

Министерство образования и науки РФ
Уральский государственный экономический университет



Н. Б. Изакова, А. Ю. Журавлева

Маркетинговые исследования с применением SPSS

Рекомендовано Учебно-методическим советом
Уральского государственного экономического университета
в качестве учебного пособия

Екатеринбург
2013

УДК 339.1+004.045(075.8)
ББК 65.291.31+32.973.26-018.2(я75)
ИЗ2

Рецензенты:

кафедра экономики и управления Уральского института Российской академии народного хозяйства и государственной службы при Президенте РФ;

Л. С. Ружанская – доктор экономических наук, заведующая кафедрой теории и практики менеджмента Высшей школы экономики и менеджмента Уральского федерального университета имени первого Президента России Б. Н. Ельцина;

Л. П. Пискунова – кандидат философских наук, доцент кафедры теории и практики менеджмента Уральского федерального университета имени первого Президента России Б. Н. Ельцина

Изакова, Н. Б.

ИЗ2 **Маркетинговые исследования с применением SPSS : учеб. пособие / Н. Б. Изакова, А. Ю. Журавлева ; М-во образования и науки РФ, Урал. гос. экон. ун-т. – Екатеринбург : Изд-во УрГЭУ, 2013. – 134 с.**

Назначение данного учебного пособия – помочь активному освоению основных методов проведения маркетинговых исследований с использованием программного комплекса SPSS. Учебное пособие содержит задания для самостоятельной и практической работы студентов и краткое описание основных методов статистического анализа, применяемых при обработке маркетинговой информации. Рассмотрены примеры практического использования каждого метода анализа; приводятся детальные пошаговые инструкции пользования программой SPSS; показано, как проводить поэтапную интерпретацию полученных результатов.

Предназначено для студентов, а также преподавателей высших учебных заведений.

**УДК 339.1+004.045(075.8)
ББК 65.291.31+32.973.26-018.2(я75)**

© Изакова Н. Б., Журавлева А. Ю.,
2013

© Уральский государственный
экономический университет, 2013

Введение

Эффективная маркетинговая деятельность невозможна без объективной маркетинговой информации, основную часть которой дают маркетинговые исследования. Они играют ключевую роль в реализации маркетингового подхода в бизнесе, поскольку охватывают факторы микро- и макросреды, оценку параметров рынка. На основе полученной информации принимаются решения о сегментировании рынка, выборе целевого сегмента, разработке комплекса маркетинга и его реализации. При проведении маркетинговых исследований необходимо не только собрать данные, но и грамотно их обработать, получить новую объективную информацию и дать содержательную интерпретацию результатов исследования.

Современное маркетинговое исследование требует умелого использования компьютерной технологии, четкой организации информационно-аналитического обеспечения и сопровождения. Особенно важным инструментом исследований становится статистика, которая позволяет отразить фактическое состояние рынка, его динамику, структуру, оценить колебания спроса, выявить влияние различных факторов, смоделировать тенденцию дальнейшего развития. Для любой компании, приступающей к производству или продаже нового товара, важно знать ответы на следующие вопросы: в чем нуждается потребитель; какое количество товара он сможет приобрести; сколько готов заплатить за него; почему он должен предпочесть наш товар, а не конкурента? Информационно-аналитический материал, который отвечает на эти вопросы, призван обеспечить бизнесу конкурентные преимущества. Широкие возможности для анализа и преобразования информационных данных, наглядного представления результатов исследования дает программа SPSS (Superior Performing Software Systems) – система статистической обработки информации.

В учебном пособии рассмотрены основные методы обработки маркетинговой информации с помощью программы SPSS; **изло-**

жены вопросы проведения углубленного статистического анализа данных, собранных в ходе маркетингового исследования – факторный, кластерный, дискриминантный анализ; приведены примеры, иллюстрирующие возможности использования каждого метода обработки маркетинговой информации. В содержании каждой темы дана краткая теоретическая информация и приведены пошаговые инструкции по реализации методов статистического анализа в SPSS. Особое внимание уделяется получаемым результатам и их интерпретации.

Студентам предлагаются практические задания для получения и отработки навыков анализа маркетинговой информации в SPSS.

Теоретический материал и практические задания рекомендуются для работы преподавателя со студентами в учебной аудитории, а также для самостоятельной работы студентов.

В учебном пособии «Введение», главы 1, 2, 3, 4, 6, 7, 8, 10, практические задания к главам 5, 9 написаны старшим преподавателем кафедры маркетинга и международного менеджмента *Изаковой Натальей Борисовной*; главы 5, 9, практическое задание к главам 8, 10 – ассистентом кафедры маркетинга и международного менеджмента *Журавлевой Александрой Юрьевной*.

1. Формирование исходной базы данных в SPSS

1.1. Подготовка данных к анализу

Анализ маркетинговой информации с применением программы SPSS включает выполнение следующих необходимых шагов (рис. 1.1).

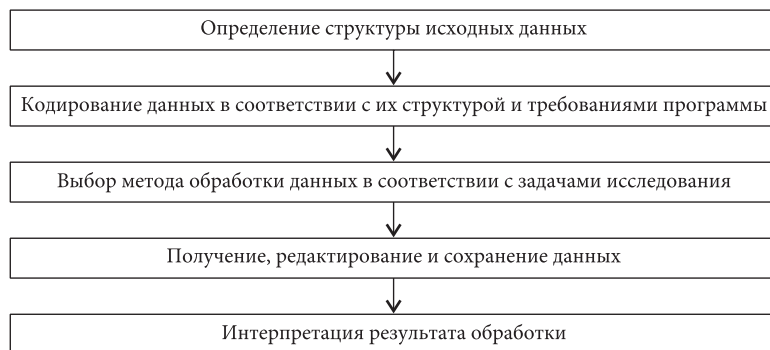


Рис. 1.1. Анализ данных с применением программы SPSS

1. На первом этапе необходимо проверить правильность составления анкет, полноту заполнения и качество интервьюирования, а также репрезентативность выборки.

2. На этапе кодирования необходимо присвоить код каждому возможному варианту ответа по каждому вопросу. С этой целью используют следующие типы шкал измерения переменных:

- номинальная;
- порядковая (ранговая);
- интервальная;
- относительная.

От выбора шкалы (табл. 1.1) зависят вся последующая методика изучения данных и определение возможности расчета статистических показателей.

Таблица 1.1

Типы шкал

Тип шкалы	Особенности построения
1. Номинальная	Каждому свойству потребителя ставится в соответствие некоторый код, позволяющий отличить одно значение от другого. По данной шкале измеряются, как правило, качественные характеристики объекта исследования. Например, значению свойства пол «мужской» присваивается код «1», «женский» – код «2»
2. Порядковая	Присваивает значения свойствам потребителя, находящимся на разных уровнях по отношению друг к другу. Ответы респондентов можно упорядочить по уровню изучаемого свойства переменной. Например, по степени предпочтений покупателей различных марок товара, «наиболее предпочитаемой» присваиваем код «1», коды 2, 3, 4 присваиваются маркам по степени убывания предпочтений
3. Интервальная	Позволяет дать количественную оценку различиям между переменными, т. е. определить, насколько одно значение схоже или отличается от другого (например, шкала Цельсия, календарь). Данная шкала также используется при кодировании ответов респондентов, полученных в результате применения в процессе анкетирования семантического дифференциала (например, когда от респондентов требуется оценить качество продукта по семибалльной шкале)
4. Относительная	Предполагает существование естественного нуля. Поэтому в данной шкале можно сравнивать значения переменной по отношению друг к другу. Это могут быть физические (вес, длина, объем и пр.) и экономические (цена, объем продаж, прибыль и пр.) характеристики

В программе SPSS интервальная и относительная шкалы объединяются в метрическую шкалу.

3. Выбор метода обработки данных основывается на итогах предыдущих этапов маркетинговых исследований, характеристиках информации, а также задачах, поставленных перед маркетинговым исследованием. Могут быть использованы следующие виды анализа: описательные статистики, корреляционный анализ, построение таблиц сопряженности, кросс-табуляция, проверка статистических

гипотез о виде распределения, дисперсионный анализ, дискриминантный анализ, кластерный анализ, многомерное шкалирование, факторный анализ, анализ соответствий, регрессионный анализ, совместный анализ.

4. В результате проведенного анализа исследователь получает массив данных, доступный осмыслению и содержательной интерпретации. Необходимо представить, отредактировать и сохранить полученные данные, так как исчерпывающий анализ обычно требует многократной обработки данных с применением разных методов.

5. Интерпретация результата обработки данных – самостоятельная задача исследователя. Опираясь на полученные статистические данные, важно выявить причинно-следственные отношения между изучаемыми признаками, факторы, оказывающие наибольшее влияние на исследуемую проблему, дать грамотную обоснованную оценку ситуации и выстроить прогноз.

1.2. Структура редактора данных

Файл исходной базы данных для проведения анализа в SPSS формируется в редакторе данных (Data Editor). Редактор данных имеет две вкладки: Переменные (Variable View) и Данные (Data View). Вкладки представляют собой таблицы, содержащие информацию о данных, собранных для проведения анализа.

Во вкладке Переменные представлена таблица с данными, которые описывают свойства переменных. Каждая строка отображает переменную (вопрос анкеты), каждый столбец – ее свойства.

В столбце Имя (Name) записывают имя переменной – это может быть номер или часть вопроса в анкете, например: переменные «пол», «занятость», «марка». Имя переменной не является произвольным, оно может содержать буквы латинского алфавита и цифры, а также некоторые символы: \$, #. Длина имени – не более 64 знаков. Не допускаются пробелы и буквы других алфавитов. Имя переменной должно начинаться с буквы и не может заканчиваться знаком подчеркивания «_» и точкой.

В столбце Тип (Type) задается тип переменной; текущим типом является Числовой (Numeric). В подавляющем большинстве случаев

лучше иметь дело с числовыми переменными. Если требуется изменить тип переменной, нужно нажать на кнопку Тип переменной (Variable Type).

В столбце Ширина (Width) задается максимальное количество знаков, которые может иметь переменная, включая дробную часть.

В столбце Десятичные (Decimal) выбирается количество десятичных знаков после запятой, если тип переменной допускает использование дробных чисел.

В столбце Метка (Label) можно задать метку переменной. Метка используется для того, чтобы более подробно отразить смысл переменной. Это своего рода комментарий к имени переменной. При задании меток переменных часто используются формулировки вопросов, содержащихся в анкете.

В столбце Значения (Values) отображаются значения меток переменных. В поле Значения указываются коды возможных вариантов ответа на этот вопрос. Для заполнения данного столбца необходимо произвести кодировку вариантов ответа. В диалоговом окне Значение меток переменных в поле Значение указываются числовые коды вариантов ответа, а в поле Метка – их формулировки.

В столбце Пропущенные значения (Missing) следует указать, какие коды вариантов ответов следует исключить из анализа. Например, отсутствие определенного ответа: «98» – не знаю, «99» – нет ответа.

В столбце Столбцы (Columns) таблицы «Переменные» указывается ширина столбца, содержащего значения соответствующей переменной в таблице другой вкладки редактора данных: Данные (Data View). По умолчанию ширина столбца задается 8.

В столбце Выравнивание (Alignment) задается положение кодов ответов в таблице «Значения переменных» во вкладке редактора данных Данные. Они могут быть выровнены по правому краю (Right), по левому краю (Left) или по центру (Center). По умолчанию задается выравнивание по правому краю.

В столбце Шкала измерения (Measure) указывается тип шкалы, по которой измеряется переменная. По умолчанию задается метрическая шкала (Scale). В случае необходимости тип шкалы можно изменить.

Основное правило создания файла данных в SPSS: переменные должны быть одновариантными, каждая переменная может иметь только одну метку. Таким образом, если вопрос может иметь несколько вариантов ответа каждого респондента, необходимо создать несколько одновариантных переменных (дихотомическая кодировка данных).

Например, на вопрос «Какую марку одежды вы предпочитаете?» ответ может быть закодирован следующим образом: «1» – предпочитаю, «0» – не предпочитаю. Следовательно, ответы респондентов распределяются так, как показано в табл. 1.2.

Таблица 1.2

Дихотомическая кодировка данных*

Респондент	Марка А	Марка В	Марка С
Респондент 1	1	0	1
Респондент 2	0	1	1
Респондент 3	0	1	1
Респондент 4	1	0	0

Примечание. * Вопрос анкеты «Какую марку одежды вы предпочитаете?».

Задания для практических работ

Задание 1. Укажите, какой тип шкал наиболее подходит для кодирования ответов на следующие вопросы анкеты:

1. Укажите ваш возраст _____.
2. Семейное положение (замужем/не замужем).
3. Выберите марки электронной продукции, которые вы обычно покупаете?
 - Sony;
 - Panasonic;
 - Samsung;
 - Sharp.
4. Укажите, согласны ли вы со следующими утверждениями:

	Согласен	Не согласен
Я всегда стремлюсь делать выгодные покупки	<input type="checkbox"/>	<input type="checkbox"/>
Я люблю проводить время вне дома	<input type="checkbox"/>	<input type="checkbox"/>

	Согласен	Не согласен
Мне нравится пользоваться новыми современными электронными приборами	<input type="checkbox"/>	<input type="checkbox"/>
Я люблю готовить	<input type="checkbox"/>	<input type="checkbox"/>

5. Проранжируйте данные марки электронной продукции с точки зрения их качества.

Марка	Рейтинг (обведите одну из цифр)						
	Очень низкое			Очень высокое			
Sony	1	2	3	4	5	6	7
Panasonic	1	2	3	4	5	6	7
Samsung	1	2	3	4	5	6	7
Sharp	1	2	3	4	5	6	7

6. Что вы можете сказать о ценах в нашем магазине?

- Они выше, чем в других магазинах.
- Они такие же, как и в других магазинах.
- Они ниже, чем в других магазинах.

7. Укажите, в каких интервалах находится ваш семейный доход из расчета на одного человека.

- Ниже 5 000 р.
- От 5 000 до 10 000 р.
- От 10 000 до 15 000 р.
- От 15 000 до 20 000 р.
- Более 20 000 р.

Задание 2. На основе результатов опроса 100 отдыхающих на базе отдыха «Солнечная» сформируйте исходную базу данных в программе SPSS с учетом шкалирования (продумайте какой тип шкал подходит для кодирования ответов на данные вопросы). Исходные данные для формирования базы данных приведены в приложении. Полная база данных результата опроса прилагается к учебному пособию в электронном виде.

Туристы, отдохнувшие на базе отдыха «Солнечная», ответили на ряд вопросов анкеты (табл. 1.3) в рамках маркетингового исследования.

Таблица 1.3

Анкета опроса отдыхающих базы отдыха «Солнечная»

Вопрос	Кодирование вариантов ответов, тип шкалы
1. Каковы ваши основные мотивы выбора места отдыха за городом?	
<input type="checkbox"/> Близость к городу	
<input type="checkbox"/> Приемлемые цены	
<input type="checkbox"/> Близость водоема, леса	
<input type="checkbox"/> Уровень комфорта	
<input type="checkbox"/> Тишина	
<input type="checkbox"/> Вкусная еда	
<input type="checkbox"/> Наличие развлечений	
<input type="checkbox"/> Комфорт отдыха с детьми	
<input type="checkbox"/> Затрудняюсь ответить	
<input type="checkbox"/> Другое	
2. Как часто вы отдыхаете за городом?	
<input type="checkbox"/> 1 раз в неделю или чаще	
<input type="checkbox"/> 1–2 раза в месяц	
<input type="checkbox"/> 1–2 раза в полгода	
<input type="checkbox"/> 1 раз в год	
<input type="checkbox"/> Реже 1 раза в год	
<input type="checkbox"/> Затрудняюсь ответить	
3. Как вы предпочитаете отдыхать за городом?	
<input type="checkbox"/> Самостоятельно, «дикарем»	
<input type="checkbox"/> На базе отдыха	
<input type="checkbox"/> В санатории	
<input type="checkbox"/> Другое	
4. Как вы добираетесь до места отдыха за городом?	
<input type="checkbox"/> На автомобиле	
<input type="checkbox"/> На поезде	
<input type="checkbox"/> На такси	
<input type="checkbox"/> Другое	
5. Хотели бы вы, чтобы ваш досуг за городом был организован?	
6. Какие развлечения вам интересны при отдыхе за городом?	
<input type="checkbox"/> Дискотека	

Продолжение табл. 1.3

Вопрос	Кодирование вариантов ответов, тип шкалы
<input type="checkbox"/> Спортивные мероприятия	
<input type="checkbox"/> Водные развлечения: лодки, скутеры и пр.	
<input type="checkbox"/> Пешие прогулки в лесу	
<input type="checkbox"/> Туристические походы	
<input type="checkbox"/> Катание на лошадях	
<input type="checkbox"/> Баня, сауна, массаж	
<input type="checkbox"/> Бассейн	
<input type="checkbox"/> Культмассовые мероприятия	
<input type="checkbox"/> Другое	
7. На сколько дней вы обычно выезжаете за город?	
8. Какую сумму вы готовы потратить в целом на отдых за городом за одни сутки на 1 чел.?	
9. Сколько при этом обычно составляют расходы на питание за одни сутки на 1 чел.?	
10. Сколько при этом обычно составляют расходы на проживание за городом за одни сутки на 1 чел.?	
11. Сколько вы готовы потратить в целом на организацию вашего досуга за одни сутки?	
12. Из каких источников вы получили информацию о базе отдыха «Солнечная»?	
<input type="checkbox"/> От знакомых, отдохнувших на базе отдыха	
<input type="checkbox"/> Из рекламы в прессе	
<input type="checkbox"/> Из рекламы по телевидению	
<input type="checkbox"/> Из почтовой рассылки	
<input type="checkbox"/> От представителя базы отдыха	
<input type="checkbox"/> От туристической компании	
<input type="checkbox"/> Другое	
13. Оцените степень своей удовлетворенности местом отдыха в целом	
14. Ваш пол:	
М <input type="checkbox"/> Ж <input type="checkbox"/>	
15. Ваш возраст _____	
16. Уровень дохода за месяц	
<input type="checkbox"/> Менее 5 000 р.	
<input type="checkbox"/> 5 000–10 000 р.	

Вопрос	Кодирование вариантов ответов, тип шкалы
<input type="checkbox"/> 10 000–20 000 р.	
<input type="checkbox"/> 20 000–30 000 р.	
<input type="checkbox"/> 30 000–50 000 р.	
<input type="checkbox"/> Более 50 000 р.	
17. Семейное положение	
<input type="checkbox"/> Не замужем/холост	
<input type="checkbox"/> Замужем/женат	
<input type="checkbox"/> Разведен/разведена	
<input type="checkbox"/> Вдова/вдовец	
18. Укажите, пожалуйста, вашу занятость (профессия)	
<input type="checkbox"/> Студент	
<input type="checkbox"/> Школьник	
<input type="checkbox"/> Предприниматель	
<input type="checkbox"/> Государственный служащий	
<input type="checkbox"/> Преподаватель	
<input type="checkbox"/> Рабочий	
<input type="checkbox"/> Пенсионер	

Пошаговая инструкция

Шаг 1. Запустить программу SPSS.

Шаг 2. Перейдите на вкладку Переменные, щелкнув на ее ярлычке мышью.

Шаг 3. Для задания имен переменных файла введите слово «пол» и перейдите на следующую строку; введите слово «возраст» и перейдите на следующую строку, словосочетание «семейное_положение» и перейдите на следующую строку, словосочетание «уровень_дохода» и перейдите на следующую строку, введите слово «занятость» и перейдите на следующую строку.

Аналогичным образом вводятся все имена переменных.

Шаг 4. Перейдите в столбец Тип и откройте диалоговое окно Тип переменной. Установите тип Числовая.

Шаг 5. Перейдите в столбец Ширина и задайте значение ширины, равное 8. При помощи мыши вставьте скопированное значение для всех остальных переменных.

Шаг 6. Перейдите в столбец Десятичные и задайте значение, равное 0. При помощи мыши вставьте скопированное значение для всех остальных переменных.

Шаг 7. Перейдите в столбец Метки переменных и в строке Пол введите словосочетание «пол респондента»; в строке Возраст введите словосочетание «возраст респондента»; и т. д.

Шаг 8

8.1. Перейдите в ячейку столбца Значения, соответствующую переменной «пол», и щелкните мышью по кнопке с многоточием. На экране появляется окно Значение меток переменных. В поле Значение введите число 1, в поле Метка введите слово «мужчины» и щелкните по кнопке Добавить. В поле Значение введите число 2, в поле Метка введите слово «женщины» и щелкните по кнопке Добавить.

Для закрытия диалогового окна нажмите на кнопку ОК.

8.2. Перейдите в ячейку столбца Значения, соответствующую переменной «возраст», и оставьте содержание ячейки без изменений (слово нет).

Для закрытия диалогового окна нажмите на кнопку ОК.

8.3. Перейдите в ячейку столбца Значения, соответствующую переменной «семейное положение», и щелкните мышью по кнопке с многоточием. На экране появляется окно Значение меток переменных. В поле Значение введите число 1, в поле Метка введите слова «замужем/женат» и щелкните по кнопке Добавить. В поле Значение введите число 2, в поле Метка введите слова «незамужем/холост» и щелкните по кнопке Добавить. В поле Значение введите число 3, в поле Метка введите слова «разведен/разведена» и щелкните по кнопке Добавить. В поле Значение введите число 4, в поле Метка введите слова «вдова/вдовец» и щелкните по кнопке Добавить. В поле Значение введите число 98, в поле Метка введите слова «затрудняюсь ответить» и щелкните по кнопке Добавить. В поле Значение введите число 99, в поле Метка введите слова «нет ответа» и щелкните по кнопке Добавить.

Для закрытия диалогового окна нажмите на кнопку ОК.

8.4. Заполните данные столбца Значения для остальных переменных аналогичным образом, согласно выбранным вами типам шкал.

Шаг 9. Перейдите в ячейку столбца Пропуски, соответствующую переменной «семейное положение», и щелкните мышью по кнопке с многоточием. На экране появляется окно Пропущенные значения, в котором необходимо ввести значения 98 и 99.

Для закрытия диалогового окна нажмите на кнопку ОК.

Заполните данные столбца Пропуски для остальных переменных аналогичным образом, согласно выбранным вами типам шкал.

Шаг 10. В столбце Столбцы указывается ширина столбца, содержащего имена переменных. По умолчанию ширина столбцов задается 8.

Шаг 11. Подведите курсор к соответствующей ячейке столбца Выравнивание, и при нажатии клавиши мыши на экране компьютера появится меню, содержащее три варианта выравнивания данных: по левому краю, по правому краю, по центру, из которых следует выбрать желаемый.

Шаг 12. В столбце Шкала измерения указывается тип шкалы, по которой измеряется переменная. По умолчанию задается метрическая шкала. В случае необходимости тип шкалы можно изменить. Для этого следует подвести курсор в соответствующую ячейку столбца Шкала измерения и нажать клавишу мыши, после чего на экране появится меню из трех типов шкалы измерения. В зависимости от вида переменной следует выбрать один из трех типов шкалы измерения: метрическую, порядковую или номинальную. Поскольку переменная «пол» измеряется по номинальной шкале, то в строке этой переменной в столбце Шкала измерения выбирается тип шкалы измерения номинальная.

Шаг 13. Заполните данные столбцов Выравнивание, Шкала измерения для остальных переменных аналогичным образом, согласно выбранным вами типам шкал.

Шаг 14. Сохраните файл под именем «Солнечная, номер группы, Ф.И.О. студента».

Шаг 15. Перейдите на вкладку Данные, щелкнув на ее ярлычке мышью.

Шаг 16. Введите все результаты опроса в ячейки электронной таблицы, используя условия кодирования информации. Результаты опроса представлены в приложении.

Шаг 17. Сохраните файл. Выйдите из программы.

Задание 3. Разработайте анкету и спроектируйте шкалу измерений (обоснуйте выбор шкалы). Цель составления анкеты – создание образа потребителя.

Задание 4. Разработайте анкету и спроектируйте шкалу измерений (обоснуйте выбор шкалы), сформируйте исходную базу данных в программе SPSS для следующих задач:

а) фирма-изготовитель детских игрушек желает знать, как маленькие дети реагируют на электронную игру «Веселые песенки», в которой ребенок должен петь совместно с известными героями мультипликационных фильмов;

б) фирма-производитель молочных продуктов испытывает три новые вкусовые добавки в молочные коктейли и желает знать мнение потребителей о их сладости, приятности и насыщенности вкуса;

в) фирма-изготовитель бакалейной продукции решает вопрос о выходе на новые географические рынки;

г) футбольный клуб желает узнать, что местные болельщики думают о состоянии стадиона и о сервисе на нем;

д) высшее профессиональное учебное заведение решает провести маркетинговое исследование рынка образовательных услуг с целью выявления своего уровня конкурентоспособности.

2. Управление данными

В процессе работы с информацией могут понадобиться преобразованные данные, являющиеся результатом некоторых действий над исходными данными. Например, для маркетингового исследования могут представлять интерес средняя цена на товар в регионе, общий объем продаж предприятия по всем товарным группам, ранг марки товара и пр.

Преобразование данных предполагает выполнение следующих операций:

1. Вычисление новых переменных на основе существующих: среднее, сумма значений и пр.

2. Ранжирование существующих переменных.

3. Отбор подмножества наблюдений для дальнейшего анализа.

4. Перекодирование переменных в другие переменные для разделения выборки наблюдений на подгруппы с целью дальнейшего анализа.

5. Сортировка наблюдений для реорганизации файла данных для удобства работы с информацией.

6. Объединение данных разных файлов; добавление наблюдений, добавление переменных.

7. Агрегирование данных с целью создания значений переменных, каждое из которых представляет собой результат объединения исходных значений, например среднее.

8. Реструктурирование данных позволяет производить сложные манипуляции со структурой файла данных (например, преобразовывать набор переменных в группы значений одной переменной).

В следующих заданиях показаны примеры выполнения наиболее часто востребованных операций в процессе работы с маркетинговой информацией.

Задания для практических работ

Задание 1. Загрузите базы данных файла «Солнечная, группа, Ф.И.О. студента».

Задание 2. Из интервальной переменной, содержащей средний доход респондентов, создайте категориальную переменную с тремя значениями: «Высокий доход», «Средний доход», «Низкий доход».

Задание 3. Из интервальной переменной, содержащей возраст респондентов, создайте категориальную переменную «Возрастные_группы» со следующими значениями: «Менее 18 лет», «18–25 лет», «26–35 лет», «36–45 лет», «46–55 лет», «56–70 лет», «Более 70 лет». Присвойте этой переменной и ее значениям метки.

Задание 4. Временно разбейте файл с данными на 2 подвыборки:
4.1. С наблюдениями, соответствующими группе респондентов, относящих себя к студентам, и всеми остальными респондентами.
4.2. С наблюдениями для мужчин и для женщин.

Задание 5. Ранжируйте отдыхающих по возрасту.

Задание 6. Проведите сортировку наблюдений по возрасту и полу.

Пошаговые инструкции к выполнению заданий 2–6

К заданию 2

Шаг 1. В меню преобразовать выберите команду Перекодировать – перекодировать в другие переменные.

Шаг 2. Щелкните по переменной Доход, а затем – по кнопке со стрелкой.

Шаг 3. Перейдите в поле Имя и введите Группы_по_доходу и щелкните по кнопке Изменить.

Шаг 4. Щелкните по кнопке Старые и новые значения. Установите переключатель Диапазон, в верхнем поле введите значение 1, в нижнем 2, перейдите в поле Новое значение и введите там число 1, затем щелкните по кнопке Добавить.

Шаг 5. Повторите предыдущее действие для старых значений 3–4 и нового значения 2.

Шаг 6. Установите переключатель Диапазон от указанного значения до наибольшего и в поле под ним введите число 5, а в поле Новое значение – число 3, затем щелкните по кнопке Добавить.

Шаг 7. Щелкните по кнопке Продолжить, чтобы вернуться в исходное диалоговое окно, в котором щелкните по кнопке ОК.

Шаг 8. После кодировки задайте метки «Высокий доход», «Средний доход», «Низкий доход» всем новым переменным в окне Редактора данных.

К заданию 3. По аналогии с заданием 2 выполните самостоятельно задание 3.

К заданию 4.1

Шаг 1. В меню Данные (Data) выберите команду Отобразить наблюдения (Select Cases). В группе Выбрать установите переключатель Если выполнено условие и щелкните по кнопке Если.

Шаг 2. В открывшемся окне сначала щелкните по переменной Занятость (Профессиональная деятельность), а затем – по кнопке со стрелкой, чтобы ввести данную переменную в условия отбора. Далее на панели калькулятора щелкните на кнопке \geq , введите число 2, щелкните по кнопке &. Снова введите переменную Занятость (Профессиональная деятельность) в условие отбора, затем щелкните по кнопке < и введите число 3.

Шаг 3. После создания условия отбора щелкните по кнопке Продолжить, затем – по кнопке ОК, чтобы закрыть диалоговое окно и вернуться в окно редактора данных.

После выполнения этого шага при любой обработке данных будут учитываться только данные для студентов.

Чтобы сделать доступными все данные, достаточно в окне Отобразить наблюдения установить переключатель Все данные.

К заданию 4.2. По аналогии с заданием 4.1 самостоятельно выполните задание 4.2.

К заданию 5. Процедура ранжирования выполняется, когда необходимо перейти от исходных значений переменных к рангам.

Шаг 1. В меню Преобразовать выберите команду Ранжировать наблюдения.

Шаг 2. В диалоговом окне щелкните по переменной Возраст, а затем – по кнопке со стрелкой.

Шаг 3. В группе Ранг 1 присвоить наблюдению установлен переключатель С минимальным значением; при желании можно изменить порядок ранжирования, установив переключатель на С максимальным значением.

Шаг 4. Для ранжирования щелкните по кнопке ОК.

В результате выполнения этого шага в конце списка появится новая переменная с именем R возраст.

К заданию 6. Данная операция позволяет расположить информацию в порядке, удобном для исследования.

Шаг 1. В меню Данные выберите команду Сортировать наблюдения.

Шаг 2. Щелкните сначала по переменной Возраст, затем – по кнопке со стрелкой; далее щелкните по переменной пол и снова по кнопке со стрелкой.

Шаг 3. Порядок сортировки установите По убыванию. Щелкните по кнопке ОК. Проведите сортировку по тем же переменным, но с порядком сортировки По возрастанию.

Шаг 4. Вернитесь к исходному порядку, проведя сортировку по переменной №.

К заданию 7. Используя базу данных опроса отдыхающих на базе отдыха «Солнечная» создать новую переменную «возрастные группы», объединив в нее информацию о принадлежности респондента к конкретному возрастному отрезку. Ранжировать отдыхающих по возрасту, провести сортировку наблюдений по возрасту и по полу.

К заданию 8. Подготовьте книгу кодов для представленной ниже анкеты. Сформируйте исходную базу данных в SPSS для рассматриваемого примера.

Рассчитайте следующие показатели:

- средний возраст респондентов;
- процентное соотношение мужчин и женщин;
- какие ролики наиболее понравились респондентам.

Для представления полученных результатов постройте диаграммы.

Сотрудниками компании X были разработаны 6 вариантов рекламного ролика для товара Y. Руководством компании было принято решение изучить предпочтения целевых потребителей и на их основании принять решение о том, какой ролик пустить в прокат. Ниже представлены бланк анкеты и результаты проведенного исследования (табл. 2.1).

Бланк анкеты

Номер анкеты: _____ (заполняется опрашивающим)

Пол

- Женский
- Мужской
- Нет данных

Возраст: _____

- Нет данных

Какой из представленных роликов вам наиболее запомнился?

- Ролик А
- Ролик Б
- Ролик В
- Ролик Г
- Ролик Д
- Ролик Е
- Нет данных

Таблица 2.1

Результаты опроса

№ п/п	Город/регион	Пол	Возраст	Ролик
1	Г	Женский	45	А
2	Г	Мужской	22	В
3	Г	Мужской	19	В
4	Г	Женский	42	А
5	Г	Мужской	34	Г
6	Г	Женский	72	Б
7	Г	Мужской	38	В
8	Г	Женский	56	В
9	Г	Мужской	61	А
10	Г	Женский	77	А
11	Г	Женский	23	Г
12	Г	Мужской	67	Е
13	Г	Мужской	79	Г
14	Г	Женский	26	В
15	Г	Мужской	59	А
16	Р	Женский	34	Г
17	Р	Мужской	18	Е
18	Р	Женский	44	А
19	Р	Мужской	68	А
20	Р	Женский	22	Д
21	Р	Мужской	Нет данных	А
22	Р	Женский	67	Нет данных
23	Р	Мужской	33	В
24	Р	Женский	44	В
25	Р	Мужской	22	Б
26	Р	Мужской	19	А
27	Р	Женский	55	В
28	Р	Женский	39	Д
29	Р	Женский	34	А
30	Р	Мужской	68	В

3. Предварительный анализ данных

3.1. Описательные статистики

После создания удобной для работы базы данных маркетолог приступает к оценке и анализу информации. Наиболее простым методом анализа является метод описательных статистик.

В рамках данного метода оцениваются:

- 1) показатели центра распределения: средняя, мода, медиана;
- 2) показатели вариации: дисперсия, стандартное квадратичное отклонение, размах вариации, максимумы и минимумы значений, стандартная ошибка выборки;
- 3) показатели формы распределения: асимметрия, эксцесс;
- 4) другие показатели: квартиль, процентиль и пр.

Используя метод описательных статистик, можно определить, например, количество респондентов, предпочитающих конкретную марку (частотный анализ); выбрать самую многочисленную группу респондентов, предпочитающих марку А (мода); разбить все данные на группы, отвечающие определенным требованиям (квартили, процентили); и др.

На следующем примере показано использование описательных статистик для предварительного анализа данных.

Пример. На основе собранной информации о ценах на товар А в 100 торговых точках города (табл. 3.1) произведите первичный анализ: рассчитайте меры центральной тенденции, вариацию, квартили цен; постройте гистограмму и обоснуйте гипотезу о виде распределения данной величины.

Таблица 3.1

Информации о ценах на товар А торговых точек города

№ п/п	Цена, р.	№ п/п	Цена, р.	№ п/п	Цена, р.	№ п/п	Цена, р.	№ п/п	Цена, р.
1	28,18	21	26,55	41	28,19	61	36,9	81	35,87
2	29,39	22	29,05	42	29,59	62	37,15	82	30,76

Окончание табл. 3.1

№ п/п	Цена, р.	№ п/п	Цена, р.	№ п/п	Цена, р.	№ п/п	Цена, р.	№ п/п	Цена, р.
3	31,60	23	34,50	43	31,60	63	30,05	83	32,09
4	27,12	24	28,22	44	27,12	64	24,97	84	25,88
5	32,01	25	37,74	45	32,12	65	24,49	85	27,15
6	35,29	26	29,41	46	35,29	66	23,17	86	29,18
7	27,11	27	31,84	47	27,66	67	34,88	87	30,11
8	24,04	28	30,61	48	24,04	68	30,55	88	33,25
9	24,34	29	27,18	49	24,87	69	32,12	89	23,09
10	31,59	30	28,93	50	31,59	70	28,22	90	27,99
11	32,50	31	25,09	51	25,97	71	29,16	91	28,05
12	31,60	32	34,03	52	28,45	72	34,05	92	32,77
13	27,12	33	33,66	53	32,86	73	37,12	93	31,05
14	32,01	34	26,42	54	34,19	74	28,18	94	22,99
15	35,29	35	23,95	55	27,23	75	24,09	95	23,19
16	27,11	36	33,00	56	29,34	76	23,46	96	29,06
17	24,04	37	29,51	57	36,02	77	27,34	97	30,25
18	24,34	38	24,61	58	35,27	78	28,13	98	25,55
19	31,59	39	27,99	59	29,16	79	23,99	99	29,45
20	32,50	40	37,09	60	28,34	80	33,18	100	31,58

Пошаговые инструкции к выполнению заданий

Шаг 1. Создайте исходную базу данных «Цены на товар А» в программе SPSS.

Шаг 2. В меню Анализ выберите команду Описательные статистики и вкладку Частоты.

Шаг 3. В диалоговом окне Частоты отметьте Вывести частотные таблицы.

Шаг 4. Во вкладке Статистики включите Вывод средней, моды, медианы, эксцесса, коэффициента асимметрии, стандартного отклонения, ошибки средней, минимального и максимального значений.

Шаг 5. Во вкладке Диаграммы выберите тип графика Гистограммы и отметьте Показать на гистограмме нормальную кривую.

Шаг 6. Во вкладке Формат выберите тип вывода данных По возрастанию значений.

Шаг 7. После создания условия отбора щелкните по кнопке Продолжить, затем по кнопке ОК, чтобы открыть Окно вывода данных.

Шаг 8. Проанализируем информацию, представленную в Окне вывода, сделаем выводы относительно информации о ценах.

Интерпретация результатов

В окне вывода представлена таблица «Статистики» с результатами анализа данных.

Таблица 3.2

Цена на товар А в магазинах города

N	Валидные	100
	Пропущенные	0
Среднее		29,6143
Медиана		29,4000
Мода		24,04*
Стд. отклонение		3,83532
Дисперсия		14,710
Асимметрия		,179
Стд. ошибка асимметрии		,241
Экссесс		-,785
Стд. ошибка эксцесса		,478
Размах		14,75
Минимум		22,99
Максимум		37,74
Процентили	25	27,1125
	50	29,4000
	75	32,1200

* Имеется несколько мод. Показана наименьшая.

Выводы

1. На основе таблицы можно сделать вывод, что средняя цена на товар А составляет $29,61 \pm 1,96 \times 3,835$ р. (для 95%-ного уровня значимости), значение медианы близко к значению средней, эксцесс и асимметрия колеблются около 0, что позволяет сделать вывод о том, что данное распределение близко к нормальному.

Квартили позволяют судить о том, какая доля наблюдений значений цены лежит ниже указанных границ: 27,11 р. – 25%; 29,4 р. – 50%; 32,12 р. – 75%.

Согласно таблице частот почти каждая цена встречается один раз, чаще встречаются цены: 24,04 р. и 27,12 р. – 3 раза.

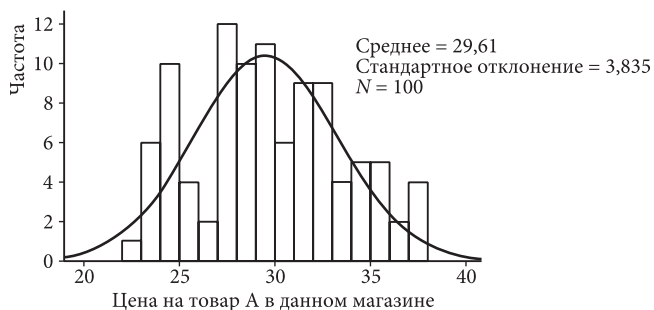


Рис. 3.1. Распределение цен на товар А в магазинах города

2. Представленная гистограмма доказывает нормальность распределения, что позволяет в дальнейшем корректно использовать различные виды статистического анализа.

Шаг 9. Отредактируйте содержимое окна вывода. Для дальнейшего использования окончательного результата все содержимое окна вывода или его фрагменты сохраните в формате Word.

3.2. Графики

Большинство методов анализа программы SPSS имеет возможность построения графиков. Если необходимо представить маркетинговую информацию графически, без дальнейшего статистического анализа можно воспользоваться командой построения диаграмм.

Меню Графика окна редактора данных включает в себя три подменю: Конструктор диаграмм, Панель выбора диаграмм и Устаревшие диалоговые окна. Последнее содержит список стандартных графиков. Подменю Конструктор диаграмм и Панель выбора диаграмм позволяют строить диаграммы в интерактивном режиме: добавлять переменные, изменять категории данных и пр.

Пример. Используя базу данных опроса отдыхающих на базе отдыха «Солнечная», постройте диаграмму уровня удовлетворенности туристов местом отдыха разных возрастных групп.

Пошаговая инструкция

Шаг 1. В меню графика выберите программу Устаревшие диалоговые окна.

Шаг 2. Выберите Столбики, затем – один из типов диаграммы, например: Простые. Поставьте переключатель в поле Итоги по группам наблюдений и нажмите кнопку Задать.

Шаг 3. В открывшемся окне Столбики выберите переключатель № наблюдения, в поле Категориальная ось выберите Довольны ли вы местом отдыха (Удовлетворенность отдыхом), в поле Строки – Возрастные группы, поле Столбцы – оставьте пустым.

Шаг 4. Щелкните по кнопке ОК, чтобы открыть Окно вывода.

Интерпретация результатов

В результате выполненных шагов в окне вывода появляется столбиковая диаграмма (рис. 3.2).

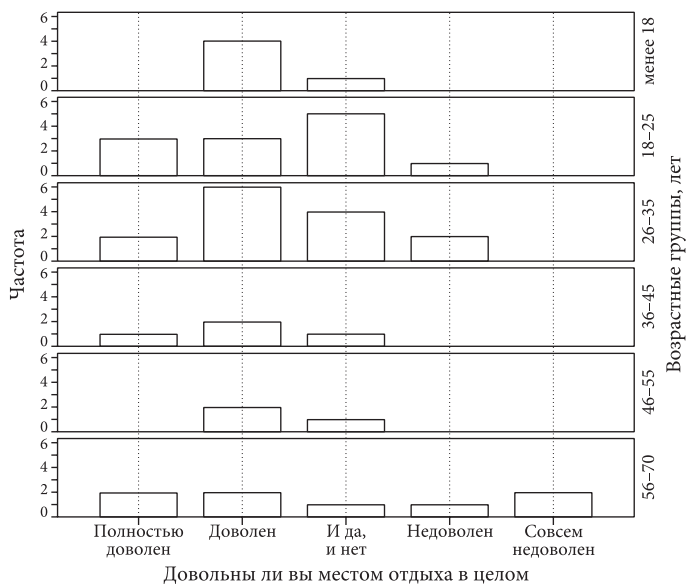


Рис. 3.2. Уровень удовлетворенности туристов разных возрастных групп

3.3. Корреляции

В процессе изучения и анализа маркетинговой информации полезно выявить взаимосвязи между различными переменными, например: зависимость объема продаж от цены, от затрат на определенный вид рекламы и отдачу от нее. С этой целью применяется метод корреляций. Корреляция, или коэффициент корреляции, – это статистический показатель вероятностной связи между двумя переменными, измеренными в количественной форме [1].

Для выявления связи между количественными переменными будем использовать следующие коэффициенты:

- линейный коэффициент корреляции Пирсона (Pearson);
- коэффициенты корреляции Спирмена и Кендалла (Spearman, Kendall);
- частный коэффициент корреляции – коэффициент корреляции между переменными, когда все остальные переменные зафиксированы на постоянном уровне.

Частную корреляцию в маркетинговых исследованиях можно использовать, например, для оценки взаимосвязи объема продаж и места расположения торговой точки, если исключить эффект влияния торговой марки на покупателя.

Если коэффициенты корреляции близки 0, можно утверждать, что связи между переменными отсутствуют. Значения коэффициентов, близких 1, показывают, что существует сильная линейная связь между переменными.

Для оценки достоверности результатов анализа необходимо учитывать уровень значимости рассчитанных коэффициентов.

Пример. Проанализировать зависимость между объемами продаж товара и расходами на рекламу (табл. 3.3) как с учетом влияния цены, так и без учета влияния цены.

Таблица 3.3

Исходная информация

№ п/п	Расходы на рекламу, тыс. р.	Объем продаж, тыс. р.	Цена, р.
1	15,1	49,2	9,8
2	25,8	60,6	13,0
3	18,9	58,2	10,1
4	24,3	63,8	13,0

Окончание табл. 3.3

№ п/п	Расходы на рекламу, тыс. р.	Объем продаж, тыс. р.	Цена, р.
5	12,0	53,1	10,6
6	10,0	53,9	9,4
7	26,7	69,6	14,1
8	30,8	55,1	11,3
9	19,7	55,0	10,3
10	23,5	57,8	12,5
11	6,4	20,2	4,0
12	22,2	75,3	14,0
13	35,1	62,9	14,0
14	28,4	80,5	16,4
15	21,9	56,4	12,1
16	3,4	29,2	6,2
17	7,9	36,8	7,8
18	33,1	75,7	14,0
19	16,9	39,9	6,6
20	18,1	35,9	7,6

Пошаговая инструкция

Шаг 1. В меню Анализ выберите команду Корреляции – Парные корреляции.

Шаг 2. В диалоговом окне щелкните по переменным Расходы на рекламу и Объем продаж, затем по стрелке, чтобы переместить их в окно Переменные.

Шаг 3. Щелкните по кнопке ОК, чтобы открыть Окно вывода. Проанализируйте информацию, представленную в Окне вывода, сделайте выводы относительно зависимости объемов продаж и затратами на рекламу.

Шаг 4. В меню Анализ выберите команду Корреляции – Частные.

Шаг 5. В диалоговом окне щелкните по переменным Реклама и Объем продаж, затем – по стрелке, чтобы переместить их в окно Переменные.

Шаг 6. Щелкните по переменной Цена, а затем – по стрелке, чтобы переместить ее в окно Контрольные переменные.

Шаг 7. Щелкните по кнопке ОК, чтобы открыть Окно вывода.

Проанализируйте информацию, представленную в Окне вывода; сделайте выводы относительно зависимости объемов продаж и затратами на рекламу с учетом влияния цены.

Интерпретация результатов

В окне вывода представлены таблицы с показателями корреляции (табл. 3.4, 3.5).

При вычислении обычного коэффициента корреляции Пирсона (Analyze-Correlate-Bivariate) получили следующие результаты, представленные в табл. 3.4.

Таблица 3.4

Корреляции, тыс. р.

		Расходы на рекламу	Объем продаж
Расходы на рекламу	Корреляция Пирсона	1	,780*
	Значимость (2-сторон.) N	20	,000 20
Объем продаж	Корреляция Пирсона	,780*	1
	Значимость (2-сторон.) N	,000 20	20

* Корреляция значима на уровне 0,01 (2-сторон.).

На основе данной таблицы можно сделать вывод, что существует значительная корреляция между расходами на рекламу и объемом продаж (коэффициент корреляции 0,78 близок к 1).

Но при учете фактора «цена» (исключаем его из анализа), рассчитывая частный коэффициент корреляции (Analyze-Correlate-Partial), получаем следующие данные (табл. 3.5).

Таблица 3.5

Корреляции. Контрольные переменные, тыс. р.

Контрольные переменные			Расходы на рекламу	Объем продаж
Цена на товар	Расходы на рекламу	Корреляция	1,000	-,089
		Значимость (2-сторон.)		,718
		ст. св.	0	17

Контрольные переменные		Расходы на рекламу	Объем продаж
Объем продаж	Корреляция	,089	1,000
	Значимость (2-сторон.)	,718	.
	ст. св.	17	0

Корреляция равна $-0,089$, что близко к 0, следовательно, реклама не повлияла на объем продаж; продажи изменились из-за изменения цены на товар.

Задания для практических работ

Задание 1. Используя базу данных опроса отдыхающих на базе отдыха «Солнечная», проведите предварительный анализ, включив в него следующие переменные: пол, возраст, доход отдыхающих, общие расходы на отдых, частота и продолжительность отдыха за городом.

1. Провести предварительный анализ данных с помощью описательных статистик. Включить в анализ:

- таблицу распределения частот;
- вывод средней, моды, медианы, эксцесса и коэффициента асимметрии, стандартного отклонения и ошибки средней, максимального и минимального значений;
- гистограмму с нанесенной на ней нормальной кривой;
- обоснуйте гипотезу о виде распределения данной величины;
- сделать выводы.

2. Результат представить в отредактированном формате Word.

Задание 2. На основе информации об объемах продаж товара в 100 торговых точках города (табл. 3.6) произведите первичный анализ данных: рассчитайте меры центральной тенденции, вариацию, квартили показателей объема продаж; постройте гистограмму и обоснуйте гипотезу о виде распределения данной величины.

1. Проведите предварительный анализ данных с помощью описательных статистик. Включите в анализ:

- таблицу распределения частот;
- вывод средней, моды, медианы, эксцесса и коэффициента асимметрии, стандартного отклонения и ошибки средней, максимального и минимального значений;
- гистограмму с нанесенной на ней нормальной кривой;
- обоснуйте гипотезу о виде распределения данной величины;
- сделайте выводы.

Таблица 3.6

Информации о объемах продаж товара в 100 торговых точках города, тыс. р.

№ п/п	Объем продаж	№ п/п	Объем продаж	№ п/п	Объем продаж	№ п/п	Объем продаж	№ п/п	Объем продаж
1	281	21	265	41	281	61	369	81	358
2	293	22	290	42	295	62	371	82	307
3	316	23	345	43	316	63	300	83	320
4	271	24	282	44	271	64	249	84	258
5	320	25	377	45	321	65	244	85	271
6	352	26	294	46	352	66	231	86	291
7	271	27	318	47	276	67	348	87	301
8	240	28	306	48	240	68	305	88	332
9	243	29	271	49	248	69	321	89	230
10	315	30	289	50	315	70	282	90	279
11	325	31	250	51	259	71	291	91	280
12	316	32	340	52	284	72	340	92	327
13	271	33	336	53	328	73	371	93	310
14	320	34	264	54	341	74	281	94	229
15	352	35	239	55	272	75	240	95	231
16	271	36	330	56	293	76	234	96	290
17	240	37	295	57	360	77	273	97	302
18	243	38	246	58	352	78	281	98	255
19	315	39	279	59	291	79	239	99	294
20	325	40	370	60	283	80	331	100	315

2. Результат представьте в отредактированном формате Word.

Задание 3. Проанализируйте зависимость между объемами продаж и ценами как с учетом влияния расходов на рекламу, так

и без учета (табл. 3.7). Постройте графики двумерного рассеяния для переменных объем продаж – цена; объем продаж – затраты на рекламу.

Таблица 3.7

Исходная информация

№ п/п	Расходы на рекламу, тыс. р.	Объем продаж, тыс. р.	Цена, р.
1	15,1	49,2	9,8
2	25,8	60,6	13,0
3	18,9	58,2	10,1
4	24,3	63,8	13,0
5	12,0	53,1	10,6
6	10,0	53,9	9,4
7	26,7	69,6	14,1
8	30,8	55,1	11,3
9	19,7	55,0	10,3
10	23,5	57,8	12,5
11	6,4	20,2	4,0
12	22,2	75,3	14,0
13	35,1	62,9	14,0
14	28,4	80,5	16,4
15	21,9	56,4	12,1
16	3,4	29,2	6,2
17	7,9	36,8	7,8
18	33,1	75,7	14,0
19	16,9	39,9	6,6
20	18,1	35,9	7,6

1. Сделайте окончательные выводы о факторах, повлиявших на изменение объема продаж.

2. Результат представьте в отредактированном формате Word.

Задание 4. Используя базу данных опроса отдыхающих на базе отдыха «Солнечная», постройте диаграммы уровня удовлетворенности отдыхом туристов с разным уровнем доход и в зависимости от их профессиональной деятельности. Сделайте выводы.

Задание 5. Используя базу данных опроса отдыхающих на базе отдыха «Солнечная», постройте диаграммы предпочтений вида отдыха за городом туристов в зависимости от их профессиональной деятельности и семейного положения. Сделайте выводы.

Задание 6. Используя базу данных опроса отдыхающих на базе отдыха «Солнечная», постройте диаграммы предпочтений частоты отдыха за городом туристов в зависимости от их пола и семейного положения. Сделайте выводы.

4. Анализ таблиц сопряженности

Таблицы сопряженности служат для описания связи двух и более номинальных переменных. Примерами номинальных переменных являются пол, местность, марка товара, вид транспорта, ответы «да» или «нет» и пр.

С помощью анализа таблиц сопряженности можно найти ответы, например, на следующие вопросы:

1. Как много женщин среди приверженцев данной марки товара?
2. Связана ли интенсивность потребления данного товара с изменением климатических условий?
3. Связана ли частота потребления товара с уровнем дохода потребителя?

Для установления степени связи между переменными используется критерий независимости χ^2 (Хи-квадрат Пирсона). Чем больше значение χ^2 , тем больше зависимость между переменными. Значения χ^2 , близкие к 0, свидетельствуют о независимости переменных.

Вместе с χ^2 вычисляется p -уровень значимости. При $p > 0,05$ считается, что переменные независимы. При $p < 0,05$ предположение о независимости переменных отклоняется и делается вывод о том, что переменные зависят друг от друга [1].

Для определения силы связи между переменными вычисляется коэффициент Крамера V . Значения этого коэффициента всегда лежат между 0 и 1. Для более точной оценки силы связи между переменными могут определяться коэффициенты Фи, Лямбда и Тау Гудмена и Краскала.

Пример. В результате опроса 50 респондентов (1 – мужчины, 2 – женщины) выявили их предпочтения в потреблении соков А и В (1 – А, 2 – В) (табл. 4.1). Выясните, есть ли зависимость между полом респондента и тем соком, который он предпочитает.

Таблица 4.1

Информация для анализа

№ п/п	Пол	Сок	№ п/п	Пол	Сок	№ п/п	Пол	Сок	№ п/п	Пол	Сок
1	1,0	1,0	14	1,0	2,0	27	1,0	1,0	40	1,0	2,0
2	1,0	2,0	15	2,0	1,0	28	2,0	1,0	41	1,0	1,0
3	2,0	2,0	16	2,0	1,0	29	2,0	1,0	42	1,0	2,0
4	1,0	1,0	17	1,0	2,0	30	1,0	2,0	43	2,0	1,0
5	1,0	1,0	18	2,0	1,0	31	1,0	1,0	44	1,0	1,0
6	2,0	2,0	19	2,0	2,0	32	2,0	2,0	45	1,0	2,0
7	2,0	2,0	20	1,0	2,0	33	2,0	1,0	46	1,0	1,0
8	1,0	1,0	21	1,0	1,0	34	1,0	2,0	47	2,0	2,0
9	2,0	1,0	22	2,0	1,0	35	1,0	2,0	48	2,0	1,0
10	1,0	1,0	23	1,0	1,0	36	1,0	2,0	49	2,0	2,0
11	2,0	1,0	24	2,0	2,0	37	2,0	1,0	50	1,0	2,0
12	1,0	2,0	25	1,0	2,0	38	2,0	1,0			
13	1,0	1,0	26	1,0	2,0	39	2,0	1,0			

Пошаговая инструкция

Шаг 1. Анализ – Описательные статистики – таблицы сопряженности.

Шаг 2. В строке – сок, в столбце – пол.

Шаг 3. Установить флажок Вывести кластеризованные столбиковые диаграммы.

Шаг 4. Статистики – отметить ХИ-квадрат, Корреляции, Фи и Лямбда.

Шаг 5. Ячейки – отметить частоты: наблюдаемые, ожидаемые; Остатки: нестандартизированные, стандартизированные, скорректированные; Проценты: по строке, по столбцу, по таблице.

Шаг 6. Точные критерии – только асимптотически.

Шаг 7. ОК.

Интерпретация результатов

Исходя из анализа табл. 4.2, можно сделать следующие выводы: среди потребителей сока А женщин больше (51,9%) , чем мужчин (48,1%); сок В, наоборот, больше потребляют мужчины (65,2%), чем

женщины (34,8%). Если рассмотреть потребление сока отдельно внутри групп, то выясняется, что женщины больше любят сок А, чем В (63,6% – А против 36,4% – В), и наоборот (53,6% – В; 46,4% – А). В целом среди потребителей сок А на первом месте (56% респондентов), В – на втором (44%).

Таблица 4.2

Таблица сопряженности «предпочитаемый сок × пол респондента»

		Пол респондента		Итого		
		Мужчины	Женщины			
Предпочитаемый сок	А	Частота	13	14	27	
		Ожидаемая частота	15,1	11,9	27,0	
		% по категории переменной «предпочитаемый сок»	48,1	51,9	100,0	
		% по категории переменной «пол респондента»	46,4	63,6	54,0	
		% по таблице (слюю)	26,0	28,0	54,0	
		Остаток	-2,1	2,1		
	В	Частота	15	8	23	
		Ожидаемая частота	12,9	10,1	23,0	
		% по категории переменной «предпочитаемый сок»	65,2	34,8	100,0	
		% по категории переменной «пол респондента»	53,6	36,4	46,0	
		% по таблице (слюю)	30,0	16,0	46,0	
		Остаток	2,1	-2,1		
	Итого		Частота	28	22	50
	Ожидаемая частота		28,0	22,0	50,0	
% по категории переменной предпочитаемый сок		56,0	44,0	100,0		

Для предварительного анализа влияния пола на потребление сока рассмотрим величины скорректированного остатка; в нашем случае он не выходит за границы стандартизованного остатка, следовательно, гипотеза о наличии связи не подтверждается.

Также показатель Хи-квадрат Пирсона (табл. 4.3) имеет малое значение 1,469, а значимость существенно превышает 0,05 (0,226), что также подтверждает отсутствие связи между полом и выбором сока.

Таблица 4.3

Критерии Хи-квадрат

	Значение	ст. св.	Асимпт. значимость (2-стор.)	Точная значимость (2-стор.)	Точная значимость (1-стор.)
Хи-квадрат Пирсона	1,469 ¹	1	,226		
Поправка на непрерывность ²	,858	1	,354		
Отношение правдоподобия	1,480	1	,224		
Точный критерий Фишера				,264	,177
Линейно-линейная связь	1,439	1	,230		
Количество валидных наблюдений ²	50				

¹ В 0 (0%) ячейках ожидаемая частота меньше 5. Минимальная ожидаемая частота равна 10,12.

² Вычисляется только для таблицы 2×2.

Коэффициенты Лямбда и Тау Гудмена и Краскала (табл. 4.4) очень малы, что также говорит об отсутствии связи.

Таблица 4.4

Направленные меры

		Значение	Асимпт. стандартная ошибка ¹	Прибл. T ²	Прибл. значимость	
Номинальная по номинальной	Лямбда	Симметричная	,067	,196	,334	,739
		Зависимая «предпочитаемый сок»	,087	,220	,379	,705
		Зависимая «пол респондента»	,045	,231	,193	,847
	Тау Гудмена и Краскала	Зависимая «предпочитаемый сок»	,029	,048		,230 ³
		Зависимая «пол респондента»	,029	,048		,230 ³

¹ Не подразумевающая истинность нулевой гипотезы.

² Используется асимптотическая стандартная ошибка в предположении истинности нулевой гипотезы.

³ На основании аппроксимации хи-квадрат.

Величины коэффициентов Φ и V Крамера (табл. 4.5) говорят о низкой связи между переменными, а значимость 0,226 также подтверждает гипотезу об отсутствии связи.

Таблица 4.5

Симметричные меры

		Значение	Асимптотическая стандартная ошибка ¹	Прибл. T^2	Прибл. значимость
Номинальная по номинальной	Φ	,171			,226
	V Крамера	,171			,226
Интервальная по интервальной	R Пирсона	,171	,139	-1,205	,234 ³
Порядковая по порядковой	Корреляция Спирмена	,171	,139	-1,205	,234 ³
Количество валидных наблюдений		50			

¹ Не подразумеваемая истинность нулевой гипотезы.

² Используется асимптотическая стандартная ошибка в предположении истинности нулевой гипотезы.

³ На основании нормальной аппроксимации.



Рис. 4.1. Предпочитаемые напитки

Таким образом, на основе проведенного анализа можно сделать вывод, что между полом респондента и соком, который он предпочитает, нет зависимости.

Задания для практических работ

Задание 1. Используя базу данных опроса отдыхающих базы отдыха «Солнечная», выясните, есть ли зависимость между:

- 1) полом респондента и выбором способа отдыха за городом;
- 2) полом респондента и удовлетворенностью отдыхом;
- 3) уровнем дохода и выбором способа отдыха за городом;
- 4) уровнем дохода и удовлетворенностью отдыхом.

Задание 2. Постройте таблицу сопряженности для данных, представленных в табл. 4.6, и выявите зависимость между полом респондента (мужчины – 1, женщины – 2) и маркой одежды, которую он предпочитает (марка А – 1, марка В – 2, марка С – 3).

Таблица 4.6

Информация для анализа

№ п/п	Пол	Марка одежды	№ п/п	Пол	Марка одежды	№ п/п	Пол	Марка одежды	№ п/п	Пол	Марка одежды
1	1	1	14	1	2	27	1	1	40	1	2
2	1	2	15	2	1	28	2	3	41	2	1
3	2	2	16	1	3	29	2	1	42	1	3
4	1	3	17	1	2	30	1	2	43	1	3
5	1	3	18	2	3	31	1	3	44	1	1
6	2	2	19	1	2	32	2	3	45	1	2
7	2	2	20	1	3	33	2	1	46	1	3
8	1	3	21	2	1	34	1	3	47	2	2
9	2	3	22	2	3	35	1	3	48	2	3
10	1	1	23	1	3	36	1	2	49	2	3
11	2	3	24	2	2	37	2	1	50	1	2
12	1	2	25	1	2	38	2	3			
13	1	1	26	1	2	39	1	1			

Задание 3. Постройте таблицу сопряженности для данных, представленных в табл. 4.7, и выявите зависимость между уровнем дохода респондента (низкий – 1, средний – 2, высокий – 3) и маркой одежды, которую он предпочитает (марка А – 1, марка В – 2, марка С – 3).

Таблица 4.7

Информация для анализа

№ п/п	Уровень дохода	Марка одежды	№ п/п	Уровень дохода	Марка одежды	№ п/п	Уровень дохода	Марка одежды	№ п/п	Уровень дохода	Марка одежды
1	3	1	14	3	2	27	1	1	40	3	2
2	1	2	15	2	1	28	3	3	41	2	1
3	2	2	16	1	3	29	2	1	42	3	3
4	3	3	17	1	2	30	3	2	43	1	3
5	1	3	18	2	3	31	1	3	44	1	1
6	2	2	19	3	2	32	2	3	45	3	2
7	3	2	20	1	3	33	2	1	46	1	3
8	1	3	21	3	1	34	1	3	47	2	2
9	2	3	22	2	3	35	1	3	48	3	3
10	3	1	23	1	3	36	1	2	49	2	3
11	2	3	24	2	2	37	3	1	50	1	2
12	1	2	25	1	2	38	2	3			
13	1	1	26	3	2	39	3	1			

Задание 4. Исследуйте зависимость между уровнем дохода респондентов (низкий – 1, средний – 2, высокий – 3) и использованием сети Интернет для покупок (использует редко – 1, использует средне – 2, использует часто – 3).

Таблица 4.8

Информация для анализа

№ п/п	Уровень дохода	Частота покупок в сети Интернет	№ п/п	Уровень дохода	Частота покупок в сети Интернет	№ п/п	Уровень дохода	Частота покупок в сети Интернет	№ п/п	Уровень дохода	Частота покупок в сети Интернет
1	3	1	14	3	2	27	1	1	40	3	2
2	1	2	15	2	1	28	3	3	41	2	1
3	2	2	16	1	3	29	2	1	42	3	3
4	3	3	17	1	2	30	3	2	43	1	3
5	1	3	18	2	3	31	1	3	44	1	1

Окончание табл. 4.8

№ п/п	Уровень дохода	Частота покупок в сети Интернет	№ п/п	Уровень дохода	Частота покупок в сети Интернет	№ п/п	Уровень дохода	Частота покупок в сети Интернет	№ п/п	Уровень дохода	Частота покупок в сети Интернет
6	2	2	19	3	2	32	2	3	45	3	2
7	3	2	20	1	3	33	2	1	46	1	3
8	1	3	21	3	1	34	1	3	47	2	2
9	2	3	22	2	3	35	1	3	48	3	3
10	3	1	23	1	3	36	1	2	49	2	3
11	2	3	24	2	2	37	3	1	50	1	2
12	1	2	25	1	2	38	2	3			
13	1	1	26	3	2	39	3	1			

5. Сравнение средних величин

Методы сравнения средних величин часто используются в маркетинговых исследованиях для выявления взаимосвязи между исследуемыми переменными. К таким методам относятся Т-тесты и дисперсионный анализ [2].

Основные термины

Нулевая гипотеза – предположение о том, что между определенными статистическими параметрами генеральной совокупности не существует связи или различия. Ее подтверждение не требует от компании каких-либо действий [1].

Альтернативная гипотеза – утверждение о том, что между определенными статистическими параметрами генеральной совокупности есть связь или различия. Ее подтверждение означает, что руководству компании следует предпринять какие-либо действия или менять свои взгляды на положение дел [1].

Параметрические методы проверки гипотез предполагают, что изучаемые переменные измерены с помощью интервальной шкалы.

Непараметрические методы проверки гипотез предполагают, что переменные измерены с помощью номинальной или порядковой шкалы.

Размах вариации R – разность между самым большим и самым малым значениями признака у единиц данной совокупности:

$$R = X_{\max} - X_{\min}.$$

Дисперсия – средний квадрат отклонений индивидуальных значений признака от его средней величины:

$$D = \frac{\sum_{i=1}^n (x - \bar{x})^2}{n},$$

где D – дисперсия;

x – анализируемый показатель;

\bar{x} – среднее значение анализируемого показателя;

n – количество значений в анализируемой совокупности данных.

Стандартное отклонение – корень из дисперсии (среднего квадрата отклонений):

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x - \bar{x})^2}{n}}.$$

Коэффициент вариации – наиболее универсальный показатель, отражающий степень разбросанности значений независимо от их масштаба и единиц измерения. Коэффициент вариации измеряется в процентах и может быть использован для сравнения вариации различных процессов и явлений:

$$V = \frac{\sigma}{\bar{x}}.$$

F-критерий или критерий Фишера – статистический критерий для проверки равенства двух дисперсий из двух совокупностей. F-статистика представляет собой отношение двух выборочных дисперсий.

T-тест – одномерный метод проверки гипотез, использующий T-распределение. Применяется для небольших выборок, когда стандартное отклонение не известно.

В ходе проведения T-теста или дисперсионного анализа проверяется исходная (нулевая) гипотеза о равенстве сравниваемых средних величин. Нулевая гипотеза формулируется следующим образом: «Взаимосвязи между исследуемыми величинами не существует». Например, формулируя исходную (нулевую) гипотезу, предполагаем равенство среднего балла успеваемости студентов разного возраста, что свидетельствует о том, что возраст студента не влияет на его успеваемость. По результатам проведения анализа данная гипотеза должна быть подтверждена или опровергнута.

SPSS предоставляет возможность сравнить средние величины с помощью разных методов (рис. 5.1) [2].

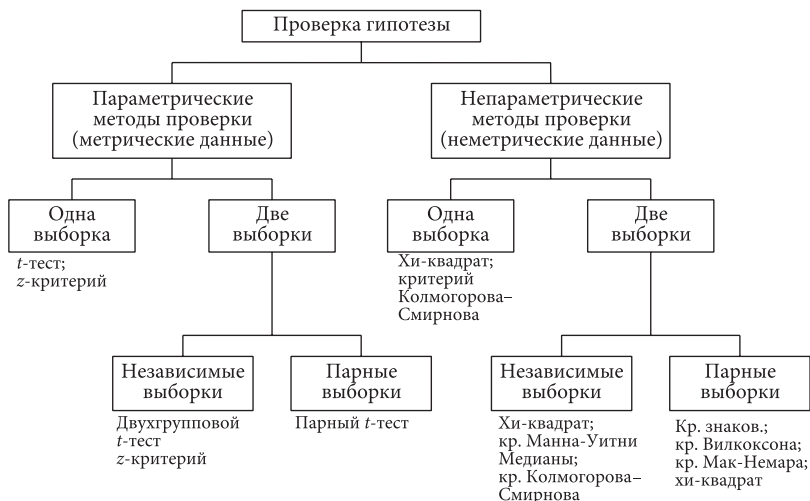


Рис. 5.1. Методы сравнения средних величин

Основным результатом Т-теста, или дисперсионного анализа, является величина «Значимость»; она характеризует вероятность ошибки, с которой может быть отклонена исходная (нулевая) гипотеза. Если «Значимость» меньше 0,05, то исходная гипотеза может быть отклонена с вероятностью ошибки меньше 5%, т. е. ее можно считать неверной.

Результаты проверки исходной (нулевой) гипотезы определяются доверительным интервалом. По умолчанию устанавливается доверительный интервал 95%.

При запуске процедуры «Сравнение средних» (рис. 5.2) определяются средние величины зависимых переменных в разных группах.

Сравнение двух средних величин осуществляется при помощи Т-теста.

Одновыборочный t-критерий используется для сравнения средней величины изучаемого признака в выборке с выбранной заданной стандартной величиной. Например, с помощью этого теста можно определить, отличается ли средняя цена на конкретный товар выбранных торговых предприятий от средней цены, рассчитанной официальным источником или ценой, которую готовы заплатить потребители за данный товар.

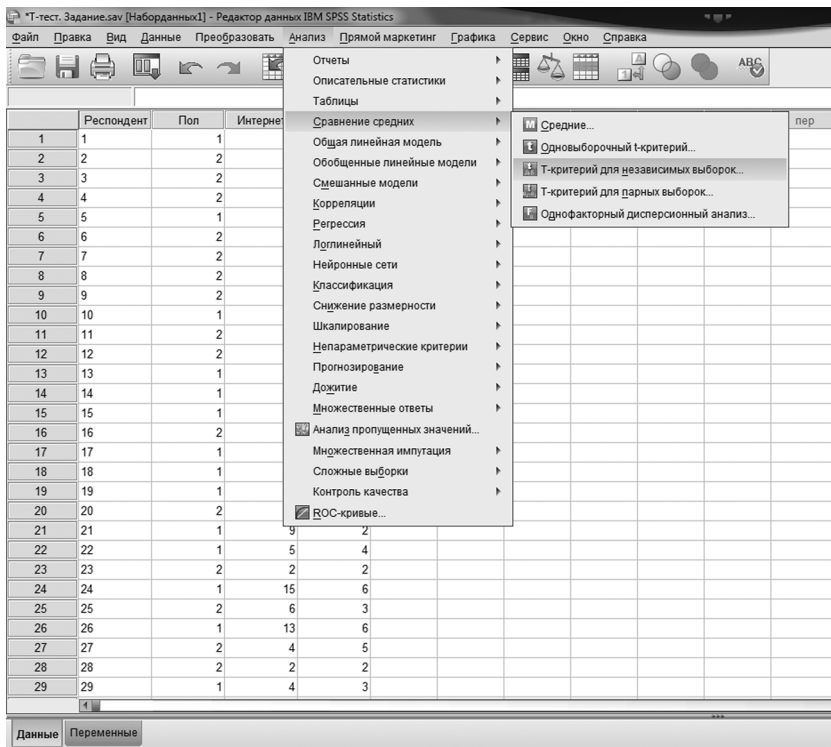


Рис. 5.2. Сравнение средних

T-критерий для независимых выборок (рис. 5.1) применяется для сравнения средних величин изучаемого признака в двух группах. При этом каждый респондент может оказаться только в одной из двух групп, например: мужчины и женщины, семейные и несемейные, покупающие или не покупающие товар «X» и т. д.

T-критерий для парных выборок (рис. 5.1) применяется для сравнения средних величин в двух группах, но при этом один и тот же респондент может одновременно оказаться в разных группах, например: 1) респонденты, покупающие товар А, и 2) респонденты, покупающие товар В.

Для сравнения средних величин в трех и более группах применяется *однофакторный дисперсионный анализ* (рис. 5.1) [2].

Пример. Предположим, что было проведено маркетинговое исследование «Использование респондентами сети Интернет для

личных целей». В табл. 5.1 представлены данные о 30 респондентах, включающие пол (1 – мужчина, 2 – женщина), использование сети Интернет (в часах в неделю).

Таблица 5.1

Результаты маркетингового исследования, ч/нед.

Номер респондента	Пол	Использование Интернета
1	1.0	14.0
2	2.0	2.0
3	2.0	3.0
4	2.0	3.0
5	1.0	13.0
6	2.0	6.0
7	2.0	2.0
8	2.0	6.0
9	2.0	6.0
10	1.0	15.0
11	2.0	3.0
12	2.0	4.0
13	1.0	9.0
14	1.0	8.0
15	1.0	5.0
16	2.0	3.0
17	1.0	9.0
18	1.0	4.0
19	1.0	14.0
20	2.0	6.0
21	1.0	9.0
22	1.0	5.0
23	2.0	2.0
24	1.0	15.0
25	2.0	6.0
26	1.0	13.0
27	2.0	4.0
28	2.0	2.0
29	1.0	4.0
30	1.0	3.0

Проблема. Мы хотим определить, действительно ли интенсивность использования Интернета мужчинами отличается от использования Интернета женщинами. Для этого выполним Т-тест для двух независимых выборок.

Т-тест позволяет проверить верность гипотезы, согласно которой средние величины тестируемого показателя в двух группах равны. В данном примере нулевую гипотезу формулируем следующим образом: «Мужчины и женщины одинаково часто пользуются Интернетом, т. е. пол не влияет на интенсивность пользования Интернетом».

Пошаговая инструкция

Шаг 1. Меню Анализ – Сравнение средних – Т-критерий для независимых выборок – диалоговое окно Т-критерий для независимых выборок.

Выбираем метку тестируемой переменной (в рассматриваемом примере это «Использование сети Интернет») и переносим ее в поле Проверяемые переменные. Далее выбираем метку группирующей переменной (в рассматриваемом примере это «Пол») и переносим ее в поле Группировать.

Шаг 2. Кодировать группирующую переменную, которая делит всех респондентов на две группы: мужчины («1») и женщины («2»).

Задать группы – мужчины «1», женщины – «2».

Метрические переменные (например, «возраст») обозначаются при помощи указания точки разрыва (*Порогового значения*) (например, старше или младше указанного возраста).

Шаг 3. Продолжить – диалоговое окно Т-критерий для независимых выборок.

Параметры – Процент доверительного интервала, который по умолчанию задается в размере 95%.

Шаг 4. Продолжить – диалоговое окно Т-критерий для независимых выборок – ОК – запускается процедура проведения Т-теста.

Интерпретация результатов Т-теста для независимых выборок

Верность этой гипотезы определяется в зависимости от величины «Значимость» (рис. 5.3), которая равна 0,000 (F-критерий имеет

Групповые статистики

Использование Internet	Пол	N	Среднее	Стд. отклонение	Стд. ошибка среднего
Использование Internet	мужчина	15	9,33	4,402	1,137
	женщина	15	3,87	1,685	,435

Критерий для независимых выборок

Использование Internet	Критерий равенства дисперсий Ливина		t-критерий равенства средних			95% доверительный интервал равености средних	
	F	Знч.	Значимость (2-сторонняя)	Разность средних	Стд. ошибка разности	Нижняя граница	Верхняя граница
Предполагается равенство дисперсий	15,507	,000	,000	5,467	1,217	2,974	7,960
Равенство дисперсий не предполагается		4,492	,000	5,467	1,217	2,910	8,024

Рис. 5.3. Интерпретация результатов для независимых выборок

вероятность меньше 0,05). Следовательно, исходная гипотеза может быть отклонена с вероятностью ошибки 0%, что существенно ниже допустимого уровня (5%). Следовательно, исходная гипотеза может быть отклонена.

Нулевая гипотеза (H₀) отклоняется. В рассматриваемом примере необходимо использовать *t*-критерий, который соответствует утверждению «Предполагается, что дисперсии не равны». Значение *t* равно 4,492, и с учетом 18,014 степеней свободы это дает значение вероятности, равное 0,000, которое меньше уровня значимости, равного 0,05. Следовательно, нулевую гипотезу о равенстве средних отклоняют. Поскольку среднее значение для степени использования Интернета для мужчин равно 9,333, а для женщин – 3,867, то мужчины используют Интернет значительно больше по сравнению с женщинами.

Обратный пример. Нулевая гипотеза – взаимосвязи между исследуемыми величинами не существует (т. е. сравниваемые средние величины равны: дисперсии распределения удовлетворенности мужчин и женщин равны) – мужчины и женщины одинаково удовлетворены местом отдыха.

Таблица 5.2

**Результаты Т-теста: статистические показатели в группах.
Групповые статистики**

	Пол	N	Среднее	Стд. отклонение	Стд. ошибка среднего
Довольны ли вы местом отдыха в целом	Мужчины	45	2,22	,850	,177
	Женщины	43	2,60	1,118	,224

Как видно из данных табл. 5.2, всего в исследованиях приняли участие 45 мужчин и 43 женщины. Средняя удовлетворенность местом отдыха у мужчин составила 2,22, а у женщин – 2,60.

Является ли значимым разница между средней удовлетворенностью местом отдыха у мужчин и женщин с точки зрения статистики определяют с помощью Т-теста.

При проведении теста Ливина проверяется следующая гипотеза: «Дисперсии распределения тестируемой величины в разных группах равны». Верность данной гипотезы определяется на основе величины «Significance» («Значимость»). В зависимости от

выполнения условия равенства дисперсий необходимо выбрать одну из строк таблицы. При заданной величине доверительного интервала 95% (вероятность ошибки в случае отклонения исходной гипотезы – 5%) значимость = 0,158 (тест Ливиня), что > 0,05, следовательно, нулевая гипотеза может быть отклонена с вероятностью ошибки 15,8%. Гипотеза не отклоняется, она верна – дисперсии равны.

Таблица 5.3

Критерии для независимых выборок

		Критерий равенства дисперсий Ливиня		t-критерий равенства средних						
		F	Знч.	t	ст. св.	Значимость (2-сторонняя)	Разность средних	Стд. ошибка разности	95% доверительный интервал разности средних	
									Нижняя граница	Верхняя граница
Довольны ли вы местом отдыха в целом	Предполагается равенство дисперсий	2,064	,158	-1,325	46	,192	-,383	,289	-,964	,198
	Равенство дисперсий не предполагается			-1,341	44,486	,187	-,383	,285	-,958	,192

Выбираем строку равенства дисперсий и проверяем верность гипотезы «Средние величины в двух группах равны». Показатель «Значимость (2-сторонняя)» составляет 0,192, что означает, что исходная гипотеза может быть отклонена с вероятностью ошибки 19,2%, а это выше допустимого уровня в 5%. Следовательно, исходная гипотеза не может быть отклонена, т. е. сравниваемые величины с точки зрения статистики равны.

Можно сделать вывод: разница между средним уровнем удовлетворенности местом отдыха у мужчин и женщин (2,22 и 2,62) не является статистически значимой.

Таким образом, в результате проведенного Т-теста доказано, что пол туриста не влияет на его удовлетворенность местом отдыха.

Задания для практических работ

Задание 1. Сформулировать нулевую гипотезу и провести Т-тест, сделать выводы.

Результаты представить в виде файла Output программы SPSS и файла в формате Word.

1.1. Используя базу данных опроса отдыхающих базы отдыха «Солнечная», выявите, как соотносится цена в 1435 р. на одного человека двухместного стандартного номера с трехразовым питанием с суммой, которую отдыхающие готовы потратить на одного человека в сутки на отдых за городом?

1.2. Используя базу данных опроса отдыхающих базы отдыха «Солнечная», выявите, как соотносится стоимость питания 380 р. в столовой базы отдыха в день с суммой, которую отдыхающие готовы тратить на питание за городом в день на одного человека?

Задание 2. Сформулировать нулевую гипотезу и провести Т-тест, сделать выводы. Результаты представить в виде файла Output программы SPSS и файла в формате Word.

2.1. Используя базу данных опроса отдыхающих базы отдыха «Солнечная», выявите, влияет ли пол туриста на его удовлетворенность местом отдыха.

2.2. Используя базу данных опроса отдыхающих базы отдыха «Солнечная», выявите, влияет ли возраст туриста на его удовлетворенность местом отдыха. (Возрастную границу сравнения определите самостоятельно.)

2.3. Используя базу данных опроса отдыхающих базы отдыха «Солнечная», выявите, влияет ли продолжительность отдыха на удовлетворенность туриста местом отдыха.

2.4. Используя базу данных опроса отдыхающих базы отдыха «Солнечная», выявите, влияет ли уровень дохода туриста на его удовлетворенность местом отдыха (объединяем низкий и средний в один).

Задание 3. Продуктовый магазин провел исследование частоты покупок мужчин и женщин в течение месяца.

Определить, действительно ли частота покупок мужчин статистически отличается от частоты покупок женщин.

Таблица 5.4

Результаты маркетингового исследования

Номер респондента	Пол	Частота покупок
1	1.0	7.0
2	2.0	2.0
3	2.0	3.0
4	2.0	3.0
5	1.0	10.0
6	2.0	6.0
7	2.0	2.0
8	2.0	1.0
9	2.0	6.0
10	1.0	5.0
11	2.0	3.0
12	2.0	4.0
13	1.0	9.0
14	1.0	8.0
15	1.0	5.0
16	2.0	13.0
17	1.0	9.0
18	1.0	4.0
19	1.0	10.0
20	2.0	6.0
21	1.0	9.0
22	1.0	5.0
23	2.0	2.0
24	1.0	12.0
25	2.0	6.0
26	1.0	11.0
27	2.0	4.0
28	2.0	12.0
29	1.0	4.0
30	1.0	3.0

6. Однофакторный дисперсионный анализ

Однофакторный дисперсионный анализ проводится для выявления влияния одной переменной на другую. При этом одна из переменных является независимой и должна быть порядковой или номинальной, а другая – зависимой и метрической. Данный вид анализа проверяет верность гипотезы, согласно которой средние величины более чем в двух группах равны [1].

С помощью однофакторного дисперсионного анализа можно найти ответы, например, на следующие вопросы:

1. Действительно ли различаются предпочтения потребителей к торговой марке в зависимости от их уровня дохода?
2. Действительно ли различаются предпочтения потребителей к торговой марке в зависимости от вида рекламного ролика, который они посмотрели?
3. Различаются ли разные сегменты рынка с точки зрения предпочтений мест приобретения товара?
4. Влияет ли уровень образования респондентов на выбор места отдыха?
5. Различаются ли географические сегменты по товарным предпочтениям потребителей?

При однофакторном дисперсионном анализе сравниваются между собой средние значения нескольких групп (выборок), на которые делятся все анализируемые данные. Независимая переменная, при помощи которой все данные разделяются на группы (категории), называется *категориальным фактором* (рис. 6.1).

Последовательность проведения однофакторного дисперсионного анализа:

1. Формулировка вопроса в соответствии с требованиями, предъявляемыми к переменным, выбор зависимой и независимой переменных.
2. Формулирование исходной (нулевой) гипотезы, согласно которой не существует связи между выбранными переменными.

В результате анализа нулевая гипотеза должна быть подтверждена или опровергнута.

3. В ходе анализа проверяются условия равенства дисперсий зависимой переменной в нескольких выбранных группах (категориях).

4. Проведение проверки неравенства средних значений зависимой переменной в сравниваемых группах для выявления взаимосвязи между переменными.

5. Определение особенностей выявленной взаимосвязи; вычленение категорий, ее обуславливающих.

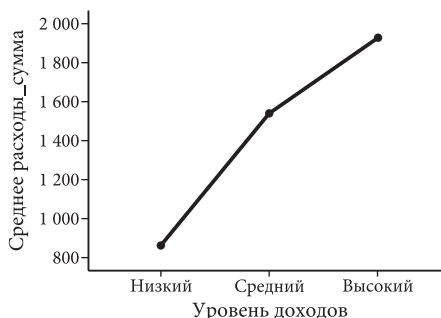


Рис. 6.1. Диалоговое окно однофакторного дисперсионного анализа

Пример. Используя базу данных опроса отдыхающих базы отдыха «Солнечная», проверим, существует ли взаимосвязь между доходом отдыхающих и суммой, которую они готовы потратить на 1 человека за одни сутки отдыха за городом.

Предварительно перед проведением однофакторного дисперсионного анализа проведем преобразование данных о среднем доходе респондентов и создадим категориальную переменную «группы по доходу» с тремя значениями «низкий доход», «средний доход», «высокий доход».

Нулевая гипотеза. Туристы с разным уровнем дохода готовы потратить в среднем одинаковую сумму на 1 человека за одни сутки отдыха за городом. (Не существует связи между доходом отдыхающих и размером суммы, которую они готовы потратить на отдых за городом.)

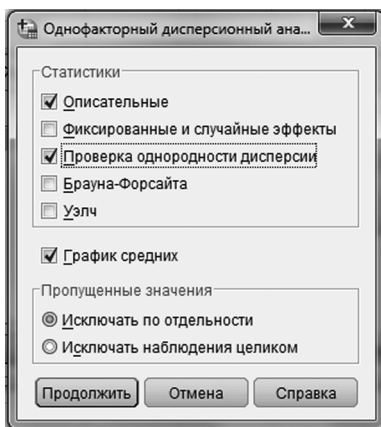
Пошаговая инструкция

Шаг 1. Анализ – Сравнение средних – Однофакторный дисперсионный анализ.

Шаг 2. «Общие расходы на отдых_сумма» в поле Список зависимых переменных.

Шаг 3. «Группы по доходу» в поле Фактор.

Шаг 4. Кнопка Параметры – в открывшемся окне выбрать: Описательные, Проверка однородности дисперсии и График средних (рис. 6.2).



Шаг 5. Кнопка Продолжить.

Шаг 6. Кнопка Апостериорные: выбрать Шеффе и T2 Тамхейна (рис. 6.3).

Шаг 7. Кнопка Продолжить и ОК.

Рис. 6.2. Диалоговое окно «Описательные статистики»

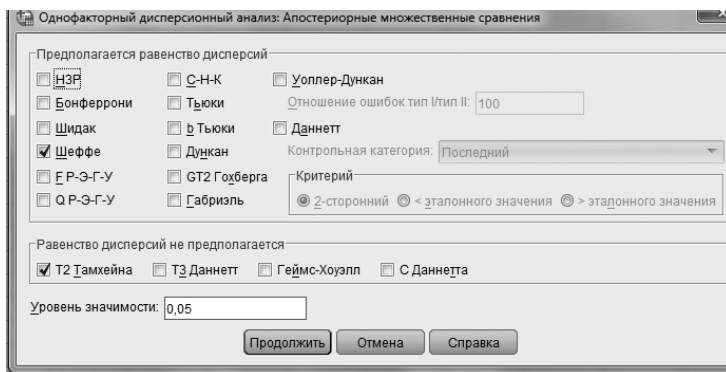


Рис. 6.3. Апостериорные множественные сравнения

Интерпретация результатов

Проверка практической значимости результатов исследования.

На экран компьютера выводится таблица «Описательные статистики», которая содержит статистические показатели, описывающие распределение зависимой переменной в разных группах. В данном примере табл. 6.1 содержит зависимую переменную «Расходы на отдых» в группах отдыхающих с разным уровнем дохода.

Таблица 6.1

Описательные статистики: расходы_сумма

	N	Среднее	Стд. отклонение	Стд. ошибка	95%-ный доверительный интервал для среднего		Минимум	Максимум
					Нижняя граница	Верхняя граница		
Низкий	13	861,54	221,880	61,538	727,46	995,62	500	1 000
Средний	25	1 540,00	379,693	75,939	1 383,27	1 696,73	1 000	2 000
Высокий	7	1 928,57	345,033	130,410	1 609,47	2 247,67	1 500	2 500
Итого	45	1 404,44	499,525	74,465	1 254,37	1 554,52	500	2 500

На данном этапе проверяется практическая значимость сформированных групп. Все группы имеют практическую значимость для исследования, так как количество объектов исследования в каждой группе больше 2. В случае, если сформируется группа с одним ответом респондента, эта группа должна быть исключена из исследования, так как является практически незначимой.

1. Проверка равенства дисперсий по тесту Ливиня

Проверяется гипотеза «Дисперсии в сравниваемых группах равны» (табл. 6.2). Значимость 0,333 означает, что гипотеза может быть отклонена с вероятностью ошибки 33%. Следовательно, гипотеза не отклоняется, и это значит, что дисперсии равны. Если значимость будет меньше 0,05, то гипотеза может быть отклонена, т. е. дисперсии не равны.

Таблица 6.2

Критерий однородности дисперсий: расходы_сумма

Статистика Ливиня	Ст. св. 1	Ст. св. 2	Знч.
1,129	2	42	,333

Критерий однородности дисперсий Ливиня со значимостью 0,333 показал, что дисперсии для каждой группы статистически достоверно не различаются. Следовательно, результаты анализа корректны; в качестве апостериорных сравнений (множественных) будем использовать тест Шеффе.

3. Проверка верности нулевой гипотезы

После проверки равенства дисперсий на экран выводятся результаты однофакторного дисперсионного анализа (табл. 6.3).

Таблица 6.3

Однофакторный дисперсионный анализ: расходы_сумма

	Сумма квадратов	Ст. св.	Средний квадрат	F	Знч.
Между группами	6214056,166	2	3107028,083	27,386	,000
Внутри групп	4765054,945	42	113453,689		
Итого	1,098E7	44			

Проверяем верность исходной нулевой гипотезы: туристы с разным уровнем дохода готовы потратить в среднем одинаковую сумму на 1 человека за одни сутки отдыха за городом. (Не существует связи между доходом отдыхающих и размером суммы, которую они готовы потратить на отдых за городом).

Она может быть отклонена с вероятностью ошибки 0% (значимость 0,000), т. е. гипотеза не верна и должна быть отклонена.

Следовательно, можно сделать вывод, что туристы с разным уровнем дохода готовы потратить разные суммы на 1 человека за одни сутки отдыха за городом (существует зависимость между доходом отдыхающих и размером суммы, которую они готовы потратить на отдых за городом).

3. Для получения более точных результатов необходимо выяснить, в каких именно группах отличия наиболее значительны

Проводим последующие многовариантные (множественные сравнения). В нашем случае дисперсии равны, поэтому анализируем данные теста Шеффе. В случае неравенства дисперсий значимыми являются данные теста T2 Тамхейна.

Пары, характеризующиеся значительной разностью средних, обозначаются звездочкой.

Как видно из табл. 6.4, разность сумм, которую готовы потратить на отдых туристы, значительна для каждой из групп.

Таблица 6.4

Множественные сравнения. Зависимая переменная: расходы_сумма

	(I) уровень дохода	(J) уровень дохода	Разность средних (I-J)	Стд. ошибка	Знч.	95%-ный доверительный интервал	
						Нижняя граница	Верхняя граница
Шеффе	Низкий	Средний	-678,462*	115,175	,000	-970,74	-386,18
		Высокий	-1067,033*	157,908	,000	-1467,75	-666,31
	Средний	Низкий	678,462*	115,175	,000	386,18	970,74
		Высокий	-388,571*	144,034	,035	-754,09	-23,06
	Высокий	Низкий	1067,033*	157,908	,000	666,31	1467,75
		Средний	388,571*	144,034	,035	23,06	754,09

* Разность средних значима на уровне 0,05.

Значимость в каждой группе меньше 0,05, что говорит о достоверности результатов анализа.

Значимость разности сумм, которую туристы с разным уровнем дохода готовы потратить на отдых за городом, можно также увидеть на графике (рис. 6.1).

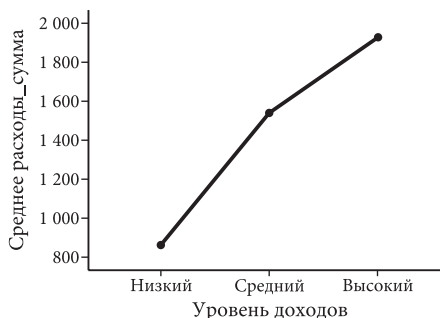


Рис. 6.1. Зависимость средних расходов на отдых и уровнем дохода респондента

В целом можно сделать вывод, что как низкий, так и средний и высокий уровень дохода туриста влияет на размер суммы, которую он готов потратить на 1 человека за один день отдыха за городом. Следовательно, менеджменту базы отдыха «Солнечная»

необходимо продумать грамотную политику ценообразования с учетом сегментирования рынка по доходу отдыхающих.

Задания для практических работ

Задание:

- сформулируйте нулевую гипотезу;
- проведите однофакторный дисперсионный анализ;
- проведите последующие многовариантные сравнения;
- представьте график зависимости;
- сделайте выводы;
- результаты представьте в виде документа в редакторе Word.

Задание 1. Используя базу данных опроса отдыхающих базы отдыха «Солнечная», проверьте, существует ли взаимосвязь между уровнем дохода отдыхающих и продолжительностью разового отдыха.

Задание 2. Используя базу данных опроса отдыхающих базы отдыха «Солнечная», проверьте, существует ли взаимосвязь между возрастом отдыхающих и суммой, которую они готовы потратить на 1 человека за одни сутки отдыха за городом.

Задание 3. Используя базу данных опроса отдыхающих базы отдыха «Солнечная» проверьте, существует ли взаимосвязь между возрастом отдыхающих и продолжительностью разового отдыха за городом.

Задание 4. Используя базу данных опроса отдыхающих базы отдыха «Солнечная», проверьте, существует ли взаимосвязь между полом отдыхающих и продолжительностью разового отдыха за городом.

Задание 5. Используя базу данных опроса отдыхающих базы отдыха «Солнечная», проверьте, существует ли взаимосвязь между полом отдыхающих и суммой, которую они готовы потратить на 1 человека за одни сутки отдыха за городом.

Задание 6. Определите, существует ли взаимосвязь между эффективностью работы торговых представителей и коэффициентом их интеллекта.

Результаты теста IQ закодированы следующим образом:

- 1 – от 80 до 89;
- 2 – от 90 до 99;
- 3 – от 100 до 109;
- 4 – от 110 и более.

Таблица 6.5

Данные о коэффициентах интеллекта торговых представителей

Торговый представитель	Индекс эффективности	Тест IQ
1	122	4
2	105	3
3	103	2
4	95	1
5	97	2
6	106	4
7	100	1
8	115	1
9	78	4
10	101	4
11	115	2
12	120	1
13	88	1
14	110	1
15	96	2
16	93	1
17	92	1
18	103	4
19	121	1
20	95	2
21	99	4
22	102	4
23	98	3
24	100	1
25	99	2

Окончание табл. 6.5

Торговый представитель	Индекс эффективности	Тест IQ
26	99	2
27	113	1
28	114	3
29	110	4
30	98	2
31	92	3
32	106	3
33	103	1
34	111	1
35	102	2
36	102	1
37	88	3
38	105	1
39	94	2
40	108	1
41	84	3

Задача 7. Предположим, что исследователь хочет выяснить, зависит ли отношение респондента к месту жительства от длительности его проживания в этом городе и влияния погодных условий. Отношение выражают следующим образом: 1 – не нравится; 2 – есть плюсы и минусы; 3 – очень нравится; влияние погодных условий выражают по 11-балльной шкале (1 – не нравится; 11 – очень нравится). Продолжительность проживания измеряют количеством лет, которые респондент прожил в этом городе [4].

Таблица 6.6

Данные выборки

№ п/п	Отношение к городу	Длительность проживания, лет	Влияние погодных условий
1	2	10	3
2	3	12	11
3	2	12	4
4	1	4	1
5	3	12	11

Окончание табл. 6.6

№ п/п	Отношение к городу	Длительность проживания, лет	Влияние погодных условий
6	1	6	1
7	2	8	7
8	1	2	4
9	3	18	8
10	3	9	10
11	3	17	8
12	1	2	5

7. Линейный регрессионный анализ

Регрессионный анализ применяется для выявления влияния одной или нескольких независимых переменных на одну зависимую переменную.

Регрессионный анализ позволяет дать математическое описание зависимости между переменными, построить модель (тренд), отражающую данную зависимость и позволяющую строить прогнозы изменения зависимой переменной.

В практике маркетинговых исследований с помощью регрессионного анализа можно найти ответы, например, на следующие вопросы:

1. Объясняются ли различия в объемах продаж товара в разные периоды различием расходом на продвижение? И какова форма этой зависимости?

2. Какую долю в изменении объема продаж можно объяснить расходами на продвижение, различием цен и, наконец, расходами на персонал?

3. Чему равен вклад расходов на персонал в изменение объема продаж при фиксированных ценах и расходах на продвижение?

4. Какие объемы продаж можно ожидать при заданных ценах, расходах на продвижение и расходах на персонал?

В регрессионном анализе могут участвовать только метрические (количественные) переменные. Если необходимо использовать номинальные переменные, их необходимо разложить на дихотомические (см. разд. 1.2 «Структура редактора данных»).

Для описания и выявления линейной зависимости между объектом исследования (зависимой переменной) и одним фактором, влияющим на него (независимой переменной), используется *уравнение простой линейной регрессии*.

Регрессионное уравнение в этом случае имеет вид:

$$y = a - bx,$$

где y – зависимая переменная;

x – независимая переменная;

a – свободный член (константа);

b – коэффициент регрессии [2].

Для выявления и описания линейной зависимости между объектом исследования (зависимой переменной) и несколькими факторами, на него влияющими (независимыми переменными), используется *множественная линейная регрессия*.

Регрессионное уравнение в этом случае имеет вид:

$$y = a - b_1x_1 - b_2x_2 - \dots - b_nx_n.$$

В результате регрессионного анализа необходимо определить свободный член a и коэффициенты регрессии b .

Кроме, того обязательным условием является расчет показателей, характеризующих практическую применимость и статистическую значимость результатов анализа и построенной регрессионной модели.

К данным показателям относятся:

- *Коэффициент детерминации (R)* – характеризует силу связи между значениями зависимой и независимой переменных. Значения коэффициента находятся в интервале от нуля до единицы. Чем ближе значение коэффициента к единице, тем сильнее линейная взаимосвязь. В общем случае он должен превышать 0,5.

- *Коэффициент R -квадрат (R Square)* – показывает, какая часть вариации признака в зависимой переменной объясняется независимой переменной. Значения коэффициента лежат в интервале от нуля до единицы. Данный показатель должен превышать 0,5. Если он равен 0,5, это говорит о том, что регрессионная модель справедлива только для 50% данных.

- *Показатель значимости p* – характеризует значимость регрессионной модели, построенной на основе данных опроса респондентов, попавших в выборку, для всей генеральной совокупности. Если $p < 0,05$, то можно говорить о справедливости регрессионной модели для всей генеральной совокупности. Если $p > 0,05$, то связь между переменными слабая или не обнаружена.

- Важной частью регрессионного анализа является анализ остатков, т. е. отклонений наблюдаемых значений от теоретически

ожидаемых. Остатки должны появляться случайно (не систематически) и подчиняться случайному распределению. Проверка наличия систематических связей между остатками может быть произведена при помощи теста Дарбина–Уотсона (Durbin–Watson) на автокорреляцию. Этот тест позволяет рассчитать коэффициент, значение которого варьирует от 0 до 4. Если значение данного коэффициента близко 2, это означает, что автокорреляция отсутствует.

7.1. Простая линейная регрессия

Как уже отмечалось, простая линейная регрессия служит для описания и выявления линейной зависимости между объектом исследования (зависимой переменной) и одним фактором, влияющим на него (независимой переменной).

Преимуществом данного вида анализа является возможность представить результат графически в виде линейного графика. Построение тренда дает возможность прогнозировать, как будет меняться одна переменная при изменении другой. Построение тренда очень часто используется в практике маркетинга.

Пример. Используя базу данных опроса отдыхающих базы отдыха «Солнечная», выявите, в какой зависимости находятся сумма расходов на отдых в целом и сумма расходов на питание. Сделайте прогноз, как будут изменяться общие расходы на отдых при увеличении (уменьшении) расходов на питание.

Последовательность действий:

1. Провести линейный регрессионный анализ.
2. Построить регрессионную модель (тренд), отражающую зависимость между переменными; составить уравнение регрессии.
3. Сделать вывод о том, как будет изменяться одна переменная при изменении другой.
4. Представить результаты анализа в виде линейного графика.
5. Сделать вывод о вероятности ошибки при использовании данной регрессионной модели.

Пошаговая инструкция

Линейный регрессионный анализ:

Шаг 1. Анализ – Регрессия – Линейная.

Шаг 2. Переменная «Общие расходы на отдых (расхода сумма)» – в окно Зависимая переменная.

Шаг 3. Переменная «Расходы на питание (расходы 2)» – в окно Независимая переменная.

Шаг 4. Кнопка Статистические показатели – Оценки коэффициентов регрессии, Доверительные интервалы, критерий Durbin-Watson.

Шаг 5. Продолжить – ОК.

Построение графика:

Шаг 1. Графики – Диаграмма рассеяния – Простая диаграмма рассеяния.

Шаг 2. Переменная «Общие расходы на отдых (расхода сумма)» – ось Y.

Шаг 3. Переменная «Расходы на питание (расходы 2)» ось X.

Шаг 4. ОК.

Обработка графика:

Шаг 1. Нажать 2 раза правой кнопкой мыши по графику.

Шаг 2. В окне редактор диаграмм выбрать меню Элементы – Приближенная линия.

Шаг 3. В новом окне Свойства во вкладке Приближенная линия отметить линейный вид графика.

Шаг 4. Закрыть.

Интерпретация результатов

Ниже приведены фрагменты данных, сгенерированные программой. На экран компьютера выводятся три таблицы: «Сводка для модели» (табл. 7.1), «Дисперсионный анализ» (табл. 7.2) и «Коэффициенты» (табл. 7.3).

1. Значение коэффициента детерминации $R = 0,695 (> 0,5)$, что свидетельствует о тесной взаимосвязи между общей суммой на отдых и суммой затрат на питание в день.

Таблица 7.1

Сводка для модели

Модель	R	R квадрат	Скорректированный R квадрат	Стд. ошибка оценки	Дурбин-Уотсон
1	,695*	,483	,471	397,433	1,246

* Предикторы: (константа) расходы на питание.

Таблица 7.2

Дисперсионный анализ

Модель	Сумма квадратов	ст. св.	Средний квадрат	F	Знч.
1 Регрессия	6336018,018	1	6336018,018	40,113	,000*
Остаток	6791981,982	43	157953,069		
Итого	1,313E7	44			

* Предикторы: (константа) расходы на питание.

Таблица 7.3

Коэффициенты

Модель	Нестандартизованные коэффициенты		Стандартизованные коэффициенты	t	Знч.
	B	Стд. ошибка	Бета		
1 (Константа)	752,252	117,285		6,414	,000
Расходы на питание	2,671	,422	,695	6,334	,000

Коэффициент R -квадрат составляет 0,483. Это означает, что построенная регрессионная модель описывает только 48,3% случаев, когда увеличение суммы затрат на питание влечет за собой увеличение общей суммы затрат на отдых. Это необходимо учитывать при применении результатов анализа в прогнозировании расходов отдыхающих.

Значение теста Дарбина–Уотсона составляет 1,246. Это существенно ниже 2, что говорит о возможной автокорреляции, т. е. отклонения от теоретически ожидаемых результатов (остатки) могут появляться систематически.

На основании результатов таблицы «Дисперсионный анализ» показатель «Значимость» равен 0,000 ($< 0,5$). Это говорит о том, что регрессионная модель, построенная на основе данных

респондентов, попавших в выборку, справедлива для всей генеральной совокупности.

Результаты регрессионного анализа см. в табл. 7.3.

2. Построение регрессионной модели. Уравнение регрессии имеет вид:

$$Y = 752,252 + 2,671X.$$

3. Построенная регрессионная модель показывает, что если отдыхающий не тратит денег на питание (его кормят друзья или он голодает, или ест дары леса), то его общие расходы на отдых за городом в среднем составят 752,252 р.

С учетом стандартной ошибки, равной 117,285 при доверительном интервале 95%, сумма общих расходов на отдых за городом в день без питания на человека составит в среднем $752,252 \pm 2 \times 117,285$ р.

Значение коэффициента регрессии в построенной модели составляет 2,671. Это значит, что увеличение затрат на питание на 1 р. влечет за собой увеличение общих затрат на отдых за городом на 2,671 р.

4. Используя диаграмму рассеяния (рис. 7.1), представленную ниже, можно прогнозировать, как изменится общая сумма расходов на отдых за городом при изменении суммы затрат на питание.

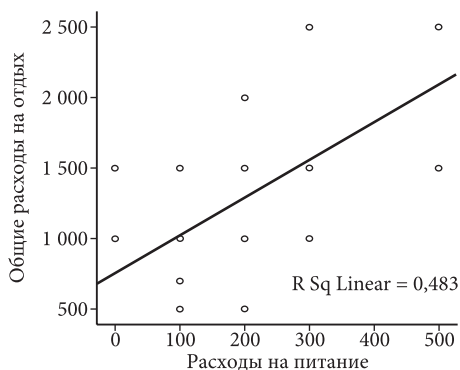


Рис. 7.1. Диаграмма рассеяния

5. Данная модель описывает только 48,3% случаев, поэтому вероятность ошибки при использовании данной регрессионной модели в целях прогнозирования достаточно велика.

7.2. Множественная линейная регрессия

Множественная регрессия является расширением простой регрессии. В отличие от простой она исследует влияние двух и более факторов на зависимую переменную.

Особенностями множественной линейной регрессии являются невозможность графического представления результатов анализа и вероятность эффекта мультиколлинеарности, т. е. существования причинно-следственной связи (корреляции) между независимыми переменными. Другими словами, независимые переменные, включенные в регрессионную модель, могут обозначать, в принципе, одно и то же.

Одним из условий построения множественной регрессионной модели является проверка отсутствия корреляции между независимыми переменными [3].

Условия получения приемлемых результатов анализа:

- Для того чтобы существующие зависимости между переменными были признаны значимыми, необходимо иметь достаточно большие размеры выборки (N не менее 50).

- Данные должны быть корректными и не содержать ошибок.

- Распределение значений независимых переменных должно быть близким к нормальному (значения асимметрий и эксцессов не превосходят 1).

- Запрет на использование зависимых переменных, корреляции между которыми близки к 1 (-1).

Пример. Используя базу данных опроса отдыхающих базы отдыха «Солнечная», выявите, в какой зависимости находятся общие расходы в день на отдых за городом одного отдыхающего и следующие статьи расходов:

- расходы на питание;
- расходы на развлечения.

Последовательность действий:

1. Проведите множественный линейный регрессионный анализ.
2. Проверьте условия получения приемлемых результатов анализа.
3. Постройте регрессионную модель, отражающую зависимость между переменными, составьте уравнение регрессии.
4. Сделайте вывод о том, как будет изменяться одна переменная при изменении других.

5. Сделайте вывод о вероятности возникновения нежелательного эффекта мультиколлинеарности.

6. Сделайте вывод о вероятности ошибки при использовании данной регрессионной модели.

Пошаговая инструкция

Шаг 1. В меню Анализ – Регрессия – Линейная.

Шаг 2. Переменная «Общие расходы на отдых (расхода сумма)» – в окно Зависимая переменная.

Шаг 3. Переменная «Расходы на питание (расходы 2 питание)» – в окно Независимая переменная.

Шаг 4. Переменная «Расходы на развлечения» – в окно Независимая переменная.

Шаг 5. В окне Метод выберите Шаговый отбор.

Шаг 6. Кнопка Статистические показатели – Описательные статистики, Оценки коэффициентов регрессии, критерий Durbin-Watson, диагностика коллинеарности.

Шаг 7. Продолжить – ОК.

Интерпретация результатов

В результате выполнения приведенной выше инструкции программа сгенерировала данные, позволяющие судить о том, какая из независимых переменных оказывает наибольшее влияние на переменную «Общие расходы на отдых» (табл. 7.4).

Таблица 7.4

Описательные статистики

	Среднее	Стд. отклонение	N
Общие расходы на отдых	194,44	107,694	90

1. В анализе участвовали ответы 90 респондентов из 100 возможных ($N > 50$). Коэффициенты корреляции между независимыми переменными меньше 1 (табл. 7.5), что говорит о выполнении условия для получения приемлемых результатов анализа.

Коэффициент детерминации (табл. 7.6) составляет $R = 0,695$ (возможные значения от 0 до 1), что свидетельствует о наличии

плотной взаимосвязи между суммой общих расходов и суммами, расходуемыми на питание и развлечения.

Таблица 7.5

Корреляции

		Общие расходы на отдых	Расходы на питание	Расходы на развлечения
Корреляция Пирсона	Общие расходы на отдых	1,000	,695	,606
	Расходы на питание	,695	1,000	,842
	Расходы на развлечения	,606	,842	1,000
Знч. (1-сторон.)	Общие расходы на отдых	.	,000	,000
	Расходы на питание	,000	.	,000
	Расходы на развлечения	,000	,000	.
N	Общие расходы на отдых	90	90	90
	Расходы на питание	90	90	90
	Расходы на развлечения	90	90	90

Таблица 7.6

Сводка для модели

Модель	R	R квадрат	Скорректированный R-квадрат	Стд. ошибка оценки	Дурбин-Уотсон
1	,695a	,483	,471	397,433	1,246

Коэффициент R-квадрат равен 0,483; это означает, что данная регрессионная модель описывает 48% случаев, т. е. ответов респондентов о структуре их расходов на отдых за городом.

Результаты теста Дарбина–Уотсона на автокорреляцию – 1,246. Это ниже, чем 2, что свидетельствует о возможных системных связях между отклонениями наблюдаемых значений от теоретически ожидаемых значений.

Таблица 7.7

Коэффициенты

Модель		Нестандартизованные коэффициенты		Стандартизованные коэффициенты	t	Знач.
		B	Стд. ошибка	Бета		
1	(Константа)	752,252	81,985		9,175	,000
	Расходы на питание	2,671	,295	,695	9,060	,000
2	(Константа)	737,260	87,311		8,444	,000
	Расходы на питание	2,433	,549	,633	4,434	,000
	Расходы на развлечения	,371	,720	,074	,516	,607

Таблица 7.8

Диагностики коллинеарности

Модель	Измерение	Собственное значение	Показатель обусловленности	Доли дисперсии	
				(Константа)	Расходы на питание
1	1	1,863	1,000	,07	,07
	2	,137	3,688	,93	,93

2. На основе рассчитанных коэффициентов (табл. 7.7) составляем уравнение множественной регрессионной модели:

$$Y = 752,252 + 2,671X_1 + 0,074X_2,$$

где Y – общие расходы туристов на отдых за городом;

X_1 – расходы на питание;

X_2 – расходы на развлечения.

3. Исходя из анализа коэффициентов B (табл. 7.7), можно сделать вывод, что расходы на питание существенно влияют на общую сумму расходов на отдых, чем расходы на развлечения: B на питание = 2,671, B на развлечения = 0,074.

Постоянный показатель «Константа» = 752,252; это существенная величина, что говорит о том, что включенные в уравнение независимые переменные не в полной мере описывают зависимую переменную, следовательно, есть еще значительные статьи расходов, оказывающие влияние на общие затраты отдыхающих за городом.

4. Из таблицы 7.8 «Диагностика коллинеарности» мы видим, что статистики коллинеарности (взаимосвязи между независимыми переменными), т. е. толерантность = 0,291 (должно превышать 0,1), КРД = 3,435 (должен быть меньше 10), свидетельствуют о невозможности возникновения нежелательного эффекта мультиколлинеарности (взаимного влияния и зависимости друг от друга размера сумм расходов на питание и развлечения).

5. Значение показателя «Значимость» = 0,000 (табл. 7.7), что меньше 0,5. Это говорит о том, что регрессионная модель справедлива для генеральной совокупности в целом. Мы можем распространить результаты анализа данных выборки отдыхающих на всех отдыхающих на базе отдыха. Регрессионная модель описывает 48% случаев, т. е. ответов респондентов о структуре их расходов на отдых за городом. Она может быть использована специалистами по маркетингу при решении задач ценообразования на базе отдыха «Солнечная», хотя вероятность ошибки при этом остается существенной. Для минимизации ошибки необходимо более четко структурировать расходы отдыхающих, выделив в них недостающие статьи, например «расходы на транспорт», «расходы на проживание».

Задания для практических работ

Задание 1. Используя базу данных опроса отдыхающих базы отдыха «Солнечная», выявите, в какой зависимости находятся суммы расходов на отдых в целом и суммы расходов на питание для мужчин и женщин. Сделайте прогноз, как будут изменяться общие расходы на отдых при увеличении (уменьшении) расходов на питание для мужчин и женщин.

Последовательность действий:

1. Провести линейный регрессионный анализ.
2. Построить регрессионную модель (тренд), отражающую зависимость между переменными, составить уравнение регрессии.
3. Сделать вывод о том, как будет изменяться одна переменная при изменении другой.
4. Представить результаты анализа в виде линейного графика.
5. Сделать вывод о вероятности ошибки при использовании данной регрессионной модели.

Задание 2. Используя базу данных опроса отдыхающих базы отдыха «Солнечная», выявите в какой зависимости находятся суммы расходов на отдых в целом и суммы расходов на развлечения для мужчин и женщин. Сделайте прогноз, как будут изменяться общие расходы на отдых при увеличении (уменьшении) расходов на развлечения для мужчин и женщин.

Последовательность действий:

1. Провести линейный регрессионный анализ.
2. Построить регрессионную модель (тренд), отражающую зависимость между переменными, составить уравнение регрессии.
3. Сделать вывод о том, как будет изменяться одна переменная при изменении другой.
4. Представить результаты анализа в виде линейного графика.
5. Сделать вывод о вероятности ошибки при использовании данной регрессионной модели.

Задание 3. Используя базу данных опроса отдыхающих базы отдыха «Солнечная», выявите, в какой зависимости находятся общие расходы в день на отдых за городом для мужчин и женщин и следующие статьи расходов:

- расходы на питание;
- расходы на развлечения для мужчин и женщин.

Последовательность действий:

1. Проведите множественный линейный регрессионный анализ.
2. Проверьте условия получения приемлемых результатов анализа.
3. Постройте регрессионную модель, отражающую зависимость между переменными, составьте уравнение регрессии.
4. Сделайте вывод о том, как будет изменяться одна переменная при изменении других.
5. Сделайте вывод о вероятности возникновения нежелательного эффекта мультиколлинеарности.
6. Сделайте вывод о вероятности ошибки при использовании данной регрессионной модели.

Задание 4. Используя результаты оценки потребителями 15 магазинов (табл. 7.9) с точки зрения уровня обслуживания и приемлемости цен, проведите регрессионный анализ и выявите

зависимость в предпочтениях от приемлемости цен и уровня обслуживания. Магазины оценивались по 10-балльной шкале. Чем выше балл, тем выше оценка.

Таблица 7.9

Оценки потребителей

№ п/п	Предпочтение	Качество обслуживания	Цена
1	6	5	3
2	9	6	10
3	8	6	4
4	3	2	1
5	10	6	10
6	4	3	1
7	5	4	7
8	2	1	4
9	10	9	8
10	9	5	10
11	10	8	8
12	2	1	5
13	9	8	5
14	5	3	2
15	3	8	3

Последовательность действий:

1. Проведите множественный линейный регрессионный анализ.
2. Проверьте условия получения приемлемых результатов анализа.
3. Постройте регрессионную модель, отражающую зависимость между переменными, составьте уравнение регрессии.
4. Сделайте вывод о том, как будет изменяться одна переменная при изменении других.
5. Сделайте вывод о вероятности возникновения нежелательно-го эффекта мультиколлинеарности.
6. Сделайте вывод о вероятности ошибки при использовании данной регрессионной модели.

Задание 5. Имеются данные об объемах продаж, расходах на рекламу и ценах торгового предприятия. Оцените влияние цен

и расходов на рекламу на изменение объемов продаж (табл. 7.10).
Постройте уравнение регрессии [5].

Таблица 7.10

Исходная информация

№ п/п	Расходы на рекламу, тыс. р.	Объем продаж, тыс. р.	Цена, р.
1	15,1	49,2	9,8
2	25,8	60,6	13,0
3	18,9	58,2	10,1
4	24,3	63,8	13,0
5	12,0	53,1	10,6
6	10,0	53,9	9,4
7	26,7	69,6	14,1
8	30,8	55,1	11,3
9	19,7	55,0	10,3
10	23,5	57,8	12,5
11	6,4	20,2	4,0
12	22,2	75,3	14,0
13	35,1	62,9	14,0
14	28,4	80,5	16,4
15	21,9	56,4	12,1
16	3,4	29,2	6,2
17	7,9	36,8	7,8
18	33,1	75,7	14,0
19	16,9	39,9	6,6
20	18,1	35,9	7,6

Последовательность действий:

1. Проведите множественный линейный регрессионный анализ.
2. Проверьте условия получения приемлемых результатов.
3. Постройте регрессионную модель, отражающую зависимость между переменными, составьте уравнение регрессии.
4. Сделайте вывод о том, как будет изменяться одна переменная при изменении других.
5. Сделайте вывод о вероятности возникновения нежелательно-го эффекта мультиколлинеарности.
6. Сделайте вывод о вероятности ошибки при использовании данной регрессионной модели.

Задание 6. Фирма провела рекламную кампанию. Через 10 недель руководство фирмы решило проанализировать эффективность рекламной кампании, сопоставив недельные объемы продаж с расходами на рекламу (табл. 7.11).

Таблица 7.11

Результаты рекламной кампании, тыс. р.

Расходы на рекламу	5	8	6	5	3	9	12	4	3	10
Объемы продаж	72	76	78	70	68	80	82	65	62	90

Последовательность действий:

1. Проведите линейный регрессионный анализ. Постройте регрессионную модель, отражающую зависимость между переменными, составьте уравнение регрессии.

2. Сделайте вывод о том, как будут изменяться объемы продаж при изменении расходов на рекламу.

3. Сделайте вывод о вероятности ошибки при использовании данной регрессионной модели.

Задание 7. Для оценки удовлетворенности отдыхающих базы отдыха было проведено исследование и опрошено 50 человек (табл. 7.12). В качестве факторов, влияющих на удовлетворенность отдыхом, выделили месторасположение базы отдыха, качество обслуживания персонала и уровень затрат на отдых. Для оценки удовлетворенности отдыхающих и каждого фактора была предложена 5-балльная шкала. Чем выше балл, тем выше оценка.

Таблица 7.12

Результаты опроса отдыхающих

№ п/п	Удовлетворенность отдыхом	Месторасположение	Персонал	Уровень затрат	№ п/п	Удовлетворенность отдыхом	Месторасположение	Персонал	Уровень затрат
1	2	2	4	2	26	4	2	4	1
2	1	3	1	3	27	5	3	3	1
3	2	4	5	5	28	2	5	1	3
4	3	3	3	4	29	2	4	3	5
5	2	1	4	4	30	1	4	5	2

Окончание табл. 7.12

№ п/п	Удовлетворенность отдыхом	Месторасположение	Персонал	Уровень загара	№ п/п	Удовлетворенность отдыхом	Месторасположение	Персонал	Уровень загара
6	3	1	3	3	31	3	3	2	3
7	3	3	1	1	32	2	1	3	5
8	2	5	1	3	33	3	3	4	4
9	2	2	3	1	34	3	5	3	4
10	1	3	5	1	35	4	2	1	3
11	3	5	2	3	36	2	3	1	1
12	2	4	3	5	37	2	4	3	1
13	2	4	5	2	38	3	3	5	3
14	4	3	4	3	39	1	1	2	5
15	3	1	4	5	40	3	1	3	2
16	1	1	3	4	41	3	3	5	3
17	2	3	1	4	43	2	5	4	4
18	2	5	3	3	43	5	2	4	1
19	3	2	1	1	44	1	3	4	5
20	1	3	3	1	45	2	5	3	3
21	5	3	5	3	46	4	4	1	4
22	3	1	2	5	47	3	4	1	3
23	2	1	3	2	48	3	3	3	1
24	2	3	5	3	49	2	1	5	3
25	1	5	4	3	50	2	1	2	5

Проведите регрессионный анализ и выявите зависимость удовлетворенности от вышеперечисленных факторов.

Последовательность действий:

1. Проведите множественный линейный регрессионный анализ.
2. Проверьте условия получения приемлемых результатов анализа.
3. Постройте регрессионную модель, отражающую зависимость между переменными, составьте уравнение регрессии.
4. Сделайте вывод о том, как будет изменяться одна переменная при изменении других.
5. Сделайте вывод о вероятности возникновения нежелательного эффекта мультиколлинеарности.
6. Сделайте вывод о вероятности ошибки при использовании данной регрессионной модели.

8. Дискриминантный анализ

Применяется для анализа различий заранее заданных групп множества объектов исследования. При этом каждый объект может быть отнесен только к одной группе. Переменная, разделяющая совокупность объектов исследования на группы, называется *группирующей*. Признаки, используемые для выявления различий между группами, называются *дискриминационными переменными* [3].

В практике маркетинговых исследований с помощью дискриминантного анализа можно найти ответы, например, на следующие вопросы:

1. Различаются ли между собой сегменты рынка по предпочтению к различным маркам товара?
2. Какие существуют различия между постоянными покупателями супермаркетов и небольших продуктовых магазинов?
3. Какими отличительными характеристиками обладают потребители, реагирующие на рекламу в сети Интернет?
4. Чем с точки зрения демографических (социально-экономических, психографических) характеристик отличаются приверженцы данной торговой марки от тех, у кого данной приверженности нет?
5. Какие психографические (демографические, культурные и пр.) характеристики отличают восприимчивых к цене товара потребителей от невосприимчивых?

Дискриминантный анализ представляет собой альтернативу множественному регрессионному анализу в том случае, когда зависимая переменная представляет собой не количественную, а номинальную переменную.

Результатом дискриминантного анализа является построение дискриминационного уравнения, которое имеет вид:

$$d = a + b_1x_1 + b_2x_2 + \dots + b_nx_n,$$

где d – группирующая (зависимая) переменная;

a – свободный член (константа);

b_n – коэффициенты дискриминантной функции;
 x_n – дискриминационные (независимые) переменные.

С помощью этой модели, зная характеристики объекта исследования, можно с определенной степенью уверенности определить его принадлежность к одной из исследованных групп.

Этапы дискриминантного анализа:

1. Выбор переменных-предикторов (дискриминантных переменных). Необходимо выяснить, все ли выбранные переменные в действительности являются отличительными характеристиками исследуемых групп.

2. Выбор параметров, построение дискриминантной модели. Оценка статистической значимости выбранных переменных. Исследователю необходимо выбрать метод, который наилучшим образом позволит рассчитать параметры дискриминантной модели.

3. Интерпретация результатов, оценка прогнозов. В результате построения дискриминантной модели можно определить по известным значениям одной выборки неизвестные значения критерия для другой выборки.

Например, используя базу данных результатов опроса отдыхающих базы отдыха «Солнечная», требуется построить дискриминантную модель, при помощи которой на основании социально-демографических признаков – пол, количество лет, уровень дохода респондентов, расходы на отдых – можно было бы причислить туриста к одной из двух групп: выбирающих в качестве развлечений дискотеки при отдыхе за городом и не выбирающих.

Пошаговая инструкция

Шаг 1. В меню Анализ – Классификация – Дискриминантный анализ.

Шаг 2. Перенести метки независимых переменных в окно Независимые: пол, количество лет, уровень дохода респондентов, расходы на отдых.

Шаг 3. Выбрать метку группирующей переменной и перенести ее в окно Группировать по: проведение вечерних дискотек.

Шаг 4. Кнопка Задать диапазон – Дискриминантный анализ – Минимум: 1 и Максимум: 2.

Шаг 5. Кнопка Статистики – установить: Средние, Однофакторный дисперсионный анализ, нестандартизованные, внутригрупповая корреляция.

Шаг 6. Кнопка Классифицировать – Поточечные результаты, итоговая таблица, Графики: Для отдельных групп.

Шаг 7. ОК.

Интерпретация результатов

1. На первом этапе дискриминантного анализа необходимо провести оценку выбора дискриминантных переменных. В таблице «Анализ обработанных переменных» (табл. 8.1) дается обзор действительных и пропущенных значений.

Таблица 8.1

Сводка результатов обработки наблюдений

Невзвешенные наблюдения		N	Процент
Валидные		45	90,0
Исключенные	Пропущенные или лежащие вне диапазона коды группирующей переменной	0	,0
	По крайней мере одна пропущенная дискриминантная переменная	5	10,0
	Оба групповых кода пропущены или лежат вне диапазона, и отсутствует по крайней мере одна дискриминантная переменная	0	,0
	Итого искл.	5	10,0
Всего набл.		50	100,0

Из 50 респондентов, участвовавших в опросе, в построении дискриминационной функции используются данные по 45 респондентам. Число респондентов, не давших ответы хотя бы по одному из выбранных признаков, составляет 5.

Следующая таблица «Групповые статистики» (Статистические показатели в группах) содержит данные о средних значениях дискриминационных переменных в каждой группе (табл. 8.2).

Из данных табл. 8.2 можно сделать вывод, что средний возраст туристов, посещающих дискотеки – 23 года; не посещающих – 45 лет. Вообще средний возраст опрошенных респондентов – 34 года.

Таблица 8.2

Групповые статистики

Проведение вечерних дискотек		Среднее	Стд. отклонение	Количество валидных (искл. целиком)	
				Невзвешенные	Взвешенные
Да	Пол	1,57	,507	23	23,000
	Возраст	1,00	,000	23	23,000
	Количество лет	22,70	5,355	23	23,000
	Уровень дохода респондентов	1,13	,344	23	23,000
Нет	Пол	1,48	,511	23	23,000
	Возраст	1,00	,000	23	23,000
	Количество лет	44,78	13,021	23	23,000
	Уровень дохода респондентов	1,91	,733	23	23,000
Итого	Пол	1,52	,505	46	46,000
	Возраст	1,00	,000	46	46,000
	Количество лет	33,74	14,885	46	46,000
	Уровень дохода респондентов	1,52	,691	46	46,000

Уровень дохода туристов, посещающих дискотеки ниже (1,13, т. е. ближе к 1, что соответствует интервалу менее 5 000 р.), чем уровень дохода, не посещающих дискотеки (1,91, т. е. ближе к 2, что соответствует интервалу 5 000–100 000 р. и выше).

В среднем мужчины и женщины одинаково посещают дискотеки (среднее по полу равно 1,52, что является серединой между номинальными переменными «1» – мужчины, «2» – женщины). При этом доля женщин среди туристов, посещающих дискотеки, составляет 57% (среднее 1,57), среди не посещающих – 48% (среднее 1,48).

2. Оценка статистической значимости выбранных переменных.

Исходя из табл. 8.3 «Критерий равенства групповых средних», необходимо выбрать независимые дискриминационные переменные. Для проведения теста на равенство средних значений в группах используется Лямбда Уилкса (Wilks' Lambda). Основным результатом определяется величиной «Significance» («Значимость»). Если

«Значимость» (p) меньше 0,05, это означает, что различия между средними значениями дискриминационных переменных в исследуемых группах являются статистически значимыми (см. раздел 5 «Сравнение средних величин»).

Таблица 8.3

Критерий равенства групповых средних

	Лямбда Уилкса	F	ст. св1	ст. св2	Знч.
Пол	,992	,336	1	44	,565
Возраст, а					
Количество лет	,437	56,605	1	44	,000
Уровень дохода респондентов	,672	21,470	1	44	,000

а. Невозможно вычислить, так как эта переменная является константой.

Значимость переменных: «Уровень дохода» – 0,000; «Количество лет» – 0,000, расходы на питание – 0,017, что меньше 0,05. Это означает, что туристы, посещающие и не посещающие дискотеки, отличаются друг от друга по данным характеристикам. Для переменной «пол» значимость признака 0,565, что больше 0,05. В случае, если значимость признака больше 0,05, данный признак необходимо исключить из дискриминационных переменных и заново сформировать задание для проведения дискриминационного анализа.

Далее следует доказать, что выбранные переменные действительно являются независимыми друг от друга. В этом случае рассчитывается коэффициент корреляции, для характеристики связи между существующими переменными. В табл. 8.4 «Объединенные внутригрупповые матрицы» показаны коэффициенты корреляции между независимыми переменными для исследуемых групп.

Коэффициент корреляции r – статистический показатель вероятностной связи между двумя переменными, измеренными в количественной шкале. Вероятностная связь характеризуется тем, что каждому значению одной переменной соответствует множество значений другой переменной, r меняется от -1 до $+1$. Чем ближе коэффициент корреляции к 0, тем слабее зависимость между переменными (см. раздел 3.3 «Корреляции»).

Таблица 8.4

Объединенные внутригрупповые матрицы

		Уровень дохода	Количество лет	Расходы на питание
Корреляция	Уровень дохода	1,000	-,379	,630
	Количество лет	-,379	1,000	-,426
	Расходы на питание	,630	-,426	1,000
		Пол	Количество лет	Уровень дохода
Корреляция	Пол	1,000	,221	-,214
	Количество лет	,221	1,000	-,516
	Уровень дохода	-,214	-,516	1,000

Коэффициент корреляции между возрастом (количество лет) и полом = 0,221, что меньше 0,5. Это доказывает отсутствие корреляционной зависимости между данными переменными. А коэффициент корреляции между переменными уровень дохода и расходы на питание = 0,630, между количеством лет и уровнем дохода = -0,516, что больше 0,5. Соответственно, данные переменные влияют друг на друга и эта пара должна быть исключена из анализа.

3. Построение дискриминационной модели.

Программа рассчитывает коэффициенты дискриминантной функции: $d = a + b_1x_1 + b_2x_2 + \dots + b_nx_n$ (табл. 8.5).

Таблица 8.5

Нормированные коэффициенты канонической дискриминантной функции

	Функция
	1
Уровень дохода	,618
Количество лет	1,028
Расходы на питание	,262

Используя данные коэффициенты, можно оценить относительный вклад каждой переменной в различие исследуемых групп.

В нашем примере возраст респондентов (количество лет) в 1,7 (1,028/0,618) раза больше влияет на желание посещать дискотеки, чем доход, и в 3,9 (1,028/0,262) раза больше, чем сумма расходов, потраченных на питание на отдыхе.

Корреляционные коэффициенты, представленные в табл. 8.6, позволяют оценить силу связи между переменными и стандартизированными значениями дискриминантной функции.

Таблица 8.6

Структурная матрица

	Функция	
	1	
Количество лет	,682	
Уровень дохода	,394	
Расходы на питание	,214	

Используя коэффициенты, представленные в табл. 8.7, составим дискриминационную модель:

$$d = -5,642 + 0,622X_1 + 0,106X_2 + 0,002X_3,$$

где X_1 – уровень дохода;

X_2 – возраст;

X_3 – расходы на питание (выделены красным, так как требуют исключения из модели в силу высокой корреляции между переменными).

Таблица 8.7

Коэффициенты канонической дискриминантной функции

	Функция	
	1	
Уровень дохода	,622	
Количество лет	,106	
Расходы на питание	,002	
(Константа)	-5,642	

Итоговая дискриминантная модель:

$$d = -5,642 + 0,622X_1 + 0,106X_2.$$

Четкость разделения исследуемых групп характеризуется расстоянием между средними значениями дискриминантной функции в исследуемых группах (табл. 8.8).

Таблица 8.8

Функции в центроидах групп

Проведение вечерних дискотек*	Функция
	1
Да	-1,696
Нет	1,773

* Ненормированные канонические дискриминантные функции вычислены в центроидах групп.

В данном случае расстояние между средними в группах выбирающих и не выбирающих дискотеку большое (от $-1,696$ до $1,773$). Чем больше расстояние, тем больше различий между группами.

Различия между исследуемыми группами показано также на графиках распределения значений дискриминантной функции (рис. 8.1 и 8.2).

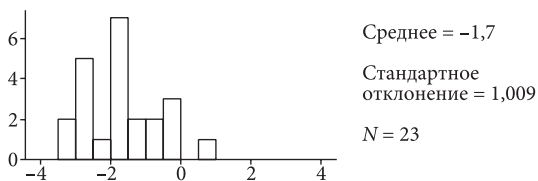


Рис. 8.1. Распределение значений дискриминантной функции для группы «посещающие дискотеки»

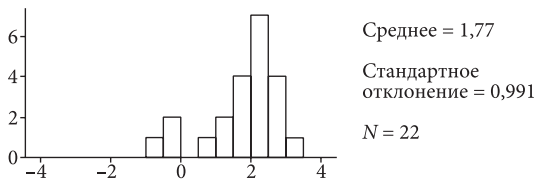


Рис. 8.2. Распределение значений дискриминантной функции для группы «не посещающие дискотеки»

Чем больше рассеяние значений дискриминантной функции в исследуемых группах, тем шире область их пересечения и слабее различия между группами, и наоборот.

На графиках показано четкое различие между исследуемыми группами респондентов, посещающих и не посещающих дискотеку.

4. Прогнозы выбора дискотек определенным туристом, исходя из его возраста и уровня дохода.

Для туриста в возрасте 20 лет и уровнем дохода от 5 000 до 10 000 р. (категория 2) значение дискриминантной функции d составит

$$d = -5,642 + 0,622 \times 2 + 0,106 \times 20 = -2,278.$$

Исходя из анализа графиков не посещающих дискотеки, значения дискриминантной функции, равное $-2,278$, не имеет ни один человек, а из посещающих имеет 1 человек; следовательно, можно сделать вывод, что туристы в возрасте 20 лет и уровнем дохода от 5 000 до 10 000 р. скорее всего будут посещать дискотеки.

Точность прогнозов оценивается по результатам классификации по исследуемым группам. В табл. 8.9 представлены результаты классификации респондентов: фактическая и предсказуемая принадлежность к исследуемым группам. При несовпадении групп значение переменной указывается как (**). В столбце указывается вероятность, с которой респондент может быть причислен к данной группе. Например, турист под номером 5 с вероятностью 83,7% может быть причислен к группе посещающих дискотеку.

Точность прогнозов дискриминантной модели определяется по данным сводной таблицы результатов классификации (табл. 8.10).

Таким образом, 22 туриста из 23, предпочитающие дискотеки корректно причислены к данной группе, а 1 – нет. Из не посещающих – только 3 корректно, а 19 – нет.

Это дает возможность сделать вывод, что 86,4% не предпочитающих дискотеки могут их посещать.

Таблица 8.9

Поточечные статистики

Номер наблюдения	Фактическая группа	Наивероятнейшая группа				Вторая вероятнейшая группа			Дискриминантные баллы		
		Предказанная группа		$P(G = g D = d)$	Квадрат расстояния Махалонобиса до центра	Группа	$P(G = g D = d)$	Квадрат расстояния Махалонобиса до центра			
		P	ст. св.								
Исходные	1	1	,769	1	,999	,087	2	,001	14,165	-1,990	
	2	2	,498	1	1,000	,459	1	,000	17,199	2,451	
	3	1	,139	1	,709	2,185	2	,291	3,965	-,218	
	4	1	,278	1	,905	1,179	2	,095	5,681	-,610	
	5	1	2**	,207	1	,837	1,595	1	,163	4,869	,510
	6	2	,643	1	,988	,214	1	,012	9,038	1,310	
	7	1	,782	1	,999	,077	2	,001	14,035	-1,973	
	8	2	,950	1	,997	,004	1	,003	11,604	1,710	
	9	2	,821	1	,995	,051	1	,005	10,522	1,548	
	10	1	,625	1	1,000	,239	2	,000	15,671	-2,185	

** Неправильно классифицированное наблюдение.

Таблица 8.10

Результаты классификации*

		Проведение вечерних дискотек	Предказанная принадлежность к группе		Итого
			Да	Нет	
Исходные	Частота	Да	22	1	23
		Нет	3	19	22
	%	Да	95,7	4,3	100,0
		Нет	13,6	86,4	100,0

* 91,1% исходных сгруппированных наблюдений классифицировано правильно.

Задания для практических работ

Задание 1. Предположим, что мы хотим определить главные характеристики семей, которые владеют фургоном. Данные получены на основе выборки, включающей 40 семей [3]. Были собраны следующие данные (табл. 8.11):

- доход в долларах;
- количество членов семьи;
- образование главы семьи (в годах обучения);
- регион проживания (1 – северный, 2 – южный);
- стиль жизни (1 – либеральный, 2 – консервативный);
- количество машин в собственности;
- владеет ли семья фургоном (1 – да, 2 – нет).

Таблица 8.11

Данные выборки

№ п/п	Доход, тыс. дол.	Количество членов семьи	Образование главы семьи	Регион проживания	Стиль жизни	Количество машин в собственности	Владеет ли семья фургоном?
1	26	3	12	1	1	1	1
2	17	4	12	1	1	1	2
3	14	2	10	1	1	1	2
4	35	4	9	1	1	1	2
5	24	3	8	1	1	1	2
6	13	4	6	1	2	1	2
7	34	3	8	1	2	1	2
8	16	3	11	1	2	1	2
9	14	2	12	1	2	1	2
10	27	2	12	1	2	2	2
11	15	4	12	2	1	1	1
12	34	4	11	2	2	1	1
13	15	3	12	2	2	1	1
14	17	2	12	1	2	1	2
15	13	3	8	1	2	1	2
16	26	2	12	1	2	1	2
17	17	4	12	2	2	1	1
18	37	3	12	2	2	1	1
19	24	4	10	2	2	1	1

Окончание табл. 8.11

№ п/п	Доход, тыс. дол.	Количество членов семьи	Образование главы семьи	Регион проживания	Стиль жизни	Количество машин в собственности	Владеет ли семья фургоном?
20	14	3	10	2	2	1	2
21	24	4	10	2	2	1	2
22	26	4	13	2	1	1	2
23	15	6	13	1	1	2	2
24	26	8	12	2	2	2	2
25	16	8	12	1	1	2	2
26	38	3	12	1	1	1	1
27	88	3	12	1	2	1	1
28	55	3	12	2	1	2	1
29	73	4	12	2	1	1	1
30	39	3	10	2	2	2	1
31	38	3	10	2	2	1	1
32	46	3	16	2	1	1	1
33	39	4	16	2	1	1	1
34	46	2	14	2	2	1	1
35	99	5	12	2	1	3	1
36	72	9	10	2	2	2	1
37	64	7	12	2	2	2	1
38	71	7	15	2	1	1	1
39	48	4	12	2	2	1	2
40	69	9	12	1	1	2	2

На основе построенной модели определите вероятность покупки фургона или, наоборот, отказа от таковой американского гражданина с доходом 40 тыс. дол., 5 членами семьи, образованием 9 лет, проживающего в северном регионе, придерживающегося консервативных взглядов и имеющего в распоряжении 2 автомобиля.

Задание 2. Используя базу данных результатов опроса отдыхающих базы отдыха «Солнечная», постройте дискриминантную модель, с помощью которой, основываясь на социально-демографических признаках, можно было бы причислить туриста к одной из двух групп: выбирающих в качестве развлечений спортивные мероприятия во время отдыха за городом и не выбирающих.

Задание 3. Используя базу данных результатов опроса отдыхающих базы отдыха «Солнечная», постройте дискриминантную модель, с помощью которой, основываясь на социально-демографических признаках, можно было бы причислить туриста к одной из двух групп: выбирающих в качестве развлечений пешие прогулки в лесу при отдыхе за городом и не выбирающих.

Задание 4. Используя базу данных результатов опроса отдыхающих базы отдыха «Солнечная», построить дискриминантную модель, с помощью которой, основываясь на социально-демографических признаках, можно было бы причислить туриста к одной из двух групп: выбирающих в качестве развлечений туристические походы во время отдыха за городом и не выбирающих.

Задание 5. Используя базу данных результатов опроса отдыхающих базы отдыха «Солнечная», постройте дискриминантную модель, с помощью которой, основываясь на социально-демографических признаках, можно было бы причислить туриста к одной из двух групп: выбирающих в качестве развлечений катание на лошадях во время отдыха за городом и не выбирающих.

Задание 6. Используя базу данных результатов опроса отдыхающих базы отдыха «Солнечная», постройте дискриминантную модель, с помощью которой, основываясь на социально-демографических признаках, можно было бы причислить туриста к одной из двух групп: выбирающих в качестве развлечений посещение бани, сауны, массаж во время отдыха за городом и не выбирающих.

Задание 7. Используя базу данных результатов опроса отдыхающих базы отдыха «Солнечная», постройте дискриминантную модель, с помощью которой, основываясь на социально-демографических признаках, можно было бы причислить туриста к одной из двух групп: выбирающих в качестве развлечений бассейн во время отдыха за городом и не выбирающих.

Задание 8. Используя базу данных результатов опроса отдыхающих базы отдыха «Солнечная», постройте дискриминантную

модель, с помощью которой, основываясь на социально-демографических признаках, можно было бы причислить туриста к одной из двух групп: выбирающих в качестве развлечений культмассовые мероприятия во время отдыха за городом и не выбирающих.

Задание 9. Используя базу данных результатов опроса отдыхающих базы отдыха «Солнечная», построить дискриминантную модель, с помощью которой, основываясь на социально-демографических признаках, можно было бы причислить туриста к одной из двух групп: мотивом выбора места отдыха за городом является близость к городу или не является.

Задание 10. Используя базу данных результатов опроса отдыхающих базы отдыха «Солнечная», постройте дискриминантную модель, с помощью которой, основываясь на социально-демографических признаках, можно было бы причислить туриста к одной из двух групп: мотивом выбора места отдыха за городом являются приемлемые цены или таковые не являются.

Задание 11. Используя базу данных результатов опроса отдыхающих базы отдыха «Солнечная», постройте дискриминантную модель, с помощью которой, основываясь на социально-демографических признаках, можно было бы причислить туриста к одной из двух групп: мотивом выбора места отдыха за городом является близость водоема, леса или не является.

Задание 12. Используя базу данных результатов опроса отдыхающих базы отдыха «Солнечная», постройте дискриминантную модель, с помощью которой, основываясь на социально-демографических признаках, можно было бы причислить туриста к одной из двух групп: мотивом выбора места отдыха за городом является уровень комфорта или не является.

Задание 13. Используя базу данных результатов опроса отдыхающих базы отдыха «Солнечная», постройте дискриминантную модель, с помощью которой, основываясь на социально-демографических признаках, можно было бы причислить туриста к одной

из двух групп: мотивом выбора места отдыха за городом является тишина или не является.

Задание 14. Используя базу данных результатов опроса отдыхающих базы отдыха «Солнечная», постройте дискриминантную модель, с помощью которой, основываясь на социально-демографических признаках, можно было бы причислить туриста к одной из двух групп: мотивом выбора места отдыха за городом является вкусная еда или не является.

Задание 15. Используя базу данных результатов опроса отдыхающих базы отдыха «Солнечная», постройте дискриминантную модель, с помощью которой, основываясь на социально-демографических признаках, можно было бы причислить туриста к одной из двух групп: мотивом выбора места отдыха за городом является наличие развлечений или не является.

Задание 16. Используя базу данных результатов опроса отдыхающих базы отдыха «Солнечная», постройте дискриминантную модель, с помощью которой, основываясь на социально-демографических признаках, можно было бы причислить туриста к одной из двух групп: мотивом выбора места отдыха за городом является комфорт отдыха с детьми или не является.

9. Кластерный анализ

Кластерный анализ используется для классификации объектов по заданным признакам. Задача кластерного анализа состоит в формировании групп:

- однородных внутри, объекты схожи между собой (условие внутренней гомогенности);
- отличных от объектов в других группах (условие внешней гетерогенности) [2].

Целью кластерного анализа в маркетинге является определение целевых групп потребителей, для которых необходимо разработать уникальный комплекс инструментов маркетинга.

Пример. Курильщики сигар, возраст и уровень доходов которых известны, исследуются на предмет возможности их разделения на однородные группы (кластеры) (рис. 9.1).

В варианте В однородные кластеры не выявлены. Следовательно, целенаправленная дифференциация торгового предложения невозможна.

В варианте А выявлены две однородные группы курильщиков сигар: «старые и бедные», «молодые и богатые», которых можно считать двумя целевыми группами потребителей. В этом случае целесообразно разработать два специальных торговых предложения – уникальных по цене, уровню качества продукции, упаковке, системе продвижения товара и т. д. [2].

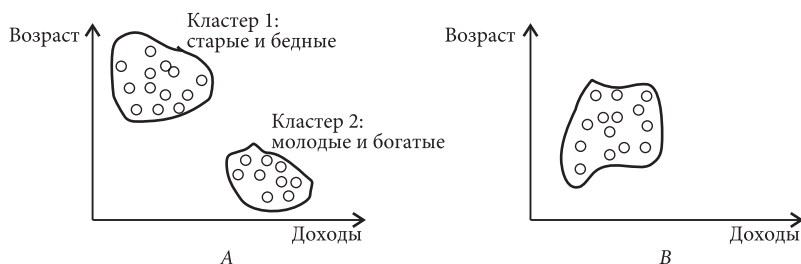


Рис. 9.1. Кластерный анализ [2]

Техника кластерного анализа заключается в выявлении уровня схожести всех исследуемых элементов и последовательном объединении элементов в порядке возрастания уровня различия между ними. Число выявленных кластеров зависит от заданного уровня схожести элементов, включаемых в один кластер [2].

Основные термины, применяемые в кластерном анализе

Евклидово расстояние – квадратный корень из суммы квадратов разностей в значениях для каждой переменной.

Иерархическая кластеризация – метод, при котором строится иерархическая или древовидная структура. Позволяет детально исследовать различия между объектами, выбрать оптимальное число кластеров.

Агломеративная, или объединяющая, кластеризация – иерархический метод формирования кластеров, при котором каждый объект сначала находится в отдельном кластере, затем объекты группируются во все более и более крупные кластеры [2].

Разделяющая, или дивизивная, кластеризация – иерархический метод кластеризации, при котором все объекты первоначально находятся в одном большом кластере. Кластеры формируются делением этого большого кластера на более мелкие [2].

Методы связи – агломеративные методы иерархической кластеризации, которые объединяют объекты в кластер на основе вычисленного расстояния между ними (рис. 9.2).

Метод одиночной связи – метод связи, в основе которого лежит минимальное расстояние между объектами, или правило ближайшего соседа.

Метод полной связи – метод связи, в основе которого лежит максимальное расстояние между объектами, или правило дальнего соседа.

Метод средней связи – метод связи, в основе которого лежит среднее значение всех расстояний, измеренных между объектами двух кластеров; при этом в каждую пару входят объекты из разных кластеров.

Дисперсионные методы – метод Варда и центроидный метод (рис. 9.3).

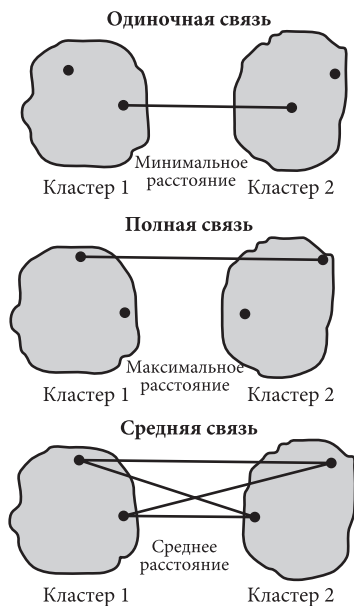


Рис. 9.2. Методы связи [2]

Метод Варда – дисперсионный метод, при котором кластеры формируют таким образом, чтобы минимизировать квадраты евклидовых расстояний до значений средних кластерных.

Центроидный метод – дисперсионный метод, при котором расстояние между двумя кластерами представляют собой расстояние между их центроидами (средними для всех переменных).

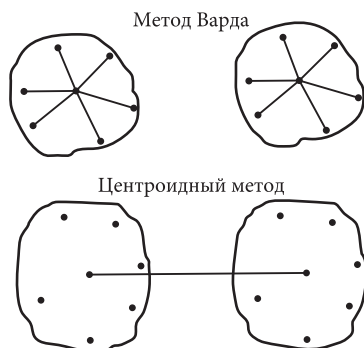


Рис. 9.3. Дисперсионные методы [2]

Метод к-средних – неиерархический метод кластеризации, при котором сначала определяют центр кластера, а затем группируют все переменные в пределах заданного порогового значения.

Последовательный пороговый метод – неиерархический метод кластеризации, при котором выбирают кластер, и все объекты, находящиеся в пределах заданного от центра порогового значения, группируют вместе.

Параллельный пороговый метод – неиерархический метод кластеризации, при котором одновременно определяют несколько кластерных центров. Все объекты, находящиеся в пределах заданного центром порогового значения, группируют вместе.

Метод оптимизирующего распределения – неиерархический метод кластеризации, который позволяет поставить объекты в соответствие другим кластерам (перераспределить объекты), чтобы оптимизировать суммарный критерий [2].

Пример. Рассмотрим кластеризацию потребителей на основе их отношения к посещению магазинов для покупки товаров. Опираясь на опыт в прошлом, маркетологи определили шесть переменных. Потребителей попросили выразить их степень согласия со следующими утверждениями по семибальной шкале (1 – не согласен, 7 – абсолютно согласен):

Фас_1 – Посещение магазинов для покупки товаров – приятный процесс;

Фас_2 – Посещение магазинов для покупки товаров плохо сказывается на нашем бюджете;

Фас_3 – Я совмещаю посещение магазинов для покупки товаров с питанием вне дома;

Фас_4 – Я стараюсь сделать лучшие покупки при посещении магазинов;

Фас_5 – Мне не нравится посещение магазинов для покупки товаров;

Фас_6 – Можно сэкономить много денег, сравнивая цены в разных магазинах [2].

*Задание 2.sav [Наборданных2] - Редактор данных IBM SPSS Statistics

Файл Правка Вид Данные Преобразовать Анализ Прямой маркетинг Графика Сервис Окно Справка

13:

	Респондент	Фас_1	Фас_2	Фас_3	Фас_4	Фас_5	Фас_6	пер	пер	пер
1	1	6	4	7	3	2	3			
2	2	2	3	1	4	5	4			
3	3	7	2	6	4	1	3			
4	4	4	6	4	5	3	6			
5	5	1	3	2	2	6	4			
6	6	6	4	6	3	3	4			
7	7	5	3	6	3	3	4			
8	8	7	3	7	4	1	4			
9	9	2	4	3	3	6	3			
10	10	3	5	3	6	4	6			
11	11	1	3	2	3	5	3			
12	12	5	4	5	4	2	4			
13	13	2	2	1	5	4	4			
14	14	4	6	4	6	4	7			
15	15	6	5	4	2	1	4			
16	16	3	5	4	6	4	7			
17	17	4	4	7	2	2	5			
18	18	3	7	2	6	4	3			
19	19	4	6	3	7	2	7			
20	20	2	3	2	4	7	2			
21										
22										
23										
24										
25										
26										
27										
28										
29										

Данные Переменные

Рис. 9.4. Данные для кластеризации

Пошаговая инструкция

Шаг 1. Подготовить данные для анализа в SPSS.

Шаг 2. Анализ – выбрать Классификация. В рассматриваемом примере применяется иерархический кластерный анализ.

Шаг 3. Анализ – Классификация – Иерархическая кластеризация – Иерархический кластерный анализ (рис. 9.5).

Из списка семи переменных исходного массива данных следует выбрать переменные, по которым будет производиться формирование кластеров, и перенести их в правое поле Переменные. В данном случае – это переменные, характеризующие отношение респондентов к посещению магазинов для покупки товаров.

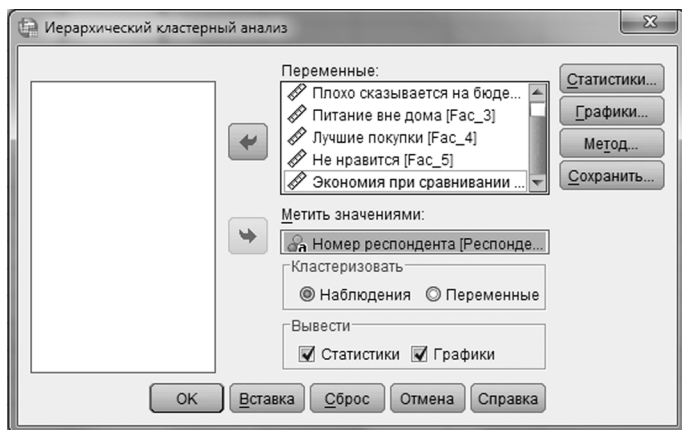


Рис. 9.5. Иерархический кластерный анализ

Также следует выбрать переменную «Номер респондента» и перенести ее в правое поле Метить значениями (только текстовые переменные!).

В поле Кластеризовать выбрать один из двух предлагаемых вариантов: Наблюдения или Переменные. В нашем примере выбирается вариант Наблюдения, в ходе кластерного анализа будут собираться в кластеры потребители, а не характеристики их отношения.

Шаг 4. Нажать кнопку Статистики – диалоговое окно Статистические показатели.

В окне Статистические показатели отметить команды Порядок агломерации и Матрица близостей. Это позволит вывести на экран таблицу, содержащую результаты сравнения объектов исследования (Матрица близостей), и таблицу, отображающую алгоритм формирования кластеров (Порядок агломерации) (рис. 9.6).

Кнопка Продолжить – возврат в главное диалоговое окно Иерархический кластерный анализ.

Шаг 5. Кнопка Графики – на экране появляется вспомогательное диалоговое окно Графики (рис. 9.7).

В диалоговом окне представлены команды на построение различных графиков и диаграмм, описывающих процедуру формирования кластеров.

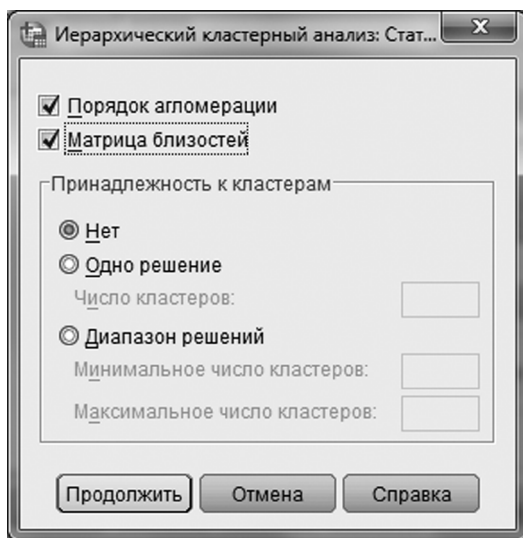


Рис. 9.6. Диалоговое окно «Статистические показатели»

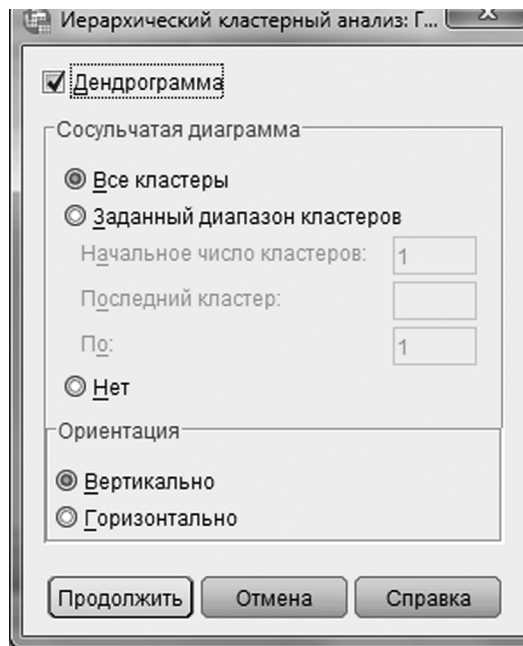


Рис. 9.7. Диалоговое окно «Графики»

Выбрать команду Дендограмма. Данная команда выводит на экран дендограмму, которая является графическим отображением выполнения алгоритма формирования кластеров.

Продолжить – Иерархический кластерный анализ.

Шаг 6. Кнопка Метод – на экране появляется одноименное вспомогательное диалоговое окно «Методы» (рис. 9.8).

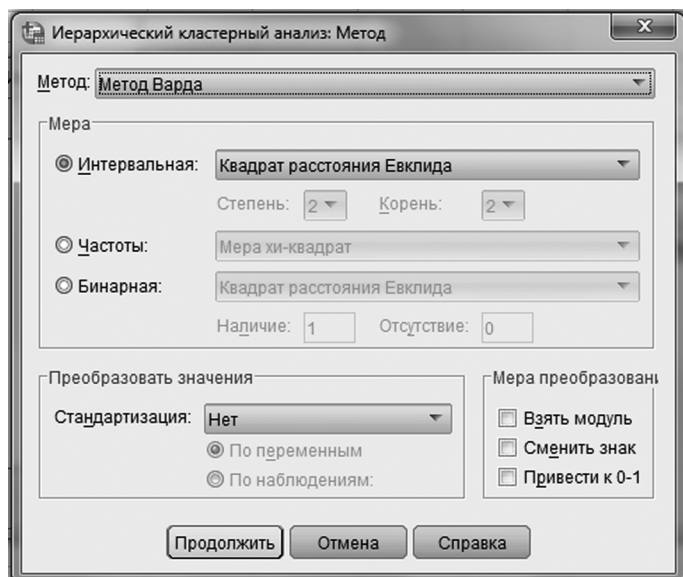


Рис. 9.8. Методы кластерного анализа

В поле Метод – выбрать метод Варда.

В поле Мера – выбрать интервальный показатель, по которому совокупность объектов исследования разделяется на кластеры.

В качестве показателя, характеризующего степень схожести (различия) объектов исследования, выбирается квадрат евклидова расстояния.

Кнопка Продолжить – возврат в главное диалоговое окно Иерархический кластерный анализ.

Шаг 7. Кнопка Сохранить – диалоговое окно Сохранить позволяет сохранить результаты кластерного анализа как новые переменные в исходной базе данных. В результате создается новая

переменная, значения которой представляют собой номера кластеров, к которым относится тот или иной объект исследования.

Шаг 8. Кнопка ОК.

Интерпретация полученных результатов

Таблица 9.1

Шаги агломерации

Этап	Кластер объединен с		Коэффициенты	Этап первого появления кластера		Следующий этап
	кластером 1	кластером 2		Кластер 1	Кластер 2	
1	14	16	1,000	0	0	6
2	6	7	2,000	0	0	7
3	2	13	3,500	0	0	15
4	5	11	5,000	0	0	11
5	3	8	6,500	0	0	16
6	10	14	8,167	0	1	9
7	6	12	10,500	2	0	10
8	9	20	13,000	0	0	11
9	4	10	15,583	0	6	12
10	1	6	18,500	0	7	13
11	5	9	23,000	4	8	15
12	4	19	27,750	9	0	17
13	1	17	33,100	10	0	14
14	1	15	41,333	13	0	16
15	2	5	51,833	3	11	18
16	1	3	64,500	14	5	19
17	4	18	79,667	12	0	18
18	2	4	172,667	15	17	19
19	1	2	328,600	16	18	0

В плане агломерации показано количество случаев или кластеров, которые нужно объединить на каждой стадии (рис. 9.9). Первая строка соответствует первой стадии, когда есть 19 кластеров. На этой стадии объединены респонденты 14 и 16, квадрат евклидового расстояния между точками, соответствующими этим двум респондентам, дан в колонке «Коэффициент». Колонка «Стадия, на

которой впервые появился кластер» показывает стадию, на которой впервые был сформирован кластер. Например, 1 на стадии 6 указывает на то, что респондента 14 впервые включили в кластер на стадии 1. Последняя колонка показывает стадию, на которой другой случай или кластер объединили с этим кластером.

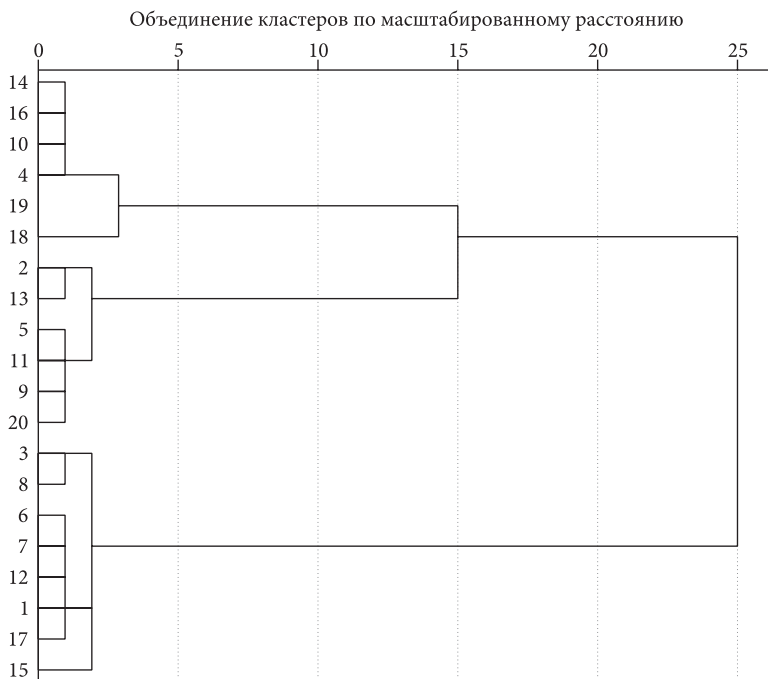


Рис. 9.9. Древоидная диаграмма, используемая в методе Варда

Древоидную диаграмму читают слева направо. Вертикальные линии показывают кластеры, объединяемые вместе. Положение линии показывает расстояния, при которых объединили кластеры.

Для определения количества кластеров можно руководствоваться следующим:

- теоретическими и практическими соображениями;
- в качестве критерия можно использовать расстояния, при которых объединяют кластеры. В нашем случае значение в колонке «Коэффициент» увеличивается более чем вдвое при переходе от стадии 17 к 18. На последних стадиях древоидной диаграммы

кластеры объединяются при больших расстояниях. Таким образом, принимаем решение о трех кластерах;

- должны быть достаточно выразительными относительные размеры кластеров. Подсчитав частоты кластерной принадлежности, видно, что выбор трех кластеров приводит к кластерам, содержащим 8, 6 и 4 элементов; если же перейти к 4 – то размеры кластеров будут 8, 6, 5, 1.

Таблица 9.2

Результаты кластерного анализа

Ward Method		Приятный процесс	Плохо сказывается на бюджете	Питание вне дома	Лучшие покупки	Не нравится	Экономия при сравнении цен
1	Среднее	5,75	3,63	6,00	3,13	1,88	3,88
	<i>N</i>	8	8	8	8	8	8
	Стд. отклонение	1,035	,916	1,069	,835	,835	,641
2	Среднее	1,67	3,00	1,83	3,50	5,50	3,33
	<i>N</i>	6	6	6	6	6	6
	Стд. отклонение	,516	,632	,753	1,049	1,049	,816
3	Среднее	3,50	5,83	3,33	6,00	3,50	6,00
	<i>N</i>	6	6	6	6	6	6
	Стд. отклонение	,548	,753	,816	,632	,837	1,549
Итого	Среднее	3,85	4,10	3,95	4,10	3,45	4,35
	<i>N</i>	20	20	20	20	20	20
	Стд. отклонение	1,899	1,410	2,012	1,518	1,761	1,496

Кластер 1 имеет относительно высокие значения по фактору 1 («Посещение магазинов – приятный процесс») и переменной 3 («Я совмещаю посещение магазинов с питанием вне дома»). Он также имеет низкое значение по фактору 5: «Меня не интересуют покупки». Следовательно, кластер 1 можно назвать так: «Любители посещать магазины и делать покупки». Этот кластер состоит из случаев: 1, 3, 6, 7, 8, 12, 15 и 17. Кластер 2 – прямая противоположность кластеру 1: он имеет низкие значения по факторам 1

и 3 – и высокое по фактору 5; значит, этот кластер можно назвать «апатичные покупатели». В кластер входят случаи 2, 5, 9, 11, 13 и 20. Кластер 3 имеет высокие значения по факторам 2 («Посещение магазинов плохо сказывается на моем бюджете»), 4 («Я стараюсь сделать лучшие покупки, посещая магазины»), 6 («Можно сэкономить много денег, сравнивая цены в разных магазинах»). Этот кластер можно назвать экономные покупатели.

Задания для практических работ

Задание 1. Проведите кластеризацию туристов на основе их мотивов при выборе места отдыха за городом.

На вопрос «Что для вас важно при выборе места отдыха за городом?» респонденты выбирали следующие варианты ответа (1 – очень важно, 2 – важно, 3 – и да и нет, 4 – не важно, 5 – совсем не важно):

- 1) близость к городу;
- 2) приемлемые цены;
- 3) близость водоема, леса;
- 4) уровень комфорта;
- 5) тишина, уединение;
- 6) хорошее питание;
- 7) наличие развлечений;
- 8) комфорт отдыха с детьми;
- 9) возможность лечения, ухода за здоровьем;
- 10) организация детского отдыха.

Таблица 9.3

Мотивы выбора отдыхающими места отдыха за городом

№ п/п	1	2	3	4	5	6	7	8	9	10
1	2.0	2.0	2.0	2.0	4.0	4.0	2.0	4.0	5.0	4.0
2	4.0	3.0	3.0	3.0	1.0	1.0	5.0	22.0	2.0	4.0
3	2.0	3.0	4.0	2.0	4.0	2.0	2.0	2.0	4.0	2.0
4	1.0	2.0	5.0	4.0	5.0	2.0	1.0	4.0	4.0	1.0
5	5.0	3.0	2.0	2.0	5.0	3.0	2.0	1.0	5.0	2.0
6	1.0	4.0	3.0	1.0	4.0	2.0	3.0	1.0	55.0	2.0
7	2.0	2.0	3.0	4.0	5.0	5.0	2.0	5.0	5.0	4.0
8	3.0	2.0	1.0	2.0	3.0	1.0	5.0	1.0	3.0	2.0

Окончание табл. 9.3

№ п/п	1	2	3	4	5	6	7	8	9	10
9	4.0	3.0	2.0	2.0	2.0	1.0	5.0	1.0	2.0	4.0
10	1.0	1.0	3.0	4.0	5.0	4.0	1.0	3.0	5.0	4.0
11	5.0	2.0	3.0	5.0	4.0	4.0	2.0	4.0	5.0	5.0
12	5.0	1.0	4.0	5.0	5.0	5.0	5.0	4.0	5.0	5.0
13	3.0	1.0	4.0	5.0	4.0	5.0	5.0	4.0	5.0	5.0
14	2.0	2.0	5.0	2.0	5.0	2.0	2.0	2.0	4.0	2.0
15	1.0	99.0	3.0	3.0	3.0	2.0	1.0	2.0	5.0	1.0
16	4.0	2.0	3.0	2.0	3.0	2.0	3.0	2.0	4.0	4.0
17	2.0	2.0	1.0	2.0	4.0	4.0	2.0	2.0	5.0	2.0
18	3.0	4.0	4.0	1.0	4.0	4.0	3.0	22.0	5.0	4.0
19	5.0	1.0	2.0	4.0	5.0	5.0	1.0	4.0	5.0	5.0
20	1.0	1.0	1.0	4.0	5.0	5.0	1.0	5.0	5.0	1.0
21	1.0	1.0	3.0	1.0	1.0	1.0	5.0	1.0	2.0	5.0
22	2.0	2.0	1.0	2.0	1.0	2.0	5.0	1.0	2.0	4.0
23	2.0	3.0	1.0	2.0	1.0	1.0	2.0	1.0	2.0	2.0
24	3.0	2.0	3.0	2.0	4.0	2.0	2.0	2.0	3.0	2.0
25	4.0	2.0	5.0	2.0	2.0	2.0	4.0	2.0	4.0	2.0
26	5.0	1.0	5.0	2.0	1.0	1.0	5.0	1.0	2.0	2.0
27	5.0	1.0	4.0	5.0	4.0	3.0	5.0	1.0	5.0	2.0
28	2.0	1.0	3.0	2.0	2.0	2.0	4.0	1.0	2.0	2.0
29	2.0	2.0	5.0	2.0	1.0	1.0	5.0	1.0	2.0	4.0
30	1.0	1.0	2.0	1.0	4.0	4.0	3.0	1.0	4.0	1.0

Задание 2. Проведите кластеризацию туристов на основе выбора источников информации о вариантах отдыха за городом.

Задание 3. Проведите кластеризацию туристов на основе выбора способа организации досуга при отдыхе за городом:

- 1) проведение вечерних дискотек;
- 2) пешие прогулки по лесу, «тропа здоровья»;
- 3) туристические походы;
- 4) катание на лошадях;
- 5) баня, сауна, массаж;
- 6) посещение бассейна;
- 7) катание на катерах, скутерах, катамаранах;

- 8) конкурсы, игры, концерты;
 9) дневные и вечерние мероприятия для детей, детская комната.

Таблица 9.4

Выбор отдыхающими способа проведения досуга*

№ п/п	1	2	3	4	5	6	7	8	9	10
1	1.0	1.0	2.0	1.0	1.0	1.0	1.0	1.0	2.0	1.0
2	1.0	2.0	2.0	1.0	1.0	1.0	1.0	1.0	2.0	1.0
3	2.0	1.0	2.0	2.0	1.0	1.0	1.0	1.0	2.0	1.0
4	2.0	1.0	2.0	2.0	2.0	1.0	1.0	1.0	1.0	2.0
5	2.0	1.0	2.0	2.0	2.0	1.0	1.0	1.0	1.0	1.0
6	2.0	2.0	2.0	2.0	2.0	1.0	1.0	2.0	2.0	1.0
7	2.0	1.0	2.0	1.0	1.0	1.0	1.0	1.0	2.0	1.0
8	2.0	2.0	1.0	2.0	2.0	2.0	1.0	2.0	2.0	1.0
9	2.0	2.0	1.0	2.0	2.0	2.0	2.0	2.0	2.0	1.0
10	2.0	1.0	2.0	2.0	2.0	1.0	1.0	1.0	1.0	1.0
11	2.0	1.0	2.0	2.0	1.0	1.0	1.0	1.0	1.0	1.0
12	1.0	1.0	2.0	2.0	1.0	1.0	1.0	1.0	2.0	2.0
13	1.0	1.0	2.0	1.0	1.0	1.0	1.0	1.0	2.0	1.0
14	2.0	2.0	2.0	2.0	2.0	1.0	1.0	1.0	2.0	1.0
15	2.0	2.0	2.0	2.0	2.0	1.0	1.0	2.0	2.0	1.0
16	2.0	2.0	2.0	2.0	2.0	1.0	1.0	2.0	2.0	1.0
17	2.0	2.0	2.0	2.0	2.0	1.0	1.0	1.0	1.0	2.0
18	2.0	2.0	2.0	1.0	2.0	2.0	2.0	1.0	1.0	2.0
19	2.0	1.0	2.0	2.0	2.0	2.0	1.0	1.0	2.0	2.0
20	2.0	1.0	2.0	2.0	2.0	1.0	1.0	1.0	2.0	2.0
21	2.0	2.0	2.0	2.0	1.0	1.0	1.0	1.0	1.0	2.0
22	2.0	2.0	1.0	2.0	2.0	1.0	1.0	2.0	2.0	1.0
23	2.0	2.0	1.0	2.0	2.0	1.0	1.0	2.0	2.0	1.0
24	1.0	1.0	2.0	1.0	1.0	1.0	1.0	1.0	1.0	2.0
25	1.0	2.0	2.0	1.0	1.0	1.0	1.0	1.0	1.0	2.0
26	2.0	2.0	2.0	2.0	1.0	1.0	1.0	1.0	1.0	1.0
27	2.0	2.0	1.0	2.0	2.0	1.0	2.0	2.0	2.0	1.0
28	2.0	2.0	1.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
29	2.0	2.0	2.0	2.0	2.0	1.0	1.0	1.0	1.0	2.0
30	2.0	1.0	2.0	2.0	1.0	1.0	1.0	1.0	1.0	2.0

* 1 – выбирает; 2 – не выбирает.

Задание 4. В исследовании взаимосвязи между семейным и покупательским поведением проводилось измерение следующих утверждений о стиле жизни респондента (1 – не согласен, 7 – согласен) (табл. 9.6).

V1 Я лучше спокойно провел бы вечер дома, чем пошел бы на вечеринку.

V2 Я всегда проверяю цены, даже на дешевый товар.

V3 Магазины меня интересуют больше, чем кинофильмы.

V4 Я не покупаю товары, которые рекламируют на билбордах.

V5 Я – домосед.

V6 Я экономлю, используя купоны для покупки товаров.

V7 Компании зря тратят большие деньги на рекламу.

Проведите кластерный анализ.

Таблица 9.5

Измерение утверждений о стиле жизни респондента

№ п/п	V1	V2	V3	V4	V5	V6	V7
1	6	2	7	6	5	3	5
2	5	7	5	6	6	6	4
3	5	3	4	5	6	6	7
4	3	2	2	5	1	3	2
5	4	2	3	2	2	1	3
6	2	6	2	4	3	7	5
7	1	3	3	6	2	5	7
8	3	5	1	4	2	5	6
9	7	3	6	3	5	2	4
10	6	3	3	4	4	6	5
11	6	6	2	6	4	4	7
12	3	2	2	7	6	1	6
13	5	7	6	2	2	6	1
14	6	3	5	5	7	2	3
15	3	2	4	3	2	6	5
16	2	7	5	1	4	5	2
17	3	2	2	7	2	4	6
18	6	4	5	4	7	3	3
19	7	2	6	2	5	2	1

Окончание табл. 9.5

№ п/п	V1	V2	V3	V4	V5	V6	V7
20	5	6	6	3	4	5	3
21	2	3	3	2	1	2	6
22	3	4	2	1	4	3	6
23	2	6	3	2	1	5	3
24	6	5	7	4	5	7	2
25	7	6	5	4	6	5	3

10. Факторный анализ

Факторный анализ – это класс методов, используемых главным образом, для сокращения количества переменных и их обобщения. Факторный анализ не предполагает разделение переменных на независимые и зависимые.

В практике маркетинговых исследований факторный анализ применяется в следующих ситуациях:

- для сегментирования рынка и выявления переменных с целью группировки потребителей;
- для определения характеристик торговой марки с целью выявления предпочтений потребителей;
- при разработке рекламной стратегии с целью выявления особенностей восприятия потребителем рекламного продукта.

Факторный анализ применяется для выделения из большого массива данных малого числа групп, состоящих из переменных, объединенных общими факторами (рис. 10.1).

В один фактор объединяются переменные, плотно коррелирующие между собой и слабо коррелирующие с переменными, объединяемыми другими факторами. Основной задачей факторного анализа является сокращение числа переменных и упрощение процедуры анализа существующей базы данных [7].

В процессе проведения факторного анализа рассчитываются и анализируются следующие показатели:

- критерий сферичности Бартлетта – статистика, проверяющая гипотезу о том, что переменные в генеральной совокупности не коррелируют между собой, каждая переменная коррелирует сама с собой ($r = 1$), но не взаимосвязана с другими переменными ($r = 0$);
- корреляционная матрица – матрица попарных корреляций r между возможными парами переменных, включенных в анализ;
- критерий адекватности выборки Кайзера–Мейера–Олкина (КМО) – коэффициент для проверки целесообразности выполне-

ния факторного анализа. Высокие значения (от 0,5 до 1) указывают, что факторный анализ целесообразен, малые значения (до 0,5) – не целесообразен.

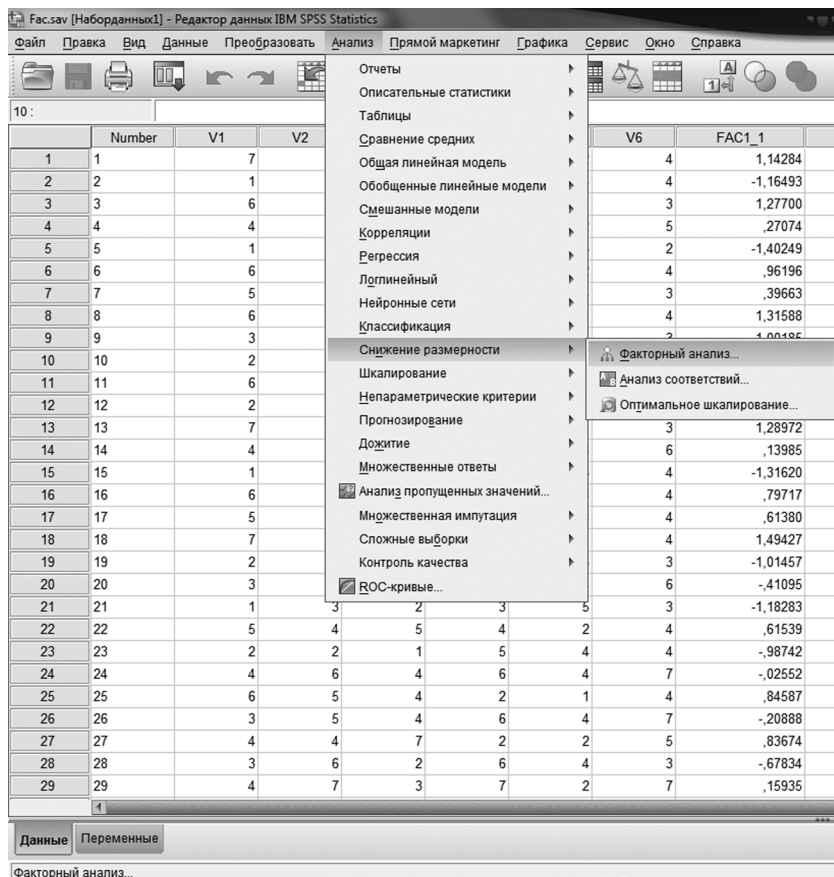


Рис. 10.1. Диалоговое окно факторного анализа

Графическое изображение критерия «каменистой осыпи» – график зависимости собственных значений от количества факторов в порядке их убывания.

Процедура факторного анализа включает следующие этапы:

1. Формулировка проблемы.
2. Проверка возможности проведения факторного анализа.
3. Определение метода факторного анализа.

4. Выбор числа факторов.
5. Вращение факторов.
6. Определение значений факторов.
7. Проведение подгонки выбранной модели.

Пример. Проведем факторный анализ с целью сокращения массива данных, содержащих информацию о мотивах туристов, при выборе места отдыха за городом. Оптимизируем структуру данных, сократив число переменных.

Основные задачи:

- оценить пригодность сходных данных для проведения факторного анализа;
- выявить корреляционные взаимосвязи между переменными исходного массива;
- определить оптимальное число факторов, т. е. групп, на которые может быть разделен существующий массив переменных;
- разделить существующий массив переменных на группы на основании значений коэффициентов корреляции;
- подобрать названия созданным переменным.

Мотивы туристов при выборе места отдыха за городом:

- 1) близость к городу;
- 2) приемлемые цены;
- 3) близость водоема, леса;
- 4) уровень комфорта;
- 5) тишина, уединение;
- 6) хорошее питание;
- 7) наличие развлечений;
- 8) комфорт отдыха с детьми;
- 9) возможность лечения, ухода за здоровьем.

Пошаговая инструкция

Шаг 1. Меню Анализ – Сокращение размерности – Факторный анализ ... Открывается диалоговое окно Факторный анализ.

Шаг 2. Из этого списка переменных выбрать необходимый массив и перенести его в поле Переменные.

В поле Переменная отбора наблюдений можно перенести переменную «пол», факторный анализ будет проводиться по двум

массивам данных – отдельно для мужчин и женщин. В рассматриваемом примере разделения анализируемых данных на подгруппы не производится.

Шаг 3. Диалоговое окно **Описательные статистики** – выбрать **КМО** и критерий **сферичности Бартлетта** для проведения тестов **КМО** и **Бартлетт**, проверяющих пригодность данных для проведения факторного анализа.

Шаг 4. Диалоговое окно **Описательные статистики – Корреляционная матрица – Коэффициенты – Продолжить**.

Шаг 5. Диалоговое окно **Извлечение** – задать условия определения количества факторов. В диалоговом окне **Извлечение** – выбрать метод **Главные компоненты – Матрица корреляций** (рис. 10.2).

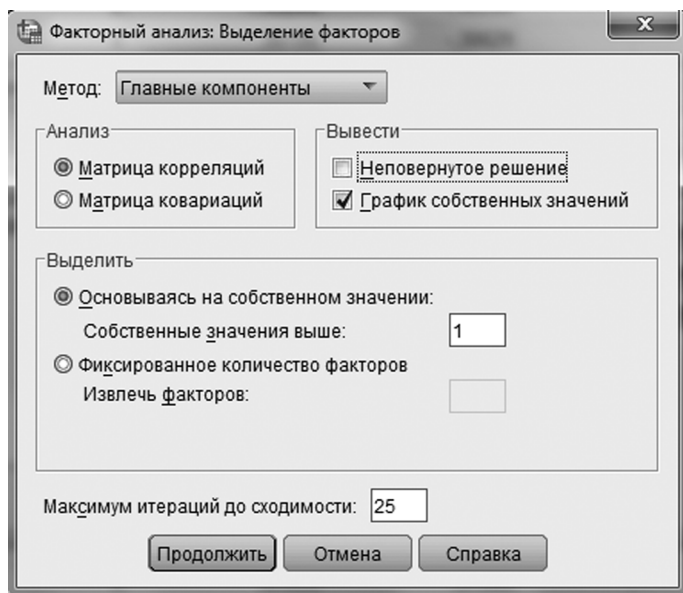


Рис. 10.2. Матрица корреляций

Шаг 6. Задать условие: **собственное значение больше «1»**. На основании выполнения этого условия будет определяться количество факторов.

Шаг 7. Вывести **график собственных значений** – **График собственных значений – Продолжить**.

Шаг 8. Выбор ротации матрицы коэффициентов: в главном диалоговом окне Факторный анализ – диалоговое окно Вращение – метод ротации Варимакс – команда вывода на экран ротированных значений матрицы коэффициентов (Повернутое решение) – Продолжить.

Шаг 9. Создание новых переменных: в диалоговом окне Факторный анализ – диалоговое окно Значения факторов – отметить команду Сохранить как переменные – метод расчета значений новых переменных Регрессионная модель. Вновь созданные факторы будут сохраняться в базе данных как новые переменные.

Шаг 10. ОК.

Интерпретация результатов

1. Результаты теста КМО и Бартлетта (табл. 10.1) позволяют сделать вывод об общей пригодности имеющихся данных для факторного анализа: значение теста КМО = 0,509 > 0,5; значимость теста Бартлетта – 0,000, что говорит о целесообразности факторного анализа в силу коррелированности факторов.

Таблица 10.1

Результаты теста КМО и Бартлетта

Мера выборочной адекватности Кайзера–Мейера–Олкина		,509
Критерий сферичности Бартлетта	Приблиз. хи-квадрат	262,296
	ст. св.	45
	Знч.	,000

Коэффициенты корреляции характеризуют плотность связи между переменными исходного массива (табл. 10.2).

2. Определение оптимального числа компонентов факторной модели (табл. 10.3).

Начальные собственные значения должны быть больше 1.

Оптимальное число факторов – 5. Такая модель сохраняет 80,77% исходной информации, при этом число факторов сокращается в 2 раза.

График показывает зависимость между характеристическими числами и числом компонент факторной модели: с 5-го по 6-й

Таблица 10.2

Корреляционная матрица

	Близость к городу	Близость к водоему, леса	Уровень комфорта	Тишина, уединение	Хорошее питание	Наличие развлечений	Комфорт отдыха с детьми	Возможность лечения, ухода за здоровьем	Организация детского отдыха
Близость к городу	1,000	,009	,519	-,172	,149	,322	,075	-,166	,241
Приемлемые цены	-,060	1,000	-,341	,270	-,247	-,121	,232	,273	-,407
Близость водоема, леса	,009	1,000	-,056	,017	-,083	,107	,114	,027	,030
Уровень комфорта	,519	-,056	1,000	-,113	,384	,070	,008	-,125	,256
Тишина, уединение	-,172	,017	-,113	1,000	,084	-,167	-,060	,989	-,129
Хорошее питание	,149	-,083	,384	,084	1,000	-,422	,191	,036	,237
Наличие развлечений	,322	,107	,070	-,167	-,422	1,000	-,045	-,116	,248
Комфорт отдыха с детьми	,075	,114	,008	-,060	,191	-,045	1,000	-,053	,202
Возможность лечения, ухода за здоровьем	-,166	,027	-,125	,989	,036	-,116	-,053	1,000	-,106
Организация детского отдыха	,241	,030	,256	-,129	,237	,248	,202	-,106	1,000

факторы происходит перелом графика. Это подтверждает, что оптимальное количество факторов – 5.

Таблица 10.3

Полная объясненная дисперсия

Компонента	Начальные собственные значения			Суммы квадратов нагрузок вращения		
	Всего	% дисперсии	Кумулятивный %	Всего	% дисперсии	Кумулятивный %
1	2,570	25,704	25,704	2,076	20,762	20,762
2	1,796	17,956	43,660	1,848	18,476	39,238
3	1,402	14,020	57,680	1,563	15,628	54,867
4	1,230	12,296	69,977	1,342	13,422	68,289
5	1,079	10,794	80,771	1,248	12,482	80,771
6	,870	8,699	89,470			
7	,431	4,310	93,780			
8	,343	3,434	97,214			
9	,270	2,702	99,916			
10	,008	,084	100,000			

Метод выделения: Анализ главных компонент.

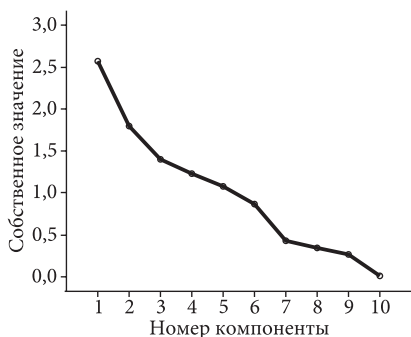


Рис. 10.3. График собственных значений

3. На основании ротированной матрицы компонент в одну группу собираются переменные, имеющие наиболее тесную связь (самое большое значение коэффициента корреляции) с данным компонентом факторной модели. На основе этого производится группировка переменных исходного массива.

Таблица 10.4

Матрица повернутых компонент

	Компонента				
	1	2	3	4	5
Близость к городу	-,088	,852	-,198	-,009	,147
Приемлемые цены	,278	-,190	-,221	-,561	,622
Близость водоема, леса	,074	-,240	-,210	,664	,215
Уровень комфорта	-,062	,793	,241	,093	-,148
Тишина, уединение	,988	-,074	,082	-,040	-9,640E-6
Хорошее питание	,059	,331	,831	,158	,054
Наличие развлечений	-,075	,335	-,797	,261	-,070
Комфорт отдыха с детьми	-,086	,078	,179	,221	,874
Возможность лечения, ухода за здоровьем	,991	-,071	,026	-,021	,006
Организация детского отдыха	-,084	,400	,087	,659	-,004

Метод выделения: Анализ методом главных компонент.

Метод вращения: Варимакс с нормализацией Кайзера.

Таблица 10.5

Матрица преобразования компонент

Компонента	1	2	3	4	5
1	-,690	,611	,026	,366	-,123
2	,552	,488	,648	,191	,033
3	,457	,355	-,750	,321	,001
4	-,103	-,006	,000	,147	,984
5	-,019	,512	-,129	-,839	,126

Метод выделения: Анализ методом главных компонент.

Метод вращения: Варимакс с нормализацией Кайзера.

Выделяем следующие факторы:

- Фактор 1 – тишина и уединение, уход за здоровьем.
- Фактор 2 – близость к городу, уровень комфорта.
- Фактор 3 – хорошее питание, наличие развлечений.
- Фактор 4 – организация детского отдыха, близость водоема.
- Фактор 5 – комфорт отдыха с детьми, приемлемые цены.

4. В базе данных автоматически переносятся новые переменные построенной факторной модели (рис. 10.4). В столбце Метка отображается номер компонента факторной модели.

	Name	Type	Width	Decimals	Label	Values	Missi
129	s_8	Numeric	2	0	Членность населения	1, 500,000 и 6, 98, 99	98, 99
130	s_9	Numeric	2	0	Доход семьи	1, менее 500	98, 99
131	FAC1_1	Numeric	11	5	REGR factor score 1 for	None	None
132	FAC2_1	Numeric	11	5	REGR factor score 2 for	None	None
133	FAC3_1	Numeric	11	5	REGR factor score 3 for	None	None
134	FAC4_1	Numeric	11	5	REGR factor score 4 for	None	None
135							
136							

Рис. 10.4. Фрагмент вкладки «Переменные»

Каждый компонент построенной факторной модели получил свое название, которое следует занести в исходный файл данных SPSS в столбец Метка таблицы «Переменные», компьютер автоматически вычисляет значения новых переменных. При этом балльные оценки данных трансформированы из интервала «от 1 до 5» в интервал «от -2 до 2».

Суть новых переменных сводится к следующему: чем больше отрицательное значение переменной, тем она важнее; чем больше положительное значение переменной, тем она менее важна.

Новые переменные могут быть использованы для дальнейших исследований, например для дальнейшего проведения кластерного анализа.

Задания для практических работ

Задание 1. В исследовании взаимосвязи между семейным и покупательским поведением проводилось измерение следующих утверждений о стиле жизни респондента (1 – не согласен, 7 – согласен) (табл. 10.5).

V1 Я лучше спокойно провел бы вечер дома, чем пошел на вечеринку.

V2 Я всегда проверяю цены, даже на дешевый товар.

V3 Магазины меня интересуют больше, чем кинофильмы.

V4 Я не покупаю товары, которые рекламируют на билбордах.

V5 Я – домосед.

V6 Я экономлю, используя купоны для покупки товаров.

V7 Компании зря тратят большие деньги на рекламу.

Проведите факторный анализ.

Оптимизируйте структуру данных, сократив число переменных:

- оцените пригодность сходных данных для проведения факторного анализа;
- выявите корреляционные взаимосвязи между переменными исходного массива;
- определите оптимальное число факторов, т. е. групп, на которые может быть разделен существующий массив переменных;
- разделите существующий массив переменных на группы на основании значений коэффициентов корреляции;
- подберите названия созданным переменным.

Таблица 10.6

Измерение утверждений о стиле жизни респондента

№ п/п	V1	V2	V3	V4	V5	V6	V7
1	6	2	7	6	5	3	5
2	5	7	5	6	6	6	4
3	5	3	4	5	6	6	7
4	3	2	2	5	1	3	2
5	4	2	3	2	2	1	3
6	2	6	2	4	3	7	5
7	1	3	3	6	2	5	7
8	3	5	1	4	2	5	6
9	7	3	6	3	5	2	4
10	6	3	3	4	4	6	5
11	6	6	2	6	4	4	7
12	3	2	2	7	6	1	6
13	5	7	6	2	2	6	1
14	6	3	5	5	7	2	3
15	3	2	4	3	2	6	5
16	2	7	5	1	4	5	2
17	3	2	2	7	2	4	6
18	6	4	5	4	7	3	3
19	7	2	6	2	5	2	1
20	5	6	6	3	4	5	3
21	2	3	3	2	1	2	6
22	3	4	2	1	4	3	6

№ п/п	V1	V2	V3	V4	V5	V6	V7
23	2	6	3	2	1	5	3
24	6	5	7	4	5	7	2
25	7	6	5	4	6	5	3

Задание 2. Проведите факторный анализ с целью сокращения массива данных, содержащих информацию о мотивах туристов, при выборе места отдыха за городом (табл. 10.7). Оптимизируйте структуру данных, сократив число переменных:

- оцените пригодность сходных данных для проведения факторного анализа;
- выявите корреляционные взаимосвязи между переменными исходного массива;
- определите оптимальное число факторов, т. е. групп, на которые может быть разделен существующий массив переменных;
- разделите существующий массив переменных на группы на основании значений коэффициентов корреляции;
- подберите названия созданным переменным.

На вопрос «Что для вас важно при выборе места отдыха за городом?» респонденты выбирали следующие варианты ответа (1 – очень важно; 2 – важно; 3 – и да, и нет; 4 – не важно; 5 – совсем не важно):

- 1) близость к городу;
- 2) приемлемые цены;
- 3) близость водоема, леса;
- 4) уровень комфорта;
- 5) тишина, уединение;
- 6) хорошее питание;
- 7) наличие развлечений;
- 8) комфорт отдыха с детьми;
- 9) возможность лечения, ухода за здоровьем;
- 10) организация детского отдыха.

Таблица 10.7

Мотивы туристов при выборе места отдыха

№ п/п	1	2	3	4	5	6	7	8	9	10
1	2.0	2.0	2.0	2.0	4.0	4.0	2.0	4.0	5.0	4.0
2	4.0	3.0	3.0	3.0	1.0	1.0	5.0	22.0	2.0	4.0
3	2.0	3.0	4.0	2.0	4.0	2.0	2.0	2.0	4.0	2.0
4	1.0	2.0	5.0	4.0	5.0	2.0	1.0	4.0	4.0	1.0
5	5.0	3.0	2.0	2.0	5.0	3.0	2.0	1.0	5.0	2.0
6	1.0	4.0	3.0	1.0	4.0	2.0	3.0	1.0	55.0	2.0
7	2.0	2.0	3.0	4.0	5.0	5.0	2.0	5.0	5.0	4.0
8	3.0	2.0	1.0	2.0	3.0	1.0	5.0	1.0	3.0	2.0
9	4.0	3.0	2.0	2.0	2.0	1.0	5.0	1.0	2.0	4.0
10	1.0	1.0	3.0	4.0	5.0	4.0	1.0	3.0	5.0	4.0
11	5.0	2.0	3.0	5.0	4.0	4.0	2.0	4.0	5.0	5.0
12	5.0	1.0	4.0	5.0	5.0	5.0	5.0	4.0	5.0	5.0
13	3.0	1.0	4.0	5.0	4.0	5.0	5.0	4.0	5.0	5.0
14	2.0	2.0	5.0	2.0	5.0	2.0	2.0	2.0	4.0	2.0
15	1.0	99.0	3.0	3.0	3.0	2.0	1.0	2.0	5.0	1.0
16	4.0	2.0	3.0	2.0	3.0	2.0	3.0	2.0	4.0	4.0
17	2.0	2.0	1.0	2.0	4.0	4.0	2.0	2.0	5.0	2.0
18	3.0	4.0	4.0	1.0	4.0	4.0	3.0	22.0	5.0	4.0
19	5.0	1.0	2.0	4.0	5.0	5.0	1.0	4.0	5.0	5.0
20	1.0	1.0	1.0	4.0	5.0	5.0	1.0	5.0	5.0	1.0
21	1.0	1.0	3.0	1.0	1.0	1.0	5.0	1.0	2.0	5.0
22	2.0	2.0	1.0	2.0	1.0	2.0	5.0	1.0	2.0	4.0
23	2.0	3.0	1.0	2.0	1.0	1.0	2.0	1.0	2.0	2.0
24	3.0	2.0	3.0	2.0	4.0	2.0	2.0	2.0	3.0	2.0
25	4.0	2.0	5.0	2.0	2.0	2.0	4.0	2.0	4.0	2.0
26	5.0	1.0	5.0	2.0	1.0	1.0	5.0	1.0	2.0	2.0
27	5.0	1.0	4.0	5.0	4.0	3.0	5.0	1.0	5.0	2.0
28	2.0	1.0	3.0	2.0	2.0	2.0	4.0	1.0	2.0	2.0
29	2.0	2.0	5.0	2.0	1.0	1.0	5.0	1.0	2.0	4.0
30	1.0	1.0	2.0	1.0	4.0	4.0	3.0	1.0	4.0	1.0

Задание 3. Проведите факторный анализ с целью сокращения массива данных, содержащих информацию о потребностях туристов организации отдыха за городом.

На вопрос «Хотели бы вы, чтобы ваш досуг за городом был организован?» отдыхающие отвечали следующим образом:

- 1) проведение спортивных состязаний, конкурсов;
- 2) проведение вечерних дискотек;
- 3) пешие прогулки по лесу, тропа здоровья;
- 4) туристические походы;
- 5) катание на лошадях;
- 6) баня, сауна, массаж;
- 7) посещение бассейна;
- 8) катание на катерах, скутерах, катамаранах;
- 9) конкурсы, игры, концерты;
- 10) дневные и вечерние мероприятия для детей, детская комната.

Таблица 10.8

Выбор отдыхающими способа проведения досуга*

№ п/п	1	2	3	4	5	6	7	8	9	10
1	1.0	1.0	2.0	1.0	1.0	1.0	1.0	1.0	2.0	1.0
2	1.0	2.0	2.0	1.0	1.0	1.0	1.0	1.0	2.0	1.0
3	2.0	1.0	2.0	2.0	1.0	1.0	1.0	1.0	2.0	1.0
4	2.0	1.0	2.0	2.0	2.0	1.0	1.0	1.0	1.0	2.0
5	2.0	1.0	2.0	2.0	2.0	1.0	1.0	1.0	1.0	1.0
6	2.0	2.0	2.0	2.0	2.0	1.0	1.0	2.0	2.0	1.0
7	2.0	1.0	2.0	1.0	1.0	1.0	1.0	1.0	2.0	1.0
8	2.0	2.0	1.0	2.0	2.0	2.0	1.0	2.0	2.0	1.0
9	2.0	2.0	1.0	2.0	2.0	2.0	2.0	2.0	2.0	1.0
10	2.0	1.0	2.0	2.0	2.0	1.0	1.0	1.0	1.0	1.0
11	2.0	1.0	2.0	2.0	1.0	1.0	1.0	1.0	1.0	1.0
12	1.0	1.0	2.0	2.0	1.0	1.0	1.0	1.0	2.0	2.0
13	1.0	1.0	2.0	1.0	1.0	1.0	1.0	1.0	2.0	1.0
14	2.0	2.0	2.0	2.0	2.0	1.0	1.0	1.0	2.0	1.0
15	2.0	2.0	2.0	2.0	2.0	1.0	1.0	2.0	2.0	1.0
16	2.0	2.0	2.0	2.0	2.0	1.0	1.0	2.0	2.0	1.0
17	2.0	2.0	2.0	2.0	2.0	1.0	1.0	1.0	1.0	2.0
18	2.0	2.0	2.0	1.0	2.0	2.0	2.0	1.0	1.0	2.0
19	2.0	1.0	2.0	2.0	2.0	2.0	1.0	1.0	2.0	2.0
20	2.0	1.0	2.0	2.0	2.0	1.0	1.0	1.0	2.0	2.0
21	2.0	2.0	2.0	2.0	1.0	1.0	1.0	1.0	1.0	2.0

Окончание табл. 10.8

№ п/п	1	2	3	4	5	6	7	8	9	10
22	2.0	2.0	1.0	2.0	2.0	1.0	1.0	2.0	2.0	1.0
23	2.0	2.0	1.0	2.0	2.0	1.0	1.0	2.0	2.0	1.0
24	1.0	1.0	2.0	1.0	1.0	1.0	1.0	1.0	1.0	2.0
25	1.0	2.0	2.0	1.0	1.0	1.0	1.0	1.0	1.0	2.0
26	2.0	2.0	2.0	2.0	1.0	1.0	1.0	1.0	1.0	1.0
27	2.0	2.0	1.0	2.0	2.0	1.0	2.0	2.0	2.0	1.0
28	2.0	2.0	1.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
29	2.0	2.0	2.0	2.0	2.0	1.0	1.0	1.0	1.0	2.0
30	2.0	1.0	2.0	2.0	1.0	1.0	1.0	1.0	1.0	2.0

Примечание. *1 – выбирает; 2 – не выбирает.

Оптимизируйте структуру данных, сократив число переменных:

- оцените пригодность сходных данных для проведения факторного анализа;
- выявите корреляционные взаимосвязи между переменными исходного массива;
- определите оптимальное число факторов, т. е. групп, на которые может быть разделен существующий массив переменных;
- разделите существующий массив переменных на группы на основании значений коэффициентов корреляции;
- подберите названия созданным переменным.

Библиографический список

1. Наследов А. SPSS 19: профессиональный статистический анализ данных. – СПб.: Питер, 2011. – 400 с.
2. Моосмюллер Г., Ребрик Н.Н. Маркетинговые исследования с SPSS: учеб. пособие. – М.: ИНФРА-М, 2009. – 160 с.
3. Туркин В.С. Методы обработки маркетинговой информации: учеб.-метод. пособие. – М.: МЭСИ, 2004. – 144 с.
4. Просветов Г.И. Маркетинговые исследования: задачи и решения: учеб.-практ. пособие. – М.: Альфа-пресс, 2013. – 240 с.
5. Котлер Ф., Келлер К. Л. Маркетинг менеджмент. – 12-е изд. – СПб.: Питер, 2012. – 816 с.
6. Малхотра Н. К. Маркетинговые исследования и эффективный анализ статистических данных. – Киев: DiaSoft, 2010. – 768 с.
7. Малхотра Н. К. Маркетинговые исследования с помощью SPSS: практ. рук. – М.: Вильямс, 2006. – 1200 с.

Приложение

Результаты опроса отдыхающих базы отдыха «Солнечная»

№ п/п	Пол респондента	Возраст, лет	Семейное положение	Доход, р.	Профессиональная деятельность
1	Мужчины	19,0	Не замужем/ холост	5 000–10 000	Студент
2	Мужчины	59,0	Вдовец/ вдова	5 000–10 000	Госслужащий
3	Женщины	28,0	Замужем/ женат	10 000–20 000	Предприниматель
4	Мужчины	32,0	Замужем/ женат	5 000–10 000	Рабочий
5	Мужчины	29,0	Разведена/ разведен	20 000–30 000	Предприниматель
6	Женщины	27,0	Замужем/ женат	30 000–50 000	Предприниматель
7	Женщины	21,0	Не замужем/ холост	5 000–10 000	Студент
8	Женщины	48,0	Разведена/ разведен	10 000–20 000	Предприниматель
9	Женщины	56,0	Замужем/ женат	5 000–10 000	Рабочий
10	Женщины	19,0	Не замужем/ холост	5 000–10 000	Студент
11	Женщины	21,0	Не замужем/ холост	5 000–10 000	Студент
12	Мужчины	17,0	Не замужем/ холост	Менее 5 000	Школьник
13	Мужчины	18,0	Не замужем/ холост	Менее 5 000	Студент
14	Мужчины	32,0	Замужем/ женат	5 000–10 000	Предприниматель

Продолжение приложения

№ п/п	Пол респондента	Возраст, лет	Семейное положение	Доход, р.	Профессиональная деятельность
15	Женщины		Нет ответа	5 000–10 000	Госслужащий
16	Мужчины	29,0	Не замужем/ холост	10 000–20 000	Предприниматель
17	Мужчины	41,0	Замужем/ женат	20 000–30 000	Предприниматель
18	Мужчины	30,0	Замужем/ женат	Более 50 000	Предприниматель
19	Женщины	18,0	Не замужем/ холост	Менее 5 000	Студент
20	Женщины	20,0	Не замужем/ холост	5 000–10 000	Студент
21	Женщины	61,0	Вдовец/ вдова	5 000–10 000	Пенсионер
22	Женщины	54,0	Замужем/ женат	10 000–20 000	Госслужащий
23	Женщины			5 000–10 000	Нет ответа
24	Мужчины	32,0	Разведена/ разведен	5 000–10 000	Преподаватель
25	Мужчины	36,0	Замужем/ женат	10 000–20000	Госслужащий
26	Женщины	59,0	Замужем/ женат	5 000–10 000	Пенсионер
27	Женщины		Нет ответа	Менее 5 000	Нет ответа
28	Женщины	55,0	Замужем/ женат	5 000–10 000	Госслужащий
29	Мужчины	60,0	Вдовец/ вдова	10 000–20 000	Пенсионер
30	Мужчины	27,0	Замужем/ женат	10 000–20 000	Предприниматель
31	Мужчины	32,0	Не замужем/ холост	30 000–50 000	Предприниматель
32	Мужчины	44,0	Не замужем/ холост	20 000–30 000	Рабочий
33	Мужчины			Нет ответа	Нет ответа
34	Женщины	22,0	Не замужем/ холост	5 000–10 000	Студент

Продолжение приложения

№ п/п	Пол респондента	Возраст, лет	Семейное положение	Доход, р.	Профессиональная деятельность
35	Женщины	25,0	Не замужем/ холост	5 000–10 000	Госслужащий
36	Мужчины	21,0	Не замужем/ холост	5 000–10 000	Студент
37	Женщины	17,0	Не замужем/ холост	Менее 5 000	Школьник
38	Женщины	19,0	Не замужем/ холост	Менее 5 000	Студент
39	Женщины	63,0	Вдовец/ вдова	5 000–10 000	Пенсионер
40	Мужчины	29,0	Замужем/ женат	10 000–20 000	Предприниматель
41	Мужчины	41,0	Замужем/ женат	20 000–30 000	Предприниматель
42	Женщины	27,0	Замужем/ женат	30 000–50 000	Госслужащий
43	Женщины	56,0	Разведена/ разведен	10 000–20 000	Госслужащий
44	Женщины	25,0	Замужем/ женат	5 000–10 000	Рабочий
45	Мужчины	34,0	Замужем/ женат	30 000–50 000	Рабочий
46	Мужчины	35,0	Замужем/ женат	5 000–10 000	Госслужащий
47	Женщины	20,0	Не замужем/ холост	Менее 5 000	Студент
48	Женщины	57,0	Замужем/ женат	10 000–20 000	Госслужащий
49	Женщины	19,0	Не замужем/ холост	5 000–10 000	Студент
50	Мужчины	18,0	Не замужем/ холост	Менее 5 000	Студент
51	Женщины		Нет ответа	5 000–10 000	Госслужащий
52	Мужчины	29,0	Не замужем/ холост	10 000–20 000	Предприниматель

Продолжение приложения

№ п/п	Пол респондента	Возраст, лет	Семейное положение	Доход, р.	Профессиональная деятельность
53	Мужчины	41,0	Замужем/ женат	20 000–30 000	Предприниматель
54	Мужчины	30,0	Замужем/ женат	Более 50 000	Предприниматель
55	Женщины	18,0	Не замужем/ холост	Менее 5 000	Студент
56	Женщины	20,0	Не замужем/ холост	5 000–10 000	Студент
57	Женщины	61,0	Вдовец/ вдова	5 000–10 000	Пенсионер
58	Женщины	54,0	Замужем/ женат	10 000–20 000	Госслужащий
59	Женщины			5 000–10 000	Нет ответа
60	Мужчины	32,0	Разведена/ разведен	5 000–10 000	Преподаватель
61	Мужчины	36,0	Замужем/ женат	10 000–20 000	Госслужащий
62	Женщины	59,0	Замужем/ женат	5 000–10 000	Пенсионер
63	Женщины		Нет ответа	Менее 5 000	Нет ответа
64	Женщины	55,0	Замужем/ женат	5 000–10 000	Госслужащий
65	Мужчины	60,0	Вдовец/ вдова	10 000–20 000	Пенсионер
66	Мужчины	19,0	Не замужем/ холост	5 000–10 000	Студент
67	Мужчины	59,0	Вдовец/ вдова	5 000–10 000	Госслужащий
68	Женщины	28,0	Замужем/ женат	10 000–20 000	Предприниматель
69	Мужчины	32,0	Замужем/ женат	50 00–10 000	Рабочий
70	Мужчины	29,0	Разведена/ разведен	20 000–30 000	Предприниматель
71	Женщины	27,0	Замужем/ женат	30 000–50 000	Предприниматель

Продолжение приложения

№ п/п	Пол респондента	Возраст, лет	Семейное положение	Доход, р.	Профессиональная деятельность
72	Женщины	21,0	Не замужем/ холост	5 000–10 000	Студент
73	Женщины	48,0	Разведена/ разведен	10 000–20 000	Предприниматель
74	Женщины	56,0	Замужем/ женат	5 000–10 000	Рабочий
75	Женщины	19,0	Не замужем/ холост	5 000–10 000	Студент
76	Женщины	21,0	Не замужем/ холост	5 000–10 000	Студент
77	Мужчины	17,0	Не замужем/ холост	Менее 5 000	Школьник
78	Мужчины	18,0	Не замужем/ холост	Менее 5 000	Студент
79	Мужчины	32,0	Замужем/ женат	5 000–10 000	Предприниматель
80	Женщины		Нет ответа	5 000–10 000	Госслужащий
81	Мужчины	21,0	Не замужем/ холост	5 000–10 000	Студент
82	Женщины	17,0	Не замужем/ холост	Менее 5 000	Школьник
83	Женщины	19,0	Не замужем/ холост	Менее 5 000	Студент
84	Женщины	63,0	Вдовец/ вдова	5 000–10 000	Пенсионер
85	Мужчины	29,0	Замужем/ женат	10 000–20 000	Предприниматель
86	Мужчины	41,0	Замужем/ женат	20 000–30 000	Предприниматель
87	Женщины	27,0	Замужем/ женат	30 000–50 000	Госслужащий
88	Женщины	56,0	Разведена/ разведен	10 000–20 000	Госслужащий
89	Женщины	25,0	Замужем/ женат	5 000–10 000	Рабочий

Окончание приложения

№ п/п	Пол респондента	Возраст, лет	Семейное положение	Доход, р.	Профессиональная деятельность
90	Мужчины	34,0	Замужем/ женат	30 000–50 000	Рабочий
91	Мужчины	35,0	Замужем/ женат	5 000–10 000	Госслужащий
92	Женщины	20,0	Не замужем/ холост	Менее 5 000	Студент
93	Женщины	57,0	Замужем/ женат	10 000–20 000	Госслужащий
94	Женщины	19,0	Не замужем/ холост	5 000–10 000	Студент
95	Мужчины	18,0	Не замужем/ холост	Менее 5 000	Студент
96	Женщины		Нет ответа	5 000–10 000	Госслужащий
97	Мужчины	29,0	Не замужем/ холост	10 000–20 000	Предприниматель
98	Мужчины	41,0	Замужем/ женат	20 000–30 000	Предприниматель
99	Мужчины	30,0	Замужем/ женат	Более 50 000	Предприниматель
100	Женщины	18,0	Не замужем/ холост	Менее 5 000	Студент

Оглавление

Введение	3
1. Формирование исходной базы данных в SPSS	5
1.1. Подготовка данных к анализу	5
1.2. Структура редактора данных	7
<i>Задания для практических работ</i>	9
<i>Пошаговая инструкция</i>	13
2. Управление данными	17
<i>Задания для практических работ</i>	18
<i>Пошаговые инструкции к выполнению заданий 2–6</i>	18
3. Предварительный анализ данных	23
3.1. Описательные статистики	23
<i>Пошаговые инструкции к выполнению заданий</i>	24
<i>Интерпретация результатов</i>	25
3.2. Графики	26
<i>Пошаговая инструкция</i>	27
<i>Интерпретация результатов</i>	27
3.3. Корреляции	28
<i>Пошаговая инструкция</i>	29
<i>Интерпретация результатов</i>	30
<i>Задания для практических работ</i>	31
4. Анализ таблиц сопряженности	35
<i>Пошаговая инструкция</i>	36
<i>Интерпретация результатов</i>	36
<i>Задания для практических работ</i>	40
5. Сравнение средних величин	43
Основные термины	43
<i>Пошаговая инструкция</i>	48
<i>Интерпретация результатов Т-теста для независимых выборок</i>	48
<i>Задания для практических работ</i>	52

6. Однофакторный дисперсионный анализ	54
<i>Пошаговая инструкция</i>	56
<i>Интерпретация результатов</i>	57
<i>Задания для практических работ</i>	60
7. Линейный регрессионный анализ	64
7.1. Простая линейная регрессия	66
<i>Пошаговая инструкция</i>	67
<i>Интерпретация результатов</i>	67
7.2. Множественная линейная регрессия	70
<i>Пошаговая инструкция</i>	71
<i>Интерпретация результатов</i>	71
<i>Задания для практических работ</i>	74
8. Дискриминантный анализ	80
<i>Пошаговая инструкция</i>	81
<i>Интерпретация результатов</i>	82
<i>Задания для практических работ</i>	90
9. Кластерный анализ	95
<i>Основные термины, применяемые в кластерном анализе</i>	96
<i>Пошаговая инструкция</i>	99
<i>Интерпретация полученных результатов</i>	103
<i>Задания для практических работ</i>	106
10. Факторный анализ	111
<i>Пошаговая инструкция</i>	113
<i>Интерпретация результатов</i>	115
<i>Задания для практических работ</i>	119
Библиографический список	125
Приложение	126

Учебное издание

Изакова Наталья Борисовна
Журавлева Александра Юрьевна

Маркетинговые исследования с применением SPSS

Учебное пособие

Редактор и корректор
М. В. Баусова

Компьютерная верстка
А. А. Гребеницковой

Издательство Уральского государственного экономического университета
620144, г. Екатеринбург, ул. 8 Марта / Народной Воли, 62 / 45

Поз. 214. Подписано в печать 17.12.2013.
Формат 60×84 / 16. Уч.-изд. л. 4,20. Усл. печ. л. 7,7.
Тираж 80 экз. Заказ 59.

Отпечатано с готового оригинал-макета
в подразделении оперативной полиграфии УрГЭУ