

А. Б. Шабденова

МЕТОДЫ ОБРАБОТКИ И АНАЛИЗА
СОЦИОЛОГИЧЕСКОЙ
ИНФОРМАЦИИ НА БАЗЕ SPSS

Учебное пособие

Алматы
«Қазақ университеті»
2023

УДК 316(075.8)

ББК 60.5я73

Ш 12

*Рекомендовано к изданию Ученым советом
факультета философии и политологии (протокол №2 от 8.10.2021)
и РИСО КазНУ им. аль-Фараби (протокол №4 от 24.11.2022)*

Рецензенты:

доктор социологических наук, доцент **Б.Н. Кылышбаева**
(КазНУ им. аль-Фараби)

доктор PhD, ассоциированный профессор **А.С. Беймишева**
(НАО «Университет Нархоз»)

кандидат социологических наук **А.А. Калиева** (КазНацЖенПУ)

Шабденова А.Б.

Ш 12 Методы обработки и анализа социологической информации на базе SPSS: учебное пособие / А.Б. Шабденова. – Алматы: Қазақ университеті, 2023. – 119 с.

ISBN 978-601-04-6137-6

Учебное пособие нацелено на формирование у обучающихся основных навыков по вводу, обработке и анализу данных с использованием возможностей статистического пакета для социальных наук – SPSS (Statistical Package for Social Science). Содержание пособия полностью соответствует программе дисциплины «Методы обработки и анализа социологической информации (на базе SPSS)», изучаемой в высших учебных заведениях, в рамках образовательных программ по социологии.

Предназначено для студентов, обучающихся по образовательным программам различных социальных и гуманитарных наук, где ожидаемыми результатами обучения является обработка и анализ данных в программе SPSS, а также для широкого круга специалистов в научных и исследовательских сферах.

УДК 316(075.8)

ББК 60.5я73

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	4
1. ОСНОВЫ ОБРАБОТКИ И АНАЛИЗА СОЦИОЛОГИЧЕСКОЙ ИНФОРМАЦИИ И ВОЗМОЖНОСТИ SPSS	7
1.1. Обзор программы SPSS: немного истории и возможности	7
1.2. Основные особенности программы и подготовка к работе в SPSS	10
1.3. Ввод данных в SPSS. Типы переменных	16
1.4. Работа с файлами данных. Модификация и преобразование данных... 21	21
Работа с файлами данных	21
Преобразование данных. Вычисление и перекодировка переменных.....	28
2. ОПИСАТЕЛЬНЫЕ СТАТИСТИКИ, МЕРЫ ЦЕНТРАЛЬНОЙ ТЕНДЕНЦИИ И РАЗБРОСА	34
2.1. Частотное распределение переменных	34
2.2. Меры центральной тенденции и разброса	36
2.3. Анализ двумерной зависимости. Таблицы сопряженности	38
3. АНАЛИЗ ВЗАИМОСВЯЗЕЙ КОЛИЧЕСТВЕННЫХ И КАЧЕСТВЕННЫХ ПЕРЕМЕННЫХ	46
3.1. Сравнение средних значений показателей в группах	46
3.2. Одновыборочный Т-критерий	51
3.3. Т-критерий для независимых выборок	53
3.4. Т-критерий для парных выборок	56
3.5. Однофакторный дисперсионный анализ	58
4. АНАЛИЗ ВЗАИМОСВЯЗЕЙ ПЕРЕМЕННЫХ. РЕГРЕССИОННЫЕ МОДЕЛИ	64
4.1. Корреляционный анализ. Парная корреляция	64
4.2. Регрессионный анализ	70
5. АНАЛИЗ СТРУКТУРЫ СОЦИОЛОГИЧЕСКИХ ДАННЫХ	77
5.1. Факторный анализ	77
5.2. Кластерный анализ	89
6. ОБРАБОТКА И АНАЛИЗ МНОЖЕСТВЕННЫХ ОТВЕТОВ	97
7. СИНТАКСИС В SPSS	107
8. ПОСТРОЕНИЕ ГРАФИКОВ В SPSS.....	112
ЗАКЛЮЧЕНИЕ	116
СПИСОК ИСТОЧНИКОВ	118

ВВЕДЕНИЕ

Современный рынок труда требует от специалистов владения различными и многосторонними навыками, в том числе навыками сбора, обработки и анализа статистической информации и результатов исследований. Таким образом, способность анализировать количественную информацию, строить аналитические модели на основе различных видов статистического анализа является актуальным и востребованным навыком. В этом помогают специализированные программы статистической обработки данных, одной из которых является программный пакет SPSS. Знание алгоритмов запуска процедур частотного, корреляционного, регрессионного, факторного, кластерного анализов, а также таблиц сопряженности, мер центральной тенденции и их интерпретации является признаком высокой квалификации и повышает шансы успешного трудоустройства. Настоящее учебное пособие способствует пониманию обозначенных аналитических процедур и усвоению алгоритмов их запуска.

Учебное пособие нацелено на формирование у обучающихся основных навыков по вводу, обработке и анализу данных с использованием возможностей статистического пакета для социальных наук – SPSS (Statistical Package for Social Science). В качестве данных для обработки и анализа в SPSS можно использовать: результаты социологических и других исследований, статистическую информацию, фиксируемую на предприятиях и организациях, данные официальной статистики и т.п.

Содержание учебного пособия полностью соответствует программе дисциплины «Методы обработки и анализа социологической информации (на базе SPSS)», изучаемой в высших учебных заведениях, в рамках образовательных программ по социологии. В результате изучения данной дисциплины обучающийся будет способен: использовать основные принципы ввода и контроля данных в SPSS, определять виды переменных и шкал; обрабатывать и анализировать социологическую информацию, строить модели на основе регрессионного, факторного и кластерного видов анализа, с использованием программного

пакета статистической обработки данных SPSS. Таким образом, обучение по обозначенной дисциплине, включая данное учебное пособие, способствует формированию квалифицированных кадров, конкурентоспособных на рынке труда.

Существуют различные учебные пособия, описывающие методы обработки и анализа данных в программе SPSS, отличительной особенностью данного учебного пособия является то, что автор уделила больше внимания практическим аспектам формирования базы данных, её обработки и запуска процедур анализа, о чём свидетельствуют подробные описания и сопровождающие их наглядные материалы в виде скриншотов описываемых процедур, диалоговых окон и результатов. Вместе с тем, по ходу представления обучающего материала по ряду процедур имеются ссылки на источники литературы с указанием страниц, в которых можно ознакомиться с дополнительной информацией по данной тематике.

Данное учебное пособие основано на методике применения анализа данных результатов социологических исследований. В работе использовалась лицензионная программа IBM SPSS Statistics, 22 версия.

При подготовке данного пособия использованы широко известные работы, написанные в рамках изучаемой области: А.О. Крыштановского «Анализ социологических данных с помощью пакета SPSS» и В.В. Пациорковского и В.В. Пациорковской «SPSS для социологов».

Обратите внимание, что цивилизованная работа в программе SPSS в исследовательских и учебных целях невозможна без приобретения официальной лицензии на использование программы. Со всей необходимой информацией по данному вопросу можно ознакомиться на сайте компании: www.spss.com, также на данном сайте доступна пробная версия SPSS, которую можно бесплатно скачать и протестировать в течение 14 дней.

Компания-производитель постоянно обновляет версии программы SPSS, при этом совершенствуется не только дизайн и интерфейс, но и вводятся новые процедуры, о чём сообщается на их сайте. Несмотря на это, данное учебное пособие знакомит читателей с основными неизменными принципами работы и

процедурами запуска команд, что поможет в работе с различными версиями программы.

Учебное пособие «Методы обработки и анализа социологической информации на базе SPSS» предназначено для студентов, обучающихся по образовательным программам различных социальных и гуманитарных наук, где ожидаемыми результатами обучения являются обработка и анализ данных в программе SPSS. Представляется, что данное учебное пособие вызовет интерес не только у обучающихся по образовательным программам социальных и гуманитарных наук, но и у широкого круга специалистов, поскольку формирование навыков работы в программах статистической обработки является востребованным во многих научных и исследовательских сферах.

1. ОСНОВЫ ОБРАБОТКИ И АНАЛИЗА СОЦИОЛОГИЧЕСКОЙ ИНФОРМАЦИИ И ВОЗМОЖНОСТИ SPSS

1.1. Обзор программы SPSS: немного истории и возможности

Основными методами сбора социологической информации являются качественный и количественный. Качественные методы исследования нацелены на получение спектра мнений и объяснение на основе них социальных явлений. Количественные методы позволяют оценить распространённость изучаемых социальных явлений, а также объяснить их с помощью статистических методов обработки информации и анализа. Для обработки данных обоих методов исследований разработаны программы, позволяющие обработать и проанализировать результаты интервью, фокус групповых дискуссий, различных опросов, наблюдений и т.д. Одной из таких программ для обработки результатов количественных исследований является программа SPSS – Statistical Package for Social Sciences. Несмотря на существование ряда программ для обработки и анализа статистической информации, программа SPSS является наиболее распространённой.

Согласно истории создания программы, хорошо описанной Ахим Бююль и Петер Цёфель, её начало было положено в 1965 году студентами Norman Nie и Dale Bent, которые не нашли удобной программы для анализа количественных данных. Имеющиеся на тот момент программы не обеспечивали удобство и наглядность представления обработанной статистической информации, в результате они решили разработать собственную программу. Далее работа была продолжена в Чикагском университете в 1970 году, к тому моменту было уже несколько версий программы SPSS, а один из её основателей Norman Nie создал соответствующую фирму. Изначально версиям

программы присваивали порядковые номера. Так, до 1981 года существовало 9 версий, работавших с помощью перфокарт на больших ЭВМ. С появлением персональных компьютеров и дальнейшим появлением операционной системы Windows программа SPSS была усовершенствована, первая версия для новой операционной системы имела порядковый номер 5, затем последовали другие версии, 7 версия была доступна для работы в начале 1996 года [1].

По состоянию на июль 2021 года разработана и предлагается для работы 27 версия программы, которая располагает множеством возможностей и помимо обработки и анализа данных позволяет создавать графический и табличный материал с хорошей визуализацией, готовой для представления в отчётах, презентациях и статьях. Новая версия программы SPSS Statistics 27 сопровождается обучающими и демонстрационными материалами, доступными на цифровом ресурсе издателей программы.

Последние версии SPSS Statistics доступны для операционных систем Microsoft Windows и Mac и, как уже отмечено, предоставляют множество возможностей для пользователей. В частности, разработчики программы на своём сайте обозначили следующие возможности SPSS Statistics:

- «Анализ и улучшение понимания данных, решение сложных исследовательских и бизнес-задач с помощью удобного пользовательского интерфейса.

- Быстрый анализ больших и сложных наборов данных с использованием передовых статистических методов для принятия точных, обоснованных решений.

- Использование расширений, языков программирования Python и R для интеграции с открытым программным обеспечением.

- Гибкие варианты развертывания позволяют легко выбирать программное обеспечение и управлять им» [2].

Поработать в последней версии программы SPSS в тестовом режиме и далее приобрести её можно на сайте <https://www.ibm.com/ru-ru/products/spss-statistics>.

В целом с помощью программы SPSS обрабатывают и анализируют доступные статистические данные, результаты различных видов опросов, наблюдений, замеров, тестов, экспериментов, опытов и т.п. Данную программу использует широкий круг исследователей и студентов, специализирующихся в социальных, медицинских, биологических, математических и других науках.

Применительно к социологической информации, чаще всего анализируются результаты опросов, наиболее распространённым из которых является стандартизированный опрос. Стандартизированный опрос – это метод, при котором используется самостоятельно заполненный респондентами анкетный опрос с четко определенными формулировками и порядком вопросов для получения сопоставимости данных, собранных разными интервьюерами.

Следует отметить, что социологическая информация не ограничивается результатами опросов, в арсенале социологов есть различные методы сбора информации, которые можно обработать и проанализировать в программе SPSS. К примеру, к ним относятся, но не ограничиваются: включенное и стороннее наблюдения; замеры посещаемости населением объектов и мероприятий; замеры трафика движений (пешеходного и транспортного); тестирования респондентов на предмет осведомленности или участия; тестирование социальной рекламы, информации, продукции и т.д. Во всех приведённых примерах данные фиксируются в специально разработанных формах или листах наблюдений и сбора информации.

В целом можно обозначить следующие общие возможности программы SPSS:

- Ввод и хранение данных статистики и исследований.
- Возможность использования переменных разных типов (качественные и количественные).
- Описательная статистика.
- Частотное распределение изучаемых признаков.
- Таблицы, графики, таблицы сопряжённости, диаграммы.

– Различные виды статистического анализа (сравнение средних, дисперсионный анализ, регрессионный, корреляционный, факторный и кластерный анализы).

Понятный и дружелюбный интерфейс программы SPSS не требует специальных знаний программирования или компьютерных навыков. Данная программа доступна для освоения любому пользователю компьютеров, что способствовало её распространению и популярности среди студентов и исследователей разных стран.

Контрольные вопросы:

1. Приведите примеры данных, которые можно обрабатывать и анализировать в программе SPSS.
2. Объясните разницу между качественными и количественными методами сбора социологической информации.
3. Результаты каких социологических исследований следует обрабатывать и анализировать в статистических программах?
4. Какие из возможностей программы SPSS вы бы использовали в своей самостоятельной работе и почему?
5. Объясните цель обработки данных в программе SPSS.

1.2. Основные особенности программы и подготовка к работе в SPSS

Основной особенностью программы SPSS является то, что данные вводятся в формате цифр. Текстовую информацию также можно вводить в эту программу, но она не может быть использована в статистическом анализе. Текстовую информацию можно обработать и получить описательные статистики в виде частотного распределения (рисунок 1) или таблиц сопряженности (рисунок 2). Представленные рисунки текстовой переменной показывают, что любые несоответствия в текстовой информации разделяют данные на отдельные категории ответов. К примеру, отвечая на открытый вопрос «Какими характеристиками должно обладать идеальное место работы?», респонденты говорили о «близком расположении к дому», при этом интервьюеры по-разному записали этот ответ, в результате получилось пять вариантов одного ответа. Поэтому, если работать с текстовой

переменной, необходимо проверить ответы и при необходимости скорректировать их, привести к единому формату.

Для проведения дальнейшего статистического анализа, т.е. других видов анализа, помимо описательной статистики, текстовую информацию необходимо перекодировать в цифровую, создав новую переменную, данная процедура будет показана в последующих темах.

→ Частоты

Статистика

Другое

N	Допустимо	500
	Пропущенные	0

Другое

	Частота	Проценты	Процент допустимых	Накопленный процент
Допустимо	483	96,6	96,6	96,6
Ближе к дому	1	,2	,2	96,8
близ дома	1	,2	,2	97,0
близко к дому	1	,2	,2	97,2
близость к дому	1	,2	,2	97,4
Близость к дому	3	,6	,6	98,0
по договору работа	1	,2	,2	98,2
удобный график	2	,4	,4	98,6
Удобство расположения	1	,2	,2	98,8
хорошая зарплата	3	,6	,6	99,4
Честный и справедливый руководитель	2	,4	,4	99,8
экология	1	,2	,2	100,0
Всего	500	100,0	100,0	

Рисунок 1. Частотное распределение текстовой переменной

Таблицы сопряженности

Комбинационная таблица Другое * Пол респондента

		Пол респондента		Всего
		Мужской	Женский	
Другое	Количество	226	257	483
	% в Пол респондента	96,6%	96,6%	96,6%
Ближе к дому	Количество	1	0	1
	% в Пол респондента	0,4%	0,0%	0,2%
близ дома	Количество	1	0	1
	% в Пол респондента	0,4%	0,0%	0,2%
близко к дому	Количество	1	0	1
	% в Пол респондента	0,4%	0,0%	0,2%
близость к дому	Количество	0	1	1
	% в Пол респондента	0,0%	0,4%	0,2%
Близость к дому	Количество	2	1	3
	% в Пол респондента	0,9%	0,4%	0,6%
по договору работа	Количество	1	0	1
	% в Пол респондента	0,4%	0,0%	0,2%
удобный график	Количество	0	2	2
	% в Пол респондента	0,0%	0,8%	0,4%
Удобство расположения	Количество	0	1	1
	% в Пол респондента	0,0%	0,4%	0,2%
хорошая зарплата	Количество	0	3	3
	% в Пол респондента	0,0%	1,1%	0,6%
Честный и справедливый руководитель	Количество	1	1	2
	% в Пол респондента	0,4%	0,4%	0,4%
экология	Количество	1	0	1
	% в Пол респондента	0,4%	0,0%	0,2%
Всего	Количество	234	266	500
	% в Пол респондента	100,0%	100,0%	100,0%

Рисунок 2. Таблицы сопряженности текстовой переменной в разрезе номинальной переменной

Таким образом, для обработки и анализа данных в программе SPSS необходимо подготовить данные в формат цифр. Если данные собраны в виде текстовых ответов, к примеру, как они представлены в ответах в Гугл формах, то их следует перекодировать в цифры, поскольку это программа для статистического анализа и обрабатывает и анализирует преимущественно информацию, представленную в цифровом формате.

Для работы в программе SPSS существует несколько подходов формирования баз данных:

– База данных формируется автоматически в формате SPSS, Excel и т.д., если опрос проводится в специально разработанных программах, приложениях по сбору данных.

– Данные вводятся непосредственно в программу SPSS.

– Данные вводятся в специально разработанных программах ввода или в Excel, затем экспортируются в программу SPSS.

После запуска программы SPSS открывается редактор данных, состоящий из двух окон:

– В окне «Представление Данные» (рисунок 3) формируется непосредственно база данных, вводятся значения результатов опросов, статистические данные и т.п. По строкам (пронумерованы программой с левой стороны) вводится каждая единица анализа, это каждая отдельная часть объекта социологического исследования (респонденты) или единица наблюдения (семья, группа). По столбцам представлены переменные, это характеристики и изучаемые признаки, к примеру, гендер, возраст, уровень образования, отношение к изучаемой проблеме, уровень удовлетворенности и т.д. Переменные и их метки значений вводятся в следующем окне «Представление Переменные».

– В окне «Представление Переменные» (рисунок 4) формируется макет базы данных, а именно вводятся имена переменных, сами переменные (вопросы анкеты, социально-демографические признаки, различные характеристики объектов исследования), значения ответов, задаётся тип переменной. Также при необходимости здесь можно задать пропущенные значения ответов, т.е. не включать в анализ определённые варианты ответов.

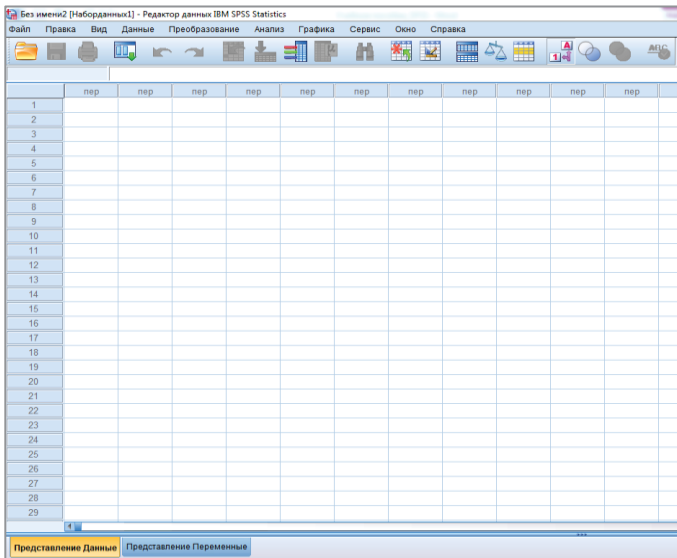


Рисунок 3. Исходный вид программы, редактор данных –
Представление Данные

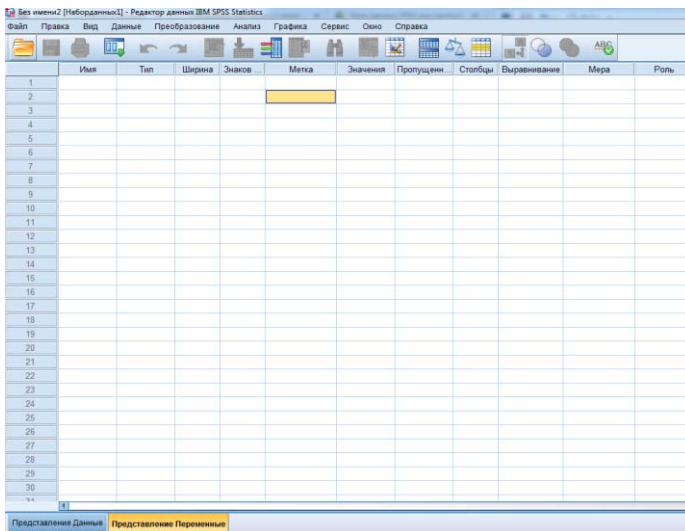


Рисунок 4. Исходный вид программы, редактор данных –
Представление Переменные

Все результаты обработки и анализа базы данных отражаются в файле «Вывод», который формируется программой автоматически (рисунок 5).

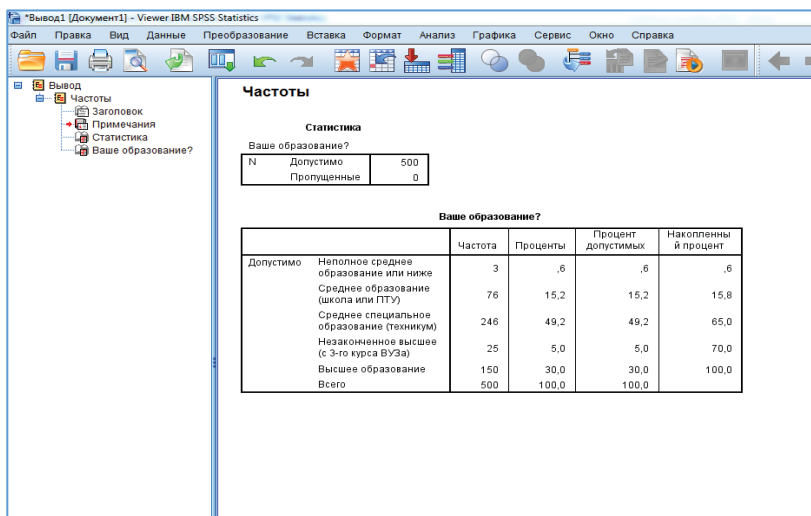


Рисунок 5. Исходный вид программы – файл «Вывод», где отражаются все результаты обработки и анализа данных

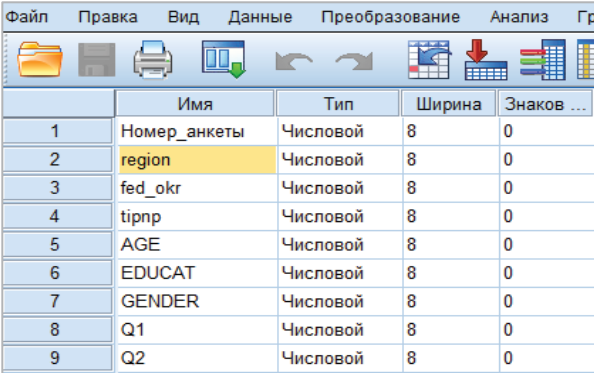
Контрольные вопросы:

1. Приведите примеры единиц анализа для обработки в программе SPSS.
2. Приведите примеры переменных.
3. В каких случаях вводятся текстовые переменные? Приведите примеры.
4. Перечислите требования к текстовым переменным для их обработки и анализа описательной статистики.
5. Изучите ресурс «Google формы» для создания своего опроса. Каким образом будет сформирована база данных для обработки и анализа в программе SPSS?

1.3. Ввод данных в SPSS. Типы переменных

Формирование базы данных в программе SPSS лучше начать с формирования макета, для этого переходим на лист Представление Переменные, продемонстрированный на рисунке 4, где проходит работа с переменными.

Начинаем формировать макет базы данных с имени переменной, здесь необходимо начинать с букв, недопустимо начинать с цифры или вставлять пробелы, вместо последних можно использовать нижнее подчёркивание. Следует отметить, что в старых версиях программы SPSS имена переменных воспринимаются только латиницей, кириллицей возможно называть переменные начиная с 19-20 версий программы SPSS. Имена переменных обозначают либо по содержанию, например, возраст называют AGE, образование – EDUCAT, либо по номеру вопроса – Q1, Q2 или Вопрос_1, Вопрос_2 и так далее (рисунок 6). Любую базу данных рекомендуется начинать с нумерации, что позволит идентифицировать каждую анкету или респондента, поэтому первой переменной является «Номер анкеты».



	Имя	Тип	Ширина	Знаков ...
1	Номер_анкеты	Числовой	8	0
2	region	Числовой	8	0
3	fed_okr	Числовой	8	0
4	tipnr	Числовой	8	0
5	AGE	Числовой	8	0
6	EDUCAT	Числовой	8	0
7	GENDER	Числовой	8	0
8	Q1	Числовой	8	0
9	Q2	Числовой	8	0

Рисунок 6. Примеры имён переменных

Далее определяем тип переменной, чаще всего стоит «числовой» тип, при необходимости можно выбрать другой тип переменной из предлагаемого списка (рисунок 7).

Следующий столбец макета показывает ширину переменной, здесь по умолчанию стоит 8, данной ширины вполне достаточно для числовых переменных (рисунок 6). В случае текстовых переменных необходимо расширить ширину до необходимого числа символов вводимой текстовой информации, что можно сделать либо в столбце «Ширина» макета, либо при выборе типа переменной в соответствующем окошке (рисунок 7).

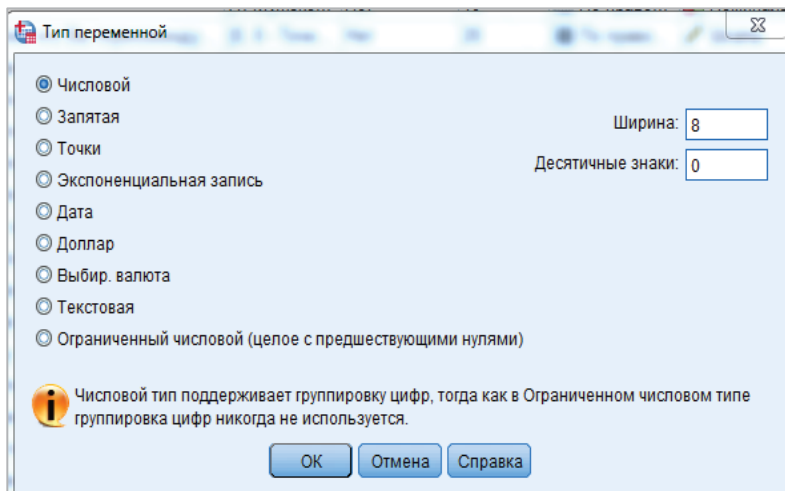


Рисунок 7. Окно для выбора типа переменных

Следующий столбец «Знаков после запятой» соответственно регулирует число знаков после запятой. Обычно по умолчанию стоит 2, т.е. возможно ввести сотые доли числа. Чаще всего в социологических исследованиях используются целые числа, без десятых и сотых долей, поэтому в данном пункте можно проставить 0 по всему макету базы данных (рисунок 6).

В разделе «Метка» следует более подробно расшифровать имя переменной или вставить вопрос анкеты целиком, здесь следует отметить, что слишком длинный текст вопросов не вмещается в строку, соответственно такие вопросы необходимо сократить (рисунок 8).

Метка
Номер анкеты
Регион опроса
Федеральный округ
Тип населенного пункта
Сколько полных лет Вам исполнилось?
Ваше образование?
Пол респондента
Насколько вероятно, что Вы порекомендуете своим друзьям и знакомым устроится на работу в ту ор...
Почему Вы поставили именно такую оценку?
Насколько вероятно, что Вы смените по собственному желанию место работы в ближайшие 6 месяц...
Какие из перечисленных проблем в наибольшей степени волнуют Вас на текущем месте работы?
Какие из перечисленных проблем в наибольшей степени волнуют Вас на текущем месте работы?
Какие из перечисленных проблем в наибольшей степени волнуют Вас на текущем месте работы?

Рисунок 8. Метки переменных

В следующем разделе «Значения» вводятся значения цифр переменной (рисунок 9). К примеру, в переменной под именем EDUCAT, в столбце «Метка» обозначаем «Уровень образования респондентов», далее в графе «Значения» обозначаем, что цифра 1 означает «Неполное среднее образование или ниже», цифра 2 означает «Среднее образование» и т.д., наибольшая цифра 5 означает «Высшее образование».

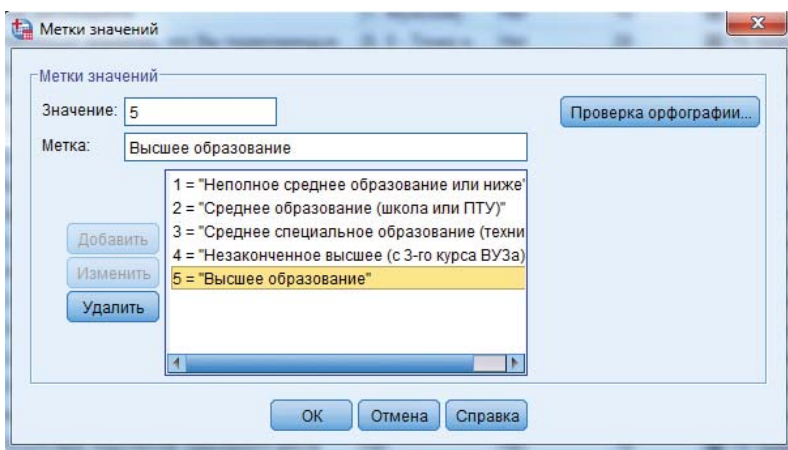


Рисунок 9. Окно для ввода значений переменных

Столбец «Пропущенные» предназначен для обозначения пропущенных значений (рисунок 10). Например, можно исключить из анализа ответы «Затрудняюсь ответить» и «Отказ от ответа», закодированные, как 98 и 99, обозначив их по отдельности в графах подраздела «Отдельные пропущенные значения», всего допускается отметить до трёх отдельных пропущенных значений.

Либо можно задать диапазон пропущенных значений, например, при анализе только молодёжи и людей старшего возраста. Так, например, для количественной переменной «Возраст» обозначим пропущенными значениями минимум 29 лет и максимум 59 лет, тогда данная возрастная группа людей среднего возраста будет исключена из анализа, будут анализироваться только молодые люди до 29 лет и люди более старшего возраста от 60 лет и старше.

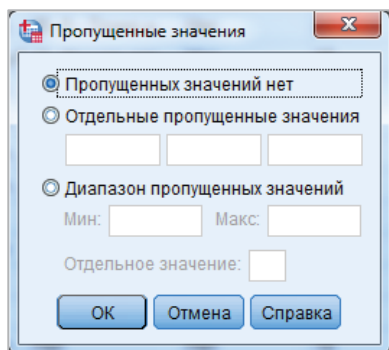


Рисунок 10. Окно для обозначения пропущенных значений переменных

Следующие столбцы технического характера, которые редактируют базу данных, можно настроить по своему желанию или требованию.

В столбце «Мера» обозначаем тип переменной в соответствии с мерой измерения (рисунок 11).

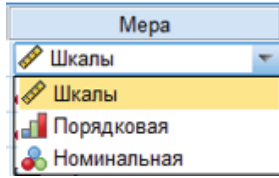


Рисунок 11. Окно для обозначения меры переменных

Обзор литературы по SPSS показывает несколько классификаций переменных, обобщая их, можно выделить два основных типа переменных: качественные и количественные. С качественными переменными невозможно провести математические операции и вычисления, поскольку в них цифры обозначают лишь разделение на группы или ответы. Количественные переменные представляют возможность для различных математических операций, поскольку в них цифры имеют значение.

Внутри обозначенных выше двух типов также есть разделение. Так, среди качественных существует два основных вида переменных:

- Номинальные, где за цифрами нет никакой математической значимости, они обозначают кодирование какой-либо характеристики или ответа. К примеру, к таким переменным относится гендер, где под 1 представлены мужчины, а под цифрой 2 – женщины, можно поменять кодировку местами или закодировать другими цифрами.

- Порядковые, где ответы ранжированы в определённом порядке. К примеру, переменная «уровень удовлетворенности» каким-либо аспектом, где под цифрой 1 значит ответ «полностью не удовлетворены», под 2 – «скорее не удовлетворены» и так далее, данный порядок также можно поменять местами, т.е. начать с 5. Таким образом, цифрам кодов присваиваются значения ответов, имеющих определённый порядок, но мы не можем утверждать, что между этими категориями равные промежутки, т.е. разница между 1 и 2 не такая же, как между 3 и 4.

В количественных переменных (на рисунке 11 они обозначены «Шкалы»), в отличие от порядковых, между цифрами есть равные промежутки, например, температура: мы можем

обозначить, что 10 градусов больше 5 градусов, так же, как и 15 больше 10 градусов, в обоих случаях разница составляет 5 градусов. Среди количественных переменных существует разделение на следующие типы переменных:

– Дискретные, когда между целыми числами существуют промежутки, например, количество членов домохозяйства, количество детей, количество страниц.

– Непрерывные переменные, когда между любыми категориями могут находиться ещё значения, например, рост (170 сантиметров и 5 миллиметров), вес (60 килограмм, 200 грамм и 50 миллиграмм).

– Интервальная переменная, характерной особенностью данной переменной является ее относительность, т.е. отсутствие абсолютного нуля, начала координат, например, коэффициент интеллекта или температура пациентов.

Шкала отношений, к этой шкале можно отнести интервальные переменные, которые имеют абсолютную нулевую точку. Например, возраст респондентов.

Контрольные вопросы:

1. Приведите примеры допустимых обозначений имён переменных в программе SPSS.
2. Объясните, с чего следует начинать базу данных.
3. Приведите примеры номинальных переменных и объясните, почему они являются номинальными.
4. Приведите примеры порядковых переменных и объясните, почему они являются порядковыми.
5. Приведите примеры количественных переменных и объясните, почему они являются количественными.
6. Создайте макет базы данных программе SPSS для анкеты своего опроса.

1.4. Работа с файлами данных.

Модификация и преобразование данных

Работа с файлами данных

Сформировав макет базы данных, описанный в предыдущем разделе 1.3, вводим сами данные (результаты опросов, статистику и т.п.) в окно Представление Данные, показанный на рисунке 3.

Данные для обработки и анализа вводятся непосредственно в программу SPSS или экспортируются из других программ, подробно об этом описано в разделе 1.2.

Создав базу данных в SPSS, также можно экспортировать их из данной программы в другие программы, к примеру, Excel. Различные работы с файлами данных, такие, как создать новую базу, открыть, сохранить в другом файле, экспортировать и т.п., реализуются в меню Файл (рисунок 12).

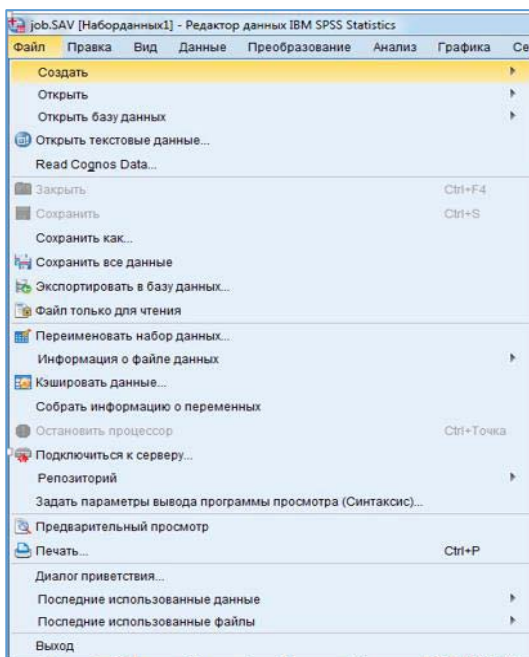


Рисунок 12. Меню «Файл» для работы с файлами данных

Следующее меню «Правка» позволяет производить различные технические операции с базой данных. К примеру, вставлять переменные или наблюдения, задавать технические параметры файлов (рисунок 13).

Меню «Правка» отличается для файла «Вывод» (рисунок 13.1) тем, что в нём есть другие операции, например, можно

выбрать все примечания, команды, заголовки и удалить их перед экспортом файла, в итоге получается файл распределений без лишней информации.

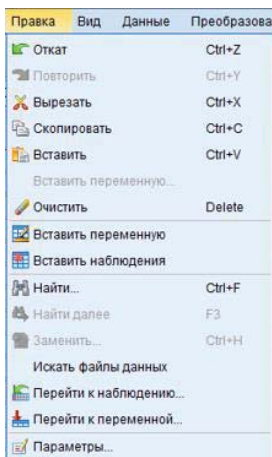


Рисунок 13. Меню «Правка» для редактирования файла базы данных

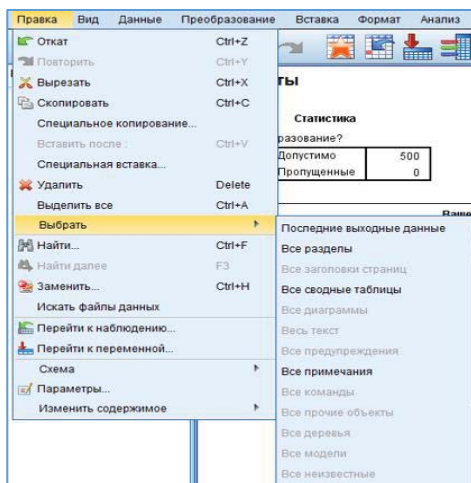


Рисунок 13.1. Меню «Правка» для редактирования файла «Вывод»

Следующее меню «Вид» предназначено для редактирования панели меню, шрифтов и других панелей, также позволяет переходить с окна данных на окно переменных (рисунок 14).



Рисунок 14. Меню «Вид» для технического редактирования файла

В меню «Данные» представляется возможность работать с самими данными, сортировать наблюдения, переменные, соединять файлы, реструктурировать их, проводить процедуру взвешивания данных (рисунок 15).

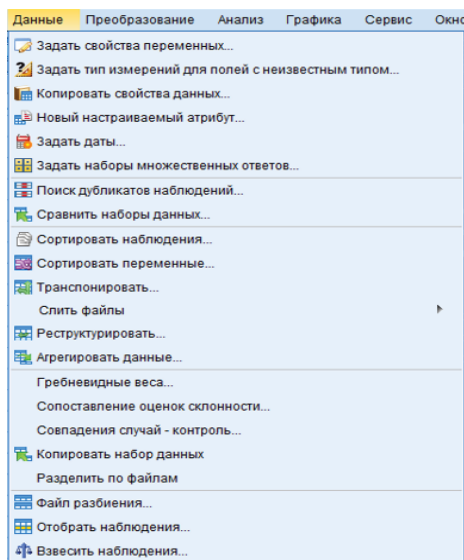


Рисунок 15. Меню «Данные» для работы с базой данных

Также есть возможность разбиения вывода данных по какому-либо признаку, к примеру, если необходимо получить распределения или анализ в разрезе каждого региона, то выбираем подраздел «Файл разбиения», в нём отмечаем «Организовать вывод по группам», выбираем переменную «Регионы» и стрелочкой перемещаем в окно «Группы образуются по:», далее нажимаем кнопку «ОК» (рисунок 16). Теперь данные будут проанализированы в разрезе выбранной переменной и вывод анализируемых данных будет сделан по каждому региону, обозначенному в переменной «Регионы».

Обратите внимание, что по каждой процедуре есть кнопка «Справка», нажав на которую, можно получить информацию о самой процедуре, её предназначении, алгоритме действий. Данная кнопка всегда расположена крайней справа, внизу открытого окна процедуры (рисунок 16).

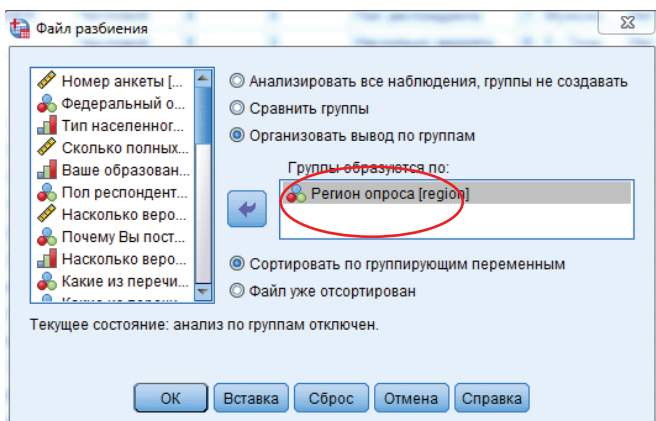


Рисунок 16. Пример работы с файлом разбиения

Рассмотрим ещё одну распространённую процедуру в меню «Данные» по работе с базой данных. Когда необходимо проанализировать только одну или несколько определённых групп исследования, можно использовать процедуру «Отобразить наблюдения». Открыв это окно, можно увидеть, что по умолчанию стоит «Все наблюдения», т.е. обрабатываются и анализируются все наблюдения (рисунок 17).

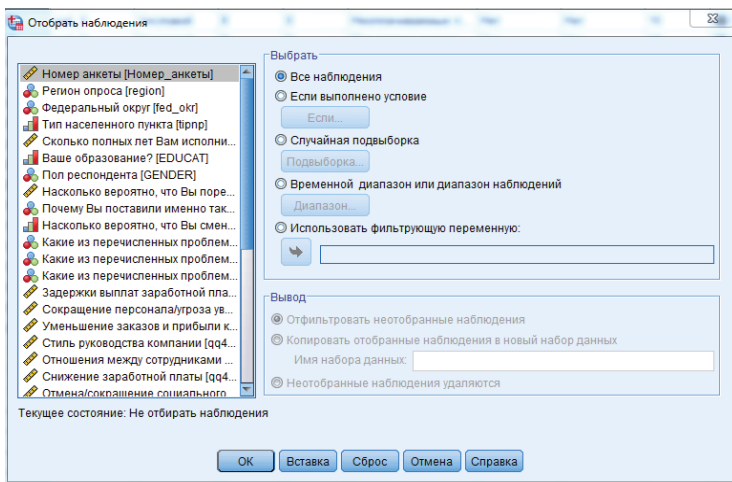


Рисунок 17. Процедура отбора наблюдений

Допустим, перед нами стоит задача проанализировать только женщин в возрасте до 35 лет. Для этого выбираем опцию «Если выполнено условие», активируется опция «Если», заходим в неё, выбираем переменную GENDER, фиксирующую гендер респондентов, переносим её в окно условия и прописываем условие GENDER=2, поскольку женщины у нас закодированы кодом 2. Далее ставим знак объединения «&» и прописываем второе условие, для этого переносим переменную возраста под именем AGE и ставим условие <35 (рисунок 18). Далее нажимаем кнопку «Продолжить» и «ОК».

В результате включен фильтр, о чём программа информирует в нижней правой строке, и в окне «Представление Данные» часть строк слева перечёркнута линией (рисунок 19). Далее в обработку и анализ будут включены только женщины в возрасте до 35 лет.

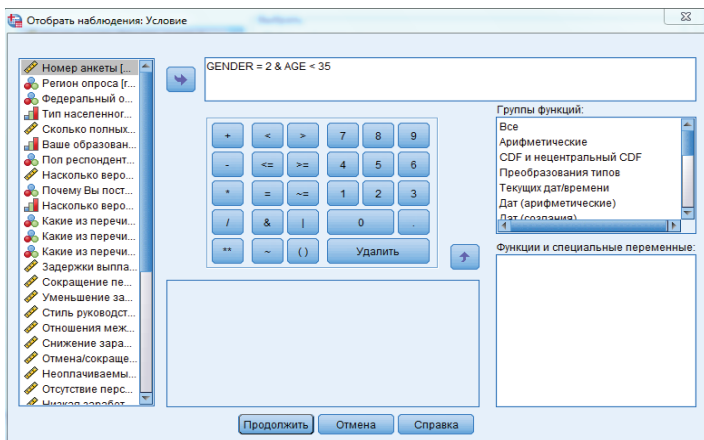


Рисунок 18. Пример составления условия для отбора наблюдений

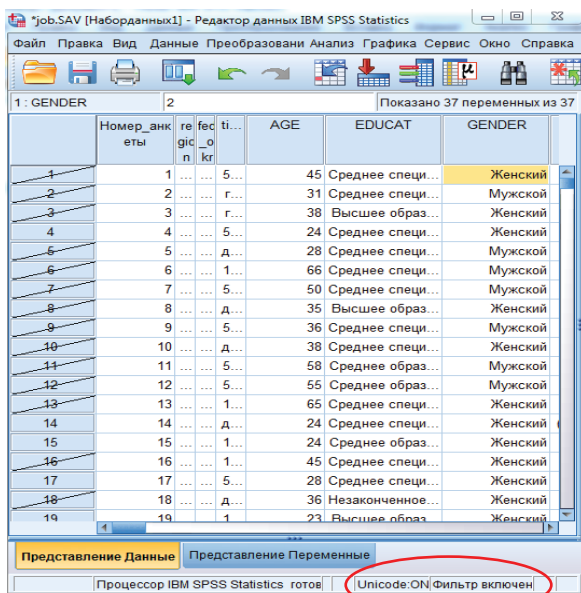


Рисунок 19. Результат отбора наблюдений

Таким образом, можно прописывать различные условия для отбора необходимых целевых групп и далее анализировать

только их. После анализа отобранных групп и перехода к анализу всей выборки не забывайте отключить фильтр. Для этого необходимо зайти в меню «Данные», «Отобразить наблюдения» и выбрать опцию «Все наблюдения».

Контрольные вопросы:

1. Сформируйте базу данных в программе SPSS на основе проведенного Вами опроса, при этом используйте ввод данных в самой программе.
2. Экспортируйте базу данных из Excel в SPSS.
3. Объясните, для чего используется процедура «Файл разбиения».
4. Объясните, для чего используется процедура «Отбор наблюдений».
5. Объясните, в чём разница между двумя процедурами «Файл разбиения» и «Отбор наблюдений».

Преобразование данных.

Вычисление и перекодировка переменных

Следующее меню «Преобразование» также предназначено для работы с базой данных. Здесь расположены процедуры для модификации и преобразования данных. Можно вычислить новую переменную, перекодировать, ранжировать наблюдения и т.д. (рисунок 20).

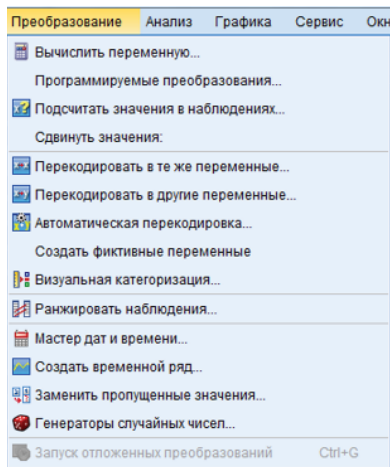


Рисунок 20. Меню «Преобразование» для работы с базой данных

Рассмотрим наиболее часто применяемые процедуры по модификации данных: вычислить переменную и перекодировать. Выбирая процедуру «Вычислить переменную», открывается окно для введения числовых выражений (рисунок 21).

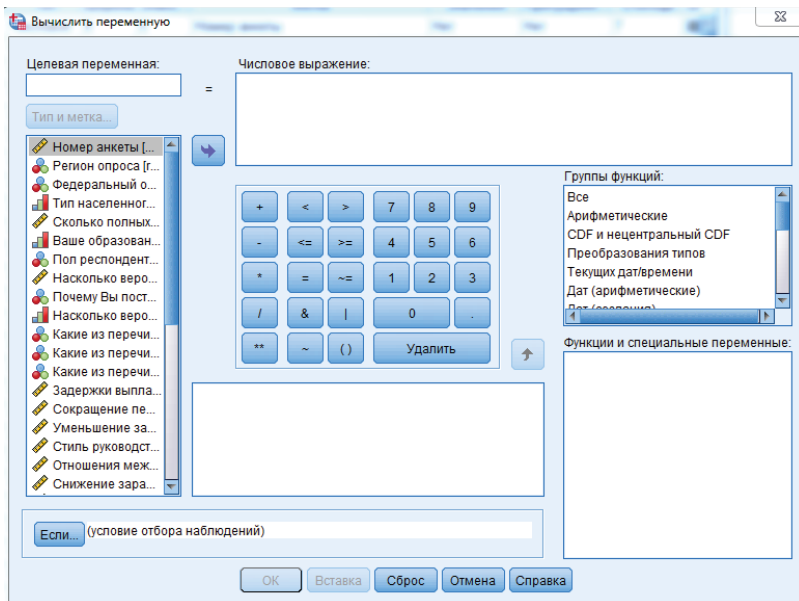


Рисунок 21. Процедура «Вычислить переменную»

Для вычисления переменной в окошке «Целевая переменная» необходимо обозначить имя вычисляемой переменной, в окошке «Числовое выражение» прописать формулу для вычисления, также можно воспользоваться функциями и специальными переменными. Для этого необходимо выбрать «Группы функций», далее выбрать конкретную функцию из предложенных в окошке «Функции и специальные переменные», щёлкнуть по ней два раза, и она будет активна в окошке «Числовое выражение». Одновременно в соседнем окошке выходит информация с объяснением выбранной функции (рисунок 22). Далее выбираем переменные, которые необходимо просуммировать.

Таким образом, можно вычислить различные переменные, к примеру, количество, т.е. сумму отмеченных характеристик респондентами, сколькими языками владеют, сколькими навыками владеют, пересчитать доходы в требуемой валюте и т.п.

Также, к примеру, можно вычислить возраст респондентов, если есть только информация о годе рождения. Для этого в окошке «Целевая переменная» обозначаем имя вычисляемой переменной, допустим «Возраст», в окошке «Числовое выражение» прописываем формулу расчёта: текущий год минус год рождения, т.е. выбираем переменную «Год рождения» и нажимаем кнопку «ОК». В конце нашей базы данных появится новая переменная «Возраст», отражающая возраст опрошенных респондентов, исходя из их года рождения.

Обратите внимание, что можно вычислить новые переменные с заданными условиями, для этого необходимо прописать условия в выбранном окошке «Если», расположенном внизу слева (рисунок 22).

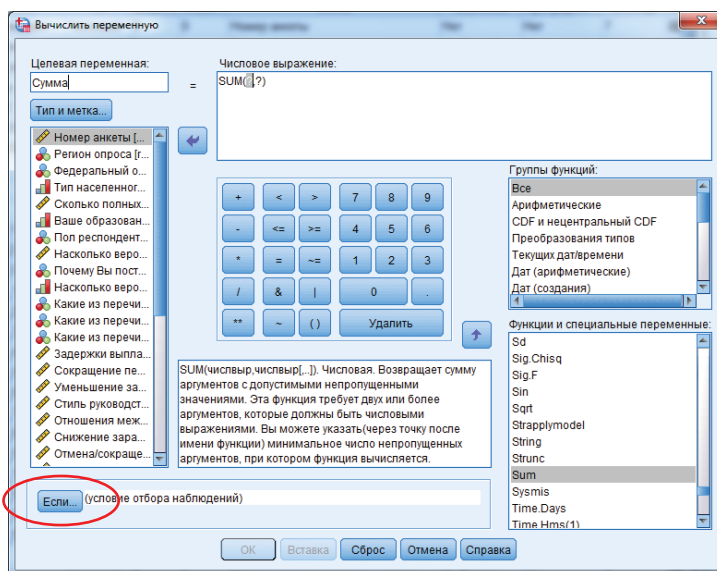


Рисунок 22. Пример вычисления переменной «Сумма» в процедуре «Вычислить переменную»

Следующей наиболее распространённой процедурой является перекодировка переменных, для этого программа SPSS предлагает две процедуры: «Перекодировать в те же переменные» и «Перекодировать в другие переменные». Выбрав первую процедуру, исходная переменная безвозвратно изменится, первоначальные коды будут перекодированы в другие коды, при этом теряется первоначальная информация. Вторая процедура «Перекодировать в другие переменные» позволяет сохранить исходную переменную без потери информации и вместе с тем получить новую переменную с новыми кодами.

Рекомендуется выбирать вторую процедуру «Перекодировать в другие переменные», поскольку сохраняются исходные данные, что позволяет в последующем создавать другие переменные с новыми кодами и перекодировками.

Рассмотрим процедуру перекодировки количественной переменной, показывающей возраст респондентов (AGE), в порядковую переменную, представляющую возрастные группы. Открыв процедуру «Перекодировать в другие переменные», выбираем переменную из списка слева, которую требуется перекодировать, например переменную AGE, отражающую возраст респондентов, переносим в окошко «Числовая переменная → Выходная переменная». Далее необходимо дать имя выходной переменной в соответствующем окошке справа, назовём новую переменную «Возрастные_группы», после этого необходимо обязательно нажать кнопку «Изменить», после этого в окошке «Числовая переменная → Выходная переменная» меняется информация на следующую: «AGE → Возрастные_группы» (рисунок 23).

Далее для перекодировки нажимаем кнопку «Старые и новые значения», открывшееся окно делится на две части: «Старое значение» и «Новое значение». В первой части можно выбрать отдельные значения, системные пропущенные, диапазоны значений, также диапазоны от наименьшего до указанного или от указанного до наибольшего – это всё данные исходной переменной, которую требуется перекодировать. Выбранным в части «Старое значение» опциям присваиваем новые коды в окошке «Значение» второй части «Новое значение» и обязательно нажимаем кнопку «Добавить», которая становится активной. В

окошке «Старое→Новое» появляется информация о перекодировке.

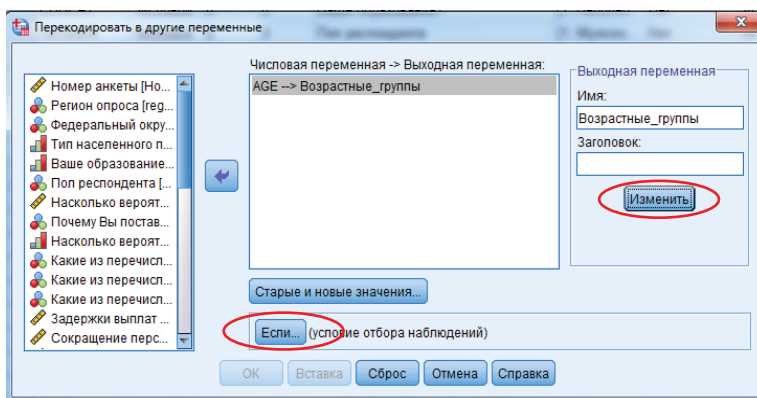


Рисунок 23. Пример процедуры «Перекодировать в другие переменные»_1

К примеру, возрастному диапазону от 18 до 29 лет ставим новый код «1», диапазону от 30 до 39 лет ставим код «2», диапазону 40-49 лет присваиваем код «3» и для последней возрастной группы выбираем опцию «Диапазон от указанного значения до наибольшего», где отмечаем возраст 50, этой группе ставим код «4». После завершения перекодировки необходимо обязательно нажать кнопку «Продолжить» (рисунок 24).

Затем следует завершить процедуру перекодировки нажатием кнопки «ОК». В базе данных появится новая переменная, которую мы назвали «Возрастные_группы», необходимо зайти в значения этой переменной и проставить метки значений, т.е. значение 1 означает возрастная группа «18-29 лет», значение 2 – «30-39 лет», 3 – «40-49 лет» и 4 – «50 лет и старше».

При перекодировке переменных есть возможность перекодировать в новые переменные с определёнными условиями, для этого необходимо прописать условия в окошке «Если», расположенном внизу слева (рисунок 23).

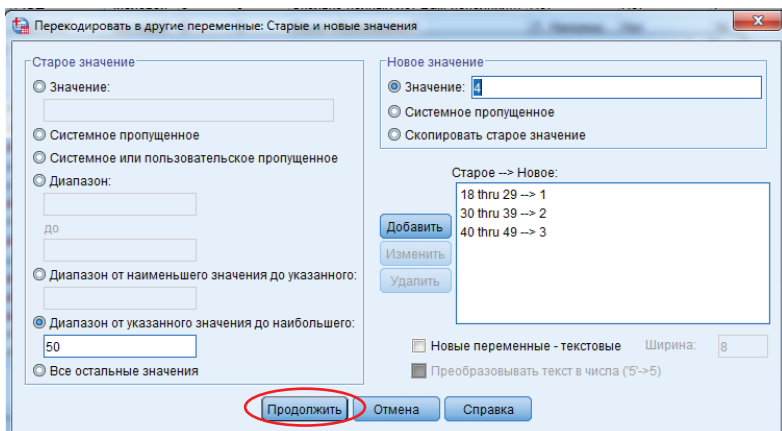


Рисунок 24. Пример процедуры «Перекодировать в другие переменные» 2

Обратите внимание, что по каждой процедуре есть кнопка «Справка», нажав на которую, можно получить информацию о самой процедуре, её предназначении, алгоритме действий. Данная кнопка всегда расположена крайней справа внизу открытого окна выбранной процедуры.

Контрольные вопросы:

1. На основе переменной «Возраст» в своей базе данных вычислите год рождения респондентов.
2. Также на основе переменной «Возраст» в своей базе данных вычислите, сколько будет лет респондентам в 2050 году.
3. Перекодируйте количественную переменную «Возраст» в порядковую переменную «Возрастные группы».
4. Также перекодируйте переменную «Уровень образования» в новую переменную, где должно быть только две категории: с высшим образованием и без высшего образования.
5. Перекодируйте количественную переменную «Возраст» в порядковую «Возрастные группы» только для женщин с высшим образованием.

2. ОПИСАТЕЛЬНЫЕ СТАТИСТИКИ, МЕРЫ ЦЕНТРАЛЬНОЙ ТЕНДЕНЦИИ И РАЗБРОСА

2.1. Частотное распределение переменных

Все методы анализа данных расположены в меню «Анализ». Прежде всего анализ данных в программе SPSS следует начинать с анализа частотного распределения переменных «Частоты», это один из методов описательной статистики и находится он в соответствующей вкладке «Описательные статистики» меню «Анализ» (рисунок 25). Частотное распределение позволяет понять, какой тип переменной, как распределены ответы, далее уже можно принять решение, какие сделать модификации с переменными и какие методы анализа применить.

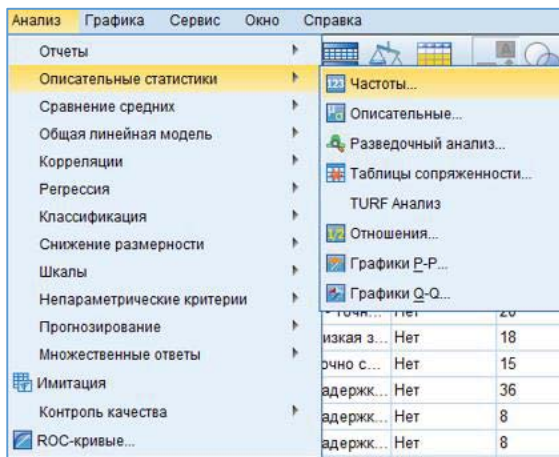


Рисунок 25. Меню «Анализ», описательные статистики

В открывшемся окне «Частоты» выбираем в левом окне переменные, частотный анализ которых необходимо сделать, переносим их в правое окно. Перенести можно либо стрелочкой

посередине, либо двойным щелчком по выбранной переменной. Затем необходимо нажать кнопку «ОК» (рисунок 26).

Нажав кнопку «ОК», автоматически открывается файл «Вывод», показывающий результат частотного анализа (рисунок 27). В частотной таблице в столбце «Частота» представлены абсолютные значения распределения ответов респондентов, в столбце «Проценты» показано распределение процентов всей выборки, включая пропущенные значения, если они есть. В следующем столбце «Процент допустимых» представлены валидные проценты без учёта пропущенных значений. Заключительный столбец «Накопленный процент» показывает поэтапную сумму ответов, т.е. суммирует первые варианты ответов со вторыми, и показывает эту сумму, далее прибавляет третий вариант и также показывает их сумму, и т.д. до ста процентов (рисунок 27).

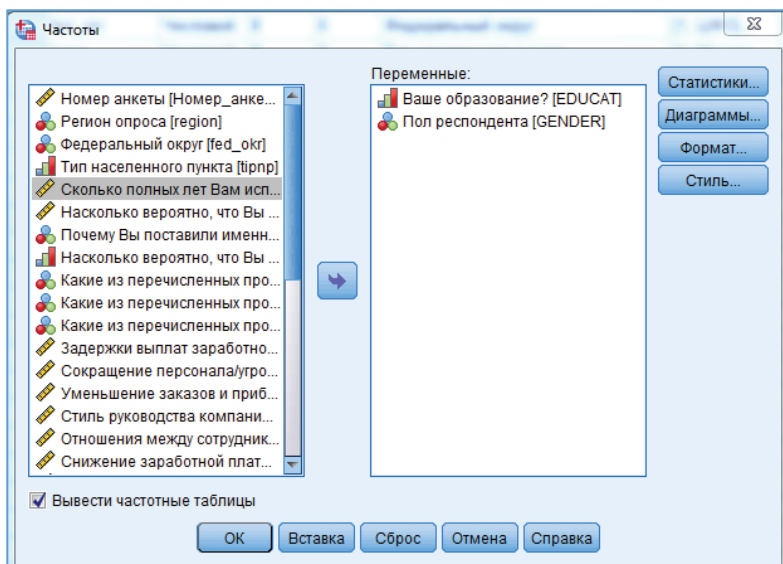


Рисунок 26. Пример частотного анализа в меню описательных статистик

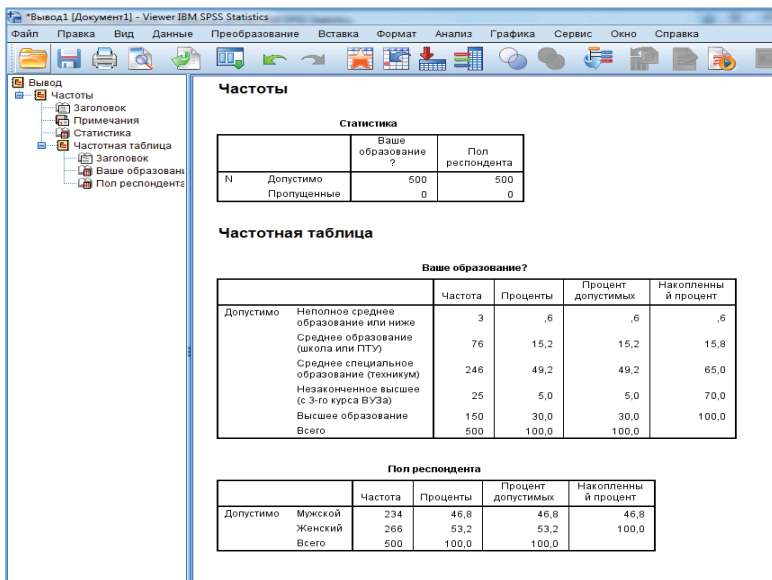


Рисунок 27. Результат частотного анализа в файле вывода

2.2. Меры центральной тенденции и разброса

Представленный выше частотный анализ хорошо подходит для номинальных и порядковых переменных. Для количественных переменных такой анализ недостаточен, для них целесообразней получить описательные статистики и меры центральной тенденции. Обозначенную информацию можно получить, отметив нужные статистики в окне «Статистики», для этого необходимо нажать эту кнопку, расположенную в правом верхнем углу окна частотного анализа (рисунок 26).

Например, нам необходимо проанализировать количественную переменную «Возраст». В частотном анализе выбираем эту переменную, переносим её в правое окно, затем заходим в «Статистики» и отмечаем нужные опции, чаще всего это: среднее значение (Mean), срединное значение (Median), наиболее часто встречающееся значение (Mode), минимальное значение (Min) и максимальное значение (Max). При необходимости можно

выбрать опции таких статистик, как: дисперсия, диапазон, сумма, среднеквадратичная ошибка среднего, асимметрия, квартили и различные процентили (рисунок 28). Далее нажимаем кнопку «Продолжить» и «ОК».

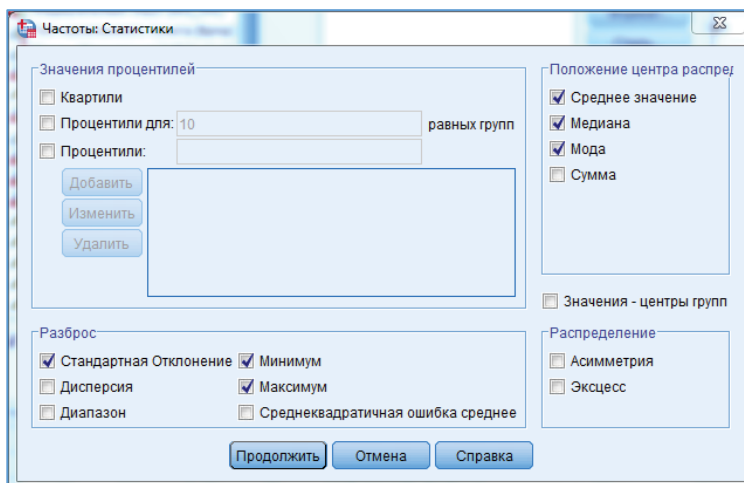


Рисунок 28. Окно «Статистики» для обозначения описательных статистик, мер центральной тенденции и разброса

Результаты выбранных описательных статистик, мер центральной тенденции и разброса выходят в файле «Вывод» в таблице «Статистика». Описывая переменную, отражающую возраст респондентов, следует отметить, что средний возраст опрошенных составляет 40,53 года, стандартное отклонение равно 12,15. Медиана равна 39 лет, т.е. половина опрошенных в возрасте до 39 лет и другая половина старше 39 лет. Наиболее часто встречаемый возраст составляет 35 лет. Минимум – 19 лет и максимальный возраст – 77 лет (рисунок 29).

→ Частоты

Статистика

Сколько полных лет Вам исполнилось?

N	Допустимо	500
	Пропущенные	0
Среднее значение		40,53
Медиана		39,00
Мода		35
Стандартная отклонения		12,151
Минимум		19
Максимум		77

Рисунок 29. Результаты выбранных для анализа статистик

Контрольные вопросы:

1. Сделайте частотное распределение нескольких переменных и проанализируйте их.
2. Объясните разницу между информацией в столбцах «Проценты» и «Процент допустимых».
3. В анализе каких переменных необходима информация в столбце «Накопленный процент»? Приведите примеры.
4. В каком случае необходимы статистики мер центральной тенденции и другие описательные статистики? Приведите примеры.
5. Сделайте анализ описательных статистик переменной, объясните полученные результаты.

2.3. Анализ двумерной зависимости.

Таблицы сопряженности

Следующим широко распространённым методом является анализ двумерных таблиц или, как ещё их называют, таблиц сопряженности. Об их распространённости в социологических исследованиях отмечено в учебном пособии «SPSS для социологов», где авторы отметили: «Традиционно в социологии построение и описание таблиц распределения (сопряженности) является одним из основных методов анализа данных выборочных исследований» [3].

Запуск данного анализа расположен в меню «Анализ» → «Описательные статистики» → «Таблицы сопряженности» (рисунок 25). Пройдя по обозначенной схеме, открывается диалоговое окно для запуска таблиц сопряженности (рисунок 30).

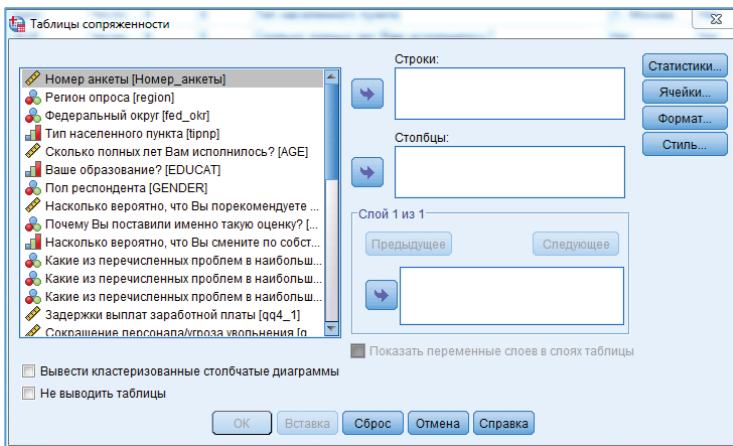


Рисунок 30. Диалоговое окно для запуска анализа таблиц сопряженности

Далее отбираются переменные в строки и столбцы, здесь необходимо определить, какой признак является зависимым, а какой независимым. Определение зависимой и независимой переменных зависит от исследовательских задач, выдвинутых гипотез и распределения переменных, вот почему рекомендуется для начала сделать частотное распределение всех переменных.

В отношении расположения переменных, чаще в столбцы ставят независимый признак, в разрезе которого анализируется зависимая переменная, которую переносят в строки. Но если независимая переменная содержит много категорий ответов, то можно поставить её и в строки. Таким образом в результате получится более удобная для восприятия таблица. Главное определить, в разрезе какой переменной анализируются данные, т.е. независимую переменную, то в этом направлении отмечаются проценты, что является обязательным при запуске таблиц сопряженности, если не будут отмечены проценты, программа выдаст результат в абсолютных значениях.

Проценты отмечаются в разделе «Ячейки», расположенном в правом верхнем углу. Нажав на кнопку «Ячейки», открывается диалоговое окно, где есть пять областей: количества, проценты, z-критерий, остатки и нецелочисленные веса (рисунок 31). По умолчанию отмечена только опция «Наблюдаемые» в области «Количества». Остальные опции следует отмечать в зависимости от целей анализа, задач исследования и потребностей исследователя – можно отметить требуемые опции и в результате получить запрашиваемую информацию, например стандартизованные остатки или ожидаемое количество ответов (рисунок 31).

Отметив необходимую информацию, нажимаем кнопку «Продолжить» и далее «ОК», и в файле «Вывод» сформируется запрашиваемая таблица сопряженности.

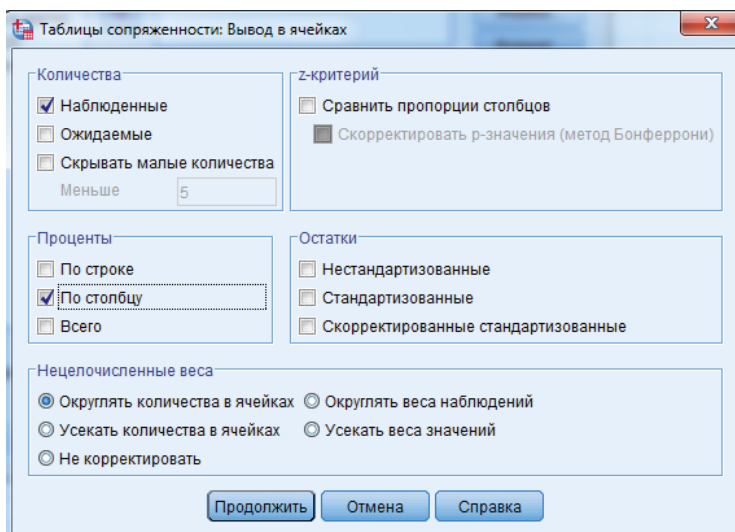


Рисунок 31. Диалоговое окно «Ячейки» таблиц сопряженности

Рассмотрим пример. Нам необходимо понять, среди кого больше распространена важность моды – среди мужчин или женщин. Ожидаемо, что женщинам более важна мода в одежде, обуви и аксессуарах, нежели мужчинам, но нам нужно подтвердить эту гипотезу эмпирическими данными. Запускаем

таблицы сопряженности по следующему алгоритму: «Анализ» → «Описательные статистики» → «Таблицы сопряженности». В столбцы переносим независимую переменную, фиксирующую пол респондентов, GENDER. В строки переносим переменную, отражающую важность моды, это ответы респондентов на вопрос «Насколько для Вас в целом важна мода в одежде, обуви, аксессуарах?» (рисунок 32).

В «Ячейках» отмечаем проценты по столбцу, нажимаем «Продолжить» и «ОК». Файл вывода показывает результат (рисунок 33). Результаты ответов показывают, что важность моды больше распространена среди женщин, чем среди мужчин: 46% женщин ответили, что мода «скорее важна» и для 41% «очень важна», что в сумме составляет 87%. Среди мужчин 26% ответили «скорее важна» и 16% «очень важна», что в сумме составляет 42%.

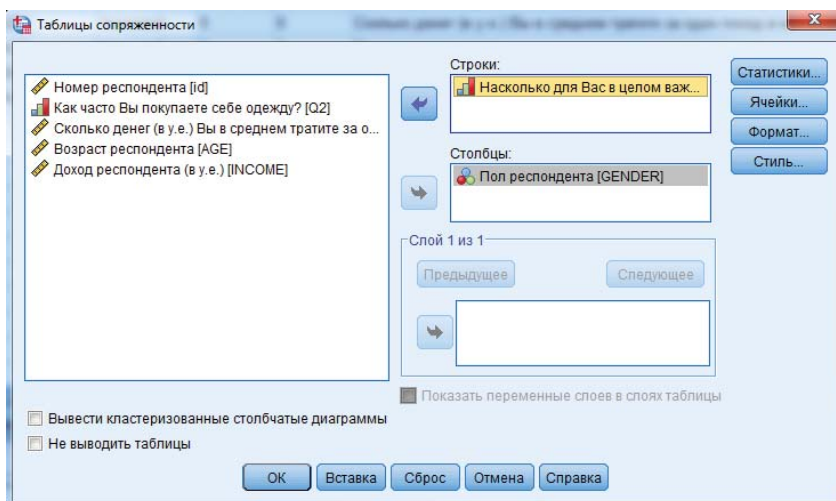


Рисунок 32. Пример запуска таблиц сопряженности

Таблицы сопряженности

	Сводный отчет по наблюдениям					
	Респонденты					
	Допустимо		Пропущенные		Всего	
	N	Проценты	N	Проценты	N	Проценты
Насколько для Вас в целом важна мода в одежде, обуви, аксессуарах? * Пол респондента	200	100,0%	0	0,0%	200	100,0%

Комбинационная таблица Насколько для Вас в целом важна мода в одежде, обуви, аксессуарах? * Пол респондента

			Пол респондента		Всего
			мужской	женский	
Насколько для Вас в целом важна мода в одежде, обуви, аксессуарах?	Совсем не важна	Количество	6	3	9
		% в Пол респондента	6,3%	2,9%	4,5%
	Скорее не важна	Количество	50	10	60
		% в Пол респондента	52,1%	9,6%	30,0%
	Скорее важна	Количество	25	48	73
		% в Пол респондента	26,0%	46,2%	36,5%
	Очень важна	Количество	15	43	58
		% в Пол респондента	15,6%	41,3%	29,0%
Всего	Количество	96	104	200	
	% в Пол респондента	100,0%	100,0%	100,0%	

Рисунок 33. Результат запуска таблиц сопряженности в файле вывода

Получив результаты двумерного распределения таблиц сопряженности, может возникнуть вопрос, есть ли здесь влияние третьей переменной? Возможно, важность моды в разрезе гендера меняется с возрастом. Чтобы проверить возникшую гипотезу, можно ещё раз запустить анализ таблиц сопряженности, но включить в него ещё одну переменную, а именно возраст. В базе данных возраст представлен количественной переменной, данная переменная не подходит для таблиц сопряженности, поэтому эту переменную необходимо перекодировать в порядковую и затем уже включать в таблицы сопряженности. Процедура перекодирования данных представлена в разделе «1.4. Преобразование и модификация данных».

Размах количественной переменной «Возраст» респондентов варьирует от 16 до 35 лет, средний возраст – 23 года. Преобразуем данную переменную в новую порядковую переменную «Возрастные_группы» с тремя группами: до 20 лет, 21-29 лет и 30 лет и старше.

Далее также запускаем таблицы сопряженности по следующему алгоритму: «Анализ» → «Описательные статистики» → «Таблицы сопряженности». В столбцы переносим независимую переменную, фиксирующую пол респондентов, GENDER. В строки переносим переменную, отражающую важность моды, это ответы респондентов на вопрос «Насколько для Вас в целом важна мода в одежде, обуви, аксессуарах?», далее в окошко ниже «Слой 1 из 1» переносим нашу новую порядковую переменную «Возрастные группы» (рисунок 34). В ячейках ставим проценты по столбцу.

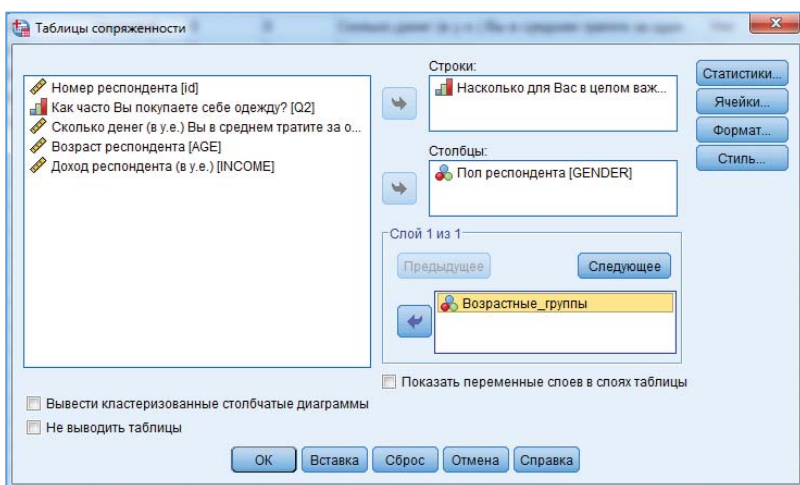


Рисунок 34. Пример запуска таблиц сопряженности с включением третьей переменной

Результат представлен в виде таблице на рисунке 35, где показано распределение ответов о важности моды в разрезе гендера и в разрезе возрастных групп. Анализируя результаты таблицы, можно сделать вывод, что у мужчин интерес к моде становится менее важным с возрастом. Для большинства женщин интерес к моде остаётся важным в любом возрасте.

Комбинационная таблица Насколько для Вас в целом важна мода в одежде, обуви, аксессуарах? * Пол респондента * Возрастные_группы						
Возрастные_группы				Пол респондента		Всего
				мужской	женский	
До 20 лет	Насколько для Вас в целом важна мода в одежде, обуви, аксессуарах?	Скорее важна	Количество	15	17	32
			% в Пол респондента	62,5%	30,4%	40,0%
		Очень важна	Количество	9	39	48
			% в Пол респондента	37,5%	69,6%	60,0%
		Всего	Количество	24	56	80
			% в Пол респондента	100,0%	100,0%	100,0%
21-29 лет	Насколько для Вас в целом важна мода в одежде, обуви, аксессуарах?	Совсем не важна	Количество	5	3	8
			% в Пол респондента	8,3%	7,5%	8,0%
		Скорее не важна	Количество	44	9	53
			% в Пол респондента	73,3%	22,5%	53,0%
		Скорее важна	Количество	8	25	33
			% в Пол респондента	13,3%	62,5%	33,0%
		Очень важна	Количество	3	3	6
		% в Пол респондента	5,0%	7,5%	6,0%	
	Всего	Количество	60	40	100	
			% в Пол респондента	100,0%	100,0%	100,0%
30 лет и старше	Насколько для Вас в целом важна мода в одежде, обуви, аксессуарах?	Совсем не важна	Количество	1	0	1
			% в Пол респондента	8,3%	0,0%	5,0%
		Скорее не важна	Количество	6	1	7
			% в Пол респондента	50,0%	12,5%	35,0%
		Скорее важна	Количество	2	6	8
			% в Пол респондента	16,7%	75,0%	40,0%
	Очень важна	Количество	3	1	4	
		% в Пол респондента	25,0%	12,5%	20,0%	
	Всего	Количество	12	8	20	
			% в Пол респондента	100,0%	100,0%	100,0%
Всего	Насколько для Вас в целом важна мода в одежде, обуви, аксессуарах?	Совсем не важна	Количество	6	3	9
			% в Пол респондента	6,3%	2,9%	4,5%
		Скорее не важна	Количество	50	10	60
			% в Пол респондента	52,1%	9,6%	30,0%
	Скорее важна	Количество	25	48	73	
		% в Пол респондента	26,0%	46,2%	36,5%	

Рисунок 35. Результат запуска таблицы сопряженности с включением третьей переменной

Являются ли показанные в таблицах сопряженности различия статистически значимыми? Возможно ли опровергнуть нулевую гипотезу об отсутствии различий и принять альтернативную гипотезу? Ответы на эти вопросы можно получить, задав необходимые статистики, в частности наиболее известный коэффициент связи Хи-квадрат. Коэффициенты связи находятся в соответствующем окне, на которое можно перейти, нажав на кнопку «Статистики» в диалоговом окне таблиц сопряженности. В открывшемся диалоговом окне «Статистики» представляется возможность задать Хи-квадрат, коэффициент корреляции,

коэффициенты сопряженности для номинальных и порядковых переменных и другие статистические показатели (рисунок 36).

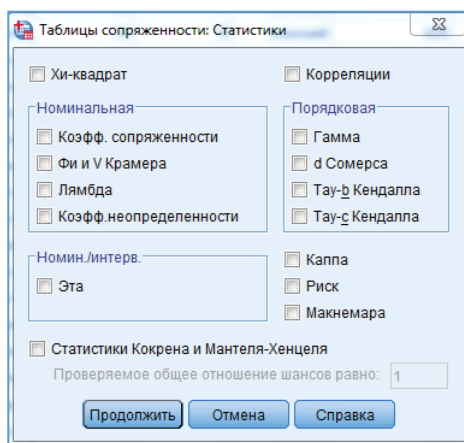


Рисунок 36. Диалоговое окно «Статистики» таблиц сопряженности

С дополнительной информацией о таблицах сопряженности и коэффициентах связи можно ознакомиться в учебном пособии Крыштановского, на стр. 39-58 [4].

Контрольные вопросы:

1. Какие переменные вы выберете в качестве независимых при запуске таблиц сопряженности? Обоснуйте свой выбор.
2. Запустите таблицы сопряженности и проанализируйте полученный результат.
3. Какие переменные следует выбирать в качестве зависимых? Объясните свой ответ.
4. В таблицах сопряженности по умолчанию выходит результат в абсолютных величинах, достаточно ли этой информации? Объясните, для чего необходимо обязательно отмечать проценты при запуске таблиц сопряженности.
5. Для каких исследовательских задач целесообразно применять таблицы сопряженности? Приведите примеры.

3. АНАЛИЗ ВЗАИМОСВЯЗЕЙ КОЛИЧЕСТВЕННЫХ И КАЧЕСТВЕННЫХ ПЕРЕМЕННЫХ

Изученные до этого методы направлены на формирование описательных моделей рассматриваемых социальных явлений. Как отмечает Крыштановский: «Действительно, даже анализ таблиц сопряженности, вычисление коэффициентов связи в этих таблицах подразумевают фиксацию статистических взаимосвязей переменных, а на основе этих выявленных взаимосвязей социолог начинает конструировать объяснительные модели, привлекая социологические теории, своё знание социальной реальности, т.е. ту информацию, которая в анализе данных отсутствует» [4, с. 82].

Анализируя результаты социологических исследований, представляет интерес изучить причинные модели, где один из признаков является следствием каких-либо причин. В настоящем разделе будут рассмотрены методы взаимосвязи количественных переменных с качественными, первые чаще считаются следствием, вторые переменные расцениваются, как причины.

3.1. Сравнение средних значений показателей в группах

Для анализа количественных переменных, таких, как возраст, доход, количество членов домохозяйства и т.п., представляет интерес изучить средние значения переменных, сравнить их в группах, сравнить по отношению к какой-либо заданной величине. Средние значения – весьма информативная мера из числа мер центральной тенденции. К примеру, можно сравнить средний уровень заработной платы в регионах или средний возраст в изучаемых группах. Для этого в программе SPSS есть процедуры «Сравнение средних», которые находятся в меню «Анализ» (рисунок 37).

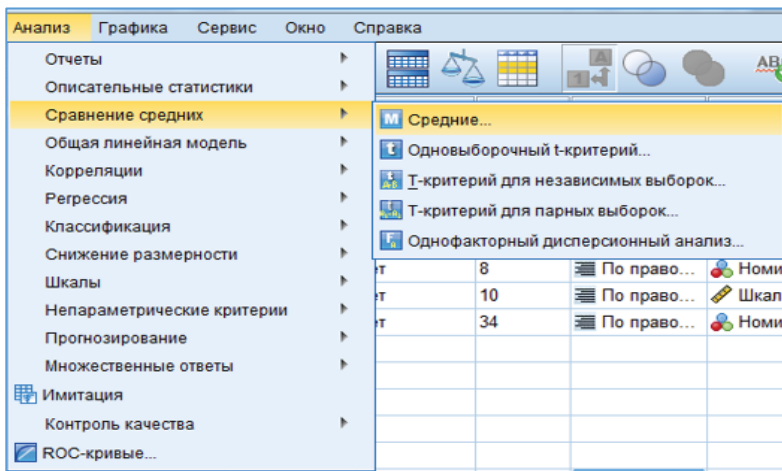


Рисунок 37. Меню вкладки «Сравнение средних»

Первая процедура «Средние» позволяет вычислить и сравнить в табличном виде средние значения интересующих переменных, данную процедуру можно запустить как для одной переменной, так и одновременно для нескольких переменных. К примеру, нам необходимо посмотреть средний доход в разрезе гендера, у кого в среднем больше доход, у мужчин или женщин. Также можно посмотреть средний уровень доходов в разрезе переменной важности моды. Для этого заходим по следующему алгоритму: «Анализ» → «Сравнение средних» → «Средние». В окошко «Список зависимых переменных» переносим переменную «Доход респондента (в условных единицах)», в окошко «Список независимых переменных» переносим переменные «Пол респондента» и «Насколько для Вас в целом важна мода?» (рисунок 38).

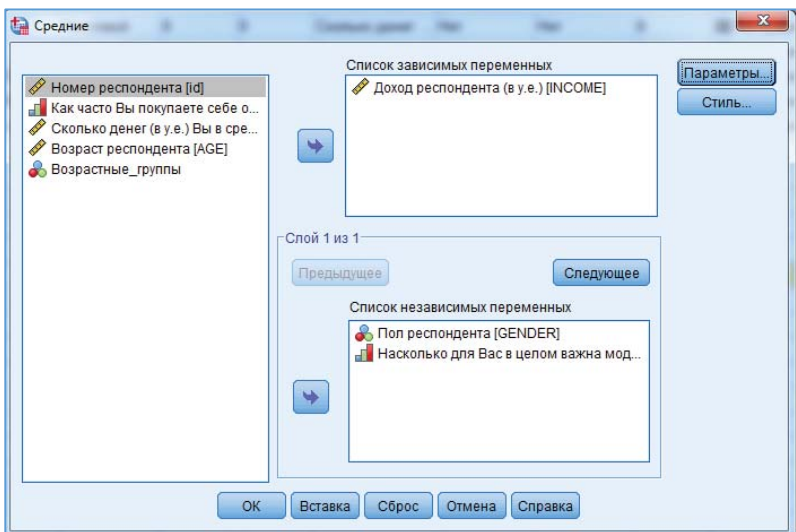


Рисунок 38. Диалоговое окно процедуры «Средние»

В результате мы получаем средние значения в разрезе заданных переменных (рисунок 39). Результаты свидетельствуют, что в изучаемой генеральной совокупности у женщин в среднем (1044 у.е.) доходы больше, чем у мужчин (870 у.е.). Также наблюдаются различия в разрезе важности моды, в среднем доходы больше у тех, кому в разной степени важна мода. Полученные результаты требуют дополнительного анализа, чтобы понять, что является причиной, а что следствием, т.е. у изучаемой социальной группы становится больше доходов и поэтому для них повышается важность моды или же наоборот, те, кому важна мода, стараются больше зарабатывать.

Средние

Сводный отчет по наблюдениям

	Респонденты					
	Включено		Исключено		Всего	
	N	Проценты	N	Проценты	N	Проценты
Доход респондента (в у.е.) * Пол респондента	200	100,0%	0	0,0%	200	100,0%
Доход респондента (в у.е.) * Насколько для Вас в целом важна мода в одежде, обуви, аксессуарах?	200	100,0%	0	0,0%	200	100,0%

Доход респондента (в у.е.) * Пол респондента

Доход респондента (в у.е.)

Пол респондента	Среднее значение	N	Стандартная отклонения
мужской	869,95	96	422,296
женский	1043,94	104	511,150
Всего	960,42	200	477,453

Доход респондента (в у.е.) * Насколько для Вас в целом важна мода в одежде, обуви, аксессуарах?

Доход респондента (в у.е.)

Насколько для Вас в целом важна мода в одежде, обуви, аксессуарах?	Среднее значение	N	Стандартная отклонения
Совсем не важна	836,67	9	276,315
Скорее не важна	729,25	60	330,133
Скорее важна	1195,48	73	509,649
Очень важна	922,93	58	461,930
Всего	960,42	200	477,453

Рисунок 39. Результат процедуры сравнения средних

По умолчанию программа выдаёт информацию о средних значениях, количестве наблюдений и стандартном отклонении, при необходимости других статистик можно нажать кнопку «Параметры», расположенную в верхнем правом углу. Также при необходимости в этом окне можно включить «Таблица дисперсионного анализа и эта», в результате чего получим информацию о значимости полученных различий средних значений (рисунок 40).

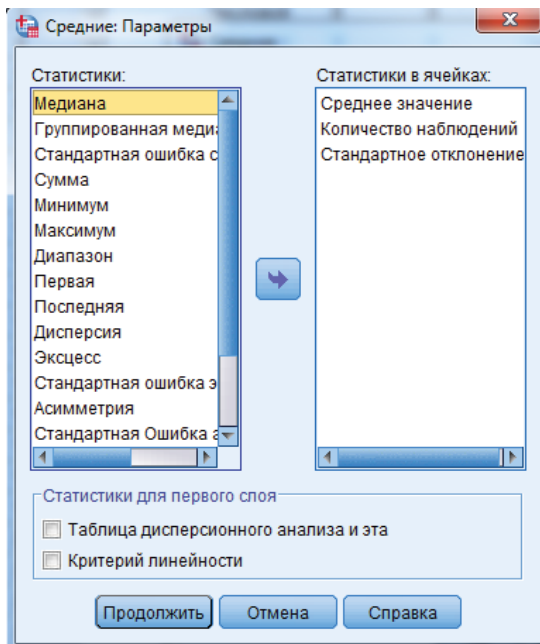


Рисунок 40. Диалоговое окно «Параметры» в процедуре «Средние»

Обратите внимание, что по каждой процедуре есть кнопка «Справка», нажав на которую, можно получить информацию о самой процедуре, её предназначении, алгоритме действий, о статистиках. Данная кнопка всегда расположена крайней справа, внизу открытого окна выбранной процедуры.

Контрольные вопросы:

1. Для каких исследовательских задач целесообразно применять процедуру «Сравнение средних»? Приведите примеры.
2. Какие переменные вы выберете в качестве независимых при запуске процедуры «Сравнение средних»? Обоснуйте свой выбор.
3. Какие переменные следует выбирать в качестве зависимых при запуске процедуры «Сравнение средних»? Объясните свой ответ.
4. Какие дополнительные статистики вы отобрали бы для своего анализа средних? Объясните, какую информацию они раскроют.
5. Запустите процедуру «Сравнение средних» и проанализируйте полученный результат.

3.2. Одновыборочный Т-критерий

Процедуры Т-тест (или тест Стьюдента) предназначены для решения задач доказательства наличия различий средних значений количественных переменных, когда имеются только две сравниваемые группы.

Процедура одновыборочного критерия Т-тест (One sample T-test) проводится для сравнения среднего значения изучаемого показателя относительно известного или существующего среднего показателя, т.е. сравнивается среднее значение переменной с заданным значением.

Данная процедура находится во вкладке «Сравнение средних», которая, в свою очередь, находится в меню «Анализ» программы SPSS (рисунок 37). К примеру, необходимо проанализировать среднее значение дохода изучаемой группы по сравнению с известным значением дохода. Для этого заходим в процедуру «Одновыборочный Т-критерий», в окошко «Проверяемые переменные» переносим переменную «Доход респондента», который был замерен в условных единицах. В нижнее окошко «Проверяемое значение» прописываем среднее значение дохода 600 условных единиц (рисунок 41). Далее нажимаем «ОК».

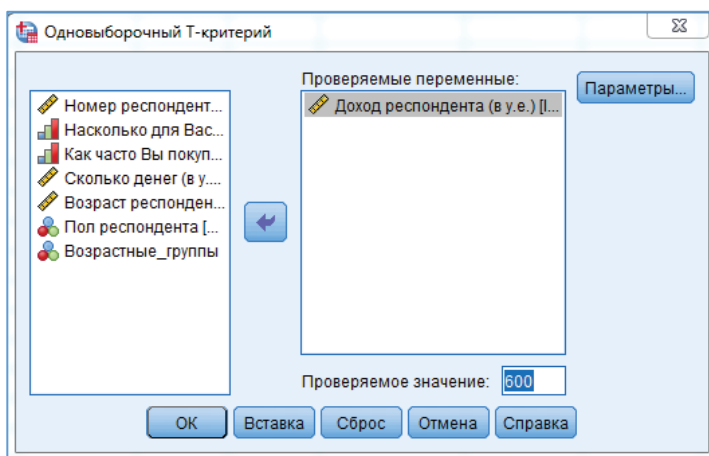


Рисунок 41. Диалоговое окно процедуры «Одновыборочный Т-критерий»

В файле «Вывод» автоматически формируется результат (рисунок 42). Представляются две таблицы, в первой из них (Одновыборочная статистика) показаны описательные статистики изучаемой переменной, во второй таблице «Одновыборочный критерий» представлены результаты проверки статистической гипотезы, утверждающей, что среднее значение дохода изучаемой группы равно фиксированному среднему значению дохода.

Т-критерий				
Одновыборочная статистика				
	N	Среднее значение	Стандартная отклонения	Среднекв. ошибка среднего
Доход респондента (в у. е.)	200	960,43	477,453	33,761

Одновыборочный критерий						
	Значение критерия = 600					
	t	ст. св.	Знач. (2-х сторонняя)	Разность средних	95% доверительный интервал для разности	
					Нижняя	Верхняя
Доход респондента (в у. е.)	10,676	199	,000	360,425	293,85	427,00

Рисунок 42. Результат процедуры «Одновыборочный Т-критерий»

Полученные результаты запущенной процедуры показывают, что среднее значение дохода нашей изучаемой группы составляет 960,43 условных единиц. Разность средних значений дохода между исследуемой группой и известным средним значением в 600 условных единиц составляет 360,425. Значение значимости (2-сторонней) (Sig.<0,05) свидетельствует, что отклоняем нулевую гипотезу и принимаем альтернативную – есть различия в средних значениях дохода между изучаемой группой и фиксированным средним значением.

Контрольные вопросы:

1. Для каких исследовательских задач целесообразно применять процедуру «Одновыборочный Т-критерий»? Приведите примеры.
2. Какие переменные вы выберете для процедуры «Одновыборочный Т-критерий»? Обоснуйте свой выбор.

3. Приведите примеры нулевой и альтернативной гипотез сравнения данных процедурой «Одновыборочный Т-критерий».

4. При каком значении значимости можно принять нулевую гипотезу?

5. Запустите процедуру «Одновыборочный Т-критерий» и проанализируйте полученный результат.

3.3. Т-критерий для независимых выборок

Процедура Т-критерий (Т-test) для независимых выборок применяется в случае необходимости сравнения среднего значения одной переменной в двух группах. Данная процедура часто применяется в анализе эмпирических данных. О независимости выборок в социологических исследованиях отмечает Крыштановский: «Для социологических исследований важное допущение о том, что эти две группы (две выборки) являются независимыми, почти всегда выполняется. Действительно, если мы сравниваем выборки в двух типах населённых пунктов либо выборки мужчин и женщин, мы знаем, что сбор данных в этих группах выполняется независимо. Другими словами, то, как отвечали женщины, никак не влияло на ответы мужчин» [4, с. 89].

Обозначенная процедура находится во вкладке «Сравнение средних», которая, в свою очередь, находится в меню «Анализ» программы SPSS (рисунок 37). Для примера проанализируем среднее значение доходов (в условных единицах) между двумя независимыми выборками мужчин и женщин. Для этого необходимо зайти в процедуру «Т-критерий для независимых выборок», в окошко «Проверяемые переменные» переносим переменную «Доход респондента», который был замерен в условных единицах. В нижнее окошко «Группировать по» переносим переменную GENDER – рядом с этой переменной возникают скобки со знаками вопросов и становится активной кнопка «Задать группы» (рисунок 43). Далее заходим в «Задать группы» и проставляем значения переменной, которые необходимо изучить. Следует отметить, что группирующая переменная может иметь более двух значений, но поскольку процедура «Т-критерий для независимых выборок» сравнивает

только две группы респондентов, необходимо выбрать из нескольких значений только две группы. В данном случае в переменной GENDER только две группы, поэтому в Группа 1 ставим 1, это мужчины, и в Группа 2 соответственно ставим 2, это женщины. Далее нажимаем «Продолжить» и «ОК».

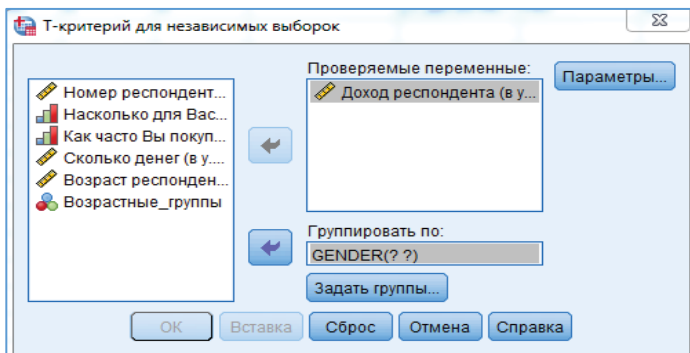


Рисунок 43. Диалоговое окно процедуры «Т-критерий для независимых выборок»

Результат запущенной процедуры «Т-критерий для независимых выборок» представлен на рисунке 44. Верхняя таблица «статистика группы» представляет информацию об описательных статистиках – данная информация идентична информации, которую мы получили в результате процедуры «Сравнение средних», представленной на рисунке 39.

Нижняя таблица «Критерий для независимых выборок» состоит из двух частей. Сначала необходимо обратить внимание на «Критерий равенства дисперсий Ливиня» – здесь проверяется нулевая гипотеза о равенстве дисперсий двух изучаемых выборок, в нашем случае мужчин и женщин. Таблица показывает, что F статистика равна 6,556; значимость равна 0,011. Значение значимости меньше наиболее употребительного в статистике уровня значимости 0,05, поэтому отвергаем нулевую гипотезу о равенстве дисперсий соответственно смотрим на статистики, обозначенные в нижней строке «Не предполагаются равные дисперсии». Во второй части таблицы «Т-критерий для равенства средних» смотрим на информацию нижней строки. Для начала

обращаем внимание на значение значимости, которое равно 0,009, это меньше, чем 0,05, соответственно отклоняем нулевую гипотезу о равенстве средних значений доходов мужчин и женщин. Таким образом, можно сделать вывод, что доход мужчин и женщин различен.

Т-критерий									
Статистика группы									
Пол респондента		N	Среднее значение	Стандартная отклоняющая	Средняя ошибка среднего				
Доход респондента (в у. е.)									
мужской		96	869,95	422,296	43,100				
женский		104	1043,94	511,150	50,122				

Критерий для независимых выборок										
		Критерий равенства дисперсий Левина			t-критерий для равенства средних					
		F	Знач.	t	ст.ев.	Знач. (2-х сторонняя)	Разность средних	Среднеквадратичная ошибка разности	95% доверительный интервал для разности	
									Нижняя	Верхняя
Доход респондента (в у. е.)	Предполагаются равные дисперсии	6,556	.011	-2,612	198	.010	-173,994	66,609	-306,348	-42,641
	Не предполагаются равные дисперсии			-2,632	195,655	.009	-173,994	66,105	-304,365	-43,624

Рисунок 44. Результат процедуры «Т-критерий для независимых выборок»

Т-критерий является одной из необходимых процедур, поскольку это статистический инструмент проверки средних изучаемых переменных, так как визуальное сходство или различие значений средних не может быть доказательством их равенства или различия.

Контрольные вопросы:

1. Для каких исследовательских задач целесообразно применять процедуру «Т-критерий для независимых выборок»? Приведите примеры.
2. Какие переменные вы выберете для процедуры «Т-критерий для независимых выборок»? Обоснуйте свой выбор.
3. Приведите примеры нулевой и альтернативной гипотез сравнения данных процедурой «Т-критерий для независимых выборок».
4. При каком значении значимости следует отвергнуть нулевую гипотезу?
5. Запустите процедуру «Т-критерий для независимых выборок» и проанализируйте полученный результат.

3.4. Т-критерий для парных выборок

Процедура «Т-критерий для парных выборок» позволяет сравнить средние значения двух отдельных переменных одной выборки между собой. Исследовательская задача состоит в том, чтобы проверить различаются ли средние значения двух количественных переменных, измеренных при опросе одной выборки.

Процедура «Т-критерий для парных выборок» также находится во вкладке «Сравнение средних», которая, в свою очередь, находится в меню «Анализ» программы SPSS (рисунок 37).

Для примера проанализируем, есть ли различия в средних значениях расходов на покупку одежды за один раз и на одно посещение кафе или ресторана. Исследовательская задача состоит в том, чтобы понять, на что в среднем больше тратят люди из исследуемой выборки: на одежду или на то, чтобы покушать в кафе, ресторане, или эти расходы не различаются.

Заходим в обозначенную процедуру по следующему алгоритму: «Анализ» → «Сравнение средних» → «Т-критерий для парных выборок». Переносим в окно «Парные переменные» наши две количественные переменные, отражающие ответы на вопросы: «Сколько денег (в условных единицах) Вы в среднем тратите за один поход в магазин за одеждой?» и «Сколько денег (в условных единицах) Вы в среднем тратите за один поход в кафе, ресторан?». Далее нажимаем «ОК».

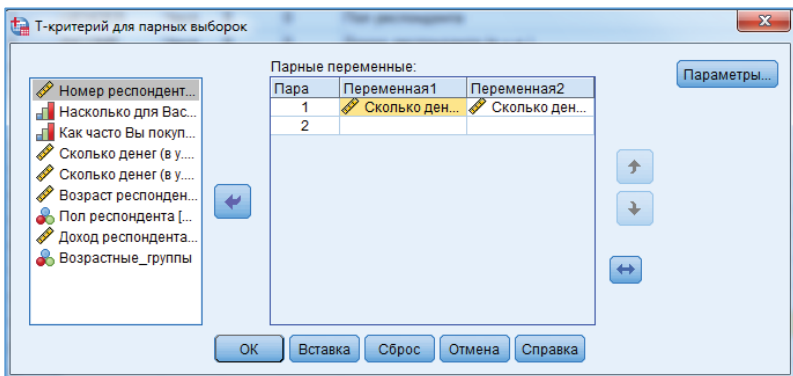


Рисунок 45. Диалоговое окно процедуры «Т-критерий для парных выборок»

Результат процедуры «Т-критерий для парных выборок» представляется в трёх таблицах. Первая таблица «Статистика парных выборок» показывает описательные статистики изучаемых переменных. Вторая таблица «Корреляции парных выборок» показывает корреляцию между двумя изучаемыми переменными. Третья таблица «Критерий парных выборок» показывает статистики процедуры. Значение значимости ($\text{Sig.} < 0,05$) свидетельствует об отклонении нулевой гипотезы. Таким образом, можно сделать вывод, что изучаемая нами социальная группа в среднем расходуют больше денег за один поход в магазин за одеждой, чем за один поход покушать в кафе или ресторане, разница в среднем составляет около 129 условных единиц (рисунок 46).

Т-критерий		Статистика парных выборок			
		Среднее значение	N	Стандартная отклонения	Средняя ошибка среднего
Пара 1	Сколько денег (в у.е.) Вы в среднем тратите за один поход в магазин за одеждой?	151,29	200	119,224	8,360
	Сколько денег (в у.е.) Вы в среднем тратите за один поход в кафе, ресторан?	22,37	200	20,782	1,469

Корреляция парных выборок				
		N	Корреляция	Знач.
Пара 1	Сколько денег (в у.е.) Вы в среднем тратите за один поход в магазин за одеждой? &	200	,166	,017
	Сколько денег (в у.е.) Вы в среднем тратите за один поход в кафе, ресторан?			

Критерий парных выборок		Парные различия				T	ст.св.	Знач. (2-х сторонняя)
		Среднее значение	Стандартная отклонения	Средняя ошибка среднего	95% доверительный интервал для разности			
							Нижняя	Верхняя
Пара 1	Сколько денег (в у.е.) Вы в среднем тратите за один поход в магазин за одеждой?	128,905	116,548	8,241	112,654	145,156	15,642	199
	Сколько денег (в у.е.) Вы в среднем тратите за один поход в кафе, ресторан?							

Рисунок 46. Результат процедуры «Т-критерий для парных выборок»

Контрольные вопросы:

1. Запустите процедуру «Т-критерий для парных выборок» и проанализируйте полученный результат.
2. Для каких исследовательских задач целесообразно применять процедуру «Т-критерий для парных выборок»? Приведите примеры.
3. Приведите примеры нулевой и альтернативной гипотез сравнения данных процедурой «Т-критерий для парных выборок».
4. При каком значении значимости следует отвергнуть нулевую гипотезу?
5. Какие переменные вы выберете для процедуры «Т-критерий для парных выборок»? Обоснуйте свой выбор.

3.5. Однофакторный дисперсионный анализ

Представленная выше процедура сравнения средних для независимых выборок (Т-критерий) имеет ограничение: сравнить можно только две группы. Зачастую в социологических исследованиях представлены номинальные или порядковые переменные с несколькими группами, например, несколько групп респондентов по уровню образования, по этническому признаку. В таких случаях применяют методы дисперсионного анализа. Данный вид анализа основан на сравнении дисперсий

количественной переменной в разрезе категорий номинальной или порядковой переменных.

В программе SPSS представлена процедура «Однофакторный дисперсионный анализ», расположенная во вкладке «Сравнение средних» в меню «Анализ» (рисунок 47).

В рамках изучаемого курса рассматривается только процедура «Однофакторный дисперсионный анализ», в которой подразумевается лишь один фактор, в разрезе которого анализируются данные. Существует ещё многофакторный дисперсионный анализ, который применяется в случае двух и более количественных зависимых переменных, где также проверяются групповые различия в разрезе нескольких зависимых переменных.

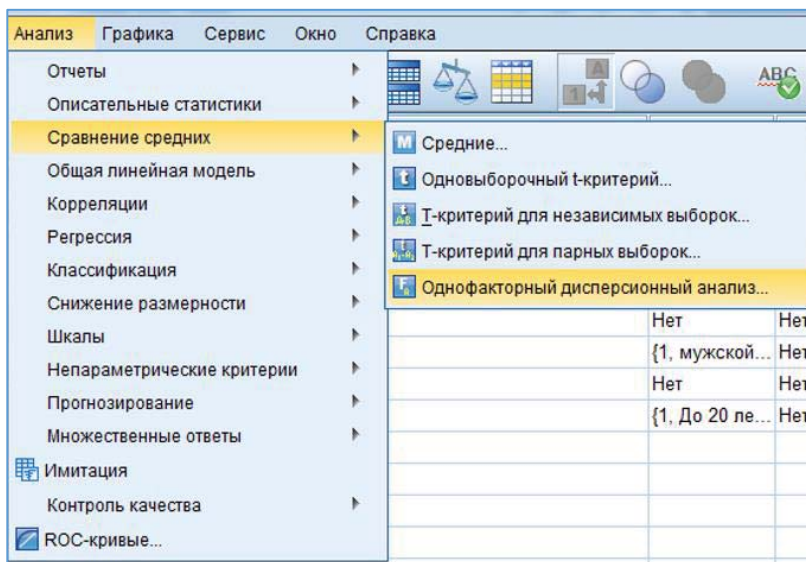


Рисунок 47. Меню для запуска процедуры «Однофакторный дисперсионный анализ»

Рассмотрим в качестве примера следующую задачу: различаются ли расходы на покупку одежды в зависимости от важности моды или никаких различий нет, и люди в среднем

расходуют на одежду одинаково в независимости от того, важна для них мода или нет. Заходим по обозначенной выше схеме: «Анализ» → «Сравнение средних» → «Однофакторный дисперсионный анализ». В окошко «Список зависимых переменных» переносим количественную переменную, отражающую расходы на одежду за один раз, это вопрос «Сколько денег (в у.е.) Вы в среднем тратите за один поход в магазин за одеждой?». В окошко «Фактор» переносим порядковую переменную, фиксирующую важность моды для человека, это вопрос «Насколько для Вас в целом важна мода в одежде, обуви, аксессуарах?». На этом можно закончить, нажав кнопку «ОК» (рисунок 48).

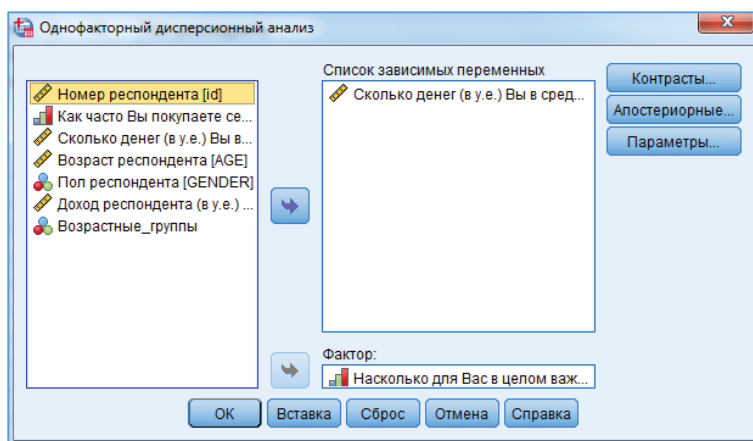


Рисунок 48. Диалоговое окно процедуры «Однофакторный дисперсионный анализ»

При выше описанном алгоритме получится результат, представленный на рисунке 49. Получаем таблицу ANOVA, название которой является аббревиатурой полного названия: Analysis of variances, что в переводе означает анализ дисперсий. Значение значимости свидетельствует о необходимости отклонить нулевую гипотезу о равенстве расходов и принять альтернативную о различии расходов. Таким образом, можно сделать вывод, что расходы на покупку одежды различаются в зависимости от важности моды.

→ Однонаправленная

ANOVA

Сколько денег (в у.е.) Вы в среднем тратите за один поход в магазин за одеждой?

	Сумма квадратов	ст. св.	Средний квадрат	F	Знач.
Между группами	632399,694	3	210799,898	19,226	,000
Внутри групп	2149000,181	196	10964,287		
Всего	2781399,875	199			

Рисунок 49. Результат запуска процедуры «Однофакторный дисперсионный анализ»

Возникает вопрос, каковы суммы расходов, кто тратит больше, кто меньше, сколько респондентов в каждой группе, каково стандартное отклонение. Чтобы получить обозначенную информацию об описательных статистиках и другие статистики, необходимо отметить их во вкладке «Параметры» (рисунок 50), нажать «Продолжить» и заново «ОК».

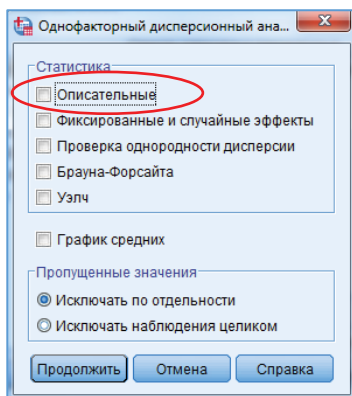


Рисунок 50. Диалоговое окно «Параметры» процедуры «Однофакторный дисперсионный анализ»

При необходимости более детального анализа методами множественных сравнений можно задать дополнительные статистики. К примеру, в некоторых литературных источниках по

изучению статистики упоминаются такие тесты, как Бонферрони, Шеффе, Дункан, Уоллер-Дункан и др. Они относятся к апостериорным множественным сравнениям (рисунок 51).

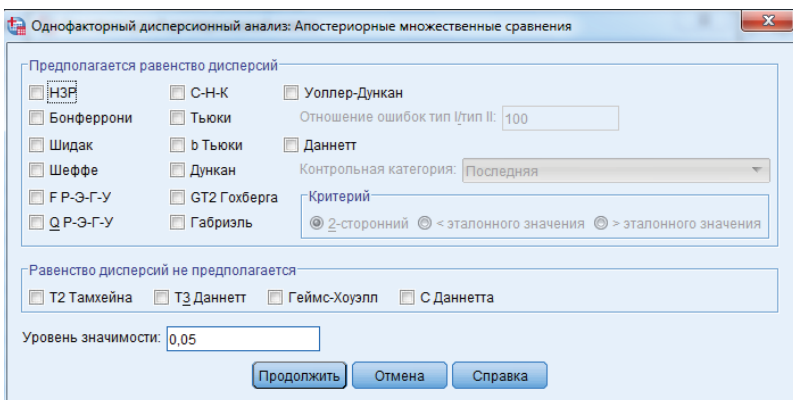


Рисунок 51. Диалоговое окно «Апостериорные» процедуры «Однофакторный дисперсионный анализ»

В своём учебном пособии Крыштановский следующим образом описывает эти методы: «Суть множественных сравнений состоит в определении различий – совпадений средних значений количественной переменной во всех возможных парах групп, определяемых градациями переменной – фактора. Иными словами, если мы проводим множественные сравнения различий в уровнях заработной платы в образовательных группах, метод множественных сравнений построит все возможные пары уровней образования и сравнит среднюю зарплату в этих парах» [4, с. 104].

Программа SPSS предлагает на выбор несколько методов множественных сравнений, при этом данные методы разделены на две категории в зависимости от проверки равенства дисперсий. То есть предполагается равенство дисперсий и равенство дисперсий не предполагается. Оптимальным вариантом будет выбрать метод из второй категории, где равенство дисперсий не предполагается, поскольку, у исследователя нет достаточных оснований полагать, что дисперсии изучаемой количественной переменной одинаковы во всех группах выбранного фактора. В

упомянутом выше учебном пособии отмечается: «Отметим, что, как показывает практика, все предлагаемые методы дают весьма близкие результаты» [4, с. 106].

С дополнительным материалом по данной теме можно ознакомиться в учебном пособии для вузов «Анализ социологических данных с помощью пакета SPSS», разработанном А.О. Крыштановским [4, с. 104-109].

Контрольные вопросы:

1. Для каких исследовательских задач целесообразно применять процедуру «Однофакторный дисперсионный анализ»? Приведите примеры.
2. Приведите примеры нулевой и альтернативной гипотез сравнения данных процедурой «Однофакторный дисперсионный анализ».
3. При каком значении значимости следует отвергнуть нулевую гипотезу?
4. Запустите процедуру «Однофакторный дисперсионный анализ» и проанализируйте полученный результат.
5. Какие переменные вы выберите для процедуры «Однофакторный дисперсионный анализ»? Обоснуйте свой выбор.

4. АНАЛИЗ ВЗАИМОСВЯЗЕЙ ПЕРЕМЕННЫХ. РЕГРЕСИОННЫЕ МОДЕЛИ

4.1. Корреляционный анализ. Парная корреляция

Изучение взаимосвязей между признаками, переменными представляет интерес в плане прослеживания причинно-следственных связей. Корреляция – это такое отношение между признаками, в котором в зависимости от изменения одного признака изменяется значение другого признака. Основной процедурой измерения корреляционной связи, используемой социологами, является парная корреляция (Bivariate correlation) [3, с. 231].

Процедура запуска корреляционного анализа реализуется по схеме «Анализ» → «Корреляции» → «Парные», представленной на рисунке 52.

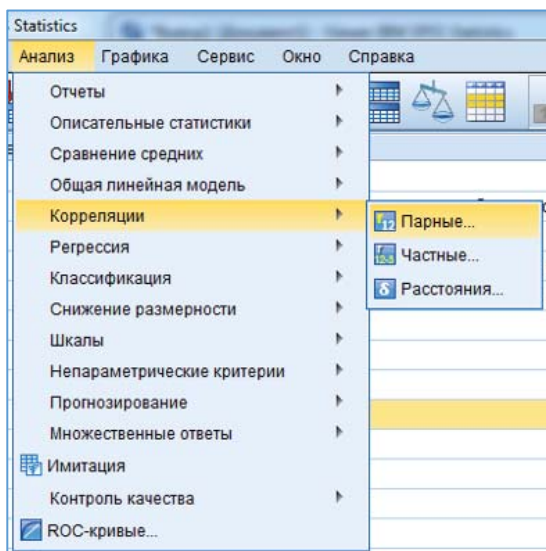


Рисунок 52. Меню «Корреляции» в общем меню «Анализ»

Основа корреляционного анализа состоит в изучении наличия связи между переменными, направления связи и силы этой взаимосвязи. Наличие связи определяется значением значимости, при значимости меньше 0,05 следует отклонить нулевую гипотезу об отсутствии взаимосвязи и принять альтернативную гипотезу о наличии взаимосвязи между переменными.

Направление связи определяется знаком коэффициентов корреляции. Положительный знак свидетельствует о прямой корреляции, то есть при увеличении одного признака (переменной) увеличивается другой признак и, наоборот, уменьшение одного признака влияет на уменьшение второго признака. Отрицательный знак показывает обратную связь, при которой при увеличении одного признака уменьшается второй признак и, наоборот, при уменьшении одного признака увеличивается другой.

Сила взаимосвязи определяется значением коэффициента корреляции, который варьирует в пределах от -1 до +1. Чем ближе значение коэффициента корреляции к нулю, тем слабее связь, чем ближе к -1, тем больше проявляется отрицательная связь и чем ближе к +1, тем сильнее проявляется положительное направление взаимосвязи между переменными.

Коэффициент корреляции – это описательная статистика, количественная мера совместной изменчивости двух переменных. Из информации, приведённой выше, следует, что сила корреляционной взаимосвязи не зависит от её направленности или знака, а определяется по модулю, т.е. абсолютному значению коэффициента корреляции. Если коэффициент корреляции равен 0, это свидетельствует о том, что обе исследуемые переменные не зависят друг от друга, между ними нет никакой взаимосвязи. В учебниках по статистике даётся интерпретация силы связи в зависимости от значения коэффициентов корреляции. В целом можно отметить о следующей интерпретации:

от 0,1 до 0,4 – слабая корреляционная взаимосвязь
от 0,4 до 0,6 – средняя корреляционная взаимосвязь
от 0,7 до 1,0 – сильная корреляционная взаимосвязь

В программе SPSS представлено три коэффициента корреляции: Пирсона, Тау-в Кендалла и Спирмана. Для количественных переменных с нормальным распределением и переменных с номинальной шкалой используется коэффициент корреляции Пирсона. В случаях, когда количественная шкала не является нормально распределённой или одна из шкал является порядковой, используются коэффициенты ранговой корреляции Спирмана или Тау-в Кендалла.

Рассмотрим пример корреляционного анализа между двумя количественными переменными: есть ли взаимосвязь между доходом и тем, сколько денег тратят люди за один поход в магазин за одеждой, обувью и аксессуарами. Если есть, то какая эта взаимосвязь?

Запускаем корреляционный анализ по следующему алгоритму: «Анализ» → «Корреляции» → «Парные». Переносим изучаемые переменные в окошко «Переменные», это вопросы «Сколько денег (в у.е.) Вы в среднем тратите за один поход в магазин за одеждой?» и «Доход респондента (в у.е.)». По умолчанию отмечен коэффициент корреляции Пирсона, так и оставляем. Далее нажимаем «ОК» (рисунок 53).

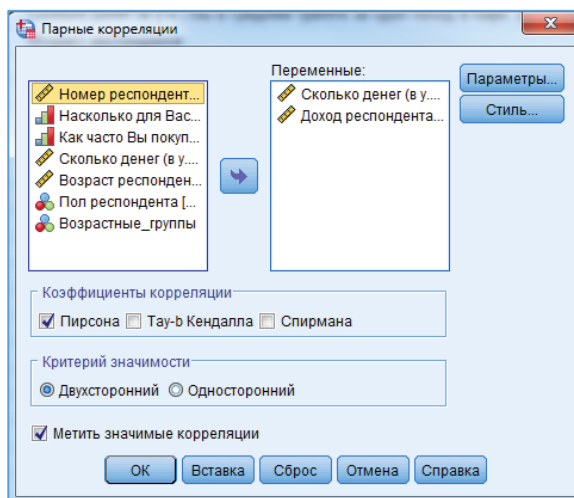


Рисунок 53. Диалоговое окно процедуры «Парные корреляции»

Результат выходит в файле «Вывод», как на рисунке 54. Здесь можно смотреть как на первую строку таблицы, так и на вторую строку, в них информация одинаковая, лишь симметрично расположена. Вначале следует обратить внимание на значимость (двухсторонняя), она меньше 0,05, значит отвергаем нулевую гипотезу и делаем вывод, что эти две переменные взаимосвязаны. Коэффициент корреляции положительный, значит взаимосвязь прямая, то есть при увеличении доходов увеличиваются и суммы затрат за один поход в магазин за одеждой. И наоборот, при уменьшении доходов уменьшаются и затраты на одежду за один поход в магазин. Коэффициент корреляции Пирсона равен 0,753, это значит, что сила взаимосвязи между изучаемыми переменными сильная.

→ **Корреляции**

Корреляции

		Сколько денег (в у.е.) Вы в среднем тратите за один поход в магазин за одеждой?	Доход респондента (в у.е.)
Сколько денег (в у.е.) Вы в среднем тратите за один поход в магазин за одеждой?	Корреляция Пирсона Знач. (двухсторонняя) N	1 200	,753** ,000 200
Доход респондента (в у.е.)	Корреляция Пирсона Знач. (двухсторонняя) N	,753** ,000 200	1 200

** . Корреляция значима на уровне 0,01 (двухсторонняя).

Рисунок 54. Результат процедуры «Парные корреляции»

Рассмотрим ещё один пример с другими коэффициентами корреляции. Рассмотрим, взаимосвязаны ли номинальная переменная «Гендер» и количественная переменная «Расходы на одежду за один поход в магазин», и, если взаимосвязаны, то каким образом. Запускаем корреляционный анализ по следующему алгоритму: «Анализ» → «Корреляции» → «Парные».

Переносим изучаемые переменные в окошко «Переменные», это вопрос «Сколько денег (в у.е.) Вы в среднем тратите за один поход в магазин за одеждой?» и «Пол респондента». По умолчанию отмечен коэффициент корреляции Пирсона, меняем галочку на коэффициент корреляции Спирмана. Далее нажимаем «ОК» (рисунок 55).

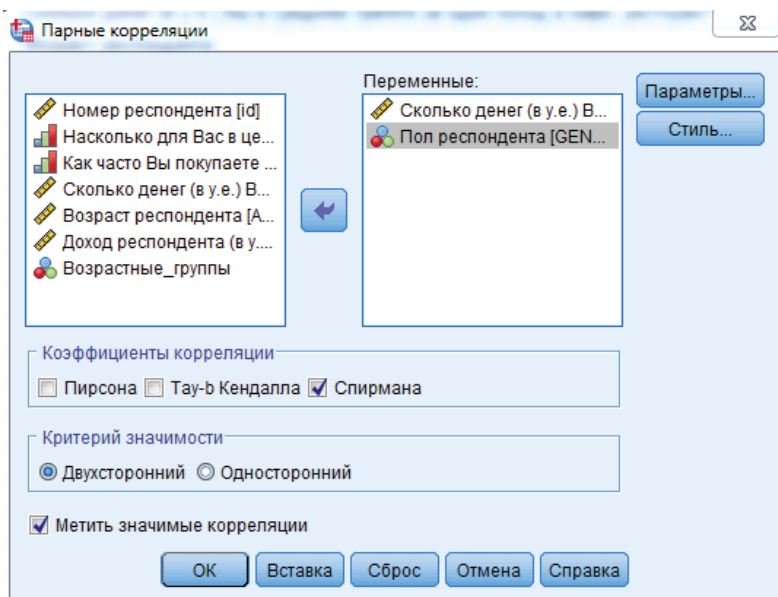


Рисунок 55. Второй пример запуска процедуры «Парные корреляции»

В первую очередь обращаем внимание на значимость (двухсторонняя), она значительно меньше 0,05, значит отвергаем нулевую гипотезу и делаем вывод, что эти две переменные взаимосвязаны (рисунок 56).

➔ Непараметрические корреляции

Корреляции			Сколько денег (в у.е.) Вы в среднем тратите за один поход в магазин за одеждой?	Пол респондента
Ро Спирмана	Сколько денег (в у.е.) Вы в среднем тратите за один поход в магазин за одеждой?	Кoeffициент корреляции Знач. (2-х сторонняя) N	1,000 . 200	,267** ,000 200
	Пол респондента	Кoeffициент корреляции Знач. (2-х сторонняя) N	,267** ,000 200	1,000 . 200

** Корреляция значима на уровне 0,01 (двухсторонняя).

Рисунок 56. Результат второго примера запуска процедуры «Парные корреляции»

Знак коэффициента корреляции положительный, значит взаимосвязь прямая, то есть при увеличении одной переменной увеличивается и другая. Переменная гендера закодирована, как 1 – мужчины и 2 – женщины, т.е. в сторону увеличения переменной представлены женщины. Таким образом, можно сделать вывод, что женщины расходуют большие суммы за один поход в магазин за одеждой. И наоборот, при уменьшении переменной гендер, а это код 1, то есть мужчины, уменьшаются и затраты на одежду за один поход в магазин. Коэффициент корреляции Спирмана равен 0,267, это значит, что сила взаимосвязи между изучаемыми переменными слабая. Это свидетельствует о том, что есть и среди мужчин те, кто тратят большие суммы на одежду за один поход в магазин.

Контрольные вопросы:

1. Запустите процедуру «Парные корреляции» между количественными переменными и проанализируйте полученный результат.
2. Запустите процедуру «Парные корреляции» между порядковой и количественной переменными и проанализируйте полученный результат
3. Для каких исследовательских задач целесообразно применять корреляционный анализ? Приведите примеры.

4. Какие переменные вы выберете для процедуры «Парные корреляции»? Обоснуйте свой выбор.

5. Приведите примеры нулевой и альтернативной гипотез корреляционного анализа, проводимого для выбранных Вами переменных.

4.2. Регрессионный анализ

Регрессионный анализ проводится с целью предсказать зависимую переменную с помощью независимых переменных, также определить их вклад влияния на определённую зависимую переменную. Предпосылкой для применения регрессионного анализа является наличие связи между переменными.

Задача регрессионного анализа – построить модель, которая позволяет оценить значение зависимой переменной на основе значений независимых переменных. Наиболее распространёнными моделями регрессионного анализа являются:

- линейная регрессия, которая позволяет оценить влияние независимых переменных на зависимую;
- логистическая регрессия, которая позволяет предсказать изменения зависимой переменной при наступлении определённого события.

В качестве зависимой переменной могут выступать количественные переменные, также распределение этой переменной должно быть близкое к нормальному. Литературные источники информируют, что в качестве независимых переменных также должны быть количественные переменные. Вместе с тем, в качестве независимых переменных могут выступать и другие типы переменных, но они предварительно преобразовываются в дихотомические переменные, их ещё называют индексные или фиктивные переменные. Например, для номинальной переменной «пол респондента» создаются две новые индексные переменные для выделения мужчин и женщин. Данные переменные будут, соответственно, равны 0 или 1 в зависимости от того, принадлежит или не принадлежит респондент к соответствующей группе. Далее, в модель включаются обе переменные.

Таким образом, прежде чем запускать регрессионный анализ, необходимо подготовить переменные к анализу. На данном

предварительном этапе следует перекодировать номинальные переменные (пол, место проживания, занятость, семейное положение, образование) в соответствующие дихотомические (индексные) переменные. Также непосредственно перед построением модели необходимо убедиться, что для всех переменных незначимые значения (затрудняюсь ответить, отказ от ответа) определены как пропущенные (missing values). Если этого не сделать, качество построенной линейной регрессии может существенно ухудшиться.

Вместе с тем для применения регрессионного анализа необходимо проверить переменные на отсутствие мультиколлинеарности, автокорреляции и гетероскедастичности. Первое означает независимость между собой переменных, выступающих предикторами, то есть отсутствие высокой корреляции между независимыми переменными в модели множественной регрессии. Данный недостаток можно решить, убрав из анализа переменные с высоким коэффициентом корреляции. Отсутствие автокорреляции является отсутствием независимости остатков, что выявляется с помощью теста Дурбина-Уотсона, который обнаруживает автокорреляцию первого порядка. Если обозначенный тест равен 0, то это означает полную положительную автокорреляцию. Если тест Дурбина-Уотсона равен 4 – означает полную отрицательную автокорреляцию, и, если тест равен 2, то это означает отсутствие автокорреляции. Отсутствие гетероскедастичности означает гомоскедастичность, то есть одинаковый разброс наблюдений вокруг линии регрессии для всех значений независимых переменных. При обнаружении гетероскедастичности необходимо разделить массив данных на несколько более-менее гомоскедастичных массивов и построить модели регрессионного анализа для каждого из них. Дополнительно о методах регрессионного анализа, также об ограничениях моделей регрессии можно ознакомиться в учебном пособии для вузов «Анализ социологических данных с помощью пакета SPSS», разработанном А.О. Крыштановским [4, с. 115-190].

Запуск регрессионных моделей реализуется по следующему алгоритму: «Анализ» → «Регрессия» → . Далее выбирается необходимый регрессионный анализ. В разных модификациях программы SPSS различен набор данных видов анализов, это

зависит от запрашиваемых характеристик при покупке лицензии на пользование программой. На рисунке 57 представлена ограниченная версия моделей регрессионного анализа.

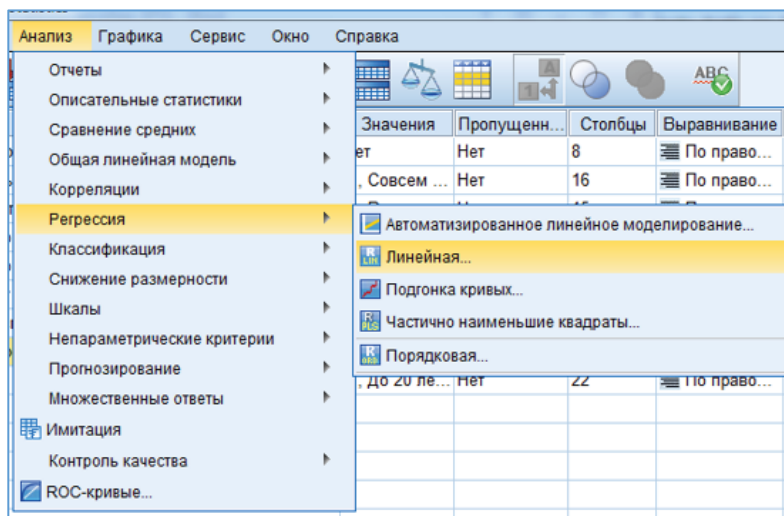


Рисунок 57. Меню «Регрессия» в общем меню «Анализ»

Рассмотрим пример регрессионного анализа. Исследовательская задача состоит в том, чтобы определить, что в большей степени влияет на то, сколько человек тратит на одежду за один поход в магазин, его доход или возраст. В базе данных «fashion.sav» есть переменные, показывающие доход изучаемой популяции, возраст и сколько денег они тратят на одежду за один поход в магазин. Все эти переменные количественные, их распределение близко к нормальному.

Запускаем регрессионный анализ по алгоритму: «Анализ» → «Регрессия» → «Линейная регрессия». Далее в окошко «Зависимые переменные» переносим переменную «Сколько денег Вы в среднем тратите за один поход в магазин за одеждой?», в окошко «Независимые переменные» переносим «Доход респондента (в у.е.)» и «Возраст респондента» (рисунок 58).

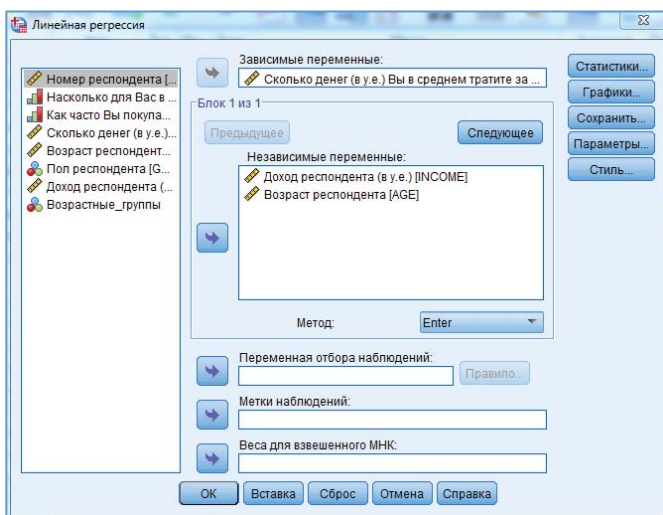


Рисунок 58. Диалоговое окно «Линейная регрессия»

Для проверки на отсутствие автокорреляции, что выявляется с помощью теста Дурбина-Уотсона, заходим в меню кнопки «Статистики» и ставим галочку в соответствующем пункте «Дурбин-Уотсон» в нижнем разделе «Остатки» (рисунок 59). Ещё раз обозначим, если данный тест равен 2, то это означает отсутствие автокорреляции. Далее нажимаем «Продолжить» и «ОК».

Результат выходит в файле «Вывод» в виде нескольких таблиц. Первая таблица информирует, какие переменные введены в модель, также отмечено в сносках, какая переменная является зависимой (рисунок 60).

В таблице «Сводка для модели» обращаем внимание на коэффициент детерминации R-квадрат, который показывает сколько процентов дисперсии зависимой переменной объясняет данная модель. Обозначенный показатель является одной из характеристик качества регрессионной прямой. В данном примере R-квадрат равен 0,572, что свидетельствует о том, что зависимая переменная на 57% объясняется данной моделью регрессионного анализа (рисунок 60). Подробнее о коэффициенте детерминации R-квадрат можно ознакомиться на стр. 123-124

учебного пособия для вузов А.О. Крыштановского «Анализ социологических данных с помощью пакета SPSS» [4]. Статистика Дурбин-Уотсон равна 2, что свидетельствует об отсутствии автокорреляции.

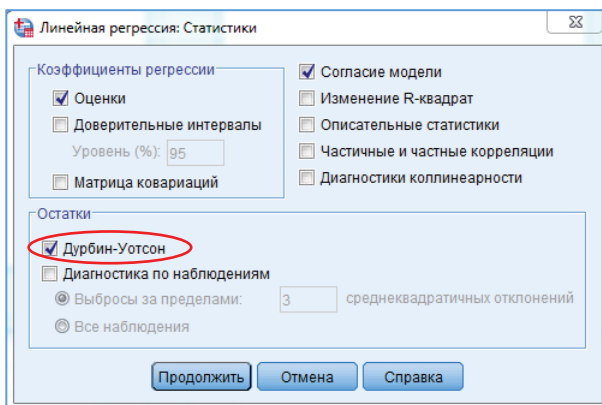


Рисунок 59. Диалоговое окно «Статистики» в процедуре «Линейная регрессия»

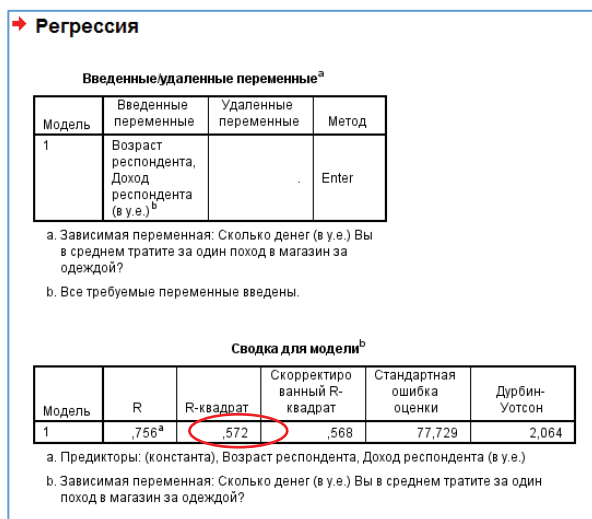


Рисунок 60. Результат запуска процедуры «Линейная регрессия»_1

Следующие таблицы результата запуска регрессионного анализа представлены на рисунке 61. В следующей таблице «ANOVA» обращаем внимание на значимость данной модели. В нашем примере значимость меньше 0,05, что свидетельствует о том, что можно отвергнуть гипотезу об отсутствии влияния, соответственно можно сделать вывод, что независимые переменные нашей регрессионной модели действительно влияют на зависимую переменную.

Какие переменные влияют в большей степени, видно из таблицы «Коэффициенты». Данные для расчёта регрессионного уравнения можно взять из столбца «B» нестандартизованных коэффициентов, где константа уравнения равна 8,912, а увеличение дохода на 1 у.е. увеличивает расходы на одежду на 0,190 у.е. В целом, доход респондентов в большей степени влияет на зависимую переменную, то есть на то, сколько они расходуют в среднем за один поход в магазин за одеждой. Об этом свидетельствует значимость данной переменной, которая меньше 0,05, и значение стандартизованного коэффициента Бета, равное 0,768. Возраст респондентов не влияет на расходы на одежду, т.е. люди в любом возрасте могут тратить примерно одинаковые суммы на одежду.

ANOVA ^a						
Модель		Сумма квадратов	ст.св.	Средний квадрат	F	Знач.
1	Регрессия	1591176,530	2	795588,265	131,682	,000 ^b
	Остаток	1190223,345	197	6041,743		
	Всего	2781399,875	199			

a. Зависимая переменная: Сколько денег (в у.е.) Вы в среднем тратите за один поход в магазин за одеждой?
b. Предикторы: (константа), Возраст респондента, Доход респондента (в у.е.)

Коэффициенты ^a						
Модель		Нестандартизованные коэффициенты		Стандартизованные коэффициенты	t	Знач.
		B	Стандартная Ошибка			
1	(Константа)	8,912	27,069		,329	,742
	Доход респондента (в у.е.)	,190	,012	,768	16,135	,000
	Возраст респондента	-1,776	1,165	-,073	-1,525	,129

a. Зависимая переменная: Сколько денег (в у.е.) Вы в среднем тратите за один поход в магазин за одеждой?

Рисунок 61. Результат запуска процедуры «Линейная регрессия» _2

Дополнительную информацию о регрессионном анализе можно получить на стр. 275-289 учебного пособия «SPSS для социологов» В.В. Пациорковского и В.В. Пациорковской [3].

Контрольные вопросы:

1. Для каких исследовательских задач целесообразно применять регрессионный анализ? Приведите примеры.
2. Какие переменные вы выберете для процедуры «Линейная регрессия»? Обоснуйте свой выбор.
3. Приведите примеры нулевой и альтернативной гипотез регрессионного анализа, проводимого для выбранных Вами переменных.
4. Запустите процедуру «Линейная регрессия» и проанализируйте полученный результат.
5. Составьте регрессионное уравнение из полученных данных и объясните его.

5. АНАЛИЗ СТРУКТУРЫ СОЦИОЛОГИЧЕСКИХ ДАННЫХ

К данной категории исследовательских процедур относятся факторный и кластерный анализы, каждый из которых изучает структуру данных, полученных в ходе количественного опроса. Изучая структуру характеристик, показателей описывающих респондентов, их взгляды или поведение, применяется факторный анализ, который нацелен на группировку переменных. При изучении структуры социальных групп применяется кластерный анализ, который нацелен на формирование групп респондентов. В обоих видах анализа методика сводится к группировке данных.

5.1. Факторный анализ

В результате факторного анализа большое количество переменных сводится к меньшему количеству вновь образованных переменных, называемых факторами. Задача исследователя – изучить и обозначить образовавшиеся факторы. В литературе по SPSS обозначено следующее объяснение: «Социологический смысл модели факторного анализа состоит в том, что измеряемые эмпирические показатели, переменные считаются следствием других, глубинных, скрытых от непосредственного измерения характеристик – латентных переменных» [4, с. 191].

Факторный анализ – это процедура, при помощи которой большое количество изучаемых переменных среди имеющихся в базе данных наблюдений, сводят к меньшему числу независимых факторов. При этом в каждый из факторов объединяются сильно коррелирующие между собой переменные. Переменные из разных факторов либо вообще не коррелируют, либо слабо коррелируют между собой. Основной смысл факторного анализа состоит в сжатии имеющейся информации.

Факторы, образующиеся в результате факторного анализа, – это скрытые, или по-другому латентные, переменные, моделируемые по признаку корреляции между переменными. Исследователю следует объективно объяснить сформировавшиеся факторы. В результате, благодаря сокращению числа переменных, представляется возможность анализировать данные в разрезе полученных факторов, количество которых меньше первоначального количества изучаемых переменных.

Порядок реализации факторного анализа: «На первом шаге процедуры факторного анализа происходит стандартизация заданных значений переменных (z -преобразование), затем при помощи стандартизованных значений рассчитывают корреляционные коэффициенты Пирсона между рассматриваемыми переменными. Исходным элементом для дальнейших расчётов является корреляционная матрица» [1, с. 368]. Математическое объяснение факторного анализа подробно представлено в учебном пособии Крыштановского «Анализ социологических данных с помощью пакета SPSS», на стр. 191-197 [4].

Условия для проведения факторного анализа:

- количество наблюдений должно быть минимум в два раза больше количества переменных;

- исходные переменные должны быть распределены неоднородно;

- переменные либо должны быть количественными, либо должны быть перекодированы в дихотомические переменные, то есть имеющие только два значения – 0 или 1, в частности это касается порядковых и номинальных переменных.

Запускается факторный анализ по следующему алгоритму: «Анализ» → «Снижение размерности» → «Факторный анализ» (рисунок 62).

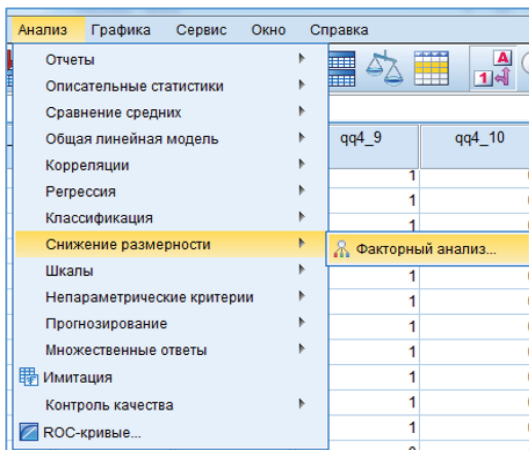


Рисунок 62. Меню запуска процедуры факторного анализа

Рассмотрим факторный анализ на конкретном примере. В файле базы данных «job» есть переменные, отражающие проблемы, которые волнуют опрошенных людей на их месте работы. Это ответы респондентов на вопрос «Какие из перечисленных проблем в наибольшей степени волнуют Вас на текущем месте работы?». Участники опроса отмечали подходящие им проблемы, предложенные в анкете опроса, вместе с тем могли писать свои проблемы в опции «Другое». Был предложен следующий набор проблем:

1. Задержки выплат заработной платы
2. Сокращение персонала/угроза увольнения
3. Уменьшение заказов и прибыли компании
4. Стиль руководства компании
5. Отношения между сотрудниками внутри коллектива
6. Снижение заработной платы
7. Отмена/сокращение социального пакета
8. Неоплачиваемые переработки и ненормированный график работы
9. Отсутствие перспектив карьерного роста
10. Низкая заработная плата
11. Нет проблем.

Исследовательская задача состоит в том, чтобы понять, какие факторы стоят за обозначенными проблемами, необходимо сократить представленные переменные до меньшего количества факторов. Для этого проходим по схеме: «Анализ» → «Снижение размерности» → «Факторный анализ», открывается диалоговое окно факторного анализа. Переносим наши переменные, отражающие проблемы на текущем месте работы, с левой стороны в правое окошко «Переменные». Кроме последней переменной «Нет проблем», поскольку она отражает отсутствие проблем, поэтому её можно проанализировать отдельно в частотном анализе и в таблицах сопряженности (рисунок 63). Все анализируемые переменные закодированы следующим образом: 1 – есть данная проблема, 0 – нет данной проблемы, то есть это дихотомические переменные.

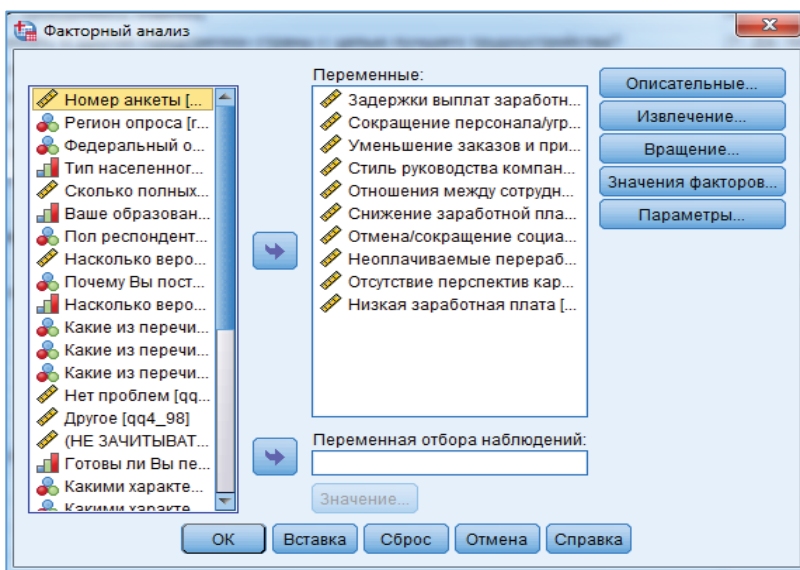


Рисунок 63. Диалоговое окно процедуры «Факторный анализ»

В меню «Извлечение» стоит по умолчанию метод главных компонент, здесь ничего не меняем. Заходим в диалоговое окошко кнопки «Вращение», в которой по умолчанию отмечена

опция «Нет», отмечаем опцию «Варимакс» – это наиболее распространённый метод вращения в факторном анализе, описанный во многих литературных источниках (рисунок 64). Нажимаем кнопку «Продолжить».

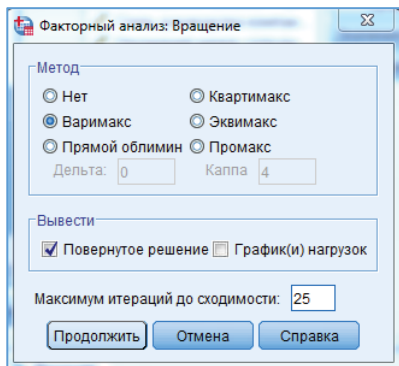


Рисунок 64. Диалоговое окно «Вращение» процедуры «Факторный анализ»

Оказавшись вновь в диалоговом окне факторного анализа, нажимаем кнопку «Значения факторов», где отмечаем опцию «Сохранить как переменные», автоматически активизируется метод «Регрессия» (рисунок 65). Далее нажимаем кнопку «Продолжить».

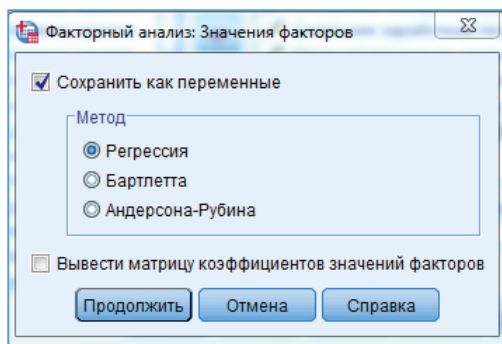


Рисунок 65. Диалоговое окно «Значения факторов» процедуры «Факторный анализ»

Затем заходим в диалоговое окошко кнопки «Параметры», где в разделе «Формат вывода коэффициентов» отмечаем опции «Отсортировать по величине» и «Не выводить коэффициенты с низкими значениями». Отметив последнюю опцию, активизируется «Абсолютное значение меньше», где меняем стоящее по умолчанию значение 0,10 на значение 0,50 и нажимаем «Продолжить» (рисунок 66).

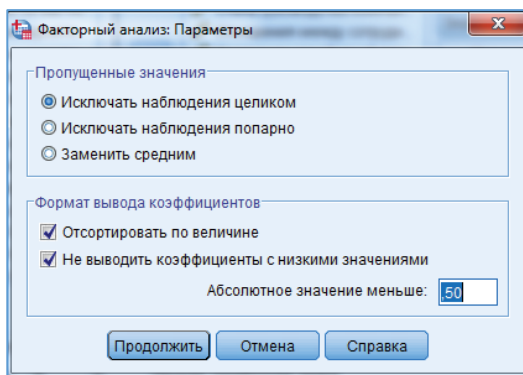


Рисунок 66. Диалоговое окно «Параметры» процедуры «Факторный анализ»

Вернувшись вновь в диалоговое окно факторного анализа, нажимаем «ОК», после чего в файле «Вывод» выходит результат запущенного анализа, который состоит из нескольких таблиц. Первая таблица (рисунок 67) под названием «Общности» показывает, какую часть дисперсии каждой переменной объясняет данная модель факторного анализа. В литературных источниках отмечается, что переменные, которые плохо объясняются построенной моделью, следует исключить из анализа [4, с. 201]. Это переменные, у которых значения в столбце «Извлечение» меньше 0,4. В представленном примере, все переменные хорошо объясняются полученной моделью факторного анализа.

Факторный анализ

Общности

	Начальная	Извлечение
Задержки выплат заработной платы	1,000	,494
Сокращение персонала/угроза увольнения	1,000	,477
Уменьшение заказов и прибыли компании	1,000	,453
Стиль руководства компании	1,000	,467
Отношения между сотрудниками внутри коллектива	1,000	,650
Снижение заработной платы	1,000	,545
Отмена/сокращение социального пакета	1,000	,712
Неоплачиваемые переработки и ненормированный график работы	1,000	,684
Отсутствие перспектив карьерного роста	1,000	,628
Низкая заработная плата	1,000	,758

Метод выделения факторов: метод главных компонент.

Рисунок 67. Результат запуска процедуры «Факторный анализ»_таблица 1

Следующая таблица «Объяснённая совокупная дисперсия» показывает процент объяснённой дисперсии (рисунок 68). Первый компонент данной таблицы всегда объясняет наибольший процент дисперсии, в данном примере он объясняет 12,5% дисперсии. Второй компонент объясняет 11,9% дисперсии, а в сумме они объясняют 24,4% и так далее. В демонстрируемом примере программой отобрано пять главных компонент, которые в сумме объясняют 58,7% совокупной дисперсии.

Объясненная совокупная дисперсия									
Компонент	Начальные собственные значения			Суммы квадратов нагрузок извлечения			Суммы квадратов загрузок вращения		
	Всего	% дисперсии	Суммарный %	Всего	% дисперсии	Суммарный %	Всего	% дисперсии	Суммарный %
1	1,343	13,428	13,428	1,343	13,428	13,428	1,255	12,547	12,547
2	1,276	12,758	26,187	1,276	12,758	26,187	1,186	11,863	24,410
3	1,137	11,369	37,556	1,137	11,369	37,556	1,171	11,707	36,117
4	1,098	10,981	48,537	1,098	10,981	48,537	1,141	11,415	47,531
5	1,014	10,142	58,679	1,014	10,142	58,679	1,115	11,147	58,679
6	,965	9,654	68,332						
7	,899	8,993	77,325						
8	,844	8,440	85,766						
9	,728	7,278	93,044						
10	,696	6,956	100,000						

Метод выделения факторов: метод главных компонент.

Рисунок 68. Результат запуска процедуры «Факторный анализ»_таблица 2

Далее следует обратить внимание на таблицу «Повернутая матрица компонентов», пропустив таблицу «Матрица компонентов». «В отличие от матрицы факторных нагрузок до вращения, матрица после вращения заметно удобнее – в ней почти все факторные нагрузки либо большие, либо маленькие, и, следовательно, такая матрица проще для интерпретации» [4, с. 201]. Таким образом, более удобная матрица факторных нагрузок получается после вращения (рисунок 69).

Программа выделила пять факторов, которые расположены в столбцах. Первый фактор объединил две переменные, отражающие проблемы «Сокращение персонала/угроза увольнения» и «Задержки выплат заработной платы» – у данных переменных довольно сильные коэффициенты корреляции с первым фактором (0,682 и 0,642). Второй фактор демонстрирует высокие корреляции с переменными: «Стиль руководства компании» (0,673), «Отношения между сотрудниками внутри коллектива» (0,611) и «Уменьшение заказов и прибыли компании» (0,542). И так далее по пяти факторам, каждый из которых содержит определённые переменные с высокими корреляциями, кроме переменной «Отсутствие перспектив карьерного роста», которая не коррелирует ни с одним из факторов (рисунок 69).

Повернутая матрица компонентов ^а					
	Компонент				
	1	2	3	4	5
Сокращение персонала/угроза увольнения	,682				
Задержки выплат заработной платы	,642				
Стиль руководства компании		,673			
Отношения между сотрудниками внутри коллектива		,611			
Уменьшение заказов и прибыли компании		-,542			
Отмена/сокращение социального пакета			,822		
Неоплачиваемые переработки и ненормированный график работы				,770	
Снижение заработной платы				,537	
Низкая заработная плата					-,843
Отсутствие перспектив карьерного роста					

Метод выделения факторов: метод главных компонент.
Метод вращения: варимакс с нормализацией Кайзера.
а. Вращение сошло за 9 итераций.

Рисунок 69. Результат запуска процедуры «Факторный анализ»_таблица 3

Задача исследователя состоит в том, чтобы объяснить полученные факторы. Первый фактор описывает проблемы сокращения персонала и задержек выплат заработной платы. Второй фактор показывает проблемы в отношении компании в целом больше психологического характера, это стиль руководства, отношения между сотрудниками и прибыли компании. Третий фактор отражает проблему, связанную с социальным пакетом на текущем рабочем месте. Четвёртый фактор описывает проблемы, связанные с несправедливой оплатой труда. Пятый фактор отражает низкую заработную плату – чем ниже заработная плата, тем выше проблема.

Поскольку мы отметили в «Значения факторов», что необходимо сохранить сформированные факторы, они были сохранены в базе данных, и далее можно проанализировать каждый из факторов в разрезе различных характеристик. Для этого можно использовать метод ранжирования данных: заходим в меню «Преобразование» и выбираем опцию «Ранжировать наблюдения» (рисунок 70).

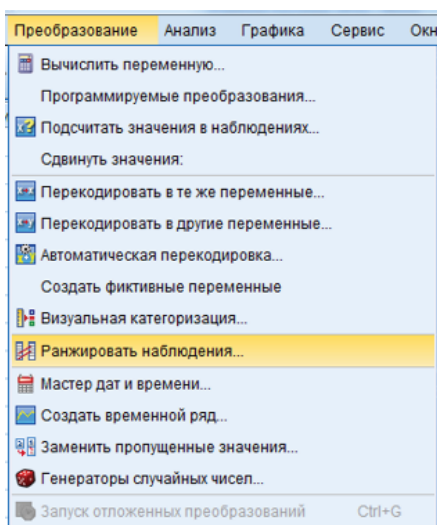


Рисунок 70. Меню «Преобразование» для запуска процедуры «Ранжировать наблюдения»

В диалоговом окне процедуры «Ранжировать наблюдения» переносим из левого сектора в правый, образованную в результате факторного анализа, переменную «REGR factor score 1 for analysis 1» (рисунок 71). Далее заходим в подраздел «Типы рангов», убираем галочку с опции «Ранг», отмечаем опцию «N разбиение» и оставляем стоящую по умолчанию цифру 4 (рисунок 72). Нажимаем «Продолжить», далее – «ОК».

В результате в базе данных формируется переменная с именем NFAC1_1 и с меткой «Percentile Group of FAC1_1» – это переменная, показывающая четыре группы перцентилей, которые необходимо обозначить в столбце «Значения». Заходим в «Значения» данной переменной и обозначаем: 1 – отсутствует, 2 – слабое, 3 – сильное и 4 – очень сильное. Также для удобства можно изменить метку данной переменной с «Percentile Group of FAC1_1» на название первого фактора «Проблемы сокращения персонала и задержек выплат заработной платы».

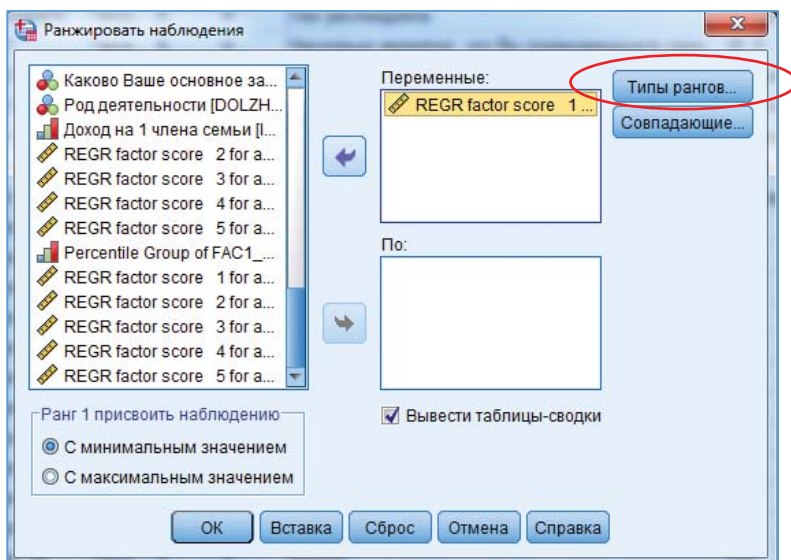


Рисунок 71. Диалоговое окно процедуры «Ранжировать наблюдения»

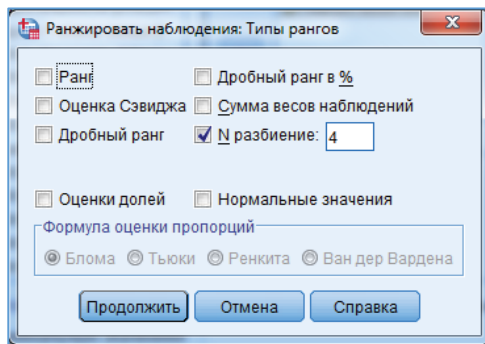


Рисунок 72. Диалоговое окошко «Типы рангов» процедуры «Ранжировать наблюдения»

Среди каких социальных групп шире распространён первый фактор, то есть проблемы сокращения персонала и задержек выплат заработной платы? С целью ответить на поставленный вопрос необходимо проанализировать новую переменную под именем NFAC1_1 «Проблемы сокращения персонала и задержек выплат заработной платы» в разрезе социально-демографических характеристик. Для этого следует применить таблицы сопряженности. К примеру, анализ в разрезе уровня образования показал, что рассматриваемая проблема сокращения персонала и задержек выплат заработной платы больше беспокоит людей без высшего образования, поскольку ответы «Сильное» и «Очень сильное» больше распространены среди населения с уровнем образования «Среднее и ниже» и «Среднее специальное» (рисунок 73).

Аналогичным образом можно проанализировать данную проблему в разрезе гендера, типа поселения и других переменных, чтобы в результате выявить, среди каких социальных групп больше распространён каждый из пяти сформировавшихся факторов проблем на текущем месте работы.

→ **Таблицы сопряженности**

Сводный отчет по наблюдениям

	Респонденты					
	Допустимо		Пропущенные		Всего	
	N	Проценты	N	Проценты	N	Проценты
Проблемы сокращения персонала и задержек выплат заработной платы * Уровень образования	500	100,0%	0	0,0%	500	100,0%

Комбинационная таблица Проблемы сокращения персонала и задержек выплат заработной платы * Уровень образования

			Уровень образования			Всего
			Среднее и ниже	Среднее специальное	Высшее	
Проблемы сокращения персонала и задержек выплат заработной платы	Отсутствует	Количество % в Уровень образования	14 17,7%	62 22,9%	44 29,3%	120 24,0%
	Слабое	Количество % в Уровень образования	23 29,1%	66 24,4%	40 26,7%	129 25,8%
	Сильное	Количество % в Уровень образования	22 27,8%	82 30,3%	24 16,0%	128 25,6%
	Очень сильное	Количество % в Уровень образования	20 25,3%	61 22,5%	42 28,0%	123 24,6%
Всего	Количество % в Уровень образования	79 100,0%	271 100,0%	150 100,0%	500 100,0%	

Рисунок 73. Результат анализа (таблица сопряженности) первого фактора в разрезе уровня образования

Вместе с тем, полученные факторы можно использовать в качестве независимых переменных в регрессионном анализе. К примеру, в представленном выше примере можно рассмотреть, какой из выявленных факторов больше влияет на уровень удовлетворенности своей работой, где зависимая переменная – уровень удовлетворенности работой, а независимыми переменными являются сформированные в результате факторного анализа пять факторов.

Контрольные вопросы:

1. Для каких исследовательских задач целесообразно применять факторный анализ? Приведите примеры.
2. Какие переменные необходимо включать в процедуру факторного анализа? Обоснуйте свой выбор.
3. В чём заключается отличие факторного анализа от кластерного анализа?
4. Запустите процедуру «Факторный анализ» и проанализируйте полученный результат.
5. Приведите примеры дальнейшего использования полученных в результате факторного анализа факторов.

5.2. Кластерный анализ

Изучая структуру социальных групп выборочной совокупности, нередко применяется кластерный анализ, целью которого является формирование групп респондентов и дальнейшее их изучение. Это ещё один вид статистического анализа, который сводится к группировке данных. Если в факторном анализе группируются переменные, то в кластерном анализе группируются единицы опроса или наблюдения, то есть респонденты или объекты исследования. По-другому можно обозначить следующим образом: в факторном анализе группируются столбцы базы данных, а в кластерном анализе группируются строки базы данных.

Представляя базу данных, как множество точек в определённом пространстве, кластерный анализ выделяет скопление точек, разделяя данные на более однородные подгруппы объектов исследования, ответивших примерно одинаковым образом.

В программе SPSS реализуется несколько методик классификации объектов исследования. В данном учебном пособии рассматривается метод кластеризации К-средними. Иерархическая кластеризация подробно рассмотрена в учебном пособии Крыштановского «Анализ социологических данных с помощью пакета SPSS», на стр. 206–212 [4]. Дискриминантный анализ подробно описан в учебном пособии В.В. Пациорковского и В.В. Пациорковской «SPSS для социологов», на стр. 317-335 [3].

Запуск различных видов кластерного анализа реализуется по схеме «Анализ» → «Классификация», далее из открывшегося списка выбирается необходимая методика кластеризации: «Двухэтапный кластерный анализ», «Кластеризация К-средними», «Иерархическая кластеризация», «Дискриминантный анализ» или «Метод ближайшего сходства» (рисунок 74).

Выбрав любой из методов кластерного анализа, открывается диалоговое окно, где можно получить информацию о выбранной процедуре, нажав кнопку «Справка», расположенную внизу справа.

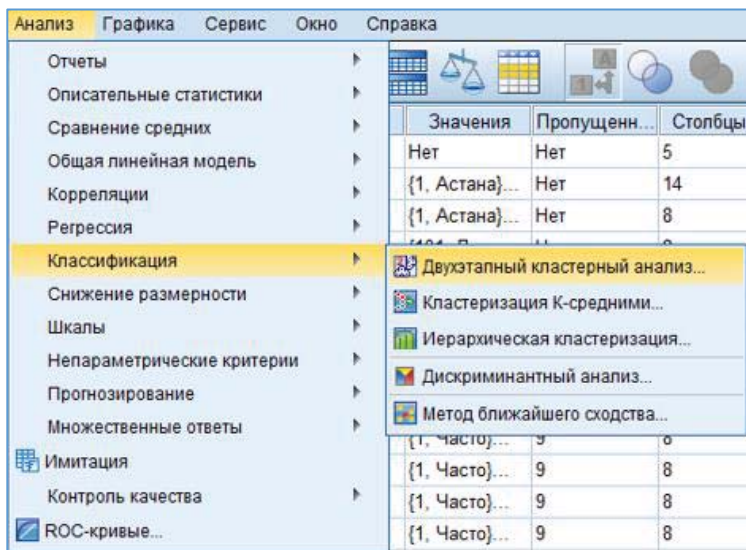


Рисунок 74. Меню классификации для запуска различных методик кластерного анализа

Рассмотрим метод кластеризации К-средними на примере базы данных социологического опроса исследования «Молодежь Центральной Азии. Казахстан», проведённого по всему Казахстану в 2014 году. В Казахстане данный проект был реализован казахстанским офисом Фонда им. Ф. Эберта совместно с исследовательским институтом «Общественное мнение». Исследование ориентировано на молодых людей в возрасте от 14 до 29 лет [5, с. 10].

Анкета исследования содержала вопрос, который задавался респондентам в ходе опроса: «Как часто вы выполняете каждый из следующих видов деятельности?», далее предлагалась таблица, где нужно было отметить подходящий ответ по каждой строке (таблица 1). К примеру, отметить 1, если часто слушает музыку, или отметить цифру 2, если иногда проводит время с друзьями и т.д. [5, с. 193].

Таблица 1 – Вопрос из анкеты социологического опроса «Как часто вы выполняете каждый из следующих видов деятельности?» с вариантами ответов

	Часто	Иногда	Редко	Никогда	Нет ответа
1. Слушаете музыку	1	2	3	4	9
2. Проводите время с друзьями	1	2	3	4	9
3. Читаете книги/журналы	1	2	3	4	9
4. Занимаетесь спортом	1	2	3	4	9
5. Смотрите телевизор	1	2	3	4	9
6. Ходите в кинотеатр	1	2	3	4	9
7. Выполняете работу по дому	1	2	3	4	9
8. Навещаете родственников	1	2	3	4	9
9. Навещаете семью (если живете отдельно)	1	2	3	4	9

Следует обратить внимание, что прежде чем запускать кластерный анализ, необходимо варианты «нет ответа» обозначить, как пропущенные значения. Для этого необходимо в базе данных по каждой переменной таблицы 1 зайти в столбец «Пропущенные» на листе «Представление Переменные», отметить опцию «Отдельные пропущенные значения» и обозначить цифру 9 в первой открытой ячейке (рисунок 75).

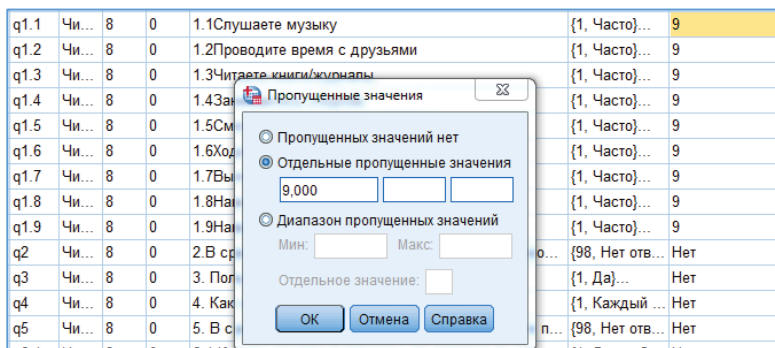


Рисунок 75. Диалоговое окно для обозначения пропущенных значений

Запуск кластеризации К-средними реализуется по схеме «Анализ» → «Классификация» → «Кластеризация К-средними».

В открывшемся диалоговом окне «Кластерный анализ методом К средних» переносим переменные вопроса, представленного в таблице 1, из правой части в левую, в окошко «Переменные» (рисунок 76).

Далее в этом же диалоговом окне обозначаем необходимое число кластеров в соответствующей опции, обычно по умолчанию стоит 2. В данном примере оставим без изменений отмеченные 2 кластера, то есть в итоге мы получим две группы респондентов (рисунок 76).

Необходимо будет сохранить распределение по группам, для этого нажимаем кнопку «Сохранить», расположенную в правом верхнем углу, и отмечаем галочкой «Принадлежность к кластеру» и «Продолжить» (рисунок 77).

Затем, нажав кнопку «Параметры», отмечаем галочкой опцию «Таблица дисперсионного анализа» в разделе «Статистика», далее – «Продолжить» (рисунок 78). Оказавшись вновь в диалоговом окне процедуры «Кластеризация К-средними», нажимаем «ОК».

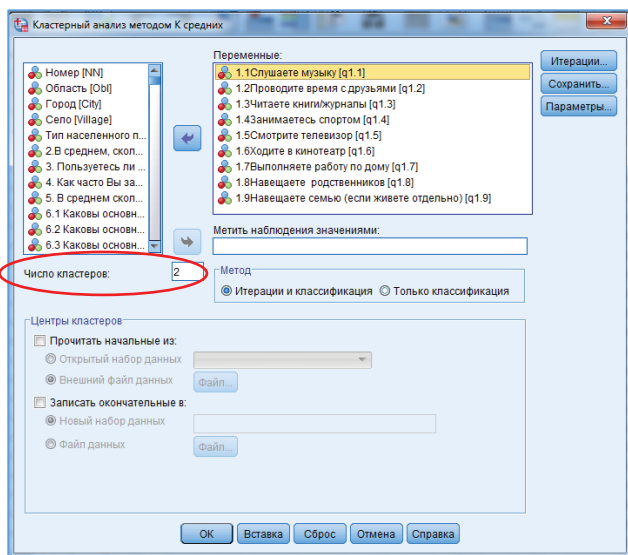


Рисунок 76. Диалоговое окно процедуры «Кластеризация К-средними»

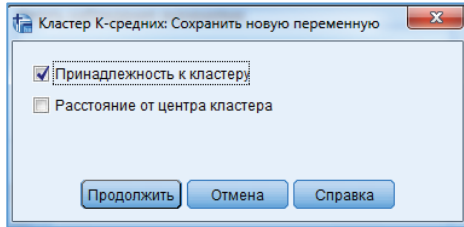


Рисунок 77. Диалоговое окно «Сохранить» процедуры «Кластеризация K-средними»

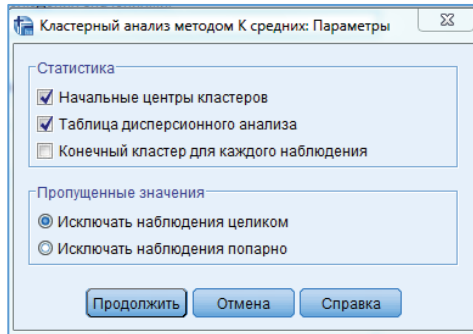


Рисунок 78. Диалоговое окно «Параметры» процедуры «Кластеризация K-средними»

В файле вывода формируется результат запущенного кластерного анализа методом «Кластеризация K-средними», состоящий из нескольких таблиц. В первой таблице представлены начальные центры кластеров, во второй таблице показана информация о хронологии итераций, в данном примере по умолчанию было отмечено 10 итераций.

Представляет интерес таблица «Конечные центры кластеров» (рисунок 79). В данной таблице представлены конечные центры двух сформированных кластеров. Первый столбец представляет кластер 1, где сгруппированы наблюдения с ответами, близкими к обозначенным. К примеру, по переменным «1.1 Слушаете музыку» и «1.2 Проводите время с друзьями» отмечена цифра 2, то есть это ответы «Иногда». Далее по переменным «1.3 Читаете книги/журналы» и «1.4 Занимаетесь спортом» отмечены цифры 3

– это ответы «Редко». Таким образом, первый кластер – это группа более пассивных людей, которые иногда слушают музыку, иногда проводят время с друзьями, также редко читают книги/журналы и редко занимаются спортом.

Второй кластер – это напротив группа более активных людей. По тем же переменным «1.1 Слушаете музыку» и «1.2 Проводите время с друзьями» отмечена цифра 1, то есть ответы «Часто». По переменным «1.3 Читаете книги/журналы» и «1.4 Занимаетесь спортом» отмечены цифры 3 – это ответы «Иногда». Эти данные свидетельствуют, что второй кластер – это группа более активных людей, которые часто слушают музыку, часто проводят время с друзьями, иногда занимаются спортом и иногда читают книги или журналы.

Конечные центры кластеров		
	Кластеризовать	
	1	2
1.1Слушаете музыку	2	1
1.2Проводите время с друзьями	2	1
1.3Читаете книги/журналы	3	2
1.4Занимаетесь спортом	3	2
1.5Смотрите телевизор	2	1
1.6Ходите в кинотеатр	3	2
1.7Выполняете работу по дому	2	1
1.8Навещаете родственников	2	1
1.9Навещаете семью (если живете отдельно)	2	1

Рисунок 79. Результат кластерного анализа методом «Кластеризация К-средними»₁

В следующей таблице «ANOVA» представлена информация о статистиках по каждой переменной. И в последней таблице показано число наблюдений в каждом кластере. К примеру, в первом кластере – 197 наблюдений и во втором кластере – 347 наблюдений (рисунок 80).

ANOVA						
	Кластеризовать		Ошибка		F	Знач.
	Средний квадрат	ст.св.	Средний квадрат	ст.св.		
1.1 Слушаете музыку	36,746	1	,362	542	101,517	,000
1.2 Проводите время с друзьями	61,792	1	,410	542	150,879	,000
1.3 Читаете книги/журналы	110,593	1	,669	542	165,241	,000
1.4 Занимаетесь спортом	215,321	1	,626	542	344,229	,000
1.5 Смотрите телевизор	8,648	1	,504	542	17,153	,000
1.6 Ходите в кинотеатр	179,472	1	,639	542	280,914	,000
1.7 Выполняете работу по дому	3,339	1	,532	542	6,274	,013
1.8 Навещаете родственников	46,333	1	,375	542	123,696	,000
1.9 Навещаете семью (если живете отдельно)	35,062	1	,421	542	83,297	,000

Число наблюдений в каждом кластере		
Кластеризовать	1	197,000
	2	347,000
Допустимо		544,000
Пропущенные		456,000

Рисунок 80. Результат кластерного анализа методом «Кластеризация К-средними»_2

Поскольку мы отметили, что необходимо сохранить принадлежность к кластеру, в базе данных была создана новая переменная под именем QCL_1 и меткой «Номер кластера наблюдения». В значениях можно обозначить метки кластеров, к примеру, обозначить их следующим образом: 1 – менее активные и 2 – более активные. Данные значения были обозначены в соответствии со смысловыми объяснениями кластеров, представленными выше.

Далее можно проанализировать полученные кластеры в разрезе различных переменных, например, чтобы понять, среди кого, в разрезе гендерного признака, больше распространены более активные (по активностям, представленным в таблице 1) люди, следует сделать таблицу сопряженности. Для этого заходим по схеме «Анализ» → «Описательные статистики» → «Таблицы сопряженности», в строки переносим новую переменную «Номер кластера наблюдения», в столбцы переносим переменную, отражающую гендер. В ячейках

обозначаем проценты по столбцу, в статистиках отмечаем критерий «Хи-квадрат» и «ОК». Результаты (рисунок 81) свидетельствуют, что статистически значимых различий не наблюдается, большинство как мужчин (63%), так и женщин (65%) более активные по пунктам, обозначенным в таблице 1.

Таблицы сопряженности					
Комбинационная таблица Номер кластера наблюдения * D1. Пол					
		D1. Пол		Всего	
		Мужской	Женский		
Номер кластера наблюдения	Менее активные	Количество	95	102	197
		% в D1. Пол	37,0%	35,5%	36,2%
	Более активные	Количество	162	185	347
		% в D1. Пол	63,0%	64,5%	63,8%
Всего	Количество	257	287	544	
	% в D1. Пол	100,0%	100,0%	100,0%	

Критерии хи-квадрат					
	Значение	ст.св.	Асимптотическая значимость (2-сторонняя)	Точная знч. (2-сторонняя)	Точная знч. (1-сторонняя)
Хи-квадрат Пирсона	,119 ^a	1	,730		
Поправка на непрерывность ^b	,065	1	,798		
Отношения правдоподобия	,119	1	,730		
Точный критерий Фишера				,789	,399
Линейно-линейная связь	,119	1	,730		
Количество допустимых наблюдений	544				

Рисунок 81. Результат таблицы сопряженности между кластерами и гендером

Контрольные вопросы:

1. В чём заключается отличие кластерного анализа от факторного анализа?
2. Для каких исследовательских задач целесообразно применять кластерный анализ? Приведите примеры.
3. Какие переменные необходимо включать в процедуру кластерного анализа? Приведите примеры.
4. Запустите процедуру кластерного анализа методом «Кластеризация К-средними» и проанализируйте полученный результат.
5. Приведите примеры дальнейшего использования полученных в результате кластерного анализа переменных.

6. ОБРАБОТКА И АНАЛИЗ МНОЖЕСТВЕННЫХ ОТВЕТОВ

Зачастую анкеты социологических опросов содержат вопросы с возможностью выбора нескольких ответов. Когда респонденту предлагается отметить все подходящие ответы или же отобрать не более 2-5 вариантов ответов на поставленный вопрос. В таком случае есть два варианта кодировки данных вопросов в базе макета программы SPSS:

- Первый вариант кодировки – создание дихотомических переменных, где возможно только два варианта ответа на каждую переменную, зачастую это кодировка 1:0. Создаётся отдельная переменная на каждый вариант ответа, и, если данный вариант ответа выбран, то вводится 1, если не выбран, то ставится 0, и так по каждой переменной, то есть по каждому варианту ответа.

- Второй вариант кодировки – создаётся переменная на основе вопроса, а варианты ответов кодируются в значениях данной переменной, таким образом создаётся такое количество кодов, сколько вариантов ответов в данном вопросе.

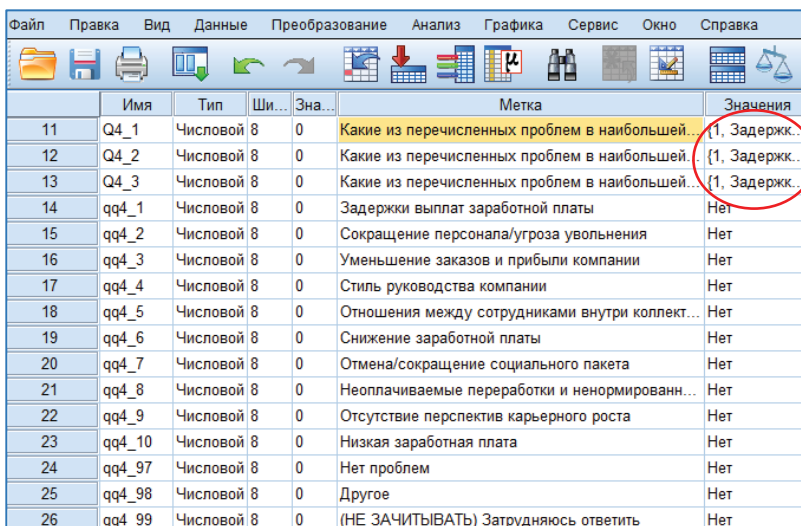
К примеру, в базе данных *job.sav, доступной на электронном ресурсе аналитического центра НАФИ [6], есть вопрос: «Какие из перечисленных проблем в наибольшей степени волнуют Вас на текущем месте работы?». Для ответа на поставленный вопрос всем респондентам предлагается отметить не более 3-х основных проблем из обозначенных ниже:

1. Задержки выплат заработной платы
2. Сокращение персонала/угроза увольнения
3. Уменьшение заказов и прибыли компании
4. Стиль руководства компании
5. Отношения между сотрудниками внутри коллектива
6. Снижение заработной платы
7. Отмена/сокращение социального пакета
8. Неоплачиваемые переработки и ненормированный график работы
9. Отсутствие перспектив карьерного роста

10. Низкая заработная плата

11. Нет проблем

На рисунке 82 представлено оба варианта кодировки данного вопроса. Переменные с qq4_1 по qq4_99 показывают первый вариант кодировки, то есть созданы отдельные переменные на каждый вариант ответа, это дихотомические переменные. Если респондент выбирает две основные проблемы, отражённые в вариантах ответов: «Задержки выплат заработной платы» и «Отмена/сокращение социального пакета», то в базе данных ответов по отмеченным вариантам ставится 1, по остальным вариантам ответов ставится 0.



	Имя	Тип	Ши...	Зна...	Метка	Значения
11	Q4_1	Числовой	8	0	Какие из перечисленных проблем в наибольшей...	{1, Задержк...
12	Q4_2	Числовой	8	0	Какие из перечисленных проблем в наибольшей...	{1, Задержк...
13	Q4_3	Числовой	8	0	Какие из перечисленных проблем в наибольшей...	{1, Задержк...
14	qq4_1	Числовой	8	0	Задержки выплат заработной платы	Нет
15	qq4_2	Числовой	8	0	Сокращение персонала/угроза увольнения	Нет
16	qq4_3	Числовой	8	0	Уменьшение заказов и прибыли компании	Нет
17	qq4_4	Числовой	8	0	Стиль руководства компании	Нет
18	qq4_5	Числовой	8	0	Отношения между сотрудниками внутри коллект...	Нет
19	qq4_6	Числовой	8	0	Снижение заработной платы	Нет
20	qq4_7	Числовой	8	0	Отмена/сокращение социального пакета	Нет
21	qq4_8	Числовой	8	0	Неоплачиваемые переработки и ненормированн...	Нет
22	qq4_9	Числовой	8	0	Отсутствие перспектив карьерного роста	Нет
23	qq4_10	Числовой	8	0	Низкая заработная плата	Нет
24	qq4_97	Числовой	8	0	Нет проблем	Нет
25	qq4_98	Числовой	8	0	Другое	Нет
26	qq4_99	Числовой	8	0	(НЕ ЗАЧИТЫВАТЬ) Затрудняюсь ответить	Нет

Рисунок 82. Варианты кодировки вопроса «Какие из перечисленных проблем в наибольшей степени волнуют Вас на текущем месте работы?»

При втором варианте кодировки созданы три переменные Q4_1, Q4_2 и Q4_3 – это сам вопрос «Какие из перечисленных проблем в наибольшей степени волнуют Вас на текущем месте работы?», а варианты ответов закодированы в значениях данных переменных (рисунок 82).

В столбце значений каждой из трёх переменных (Q4_1, Q4_2 и Q4_3) закодированы ответы на обозначенный вопрос, как

показано на рисунке 83. Если, как в выше описанном примере, респондент выбирает две основные проблемы, отражённые в вариантах ответов: «Задержки выплат заработной платы» и «Отмена/сокращение социального пакета», то в базе данных ответов в переменной Q4_1 ставится цифра 1, это значение первой отмеченной проблемы, а в переменной Q4_2 соответственно ставится значение 7, отражающее второй выбранный вариант ответа.

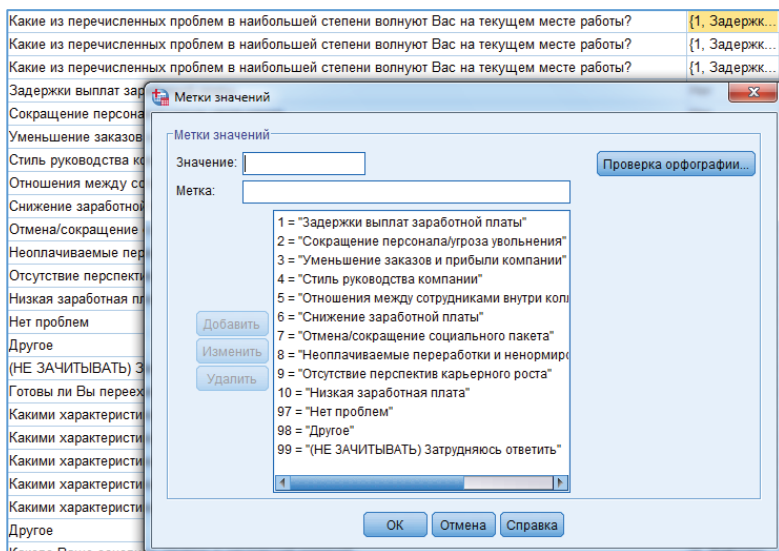


Рисунок 83. Значения ответов на вопрос «Какие из перечисленных проблем в наибольшей степени волнуют Вас на текущем месте работы?»

Анализируя переменные, закодированные первым вариантом кодировки, то есть дихотомические переменные, возможно проводить описательные статистики, а именно частотный анализ и таблицы сопряженности – данные процедуры необходимо проводить с каждой переменной.

С целью анализа переменных, закодированных вторым способом кодировки, следует проделать предварительную обработку переменных, чтобы объединить ответы всех переменных. Для этого в нашем примере заходим по схеме «Анализ» →

«Множественные ответы» → «Задать наборы множественных ответов» (рисунок 84).

Открывается диалоговое окно «Задать наборы множественных ответов». В левом окне «Параметры набора» выделяем три переменные Q4_1, Q4_2 и Q4_3 и переносим их в правое окно «Переменные в наборе». Отмечаем опцию «Категории» и указываем диапазон от 1 до максимального количества ответов. Можно обозначать верхний предел цифрой больше, чем общее количество ответов, как в данном примере, где поставлена цифра 20, хотя вариантов ответов 13. Далее в строке «Имя» прописываем имя задаваемого набора множественных ответов, в данном примере обозначено имя Q4all, после этого становится активной кнопка «Добавить», следует её нажать (рисунок 85).

После нажатия кнопки «Добавить» обозначенное имя переменной появится в окне «Наборы множественных ответов», далее следует нажать кнопку «Заккрыть», расположенную внизу диалогового окна «Задать наборы множественных ответов».

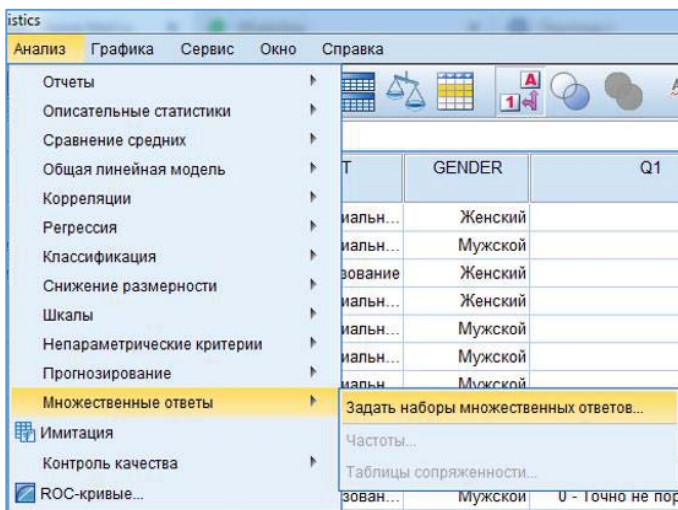


Рисунок 84. Меню для обработки множественных ответов

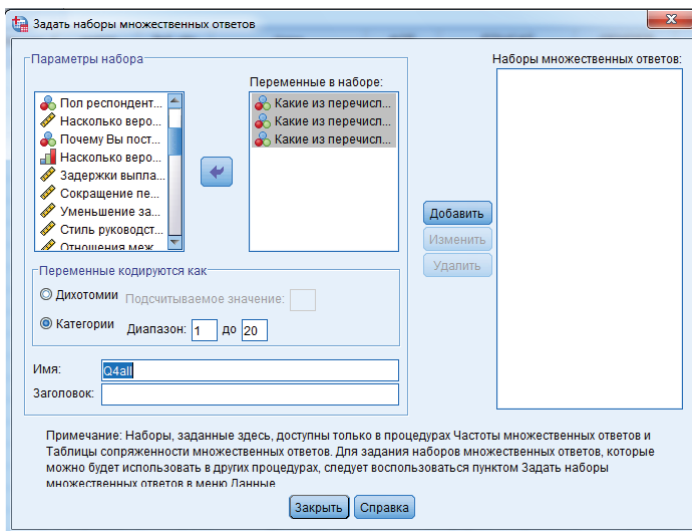


Рисунок 85. Диалоговое окно «Задать наборы множественных ответов»

Далее для анализа созданного набора множественных ответов следует вновь зайти по алгоритму «Анализ» → «Множественные ответы» → «Задать наборы множественных ответов», представленному на рисунке 84. После создания переменной множественных ответов процедуры «Частоты» и «Таблицы сопряженности» становятся активными.

Для проведения частотного анализа необходимо зайти в процедуру «Частоты», откроется диалоговое окно «Частоты для множественных ответов», где следует перенести созданную переменную, в данном примере Q4all, из левого окошка «Наборы множественных ответов» в правое «Таблицы для» и нажать «ОК» (рисунок 86).

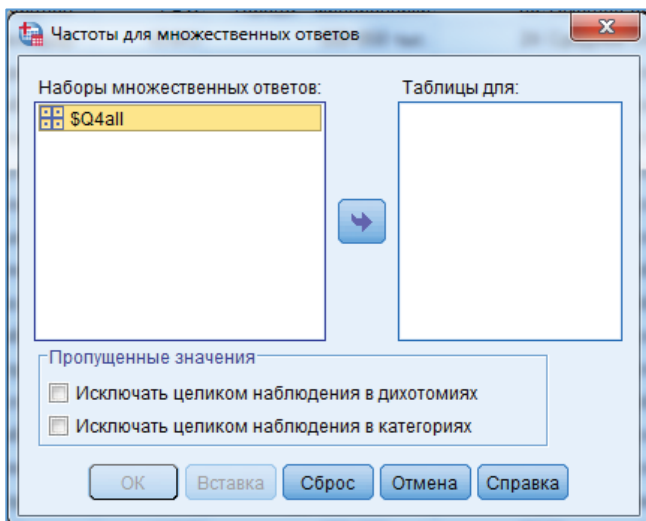


Рисунок 86. Диалоговое окно «Частоты для множественных ответов»

В файле вывода выходит сводный результат частотного анализа множественных ответов, где возможно определить, что наиболее распространённой проблемой является «Снижение заработной платы», об этом отметили 38% респондентов, далее следует «Уменьшение заказов и прибыли компании» (25,3%) и «Неоплачиваемые переработки и ненормативный график работы» (24,8%) (рисунок 87).

В таблице частотного анализа множественных ответов следует анализировать последний столбец «Процент наблюдений», поскольку данный процент учитывает все множественные ответы. Для удобства можно скопировать полученную таблицу в файл Excel и отсортировать данные ответов по последнему столбцу.

Множественные ответы

Периодичность \$Q4all

		Ответы		Процент наблюдений
		N	Проценты	
\$Q4all ^a	Задержки выплат заработной платы	81	9,8%	19,7%
	Сокращение персонала/угроза увольнения	83	10,1%	20,2%
	Уменьшение заказов и прибыли компании	104	12,6%	25,3%
	Стиль руководства компании	89	10,8%	21,7%
	Отношения между сотрудниками внутри коллектива	58	7,0%	14,1%
	Снижение заработной платы	156	18,9%	38,0%
	Отмена/сокращение социального пакета	44	5,3%	10,7%
	Неоплачиваемые переработки и ненормированный график работы	102	12,4%	24,8%
	Отсутствие перспектив карьерного роста	94	11,4%	22,9%
	Низкая заработная плата	14	1,7%	3,4%
Всего		825	100,0%	200,7%

а. Сгруппировать

Рисунок 87. Результат частотного анализа множественных ответов

Аналогичная процедура для запуска таблиц сопряженности производится по алгоритму: «Анализ» → «Множественные ответы» → «Таблицы сопряженности», представленному на рисунке 84 (после создания переменной множественных ответов процедура «Таблицы сопряженности» будет активной). Откроется диалоговое окно «Таблицы сопряженности для множественных ответов». В данном окне следует перенести созданную переменную, в рассматриваемом примере это Q4all, из нижнего левого окошка «Наборы множественных ответов» в правое верхнее окошко «Строки». В окошко «Столбцы» необходимо перенести переменную, в разрезе которой следует сделать таблицу сопряженности. В данном примере перенесём переменную гендер, далее в скобках возникают вопросы, то есть

необходимо задать диапазон, для чего заходим в соответствующую кнопку «Задать диапазон» и обозначаем минимум 1, максимум 2. В данном примере, в строке «Максимум» обозначено 2, поскольку переменная гендер содержит только две категории (Мужчины и Женщины), если переменная содержит больше категорий, например, «Уровень образования» содержит 5 категорий, то в строке «Максимум» ставим цифру 5. Сколько категорий в анализируемой переменной, которую перемещаем в «Столбцы», такую цифру и обозначаем в строке «Максимум» диалогового окна «Задать диапазон переменной». Далее в обозначенном окне необходимо нажать кнопку «Продолжить» (рисунок 88).

Затем следует зайти во вкладку «Параметры» (отмечено красным кругом на рисунке 88) и отметить проценты в ячейках. В зависимости от целей и задач анализа проценты отмечаются либо по строке, либо по столбцу, а также всего, то есть по всем респондентам таблицы.

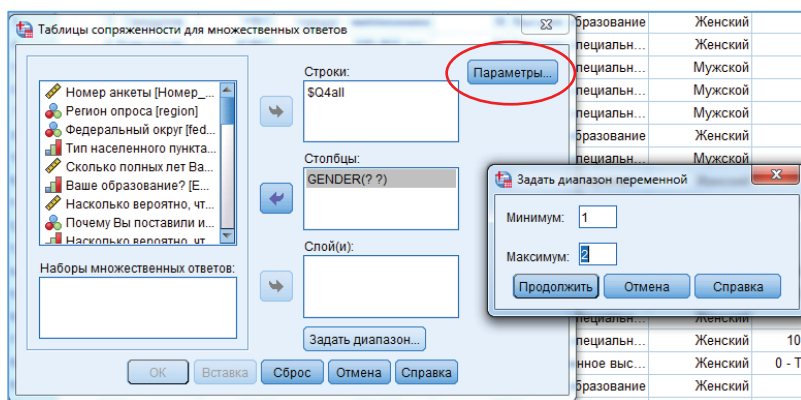


Рисунок 88. Запуск анализа таблиц сопряженности множественных ответов

На рисунке 89 представлены результаты запуска описательного анализа таблиц сопряженности между вопросом с множественными ответами «Какие из перечисленных проблем в наибольшей степени волнуют Вас на текущем месте работы?» и

гендером. В данном примере проценты были заданы по столбцу, то есть в разрезе гендера.

Множественные ответы

Комбинационная таблица \$Q4all * GENDER

			Пол респондента		Всего
			Мужской	Женский	
\$Q4all ^a	Задержки выплат заработной платы	Количество	45	36	81
		% в GENDER	23,2%	16,6%	
	Сокращение персонала/угроза увольнения	Количество	41	42	83
		% в GENDER	21,1%	19,4%	
	Уменьшение заказов и прибыли компании	Количество	46	58	104
		% в GENDER	23,7%	26,7%	
	Стиль руководства компании	Количество	52	37	89
		% в GENDER	26,8%	17,1%	
	Отношения между сотрудниками внутри коллектива	Количество	23	35	58
		% в GENDER	11,9%	16,1%	
	Снижение заработной платы	Количество	69	87	156
		% в GENDER	35,6%	40,1%	
	Отмена/сокращение социального пакета	Количество	17	27	44
		% в GENDER	8,8%	12,4%	
	Неоплачиваемые переработки и ненормированный график работы	Количество	53	49	102
		% в GENDER	27,3%	22,6%	
	Отсутствие перспектив карьерного роста	Количество	45	49	94
		% в GENDER	23,2%	22,6%	
	Низкая заработная плата	Количество	6	8	14
		% в GENDER	3,1%	3,7%	
Всего		Количество	194	217	411

Процентные отношения и итоги основаны на респондентах
а. Сгруппировать

Рисунок 89. Результат анализа таблиц сопряженности множественных ответов в разрезе гендера

Представленные результаты показывают, что среди женщин чаще распространена проблема «Снижение заработной платы» (40,1%), чем среди мужчин (35,6%). Также женщины чаще отмечали проблему «Отношения между сотрудниками внутри коллектива» – о данной проблеме отметили 16,1% женщин и 11,9% мужчин. Мужчины чаще, чем женщины отмечали такую проблему, как «Задержки выплат заработной платы», – о данной проблеме отметили 23,2% мужчин и 16,6% женщин. Вместе с тем, мужчины чаще, чем женщины, отмечали, что их волнует, как

происходит руководство компанией, то есть мужчины чаще называли проблему «Стиль руководства компании» – 26,8% мужчин и 17,1% женщин.

Контрольные вопросы:

1. Объясните разницу двух способов кодировки множественных ответов.
2. Для каких исследовательских целей целесообразнее использовать каждый из двух видов кодировки множественных ответов?
3. Задайте набор множественных ответов, сделайте частотное распределение и объясните полученные результаты.
4. Задайте набор множественных ответов, запустите анализ таблиц сопряженности в разрезе какой-нибудь переменной и объясните полученные результаты.
5. Для каких исследовательских задач целесообразно обозначать в таблицах сопряженности проценты по строке, а для каких задач – по столбцу? Приведите примеры.

7. СИНТАКСИС В SPSS

В программе SPSS для программирования команд и других операций используются специальные операции, прописываемые в редакторе синтаксиса, – Syntax. С помощью данного редактора предоставляется возможность запускать различные виды анализа, перекодировать или вычислять переменные. Одним из преимуществ редактора синтаксиса является то, что в нём можно сохранить историю проделанных операций и видов анализа. Можно поделиться данным файлом с коллегами или открыть самому через некоторое время и увидеть процедуры, проделанные с базой данных, вновь воспроизвести их или изменить.

Открыть редактор синтаксиса можно несколькими способами:

- В случае, если создаётся новый файл синтаксиса, – то через меню «Файл» на верхней панели базы данных в программе SPSS, по алгоритму: «Файл» → «Создать» → «Редактор синтаксиса» (рисунок 90).

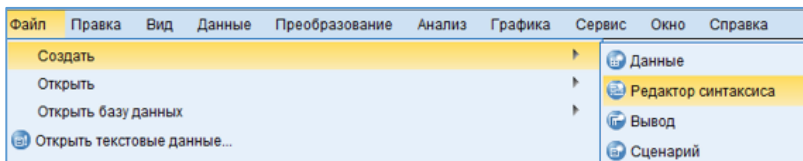


Рисунок 90. Алгоритм создания нового файла редактора синтаксиса

- В случае открытия уже существующего файла синтаксиса необходимо пройти по алгоритму: «Файл» → «Открыть» → «Синтаксис», тогда откроется окно для выбора файла редактора синтаксиса (рисунок 91).

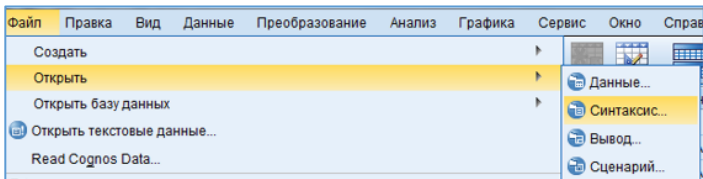


Рисунок 91. Алгоритм открытия существующего файла редактора синтаксиса

- При запуске любого анализа в диалоговом окне процедуры внизу есть ряд кнопок, рядом с «ОК» есть кнопка «Вставка». Для открытия редактора синтаксиса следует вместо «ОК» нажать «Вставка» (рисунок 92). После этого автоматически создаётся и открывается файл редактора синтаксиса (рисунок 93).

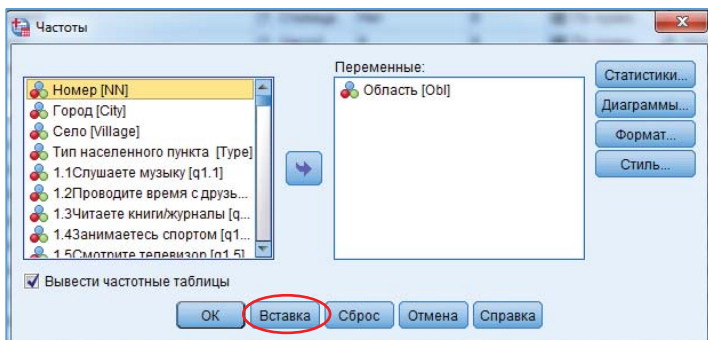


Рисунок 92. Запуск редактора синтаксиса через диалоговое окно процедуры

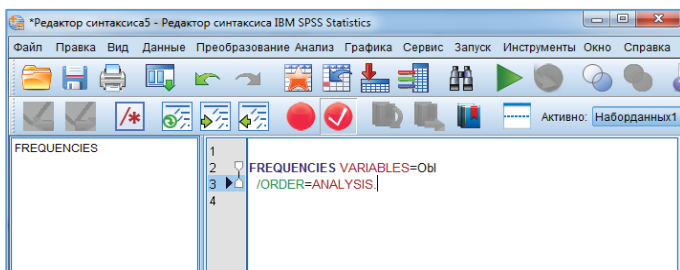


Рисунок 93. Автоматически созданный файл редактора синтаксиса после нажатия «Вставка» в диалоговом окне процедуры

- В последних версиях программы SPSS синтаксис автоматически записывается в файл вывода, это делается по умолчанию без необходимости что-либо указывать (рисунок 94). Далее обозначенный синтаксис можно из файла вывода скопировать в файл редактора синтаксиса.

```

CROSSTABS
  /TABLES=Type BY d1
  /FORMAT=AVALUE TABLES
  /STATISTICS=CHISO
  /CELLS=COUNT COLUMN
  /COUNT ROUND CELL.
  
```

→ **Таблицы сопряженности**

Сводный отчет по наблюдениям

	Респонденты					
	Допустимо		Пропущенные		Всего	
	N	Проценты	N	Проценты	N	Проценты
Тип населенного пункта * D1. Пол	1000	100,0%	0	0,0%	1000	100,0%

Комбинационная таблица Тип населенного пункта * D1. Пол

			D1. Пол		Всего
			Мужской	Женский	
Тип населенного пункта	Столица	Количество	26	27	53
		% в D1. Пол	5,2%	5,4%	5,3%
	Город республиканского значения	Количество	47	50	97
		% в D1. Пол	9,4%	10,0%	9,7%

Рисунок 94. Редактор синтаксиса в файле «Вывод»

С целью запустить процедуру из файла редактора синтаксиса следует выделить команду, далее либо нажать на зелёную стрелку на панели инструментов, либо нажать правую кнопку мыши и в открывшемся меню выбрать «Запустить выделенный фрагмент» (рисунок 95).

В представленном на рисунке примере запустится процедура частотного анализа переменной под именем «Obl», отражающей области опроса. После нажатия «Запустить выделенный фрагмент» в файле вывода появится результат данного частотного анализа.

Аналогично можно запускать процедуры по перекодировке или вычислению переменных, а также другие аналитические процедуры. Следует обратить внимание, что при выделении фрагментов в редакторе синтаксиса важно захватить строки полностью, включая точки, поскольку каждый знак имеет значение.

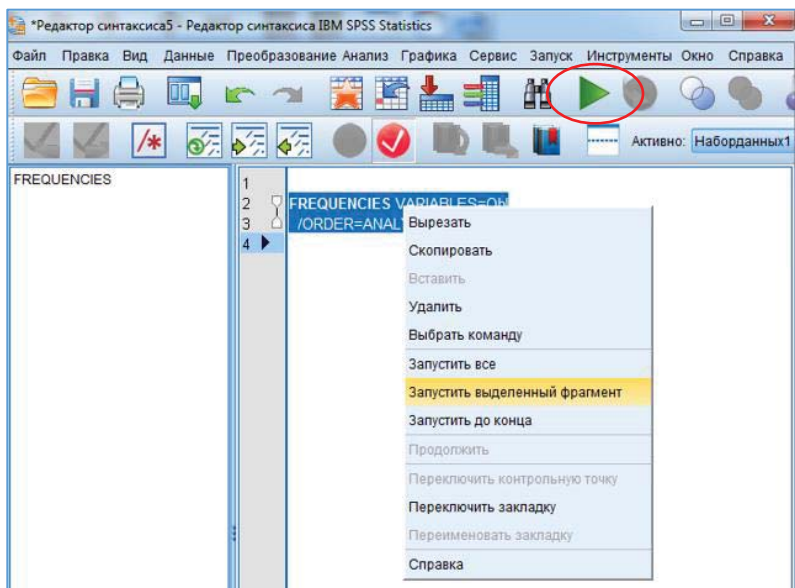


Рисунок 95. Запуск редактора синтаксиса для получения процедуры частотного анализа

С дополнительной информацией по работе с редактором синтаксиса в программе SPSS можно ознакомиться на сайте аналитического центра НАФИ по ссылке: <https://naf1.ru/upload/spss/NAFI%20-%2013.lecture.pdf>, а также на сайте специального образовательного проекта Raynald's SPSS Tools по ссылке: <https://www.spsstools.net/ru/syntax/learning-syntax/>.

Контрольные вопросы:

1. Что такое редактор синтаксиса в программе SPSS?
2. В чём заключается польза для исследователя от работы в редакторе синтаксиса?
3. Для каких исследовательских задач целесообразно применять редактор синтаксиса? Приведите примеры.
4. Запустите процедуру любого анализа через редактор синтаксиса.
5. Сделайте модификацию переменных через редактор синтаксиса в SPSS.

8. ПОСТРОЕНИЕ ГРАФИКОВ В SPSS

Представление данных в программе SPSS возможно таблицами, формирующимися в файле «Вывод», которые можно копировать и вставлять в различных форматах вставки. Также данные возможно представлять в виде графиков. Помимо визуального представления данных графики дают возможность изучить данные, сделать на их основе анализ.

Существует два способа формирования графиков. Один из этих способов создания графиков в программе SPSS – это использование дополнительных кнопок «Графики», расположенных в процедурах анализа, к примеру, при запуске частотного анализа, таблиц сопряженности, регрессионного анализа. Следует отметить, что в случае таблиц сопряженности необходимо отметить галочкой опцию «Вывести кластеризованные столбчатые диаграммы».

Второй способ – это задать графики в меню Графика – Мастер диаграмм (рисунок 96).

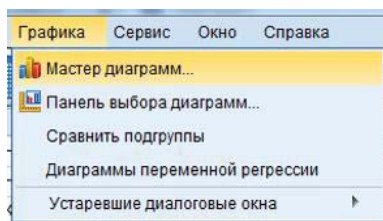


Рисунок 96. Меню «Графика» для запуска «Мастер диаграмм» с целью построения графиков

В открывшемся диалоговом окне «Мастер диаграмм» представляется возможность выбрать различные виды графиков: столбцы, линии, круги, точки, гистограммы, максимум-минимум, ящики с усами и двойные оси. Список для выбора графиков расположен внизу. Выбрав нужную диаграмму, по ней следует щёлкнуть два раза для перемещения в верхнее окно. Для примера

выберем график «Простые столбики ошибок», который позволяет сравнить средние значения изучаемого параметра в разрезе различных выборок, то есть в социальных группах. Далее в верхнем окне необходимо выбрать переменную для расположения по оси X, к примеру «Пол респондента», и переменную для расположения по оси Y, к примеру «Сколько денег в среднем Вы тратите за один поход в магазин за одеждой?», как показано на рисунке 97. Далее нажимаем ОК и в файле «Вывод» формируется график.

Полученный график, представленный на рисунке 90, свидетельствует, что женщины в среднем тратят за один поход в магазин за одеждой больше, чем мужчины.

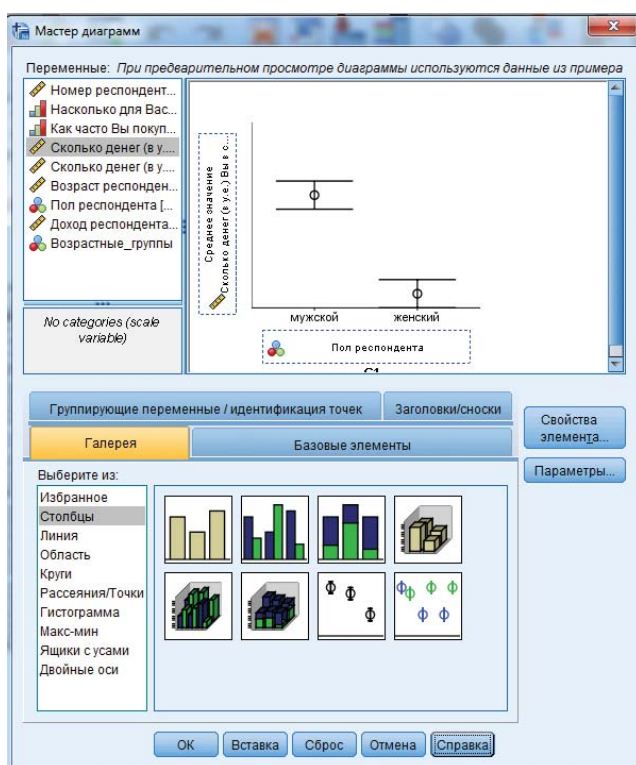


Рисунок 97. Диалоговое окно «Мастер диаграмм» для построения графиков

Для редактирования полученного графика, представленного на рисунке 98, необходимо в файле «Вывод» щёлкнуть по нему два раза, в результате чего откроется «Редактор диаграмм» – в данном диалоговом окне есть своя панель инструментов, предназначенных для редактирования созданного графика, изменения его вида, шрифтов и т.д. Для решения возникших вопросов можно зайти в раздел «Справка», где представлена информация по различным вопросам в программе SPSS.

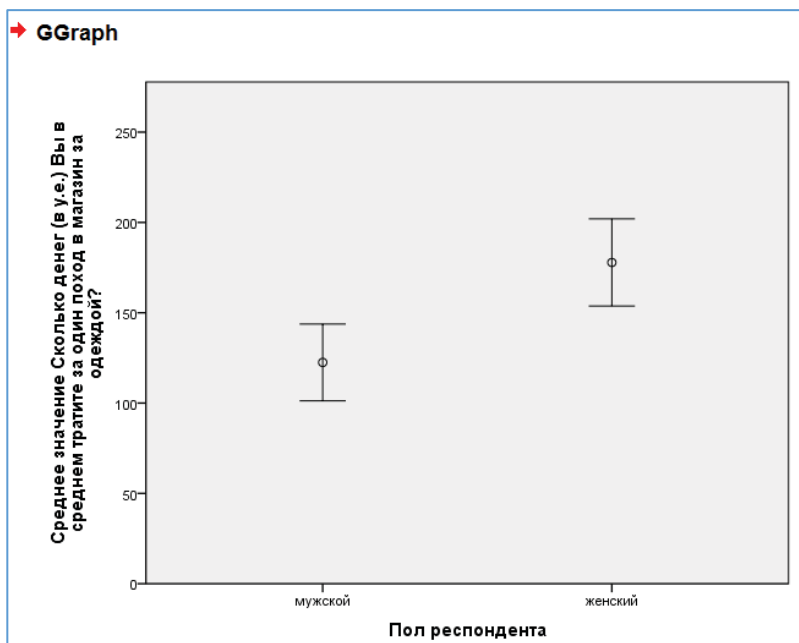


Рисунок 98. График, полученный в результате запуска через диалоговое окно «Мастер диаграмм»

С дополнительной информацией о графическом представлении данных в программе SPSS можно ознакомиться в главе 12 учебного пособия «SPSS для социологов», на стр. 253-273 [3].

Контрольные вопросы:

1. Постройте график частотного распределения и объясните полученные данные.
2. Запустите график таблиц сопряженности и объясните полученные данные.
3. Для каких исследовательских задач целесообразно применять график гистограммы? Приведите примеры.
4. Приведите примеры, какие переменные необходимо использовать в графике линий, объясните почему.
5. В каких переменных следует применять круговой график? Приведите примеры.

ЗАКЛЮЧЕНИЕ

Современные цифровые технологии предоставляют много возможностей для проведения исследований по изучению социального мира и соответственно являются востребованными профессионалы по обработке и анализу данных. Предполагается, что ценность обозначенных специалистов будет и дальше актуальной, поскольку каждая организация стремится максимально использовать всевозможные внутренние и внешние источники информации для понимания своих целевых групп из числа клиентов и заказчиков, а также различных социальных групп в целом.

Настоящее учебное пособие способствует не только лучшему пониманию материала, изучаемого в рамках курса «Методы обработки и анализа социологической информации (на базе SPSS)», но также способствует формированию навыков работы в программе SPSS при самостоятельном изучении. Ожидается, что по результатам освоения материалов, представленных в данном пособии, читатель будет способен работать в обозначенной программе, а именно:

- формировать макеты базы данных, определять виды переменных и шкал;
- осуществлять ввод результатов исследований или статистических данных;
- модифицировать и вычислять переменные;
- обрабатывать и анализировать информацию.

Обозначенные компетенции расширяют профессиональные возможности обучающихся и позволяют быть конкурентоспособными на рынке труда.

Следует отметить, что данное учебное пособие сфокусировано на основном материале, овладение которым позволит обучающимся начать обрабатывать и анализировать результаты своих исследований или доступные базы данных исследований, проведённых другими исследователями. С дополнительной информацией по процедурам в программе SPSS, не представленной

в настоящем учебном пособии, возможно ознакомиться в справочном разделе, перейти на который можно, нажав на соответствующую кнопку «Справка», расположенную в диалоговом окне каждой процедуры.

Для совершенствования навыков работы в программе SPSS рекомендуется наращивание опыта, то есть практическое применение представленного материала, а также ознакомление с дополнительной литературой по данной теме. Совершенные навыки по обработке и анализу данных позволят обучающимся повысить свою квалификацию, объективно анализировать ситуацию, основываясь не на предположениях и догадках, а на основе эмпирических и статистических данных. В свою очередь это будет способствовать развитию науки в академической сфере, а также принятию информированных управленческих решений в сфере бизнеса и государственного администрирования.

СПИСОК ИСТОЧНИКОВ

1. Бююль Ахим, Цёфель Петр. SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей: Пер. с нем. / Ахим Бююль, Петр Цёфель. – СПб.: ООО «ДиаСофтЮП», 2005 – 608 с.
2. Электронный ресурс IBM SPSS Statistics, режим доступа <https://www.ibm.com/ru-ru/products/spss-statistics/resources>.
3. Пациорковский В.В., Пациорковская В.В. SPSS для социологов: Учебное пособие / ИСЭПН РАН. – М., 2005 – 433 с.
4. Крыштановский А.О., Анализ социологических данных с помощью пакета SPSS: Учебное пособие для вузов. – М.: Изд. дом ГУ ВШЭ, 2006. – 281 с.
5. Молодежь Центральной Азии. Казахстан. На основе социологического опроса. Под научным руководством проф. Клауса Хуррельманн (Германия, Берлин). – Издание второе. – Алматы, 2016. – 281 с.
6. Аналитический центр НАФИ, электронный ресурс, режим доступа: <https://nafi.ru/>.

Учебное издание

Шабденова Айжан Базархановна

**МЕТОДЫ ОБРАБОТКИ И АНАЛИЗА
СОЦИОЛОГИЧЕСКОЙ
ИНФОРМАЦИИ НА БАЗЕ SPSS**

Учебное пособие

Редактор *Г. Бекбердиева*
Компьютерная верстка *А. Алдашевой*
Дизайн обложки *А. Аймбетова*

В оформлении обложки использованы иллюстрации с сайта [freepik.com](https://www.freepik.com)

ИБ № 14892

Подписано в печать 20.02.2023. Формат 60x84 1/16.

Бумага офсетная. Печать цифровая. Объем 7,375 п.л.

Тираж 50 экз. Заказ № 33.

Издательский дом «Қазақ университеті»
Казахского национального университета им. аль-Фараби.
050040, г. Алматы, пр. аль-Фараби, 71. КазНУ

Отпечатано в типографии издательского дома «Қазақ университеті»