



**Е.Е. Фомина**

## **ПРИМЕНЕНИЕ МЕТОДА АНАЛИЗА СООТВЕТСТВИЙ В СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ИССЛЕДОВАНИЯХ**

Математическое моделирование широко используется в экономике, социологии, маркетинге, политологии и других науках для анализа данных и получения содержательных и обоснованных выводов об их структуре и зависимостях между переменными, описывающих тот или иной процесс или явление. Разработан достаточно большой набор математических методов для исследования числовых данных, представленных в интервальной шкале или шкале отношений. К таким методам относятся деревья классификации, дискриминантный, факторный, кластерный анализ и др. Однако часто в социально-экономических исследованиях возникает необходимость изучения категориальных переменных или переменных нечисловой природы, основной формой представления которых являются частотные таблицы. Для численного и визуального анализа таблиц сопряженности большого размера может использоваться метод анализа соответствий, основанный на статистике хи-квадрат. Задача настоящей работы – демонстрация возможностей метода при анализе и интерпретации первичных результатов социально-экономических исследований. В статье рассмотрены математические основы простого анализа соответствий и практическое использование метода при обработке результатов анкетирования. Метод применен для изучения динамики финансовой активности населения. Проанализированы предпочтения респондентов в сфере услуг, предоставляемых различными финансовыми организациями, в зависимости от возраста. Построена двумерная карта соответствий и таблица со статистиками решения, позволившая провести визуальный и численный анализ взаимосвязи между возрастными категориями и услугами финансовых организаций. Выявлены наиболее предпочтительные услуги для каждой возрастной группы. Проведенный анализ результатов может быть использован для разработки рекомендаций по формированию потребительских предпочтений на рынке финансовых услуг.

Ключевые слова: *анализ соответствий, частотные таблицы, анализ категориальных переменных, анкетирование.*

**Введение.** Для анализа данных, представленных в интервальной шкале или шкале отношений, разработан достаточно большой набор математических методов, к которым можно отнести корреляционно-регрессионный анализ [1–4], кластерный, дискриминантный и факторный анализ [5–7], деревья классификации [8–10], а также ряд других методов добычи данных, позволяющих осуществлять поиск закономерностей в массиве информации.

Однако в социологических, экономических и маркетинговых исследованиях часто возникает необходимость анализа номинальных и порядковых

---

© Фомина Е.Е., 2019

**Фомина Елена Евгеньевна** – канд. техн. наук, доцент кафедры информатики и прикладной математики ФГБОУ ВО «Тверской государственной технической университет», e-mail: f-elena2008@yandex.ru.

переменных или переменных нечисловой природы. Так, например, при изучении общественного мнения по той или иной теме (проблеме) прибегают к анкетированию, результатом которого является база данных. Поля базы данных – категориальные переменные, содержащие варианты ответа на тот или иной вопрос.

Простейшим инструментом исследования взаимосвязи между двумя переменными такого типа является расчет критерия хи-квадрат [4, 11]. Поиск статистически значимых взаимодействий в большом наборе данных может быть осуществлен с использованием метода логлинейного анализа, который позволяет обнаруживать сложные внутренние взаимосвязи между категориальными переменными [12–14]. Однако данные методы не позволяют визуализировать результаты и более содержательно интерпретировать их. Актуальной задачей является задача внедрения эффективных методов анализа информации такого рода.

Для решения этой проблемы может использоваться анализ соответствий. Анализ соответствий – это метод многомерного статистического анализа, предназначенный для численного и визуального исследования таблиц сопряженности большой размерности [15–18]. Он основан на расчете расстояния хи-квадрат, которое позволяет учесть «вес» каждого объекта. Результатом применения метода будет являться двумерная карта соответствий между переменными, что существенно упростит восприятие данных и интерпретацию решения.

В статье продемонстрированы возможности метода при анализе и интерпретации первичных результатов социально-экономических исследований.

**Метод исследования. Математические основы анализа соответствий.** На практике используется две разновидности анализа соответствий – простой и множественный. Множественный анализ соответствий предполагает применение алгоритма простого анализа соответствий к индикаторной матрице, строками которой являются номера рассмотренных случаев (например, номера респондентов), а столбцами – значения категорий всех переменных. В связи с этим рассмотрим математические основы только простого анализа соответствий [18].

Исходными данными метода является таблица сопряженности двух переменных  $A$  и  $B$ , принимающих дискретные значения  $A = \{A_1, A_2, \dots, A_N\}$  и  $B = \{B_1, B_2, \dots, B_M\}$  с частотами  $n_{ij}$ ,  $i = \overline{1, N}$ ,  $j = \overline{1, M}$ :

	$B_1$	$B_2$	...	$B_M$	<i>Маргинальные суммы по строкам</i>
$A_1$	$n_{11}$	$n_{12}$	...	$n_{1M}$	$n_{1+} = \sum_{j=1}^M n_{1j}$
$A_2$	$n_{21}$	$n_{22}$	...	$n_{2M}$	$n_{2+} = \sum_{j=1}^M n_{2j}$
...	...	...	...	...	...
$A_N$	$n_{N1}$	$n_{N2}$	...	$n_{NM}$	$n_{N+} = \sum_{j=1}^M n_{Nj}$
<i>Маргинальные суммы по столбцам</i>	$n_{+1} = \sum_{i=1}^N n_{i1}$	$n_{+2} = \sum_{i=1}^N n_{i2}$	...	$n_{+M} = \sum_{i=1}^N n_{iM}$	$n = \sum_{j=1}^M n_{+j} = \sum_{i=1}^N n_{i+}$

Поскольку каждая строка и столбец таблицы содержат разное число случаев, сравнение данных невозможно, для этого их следует унифицировать, составив матрицу соответствий  $P$ :

$$P = \left\{ p_{ij} = \left( \frac{n_{ij}}{n} \right) \right\}_{i=1, \overline{N}, j=1, \overline{M}}.$$

Полученные в матрице  $P$  относительные частоты дают возможность сравнивать значения строк и столбцов между собой.

Далее рассчитываются профили строк (1) и столбцов (2), т.е. наборы относительных частот, вычисленные для строки / столбца:

$$\left( \begin{array}{cccc} \frac{n_{11}}{n_{1+}} & \frac{n_{12}}{n_{1+}} & \dots & \frac{n_{1M}}{n_{1+}} \\ \frac{n_{21}}{n_{2+}} & \frac{n_{22}}{n_{2+}} & \dots & \frac{n_{2M}}{n_{2+}} \\ \dots & \dots & \dots & \dots \\ \frac{n_{N1}}{n_{N+}} & \frac{n_{N2}}{n_{N+}} & \dots & \frac{n_{NM}}{n_{N+}} \end{array} \right), \tag{1}$$

$$\left( \begin{array}{cccc} \frac{n_{+1}}{n} & \frac{n_{+2}}{n} & \dots & \frac{n_{+M}}{n} \\ \frac{n_{21}}{n_{+1}} & \frac{n_{22}}{n_{+1}} & \dots & \frac{n_{2M}}{n_{+1}} \\ \dots & \dots & \dots & \dots \\ \frac{n_{N1}}{n_{+1}} & \frac{n_{N2}}{n_{+1}} & \dots & \frac{n_{NM}}{n_{+1}} \end{array} \right). \tag{2}$$

Профили представляют собой точки в многомерном пространстве, а средний профиль – точку в центре облака точек, представленного остальными профилями.

Для того чтобы учесть значимость каждого профиля или то количество случаев, которое он определяет, рассчитываются их массы:

$$r_i = \frac{n_{i+}}{n}, \quad i = \overline{1, N} \text{ – массы строк;}$$

$$c_j = \frac{n_{+j}}{n}, \quad j = \overline{1, M} \text{ – массы столбцов.}$$

Тогда профили (1) и (2) можно записать в матричном виде соответственно:

$$D_r^{-1}P, \tag{3}$$

где  $D_r$  – диагональная матрица масс строк;

$$PD_c^{-1}, \tag{4}$$

где  $D_c$  – диагональная матрица масс столбцов.

Близость между профилями-точками в пространстве определяется путем вычисления хи-квадрат расстояния, которое учитывает «вес» каждого профиля и служит неким инструментом стандартизации:

$$d(i_1, i_2) = \sqrt{\sum_{j=1}^M \left( \frac{p_{i_1j}}{r_{i_1}} - \frac{p_{i_2j}}{r_{i_2}} \right)^2 / c_j} \text{ – хи-квадрат расстояние между профилями } i_1 \text{ и } i_2 \text{ строки;}$$

ми  $i_1$  и  $i_2$  строки;

$$d(j_1, j_2) = \sqrt{\sum_{i=1}^N \left( \frac{p_{ij_1}}{c_{j_1}} - \frac{p_{ij_2}}{c_{j_2}} \right)^2 / r_i} \text{ – хи-квадрат расстояние между профилями } j_1 \text{ и } j_2 \text{ столбцов.}$$

ми  $j_1$  и  $j_2$  столбцов.

Расстояния между профилями используются для вычисления инерции ( $I$ ) – суммы моментов инерции всех профилей системы, в которой центром тяжести является средний профиль:

$$I = \sum m \times d^2,$$

где  $m$  – масса профиля;  $d$  – расстояние от профиля до центра тяжести (среднего профиля).

Чем больше инерция, тем больше разброс профилей относительно центра тяжести. Это понятие эквивалентно понятию дисперсии в факторном анализе.

В основе исследования частотных таблиц большой размерности методом анализа соответствий лежит идея изучения взаимосвязи между профилями, которые имеют некоторые массы, позволяющие учесть их значимость, в пространстве с метрикой хи-квадрат расстояние.

Поскольку размерность этого пространства, как правило, больше трех, невозможна его визуализация. Задача анализа соответствий – это снижение размерности исходного пространства с целью представления профилей строк и столбцов одновременно на одной плоскости в более удобном восприятии для их последующего визуального и численного исследования [19].

Снижение размерности исходного пространства сопряжено с поиском двумерного пространства (плоскости), которое наиболее точно отражает расстояние между точками и минимизирует остаточную инерцию [19].

Для решения задачи снижения размерности пространства строк используется сингулярное разложение матрицы стандартизированных остатков  $A_1 = D_r^{-1/2}(P - rc^T)D_c^{-1/2}$ . Для снижения размерности пространства столбцов – сингулярное разложение матрицы стандартизированных остатков  $A_2 = D_c^{1/2}(P^T - cr^T)D_r^{-1/2}$  [18]. При этом  $A_2 = A_1^T$ . Таким образом, задача строк и столбцов решается путем разложения одной и той же матрицы, что дает основание для совместного представления профилей-строк и столбцов на одной плоскости [15].

Идея снижения размерности базируется на методе главных компонент [20], согласно которому сокращение размерности пространства происходит посредством «отсечения» неинформативных переменных, используя для этого следующие критерии:

- критерий, связанный с долей объясненной инерции: суммарная объясненная инерция должна быть не менее заданной доли;
- критерий Кеттела (критерий «каменистой осыпи»): согласно которому сингулярные числа отображаются на графике, где по оси абсцисс откладываются их номера, а по оси ординат – их значения. Далее ищется точка на графике, где убывание сингулярных чисел максимально замедляется. Номер, соответствующий этому числу, и определяет оптимальное число измерений.

**Материалы исследования. Применение анализа соответствий для обработки результатов анкетирования.** Возможности метода продемонстрируем на примере анализа результатов исследования, посвященного динамике финансовой активности населения как одного из основных показателей общего социально-экономического уровня развития регионов и страны в целом.

Одной из задач исследования являлась задача анализа спроса на услуги различных финансовых организаций в зависимости от возраста респондента. Такое исследование поможет дать более глубокое представление о доступе к услугам, о пользовании ими, включая в анализ информацию о развивающихся услугах, таких, например, как мобильный банкинг.

В частности, респондентам предлагалось ответить на вопрос: «Какими услугами различных финансовых организаций Вы пользовались на протяжении последних 12 месяцев?». Варианты: «обмен валюты»; «денежный перевод»; «текущий банковский счет / вклад до востребования»; «срочный вклад в банке / депозит»; «потребительский кредит»; «ипотечный кредит»; «пластиковая карта без возможности кредитования»; «кредитная пластиковая карта»; «полис обязательного страхования (медицинского, ОСАГО и т.п.)»; «полис добровольного страхования»; «вклад в негосударственный пенсионный фонд»; «покупка и продажа ценных бумаг»; «денежные игры / лотереи»; «мобильный банк».

В опросе приняли участие 996 жителей г. Твери и Тверской области. Возрастные категории представлены следующим количеством респондентов: «Возраст: 18–24» – 73 чел.; «Возраст: 25–34» – 299 чел.; «Возраст: 35–44» – 179 чел.; «Возраст: 45–55» – 299 чел.; «Возраст: 55 и старше» – 146 чел.

Фрагмент таблицы сопряженности вариантов ответов и возрастных категорий приложен на рис. 1. Каждый респондент мог выбрать несколько вариантов ответов, кроме того, каждая услуга была выбрана разным количеством человек, поэтому сравнение представленных данных в исходном виде затруднительно.

Услуга	Возраст: 18–24	Возраст: 25–34	Возраст: 35–44	Возраст: 45–55	Возраст: 55 и старше
Обмен валюты	49	53	44	51	27
Ден. перевод	81	67	61	61	59
Текущ. БС / до востр.	39	58	39	51	95
Срочн. счет / депозит	15	31	28	36	47
Потр. кредит	59	91	92	90	42
Ип. кредит	3	13	13	4	1
Пл. карта (без возм. кред.)	125	162	149	181	141

Рис. 1. Фрагмент таблицы сопряженности

Используем метод анализа соответствий для исследования зависимости между переменными. Расчеты производились в пакете Statistica 10 [21].

Применение метода позволило выделить четыре измерения, которым соответствуют сингулярные числа и инерция, представленные в табл. 1.

Анализ табл. 1 позволяет сделать вывод, что две оси объясняют 89,985 % инерции, что говорит о хорошем соответствии двумерной модели исходным данным.

На рис. 2 представлена карта соответствий для двумерного решения, а в табл. 2 – его основные статистики.

Таблица 1

## Сингулярные значения и инерция

№ п/п	Сингул. – значения	Собств. – значения	Процент – Инерция	Кумулят. – Процент	Хи-квадрат
1	0,202	0,041	69,721	69,721	135,695
2	0,109	0,012	20,264	89,986	39,439
3	0,063	0,004	6,850	96,836	13,333
4	0,043	0,002	3,164	100,000	6,158

На карте соответствий изображены точки, соответствующие строкам и столбцам одновременно. Сравнение таких разных по типу данных некорректно, поэтому далее будем рассматривать положение столбцов (возрастных категорий) в пространстве строк (услуг).

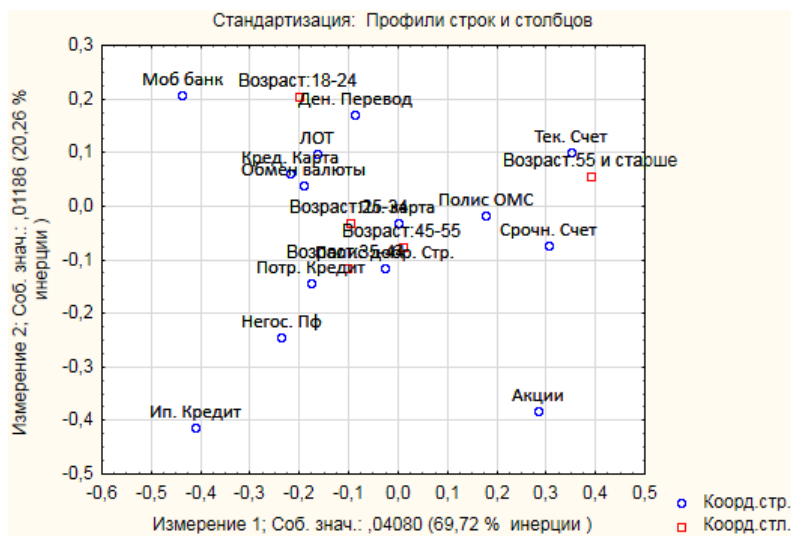


Рис. 2. Карта соответствий

Проанализируем более детально таблицу с характеристиками первого и второго измерений (см. табл. 2).

Одним из показателей соответствия модели исходным данным является столбец «*Качество*». При корректном решении основная масса точек строк и столбцов должна быть правильно представлена на карте соответствий, т.е. расстояния между ними не должны искажаться. Как можно видеть по данным табл. 2, значения в столбце «*Качество*» достаточно высоки для всех строк и столбцов, за исключением строк «Пластиковая карта без возможности кредитования», «Полис добровольного страхования», «Покупка и продажа ценных бумаг», а также столбца «*Возраст: 45–55 лет*».

Таблица 2

## Статистики двумерного решения

Строки/ столбцы	Координ. измерения 1	Координ. измерения 2	Масса	Качество	Относит. инерция	Инерция измерения 1	Косинус <sup>2</sup> измерения 1	Инерция измерения 2	Косинус <sup>2</sup> измерения 2
<i>Услуги</i>									
Обмен валю- ты	-0,193	0,038	0,067	0,928	0,048	0,061	0,893	0,008	0,035
Ден.перевод	-0,087	0,171	0,099	0,915	0,068	0,018	0,186	0,245	0,729
Текущ. БС / до востр.	0,349	0,101	0,085	0,963	0,199	0,254	0,888	0,073	0,074
Срочн. счет / депозит	0,305	-0,073	0,047	1,000	0,079	0,108	0,946	0,021	0,054
Потр. кредит	-0,176	-0,144	0,112	0,984	0,101	0,085	0,591	0,196	0,394
Ип. кредит	-0,410	-0,413	0,010	0,665	0,089	0,042	0,330	0,147	0,335
Пл. карта без возм. кред.	-0,001	-0,031	0,228	0,241	0,015	0,000	0,000	0,018	0,241
Кред. пл. кар- та	-0,220	0,061	0,055	0,915	0,054	0,066	0,850	0,017	0,065
Полис ОС	0,178	-0,018	0,202	0,993	0,112	0,157	0,982	0,006	0,010
Полис ДС	-0,027	-0,116	0,021	0,382	0,014	0,000	0,020	0,024	0,362
Вкл. в негос. ПФ	-0,237	-0,244	0,012	0,816	0,030	0,017	0,396	0,062	0,420
Покупка цен.бумаг	0,283	-0,381	0,003	0,491	0,021	0,005	0,174	0,033	0,316
Ден.игры / Лот.	-0,165	0,099	0,020	0,824	0,016	0,014	0,606	0,017	0,217
Моб. банк	-0,438	0,207	0,037	0,951	0,154	0,172	0,778	0,132	0,173
<i>Возрастные категории</i>									
Возраст: 18–24	-0,202	0,205	0,171	0,992	0,245	0,171	0,488	0,608	0,503
Возраст: 25–34	-0,098	-0,031	0,226	0,590	0,070	0,054	0,536	0,019	0,054
Возраст: 35–44	-0,104	-0,114	0,195	0,789	0,101	0,052	0,359	0,215	0,431
Возраст: 45–55	0,008	-0,077	0,214	0,324	0,068	0,000	0,004	0,107	0,320
Возраст: 55 и старше	0,391	0,056	0,193	0,996	0,516	0,722	0,976	0,051	0,020

Столбец «Масса» позволяет проанализировать распределение массы профилей по точкам пространства.

Проинтерпретируем решение, рассмотрев положение возрастных категорий в пространстве услуг. Поскольку переменная «Услуги» содержит



14 значений, то средний вклад каждой услуги составит  $1/14=0,071$ . Таким образом, с каждым измерением (осью) будут связаны те услуги, для которых вклады в его инерцию больше 0,071.

Дополнительным критерием содержательной связи точки с той или иной осью является значение квадратов корреляций (столбец «Косинус<sup>2</sup>»). С каждой осью будем ассоциировать те точки, квадрат корреляции которых больше 0,7.

Горизонтальная ось наиболее тесно связана с такими точками-строками, как «Текущий банковский счет / счет до востребования» (инерция 0,254; квадрат корреляции 0,888), «Срочный вклад в банке / депозит» (инерция 0,108; квадрат корреляции 0,946) и «Полис обязательного страхования» (инерция 0,157; квадрат корреляции 0,982) – точки ассоциируются с положительной частью оси, так как координаты этих точек по первому измерению положительны (см. табл. 2) и «Потребительский кредит» (инерция 0,085; квадрат корреляции 0,591), «Кредитная пластиковая карта» (инерция 0,066; квадрат корреляции 0,850), «Мобильный банк» (инерция 0,172; квадрат корреляции 0,778), «Обмен валюты» (инерция 0,061; квадрат корреляции 0,893) – ассоциируются с отрицательной частью оси, так как координаты этих точек по первому измерению отрицательны (см. табл. 2). Нужно отметить, что знак координаты не несет содержательной нагрузки, а позволяет связать точку с направлением оси.

Вертикальная ось наиболее тесно связана с такими точками-строками, как «Денежный перевод» (инерция 0,245; квадрат корреляции 0,729), «Текущий банковский счет / счет до востребования» (инерция 0,073; квадрат корреляции 0,074) и «Мобильный банк» (инерция 0,132; квадрат корреляции 0,173) – точки ассоциированы с положительной частью оси, так как имеют положительные координаты по второму измерению (см. табл. 2) и «Потребительский кредит» (инерция 0,196; квадрат корреляции 0,394), «Ипотечный кредит» (инерция 0,147; квадрат корреляции 0,335) – ассоциированы с отрицательной частью оси, так как имеют отрицательные координаты по второму измерению (см. табл. 2).

При анализе взаимосвязей осей и точек-профилей установлено, что некоторые профили связаны с двумя измерениями одновременно – «Текущий банковский счет / счет до востребования», «Потребительский кредит», «Мобильный банк» (для этих точек значение инерции превышает порог 0,071). В этом случае при интерпретации связи будем руководствоваться значением квадрата корреляции и ассоциировать точку с той осью, для которой значение «Косинус<sup>2</sup>» больше.

Карта соответствий расположения возрастных категорий в пространстве услуг представлена на рис. 3. Для более детального анализа проведены биссектрисы углов I–IV четвертей, разделяющие плоскость на четыре сектора. Точки каждого сектора максимально коррелируют с теми частями осей, которые в нем содержатся.

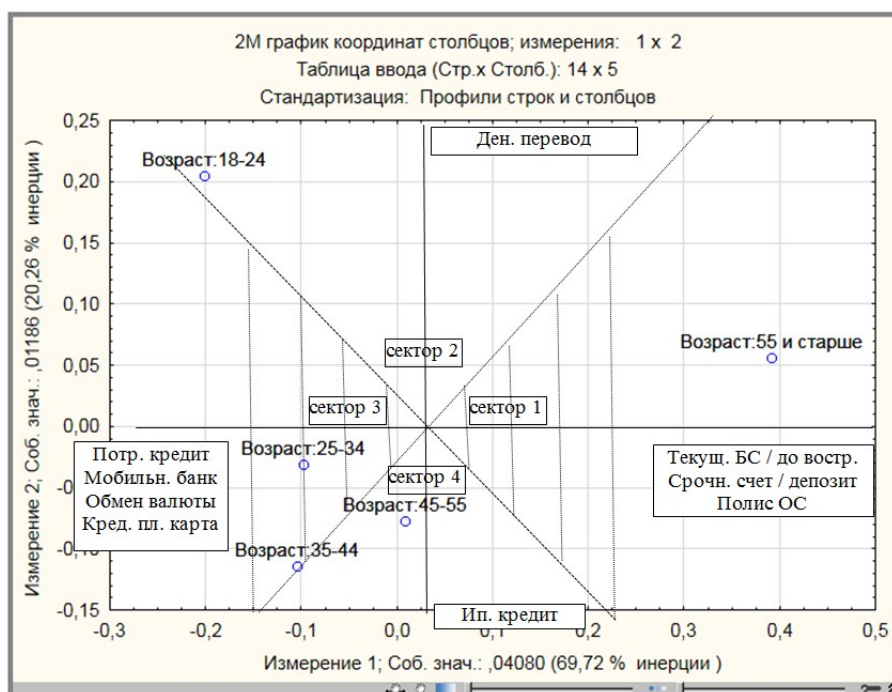


Рис. 3. Карта соответствий возрастных категорий в пространстве услуг

Возрастная категория «55 лет и старше» расположена в первом секторе и наиболее тесно связана с положительным направлением горизонтальной оси. Таким образом, для респондентов этой возрастной категории наиболее востребованными являются банковские услуги по использованию текущего и срочного вкладов и полисы обязательного страхования.

Возрастная категория «18–24 года» – самая молодая группа респондентов – располагается практически на биссектрисе угла, соответствующего II четверти, следовательно, имеет приблизительно равную корреляцию с обеими осями. Кроме того, эта точка ближе всего расположена к положительному полюсу вертикальной оси и отрицательному полюсу горизонтальной оси. Следовательно, для респондентов этой группы наиболее предпочтительны услуги пользования денежными переводами, потребительскими кредитами, мобильным банком, кредитной пластиковой картой и обменом валюты.

Возрастная категория «35–44 года» лежит на биссектрисе угла, соответствующего III четверти, т.е. скоррелирована с обеими осями. Следовательно, для респондентов этой группы наиболее востребованными являются услуги, относящиеся к отрицательным полюсам обеих осей.

Возрастная категория «45–55 лет» наиболее связана с отрицательной частью вертикальной оси, которая сопоставлена с такой услугой, как

ипотечный кредит. Нужно отметить, что качество представления этой возрастной категории на карте соответствий не высоко (см. табл. 2), поэтому для более корректной интерпретации этой точки следует рассмотреть третье измерение.

Возрастная категория «25–34 года» связана с отрицательной частью горизонтальной оси, соотнесенной с такими услугами, как потребительский кредит, кредитная пластиковая карта, мобильный банк и обмен валюты.

Нужно отметить, что имеются услуги, которые не соотнесены ни с одной из осей. К ним относятся такие услуги, как «Пластиковая карта без возможности кредитования», «Полис добровольного страхования» и «Вклад в негосударственный пенсионный фонд», «Покупка и продажа ценных бумаг» и «Денежные игры/лотереи». Этот факт можно объяснить тем, что данные услуги одинаково востребованы или не востребованы для каждой возрастной категории.

Проведенный анализ может способствовать разработке рекомендаций по формированию потребительских предпочтений на рынке финансовых услуг для различных возрастных категорий респондентов.

**Выводы.** Показаны возможности метода анализа соответствий при изучении социально-экономической информации и получении первичных результатов о взаимосвязи переменных, описывающих процессы или явления.

Анализ соответствий позволяет производить визуальное и численное исследование таблиц сопряженности больших размерностей и выявлять наличие взаимосвязей между категориальными переменными.

Метод был применен для изучения спроса на услуги различных финансовых организаций в зависимости от возрастной категории.

Построена двумерная карта соответствий, позволившая рассмотреть и проанализировать расположение возрастных групп в пространстве услуг. Выявлены наиболее предпочтительные услуги для каждой категории.

## Список литературы

1. Эконометрика: учеб. / под ред. И.И. Елисеевой. – М.: Финансы и статистика, 2003. – 344 с.
2. Мелихова Е.В., Рогачев А.Ф. Эконометрика: учеб. пособие / Волгогр. ГАУ. – Волгоград, 2014. – 96 с.
3. Теория статистики: учеб. / под ред. проф. Г.Л. Громько. – 2-е изд., перераб. и доп. – М.: ИНФРА-М, 2005. – 476 с.
4. Фомина Е.Е., Жиганов Н.К. Математические методы анализа данных в социологии с использованием пакетов MS Excel и Statistica: учеб. пособие / ТвГТУ. – Тверь, 2017. – 68 с.

5. Бессокирная Г.П. Факторный анализ: традиции использования и новые возможности // Социология: методология, методы, математическое моделирование. – 2000. – № 12. – С. 142–153.

6. Буреева Н.Н. Многомерный статистический анализ с использованием ППП «STATISTICA»: учебно-методический материал по программе повышения квалификации «Применение программных средств в научных исследованиях и преподавании математики и механики». – Нижний Новгород, 2007. – 112 с.

7. Фомина Е.Е. Методы многомерной статистики в социологических и социально-экономических исследованиях: учеб. пособие / ТвГТУ. – Тверь, 2019. – 112 с.

8. Бова А. Деревья решений как техника добычи данных // Социология: теория, методы, маркетинг. – 2002. – № 1. – С. 128–136.

9. Фомина Е.Е. Возможности метода деревьев классификации при обработке социологической информации // Гуманитарный вестник. – 2018. – № 11 (73). – С. 5.

10. Анализ статистических данных с использованием деревьев решений [Электронный ресурс]. – URL: <http://math.nsc.ru/AP/datamine/decisiontree.htm> (дата обращения: 22.04.2018).

11. Толстова Ю.Н. Математическая статистика для социологов: задачник: учеб. пособие для вузов. – М.: Изд. дом ВШЭ, 2010. – 185 с.

12. Толстова Ю.Н. Анализ социологических данных. Методология, дескриптивная статистика, изучение связей между номинальными признаками. – М.: Научный мир, 2000. – 352 с.

13. Трофимов Д.А. Логлинейный анализ таблиц мобильности: обзор основных моделей // Социология: 4М. – 2008. – № 26. – С. 119–138.

14. Фомина Е.Е. Возможности логлинейного анализа при обработке результатов анкетирования // Вестник Пермского национального исследовательского политехнического университета. Социально-экономические науки. – 2018. – № 3. – С. 197–211.

15. Клишина Ю.И. Применение анализа соответствий в обработке нечисловой информации // Социология: методология, методы, математические модели. – 1991. – № 2. – С. 105–119.

16. Крипак Е.М. Применение метода анализа соответствий в процессе стратегического конкурентного анализа // Вестник Оренбургского государственного университета. – 2014. – № 14 (175). – С. 438–441.

17. Скрипникова Н.И., Чудинова О.С. Использование методов анализа соответствий в маркетинговых исследованиях // Научное сообщество студентов XXI столетия. Экономические науки: материалы IV студ. междунар. заоч. науч.-практ. конф. (18 октября 2012 г.). – Новосибирск: Сибирская ассоциация консультантов, 2012. – С. 104.

18. Шафир М.А. Анализ соответствий: представление метода // Социология: 4М. – 2009. – № 28. – С. 29–44.

19. Прикладная статистика: Классификации и снижение размерности [Электронный ресурс]. – URL: [http://stu.sernam.ru/book\\_stat3.php](http://stu.sernam.ru/book_stat3.php) (дата обращения: 05.02.2019).

20. Общий алгоритм и теоретические проблемы факторного анализа [Электронный ресурс]. – URL: <http://www.studfiles.ru/preview/1938850/page:3/> (дата обращения: 01.03.2017).

21. Боровиков В.П. STATISTICA: искусство анализа данных на компьютере. – СПб.: Питер, 2001. – 656 с.

## References

1. *Ekonometrika* [Econometrics]. Ed. I.I. Eliseeva. Moscow, Finansy i statistika, 2003, 344 p.

2. Melikhova E.V., Rogachev A.F. *Ekonometrika* [Econometrics]. Volgograd, Volgograd State Agrarian University, 2014, 96 p.

3. *Teoriia statistiki* [Statistics theory]. Ed. G.L. Gromyko. 2nd ed., Moscow, INFRA-M, 2005, 476 p.

4. Fomina E.E., Zhiganov N.K. *Matematicheskie metody analiza dannykh v sotsiologii s ispol'zovaniem paketov MS Excel i Statistica* [Mathematical data analysis methods in sociology using software packages MS Excel and Statistica]. Tver, Tver State Technical University, 2017, 68 p.

5. Bessokirnaia G.P. *Faktornyĭ analiz: traditsii ispol'zovaniia i novye vozmozhnosti* [Factor analysis: Traditions of using and new opportunities]. *Sotsiologĭia: metodologĭia, metody, matematicheskoe modelirovanie*, 2000, no. 12, pp. 142–153.

6. Bureeva N.N. *Mnogomernyi statisticheskii analiz s ispol'zovaniem PPP "STATISTICA"* [Multidimensional statistical analysis using software package "Statistica"]. *Uchebno-metodicheskii material po programme povysheniia kvalifikatsii "Primenenie programmnykh sredstv v nauchnykh issledovaniakh i prepodavanii matematiki i mekhaniki"*. Nizhny Novgorod, 2007, 112 p.

7. Fomina E.E. *Metody mnogomernoi statistiki v sotsiologicheskikh i sotsial'no-ekonomicheskikh issledovaniakh* [Methods of multidimensional statistics in sociological and socio-economic studies]. Tver, Tver State Technical University, 2019, 112 p.

8. Bova A. *Derev'ia reshenii kak tekhnika dobychi dannykh* [Decision trees as a data mining technique]. *Sotsiologĭia: teoriia, metody, marketing*, 2002, no. 1, pp. 128–136.

9. Fomina E.E. *Vozmozhnosti metoda derev'ev klassifikatsii pri obrabotke sotsiologicheskoi informatsii* [The potential of the method of classification trees in the sociological information processing]. *Gumanitarnyi vestnik*, 2018, no. 11(73), p. 5.

10. Analiz statisticheskikh dannykh s ispol'zovaniem derev'ev reshenii [Analysis of statistical data using decision trees]. Available at: <http://math.nsc.ru/AP/datamine/decisiontree.htm> (accessed 22 April 2018).
11. Tolstova Iu.N. Matematicheskaia statistika dlia sotsiologov [Mathematical statistics for sociologists]. Moscow, Higher School of Economics, 2010, 185 p.
12. Tolstova Iu.N. Analiz sotsiologicheskikh dannykh [Analysis of sociological data]. *Metodologiya, deskriptivnaia statistika, izuchenie svyazei mezhdru nominal'nymi priznakami*. Moscow, Nauchnyi mir, 2000, 352 p.
13. Trofimov D.A. Loglineinyi analiz tablits mobil'nosti: obzor osnovnykh modelei [Log-linear analysis of mobility tables: an overview of the basic models]. *Sotsiologiya: 4M*, 2008, no. 26, pp. 119–138.
14. Fomina E.E. Vozmozhnosti loglineinogo analiza pri obrabotke rezul'tatov anketirovaniia [The possibility of log-linear analysis for survey results processing]. *PNRPU Sociology and Economics Bulletin*, 2018, no. 3, pp. 197–211.
15. Klishina Iu.I. Primenenie analiza sootvetstviu v obrabotke nechislovoi informatsii [Application of the correspondence analysis in the processing of non-numeric information]. *Sotsiologiya: 4M*, 1991, no. 2, pp. 105–119.
16. Kripak E.M. Primenenie metoda analiza sootvetstviu v protsesse strategicheskogo konkurentnogo analiza [Application of the correspondence analysis in the process of strategic competitive analysis]. *Vestnik Orenburgskogo gosudarstvennogo universiteta*, 2014, no. 14(175), pp. 438–441.
17. Skripnikova N.I., Chudinova O.S. Ispol'zovanie metodov analiza sootvetstviu v marketingovykh issledovaniiah [The use of methods of the correspondence analysis in marketing researches]. “*Nauchnoe soobshchestvo studentov XXI stoliia. Ekonomicheskie nauki*”. Proceedings of IV Stud. Int. Virtual Sci.-Pract. Conf. October 18, 2012, Novosibirsk, Sibirskaiia assotsiatsiia konsul'tantov, 2012, p. 104.
18. Shafir M.A. Analiz sootvetstviu: predstavlenie metoda [Correspondence analysis: Presentation of the method]. *Sotsiologiya: metodologiya, metody, matematicheskoe modelirovanie*, 2009, no. 28, pp. 29–44.
19. Prikladnaia statistika: Klassifikatsii i snizhenie razmernosti [Applied statistics: Classification and dimension reduction]. Available at: [http://stu.sernam.ru/book\\_stat3.php](http://stu.sernam.ru/book_stat3.php) (accessed 05 February 2019).
20. Obshchii algoritm i teoreticheskie problemy faktornogo analiza [General algorithm and theoretical problems of factor analysis]. Available at: <http://www.studfiles.ru/preview/1938850/page:3/> (accessed 01 March 2017).
21. Borovikov V.P. STATISTICA: iskusstvo analiza dannykh na komp'yutere [Statistica: The art of data analysis with computer.]. St. Petersburg, Piter, 2001, 656 p.

Оригинальность 91 %

Получено 11.02.2019

Принято 11.03.2019

Опубликовано 04.10.2019

**E.E. Fomina**

## **THE APPLICATION OF THE METHOD OF CORRESPONDENCE ANALYSIS IN SOCIO-ECONOMIC RESEARCH**

Mathematical modeling is widely used in economics, sociology, marketing, political science and other sciences to analyze the data and to obtain meaningful and well-grounded conclusions about their structure and relationships between variables describing a process or phenomenon. A fairly large set of mathematical methods for the study of numerical data presented in the interval scale or the scale of relations has been developed. Such methods include classification trees, discriminant, factor, cluster analysis, and others. However, it is often necessary to study categorical or non-numeric variables in socio-economic research, the main form of which is frequency tables. For numerical and visual analysis of large-size frequency tables can be used a correspondence analysis method based on chi-square statistics. The aim of this article is to demonstrate the capabilities of the method in the analysis and interpretation of the primary results of socio-economic research. The article deals with the mathematical foundations of a simple correspondence analysis and the practical use of the method in the processing of the survey results. The method is used to study the dynamics of financial activity of the population. Preferences of respondents in the sphere of services provided by various financial institutions, depending on age, are analyzed. A two-dimensional map of correspondences and a table with decision statistics were built, which allowed to carry out a visual and numerical analysis of the relationship between age categories and services of financial institutions. The most preferred services for each age group were identified. The results of analysis can be used to develop recommendations for the formation of consumer preferences in the financial services market.

*Keywords: correspondence analysis, frequency tables, categorical variables analysis, questionnaire.*

**Elena E. Fomina** – Candidate of Technical Sciences, Associate Professor, Department of Informatics and Applied Mathematics, Tver State Technical University, e-mail: f-elena2008@yandex.ru.

Received 11.02.2019

Accepted 11.03.2019

Published 04.10.2019