

Министерство образования и науки Российской Федерации
ФГБОУ ВПО Уральский государственный экономический университет

Н.Б.Изакова, А.Ю.Журавлева

МАРКЕТИНГОВЫЕ ИССЛЕДОВАНИЯ С ПРИМЕНЕНИЕМ SPSS

Екатеринбург

2013

Авторы:

Журавлева Александра Юрьевна - ассистент кафедры маркетинга и международного менеджмента - главы 5, 9, практическое задание в главе 8, 10.

Изакова Наталья Борисовна - старший преподаватель кафедры маркетинга и международного менеджмента – введение, главы 1, 2, 3, 4, 6, 7, 8, 10, практические задания в главе 5, 9.

Рецензенты

И.Д.Тургель – д.э.н., профессор, зав.кафедрой экономики и управления Уральского института Российской академии народного хозяйства и государственной службы при Президенте Российской Федерации.

Л.П.Пискунова – к.ф.н., доцент кафедры теории и практики менеджмента Уральского федерального университета имени первого Президента России Б.Н.Ельцина

АННОТАЦИЯ

Назначение данного учебного пособия – помочь активному освоению основных методов проведения маркетинговых исследований с использованием программного комплекса SPSS. Практикум содержит задания для самостоятельной 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, преподавателей высших профессиональных учебных заведений.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	5
1 ФОРМИРОВАНИЕ ИСХОДНОЙ БАЗЫ ДАННЫХ В SPSS	7
1.1 ПОДГОТОВКА ДАННЫХ К АНАЛИЗУ	7
1.2 СТРУКТУРА РЕДАКТОРА ДАННЫХ	9
ЗАДАНИЯ ДЛЯ ПРАКТИЧЕСКИХ РАБОТ	11
2 УПРАВЛЕНИЕ ДАННЫМИ	19
ЗАДАНИЯ ДЛЯ ПРАКТИЧЕСКИХ РАБОТ	20
3 ПРЕДВАРИТЕЛЬНЫЙ АНАЛИЗ ДАННЫХ.....	26
3.1 ОПИСАТЕЛЬНЫЕ СТАТИСТИКИ	26
3.2 ГРАФИКИ	29
3.3 КОРРЕЛЯЦИИ.....	31
ЗАДАНИЯ ДЛЯ ПРАКТИЧЕСКИХ РАБОТ	35
4 АНАЛИЗ ТАБЛИЦ СОПРЯЖЕННОСТИ.....	38
ЗАДАНИЯ ДЛЯ ПРАКТИЧЕСКИХ РАБОТ	43
5 СРАВНЕНИЕ СРЕДНИХ ВЕЛИЧИН	46
ЗАДАНИЯ ДЛЯ ПРАКТИЧЕСКИХ РАБОТ	55
6 ОДНОФАКТОРНЫЙ ДИСПЕРСИОННЫЙ АНАЛИЗ	57
ЗАДАНИЯ ДЛЯ ПРАКТИЧЕСКИХ РАБОТ	63
7 ЛИНЕЙНЫЙ РЕГРЕССИОННЫЙ АНАЛИЗ.....	67
7.1 ПРОСТАЯ ЛИНЕЙНАЯ РЕГРЕССИЯ	69
7.2 МНОЖЕСТВЕННАЯ ЛИНЕЙНАЯ РЕГРЕССИЯ.....	73
ЗАДАНИЯ ДЛЯ ПРАКТИЧЕСКИХ РАБОТ	78
8 ДИСКРИМИНАНТНЫЙ АНАЛИЗ.....	85
ЗАДАНИЯ ДЛЯ ПРАКТИЧЕСКИХ РАБОТ	96
9 КЛАСТЕРНЫЙ АНАЛИЗ	101
ЗАДАНИЯ ДЛЯ ПРАКТИЧЕСКИХ РАБОТ	113
10 ФАКТОРНЫЙ АНАЛИЗ.....	119

ЗАДАНИЯ ДЛЯ ПРАКТИЧЕСКИХ РАБОТ	128
ЛИТЕРАТУРА	134
ПРИЛОЖЕНИЯ.....	135

ВВЕДЕНИЕ

Эффективная маркетинговая деятельность не возможна без объективной маркетинговой информации, основную часть которой дают маркетинговые исследования. Они играют ключевую роль в реализации маркетингового подхода в бизнесе. Ими охватываются факторы микро и макросреды, оценка 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 результатам исследования. Современное маркетинговое исследование требует умелого использования компьютерной технологии, 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 его дальнейшего развития. Для любой компании, приступающей к производству или продаже нового товара, важно знать 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 предоставляет программа *SPSS (Superior Performing Software Systems)* - система статистической обработки информации.

В данном учебном пособии рассмотрены основные методы обработки маркетинговой информации с помощью программы SPSS, изложены

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.

К пособию прилагается электронная база данных результатов опроса 100 респондентов, отдыхающих на базе отдыха для использования при выполнении практических заданий.

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

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, представленных на рисунке 1.1:



Рисунок 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 используются следующие типы шкал измерения переменных:

- Номинальная
- Порядковая (ранговая)
- Интервальная
- Относительная

От выбора шкалы (таблица 1.1) зависит 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 характеристики (вес, длина, объем и пр.) и экономические характеристики (цена, объем продаж, прибыль и пр.).

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. На данном этапе необходимо представить, отредактировать и сохранить полученные данные, так как исчерпывающий анализ требует повторяющейся обработки данных и одновременного применения различных 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

Формирование исходного файла базы данных в *SPSS* для последующего анализа производится в окне редактора данных, который имеет две вкладки: «Переменные» и «Данные». В верхней части окна расположены строка меню и панель инструментов.

Вкладка «Переменные» представляет собой таблицу, и позволяет задать структуру файла данных, описать свойства переменных, определить

их имена, метки и тип. В каждой строке задается переменная (например, вопрос анкеты), а в каждом столбце – свойства данной переменной (ответы на вопрос анкеты).

В столбце «Имя» (*Name*) записывают имя переменной — это может быть номер или часть вопроса в анкете. Например, переменная «пол», «занятость», «марка». Имя переменной должно удовлетворять следующим требованиям: может содержать любые буквы, цифры и некоторые символы: \$, #, длина имени не более 64 знаков, должно начинаться с буквы и не может заканчиваться знаком подчеркивания «_» и точкой. Не допускаются пробелы и буквы других алфавитов.

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.

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. Можно использовать формулировки вопросов, содержащихся в анкете.

В столбце «Значения» (*Values*) отображаются значения меток переменных. Чтобы заполнить данный столбец необходимо присвоить различным ответам респондентов коды и внести их в поле «Значение», а в поле «Метка» — внести формулировки ответов респондентов.

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

В столбце «Столбцы» (*Columns*) таблицы «Переменные» задается ширина столбца, которая позволит полностью видеть имена переменных в таблице редактора данных: «Данные» (*Date View*). По умолчанию ширина столбца задается «8».

Параметр «Выравнивание» (*Alignment*) позволяет управлять расположением данных внутри ячейки. 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*) необходимо выбрать одну из трех шкал: количественную, порядковую или номинальную. По умолчанию задана метрическая шкала (*Scale*).

Создание файла данных в SPSS должно подчиняться следующему требованию: 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?» может быть закодирован следующим образом: «1» - предпочитаю, «0» - не предпочитаю. Следовательно, ответы респондентов так, как показано в таблице 1.2.

Таблица 1.2 – Дихотомическая кодировка данных. Вопрос анкеты «Какую марку одежды Вы предпочитаете?»

Респонденты	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

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:

Утверждение

Согласен

Не согласен

- Я люблю ходить по магазинам
- Я люблю посещать новые места
- Мне нравится пользоваться новыми современными электронными приборами
- Я люблю готовить

5. Пожалуйста, 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 скажите 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 15000 руб.

- от 15000 до 20000 руб.
- от 20000 до 25000 руб.
- от 25000 до 30000 руб.
- более 30000 руб.

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 А. Полная база данных результата опроса прилагается к данному учебному пособию в электронном виде.

Туристы, отдохавшие на базе отдыха «Солнечная» ответили на ряд вопросов анкеты (таблица 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?	
	- Близость к городу	
	- приемлемые цены	
	- близость водоема, леса	
	- уровень комфорта	
	- тишина	
	- вкусная еда	
	- наличие развлечений	
	- комфорт отдыха с детьми	
	- затрудняюсь ответить	
	- другое	
2.	Как часто Вы отдыхаете за городом?	
	- 1 раз в неделю или чаще	
	- 1-2 раза в месяц	
	- 1-2 раза в полгода	
	- 1 раз в год	
	- реже 1 раза в год	
	- затрудняюсь ответить	
3.	Как Вы предпочитаете отдыхать за городом?	
	- самостоятельно, «дикарем»	
	- на базе отдыха	
	- в санатории	
	- другое	
4.	Как Вы добираетесь до места отдыха за городом?	

	- на автомобиле	
	- на поезде	
	- на такси	
	- другое	
5.	Хотели бы Вы, чтобы Ваш досуг за городом был организован?	
6.	Какие развлечения Вам интересны при отдыхе за городом?	
	- дискотека	
	- спортивные мероприятия	
	- водные развлечения: катание на лодках, скутерах...	
	- пешие прогулки в лесу	
	- туристические походы	
	- катание на лошадях	
	- баня, сауна, массаж	
	- бассейн	
	- культмассовые мероприятия	
	- другое	
7.	На сколько дней Вы обычно выезжаете за город?	
8.	Сколько всего Вы готовы потратить на отдых за городом за одни сутки на 1 человека?	
9.	Сколько при этом обычно составляют расходы на питание за одни сутки на 1 человека?	
10.	Сколько при этом обычно составляют расходы на проживание за городом за одни сутки на 1 человека?	
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»?	
	- от знакомых, отдыхавших на базе отдыха	
	- из рекламы в прессе	
	- из рекламы по телевидению	
	- из почтовой рассылки	
	- от представителя базы отдыха	
	- от туристической компании	
	- другое	
13.	Оцените степень своей удовлетворенности местом отдыха в целом	
14.	Ваш пол	
	м ж	
15.	Ваш возраст	
16.	Уровень дохода за месяц	
	менее 5000 руб	
	5000 - 10000 руб	
	10000 - 20000 руб	
	20 000 - 30000 руб	
	30000 - 50000 руб	
	более 50000 руб	
17.	Семейное положение	
	- не замужем/холост	

	- замужем/женат	
	- разведен/разведена	
	- вдова/вдовец	
18.	Укажите, пожалуйста, Вашу занятость (профессия)	
	- студент	
	- школьник	
	- предприниматель	
	- государственный служащий	
	- преподаватель	
	- рабочий	
	- 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. В окне «**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. В поле *Значение* окна «**Значение меток переменных**» 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 типов 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 типов шкал.

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

Шаг 11. Нажмите клавишей мыши на ячейку «**Выравнивание**», в меню выберите один из трех вариантов выравнивания данных: по левому краю, по правому краю, по центру.

Шаг 12. В столбце «Шкала измерения» выберите необходимый тип шкалы, в которой закодирована переменная: метрическую, порядковую или номинальную. Так как переменная «пол» относится к номинальной, ей присвоен определенный код (1 или 2), то выбирается тип шкалы измерения - *номинальная*.

Шаг 13. Заполните данные столбцов «**Выравнивание**», «**Шкала измерения**» для остальных переменных аналогичным образом, согласно выбранным Вами типов шкал.

Шаг 14. Сохраните файл под именем «Солнечная» номер группы и 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. Данные результатов опроса представлены в приложении А.

Шаг 17. Сохраните файл. Выйдите из программы.

Задание 3. Разработать 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-производитель игрушек для детей хотела бы узнать, 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, которые представляют собой результат обработки исходных данных. Например, для маркетингового исследования может представлять интерес средняя цена на товар в регионе, общий объем продаж предприятия по всем товарным группам, ранг марки товара и пр.

В процессе преобразования данных могут быть выполнены следующие операции:

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 для 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

1. Загрузите базы данных файла «Солнечная, группа, Ф.И.О.студента».
2. Из интервальной переменной, содержащей средний доход респондентов, создайте категориальную переменную с тремя значениями: «высокий доход», «средний доход», «низкий доход».
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.
4. Временно разбейте файл с данными на две 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.
5. Проранжируйте отдыхающих по возрасту.
6. Проведите сортировку наблюдений по возрасту и по полу.

Пошаговые инструкции к выполнению заданий

Задание 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 на кнопке **Изменить.**

Шаг 4. Щелкните на 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. Проведите аналогичные действия для значений 3-4 и вновь полученного 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**, затем в исходном диалоговом окне щелкните по кнопке **ОК**.

Шаг 8. После кодировки задайте метки «**высокий доход**», «**средний доход**», «**низкий доход**» всем новым переменным в окне **Редактора данных**.

Задание 3. По аналогии с заданием 2 выполните самостоятельно задание 3.

Задание 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. Затем на калькуляторе выберите кнопку «>=», наберите «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**, и затем по кнопке **ОК**.

После того, как выполнен этот шаг при других вариантах обработки данных будут учитываться только те, которые имеют отношения к переменной студенты. Чтобы вернуть доступ к обработке всех данных, достаточно в окне **Отобразить наблюдения** установить переключатель **Все данные**.

Задание 4.2. По аналогии с заданием 4.1, выполните самостоятельно задание 4.2.

Задание 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 данных действий в конце списка появится новая переменная с именем **Rвозраст**.

Задание 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 №.

Задание 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. Проранжировать отдыхающих по возрасту, провести сортировку наблюдений по возрасту и по полу.

Задание 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 (таблица 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

№	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, процентиля) 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 (таблица 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 точках города

№ п/п	Цена, руб.								
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
3	31.6	23	34.5	43	31.6	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	32.5	31	25.09	51	25.97	71	29.16	91	28.05
12	31.6	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	33.0	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	32.5	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 – Статистики.

Цена на товар А в магазинах города

N	Валидные	100
	Пропущенные	0
Среднее		29,61
Медиана		29,40
Мода		24,04
Стд. отклонение		3,83
Дисперсия		14,7
Асимметрия		,17
Стд. ошибка асимметрии		,24
Эксцесс		-,78
Стд. ошибка эксцесса		,47
Размах		14,7
Минимум		22,9
Максимум		37,7
Процентили	25	27,11
	50	29,4
	75	32,12

а. Имеется несколько мод. Показана наименьшая.

Выводы:

1. На основе таблицы можно сделать вывод, что средняя цена на товар А составляет $29,61 \pm 1,96 \cdot 3,83$ руб. (для 95% 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 руб. – 25%; 29,4 руб. – 50%; 32,12 руб. – 75%.

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

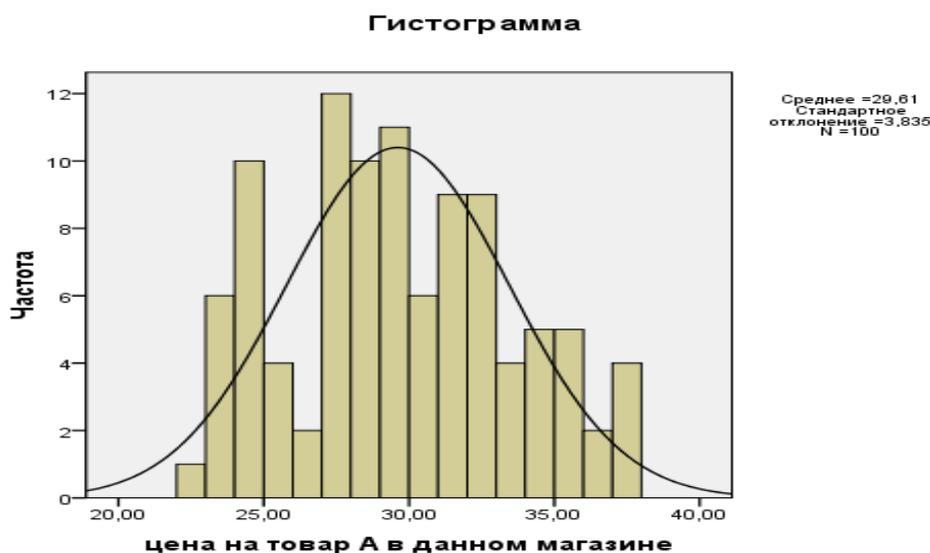


Рисунок 3.1 – Распределение цен на товар А в магазинах города

2. Представленная гистограмма доказывает нормальность распределения, что позволяет в дальнейшем корректно использовать различные виды стат. 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. В случае, 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* дают возможность строить и редактировать различные диаграммы: добавлять переменные, менять категории данных и пр.

Пример. Используя базу данных опроса отдыхающих на базе отдыха «Солнечная» постройте диаграмму уровня удовлетворенности туристов 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** представляют выберите переключатель **№ наблюдения**, в поле **Категориальная ось** выберите – **Довольны ли Вы местом отдыха (Удовлетворенность отдыхом)**, в поле **Строки - Возрастные группы**, в поле **Столбцы** – оставьте пустым.

Шаг 4. Щелкните по кнопке **OK**, чтобы открыть **Окно вывода**.

Интерпретация результатов

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

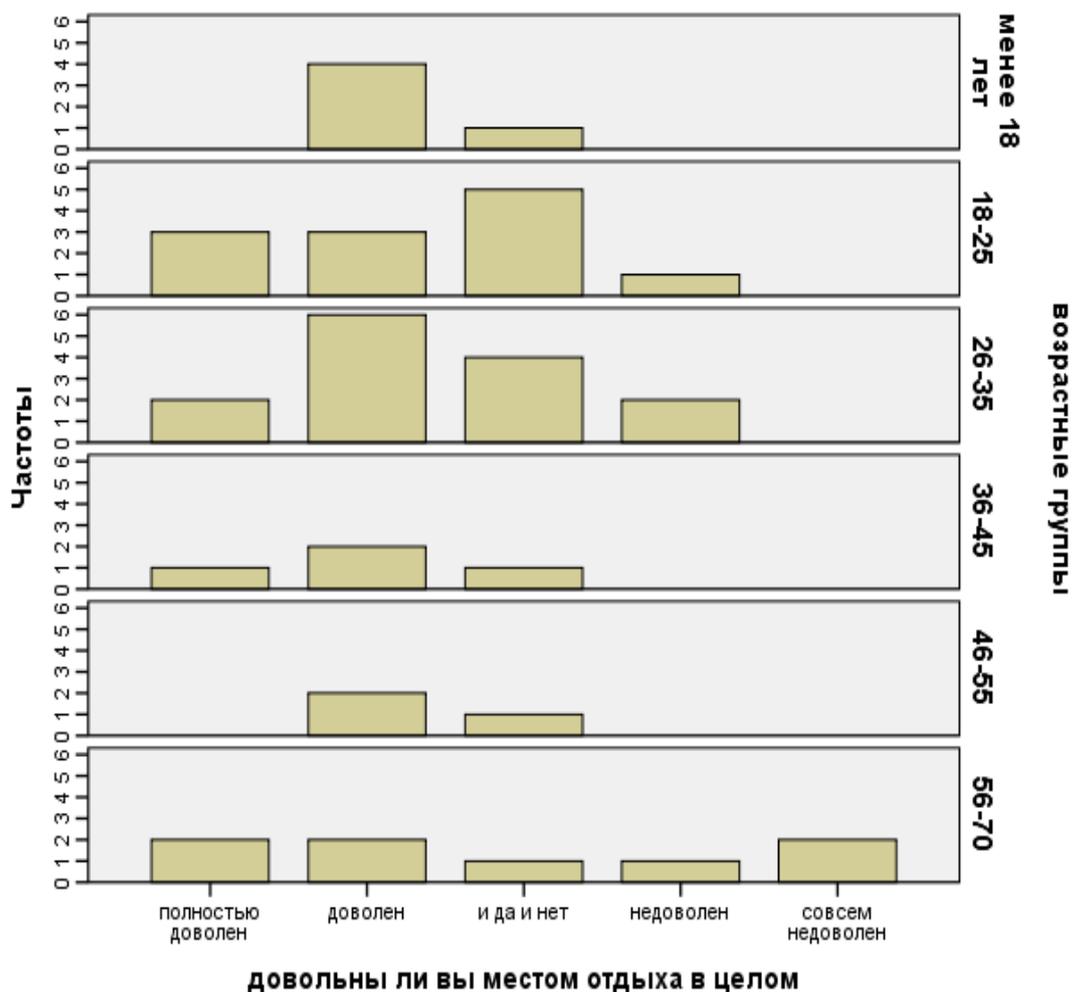


Рисунок 3.2 – Уровень удовлетворенности туристов разных возрастных групп

3.3 КОРРЕЛЯЦИИ

В процессе изучения и анализа маркетинговой информации полезно выявить взаимосвязи между различными переменными, например зависимость объема продаж от цены, зависимость 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 к 0, можно утверждать об отсутствии связи между переменными. Значения коэффициентов близких к 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 как с учетом влияния цены, так и без учета влияния цены.

Таблица 3.3 – Исходная информация

№	Расходы на рекламу, тыс.руб.	Объем продаж, тыс.руб.	Цена, руб.
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

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 следующие таблицы с показателями корреляции.

При вычислении обычного коэффициента корреляции Пирсона (Analyze-Correlate-Bivariate) получили следующий результат, представленный в таблице 3.4 - Корреляции.

Таблица 3.4 - Корреляции

	Расходы на рекламу	на	Объем продаж в тыс.руб.
Расходы на рекламу, тыс.руб.	Корреляция Пирсона	1	,78
	Знч.(2-сторон)		,00
	N	20	20
Объем продаж в тыс.руб.	Корреляция Пирсона	,78	1
	Знч.(2-сторон)	,00	
	N	20	20

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 (таблица 3.5):

Таблица 3.5 – Корреляции. Контрольные переменные

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

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» провести предварительный анализ, включив в него следующие переменные: пол, возраст, доход отдыхающих, общие расходы на отдых, частоту 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 (таблица 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.

Таблица 3.6 - Информации о объемах продаж товара в 100 торговых точках города

№ п/п	Объем продаж, тыс.руб.								
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

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 3. Проанализируйте зависимость между объемами продаж и ценами как с учетом влияния расходов на рекламу, так и без учета (таблица 3.7). Постройте графики двумерного рассеяния для переменных **объем продаж – цена; объем продаж – затраты на рекламу.**

Таблица 3.7 – Исходная информация

№	Расходы на рекламу, тыс.руб.	Объем продаж, тыс.руб.	Цена, руб.
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

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

Результат представить 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 р-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) (таблица 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									
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

Таблица 4.2 – Таблица сопряженности предпочитаемый сок * пол респондента

		пол респондента		Итого
		мужчины	женщины	
предпо читаем ый сок	Частота	13	14	27
	Ожидаемая частота	15	11,9	27
	% по категории переменной предпочитаемый сок	48,1%	51,9%	100,0%

Таблица 4.3 - Критерии хи-квадрат

	Значение	ст.св.	Асимпт. значимость (2-стор.)	Точная значимость (2-стор.)	Точная значимость (1- стор.)
--	----------	--------	---------------------------------	-----------------------------------	------------------------------------

	% по категории переменной пол респондента	46,4%	63,6%	54,0%
	% по таблице (слою)	26,0%	28,0%	54,0%
	Остаток	-2,1	2,1	
В	Частота	15	8	23
	Ожидаемая частота	12,9	10	23
	% по категории переменной предпочитаемый сок	65,2%	34,8%	100,0%
	% по категории переменной пол респондента	53,6%	36,4%	46,0%
	% по таблице (слою)	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%

Хи-квадрат Пирсона	1,469 ^a	1	,226		
Поправка на непрерывность ^b	,858	1	,354		
Отношение правдоподобия	1,480	1	,224		
Точный критерий Фишера				,264	,177
Линейно-линейная связь	1,439	1	,230		
Кол-во валидных наблюдений ^b	50				

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

b. Вычисляется только для таблицы 2x2.

Таблица 4.4 - Направленные меры

		Значение	Асимпт. станд ндртная ошибка ^a	Прибл. T ^b	Прибл. значимость
Номинал Лямбда бная по номинал бной	Симметричная	,06	,19	,33	,73
	Зависимая предпочитаемый сок	,087	,220	,379	,705
	Зависимая пол респондента	,045	,231	,193	,847
Тау Гудмена и Краскала	Зависимая предпочитаемый сок	,029	,048		,230 ^c
	Зависимая пол респондента	,029	,048		,230 ^c

a. Не подразумевающая истинность нулевой гипотезы.

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

- также показатель Хи-квадрат Пирсона (таблица 4.4) имеет малое значение 1,469, а значимость существенно превышает 0,05 (0,226), что также подтверждает отсутствие связи между полом и выбором сока.

- коэффициенты Лямбда и Тау Гудмена и Краскала (таблица 4.5) очень малы, что также говорит об отсутствии связи;

- величины коэффициентов Фи и V Крамера (таблица 4.3) также 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 ^a	Прибл. T ^b	Прибл. значимость
Номинальная по номинальной	Фи	-,17			,22
	V Крамера	,17			,22
Интервальная по интервальной	R Пирсона	-,17	,139	-1,20	,23
Порядковая по порядковой	Корреляция Спирмена.	-,17	,139	-1,20	,23
Кол-во валидных наблюдений		50			

a. Не подразумевая истинность нулевой гипотезы.

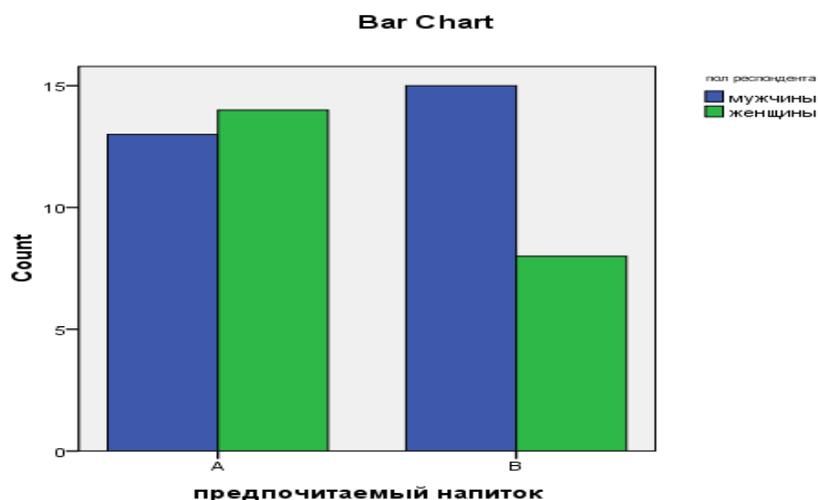


Рисунок 4.1 – График предпочитаемых напитков

Таким образом, на основе проведенного анализа можно сделать вывод о том, 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:

- 1.1. - полом респондента и выбором способа отдыха за городом
- 1.2.– полом респондента и удовлетворенностью отдыхом
- 1.3.– уровнем дохода и выбором способа отдыха за городом
- 1.4.– уровнем дохода и удовлетворенностью отдыхом

Задание 2. Постройте таблицу сопряженности для данных, представленных в таблице 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

№	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 таблице 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 Интернет	№ п/п	Уровень дохода	Частота покупок в Интернет	№ п/п	Уровень дохода	Частота покупок в Интернет	№ п/п	Уровень дохода	Частота покупок в 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
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 – применяются для переменных, измеренных с помощью интервальных шкал.

Непараметрические методы проверки гипотез – применяются для переменных, измеренных с помощью номинальных или порядковых шкал.

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

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

Дисперсия – это 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, с черточкой сверху – среднее значение 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, с помощью которого проверяется гипотеза о равенстве дисперсий двух разных совокупностей. F-критерий это отношение двух выборочных дисперсий.

T-критерий – Одномерный метод проверки гипотез, использующий T-распределение. T-критерий применяется для небольших выборок, когда стандартное отклонение не известно.

Метод T-теста используется для проверки достоверности различия двух выборок по количественной переменной. Нулевая гипотеза формулируется следующим образом: «Взаимосвязи между исследуемыми переменными нет». Например, формулируя исходную (нулевую) гипотезу предполагаем равенство среднего чека покупателей разного возраста, это говорит о том, что люди разного возраста в среднем тратят одинаковую сумму на покупки. В результате проведенного исследования данная гипотезе либо подтверждается, либо опровергается.

SPSS предлагает сравнивать средние величины при помощи нескольких методов (рисунок 5.1).



Рисунок 5.1 – Методы сравнения средних величин

Результат сравнения средних значений с применением Т-теста оценивается по уровню значимости (*p-уровень*). Значимость является мерой достоверности вычисленных результатов. Если «Значимость» не превышает 0,05, это означает, что нулевая (исходная) гипотеза может быть отклонена с вероятностью ошибки меньше 5%, т.е. ее можно отклонить и, следовательно, она неверна.

При использовании Т-теста устанавливается доверительный интервал равный 95%.

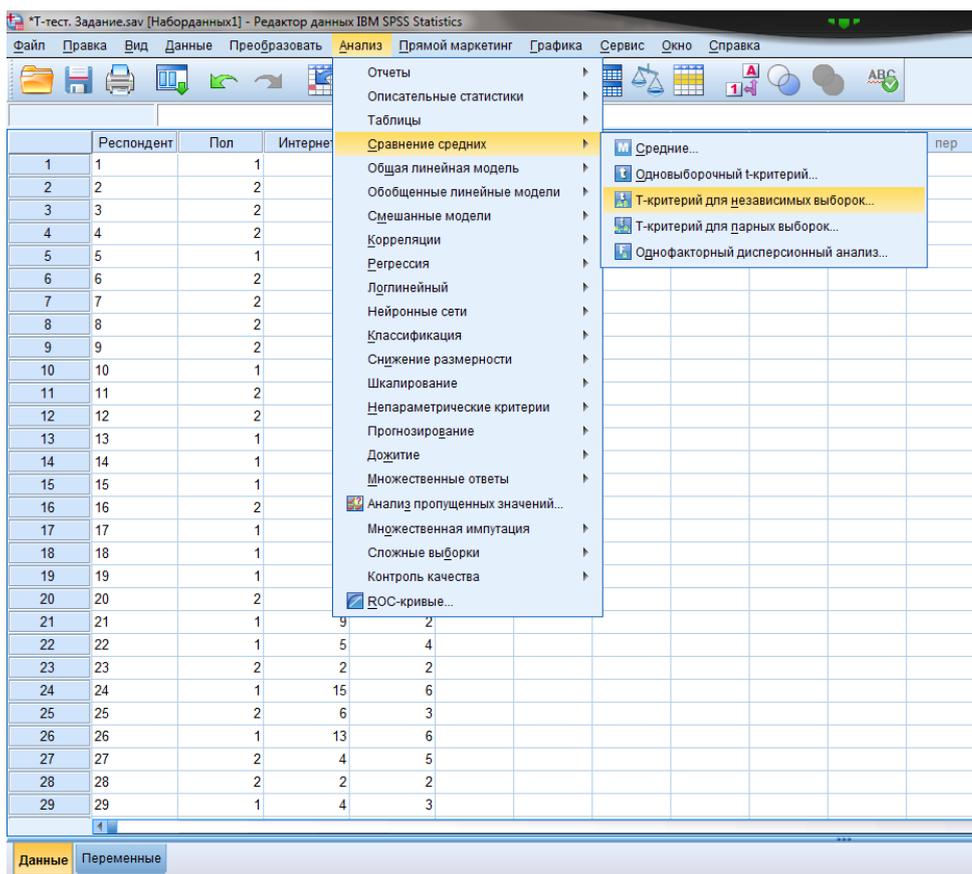


Рисунок 5.2 – Сравнение средних

При запуске процедуры «Сравнение средних» (рисунок 5.2) программа определяет средние показатели выбранных для анализа зависимых переменных в разных группах при помощи Т-критерия.

Одновыборочный t-критерий используется для сравнения средней величины изучаемого признака в выборке с некоторой эталонной величиной. Например, 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.

Т-критерий для независимых выборок (рисунок 5.1) используется для сравнения средних значений двух выборок. Для данных выборок должны быть определены значения одной и той же переменной. Например:

удовлетворенность покупкой мужчин и женщин, семейных и несемейных, предпочитающих или не предпочитающих марку «АК» и т.д.

T-критерий для парных выборок (рисунок 5.1) позволяет сравнить средние значения двух измерений одного признака для одной и той же выборки. Например, результаты опроса потребителей в начале и в конце месяца.

В случае, когда необходимо сравнивать величины более чем в двух группах, применяется *однофакторный дисперсионный анализ* (рисунок 5.1).

Пример. Предположим, что было проведено маркетинговое исследование «Использование респондентами Интернета для личных целей». В таблице 1 представлены данные о 30 респондентах, включающие пол (1-мужчина, 2-женщина), использование Интернета (в часах в неделю).

Таблица 5.1 – Результаты маркетингового исследования

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 мужчины более интенсивно используют Интернет, чем женщины. С этой целью проведем Т-тест для двух независимых выборок.

Т-тест позволяет проверить равенство средних значений тестируемого показателя в двух группах. 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». В окне «Т-критерий для независимых выборок» выбираем метку тестируемой переменной (в рассматриваемом примере это «Использование 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. Кодуем группирующую переменную, создаем две группы: мужчины – присваиваем код «1», женщины – код «2».

Метрические переменные, например «возраст», выбираются при помощи порогового значения, например, «Пороговое значение» – возраст 35 лет.

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» .

Интерпретация результатов

Групповые статистики

	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

		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
Использование Internet	Предполагается равенство дисперсий	15,507	,000	4,492	28	,000	5,467	1,217	2,974	7,960
	Равенство дисперсий не предполагается			4,492	18,014	,000	5,467	1,217	2,910	8,024

Рисунок 5.2 – Интерпретация результатов для независимых выборок

Верность выдвинутой гипотезы можно проверить с помощью величины «Значимость» по критерию Ливиня (рисунок 5.2), которая равна 0,000 (F-критерий имеет вероятность меньше 0,05). Следовательно, гипотезу о равенстве дисперсий отклоняем с вероятностью ошибки 0%, что ниже порогового значения 5%. Нулевая гипотеза (H₀) отклоняется. Следовательно, в данном случае используется t-критерий, соответствующий утверждению «Равенство дисперсий не предполагается». На рисунке 5.2 можно видеть, что t равно 4,492, степеней свободы - 18,014, двусторонняя

значимость равна 0,000, которое меньше допустимого уровня, равного 0,05. Следовательно, гипотезу о том, что мужчины и женщины одинаково часто пользуются Интернет отклоняем. Поскольку среднее значение частоты использования Internet мужчинами равно 9,333, а женщинами – 3,867, то мужчины используют Internet значительно больше по сравнению с женщинами.

Обратный пример.

Нулевая гипотеза – дисперсии распределения удовлетворенности мужчин и женщин равны (сравниваемые средние величины 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	48	2,34	,85	,178
	женщины	42	2,87	1,119	,225

В таблице 5.2 выведены рассчитанные статистические показатели в исследуемых группах, которые представили 45 мужчин и 43 женщины. В среднем удовлетворенность местом отдыха у мужчин равна 2,34, а у женщин — 2,87.

Таблица 5.3 – Критерии для независимых выборок

Критерий равенства дисперсий Ливиня		t-критерий равенства средних					
F	Знч.	t	ст.св.	Значимость (2-сторо)	Разность средних	Стд. ошибка ка разно	95% доверительный интервал разности средних

						нная)		сти	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

С помощью Т-теста определим с точки зрения статистики значимость разницы между уровнем удовлетворенности местом отдыха у мужчин и женщин.

При проведении теста Ливина проверяется следующая гипотеза: «Дисперсии распределения тестируемой величины в разных группах равны» [1]. 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 $> 0,05$. Таким образом, гипотеза о равенстве дисперсий может быть отклонена с вероятностью ошибки 15,8%, что значительно больше допустимых 5%. Гипотеза не отклоняется, она верна – дисперсии равны.

Выбираем строку равенства дисперсий и проверяем верность гипотезы «Средние величины в двух группах равны». Показатель Значимость (2-сторонняя) составляет 0,192, что означает, что гипотеза о том, что мужчины и женщины одинаково удовлетворены местом отдыха, может быть отклонена с вероятностью ошибки 19,2%, а это выше

допустимого уровня в 5%. Следовательно, исходную гипотезу не отклоняем, сравниваемые величины с точки зрения статистики равны, и разница между средними уровнями удовлетворенности отдыхом у мужчин и женщин не является статистически значимой.

Таким образом, в результате проведенного Т-теста доказано, что удовлетворенность местом отдыха не зависит от пола туриста.

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» выявить, как соотносится цена в 1435 руб. на одного человека двухместного стандартного номера с 3-х разовым 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» выявить, как соотносится стоимость питания 380 рублей 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» выявить, влияет ли возраст туриста на его удовлетворенность местом отдыха? (Возрастную границу сравнения определите самостоятельно)

2.2 Используя базу данных опроса отдыхающих базы отдыха «Солнечная» выявить, влияет ли продолжительность отдыха на удовлетворенность туриста местом отдыха?

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

Задание 3. Продуктовый магазин провел исследование о частоте покупок мужчин и женщин в течении 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 группы потребителей по предпочтениям 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* (Рисунок 6.1).

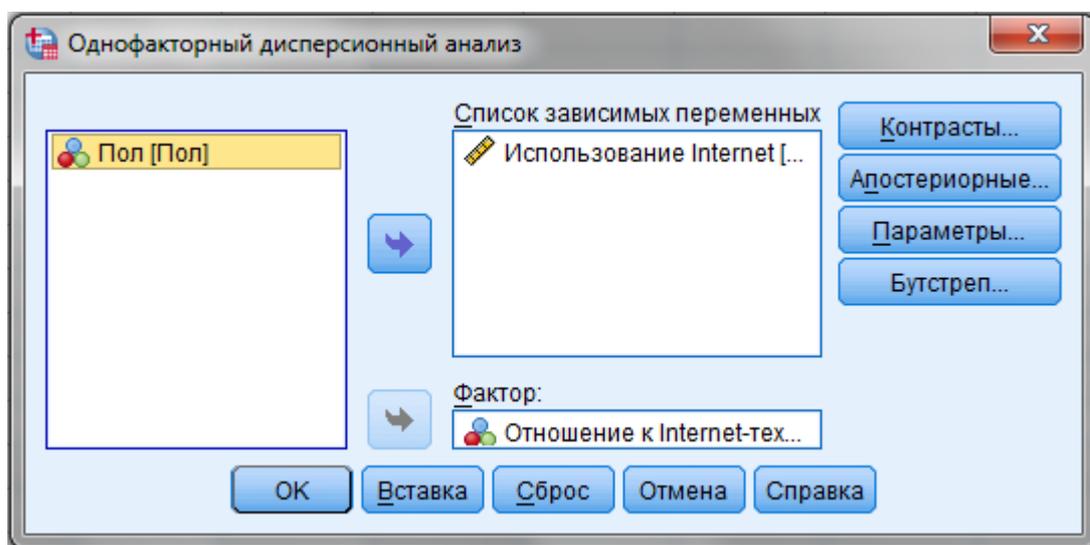


Рисунок 6.1 – Диалоговое окно Однофакторного дисперсионного анализа

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 нет связи между выбранными переменными. В результате анализа нулевая гипотеза должна быть подтверждена или опровергнута.

3. В ходе проведения 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, выявление категорий ее обуславливающих.

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 1 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 1 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 (Рисунок 6.2)

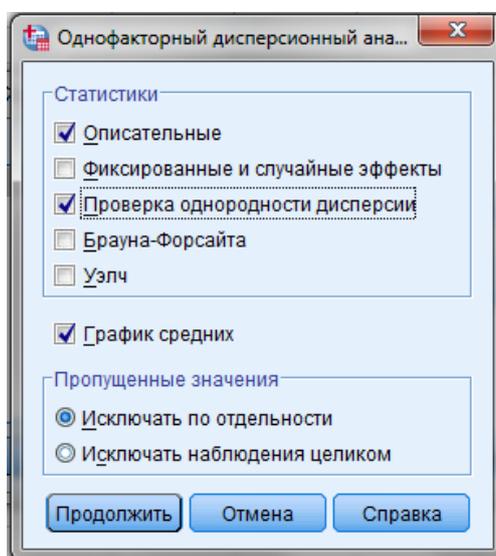


Рисунок 6.2 – Диалоговое окно «Описательные статистики»

Шаг 5. Кнопка «Продолжить»

Шаг 6. Кнопка «Апостериорные»: выбрать Шефе и T2 Тамхейна (Рисунок 6.3)

Шаг 7. Кнопка «Продолжить» и ОК.

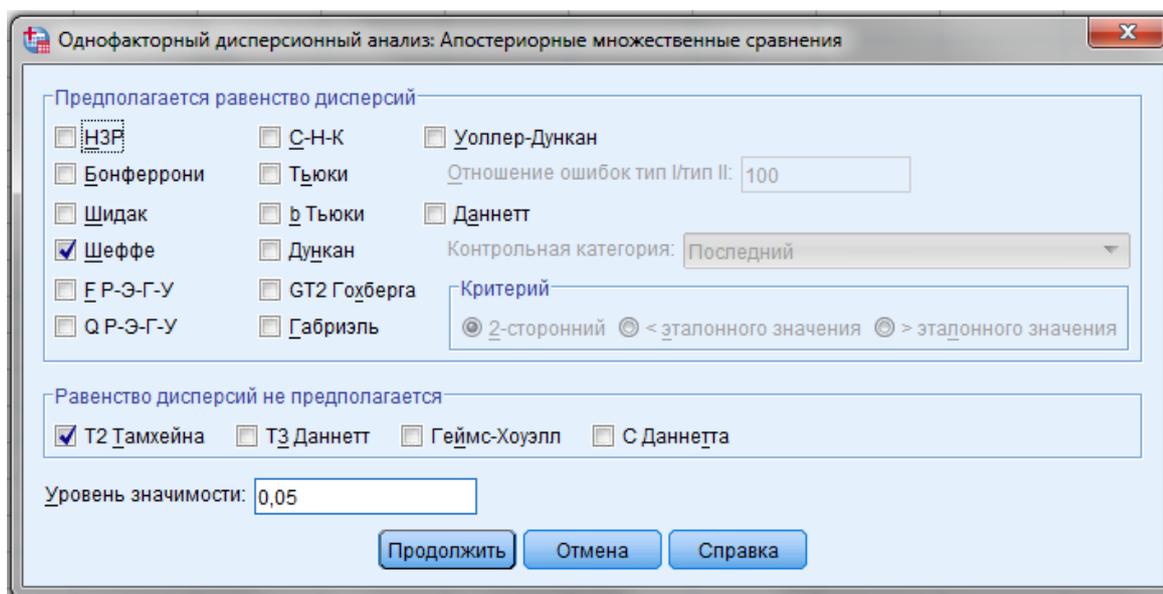


Рисунок 6.3 – Апостериорные множественные сравнения

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 таблица 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% доверительный интервал для среднего		Минимум	Максимум
					Нижняя граница	Верхняя граница		
низкий	13	861,54	221,880	61,538	727,46	995,62	500	1000
средний	25	1540,00	379,693	75,939	1383,27	1696,73	1000	2000
высокий	7	1928,57	345,033	130,410	1609,47	2247,67	1500	2500
Итого	45	1404,44	499,525	74,465	1254,37	1554,52	500	2500

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». 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, гипотеза может быть отклонена, то есть 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 из групп 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.

Таблица 6.3 - Однофакторный дисперсионный анализ: расходы_сумма

	Сумма квадратов	Ст. св.	Средний квадрат	F	Знч.
Между группами	6214056,16	2	3107028,08	27,38	,000
Внутри групп	4765054,94	42	113453,68		
Итого	1,098E7	44			

Проверяем верность исходной нулевой гипотезы: туристы с разным уровнем дохода готовы потратить в среднем одинаковую сумму на 1

человека за 1 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), то есть 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 1 сутки отдыха за городом (существует зависимость между доходом отдыхающих и размером суммы, которую они готовы потратить на отдых за городом).

4. Для того, чтобы получить более точные результаты определим группы, в которых 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 таблицы 6.4 разность сумм, которую готовы потратить на отдых туристы, значительна для каждой из групп.

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

Таблица 6.4 - Множественные сравнения. Зависимая переменная: расходы_сумма

	(I) уровень дохода	(J) уровень дохода	Разность средних (I-J)	Стд. ошибка	Знч.	95% 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 разности сумм, которую туристы с разным уровнем дохода готовы потратить на отдых за городом можно также увидеть на графике (рисунок 6.1).

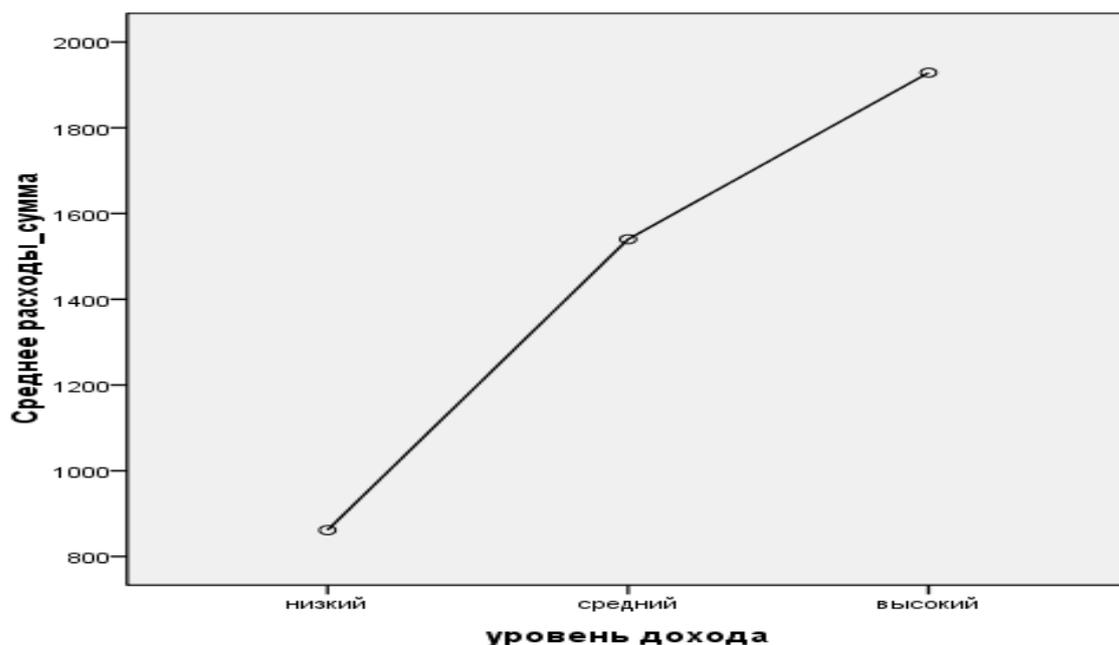


Рисунок 6.1 – Зависимость средних расходов на отдых и уровнем дохода респондента

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 1 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 1 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 1 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 коэффициентов 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
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. Проведите исследование и выясните, есть ли зависимость между отношением респондента к месту жительства от длительности его проживания в этом городе и влияния погодных условий. Отношение выражают (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).

Таблица 6.6 – Данные выборки

Номер	Отношение к месту жительства	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
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), отражающий данную зависимость и позволяющий 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 расходов на продвижение? И какова форма этой зависимости?

2. Какую долю в изменении объема продаж можно объяснить расходами на продвижение, различием цен и 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 регрессионный анализ могут быть включены только количественные переменные. Если необходимо использовать номинальные переменные, их необходимо представить как дихотомические (см.1.2 «Структура редактора данных»).

Для описания и выявления соответствия между зависимой переменной и одним независимым параметром, влияющим на наблюдаемую переменную, используется *уравнение простой линейной регрессии*:

$$y = a - bx, \text{ где}$$

y — зависимая переменная;

x — независимая переменная;

a — свободный член (константа);

b — коэффициент при независимой переменной или коэффициент регрессии [2].

Для описания и выявления соответствия между зависимой переменной и несколькими независимыми параметрами влияющими на наблюдаемую переменную, используется *уравнение множественной линейной регрессии*:

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

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

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

К данным показателям относятся:

- *Коэффициент корреляции (R)* — отражает связь между значениями зависимой и независимой переменных. Значения коэффициента корреляции могут изменяться в интервале от нуля до единицы. При приближении коэффициента к единице, взаимосвязь усиливается. Если R превышает 0,5, то говорят о наличии связи между переменными, и наоборот.

- *Коэффициент R -квадрат (R Square)* — коэффициент детерминации равен той доле результатов наблюдений, которая объясняется уравнением регрессии. Его значения также изменяются в интервале от нуля до единицы. Если R -квадрат равен 0,5, то только для 50% всех наблюдений справедлива построенная регрессионная модель. Для значимых результатов исследования значение коэффициента R -квадрат должно превышать 0,5.

- *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. Результат $p > 0,05$ говорит о том, что связь между переменными слабая или не обнаружена.

- Точки не находятся точно на построенной прямой. Помимо x на y оказывают влияние и другие факторы. Тест Дарбина—Уотсона (*Durbin—Watson*) позволяет сказать, насколько сильно влияние неучтенных факторов, действительно ли модель линейна. Отклонения от построенной прямой должны появляться случайно, между ними не должно быть систематической связи. Данный факт проверяется с помощью расчета коэффициента Дарбина—Уотсона. Значения коэффициента меняются в интервале от 0 до 4. При значении коэффициента от 0 до 2 можно утверждать, что автокорреляция отсутствует.

7.1 ПРОСТАЯ ЛИНЕЙНАЯ РЕГРЕССИЯ

Как уже отмечалось, простая линейная регрессия служит для описания и выявления соответствия между зависимой переменной и одним независимым параметром, влияющим на наблюдаемую переменную

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» выявить, в какой зависимости находятся сумма расходов на отдых в целом и сумма расходов на питание. Сделать прогноз изменения общих расходов 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: таблица 7.1 «Сводка для модели», таблица 7.2 «Дисперсионный анализ» и таблица 7.3 «Коэффициенты».

Таблица 7.1 - Сводка для модели

Модель	R	R квадрат	Скорректированный R квадрат	Стд. ошибка оценки	Дурбин-Уотсон
1	,695 ^a	,48	,471	397,433	1,246

a. Предикторы: (константа) расходы на питание

b. Зависимая переменная: общие расходы на отдых

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.

Значение коэффициента детерминации (R квадрат) равно 0,48, что говорит о том, что построенная модель верна только для 48% случаев, в результате которых увеличение затрат на питание влечет за собой увеличение общей суммы затрат на отдых. Данный факт руководству базы отдыха нужно учитывать, планируя расходов отдыхающих.

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

Таблица 7.2 - Дисперсионный анализ

Модель	Сумма квадратов	ст.св.	Средний квадрат	F	Знч.
1 Регрессия	6336018,018	1	6336018,018	40,113	,000 ^a
Остаток	6791981,9	43	157953,06		
Итого	1,313E7	44			

а. Предикторы: (константа) расходы на питание

б. Зависимая переменная: общие расходы на отдых

На основании результатов таблицы «Дисперсионный анализ» показатель Значимость равен 0,000 (меньше 0,5). Таким образом, можно утверждать, что построенная регрессионная модель верна для всех отдыхающих на базе отдыха Солнечная.

Результаты регрессионного анализа:

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

а. Зависимая переменная: общие расходы на отдых

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 руб.

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

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

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

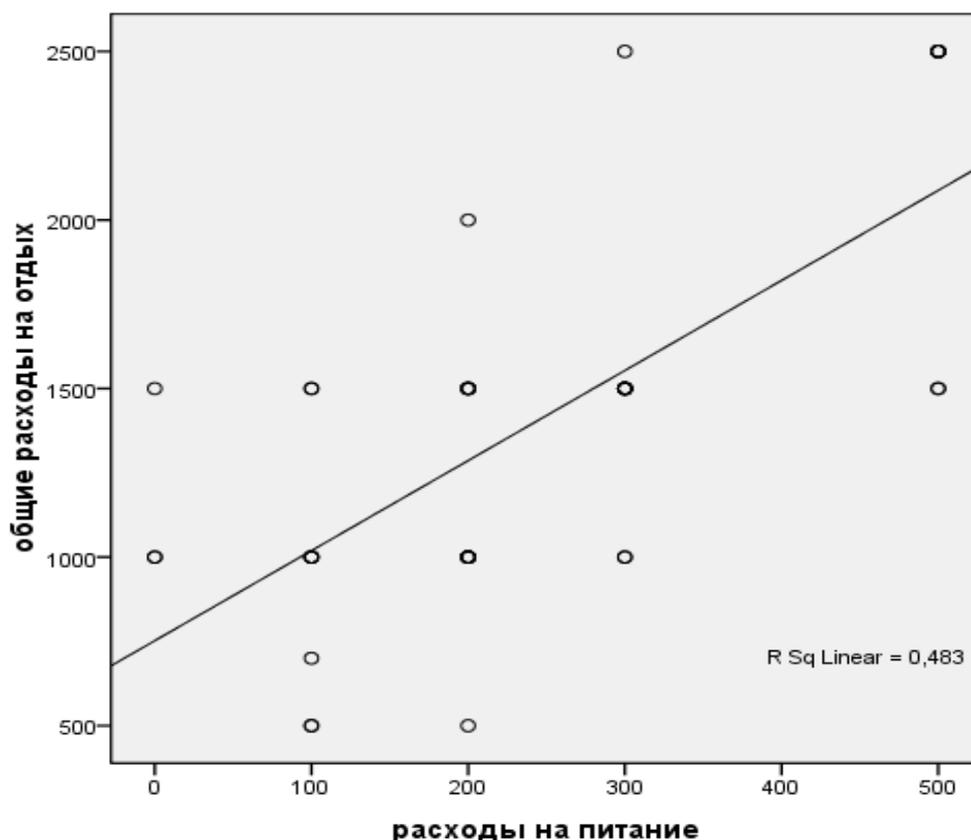


Рисунок 7.1 – Диаграмма рассеяния

5. Построенное уравнение верно лишь для 48,3% случаев, что необходимо учитывать при использовании данной регрессионной модели.

7.2 МНОЖЕСТВЕННАЯ ЛИНЕЙНАЯ РЕГРЕССИЯ

Множественная регрессия является расширением простой линейной 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, то есть существования причинно-следственной связи (корреляции) между независимыми переменными.

При построении множественной регрессионной модели необходимо проверить отсутствие или наличие корреляции между выбранными независимыми переменными.

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 необходимо проверить на корректность и наличие ошибок.
- Необходимо проверить распределение на нормальность (показатели асимметрии и эксцесса не превосходят 1).
- Запрет на использование зависимых переменных, корреляции между которыми близки к 1 (-1).

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

- расходы на питание и 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. Сделать вывод о вероятности возникновения нежелательного эффекта мультиколлинеарности.

6. Сделать 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 информацию, на основе которой можно выявить переменную, оказывающую максимальное влияние на переменную «Общие расходы на отдых».

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

Таблица 7.4 - Описательные статистики

	Среднее	Стд. Отклонение	N
общие расходы на отдых	194,44	107,694	90

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	,695 ^a	,48	,471	397,433	1,246

Коэффициент детерминации (таблица 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.

Значение коэффициента детерминации (R квадрат) равно 0,48, что говорит о том, что построенная модель верна только для 48% случаев для обработанной выборки ответов отдыхающих о том, как они тратят деньги на отдыхе 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	Знч.

		В	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

а. Зависимая переменная: общие расходы на отдых

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 (таблица 7.7) составляем уравнение множественной регрессионной модели:

$$Y = 737,26 + 2,433X_1 + 0,371X_2,$$

где Y - общие расходы туристов на отдых за городом,

X₁ – расходы на питание,

X₂ – расходы на развлечения.

3. Исходя из анализа коэффициентов В (таблица 7.7), можно сделать вывод, что расходы на питание значительно существеннее влияют на общую сумму расходов на отдых, чем расходы на развлечения: В на питание = 2,433, В на развлечения = 0,371.

Также, постоянный показатель «константа» = 737,26, это существенная величина, что говорит о том, что 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): 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 $Z_{\text{значимость}} = 0,000$ (таблица 7.7), что меньше 0,5. Следовательно, мы можем использовать построенную регрессионную модель для всех отдыхающих на базе отдыха «Солнечная». Модель можно использовать руководству базы отдыха «Солнечная», планируя вопросы ценообразования. 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» выявить, в какой зависимости отдельно для мужчин и женщин находятся суммы расходов на отдых в целом и суммы расходов на питание. Сделать прогноз изменения общих расходов 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» выявить для мужчин и женщин зависимость суммы расходов на отдых в целом и суммы расходов на развлечения. Сделать прогноз изменения общих расходов 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» выявить, в какой зависимости между собой находятся общие расходы в день на отдых за городом для мужчин и женщин, и следующие статьи расходов:

- расходы на питание и 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. Сделать вывод о вероятности возникновения нежелательного эффекта мультиколлинеарности.

6. Сделать вывод о вероятности ошибки, при использовании данной регрессионной модели.

Задание 4. Используя результаты оценки потребителями 15 магазинов (таблица 7.9) с точки зрения уровня обслуживания и приемлемости цен, проведите регрессионный анализ и выявите зависимость в предпочтениях от приемлемости цен и уровня обслуживания. При оценке магазинов использовалась 10-балльная шкала. 10 – соответствует максимальной оценке, 1 – минимальной.

Таблица 7.9 – Оценки потребителей

№	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. Сделать вывод о вероятности возникновения нежелательного эффекта мультиколлинеарности.

6. Сделать вывод о вероятности ошибки, при использовании данной регрессионной модели.

Задание 5. Имеются данные о объемах продаж, расходах на рекламу и ценах торгового предприятия. Оцените влияние цен и расходов на рекламу на изменение объемов продаж (таблица 7.10). Постройте уравнение регрессии.

Таблица 7.10 – Исходная информация

№	Расходы на рекламу, тыс.руб.	Объем продаж, тыс.руб.	Цена, руб.
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 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. Сделать вывод о вероятности возникновения нежелательного эффекта мультиколлинеарности.

6. Сделать вывод о вероятности ошибки, при использовании данной регрессионной модели.

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

Таблица 7.11 – Результаты рекламной компании

Расходы на рекламу, тыс.руб.	5	8	6	5	3	9	12	4	3	10
Объемы продаж, тыс.руб.	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 об изменении объемов продаж в зависимости от изменений 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 (таблица 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.

Проведите регрессионный анализ и выявите зависимость в удовлетворенности от вышеперечисленных факторов.

Таблица 7.12 – Результаты опроса отдыхающих

№	Удовлетворенность отдыхом	Место-расположение	Персонал	Уровень затрат	№	Удовлетворенность отдыхом	Место-расположение	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
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

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. Сделать вывод о вероятности возникновения нежелательного эффекта мультиколлинеарности.

6. Сделать 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*[2].

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

1. Существуют ли различия между сегментами рынка по предпочтению к различным маркам товара?

2. Какие существуют различия между постоянными покупателями супермаркетов и небольших продуктовых магазинов?

3. Какие характеристики присущи потребителям, реагирующим на рекламу в 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 используется зависимая номинальная переменная, а не количественная.

В результате проведения дискриминантного анализа должно быть построено дискриминационное уравнение:

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

где d — зависимая номинальная переменная;

b_n — коэффициенты при независимых переменных;

a — постоянный член (Константа);

x_n — независимые дискриминационные признаки (предикторы).

Данная модель позволяет предсказать принадлежность объектов исследования к двум или более непересекающимся группам.

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», построить дискриминантную модель, с помощью которой, основываясь на социально-демографических признаках, разделить всех туристов на две группы: выбирающих в качестве развлечений дискотеки на 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» (таблица 8.1) дается обзор действительных и пропущенных значений.

Таблица 8.1 - Сводка результатов обработки наблюдений

Невзвешенные наблюдения	N	Процент
Валидные	46	90,0
Исключены		
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.	51	100,0

Из 51 респондента, участвовавшего в опросе, в построении дискриминационной функции используются данные по 46 респондентам.

Число респондентов, не давших ответы хотя бы по одному из выбранных признаков, составляет 5.

В следующей таблице «Групповые статистики» (Статистические показатели в группах) показаны результаты расчета средних значений независимых переменных в каждой группе (таблица 8.2).

Таблица 8.2 - Групповые статистики

проведение вечерних дискотек	Среднее	Стд.отклонение	Кол-во 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

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

Уровень дохода туристов, посещающих дискотеки ниже (1,13, то есть ближе к 1, что соответствует интервалу менее 5000 руб), чем уровень дохода, не посещающих дискотеки 1,91, то есть ближе к 2, что соответствует интервалу в 5000 -100000 руб.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 среди туристов,

посещающих дискотеки доля женщин - 57% (среднее 1,57), среди туристов, не посещающих дискотеки – 48% (среднее 1,48).

2. Оценка статистической значимости выбранных переменных.

Исходя из таблицы 8.3 «Критерий равенства групповых средних» необходимо выбрать независимые дискриминационные переменные. Для проведения теста на равенство средних значений в группах используется коэффициент Лямбда Уилкса (*Wilks' Lambda*), значение которого должно быть минимальным. О состоятельности построенной дискриминантной модели свидетельствует статистическая значимость «*Significance*» (p), значение которой должно быть меньше 0,05.

Таблица 8.3 - Критерий равенства групповых средних

	Лямбда Уилкса	F	ст.св1	ст.св2	Знч.
Пол	,992	,336	1	44	,565
Возраст					,000
Количество лет	,437	56,605	1	44	,000
Уровень дохода респондентов	,672	21,470	1	44	,000
расходы на питание					0,017

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 отличия туристов, посещающие и не посещающие дискотеки по данным характеристикам статистически значимы. 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 таблице 8.4 «Объединенные внутригрупповые матрицы» показаны коэффициенты корреляции между независимыми переменными для исследуемых групп.

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
Корреляция	Пол	Количество лет	Уровень дохода респондентов
Пол	1,000	,221	-,214
Количество лет	,221	1,000	-,516
Уровень дохода респондентов	-,214	-,516	1,000

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 [1], r меняется от -1 до +1. Чем ближе коэффициент корреляции к 0, тем слабее зависимость между переменными (см. раздел 3.3 «Корреляции»).

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.

Необходимо рассчитать нормированные коэффициенты канонической дискриминантной функции:

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

где d – зависимая (группирующая переменная),

a – свободный член (константа),

b_n – нормированные коэффициенты,

x_n – независимые переменные.

Используя коэффициенты канонической дискриминантной функции (таблица 8.5) можно оценить относительный вклад каждой переменной в различие исследуемых групп.

Таблица 8.5 - Нормированные коэффициенты канонической дискриминантной функции

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

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

Корреляционные коэффициенты, представленные в таблице 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 таблице 8.7, составляем дискриминационную модель:

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

где X_1 – уровень дохода,

X_2 – возраст,

X3 – 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

*Ненормированные коэффициенты

Итоговая дискриминантная модель:

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

Насколько четко разделены группы респондентов на предпочитающих и не предпочитающих дискотеки можно определить на основе расстояний между средними в группах (центроиды групп) (таблица 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) (пограничное значение - 0). Чем больше расстояние, тем больше различий между группами.

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

Каноническая дискриминантная функция 1

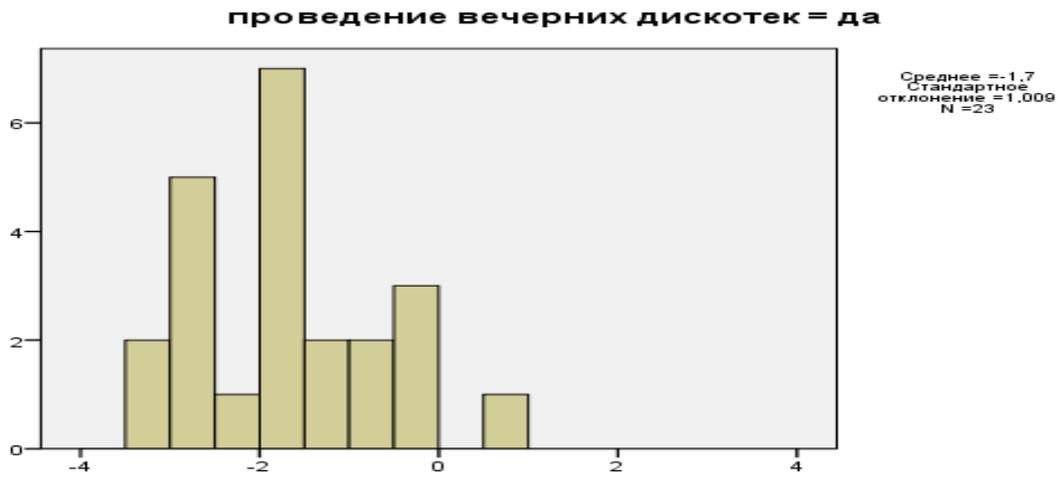


Рисунок 8.1 – Распределение значений дискриминантной функции для группы «посещающие дискотеки»

Каноническая дискриминантная функция 1

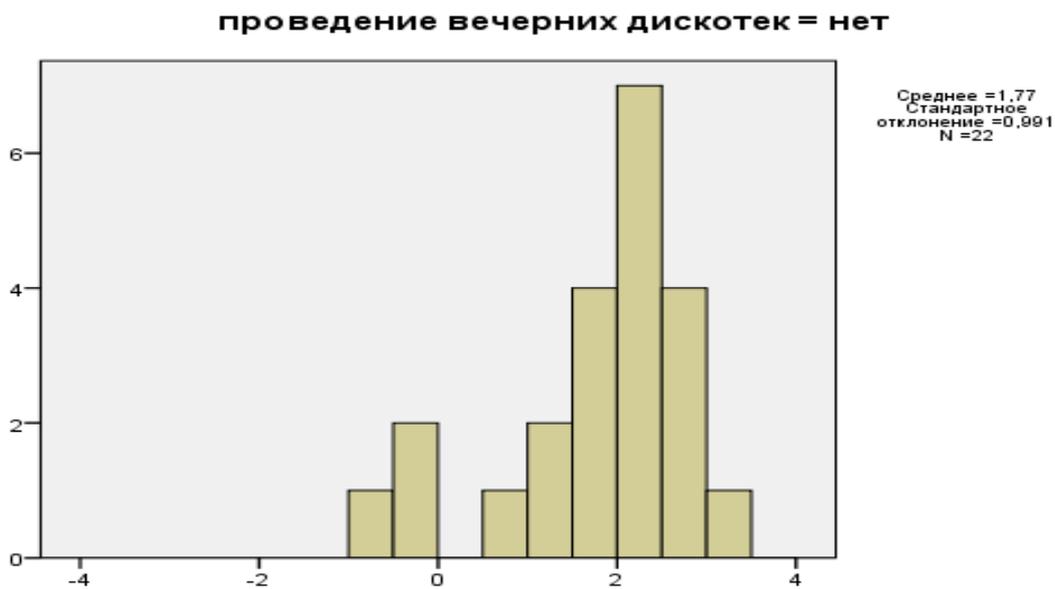


Рисунок 8.2 – Распределение значений дискриминантной функции для группы «не посещающие дискотеки»

Различия между группами ослабевают при увеличении рассеяния значений дискриминантной функции и области их пересечения в исследуемых группах, и наоборот.

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

Построим прогноз предпочтения туристом посещения дискотек, на основе возраста и уровня дохода.

Если возраст туриста равен 20 годам и уровень его дохода составляет от 5000 до 10000 руб. (категория 2), то значение дискриминантной функции имеет вид:

$$D = -5,642 + 0,622*2+0,106*20=-2,278.$$

Исходя из анализа графиков не посещающих дискотеки значения дискриминационной функции $-2,278$ не имеет ни один человек, а посещающие имеет 1 человек, следовательно можно сделать вывод, что туристы в возрасте 20 лет и уровнем дохода от 5000 до 10000 руб. скорее всего будут посещать дискотеки.

Точность прогнозов оценивается по результатам классификации по исследуемым группам. В таблице 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.

Таблица 8.9 – Поточечные статистики

Номер наблюдения	Фактическая группа	Наивероятнейшая группа			Вторая вероятнейшая группа			Дискриминантные баллы
		Предсказанная группа	$P(D>d G=g)$	$P(G=g D=d)$	Группа	$P(G=g D=d)$	Функция 1	
				Ковария расстояния Махалонобиса до центра				

			Р	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	2	,643	1	,988	,214	1	,012	9,038	1,310
	7	1	1	,782	1	,999	,077	2	,001	14,035	-1,973
	8	2	2	,950	1	,997	,004	1	,003	11,604	1,710
	9	2	2	,821	1	,995	,051	1	,005	10,522	1,548
	10	1	1	,625	1	1,000	,239	2	,000	15,671	-2,185

** - Неправильно классифицированное наблюдение

Точность сделанных прогнозов можно определить по данным таблицы 8.10 - Результаты классификации.

Таблица 8.10 – Результаты классификации

Проведение вечерних дискотек	Предсказанная 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.

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 о том, что 86,4% отдыхающих на базе отдыха и не предпочитающих дискотеки, могут их посещать. Следовательно, руководству базы отдыха необходимо продумать политику продвижения дискотек среди отдыхающих.

ЗАДАНИЯ ДЛЯ ПРАКТИЧЕСКИХ РАБОТ

Задание 1. Используя базу данных результатов опроса отдыхающих базы отдыха «Солнечная», построить дискриминантную модель, с помощью которой, основываясь на социально-демографических признаках, разделить всех туристов на две группы: выбирающих в качестве развлечений спортивные мероприятия при отдыхе за городом и не выбирающих.

Задание 2. Используя базу данных результатов опроса отдыхающих базы отдыха «Солнечная», построить дискриминантную модель, с помощью которой, основываясь на социально-демографических признаках, разделить всех туристов на две группы: выбирающих в качестве развлечений пешие прогулки в лесу при отдыхе за городом и не выбирающих.

Задание 3. Определить особенности семей, которые владеют автомобильным фургоном. Данные получены на основе выборки, включающей 40 семей. Были собраны следующие данные (таблица 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).

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

Таблица 8.11 – Данные выборки

№	Доход, тыс. долл.	Кол-во членов семьи	Образов ание главы семьи	Регион прожив ания	Стиль жизни	Кол-во машин в собстве нности	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
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

Задание 4. Используя базу данных результатов опроса отдыхающих базы отдыха «Солнечная», построить дискриминантную модель, с помощью которой, основываясь на социально-демографических признаках, разделить всех туристов на две группы: выбирающих в качестве развлечений туристические походы при отдыхе за городом и не выбирающих.

Задание 5. Используя базу данных результатов опроса отдыхающих базы отдыха «Солнечная», построить дискриминантную модель, с помощью которой, основываясь на социально-демографических признаках, разделить всех туристов на две группы: выбирающих в качестве развлечений катание на лошадях при отдыхе за городом и не выбирающих.

Задание 6. Используя базу данных результатов опроса отдыхающих базы отдыха «Солнечная», построить дискриминантную модель, с помощью которой, основываясь на социально-демографических признаках, разделить всех туристов на две группы: выбирающих в качестве развлечений баня, сауна, массаж при отдыхе за городом и не выбирающих.

Задание 7. Используя базу данных результатов опроса отдыхающих базы отдыха «Солнечная», построить дискриминантную модель, с помощью которой, основываясь на социально-демографических признаках, разделить всех туристов на две группы: выбирающих в качестве развлечений бассейн при отдыхе за городом и не выбирающих.

Задание 8. Используя базу данных результатов опроса отдыхающих базы отдыха «Солнечная», построить дискриминантную модель, с помощью которой, основываясь на социально-демографических признаках, разделить всех туристов на две группы: выбирающих в качестве развлечений культмассовые мероприятия при отдыхе 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, разделить всех туристов на две группы: мотивом выбора места отдыха за городом является близость к городу или не является.

Задание 10. Используя базу данных результатов опроса отдыхающих базы отдыха «Солнечная», построить дискриминантную модель, с помощью которой, основываясь на социально-демографических признаках, разделить всех туристов на две группы: мотивом выбора места отдыха за городом являются приемлемые цены или не являются.

Задание 11. Используя базу данных результатов опроса отдыхающих базы отдыха «Солнечная», построить дискриминантную модель, с помощью которой, основываясь на социально-демографических признаках, разделить всех туристов на две группы: мотивом выбора места отдыха за городом является близость водоема, леса или не является.

Задание 12. Используя базу данных результатов опроса отдыхающих базы отдыха «Солнечная», построить дискриминантную модель, с помощью которой, основываясь на социально-демографических признаках, разделить всех туристов на две группы: мотивом выбора места отдыха за городом является уровень комфорта или не является.

Задание 13. Используя базу данных результатов опроса отдыхающих базы отдыха «Солнечная», построить дискриминантную модель, с помощью которой, основываясь на социально-демографических признаках, разделить всех туристов на две группы: мотивом выбора места отдыха за городом является тишина или не является.

Задание 14. Используя базу данных результатов опроса отдыхающих базы отдыха «Солнечная», построить дискриминантную модель, с помощью которой, основываясь на социально-демографических признаках, разделить всех туристов на две группы: мотивом выбора места отдыха за городом является вкусная еда или не является.

Задание 15. Используя базу данных результатов опроса отдыхающих базы отдыха «Солнечная», построить дискриминантную модель, с помощью которой, основываясь на социально-демографических признаках, разделить всех туристов на две группы: мотивом выбора места отдыха за городом является наличие развлечений или не является.

Задание 16. Используя базу данных результатов опроса отдыхающих базы отдыха «Солнечная», построить дискриминантную модель, с помощью которой, основываясь на социально-демографических признаках, разделить всех туристов на две группы: мотивом выбора места отдыха за городом является комфорт отдыха с детьми или не является.

9 КЛАСТЕРНЫЙ АНАЛИЗ

Кластерный анализ – позволяет выявить группы (кластеры) объектов по заданным переменным. Программа при этом автоматически определяет количество существующих кластеров.

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].

С помощью кластерного анализа в маркетинге выявляются целевые группы потребителей, для которых необходимо разработать уникальный комплекс инструментов маркетинга.

Пример. Необходимо разделить всех курящих по возрасту и уровню дохода на однородные группы (кластеры) (рисунок 9.1).

На рисунке 9.1. вариант *B* не выявлено однородных кластеров, все курильщики представляют единую группу, следовательно, разработка уникальных торговых предложений не имеет смысла.

На рисунке 9.1. вариант *A* выявлены два однородных кластера курящих: «старые и бедные», «молодые и богатые». Данные группы могут быть представлены, как две целевые группы потребителей табачных изделий. Следовательно. Для каждой из них необходимо разработать уникальные торговые предложения по цене, качеству, особым свойствам, дизайну упаковки, особенностям продвижения, системе распределения товара.

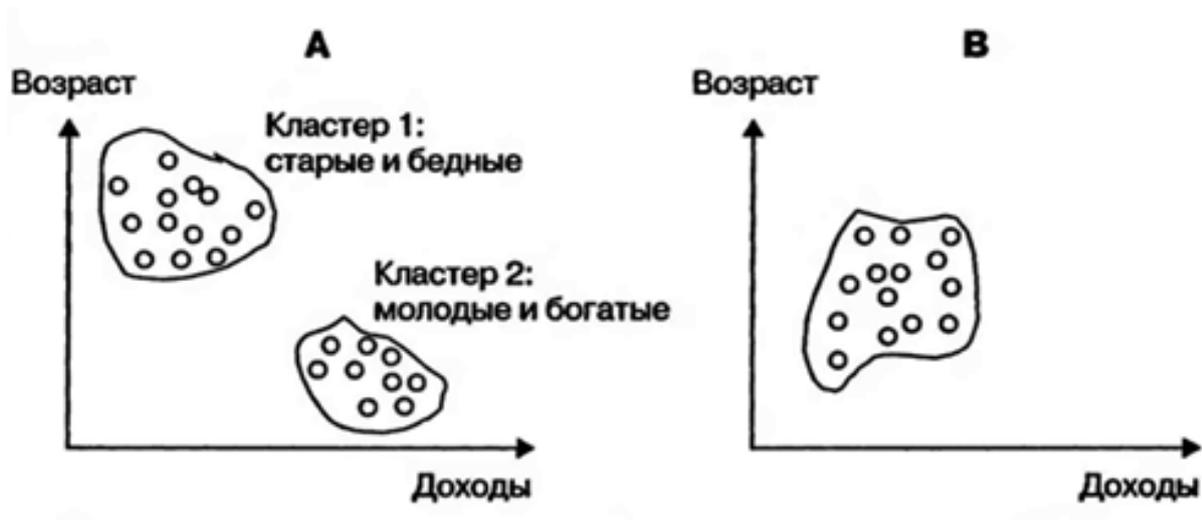


Рисунок 9.1 – Кластерный анализ [2]

В процессе выполнения кластерного анализа выявляются статистические связи между анализируемыми переменными, которые указывают на схожесть переменных и, затем объединение выявленных факторов в группы на основе уровня различий между ними. Количество кластеров зависит от задаваемых параметров схожести переменных, объединяемых в один кластер.

Основные термины, применяемые в кластерном анализе

Евклидово расстояние – расстояние между объектами, равное сумме квадратов разностей между значениями одноименных переменных объектов.

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 в значительно более крупные кластеры.

Разделяющая или дивизиональная кластеризация – иерархический метод формирования кластеров, при котором один общий большой все кластер делится на более мелкие для создания однородных кластеров.

Методы связи – методы формирования кластеров, при которых объекты объединяются в группу на основе рассчитанного между ними расстояния (рисунок 9.2).

Метод одиночной связи – или правило ближайшего соседа – в основе лежит выбор переменных, расстояние между которыми минимально.

Метод полной связи – или правило дальнего соседа - в основе лежит выбор переменных, расстояние между которыми максимально.

Метод средней связи – в основе лежит среднее значение расстояния между переменными разных кластеров.

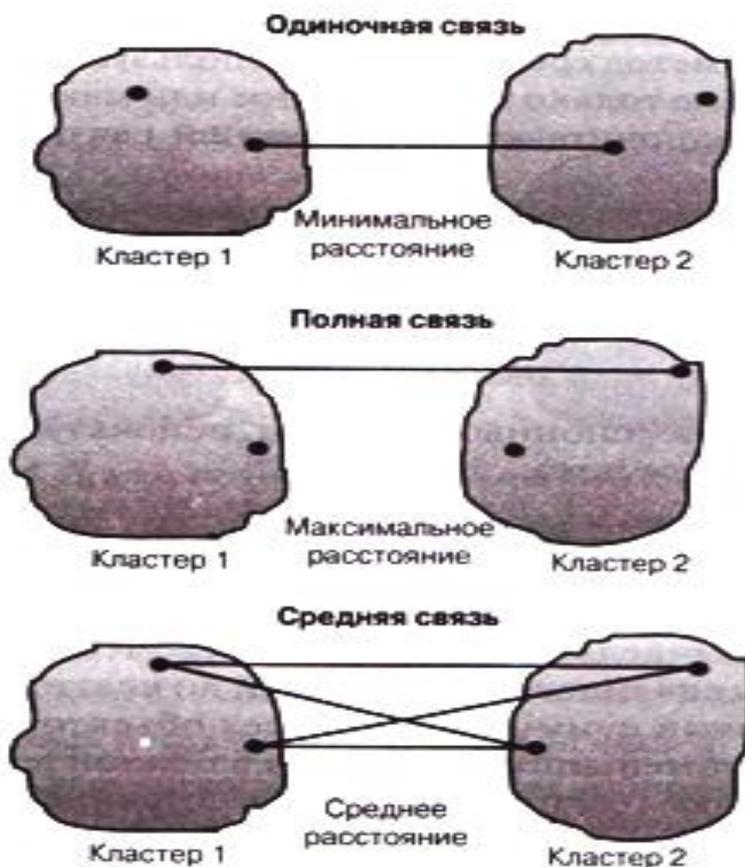


Рисунок 9.2 – Методы связи

Дисперсионные методы – Метод Варда и Центроидный метод (Рисунок 9.3).

Метод Варда – в качестве меры используется квадрат евклидового расстояния, который должен быть минимальным.

Центроидный метод – оценивается расстояние между центроидами (средними) групп переменных.

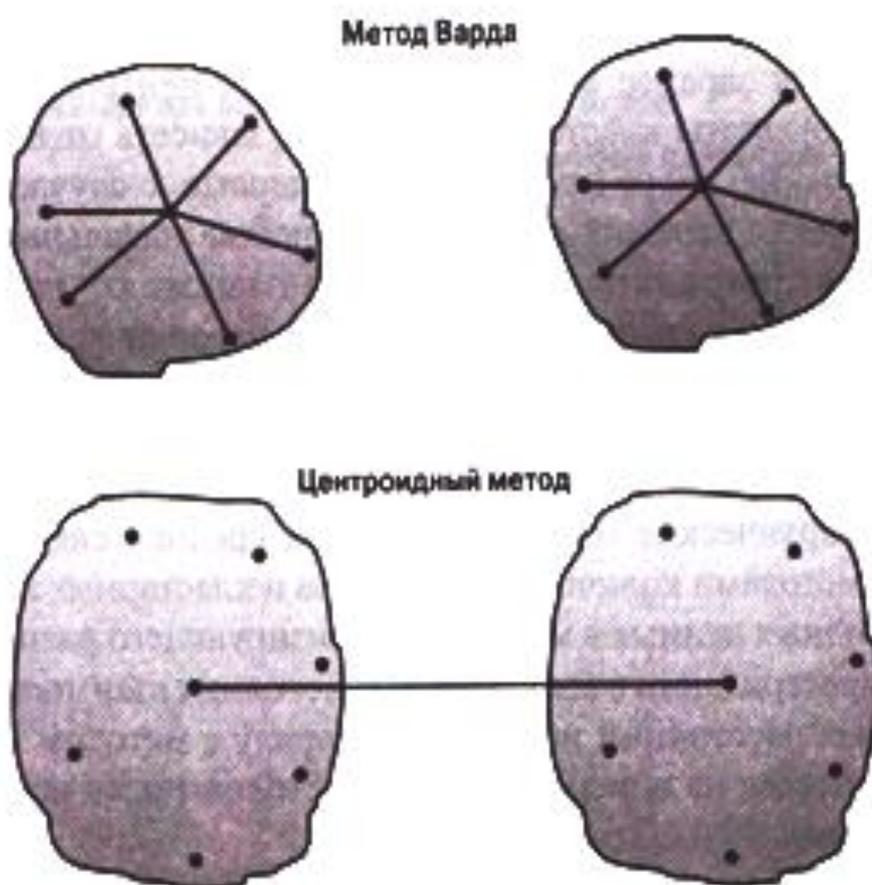


Рисунок 9.3 – Дисперсионные методы

Метод k – средних – переменные группируются в кластер в пределах порогового значения, которое задается исходя из определенного центра кластера.

Пример. 20 респондентов попросили выразить свое отношение к посещению торговых центров для приобретения товаров. Выделили шесть

утверждений, степень согласия с которыми респондентов попросили выразить на основе семибалльной шкалы (1 – совсем не согласен, 7 – абсолютно согласен).

Фас_1 – Считаю посещение торговых центром приятным процессом.

Фас_2 – Посещение торговых центров подрывает бюджет моей семьи.

Фас_3 – Мне нравится обедать и ужинать в торговых центрах.

Фас_4 – В торговых центрах мне нравится делать хорошие покупки.

Фас_5 – Меня очень раздражает посещение торговых центров.

Фас_6 – Посещая разные магазины, можно сравнивать цены и выбирать более дешевый товар.

	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										

Рисунок 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» (Рисунок 9.5).

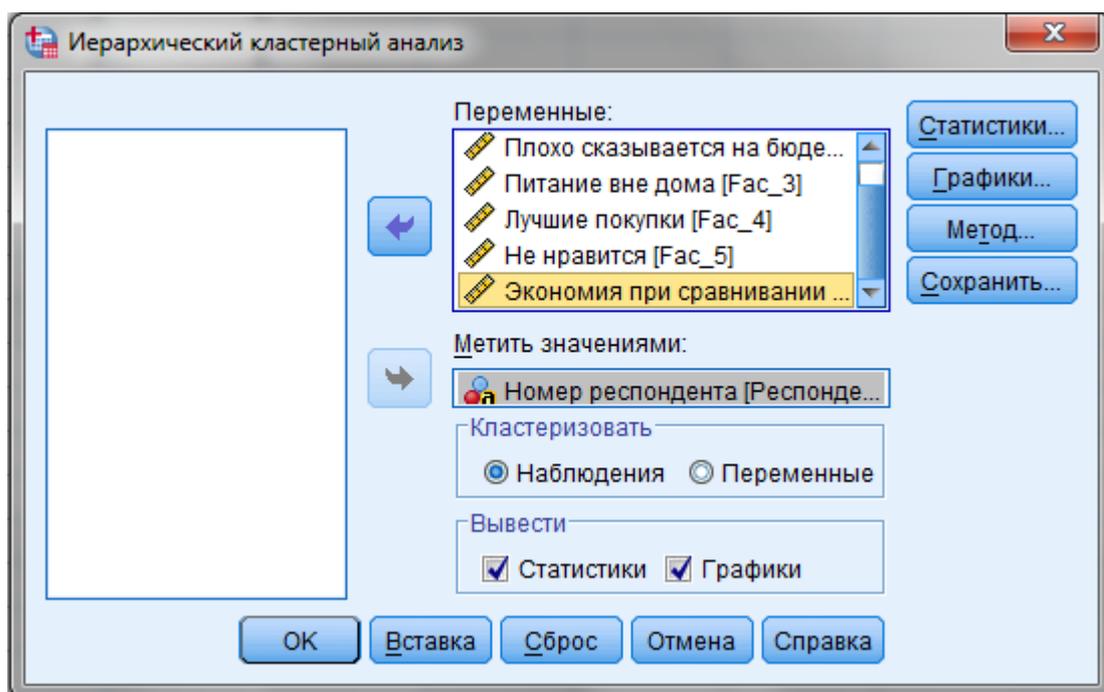


Рисунок 9.5 - Иерархический кластерный анализ

Выбираем все шесть факторов, отражающих отношение респондентов к посещению торговых центров, и переносим их в поле «Переменные».

В поле «Метить значениями» необходимо перенести переменную «Номер респондента». В данном случае могут использоваться только поле текстовые переменные, поэтому в столбце «Тип переменной» необходимо выбрать «Текстовая».

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» (Рисунок 9.6).

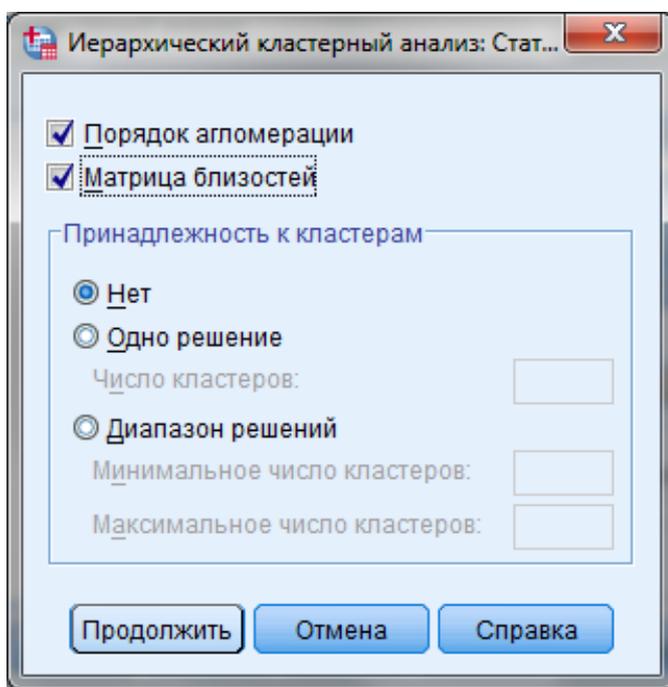


Рисунок 9.6 – Диалоговое окно «Статистические показатели»

В группе «Принадлежность к кластерам» можно выбрать три варианта:

- «Нет» - в результаты анализа включаются все кластеры.
- «Одно решение» - задается точное количество кластеров.
- «Диапазон решений» - дает возможность получить несколько решений с разным количеством кластеров.

Кнопка «Продолжить» - возвращаемся в основное диалоговое окно.

ШАГ 5. Кнопка «Графики» - на экране появляется вспомогательное диалоговое окно «Графики» (Рисунок 9.7).

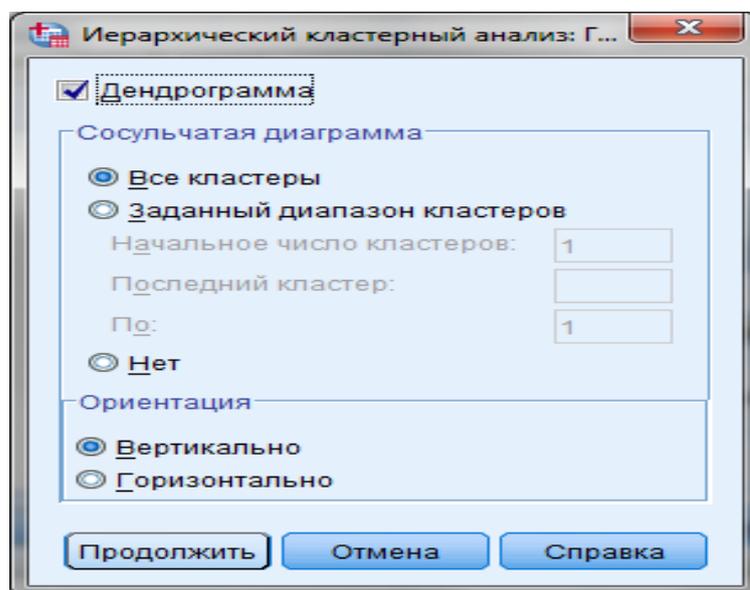


Рисунок 9.7 - Диалоговое окно «Графики»

В данном окне можно выбрать в качестве выводимых результатов анализа сосульчатую диаграмму, которая демонстрирует процесс формирования кластеров на основе величины разности между переменными.

Далее выбираем «Продолжить» - «Иерархический кластерный анализ».

ШАГ 6. Кнопка «Метод» - позволяет нам работать с функциями окна «Методы» (Рисунок 9.8).

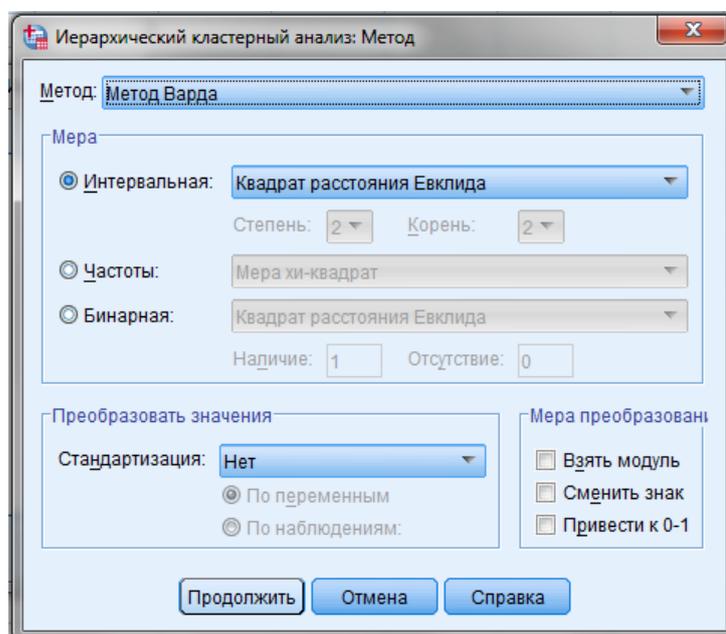


Рисунок 9.8 – Методы кластерного анализа

В поле «Метод» - выбрать метод «Варда».

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

Из предлагаемых программой интервальных показателей выбираем квадрат расстояния Евклида.

Кнопка «Продолжить» - «Иерархический кластерный анализ».

ШАГ 7. Кнопка «Сохранить» - диалоговое окно «Сохранить». В данном окне выбираем вариант сохранения результатов анализа в исходной базе данных. Выбираем кнопку «Диапазон решений» и устанавливаем диапазон от 3 до 5. В результате в конце списка переменных базы данных появятся новые переменные, которые были созданы в результате кластеризации. Данным переменным можно дать имена, которые соответствуют их общим характеристикам.

ШАГ 8. Кнопка «ОК».

Интерпретация полученных результатов

В таблице 9.2 «Шаги агломерации» показаны номера кластеров, которые объединяются на каждом этапе. В колонке «Коэффициенты» показаны эвклидовы расстояния между кластерами. В колонке «Следующий этап» показан шаг, на котором впервые был создан данный кластер.

С помощью данной таблицы можно провести предварительную оценку количества создаваемых кластеров, определив на каком этапе, резко возрастает евклидово расстояние между переменными.

Таблица 9.2 - Шаги агломерации

Этап	Кластер объединен с		Коэффициенты	Этап первого появления кластера		Следующий этап
	Кластер 1	Кластер 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

Следующим элементов окна вывода является древовидная диаграмма (рисунок 9.5). Данная диаграмма позволяет оценить расстояние между объектами и их принадлежность к кластерам на любом уровне.

Читать древовидную диаграмму нужно слева направо. Кластеры, которые объединяются в группу соединены вертикальными линиями. Шкала расстояний от 0 до 25 показывает величину квадрата расстояния Евклида, при которой кластеры объединяются. 0 – наименьшее расстояние первого этапа, 25 – наибольшее расстояние последнего этапа.

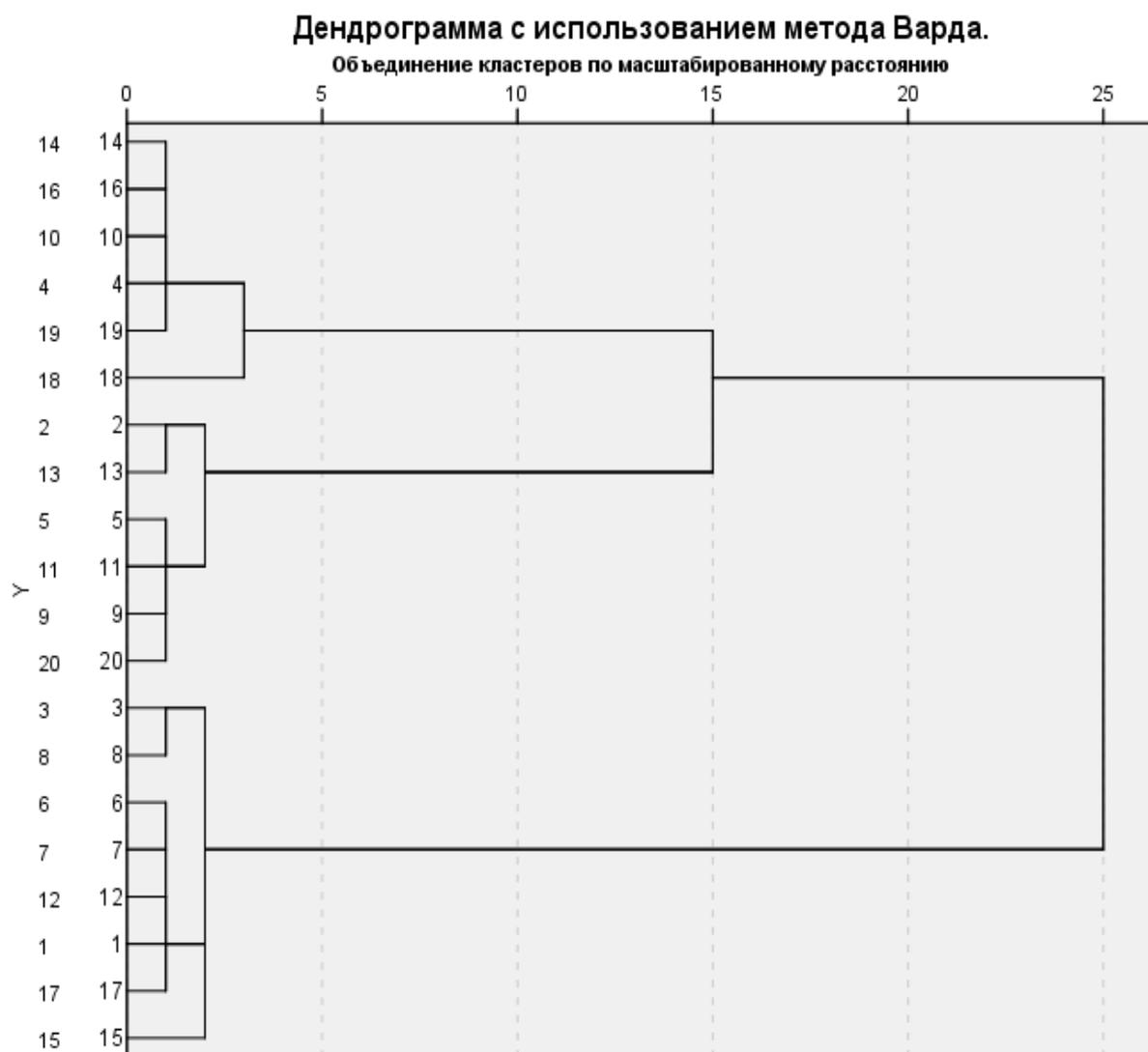


Рисунок 9.5 – Древоподобная диаграмма

Для определения количества кластеров можно руководствоваться следующим:

- теоретическими и практическими соображениями.
- этапом, на котором резко возрастает евклидово расстояние между переменными. В данном случае в два раза увеличивается расстояние между шагами 17 и 18. Следовательно, из общих 20 переменных вычитаем 17, получаем 3. Можно остановиться на 3 кластерах.
- размеры кластеров должны быть существенными, включать достаточное количество переменных.

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

Таблица 9.3 – Результаты кластерного анализа

Отчет

Ward Method	Приятный процесс	Подрывает бюджет	Питание вне дома	Хорошие покупки	Раздражение	Экономия при сравнении цен	
1	Среднее	5,75	3,63	6,00	3,13	1,88	3,88
	N	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
	N	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
	N	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
	N	20	20	20	20	20	20
	Стд.Отклонение	1,899	1,410	2,012	1,518	1,761	1,496

В таблице «Результаты кластерного анализа» выбираем переменные, включаемые в каждый из трех кластеров. В кластер 1 могут попасть респонденты, предпочитающие питаться вне дома и ответившие, что им нравится посещать магазины (среднее 6,00 и 5,75). Следовательно, данный кластер можно назвать «любители торговых центров». В данный кластер попадают переменные: 1, 3, 6, 7, 8, 12, 15 и 17. В кластер 2 попадают случаи

2, 5, 9, 11, 13 и 20, которые соответствуют варианту ответов, связанных с раздражением от торговых центров. Назовем его «раздраженные покупатели». Кластер 3 можно назвать «экономные покупатели», так как он имеет высокие средние показатели по факторам 2 (Подрывает бюджет), 4 (Хорошие покупки), 6 (Экономия при сравнении цен).

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

номер	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

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):

11. близость к городу

12. приемлемые цены

13. близость водоема, леса

14. уровень комфорта

15. тишина, уединение

16. хорошее питание

17. наличие развлечений

18. комфорт отдыха с детьми

19. возможность лечения, ухода за здоровьем

20. организация детского отдыха

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

Задание 2. 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 отдыхающих способа проведения досуга (1 – выбирает, 2 – не выбирает).

номер	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

Задание 3. В исследовании взаимосвязи между семейным и покупательским поведением проводилось измерение следующих утверждений о стиле жизни респондента (1 – не согласен, 7 – согласен) (Таблица 10.5).

V1 Я люблю проводить время дома, чем ходить на вечеринки

V2 Покупая товар, я всегда смотрю на цены

V3 Мне больше нравится ходить в магазины, чем в кино

V4 я не доверяю рекламе товаров

V5 Я – любитель сидеть дома.

V6 Я часто пользуюсь купонами скидок.

V7 на рекламу зря тратятся такие большие деньги

Проведите кластерный анализ.

Таблица 9.4 - Измерение утверждений о стиле жизни респондента

Номер	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
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 – представляет собой метод обобщения или сокращения большого количества переменных, объединение их в группы на основе характерных связей. В факторном анализе переменные не делятся на независимые и зависимые.

В практике маркетинговых исследований с 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 малое число групп, состоящих из переменных, объединенных общими факторами (Рисунок 10.1).

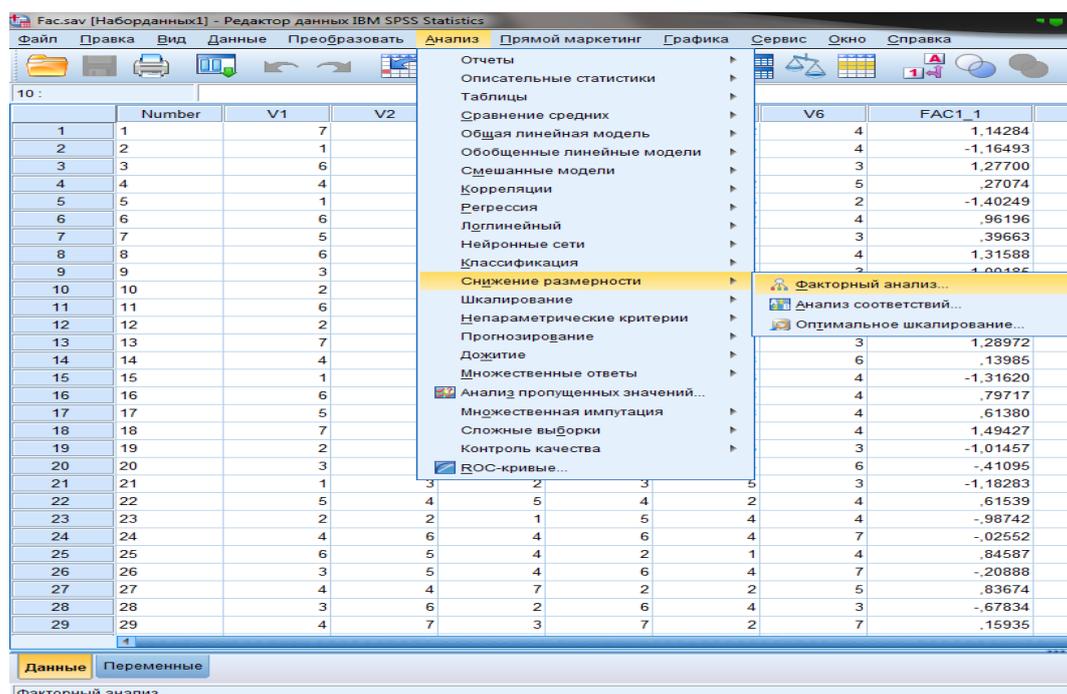


Рисунок 10.1 –Диалоговое окно факторного анализа

В один фактор объединяются переменные, плотно коррелирующие между собой и слабо коррелирующие с переменными, которые объединяются на основе других факторов. Факторный анализ проводится с целью сокращения числа переменных и упрощение процедуры анализа существующей базы данных.

В процессе проведения факторного анализа рассчитываются и анализируются следующие показатели:

- Критерий сферичности Бартлетта – показатель, с помощью которого проверяют, отличаются ли корреляции от 0. Если γ близко к нулю, то выбранная переменная не взаимосвязана с другими. Значимость меньше 0,05 указывает на то что проведение факторного анализа приемлемо.

- Корреляционная матрица – матрица, включающая в себя все возможные коэффициенты корреляций r между анализируемыми переменными.

- КМО - мера адекватности выборки Кайзера-Мейера-Олкина – величина, используемая для оценки применимости факторного анализа. Значения от 0,5 до 1 говорят об адекватности факторного анализа, значения до 0,5 указывают на то, что факторный анализ неприменим к выборке.

Графическое изображение критерия «каменистой осыпи» – график собственных значений факторов, расположенных в порядке убывания, используется для определения достаточного числа факторов.

Процедура факторного анализа включает следующие этапы:

- 1.Формулировка проблемы.
- 2.Проверка возможности проведения, вычисление корреляционной матрицы.
- 3.Выбор метода факторного анализа.
- 4.Извлечение 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 возможность проведения и адекватность факторного анализа для данной выборки;
- вычислить корреляционную матрицу и выявить взаимосвязи между переменными базы данных;
- выявить и извлечь необходимое количество факторов для создания упрощенной структуры;
- разбить базу данных на группы факторов на основе значений совместной 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».

Если есть необходимость провести факторный анализ отдельно для двух переменных, например мужчин и женщин, то в поле «Переменная отбора наблюдений» вносится переменная «пол». В данном случае нет необходимости проводить такое деление.

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» (Рисунок 10.2).

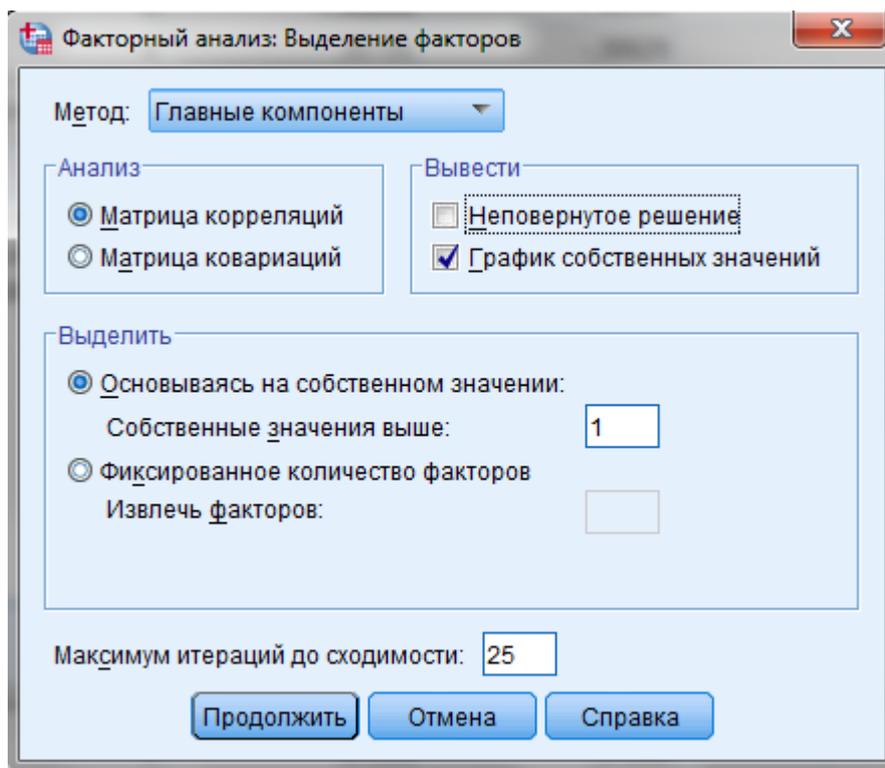


Рисунок 10.2 – Матрица корреляций

ШАГ 6. Задать условие: *собственное значение больше «1»*. При данном условии программа определит факторы в количестве больше 1.

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» - «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»*. В итоге создаются новые переменные, которые можно будет использовать в дальнейшем анализе.

ШАГ 10. «ОК».

Интерпретация результатов

1. Величина КМО показывает приемлемую адекватность выборки для факторного анализа $KMO = 0,512 > 0,5$. Критерий Бартлетта ($p < 0,05$), что говорит о целесообразности факторного анализа в силу коррелированности факторов.

Таблица 10.1 - Результаты теста КМО и Бартлетта

Мера адекватности и критерий Бартлетта

Мера выборочной адекватности Кайзера-Мейера-Олкина.		,512
Критерий сферичности	Приблиз. хи-квадрат	262,3
Бартлетта	ст.св.	45
	Знч.	,000

Таблица 10.2 – Корреляционная матрица.

Корреляционная матрица

	близос ть к городу	приемл емые цены	близос ть водое ма, леса	урове нь комф орта	тиши на, уеди нение	хоро шее питан ие	налич ие развле чений	комф орт отды ха с деть ми	возможн ость лечения, ухода за здоровье м	организа ция детск. отдыха
Кор реля ция	1,000	-,060	,009	,519	-,172	,149	,322	,075	-,166	,241
близость к городу										
приемле мые цены	-,060	1,000	-,064	-,341	,270	-,247	-,121	,232	,273	-,407
близость водоема, леса	,009	-,064	1,000	-,056	,017	-,083	,107	,114	,027	,030
уровень комфорта	,519	-,341	-,056	1,000	-,113	,384	,070	,008	-,125	,256
тишина, уедине ние	-,172	,270	,017	-,113	1,000	,084	-,167	-,060	,989	-,129
хорошее питание	,149	-,247	-,083	,384	,084	1,000	-,422	,191	,036	,237
наличие развлече ний	,322	-,121	,107	,070	-,167	-,422	1,000	-,045	-,116	,248
комфорт отдыха с детьми	,075	,232	,114	,008	-,060	,191	-,045	1,000	-,053	,202
возможн ость лечения, ухода за здоровье м	-,166	,273	,027	-,125	,989	,036	-,116	-,053	1,000	-,106
организа ция детского отдыха	,241	-,407	,030	,256	-,129	,237	,248	,202	-,106	1,000

Коэффициенты корреляции характеризуют плотность связи между переменными исходного массива.

2. Выявление и извлечение необходимого количества факторов для создания упрощенной структуры

Таблица 10.4 - Полная объясненная дисперсия

Компонента	Начальные собственные значения			Суммы квадратов нагрузок вращения		
	Всего	% дисперсии	Кумулятивный %	Всего	% дисперсии	Кумулятивный %
1	2,57	25,7	25,7	2,07	20,76	20,76
2	1,79	17,95	43,66	1,84	18,47	39,23
3	1,4	14,02	57,68	1,56	15,62	54,86
4	1,23	12,29	69,97	1,34	13,42	68,28
5	1,07	10,79	80,77	1,24	12,48	80,77
6	,87	8,69	89,47			
7	,43	4,31	93,78			
8	,34	3,43	97,21			
9	,27	2,7	99,91			
10	,008	,08	100,0			

Метод выделения: Анализ главных компонент.

Начальные собственные значения должны быть больше 1.

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

График собственных значений

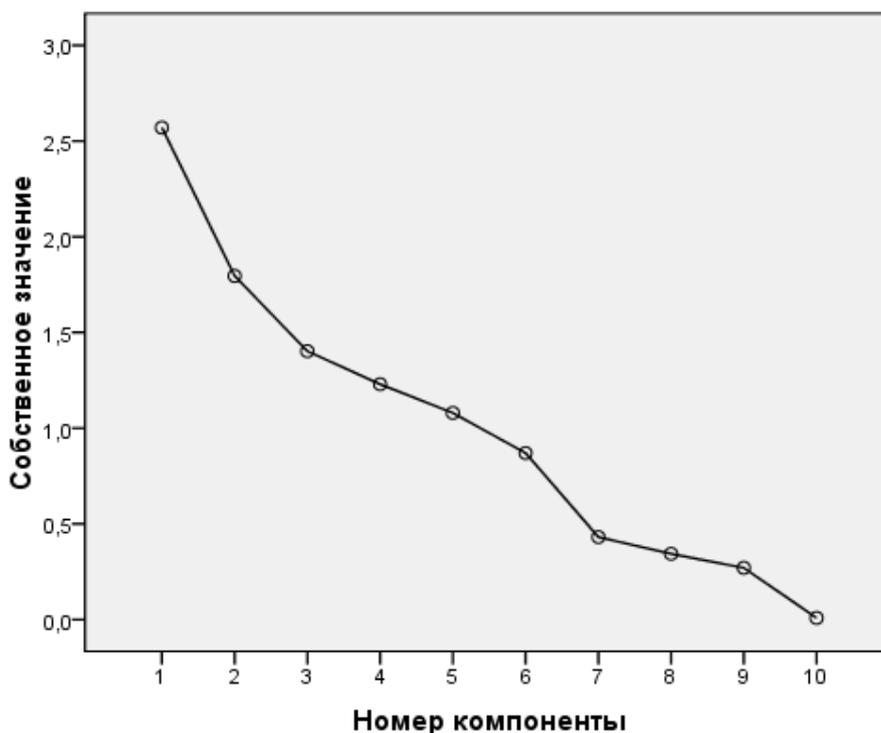


Рисунок 10.1 – График собственных значений

График показывает соответствующие собственные значения в системе координат: с 5 по 6 факторы происходит перелом графика. Это подтверждает, что оптимальное количество факторов 5.

2. На основании ротированной матрицы (таблица 10.5) компонентов в одну группу собираются переменные, которые наиболее тесно взаимосвязанные между собой (наиболее высокое значение коэффициента корреляции). В результате программа группирует переменные исходного массива и создает матрицу преобразования компонент (таблица 10.6)

Таблица 10.5 - 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 10.6 - Матрица преобразования компонент

Компонента	1	2	3	4	5
1	-,69	,61	,023	,36	-,12
2	,55	,48	,64	,19	,03
3	,45	,35	-,75	,32	,002
4	-,101	-,008	,000	,14	,98
5	-,020	,51	-,130	-,83	,123

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

3. В базе данных автоматически переносятся новые переменные построенной факторной модели (Рисунок 10.2). В столбце «Метка» отображается номер компонента факторной модели.

Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measurement	
28	транспорт	Numeric	8	0	вид транспорта	{1, автомобиль}...	None	8	Left	Nominal
29	досуг	Numeric	8	0	хотели бы Вы, чтобы ...	{1, да}...	None	8	Left	Nominal
30	спорт	Numeric	8	0	проведение спортивн...	{1, да}...	None	8	Left	Nominal
31	дискотека	Numeric	8	0	проведение вечерни...	{1, да}...	None	8	Left	Nominal
32	прогулки	Numeric	8	0	пешие прогулки по л...	{1, да}...	None	8	Left	Nominal
33	походы	Numeric	8	0	туристические походы	{1, да}...	None	8	Left	Nominal
34	кони	Numeric	9	0	катание на лошадях	{1, да}...	None	8	Left	Nominal
35	баня	Numeric	8	0	баня, сауна, массаж	{1, да}...	None	8	Left	Nominal
36	бассейн	Numeric	8	0	посещение бассейна	{1, да}...	None	8	Left	Nominal
37	катер	Numeric	8	0	катание на катерах, с...	{1, да}...	None	8	Left	Nominal
38	культурмассовые	Numeric	8	0	конкурсы, игры, конц...	{1, да}...	None	8	Left	Nominal
39	детские	Numeric	8	0	дневные и вечерние ...	{1, да}...	None	8	Left	Nominal
40	отдыхавшие	Numeric	8	0	знакомые, отдыхавш...	{1, да}...	None	8	Left	Nominal
41	пресса	Numeric	8	0	реклама в прессе	{1, да}...	None	8	Left	Nominal
42	телевидение	Numeric	8	0	реклама по телевиде...	{1, да}...	None	8	Left	Nominal
43	представитель	Numeric	8	0	от представителя баз...	{1, да}...	None	8	Left	Nominal
44	тур_компания	Numeric	8	0	от представителя тур...	{1, да}...	None	8	Left	Nominal
45	важность_1	Numeric	8	0	спокойный отдых	{1, очень важно}...	None	8	Right	Nominal
46	важность_2	Numeric	8	0	спорт	{1, очень важно}...	None	8	Right	Nominal
47	важность_3	Numeric	8	0	развлечения, дискот...	{1, очень важно}...	None	8	Right	Nominal
48	важность_4	Numeric	8	0	общение с природой	{1, очень важно}...	None	8	Right	Nominal
49	важность_5	Numeric	8	0	туристические походы	{1, очень важно}...	None	8	Right	Nominal
50	важность_6	Numeric	8	0	катание на лошадях	{1, очень важно}...	None	8	Right	Nominal
51	важность_7	Numeric	8	0	й	{1, очень важно}...	None	8	Right	Nominal
52	важность_8	Numeric	8	0	посещение бассейна	{1, очень важно}...	None	8	Right	Nominal
53	важность_9	Numeric	8	0	культурные программы	{1, очень важно}...	None	8	Right	Nominal
54	тишина_лечение	Numeric	11	5	REGR factor score 1...	None	None	13	Right	Ordinal
55	комфорт	Numeric	11	5	REGR factor score 2...	None	None	13	Right	Ordinal
56	развлечение_питание	Numeric	11	5	REGR factor score 3...	None	None	13	Right	Ordinal
57	детский_отдых	Numeric	11	5	REGR factor score 4...	None	None	13	Right	Ordinal
58	приемлимые_цены	Numeric	11	5	возрастные группы т...	None	None	13	Right	Ordinal
59	FAC1_1	Numeric	11	5	REGR factor score 1...	None	None	13	Right	Scale
60	FAC2_1	Numeric	11	5	REGR factor score 2...	None	None	13	Right	Scale
61	FAC3_1	Numeric	11	5	REGR factor score 3...	None	None	13	Right	Scale
62	FAC4_1	Numeric	11	5	REGR factor score 4...	None	None	13	Right	Scale
63	FAC5_1	Numeric	11	5	REGR factor score 5...	None	None	13	Right	Scale
64										

Рисунок 10.2 – Фрагмент вкладки «Переменные»

Названия новых компонент необходимо занести в исходную базу данных в столбец «Метка» таблицы «Переменные», компьютер автоматически вычисляет значения новых переменных.

Суть новых переменных сводится к следующему: наибольшее отрицательное значение говорит о большей значимости переменной, и

наоборот, наибольшее положительное значение говорит о наименьшей значимости переменной.

Созданные переменные в дальнейшем могут использоваться для анализа, например для 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) (Таблица 10.5).

V1 Я люблю проводить время дома, чем ходить на вечеринки

V2 Покупая товар, я всегда смотрю на цены

V3 Мне больше нравится ходить в магазины, чем в кино

V4 я не доверяю рекламе товаров

V5 Я – любитель сидеть дома.

V6 Я часто пользуюсь купонами скидок.

V7 На рекламу зря тратятся такие большие деньги

Проведите факторный анализ.

Оптимизируйте структуру данных, сократив число переменных.

- оцените возможность проведения и адекватность факторного анализа для данной выборки;
- вычислите корреляционную матрицу и выявите взаимосвязи между переменными базы данных;
- выявите и извлеките необходимое количество факторов для создания упрощенной структуры;
- разбейте базу данных на группы факторов на основе значений совместной корреляции;
- подберите названия созданным переменным.

Таблица 10.5 - Измерение утверждений о стиле жизни респондента

Номер	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
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. Оптимизируйте структуру данных, сократив число переменных.

- оцените возможность проведения и адекватность факторного анализа для данной выборки;
- вычислите корреляционную матрицу и выявите взаимосвязи между переменными базы данных;
- выявите и извлеките необходимое количество факторов для создания упрощенной структуры;
- разбейте базу данных на группы факторов на основе значений совместной 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):

21. близость к городу
22. приемлемые цены
23. близость водоема, леса
24. уровень комфорта
25. тишина, уединение
26. хорошее питание
27. наличие развлечений
28. комфорт отдыха с детьми
29. возможность лечения, ухода за здоровьем
30. организация детского отдыха

номер	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 10.6 – Выбор отдыхающих способа проведения досуга (1 – выбирает, 2 – не выбирает).

номер	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

Оптимизируйте структуру данных, сократив число переменных.

- оцените возможность проведения и адекватность факторного анализа для данной выборки;
- вычислите корреляционную матрицу и выявите взаимосвязи между переменными базы данных;
- выявите и извлеките необходимое количество факторов для создания упрощенной структуры;
- разбейте базу данных на группы факторов на основе значений совместной корреляции;
- подберите названия созданным переменным.

ЛИТЕРАТУРА

1. Наследов А. SPSS 19: профессиональный статистический анализ данных. – СПб.: Питер, 2011. - 400 с.: ил.
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: Учебно-методическое пособие/Московский государственный университет экономики, статистики и информатики. - М.: 2004. – 144 с.
4. Просветов Г.И. Маркетинговые исследования: задачи и решения: Учебно-практическое пособие. – М.: Издательство «Альфа-пресс», 2013. – 240 с.
5. Котлер Ф., Келлер К.Л. Маркетинг менеджмент. 12-е изд. – СПб.: Питер, 2012 – 816 с.
6. Нэреш К. Малхотра Маркетинговые исследования и эффективный анализ статистических данных. К.: DiaSoft, 2010.- 768 с.
7. Нэреш К. Малхотра Маркетинговые исследования с помощью SPSS. Практическое руководство. /Издательский дом «Вильямс», 2006. - 1200 с.

ПРИЛОЖЕНИЯ

ПРИЛОЖЕНИЕ А

Таблица 1 – Результаты опроса отдыхающих базы отдыха «Солнечная»

№ п/п	Пол респондента	Возраст, лет	Семейное положение	Доход	Профессиональная деятельность
1.	мужчины	19.0	не замужем/холост	5000 - 10000 руб	студент
2.	мужчины	59.0	вдовец/вдова	5000 - 10000 руб	гос.служащий
3.	женщины	28.0	замужем/женат	10000 - 20000 руб	предприниматель
4.	мужчины	32.0	замужем/женат	5000 - 10000 руб	рабочий
5.	мужчины	29.0	разведена/разведен	20 000 - 30000 руб	предприниматель
6.	женщины	27.0	замужем/женат	30000 - 50000 руб	предприниматель
7.	женщины	21.0	не замужем/холост	5000 - 10000 руб	студент
8.	женщины	48.0	разведена/разведен	10000 - 20000 руб	предприниматель
9.	женщины	56.0	замужем/женат	5000 - 10000 руб	рабочий
10.	женщины	19.0	не замужем/холост	5000 - 10000 руб	студент
11.	женщины	21.0	не замужем/холост	5000 - 10000 руб	студент
12.	мужчины	17.0	не замужем/холост	менее 5000 руб	школьник
13.	мужчины	18.0	не замужем/холост	менее 5000 руб	студент
14.	мужчины	32.0	замужем/женат	5000 - 10000 руб	предприниматель
15.	женщины		нет ответа	5000 - 10000 руб	гос.служащий
16.	мужчины	29.0	не замужем/холост	10000 - 20000 руб	предприниматель
17.	мужчины	41.0	замужем/женат	20 000 - 30000 руб	предприниматель
18.	мужчины	30.0	замужем/женат	более 50000 руб	предприниматель
19.	женщины	18.0	не замужем/холост	менее 5000 руб	студент
20.	женщины	20.0	не замужем/холост	5000 - 10000 руб	студент
21.	женщины	61.0	вдовец/вдова	5000 - 10000 руб	пенсионер
22.	женщины	54.0	замужем/женат	10000 - 20000 руб	гос.служащий
23.	женщины			5000 - 10000 руб	нет ответа
24.	мужчины	32.0	разведена/разведен	5000 - 10000 руб	преподаватель
25.	мужчины	36.0	замужем/женат	10000 - 20000 руб	гос.служащий
26.	женщины	59.0	замужем/женат	5000 - 10000 руб	пенсионер
27.	женщины		нет ответа	менее 5000 руб	нет ответа
28.	женщины	55.0	замужем/женат	5000 - 10000 руб	гос.служащий
29.	мужчины	60.0	вдовец/вдова	10000 - 20000 руб	пенсионер
30.	мужчины	27.0	замужем/женат	10000 - 20000 руб	предприниматель
31.	мужчины	32.0	не замужем/холост	30000 - 50000 руб	предприниматель
32.	мужчины	44.0	не замужем/холост	20 000 - 30000	рабочий

				руб	
33.	мужчины			нет ответа	нет ответа
34.	женщины	22.0	не замужем/холост	5000 - 10000 руб	студент
35.	женщины	25.0	не замужем/холост	5000 - 10000 руб	гос.служащий
36.	мужчины	21.0	не замужем/холост	5000 - 10000 руб	студент
37.	женщины	17.0	не замужем/холост	менее 5000 руб	школьник
38.	женщины	19.0	не замужем/холост	менее 5000 руб	студент
39.	женщины	63.0	вдовец/вдова	5000 - 10000 руб	пенсионер
40.	мужчины	29.0	замужем/женат	10000 - 20000 руб	предприниматель
41.	мужчины	41.0	замужем/женат	20 000 - 30000 руб	предприниматель
42.	женщины	27.0	замужем/женат	30000 - 50000 руб	гос.служащий
43.	женщины	56.0	разведена/разведен	10000 - 20000 руб	гос.служащий
44.	женщины	25.0	замужем/женат	5000 - 10000 руб	рабочий
45.	мужчины	34.0	замужем/женат	30000 - 50000 руб	рабочий
46.	мужчины	35.0	замужем/женат	5000 - 10000 руб	гос.служащий
47.	женщины	20.0	не замужем/холост	менее 5000 руб	студент
48.	женщины	57.0	замужем/женат	10000 - 20000 руб	гос.служащий
49.	женщины	19.0	не замужем/холост	5000 - 10000 руб	студент
50.	мужчины	18.0	не замужем/холост	менее 5000 руб	студент
51.	женщины		нет ответа	5000 - 10000 руб	гос.служащий
52.	мужчины	29.0	не замужем/холост	10000 - 20000 руб	предприниматель
53.	мужчины	41.0	замужем/женат	20 000 - 30000 руб	предприниматель
54.	мужчины	30.0	замужем/женат	более 50000 руб	предприниматель
55.	женщины	18.0	не замужем/холост	менее 5000 руб	студент
56.	женщины	20.0	не замужем/холост	5000 - 10000 руб	студент
57.	женщины	61.0	вдовец/вдова	5000 - 10000 руб	пенсионер
58.	женщины	54.0	замужем/женат	10000 - 20000 руб	гос.служащий
59.	женщины			5000 - 10000 руб	нет ответа
60.	мужчины	32.0	разведена/разведен	5000 - 10000 руб	преподаватель
61.	мужчины	36.0	замужем/женат	10000 - 20000 руб	гос.служащий
62.	женщины	59.0	замужем/женат	5000 - 10000 руб	пенсионер
63.	женщины		нет ответа	менее 5000 руб	нет ответа
64.	женщины	55.0	замужем/женат	5000 - 10000 руб	гос.служащий
65.	мужчины	60.0	вдовец/вдова	10000 - 20000 руб	пенсионер
66.	мужчины	19.0	не замужем/холост	5000 - 10000 руб	студент
67.	мужчины	59.0	вдовец/вдова	5000 - 10000 руб	гос.служащий
68.	женщины	28.0	замужем/женат	10000 - 20000 руб	предприниматель
69.	мужчины	32.0	замужем/женат	5000 - 10000 руб	рабочий

70.	мужчины	29.0	разведена/разведен	20 000 - 30000 руб	предприниматель
71.	женщины	27.0	замужем/женат	30000 - 50000 руб	предприниматель
72.	женщины	21.0	не замужем/холост	5000 - 10000 руб	студент
73.	женщины	48.0	разведена/разведен	10000 - 20000 руб	предприниматель
74.	женщины	56.0	замужем/женат	5000 - 10000 руб	рабочий
75.	женщины	19.0	не замужем/холост	5000 - 10000 руб	студент
76.	женщины	21.0	не замужем/холост	5000 - 10000 руб	студент
77.	мужчины	17.0	не замужем/холост	менее 5000 руб	школьник
78.	мужчины	18.0	не замужем/холост	менее 5000 руб	студент
79.	мужчины	32.0	замужем/женат	5000 - 10000 руб	предприниматель
80.	женщины		нет ответа	5000 - 10000 руб	гос.служащий
81.	мужчины	21.0	не замужем/холост	5000 - 10000 руб	студент
82.	женщины	17.0	не замужем/холост	менее 5000 руб	школьник
83.	женщины	19.0	не замужем/холост	менее 5000 руб	студент
84.	женщины	63.0	вдовец/вдова	5000 - 10000 руб	пенсионер
85.	мужчины	29.0	замужем/женат	10000 - 20000 руб	предприниматель
86.	мужчины	41.0	замужем/женат	20 000 - 30000 руб	предприниматель
87.	женщины	27.0	замужем/женат	30000 - 50000 руб	гос.служащий
88.	женщины	56.0	разведена/разведен	10000 - 20000 руб	гос.служащий
89.	женщины	25.0	замужем/женат	5000 - 10000 руб	рабочий
90.	мужчины	34.0	замужем/женат	30000 - 50000 руб	рабочий
91.	мужчины	35.0	замужем/женат	5000 - 10000 руб	гос.служащий
92.	женщины	20.0	не замужем/холост	менее 5000 руб	студент
93.	женщины	57.0	замужем/женат	10000 - 20000 руб	гос.служащий
94.	женщины	19.0	не замужем/холост	5000 - 10000 руб	студент
95.	мужчины	18.0	не замужем/холост	менее 5000 руб	студент
96.	женщины		нет ответа	5000 - 10000 руб	гос.служащий
97.	мужчины	29.0	не замужем/холост	10000 - 20000 руб	предприниматель
98.	мужчины	41.0	замужем/женат	20 000 - 30000 руб	предприниматель
99.	мужчины	30.0	замужем/женат	более 50000 руб	предприниматель
100.	женщины	18.0	не замужем/холост	менее 5000 руб	студент