

Для цитирования: Экономика региона. — 2015. — №4. — С. 145-157.
doi 10.17059/2015-4-12
УДК 681.3.06

М. Б. Гузаиров, И. В. Дегтярева, Е. А. Макарова

Уфимский государственный авиационный технический университет (Уфа, Российская Федерация)

РАСХОДЫ НАСЕЛЕНИЯ РЕГИОНОВ РОССИЙСКОЙ ФЕДЕРАЦИИ НА ПОКУПКУ ПРОДУКТОВ ПИТАНИЯ: КОМПОНЕНТНЫЙ И КЛАСТЕРНЫЙ АНАЛИЗ¹

В статье рассматривается решение задачи проведения компонентного и кластерного анализа данных о расходах на приобретение продуктов питания как одной из важнейших составляющих уровня жизни населения. Цель анализа состоит в построении кластеров регионов Российской Федерации, различающихся по структуре расходов домашних хозяйств на покупку продуктов питания, представленных в абсолютных единицах, с учетом обобщенных показателей уровня жизни населения. Для исследования применяются такие методы интеллектуального анализа данных, как компонентный и кластерный анализ. Предложена процедура проведения интеллектуального анализа данных, которая основана на взаимосвязанном выполнении методов компонентного и кластерного анализа. Процедура интеллектуального анализа данных учитывает взаимосвязь между результатами, полученными разными методами, а также возможность возврата к предыдущему методу с целью его повторного применения для последовательного уточнения состава формируемых кластеров состояний домохозяйств. Выявлены малочисленные кластеры благополучных регионов, характеризующихся высоким и средним уровнем расходов домохозяйств на покупку основных продуктов питания, а также довольно многочисленные кластеры недостаточно благополучных и неблагополучных регионов, характеризующихся низким уровнем расходов домохозяйств. Показано, что рост уровня жизни, характеризующийся величиной валового регионального продукта на душу населения, сопровождается ростом коэффициента Джини, характеризующего неравенство в распределении доходов, а также снижением расходов на приобретение малоценных продуктов питания. Результаты анализа могут быть применены для разработки системы поддержки принятия решений, предназначенной для анализа сценариев макроэкономического регулирования в области политики доходов с целью повышения уровня жизни населения. Проведенный анализ данных о расходах населения на приобретение продуктов питания позволил выявить кластерную структуру регионов РФ, визуализировать ее в пространстве обобщенных признаков, описываемых главными компонентами, и сформулировать отличительные характеристики кластеров регионов, важные для принятия управленческих решений.

Ключевые слова: расходы домохозяйств, компонентный и кластерный анализ, кластеры регионов, диаграмма рассеивания

Введение

Современные условия развития российской экономики характеризуются наличием значительных проблем, связанных с обеспечением роста уровня и качества жизни населения, что в первую очередь, связано с ростом валового внутреннего продукта (ВВП).

Согласно прогнозу социально-экономического развития Российской Федерации (РФ) на 2013–2015 гг., опубликованному Министерством экономического развития, в качестве одного из основных путей стимулирования ро-

ста ВВП рассматривается сценарий стимулирования внутреннего потребительского спроса, который формируется на основе потребительских расходов домохозяйств². Сценарий потребительски ориентированного роста, который необходим в условиях рецессии, должен при появлении положительных тенденций в динамике ВВП преобразоваться в сценарий инвестиционно ориентированного роста [1–4]. Согласно предложенным в прогнозе Министерством экономического развития РФ сценариям, рост потребительского спроса необходимо обеспечить, в первую оче-

¹ © Гузаиров М. Б., Дегтярева И. В., Макарова Е. А. Текст. 2015.

² Прогноз социально-экономического развития Российской Федерации на 2013 год и плановый период 2014–2015 годов. [Электронный ресурс]. Доступ из справ.-правовой системы «КонсультантПлюс» (дата обращения 18.06.2015).

редь, за счет достижения устойчивых темпов роста потребления домашних хозяйств (на уровне 5,7 % в год), которые, в свою очередь, планируется получить за счет ускорения роста реальных располагаемых доходов населения. В этой связи актуальны исследования динамики процессов управления расходами и доходами сектора домохозяйств многосекторной макроэкономической системы, в рамках которых разрабатывается программное обеспечение для интеллектуальной системы поддержки принятия решений и имитационного моделирования процессов формирования доходов и расходов сектора домохозяйств многосекторной макроэкономической системы [5, 6].

Основными компонентами интеллектуальной системы поддержки принятия решений и имитационного моделирования являются модуль имитационного моделирования динамики доходов и расходов населения в рамках воспроизводственного процесса многосекторной макроэкономической системы и модуль интеллектуального анализа данных и поддержки принятия решений при управлении макроэкономической системой, особенностью которого является формирование обучающих выборок на основе результатов проведения имитационных экспериментов с динамической моделью многосекторной макроэкономической системы. Статистические данные использовались только для настройки параметров динамической модели, формирование кластеров домохозяйств выполнялось на основе анализа экспериментальных данных в динамике, полученных в результате проведения имитационных экспериментов. Однако целесообразно было бы на этапе формирования кластеров использовать не только экспериментальные данные, но и статистические данные о финансах сектора домохозяйств (в частности потребительских расходах), и только потом переходить к окончательному формированию кластеров домохозяйств. В связи с этим необходимым становится решение задачи разработки системы интеллектуального анализа данных о потребительских расходах домохозяйств, которая использует статистические данные о расходах домохозяйств, а также позволяет формировать кластеры домохозяйств и выявить их структурные особенности с учетом расходов населения по основным направлениям.

Методы

В статье рассматривается решение задачи интеллектуального анализа данных только о части потребительских расходов, направлен-

ных на приобретение продуктов питания как одного из важнейших составляющих уровня жизни населения.

Для исследования применяются такие методы интеллектуального анализа данных, как компонентный и кластерный анализ данных, которые позволяют визуализировать данные в пространстве меньшей размерности и выявить кластерную структуру регионов с целью формирования отличительных характеристик сформированных кластеров для последующего принятия решений в области управления доходами населения.

Выборка составлена на основе официальных статистических данных о структуре расходов на покупку продуктов питания за первый квартал 2012 г., публикуемых на сайте Федеральной службы государственной статистики¹. В качестве объекта исследования рассматриваются расходы домашних хозяйств на покупку продуктов питания, которые исследуются в разрезе регионов РФ.

Цель проводимого интеллектуального анализа состоит в построении кластеров регионов РФ, различающихся по структуре расходов домашних хозяйств на покупку продуктов питания, представленных в абсолютных единицах, с учетом обобщенных показателей уровня жизни населения. Выполнен анализ 83 объектов, представленных регионами РФ, то есть количество строк в выборке $n = 83$. Каждый объект (регион) характеризуется тринадцатью признаками, то есть количество столбцов в выборке $m = 13$. Особенность состава анализируемых (исходных) признаков состоит в том, что они включают две группы признаков.

К первой группе относятся признаки, характеризующие расходы на приобретение различных продуктов питания, представленные в абсолютных (денежных) единицах и соответствующие десяти направлениям потребительских расходов на приобретение продуктов питания: расходы на покупку хлебобулочных изделий и крупы; расходы на покупку мяса; расходы на покупку рыбы и морепродуктов; расходы на молочные изделия, сыр и яйца; расходы на покупку масла и жиров; расходы на фрукты, овощи; расходы на сахар (джем, мед, шоколад

¹ Доходы, расходы и потребление домашних хозяйств в I квартале 2012 года (по итогам выборочного обследования бюджетов домашних хозяйств). Статистический сборник. 2012 г. // Федеральная служба государственной статистики. [Электронный ресурс] URL: http://www.gks.ru/free_doc/doc_2012/bul_dr/dox_rasx12/dox1.rar (дата обращения 18.06.2015).

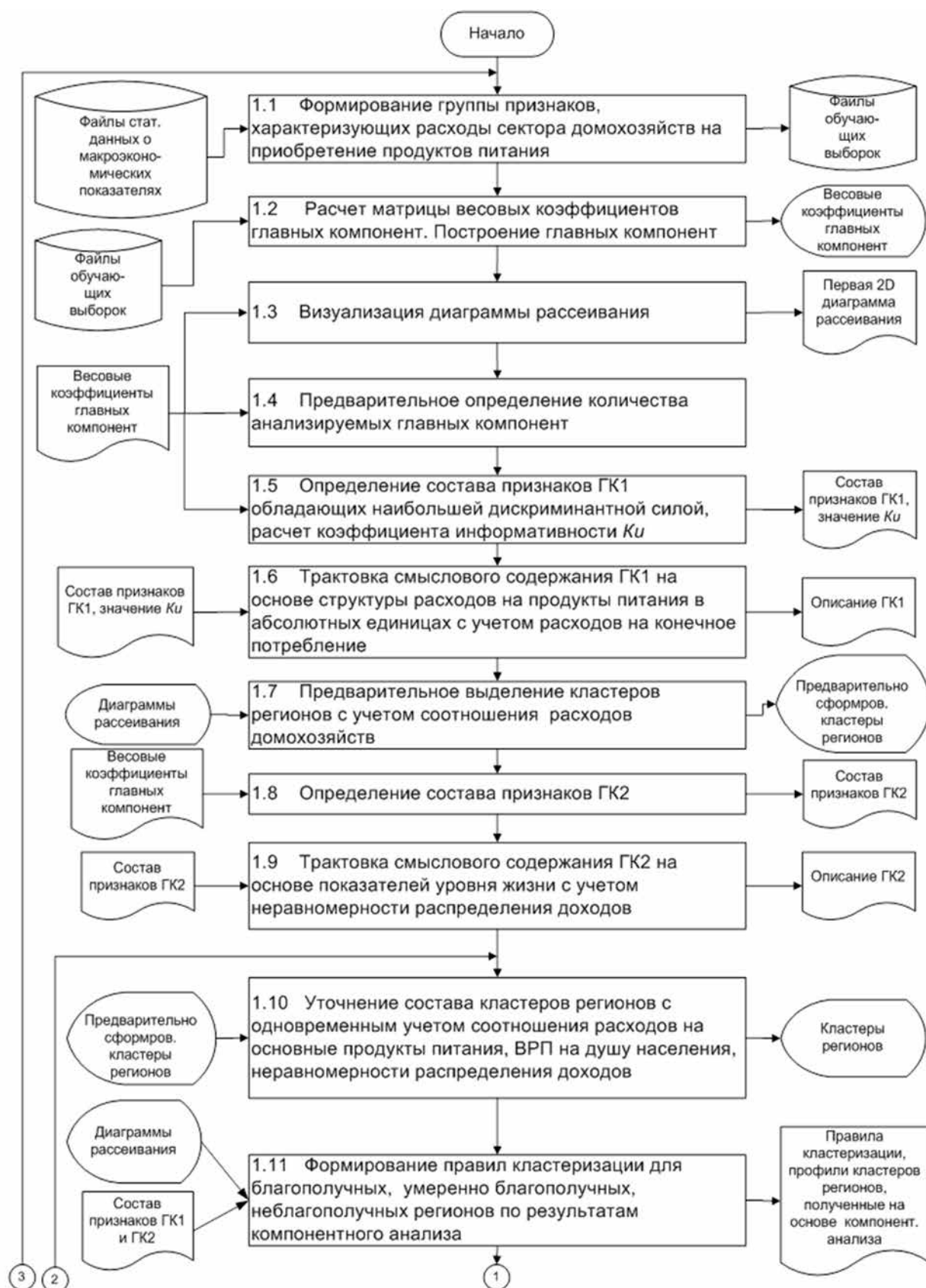


Рис. 1. Процедура взаимосвязанного проведения компонентного и кластерного анализа структуры расходов на продукты питания (продолжение на след. стр.)

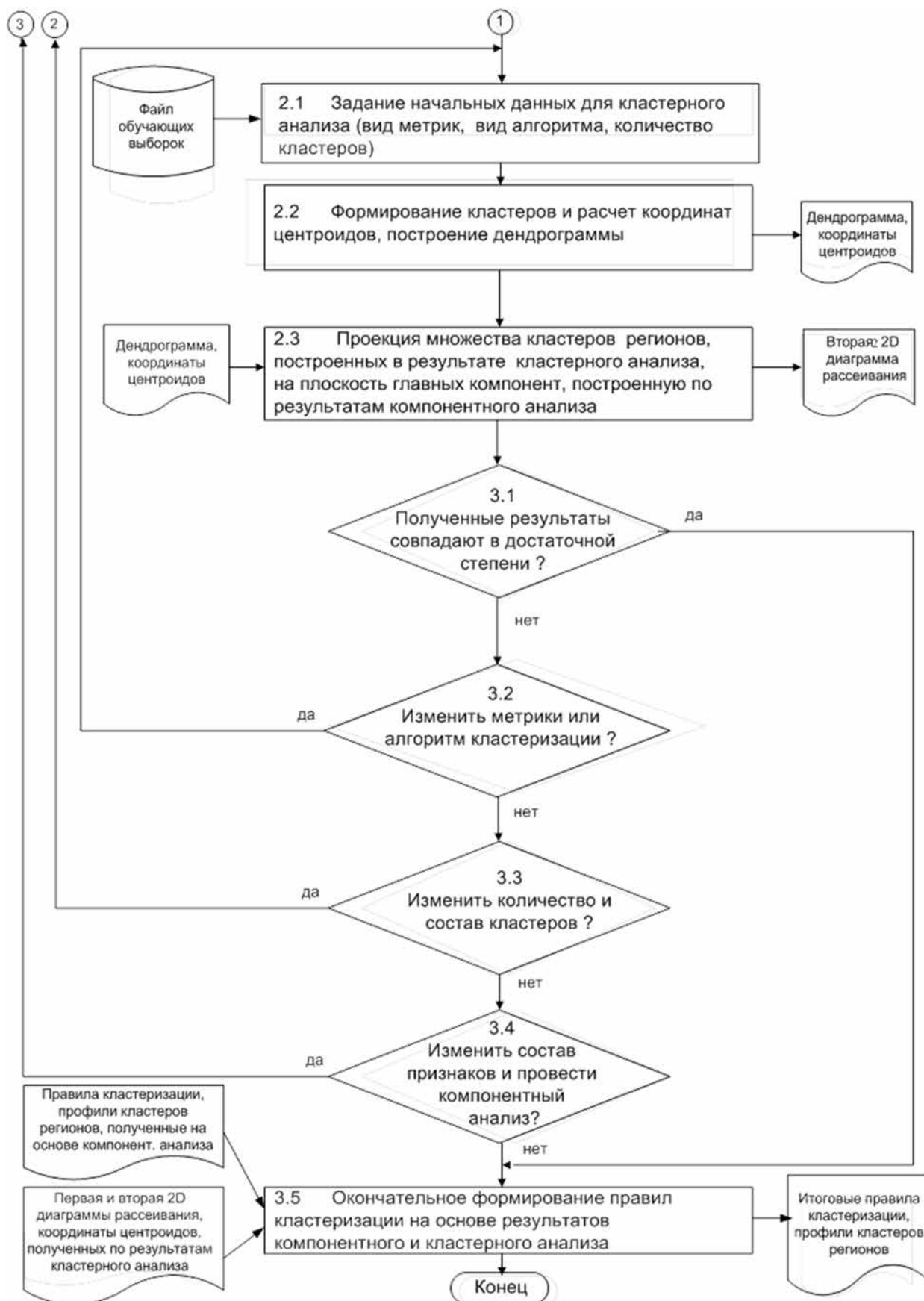


Рис. 1. (продолжение) Процедура взаимосвязанного проведения компонентного и кластерного анализа структуры расходов на продукты питания (ГК1 и ГК2 — главная компонента 1 и главная компонента 2 соответственно)

и конфеты); расходы на безалкогольные напитки; расходы на другие продукты питания.

Ко второй группе отнесены следующие три признака: расходы на конечное потребление (в среднем на члена домашнего хозяйства в месяц, рублей), валовой региональный продукт (ВРП) на душу населения (руб.); а также коэффициент Джини. Эти признаки предназначены для описания уровня жизни населения региона в целом с учетом неравномерности распределения доходов. Включение этих обобщенных характеристик позволяет дать оценку структуры расходов домохозяйств на продукты питания по кластерам регионов с различным уровнем жизни.

Процедура исследования с помощью методов интеллектуального анализа данных включает три этапа, каждый из которых предполагает выполнение множества шагов (рис. 1). Первый этап предназначен для проведения компонентного анализа, на втором этапе выполняется кластерный анализ данных о расходах на покупку продуктов питания сектора домохозяйств регионов РФ [7, 8]. Третий этап служит, во-первых, для согласования результатов, полученных двумя методами, во-вторых, для возврата на предыдущие шаги и повторного применения методов интеллектуального анализа данных с учетом возможных корректировок параметров. Процедура может выполняться с помощью различных программных продуктов, например, в средах Statgraphics или Statistica [9, 10].

Особенности выполнения компонентного анализа (этап 1) состоят в следующем.

Формирование группы признаков для проведения анализа на шаге 1.1 целесообразно выполнять не только методом экспертного оценивания, но и с привлечением компонентного анализа как метода предобработки данных на этапе их подготовки [7, 11, 12].

По результатам выполнения шагов 1.2 и 1.3 автоматически формируются главные компоненты, которые являются координатными осями нового пространства и математически представлены в виде линейной комбинации исходных признаков [7, 9]. Диаграмма рассеивания представляет собой проекцию множества объектов (регионов) в новое координатное пространство меньшей размерности. Более предпочтительным является анализ расположения регионов в двумерном пространстве с помощью 2D-диаграмм рассеивания. В случае необходимости, когда интерес представляет расположение регионов в трех- или четырехмерном пространстве, также целесообразно

использовать 2D-диаграммы, получаемые путем сочетания различных координатных осей. Все последующие шаги компонентного анализа выполняются в автоматизированном режиме с участием системного аналитика.

Количество анализируемых главных компонент, определяемое на шаге 1.4, может варьироваться от 1 до k , где $k < m$. Как правило, анализу подвергаются две или три главные компоненты в случае, если предполагается визуальный анализ пространства признаков. В проведенном исследовании рассматриваются первые две главные компоненты: первая главная компонента ГК1 и вторая главная компонента ГК2.

Шаги 1.5–1.7 посвящены формированию названия для главной компоненты 1 как новой интегральной характеристики, присущей данным и являющейся латентной. Формирование названий главных компонент выполняется на основе расчета коэффициента информативности k_i , значение которого должно приближаться к 0,75 справа [7, 9]. Процедура расчета коэффициента информативности k_i предполагает, что выбираются для анализа те признаки, которые обладают наибольшей дискриминантной силой. По результатам расчета k_i определяется тот набор признаков, который является достаточным для описания первой главной компоненты.

Предварительное выделение кластеров домохозяйств регионов выполняется на основе различий только по значениям первой главной компоненты ГК1.

Шаги 1.8–1.9 выполняются для анализа второй главной компоненты 2 и аналогичны предыдущим. Совмещение результатов деления на кластеры по первой и второй главным компонентам выполняется на шаге 1.10.

На шаге 1.11 формулируются закономерности, то есть характерные свойства, присущие выделенным кластерам домохозяйств регионов, что позволяет составить их профили. Кластеры объединяются в группы кластеров благополучных, умеренно благополучных и неблагополучных регионов.

Этап 2 предназначен для проведения кластерного анализа данных о расходах на покупку продуктов питания сектора домохозяйств регионов РФ [7, 13–15].

На шаге 2.1 выполняется настройка начальных данных для кластерного анализа. Для этого осуществляется выбор мер сходства и вида алгоритма кластеризации. Необходимо также задать количество кластеров для первоначального построения дендрограммы.

Формирование кластеров, расчет координат центроидов, построение дендрограммы выполняются автоматически (шаг 2.2).

Интерес представляет шаг 2.3, на котором предлагается выполнить нанесение кластеров, выделенных по результатам кластерного анализа, на плоскость главных компонент, построенную по результатам компонентного анализа. Итогом выполнения шага 2.3 является вторая 2D-диаграмма рассеивания. Построение второй 2D-диаграммы рассеивания имеет целью последующее сравнение результатов построения кластеров алгоритмами компонентного и кластерного анализа.

Этап 3 предполагает сравнение результатов кластеризации, полученных компонентным и кластерным анализом. В случае положительного результата сопоставления кластеров и их профилей (шаг 3.1), процедура завершается формированием окончательных выводов по кластеризации домохозяйств регионов и их профилей (шаг 3.5). В противном случае возможен возврат к повторному выполнению кластерного анализа (шаг 3.2) при условии корректировки параметров кластеризации (метрик, алгоритмов кластеризации).

В ситуации недостаточности варьирования параметров кластерного анализа выполняется возврат к повторному выполнению компонентного анализа. Может быть выполнена либо «мягкая» корректировка путем изменения количества и состава кластеров без перерасчета весовых коэффициентов, (шаг 3.3), либо «жесткая корректировка» путем изменения состава признаков и перерасчет весовых коэффициентов (шаг 3.4).

Таблица 1

Исходная сводка метода главных компонент

№ компоненты	Собственное значение	Процент отклонения	Одинарный процент
1	7,47705	57,516	57,516
2	1,42005	10,923	68,439
3	0,993111	7,639	76,079
4	0,738998	5,685	81,763
5	0,672779	5,175	86,938
6	0,492857	3,791	90,730
7	0,312159	2,401	93,131
8	0,246566	1,897	95,028
9	0,221044	1,700	96,728
10	0,156331	1,203	97,930
11	0,113087	0,870	98,800
12	0,0849406	0,653	99,454
13	0,0710202	0,546	100,000

Результаты

Далее представлены результаты выполнения предложенной процедуры интеллектуального анализа данных о структуре расходов на продукты питания.

По результатам выполнения этапа 1 построены три главные компоненты с использованием пакета StatGraphics, получена сводная характеристика для всех компонент с указанием собственных значений (табл. 1).

Приведенные цифры говорят о том, что первые две главные компоненты описывают 68,5 % дисперсии исходных данных. Третья главная компонента добавляет еще 7,5 % дисперсии, так что в сумме получается 76 % дисперсии, что достаточно для анализа пространственного распределения объектов.

Таблица 2

Весовые коэффициенты признаков для трех главных компонент

№	Признак (расходы на различные виды продуктов питания)	Компонента 1	Компонента 2	Компонента 3
1	Хлебобулочные изделия и крупы	0,319126	0,20463	0,0687145
2	Мясо	0,318424	-0,0015162	-0,162666
3	Рыба и морепродукты	0,255102	-0,130902	-0,339477
4	Молочные изделия (сыр и яйца)	0,344796	0,0512245	-0,0662633
5	Масла и жиры	0,225393	-0,0432492	-0,526929
6	Фрукты	0,294104	0,0680102	0,17974
7	Овощи	0,289657	-0,243074	-0,270198
8	Сахар (джем, мед, конфеты)	0,119769	0,557114	0,341807
9	Другие продукты питания	0,283394	0,356343	0,0350852
10	Безалкогольные напитки	0,332408	0,106361	0,120093
11	ВРП на душу населения	0,2177	-0,370007	0,371906
12	Расходы на конечное потребление (на члена домохозяйства).	0,33383	-0,0693131	0,123349
13	Коэффициент Джини	0,16618	-0,531418	0,4245

Весовые коэффициенты всех признаков для трех главных компонент представлены в таблице 2. На основе полученных данных о весовых коэффициентах выполнен расчет коэффициентов информативности для всех трех главных компонент. Для каждой компоненты определен состав признаков, определяющих ее смысловое содержание (выделены жирным курсивом).

Показано, что первая главная компонента 1 характеризует в целом расходы на приобретение основных продуктов питания (расходы на покупку мяса, молочных изделий, хлебобулочных изделий, фруктов, овощей) с учетом суммарных расходов на конечное потребление. В области больших значений главной компоненты 1 находятся регионы, которые характеризуются значительными расходами в абсолютных величинах на приобретение этих продуктов питания, являющихся ценными и важными (рис. 2). Это регионы, расположенные в крайней правой области, находящиеся в кластерах пять, шесть, семь, восемь, девять (г. Москва, Ямало-Ненецкий автономный округ, г. Санкт-Петербург, Сахалинская область, Магаданская область, Чукотский автономный округ и другие).

Главная компонента 2 характеризует совокупные расходы на продукты питания, содержащие углеводы (сахар), с учетом ВРП на душу населения и неравномерности распределения доходов (отраженной в индексе Джини).

Необходимо выделить характерную особенность для регионов, расположенных в крайних областях (высоких или низких значениях

главной компоненты 2). Согласно полученным знакам при весовых коэффициентах наиболее значимых признаков показано, что рост уровня жизни сопровождается ростом степени неравномерности распределения доходов и снижением расходов на покупку углеводосодержащих продуктов питания.

Регионам, которые находятся в области малых значений главной компоненты 2 (например, г. Москва), присущ более высокий уровень жизни, однако и большая дискриминация в распределении доходов, а также малые расходы на продукты, содержащие углеводы. Для регионов в области больших значений главной компоненты 2, например, для Чеченской республики, присущ малый ВРП на душу населения (низкий уровень жизни), а также малая степень неравномерности распределения доходов (малый коэффициент Джини) и достаточно высокие расходы на приобретение продуктов, содержащих углеводы.

Третья главная компонента характеризуется значением ВРП на душу населения, коэффициентом Джини и расходами на продукты питания, содержащие жиры. Особенность взаимодействия выделенных признаков состоит в том, что с ростом ВРП на душу населения увеличивается и неравномерность распределения доходов, а также снижаются расходы на покупку продуктов питания, содержащих жиры.

По результатам компонентного анализа выделены девять кластеров регионов. В пятый, седьмой, восьмой, девятый кластер вошли регионы, характеризующиеся высоким уровнем расходов на ценные продукты питания и ма-

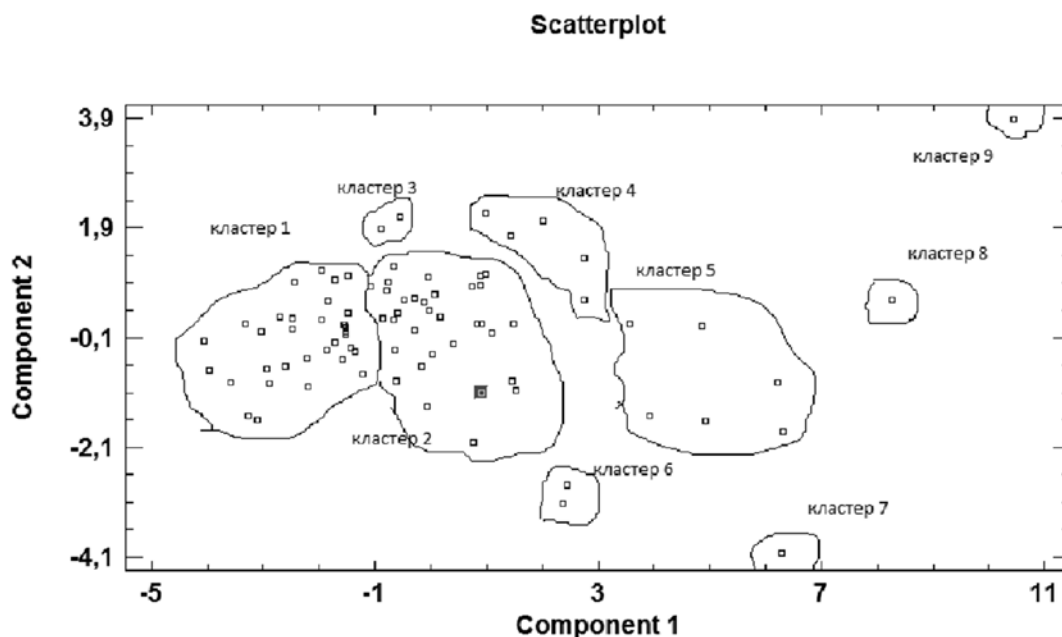


Рис. 2. Первая 2D-диаграмма рассеивания (результат компонентного анализа)

лыми расходами на углеводсодержащие продукты питания. В третий, четвертый, шестой кластеры вошли регионы, характеризующиеся средним уровнем расходов на ценные продукты питания. Во второй кластер вошли регионы, характеризующиеся низким или средним уровнем расходов на ценные продукты питания.

И, наконец, в первый кластер вошли регионы, характеризующиеся низким уровнем расходов на продукты питания, а также низким уровнем расходов на конечное потребление в целом (Республика Ингушетия и Дагестан).

Структурные особенности расположения кластеров в новом двумерном пространстве состоят в следующем. Справа на 2D-диаграмме расположены кластеры с малым числом регионов с высокими расходами на приобретение продуктов питания и ВРП на душу населения. По мере движения влево населенность (число объектов в кластерах) растет, а характеристики по расходам на приобретение продуктов питания и ВРП на душу населения снижаются. В крайней левой части на диаграмме расположен довольно многочисленный кластер регионов с низким в целом уровнем жизни.

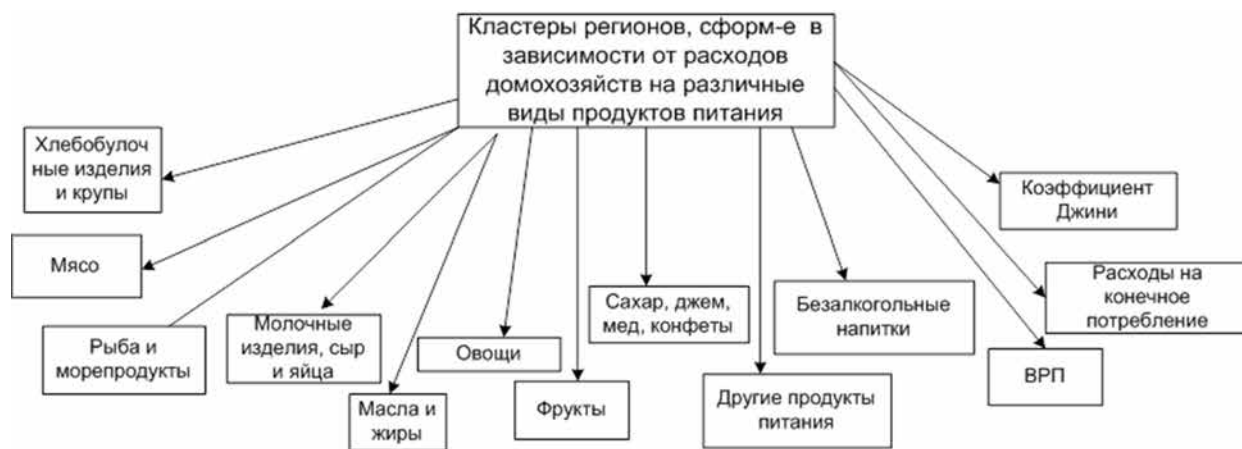


Рис. 3. Дерево зависимости номера кластера от множества анализируемых признаков расходов на покупку продуктов питания (двухуровневое)



Рис. 4. Дерево зависимости номера кластера от множества анализируемых признаков расходов на покупку продуктов питания (трехуровневое)

Таблица 3
Профили кластеров в пространстве главных компонент

№ кластера	Значение главной компоненты 1	Значение главной компоненты 2
Кластер 1	Низкое	Среднее или высокое
Кластер 2	Среднее	Среднее или высокое
Кластер 3	Среднее	Очень высокое
Кластер 4	Среднее	Высокое или очень высокое
Кластер 5	Высокое	Среднее или высокое
Кластер 6	Среднее	Низкое
Кластер 7	Высокое	Низкое
Кластер 8	Очень высокое	Высокое
Кластер 9	Очень высокое	Очень высокое

На рисунке 3 представлено двухуровневое зависимости номера кластера регионов от множества анализируемых признаков расходов на покупку продуктов питания для ситуации, когда еще не проведен компонентный анализ. Далее, на рисунке 4 представлено трехуровневое дерево зависимости номера кластера регионов от множества анализируемых признаков расходов на покупку продуктов питания для ситуации, когда по результатам проведения компонентного анализа выполнена структуризация не только объектов в пространстве признаков, но и самих признаков. Результатом структуризации признаков является их объединение в новые интегральные признаки — главные компоненты.

Профили кластеров, составленные с учетом выделенных зон в пространстве главных компонент (рис. 5), представлены в таблице 3.

Правила кластеризации составляются на основе данных таблицы 2 и с учетом состава признаков, определяющих главные компоненты. Пример правила для первого кластера имеет следующий вид: «Если расходы на покупку молочных изделий, безалкогольных напитков, хлебобулочных изделий и крупы, мяса, фруктов, овощей, других продуктов питания низкие, и при этом расходы на конечное потребление низкие, расходы на покупку сахара (джема, меда, конфет) средние или высокие, коэффициент Джини низкий, ВРП на душу населения низкий или средний, то этот субъект относится к кластеру 1».

По результатам этапа 2 предложенной процедуры, на котором выполняется кластерный анализ, принято окончательное решение разделить все регионы на семь кластеров. Это решение принято на основе сопоставления диаграммы рассеивания компонентного анализа и дендрограммы кластерного анализа. Для наглядного представления окончательного разделения множества регионов на кластеры приведена вторая 2D-диаграмма рассеивания в пространстве главной компоненты, составленной на основе дендрограммы кластерного анализа (рис. 6). Отметим, что выделение кластеров соответствует в большей степени делению объектов по главной компоненте 1 как обладающей наибольшей дискриминантной силой.

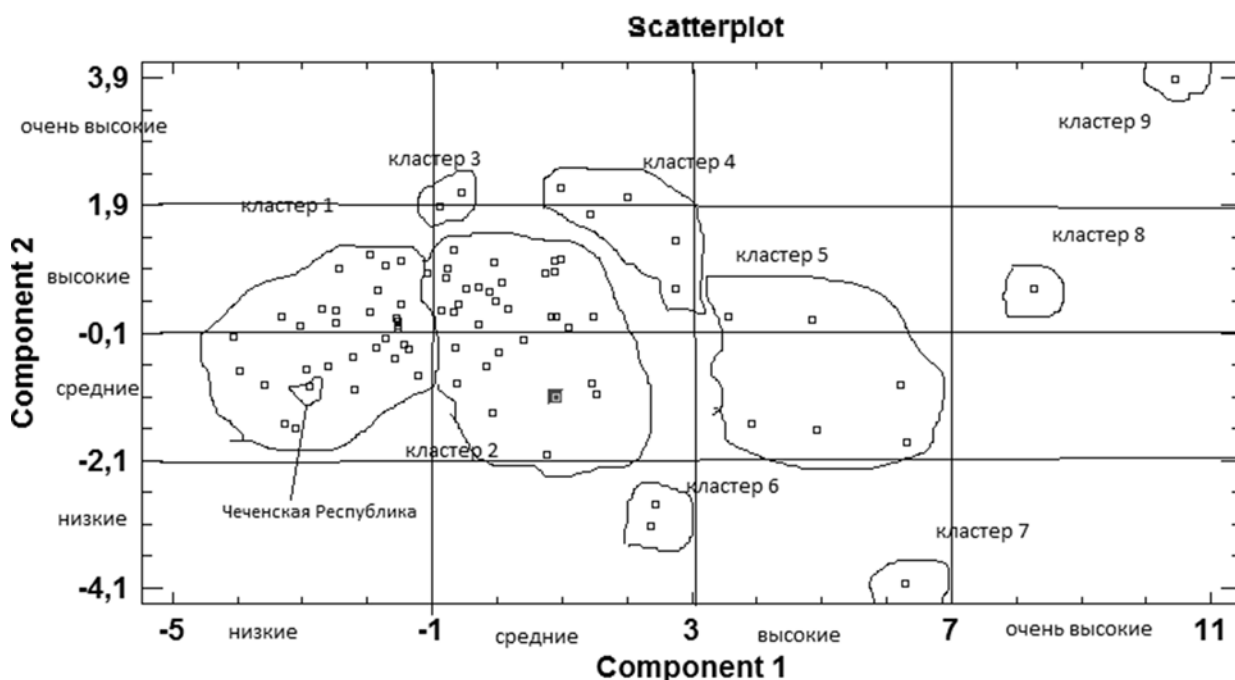


Рис. 5. Первая 2D-диаграмма рассеивания с учетом выделенных зон (результат компонентного анализа)

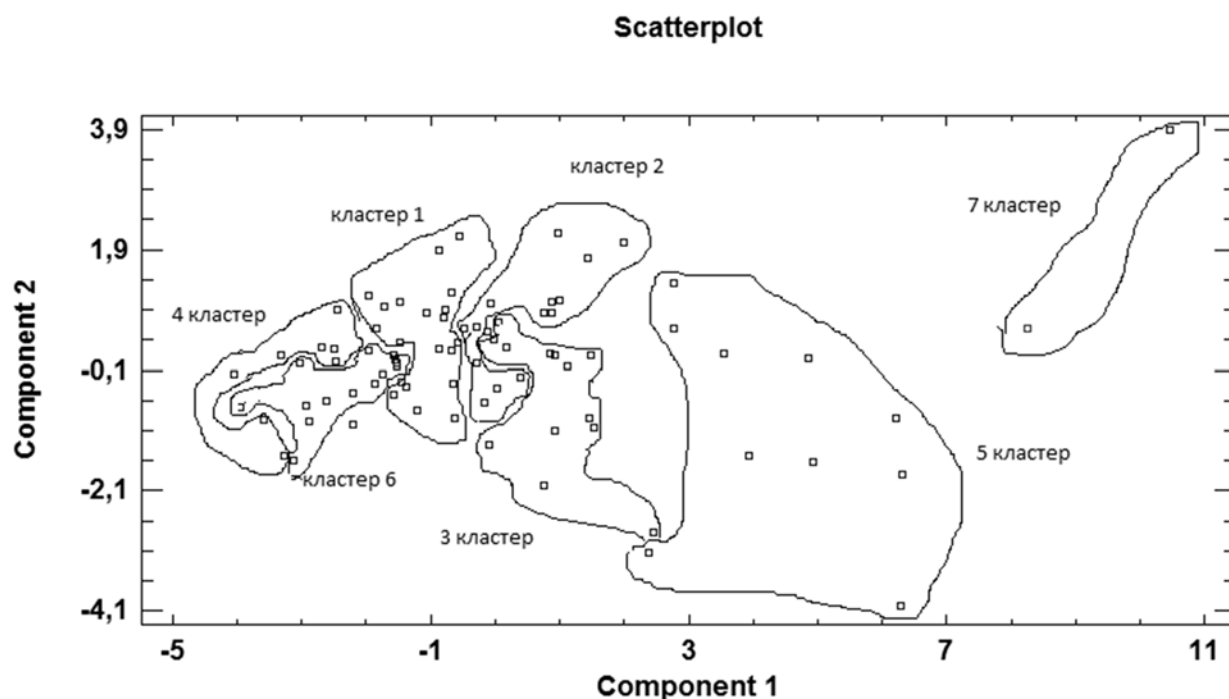


Рис. 6. Вторая 2D-диаграмма рассеивания (результат кластерного анализа)

В седьмом кластере, построенном по результатам кластерного анализа, находятся два региона, которые в компонентном анализе составляли два самостоятельных отдельных кластера. Поскольку они находятся близко друг к другу, в кластерном анализе произошло их объединение в один кластер, это Ямало-Ненецкий автономный округ и Камчатский край.

Пятый кластер, построенный по результатам кластерного анализа, включает такие регионы, которые имеют большое значение главной компоненты 1. Эти регионы соответствуют пятому, шестому, седьмому кластерам, и близко расположенной к нему части четвертого кластера, которые построены в результате проведения компонентного анализа. К пятому кластеру относятся десять регионов: Мурманская область, г. Санкт-Петербург, Приморский край, Хабаровский край, Сахалинская область, Республика Саха, г. Москва, Ненецкий автономный округ, Магаданская обл., Чукотский автономный округ.

Третий кластер, построенный в кластерном анализе, включает в себя всю оставшуюся часть второго кластера и близлежащую часть шестого кластера, построенных по результатам компонентного анализа, ему соответствует среднее значение главной компоненты 1. К третьему кластеру относятся десять регионов: Московская область, Республика Коми, Пермский край, Самарская область, Свердловская область, Тюменская область, Забайкальский край, Красноярский край, Иркутская

область, Ханты-Мансийский автономный округ.

Второй кластер включает в себя части второго и четвертого кластеров, построенных по результатам компонентного анализа. Ко второму кластеру относятся двадцать регионов: Брянская область, Ивановская область, Калужская область, Костромская область, Тверская область, Тульская область, Ярославская область, Республика Карелия, Архангельская область, Волгоградская область и др.

Первый кластер включает частично кластеры один, два, три, построенные по результатам компонентного анализа. К первому кластеру относятся двадцать четыре региона: Белгородская область, Владимирская область, Воронежская область, Курская область, Липецкая область, Орловская область, Рязанская область, Смоленская область, Волгоградская область, Республика Башкортостан и др.

Четвертый и шестой кластеры, построенные в результате кластерного анализа, характеризуются малыми значениями главной компоненты 2 и включают в основном первый кластер и близко расположенный к нему второй кластер, построенные в результате компонентного анализа. К шестому кластеру относятся восемь регионов: Республика Адыгея, Республика Дагестан, Республика Ингушетия, Кабардино-Балкарская Республика, Карачаево-Черкесская Республика, Осетия-Алания, Чеченская Республика, Ставропольский край. К четвертому кластеру относятся девять реги-

онов: Тамбовская обл., Республика Калмыкия, Республика Марий Эл, Мордовия, Удмуртская Республика, Республика Бурятия, Республика Тыва и др.

Обсуждение

Предложенная процедура проведения интеллектуального анализа данных о структуре расходов на продукты питания основана на взаимосвязанном и циклическом выполнении методов компонентного и кластерного анализа данных о расходах населения на приобретение продуктов питания. Особенность предложенной процедуры анализа данных состоит, во-первых, в учете взаимосвязей между результатами, полученными разными методами; во-вторых, в возможности возврата к предыдущему методу с целью его повторного выполнения для последовательного уточнения состава формируемых кластеров состояний домохозяйств. Процедура может быть применена многократно для нескольких обучающих выборок, полученных путем структуризации статей расходов на продукты питания. Это позволяет выполнить построение кластеров домохозяйств и дать им интегральную характеристику за счет снижения размерности исходного признакового пространства.

По результатам компонентного анализа выделены девять кластеров регионов, отличающиеся уровнем расходов на ценные продукты питания, расходами на углеводсодержащие продукты питания, а также уровнем расходов на конечное потребление в целом.

Кластерный анализ позволил уточнить и скорректировать количество кластеров и их

границы, которые были получены на этапе предварительного анализа методом главных компонент. Правила классификации, которые сформулированы на основе данных о координатах центроидов, соответствуют результатам компонентного анализа.

По результатам анализа данных о расходах домохозяйств регионов на покупку различных продуктов питания выявлены малочисленные кластеры благополучных регионов РФ, характеризующихся высоким и средним уровнем расходов домохозяйств на покупку основных продуктов питания, а также довольно многочисленные кластеры недостаточно благополучных и неблагополучных регионов РФ, характеризующихся низким уровнем анализируемых расходов домохозяйств.

Показано, что рост уровня жизни, характеризующийся объемом ВРП на душу населения, сопровождается ростом степени неравномерности распределения доходов и снижением расходов на приобретение малоценных продуктов питания. Население регионов, характеризующихся высоким уровнем жизни, имеет в структуре расходов на питание значительную долю ценных продуктов питания, которая снижается и уступает место малоценным продуктам питания в регионах с низким уровнем жизни.

Результаты проведенного анализа могут быть применены для разработки системы поддержки принятия решений, предназначенной для формирования рекомендаций в области государственного регулирования продовольственного обеспечения с целью повышения уровня жизни населения.

Список источников

1. Иванов В. Н., Суворов А. В. Неравенство и бедность населения. Опыт решения проблемы в России и за рубежом // Проблемы прогнозирования. — 2006. — №3. — С. 132–149.
2. Колмаков И. Б. Прогнозирование показателей дифференциации денежных доходов населения // Проблемы прогнозирования. — 2006. — №1 — С. 136–163.
3. Римашевская Н. М. Некоторые проблемы социального реформирования в России // Проблемы прогнозирования. — 2006. — № 2.– С. 3–17.
4. Шевяков А. Ю. Неравенство доходов как фактор экономической и демографической динамики. — М.: ИСЭПН РАН, 2010 — 43 с.
5. Системное моделирование динамики формирования доходов и расходов населения с учетом их дифференциации / Ильясов Б. Г., Дегтярева И. В., Макарова Е. А., Валитов Р. Р. // Проблемы управления и моделирования в сложных системах : тр. XV междунар. конф. (19–22 июня 2012 г.). — Самара : Самарск. науч. центр РАН, 2013. — С. 179–193.
6. Система интеллектуальной поддержки принятия решений при управлении макроэкономическим воспроизводственным процессом на основе имитационного моделирования / Ильясов Б. Г., Дегтярева И. В., Макарова Е. А., Валитов Р. Р. // Вестник УГАТУ. — 2012. — № 3. — С. 217–229.
7. Анализ данных и процессов: учеб. пособие / А. А. Барсебян, М. С. Куприянов, И. И. Холод, М. Д. Тесс, С. И. Елизаров. — 3-е изд., перераб. и доп. — СПб. : БХВ-Петербург, 2009. — 512 с.
8. Демидова Л. А., Кираковский В. В., Пылькин А. Н. Принятие решений в условиях неопределенности. — М. : Горячая линия — Телеком, 2012. — 288 с.

9. Кулаицев А. П. Методы и средства комплексного анализа данных: учебное пособие; 4-е изд., перераб. и доп. — М. : ФОРУМ ; Инфра-М, 2013. — 312 с.
10. Компьютерный анализ и интерпретация эмпирических зависимостей : учеб. пособие / С. В. Поршневу, Е. В. Овечкина, М. В. Мащенко и др. — М. : ООО «Бином-Пресс», 2010. — 336 с.
11. Mitra S., Acharya T. Data Mining. Multimedia, Soft Computing, and Bioinformatics. Hoboken, New Jersey, John Wiley & Sons, Inc. 2003. 401 p.
12. Han J., Kamber M. Data mining: Concepts and Techniques. — San Mateo: Morgan Kaufmann Publishers, 2001 — 550 p.
13. Kanungo T., Mount D. M., Netanyahu N. S., Piatko Cr. D., Silverman R., Wu An Y. An Efficient k-Means Clustering Algorithm // Analysis and Implementation IEEE transactions on pattern analysis and machine intelligence, 2002. Vol. 24, No. 7, July. P. 881–892.
14. Ruppert D. Statistics and Data Analysis for Financial Engineering. / Ruppert, D. — New York: Springer Science+Business Media, LLC 2011. — p. 638.
15. Vercellis C. Business Intelligence: Data Mining and Optimization for Decision Making. / Vercellis, C. — New York: John Wiley & Sons Ltd, 2009. — p. 436.

Информация об авторах

Гузайров Мурат Бакеевич — доктор технических наук, профессор кафедры вычислительной техники и защиты информации, Уфимский государственный авиационный технический университет (Российская Федерация, 450000, Республика Башкортостан, г. Уфа, ул. Карла Маркса, 12; e-mail: guzairov@mail.rb.ru)

Дегтярева Ирина Викторовна — доктор экономических наук, профессор, заведующий кафедрой экономической теории, Уфимский государственный авиационный технический университет (Российская Федерация, 450000, Республика Башкортостан, г. Уфа, ул. Карла Маркса, 12; e-mail: degtjareva@mail.rb.ru)

Макарова Елена Анатольевна — доктор технических наук, доцент, профессор кафедры технической кибернетики, Уфимский государственный авиационный технический университет (Российская Федерация, 450000, Республика Башкортостан, г. Уфа, ул. Карла Маркса, 12; e-mail: ea-makarova@mail.ru)

For citation: *Ekonomika regiona* [Economy of Region]. — 2015. — No 4. — pp. 145-157.

M. B. Guzairov, I. V. Degtyareva, E. A. Makarova

Regional Population Expenditure for Foodstuffs in the Russian Federation: Componential and Cluster Analyses

The article describes the solving of the problem of conducting the componential and cluster analyses of population expenditure for foodstuffs as one of the most important components of the standard of living. The purpose of the analysis is to develop the regional clusters of the Russian Federation, which vary in the structure of household expenditure for foodstuffs. The foodstuffs are presented in absolute units taking into integral account the standard of living index. The methods of intellectual analysis such as component and cluster analyses are applied as the research methods. The procedure for the data intellectual analysis based on the interconnected performance of component and cluster analyses is proposed. The procedure of the data intellectual analysis considers the interrelation between the results received by different methods, and also the possibility to return to the previous method for the purpose of repeating the analysis to specify consistently the clusters composition.

Few clusters of the wealthy regions characterized by the high and average level of expenditure for foodstuffs are revealed as well as the quite many clusters of not enough wealthy and not wealthy regions characterized by the low level of expenditure for foodstuffs. It is shown that the growth of standard of living characterized by the size of a gross regional product per capita is followed by the growth of the Gini coefficient, which indicates both the inequality of income distribution and reduction in expenditure for low-value foodstuffs.

The results of the analysis can be applied for the development of the decision making support system intended for the analysis of the scenarios of macroeconomic regulation in the field of income policy for the purpose of increasing the standard of living of the population. The analysis of the population expenditure for foodstuffs has allowed to reveal the cluster structure of the regions of the Russian Federation, to show it according to the generalized indications, to formulate the specific characteristics of the clusters of the regions and important management decisions.

Keywords: expenditure of households, component and cluster analyses, clusters of regions, scatterplot

References

1. Ivanov, V. N. & Suvorov, A. V. (2006). Neravenstvo i bednost naseleniya. Opyt resheniya problemy v Rossii i za rubezhom [Inequality and poverty of the population: the experience of solving problems in Russia and abroad]. *Problemy prognozirovaniya* [Problems of forecasting], 3, 132–149.
2. Kolmakov, I. B. (2006). Prognozirovaniye pokazateley differentsiatsii denezhnykh dokhodov naseleniya [Forecasting of cash income differentiation]. *Problemy prognozirovaniya* [Problems of forecasting], 1, 136–163.
3. Rimashevskaya, N. M. (2006). Nekotoryye problemy sotsialnogo reformirovaniya v Rossii [Some of the problems of social reform in Russia]. *Problemy prognozirovaniya* [Problems of forecasting], 2, 3–17.

4. Shevyakov, A. Yu. (2010). *Neravenstvo dokhodov kak faktor ekonomicheskoy i demograficheskoy dinamiki [Income inequality as a factor of economic and demographic dynamics: a monograph]*. Moscow: ISEPN RAN Publ., 43.
5. Ilyasov, B. G., Degtyareva, I. V., Makarova, E. A. & Valitov, R. R. (2013). Sistemnoye modelirovaniye dinamiki formirovaniya dokhodov i raskhodov naseleniya s uchyotom ikh differentsiatsii [System modeling of the dynamics of revenues and expenditures of the population with regard to their differentiation]. *Problemy upravleniya i modelirovaniya v slozhnykh sistemakh: tr. XV mezhdunar. konf. (19–22 iyunya 2012 g.) [Issues of management and modelling in the complex systems: proceedings of the 15th international conference (19–22 of June, 1012)]*. Samara: Samarsk. nauch. tsentr RAN Publ., 179–193.
6. Ilyasov, B. G., Degtyareva, I. V., Makarova, E. A. & Valitov, R. R. (2012). Sistema intellektualnoy podderzhki prinyatiya resheniy pri upravlenii makroekonomicheskim vosproizvodstvennym protsessom na osnove imitatsionnogo modelirovaniya [System intelligent decision support in the management of macro-economic reproduction process on the basis of simulation]. *Vestnik UGATU [Bulletin of the Ufa State Aviation Technical University]*, 3, 217–229.
7. Barsegyan, A. A., Kupriyanov, M. S., Kholod, I. I., Tess, M. D. & Elizarov, S. I. (2009). *Analiz dannykh i protsessov: ucheb. posobie [Analysis of data and processes: proc. manual]*. 3rd ed., Rev. and ext. St. Petersburg: BKhV-Peterburg Publ., 512.
8. Demidova, L. A., Kirakovskiy, V. V. & Pylkin, A. N. (2012). *Prinyatie resheniy v usloviyakh neopredelennosti [Decision-making under uncertainty]*. Moscow: Goryachaya liniya — Telekom Publ., 288.
9. Kulaichev, A. P. (2013). *Metody i sredstva kompleksnogo analiza dannykh: uchebnoe posobie; 4-e izd., pererab. i dop. [Methods and tools for complex data analysis: a tutorial]*, 4th ed., rev. and ext. Moscow: FORUM Publ.; Infra-M Publ., 312.
10. Porshnev, S. V., Ovechikina, E. V., Mashchenko, M. V. et al. (2010). *Kompyuternyy analiz i interpretatsiya empiricheskikh zavisimostey: ucheb. posobie [Computer analysis and interpretation of empirical relationships: textbook]*. Moscow: Binom-Press Publ., 336.
11. Mitra, S. & Acharya, T. (2003). *Data Mining. Multimedia, Soft Computing, and Bioinformatics*. Hoboken, New Jersey, John Wiley & Sons, Inc., 401.
12. Han, J. & Kamber, M. (2001). *Data mining: Concepts and Techniques*. San Mateo: Morgan Kaufmann Publishers, 550.
13. Kanungo, T., Mount, D. M., Netanyahu, N. S., Piatko, Cr. D., Silverman, R., Wu, A. Y. (2002, July). An Efficient k-Means Clustering Algorithm. *Analysis and Implementation IEEE transactions on pattern analysis and machine intelligence*, 24(7), 881–892.
14. Ruppert, D. (2011). *Statistics and Data Analysis for Financial Engineering*. New York: Springer Science+Business Media, LLC, 638.
15. Vercellis, C. (2009). *Business Intelligence: Data Mining and Optimization for Decision Making*. New York: John Wiley & Sons Ltd, 436.

Authors

Guzairov Murat Bakeevich — Doctor of Technical Science, Professor, Faculty of computer science and robotics, Ufa State Aviation Technical University (12, Karla Marksa St., Ufa, Republic of Bashkortostan, 450000, Russian Federation; e-mail: guzairov@mail.rb.ru).

Degtyareva Irina Viktorovna — Doctor of Economics, Professor, Head of the Chair of Economic Theory, Ufa State Aviation Technical University (12, Karla Marksa St., Ufa, Republic of Bashkortostan, 450000, Russian Federation; e-mail: degtyareva@mail.rb.ru).

Makarova Elena Anatolyevna — Doctor of Technical Sciences, Associate Professor, Professor of the Chair of Engineering Cybernetics, Ufa State Aviation Technical University (12, Karla Marksa St., Ufa, Republic of Bashkortostan, 450000, Russian Federation; e-mail: ea-makarova@mail.ru).