

Министерство науки и высшего образования Российской Федерации
АМУРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Леонов А. К.

**АНАЛИЗ
СОЦИОЛОГИЧЕСКИХ ДАННЫХ:
методология, методы, техники в SPSS**

Благовещенск
Издательство АмГУ
2022

ББК 60.5
Л 47

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

Рецензенты: А.В. Щекотуров, *заведующий социологической лабораторией БФУ им. И. Канта, канд. социол. наук*
Н.Ю. Щека, *доцент кафедры социальной работы АмГУ, канд. социол. наук*

Леонов, А.К.

Анализ социологических данных: методология, методы, техники в SPSS : учеб. пособие / А.К. Леонов. – Благовещенск : Изд-во Амур. гос. ун-та, 2022. – 137 с.

Учебное пособие включает изложение основных принципов анализа социологических данных, в том числе с применением статистического пакета SPSS, а также комплекс практических заданий и контролирующие материалы, и предназначено для студентов, обучающихся по направлению подготовки 39.03.01 – «Социология».

СОДЕРЖАНИЕ

| | |
|---|-----|
| Введение | 4 |
| Глава 1. Методологический аспект анализа социологических данных | 6 |
| 1.1. Основные понятия анализа социологических данных | 6 |
| 1.2. Модель изучения свойства объекта и типы эмпирических данных | 12 |
| 1.3. Математическая статистика в анализе данных: теоретический и эмпирический аспекты | 19 |
| 1.4. Описание и объяснение социальных явлений и процессов как основные задачи социолога-аналитика | 24 |
| Глава 2. Анализ социологических данных: методы и техники в SPSS | 30 |
| 2.1. Основы статистики и исследование данных | 30 |
| 2.2. Статистические методы исследования данных | 40 |
| 2.3. Анализ одномерного распределения | 44 |
| 2.4. Статистические характеристики | 59 |
| 2.5. Анализ двумерного распределения | 75 |
| 2.6. Обработка множественных ответов в SPSS | 92 |
| 2.7. Основы многомерного анализа в социологии | 99 |
| 2.8. Визуализация в SPSS: графики, редактирование таблиц, модуль Tables | 112 |
| Итоговый тест | 126 |
| Рекомендуемая литература | 134 |

ВВЕДЕНИЕ

Целями данного учебного пособия является изучение основных принципов анализа социологических данных, формирование у студентов практических навыков анализа данных в статистическом пакете SPSS, полученных в конкретных социологических исследованиях. Учебное пособие может использоваться в числе базовой литературы при подготовке лекций и проведении практических занятий по дисциплинам «Анализ данных в социологии», «Основы применения прикладных статистических программ в социологических исследованиях».

Данный учебный курс представляет собой последовательное изложение методологических и методических принципов анализа социологических данных и включает в себя следующие задачи:

- формирование общих представлений об анализе социологических данных, их особенностях и типах;
- уяснение связи между измерением свойств социального объекта, типами социологических данных и их анализом;
- понимание роли математической статистики в анализе данных;
- формирование умения описывать социальные явления и процессы, выявлять закономерности и объяснять их;
- формирование умения анализировать количественные социологические данные;
- уяснение основ многомерного анализа социологических данных.

Кроме того, данный учебный курс излагает базовые функциональные возможности системы статистического анализа и управления данными – SPSS:

- 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 практических и лабораторных занятий студенты выполняли предложенные в настоящем учебном пособии практические задания анализа данных реальных эмпирических исследований, отработывая полученные на лекции знания и умения.

В учебном пособии приведены статистические таблицы и диаграммы, иллюстрирующие результаты социологических исследований, проведенных при участии автора либо сторонними организациями. Каждый параграф пособия имеет следующую структуру: изложение основных принципов анализа данных, в том числе в аспекте интерпретации данных и техник статистической обработки в SPSS (первая часть параграфа посвящена техникам работы в интерфейсе SPSS (включая алгоритмы и снимки экрана основных процедур), вторая часть – интерпретации получаемых результатов); контрольные вопросы по лекции; практические задания, представляющие собой наборы эмпирического материала для анализа. Также в книге приводятся тестовые задания (два варианта) для проверки усвоения теоретических основ анализа социологических данных и рекомендуемая литература для углубленного изучения предмета, рубрицированная на два раздела: методология и методы анализа социологических данных и статистическая обработка данных в SPSS.

Отличительными чертами данного учебного пособия являются: наличие указанных дидактических материалов, направленных на проверку теоретических знаний и формирование практических навыков анализа социологических данных; изложение не только методологии и методов анализа данных, но и техник статистической обработки в SPSS.

ГЛАВА 1. МЕТОДОЛОГИЧЕСКИЙ АСПЕКТ АНАЛИЗА СОЦИОЛОГИЧЕСКИХ ДАнных

1.1. Основные понятия анализа социологических данных

Анализ собранной информации – самый увлекательный этап исследования. Мы проверяем, насколько верны были исходные предположения, получаем ответы на заданные вопросы и выявляем новые проблемы. Методологический инструмент анализа – гипотезы, сформулированные в программе, и те, что возникают по мере их проверки и отвержения уже в процессе анализа собранных данных¹.

Роль эмпирических данных в изучении социальных явлений огромна. Достаточно глубокое изучение интересующих социолога закономерностей невозможно без опоры на анализ конкретных фактов, в которых эти закономерности, собственно говоря, и проявляются. В литературе редко встречается словосочетание «методология анализа данных в социологии», а если и встречается, то понимается слишком узко. Поэтому мы обязаны пояснить специфику этой области знания. Более того, в этом разделе речь идет о структурных элементах как бы «языка» анализа в социологическом исследовании, что требует введения хотя бы основных понятий – элементов такого языка².

Методология – это система знаний о способах достижения нового знания. В широком смысле методология данной науки включает в себя теорию, общенаучные и специальные методы исследования ее предмета, в узком смысле – систему методов получения информации, ее анализа, интерпретации и объяснения.

Будем рассматривать понятие «методология» в узком смысле его толкования. Какую же часть так понимаемой методологии занимает

¹ Ядов В.А. Стратегия социологического исследования. Описание, объяснение, понимание социальной реальности / В.А. Ядов. – 3-е изд., испр. – Москва: Омега-Л, 2007.

² Толстова Ю.Н. Анализ социологических данных. – М. : Научный мир, 2000.

методология анализа данных? Исходим из того, что нас интересует процедура получения знания в эмпирической социологии,

Разумеется, эта процедура включает в себя отдельные приемы, способы, методы работы с эмпирическими данными или с социологической информацией.

Мы упомянули два понятия: эмпирическая социология и социологическая информация. Под *эмпирической социологией* будем понимать совокупность исследований, основанных на методических и технических приемах и методах сбора, обработки и анализа первичной социологической информации.

Рассмотрим второе из обозначенных выше понятий. *Социологической информацией* называются любые эмпирические данные, которые содержат информацию о социальной реальности: социальных явлениях, социальных процессах, социальных общностях, социальных институтах, социальных системах, социальных группах и других социальных феноменах. Первичная социологическая информация исходит из вещественных источников: респондентов, наблюдаемых, документов, статистики, испытуемых.

Мы намеренно перечисляем некоторый понятийный контекст (вам знакомый), в рамках которого социолог сталкивается с эмпирией. Во-первых, социолог имеет дело с «готовой» информацией (государственная статистика; результаты выборов; текстовая информация различного вида; данные, полученные другими исследователями). Во-вторых, социолог сам порождает (осуществляет сбор) данные в процессе анкетирования, интервьюирования, наблюдения.

Данные могут представлять перед исследователем в виде:

- совокупности чисел, характеризующих те или иные объекты (в качестве таких совокупностей могут выступать, например, производственные характеристики предприятий, возраст респондентов, оценки выпускниками школ престижности некоторых профессий и т.д.);

- множества индикаторов определенных отношений между рассматриваемыми объектами (к примеру, при изучении производственных

бригад такими индикаторами могут служить указания каждого члена бригады на то, нравится ли ему работать вместе с любым другим членом той же бригады, такие данные часто используются при изучении малых групп);

- результатов попарных сравнений респондентами каких-либо объектов;
- совокупности определенных высказываний (например, ответов респондентов на вопрос об их профессии, о том, что им нравится в политике правительства; письма читателей газеты в редакцию; фрагменты из журнальных статей и т.д.);
- текстов документов;
- так или иначе зафиксированных результатов наблюдения за невербальным поведением каких-либо людей и т.п.

Наиболее часто в социологических исследованиях данные представляют собой совокупность значений каких-либо признаков (характеристик, переменных, величин; будем считать эти термины синонимами), измеренных для каждого из изучаемых объектов.

Признак – это некоторое общее для всех объектов качество, конкретные проявления которого (значения признака; их называют также альтернативами, градациями), вообще говоря, могут меняться от объекта к объекту. Примеры признаков – пол, возраст респондентов, их удовлетворенность своим трудом и т.д.

Теперь можно дать и некое «рабочее» определение *методологии анализа данных*. Исходим из того, что это научная рефлексия по поводу существования различных типов информации, специфики приемов ее измерения, логики и методов ее анализа.

В процессе освоения области, обозначенной как методология анализа данных, возникают определенные трудности. Они связаны в основном с существованием нескольких проблем, в определенной мере это существование постулатов, важных в эмпирической социологии. Мы рассмотрим их как контекст для того, чтобы познакомиться с основными понятиями методологии анализа данных. Речь пойдет, конечно же, об основных, но самых важных

понятия – некоторых элемента «языка» анализа в социологическом исследовании.

Первая проблема. Многозначный характер использования понятий в социологии. Например, в эмпирической социологии «анализ» включает в себя как процедуру разложения, так и объединения: цель – задачи – гипотезы – аналитическая операционализация – индикаторы; а затем (после проведения сбора данных) происходит обобщение – поиск эмпирических закономерностей.

Во-вторых, понятие «анализ», которое мы рассматриваем, на различных этапах социологического исследования трактуется по-разному. В целом же любое социологическое исследование есть анализ фрагмента социальной реальности.

Обратимся к понятию «метод». Пожалуй, это одно из самых важных понятий методологии социологических исследований. Пока под методами анализа данных (информации) будем понимать совокупность приемов, способов, участвующих в систематизации, обобщении, «переваривании» эмпирии.

Вторая проблема. Существование «статистического» (количественного) и «гуманитарного» (качественного) подходов при исследовании социальной реальности. Первый подход. Массовые явления имеют статистический характер, т. е. если изучить достаточно большое количество проявлений изучаемого социального явления, то само явление будет познано. В рамках этого подхода, который мы условно называем статистическим подходом, статистической традицией, индивид – представитель некоторой общности, Носитель информации о социальном феномене. Индивиды взаимозаменяемы, и их индивидуальные особенности как таковые не представляют для социолога особого интереса. В рамках статистической традиции основной метод сбора информации – метод опроса посредством формализованного (жестко структурированного) вопросника. Для формирования выборочных совокупностей используются в основном данные государственной статистики. Опираясь на эту традицию, социолог решает такие проблемы, как

доказательное описание социальной реальности, проверка описательных и объяснительных гипотез, поиск и объяснение эмпирических закономерностей модального (распространенного) характера. Применение математического анализа в социологии обычно и не всегда оправданно связывают только с этим первым подходом. Второй подход, сторонники этого подхода считают, что с помощью жестко формализованных методов опроса нельзя заглянуть в глубину социального. Индивид неповторим, и он не проявление социального, а как бы само социальное явление. Отсюда так называемые качественные подходы (качественные методы, качественная методология). Социолог ищет ответы на вопросы: Что это? Зачем это? Откуда это! В каких формах существует это? Каковы причины этого? Функции понимания и объяснения выходят на первый план, но относятся к индивидуальному уровню, а не к эмпирическим закономерностям на общностях. Самым трудным в рамках этого подхода является переход от глубины индивидуального, от исследования случая (кейс-стади) как типики к общим закономерностям.

В этом случае методами сбора информации являются различные виды неформализованного интервью, наблюдение. Процесс получения знания – искусство социолога. Если в рамках первого подхода социологи увлекаются описанием эмпирических закономерностей модального характера, то в рамках второго – описанием эмпирических закономерностей антимодального (мало распространенного) характера.

Самый важный вывод, вытекающий из рассмотрения двух подходов (двух традиций, двух парадигм): выбор подхода или логика сочетания подходов обусловлены целью исследования, исследовательскими задачами, исследовательской ситуацией.

Третья проблема. Сложность процесса получения знания в эмпирической социологии, специфика нашей науки еще и в том, что перед ней, в отличие от многих других наук, четыре цели: понимание, описание, объяснение, предсказание. Социальная реальность настолько сложна, что, видимо, процесс познания не может опираться на одну-единственную социологическую теорию.

Поэтому одна и та же исследовательская задача может быть решена с использованием множества подходов, приемов, способов, методов, техник.

Четвертая проблема. Необходимость постоянного соотнесения научного и обыденного, понимание понятий, терминов. Поэтому будем вводить научную трактовку многих понятий, употребляемых в обыденной речи. При этом эмпирическая интерпретация этих понятий носит многозначный характер. Мы упомянули еще раз понятие «эмпирическая интерпретация». Это еще одно из важных понятий методологии анализа. Эмпирическая интерпретация – это процедура перехода от теоретических понятий к эмпирическим индикаторам или построение модели изучения чего-либо, доведенной до совокупности эмпирических индикаторов. Эмпирическая интерпретация – это переход от теоретических конструктов (понятий, гипотез, задач) к эмпирическим конструктам.

Пятая проблема. В рамках эмпирической социологии вы можете столкнуться с так называемыми ложными проблемами. К таковым относятся проблемы, которые социолог не должен ставить перед собой. Это проблемы-искушения, и к ним относятся:

- поиск одного решающего фактора для описания и объяснения некоторого социального феномена;
 - представление о целом как о сумме частей, т. е. познание социального по частям не дает знания о целом;
 - противопоставление качества и количества;
 - поиск глобальных законов развития общества, а не закономерностей;
- поиск ответа на вопрос: куда идет страна, мир?

Все выделенные понятия при рассмотрении описанных выше проблем – элементы «языка» анализа эмпирических данных. Естественно, эта совокупность необходима социологу, но недостаточна. Другие элементы языка анализа будут введены в процессе дальнейшего изложения³.

³ Татарова Г.Г. Методология анализа данных в социологии (введение) / Учебник для вузов. – М. : NOTA BENE, 1999.

Использованная литература

Татарова Г.Г. Методология анализа данных в социологии (введение) / Учебник для вузов. – М. : NOTA BENE, 1999. – 224 с.

Толстова Ю.Н. Анализ социологических данных. – М. : Научный мир, 2000. – 352 с.

Ядов В.А. Стратегия социологического исследования. Описание, объяснение, понимание социальной реальности / В.А. Ядов. – 3-е изд., испр. – Москва: Омега-Л, 2007. – 567 с.

Контрольные вопросы

1. Что такое «методология анализа социологических данных»?
2. С какими данными имеет дело социолог как аналитик?
3. Каковы основные проблемы анализа социологических данных?

Практические задания

Задание 1. Провести сравнительный анализ количественной и качественной парадигмы в социологическом исследовании по учебнику: Готлиб, А.С. Введение в социологическое исследование: Качественный и количественный подходы. Методология. Исследовательские практики : учеб. пособие : рек. УМО / А.С. Готлиб. – М. : Флинта ; М. : Моск. психол.-соц. ин-т, 2005. – 383 с.

Задание 2. Выбрать некую социальную проблему. В ее границах обозначить социальные феномены, которые могут быть изучены в рамках: только статистического подхода, только гуманитарного подхода, при сочетании того и другого в разной последовательности.

1.2. Модель изучения свойства объекта и типы эмпирических данных

Модель изучения свойств объекта.

В социологии существует вечная проблема взаимодействия двух уровней знания: теоретического и эмпирического. Предметом изучения методологии анализа и является это взаимодействие. Любое исследование как-то

планируется независимо от того, в рамках какого подхода оно проводится. Для этого введем логическую цепочку вопросов.

Что изучать? Зачем и для достижения каких целей изучать? Где и при каких условиях изучать? С помощью каких средств изучать?

На каждом из этапов социологического исследования упомянутая выше логическая цепочка вопросов играет определенную функциональную роль, уточняя горизонтальные связи в исследовании. Например, познавательная возможность даже отдельно взятого эмпирического индикатора вопросника может быть проанализирована с помощью обозначенной логической цепочки. Тогда мы задаем себе следующие вопросы: Что изучается посредством этого вопроса? Зачем и для достижения каких целей задается вопрос? Где и при каких условиях работает вопрос? В какой форме его задать респонденту?

В целом же поиск ответов на вопросы, входящие в логическую цепочку, позволяет структурировать исследовательскую ситуацию, как бы выработать методику решения исследовательских задач. При этом представляется возможность и решения на альтернативной основе. Что означает альтернативность? Готовясь к проведению социологического исследования, социолог должен «видеть» несколько вариантов его проведения, несколько вариантов выбора методов сбора, несколько вариантов анализа эмпирии, несколько вариантов интерпретации эмпирических закономерностей. Альтернативность и порождает различные логические схемы анализа. В свою очередь эти схемы подлежат эмпирической интерпретации так же, как гипотезы, задачи, понятия исследования. Затем возможна и математическая формализация.

На каком бы «языке» социолог ни изучал социальную реальность, какой бы срез этой реальности он ни исследовал, от биографии отдельного индивида до социетального уровня, он ее изучает через призму проявления свойств либо социального явления, либо социального объекта, либо социального процесса, либо отдельно взятых индивидов.

Свойство – сугубо теоретическое и очень широкое понятие. Мы остановимся только на свойствах особого, достаточно простого вида. Какие же это свойства? Только такие, относительно которых предположительно возможны высказывания типа: у объекта А заданного свойства больше, чем у объекта В, или: объекты А и В неразличимы с точки зрения этого свойства. Здесь под объектами понимаются эмпирические объекты в соответствии с объектами анализа. Можно считать, что объект анализа и единица исследования – это одно и то же. Например, в упомянутой работе выделяются такие единицы исследования, как общности, институции, тексты, события.

В нашем случае речь идет о сравнимых эмпирических объектах, в роли которых могут выступать респонденты, эксперты, семьи, студенческие группы, регионы, районы, страны, социальные общности и т. д. Эмпирические объекты – объекты сравнения. Эмпирические объекты – объекты анализа. Это необязательно непосредственные носители информации. Например, эмпирическими объектами являются семьи, бригады, группы, а носителем (источником) информации для их изучения – отдельные респонденты. Далее будем пользоваться термином «объект», имея в виду эмпирический объект. Каждый такой объект обладает различными свойствами. Итак, рассматриваем те и только те свойства, относительно которых можно говорить в терминах «равно» – «не равно», «есть» – «нет», «больше» – «меньше».

Будем считать, что в результате построения модели изучения свойства доходим до эмпирических индикаторов.

Примеры: материальная обеспеченность, политическая активность, престижность городских районов.

Краткие выводы. Модель изучения свойства включает в свою очередь свойства более низкого уровня общности. Модели бывают простые и сложные. Эмпирические индикаторы в них иногда легко придумать, а иногда очень трудно. Конкретный вид модели обусловлен целями, ради достижения которых эта модель строится. Цели могут быть описательные и объяснительные. Одному и тому же свойству соответствует разная совокупность эмпирических

индикаторов. Форма существования индикатора обусловлена методами сбора информации. Определенная модель требует обращения к разным типам информации⁴.

Типы эмпирических данных.

Выделим пять типов информации, с которыми сталкивается социолог на практике, и, в определенной мере условно, обозначим их как:

1. Данные типа «государственная статистика».
2. Данные, полученные с помощью анкет «простой» структуры.
3. Данные, полученные с помощью анкет «сложной» структуры.
4. Данные об использовании бюджета времени.
5. Текстовые данные.

Тип 1: Данные типа «государственная статистика».

Этот тип данных происходит из различных видов документальных источников. Общей характеристикой рассматриваемого типа является то, что это – совокупность данных объективного характера, описывающая некий социум (государство, регион, область, предприятие, социальную общность и т.д.) в пространстве и во времени. В свою очередь данный тип состоит из нескольких специфических частей, обусловленных различиями в источниках информации и в объектах анализа. Рассмотрим их:

1. Информация, которая собирается регулярно (ежегодно) о различных сторонах жизнедеятельности общества и хранится в архивах различного профиля. В них содержатся значения социальных показателей по всем странам мира. Эти данные содержат в себе социальные показатели в явном и в неявном виде. В явном виде – когда показатель рассчитан как некий индекс на основе эмпирических индикаторов. Например, уровень скрытой безработицы. В неявном виде – когда есть основа для формирования социального показателя.

Объектами анализа для социолога являются страны, области жизнедеятельности, социальные общности, социальные показатели.

⁴ Татарова Г.Г. Методология анализа данных в социологии (введение) / Учебник для вузов. – М.: NOTA BENE, 1999.

2. Данные переписи населения. Источником информации в данном случае является индивид, а объектами анализа – социальные общности, регионы, социальные показатели.

3. Данные специальных исследований об условия жизни. В данном случае источником информации является семья, а объектами анализа – семья, статьи бюджета.

4. Результаты выборов, референдумов. Это особый вид статистики. Источник информации – территориальное образование (регион, область, район, округ, участок). Объекты анализа – кандидаты, территориальные образования, электоральное поведение.

5. Специфические документальные источники в виде архивов предприятий и т. д.

Таким образом, рассматриваемый нами тип социологической информации содержит характеристики социальных процессов на разном уровне их исследования: на социетальном, институциональном, групповом и личностном, индивидуальном. Это во-первых. Во-вторых, в них (в типе данных) отражена динамика социальных процессов. В-третьих, это объективные характеристики жизнедеятельности людей, не зависящие от их субъективны оценок.

Социолог обращается к такого рода данным и работает с ними в нескольких ситуация, важных с позиции методологии анализа.

1. Тогда, когда социолог изучает непосредственно те социальные показатели, которые содержатся в этом типе информации.

2. Вторая ситуация, когда социолог обращается к государственной статистике заключается в следующем. Этот тип информации может не содержать необходимых для социолога социальных показателей, но в этой информации есть характеристики, играющих роль эмпирических индикаторов для формирования социальных показателей.

3. Третья ситуация возникает при обращении к государственной статистике с целью формирования выборочных совокупностей.

Рассмотренный выше тип информации, с точки зрения социолога-пользователя, обладает двумя недостатками: возможными неполнотой и недостоверностью.

Тип 2: Данные, полученные с помощью анкет «простой» структуры.

Получаются в результате изучения общественного мнения (маркетинговые исследования, мониторинга разного вида). Эти исследования, как правило, носят описательный характер, и в них используются прямые, простые, однозначно интерпретируемые вопросы и ответы на них. Объектами анализа в рамках этого типа информации являются социальные общности, общественное мнение по различным проблемам.

Тип 3: Данные полученные посредством анкет «сложной» структуры.

К этому типу данных относим данные, полученные в исследованиях аналитического характера. Инструментарий сбора эмпирических данных в этих исследованиях, как правило, состоит из нескольких частей. Для сбора данных используются специфические, достаточно сложные приемы, методы, техники. Анкетирование может сопровождаться тестированием. Для изучения проблем конфликтности, лидерства в малых группах используется так называемая социометрическая техника.

Тип 4: Данные об использовании бюджета времени.

Информация о времяпрепровождении или использовании бюджета времени является данными о реальном поведении людей. Объектами анализа в исследованиях времяпрепровождения являются отдельно взятый респондент и его времяпрепровождение за заданный промежуток времени, социальная общность и ее времяпрепровождение, отдельно взятое занятие, вид деятельности (совокупность занятий, например, труд в домашнем хозяйстве, уход за детьми, покупки, передвижения и т. д.), группы видов деятельности (рабочее время, вне рабочее время, свободное время и т. д.), структура времяпрепровождения.

Тип 5: Текстовые данные.

Можно выделить несколько ситуаций, когда социологу приходится работать с текстовыми данными. Эти ситуации следует рассматривать как три источника данного типа социологической информации. Во-первых, социолог сталкивается с такого рода информацией при работе с открытыми вопросами анкеты. Во-вторых, при работе с текстами различного рода неформализованных интервью, т. е. при работе с мягкими/гибкими методами сбора информации. Это результаты нарративного интервью, в основном использующиеся в биографическом к подходе изучению социальной реальности. Это результаты лейтмотивного интервью: индивидуального и группового (фокус-группа). В-третьих, при работе с материалами средств массовой информации. В четвертых, при сборе эмпирической информации посредством так называемых проективных методов. Например, для изучения социальных норм используется метод неоконченных предложений, а для изучения идентичности методика двадцати самоопределений⁵.

Использованная литература

Татарова Г.Г. Методология анализа данных в социологии (введение) / Учебник для вузов. – М.: NOTA BENE, 1999. – 224 с.

Контрольные вопросы

1. Какие элементы включает в себя логическая цепочка анализа данных?
2. Какими свойствами обладает модель изучения свойств объекта?
3. Каковы особенности каждого типа социологических данных?

Практические задания

Задание 1. Выделить пару свойств в рамках некоторой социальной проблемы (той, которая была обозначена при выполнении предыдущего задания). Составить модель (нарисовать схему) изучения каждого свойства с

⁵ Татарова Г.Г. Методология анализа данных в социологии (введение) / Учебник для вузов. – М.: NOTA BENE, 1999.

тем, чтобы «спуститься» до того уровня, на котором становятся ясными источники эмпирически данных.

Задание 2. По каждому типу информации (тип 1, тип 2, тип 3, тип 4, тип 5) ответьте на следующие вопросы: Участвует или не участвует социолог в сборе информации? Откуда она берется? Кто или что является источником информации? Какие методы сбора этого типа информации существуют? Что анализирует социолог посредством этой информации?

Задание 3. В рамках некоторой социальной проблемы (той, которая была обозначена при выполнении предыдущего задания) определить виды информации, необходимые для изучения этой проблемы.

Задание 4. Ознакомившись с содержанием статистического сборника (Регионы России. Социально-экономические показатели. 2011 : Стат. сб. / Росстат. – М. : 2011. – 990 с.), выбрать те статистические характеристики, которые могут быть проанализированы при исследовании выбранной в предыдущем задании проблемы.

1.3. Математическая статистика в анализе данных: теоретический и эмпирический аспекты

О роли математизации научного знания в литературе говорится довольно много. Ниже мы коснемся лишь некоторых аспектов соответствующей проблемы, играющих, на наш взгляд, серьезную роль в организации эмпирического социологического исследования.

Математические методы как средство познания социальных явлений.

Возможность применения математики возникает тогда, когда исследователь абстрагируется от многих конкретных черт изучаемого объекта и предполагает адекватной сути решаемой задачи определенную формализацию рассматриваемого явления.

Переходя к более подробному логическому анализу рассматриваемого процесса, можно сказать следующее. Применение математики опирается на то,

что мы считаем возможным (1) выделить некоторый фрагмент реальности; (2) построить (посредством измерения) его математическую модель (т.е. получить исходные данные); (3) изучить эту модель традиционными для математики способами (в нашем случае – применить тот или иной алгоритм анализа данных) и прийти к некоторым выводам о ее «устройстве» (в результате анализа данных получить какой-то математический результат: вычислить точное значение коэффициента корреляции, найти параметры уравнения регрессии и т.д.); (4) проинтерпретировать эти выводы на содержательном языке (т.е., как говорят обычно, проинтерпретировать результаты анализа данных) и получить таким образом новое знание о реальности. Первые два этапа обычно относят к области измерения (шкалирования), последние два – к области собственно анализа данных.

Использование математических методов в процессе проведения научного исследования позволяет достичь следующих целей. Во-первых, применение математики побуждает исследователя четко сформулировать свои представления об изучаемом объекте. Во-вторых, использование математических методов позволяет четко выдержать обсужденное выше абстрагирование от неисчислимого количества реальных свойств изучаемых объектов, не дает уйти в сторону от принятого исследователем понимания изучаемого явления. В-третьих, с помощью математики можно получить содержательные выводы, не лежащие «на поверхности», за счет расширения круга используемых логических умозаключений. В-четвертых, не лежащие на поверхности выводы могут быть получены за счет создания возможности анализа огромных массивов информации (с которыми обычно и имеет дело социолог), учета огромного количества факторов (определяющих практически любое общественное явление).

В заключение отметим, что без применения математического аппарата трудно обойтись при решении практически любой социологической задачи. А поскольку главной целью анализа данных является выявление статистических закономерностей, то из всех ветвей математики для социолога естественным

образом на первое место выходит та ветвь, которая направлена именно на поиск таких закономерностей – математическая статистика (и, конечно, лежащая в ее основе теория вероятностей).

Актуальность для социологии задач, решаемых математической статистикой.

Итак, главной задачей анализа данных является сжатие собранной эмпирической информации, направленное на «вычленение» скрытых в ней статистических (т.е. имеющих место «в среднем») закономерностей. Примерно так же формулируется и основная задача математической статистики. Вряд ли можно подвергнуть сомнению то, что при изучении статистических закономерностей социолога практически всегда интересует задача перенесения полученных им результатов с той совокупности объектов, которая непосредственно была обследована (с выборки) на более широкую совокупность (генеральную). Это делает использование математической статистики еще более привлекательным для социолога: ведь с помощью соответствующих подходов можно осуществлять анализ выборочных данных именно с намерением обобщения получаемых результатов на соответствующую генеральную совокупность⁶.

Связь понятий математической статистики и понятий социологического исследования лучше всего проиллюстрировать следующим образом (таблица 1):

Таблица 1

| Понятие математической статистики | Понятие социологического исследования |
|--|--|
| случайная величина | признак |
| случайное событие | отбор респондентов |
| вероятность | проценты |
| распределение вероятностей | одномерное и двумерное распределение |

Непосредственное применение математической статистики в социологии, зачастую, бывает весьма проблематично. Условия, предполагаемые строгими теоремами математической статистики, отнюдь не всегда выполняются на

⁶ Толстова Ю.Н. Анализ социологических данных. – М. : Научный мир, 2000.

практике. И тогда вместо строгой математической статистики на сцену выступает не совсем строгое ее «приближение» – анализ данных⁷.

Математическая статистика и анализ данных.

Ниже, параллельно анализу рассматриваемых проблем, будем четко выделять причины, приведшие к необходимости введения наряду с термином «математическая статистика» термина «анализ данных».

1. На практике нередко нарушаются условия вероятностного порождения данных: не всегда ясно повториться ли выявленная закономерность при схожих условиях.

2. Отнюдь не всегда бывает ясно, какова изучаемая генеральная совокупность.

3. Для многих методов отсутствуют разработанные способы перенесения результатов их применения с выборки на генеральную совокупность.

4. Перенос результатов с выборки на генеральную совокупность может быть затруднен из-за осуществления «ремонта» выборки (например, ее перевзвешивания), что нередко делает социолог.

5. Для некоторых методов, показавших свою эффективность при решении практических задач, отсутствуют строгие доказательства корректности их использования.

6. Необходимость анализа «чисел», полученных по шкалам низких типов, к которым неприменимы большинство методов математической статистики.

7. Сложность изучаемых с помощью анализа данных явлений – необходимость постоянного вмешательства исследователя в процесс анализа.

Чтобы не оставлять за математикой лишь роль средства придания некоторого наукообразия работе социолога, необходимо учитывать, что корректность использования математического аппарата на любом из этапов исследования тесно связана с принципами реализации других этапов, в том числе и таких, в которых не задействованы никакие математические методы. Это требование конкретизируется в виде целого ряда положений.

⁷ Толстова Ю.Н. Анализ социологических данных. – М. : Научный мир, 2000.

Первое – связь измерения и анализа данных.

Второе – связь интерпретации результатов и концептуальных установок исследователя, стоящих перед ним целей⁸.

Использованная литература

Толстова Ю.Н. Анализ социологических данных. – М. : Научный мир, 2000. – 352 с.

Толстова Ю.Н. Измерение в социологии: Курс лекций. – М. : ИНФРА-М, 1998. – 224 с.

Контрольные вопросы

1. Чего позволяет достичь применение математических методов в анализе социологических данных?

2. Какова связь понятий математической статистики и социологического исследования?

3. Почему наряду с понятием «математическая статистика» используется понятие «анализ данных»?

Практические задания

Задание 1. На сайтах служб изучения общественного мнения (ВЦИОМ, ФОМ, Левада-Центр), а также в социологических журналах, найти примеры эмпирических исследований. В этих примерах найти элементы математической обработки данных, обозначенные в материалах параграфа.

Задание 2. Воспользовавшись статистическим сборником (Индикаторы науки: 2016 : статистический сборник / Н.В. Городникова, Л.М. Гохберг, К.А. Дитковский и др.; Нац. исслед. ун-т «Высшая школа экономики». – М. : НИУ ВШЭ, 2016. – 304 с.), проанализируйте следующие данные:

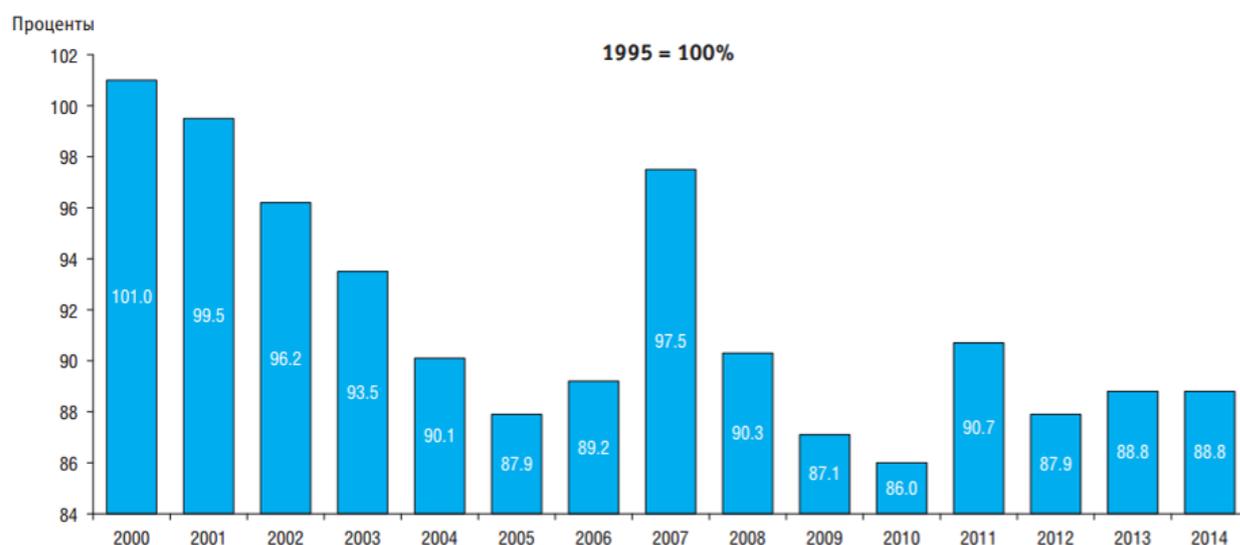
⁸ Толстова Ю.Н. Измерение в социологии: Курс лекций. – М. : ИНФРА-М, 1998.

1.1. ОРГАНИЗАЦИИ, ВЫПОЛНЯВШИЕ ИССЛЕДОВАНИЯ И РАЗРАБОТКИ

| | 1995 | 2000 | 2004 | 2005* | 2006* | 2007* | 2008* | 2009* | 2010* | 2011* | 2012* | 2013* | 2014* |
|--|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| Всего | 4059 | 4099 | 3656 | 3566 | 3622 | 3957 | 3666 | 3536 | 3492 | 3682 | 3566 | 3605 | 3604 |
| Научно-исследовательские организации | 2284 | 2686 | 2464 | 2115 | 2049 | 2036 | 1926 | 1878 | 1840 | 1782 | 1744 | 1719 | 1689 |
| Конструкторские организации | 548 | 318 | 194 | 489 | 482 | 497 | 418 | 377 | 362 | 364 | 338 | 331 | 317 |
| Проектные и проектно-исследовательские организации | 207 | 85 | 63 | 61 | 58 | 49 | 42 | 36 | 36 | 38 | 33 | 33 | 32 |
| Опытные заводы | 23 | 33 | 31 | 30 | 49 | 60 | 58 | 57 | 47 | 49 | 60 | 53 | 53 |
| Образовательные организации высшего образования | 395 | 390 | 402 | 406 | 417 | 500 | 503 | 506 | 517 | 581 | 562 | 673 | 700 |
| Организации промышленного производства | 325 | 284 | 244 | 231 | 255 | 265 | 239 | 228 | 238 | 280 | 274 | 266 | 275 |
| Прочие организации | 277 | 303 | 258 | 234 | 312 | 550 | 480 | 454 | 452 | 588 | 555 | 530 | 538 |

* С 2005 г. в связи с отменой Общероссийского классификатора отраслей народного хозяйства (ОКОНХ) изменена классификация типов организаций, выполнявших исследования и разработки.

1.2. ДИНАМИКА ЧИСЛА ОРГАНИЗАЦИЙ, ВЫПОЛНЯВШИХ ИССЛЕДОВАНИЯ И РАЗРАБОТКИ



1.4. Описание и объяснение социальных явлений и процессов как основные задачи социолога-аналитика

Еще одним ключевым понятием анализа социологических данных является интерпретация – это наше видение того, что за этими данными стоит, наше понимание смысла чисел, полученных в результате измерения⁹. Результаты исследования, полученные путем компьютерных вычислений, сведенные в таблицы или изображенные в виде графиков и рисунков, представляют собой числовые величины, пригодные для выводов о характере и признаках изучаемого явления. При этом главное условие, позволяющее

⁹ Толстова Ю.Н. Измерение в социологии: Курс лекций. – М. : ИНФРА-М, 1998.

должным образом использовать социологические данные в социальной практике, – их всестороннее и правильное объяснение – интерпретация.

Единых правил интерпретации результатов социологического исследования не существует. В каждом конкретном случае она связывается с объективными и субъективными факторами, действующими в обследованной группе населения или регионе, опирается не только на полученные данные, но и на статистический материал, результаты других опросов¹⁰.

Рассмотрим более подробно, что стоит за интерпретацией данных.

Ключевые задачи социолога в анализе социологических данных.

Описание и объяснение – взаимосвязанные функции социального познания, осуществляемые на различных его уровнях (эмпирическом и теоретическом). *Описание* представляет собой фиксацию результатов эмпирического исследования в соответствующих терминах. В процессе описания данные наблюдения, эксперимента, измерения подвергаются определенному уплотнению, сжатию, упорядочению, в результате чего они подготавливаются к последующему описанию. Описание... ограничивается перечислением некоторых свойств объектов, группировкой по их внешним признакам, установлением наблюдаемых зависимостей.

Объяснение – это процесс раскрытия сущности объектов, охватывающий их структурные, функциональные, генетические, причинно-следственные связи и отношения, тенденции развития, которые отображаются в гипотезах, законах, теориях. В методологической литературе принято различать дедуктивное и индуктивное объяснение; в зависимости от объясняющих положений – объяснение факта, закона или теории; в зависимости от характера объясняемых положений – причинное, функциональное, структурное и телеологическое виды объяснений. Отмечается, что объяснение социологических фактов происходит либо через собственный (социологический) закон, либо через модельное

¹⁰ Горшков М.К. Прикладная социология: методология и методы: Учебное пособие / М.К. Горшков, Ф.Э. Шереги. – М.: Альфа-М: ИНФРА-М, 2009.

построение – и в этом объяснения различаются уже по характеру их механизма¹¹.

«Под *типологией* обычно понимается метод научного познания, в основе которого лежит обладающее определенными свойствами разбиение изучаемой совокупности объектов на группы»¹².

Принято различать понятия эмпирической и теоретической типологии. *Эмпирическая типология* – «поиск устойчивых сочетаний свойств социальных объектов (или явлений), рассматриваемых в соответствии с описательными гипотезами в нескольких измерениях одновременно». *Теоретическая типология* – «обобщение признаков социальных явлений на основе идеальной теоретической модели и по теоретически обоснованным критериям»¹³.

Считается, что если типология тесно связана с содержательным характером разбиения совокупности на группы, с определенным уровнем познания, то *классификация* подобными свойствами не обладает и в нее не вкладывают гносеологического смысла. К методам классификации относят «любые формальные методы, с помощью которых можно определить, принадлежит или не принадлежит объект к тому или иному классу»¹⁴.

Прогнозирование в социологии – это способ научного предвидения с той или иной степенью вероятности итога, направленности, или характера протекания социальных процессов в течение некоторого промежутка времени. По методам в прикладной социологии различают три вида прогноза: 1) аналогия; 2) экспертиза; 3) статистический прогноз¹⁵.

Основные принципы и процедура описания и объяснения в социологии.

Принципы объяснения результатов исследования:

¹¹ Казанцев В.И. Методология и методика социологических исследований: Учебно-методическое пособие / В.И. Казанцев, М.Г. Светульников. – Ульяновск: Средневолжский научный центр, 2003. – 160 с.

¹² Ядов В.А. Социологическое исследование: методология, программа, методы – М.: Наука, – 1972.

¹³ Там же.

¹⁴ Там же.

¹⁵ Горшков М.К. Прикладная социология: методология и методы: Учебное пособие / М.К. Горшков, Ф.Э. Шереги. – М.: Альфа-М: ИНФРА-М, 2009.

- predeterminedness by research tasks;
- «work» within the conceptual and operational model of research, connection with the research problem;
- interpretation of research indicators in their interconnection;
- attraction of «external» qualitative and quantitative information, related to the research problem and others.

Процедура объяснения результатов исследования:

1. Transition from indicators to indicators.
2. Formation of indices, scale ratings.
3. Typologization.
4. Ranking by significance.
5. Identification and ranking of conditioning factors (causes).
6. Use of hypotheses in explaining results.
7. Evaluation of research results.
8. Comparison of data as a form of evaluation: internal and external comparison.
9. Identification of interconnections of indicator states¹⁶.

Связь социологической теории с анализом эмпирических данных: sociological laws and their role in explaining the results obtained, consists in the fact that to explain the revealed regularity, it is necessary to subordinate it to a certain social law. In its turn, the sociologist draws these laws from general sociological theory and theories of the middle level. We note that when analyzing data, it is completely impossible to exclude the influence of the sociologist on the process of explaining the results of research practically¹⁷.

Практические советы:

1. Approaching the analysis of data, we will strictly adhere to program hypotheses, avoiding two extremes: hasty conclusions

¹⁶ Васильева Е.Г. Методика и техника конкретного социологического исследования: Учебно-методическое пособие. – Волгоград: Издательство ВолГУ, 2000.

¹⁷ Там же.

относительно их подтверждения, если факты «укладываются» в гипотезу, но вместе с тем соблазна увлечься самим процессом анализа как таковым, что нередко уводит в сторону от целевой ориентации исследования.

2. Первоначальные группировки и классификации разумнее всего производить исходя из элементарных описательных гипотез, а последующие – предварять уточняющими и интерпретационными предположениями, продвигаясь к объяснительным.

Чем дальше мы углубляемся в анализ данных, тем большее значение приобретают объяснительные гипотезы, непосредственно связанные с программными задачами исследования.

3. Если гипотезы не тривиальны, особое внимание следует уделять заключениям, которые с ними не согласуются. В результате перепроверок, использования различных приемов анализа мы вводим ограничения, уточнения исходных гипотез и либо обнаруживаем их справедливость, либо уверенное опровержение, что побуждает выдвинуть новые гипотезы и осуществить их последовательную проверку.

4. никоим образом не следует смешивать уточнение и интерпретацию данных с их объяснением. Последнее является главной задачей анализа, позволяет установить причинные зависимости, истолковать найденные связи в понятиях более общих тенденций и закономерностей, дает основание для прогноза и, следовательно, для перехода к обоснованию практических решений социальных проблем¹⁸.

Использованная литература

Васильева Е.Г. Методика и техника конкретного социологического исследования: Учебно-методическое пособие. – Волгоград: Издательство ВолГУ, 2000. – 64 с.

¹⁸ Ядов В.А. Стратегия социологического исследования. Описание, объяснение, понимание социальной реальности / В.А. Ядов. – 3-е изд., испр. – Москва: Омега-Л, 2007.

Казанцев В.И. Методология и методика социологических исследований: Учебно-методическое пособие / В.И. Казанцев, М.Г. Светуныков. – Ульяновск: Средневолжский научный центр, 2003. – 160 с.

Толстова Ю.Н. Измерение в социологии: Курс лекций. – М. : ИНФРА-М, 1998. – 224 с.

Горшков М.К. Прикладная социология: методология и методы: Учебное пособие / М.К. Горшков, Ф.Э. Шереги. – М.: Альфа-М: ИНФРА-М, 2009. – 416 с.

Ядов В.А. Социологическое исследование: методология, программа, методы – М.: Наука, – 1972.

Ядов В.А. Стратегия социологического исследования. Описание, объяснение, понимание социальной реальности / В.А. Ядов. – 3-е изд., испр. – Москва: Омега-Л, 2007. – 567 с.

Контрольные вопросы

1. Что такое интерпретация социологических данных?
2. Каково место описания и объяснения в интерпретации данных?
3. На чем основывается объяснение социологических данных?

Практические задания

Задание 1. На сайтах ФОМ, ВЦИОМ, Левада-Центра а также в социологических журналах, найти результаты эмпирических социологических исследований. В этих материалах:

- найти примеры описания и объяснения, типологии и классификации, прогнозирования;
- определить, на какие принципы описания и объяснения явно опирался исследователь;
- выявить используемые исследователем процедуры описания и объяснения;
- найти примеры описания без объяснения и попытаться самостоятельно дать объяснение выявленным закономерностям.

ГЛАВА 2. АНАЛИЗ СОЦИОЛОГИЧЕСКИХ ДАННЫХ: МЕТОДЫ И ТЕХНИКИ В SPSS

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. 2):

Таблица 2

| 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. 3):

Таблица 3

| <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 шкал¹⁹.

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. 1).

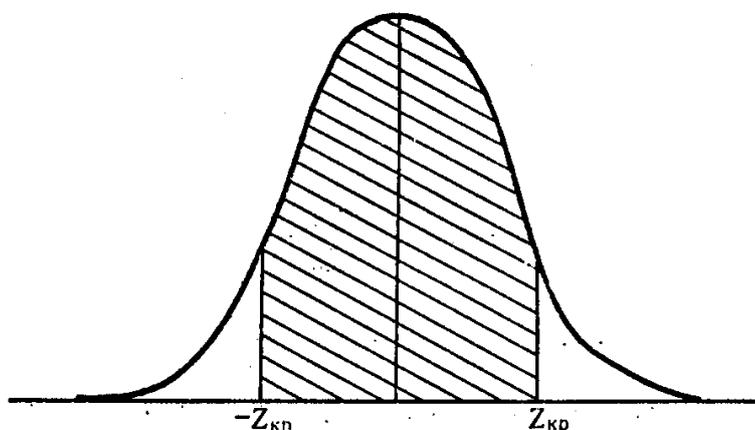


Рис. 1

Перед применением любого метода, который предполагает существование нормального распределения, наличие последнего нужно проверять в первую очередь (рис. 2).

Анализ

Описательные статистики

Частоты

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

¹⁹ 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

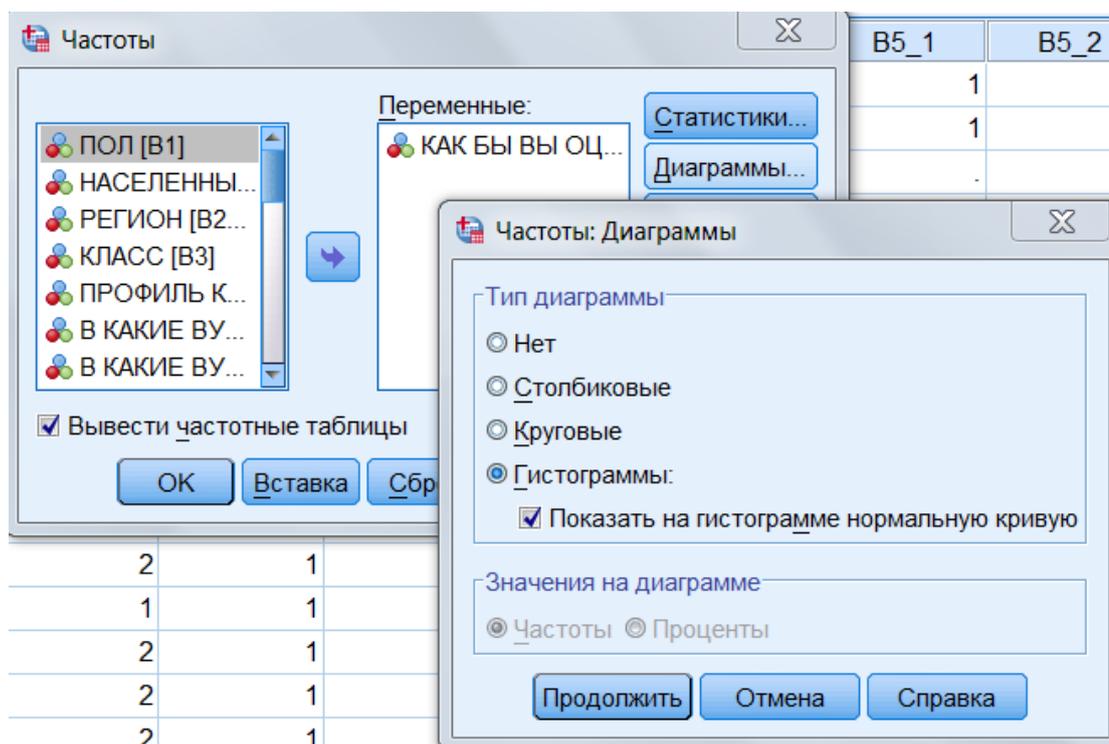


Рис. 2

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. 3)²⁰.

Анализ

Непараметрические критерии

Устаревшие диалоговые окна

Одновыборочный Колмогорова-Смирнова...

²⁰ Там же. С. 85

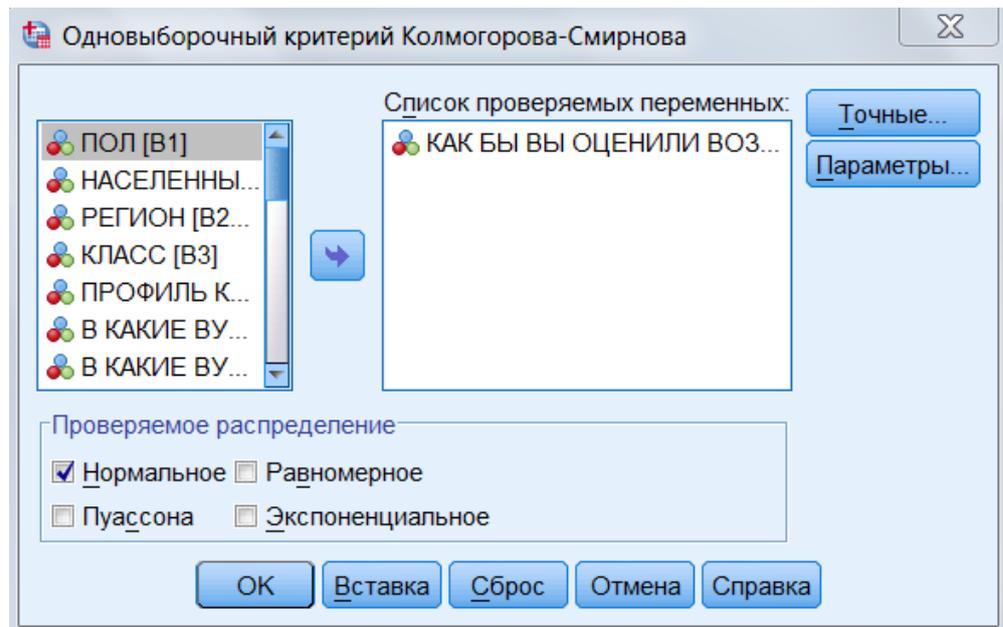


Рис. 3

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 образом²¹.

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

²¹ 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 надежности²².

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

²² 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 расчетов²³.

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.

²³ 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. 4):

Анализ

Описательные статистики

Частоты...

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

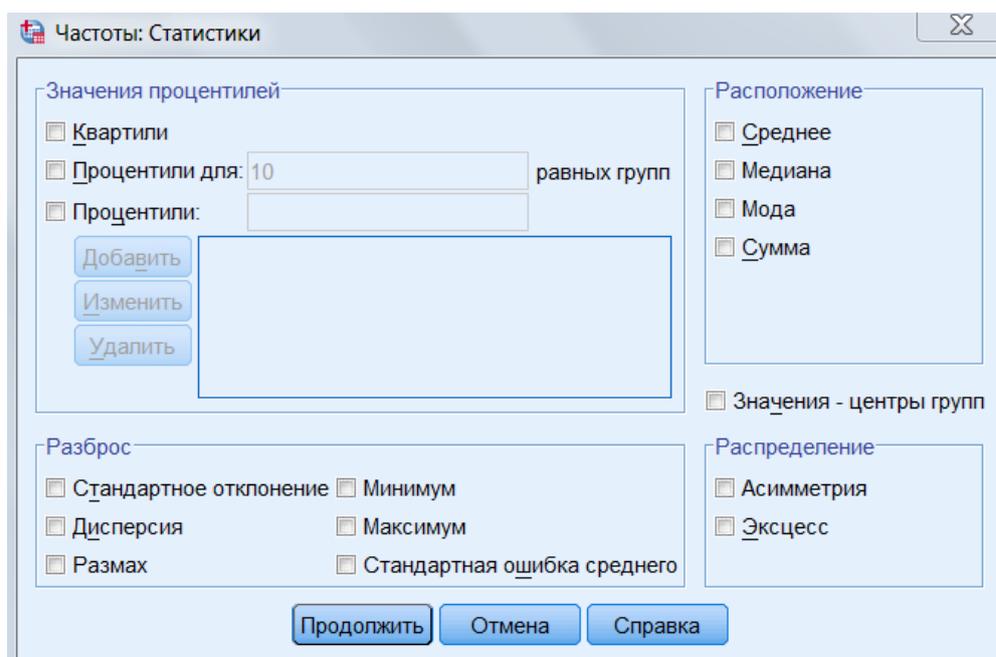


Рис. 4

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 нормального²⁴.

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»

²⁴ 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 2²⁵.

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

²⁵ 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 факторами²⁶.

²⁶ 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. Анализ одномерного распределения

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 переменной²⁷.

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.

²⁷ 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. 5):

Анализ

Описательные статистики

Частоты

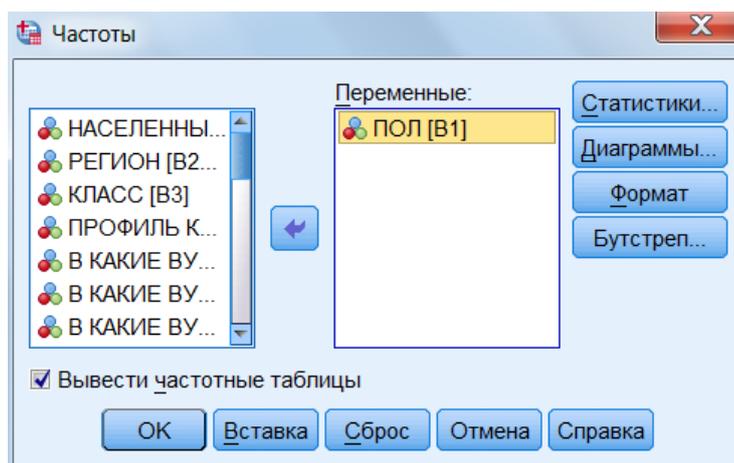


Рис. 5

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 просмотра²⁸.

Результаты появятся в окне просмотра результатов. Перед самой частотной таблицей выводится небольшая таблица с обзором допустимых и отсутствующих значений (рис. 6).

²⁸ 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. 6

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) процент²⁹.

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.

²⁹ 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 голосования³⁰.

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

³⁰ 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. 37):

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

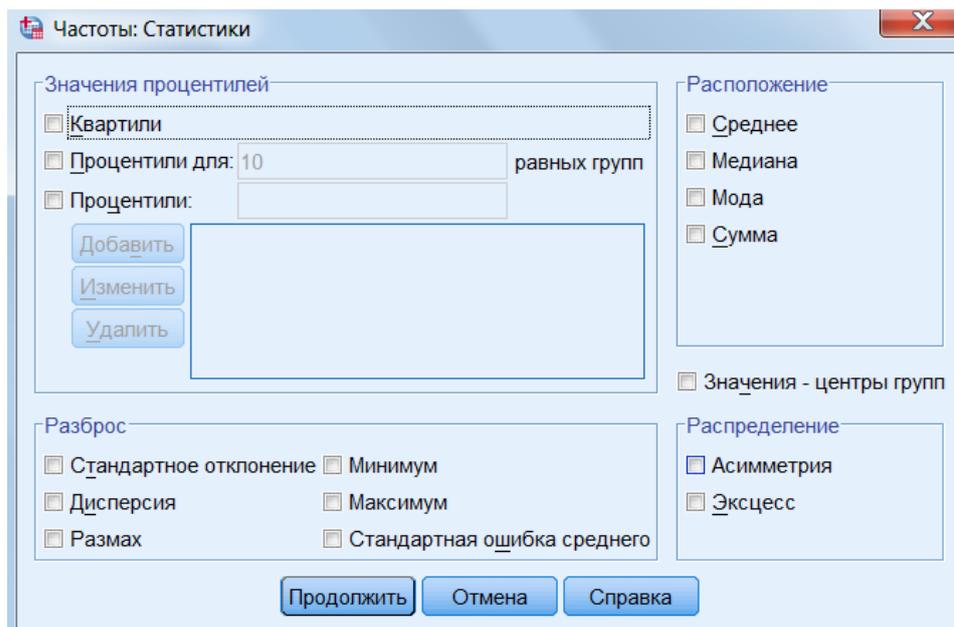


Рис. 7

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. 8).

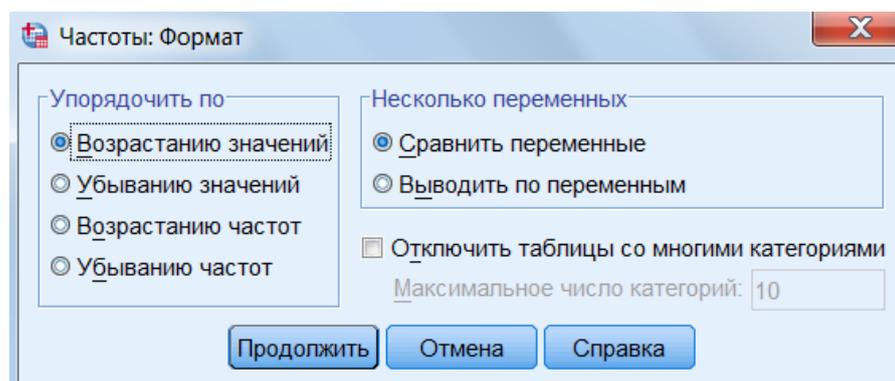


Рис. 8

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 таблиц³¹.

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

³¹ 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 исследований³².

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. 9).

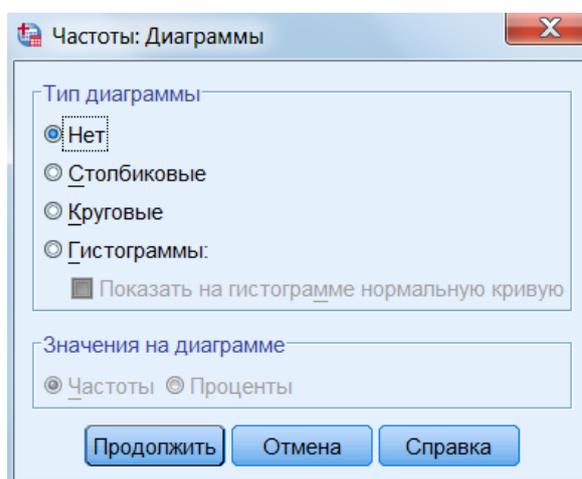


Рис. 9

- 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

³² 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.).

Основные задачи одномерного анализа.

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.

Одномерный описательный анализ решает поставленную задачу взаимодополняющими методами³³:

- построения частотных распределений (таблица 4);

Таблица 4

| В своей жизни Вы лично сталкивались с проблемой наркомании?³⁴ | Процент |
|---|----------------|
| Да | 59 |
| Нет | 41 |
| ИТОГО | 100 |

(больше половины опрошенных сталкивались с проблемой наркомании, что говорит о большом распространении данной проблемы)

- графического представления поведения анализируемой переменной (рисунок 10);

³³ Крыштановский А.О. Анализ социологических данных с помощью пакета SPSS : учебное пособие для вузов / А.О. Крыштановский. – 2-е изд. – М. : Изд. дом ГУ ВШЭ, 2007.

³⁴ Опрос населения г. Благовещенска в марте 2012 г. (Центр социологических исследований при кафедре социологии АмГУ).

В своей жизни Вы лично
сталкивались с проблемой
наркомании?



Рис. 10

- получения статистических характеристик распределения анализируемой переменной (например, средней арифметической, медианы, моды и т.п.).

Результаты контент-анализа и структурированного наблюдения представляют собой также таблицы частотных распределений (примеры посмотреть в практических заданиях) и анализируются на тех же принципах, что изложены в предыдущем параграфе.

Проблема пропущенных значений (проблема «неответов»).

Социолог постоянно сталкивается с ситуацией, когда значительная часть респондентов не дает ответа на какие-то вопросы анкеты.

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. 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».

Первый подход связан с искусственным заполнением пропусков. Известно много способов, позволяющих это сделать. Один из самых распространенных способов – заполнение пропуска средним значением рассматриваемого признака. И исследователь должен понимать, что, поступая так, он рискует придать данным более ровный, «серый» характер, чем это имеет место в действительности. Другой способ использование вместо среднего арифметического – медиану. Можно поступать по-другому: проанализировать распределение признака для тех респондентов, которые ответили на соответствующий вопрос, и заполнять пропуски таким образом, чтобы получающееся в результате распределение имело тот же характер.

Второй подход к рассмотрению кодов пропущенных данных рассматривает эти коды как равноправные остальным числовым кодам, которые приписаны всем другим типам ответов.

Третьим вариантом построения таблицы одномерного частотного распределения выступает возможность исключения из дальнейшего анализа тех респондентов, которые затруднились дать ответ.

Вопрос о том, какой из показателей – процент опрошенных, либо процент ответивших (валидный процент) – 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.

Четвертый вариант задания кодов пропущенных данных чаще всего встречается в ситуации, когда анализируемая переменная выражена количественно. Иногда в ответах респондентов на вопросы, например, о размере получаемых доходов встречаются данные, которые, строго говоря, не могут быть признаны ошибочными, однако, скорее всего, являются недостоверными. Например, если респондент сказал, что у него 15 детей или что его зарплата 5 млн. руб., эти ответы едва ли корректны. Иными словами, для многих показателей мы можем указать границы, допустимых значений, а те данные, которые выходят за эти границы целесообразно признать пропущенными данными³⁵.

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 : учебное пособие для вузов / А.О. Крыштановский. – 2-е изд. – М. : Изд. дом ГУ ВШЭ, 2007. – 281 с.

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

Толстова Ю.Н. Анализ социологических данных. – М. : Научный мир, 2000. – 352 с.

Контрольные вопросы

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

³⁵ Крыштановский А.О. Анализ социологических данных с помощью пакета SPSS : учебное пособие для вузов / А.О. Крыштановский. – 2-е изд. – М. : Изд. дом ГУ ВШЭ, 2007.

5. Что такое «одномерный описательный анализ данных» и как он реализуется в социологии?

6. Как в социологии решается проблема пропущенных вопросов?

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).

6. Проанализируйте результаты социологических опросов (сделайте вывод по таблице / диаграмме, поясните, на что указывают данные и постарайтесь их объяснить).

6.1. Альтернативные вопросы.

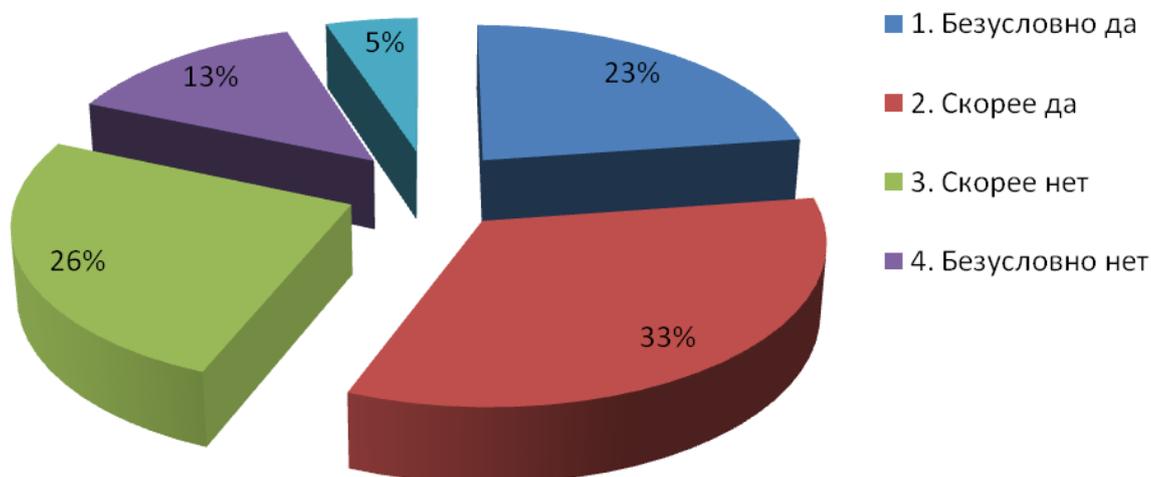
А) опрос населения Амурской области в январе-феврале 2013 г. (ЦСИ АмГУ³⁶):

³⁶ Центр социологических исследований при кафедре социологии АмГУ

| Смотрите ли вы выпуски новостей местных телекомпаний, если да, то, как регулярно? | Процент |
|--|----------------|
| Смотрю регулярно, каждый день | 37,6 |
| Смотрю периодически, когда удастся (2-3 раза в неделю) | 32,4 |
| Смотрю редко, случайно (1 раз в неделю) | 13,0 |
| Вообще не смотрю местные новости | 16,4 |
| Затрудняюсь ответить | 0,5 |
| Итого | 100,0 |

Б) опрос жителей Амурской области в августе 2012 г. (ЦСИ АмГУ, Мониторинг³⁷):

9. Скажите, пожалуйста, в последнее время чувствуете ли Вы себя в безопасности?



В) опрос жителей Амурской области в августе 2012 г. (ЦСИ АмГУ, Мониторинг):

| 10. Кто лучше всего может Вам помочь в обеспечении личной безопасности по месту жительства? | % |
|--|------------|
| 7. Родственники, друзья | 33,4 |
| 9. Я сам(а) | 32,3 |
| 1. Наряд полиции, участковый | 13,3 |
| 2. Бдительные соседи | 8,1 |
| 11. Никто | 4,3 |
| 5. Местные власти | 2,7 |
| 12. Затрудняюсь ответить | 1,4 |
| 6. Частные охранные структуры | 1,3 |
| 10. другое | 1,2 |
| 8. Незнакомые люди, прохожие | 0,9 |
| 3. Старший по дому, подъезду | 0,7 |
| 4. Общественные организации по месту жительства | 0,3 |
| ИТОГО | 100 |

³⁷ Центр изучения общественного мнения «Мониторинг» (г. Благовещенск)

6.2. Множественные вопросы.

А) опрос руководителей общественных организаций Амурской области в марте 2013 г. (ЦСИ АмГУ):

| Что мешает деятельности Вашей организацией?» | Процент³⁸ |
|---|-----------------------------|
| Недостаток денежных средств, слабая материально-техническая база | 66,7% |
| Отсутствие поддержки со стороны бизнеса, возможных спонсоров | 38,1% |
| Нехватка специалистов для работы в организации | 33,3% |
| Отсутствие интереса, поддержки со стороны органов местного самоуправления | 33,3% |
| Отсутствие интереса, поддержки со стороны органов власти | 28,6% |
| Отсутствие помещения | 28,6% |
| Низкая активность членов организации, недостаток энтузиазма | 23,8% |
| Нехватка транспортных средств | 23,8% |
| Отсутствие интереса со стороны целевой аудитории, потребителей услуг | 19,0% |
| Нехватка рекламы, доступа к средствам массовой информации | 14,3% |
| Недостаток информации, необходимой для работы | 14,3% |
| Низкая квалификация сотрудников организации | 9,5% |
| Создание препятствий со стороны надзорных, контролирующих органов | 9,5% |
| Ничто не мешает | 9,5% |
| Несовершенство законодательства | 4,8% |
| Нежелание взаимодействовать со стороны аналогичных организаций города | 4,8% |

Б) опрос первокурсников АмГУ в сентябре 2010 г. (ЦСИ АмГУ):

| Что больше всего повлияло на Ваше решение учиться в АмГУ, а не в другом вузе? | Процент³⁹ |
|--|-----------------------------|
| Хорошие отзывы друзей, родственников | 26% |
| Достаточная информация о вузе и факультете | 17% |
| Наличие рекламы вуза в СМИ и Интернете | 3% |
| Возможность получить хорошее образование | 68% |
| Простота процедуры поступления | 5% |
| Престижность диплома вуза | 39% |
| Контингент студентов | 7% |
| Материально-техническая оснащенность вуза | 3% |
| Высокий уровень профессионализма преподавательского состава | 19% |
| Возможность международного обмена, стажировок | 12% |
| Возможность включения в научно-исследовательскую деятельность | 3% |
| Хорошие перспективы трудоустройства | 27% |
| Возможность интересной внеучебной деятельности | 5% |

³⁸ Суммарный процент более 100%, т.к. респондент мог дать более одного ответа.

³⁹ Суммарный процент более 100%, т.к. респондент мог дать более одного ответа.

| | |
|--|-----------------------|
| Что больше всего повлияло на Ваше решение учиться в АмГУ, а не в другом вузе? | Процент ³⁹ |
| Наличие общежития | 3% |
| Удобное территориальное расположение здания | 10% |
| Возможность поступления с моими результатами ЕГЭ | 12% |
| ИТОГО | 260% |

6.3. Проблема «неответов».

А) пилотажный экспресс-опрос жителей Амурской области в июле 2011 г. (ЦСИ АмГУ):

| ПРЕДСТАВЬТЕ СЕБЕ, ЧТО В СЛЕДУЮЩЕЕ ВОСКРЕСЕНЬЕ СОСТОЯТСЯ ВЫБОРЫ ДЕПУТАТОВ ГОСУДАРСТВЕННОЙ ДУМЫ РФ И ЗАКОНОДАТЕЛЬНОГО СОБРАНИЯ АМУРСКОЙ ОБЛАСТИ. СКАЖИТЕ, ПОЖАЛУЙСТА, ЗА КАКУЮ ПАРТИЮ ВЫ БЫ ПРОГОЛОСОВАЛИ? | | | |
|---|----------------------|---------|------------------|
| | | Процент | Валидный процент |
| Валидные | Единая Россия | 42,0 | 61,7 |
| | КПРФ | 15,1 | 22,2 |
| | ЛДПР | 7,6 | 11,1 |
| | Справедливая Россия | 3,4 | 4,9 |
| | Итого | 68,1 | 100,0 |
| Пропущенные | Не пойду на выборы | 16,0 | |
| | Затрудняюсь ответить | 10,1 | |
| | Отказ от ответа | 5,9 | |
| | Итого | 31,9 | |
| Итого | | 100,0 | |

2.4. Статистические характеристики

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 умолчанию⁴⁰.

Чтобы получить описательную статистику числовых переменных, можно щелкнуть в диалоге **Частоты** на кнопке **Статистики**. Откроется диалоговое окно *Частоты: Статистики* (рис. 11).

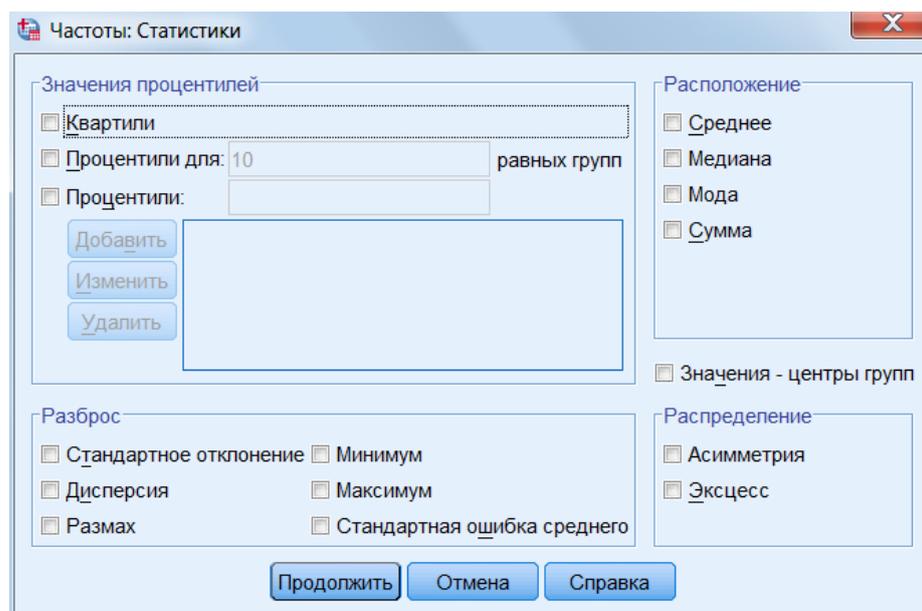


Рис. 11

В группе **Значения процентов** 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.

Квартиль является естественным развитием медианы, с той разницей, что квартильное разбиение делит всех респондентов не на 2, а на 4 части. Первый

⁴⁰ Наследов А.Д. SPSS: Компьютерный анализ в психологии и социальных науках : самоучитель / А.Д. Наследов. – СПб. : Питер, 2005. – С. 105

квартиль – это такая точка на шкале, значения меньше (либо равные) которой отметили 25 % опрошенных. Второй квартиль – точка, меньше которой отметили 50 % опрошенных (следовательно, второй квартиль совпадает с медианой). Наконец, третий квартиль – точка, градации меньше которой отметили 75 % опрошенных⁴¹.

■ *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

⁴¹ Толстова Ю.Н. Анализ социологических данных. – М. : Научный мир, 2000.

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 $x = 600 : 12 = 50$.

Одной из важнейших характеристик при описании поведения отдельных переменных является **показатель средней тенденции**. Среднее арифметическое значение признака, вычисленное для какой-либо группы респондентов, чаще всего интерпретируется как значение для наиболее типичного для этой группы человека. Несомненно, что среднее арифметическое переменной представляет совокупность значений этой переменной неполно и с возможными ошибками. Следовательно, для нас важно знать не только значение данной модели, но и степень точности, качества этой модели.

Прежде всего отметим, что, используя для описания выборки только ту или иную меру средней тенденции, исследователь рискует сильно ошибиться в своей оценке характера изучаемой совокупности респондентов. Например, если изучаемый признак - возраст, то две совокупности людей из 6-ти человек каждая, характеризующиеся следующими значениями возраста, будут иметь одинаковое среднее арифметическое:

10, 10, 10, 50, 50, 50

30, 30, 30, 30, 30, 30.

В то же время совершенно ясно, что практически для любой социологической задачи это будут совсем разные совокупности. И узнать это можно, только как-то оценив степень разброса значений возраста в каждой из них: в первой – разброс большой, во второй – он отсутствует.

Наряду со средней характеристикой, которая удобна тем, что дает нам картину (вернее, часть картины) поведения значений переменной, целесообразно иметь и еще одно число, которое оценивало бы качество средней как модели. Функции такой характеристики выполняют меры разброса: дисперсия, стандартное отклонение, стандартная ошибка среднего. Общая логика здесь такова: чем меньше разброс значений, тем лучше среднее арифметическое характеризует распределение признака; чем стандартное отклонение меньше среднего арифметического, тем лучше последнее

характеризует распределение признака. Стандартная ошибка среднего ($\sigma_{\bar{x}}$) тоже широко используется для решения задачи оценки качества среднего как модели с несколько иной стороны: она дает возможность соотнести величину \bar{x} с генеральным математическим ожиданием. Последнее с вероятностью 0,95 лежит в интервале $(\bar{x} \pm 2\sigma_{\bar{x}})$.

Подводя итог, необходимо подчеркнуть, что использование среднего арифметического без указания одного из показателей качества среднего как модели (дисперсии, стандартного отклонения, либо стандартной ошибки среднего) не дает возможности удовлетворительной интерпретации полученного среднего.

- *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: $(6 + 7) : 2 = 6,5$.

Медиана – это значение рассматриваемого признака, которое делит отвечающий этому признаку вариационный ряд (т.е. последовательность значений признака, расположенных в порядке их возрастания) пополам. Иначе говоря, медиана обладает тем свойством, что половина всех выборочных значений признака меньше нее, а половина – больше⁴².

Пример. Доход девяти респондентов таков (в тысячах рублей):

⁴² Толстова Ю.Н. Анализ социологических данных. – М. : Научный мир, 2000.

12 15 11 30 45 11 10 30 36

Упорядочим по возрастанию:

10 11 11 12 15 30 30 36 45

Находим значение, левее которого расположена одна половина респондентов, а правее – другая. В данном случае – 15. Отметим, некоторую особенность медианы по сравнению со средним арифметическим. Средний доход указанной группы респондентов равен примерно 22. Представим себе, что доход пятого респондента не 45, а 450, то есть резко выбивается из общей массы. Тогда среднее арифметическое уже будет 67 тыс. рублей. Но у оставшихся 8 респондентов доход намного меньше среднего. Медиана же остается прежней – 15 тыс. рублей. Следовательно, медиана становится в некотором смысле более «демократичным» показателем.

■ *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.

■ *Эксцесс*: Коэффициент вариации (эксцесс) 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 вариации 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 данных⁴³.

Исследование данных.

⁴³ 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

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**).

Для выполнения рассматриваемой процедуры необходимо реализовать последовательность команд:

Анализ

Описательные статистики

Разведочный анализ

В результате выполнения этой последовательности команд укромно скрывается главное диалоговое окно процедуры **Исследовать** (рис. 12). 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.

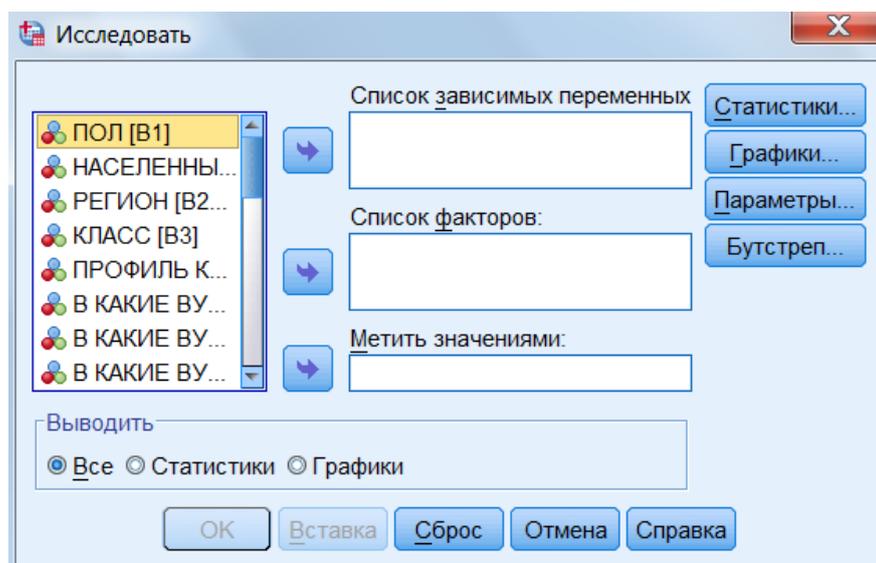


Рис. 12

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...* ⁴⁴.

Получение сводки для наблюдений.

Команда **Итоги по наблюдениям** предназначена для получения с различной степенью детализации упорядоченного списка всех данных файла или их подмножества. Это обычно необходимо при определении состава и качества данных для их дальнейшего редактирования или анализа (рис. 13).

Анализ

Отчеты

Итоги по наблюдениям...

⁴⁴ Пациорковский В.В., Пациорковская В.В. SPSS для социологов: Учебное пособие. ИСЭПН РАН. – М., 2005. – С. 162–164

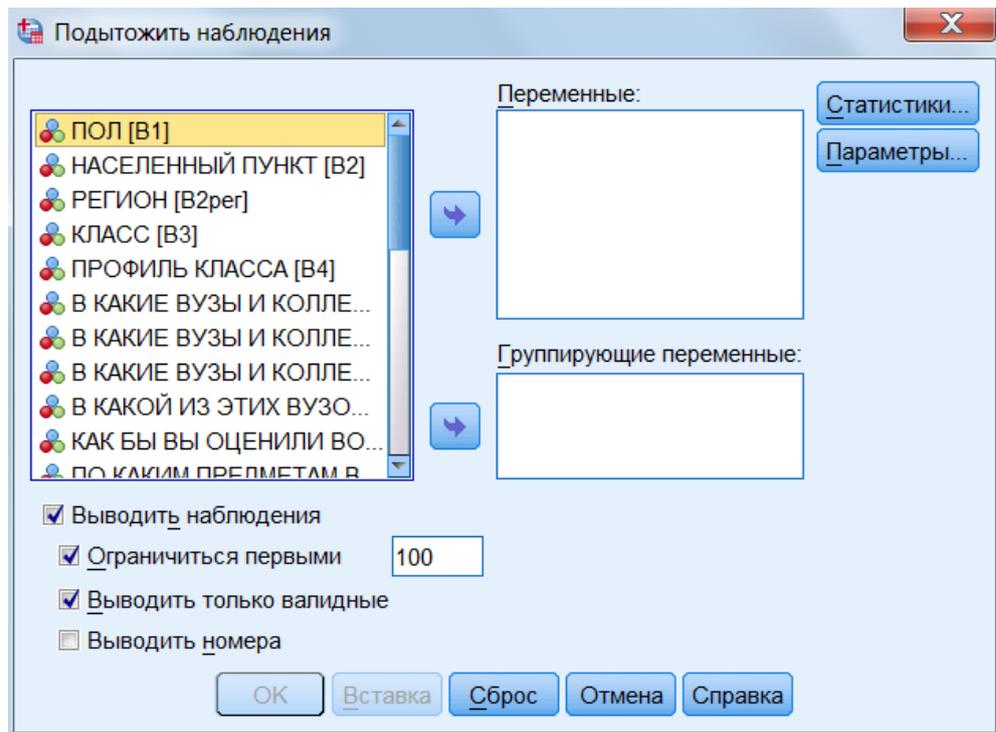


Рис. 13

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 подзаголовков⁴⁵.

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-критерия⁴⁶:

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

Анализ

Сравнение средних

T-критерий для независимых выборок

⁴⁵ 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

⁴⁶ Наследов А.Д. SPSS: Компьютерный анализ в психологии и социальных науках : самоучитель / А.Д. Наследов. – СПб. : Питер, 2005. – С. 145–146

> 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.

Крыштановский А.О. Анализ социологических данных с помощью пакета SPSS : учебное пособие для вузов / А.О. Крыштановский. – 2-е изд. – М. : Изд. дом ГУ ВШЭ, 2007. – 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.

Толстова Ю.Н. Анализ социологических данных. – М. : Научный мир, 2000. – 352 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?
5. Как соотносятся меры средней тенденции и типы шкал?
6. Какие статистические характеристики указывают на качество среднего арифметического?

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.

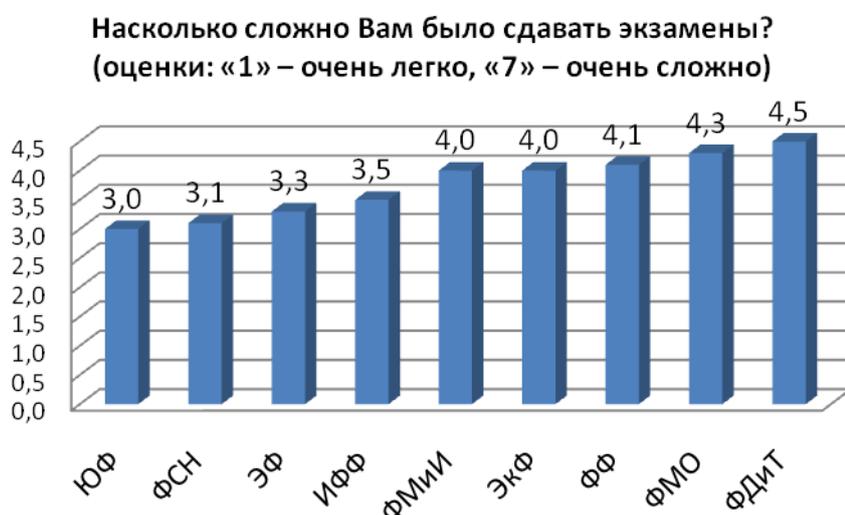
6. Выведите все средние для табличного вопроса, предварительно отобрав конкретную группу респондентов (например, юношей, девушек, студентов конкретного факультета – на ваш выбор).

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

8. Проанализируйте приведенные ниже данные.

8.1. Среднее арифметическое по порядковым шкалам.

А) опрос абитуриентов АмГУ в сентябре 2012 г. (ЦСИ АмГУ):



8.2. Меры средней тенденции и разбросов.

А) опрос руководителей общественных организаций Амурской области в марте 2013 г. (ЦСИ АмГУ):

| Количество ресурсов материально-технической базы в общественных организациях | |
|---|---------|
| Среднее | 3,0000 |
| Стд. ошибка среднего | 0,49281 |
| Медиана | 2,0000 |
| Мода | 2,00 |
| Стд. Отклонение | 2,25832 |

9. Проанализируйте меры средней тенденции и разброса. Сравните материально-техническую базу НКО и сделайте вывод

А) опрос руководителей общественных организаций Амурской области в марте 2013 г. (ЦСИ АмГУ):

| Количество ресурсов материально-технической базы в общественных организациях | НКО областного центра | НКО периферии |
|--|-----------------------|---------------|
| Среднее | 3,8750 | 2,4615 |
| Стд. ошибка среднего | 0,89518 | 0,55024 |
| Медиана | 4,0000 | 2,0000 |
| Мода | 2,00 | 1,00 |
| Стд. отклонение | 2,53194 | 1,98391 |

2.5. Анализ двумерного распределения

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 переменными⁴⁷.

⁴⁷ 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

К наиболее часто используемым инструментам изучения взаимосвязи двух переменных относятся методы анализа таблицы сопряженности. Анализ таблицы является весьма простым и наглядным, и вместе с тем эффективным инструментом изучения одновременно двух переменных⁴⁸.

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. 14).

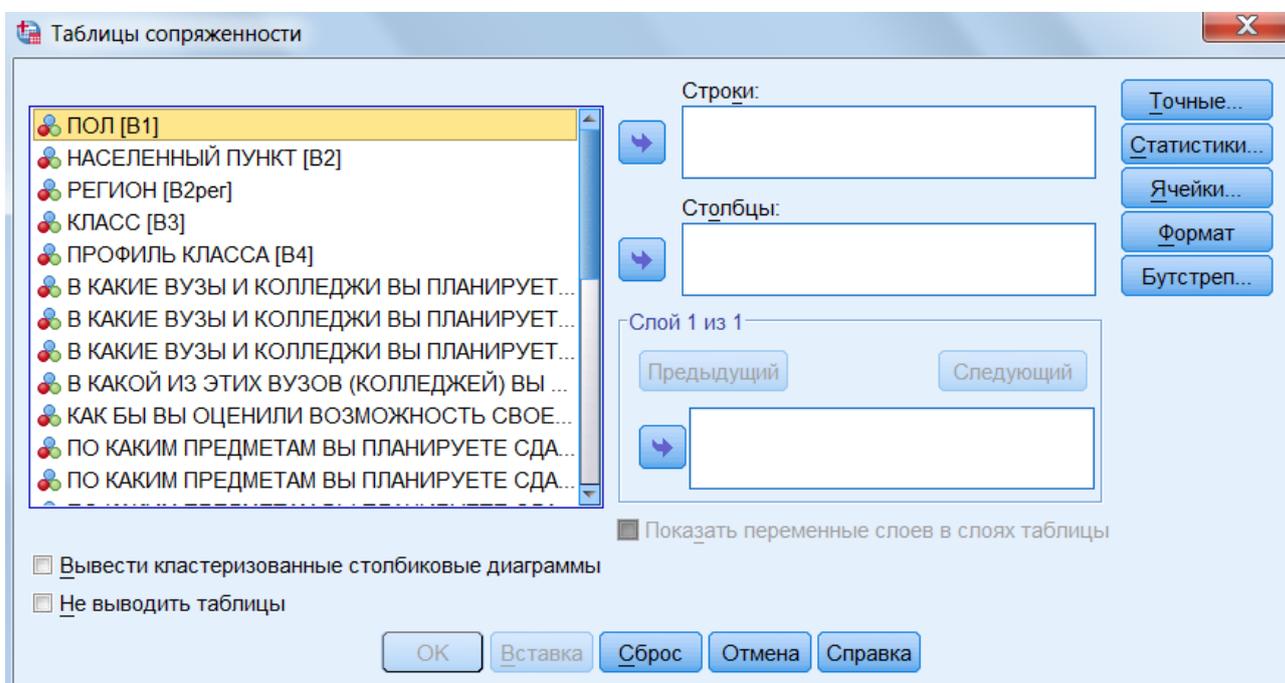


Рис. 14

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

⁴⁸ 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

сопряженности⁴⁹. 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. 15).

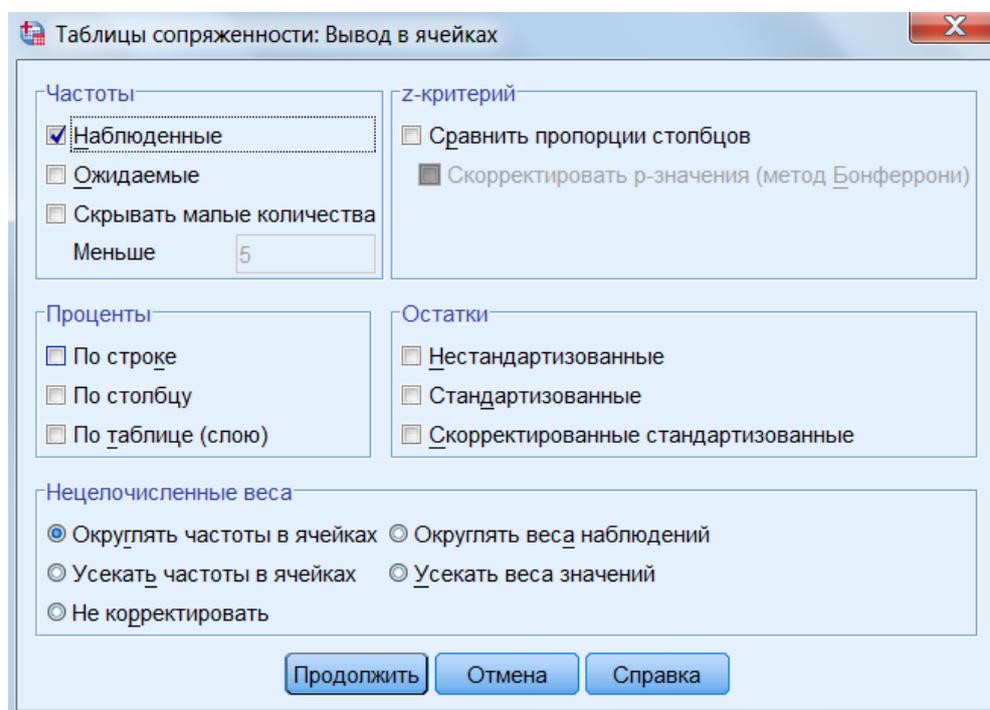


Рис. 15

⁴⁹ 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 просмотра⁵⁰.

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. 16).

⁵⁰ 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. 16

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 таблице⁵¹.

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

⁵¹ 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.

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. 17).

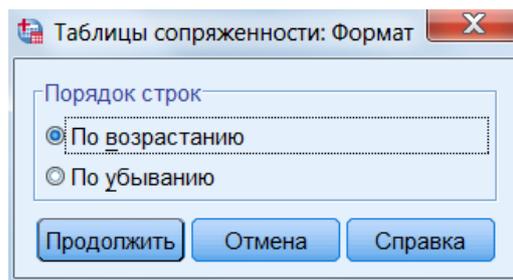


Рис. 17

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 наименьшему⁵².

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

⁵⁸ Там же. С. 186–187

Вывести кластеризованные столбиковые диаграммы. Тогда на диаграмме будут показаны две группы столбцов для двух переменных строк⁵³.

Чтобы получить статистические критерии для таблиц сопряженности, щелкните на кнопке *Статистики* в диалоговом окне **Таблицы сопряженности**. Откроется диалоговое окно Таблицы сопряженности: Статистики (рис. 18).

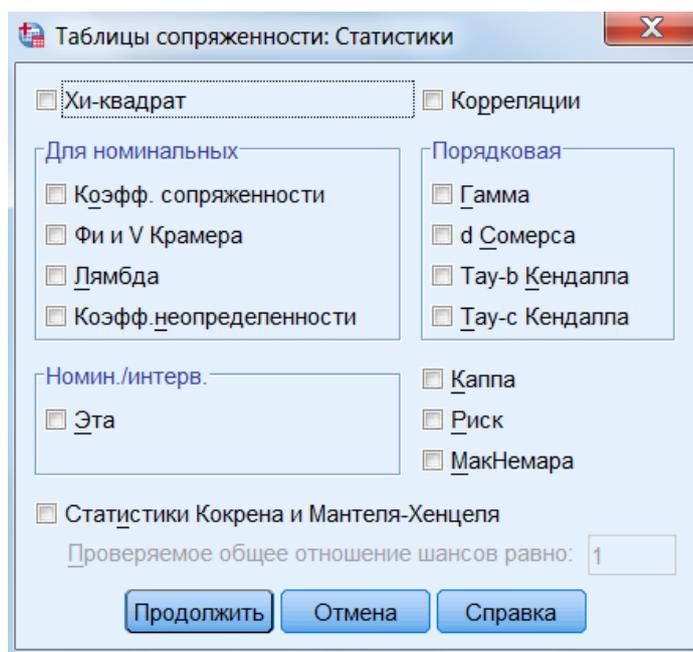


Рис. 18

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

- Тест хи-квадрат (χ^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 (κ)
- Мера риска
- Тест Мак-Немара

⁵³ Там же. С. 188

■ Статистики Кохрана и Мантеля-Хэнзеля

Критерий хи-квадрат имеет большое значение в статистических вычислениях, поэтому ему уделим особое внимание⁵⁴.

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 χ^2 и является установление степени соответствия между наблюдаемыми и ожидаемыми значениями ячеек.

В основе критерия независимости лежит вычисление величины χ^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 χ^2 также становится большой. Вместе с χ^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

⁵⁴ Там же. С. 190–191

переменных отклоняется и делается вывод о том, что две классификации (переменные) зависят друг от друга⁵⁵.

Пример. В ходе исследования были получены данные, описывающие распределение пользователей интернетом в зависимости от пола и возраста (таблица 6)⁵⁶.

Таблица 6

| | | ПОЛ | | Итого |
|----------------------------------|-----|---------|---------|--------|
| | | Мужской | Женский | |
| ПОЛЬЗУЕТЕСЬ ЛИ ВЫ ИНТЕРНЕТОМ? | Да | 64,7% | 63,2% | 63,8% |
| | Нет | 35,3% | 36,8% | 36,2% |
| Итого | | 100,0% | 100,0% | 100,0% |

$$\chi^2 = 0,137 \quad p = 0,711$$

Исходя из данных перекрестной таблицы и рассчитанного значения $p = 0,711$ (больше 0,05), делаем вывод об отсутствии связи между полом и использованием интернета: среди мужчин и женщин примерное равное количество интернет-пользователей (таблица 7).

Таблица 7

| | | СКОЛЬКО ЛЕТ ВАМ ИСПОЛНИЛОСЬ? | | | | | | Итого |
|-------------------------------------|-----|------------------------------|--------|--------|--------|--------|-------------|--------|
| | | 18-24 | 25-34 | 35-44 | 45-54 | 55-64 | 65 и старше | |
| ПОЛЬЗУЕТЕСЬ ЛИ ВЫ ИНТЕРНЕТОМ? | Да | 97,7% | 87,4% | 73,4% | 59,8% | 40,4% | 6,8% | 63,9% |
| | Нет | 2,3% | 12,6% | 26,6% | 40,2% | 59,6% | 93,2% | 36,1% |
| Итого | | 100,0% | 100,0% | 100,0% | 100,0% | 100,0% | 100,0% | 100,0% |

$$\chi^2 = 202,030 \quad p = 0,000$$

Исходя из данных перекрестной таблицы и рассчитанного значения p близкого к нулю (меньше 0,05), делаем вывод о наличии связи между возрастом и использованием интернета: чем старше возрастная группа, тем меньше в ней интернет-пользователей. Это несложно объяснить расположенностью молодого поколения к разного рода гаджетам, поиску информации преимущественно в интернете.

Основные задачи двумерного анализа.

⁵⁵ Наследов А.Д. SPSS: Компьютерный анализ в психологии и социальных науках : самоучитель / А.Д. Наследов. – СПб. : Питер, 2005. – С. 114–115

⁵⁶ Опрос населения Амурской области в январе-февреле 2013 г. (ЦСИ АмГУ).

Независимо от выбранной стратегии анализа (восходящей или нисходящей) и после изучения, условно говоря, «поведения» отдельно взятых признаков, естественным образом возникает необходимость анализа взаимосвязи, взаимодействия между признаками. Будем рассматривать только случай двух признаков. Анализ «поведения» двух признаков – совместного или относительно друг друга – социологу необходим для поиска ответа на вопросы типа: существует ли связь между этими признаками; влияет ли один признак на другой; можно ли, зная значение одного из них, сделать вывод относительно значения другого и т. д. Если гипотезы о взаимосвязях были предварительно сформулированы, то речь может пойти по проверке этих гипотез⁵⁷.

Представляется естественным использовать для оценки связей между признаками т. н. таблицы сопряженности или перекрестные таблицы (кросс-таблицы).

Анализ таблиц сопряженности.

К наиболее часто используемым инструментам изучения взаимосвязи двух переменных относятся методы анализа таблицы сопряженности. Анализ таблицы является весьма простым и наглядным, и вместе с тем эффективным инструментом изучения одновременно двух переменных (таблица 7):

⁵⁷ Татарова Г.Г. Методология анализа данных в социологии (введение) / Учебник для вузов. – М. : NOTA BENE, 1999.

Таблица 7

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

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 таблице⁵⁸.

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. От ответа на данный вопрос будет зависеть, где рассчитывать проценты: по столбцу или по строке. Это зависит от

⁵⁸ 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.

двух обстоятельств: от характера выборки обследованных и от логики анализа. Выборка может быть либо репрезентативной (выборочная совокупность есть микромодель генеральной совокупности), либо нерепрезентативной. В последнем случае нам как минимум неизвестны пропорции существенных характеристик в генеральной совокупности, или мы знаем, что эти пропорции в выборке не соблюдаются⁵⁹. Возможна двоякая логика анализа «от причин к следствию» или «от следствий к причинам», что определяется гипотезой и содержанием данных. Если выборка представительна и отражает пропорции изучаемых групп в генеральной совокупности, тогда можно вести двоякий анализ данных: по логике «от причин к следствию» и «от следствия к причинам»⁶⁰. Как правило (но не всегда), процент рассчитывается там, где находится влияющая переменная, или фактор. Частоты именно по его градациям и берутся за 100 процентов. Разумеется, присутствующая в каждом опросе «объективка» (пол, возраст, образование, происхождение и т. д.) порождает признаки, трактуемые как независимые, то есть факторы.

Также решение вопроса, где рассчитывать проценты: в строках или столбцах, – зависит не только от предварительных гипотез о причинно-следственных связях, но и от знания характеристик первичных распределений (чаще всего квотных признаков: пола, возраста, образования...) анализируемого массива данных. Второй случай возникает при анализе гендерных и когортных особенностей. Например, в гендерных исследованиях опрашивается *равное* количество мужчин и женщин, и в этом случае возможен расчет процентов не внутри ПОЛа, а внутри содержательного вопроса.

Что подвергается анализу?

Для социолога важно осознание необходимости определенной диалектики в понимании признака и его значений: выделение ситуаций, когда

⁵⁹ Крыштановский А.О. Анализ социологических данных с помощью пакета SPSS : учебное пособие для вузов / А.О. Крыштановский. – 2-е изд. – М. : Изд. дом ГУ ВШЭ, 2007.

⁶⁰ Ядов В.А. Социологическое исследование: методология, программа, методы – М.: Наука, 1972.

отдельной альтернативе имеет смысл придать статус самостоятельного признака.

1. Признак в целом («профессия» как таковая).
2. Одна градация признака как самостоятельный признак (профессия «учитель»).
3. Сочетание градаций признака как самостоятельный признак (интеллигенция: «учитель» + «врач»).
4. Сочетание исходных признаков как новый самостоятельный признак.
5. Все признаки как система⁶¹.

Какие задачи решаются?

Во-первых, социолог может сравнивать структуру конкретного признака в различных группах респондентов и выделять на этой основе типы респондентов. Эти типы и есть статистические закономерности, которые требуют от социолога объяснения.

Пример. В ходе исследования были получены данные, описывающие распределение ожидаемого уровня безработицы в зависимости от материального положения (таблица 8)⁶².

Таблица 8

| Как Вы думаете, в ближайшие 12 месяцев безработных (людей, которые не имеют работу и ищут ее) будет больше, примерно столько же или меньше, чем сейчас? | Как бы Вы оценили в настоящее время материальное положение Вашей семьи? | | | | | |
|---|---|---------|---------|--------|--------------|--------|
| | Очень хорошее | Хорошее | Среднее | Плохое | Очень плохое | Итого |
| больше | 12,5% | 19,0% | 26,1% | 37,2% | 47,0% | 29,7% |
| примерно столько же | 50,0% | 44,2% | 48,3% | 38,3% | 28,9% | 44,1% |
| меньше | 12,5% | 15,5% | 8,5% | 5,0% | 6,7% | 8,1% |
| затрудняюсь ответить | 25,0% | 21,2% | 17,1% | 19,5% | 17,4% | 18,1% |
| Итого | 100,0% | 100,0% | 100,0% | 100,0% | 100,0% | 100,0% |

Наблюдается зависимость между оценкой материального положения своей семьи и ожиданиями относительно безработицы: чем хуже оценивается материальное положение, тем пессимистичнее ожидания безработицы. Это

⁶¹ Толстова Ю.Н. Анализ социологических данных. – М. : Научный мир, 2000.

⁶² Перекрестная таблица взята из кейс-задания ФИЭБ для направления подготовки 39.03.01 Социология

можно объяснить тем, что хорошее материальное положение семьи респондента является залогом ее экономической стабильности, и по-видимому, эта стабильность экстраполируется (переносится) и на более широкие структуры, например, как в данном случае, на рынок труда.

Другой из этих типов задач заключается в сравнении социальной (социально-демографической) структуры в различных группах респондентов, то есть для составления социального (социокультурного) портрета (подробнее в следующем параграфе).

Третья задача: поиск гендерных и когортных особенностей.

Примечание: При анализе таблиц сопряженности крайне важно помнить, что мы, по сути дела, ищем наличие (или отсутствие) определенных статистических, а не причинно-следственных зависимостей. Вопрос о том, какая из переменных является причиной, т.е. оказывает влияние, а какая меняется вследствие этой причины, не может быть решен не только с помощью анализа таблиц, но и любым другим формально-статистическим методом⁶³.

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 : учебное пособие для вузов / А.О. Крыштановский. – 2-е изд. – М. : Изд. дом ГУ ВШЭ, 2007. – 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.

⁶³ Крыштановский А.О. Анализ социологических данных с помощью пакета SPSS : учебное пособие для вузов / А.О. Крыштановский. – 2-е изд. – М. : Изд. дом ГУ ВШЭ, 2007.

Татарова Г.Г. Методология анализа данных в социологии (введение) / Учебник для вузов. – М. : NOTA BENE, 1999. – 224 с.

Толстова Ю.Н. Анализ социологических данных. – М. : Научный мир, 2000. – 352 с.

Ядов В.А. Социологическое исследование: методология, программа, методы – М.: Наука, – 1972.

Контрольные вопросы

1. Что такое Двумерные таблицы? Их назначение и роль в социологии?
2. Каковы Форматы таблиц сопряженности?
3. Что такое критерий хи-квадрат? И для чего он нужен при построении таблиц сопряженности?
4. Какие задачи решает социолог, прибегая к двумерному анализу?
5. Что подвергается анализу в таблицах сопряженности?
6. Каково назначение критерия Хи-квадрат при анализе 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.
 - 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.

6. Проанализируйте таблицы (диаграммы) сопряженности. Дайте объяснение полученным результатам.

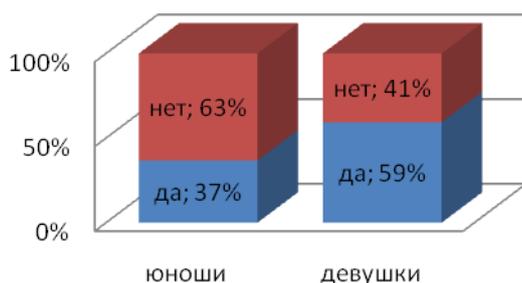
А) опрос населения Амурской области в январе-февреле 2013 г. (ЦСИ АмГУ):

| КАК ЧАСТО ВЫ ЛИЧНО ПОЛЬЗУЕТЕСЬ ДОМА ИНТЕРНЕТОМ? * ПОЛ | | |
|--|---------|---------|
| КАК ЧАСТО ВЫ ЛИЧНО ПОЛЬЗУЕТЕСЬ ДОМА ИНТЕРНЕТОМ? | ПОЛ | |
| | Мужчины | Женщины |
| Каждый день/почти каждый день (6-7 раз в неделю) | 69,3% | 68,5% |
| 4-5 раз в неделю | 8,5% | 8,1% |
| 1-3 раза в неделю | 18,3% | 18,3% |
| 1-3 раза в месяц | 2,6% | 3,6% |
| Реже, чем 1 раз в месяц | 0,7% | 1,5% |
| Никогда | 0,7% | 0,0% |
| Итого | 100,0% | 100,0% |

$$\chi^2 = 2,12 \quad p = 0,832$$

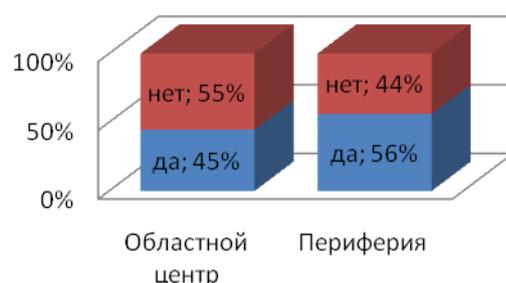
Б) опрос первокурсников АмГУ в сентябре 2010 г. (ЦСИ АмГУ):

Собираетесь ли Вы продолжить обучение в магистратуре?



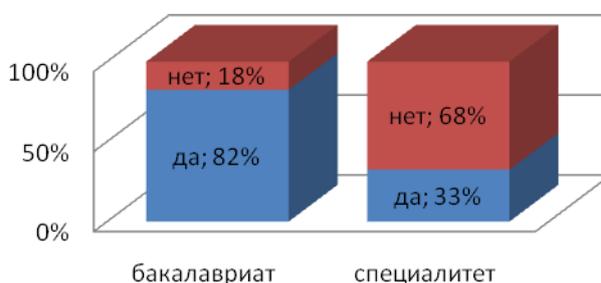
$\chi^2 = 11,36$ $p = 0,01$

Собираетесь ли Вы продолжить обучение в магистратуре?



$\chi^2 = 2,53$ $p = 0,111$

Собираетесь ли Вы продолжить обучение в магистратуре?



$\chi^2 = 55,19$ $p = 0,000$

В) опрос руководителей общественных организаций Амурской области в марте 2013 г. (ЦСИ АмГУ):

Оценка финансового состояния общественных организаций (НКО)

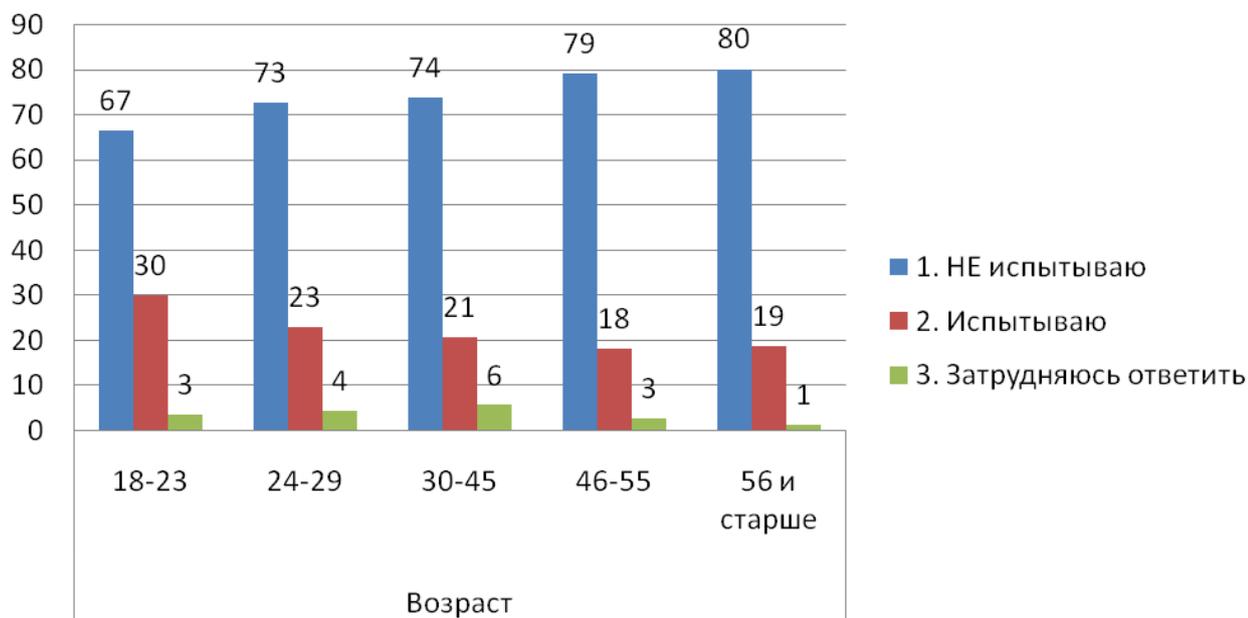


$\chi^2 = 7,07$ $p = 0,07$

| Имеет ли Ваша организация: | Центр / Периферия | |
|---|-----------------------|---------------|
| | НКО областного центра | НКО периферии |
| Собственную компьютерную технику | 71,4% | 41,7% |
| Собственную копировальную технику | 42,9% | 33,3% |
| Собственный транспорт | 14,3% | 8,3% |
| Собственную библиотеку (более 50 изданий) | 42,9% | 25,0% |
| Доступ в Интернет | 85,7% | 75,0% |
| Свой сайт | 57,1% | 33,3% |
| Собственные базы данных | 57,1% | 25,0% |
| Методические разработки | 71,4% | 25,0% |

Г) опрос населения Амурской области в августе 2012 г. (ЦСИ АмГУ, Мониторинг):

28. Некоторые люди испытывают раздражение или неприязнь по отношению к представителям той или иной национальности. а Вы лично испытываете или не испытываете раздражение или неприязнь по отношению к представителям какой-либо национальности?



Д) опрос населения г. Благовещенска в июне 2012 г. (ЦСИ АмГУ):

| Чувствуете ли вы себя защищенным от террористических актов? | По вашему мнению, представляет ли в настоящее время идеология терроризма и насильственного экстремизма реальную угрозу для государства и российского общества? | |
|---|--|--------|
| | Да | Нет |
| Да | 19,8% | 63,6% |
| Нет | 80,2% | 36,4% |
| Итого | 100,0% | 100,0% |

$$\chi^2 = 23,32 \quad p = 0,000$$

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. 19).

Анализ

Множественные ответы

Задать наборы множественных ответов...

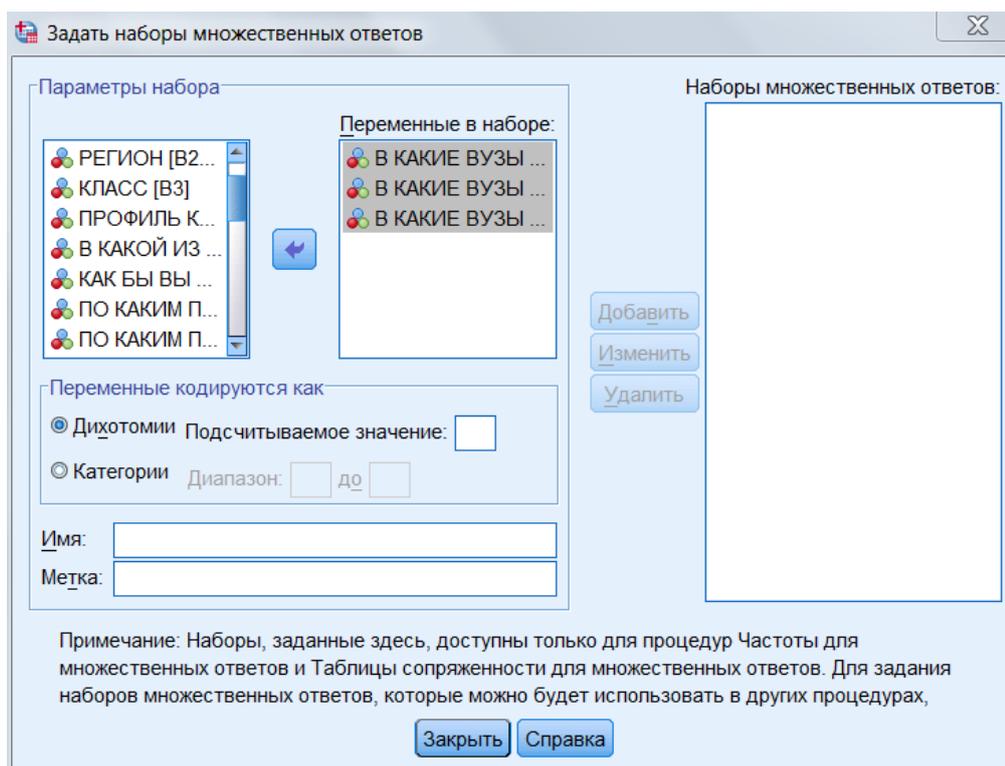


Рис. 19

- 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. 20):

Анализ

Множественные ответы

Частоты...

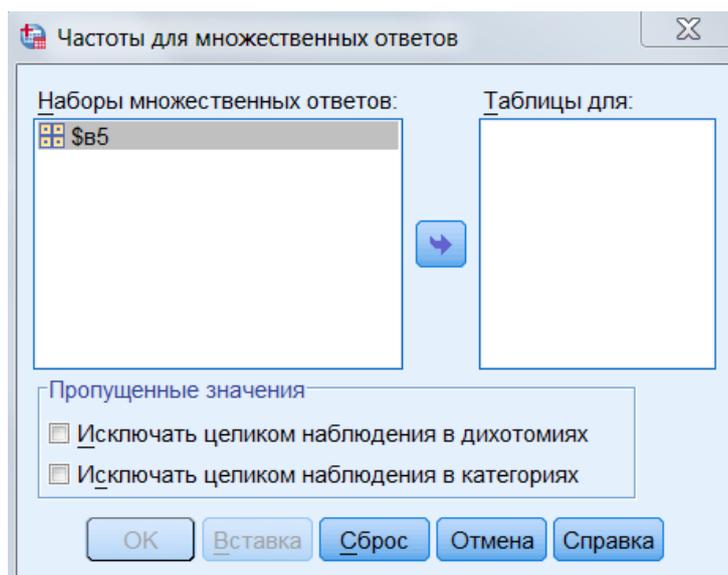


Рис. 20

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. 21).

| | | Ответы | | Процент наблюдений |
|-------------------------------|---|--------|---------|--------------------|
| | | N | Процент | |
| Вузы ^a | 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. 21

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. 22):

Анализ

Множественные ответы

Таблицы сопряженности...

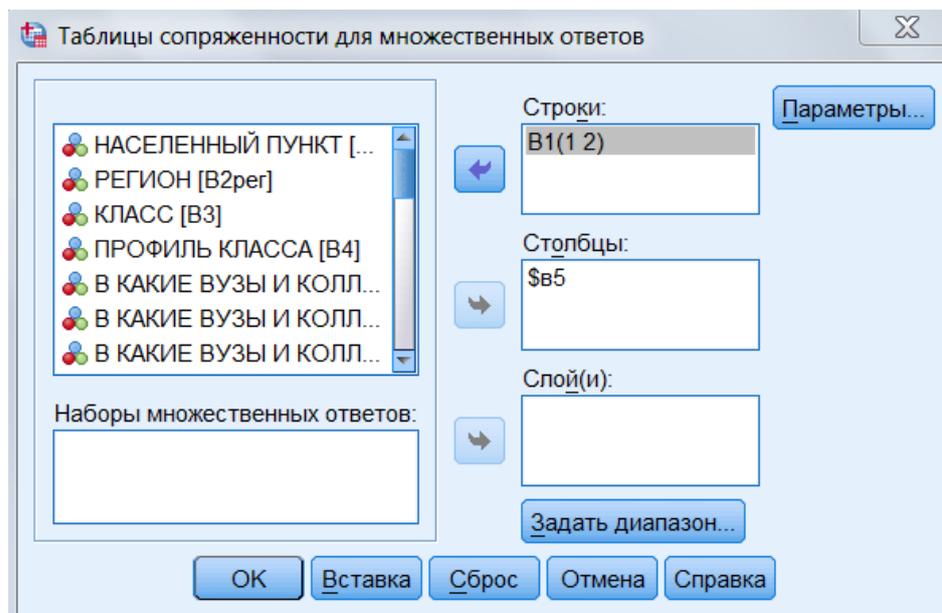


Рис. 22

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. 23):

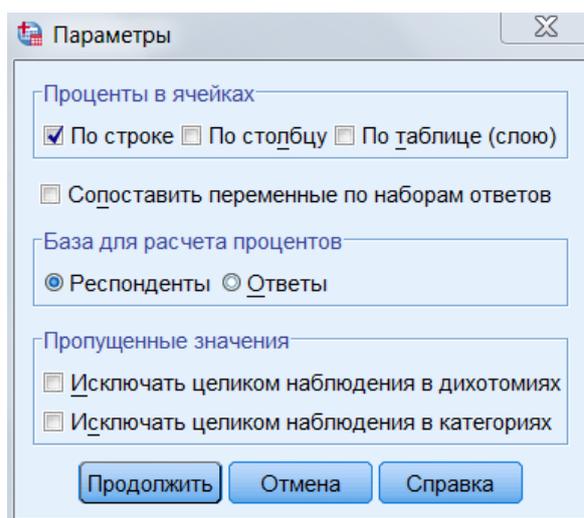


Рис. 23

- 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 респондентов⁶⁴.

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.

⁶⁴ 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 параграфа⁶⁵.

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.

⁶⁵ 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. Основы многомерного анализа в социологии

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. 9:

Таблица 9

| 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. 24)

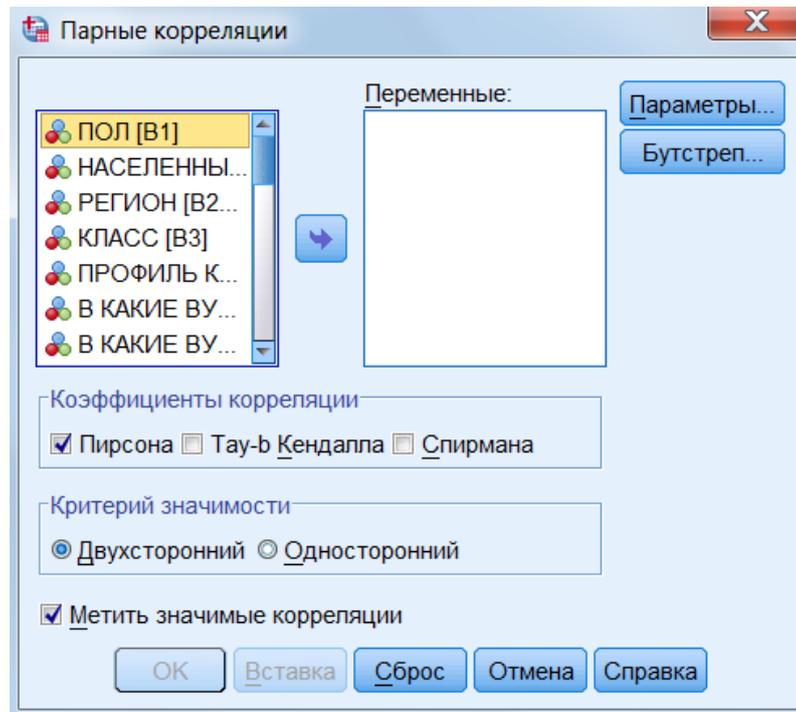


Рис. 24

Более глубокий математический анализ характеристик изучаемого явления – выяснение их взаимодействия и тенденций изменения – осуществляется путем сравнения и сопоставления рядов распределения, построенных на основании группировок по различным признакам. Для решения такой задачи разработаны специальные коэффициенты – коэффициенты корреляции.

Корреляцией называют наличие статистической взаимосвязи признаков⁶⁶. 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): порядок распределения ответов по одной сравниваемой группе респондентов и по другой в определенной степени совпадают; увеличение (уменьшение) параметров факторного признака приводит к увеличению (уменьшению) параметров результативного признака

⁶⁶ Горшков М.К. Прикладная социология: методология и методы: Учебное пособие / М.К. Горшков, Ф.Э. Шереги. – М.: Альфа-М: ИНФРА-М, 2009.

- обратная (отрицательная): порядок распределения ответов по одной сравниваемой группе респондентов в некоторой степени противоположен порядку распределения ответов по второй группе; увеличение (уменьшение) параметров факторного признака приводит к уменьшению (увеличению) параметров результативного признака⁶⁷.

В процессе измерения корреляции между признаками решаются две задачи – определение формы и тесноты связи признаков.

Формы корреляции: линейная и криволинейная.

Теснота связи – степень сопряженности между признаками.

Интерпретация значений коэффициентов представлена ниже (таблица 10):

Таблица 10

| 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.

Пример. Проверим, есть ли связь между таким показателем развития экономики региона (субъекта РФ) как внутренний региональный продукт (обобщающий показатель экономической деятельности региона, характеризующий процесс производства товаров и услуг для конечного использования) и числом научных организаций в регионе⁶⁸. Рассчитаем коэффициент корреляции Пирсона. Он равен 0,909 при уровне значимости 99%. Следовательно, связь между состоянием региональной экономики и уровнем развития науки в регионе очень сильная.

Пример. Предположим, что с ростом дохода респондента растет и его возможность покупать более крупные (с большим числом каналов) пакеты телевидения. Рассчитаем коэффициент корреляции между личным доходом

⁶⁷ Там же.

⁶⁸ Регионы России. Социально-экономические показатели. 2011 : Стат. сб. / Росстат. – М. : 2011.

респондента и количеством каналов, которые может принимать его телевизор (со слов респондента)⁶⁹. Он равен 0,206 при уровне значимости 99%. Следовательно, связь между доходом респондента и его возможностью покупать более дорогие пакеты телевидения – слабая.

Регрессионный анализ.

Регрессионный анализ позволяет прогнозировать значение зависимой переменной на основе значения независимой переменной. То есть основная задача такого анализа – в определении зависимой переменной как линейной функции переменной независимой⁷⁰.

Анализ

Регрессия

Линейная... (рис. 25)

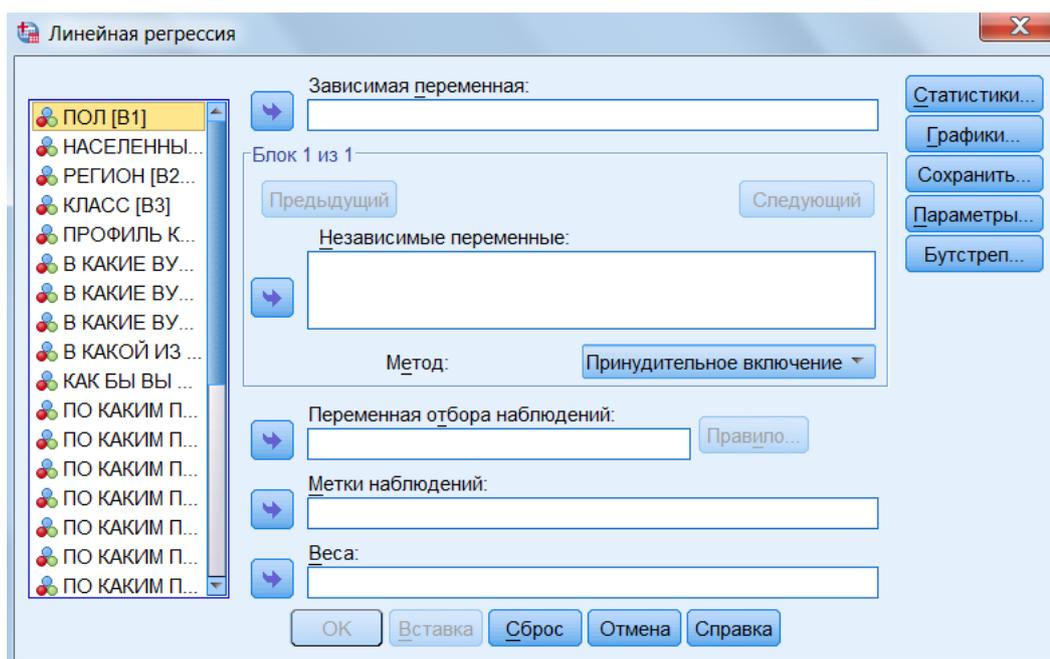


Рис. 25

Данная задача сводится к построению *уравнения регрессии* (формула 2):

$$y = bx + a, \tag{2}$$

где y – зависимая переменная (прогнозируемая)

b – коэффициент регрессии

⁶⁹ Опрос в Амурской области в январе-феврале 2013 г., ЦСИ АмГУ.

⁷⁰ Гуц А.К., Фролова Ю.В. Математические методы в социологии / предисл. Г.Г. Малинецкого. – М. : Издательство ЛКИ, 2007. – 216 с.

x – независимая переменная

a – константа.

R-квадрат (коэффициент детерминации) – величина, характеризующая качество регрессионного уравнения, его способность к прогнозу. Иными словами, показывает, на сколько процентов полученная модель объясняет изменение зависимой переменной.

Пример. В рамках опроса жителей Исландии (ESS, 2013 год, 6 волна) респондентам предлагалось оценить свою политическую ориентацию, доверие политикам и удовлетворенность правительством страны. Измерение во всех случаях проводилось по 11-балльным шкалам, в которых 0 соответствовал крайне левой политической ориентации/абсолютному недоверию политикам/крайней неудовлетворенности правительством, а 10 – крайне правой ориентации/абсолютному доверию/полной удовлетворенности. Была построена регрессионная модель, отражающая зависимость удовлетворенности правительством от политической ориентации и доверия политикам. Коэффициент детерминации R^2 оказался равен 31,3 %, а значения коэффициентов регрессионного уравнения представлены ниже в таблице 11:

Таблица 11

| Коэффициенты | | | | | | |
|--------------|------------------------------|----------------------------------|-------------|--------------------------------|---------|-------|
| Модель | | Нестандартизованные коэффициенты | | Стандартизованные коэффициенты | t | Знч. |
| | | B | Стд. Ошибка | Бета | | |
| 1 | (Константа) | 4,137 | 0,248 | | 16,670 | 0,000 |
| | Политическая ориентация | -0,423 | 0,038 | -0,357 | -11,120 | 0,000 |
| | Доверие политическим партиям | 0,515 | 0,036 | 0,454 | 14,155 | 0,000 |

Коэффициент детерминации говорит нам о том, что полученная модель объясняет удовлетворенность правительством на 31,3 %. Построим уравнение регрессии:

Удовлетворенность правительством = -0,423(политическая ориентация) + 0,515(доверие политическим партиям) + 4,137

Обратим внимание на уровень значимости p , рассчитанный для каждого коэффициента и константы. Все уровни значимости не превышают 0,05, следовательно все коэффициенты и константа статистически достоверны. Константа 4,137 нам показывает, насколько будет удовлетворен правительством респондент, если он придерживается крайне левых взглядов и абсолютно не доверяет политикам. Сдвиг политических взглядов вправо на 1 пункт приведет к снижению удовлетворенности на 0,423, а повышение доверия к политикам на 1 пункт приведет к повышению удовлетворенности на 0,515. Обратим внимание на стандартизированные коэффициенты (Бета). Данный параметр позволяет сопоставить влияние независимых переменных на зависимую (особенно, когда размерности анализируемых переменных различны). Так, доверие политикам влияет на удовлетворенность правительством немного сильнее, чем политическая ориентация.

Классификация данных.

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

Анализ

Классификация

Иерархическая кластеризация... (рис. 26)

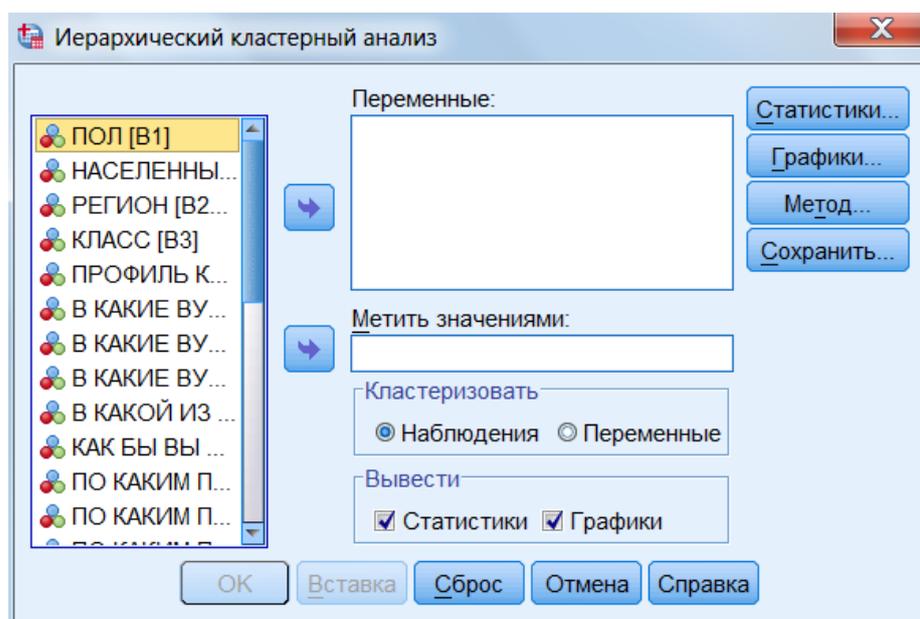


Рис. 26

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**.

Для определения *расстояния* между парой кластеров могут использоваться разные подходы: среднее расстояние между кластерами, среднее расстояние между всеми объектами пары кластеров с учетом расстояний внутри кластеров, расстояние между ближайшими соседями – ближайшими объектами кластеров, расстояние между самыми далекими соседями, расстояние между центрами кластеров (центроидный метод), метод медиан, метод Варда.

Определим основные *виды расстояний* в процедуре кластерного анализа: евклидово расстояние, квадрат евклидова расстояния, линейное расстояние и др.⁷¹

⁷¹ 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.;

Если данные разнородны (например, возраст респондентов меняется от 18 до 74 лет, а его доход от 0 до 150 тыс. рублей), то требуется их *стандартизация*, то есть сведение к однородному виду. Выделяют несколько методов стандартизации: Z-вклады, разброс от -1 до +1, от 0 до 1, максимум 1, среднее 1, стандартное отклонение 1.

Графическим представлением результата кластеризации являются *дендрограмма* (древовидная диаграмма) и *диаграмма рассеивания*.

Пример. На основе проведенного в 2009 году опроса горожан, попытаемся классифицировать «потенциальных предпринимателей», т.е. тех, кто выразил желание открыть свое дело⁷². Определим признаки, по которым будет вестись кластеризация: «Пол», «Наличие возможности заниматься МП», «Представление о материальном положении малых предпринимателей», «Отношение к развитию малого бизнеса» и «Желание открыть свое дело». Все переменные стандартизованы z-преобразованием. В качестве расстояния выбран квадрат евклидова расстояния. Методом объединения кластеров выбран метод среднего расстояния между кластерами. На дендрограмме четко выражены три кластера (рис. 27):

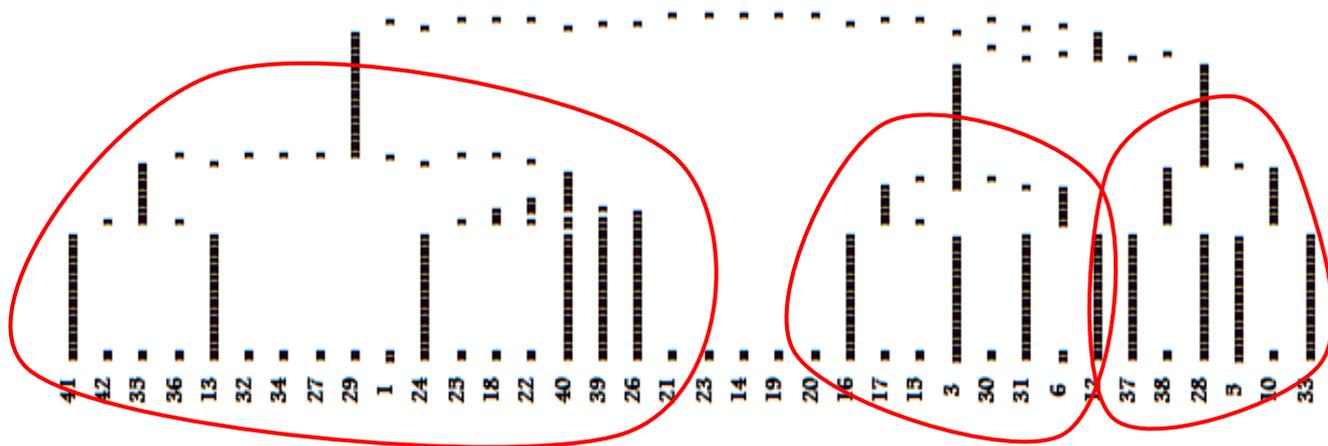


Рис. 27

Гуц А.К., Фролова Ю.В. Математические методы в социологии / предисл. Г.Г. Малинецкого. – М. : Издательство ЛКИ, 2007. – 216 с.

⁷² Опрос в г. Райчихинске, июль 2009 г., ЦСИ АмГУ.

В дальнейшем анализе составим краткий социальный портрет «потенциального предпринимателя» из каждого кластера.

Первый кластер: есть возможность и желание заниматься малым предпринимательством (МП). Данный кластер составляют мужчины среднего возраста со средним доходом, считающие, что малое предпринимательство в России должно развиваться свободно и беспрепятственно, положительно относящиеся к МП и малым предпринимателям. Условно, их можно назвать «энтузиастами».

Второй кластер: нет возможности, но есть выраженное желание заниматься МП. Данный кластер составляют как мужчины, так и женщины всех возрастов (с преобладанием молодёжи 18-24 лет), в основном со средним специальным образованием, доходом средним или ниже среднего, считающие, что МП должно развиваться сдержанно (с участием государства как регулятора), в целом положительно относящиеся к МП и малым предпринимателям. Условно, назовём их «смелыми потенциальными предпринимателями».

Третий кластер: нет возможности, но есть невыраженное желание заниматься МП. Последний кластер также составляют как мужчины, так и женщины всех возрастов со средним или средним специальным образованием и доходом средним или ниже среднего, очень высоко оценивающие доходы и материальное положение малых предпринимателей, считающие, что МП должно развиваться сдержанно (с участием государства как регулятора), в целом положительно относящиеся к МП и малым предпринимателям. Условно, можно назвать этот кластер «осторожными потенциальными предпринимателями».

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

Анализ

Снижение размерности

Факторный анализ... (рис. 28)

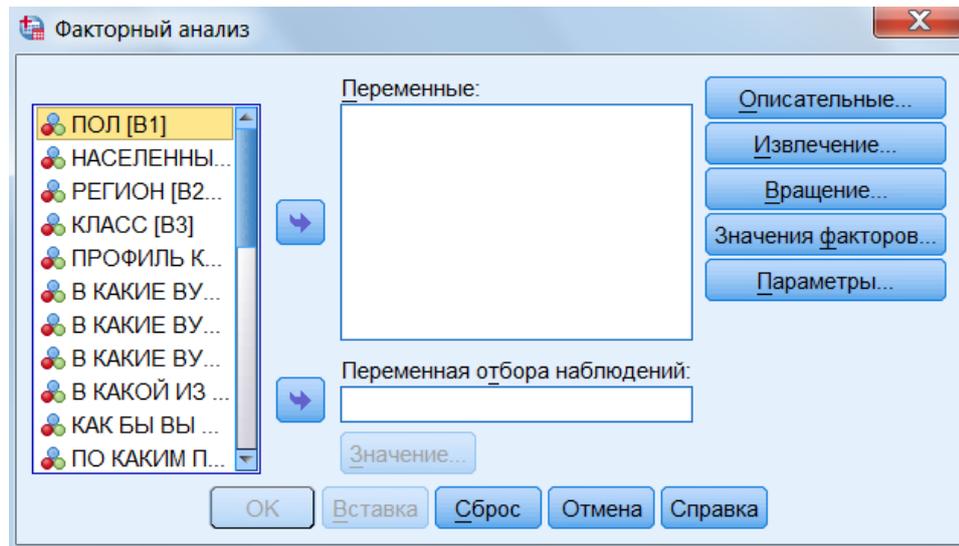


Рис. 28

Извлечение... – позволяет задать метод выделения факторов.

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

Социологический смысл модели факторного анализа состоит в том, что измеряемые эмпирические показатели, переменные считаются следствием других, глубинных, скрытых от непосредственного измерения характеристик – латентных переменных⁷³. *Факторный анализ* позволяет большое число переменных (признаков) свести к меньшему количеству факторов, то есть группирует признаки. За каждой такой группой признаков стоит глубинный латентный фактор. Иными словами, если в кластерном анализе мы группируем строки матрицы данных, в факторном анализе группируются столбцы.

Один из наиболее распространенных методов поиска факторов, *метод главных компонент*, заключается в последовательном поиске факторов. Вначале определяется первый фактор, который объясняет наибольшую часть дисперсии, затем независимый от него второй фактор, объясняющий наибольшую часть оставшейся дисперсии, и т.д.

Интерпретация факторов. Как же можно понять смысл того, что скрыто в найденных факторах? Основной информацией, которую использует для этого

⁷³ 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.

исследователь, являются факторные нагрузки. Для интерпретации необходимо приписать каждому фактору какой-то термин, понятие. Этот термин появляется на основе анализа корреляций фактора с исходными переменными. Например, если при анализе успеваемости школьников фактор имеет высокую положительную корреляцию с оценкой по алгебре, геометрии и большую отрицательную корреляцию с оценками по рисованию, можно предположить, что этот фактор характеризует точное мышление.

Не всегда такая интерпретация возможна. Для повышения интерпретируемости факторов добиваются большей контрастности матрицы факторных нагрузок. Такое улучшение результата называется методом вращения факторов⁷⁴. Его используют для того, чтобы наиболее ярко выявить те факторы, которые наилучшим образом объясняют изучаемые переменные⁷⁵.

Проблема определения числа факторов. Есть несколько методов определения числа факторов. Наиболее простой базируется на самостоятельном отборе числа факторов, ориентируясь на то, чтобы это число факторов объясняло требуемый процент общей исходной дисперсии⁷⁶.

Пример. На основе результатов пилотажного экспресс-опроса жителей Амурской области в июле 2011 г. (ЦСИ АмГУ) попытаемся выделить латентные факторы электорального поведения. Воспользуемся методом главных компонент с вращением факторов, максимизирующем дисперсию квадратов факторных нагрузок. Матрица повернутых компонент (факторов) представлена ниже (таблица 12):

⁷⁴ 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.

⁷⁵ Гуц А.К., Фролова Ю.В. Математические методы в социологии / предисл. Г.Г. Малинецкого. – М. : Издательство ЛКИ, 2007. – 216 с.

⁷⁶ 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.

| Матрица повернутых компонент ^а | | | | | |
|--|-------------|-------------|-------------|-------------|-------------|
| | Компонента | | | | |
| | 1 | 2 | 3 | 4 | 5 |
| НАСКОЛЬКО ВЫ УДОВЛЕТВОРЕНЫ СВОИМ УРОВНЕМ ЖИЗНИ? | -,238 | -,150 | ,183 | ,567 | ,525 |
| КАК ИЗМЕНИЛОСЬ ВАШЕ МАТЕРИАЛЬНОЕ ПОЛОЖЕНИЕ ЗА ПОСЛЕДНИЙ ГОД? | -,072 | -,017 | ,007 | -,050 | ,870 |
| КАК БЫ ВЫ ОЦЕНИЛИ ПОЛИТИЧЕСКУЮ СИТУАЦИЮ В СОВРЕМЕННОЙ РОССИИ? | -,206 | -,173 | ,719 | -,114 | ,024 |
| КАК ВЫ ОЦЕНИВАЕТЕ РАБОТУ ЗАКОНОДАТЕЛЬНОГО СОБРАНИЯ АМУРСКОЙ ОБЛАСТИ? | ,727 | ,093 | -,087 | ,320 | -,328 |
| КАК ВЫ ОЦЕНИВАЕТЕ РАБОТУ ГУБЕРНАТОРА АМУРСКОЙ ОБЛАСТИ? | ,509 | -,103 | ,038 | ,579 | -,189 |
| КАК ВЫ ОЦЕНИВАЕТЕ РАБОТУ ПРАВИТЕЛЬСТВА АМУРСКОЙ ОБЛАСТИ? | ,776 | ,019 | ,012 | ,303 | -,221 |
| КАК ВЫ ОЦЕНИВАЕТЕ РАБОТУ ГОСУДАРСТВЕННОЙ ДУМЫ? | ,776 | -,076 | -,068 | ,150 | -,218 |
| КАК ВЫ ОЦЕНИВАЕТЕ РАБОТУ СОВЕТА ФЕДЕРАЦИИ? | ,849 | -,127 | -,080 | -,014 | -,129 |
| КАК ВЫ ОЦЕНИВАЕТЕ РАБОТУ ПРАВИТЕЛЬСТВА РОССИЙСКОЙ ФЕДЕРАЦИИ? | ,885 | ,061 | -,079 | -,033 | ,128 |
| КАК ВЫ ОЦЕНИВАЕТЕ РАБОТУ ПРЕДСЕДАТЕЛЯ ПРАВИТЕЛЬСТВА РФ? | ,832 | ,254 | -,152 | -,083 | ,186 |
| КАК ВЫ ОЦЕНИВАЕТЕ РАБОТУ ПРЕЗИДЕНТА РОССИИ? | ,797 | ,102 | -,441 | ,094 | ,016 |
| НАМЕРЕНЫ ЛИ ВЫ ПРИНЯТЬ УЧАСТИЕ В ВЫБОРАХ ДЕПУТАТОВ ГОСУДАРСТВЕННОЙ ДУМЫ РФ И ЗАКОНОДАТЕЛЬНОГО СОБРАНИЯ АМУРСКОЙ ОБЛАСТИ, КОТОРЫЕ СОСТОЯТСЯ В ДЕКАБРЕ ЭТОГО ГОДА? | -,093 | ,838 | -,180 | ,093 | -,171 |
| ПРЕДСТАВЬТЕ СЕБЕ, ЧТО В СЛЕДУЮЩЕЕ ВОСКРЕСЕНЬЕ СОСТОЯТСЯ ВЫБОРЫ ДЕПУТАТОВ ГОСУДАРСТВЕННОЙ ДУМЫ РФ И ЗАКОНОДАТЕЛЬНОГО СОБРАНИЯ АМУРСКОЙ ОБЛАСТИ. СКАЖИТЕ, ПОЖАЛУЙСТА, ЗА КАКУЮ ПАРТИЮ ВЫ БЫ ПРОГОЛОСОВАЛИ? | -,052 | ,144 | ,852 | ,007 | ,043 |
| НАМЕРЕНЫ ЛИ ВЫ ПРИНЯТЬ УЧАСТИЕ В ВЫБОРАХ ПРЕЗИДЕНТА РОССИИ, КОТОРЫЕ СОСТОЯТСЯ В МАРТЕ 2012 ГОДА? | ,140 | ,900 | ,145 | -,026 | ,095 |
| ПРЕДСТАВЬТЕ СЕБЕ, ЧТО В СЛЕДУЮЩЕЕ ВОСКРЕСЕНЬЕ СОСТОЯТСЯ ВЫБОРЫ ПРЕЗИДЕНТА РОССИИ. СКАЖИТЕ, ПОЖАЛУЙСТА, ЗА КАКУЮ КАНДИДАТУРУ ВЫ БЫ ПРОГОЛОСОВАЛИ? | ,197 | ,251 | -,336 | ,708 | -,006 |
| Метод выделения: Анализ методом главных компонент. Метод вращения: Варимакс с нормализацией Кайзера. | | | | | |
| а. Вращение сошлось за 9 итераций. | | | | | |

Дальнейший анализ матрицы заключается в следующем: в первую компоненту группируются переменные (вопросы анкеты) имеющие наибольшие веса. Затем аналогично поступают с остальными компонентами (факторами). Таким образом, выделяется 5 латентных факторов, группирующих переменные. Условно им можно дать следующую интерпретацию:

1. Оценка работы органов государственной власти.
2. Готовность участвовать в выборах.
3. Восприятие политической ситуации.
4. Восприятие ключевых политических персоналий.

5. Оценка собственного материального положения.

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.

Гуц А.К., Фролова Ю.В. Математические методы в социологии / предисл. Г.Г. Малинецкого. – М. : Издательство ЛКИ, 2007. – 216 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.

Горшков М.К. Прикладная социология: методология и методы: Учебное пособие / М.К. Горшков, Ф.Э. Шереги. – М.: Альфа-М: ИНФРА-М, 2009. – 416 с.

Контрольные вопросы

1. Каково назначение корреляционного анализа в социологии?
2. Что позволяет получить регрессионная модель?
3. Каково применение кластерного анализа в социологии?
4. В чем назначение факторного анализа и что его объединяет с кластерным анализом?

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.

2.8. Визуализация в SPSS: графики, редактирование таблиц, модуль Tables

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 групп⁷⁷.

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

⁷⁷ 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 графики⁷⁸.

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. 29).

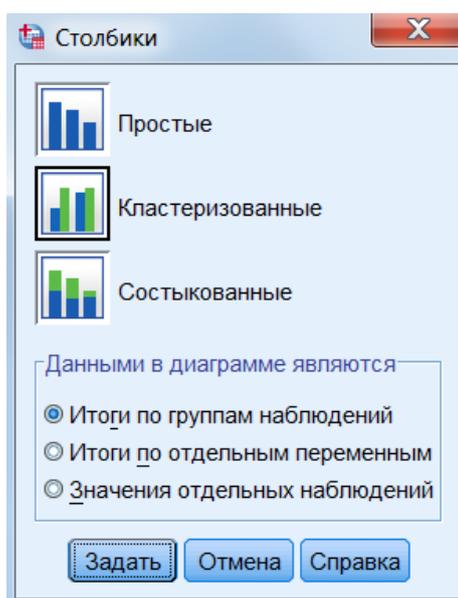


Рис. 29

^ 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

⁷⁸ 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 подмножество⁷⁹.

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. 30).

⁷⁹ Наследов А.Д. SPSS 15: профессиональный статистический анализ данных. – СПб.: Питер, 2008. – С. 98–99

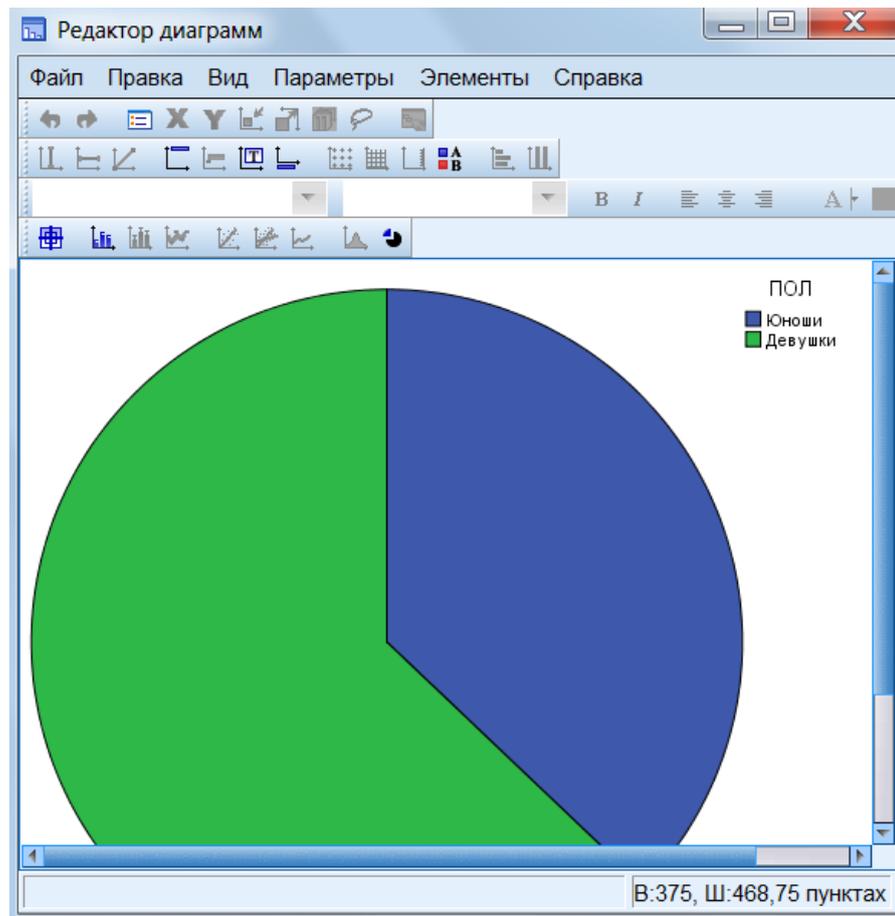


Рис. 30

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. 31).

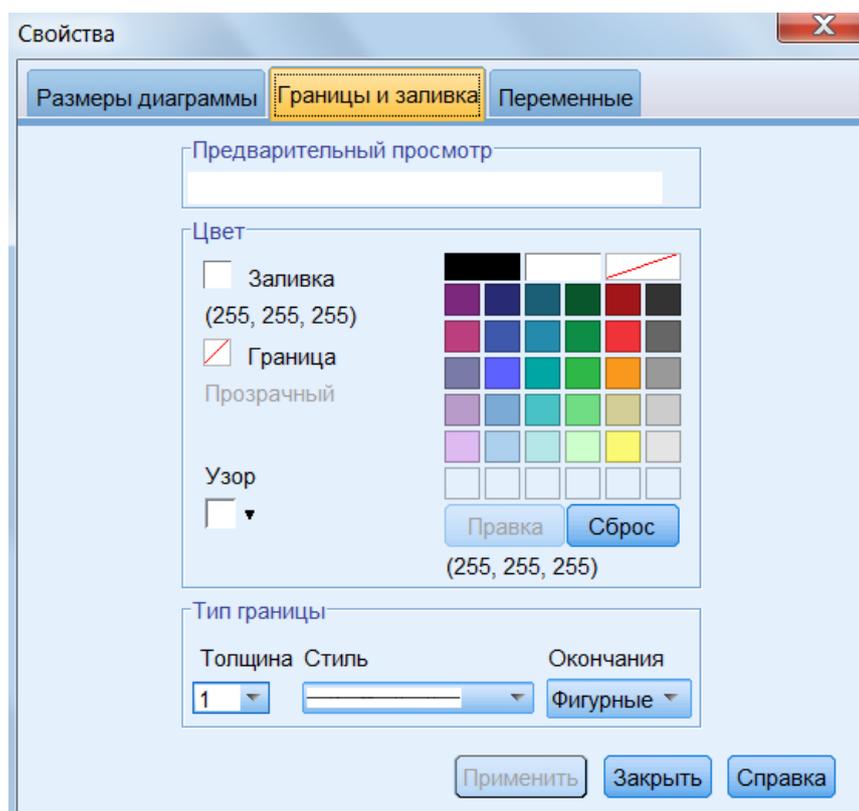


Рис. 31

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 диаграммы⁸⁰.

Редактирование таблиц.

⁸⁰Там же. С. 100–101

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. 32).

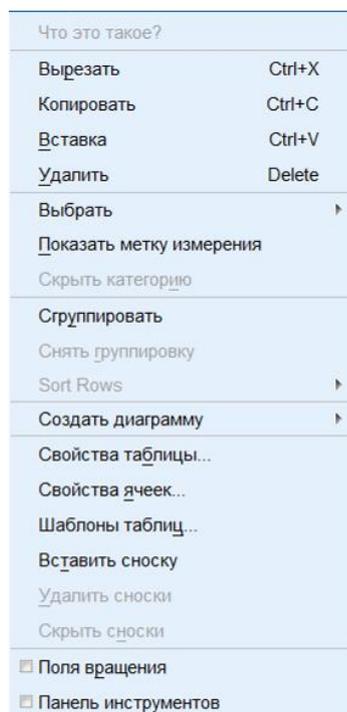


Рис. 32

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

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. 33).

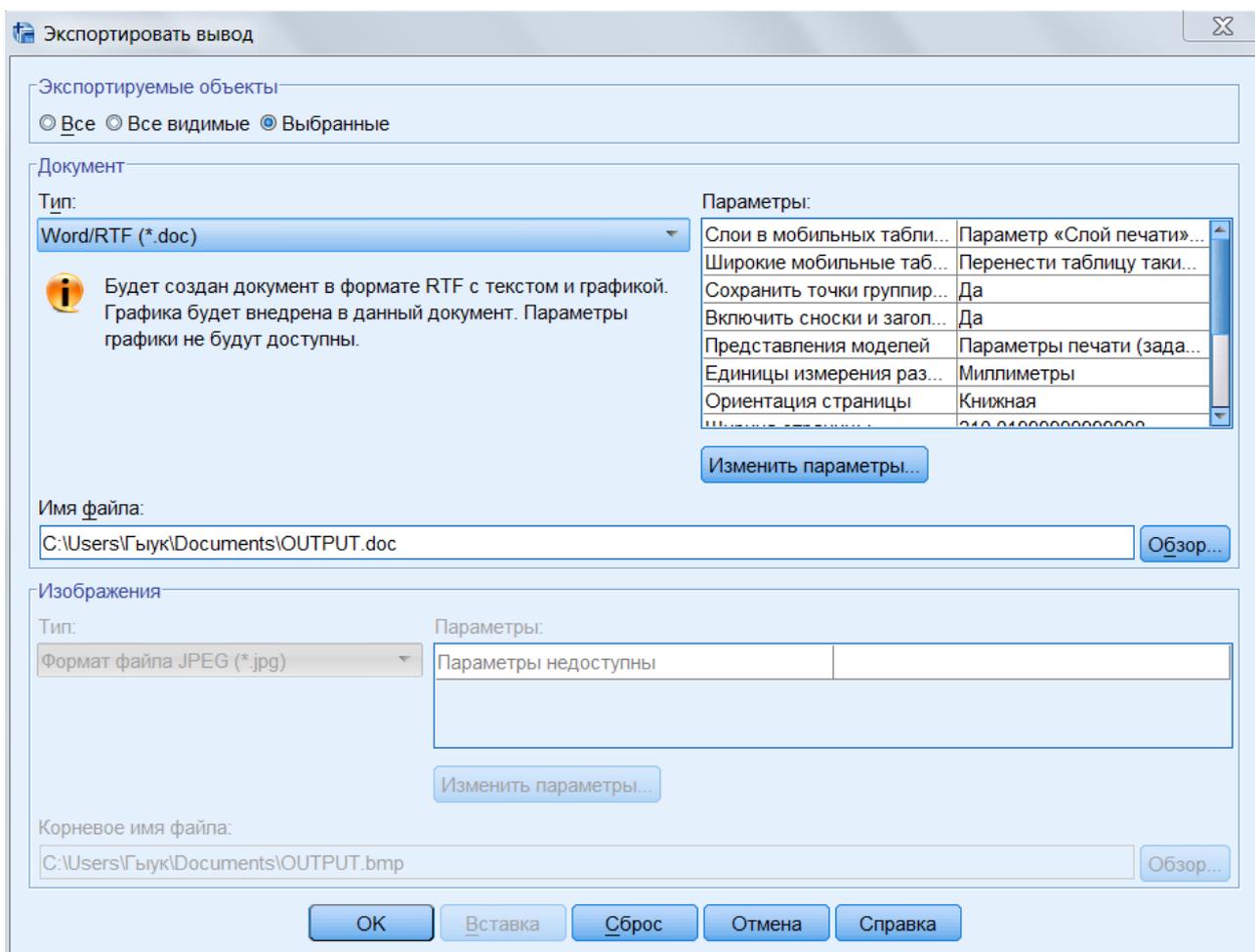


Рис. 33

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*).

Использование модуля Tables.

Модуль Tables служит для создания таблиц, готовых к презентации (рис. 34). По сравнению с режимом построения частотных таблиц и таблиц сопряженности, а также таблиц средних значений, в этом модуле пользователю предоставляются более широкие возможности⁸¹.

Анализ

Таблицы

Настраиваемые таблицы...

⁸¹ 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

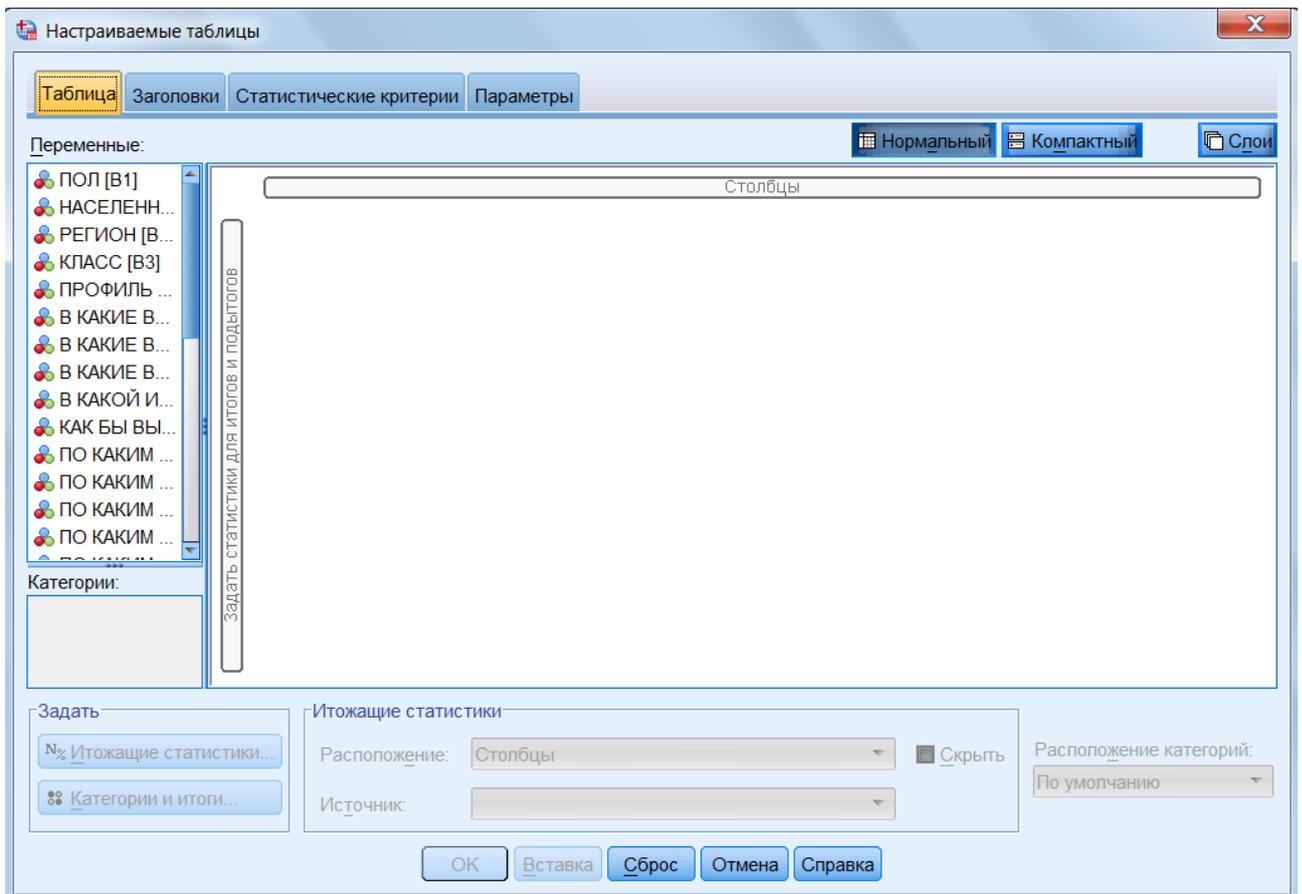


Рис. 34

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 (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. 35):

Анализ

Таблицы

Наборы множественных ответов...

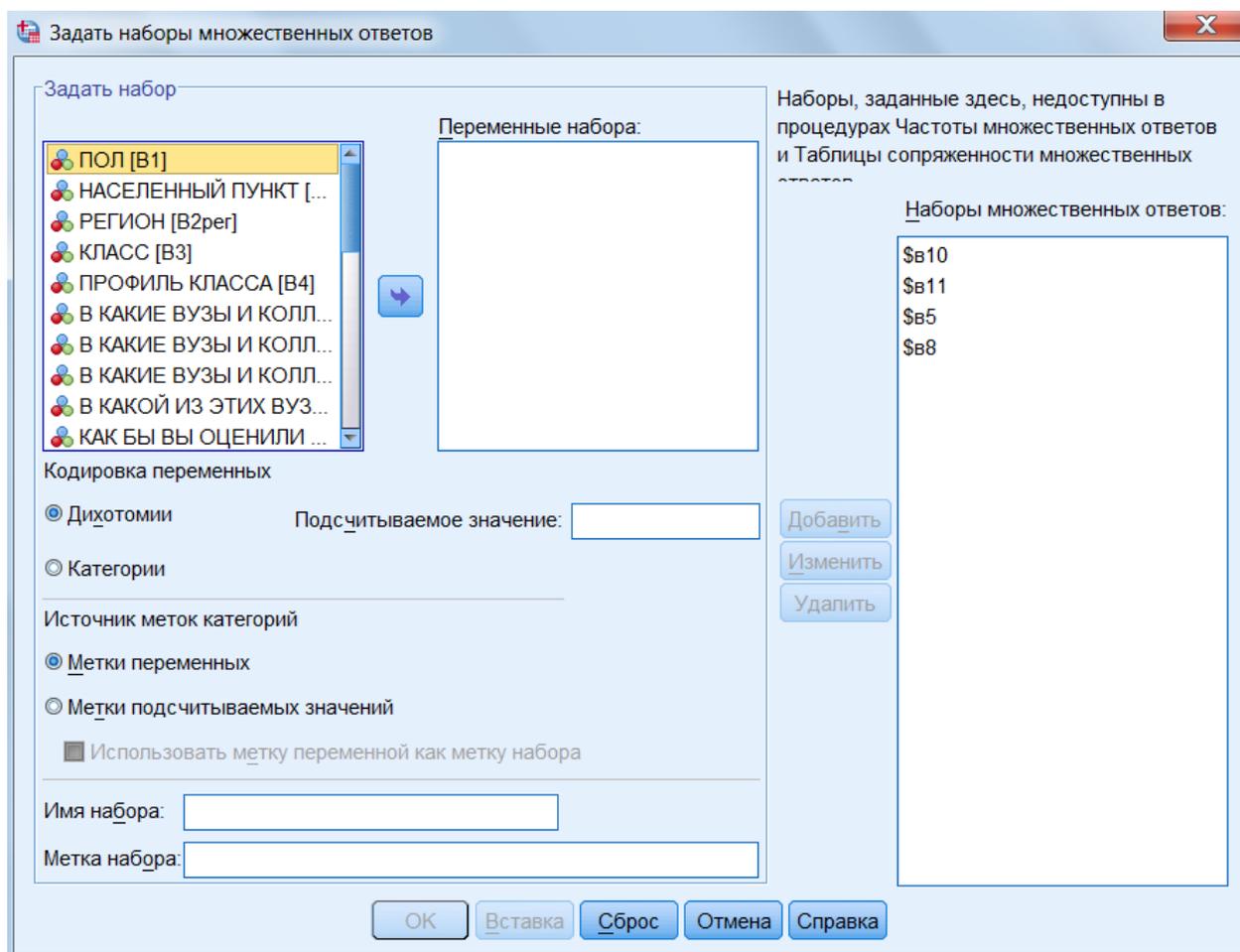


Рис. 35

Затем строится требуемая таблица описанным выше способом.

Модуль Tables также используется для построения сводных таблиц. Фрагмент такой таблице представлен на рис. 36:

| | 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% | |
| | МЭИ | 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. 36

Использованная литература

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.

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. Каковы Виды стандартных графиков?
2. Опишите возможности SPSS редактирования графиков
3. Что такое Интерактивные графики? Каковы их виды?
4. Как вызвать Редактор мобильных таблиц?
5. Как Экспортировать таблицы в Word, Excel?
6. Для чего нужен Модуль **Tables**?

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 обеим переменным;

В) постройте диаграмму рассеивания по обеим переменным;

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

9. 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...*

10. 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.

11. 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.

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

ИТОГОВЫЙ ТЕСТ

Вариант 1

1. Совокупность исследований, основанных на методических и технических приемах и методах сбора, обработки и анализа первичной социологической информации – это...

- А) Общая социология
- Б) Социологические теории среднего уровня
- В) Эмпирическая социология

2. Степень разброса значений признака показывает:

- А) Ряд распределения
- Б) Среднее арифметическое
- В) Дисперсия
- Г) Коэффициенты корреляции

3. Упорядочение данных по одному признаку – это:

- А) Простая группировка
- Б) Перекрестная группировка
- В) Эмпирическая типологизация
- Г) Теоретическая типологизация

4. Соотнесите понятия математической статистики и понятия социологического исследования:

| | |
|-------------------------------|---|
| А) случайная величина | Б) проценты |
| В) случайное событие | Г) одномерное и двумерное распределение |
| Д) вероятность | Е) признак |
| Ж) распределение вероятностей | З) отбор респондентов |

5. Что является первым этапом статистического анализа данных?

- А) описательный анализ
- Б) сравнительный анализ
- В) частотный анализ
- Г) факторный анализ

6. Соотнесите шкалу с ее типом:

| | |
|--|-----------------|
| А) Слушаете ли Вы радио? 1. Да, слушаю 2. Нет, не слушаю | Б) порядковая |
| В) Сколько всего каналов, может принимать Ваш телевизор? _____ | Г) номинальная |
| Д) Как часто Вы пользуетесь Интернетом? 1. Часто 2. Иногда 3. Очень редко | Е) интервальная |

7. Мода – это:

- А) Наиболее часто встречающееся значение признака
- Б) Значение признака, который делит вариационный ряд пополам
- В) Мера разброса признака
- Г) Среднее значение признака

8. Квадрат стандартного отклонения – это...

- А) размах
- Б) мода
- В) медиана
- Г) дисперсия

9. Всестороннее объяснение социологических данных именуется:

- А) группировка
- Б) типологизация
- В) интерпретация
- Г) анализ

10. Если в перекрестной таблице градации признака «пол респондента» расположены в столбцах, то где следует производить расчет процентов?

- А) внутри столбцов
- Б) внутри строк
- В) и внутри столбцов и внутри строк
- Г) это зависит от исследовательских задач и особенностей выборки

Д) в целом по таблице

Дайте обоснование своему ответу _____

13. Соотнесите метод многомерного анализа с его содержанием:

| | |
|--------------------------|---|
| А) Корреляционный анализ | Б) позволяет прогнозировать значение зависимой переменной на основе значения независимой переменной |
| В) Регрессионный анализ | Г) выявление меры зависимости двух или нескольких величин |
| Д) Классификация | Е) позволяет большое число переменных (признаков) свести к меньшему количеству факторов, то есть группирует признаки. |
| Ж) Факторный анализ | З) позволяет классифицировать объекты (респондентов, регионы и т.д.), т.е. объединить их в несколько групп. |

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. Соотнесите тип социологических данных с их содержанием:

| | |
|--|---|
| А) Государственная статистика | Б) Получаются в результате изучения общественного мнения. Эти исследования, как правило, носят описательный характер, и в них используются прямые, простые, однозначно интерпретируемые вопросы и ответы на них |
| В) Данные, полученные с помощью анкет «простой» структуры | Г) Совокупность данных объективного характера, описывающая некий социум в пространстве и во времени |
| Д) Данные полученные посредством анкет «сложной» структуры | Е) Данные, полученные в исследованиях аналитического характера. Инструментарий сбора эмпирических данных в этих исследованиях, как правило, состоит из нескольких частей |
| Ж) Данные об использовании бюджета времени | З) Информация о времяпрепровождении, являются данными о реальном поведении людей |
| И) Текстовые данные | К) Информация при работе с открытыми вопросами анкеты; неформализованные интервью; материалы СМИ |

2. Способ сбора, обработки, анализа социологической информации –

это:

- А) Процедура
- Б) Метод
- В) Техника
- Г) Методика

3. «Философской колыбелью» количественного подхода в социологии

является:

- А) Позитивизм
- Б) Антипозитивизм
- В) Социологизм
- Г) Сциентизм

4. Соотнесите уровень измерения и допустимые при нем меры средней тенденции:

| | |
|-------------------------------|---|
| А) номинальный | Б) мода |
| В) порядковый | Г) мода, медиана, среднее арифметическое |
| Д) интервальный (метрический) | Е) мода, медиана, в ряде случаев среднее арифметическое |

5. Видение того, что за данными стоит, понимание смысла чисел, полученных в результате измерения – это...

- А) конструирование
- Б) интерпретация
- В) изучение

6. Соотнесите задачи социолога при анализе данных с их содержанием:

| | |
|---------------|--|
| А) Описание | Б) процесс раскрытия сущности объектов, охватывающий их структурные, функциональные, генетические, причинно-следственные связи и отношения, тенденции развития, которые отображаются в гипотезах, законах, теориях |
| В) Объяснение | Г) способ научного предвидения с той или иной степенью вероятности итога, направленности, или характера протекания социальных процессов в течение некоторого промежутка времени |

| | |
|----------------------------------|--|
| Д) Типологизация и классификация | Е) метод научного познания, в основе которого лежит обладающее определенными свойствами разбиение изучаемой совокупности объектов на группы |
| Ж) Прогнозирование | З) представляет собой фиксацию результатов эмпирического исследования в соответствующих терминах и ограничивается перечислением некоторых свойств объектов, группировкой по их внешним признакам, установлением наблюдаемых зависимостей |

7. 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

8. При помощи каких методов в описательном анализе решается задача получения общих представлений об изучаемых социальных группах?

- А) построение таблиц частотных распределений
- Б) графическое представление частотных распределений
- В) статистические характеристики распределений
- Г) все варианты верны

9. К мерам средней тенденции относится:

- А) Энтропия
- Б) Медиана
- В) Индекс
- Г) Регрессия

10. Чем отличаются процент и валидный процент? (найдите соответствие)

| | |
|---------------------|---|
| А) Процент | Б) Учитывает только давших содержательный ответ |
| В) Валидный процент | Г) Учитывает всех опрошенных |

11. Связывание данных предварительно упорядоченных по двум признакам (свойствам, переменным):

- А) Простая классификация
- Б) Перекрестная группировка
- В) Эмпирическая типологизация
- Г) Теоретическая типологизация

15. Соотнесите метод многомерного анализа с его содержанием:

| | |
|--------------------------|---|
| А) Корреляционный анализ | Б) позволяет прогнозировать значение зависимой переменной на основе значения независимой переменной |
| В) Регрессионный анализ | Г) выявление меры зависимости двух или нескольких величин |
| Д) Классификация | Е) позволяет большое число переменных (признаков) свести к меньшему количеству факторов, то есть группирует признаки. |
| Ж) Факторный анализ | З) позволяет классифицировать объекты (респондентов, регионы и т.д.), т.е. объединить их в несколько групп. |

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

Методология и методы анализа социологических данных

Васильева Е.Г. Методика и техника конкретного социологического исследования: Учебно-методическое пособие. – Волгоград: Издательство ВолГУ, 2000. – 64 с.

Горшков М.К. Прикладная социология: методология и методы: Учебное пособие / М.К. Горшков, Ф.Э. Шереги. – М.: Альфа-М: ИНФРА-М, 2009. – 416 с.

Готлиб А.С. Введение в социологическое исследование. Качественный и количественный подходы. Методология. Исследовательские практики : учеб. пособие / А.С. Готлиб. – 2-е изд., перераб. и доп. – М.: Флинта : МПСИ, 2005. – 384 с.

Гуц А.К., Фролова Ю.В. Математические методы в социологии / предисл. Г.Г. Малинецкого. – М. : Издательство ЛКИ, 2007. – 216 с.

Девятко И.Ф. Методы социологического исследования. – Екатеринбург: Изд-во Урал. ун-та, 1998. – 208 с.

Казанцев В.И. Методология и методика социологических исследований: Учебно-методическое пособие / В.И. Казанцев, М.Г. Светуньков. – Ульяновск: Средневолжский научный центр, 2003. – 160 с.

Методология социологии. Учебно-методический комплекс для студентов специальности 040201 «Социология» / Проказин В.В. (составитель). – Благовещенск: Амурский гос. ун-т, 2007. – 83 с.

Рабочая книга социолога / Редкол.: Г.В. Осипов (отв. ред.) и др. – 3-е изд. – М.: УРСС, 2003. – 477 с.

Татарова Г.Г. Методология анализа данных в социологии (введение) / Учебник для вузов. – М. : NOTA BENE, 1999. – 224 с.

Толстова Ю.Н. Анализ социологических данных. – М. : Научный мир, 2000. – 352 с.

Толстова Ю.Н. Измерение в социологии: Курс лекций. – М. : ИНФРА-М, 1998. – 224 с.

Ядов В.А. Социологическое исследование: методология, программа, методы – М.: Наука, – 1972.

Ядов В.А. Стратегия социологического исследования. Описание, объяснение, понимание социальной реальности / В.А. Ядов. – 3-е изд., испр. – Москва: Омега-Л, 2007. – 567 с.

Статистическая обработка данных в SPSS

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.

Иллюстрированный самоучитель по SPSS [Электронный ресурс]. URL: www.learnspss.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.

Крыштановский А.О. Анализ социологических данных с помощью пакета SPSS : учебное пособие для вузов / А.О. Крыштановский. – 2-е изд. – М. : Изд. дом ГУ ВШЭ, 2007. – 281 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.

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.

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

Основы применения SPSS в социологии [Электронный ресурс] : учеб. пособие / АмГУ, ФСН ; сост. А. К. Леонов. – Благовещенск : Изд-во Амур. гос. ун-та, 2016. – 154 с. – Точка доступа : http://irbis.amursu.ru/DigitalLibrary/AmurSU_Edition/7337.pdf

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

SPSS: обработка статистической информации [Электронный ресурс]. URL: www.ispss.ru

Леонов Аркадий Константинович
доцент, кандидат социологических наук

Анализ социологических данных: методология, методы, техники в SPSS.
Учебное пособие