

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Оренбургский государственный университет»

Кафедра математических методов и моделей в экономике

А.В. Раменская, О.С. Чудинова

МЕТОДЫ АНАЛИЗА СООТВЕТСТВИЙ В ИССЛЕДОВАНИИ РЫНКА

Методические указания

Рекомендовано к изданию редакционно-издательским советом федерального государственного бюджетного образовательного учреждения высшего образования «Оренбургский государственный университет» для обучающихся по образовательным программам высшего образования по направлениям подготовки 01.03.04 Прикладная математика, 38.03.05 Бизнес-информатика

Оренбург
2019

УДК 330.4(076.5)
ББК 65в631я7
Р 21

Рецензент – кандидат экономических наук, доцент О.И. Бантикова

Раменская, А. В.

Р 21

Методы анализа соответствий в исследовании рынка: методические указания / А.В. Раменская, О.С. Чудинова; Оренбургский гос. ун-т. – Оренбург: ОГУ, 2019. – 39 с.

Методические указания содержат описание теоретических аспектов методов анализа соответствий, постановку задачи по использованию методов анализа соответствий в исследовании рынка и пример её решения с помощью ППП STATISTICA. Методические указания предназначены для подготовки к лабораторным занятиям по дисциплине «Методы анализа рынка и организация продаж», могут также использоваться для самостоятельной работы студентов по дисциплине «Анализ данных» и другим дисциплинам, связанным с многомерным статистическим анализом данных.

Методические указания предназначены для обучающихся по образовательным программам высшего образования по направлениям подготовки 01.03.04 Прикладная математика, 38.03.05 Бизнес-информатика, профиль «Моделирование бизнес-систем» всех формы обучения. Методические указания могут использоваться студентами других направлений подготовки для проведения многомерного статистического анализа данных.

УДК 330.4(076.5)
ББК 65в631я7

© Раменская А.В.,
Чудинова О.С., 2019
© ОГУ, 2019

Содержание

Введение	4
1 Теоретические аспекты методов анализа соответствий.....	6
1.1 Общий вид выборочной двухфакторной таблицы сопряженности	6
1.2 Основные этапы решения задачи исследования связи двух номинальных признаков	7
1.3 Основные понятия анализа соответствий.....	9
1.4 Алгоритм простого анализа соответствий.....	12
1.5 Анализ многомерных данных и множественный анализ соответствий	16
2 Описание лабораторной работы	18
3 Постановка задачи.....	19
4 Порядок выполнения работы	20
4.1 Предварительный анализ результатов анкетирования.....	20
4.2 Реализация простого анализа соответствий в ППП STASTICA	25
5 Содержание письменного отчета.....	36
6 Контрольные вопросы.....	37
Список использованных источников	38
Приложение А.....	39

Введение

Исследование рынка включает в себя широкий спектр задач, для решения которых используется различный математический инструментарий. Информационной базой зачастую выступают данные маркетинговых исследований. Помимо анализа взаимосвязи между факторами, прогнозирования будущих значений ключевых показателей, также рассматриваются задачи разведывательного анализа данных. Для анализа и визуализации информации, представленной в таблицах сопряженности большой размерности, применяется анализ соответствий.

Целью методических указаний является продемонстрировать реализацию метода простого анализа соответствий в пакете прикладных программ Statistica на примере задачи анализа конкурентов и анализа образа бренда.

Выполнение лабораторных работ и самостоятельной работы по дисциплине «Математические методы анализа рынка и организация продаж», относящейся к обязательным дисциплинам (модулям) вариативной части блока 1 «Дисциплины (модули)», направлено на формирование у обучающихся по направлению подготовки 38.03.05 Бизнес-информатика, профиль «Моделирование бизнес-систем» следующих профессиональных компетенций: ПК-2 проведение исследования и анализа рынка информационных систем и информационно-коммуникативных технологий; ПК-3 выбор рациональных информационных систем и информационно-коммуникативных решения для управления бизнесом; ПК-4 проведение анализа инноваций в экономике, управлении и информационно-коммуникативных технологиях; ПК-18 способность использовать соответствующий математический аппарат и инструментальные средства для обработки, анализа и систематизации информации по теме исследования; ПК-19 умение готовить научно-технические отчеты, презентации, научные публикации по результатам выполненных исследований.

Выполнение лабораторных работ и самостоятельной работы по дисциплине «Анализ данных», относящейся к базовой части блока Д «Дисциплины (модули)», направлено на формирование у обучающихся по направлению подготовки 01.03.04

Прикладная математика следующих общепрофессиональных компетенций: ОПК-2 Способен обоснованно выбирать, дорабатывать и применять для решения исследовательских и проектных задач математические методы и модели, осуществлять проверку адекватности моделей, анализировать результаты, оценивать надежность и качество функционирования систем; ОПК-3 Способен использовать и развивать методы математического моделирования и применять аналитические и научные пакеты прикладных программ.

1 Теоретические аспекты методов анализа соответствий

1.1 Общий вид выборочной двухфакторной таблицы сопряженности

Исследование взаимосвязи номинальных признаков осуществляется на основе таблиц сопряженности. Рассмотрим два категоризованных номинальных признака X и Y . Признак X может принимать значения x_1, x_2, \dots, x_r ; признак Y – y_1, y_2, \dots, y_s . Пусть имеется выборка, сформированная методом перекрестного отбора, объема n из генеральной совокупности (X, Y) . Тогда эмпирическое распределение генеральной совокупности (X, Y) может быть представлено в виде двухфакторной таблицы сопряженности признаков X и Y размерности $r \times s$, имеющей вид:

$x_i \backslash y_j$	y_1	y_2	\dots	y_s	n_{i*}
x_1	n_{11}	n_{12}	\dots	n_{1s}	n_{1*}
x_2	n_{21}	n_{22}	\dots	n_{2s}	n_{2*}
\dots	\dots	\dots	\dots	\dots	\dots
x_r	n_{r1}	n_{r2}	\dots	n_{rs}	n_{r*}
n_{i*}	n_{*1}	n_{*2}	\dots	n_{*s}	n

где n_{ij} – наблюдаемая частота, т.е. число объектов в выборочной совокупности, характеризующихся i -ой категорией признака X и j -ой категорией признака Y , $i = \overline{1, r}$, $j = \overline{1, s}$;

n_{i*} – маргинальная частота, характеризующая сколько объектов в выборочной совокупности обладают i -ой категорией признака X , $n_{i*} = \sum_{j=1}^s n_{ij}$, $i = \overline{1, r}$;

n_{*j} – маргинальная частота, характеризующая сколько объектов в выборочной совокупности обладают j -ой категорией признака Y , $n_{*j} = \sum_{i=1}^r n_{ij}$, $j = \overline{1, s}$;

$$\sum_{i=1}^r n_{i*} = \sum_{j=1}^s n_{*j} = \sum_{i=1}^r \sum_{j=1}^s n_{ij} = n.$$

1.2 Основные этапы решения задачи исследования связи двух номинальных признаков

Основная задача, возникающая при анализе таблиц сопряженности, состоит в выяснении, есть ли связь между признаками X и Y . Можно выделить три этапа решения этой задачи.

1) Проверка гипотезы о независимости признаков X и Y . Для проверки гипотезы используются критерий Пирсона, критерий Пирсона с поправкой Йетса на непрерывность, критерий максимального правдоподобия, точный критерий Фишера. Описание критериев приведено в работах [1,2].

Статистика критерия Пирсона χ^2 имеет вид:

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^s \frac{(n_{ij} - n_{ij}^*)^2}{n_{ij}^*}, \quad (1.1)$$

где n_{ij}^* – теоретические частоты, т.е. те частоты, которые были бы при независимости признаков X и Y , $n_{ij}^* = \frac{n_{i*} \cdot n_{*j}}{n}$.

Статистика (1.1) при независимости признаков, $n \rightarrow \infty$ и отсутствии малых теоретических частот имеет распределение «Хи-квадрат» с числом степеней свободы $\nu = (r-1)(s-1)$.

Если гипотеза о независимости признаков отклоняется, то переходят ко второму и третьему этапам решения задачи.

2) Расчет оценок коэффициентов связи. На основе таблиц сопряженности 2×2 рассчитываются следующие коэффициенты связи:

– меры связи, основанные на статистике χ^2 (φ -коэффициент Чупрова-Крамера, коэффициент сопряженности Пирсона, коэффициент контингенции Крамера, τ - коэффициент Гудмена и Краскала);

– меры связи, основанные на отношении преобладаний (коэффициент ассоциации Юла, коэффициент коллигации Юла, отношение преобладаний).

На основе таблиц сопряженности $r \times s$ рассчитываются следующие коэффициенты связи:

– меры связи, основанные на статистике χ^2 (φ -коэффициент Чупрова-Крамера, коэффициент сопряженности Пирсона, коэффициент Чупрова, коэффициент Крамера);

– коэффициенты Гудмена и Краскала.

Формулы для расчета оценок и свойства перечисленных коэффициентов связи приведены в работах [1, 2].

3) Интерпретации связи. Учитывая диапазон возможных значений коэффициентов связи, можно дать характеристику силы связи. «Направление» связи можно охарактеризовать лишь для таблиц 2×2 на основе оценок коэффициентов контингенции, ассоциации, коллигации. Если эти коэффициенты принимают положительные значения, то связь интерпретируется следующим образом: если признак X принимает значение x_1 , то вероятнее всего признак Y принимает значение y_1 ; если $X = x_2$, то вероятнее всего $Y = y_2$. Если эти коэффициенты принимают отрицательные значения, то связь интерпретируется следующим образом: если признак $X = x_1$, то вероятнее всего $Y = y_2$; если $X = x_2$, то вероятнее всего $Y = y_1$.

Для таблиц сопряженности $r \times s$ интерпретировать связь двух признаков можно на основе визуального представления таблиц сопряженности с помощью метода анализа соответствий.

1.3 Основные понятия анализа соответствий

Анализ соответствий – разведочный метод анализа, позволяющий визуально и численно исследовать структуру таблиц сопряженности большой размерности путем графического представления строк и столбцов таблицы в виде точек в пространстве низкой размерности с последующей интерпретацией карт соответствия.

Анализ соответствий позволяет категории признаков изобразить в одном пространстве, при этом позиции строк и столбцов на полученном графике соотносятся с силой и типом их связи в исходной таблице. В результате можно получить полезную информацию для интерпретации данных таблиц сопряженности.

Рассмотрим основные понятия, используемые в методах анализа соответствий [3].

1) Матрица соответствий P – это матрица относительных частот признаков X и Y , с элементами $\hat{p}_{ij} = \frac{n_{ij}}{n}$, $i = \overline{1, r}$, $j = \overline{1, s}$. Для удобства представим матрицу соответствий в виде таблицы относительных частот, рассчитав суммы по строкам и столбцам этой таблицы:

$x_i \backslash y_j$	y_1	y_2	...	y_s	\hat{p}_{i*}
x_1	\hat{p}_{11}	\hat{p}_{12}	...	\hat{p}_{1s}	\hat{p}_{1*}
x_2	\hat{p}_{21}	\hat{p}_{22}	...	\hat{p}_{2s}	\hat{p}_{2*}
...
x_r	\hat{p}_{r1}	\hat{p}_{r2}	...	\hat{p}_{rs}	\hat{p}_{r*}
\hat{p}_{*j}	\hat{p}_{*1}	\hat{p}_{*2}	...	\hat{p}_{*s}	1

Суммы относительных частот по столбцам $\hat{p}_{i*} = \sum_{j=1}^s \hat{p}_{ij}$, $i = \overline{1, r}$, называются

массами строк. Суммы относительных частот по строкам $\hat{p}_{*j} = \sum_{i=1}^r \hat{p}_{ij}$, $j = \overline{1, s}$,

называются массами столбцов.

2) Матрица профилей строк и матрица профилей столбцов. Матрица профилей

строк содержит элементы $\frac{\hat{p}_{ij}}{\hat{p}_{i*}}$ – оценка вероятности того, что элемент с i -ой

категорией признака X характеризуется j -ой категорией признака Y . Матрица

профилей столбцов содержит элементы $\frac{\hat{p}_{ij}}{\hat{p}_{*j}}$ – оценка вероятности того, что элемент

с j -ой категорией признака Y характеризуется i -ой категорией признака X .

3) Средний профиль строк, средний профиль столбцов. Вектор

$\hat{i}_{cp} = (\hat{p}_{*1}, \hat{p}_{*2}, \dots, \hat{p}_{*s})$, элементами которого являются массы столбцов, называется

средним профилем строк. Вектор $\hat{j}_{cp} = (\hat{p}_{1*}, \hat{p}_{2*}, \dots, \hat{p}_{r*})$, элементами которого

являются массы строк, называется средним профилем столбцов.

4) Расстояние между профилями. Различие профилей в анализе соответствий

измеряется с помощью расстояния χ^2 , учитывающего массу строк и столбцов.

Расстояние между профилями строк i_1 и i_2 рассчитывается по формуле:

$$d(i_1, i_2) = \sqrt{\sum_{j=1}^s \frac{\left(\frac{\hat{p}_{i_1j}}{\hat{p}_{i_1*}} - \frac{\hat{p}_{i_2j}}{\hat{p}_{i_2*}} \right)^2}{\hat{p}_{*j}}} . \quad (1.2)$$

Аналогичным образом рассчитывается расстояние между профилями столбцов, между профилем строки и средним профилем строк, между профилем столбца и средним профилем столбцов.

5) Инерция. Любой физический объект имеет центр тяжести, для каждой точки объекта можно рассчитать момент инерции. Для этого нужно массу элемента умножить на квадрат расстояния до центра тяжести. Если сложить моменты инерции всех точек объекта, получится общий момент инерции объекта. Эти понятия из физики используются в анализе соответствий: есть набор профилей (точек), у каждого профиля есть своя масса. Средний профиль аналогичен центру тяжести объектов. Расстояние между профилями – это расстояние между точками (объектами). Каждый профиль будет вносить свой вклад в общий момент инерции. Чтобы найти момент инерции профиля, нужно умножить массу профиля на квадрат расстояния до среднего профиля. Чем больше расстояние от профиля до среднего профиля, тем больший вклад в инерцию будет вносить этот профиль. Таким образом, инерция – мера разбросов профиля.

Инерция профиля i -ой строки рассчитывается по формуле:

$$In_i = p_{i*} \cdot d^2(i, i_{cp}). \quad (1.3)$$

Тогда общая инерция строк равна $\sum_{i=1}^r In_i$, относительная инерция профиля i -ой

строки равна $\frac{In_i}{\sum_{i=1}^r In_i}$.

Инерция профиля j -го столбца рассчитывается по формуле:

$$In_j = p_{*j} \cdot d^2(j, j_{cp}). \quad (1.4)$$

Тогда общая инерция столбцов равна $\sum_{j=1}^s \ln_j$, относительная инерция профиля

$$j\text{-го столбца равна } \frac{\ln_j}{\sum_{j=1}^s \ln_j}.$$

1.4 Алгоритм простого анализа соответствий

1) Рассчитываются матрица соответствий P , массы строк (\hat{p}_{i*}) и столбцов (\hat{p}_{*j}) , $i = \overline{1, r}, j = \overline{1, s}$.

2) Строятся диагональные матрицы D_r и D_s :

$$D_r = \begin{pmatrix} \frac{1}{\sqrt{p_{1*}}} & 0 & \dots & 0 \\ 0 & \frac{1}{\sqrt{p_{2*}}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{\sqrt{p_{r*}}} \end{pmatrix}, D_s = \begin{pmatrix} \frac{1}{\sqrt{p_{*1}}} & 0 & \dots & 0 \\ 0 & \frac{1}{\sqrt{p_{*2}}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{\sqrt{p_{*s}}} \end{pmatrix}. \quad (1.5)$$

3) Рассчитывается матрица стандартизованных остатков Z и осуществляется её сингулярное разложение:

$$Z = D_r P D_s - D_r^{-1} 1_r 1_s^T D_s^{-1} = U \Lambda V^T, \quad p = \min(r, s), \quad (1.6)$$

где U – матрица из собственных векторов матрицы $Z \cdot Z^T$;

V – матрица из собственных векторов матрицы $Z^T \cdot Z$;

Λ – диагональная матрица с сингулярными числами равными корню

квадратному из собственных чисел матрицы $Z \cdot Z^T$ или $Z^T \cdot Z$, упорядоченных по убыванию;

$$\mathbf{1}_s = (\mathbf{1}, \dots, \mathbf{1})^T, \mathbf{1}_r = (\mathbf{1}, \dots, \mathbf{1})^T.$$

$s \times 1$ $r \times 1$

Элементы матрицы стандартизованных остатков z_{ij} имеют вид:

$$\begin{aligned} z_{ij} &= \frac{\mathbf{f}_{ij}}{\sqrt{\mathbf{f}_{i*} \mathbf{f}_{*j}}} - \sqrt{\mathbf{f}_{i*} \mathbf{f}_{*j}} = \frac{n_{ij}}{\sqrt{n_{i*} n_{*j}}} - \frac{\sqrt{n_{i*} n_{*j}}}{n} = \\ &= \frac{n_{ij} - \frac{n_{i*} n_{*j}}{n}}{\sqrt{n_{i*} n_{*j}}} = \sqrt{\frac{1}{n} \frac{\left(n_{ij} - \frac{n_{i*} n_{*j}}{n} \right)^2}{\frac{n_{i*} n_{*j}}{n}}}. \end{aligned} \quad (1.7)$$

Квадраты элементов матрицы стандартизованных остатков представляют собой элементы статистики χ^2 Пирсона проверки гипотезы о независимости признаков. Значение статистики χ^2 можно рассчитать следующим образом:

$$\chi^2 = n \sum_{i=1}^r \sum_{j=1}^s z_{ij}^2. \quad (1.8)$$

Анализ соответствий рассматривается как декомпозицию статистики χ^2 , обеспечивающую 2 вида шкалирования: для категорий признака X (строк таблицы сопряженности) и категорий признака Y (столбцов таблицы сопряженности).

4) Расчет общей инерции таблицы сопряженности. Общая инерция равна сумме квадратов элементов матрицы стандартизованных остатков:

$$\text{In} = \sum_{i=1}^r \sum_{j=1}^s z_{ij}^2 = \frac{\chi^2}{n}. \quad (1.9)$$

Квадраты сингулярных чисел матрицы Z (собственные числа матриц $Z \cdot Z^T$ или $Z^T \cdot Z$, обозначим их $\lambda_1, \lambda_2, \dots, \lambda_p$) также разлагают общую инерцию на составные части. Их называют также главными инерциями осей и выражают в процентах от общей инерции. Выбор размерности осуществляется по доле вклада собственных чисел или доле инерции строк ($r < s$) и столбцов ($r > s$) в общую инерцию. Максимальная размерность признакового пространства: $\min(r, s) - 1$.

5) Рассчитываются координаты строк $F = D_r U \Lambda$ и координаты столбцов $G = D_s V \Lambda$ таблицы сопряженности. Получив координаты точек, можно изобразить их графически, построив карты соответствия строк и столбцов таблицы сопряженности.

6) Расчет показателей качества решения: абсолютных и относительных вкладов. Абсолютный вклад i -ой строки в главную инерцию k -ой главной оси

рассчитывается по формуле: $\frac{\hat{p}_{i*} \cdot f_{ik}^2}{\lambda_k}$, $i = \overline{1, r}$, $k = \overline{1, p-1}$. Этот показатель

характеризует, в какой степени геометрическая ориентация k -ой оси определяется i -ой категорией признака X . Эта информация используется при интерпретации осей: строки с высокими абсолютными вкладами являются наиболее важными в контексте данной оси. Сумма абсолютных вкладов по каждой оси равна 1 (100%). Наиболее существенными для интерпретации оси являются точки (строки), чей вклад выше среднего, определяемого как единица, деленная на количество строк.

Относительный вклад k -ой главной оси в общую инерцию i -ой строки

рассчитывается по формуле: $\frac{\hat{p}_{i*} \cdot f_{ik}^2}{\sum_{k=1}^{p-1} \hat{p}_{i*} \cdot f_{ik}^2}$. Общую инерцию i -ой строки (знаменатель

последней дроби) можно рассчитать также как сумму квадратов элементов i -ой

строки матрицы стандартизованных остатков $\sum_{j=1}^s z_{ij}^2$. Относительный вклад k -ой

главной оси в общую инерцию i -ой строки представляет собой квадрат

коэффициента корреляции i -ой строки и k -ой главной оси. Геометрически это квадрат косинуса угла между профилем i -ой строки и k -ой главной осью. Относительный вклад точки не зависит от её массы и показывает, насколько хорошо точка объясняется выбранной осью. Если относительный вклад оси в инерцию точки высокий, то угол между вектором точки и этой осью мал. В этом смысле точка «лежит» в направлении оси, что предполагает высокую корреляцию между вектором точки и осью.

Корни квадратные из относительных вкладов представляют собой корреляцию между профилями и главными осями. При этом знак коэффициента корреляции должен соответствовать знаку соответствующей координаты.

Для характеристики качества представления каждой строки в пространстве сниженной размерности p^* рассчитывается мера качества представления каждой

строки $\frac{\sum_{k=1}^{p^*} p_{i*} \cdot f_{ik}^2}{\sum_{j=1}^s z_{ij}^2}$, $i = \overline{1, r}$. Геометрически показатель качества представляет собой

косинус угла между точкой и плоскостью (подпространством), образованной p^* главными осями. При выборе количества осей можно использовать следующие критерии:

- формально статистический критерий (критерий Кайзера): количество осей равно количеству собственных чисел, превышающих 1 или среднее значение по всем собственным числам);
- критерий достижения требуемого уровня объясненной инерции (например, не менее 70);
- критерий «каменистой осыпи»: количество осей равно количеству точек на графике собственных чисел, расположенных до момента резкого снижения графика.

Аналогичным образом осуществляется анализ результатов и оценка качества решения для столбцов таблицы сопряженности.

1.5 Анализ многомерных данных и множественный анализ соответствий

Простой анализ соответствий может использоваться, если качественных признаков больше двух. В этом случае используются следующие способы построения исходной матрицы данных:

- 1) таблица множественных ответов;
- 2) «составленные» таблицы;
- 3) индикаторная матрица и матрица Бёрта.

Для построения таблицы множественных ответов каждому респонденту предлагается заполнить таблицу, в столбцах которой представлены объекты (торговые бренды, продукты и т.п.), а в строках – характеристики или атрибуты. На пересечении i -ой строки и j -го столбца ставится единица, если j -ый объект обладает i -ой характеристикой. При этом j -ый объект может обладать несколькими или даже всеми атрибутами. Суммируя такие таблицы по всем респондентам, получают сводную таблицу множественных ответов.

Если в анализе участвуют k номинальных признаков X_1, X_2, \dots, X_k , из которых один, например, X_k является результирующим, то можно построить «составленную» таблицу. Количество столбцов этой таблицы будет равно количеству возможных значений признака X_k , а количество строк будет равно сумме возможных значений всех остальных (факторных) признаков. В результате получают $k-1$ двухфакторных таблиц сопряженности, составленных друг под другом.

Для построения матрицы Бёрта используется бинарная (индикаторная) матрица D размерности $n \times k$, где n – число респондентов (объектов), k – количество возможных значений всех признаков. Элемент матрицы $d_{ij} = 1$, если i -ый объект обладает j -ым свойством. Тогда матрица Бёрта рассчитывается по формуле $D^T \cdot D$. Элемент с индексами i и j матрицы Бёрта показывает число наблюдений, одновременно обладающих i -ым и j -ым свойствами.

Классическое определение множественного анализа соответствий подразумевает использование простого анализа соответствий к индикаторной матрице D . Однако, учитывая связь сингулярных чисел и сингулярных векторов

матрицы D и матрицы Бёрта, множественных анализ соответствий можно определить как простой анализ соответствий матрицы Бёрта.

2 Описание лабораторной работы

Лабораторная работа включает в себя следующие этапы:

- постановку задачи;
- ознакомление с порядком выполнения работы;
- выполнение расчетов индивидуальных задач на компьютере и анализ результатов;
- подготовку письменного отчета с выводами по работе;
- защиту лабораторной работы.

3 Постановка задачи

1) В соответствии с темой (темы по вариантам приведены в приложении А) разработать анкету для анализа рынка товаров (услуг).

2) Провести анкетирование не менее 50 респондентов.

3) Провести предварительную обработку результатов анкетирования.

4) Сформировать таблицу множественных ответов об основных игроках на рынке.

5) Для построенной таблицы провести простой анализ соответствий. Дать название главным осям и провести содержательную интерпретацию построенной карты восприятия компаний/брендов.

Дадим комментарии по ответу на вопрос № 7. Респондентам предлагают отметить характеристики (атрибуты), присущие основным Интернет-провайдерам Оренбургской области. Опрашиваемый может выбрать несколько атрибутов для одного провайдера, может выбрать все атрибуты, может не выбрать ни один из атрибутов. В результате ответа на этот вопрос формируется таблица множественных ответов, которая будет использована для анализа конкурентной среды.

Проанализируем собранные результаты. Рассмотрим, каким способом респонденты пользуются для выхода в Интернет (вопрос № 9). Гистограмма распределения респондентов по ответу на вопрос № 9 представлена на рисунке 4.2.

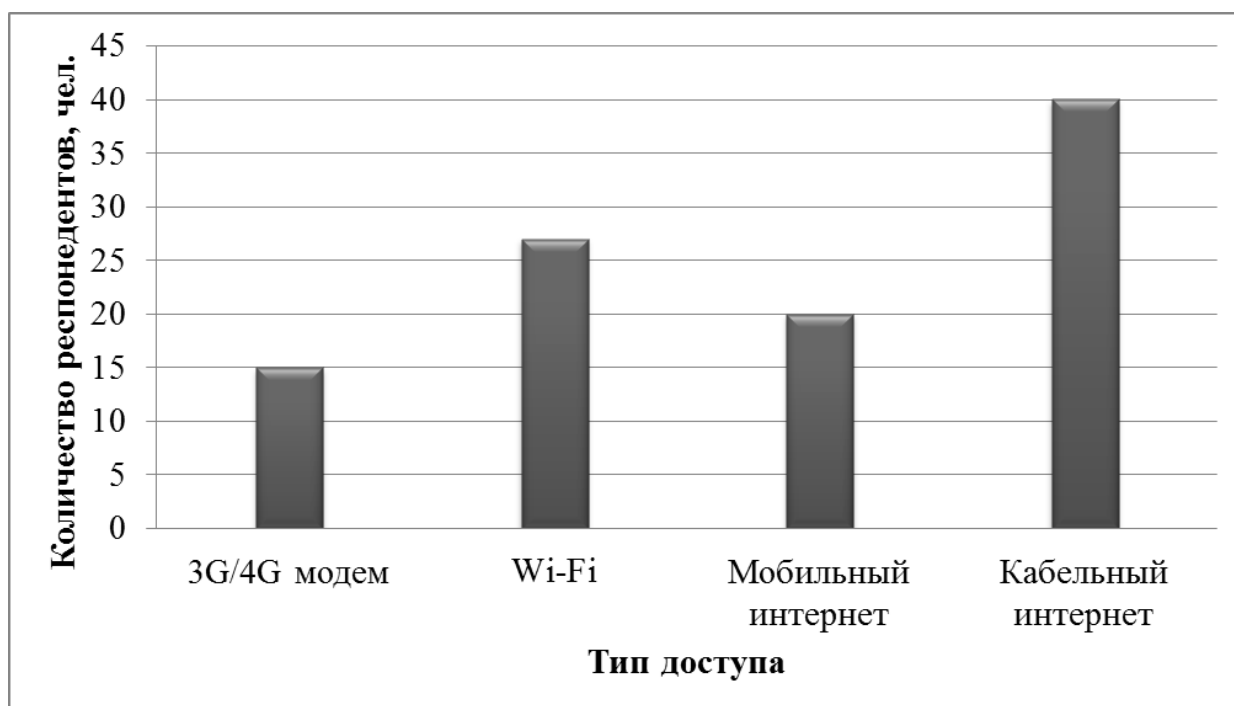


Рисунок 4.2 – Распределение респондентов по способу доступа к сети Интернет

Из рисунка 4.2 следует, что наиболее популярным является кабельное подключение, а наименее популярным – 3G/4G модемы. При этом один пользователь может использовать одновременно несколько устройств и способов подключения. Диаграмма распределения респондентов по числу используемых устройств представлена на рисунке 4.3 (вопрос № 8).

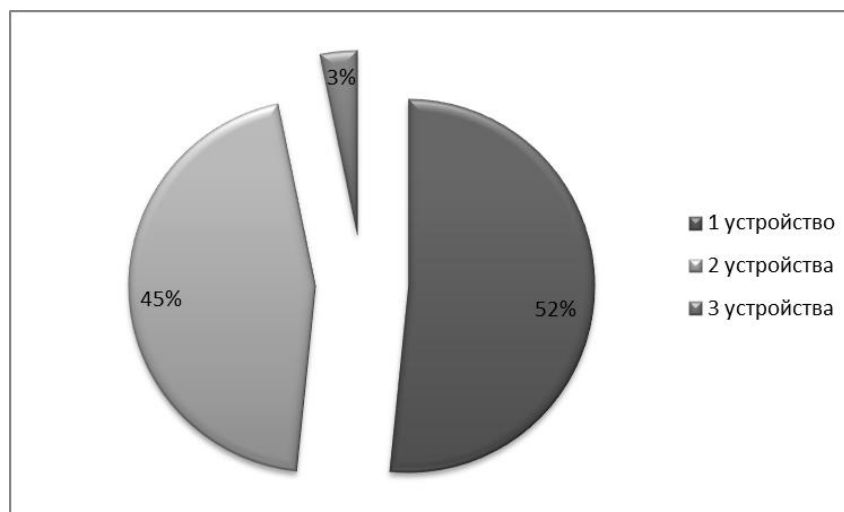


Рисунок 4.3 – Диаграмма распределения респондентов по количеству используемых устройств для доступа к сети Интернет

Рассмотрим распределение респондентов в зависимости от возраста (вопрос №1) и используемого устройства для выхода в сеть Интернет (вопрос № 8) (рисунок 4.4).

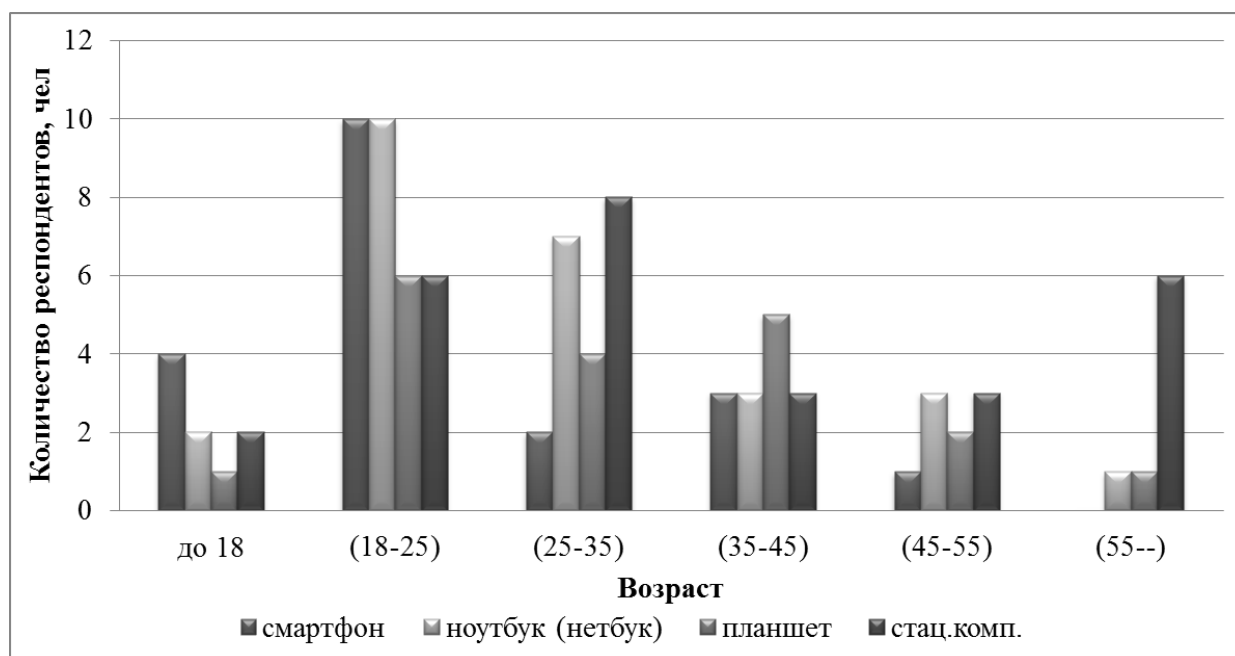


Рисунок 4.4 – Распределение респондентов в зависимости от возраста и используемого устройства для выхода в сеть Интернет

На основании графика можно сказать, что респонденты в возрасте до 18 лет чаще используют смартфоны для выхода в Интернет; респонденты в возрасте от 18 до 25 лет – чаще используют смартфоны и ноутбуки, предпочитая мобильность; респонденты в возрасте от 25 до 35 лет – стационарные компьютеры; респонденты в возрасте от 35 до 45 лет – планшеты; респонденты в возрасте от 45 до 55 – ноутбуки и стационарные компьютеры; респонденты в возрасте от 55 лет – стационарные компьютеры. Проверим, является ли эта связь значимой, для этого проведем анализ таблицы сопряженности (таблица 4.1).

Таблица 4.1 – Таблица сопряженности признаков «Возраст» и «Используемое устройство для выхода в Интернет» (человек)

Возрастная группа	Устройство, используемое для выхода в Интернет			
	смартфон	ноутбук (нетбук)	планшет	стац. компьютер
до 18 лет	4	2	1	2
18-25 лет	10	10	6	6
25-35 лет	2	7	4	8
35-45 лет	3	3	5	3
45-55 лет	1	3	2	3
старше 55 лет	0	1	1	6

Для проверки гипотезы о независимости признаков воспользуемся критерием согласия Пирсона χ^2 . Для этого рассчитаем маргинальные и теоретические частоты. Результаты расчетов приведены на рисунке 4.5.

	A	B	C	D	E	F	G
1	Исходные данные	смартфон	ноутбук (нетбук)	планшет	стац. комп.	Сумма	
2	до 18	4	2	1	2	9	
3	(18-25)	10	10	6	6	32	
4	(25-35)	2	7	4	8	21	
5	(35-45)	3	3	5	3	14	
6	(45-55)	1	3	2	3	9	
7	(55--)	0	1	1	6	8	
8	Сумма	20	26	19	28	93	
11	теоретические частоты	смартфон	ноутбук (нетбук)	планшет	стац. комп.		
12	до 18	=B\$8*\$F2/\$F\$8	2,516129032	1,838709677	2,7096774		
13	(18-25)	6,88172043	8,946236559	6,537634409	9,6344086		
14	(25-35)	4,516129032	5,870967742	4,290322581	6,3225806		
15	(35-45)	3,010752688	3,913978495	2,860215054	4,2150538		
16	(45-55)	1,935483871	2,516129032	1,838709677	2,7096774		
17	(55--)	1,720430108	2,23655914	1,634408602	2,4086022		

Рисунок 4.5 – Вспомогательные расчеты в MS Excel

Рассчитаем наблюдаемое значение статистики. Результаты расчетов приведены на рисунке 4.6. Для нахождения критического значения воспользуемся встроенной функцией MS Excel ХИ2ОБР(0,05; 15), где первый аргумент это уровень значимости, второй – число степеней свободы. Критическое значение составило 7,84. Наблюдаемое значение статистики 18,67 больше критического, следовательно нулевая гипотеза отвергается, рассматриваемые признаки зависимы. Таким образом, возраст оказывает значимое влияние на выбор устройства для доступа в Интернет.

	А	В	С	Д	Е	Ф	Г	Н
1	Исходные данные	смартфон	ноутбук (нетбук)	планшет	стац.комп.	Сумма		
2	до 18	4		2	1	2		9
3	(18-25)	10		10	6	6		32
4	(25-35)	2		7	4	8		21
5	(35-45)	3		3	5	3		14
6	(45-55)	1		3	2	3		9
7	(55--)	0		1	1	6		8
8	Сумма	20		26	19	28		93
9								
10								
11	теоретические частоты	смартфон	ноутбук (нетбук)	планшет	стац.комп.			
12	до 18	1,935483871	2,516129032	1,838709677	2,7096774			
13	(18-25)	6,88172043	8,946236559	6,537634409	9,6344086			
14	(25-35)	4,516129032	5,870967742	4,290322581	6,3225806			
15	(35-45)	3,010752688	3,913978495	2,860215054	4,2150538			
16	(45-55)	1,935483871	2,516129032	1,838709677	2,7096774			
17	(55--)	1,720430108	2,23655914	1,634408602	2,4086022			
18								
19	Критерий Пирсона	смартфон	ноутбук (нетбук)	планшет	стац.комп.			
20	до 18	2,202150538	0,105872622	0,382569327	0,1858679			
21	(18-25)	1,41297043	0,124121175	0,044213356	1,3710157			
22	(25-35)	=((B4-B14)^2)/B14		0,019645889	0,4450296			
23	(35-45)	3,84025E-05	0,213429044	1,600816558	0,3502578			
24	(45-55)	0,452150538	0,093052109	0,014148274	0,031106			
25	(55--)	1,720430108	0,683674524	0,246250707	5,3550307			
26						18,67280633	Наблюдаемое значение	
27								
28								

Рисунок 4.6 – Результаты расчета наблюдаемого значения статистики Пирсона

Рассмотрим распределение респондентов по ответу на вопрос № 10 «Услугами каких Интернет-провайдеров пользуетесь для выхода в Интернет?».

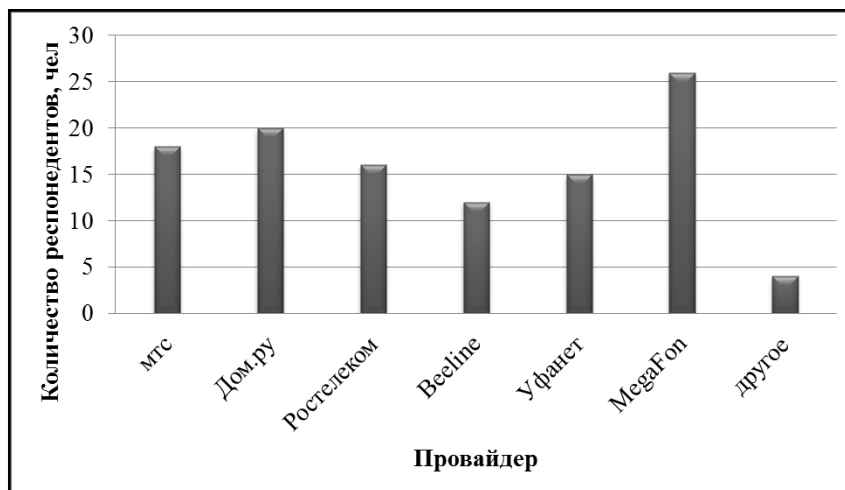


Рисунок 4.7 – Гистограмма распределения респондентов по ответу на вопрос № 10

По рисунку 4.7 можно сделать вывод, что наиболее часто используемым провайдером является Мегафон, далее, по степени убывания, – это Дом.ру, МТС, Ростелеком, Уфанет, Билайн и другие провайдеры.

4.2 Реализация простого анализа соответствий в ППП STASTICA

Рассмотрим данные о соответствии восьми имиджевых характеристик рассматриваемых провайдеров. Каждый респондент отмечал, соответствует та или иная характеристика провайдеру, после чего результаты по всем респондентам были просуммированы. Итоговая таблица множественных ответов представлена на рисунке 4.8.

	Высокая скорость	Высокая цена	Качество подключения	Низкая цена тарифа	Высокий сервис	Мобильность	Широкий спектр тарифов	Хорошая зона покрытия
МТС	15	1	9	20	9	19	16	13
Домру	20	7	21	12	18	9	14	17
Ростелеком	12	18	16	8	11	5	9	13
Билайн	9	5	11	5	7	10	6	7
Уфанет	14	5	11	8	9	6	7	8
Мегафон	16	9	12	8	12	17	9	9

Рисунок 4.8 – Таблица множественных ответов

Введем указанные данные в ППП STATISTICA, при этом переименуем переменные (столбцы-var) и наблюдения (строчки-case) согласно введенным обозначениям. Результаты представлены на рисунке 4.9.

	1	2	3	4	5	6	7	8
	Высокая скорость	Высокая цена	Качество подключения	Низкая цена тарифа	Высокий уровень сервиса	Мобильность	Широкий спектр тарифов	Хорошая зона покрытия
МТС	15	1	9	20	9	19	16	13
Домру	20	7	21	12	18	9	14	17
Ростелеком	12	18	16	8	11	5	9	13
Билайн	9	5	11	5	7	10	6	7
Уфанет	14	5	11	8	9	6	7	8
Мегафон	16	9	12	8	12	17	9	9

Рисунок 4.9 – Исходные данные в ППП STATISTICA

Далее вызываем модуль анализа соответствий: Statistics-Multivariate Exploratory Techniques-Correspondence Analysis (рисунок 4.10).

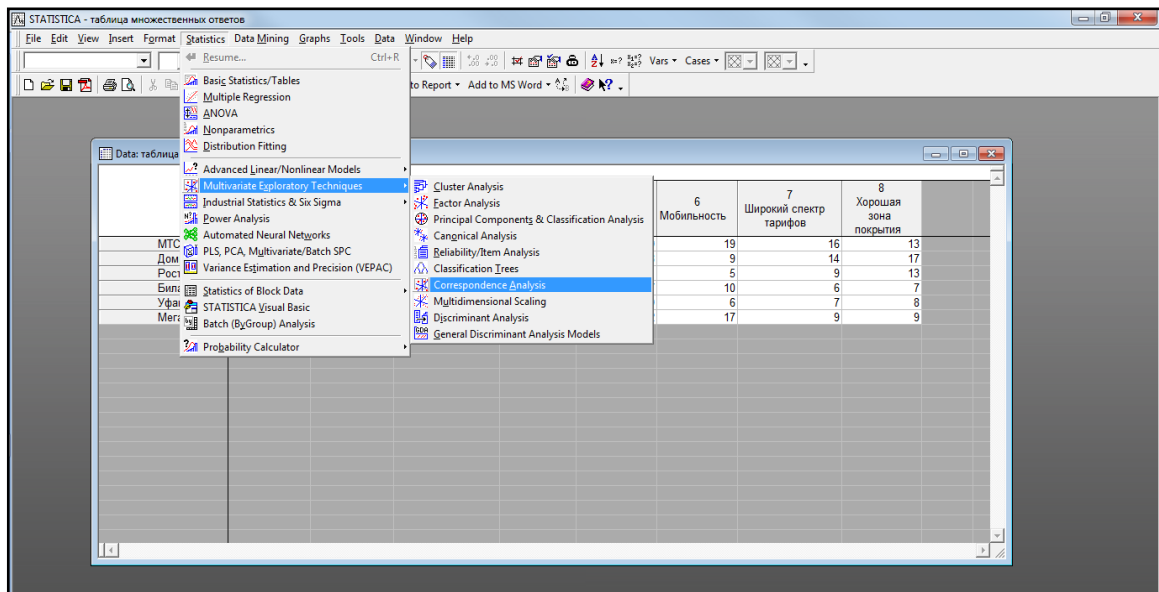


Рисунок 4.10 – Вызов модуля «Анализ соответствий» ППП STATISTICA

Работаем с простым анализом, используем первую вкладку. Выбираем в качестве исходных данных таблицу частот («Frequencies w/out grouping vars») и нажимаем кнопку «Variables with frequencies» (рисунок 4.11).

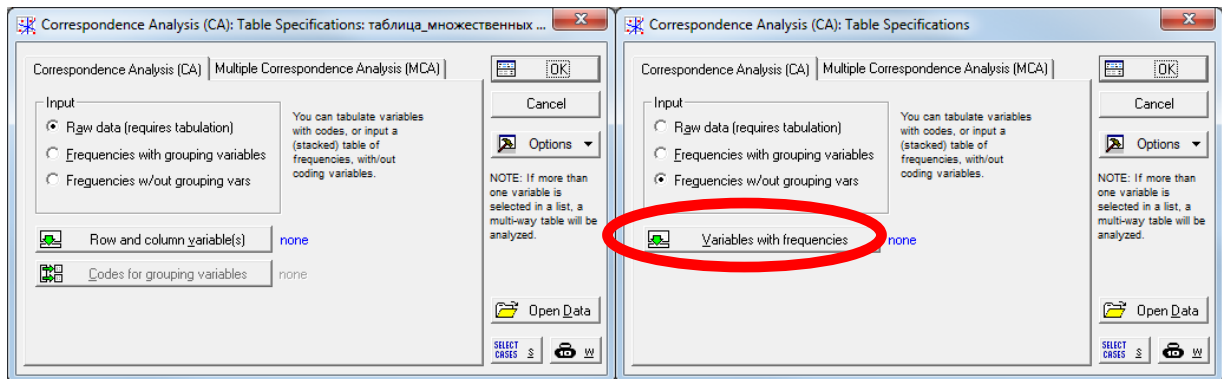


Рисунок 4.11 – Модуль простого анализа соответствий

В появившемся окне выбираем переменные, которые содержат таблицу множественных ответов и нажимаем «Ok».

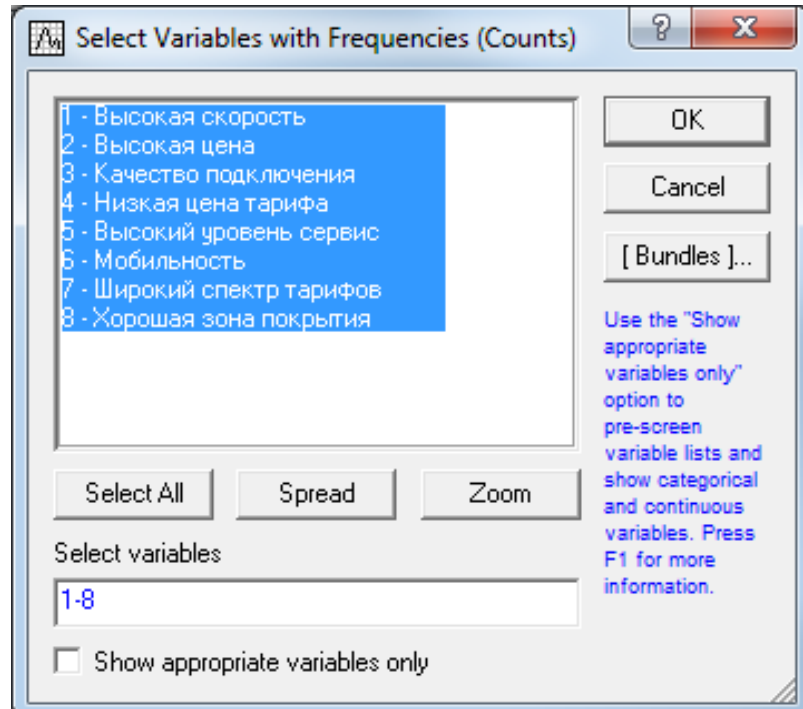


Рисунок 4.12 – Выбор признаков

Вернувшись в форму на рисунке 4.11 нажимаем кнопку «Ok». Переходим к форме с результатами, на которой выбираем вкладку «Advanced» (рисунок 4.13).

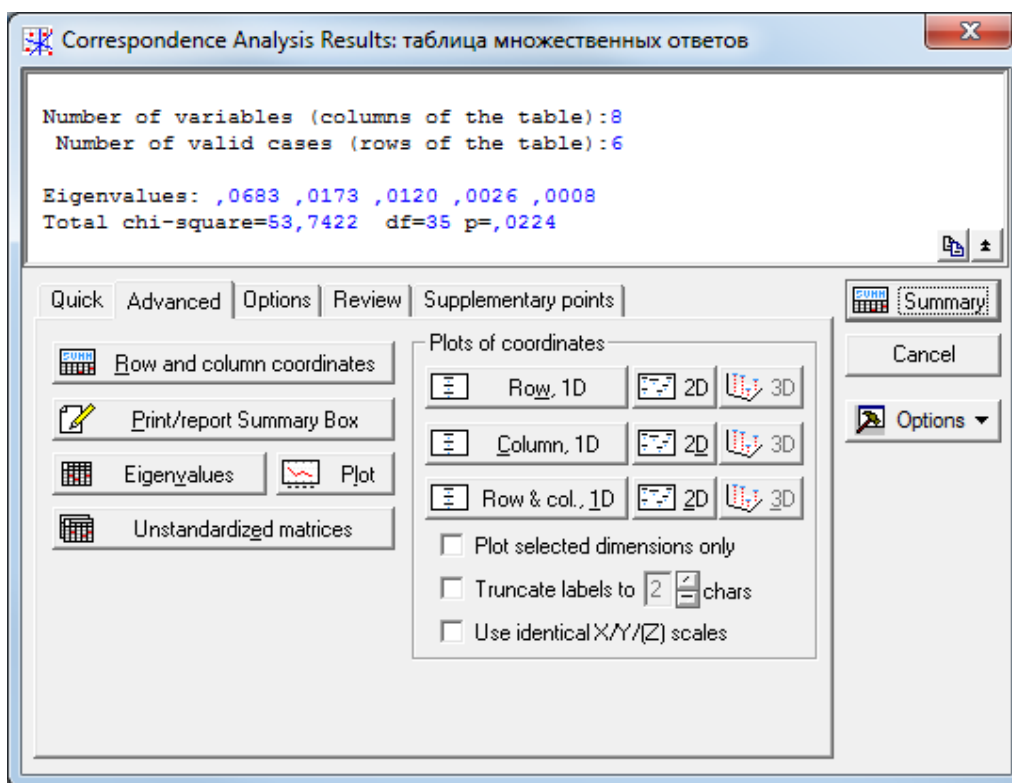


Рисунок 4.13 – Форма с результатами анализа соответствий (вкладка «Advanced»)

Проанализируем сингулярные числа, количество главных осей и их вклад в общую инерцию. Для просмотра сингулярных чисел и вклада каждой из осей в общую инерцию нажмем кнопку «Eigenvalues», результаты будут сохраняться в новой рабочей книге, как представлено на рисунке 4.14. Таким образом, видим, что было получено 5 главных осей, при этом первый две объясняют 84,75% общей инерции.

Number of Dims.	Singular Values	Eigen-Values	Perc. of Inertia	Cumulativ Percent	Chi Squares
1	0,261346	0,068302	67,61281	67,6128	36,33664
2	0,131558	0,017308	17,13293	84,7457	9,20762
3	0,109552	0,012002	11,88052	96,6263	6,38486
4	0,050729	0,002573	2,54749	99,1738	1,36908
5	0,028891	0,000835	0,82624	100,0000	0,44404

Рисунок 4.14 – Оценки сингулярных чисел и главной инерции осей

Для подтверждения предположения о достаточности двух главных осей рассмотрим график «каменистой осыпи», для его вывода необходимо нажать кнопку «Plot». Согласно рисунку 4.15 относительно резкий скачок происходит при переходе от двух осей к трем, что указывает на целесообразность выбора двумерного решения. Для просмотра полученной карты нажимаем на кнопку «Rows&col»- «2D» на форме (рисунке 4.11). Результаты построения карты приведены на рисунке 4.16.

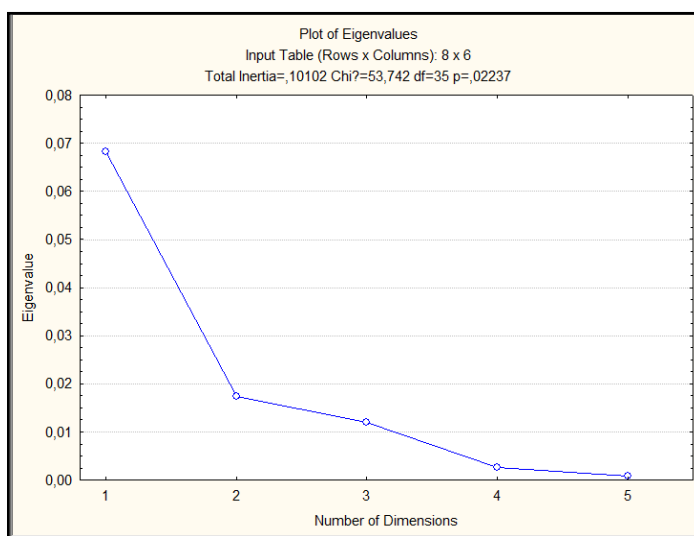


Рисунок 4.15 – График «каменистой осыпи»

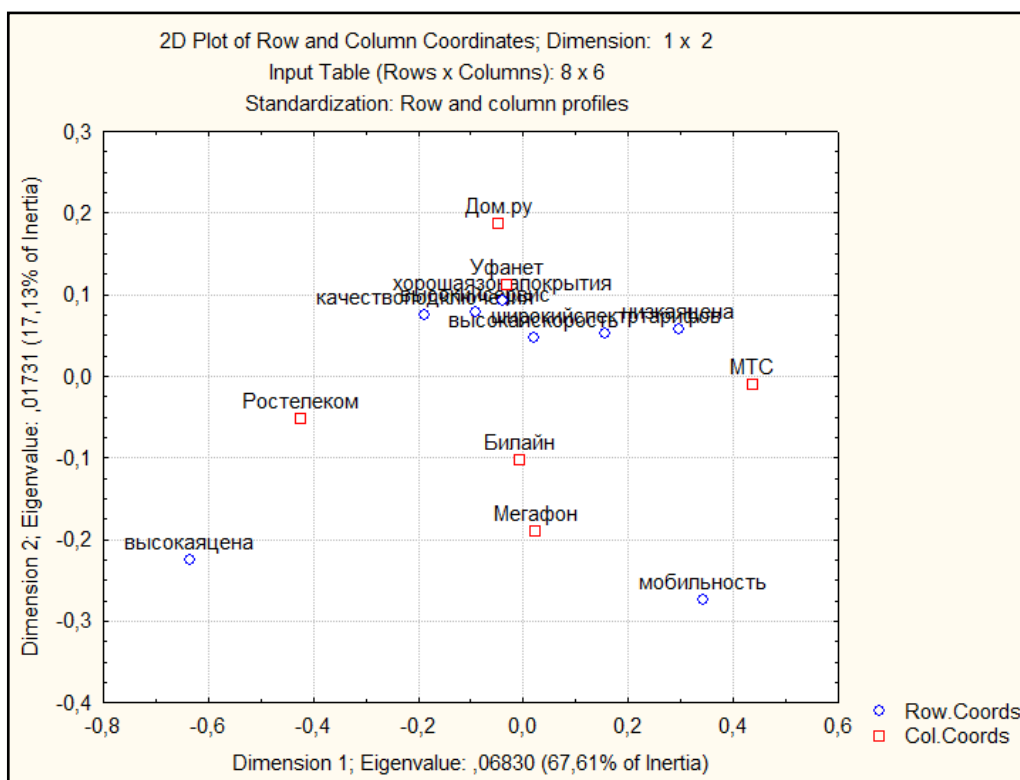


Рисунок 4.16 – Карта соответствий в пространстве двух главных осей

Для того чтобы получить наиболее корректную и последовательную интерпретацию результатов анализа соответствий, необходимо вначале определить, какие точки мы будем считать описываемыми, а какие – описывающими. Иначе говоря, будем ли мы рассматривать положение провайдеров в пространстве характеристик, или положение характеристик в пространстве провайдеров. Будет целесообразным расположить провайдеры в пространстве характеристик, предварительно интерпретировав оси и дав им названия.

Построим карту расположения провайдеров в пространстве характеристик, образованных двумя главными осями. Для построения карты расположения провайдеров в пространстве характеристик, образованном двумя главными осями, необходимо на вкладке «Advanced» нажать кнопку «Column»-«2D». Результаты представлены на рисунке 4.17.

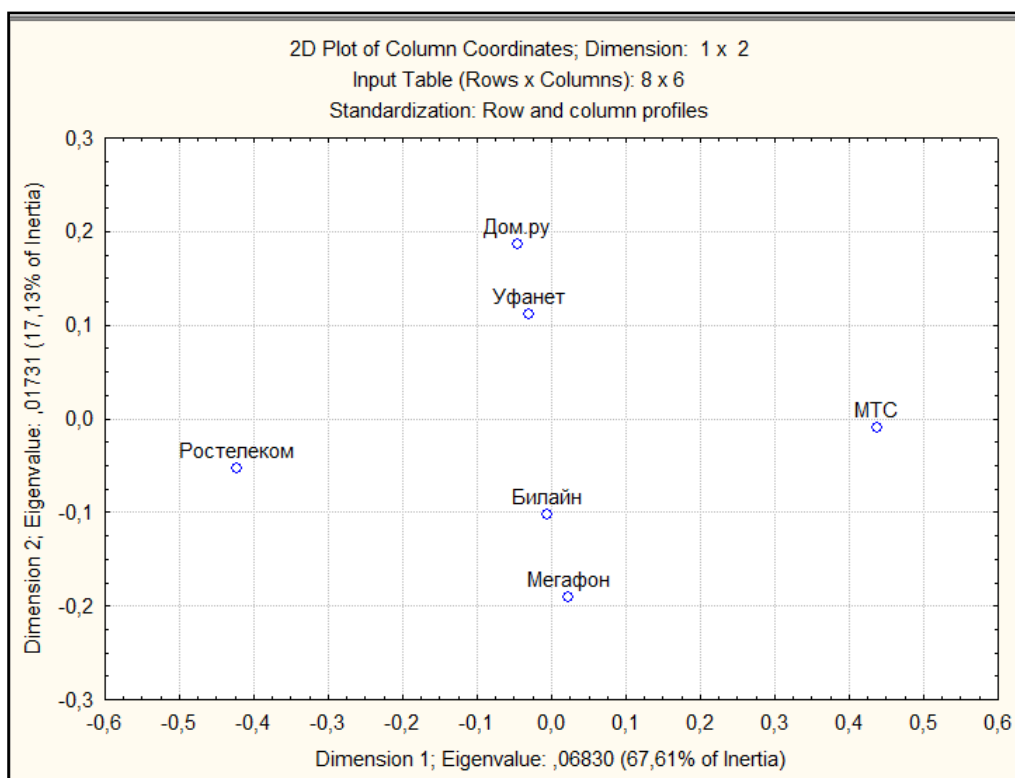


Рисунок 4.17 – расположение провайдеров в пространстве характеристик

Интерпретация осей осуществляется на основе вкладов, которые каждая точка вносит в инерцию рассматриваемых осей (абсолютных вкладов). Для того чтобы оценить вклады, нажмем на форме результатов кнопку «Summary». В рабочей книге появятся две таблицы, представленные на рисунках 4.18 и 4.19.

Column Coordinates and Contributions to Inertia (Spreadsheet1)										
Input Table (Rows x Columns): 8 x 6										
Standardization: Row and column profiles										
Column Name	Column Number	Coordin. Dim. 1	Coordin. Dim. 2	Mass	Quality	Relative Inertia	Inertia Dim. 1	Cosine? Dim. 1	Inertia Dim. 2	Cosine? Dim. 2
МТС	1	0,437111	-0,010218	0,191729	0,919012	0,394806	0,536338	0,918510	0,001157	0,000502
Дом.ру	2	-0,045551	0,186716	0,221805	0,876345	0,092547	0,006738	0,049226	0,446786	0,827119
Ростелеком	3	-0,423347	-0,052972	0,172932	0,883959	0,352516	0,453769	0,870333	0,028037	0,013627
Билайн	4	-0,005787	-0,102463	0,112782	0,307309	0,038263	0,000055	0,000977	0,068413	0,306332
Уфанет	5	-0,030323	0,111534	0,127820	0,492420	0,034327	0,001721	0,033892	0,091870	0,458527
Мегафон	6	0,023334	-0,190798	0,172932	0,722534	0,087540	0,001379	0,010648	0,363737	0,711886

Рисунок 4.18 – Основная статистика для столбцов

		Row Coordinates and Contributions to Inertia (Spreadsheet1)								
		Input Table (Rows x Columns): 8 x 6								
		Standardization: Row and column profiles								
Row Name	Row Number	Coordin. Dim.1	Coordin. Dim.2	Mass	Quality	Relative Inertia	Inertia Dim.1	Cosine? Dim.1	Inertia Dim.2	Cosine? Dim.2
высокая скорость	1	0,020566	0,047016	0,161654	0,133793	0,031497	0,001001	0,021488	0,020646	0,112305
высокая цена	2	-0,635387	-0,224409	0,084586	0,957475	0,397098	0,499970	0,851286	0,246120	0,106189
качество подключения	3	-0,187171	0,075227	0,150376	0,758680	0,079841	0,077130	0,653170	0,049168	0,105510
низкая цена	4	0,296321	0,058073	0,114662	0,711208	0,145516	0,147404	0,684903	0,022342	0,026306
высокий сервис	5	-0,091376	0,078687	0,124060	0,495710	0,036025	0,015166	0,284636	0,044382	0,211074
мобильность	6	0,344098	-0,273821	0,124060	0,961494	0,246999	0,215061	0,588703	0,537439	0,372791
широкий спектр тарифов	7	0,157380	0,052657	0,114662	0,746877	0,041855	0,041580	0,671683	0,018369	0,075193
хорошая зона покрытия	8	-0,038178	0,091958	0,125940	0,583848	0,021169	0,002688	0,085840	0,061533	0,498008

Рисунок 4.19 – Основная статистика для строк

Проведем на основе рисунка 4.18 содержательную интерпретацию осей. Мы имеем 6 высказываний, и, следовательно, средний абсолютных вклад для строк будет равен $1/8 = 0,125$. Таким образом, в качестве точек, в наибольшей степени определяющих геометрическую ориентацию осей, будем рассматривать те точки, абсолютный вклад которых превышает порог 0,125.

Применительно к горизонтальной оси это точки (в порядке убывания значения вклада): «высокая цена» (абсолютный вклад 0,5), «мобильность» (0,215), «низкая цена» (0,147). При этом первая из названных точек имеет отрицательную координату по первой оси, а вторая и третья положительные координаты (понятия «положительный» и «отрицательный» не несут никакой содержательной нагрузки, а используются для обозначения направления осей). Очевидно, что горизонтальную ось можно определить как «высокая цена» – «низкая цена». При этом в нашем случае геометрически противоположные полюса являются и логически противоположными.

Аналогичные рассуждения проведем относительно вертикальной оси. Наибольший вклад характерен для точки «мобильность» (абсолютный вклад 0,53). Кроме того, при интерпретации вертикальной оси следует также принять во внимание положение точки «хорошая зона покрытия», которая, хотя и имеет невысокий абсолютный вклад (всего лишь 0,062), но коррелирует с осью (квадрат корреляции равен 0,5) и располагается в направлении положительного полюса.

В итоге, с помощью анализа соответствий определили два ключевых фактора, определяющих основные особенности провайдеров: «высокая цена» – «низкая цена»

и «мобильность – хорошая зона покрытия». Эти два фактора образуют пространство, в котором теперь можно изобразить провайдеров, рисунок 4.20.

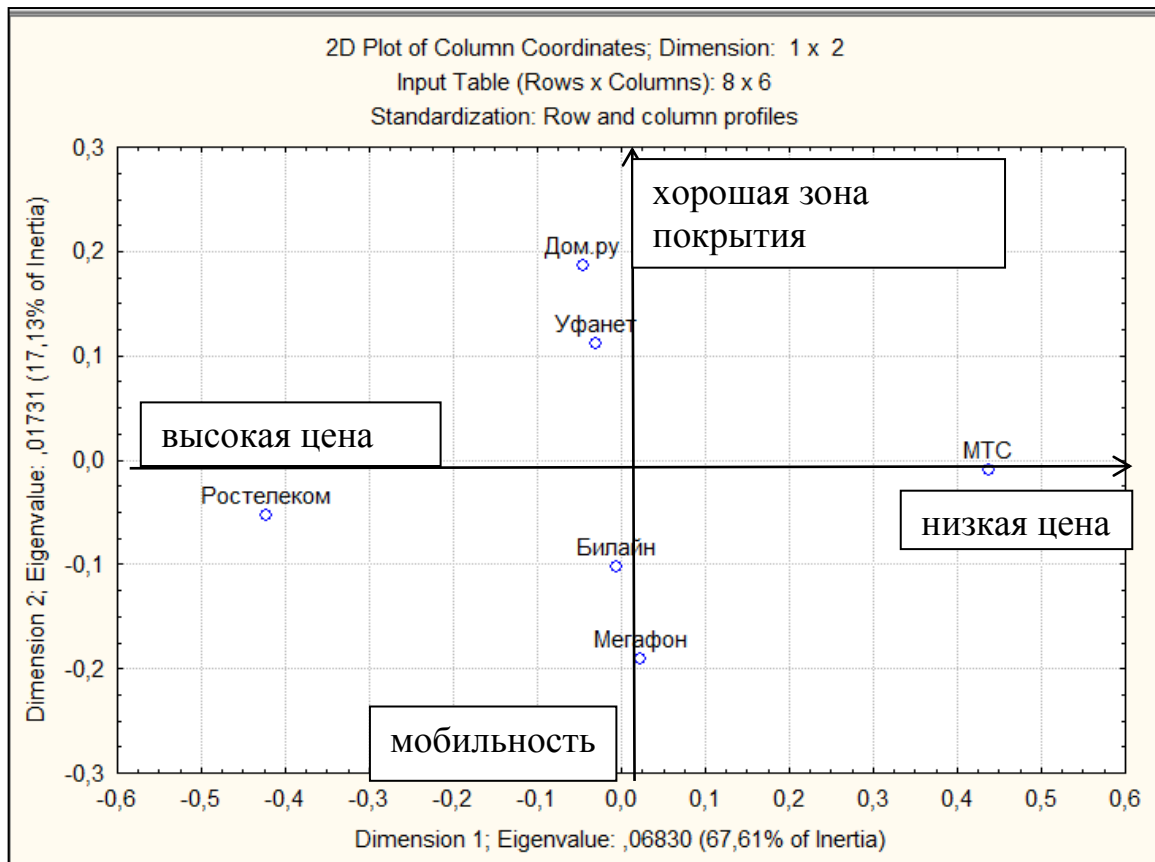


Рисунок 4.20 – Расположение провайдеров в пространстве характеристик

Начнем интерпретацию положения провайдеров с точки «МТС». Это – единственная точка, расположенная в правой полуплоскости; все остальные находятся слева от центра (кроме «Мегафон»). «МТС» воспринимается как самый выгодный провайдер. При этом, что примечательно, относительно МТС нельзя сказать, он ближе к полюсу «хорошая зона покрытия» или же к полюсу «мобильность», потому что точка практически лежит на горизонтальной оси, и, следовательно, ее координата по вертикальной оси и относительный вклад близки к нулю. Это означает, что для МТС в одинаковой степени характерны атрибуты и верхнего, и нижнего полюса; или же в одинаковой степени не характерны ни те, ни другие. Фактически, в контексте вертикальной оси характеристики «МТС» мало чем отличаются от средних показателей по всем провайдерам. Таким образом, он имеет «хорошую зону покрытия» и «мобильность» ровно настолько, насколько все

провайдеры в среднем; следовательно, по этому фактору он ничем от среднего не отличается.

Далее, чем левее расположена точка, тем в большей степени соответствующий ей провайдер воспринимается как дорогостоящий. Так, «Ростелеком», находящийся левее начала координат и достаточно тесно коррелирующий с «отрицательным» направлением горизонтальной оси, является наиболее дорогостоящим из рассматриваемых провайдеров. Также этот провайдер обладает мобильностью, так как расположен ниже горизонтальной оси.

Далее рассмотрим провайдеров «Дом.ру» и «Уфанет». Они расположены немного левее вертикальной оси, что свидетельствует о том, что эти провайдеры предоставляют цену несколько выше средней. Так же провайдеры находятся на достаточно высоком уровне относительно горизонтальной оси, то есть они имеют хорошую зону покрытия, но мало удовлетворяют условию мобильности, причем «Дом.ру» предоставляет большую зону покрытия, чем «Уфанет».

Теперь сравним «Билайн» и «Мегафон». Эти провайдеры расположены ниже других относительно вертикальной оси, то есть они обладают высокой мобильностью. Причем «Мегафон» обладает большей мобильностью, чем «Билайн». Относительно «Билайн» нельзя сказать, он ближе к полюсу «высокая цена» или же к полюсу «низкая цена», потому что точка практически лежит на вертикальной оси, и, следовательно, ее координата по горизонтальной оси и относительный вклад близки к нулю. Это означает, что для провайдеру «Билайн» в одинаковой степени характерны атрибуты и правого, и левого полюса; или же в одинаковой степени не характерны ни те, ни другие. Фактически, в контексте горизонтальной оси характеристики «Билайн» мало чем отличаются от средних показателей по всем провайдерам. Таким образом, он имеет «высокую цену» и «низкую цену» ровно настолько, насколько все провайдеры в среднем; следовательно, по этому фактору он ничем от среднего не отличается. «Мегафон» характеризуется ценой немного ниже, чем «Билайн», так как он расположен правее вертикальной оси координат.

По рисунку 4.20 можно сделать вывод, что каждый провайдер для того, чтобы сохранить свою конкурентоспособность на рынке, вырабатывает свою стратегию

развития по отношению к построенным характеристикам. Так, например, «МТС» предлагает более низкую цену, «Билайн» и «Мегафон» – мобильность, а «Дом.ру» и «Уфанет» – хорошую зону покрытия. «Ростелеком» отличается высокой ценой и в связи с этим, очевидно, проигрывает остальным Интернет-провайдерам.

5 Содержание письменного отчета

Отчет по лабораторной работе оформляется на листах формата А4 и должен иметь следующую структуру:

- 1) титульный лист;
- 2) задание на лабораторную работу;
- 3) краткие теоретические сведения, необходимые для решения поставленных задач;
- 4) краткая характеристика объекта исследования, описание информационной базы;
- 5) результаты построения карты соответствий;
- 6) анализ полученных результатов и выводы;
- 7) вопросы анкеты приводятся в приложении.

6 Контрольные вопросы

1. Дайте определение таблицы сопряженности.
2. Какие методы используются для проверки взаимосвязи между номинальными и порядковыми признаками?
3. Опишите проверки гипотезы о независимости признаков с помощью критерия Пирсона.
4. В чем основная идея метода анализа соответствий?
5. Что выступает в качестве исходных данных для реализации простого анализа соответствия?
6. Дайте определение инерции.
7. Перечислите основные шаги алгоритма простого анализа соответствий.
8. Дайте определение профиля строки и столбца.
9. Как определяется количество главных осей?
10. В каком случае главным осям рекомендуется давать название?
11. Как оценивается качество представления каждой строки в пространстве сниженной размерности?
12. Чему равен коэффициент корреляции между профилями и главными осями? Как интерпретируется эта характеристика?
13. Опишите построение матрицы Бёрта.

Список использованных источников

1. Чудинова, О.С. Анализ таблиц сопряженности в пакетах Statistica, СНИИ, Excel: методические указания к лабораторным работам, практическим занятиям и самостоятельной работе студентов / О.С. Чудинова; Оренбургский гос. ун-т.– Оренбург: ОГУ, 2014. – 57 с.
2. Методы и модели эконометрики [Электронный ресурс]: учебное пособие для студентов, обучающихся по программам высшего образования по направлениям подготовки 01.03.04 Прикладная математика, 38.04.01 Экономика, 38.03.05 Бизнес-информатика / О.И. Бантикова, В.И. Васянина, Ю.А. Жемчужникова, А.Г. Реннер, Е.Н. Седова, О.И. Стебунова, Л.М. Туктамышева, О.С. Чудинова /под ред. А. Г. Реннера. – Ч. 1. Анализ данных. – Оренбург : ОГУ, 2015.
3. Прикладная статистика: Классификация и снижение размерности: Справ.изд. / С.А Айвазян, В.М. Бухштабер, И.С. Енюков, Л.Д. Мешалкин; Под ред. С.А. Айвазяна. – М.: Финансы и статистика, 1989. – 607 с.
4. Раменская, А. В. Анализ рынка Интернет-провайдеров г. Оренбурга / Раменская А. В. // Вестник Оренбургского государственного университета, 2014. – № 14, декабрь. – С. 442-446.
5. Раменская, А. В. Проведение анализа рынка на примере комбикормовых предприятий / А. В. Раменская, Т. П. Негорожина // Вестник магистратуры, 2017. - № 6-1 (69). - С. 36-39.

Приложение А (обязательное)

Варианты исходных данных

Вариант	Тематика
1	Рынок мобильной связи Оренбургской области (города Оренбурга)
2	Аптечные сети Оренбургской области (города Оренбурга)
3	Магазины бытовой техники Оренбургской области (города Оренбурга)
4	Рынок смартфонов Оренбургской области (города Оренбурга)
5	Кинотеатры города Оренбурга
6	Рынок телекоммуникационных услуг Оренбургской области (города Оренбурга)
7	Рынок антивирусного программного обеспечения Оренбургской области (города Оренбурга)
8	Рынок планшетных компьютеров Оренбургской области (города Оренбурга)
9	Рынок портативных компьютеров Оренбургской области (города Оренбурга)
10	Кофейни города Оренбурга
11	Торгово-развлекательные центры города Оренбурга
12	Гостиничные комплексы города Оренбурга