

Кластеризация субъектов Российской Федерации по ресурсной обеспеченности населения в сфере здравоохранения и использованию коечного фонда

Цель исследования заключается в анализе и выявлении структуры параметров ресурсной обеспеченности населения в сфере здравоохранения и эффективности использования коечного фонда в субъектах Российской Федерации. Такое структурирование позволит классифицировать как субъекты Российской Федерации, так и выявить среди анализируемых параметров факторы, в большей степени определяющие региональные различия и межрегиональные диспропорции с целью приоритетного учета этих факторов при разработке основных направлений государственного регулирования межрегиональных различий субъектов Российской Федерации в сфере здравоохранения.

Материалы и методы. Для достижения поставленной цели на основе данных Единой межведомственной информационно-статистической системы был проведен кластерный анализ субъектов Российской Федерации на основе итеративного метода *k*-средних, реализованного в программном пакете Statistica.

Результаты. В результате кластерного анализа субъектов Российской Федерации по шести признакам были выделены девять кластеров и определены их профильные характеристики. Основные результаты проведенного многомерного статистического анализа включают: параметры дисперсионного анализа; величины суммарных внутрикластерных дисперсий; значения межкластерных сумм квадратов расстояний; график средних значений по шести признакам для каждого из девяти кластеров, а также описательные статистики для каждого объекта в пределах выделенных кластеров. По результатам дисперсионного анализа для всех вариантов кластеризации, по величине (и уровню значимости) *F*-значений и величине эмпирического коэффициента детерминации выявлена боль-

шая значимость для распределения объектов по кластерам следующих признаков: обеспеченность больничными койками на 10 тыс. населения; обеспеченность населения средними медицинскими работниками, работающими в государственных и муниципальных медицинских организациях (человек на 10 тыс. населения) и обеспеченность населения врачами, работающими в государственных и муниципальных медицинских организациях (человек на 10 тыс. населения).

Заключение. Описательные статистики для каждого российского региона в рамках выделенных девяти кластеров показали, что внутри кластеров наблюдается относительно низкая вариабельность значений признаков. Это свидетельствует о высокой степени однородности и компактности построенных кластеров. Полученные результаты кластеризации позволяют определить ресурсные профили российских регионов и сформировать основные приоритеты политики развития регионального здравоохранения России, нацеленные на повышение степени однородности регионов страны по качеству и доступности медицинской помощи. Практическая значимость исследования заключается в том, что его результаты позволят более эффективно разрабатывать региональную политику и проводить государственное регулирование сферы здравоохранения Российской Федерации, а также определять меры для поддержки отрасли в регионах с низкими показателями эффективности использования коечного фонда.

Ключевые слова: интеллектуальный анализ данных, кластерный анализ, дисперсионный анализ, программный пакет Statistica, метод *k*-средних, метод Elbow

Igor A. Kirshin, Alice I. Kasimova

Kazan (Volga Region) Federal University, Kazan, Russia

Clustering of the Russian Federation Regions by Resource Provision of the Population in the Sphere of Healthcare and Use of Hospital Stock

The purpose of the study is to analyze and identify the structure of the parameters of resource provision of the population in the field of healthcare and the efficiency of using hospital stock in the regions of the Russian Federation. Such structuring will allow classifying both the regions of the Russian Federation and identifying among the analyzed parameters the factors that largely determine regional differences and interregional disparities in order to prioritize these factors when developing the main directions of state regulation of interregional differences in the regions of the Russian Federation in the field of healthcare.

Materials and methods. To achieve this purpose, a cluster analysis of the regions of the Russian Federation was carried out on the basis of data from the Unified Interdepartmental Information and

Statistical System using the iterative *k*-means method implemented in the Statistica software package.

Results. As a result of the cluster analysis of the regions of the Russian Federation by six features, nine clusters were identified and their profile characteristics were determined. The main results of the multivariate statistical analysis include: analysis of variance parameters; values of total intra-cluster variances; values of inter-cluster sums of squared distances; a graph of average values for six features for each of the nine clusters, as well as descriptive statistics for each object within the selected clusters. Based on the results of the variance analysis for all clustering options, the magnitude (and significance level) of the *F*-values and the magnitude of the empirical coefficient of determination revealed a high significance

for the distribution of objects into clusters of the following features: provision of hospital beds per 10 thousand population; provision of the population with mid-level medical workers working in state and municipal medical organizations (people per 10 thousand population) and provision of the population with doctors working in state and municipal medical organizations (people per 10 thousand population).

Conclusion. Descriptive statistics for each Russian region within the nine selected clusters showed that there is a relatively low variability of feature values within the clusters. This indicates a high degree of homogeneity and compactness of the constructed clusters. The obtained clustering results allow us to determine the resource profiles

of Russian regions and formulate the main priorities of the regional healthcare development policy in Russia, aimed at increasing the degree of homogeneity of the country's regions in terms of quality and accessibility of medical care. The practical significance of the study is that its results will allow us to more effectively develop regional policies and carry out state regulation of the healthcare sector in the Russian Federation, as well as to determine measures to support the industry in regions with low indexes of hospital stock efficiency.

Keywords: data mining, cluster analysis, analysis of variance, Statistica software package, k-means method, Elbow method.

Введение

Выявление структуры связей и спецификация зависимостей в социально-экономических процессах и явлениях всегда непростая задача уже потому, что выбор метода решения зависит от большого числа факторов разной природы. Решения в таких ситуациях, как правило, принимаются на основании анализа ограниченного статистического материала с использованием методов многомерного статистического анализа. К таким методам относятся многомерные статистические методы классификации, предназначенные для разделения наблюдаемой совокупности объектов, признаков или явлений на гомогенные в определенном смысле группы. Такая классификация производится посредством машинного обучения без учителя (Unsupervised Learning) методами кластерного (естественной классификации) и дискриминантного анализа. Большинство алгоритмов обучения без учителя используются для выявления скрытых структур данных. По сути, происходит решение задачи классификации данных на основе выявления неявной структуры данных. Если рассматривать отдельно классификацию объектов и классификацию признаков, то классификация признаков (группировка столбцов матрицы данных) происходит посредством факторного анализа, когда выявляются обобщенные факторы. Анализ структуры множества объектов (группировка строк матрицы

данных) осуществляется в кластерном анализе. В последние годы в научной литературе наблюдается повышенный интерес к вопросам кластеризации регионов по различным признакам [1, 2, 3, 4].

Говоря формальным языком, задача кластерного анализа состоит в разделении множества объектов X на k (k – целое) кластеров (подмножеств) $Q_1, Q_2, \dots, Q_k$, так, чтобы каждый объект X_j принадлежал одному и только одному подмножеству. Объекты, составляющие каждый отдельный кластер, должны быть схожими или однородными, в то время как объекты, формирующие разные кластеры, разнородными [5]. Для определения количественной степени сходства или близости классифицируемых объектов в кластерном анализе применяются меры близости (подобия), оценивающие метрическое расстояние между ними.

Каждый объект может быть описан несколькими, на-

пример, m признаками, графически представляющими его «точкой» m -мерного пространства, в котором степень его сходства (однородности) с другими объектами будет определяться как расстояние между ними. Важным шагом кластеризации является определение мощности множества кластеров или их числа. Количество кластеров в зависимости от целей и методов решения задачи кластеризации, может быть либо определено заранее, либо априори неизвестно. Традиционно выделяют следующие методы кластеризации (рис. 1) [6].

Кластерный анализ производит кластеризацию объектов практически любой природы. Однако, предварительно данные, как правило, масштабируют одним из методов нормализации данных. В результате кластерного анализа происходит сокращение изначально большого объема информации, сжатие массивов больших данных и выявление их структуры.



Рис. 1. Методы кластеризации

Fig.1. Clustering methods

Материалы и методы

Классификация неравенства регионального развития является важным фактором в стратегии социально-экономического развития любой страны. Однако при группировке данных часто не учитываются скрытые информационные паттерны данных, поэтому такие группировки не способны обеспечить достижения целевых результатов. Социально-экономическое развитие российской региональной экономики, в том числе сферы здравоохранения, традиционно характеризуется неоднородностью и межрегиональными дисбалансами [7]. Исследователи регионального развития нашей страны доказывают, что спецификация «межрегиональных различий субъектов Российской Федерации посредством кластеризации позволит определить их профили и сформировать основные приоритеты стратегии развития здравоохранения РФ, нацеленные на повышение степени однородности регионов России по качеству и доступности медицинской помощи» [8]. В качестве основных инструментов анализа эффективности деятельности сферы здравоохранения применяют способы группировки (деление изучаемой совокупности объектов на количественно-однородные группы по соответствующим признакам) для выявления особенных и типичных взаимодействий процессов, отсеивающих случайные отклонения [9].

Исходя из тезиса, что одной из основных причин неравенства является неравномерное распределение ресурсов между регионами [10], проведем кластеризацию российских регионов (субъектов РФ) на основе показателей обеспеченности базовыми ресурсами здравоохранения и эффективности использования коечного фонда за 2024 г. [11]. Целью анализа является

определение кластеров субъектов РФ со схожими характеристиками и приоритизация значимости выбранных признаков для кластеризации объектов.

В качестве признаков, характеризующих ресурсное обеспечение субъектов РФ в сфере здравоохранения, были выбраны 4 показателя из базы данных Единой межведомственной информационно-статистической системы (ЕМИСС). Это параметры, определяющие потенциал системы здравоохранения по обеспечению релевантного объема медицинской помощи:

- обеспеченность населения врачами, работающими в государственных и муниципальных медицинских организациях (человек на 10 тыс. населения) (D);
- мощность амбулаторно-поликлинических учреждений, посещений в смену на 10 тыс. населения (P);
- обеспеченность населения средними медицинскими работниками, работающими в государственных и муниципальных медицинских организациях (человек на 10 тыс. населения) (M);
- обеспеченность больничными койками на 10 тыс. населения (B).

В качестве метрик, характеризующих эффективность использования коечного фонда, отобраны следующие показатели из базы данных ЕМИСС:

- средняя занятость койки в году в днях (O);
- средняя длительность пребывания больного на койке в днях (T).

Такой набор характеристик регионов обеспечивает аналитикам и практикам возможность оценки рационального использования внутренних ресурсов для повышения медицинской, социальной и экономической эффективности объектов сферы здравоохранения регионов РФ. Задача эффективного использования коечного фонда является од-

ной из приоритетных при организации медицинской помощи надлежащего качества в условиях ограниченного финансирования сферы общественного здравоохранения.

Для обеспечения сбалансированности выборки данных из анализа были исключены субъекты Южного федерального округа и Северо-Кавказского федерального округа, т.к. по ним отсутствуют данные по показателям эффективности использования коечного фонда за анализируемый период.

Анализ был проведен с использованием программного математического пакета Statistica. Предварительно средствами пакета Statistica было проведено нормирование данных методом нормализации средним (Z -нормализацией) по формуле:

$$x'_i = (x_i - \bar{x}) / \sigma_x$$

В модуле кластерного анализа данного пакета заложены три метода: иерархическая кластеризация (объединение, древовидная кластеризация (joining/tree clustering)), кластеризация методом k -средних (k -means clustering) и двухходовая кластеризация (two-way joining).

Наиболее распространенным среди неиерархических является метод k -средних, предполагающий априорное знание конечного числа кластеров k и минимизирующий сумму квадратов внутрикластерных расстояний до центра кластера (Within-cluster sum of squares, WCSS) или, другими словами, величину суммарной внутрикластерной дисперсии (функции потерь). В результате исходное множество разбивается на k максимально удаленных друг от друга кластеров [12, 13].

Выбор числа кластеров при использовании данного метода эксперт задает самостоятельно. Для более обоснованного выбора можно воспользоваться критерием расстояния между

объектами. Этот приём рекомендуется использовать при первоначальной оценке данных, когда отсутствует представление о структуре исходного датасета. Наиболее часто применяемыми методами определения мощности множества кластеров выступают метод локтя (Elbow method) и экспертный метод. В методе локтя график зависимости суммы квадратов внутрикластерных расстояний от числа кластеров позволяет определить точку, в которой добавление новых кластеров не приводит к значительному уменьшению WCSS. Координата этой точки и определяет оптимальное число кластеров. В экспертном методе выбор количества кластеров определяется знаниями о предметной области наблюдаемых объектов (domain knowledge).

Результаты исследования

Для кластеризации российских регионов нами был использован метод k -средних с применением метода локтя, позволяющим определить оптимальное число кластеров. По результатам дисперсионного анализа (ANalysis Of VAriance, ANOVA) [14] были определены WCSS для различных значений числа кластеров и построен график, иллюстрирующий метод локтя (рис. 2).

Исходя из полученного графика, отметим, что для нашей выборки данных наблюдается заметный излом графика при числе кластеров, равном девяти. Это свидетельствует о том, что множество из девяти кластеров определяется как наиболее релевантно структурированное разделение исходного множества объектов.

В соответствии с методологией ANOVA и логикой метода Elbow нами был разработан новый критерий определения релевантного числа кластеров. Как известно, в методе k -средних межкластерная сумма квадратов (BCSS) выступает мерой

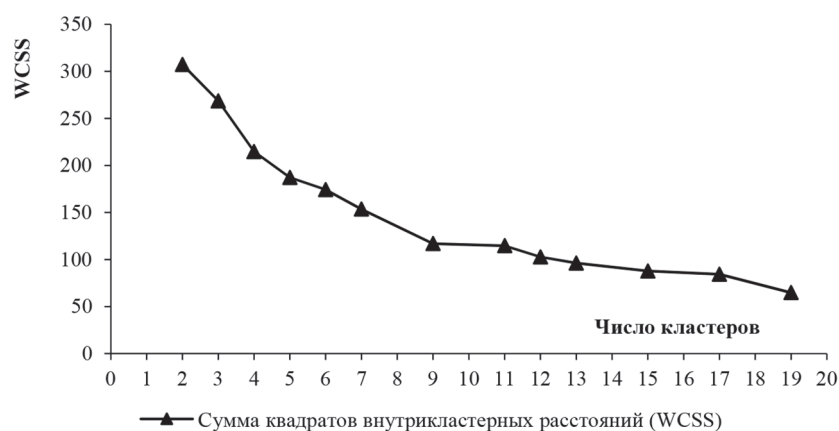


Рис. 2. График Elbow

Fig. 2. Elbow graph

разброса между кластерами. Внутрикластерная сумма квадратов (WCSS) является мерой сходства объектов или характеристикой компактного расположения объектов без значительных промежутков внутри пространства каждого кластера. Следовательно, критерием оптимальности кластеризации может служить максимизация BCSS и минимизация WCSS. Иными словами, близкое к единице значение отношения BCSS/TSS будет характеризовать полученную кластеризацию как оптимальную, т.к. большая часть общей дисперсии будет объясняться межкластерными различиями.

Развивая логику данного методического подхода, можно констатировать, что кри-

терием оптимального числа кластеров может выступить снижение темпа прироста показателя «Отношение межкластерной суммы квадратов к общей сумме квадратов (BCSS/TSS)» при увеличении числа кластеров на единицу по сравнению с предыдущей мощностью множества кластеров. На построенном графике (рис. 3) отчетливо выделяется количество кластеров, равное 9, преодоление которого приводит к снижению темпов прироста критерия BCSS/TSS. Выявленное с помощью этого метода количество кластеров, равное девяти, совпадает с определенным ранее по методу локтя.

Полученные результаты кластеризации российских регионов представлены в таблице 1.

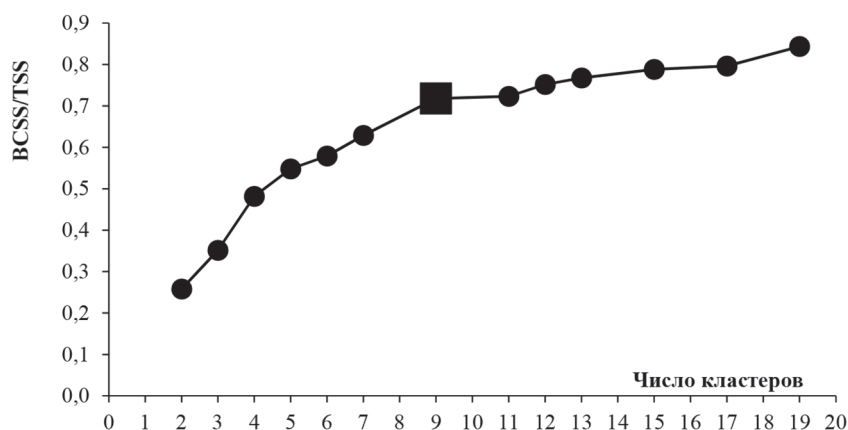


Рис. 3. Отношение межкластерной суммы квадратов к общей сумме квадратов (BCSS/TSS)

Fig. 3. Ratio of inter-cluster sum of squares to total sum of squares (BCSS/TSS)

Таблица 1 / Table 1

Состав девяти кластеров
Composition of nine clusters

№ кластера	Число субъектов РФ	Субъекты РФ
1	12	Курская область, Орловская область, Рязанская область, Смоленская область, Тамбовская область, Ярославская область, Удмуртская Республика, Нижегородская область, Ульяновская область, Иркутская область, Томская область, Амурская область
2	5	Республика Коми, Республика Тыва, Камчатский край, Магаданская область, Сахалинская область
3	15	Белгородская область, Калининградская область, Новгородская область, Республика Башкортостан, Республика Татарстан, Самарская область, Саратовская область, Тюменская область, Челябинская область, Республика Хакасия, Красноярский край, Кемеровская область, Новосибирская область, Республика Бурятия, Хабаровский край
4	1	Чукотский автономный округ
5	3	Костромская область, Тверская область, Еврейская автономная область
6	10	Брянская область, Владимирская область, Липецкая область, Вологодская область, Псковская область, Республика Марий Эл, Оренбургская область, Курганская область, Республика Алтай, Алтайский край
7	9	Воронежская область, Тульская область, Республика Мордовия, Пермский край, Пензенская область, Свердловская область, Омская область, Омская область, Забайкальский край, Приморский край
8	4	Калужская область, Московская область, г. Москва, Ленинградская область
9	11	Ивановская область, Республика Карелия, Архангельская область, Ненецкий автономный округ, Мурманская область, г. Санкт-Петербург, Чувашская Республика, Кировская область, ХМАО, ЯНАО, Республика Саха (Якутия)

Таблица 2 / Table 2

Результаты дисперсионного анализа для $k = 9$
ANOVA results for $k = 9$

Variable	BCSS	df	WCSS	df	F	signif.
D	52,55801	8	16,44199	61	16,91214	0,000000
P	46,03658	8	22,96342	61	11,88755	0,000000
M	52,71965	8	16,28035	61	24,69157	0,000000
B	55,31485	8	13,68515	61	30,81996	0,000000
O	42,51755	8	26,48245	61	12,24193	0,000000
T	47,99149	8	21,00851	61	15,24153	0,000000

Результаты кластеризации показывают, что самым большим по числу входящих в него регионов (15 регионов) является кластер 3. Самым малочисленным является кластер 4, образованный Чукотским автономным округом.

Дисперсионный анализ показал статистическую значимость всех выбранных для кластеризации шести переменных (Таблицы 2 и 3).

По величине F-значений

можно судить о значимости признаков для распределения объектов по кластерам. В нашем случае к первой тройке значимых параметров кластеризации по результатам ANOVA для всех вариантов кластеризации относятся признаки: «обеспеченность больничными койками на 10 тыс. населения (B)»; «обеспеченность населения средними медицинскими работниками, работающими в государственных и муници-

пальных медицинских организациях (человек на 10 тыс. населения) (M)»; «обеспеченность населения врачами, работающими в государственных и муниципальных медицинских организациях (человек на 10 тыс. населения) (D)».

Этот вывод подтверждается и проверкой по другому общепризнанному критерию существенности влияния факторов, по которым производится кластеризация – критерию эмпирического коэффициента детерминации (причинности). Он рассчитывается, как отношение межгрупповой дисперсии к общей дисперсии. Данный критерий также выявил большую значимость для кластеризации тех же признаков.

В таблице 3 указаны расстояния между кластерами. Выше диагонали приведены квадраты расстояний, а ниже – евклидово расстояние. Был также проведен анализ расстояний от объектов до центров кластеров для оценки степени принадлежности каждого объекта к соответствующему кластеру. В среднем, объекты находятся на небольшом расстоянии от центров своих кластеров, что указывает на высокую степень внутренней однородности кластеров.

Используя опцию пакета Statistica «График средних» (Graph of means) (рис. 4), оценим различия в средних значениях анализируемых признаков (переменных) для каждого кластера. График средних значений демонстрирует, что полученные девять кластеров имеют различные средние значения признаков, что подтверждает различия профилей кластеров по анализируемым признакам.

Для более полной оценки эффективности использования коечного фонда нами был дополнительно рассчитан показатель «Оборот койки» (BT) путем деления среднегодовой занятости койки на среднюю длительность пребывания па-

Таблица 3 / Table 3

Расстояния между кластерами (Эвклидово расстояние)
Distances between clusters (Euclidean distance)

№ кластера	1	2	3	4	5	6	7	8	9
1	0,0000	1,6573	0,4676	10,220	1,7050	0,8284	0,5193	2,6247	0,6149
2	1,2874	0,0000	3,2495	4,7571	2,9247	2,1659	2,6939	6,6983	1,2896
3	0,6838	1,8026	0,0000	14,111	2,0985	0,6427	0,3877	1,0923	0,9779
4	3,1969	2,1811	3,7566	0,0000	10,909	11,336	12,992	20,534	10,228
5	1,3058	1,7102	1,4486	3,3030	0,0000	0,9457	1,0704	3,7220	2,9894
6	0,9101	1,4717	0,8017	3,3670	0,9725	0,0000	0,5266	2,0301	1,2318
7	0,7206	1,6413	0,6227	3,6045	1,0346	0,7257	0,0000	1,2327	0,9696
8	1,6201	2,5881	1,0451	4,5315	1,9292	1,4248	1,1103	0,0000	2,6759
9	0,7841	1,1356	0,9889	3,1983	1,7290	1,1099	0,9847	1,6358	0,0000

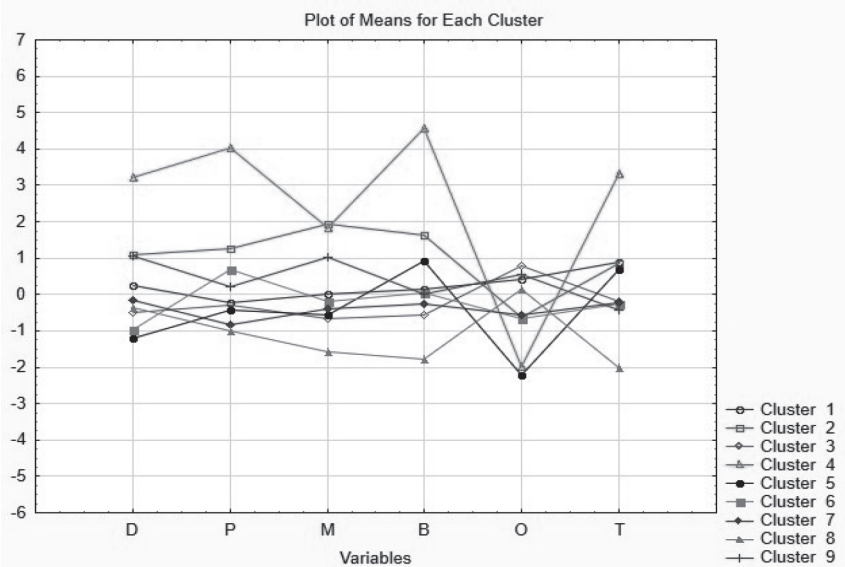


Рис. 4. Средние значения параметров кластеров по методу *k*-средних
Fig. 4. Average values of cluster parameters using the *k*-means method

пациента на койке для каждого кластера. Данный показатель выступает интегральным результатом оценки эффективности использования коечного фонда, комбинирующим среднегодовую занятость койки и среднюю длительность пребывания пациента на койке.

Среди субъектов РФ выделяется своими максимальными значениями признаков четвертый кластер, состоящий из единственного региона - Чукотского автономного округа. Параметры Чукотского автономного округа можно определить как статистические выбросы и основное внимание уделить выявлению профилей других кластеров.

Тогда, российские регионы, составляющие кластер 2 будут

иметь наивысшие значения по всем признакам ресурсной обеспеченности. Этот вывод совпадает с данными Рейтинга регионов по показателям обеспеченности медицинских организаций врачами, койками, средним медперсоналом на 10 000 человек в 2024 г [15]. Высокую ресурсную обеспеченность (на 10 000 человек) логично объяснить малочисленностью населения регионов кластера 2. Республика Коми, Республика Тыва, Камчатский край, Магаданская и Сахалинская области по численности населения входят в последнюю десятку российских регионов. Однако, по показателям эффективности использования коечного фонда данные регионы выступают одними из

аутсайдеров. Они отличаются низкой среднегодовой занятостью койки и одной из самых высоких средней длительностью пребывания больного на койке.

Принимая во внимание методику расчета среднегодовой занятости койки, как результат деления числа койко-дней, проведенных всеми больными, на число среднегодовых коек, высокими значениями параметра «В» можно объяснить низкое значение среднего числа дней занятости койки в году в регионах кластера 2. Продолжительная длительность пребывания больного на койке может свидетельствовать о несовершенстве методов обследования больных, недостаточной преемственностью между амбулаторно-поликлиническими учреждениями и стационарами, большем удельном весе длительно протекающих заболеваний в этих регионах.

Взаимодействие всех этих факторов определяет прогноз динамики временного ряда средней длительности стационарного лечения и, в конечном итоге, профиль данного кластера.

На противоположном полюсе лидеров кластерного рейтинга находится кластер 8, который формируют г. Москва, Московская, Ленинградская и Калужская области. Действительно, значения параметров ресурсной обеспеченности данных регионов сравнительно низкие и в то же время показатели эффективности использования коечного фонда — высокие (Таблица 4). Так, по мощности амбулаторно-поликлинических учреждений Московская область занимает последнее место в списке, Калужская область — 64 место из 70, г. Москва — 39. По обеспеченность населения средними медицинскими работниками, работающими в государственных и муниципальных медицинских организациях и обе-

Таблица 4 / Table 4

Сравнительный анализ значений параметров субъектов РФ кластера 8 со среднероссийским уровнем
Comparative analysis of the values of parameters of the Russian Federation regions of cluster 8 with the average Russian level

	D	P	M	B	O	T	BT
Российская Федерация	42,14	233,9	88,25	66,8	293,9	9,6	30,61
г. Москва	54	248,5	69,2	44,8	289,5	6,2	46,69
Московская область	36,3	180,2	64,4	46,6	301,8	8,3	36,36
Ленинградская область	33,7	212,5	67,3	52	284,3	8,7	32,68
Калужская область	33,57	196,5	80,32	57	297,3	9,1	32,67

Таблица 5 / Table 5

Рейтинг кластеров по показателю «Оборот койки» (BT)
Cluster rating by the index “Bunk turnover” (BT)

Ранг кластера	Кластеры	Среднее значение параметров по кластерам						
		D	P	M	B	O	T	BT
1	Кластер 8	39,39	209,4	70,31	50,1	293,2	8,075	36,31
2	Кластер 9	53,47	267,7	118,9	74,06	302,1	9,75	30,98
3	Кластер 3	38,33	244,7	87,24	66,68	307,1	10,02	30,65
	Российская Федерация	42,14	233,9	88,25	66,8	293,9	9,6	30,61
4	Кластер 6	33,4	291,8	96	74,6	275,2	9,9	27,8
5	Кластер 7	41,6	218,3	92,4	70,5	277,7	10	27,77
6	Кластер 1	45,58	248,1	99,77	75,88	299,2	11,2	26,72
7	Кластер 2	53,82	319,4	135,7	95,96	276	11,16	24,73
8	Кластер 5	31,3	237	88,8	86,5	240,3	11	21,85
9	Кластер 4	74,75	452,2	134,1	135,1	246	13,8	17,83

спеченности больничными койками Московская область, Ленинградская область и г. Москва занимают последние места в списке субъектов РФ. Однако, оборот койки в этих регионах демонстрирует максимальное значение.

Структура этих факторов в конечном итоге определяет более высокую степень организации общественного здравоохранения в г. Москва, Московской, Ленинградской и Калужской областях по сравнению с прочими субъектами РФ.

В таблице 5 представлен рейтинг кластеров по показателю «Оборот койки». Наилучшие значения средних среди остальных кластеров с номерами 1, 3, 5–7, 9 демонстрирует кластер 9, который близок по значениям ресурсной обеспеченности и эффективности использования коечного фонда к кластеру 8. Следующий по иерархии кластер, регионы которого демонстрируют высокую

эффективность использования коечного фонда по сравнению с другими кластерами, третий. В то же время он отличается меньшими значениями ресурсного обеспечения по сравнению с кластером 9.

Ресурсная обеспеченность кластеров 6 и 7 по сравнению с прочими кластерами в меньшей степени отличается от общероссийской. Однако по показателям эффективности использования коечного фонда регионы этих кластеров уступают общероссийскому уровню и занимают 4 и 5 место в рейтинге. Характеризуя кластер 1, отметим, что он отличается большей ресурсной обеспеченностью по сравнению со среднероссийским уровнем. При этом по показателю оборота койки он уступает общероссийскому уровню (6 место в рейтинге). Пятый кластер характеризуется самой низкой средней занятостью койки, оборотом койки и большой длительностью пребывания

больного на койке, если не учитывать параметры кластера 4. В 2024 году в регионах данного кластера каждая койка в среднем 125 дней в году была не занята.

При проведении кластерного анализа методом k-средних важным критерием качества кластеризации и идентификации структуры данных является оценка компактности полученных кластеров. Компактность кластеров отражает степень близости объектов кластера друг к другу и может быть оценена с использованием различных критериев, основанных на описательных статистиках. В пакете Statistica для каждого из девяти кластеров были рассчитаны следующие описательные статистики по анализируемым шести признакам: среднее значение (Mean), стандартное отклонение (Standard Deviation) и дисперсия (Variance). На основе этих описательных статистик можно рассчитать несколько критериев для оценки компактности кластеров:

1. Среднее расстояние от каждого объекта кластера до его центра. Центр класса - средние величины по всем признакам для этого класса. Чем меньше расстояния до центра, тем типичнее объект для данного кластера, т.е. чем меньше внутрикластерная дисперсия, тем более «упорядочены» и «однородны» кластеры.
2. Стандартное отклонение (Standard Deviation) показывает, насколько сильно значения признаков внутри кластера отклоняются от среднего значения. Низкое значение стандартного отклонения указывает на то, что объекты внутри кластера расположены близко друг к другу, что свидетельствует о высокой компактности кластера.
3. Относительное стандартное отклонение: коэффициент вариации (CV) — это статистическая мера, используемая для оценки относительной измен-

чивости набора данных. Он представляет собой отношение стандартного отклонения к среднему значению и обычно выражается в процентах. Коэффициент вариации полезен для сравнения степени вариативности данных в различных наборах данных, особенно когда средние значения этих наборов различаются. Кластер с наименьшим значением CV может считаться наиболее компактным.

4. Сумма квадратов внутрикластерных отклонений является одним из основных критериев оценки компактности кластеров в методе k-средних. Она рассчитывается как сумма квадратов расстояний между каждым объектом кластера и его центром. Чем меньше значение суммы, тем более компактным является кластер.

5. Среднее значение суммы квадратов внутрикластерных отклонений. Этот критерий позволяет более объективно оценить компактность кластеров, учитывая их различную размерность. Критерий характеризует *меру разброса* значений объектов кластера относительно его центроида.

6. Относительная компактность (Relative Compactness) может быть оценена путем сравнения значений описательных статистик одного кластера с аналогичными значениями других кластеров. Например, кластер с наименьшим значением стандартного отклонения или дисперсии может считаться наиболее компактным.

7. Индексы кластеризации. Существуют различные индексы и метрики, например, индекс силуэта (silhouette score), который оценивает качество кластеризации и компактность кластеров на основе не только стандартного отклонения, но и других факторов.

Описательные статистики для каждого российского региона в пределах выделенных девяти кластеров показали,

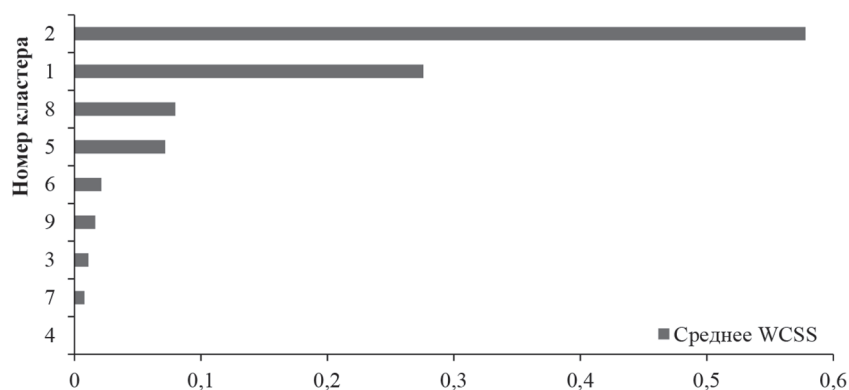


Рис. 5. Среднее WCSS по кластерам

Fig. 5. Average WCSS by clusters

что внутри кластеров наблюдается относительно низкая вариабельность значений признаков. Это свидетельствует о высокой степени внутренней гомогенности и компактности построенных кластеров.

Для оценки степени компактности был применен один из наиболее объективных критериев — среднее значение WCSS для каждого числа кластеров (рис. 5).

Наименьшей получилась средняя WCSS по кластеру 7. Это означает, что регионы этого кластера обладают самыми схожими признаками. Максимальная средняя WCSS — во 2-м кластере. Это означает, что объекты этого кластера расположены на большем расстоянии от центра и поэтому этот кластер характеризуется самой большой неоднородностью объектов.

Основываясь на расчете средней WCSS, можно утверждать, что кластеры с номерами 3, 6, 7, 9 сгруппированы более компактно. Этот вывод справедлив и для кластера 5, за исключением Тверской области. Образующие его Еврейская автономная и Костромская области располагаются на практически равных расстояниях от его центра. В то время как Тверская область удалена от центра на большее расстояние. Такой же вывод справедлив и для кластера 8 в отношении г. Москва, которая, являясь лидером по анализи-

руемым параметрам, удалена от центра на расстояние примерно в 2,5 раза превышающее расстояния от центра трех других регионов данного кластера.

Заключение

В ходе проведенного исследования была успешно выполнена кластеризация субъектов РФ и установлены качественные взаимосвязи между группами российских регионов с близкими значениями признаков обеспеченности базовыми ресурсами здравоохранения и эффективности использования коечного фонда методом k-средних с использованием пакета Statistica (Statsoft). Полученные результаты свидетельствуют о том, что выделенные кластеры однородны, компактны и различаются по анализируемым признакам, что обосновывается результатами дисперсионного анализа. Внутренняя однородность кластеров была также подтверждена анализом расстояний от объектов до центров кластеров и описательными статистиками.

По результатам дисперсионного анализа для всех вариантов кластеризации, величине (и уровню значимости) F-значений и величине *эмпирического коэффициента детерминации* выявлена большая значимость для разделения объектов по кластерам следующих признаков: «обеспеченность боль-

ничными койками на 10 тыс. населения»; «обеспеченность населения средними медицинскими работниками, работающими в государственных и муниципальных медицинских организациях (человек на 10 тыс. населения)»; «обеспеченность населения врачами, работающими в государственных и муниципальных медицинских организациях (человек на 10 тыс. населения)». В гораздо меньшей степени распределение объектов по кластерам определяется параметрами «мощности амбулаторно-поликлинических учреждений, посещений в смену на 10 тыс. населения»; «средней занятости койки в году в днях» и «средней длительности пребывания больного на койке в днях».

Для оценки оптимальности разбиения на кластеры были использованы следующие критерии: метод Elbow, метрика «Внутренняя однородность кластеров», оцениваемая по средним расстояниям от объектов до центров кластеров; результаты дисперсионного анализа; профили средних значений параметров; показатель «Вариабельность внутри кластеров». Итоговый рейтинг кластеров составлен с учетом интегрального показателя «Оборот койки».

Результаты кластеризации показали, что проведенный кластерный анализ методом k-средних является эффективным инструментом кластеризации субъектов РФ по ресурсной обеспеченности населения в сфере здравоохранения и эф-

фективности использованию коечного фонда. Полученные результаты могут быть использованы для дальнейшего анализа и интерпретации данных в целях эффективной разработки региональной политики и государственного регулирования сферы здравоохранения Российской Федерации, а также определения мер для поддержки отрасли в регионах с низкими показателями эффективности использования коечного фонда.

Особое внимание при разработке политики развития регионального здравоохранения России, нацеленной на повышение степени однородности регионов страны по качеству и доступности медицинской помощи, следует уделить субъектам РФ, формирующим 1, 2, 4–7 кластерам, характеризующимся меньшим по сравнению со среднероссийским показателем оборота койки.

Низкий показатель оборота коек свидетельствует, во-первых, о недостаточной экономической неэффективности деятельности и невысоком качестве менеджмента медицинских организаций региона. Медицинское оборудование, материальные ресурсы и медицинский персонал используются не в полной мере, что приводит к увеличению себестоимости лечения пациента. Во-вторых, низкий оборот коек может ограничивать доступность, качество медицинской помощи и приводить к росту листа ожидания. В-третьих, продолжительное пребывание пациентов в стационаре

чревато потерей им работы, устойчивым дефицитом семейного бюджета, формированием атмосферы психологического давления на пациентов и их семьи и другими негативными социальными последствиями.

Однако для более точной идентификации профиля данных кластеров необходимо собрать и проанализировать дополнительную информацию. Так, например, средняя длительность пребывания больного на койке обусловлена действием двух разных групп факторов. Первая группа определяется качеством организации медпомощи. Вторая группа практически не зависит от неё. Первая группа включает следующие факторы: своевременность проведения обследований пациента, продолжительность обследования; эффективность и интенсивность лечения; обеспеченность всем необходимым оборудованием для диагностики состояния пациента. Ко второй группе относятся характер заболевания, половозрастной состав пациентов и т.д.

Релевантная оценка эффективности использования коечного фонда позволит рационально использовать ресурсы сферы здравоохранения при оказании медицинской помощи, а также принимать доказательные управленческие решения на основе методов интеллектуального анализа данных как на уровне первичного медицинского звена, так и при управлении системой здравоохранения на региональном уровне РФ и страны в целом.

Литература

1. Меркулова Е.Ю., Богопольский А.С. Дифференциация регионов России: экономические контрасты и социальные различия // Статистика и Экономика. 2025. № 22(3). С. 39–52. DOI: 10.21686/2500-3925-2025-3-39-52.
2. Залманов И.А. Вариативный кластерный анализ занятости населения в городах Российской Федерации // Статистика и Экономика. 2024. № 21(5). С. 15–25. DOI: 10.21686/2500-3925-2024-5-15-25.

3. Горпинченко К.Н., Мороз А.И., Спаниди Д.А. Анализ экономического развития Краснодарского края с применением кластерного анализа // Естественно-гуманитарные исследования. 2024. № 5(53). С. 119–123.
4. Delamater Paul., Shortridge Ashton., Messina Joseph. Regional health care planning: A methodology to cluster facilities using community utilization patterns. BMC health services research. 2013. DOI: 10.1186/1472-6963-13-333.

5. Duran B., Odell P. Cluster analysis. A survey. Springer-Verlag. Berlin – Heidelberg – N.Y. 1974.
6. Калинина В.Н., Соловьев В.И. Введение в многомерный статистический анализ. М.: ГУУ, 2003. 66 с.
7. Трейвиш А. И. Неравномерность и структурное разнообразие пространственного развития экономики как научная проблема и российская реальность // Пространственная экономика. 2019. Т. 15. №4. С. 13–35.
8. Габдуллин Н.М., Киршин И.А., Шулаев А.В. Регулирование межрегиональных различий субъектов Российской Федерации в контексте национальных проектов «Здравоохранение» и «Демография» // Уровень жизни населения регионов России. 2020. Т. 16. № 3. С. 59–69. DOI: 10.19181/lsprr.2020.16.3.5.
9. Железнякова И. А., Ковалева Л. А., Хелисупали Т. А., Войнов М. А., Омеляновский В. В. Методология оценки эффективности использования коечного фонда медицинских организаций // ФАРМАКОЭКОНОМИКА. Современная фармакоэкономика и фармакоэпидемиология. 2017. № 10(4). С. 37–43. DOI: 10.17749/2070-4909.2017.10.4.037-043.
10. Бондаренко Н.Е., Губарев Р.В. Проблема регионального неравенства в социально-экономическом развитии Российской Федерации // Вестник Российского экономического университета имени Г.В. Плеханова. 2020. Т. 17. № 5(113). С. 56–68.
11. ЕМИСС. Ключевые показатели обеспеченности медицинских организаций в статистической системе Минздрава/Росстата [Электрон. ресурс]. Режим доступа: <https://www.fedstat.ru/indicator/61875>.
12. Jain A.K., Murty M.N., Flynn P.J. Data clustering: A Review [Электрон. ресурс] // ACM Computing Surveys. 1999. Т. 31. № 3. Режим доступа: http://users.eecs.northwestern.edu/~yingliu/datamining_papers/survey.pdf.
13. Alex Smola and S.V.N. Vishwanathan. Introduction to machine learning. Cambridge University Press, 2008. 234 с.
14. Bykowska-Derda A., Zielinska-Dawidziak M., Czlapka-Matyasik M. Dietary-Lifestyle Patterns Associated with Bone Turnover Markers, and Bone Mineral Density in Adult Male Distance Amateur Runners – A Cross-Sectional Study // Nutrients. 2022. № 14(10). 2048. DOI: 10.3390/nu14102048.
15. Рейтинг регионов – 2024. Показатели системы здравоохранения [Электрон. ресурс]. Режим доступа: <https://expertnw.com/from-editors/rejting-regionov-za-2024-g-pokazateli-sistemy-zdravookhraneniya/>.

References

1. Merkulova Ye.Yu., Bogopol'skiy A.S. Differentiation of Russian Regions: Economic Contrasts and Social Differences. Statistika i Ekonomika = Statistics and Economics. 2025; 22(3): 39–52. DOI: 10.21686/2500-3925-2025-3-39-52. (In Russ.)
2. Zalmanov I.A. Variable Cluster Analysis of Employment in the Cities of the Russian Federation. Statistika i Ekonomika = Statistics and Economics. 2024; 21(5): 15–25. DOI: 10.21686/2500-3925-2024-5-15-25. (In Russ.)
3. Gorpinchenko K.N., Moroz A.I., Spanidi D.A. Analysis of Economic Development of Krasnodar Krai Using Cluster Analysis. Yestestvenno-gumanitarnyye issledovaniya = Research in Natural Sciences and Humanities. 2024; 5(53): 119–123. (In Russ.)
4. Delamater Paul., Shortridge Ashton., Messina Joseph. Regional health care planning: A methodology to cluster facilities using community utilization patterns. BMC health services research. 2013. DOI: 10.1186/1472-6963-13-333.
5. Duran B., Odell P. Cluster analysis. A survey. Springer-Verlag. Berlin – Heidelberg – N.Y. 1974.
6. Kalinina V.N., Solov'yev V.I. Vvedeniye v mnogomernyy statisticheskiy analiz = Introduction to multivariate statistical analysis. Moscow: GUU; 2003. 66 p. (In Russ.)
7. Treyvish A.I. Unevenness and structural diversity of spatial economic development as a scientific problem and Russian reality. Prostranstvennaya ekonomika = Spatial economy. 2019; 15; 4: 13–35. (In Russ.)
8. Gabdullin N.M., Kirshin I.A., Shulayev A.V. Regulation of interregional differences in the constituent entities of the Russian Federation in the context of the national projects “Healthcare” and “Demography”. Uroven' zhizni naseleniya regionov Rossii = Standard of Living of the Population of Russian Regions. 2020; 16; 3: 59–69. DOI: 10.19181/lsprr.2020.16.3.5. (In Russ.)
9. Zheleznyakova I.A., Kovaleva L.A., Khelisupali T.A., Voynov M.A., Omel'yanovskiy V.V. Methodology for assessing the effectiveness of using the bed capacity of medical organizations. FARMAKOEKONOMIKA. Sovremennaya farmakoeconomika i farmakoepidemiologiya = PHARMACOECONOMIKA. Modern pharmacoeconomics and pharmacoepidemiology. 2017; 10(4): 37–43. DOI: 10.17749/2070-4909.2017.10.4.037-043. (In Russ.)
10. Bondarenko N.Ye., Gubarev R.V. The Problem of Regional Inequality in the Socio-Economic Development of the Russian Federation. Vestnik Rossiyskogo ekonomicheskogo universiteta imeni G.V. Plekhanova = Bulletin of the Plekhanov Russian University of Economics. 2020; 17; 5(113): 56–68. (In Russ.)
11. YEMISS. Klyuchevyye pokazateli obespechennosti meditsinskikh organizatsiy v

statisticheskoy sisteme Minzdrava/Rosstata = EMISS. Key Indicators of the Provision of Medical Organizations in the Statistical System of the Ministry of Health/Rosstat [Internet]. Available from: <https://www.fedstat.ru/indicator/61875>. (In Russ.)

12. Jain A.K., Murty M.N., Flynn P.J. Data clustering: A Review [Internet]. ACM Computing Surveys. 1999; 31: 3. Available from: http://users.eecs.northwestern.edu/~yingliu/datamining_papers/survey.pdf.

13. Alex Smola and S.V.N. Vishwanathan. Introduction to machine learning. Cambridge University Press; 2008. 234 p.

14. Bykowska-Derda A., Zielinska-Dawidziak M., Czlapka-Matyasik M. Dietary-Lifestyle Patterns Associated with Bone Turnover Markers, and Bone Mineral Density in Adult Male Distance Amateur Runners - A Cross-Sectional Study. Nutrients. 2022; 14(10): 2048. DOI: 10.3390/nu14102048.

15. Reyting regionov - 2024. Pokazateli sistemy zdravookhraneniya = Regional Rating - 2024. Healthcare System Indicators [Internet]. Available from: <https://expertnw.com/from-editors/rejting-regionov-za-2024-g-pokazateli-sistemy-zdravookhraneniya/>. (In Russ.)

Сведения об авторах

Игорь Александрович Кишин

Д.э.н., профессор, профессор Высшей школы бизнеса

Казанский (Приволжский) федеральный университет, Казань, Россия

Эл. почта: kia1125@mail.ru

Алиса Ильфатовна Касимова

Магистрант кафедры общей гигиены

Эл. почта: ilfatovna11.11.1999@mail.ru

Information about the authors

Igor A. Kirshin

Dr Sci. (Economics), Professor, Professor of the Higher School of Business

Kazan (Volga Region) Federal University, Kazan, Russia

E-mail: kia1125@mail.ru

Alice I. Kasimova

Master's student of the Department of General Hygiene

E-mail: ilfatovna11.11.1999@mail.ru