

DOI: 10.18287/2542-0461-2020-11-2-166-178

УДК 519.24



Научная статья / Scientific article

Дата: поступления статьи / Submitted: 12.02.2020

после рецензирования / Revised: 02.04.2020

принятия статьи / Accepted: 25.05.2020

**А.Ю. Трусова**

Самарский национальный исследовательский университет  
имени академика С.П. Королева, г. Самара, Российская Федерация  
E-mail: a\_yu\_ssu@mail.ru. ORCID: <https://orcid.org/0000-0001-7679-9902>

## Методология многомерной весовой кластеризации

**Аннотация:** Комбинированный подход при анализе социально-экономической сферы базируется на применении информационно-математических методов. Это позволяет расширить возможности монотетического подхода, например при структурировании объектов исследования или показателей, описывающих изучаемые объекты. В настоящее время такой подход является частью бизнес-аналитики. В статье применен многомерный подход при классификации многомерных данных. Классические методы классификации дополнены введением латентных факторов. Описана процедура классификации на латентных факторах без учета весовых коэффициентов. В дополнение к экспертным методам предложена методика оценки весовых коэффициентов при многомерной кластеризации. С учетом структуры латентных интегральных показателей рассматривается методология формирования весовых коэффициентов. Предлагается структуризация с применением методики неметрического многомерного шкалирования, в том числе с учетом весовых коэффициентов. Учитывая ранговые позиции объектов кластеризации, автор проводит группировку в теоретическом пространстве стимулов. Долевое соотношение показателей как в латентных факторах, так и в теоретическом шкальном пространстве позволяет исключить применение экспертных оценок и их субъективность. В статье также отмечается возможность совмещения и динамического анализа при исследовании многомерных массивов данных. В качестве объекта изучения рассматриваются показатели непродушенной сферы Приволжского федерального округа.

**Ключевые слова:** классификации многомерных данных, латентные факторы, формирование весовых коэффициентов, неметрическое многомерное шкалирование, монотетический подход при кластеризации, политетический подход при кластеризации, динамический анализ, показатели социально-экономической сферы.

**Цитирование.** Трусова А.Ю. Методология многомерной весовой кластеризации // Вестник Самарского университета. Экономика и управление. 2020. Т. 11. № 2. С. 166–178. DOI: <http://doi.org/10.18287/2542-0461-2020-11-2-166-178>.

**Информация о конфликте интересов:** автор заявляет об отсутствии конфликта интересов.

**A.Yu. Trusova**

Samara National Research University, Samara, Russian Federation  
E-mail: a\_yu\_ssu@mail.ru. ORCID: <https://orcid.org/0000-0001-7679-9902>

## Multi-dimensional weight clustering methodology

**Abstract:** The combined approach to the analysis of the socio-economic sphere is based on the use of information-mathematical methods. This allows you to expand the possibilities of a monothetic approach, for example, when structuring objects of research or indicators describing the studied objects. This approach is currently part of business intelligence. A multidimensional approach is applied in the classification of multidimensional data. Classical classification methods are complemented by the introduction of latent factors. The classification procedure based on latent factors without weight factors is described. In addition to expert methods, a methodology for estimating weight coefficients in multidimensional clustering is proposed. Given the structure of latent integrated indicators, the methodology of formation of weighting factors is considered. The paper proposes structuring using non-metric multidimensional scaling techniques, including taking into account weighting factors. Given the ranking positions of clustering objects, a grouping is carried out in the theoretical space of stimuli. The share ratio of indicators both in latent factors and in the theoretical scale space allows us to exclude the use of expert estimates and their subjectivity. The paper also notes the possibility of combining and dynamic analysis in the analysis of multidimensional data arrays. As an object of study, indicators of the non-production sphere of the Volga Federal District are considered.

**Key words:** classification of multidimensional data, latent factors, formation of weighting coefficients, non-metric multidimensional scaling, monothetic approach to clustering, polythetic approach to clustering, dynamic analysis, indicators of the socio-economic sphere.

**Citation.** Trusova A. Yu. Multi-dimensional weight clustering methodology. *Vestnik Samarskogo universiteta. Ekonomika i upravlenie = Vestnik of Samara University. Economics and Management*, 2020, vol. 11, no. 2, pp. 166–178. DOI: <http://doi.org/10.18287/2542-0461-2020-11-2-166-178>. (In Russ.)

**Information on the conflict of interest:** author declares no conflict of interest.

#### ИНФОРМАЦИЯ ОБ АВТОРЕ / INFORMATION ABOUT THE AUTHOR

© Алла Юрьевна Трусова – кандидат физико-математических наук, доцент, доцент кафедры математики и бизнес-информатики, Самарский национальный исследовательский университет имени академика С.П. Королева, 443086, Российская Федерация, г. Самара, Московское шоссе, 34.

© Alla Yu. Trusova – Candidate of Physical and Mathematical Sciences, associate professor, associate professor of the Department of Mathematics and Business Informatics, Samara National Research University, 34, Moskovskoye shosse, Samara, 443086, Russian Federation.

#### Введение

Современная научная мысль в экономической сфере постоянно требует мониторинга и анализа текущего состояния и развития показателей различных направлений, выявления их особенностей и тенденций [1–3]. Сложность анализа экономических показателей очевидна, так как экономическая сфера связана достаточно тесно с социальной сферой, они не могут быть стационарными априори. В научной литературе анализируются в большинстве случаев отдельные показатели, представляются их точечные и интервальные оценки для объектов различного уровня [4; 5]. Комбинированный подход при анализе социально-экономической сферы базируется на применении информационно-математических методов. Это позволяет расширить возможности монотетического подхода, например, при структурировании объектов исследования или показателей, описывающих изучаемые объекты [6; 7]. Сегодня такой подход является частью бизнес-аналитики. Настоящее исследование представляет собой изучение показателей непродовольственной сферы на примере субъектов Приволжского федерального округа [8; 9]. Основным инструментом анализа является многомерный подход. Методы многомерной статистики позволяют решать многопланово задачи анализа больших объемов данных [10; 11]. Одной из важнейших задач анализа развития субъектов является их структуризация на однородные кластеры. Методы кластерного анализа обширны и хорошо запрограммированы. В работе обращается внимание на возможность сочетания методов кластерного анализа с методами факторного анализа, что может быть выражено в совершенствовании экспертных оценок при расчете весовых коэффициентов.

#### Основная часть

Актуальность проведенного анализа производства и реализации непродовольственных товаров Приволжского федерального округа состоит в том, что стабильное развитие комплекса производства хозяйственных товаров способствует обеспечению потребностей населения в качественных и недорогих материалах и развитию бизнеса, повышению уровня импортозамещения в стране, что способствует росту национальной экономики России. Из этого следует, что данная тема требует большого внимания, ее актуально всесторонне изучать и исследовать.

Практическая значимость состоит в том, что результаты анализа показателей непродовольственной сферы с применением комбинированного подхода способствуют формированию рекомендательных решений по развитию показателей импортозамещения в Приволжском федеральном округе, а также разработке стратегий развития отдельных субъектов в условиях углубленного бизнес-аналитического подхода, усовершенствованию маркетинговой политики компаний на рынке непродовольственных товаров.

Товары непродовольственного рынка занимают одно из важнейших мест в экономике любой страны, потому что это – все те товары, которые используются каждым человеком в повседневной жизни помимо пищи. Таким образом, изучение показателей непродовольственного рынка Приволжского

федерального округа является важным, так как обрабатывающая промышленность и рынок региона оказывают сильное влияние на показатели Российской Федерации в целом. Изучение данных показателей и их взаимное влияние друг на друга возможны с использованием статистического инструментария.

В статье сфера товаров непродовольственного рынка изучается с использованием статистических показателей данной отрасли в Приволжском федеральном округе и Российской Федерации за 2016, 2017 и 2018 годы [12]. В качестве информационного инструментария используются различные программные среды, одной из таких программ является программа SPSS Statistics (Statistical Package for the Social Sciences). Источником статистических показателей служат официальные издания Федеральной службы государственной статистики (сокращенно – Росстат).

В работе исходные экономические показатели взяты из разделов «Строительство» и «Предприятия и организации», а также из разделов «Инвестиции», «Население», «Доходы и расходы населения» и «Торговля и услуги населению». Необходимые для проведения кластерного и факторного анализа показатели имеют следующие обозначения и единицы измерения:  $X_1$  – Численность населения (тыс. чел.);  $X_2$  – Реальные доходы населения (% к пред. году.);  $X_3$  – Потребительские расходы населения в среднем на душу населения (руб. в мес.);  $X_4$  – Индексы потребительских цен на непродовольственные товары (% к пред. году);  $X_5$  – Индексы цен производителей на строительную продукцию (% к пред. году);  $X_6$  – Число действующих строительных организаций;  $X_7$  – Оборачиваемость розничной торговли по торговым сетям, (%);  $X_8$  – Оборачиваемость оптовой торговли (млн руб.);  $X_9$  – Инвестиции в производство товаров непродовольственного назначения (млн руб.).

Классификация без обучения исходных показателей Приволжского федерального округа методом  $k$ -средних проводилась за указанные выше периоды. В качестве базового количества кластеров рассматривались разбиения на два, три и четыре кластера. В таблице 1 представлены субъекты ПФО и их принадлежность к кластерам при кластеризации на два кластера и при кластеризации на три кластера за изучаемый временной период.

**Таблица 1 – Результаты кластеризации субъектов ПФО методом  $k$ -средних**

**Table 1 – Results of clustering of constituent entities of the Volga Federal District using the  $k$ -means method**

№	Название субъекта ПФО	2016	2017	2018	2016	2017	2018
1	Республика Башкортостан	1	1	1	1	1	1
2	Республика Марий Эл	2	2	2	2	3	2
3	Республика Мордовия	2	2	2	2	3	2
4	Республика Татарстан	1	1	1	1	2	1
5	Удмуртская Республика	2	2	2	2	3	2
6	Чувашская Республика	2	2	2	2	3	2
7	Пермский край	1	1	2	1	1	1
8	Кировская область	2	2	2	2	3	3
9	Нижегородская область	1	1	1	2	1	1
10	Оренбургская область	2	2	2	3	3	3
11	Пензенская область	2	2	2	3	3	2
12	Самарская область	1	1	2	3	2	2
13	Саратовская область	2	2	2	2	3	3
14	Ульяновская область	2	2	2	2	3	3

Как видно из таблицы 1, при кластеризации на два кластера сохраняется структура кластеров, т. е. субъекты ПФО практически не меняют свои показатели. При кластеризации на три кластера не наблюдается такой устойчивой тенденции сохранения однородной структуры. В работе проведена проверка многомерных гипотез о равенстве векторов средних значений с использованием математиче-

ского инструмента проверки многомерных статистических гипотез. Для каждого временного периода рассчитывались средние значения по выбранным в исследовании показателям. Нулевая гипотеза, выдвинутая в исследовании, утверждает об отсутствии статистически значимых различий многомерных векторов средних значений. Наблюдаемые и критические значения статистики критерия рассчитывались с помощью соотношений:

$$T_{набл}^2 = \frac{n_1 n_2}{n_1 + n_2} (\bar{X}_1 - \bar{X}_2)^T \hat{S}^{-1} (\bar{X}_1 - \bar{X}_2),$$
$$T_{кр}^2(\alpha; k_1; k_2) = \frac{(n_1 + n_2 - 2)m}{(n_1 + n_2 - m - 1)} F(\alpha; k_1; k_2).$$

При классификации на два кластера было установлено отсутствие статистически значимых различий между средними показателями по каждому кластеру за 2016 и 2017 годы. На уровне значимости 0,11 установлено статистически значимое различие между средними показателями за 2016 и 2018 годы. Проверка частных критериев свидетельствует о статистически значимом различии между показателями «Потребительские расходы населения в среднем на душу населения» и «Оборот оптовой торговли» за 2016 и 2018 годы на уровне 0,07. Кластеризация на три кластера способствует более углубленной детализации при выявлении однородных субъектов в структуре ПФО. Кластер 1, включающий субъекты, имеющие высокие показатели, практически сохраняет свою структуру и при кластеризации на три группы. Кластер 2 и кластер 3 образуют самостоятельные региональные однородные структуры. Для них обнаруживается статистически значимое различие в средних показателях за 2016 и 2017, 2016 и 2018 годы. Проверка частных критериев показывает, что наблюдается статистически значимое различие только для показателей «Потребительские расходы населения в среднем на душу населения», «Оборот оптовой торговли», «Оборот розничной торговли по торговым сетям» и «Инвестиции в производство товаров непродовольственного назначения». Но для всех кластеров характерно отсутствие статистически значимых различий средних по кластеру показателей за 2017 и 2018 годы. Уровень значимости составляет 0,05.

В качестве особенности для субъектов первого кластера можно отметить высокий уровень развития компаний и предприятий оптовой и розничной торговли направления строительных товаров из группы непродовольственных товаров. Кластеризация на три кластера позволила уточнить уровень развития показателей непродовольственной сферы для отдельных групп регионов, имеющих свои показатели ниже, чем средние значения по ПФО в целом. Для этих регионов важным является развитие большинства изучаемых показателей. Таким образом, классификация методом  $k$ -средних позволила структурировать субъекты ПФО на однородные группы в условиях минимизации различий значений показателей внутри кластера, т. е. при минимизации внутриклассовой дисперсии. Далее предлагается методика кластеризации с переходом в пространство латентных факторов, а также кластеризация с учетом весовых коэффициентов, определяемых из интегрального показателя, исчисленного методом главных компонент и методом главных факторов.

Методом главных компонент выделены два главных фактора, и методом главных факторов алгоритмом Хоттелинга были выделены три главных фактора. Варимаксное вращение использовалось в методе главных компонент. В таблице 2 представлена матрица факторного отображения метода главных компонент после варимаксного вращения для изучаемого временного периода. Как видно, максимальная доля вариации приходится на первый латентный фактор. Коэффициенты корреляции между изучаемыми показателями и латентными факторами очень высокие, т. е. выделенные латентные факторы могут быть использованы в дальнейшем исследовании в качестве интегрального показателя, включающего в себя группу изучаемых признаков. Кроме того, из таблицы 2 видно, что для разных моментов времени наблюдается сохранение степени тесноты показателей с каждым латентным фактором, что позволяет также использовать латентные факторы интегрально, например для визуализации многомерных данных, а также для применения их при кластеризации на интегральных показателях без весовых коэффициентов.

**Таблица 2 – Матрица факторного отображения метода главных компонент после варимаксного вращения для 2016, 2017 и 2018 годов**

**Table 2 – Matrix of factor mapping of the principal component method after varimax rotation for 2016, 2017 and 2018 years**

Показатели	2016 год		2017 год		2018 год	
X <sub>1</sub>	0,978	0,033	0,978	0,166	0,974	0,008
X <sub>2</sub>	-0,383	-0,565	-0,211	0,552	-0,016	-0,599
X <sub>3</sub>	0,946	0,156	0,962	-0,061	0,961	0,109
X <sub>4</sub>	0,962	-0,032	0,961	0,241	0,961	-0,048
X <sub>5</sub>	-0,0474	0,488	0,088	0,231	-0,117	0,465
X <sub>6</sub>	0,093	0,607	0,094	-0,335	-0,009	0,654
X <sub>7</sub>	0,921	-0,112	0,987	-0,116	0,897	0,033
X <sub>8</sub>	-0,00033	-0,694	-0,429	0,281	-0,265	-0,035
X <sub>9</sub>	0,947	-0,027	0,948	0,0452	0,928	-0,157

Коэффициент информативности в исследовании составил 0,79–0,84. Проверка гипотезы о достаточности выделенных латентных факторов проводилась расчетом критерия Бартлетта. На уровне значимости 0,05 установлено, что количество выделенных латентных факторов достаточно. Весовые коэффициенты матрицы факторного отображения являются коэффициентами корреляции с соответствующими им изучаемыми показателями. Следовательно, первый латентный фактор включает в себя такие показатели, как X<sub>1</sub>, X<sub>3</sub>, X<sub>4</sub>, X<sub>7</sub> и X<sub>9</sub>. ЛАТФ<sub>1</sub> = {Численность населения, Потребительские расходы населения в среднем на душу населения, Индексы потребительских цен на непродовольственные товары, Оборот розничной торговли по торговым сетям, Инвестиции в производство товаров непродовольственного назначения}. Этот фактор интерпретируется как показатель развития потребительских отношений непродовольственной сферы. Второй латентный фактор включает в себя такие показатели, как X<sub>2</sub>, X<sub>5</sub>, X<sub>6</sub>, X<sub>8</sub>. ЛАТФ<sub>2</sub> = {Реальные доходы населения, Индексы цен производителей на строительную продукцию, Число действующих строительных организаций; Оборот оптовой торговли}. Этот фактор интерпретируется как показатель развития производственных отношений непродовольственной сферы. Используя формулы перехода из одного пространства размерности 14x9 в пространство размерности 14x2, были рассчитаны координаты субъектов ПФО в пространстве выделенных латентных факторов. На рис. 1 представлены изучаемые субъекты ПФО в теоретическом пространстве латентных факторов с учетом варимаксного вращения.

Как видно из рис. 1, Республика Татарстан и Республика Башкортостан имеют положительные значения выделенных латентных факторов, что интегрально может свидетельствовать о тенденции роста как потребительских отношений, так и производственных отношений непродовольственной сферы. Также положительная тенденция роста уровня развития потребительских отношений непродовольственной сферы наблюдается у таких субъектов ПФО, как Пермский край, Самарская область, Саратовская область и Нижегородская область. Однако по фактору развития производственных отношений непродовольственной сферы наблюдается отрицательная тенденция, т. е. снижение уровня развития производственных отношений непродовольственной сферы. Оно достаточно сильно в Саратовской области и Нижегородской области, средняя степень снижения присуща Пермскому краю, и незначительная тенденция – в Самарской области. Чувашская Республика, Пензенская и Оренбургская области сохраняют положительную тенденцию развития производственных отношений непродовольственной сферы, но имеют устойчивую тенденцию к снижению по первому латентному фактору.



Рис. 1 – Субъекты ПФО в пространстве латентных факторов (2018 год)

Fig. 1 – Constituent entities of the Volga Federal District in the space of latent factors (2018)

Следующая группа субъектов ПФО имеет отрицательные тенденции, т. е. снижение уровня развития как потребительских отношений, так и производственных по показателям непродовольственной сферы. К ним относятся Удмуртская Республика, Республика Марий Эл и Республика Мордовия. Ульяновская и Кировская области показывают отсутствие как положительной, так и отрицательной тенденции по развитию второго латентного фактора, однако по показателям развития потребительских отношений тенденция к снижению наблюдается сильная в Ульяновской области. Таким образом, все субъекты Приволжского федерального округа могут быть классифицированы на четыре кластера. Кроме того, из рис. 1 ясно, что если в основе кластеризации используется первый латентный фактор, характеризующий уровень развития потребительских отношений, то выделяются два кластера. Аналогично, если в основе кластеризации используется второй латентный фактор, характеризующий развитие производственных отношений непродовольственной сферы, то также выделяются два или три кластера. В таблице 3 представлены сводные результаты кластеризации субъектов ПФО по показателям 2018 года на латентных факторах, выделенных методом главных компонент с использованием варимаксного вращения. В таблице 3 цифрами 1, 2 или 3 обозначается номер кластера. Таким образом, используя интегральный показатель, можно осуществить монотетический подход классификации без обучения. Причем для характеристики кластеров может использоваться конкретный интегральный показатель, его численное значение, а также и исходные показатели, выбранные в исследовании. Так, при кластеризации на два кластера субъекты первого кластера имеют отчетливую положительную тенденцию роста фактора развития потребительских отношений непродовольственной сферы, а субъекты второго класса описываются тенденцией снижения данного фактора. Причем такие показатели, как «Потребительские расходы населения в среднем на душу населения», «Индекс потребительских цен на непродовольственные товары», «Оборот розничной торговли по торговым сетям» и «Инвестиции в производство товаров непродовольственного назначения», дают существенный вклад в имеющуюся тенденцию. Отмечается также важность поддержки демографической политики в данных субъектах на высоком уровне. При кластеризации по второму латентному фактору, а именно – фактору развития производственных отношений непродовольственной сферы, тен-

денцию снижения определяют реальные доходы населения. При кластеризации на три кластера по второму латентному фактору такие показатели, как «Индексы цен производителей на строительную продукцию» и «Число действующих строительных организаций», позволяют удерживать регионы второго кластера на позиции, которая может быть охарактеризована как не имеющая определенной тенденции.

**Таблица 3 – Результаты кластеризации субъектов ПФО по показателям 2018 года на латентных факторах, выделенных методом главных компонент с использованием варимаксного вращения**

**Table 3 – Clusterization results of constituent entities of the Volga Federal District according to the indicators of 2018 on latent factors identified by the principal component method using varimax rotation**

№	Субъект ПФО (2018 год)	ЛФ <sub>1</sub>	ЛФ <sub>2</sub>	ЛФ <sub>2</sub>
1	Республика Башкортостан	1	1	1
2	Республика Марий Эл	2	2	3
3	Республика Мордовия	2	2	3
4	Республика Татарстан	1	1	1
5	Удмуртская Республика	2	2	3
6	Чувашская Республика	2	1	1
7	Пермский край	1	1	3
8	Кировская область	2	2	2
9	Нижегородская область	1	1	3
10	Оренбургская область	2	2	2
11	Пензенская область	2	2	2
12	Самарская область	1	1	2
13	Саратовская область	1	2	3
14	Ульяновская область	2	1	2

Изучая аналогичные показатели за 2016 и 2017 годы, можно утверждать, что большинство субъектов ПФО сохраняют свою позицию в кластерах, но есть субъекты, которые ухудшают или улучшают свои показатели с течением времени. Таким образом, проведенная кластеризация позволяет провести комбинированный и динамический анализ показателей, характеризующих кластеры, а также проанализировать их средние значения, медианные значения, квартильные и децильные значения в динамике. Кроме того, сравнение средних значений показателей отдельных субъектов ПФО со средними значениями показателей по Приволжскому федеральному округу в целом и по РФ позволяет проводить динамический анализ, т. е. динамически отслеживать уровень развития показателей непродовольственной сферы. Это способствует своевременному регулированию уровней развития показателей в их текущей тенденции.

Как известно, при кластеризации важным является выделение групп показателей, играющих первостепенную роль в развитии изучаемых объектов. Четкой методики расчета весовых коэффициентов в научной литературе нет в настоящее время. Приводятся методики экспертного оценивания, методики вероятностного подхода, учитывающие опыт прошлых периодов. Следует отметить, что серьезная трудность оценки весовых коэффициентов определяется субъективностью экспертов. Человеческий фактор, как правило, сильно влияет на конечные результаты. И кроме того, многообразие экспертных методов также не упрощает процедуру окончательного формирования весовых коэффициентов. Математический аппарат факторного анализа позволяет максимально исключить субъективное влияние человеческого фактора. В основе формирования компонент матрицы факторного отображения заложена мера вариации показателя. Учитывая свойства компонент матрицы факторного отображения,

можно провести оценку весовых коэффициентов по расчету доли показателей, максимально коррелирующих с выбранным латентным фактором. Так, в статье соответствующие весовые коэффициенты могут быть оценены как 0,56 и 0,44 для первого и второго латентных факторов для изучаемых моментов времени 2016, 2017 и 2018 годы. Кластеризация на латентных факторах с весовыми коэффициентами позволяет, кроме того, усилить влияние групп показателей, участвующих в кластеризации, и ослабить влияние другой группы показателей. Несложными математическими расчетами можно каждый весовой коэффициент перевести в весовые коэффициенты исходных показателей. Вес также определяется коэффициентом корреляции показателя по матрице факторного отображения. Таким образом, легко исключается субъективность человеческого фактора, и, кроме того, весовая кластеризация на компонентах матрицы факторного отображения дает возможность исследователю явно увидеть влияние каждого показателя в факторе.

Далее в работе рассматривается классификация без обучения показателей рынка непродовольственных товаров с весовыми коэффициентами на латентных факторах, выделенных алгоритмом Хотеллинга. Эта процедура имеет такое достоинство, как простота в использовании информационного обеспечения, и может широко применяться, так как MSExcel в настоящее время имеет широкое распространение. Исследователь также сам задает, какое количество латентных факторов важно использовать в исследовании, с помощью коэффициента информативности. Выбор способа формирования редуцированной матрицы корреляции также остается за исследователем. В настоящей статье используется метод максимальной корреляции при формировании редуцированной матрицы. В таблице 4 представлены компоненты выделенных главных факторов для показателей 2016 и 2018 годов. Автором были выполнены 3 итерационные процедуры по выделению каждого фактора отдельно. Согласно алгоритму Хотеллинга, редуцированная матрица корреляций возводится в степень достаточное число раз. Коэффициент информативности оставленных в исследовании латентных факторов составил 0,91 и 0,97 в 2016 и 2018 годах соответственно. В таблице 4 представлены компоненты матрицы факторного отображения, рассчитанные алгоритмом Хотеллинга.

**Таблица 4 – Матрица факторного отображения метода главных факторов алгоритм Хотеллинга (2016 и 2018 годы)**

**Table 4 – Matrix of factorial mapping of the method of principal factors Hotelling algorithm (2016 and 2018 years)**

Показатели	2016 год			2018 год		
	X <sub>1</sub>	0,945	0,026	0,006	0,978	0,033
X <sub>2</sub>	-0,384	-0,542	0,043	-0,383	-0,564	0,122
X <sub>3</sub>	0,924	0,139	-0,006	0,946	0,156	0,125
X <sub>4</sub>	0,927	-0,030	-0,065	0,962	-0,032	-0,053
X <sub>5</sub>	-0,476	0,484	-0,143	-0,473	0,488	0,296
X <sub>6</sub>	0,091	0,340	0,509	0,093	0,207	-0,632
X <sub>7</sub>	0,237	0,618	0,320	0,237	0,635	0,361
X <sub>8</sub>	0,051	-0,242	-0,445	0,175	-0,275	0,428
X <sub>9</sub>	0,897	-0,102	0,214	0,921	-0,111	-0,088

Как видно из таблицы 4, количество максимально коррелирующих с первым латентным фактором показателей равно четырем, второй латентный фактор имеет два три показателя, а третий фактор, соответственно, два показателя. По таблице 4 были получены значения весовых коэффициентов, которые составили 0,44, 0,33 и 0,23 соответственно как в 2016, так и в 2018 году. Матрица расстояний рассчитывалась в работе взвешенной Евклидовой метрикой. Алгоритм кластеризации базируется на методе «Средней связи». В результате применения весовой кластеризации с весовыми коэффициентами из матрицы факторного отображения субъекты ПФО были разбиты на кластеры. В основе кластеризации использовалась матрица расстояний. На каждом шаге по минимальному расстоянию между объектами кластеризации формировались центры кластеров, а присоединение объектов проводилось методом «Средней связи». Процедура агломеративной кластеризации визуализирована дендрограммой. В качестве примера на рис. 2 представлена дендрограмма по показателям 2018 года.

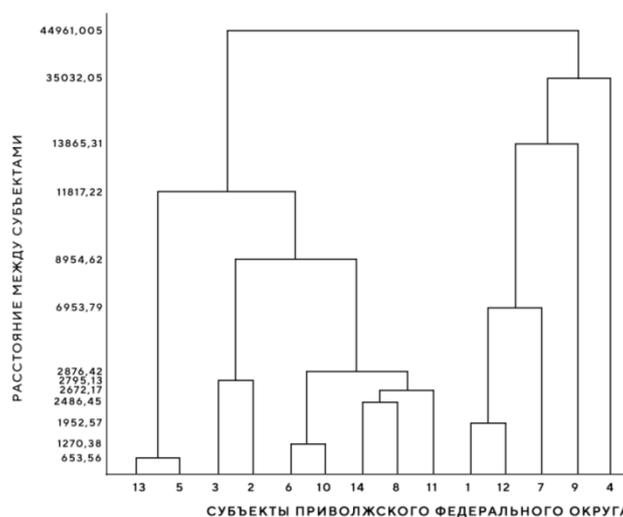


Рис. 2 – Дендрограмма субъектов ПФО (2018 год)  
 Fig. 2 – Dendrogram of the constituent entities of the Volga Federal District (2018 year)

Как видно из дендрограммы, оптимальное число кластеров может быть два или три. Критерии качества кластеризации показывают, что оптимальное число кластеров равно трем. В качестве критериев качества кластеризации использовались два: базирующиеся на сумме квадратов отклонений от центра кластера, а также на суммарной внутриклассовой вариации.

Также изучался дивизимный подход при кластеризации. Классификация проводилась с учетом весовых коэффициентов и без учета весовых коэффициентов.

Данные результаты сравнивались также со структуризацией по методике неметрического многомерного шкалирования. Как известно, неметрическое многомерное шкалирование базируется на ранговых данных. Субъекты ПФО были ранжированы по убыванию. В качестве определяющего критерия, обеспечивающего адекватность аналитических выводов, берется сохранение монотонных связей эмпирических и теоретических данных. Вид линейной монотонности используется в исследовании. В качестве меры расстояния также использовалась евклидова метрика, в том числе и взвешенная. Пошаговый поиск двумерного шкального пространства базировался на соотношениях:

$$\delta_{ij} = f(d_{ij}) = f\left(\sum_k Wk(x_{ik} - x_{jk})^2\right)^{1/2}$$

$$\delta_{ij} = f(d_{ij}) = f\left(\sum_k (x_{ik} - x_{jk})^2\right)^{1/2}$$

где  $f$  – линейная функция.

В таблице 5 представлены собственные значения выделенных компонент без учета веса показателя для значений 2018 года.

Таблица 5 – Собственные значения компонент теоретического пространства по показателям 2018 года

Table 5 – Eigenvalues of the components of the theoretical space by indicators of 2018 year

Компонента	Собственные значения		
	Итого	% дисперсии	Кумулятивный %
1	5,321	59,1	59,1
2	1,764	19,6	78,7

Как видно из таблицы 5, коэффициент информативности по первому фактору составил 0,591, по второму фактору – 0,196. Суммарное значение по двум выделенным факторам составляет 0,787. В таблице 6 представлена матрица компонент факторов в теоретическом пространстве, сформированном по мере различия между изучаемыми субъектами Приволжского федерального округа.

**Таблица 6 – Матрица компонент факторов субъектов ПФО в теоретическом пространстве в метрике различия (2018 год)**

**Table 6 – Matrix of the components of the factors of the constituent entities of the Volga Federal District in the theoretical space in the difference metric (2018 year)**

Изучаемые признаки	Теоретическая компонента	
	F <sub>1</sub>	F <sub>2</sub>
Численность населения (тыс. чел.)	-0,956	-0,037
Реальные доходы населения (% к пред. году.)	-0,439	0,874
Потребительские расходы населения в среднем на душу населения (руб. в мес.)	0,867	0,493
Индексы потребительских цен на непродовольственные товары (% к пред. году)	-0,657	0,406
Индексы цен производителей на строительную продукцию (% к пред. году)	0,911	0,132
Число действующих строительных организаций (шт.)	0,976	0,174
Оборот розничной торговли по торговым сетям (%)	0,651	0,756
Оборот оптовой торговли (млн руб.)	0,985	-0,036
Инвестиции в производство товаров непродовольственного назначения (млн руб.)	0,982	-0,048

Таблица 6 позволяет оценить долю показателей, имеющих максимальное по абсолютной величине значение в каждом факторе. Первый фактор имеет 0,78, второй фактор – 0,22. На рис. 3 субъекты ПФО представлены в теоретическом двумерном шкальном пространстве с учетом веса показателей, формирующих теоретические шкалы.

Полученные координаты субъектов в теоретическом пространстве по методике многомерного шкалирования с учетом веса позволяют также графически изобразить субъекты Приволжского федерального округа. Это способствует углубленному анализу, в том числе и при формировании групп однородных показателей. В основе кластеризации используются фактор торгового оборота товаров непродовольственной сферы и фактор уровня доходов населения. При этом учитывается максимальная степень различия между изучаемыми объектами по метрике Евклида с учетом веса. В результате в статье представляется следующая кластеризация:

- первый кластер формируют такие субъекты, как Республика Татарстан, Республика Мордовия, Нижегородская область и Самарская область;
- второй кластер образуют Республика Марий Эл, Удмуртская Республика, Чувашская Республика, Пермский край, Кировская область и Саратовская область;
- третий кластер формируют Республика Башкортостан, Оренбургская область, Пензенская область и Ульяновская область.

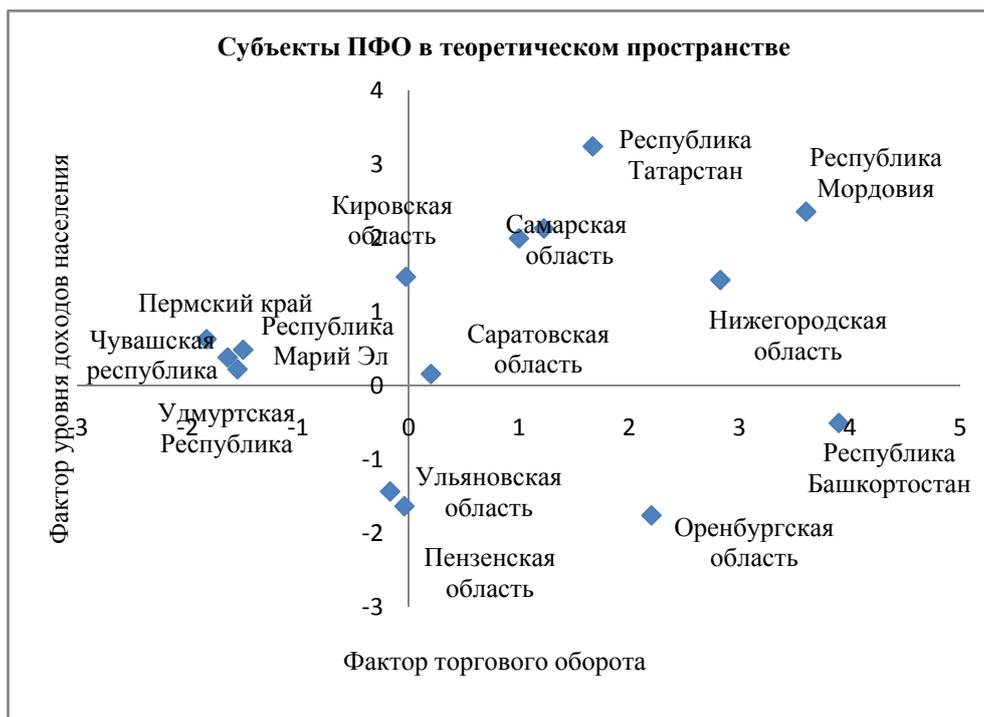


Рис. 3 – Субъекты ПФО в теоретическом двумерном шкальном пространстве

Fig. 3 – Constituent entities of the Volga Federal District in the theoretical two-dimensional scale space

Если в основу кластеризации поставить один фактор, например торговый оборот товаров непродовольственной сферы, то легко видна структура на два кластера. В частности, субъекты, геометрически расположенные в первом и четвертом квадрантах, образуют первый кластер, остальные – второй кластер. Аналогично на два кластера можно разделить субъекты ПФО по фактору уровня доходов населения. Следовательно, используя инструмент неметрического многомерного шкалирования, несложно структурировать субъекты исследования с целью визуализации и кластеризации.

### Заключение

Таким образом, в работе применен многомерный подход при классификации многомерных данных. Кроме известных методов кластеризации предложена классификация на латентных факторах без учета весовых коэффициентов, а также рассматривается методология исключения субъективного фактора при формировании весовых коэффициентов. Также предлагается структуризация с применением методики неметрического многомерного шкалирования, в том числе с учетом весовых коэффициентов. Долевое соотношение показателей как в латентных факторах, так и в теоретическом шкальном пространстве позволяет исключить применение экспертных оценок и их субъективность. В статье также отмечается возможность совмещения и динамического анализа при исследовании многомерных массивов данных.

### Библиографический список

1. Одинцова, Е.С. Динамика развития регионов РФ: оценка тенденций конвергенции // Российское предпринимательство. 2008. Т. 9. № 5. С. 10–15. URL: <https://cyberleninka.ru/article/n/dinamika-razvitiya-regionov-rf-otsenka-tendentsiy-konvergentsii>.
2. Татаркин А.И. Региональная направленность экономической политики Российской Федерации как института пространственного обустройства территорий // Экономика региона. 2016. Т. 12. Вып. 1. С. 9–27. URL: <https://cyberleninka.ru/article/n/regionalnaya-napravlennost-ekonomicheskoy-politiki-rossiyskoy-federatsii-kak-instituta-prostranstvennogo-obustroystva-territoriy>.

3. Растворцева С.Н. Основные направления влияния внешнеэкономических факторов на развитие регионов // Российское предпринимательство. 2015. Т. 16. № 23. С. 4193–4198. DOI: <https://doi.org/10.18334/rp.16.23.2170>.
4. Макарова Е.С. Классификация показателей инновационного потенциала региона // Экономика и менеджмент инновационных технологий. 2014. № 2. С. 25–29.
5. Марковская Е.А. Эффективное управление экономикой – необходимое условие успешного социально-экономического развития региона // Российское предпринимательство. 2008. Т. 9. № 10. С. 63–68. URL: <https://creativeconomy.ru/lib/2828>.
6. Назаренко В.С. Анализ тенденций комплексного социально-экономического развития городов Липецкой области // Экономика, предпринимательство и право. 2018. Т. 8. № 1. С. 51–63. DOI: <https://doi.org/10.18334/epp.8.1.38729>.
7. Kireyeva A.A., Mussabalina D.S., Tolysbaev B.S. Assessment and Identification of the Possibility for Creating IT Clusters in Kazakhstan Regions // *Ekonomika regiona [Economy of Region]*. 2018. no. 14 (2). P. 463–473. DOI: <https://doi.org/10.17059/2018-2-10>.
8. Петина П.В. Оценка эффективности реализации стратегии социально-экономического развития Самарской области // *Промышленная политика: глобализация, инновации, устойчивость: сб. материалов Всерос. научно-практич. конф. / под общ. ред. Н.М. Тюкавкина. Самара: АНО «Издательство СНЦ», 2018. С. 300–303. URL: <http://repo.ssau.ru/bitstream/PROMYShLENNAYa-POLITIKA-GLOBALIZACIYa-INNOVACII-USTOICHIVOST/OCENKA-EFFEKTIVNOSTI-REALIZACII-STRATEGII-SOCIALNOEKONOMICHESKOGO-RAZVITIYa-SAMARSKOI-OBLASTI-72143/1>.*
9. Ильина А.И., Трусова А.Ю., Люкшин Д.Е. Приволжский федеральный округ и его социально-экономические показатели в фокусе многомерного шкалирования // *Промышленная политика: глобализация, инновации, устойчивость: сб. материалов Всерос. научно-практич. конф. / под общ. ред. Н.М. Тюкавкина. Самара: АНО «Издательство СНЦ», 2018. С. 68–73.*
10. Дубров А.М., Мхитарян В.С., Трошин Л.И. Многомерные статистические методы. Москва: Финансы и статистика, 2003. 122 с.
11. Айвазян С.А. Прикладная статистика. Основы моделирования и первичная обработка данных: справочное издание. Москва: Финансы и статистика, 2013. 472 с.
12. Федеральная служба государственной. URL: <http://www.gks.ru> (дата обращения: 20.01.2020).

## References

1. Odintsova E.S. Growth dynamics of the Russian Federation territories: assessment of convergence trends. *Rossiyskoe predprinimatelstvo = Russian Journal of Entrepreneurship*, 2008, vol. 9, no. 5, pp. 10–15. Available at: <https://cyberleninka.ru/article/n/dinamika-razvitiya-regionov-rf-otsenka-tendentsiy-konvergensii>. (In Russ.)
2. Tatarkin A.I. Regional targeting of the economic policy of the Russian Federation as an institution of regional spatial development. *Economy of the region*, 2016, vol. 12, issue 1, pp. 9–27. Available at: <https://cyberleninka.ru/article/n/regionalnaya-napravlennost-ekonomicheskoy-politiki-rossiyskoy-federatsii-kak-instituta-prostranstvennogo-obustroystva-territoriy>. (In Russ.)
3. Rastvortseva S.N. The main directions of the influence of external economic factors on the regions' development. *Rossiyskoe predprinimatelstvo = Russian Journal of Entrepreneurship*, 2015, vol. 16, no. 23, pp. 4193–4198. DOI: <https://doi.org/10.18334/rp.16.23.2170>. (In Russ.)
4. Makarova E.S. Classification of indicators of innovative potential of the region. *Economics and innovations management*, 2014, no. 2, pp. 25–29. (In Russ.)
5. Markovskaya E.A. Effective economic management – a necessary condition for successful socio-economic development of the region. *Rossiyskoe predprinimatelstvo = Russian Journal of Entrepreneurship*, 2008, vol. 9, no. 10, pp. 63–68. Available at: <https://creativeconomy.ru/lib/2828>. (In Russ.)
6. Nazarenko V.S. Analysis of trends in complex social and economical development of the cities of Lipetsk region. *Journal of Economics, Entrepreneurship and Law*, 2018, vol. 8, no. 1, pp. 51–63. DOI: <https://doi.org/10.18334/epp.8.1.38729>. (In Russ.)

7. Kireyeva A.A., Mussabalina D.S., Tolysbaev B.S. Assessment and Identification of the Possibility for Creating IT Clusters in Kazakhstan Regions. *Ekonomika regiona = Economy of Region*, 2018, no. 14 (2), pp. 463–473. DOI: <https://doi.org/10.17059/2018-2-10>.
8. Petina P.V. Evaluation of effectiveness of the implementation of the strategy of socio-economic development of the Samara Region. In: *Industrial policy: globalization, innovation, sustainability: collection of materials of the all-Russian research and practical conference: N.M. Tyukavkin (Ed.)*. Samara: ANO «Izdatel'stvo SNTs», 2018, pp. 300–303. Available at: <http://repo.ssau.ru/bitstream/PROMYShLENNAYa-POLITIKA-GLOBALIZACIYa-INNOVACII-USTOICHIVOST/OCENKA-EFFEKTIVNOSTI-REALIZACII-STRATEGII-SOCIALNOEKONOMICHESKOGO-RAZVITIYa-SAMARSKOI-OBLASTI-72143/1>. (In Russ.)
9. Ilyina A. I., Trusova A.Yu., Lyukshin D.E. Privolzhsky Federal District and its socio-economic indicators in the focus of multidimensional scaling. In: *Industrial policy: globalization, innovations, sustainability: proceedings of the all-Russian research and practical conference. N.M. Tyukavkin (Ed.)*. Samara: ANO «Izdatel'stvo SNTs», 2018, pp. 68–73. Available at: <http://repo.ssau.ru/bitstream/PROMYShLENNAYa-POLITIKA-GLOBALIZACIYa-INNOVACII-USTOICHIVOST/PRIVOLZhSKII-FEDERALNYI-OKRUG-I-EGO-SOCIALNO-EKONOMICHESKIE-POKAZATELI-V-FOKUSE-MNOGOMERNOGO-ShKALIROVA NIYa-72065/1>. (In Russ.)
10. Dubrov A.M., Mkhitaryan V.S., Troshin L.I. Multivariate statistical methods. Moscow: Finansy i statistika, 2003, 122 p. (In Russ.)
11. Aivazyan S.A., Yenukov I.S., Meshalkin L.D. Applied statistics. Bases of modeling and initial data processing. Reference edition. Moscow: Finance and statistics, 2013, 472 p. Available at: <http://bookre.org/reader?file=448628&pg=1>. (In Russ.)
12. Federal State Statistics Service. Available at: <http://www.gks.ru> (accessed 20.01.2020) (In Russ.)