для обработки их подмножество<sup>67</sup>.

## **2. Редактирование диаграмм.**

После того как диаграмма создана, программа предоставляет вам широкий набор команд ее редактирования. Перед тем как выполнять какое-либо редактирование графического объекта, *необходимо дважды щелкнуть на нем мышью*. При этом на экране появится окно графического редактора, содержащее строку меню с полным набором команд и панель инструментов, на которой находятся кнопки, соответствующие наиболее часто используемым командам (рис. 47).

---

<sup>67</sup> Наследов А.Д. SPSS 15: профессиональный статистический анализ данных. – СПб.: Питер, 2008. – С. 98–99

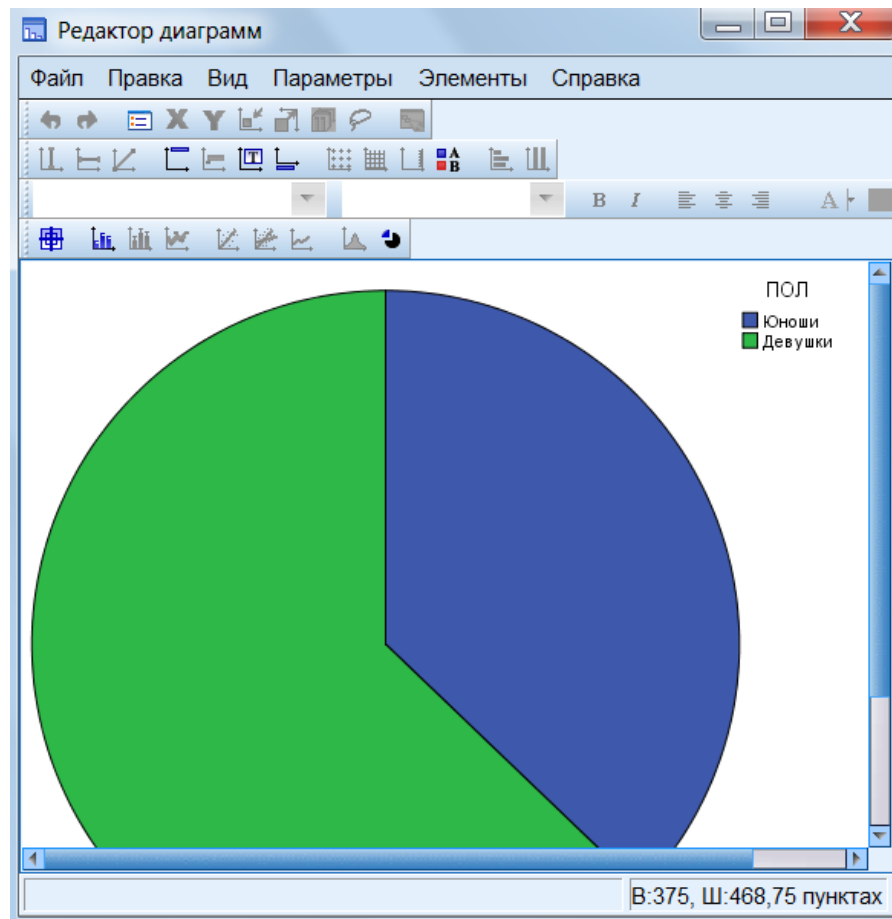


Рис. 47

Команды редактирования могут применяться к диаграмме только в том случае, если она отображена на экране в режиме редактирования. Как только вы дважды щелкнете на диаграмме, SPSS откроет новое окно с ее изображением и строкой меню в верхней части; это означает, что диаграмма доступна для редактирования. Строка меню расположена над панелью инструментов и содержит полный перечень команд редактирования диаграмм.

Как уже упоминалось, для редактирования диаграмм используются либо команды меню, либо кнопки панели инструментов. Возможны две ситуации. В первом случае выбор команды или щелчок на кнопке сразу приводит к изменению диаграммы. Примером могут служить команда и кнопка *Смена осей*. Во втором случае на экране появляется диалоговое окно, предназначенное для задания дополнительных параметров команды (рис. 48).



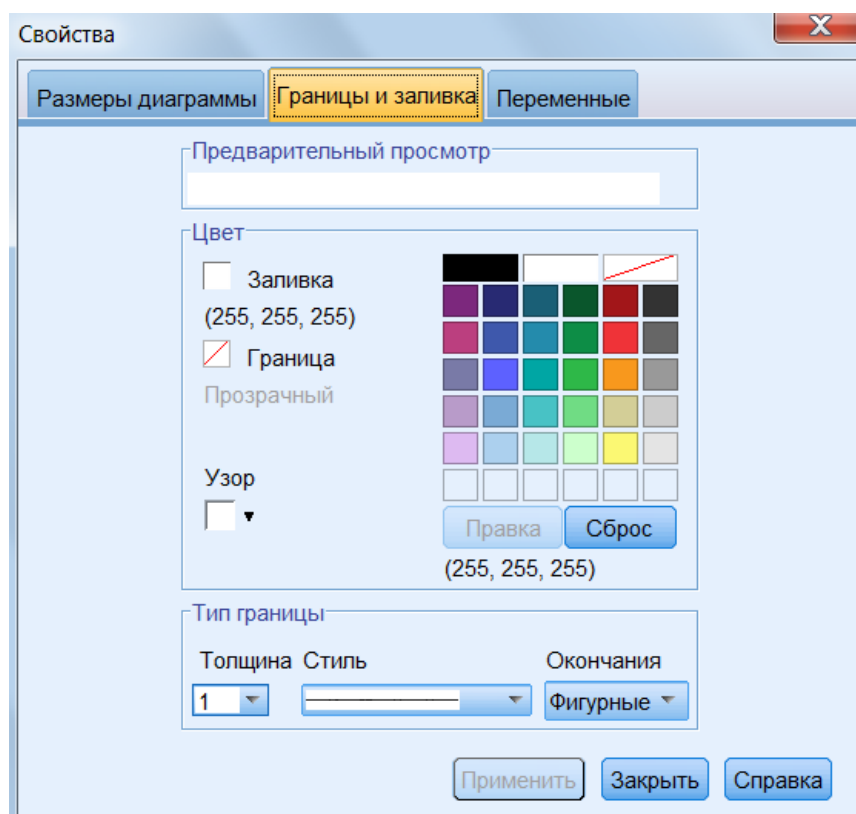


Рис. 48

Как правило, параметры могут применяться к диаграмме одним из трех способов:

^ При установке параметра сразу происходит изменение диаграммы, после чего достаточно щелкнуть по кнопке *Закрыть*, чтобы закрыть диалоговое окно редактирования.

^ Чтобы установленные параметры вступили в силу, требуется щелкнуть на кнопке *Применить* или *Применить ко всем*. Происходит изменение диаграммы, после чего вы можете закрыть диалоговое окно редактирования щелчком на кнопке *Закрыть*.

^ Кнопки *Применить* и *Применить ко всем* недоступны. Это может быть обусловлено двумя причинами: либо параметр к данной диаграмме действительно неприменим, либо для выполнения операции необходимо предварительно выделить какой-либо фрагмент диаграммы<sup>68</sup>.

## Использованная литература

<sup>68</sup>Там же. С. 100–101

Наследов А.Д. SPSS: Компьютерный анализ в психологии и социальных науках : самоучитель / А.Д. Наследов. – СПб. : Питер, 2005. – 416 с.

Наследов А.Д. SPSS 15: профессиональный статистический анализ данных. – СПб.: Питер, 2008. – 416 с.

### **Контрольные вопросы**

1. Каковы Виды стандартных графиков?
2. Опишите возможности SPSS редактирования графиков
3. Что такое Интерактивные графики? Каковы их виды?

### **Практические задания**

1. Откройте на выбор одну из двух баз данных: «первокурсники 2010» или «родители 2010». Просмотрите все вопросы анкеты и определите, какие диаграммы целесообразно использовать для каждого вопроса (запишите в тетрадь).

2. Постройте круговую диаграмму тремя способами:

2.1 при помощи подменю *Графика*;

2.2 при помощи программы *Excel*, взяв данные из частотной таблицы.

3. Постройте простую столбиковую диаграмму тремя способами:

3.1 при помощи подменю *Графика*;

3.2 при помощи программы *Excel*, взяв данные из частотной таблицы.

4. Постройте состыкованную столбиковую диаграмму при помощи программы *Excel*, взяв данные из таблицы сопряженности.

5. Постройте при помощи подменю *Графика* простую столбиковую диаграмму, в которой отражались бы средние значения по одному из табличных вопросов.

6. Постройте в программе *Excel* диаграмму для любого множественного вопроса.

7. Постройте круговую или простую столбиковую диаграмму и отредактируйте ее на свое усмотрение в редакторе диаграмм SPSS.

8. Создайте условную базу данных SPSS, в которой были бы отражены баллы ЕГЭ по обществознанию 20 студентов и их же результаты Интернет-тестирования по социологии. Затем при помощи подменю *Графика*:

А) постройте гистограмму для любой из полученных переменных («Балл ЕГЭ» или «Тестирование»);

Б) постройте коробчатую диаграмму по обоим переменным;

В) постройте диаграмму рассеивания по обоим переменным;

Г) постройте линейный график, который бы показывал взаимосвязь между баллом ЕГЭ и результатом тестирования.

## 2.8. Редактирование таблиц. Модуль Tables

### 1. Редактирование таблиц.

Программа SPSS позволяет проводить простейшее редактирование таблиц с выходными данными. Аналогично редактированию диаграмм, чтобы войти в режим редактирования таблиц необходимо дважды щелкнуть по выбранной таблице. В результате она выделится пунктирной рамкой, либо откроется отдельное окно Мобильная таблица (при большом формате таблицы). Чтобы отредактировать конкретный элемент таблицы (общий заголовок, заголовки столбцов, названия категорий, конкретные численные значения), необходимо также дважды щелкнуть по нему. Подтверждение редактирования происходит путем нажатия кнопки *ВВОД* или щелчка по другому элементу таблицы. Общее редактирование позволяет корректировать надписи и удалять их. Чтобы удалить надпись, нужно одним щелчком выделить ее и нажать кнопку *DELETE*.

Нажатие правой кнопки мыши по таблице в режиме редактирования вызывает всплывающее меню, с помощью которого становится возможным изменить свойства таблицы, свойства ячеек, шаблон таблицы,

вставить/удалить/скрыть сноску, вызвать панель инструментов, создать диаграмму (рис. 49).

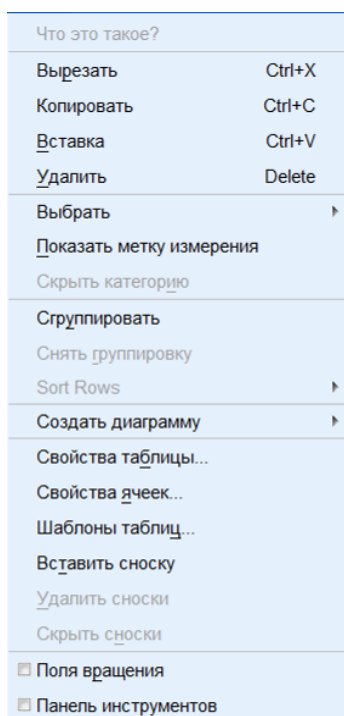


Рис. 49

Фиксация изменений в таблице происходит путем щелчка по пустому полю **Окна вывода** или закрытия окна *Мобильная таблица*.

## 2. Экспортирование таблиц в Word, Excel.

Есть два способа экспортирования таблиц из SPSS в документы иных форматов:

- 1) путем команд *Копировать* и *Вставить*;
- 2) с помощью процедуры *Экспортировать*.

Последняя позволяет перенести либо все таблицы, представленные в **Окне вывода**, либо конкретную таблицу отдельно. В первом случае нужно щелкнуть правой кнопкой мыши по пустому полю **Окна вывода** и выбрать в всплывающем меню команду *Экспортировать*, во втором – щелчок производится по конкретной таблице (рис. 50).

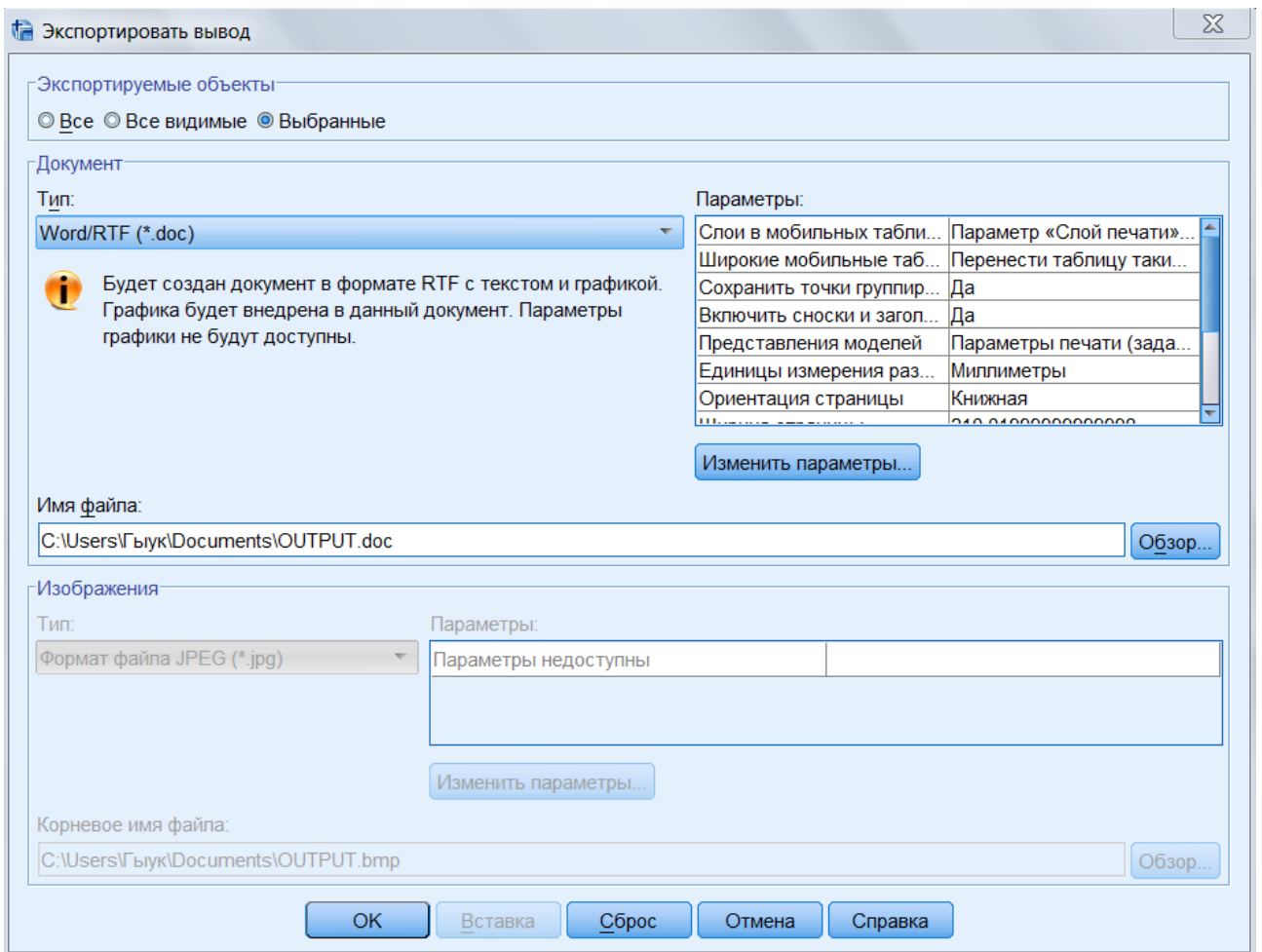


Рис. 50

В окне *Экспортирования* необходимо выбрать, какие объекты экспортировать:

- 1) Все (переносятся все объекты, включая дату и время создания **Окна вывода**, полное имя файла данных, наличие фильтра и расщепления и т.д.);
- 2) Все видимые (переносятся только те объекты, которые визуально отражены в **Окне вывода**);
- 3) Выбранные (переносятся только выделенные объекты).

На практике чаще всего используется 2 или 3 опция.

Далее указывается тип документа, в который требуется экспортировать объекты (форматы документов Word, Excel, веб-страниц, PDF и т.д.), и имя файла, в который экспортируются объекты. Также, возможно задать частные параметры экспортирования (кнопка *Изменить параметры*).

### 3. Использование модуля Tables.

Модуль Tables служит для создания таблиц, готовых к презентации (рис. 51). По сравнению с режимом построения частотных таблиц и таблиц сопряженности, а также таблиц средних значений, в этом модуле пользователю предоставляются более широкие возможности<sup>69</sup>.

## Анализ

### Таблицы

#### Настраиваемые таблицы...

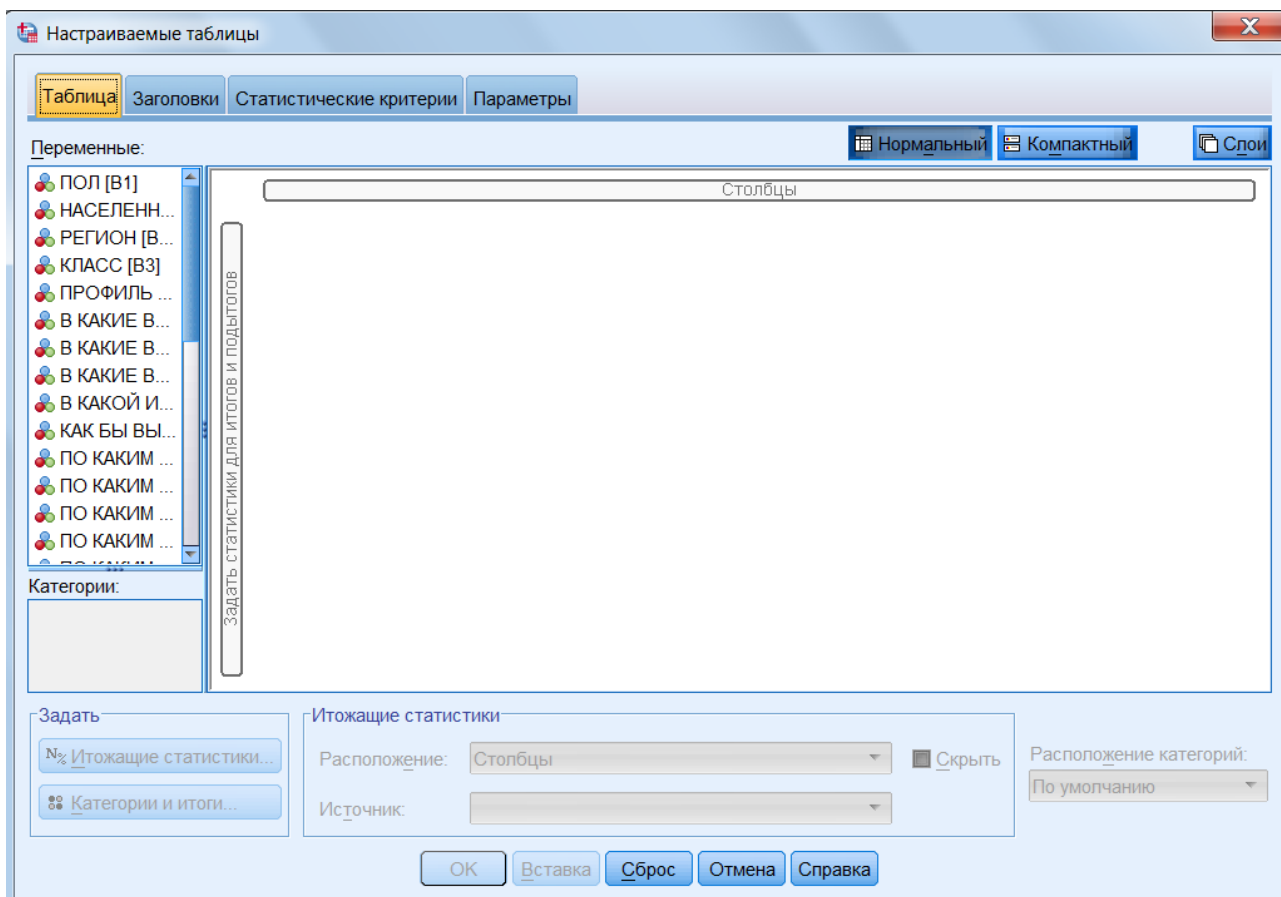


Рис. 51

Вкладка **Таблица** позволяет выбрать конкретные вопросы для статистической обработки (частотной или перекрестной), определить *Итожащие статистики* (частоты, проценты по столбцу/строке, валидные проценты по столбцу/строке и т.д.), а также выбрать конкретные варианты ответа на вопрос анкеты, которые необходимо учитывать при анализе

<sup>69</sup> Бююль А., Цёфель П. SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей : Пер. с нем. / А. Бююль, П. Цёфель. – СПб.: ООО «ДиаСофтЮП», 2005. – С. 499

(например, в случае если есть необходимость убрать из анализа Отказ от ответа и Затруднившихся ответить).

Вкладка **Заголовки** позволяет задать Заголовок таблицы, Подпись к ней и надпись в Углу таблицы. Также можно установить автоматическую Дату, Время и Табличное выражение.

Вкладка **Статистические критерии** служат для вывода критерия хи-квадрат, t-критерия и z-критерия.

Вкладка **Параметры** позволяет установить вид пустых ячеек, ширину столбцов и другое.

Для анализа множественных ответов в модуле Tables необходимо сначала задать наборы переменных с помощью следующей команды (рис. 52):

## Анализ

### Таблицы

#### Наборы множественных ответов...

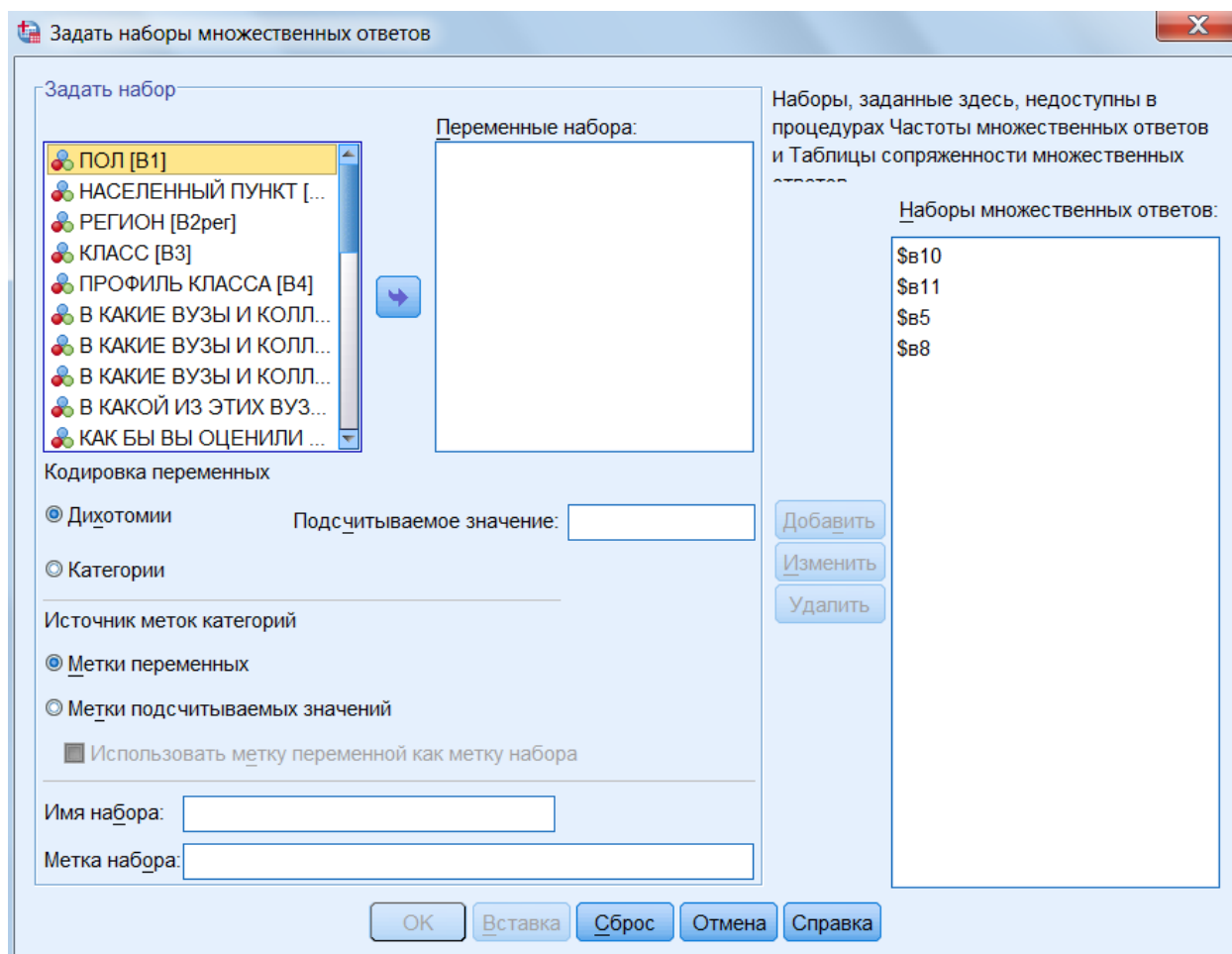


Рис. 52

Затем строится требуемая таблица описанным выше способом.

Модуль Tables также используется для построения сводных таблиц.

Фрагмент такой таблицы представлен на рис. 53:

	ПОЛ		РЕГИОН				без профиля / профиль не указан	социально-гуманитарный	социально-экономический	МЧС
	Юноши	Девушки	Благовещенск	Районы						
	% по столбцу	% по столбцу	% по столбцу	% по столбцу	% по столбцу	% по столбцу				
ПОДАЧА ДОКУМЕНТОВ В ВУЗЫ И КОЛЛЕДЖИ	АмГУ	89%	95%	94%	89%	100%	100%	88%	100%	
	АГМА	3%	5%	5%	0%	0%	0%	0%	0%	
	БГПУ	6%	36%	27%	11%	13%	28%	41%	25%	
	ДальГАУ	3%	9%	6%	11%	0%	11%	12%	0%	
	ДВФУ	14%	26%	19%	44%	7%	22%	18%	75%	
	МГУ	0%	2%	1%	0%	7%	0%	0%	0%	
	СПбГУ	3%	8%	6%	0%	7%	6%	0%	0%	
	юридические при МВД (ОрЮИ, ДВЮИ...)	3%	9%	8%	0%	13%	0%	6%	75%	
	МГТУ	3%	0%	1%	0%	0%	0%	0%	0%	
	ДВВКУ	3%	0%	1%	0%	0%	0%	6%	0%	
	ДВГМУ	0%	2%	1%	0%	0%	0%	0%	0%	
	СПбГАВМ	0%	2%	1%	0%	0%	0%	0%	0%	
	ТОГУ	3%	8%	5%	11%	0%	17%	0%	0%	
	ВГУЭС	0%	2%	0%	11%	0%	0%	0%	0%	
	СФУ	0%	2%	0%	11%	0%	0%	0%	0%	
	ДВГУПС	0%	0%	0%	0%	0%	0%	0%	0%	
	СПбГИКИТ	3%	0%	1%	0%	0%	0%	0%	0%	
	Академия художеств им. Репина	0%	2%	1%	0%	0%	0%	0%	0%	
	Иннополис	3%	0%	0%	11%	7%	0%	0%	0%	
	МЗИ	0%	2%	1%	0%	0%	0%	0%	0%	
	ХПИ	3%	0%	1%	0%	0%	0%	0%	0%	
	РАП (РГУП)	3%	0%	1%	0%	0%	0%	0%	0%	

Рис. 53

### Использованная литература

Бююль А., Цёфель П. SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей : Пер. с нем. / А. Бююль, П. Цёфель. – СПб.: ООО «ДиаСофтЮП», 2005. – 608 с.

### Контрольные вопросы

1. Как вызвать Редактор мобильных таблиц?
2. Как Экспортировать таблицы в Word, Excel?
3. Для чего нужен Модуль Tables?

### Практические задания

1. Откройте на выбор одну из трех баз данных: «куратор 2011», «первокурсники 2010» или «родители 2010». Создайте одну простую таблицу для альтернативного вопроса, используя команды *Анализ – Описательные статистики – Частоты...*, и одну таблицу сопряженности для альтернативных



вопросов, используя команды *Анализ – Описательные статистики – Таблицы сопряженности...*

2. Отредактируйте полученные таблицы средствами SPSS так, чтобы они приобрели следующие параметры:

2.1 для простой таблицы:

- применен шаблон AvantGarde;
- удалены столбцы с частотами, обычными и кумулятивными процентами

(должен остаться только валидный процент);

- выделена красным цветом ячейка с наибольшим процентом;

2.2 для таблицы сопряженности:

- все элементы, кроме названия таблицы, имеют шрифт Times New Roman с размером 12 пунктов;

- название таблицы имеет шрифт Times New Roman с размером 14 пунктов, курсив;

- все ячейки с процентами имеют зеленый цвет;

- вставлена сноска с названием базы данных.

3. Постройте сводную таблицу с помощью модуля *Таблицы*, исходя из следующего условия: в столбцах размещены два-три социально-демографических вопроса, в строках – три альтернативных вопроса и один множественный.

4. Переместите с помощью процедуры *Экспортировать...* все видимые объекты в файл World с названием «экспорт таблиц», поместив его в папке Мои документы.

## **2.9. Многомерные методы анализа данных: обзор**

### **1. Корреляционный анализ.**

Корреляция – мера зависимости двух или нескольких величин; отношение между признаками, в котором в зависимости от изменения одного признака изменяется значение другого признака.

Типы корреляции: прямая (положительная) и обратная (отрицательная).

Формы корреляции: линейная и криволинейная.

Теснота связи – степень сопряженности между признаками.

Основные коэффициенты корреляции и условия их применения приведены в табл. 6:

Таблица 6

Коэффициент корреляции	Границы применимости
Коэффициент Пирсона	Количественные и номинальные переменные, а также для линейной связи
Коэффициент Кендалла	Порядковые переменные, а также для криволинейной связи
Коэффициент Спирмена	Порядковые переменные, а также для криволинейной связи

## Анализ

### Корреляции

#### Парные... (рис. 54)

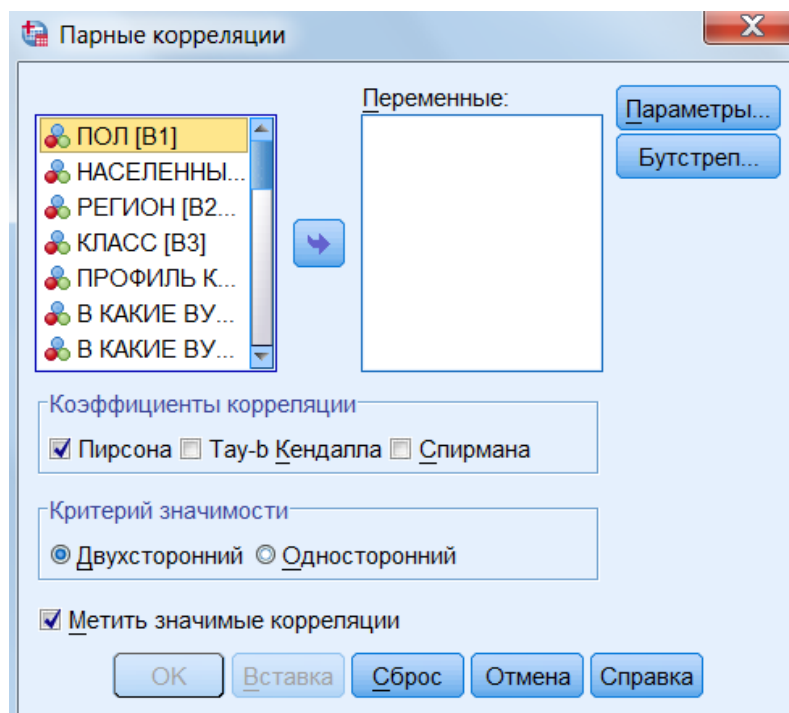


Рис. 54

Интерпретация значений коэффициентов представлена ниже (табл. 7):

Таблица 7

Значение коэффициента	Вербальная интерпретация
До 0,2	Очень слабая корреляция
До 0,5	Слабая корреляция
До 0,7	Средняя корреляция
До 0,9	Сильная корреляция
Свыше 0,9	Очень сильная корреляция

Уровень значимости – мера достоверности результата.

## 2. Регрессионный анализ.

Регрессионный анализ позволяет прогнозировать значение зависимой переменной на основе значения независимой переменной.

Уравнение регрессии (формула 2):

$$y = bx + a, \quad (2)$$

где  $y$  – зависимая переменная (прогнозируемая)

$b$  – коэффициент регрессии

$x$  – независимая переменная

$a$  – константа.

### Анализ

#### Регрессия

##### Линейная... (рис. 55)

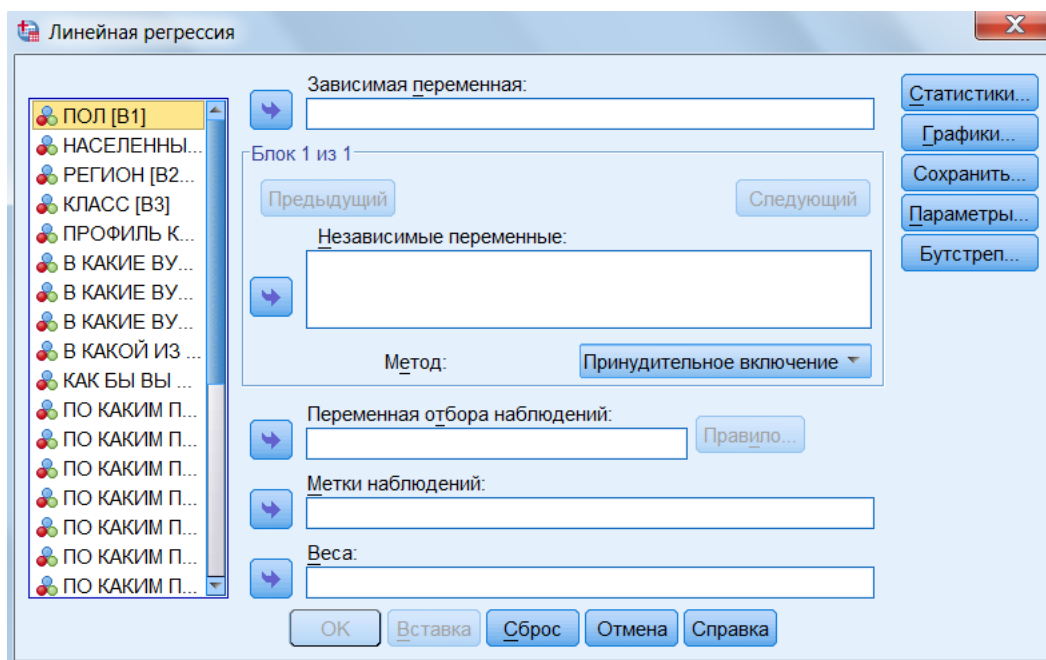


Рис. 55

R-квадрат – величина, характеризующая качество регрессионного уравнение, его способность к прогнозу.

## 3. Классификация данных.

Кластерный анализ позволяет классифицировать объекты (респондентов, регионы и т.д.), т.е. объединить их в несколько групп.

### Анализ

## Классификация

### Иерархическая кластеризация... (рис. 56)

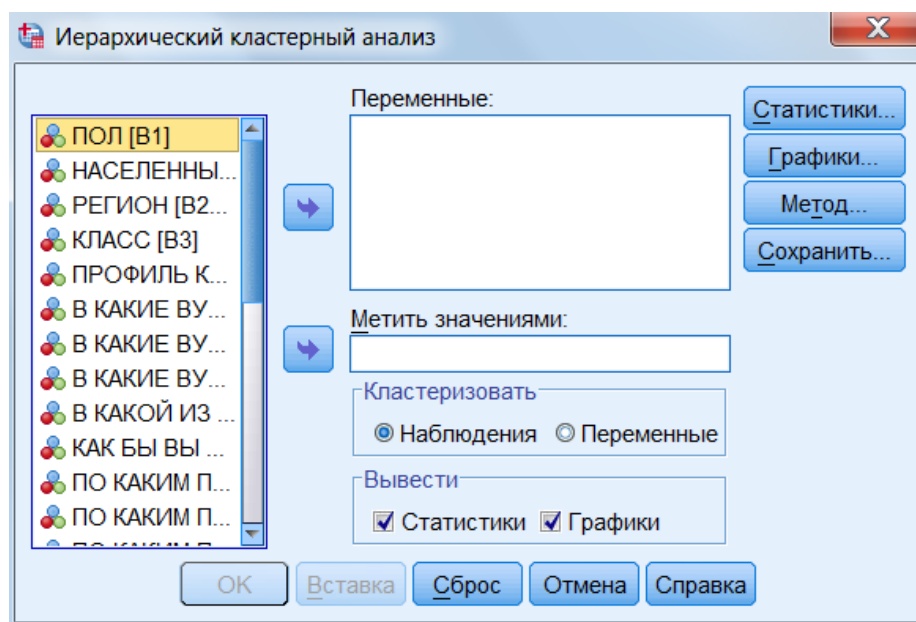


Рис. 56

*Статистики...* – позволяет задать количество кластеров.

*Графики...* – позволяет вывести дендрограмму.

*Метод...* – позволяет задать метод связи объектов, меру расстояния между ними, а также стандартизировать данные.

*Сохранить...* – позволяет сохранить результаты кластеризации в *Редакторе данных*.

#### 4. Факторный анализ.

Факторный анализ позволяет большое число переменных (признаков) свести к меньшему количеству факторов, то есть группирует признаки.

#### Анализ

##### Снижение размерности

### Факторный анализ... (рис. 57)

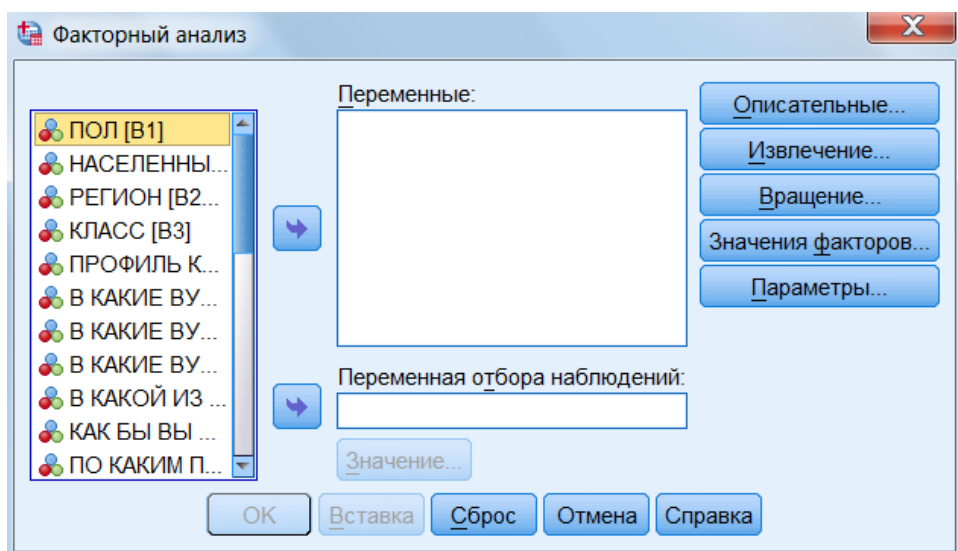


Рис. 57

*Извлечение...* – позволяет задать метод выделения факторов.

*Вращение...* – позволяет задать метод вращения факторов для оптимизации факторной модели.

### Использованная литература

Бююль А., Цёфель П. SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей : Пер. с нем. / А. Бююль, П. Цёфель. – СПб.: ООО «ДиаСофтЮП», 2005. – 608 с.

Крыштановский А.О. Анализ социологических данных с помощью пакета SPSS : учеб. пособие для вузов / А.О. Крыштановский; Гос. ун-т – Высшая школа экономи-ки. – М. : Изд. дом ГУ ВШЭ, 2006. – 281 с.

### Контрольные вопросы

1. Что такое корреляционно-регрессионный анализ?
2. Каковы Основные методы классификации?
3. Для чего нужно Снижение размерности данных?

### Практические задания

1. Воспользовавшись статистическим сборником (Регионы России. Социально-экономические показатели. 2011 : Стат. сб. / Росстат. – М. : 2011. –

990 с.), скопируйте в пустую базу данных SPSS показатели развития науки и инноваций по всем регионам.

2. Определите наличие и силу связи между показателями. Выделите наиболее сильно связанные показатели.

3. Постройте регрессионную модель, определив в качестве зависимой переменной результативность научных организаций в регионе. Независимые переменные задайте самостоятельно.

4. Классифицируйте регионы России по научному потенциалу.

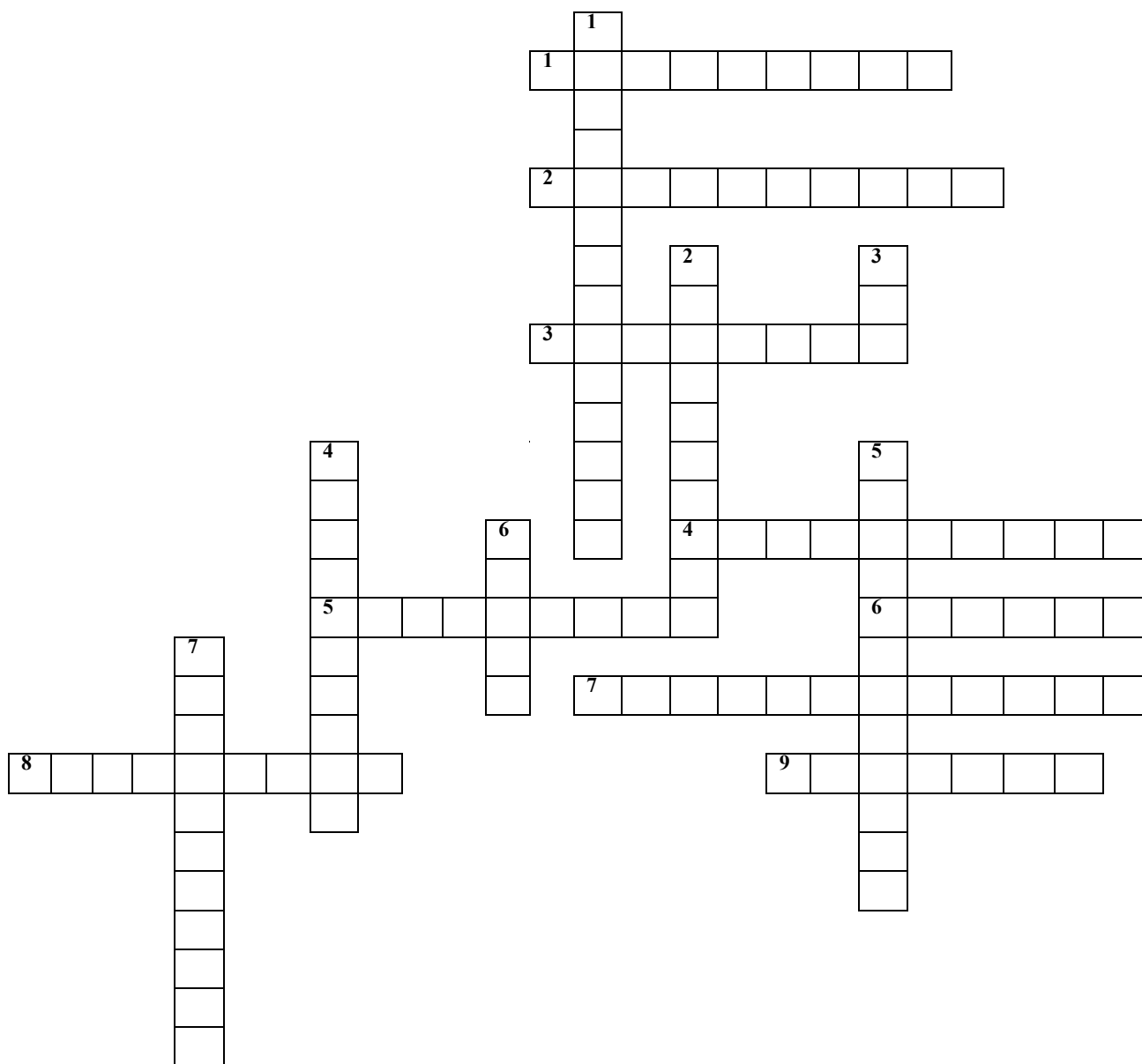
5. Определите факторы научного развития регионов России.

## **ПРОМЕЖУТОЧНЫЕ ПРОВЕРОЧНЫЕ РАБОТЫ И КРОССВОРДЫ К НИМ**

Проверочная работа предполагает проверку: 1) знания студентом основных принципов и назначения отдельной процедуры статистической обработки данных и 2) умения ее применять при работе с базой данных.

### **Проверочная работа по главе 1**

- Определение переменных. Просмотр и редактирование описаний переменных
- Ввод и редактирование данных
- Объединение данных. Слияние файлов данных
- Оценка качества первичных данных. Чистка массива данных
- Сортировка наблюдений
- Разделение наблюдений на группы
- Выбор наблюдений. Построение условного выражения
- Извлечение случайной выборки
- Перекодирование значений переменной
- Создание и вычисление новых переменных
- Подсчет частоты появлений определенных значений
- Построение индекса
- Агрегирование данных
- Ранговые преобразования. Типы рангов
- Взвешивание наблюдений
- Типы статистических шкал
- Проверка закона распределения. Нормальное распределение



*По горизонтали:* 1. Ошибки ввода, характерные для квалифицированных интервьюеров, кодировщиков, операторов и т.д. 2. Признак, свойственный каждому респонденту. 3. Вопрос в анкете, который не имеет готовых вариантов ответа. 4. Метод диагностики базы данных, который заключается в визуальном просмотре 10 % полученных от интервьюера анкет на предмет ошибок, неточностей, нелогичности в фиксировании ответов респондента. 5. Вопрос в анкете, представляющий собой совокупность подвопросов, каждый из которых описывается отдельной переменной. 6. Как называется определенное количество наблюдений с одинаковым значением признака? 7. Как называются переменные, по которым расщепляют базу данных? 8. Процедуры, решающие

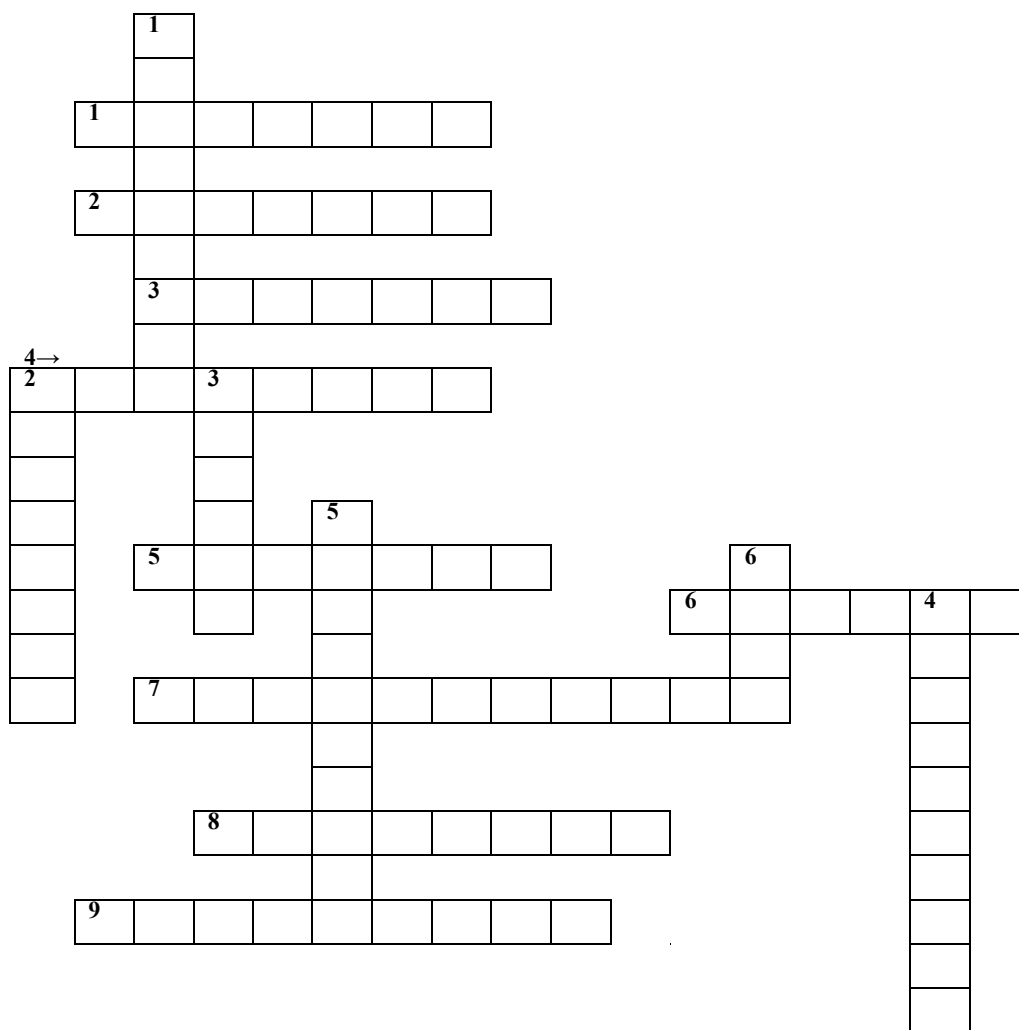


задачу эффективного управления данными. 9. Тип ранга, который используется чаще всего в ранговых преобразованиях.

*По вертикали:* 1. Вопрос в анкете, который предполагает выбор респондентом только одного варианта ответа. 2. Тип переменных, значение которых поддаются сортировке от меньшего к большему, и наоборот. 3. Один из создателей программного комплекса SPSS. 4. Какой параметр предназначен для задания числа знаков после запятой в случае, если переменная допускает использование дробных чисел? 5. Переменные, на основе которых вычисляются значения новой переменной. 6. Как называется процедура, которая помогает найти конкретное расположение ошибки ввода? 7. На основе первоначально собранных данных можно создать новые переменные и изменить кодирование. Как называют подобные преобразования данных?

## **Проверочная работа по главе 2**

- Частотные таблицы. Расчет частот
- Описательные статистик
- Исследование данных
- Сводка по наблюдениям
- Построение двумерных таблиц
- Анализ множественных ответов. Дихотомный и категориальный методы
- Редактирование графиков. Стандартные и интерактивные графики
- Редактирование мобильных таблиц
- Многомерные методы анализа данных



*По горизонтали:* 1. Какая процедура позволяет строить статистические ряды распределения, с построения которых начинается анализ данных, прошедших контроль в системе SPSS? 2. В какой графе частотной таблицы SPSS учитывает долю всех опрошенных? 3. Точка на шкале измеренных значений, выше и ниже которой лежит по половине всех измеренных значений. 4. Точка на шкале измеренных значений, меньше которой располагаются 25 %, 50 %, 75 % измеренных значений. 5. Какая вкладка позволяет выбрать конкретные вопросы для обработки, определить итожащие статистики, выбрать конкретные варианты ответа на вопрос анкеты, построить сводные таблицы? 6. С помощью какой кнопки можно вывести частотную таблицу для интересующей переменной с сортировкой ответов? 7. С помощью чего иллюстрируют распределение наблюдений по диапазону значений интервальных переменных? 8. Какой график чаще всего используется для

отображения зависимости между переменными, а также для отображения динамики показателей? 9. Квадрат стандартного отклонения.

*По вертикали:* 1. Что повышает наглядность, результатов частотного распределения и может использоваться для презентации результатов социологического исследования? 2. Какие диаграммы целесообразно использовать только для альтернативных вопросов? 3. Как называется разница между наибольшим и наименьшим значениями? 4. Какая статистика показывает, в какую сторону относительно среднего сдвинуто большинство значений распределения? 5. Один из типов столбчатых диаграмм. 6. Значение, которое наиболее часто встречается в выборке.

# ИТОГОВЫЙ ТЕСТ

## Вариант 1

### 1. Кем была создана программа SPSS?

- А) Стив Джобс и Билл Гейтс
- Б) Норман Най и Дейл Вент

### 2. Расставьте по порядку этапы обработки данных:

- А) подготовительный этап
- Б) анализ данных и подготовка отчета
- В) контроль данных
- Г) получение результатов статистических процедур
- Д) ввод и корректировка данных

### 3. Какая переменная является количественной?

- А) пол респондента
- Б) средний балл ЕГЭ абитуриента
- В) номер студенческой группы
- Г) название населенного пункта

### 4. Параметр ... позволяет ввести формулировку вопроса из анкеты.

- А) значение
- Б) пропуски
- В) столбцы
- Г) метка

### 5. Соотнесите операторы и их значения:

А) $\sim =$	Б) не равно
В) $\&$	Г) И
Д) $ $	Е) ИЛИ
Ж) $\sim$	З) НЕ

### 6. Какого окна в SPSS не существует?

- А) окно приветствия
- Б) диалоговое окно
- В) окно чата
- Г) окно редактора данных

**7. Альтернативный вопрос предполагает:**

- А) выбор респондентом одного варианта ответа
- Б) выбор респондентом нескольких вариантов ответа
- В) предложение респондентом своего варианта ответа

**8. Где фиксируется объект?**

- А) в строках
- Б) в столбцах
- В) в таблице

**9. Какое действие позволяет управлять расположением данных внутри ячейки?**

- А) выравнивание
- Б) шкалирование
- В) расширение

**10. Для чего нужен контроль данных?**

- А) для исправления ошибок
- Б) для добавления новых переменных
- В) чтобы избежать повторений

**11. Какая «иконка» используется для команды «Найти»?**

- А) лупа
- Б) компас
- В) бинокль

**12. Обязательно ли соответствие каждой переменной рабочего файла данным внешнего файла?**

- А) да
- Б) нет
- В) только для тех файлов, для которых создана резервная копия

**13. Для визуализации каких вопросов используются только столбчатые диаграммы?**

- А) табличных
- Б) открытых

В) множественных

**14. На какие группы делятся статистические процедуры?**

А) аналитические статистики

Б) описательные статистики

В) множественный статистики

Г) факторные статистики

**15. Какая процедура позволяет строить статистические ряды распределений?**

А) «частоты»

Б) «статистики»

В) «анализ»

**16. Какой пункт подменю позволяет получить описательную статистику для числовых переменных?**

А) «анализ»

Б) «статистики»

В) «частоты»

Г) «таблицы»

**17. Выберите правильный алгоритм построения таблицы сопряженности:**

А) описательные статистики – таблицы сопряженности – анализ

Б) анализ – описательные статистики – таблицы сопряженности

В) анализ – таблицы сопряженности – описательные статистики

**18. Квадрат стандартного отклонения – это...**

А) размах

Б) мода

В) медиана

Г) дисперсия

**19. Мера сглаженности нормального распределения – ...**

А) размах

Б) среднее

- В) эксцесс
- Г) медиана

## Вариант 2

**1. В каком году была создана программа SPSS?**

- А) 1965
- Б) 1988
- В) 2001
- Г) 2005

**2. Какой из вопросов предполагает выбор респондентом только одного ответа?**

- А) альтернативный
- Б) множественный
- В) открытый
- Г) полузакрытый

**3. Округление до ближайшего целого числа обозначается в SPSS...**

- А) abs
- Б) rnd
- В) sqrt

**4. Расставьте в правильном порядке шаги выполнения процедуры «Итоги по наблюдениям»:**

- А) отчеты
- Б) анализ
- В) итоги по наблюдениям

**5. Какой тип шкалы SPSS не выделяет как отдельный?**

- А) номинальный
- Б) порядковый
- В) дихотомический

**6. Соотнесите тип шкалы с допустимыми мерами средней тенденции:**

А) номинальная	Б) мода, медиана, среднее арифметическое
В) порядковая	Г) мода, медиана
Д) интервальная	Е) мода

**7. Вопрос, представляющий собой единичную переменную, называется:**

- А) табличный
- Б) простой
- В) закрытый

**8. Какой из типов переменной чаще всего используется как основание для деления объектов на группы?**

- А) количественный
- Б) порядковый
- В) категориальный

**9. Какой из способов описания множественных вопросов НЕ существует?**

- А) категориальный
- Б) дихотомический
- В) ранговый

**10. Какой алгоритм вывода Итогов по наблюдениям?**

- А) отчеты – анализ – итог по наблюдениям
- Б) анализ – отчет – итог по наблюдениям
- В) контроль данных – отчет – итог по наблюдениям

**11. Для чего нужны сервисные процедуры?**

- А) для эффективного управления данными
- Б) для анализа наблюдений
- В) для решения задач

**12. Для чего нужно агрегирование данных?**

- А) для подсчета повторений одного или нескольких значений в списке переменных



Б) для создания таких значений переменных, которые представляют собой результат объединения группы исходных значений

В) для создания переменных, по которым ранжируют базу данных

**13. Какой из типов столбчатых диаграмм используется для визуализации таблиц сопряженности?**

А) простые

Б) составные

В) кластерные

**14. Что является первым этапом статистического анализа данных?**

А) описательный анализ

Б) сравнительный анализ

В) частотный анализ

Г) факторный анализ

**15. Чем отличаются процент и валидный процент? (найдите соответствие)**

А) Процент	Б) Учитывает только давших содержательный ответ
В) Валидный процент	Г) Учитывает всех опрошенных

**16. Выберите вариант, который не относится к синонимам понятия «таблицы сопряженности»:**

А) перекрестные таблицы

Б) кросс-таблицы

В) сравнительные таблицы

Г) двумерные таблицы

**17. Расставьте шаги алгоритма анализа множественных ответов в правильном порядке:**

А) таблицы

Б) наборы множественных ответов

В) анализ

Г) настраиваемые таблицы

**18. Нормальное распределение симметрично тогда, когда...**

А) асимметрия  $> 0$

Б) асимметрия  $= 0$

В) асимметрия  $< 0$

**19. Какие типы диаграмм целесообразно использовать для визуализации перекрестных таблиц?**

А) простые и кластерные

Б) кластерные и состыкованные

В) простые и состыкованные

## РЕКОМЕНДУЕМАЯ ЛИТЕРАТУРА

### Основная

1. Бююль А., Цёфель П. SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей : Пер. с нем. / А. Бююль, П. Цёфель. – СПб.: ООО «ДиаСофтЮП», 2005. – 608 с.

2. Иллюстрированный самоучитель по SPSS [Электронный ресурс]. URL: [www.learnspss.ru](http://www.learnspss.ru)

3. Крыштановский А.О. Анализ социологических данных с помощью пакета SPSS : учеб. пособие для вузов / А.О. Крыштановский; Гос. ун-т – Высшая школа экономи-ки. – М. : Изд. дом ГУ ВШЭ, 2006. – 281 с.

4. Наследов А.Д. SPSS : Компьютерный анализ в психологии и социальных науках : самоучитель / А.Д. Наследов. – СПб. : Питер, 2005. – 416 с.

5. Наследов А.Д. SPSS 15: профессиональный статистический анализ данных. – СПб.: Питер, 2008. – 416 с.

6. Наследов А.Д. SPSS 19: профессиональный статистический анализ данных / А.Д. Наследов. – СПб. : Питер, 2011. – 400 с.

7. Пациорковский В.В., Пациорковская В.В. SPSS для социологов: Учебное пособие. ИСЭПН РАН. – М., 2005. – 433 с.

8. SPSS: обработка статистической информации [Электронный ресурс]. URL: [www.ispss.ru](http://www.ispss.ru)

### Дополнительная

Иллюстрированный самоучитель по SPSS [Электронный ресурс] / Компьютерная документация Hardline.RU. – М., 2006. – URL : <http://www.hardline.ru>.

Крамер Д. Математическая обработка данных в социальных науках : уч. пособие : рек. УМО / Д. Крамер; пер. с англ. И.В. Тимофеева, Я.И. Киселевой; ред. О.В. Митин. – М. : Академия, 2007. – 288 с.

Моосмюллер Г. Маркетинговые исследования с SPSS : уч. пособие : рек. УМО / Г. Моосмюллер, Н.Н. Ребик. – М. : Инфра-М, 2007. – 160 с.

Мхитарян С.В. Применение SPSS в маркетинговых проектах [Электронный ресурс] : учеб.-практ. пособие / С.В. Мхитарян. – М. : Евразийский открытый институт, 2011. – 272 с.

Таганов Д.Н. SPSS : статистический анализ в маркетинговых исследованиях / Д.Н. Таганов. – СПб. : Питер, 2005. – 192 с.

Электронный учебник StatSoft [Электронный ресурс] : Электронный учебник по статистике. – М. : StatSoft, 2001. – URL : <http://www.statsoft.ru/home/textbook/default.htm>.

## ПРИЛОЖЕНИЕ ПАМЯТКА ПО РАБОТЕ В SPSS

**Этапы статистического анализа:** 1) подготовительный; 2) ввод и корректировка данных; 3) контроль данных; 4) статистический обработка данных; 5) анализ и интерпретация результатов стат.обработки.

Условно работу в SPSS можно разделить на **два блока:** 1) ввод данных и их предварительная обработка; 2) статистическая обработка данных.

**Окна** в программе SPSS: 1) окно приветствия; 2) окно *Редактора данных*; 3) окно открытия файла; 4) диалоговые окна процедур; 5) окно вывода результатов.

### **РАЗДЕЛ 1. Ввод данных и их предварительная обработка**

Ввод данных осуществляется в окне *Редактора данных*, которое имеет две вкладки: *Переменные* и *Данные*. Работа начинается с создания макета анкеты во вкладке *Переменные*:

**ИМЯ** – цифровое имя вопроса анкеты (часто совпадает с номером вопроса и должно начинаться с буквы, например, В1, то есть вопрос №1).

**ТИП** – тип вводимой информации (чаще всего числовая (по умолчанию), реже – текстовая).

**ШИРИНА** – количество цифр в коде ответа в анкете (по умолчанию 8).

**ДЕСЯТИЧНЫЕ** – количество знаков после запятой во вводимых числах (чаще всего обнуляется).

**МЕТКА** – формулировка вопроса в анкете.

**ЗНАЧЕНИЯ** – формулировка ответов на данный вопрос анкеты (*Значение* – цифровой код ответа из анкеты, *Метка* – формулировка ответа с данным кодом).

**ПРОПУСКИ** – фиксация отсутствия ответа, отказа от ответа, а также затруднившихся ответить.

**СТОЛБЦЫ** – ширина столбца в *Редакторе данных* (по умолчанию 8).

**ВЫРАВНИВАНИЕ** – выравнивание вводимой информации в ячейках таблицы *Редактора данных*.

**ШКАЛА** – тип шкалы (номинальная, порядковая, количественная).

***Особенности описания отдельных типов вопросов:***

- *Альтернативный вопрос* – описывается с помощью одной переменной (для ответа на такой вопрос резервируется только одна ячейка).
- *Множественный вопрос* – 1) либо описывается столькими переменными, сколько респондент может дать ответов на данный вопрос (категориальный метод), либо 2) создают отдельную переменную для каждого возможного варианта ответа (дихотомный метод).
- *Табличный вопрос* – это совокупность подвопросов, каждый из которых описывается отдельной переменной.

После создания макета анкеты, то есть ее описания во вкладке *Переменные*, приступают к вводу массива анкет во вкладке *Данные*.

***Особенности ввода ответов на отдельные типы вопросов:***

- Ввод ответа на *альтернативный вопрос* – это ввод одной цифры, кодирующей данный ответ респондента.
- В случае с *множественным вопросом* поступают следующим образом: либо 1) каждый отдельный ответ респондента на такой тип вопроса кодируется в отдельную ячейку (категориальный метод), либо 2) кодируется каждый возможный выбор варианта ответа респондентом (дихотомный метод).
- *Табличный вопрос* – это совокупность подвопросов. Каждый такой подвопрос кодируется в базе данных и вводится отдельно.

- Ответы, данные респондентом на *открытый вопрос*, необходимо закодировать вручную, то есть присвоить каждому новому ответу соответствующий код.

## **РАЗДЕЛ 2. Статистическая обработка данных**

Методы статистической обработки данных в SPSS условно делятся на описательные и аналитические. Первичными и наиболее распространенными являются методы описательной статистики.

### ***1. Одномерные таблицы (простые таблицы) для альтернативных вопросов.***

**Анализ – Описательные статистики – Частоты...** (кнопка *Статистики...* позволяет вычислить при необходимости и целесообразности различные статистические характеристики; кнопка *Диаграммы...* позволяет построить различные диаграммы).

*Результат:* простая таблица с частотой ответов, их процентом (вычисляется от всех респондентов), валидным процентом (вычисляется только от тех респондентов, кто дал содержательный ответ на данный вопрос, то есть не пропустил вопрос и не затруднился ответить), кумулятивным (накопленным) процентом.

### ***2. Перекрестные таблицы (кросс-таблицы, таблицы сопряженности) для альтернативных вопросов.***

**Анализ – Описательные статистики – Таблицы сопряженности...**

В данном случае анализ ведется одновременно по двум вопросам (чаще всего по содержательному и демографическому). Для этого один из них вводится в окошко *Строки*, а другой – в окошко *Столбцы*. Далее с помощью кнопки *Ячейки...* определяется способ подсчета процентов (как правило, выбирается в зависимости от того, где находится демографический вопрос: если в *Строках*, то «по строке», если в *Столбцах*, то «по столбцу»).

*Результат:* перекрестная таблица с подсчетом процента по одному вопросу внутри другого.

### **3. Статистическая обработка множественных вопросов.**

Предварительно необходимо задать множественные ответы:

#### ***Анализ – Множественные ответы – Задать наборы переменных...***

Далее, выбирают все переменные, описывающие данный множественный вопрос. Затем, в зависимости от метода ввода ответов:

- категориальный метод: задается диапазон кодов ответов (например, если ответы в анкете кодируются так: 1, 2, 3, 97, – то следует указать диапазон от 1 до 100)
- дихотомный метод: задается значение, которое необходимо подсчитать (если факт ответа респондента отмечался цифрой 1, то следует задать 1).

Затем задается имя переменной (см. выше ИМЯ) и метка (см. выше МЕТКА).

#### ***Для частотного анализа множественных вопросов:***

##### ***Анализ – Множественные ответы – Частоты...***

*Результат:* простая таблица с частотой ответов, их процентом (от числа ответов) и процентом наблюдений (от числа респондентов). Как правило, для анализа используется процент наблюдений (его сумма больше 100%, т.к. респондент мог выбрать более одного варианта ответа).

#### ***Для перекрестных таблиц с множественным вопросом (множественными вопросами):***

##### ***Анализ – Множественные ответы – Таблицы сопряженности...***

Способ расчета процентов задается с помощью кнопки *Параметры...* аналогично перекрестной таблицы для альтернативного вопроса.

**!Примечание!** в данном случае необходимо задать диапазон кодов ответов и для альтернативного вопроса, участвующего в анализе.



*Результат:* перекрестная таблица.

***Аналитические методы:***

Корреляции. *Анализ – Корреляции*

Регрессия. *Анализ – Регрессия*

Классификация. *Анализ – Классификация*

Факторный анализ. *Анализ – Снижение размерности – Факторный анализ...*

***Некоторые сервисные процедуры:***

*Данные – Расщепить файл...* (позволяет разбить базу на отдельные группы (например, по полу, факультету...) и вести анализ отдельно по каждой).

*Данные – Отобратить наблюдения...* (позволяет выбрать для анализа только одну группу респондентов (например, только мужчин, только студенток ФСН...))

***Основы модификации данных:***

*Преобразовать – Перекодировать в те же переменные...* (позволяет улучшить кодировку существующей переменной).

*Преобразовать – Перекодировать в другие переменные...* (позволяет создать новую переменную на основе перекодировки существующей).

*Преобразовать – Вычислить переменную...* (позволяет вычислить новую переменную на основе имеющихся данных).

**Леонов Аркадий Константинович**  
*доцент, кандидат социологических наук*

**Основы применения SPSS в социологии. Учебное пособие**