

Ю.Е. ГАВРИЛЕНКО

## МЕТОДЫ УСТОЙЧИВОЙ КЛАСТЕРИЗАЦИИ РЕГИОНОВ РОССИИ ПО ЗАНЯТОСТИ НАСЕЛЕНИЯ

*Проблема дисбаланса на рынке труда Российской Федерации не может быть решена без выравнивания неоднородности ее регионов по социально-экономическим и демографическим характеристикам, поскольку рынок труда — это динамическая сложная система, на которую оказывают влияние множество разнообразных факторов, таких как экономическая, демографическая ситуация, качество образования, интересы участников рынка, технический прогресс и цифровизация, психологические аспекты и др. В статье рассматривается применение методов кластерного и дискриминантного анализа на социально-экономических данных, выделяются региональные особенности рынка труда в России. Кластерный анализ проведен традиционными иерархическими и итерационными методами: методом «ближайшего соседа», методом «дальнего соседа», методом «Уорда» и методом «k-средних», а также методом нечеткой кластеризации fuzzy. Полученные этими пятью методами результаты были оценены на согласованность. Проведенный дискриминантный анализ позволил получить устойчивую кластерную структуру по численности занятого населения по видам экономической деятельности, разделяющую регионы России на четыре основных группы, характеризующиеся позитивным, средним, нейтральным и негативным поведением. Благодаря построению профилей полученных кластеров были выделены слабоинформативные виды экономической деятельности, занятость в которых мало влияет на разделение регионов по группам. В статье оценены ошибки методов кластерного анализа по итоговой устойчивой кластеризации. Проанализированы регионы с высокими и низкими уровнями занятости населения, выявлены нетипичные субъекты Российской Федерации и рассмотрена их отраслевая специализация. Проведен сопоставительный анализ сформированных групп и нетипичных регионов, выделены регионы, которые могут быть условно отнесены к какому-либо кластеру. Итоговая типологизация регионов России по численности занятых по видам экономической деятельности разработана с учетом территориальных, социальных, отраслевых и климатических особенностей.*  
**Ключевые слова:** регион, кластерный анализ, дискриминантный анализ, нечеткая кластеризация, рынок труда, занятость по видам экономической деятельности, R.

JEL: C59, C69, J21

В настоящее время одной из важных проблем, осложняющих развитие нашей страны, является *дисбаланс на рынке труда*. Данная проблема не может быть решена без учета всего комплекса причин ее определяющих. Кроме того, ее решение невозможно без выравнивания существующей неоднородности *регионов* Российской Федерации по социально-экономическим и демографическим характеристикам. Задача анализа такого большого массива неоднородных данных может быть решена с помощью *кластеризации* объектной структуры.

### Методы кластеризации

Основной целью кластерного анализа является выделение в исходных многомерных данных однородных групп, таких, чтобы объекты внутри групп были бы близки по выбранной метрике в многомерном признаковом пространстве, а объекты разных групп были бы удалены друг от друга. Мерой близости объектов в признаковом пространстве является метрика [1]. Для проведения кластерного анализа была выбрана такая метрика, как квадрат евклидова расстояния (1), т.к. данная метрика позволяет лучше учесть наиболее отдаленные показатели:

$$d_{lk} = d_4(X_l, X_k) = \sum_{j=1}^m (x_{lj} - x_{kj})^2, \quad (1)$$

где  $X_l, X_k$  — координаты  $l$ -го и  $k$ -го объектов.

С точки зрения результатов *классификации* важным является не только выбор метрики, но и сама процедура кластерного анализа. Среди всей совокупности методов кластеризации наиболее распространены иерархический и итерационный кластерные анализы: метод «ближайшего соседа», метод «дальнего соседа», метод «Уорда» и метод  $k$ -средних.

Метод «ближайшего соседа», или метод *одиночной связи*, заключается в нахождении двух наиболее близких объектов с последующим присоединением новых объектов с максимальной мерой сходства к образовавшемуся кластеру. Расстояние между кластерами  $D$  (2) в таком случае оценивается как минимальное из дистанций между парами объектов, один из которых входит в первый кластер, а другой — во второй.

$$D(I, J) = \min\{d(X_i, Y_j)\}, \quad (2)$$

где  $X, Y$  — кластеры, состоящие из объектов  $\{x_1, \dots, x_k\}$  и  $\{y_1, \dots, y_l\}$ ;  
 $d(X_i, Y_j)$  — расстояние между объектами из разных кластеров.

Метод «ближайшего соседа» обладает цепочечным эффектом, когда независимо от формы кластера к нему присоединяются ближайшие к границе объекты. Поэтому данный метод хорош для классификации данных, но не для кластеризации из-за отсутствия группировки [1].

Метод «дальнего соседа», или метод *полной связи*, подразумевает определение максимально допустимого порогового значения диаметра

подмножества. Расстояние между кластерами  $D$  (3) определяется *наибольшим* расстоянием между двумя объектами в различных кластерах [2]:

$$D(I, J) = \max\{d(X_i, Y_j)\}, \quad (3)$$

где  $X, Y$  – кластеры, состоящие из объектов  $\{x_1, \dots, x_k\}$  и  $\{y_1, \dots, y_l\}$ ;  
 $d(X_i, Y_j)$  – расстояние между объектами из разных кластеров.

Метод «дальнего соседа» не обладает цепочечным эффектом. Однако на ранних этапах может объединять довольно несхожие группы.

Метод «Уорда» направлен на объединение близко расположенных объектов, т.к. на каждом шаге соединяются такие два объекта, которые приводят к минимальному увеличению суммы квадратов расстояний между объектом и средней по кластеру [3]. Метод «Уорда» чаще других методов восстанавливает интуитивно наилучшую кластеризацию. Расстояние между кластерами  $D$  (4) определяется следующим образом:

$$D(I, J) = \frac{n_1 \cdot n_2}{n_1 + n_2} (\bar{X} - \bar{Y})(\bar{X} - \bar{Y})^T, \quad (4)$$

где  $X, Y$  – кластеры, состоящие из объектов  $\{x_1, \dots, x_{n_1}\}$  и  $\{y_1, \dots, y_{n_2}\}$ ;  
 $n_1, n_2$  – количество объектов в кластерах.

Метод « $k$ -средних» вычисляет расстояние между объектом и центрами кластеров и присоединяет рассматриваемый объект к наиболее близкому кластеру. Данный алгоритм отличается от предыдущих тем, что является итерационным, и для его выполнения необходимо заранее знать количество кластеров, на которые будет разбито объектное пространство. Результаты метода  $k$ -средних сильно зависят от начального выбора центров кластеров. Оптимальное количество кластеров можно определить с помощью *силуэтного анализа* – метода, использующего графические инструменты для измерения меры связанности кластеризации, проверяющего степень согласованности данных внутри одного кластера [4].

### **Методы уточнения кластеризации**

Для формирования *устойчивой кластерной структуры* уровня занятости по видам экономической деятельности в субъектах Российской Федерации рассмотрим следующие методы уточнения разделения кластеров: метод *fanny* и *дискриминантный анализ*.

Метод *fanny* относится к *fuzzy-clustering* методам. *Fuzzy clustering* – нечеткая кластеризация, в которой каждый элемент имеет вероятность принадлежности к каждому кластеру. Другими словами, каждый элемент имеет набор коэффициентов принадлежности, соответствующих степени нахождения в конкретном кластере [5]. В итоге получаем таблицу вероятностей попадания каждого субъекта Российской Федерации в один из четырех кластеров. Включение региона в конкретный кластер происходит в случае, если вероятность попадания в него больше 75%.

Данный метод позволяет просто решить проблему объектов, расположенных на границе двух кластеров — им назначают степени принадлежности, равные 0,5, что, однако, не приводит к четкому распределению объектов по конкретным кластерам. Для решения данной проблемы необходимо использование других *уточняющих методов*. Вторым недостатком нечеткого разбиения проявляется при работе с объектами, *удаленными от центров* всех кластеров. Удаленные объекты имеют мало общего с любым из кластеров, поэтому интуитивно хочется назначить для них малые степени принадлежности. Однако сумма их степеней принадлежности такая же, как и для объектов, близких к центрам кластеров, т.е. равна единице.

*Дискриминантный анализ* включает в себя методы классификации многомерных наблюдений по принципу максимального сходства при наличии *обучающих признаков*. В отличие от кластерного анализа новые кластеры не образуются, а являются правилом, по которому объекты относятся к определенной группе. Дискриминантный анализ позволяет решить следующие задачи: определение классифицирующего правила для изучения структуры в исследуемой совокупности и распределение объектов по существующим классам. Таким образом, с помощью данного метода можно определить значимые признаки и степень их влияния на классификацию объектов [4].

### **Результаты исследования и их обсуждение**

Для выявления ситуации на региональных рынках труда проведем кластерный анализ на данных об *уровне занятости в субъектах Российской Федерации по видам экономической деятельности*, собранных за период с 2000 по 2021 г.<sup>12</sup> Основным преимуществом использования данного метода для анализа ситуации на региональных рынках труда является территориальная и отраслевая дифференциация занятости населения [6; 7].

На первом этапе проведем анализ многомерных данных на наличие *нетипичных регионов* — регионов со значениями численности занятого населения по видам экономической деятельности, существенно отличающимся от общероссийских. Для этого на стандартизированных данных проведем кластерный анализ методом *«ближайшего соседа»*. Это позволит, во-первых, объединить по мере сходства (квадрату евклидова расстояния) наиболее близкие регионы. Во-вторых, более выпукло отобразить субъекты Российской Федерации, у которых данная мера сходства наименьшая с общим кластером [8]. Исследование показало,

<sup>1</sup> Трудовые ресурсы, занятость, безработица // Федеральная служба государственной статистики. URL: [https://rosstat.gov.ru/labour\\_force](https://rosstat.gov.ru/labour_force) (дата обращения: 01.03.2022).

<sup>2</sup> Единое хранилище данных. URL: [https://ehd.moscow/index.php?DIM\\_3087\\_1=22,6,10,11,13,18&id\\_src=355&id\\_ind=345&id\\_tab=1&action=show\\_details\\_open&show=inds&show\\_full=1&exist=1&source=-9993&id\\_root\[0\]=355\\_355&id\\_root\[1\]=355\\_345](https://ehd.moscow/index.php?DIM_3087_1=22,6,10,11,13,18&id_src=355&id_ind=345&id_tab=1&action=show_details_open&show=inds&show_full=1&exist=1&source=-9993&id_root[0]=355_355&id_root[1]=355_345) (дата обращения: 01.03.2022).

что такими нетипичными регионами являются Московская область, г. Москва, г. Санкт-Петербург, Краснодарский край, Ростовская область, Республика Башкортостан, Республика Татарстан, Свердловская область.

Рассмотрим данные регионы подробнее и отметим их *отраслевую специализацию*. Наибольшую долю в ВВП России составляет торговля (15,9%). При этом Москва, Московская область и Санкт-Петербург характеризуются как центры оптовой и розничной торговли. В них также развиты такие отрасли, как медицинские услуги, образование, деловые и IT-услуги, транспорт и логистика, строительство, и менее развита добыча полезных ископаемых (в силу отсутствия крупных месторождений). По объему сельскохозяйственной продукции Республика Башкортостан занимает 3-е место среди регионов России. Данный субъект занимает 1-е место по переработке и производству нефтепродуктов и 9-е место по объему добычи нефти. В Республике Татарстан основную долю ВРП составляют добыча полезных ископаемых (21,5%) и обрабатывающие производства (18,5%). В свою очередь в общероссийском производстве доля Республики Татарстан по производству грузовых автомобилей составляет 30,5%, полиэтилена — 51,9%, синтетического каучука — 41,9%. Краснодарский край и Ростовская область также характеризуются высокой степенью развитости сельского хозяйства. Кроме того, в этих регионах велика численность занятых в отрасли туризма и производства продуктов питания. Ведущими отраслями промышленности Свердловской области являются черная и цветная металлургия, машиностроение [9]. Таким образом, отраслевая специализация отмеченных регионов показывает их лидирующие позиции по отдельным видам экономической деятельности по вкладу в ВВП страны и, следовательно, высокий уровень занятости по выделенным секторам экономики.

Исключим выявленные *нетипичные субъекты Российской Федерации* из общего объектного пространства и продолжим формирование устойчивой кластерной структуры. Проведем кластеризацию уровня занятости населения по видам экономической деятельности по регионам Российской Федерации традиционными иерархическими и итерационными методами: методом «ближайшего соседа», методом «дальнего соседа», методом «Уорда» и методом «*k*-средних», а также уточняющим методом *fanny* (см. рис. 1).

Результатом применения данных методов является объектная структура, разделенная на *четыре кластера*. Составим таблицу частот попадания региона в кластер по проведенным различным методам кластеризации. Фрагмент этой таблицы частот представлен в *таблице 1*. Заметим, что не все субъекты Российской Федерации можно однозначно отнести к одному из четырех кластеров, например, Костромскую, Орловскую области, Республику Карелию и др. (всего 24 региона).

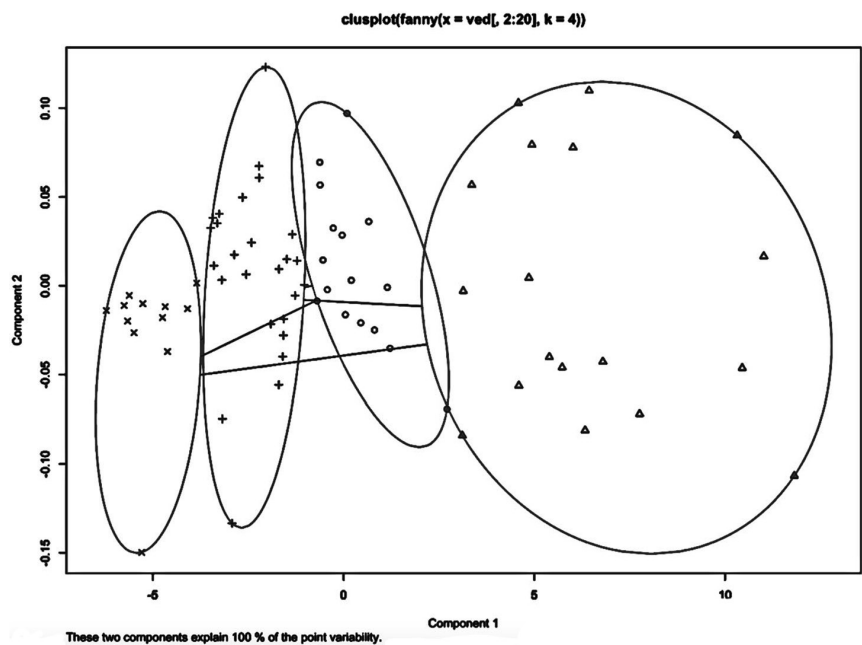


Рис. 1. Кластеризация регионов Российской Федерации, построенная по методу «fanny»

Источник: составлено автором.

Т а б л и ц а 1

Фрагмент таблицы частот попадания в кластеры

| Субъект РФ           | Ward | NN | FN | k-means | КА | КА с вер-тью 0,75 |
|----------------------|------|----|----|---------|----|-------------------|
| 1                    | 2    | 3  | 4  | 5       | 6  | 7                 |
| Белгородская область | 1    | 1  | 1  | 3       |    | 1                 |
| Брянская область     | 1    | 1  | 1  | 3       |    | 1                 |
| Владимирская область | 1    | 1  | 1  | 3       |    | 1                 |
| Воронежская область  | 2    | 2  | 2  | 2       | 2  |                   |
| Ивановская область   | 1    | 1  | 1  | 3       |    | 1                 |
| Калужская область    | 1    | 1  | 1  | 3       |    | 1                 |
| Костромская область  | 1    | 1  | 3  | 4       |    |                   |
| Курская область      | 1    | 1  | 1  | 3       |    | 1                 |
| Липецкая область     | 1    | 1  | 1  | 3       |    | 1                 |
| Орловская область    | 1    | 1  | 3  | 4       |    |                   |
| Рязанская область    | 1    | 1  | 1  | 3       |    | 1                 |
| Смоленская область   | 1    | 1  | 1  | 3       |    | 1                 |

О к о н ч а н и е   т а б л .   1

| 1                   | 2 | 3 | 4 | 5 | 6 | 7 |
|---------------------|---|---|---|---|---|---|
| Тамбовская область  | 1 | 1 | 1 | 3 |   | 1 |
| Тверская область    | 1 | 1 | 1 | 3 |   | 1 |
| Тульская область    | 1 | 1 | 1 | 3 |   | 1 |
| Ярославская область | 1 | 1 | 1 | 3 |   | 1 |
| Республика Карелия  | 1 | 1 | 3 | 4 |   |   |
| Республика Коми     | 1 | 1 | 1 | 3 |   | 1 |

Источник: составлено автором.

Сформируем обучающее правило и проведем дискриминантный анализ с помощью языка программирования R [10]. Вероятности попадания неопределенного региона Российской Федерации в конкретный кластер равны: 48,3% в первый кластер, 25% – во второй, 20% – в третий и 6,7% – в четвертый. Поскольку первый кластер по всем методам выявлен как самый многочисленный, вероятность попадания неопределенных регионов в него выше всего.

Коэффициенты полученной линейной дискриминантной функции представлены в *таблице 2*.

Т а б л и ц а   2

**Коэффициенты линейной дискриминантной функции**

| Виды экономической деятельности                                                                              | Коэффициент |
|--------------------------------------------------------------------------------------------------------------|-------------|
| 1                                                                                                            | 2           |
| Сельское, лесное хозяйство, охота, рыболовство и рыбоводство                                                 | -6,0873     |
| Добыча полезных ископаемых                                                                                   | -0,6275     |
| Обрабатывающие производства                                                                                  | 10,9385     |
| Обеспечение электроэнергией, газом и паром, кондиционирование воздуха                                        | -100,1325   |
| Водоснабжение, водоотведение, организация сбора и утилизация отходов, деятельность по ликвидации загрязнений | -45,0140    |
| Строительство                                                                                                | 26,8198     |
| Торговля оптовая и розничная, ремонт автотранспортных средств и мотоциклов                                   | 13,1141     |
| Транспортировка и хранение                                                                                   | -4,4471     |
| Деятельность гостиниц и предприятий общественного питания                                                    | -92,6955    |
| Деятельность в области информации и связи                                                                    | 73,6848     |

О к о н ч а н и е   т а б л .   2

| 1                                                                                                                         | 2         |
|---------------------------------------------------------------------------------------------------------------------------|-----------|
| Деятельность финансовая и страховая                                                                                       | -41,5876  |
| Деятельность по операциям с недвижимым имуществом                                                                         | -92,1192  |
| Деятельность профессиональная, научная и техническая; деятельность административная и сопутствующие дополнительные услуги | -7,4873   |
| Государственное управление и обеспечение военной безопасности, социальное обеспечение                                     | -7,8180   |
| Образование                                                                                                               | 10,4554   |
| Деятельность в области здравоохранения и социальных услуг                                                                 | 10,5240   |
| Деятельность в области культуры, спорта, организации досуга и развлечений                                                 | -119,8138 |
| Предоставление прочих видов услуг                                                                                         | 133,2188  |
| Другие виды экономической деятельности                                                                                    | 10,9671   |

Источник: составлено автором по результатам программы на языке R.

В результате дискриминантного анализа получена устойчивая кластеризация регионов по уровню занятости населения и по видам экономической деятельности по регионам Российской Федерации (диаграмма отклонений центроидов кластеров от общероссийского среднего значения, см. рис. 2).

Первый кластер итоговой классификации, в состав которого вошли Белгородская, Брянская, Владимирская, Ивановская, Рязанская, Тульская, Тверская области и некоторые другие (всего 29 регионов), характеризуется нейтральным поведением субъектов Российской Федерации по выбранным видам экономической деятельности, т.к. разница между центроидами групп и средними значениями показателей по всей России близка к нулю.

Во второй кластер вошли регионы, в которых уровни рассматриваемых показателей примерно на 114–117% выше общероссийских, здесь преобладают такие виды экономической деятельности, как торговля розничная и оптовая, обрабатывающие производства, образование, транспортировка и хранение. В его состав вошли Воронежская, Ленинградская, Волгоградская, Оренбургская, Саратовская, Новосибирская, Омская области и некоторые другие (всего 14 регионов).

Третий кластер характеризуется негативным поведением регионов по выбранным видам экономической деятельности: добыча полезных ископаемых, деятельность профессиональная, научная и техническая, деятельность административная и сопутствующие дополнительные услуги, деятельность в области культуры и спорта показывают самый высокий отрицательный эффект (уровни рассматриваемых показателей примерно на 45–46% ниже общероссийских). В составе этого

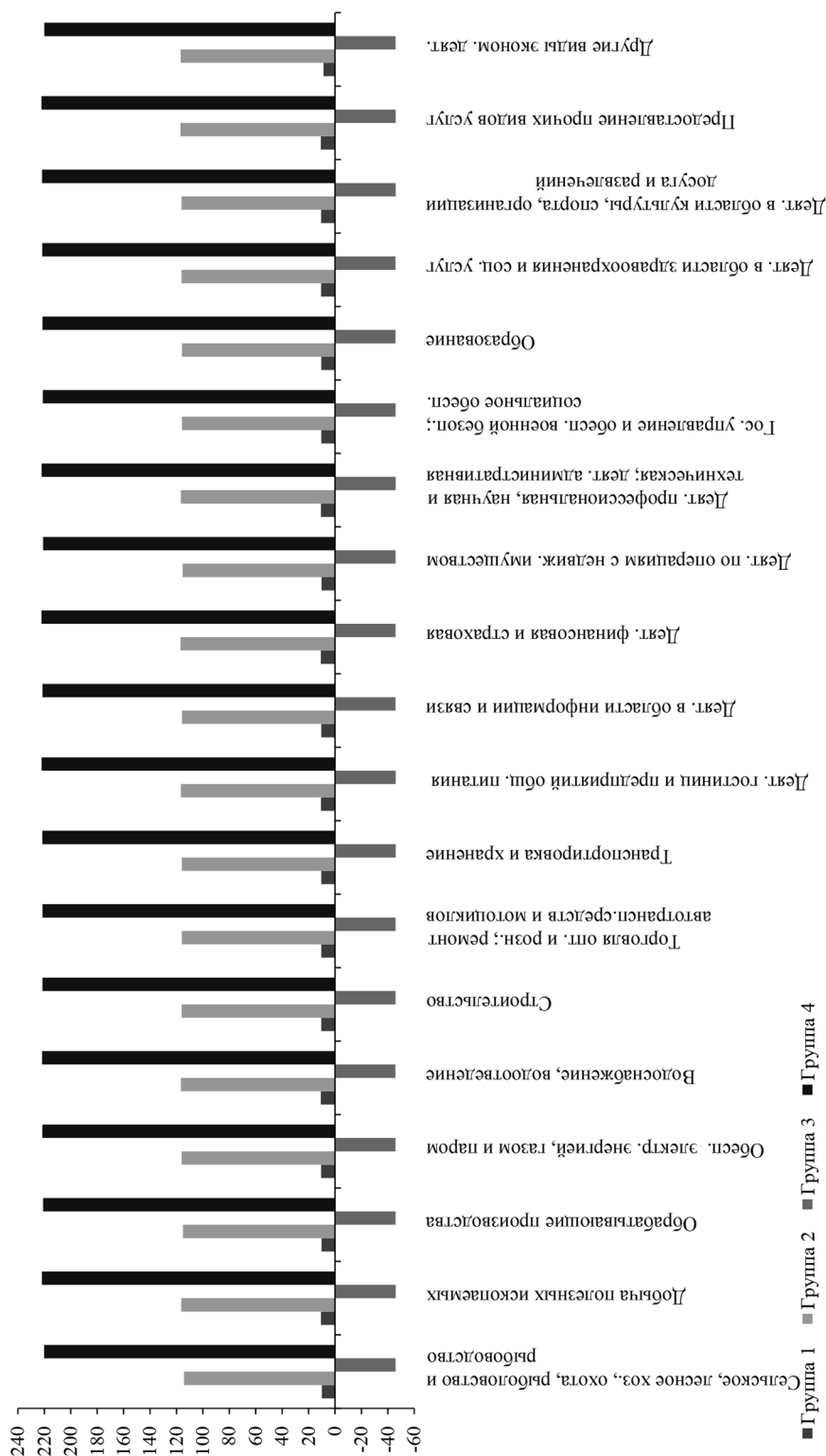


Рис. 2. Отклонения центров кластеров от медианного уровня в Российской Федерации

Источник: составлено автором.

кластера выделим такие регионы, как Республики Адыгея, Калмыкия, Ингушетия, Камчатский край, Амурская, Магаданская, Сахалинская области и некоторые другие (всего 24 региона).

Уровни рассматриваемых показателей в регионах четвертого кластера показывают позитивную тенденцию: значения по всем видам экономической деятельности в два раза превышают средний общероссийский уровень. В этот кластер вошли Нижегородская, Самарская, Тюменская, Челябинская области, Красноярский край.

Построим *итоговую кластеризацию* на карте Российской Федерации. На *рисунке 3* изображена карта России с выделенными *регионами*, принадлежащими к разным кластерам.



**Рис. 3. Карта регионов Российской Федерации в зависимости от их принадлежности к кластеру**

Источник: составлено автором.

Теперь выявим слабоинформативные виды экономической деятельности согласно профилям полученных кластеров (см. *рис. 4*). Таким образом, слабоинформативными видами экономической деятельности являются: сельское, лесное хозяйство, охота, рыболовство и рыбоводство; обрабатывающие производства; торговля оптовая и розничная; ремонт автотранспортных средств и мотоциклов; государственное управление и обеспечение военной безопасности; социальное обеспечение; образование; деятельность в области здравоохранения и социальных услуг; предоставление прочих видов услуг. Выделенные виды экономической деятельности слабо влияют на разделение субъектов Российской Федерации по кластерам, т.к. расстояние между вершинами, характеризующими данные виды экономической деятельности, является достаточно большим, в то время как остальные вершины тяготеют к приблизительно одинаковым значениям.

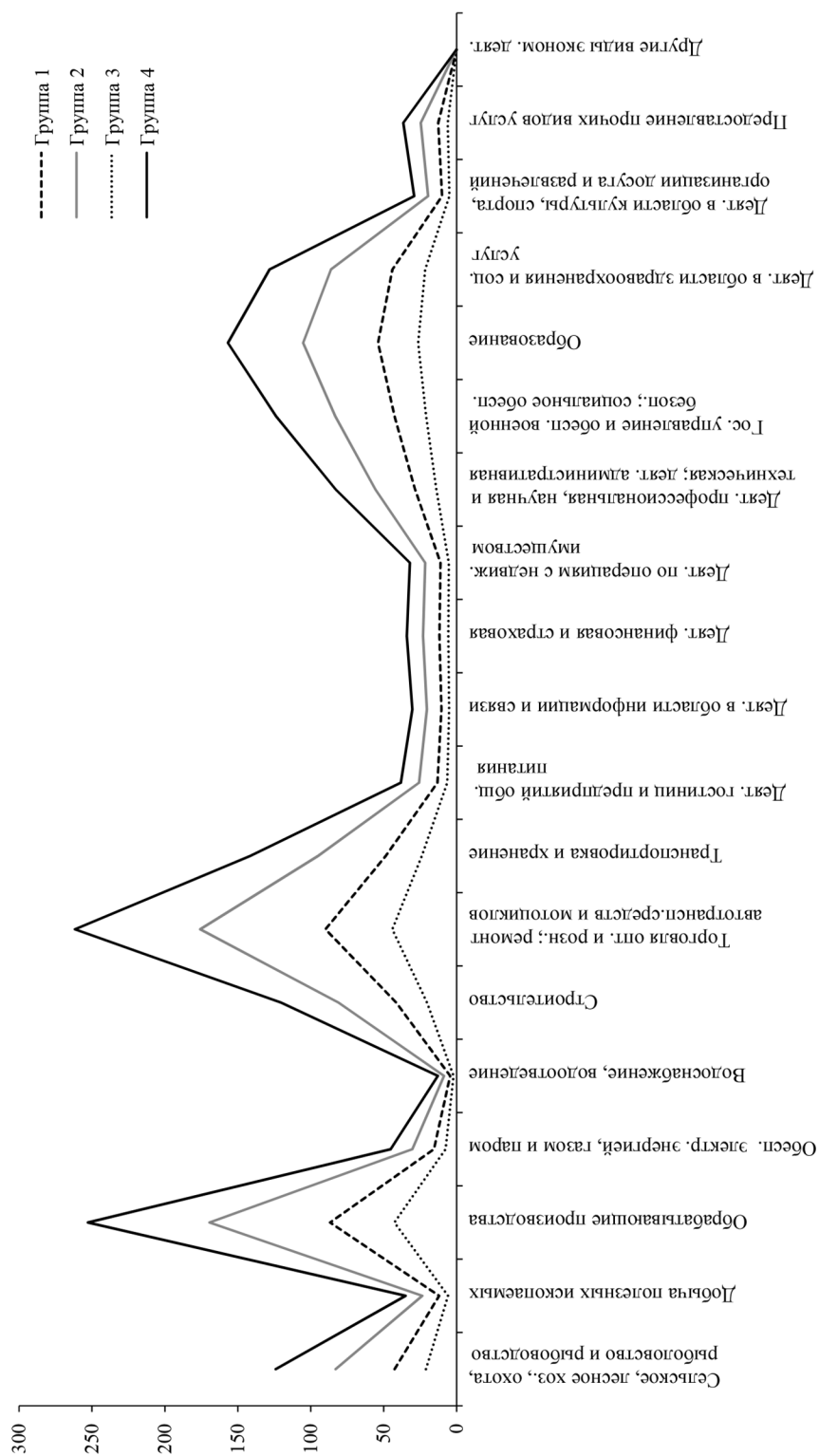
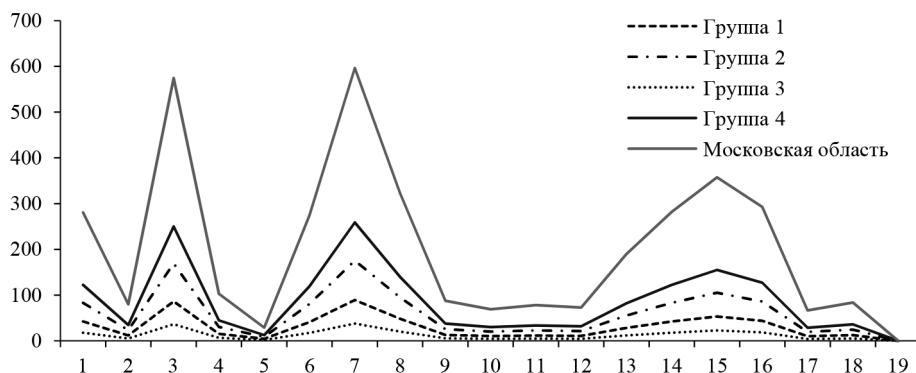


Рис. 4. Профили кластеров

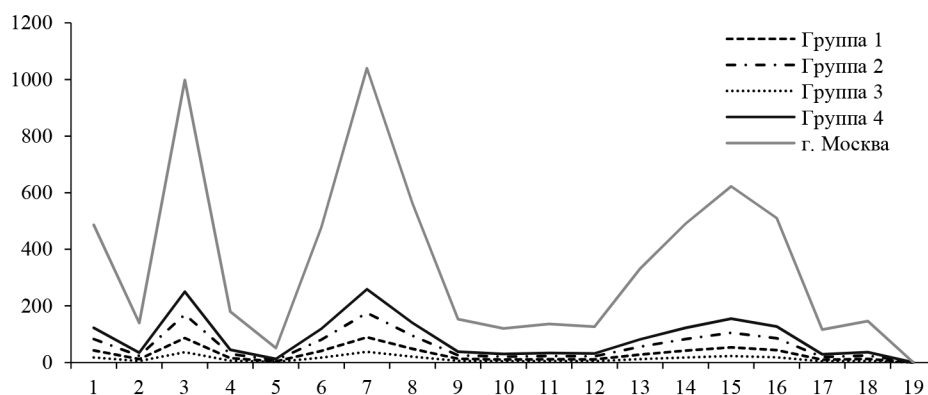
Источник: составлено автором.

Рассмотрим возможность условного отнесения нетипичного региона к определенному кластеру с помощью построения профилей кластеров совместно с профилями выбросов (см рис. 5–12).



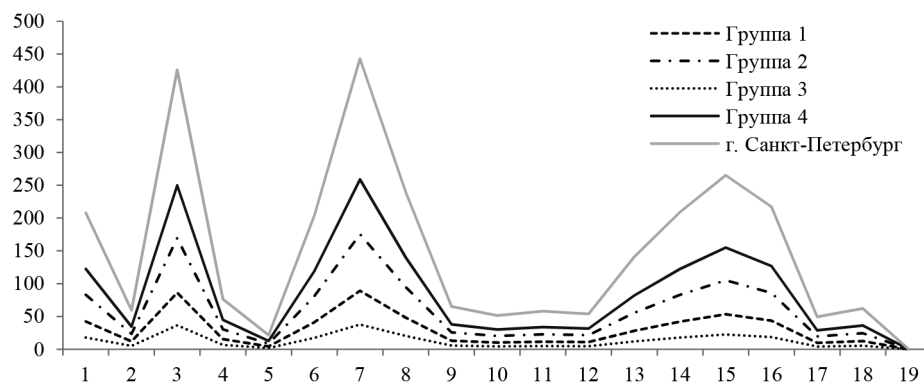
**Рис. 5. Профили кластеров с Московской областью**

Источник: составлено автором.



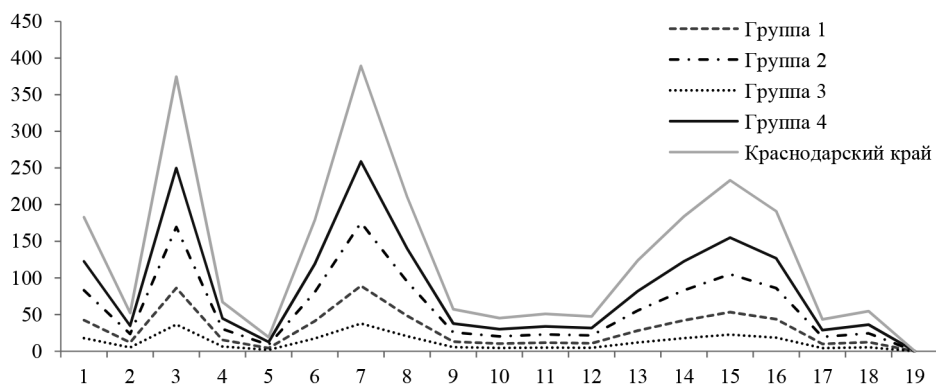
**Рис. 6. Профили кластеров с г. Москвой**

Источник: составлено автором.



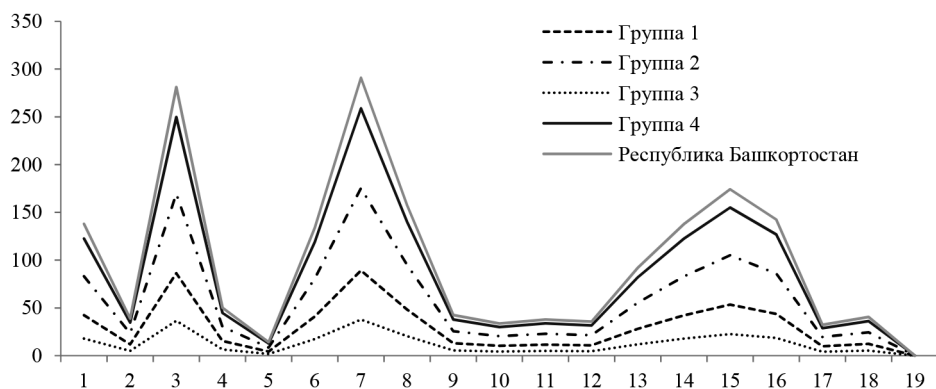
**Рис. 7. Профили кластеров с г. Санкт-Петербургом**

Источник: составлено автором.



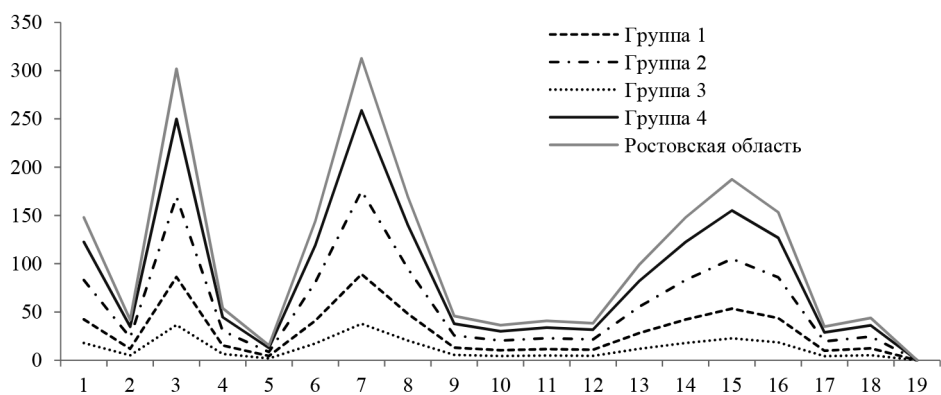
**Рис. 8. Профили кластеров с Краснодарским краем**

Источник: составлено автором.



**Рис. 9. Профили кластеров с Республикой Башкортостан**

Источник: составлено автором.



**Рис. 10. Профили кластеров с Ростовской областью**

Источник: составлено автором.

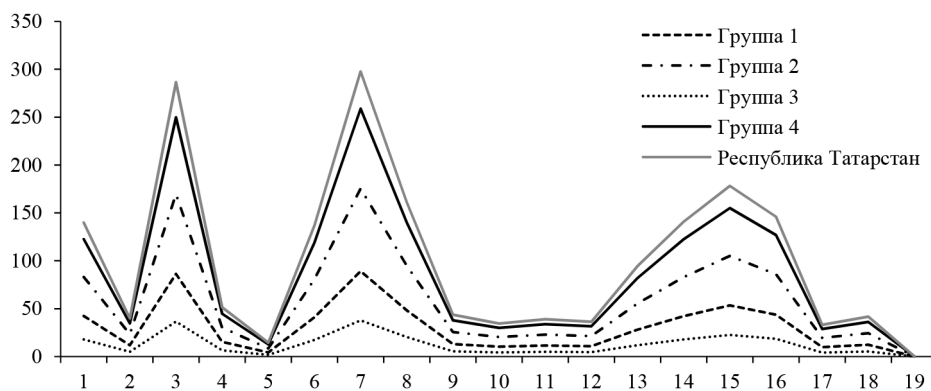


Рис. 11. Профили кластеров с Республикой Татарстан

Источник: составлено автором.

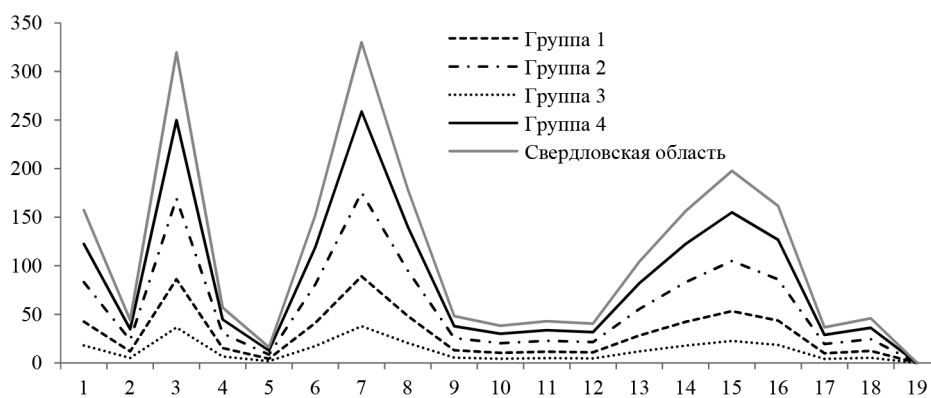


Рис. 12. Профили Кластеров со Свердловской областью

Источник: составлено автором.

Согласно полученным графикам, г. Санкт-Петербург, Краснодарский край, Республики Башкортостан и Татарстан, Ростовскую и Свердловскую области можно условно отнести к четвертому кластеру. Остальные нетипичные регионы — г. Москву и Московскую область — нельзя условно отнести ни к какому кластеру, т.к. их профили находятся слишком далеко от профилей выделенных кластеров.

Оценим согласованность результатов кластерного анализа разными методами с помощью программы на языке R. Индексы согласованности результатов следующие: методом «Уорда» и методом «ближайшего соседа» — 0,7903; методом «Уорда» и методом «дальнего соседа» — 0,8075; методом «Уорда» и методом «к-средних» — 0,8122; методом «ближайшего соседа» и методом «дальнего соседа» — 0,7105; методом «ближайшего соседа» и методом «к-средних» — 0,7089; методом «дальнего соседа» и методом «к-средних» — 0,9726. Как и было замечено при проведении кластерного анализа, самые схожие результаты показали методы «даль-

него соседа» и «к-средних», метод «ближайшего соседа» демонстрирует наиболее отличный от остальных методов результат.

Далее оценим ошибки методов кластерного анализа по итоговой устойчивой кластеризации следующим способом: вычтем из единицы отношение количества не совпавших регионов данного метода с итоговой кластеризацией и количества не совпавших регионов по всем методам. Минимальную схожесть с итоговой кластерной структурой показывает метод «к-средних» — 56,82%; метод нечеткой кластеризации «fanny» схож с устойчивой кластеризацией на 75,76%; метод «ближайшего соседа» на 78,03% схож с полученной классификацией регионов; метод «Уорда» оказался близок к итоговой кластерной структуре на 90,15%; самый достоверный результат показал метод «дальнего соседа», описывающий итоговую устойчивую кластерную структуру на 99,24%.

Как мы полагаем, полученная кластеризация позволяет выявить лидирующие и отстающие регионы Российской Федерации по численности занятого населения по видам экономической деятельности. Показано, что наименьшие значения численности занятых наблюдаются в Чукотском автономном округе — 31,3 тыс. чел., Еврейской автономной области — 77,6 тыс. чел. и в Республике Алтай — 86,8 тыс. чел.; наибольшие — в г. Москве — 6 688,9 тыс. чел., Московской области — 3 840,3 тыс. чел. и в г. Санкт-Петербурге — 2 850 тыс. чел. Основные виды экономической деятельности, влияющие на территориальное разделение субъектов Российской Федерации в схожие группы: добыча полезных ископаемых; обеспечение электрической энергией, газом и паром; водоснабжение; деятельность гостиниц и предприятий общественного питания; деятельность в области информации и связи; деятельность финансовая и страховая; деятельность по операциям с недвижимым имуществом; деятельность в области культуры и спорта. Такая устойчивая кластерная структура в дальнейшем может быть применена для разработки мер регулирования дисбаланса на рынке труда.

### Список литературы

1. Тихомирова Т.М., Галочкина Ж.С. Методы устойчивой классификации регионов РФ с учетом динамики медико-демографической ситуации // Экономика природопользования. 2012. № 4. С. 132–142.
2. Лапа Е.А., Лапа Е.И. Типологизация региональных рынков труда на основе кластерного анализа // Теория и практика общественного развития. 2016. № 2. С. 68–71.
3. Портнова Л.В. Применение метода кластерного анализа в оценке и прогнозировании уровня безработицы в регионе // Вестник Оренбургского государственного университета. 2012. № 4. С. 158–163.
4. Тихомиров Н.П., Тихомирова Т.М., Ушаев О.С. Методы эконометрики и многомерного статистического анализа. М.: Экономика, 2011. 647 с.
5. Шитиков В.К., Мостицкий С.Э. Классификация, регрессия и другие алгоритмы Data mining с использованием R. URL: <https://github.com/ranalytics/data-mining> (дата обращения: 07.04.2022).

6. Орлов Д.А., Постников Е.А. Кластеризация региональных рынков труда с учетом оценок NAIRU // Вестник Южно-Уральского государственного университета. Серия «Экономика и менеджмент». 2021. Т. 15. № 3. С. 34–44.

7. Шубат О.М., Караева А.П. Кластерный анализ в исследовании социально-экономических процессов: опыт критического // Проблемы моделирования социальных процессов: Россия и страны АТР: материалы Второй всероссийской научно-практической конференции с международным участием. — Владивосток, 2016. С. 325–328.

8. Маратканова И.В. Применение метода кластерного анализа для оценки сберегательно-инвестиционного потенциала населения Сибирского Федерального округа // Вестник Югорского государственного университета. 2021. Вып. 1 (60). С. 48–61

9. Регионы России. Социально-экономические показатели: стат. сб. / Росстат. М., 2020. 1242 с.

10. Уикем Х., Гроулмунд Г. Язык R в задачах науки о данных. Импорт, подготовка, обработка, визуализация и моделирование данных. М.: Вильямс, 2017. 592 с.

## References

1. Tikhomirova T.M., Galochkina Zh.S. Metody ustoichivoi klassifikatsii regionov RF s uchetom dinamiki mediko-demograficheskoi situatsii [Methods of Stable Classification of Regions of the Russian Federation Taking Into Account the Dynamics of the Medical and Demographic Situation], *Ekonomika prirodopol'zovaniia* [The Economics of Environmental Management], 2012, No. 4, pp. 132–142. (In Russ.).

2. Lapa E.A., Lapa E.I. Tipologizatsiia regional'nykh rynkov truda na osnove klasterного анализа [Typologization of Regional Labor Markets Based on Cluster Analysis], *Teoriia i praktika obshchestvennogo razvitiia* [Theory and Practice of Social Development], 2016, No. 2, pp. 68–71. (In Russ.).

3. Portnova L.V. Primenenie metoda klasterного анализа v otsenke i prognozirovanii urovnia bezrabotitsy v regione [Application of the Cluster Analysis Method in Assessing and Forecasting the Unemployment Rate in the Region], *Vestnik Orenburgskogo gosudarstvennogo universiteta* [Bulletin of Orenburg State University], 2012, No. 4, pp. 158–163. (In Russ.).

4. Tikhomirov N.P., Tikhomirova T.M., Ushmaev O.S. Metody ekonometriki i mnogomernogo statisticheskogo analiza [Methods of Econometrics and Multidimensional Statistical Analysis]. Moscow, Ekonomika, 2011, 647 p. (In Russ.).

5. Shitikov V.K., Mastitskii S.E. Klassifikatsiia, regressiia i drugie algoritmy Data mining s ispol'zovaniem R [Classification, Regression and Other Data Mining Algorithms Using R]. (In Russ.). Available at: <https://github.com/ranalytics/data-mining> (accessed 07 April 2022).

6. Orlov D.A., Postnikov E.A. Klasterizatsiia regional'nykh rynkov truda s uchetom otsenok NAIRU [Clustering of Regional Labor Markets Taking Into Account NAIRU Estimates], *Vestnik Yuzhno-Ural'skogo gosudarstvennogo universiteta. Seriya "Ekonomika i menedzhment"* [Bulletin of South Ural State University. The Series "Economics and Management"], 2021, Vol. 15, No. 3, pp. 34–44. (In Russ.).

7. Shubat O.M., Karaeva A.P. Klasternyi analiz v issledovanii sotsial'no-ekonomicheskikh protsessov: opyt kriticheskogo [Cluster Analysis in the Study of Socio-Economic Processes: the Experience of Critical Analysis], *Problemy modelirovaniia sotsial'nykh protsessov: Rossiia i strany ATR: materialy Vtoroi vserossiiskoi nauchno-prakticheskoi konferentsii s mezhdunarodnym uchastiem* [Problems of Modeling Social

Processes: Russia and the Countries of the Asia-Pacific Region : Materials of the Second All-Russian Scientific and Practical Conference with International Participation], Vladivostok, 2016, pp. 325–328. (In Russ.).

8. Maratkanova I.V. Primenenie metoda klasternogo analiza dlia otsenki sberegatel'no-investitsionnogo potentsiala naseleniia Sibirskogo Federal'nogo okruga [Application of the Cluster Analysis Method to Assess the Savings and Investment Potential of the Population of the Siberian Federal District], *Vestnik Iugorskogo gosudarstvennogo universiteta* [Bulletin of the Ugra State University], 2021, Issue 1 (60), pp. 48–61. (In Russ.).

9. Regiony Rossii. Sotsial'no-ekonomicheskie pokazateli: stat. sb. [Regions of Russia. Socio-economic indicators, Statistical Collection], Rosstat. Moscow, 2020, 1242 p. (In Russ.).

10. Uikem Kh., Groulmund G. Iazyk R v zadachakh nauki o dannykh. Import, podgotovka, obrabotka, vizualizatsiia i modelirovanie dannykh [The R Language in Data Science Problems. Import, Preparation, Processing, Visualization and Modeling of Data], Moscow, Vil'iams, 2017, 592 p. (In Russ.).

## METHODS OF SUSTAINABLE CLUSTERING OF RUSSIAN REGIONS BY EMPLOYMENT

*The problem of the imbalance in the labor market of the Russian Federation cannot be solved without leveling the heterogeneity of its regions by socio-economic and demographic characteristics, since the labor market is a dynamic complex system that is influenced by a variety of factors, such as the economic, demographic situation, quality of education, interests of market participants, technological progress and digitalization, psychological aspects, etc. The article discusses the application of cluster and discriminant analysis methods on socio-economic data, highlights the regional features of the labor market in Russia. Cluster analysis was carried out using traditional hierarchical and iterative methods: the “Nearest Neighbor” method, the “Far Neighbor” method, the “Ward” method and the k-means method, as well as the fuzzy clustering method. The results obtained by these five methods were evaluated for consistency. The conducted discriminant analysis allowed us to obtain a stable cluster structure in terms of the number of employed people by type of economic activity, dividing the regions of Russia into four main groups characterized by positive, average, neutral and negative behavior. Thanks to the construction of profiles of the obtained clusters, poorly informative types of economic activity were identified, employment in which has little effect on the division of regions into groups. The article evaluates the errors of cluster analysis methods for the final stable clustering. The regions with high and low levels of employment are analyzed, atypical subjects of the Russian Federation are identified and their industry specialization is considered. A comparative analysis of the formed groups and atypical regions was carried out, regions that can be conditionally assigned to any cluster were identified. The final typologization of the regions of Russia by the number of employed by type of economic activity has been developed taking into account territorial, social, sectoral and climatic features.*

**Keywords:** region, cluster analysis, discriminant analysis, fuzzy clustering, labor market, employment by type of economic activity, R.

JEL: C59, C69, J21

*Дата поступления – 23.09.2022 г.*

**ГАВРИЛЕНКО Юлия Евгеньевна**

аспирантка; лаборант-исследователь научной лаборатории «Облачных технологий и аналитики Больших данных»;

Федеральное государственное бюджетное образовательное учреждение высшего образования «Российский экономический университет имени Г.В. Плеханова» / Стремянный пер., д. 36, г. Москва, 117997.

e-mail: Gavrilenko.YE@rea.ru

**GAVRILENKO Iuliia E.**

Postgraduate Student; Research Assistant at the Scientific Laboratory of “Cloud Technologies and Big Data Analytics”;

Federal State Budget Educational Institution of Higher Education “Plekhanov Russian University of Economics” / 36, Stremyanny Lane, Moscow, 117997.

E-mail: Gavrilenko.YE@rea.ru

**Для цитирования:**

Гавриленко Ю.Е. Методы устойчивой кластеризации регионов России по занятости населения // Федерализм. 2022. Т. 27. № 3 (107). С. 160–177. DOI: <http://dx.doi.org/10.21686/2073-1051-2022-3-160-177>