

КЛАСТЕРНОЕ ЗОНИРОВАНИЕ ПО ФАКТОРАМ РАЗВИТИЯ ЧЕЛОВЕЧЕСКОГО ПОТЕНЦИАЛА

В.П. Герасенко, канд. экон. наук,

П.В. Герасенко,

Белорусский торгово-экономический университет потребительской кооперации

Постановка задачи. В практической деятельности при принятии решений на региональном уровне широко используется экономическое зонирование по какому-либо одному приоритетному признаку (климатические зоны, плодородие земель, глубина промерзания почвы, доходы населения, товарные рынки и т. д.). Отдельные показатели социально-экономического развития могут иметь изменения, связанные с возникновением опасных ситуаций (загрязнение природной среды, изменение климата и т. д.). Следовательно, возникает необходимость непрерывного отслеживания динамики изменения социально-экономических показателей развития региона и уточнения границ выделяемых экономических зон.

Территориальная дифференциация изучаемых параметров, характеризующих количественную и качественную экономическую природу изучаемых объектов, явлений и т. д., может быть выявлена на основе более глубокого изучения комбинаций факторных признаков. При этом возникает необходимость выделения однородных региональных зон, для каждой из которых значения рассмотренных переменных похожи друг на друга и расположены близко на диаграмме рассеяния. Такой подход позволяет расширить содержание проводимого многомерного анализа. Наиболее существенные методологические чер-

ты кластерного анализа, полезные для регионального исследования, могут быть отражены по следующим двум направлениям: образование единой меры, охватывающей ряд характеризующих признаков, и чисто количественное решение вопроса о группировке объектов наблюдения.

Идея классификации по сочетанию ряда признаков давно привлекает внимание специалистов по региональной экономике. Например, в одной из наиболее популярных задач кластерного анализа - группировке районов - она давно признана. Еще в 1920 г., анализируя связь между элементами крестьянского хозяйства, Б. С. Ястребский рассматривал 34 характеристики уездов, влиявшие на эту связь [10, С. 48-69]. Можно привести и другие примеры группировки территориальных единиц по комплексу признаков, неизменно имевших место в задачах районирования. Но в кластерном анализе признаки объединяются с помощью некоторой «метрики» в один количественный показатель сходства (различия) группируемых объектов [4, 5, 9].

В кластерном региональном анализе группировочные признаки могут подвергаться объединению с помощью различных функций расстояния. В данном исследовании в качестве такой функции принятая функция евклидова расстояния, которая является самой распространенной

ной мерой для определения расстояния между двумя точками на плоскости. Такой подход отражает специфику алгоритма решения нашей региональной задачи и, следовательно, требует логического пояснения основных элементов алгоритма.

В простейшем случае для определения расстояния между двумя точками на плоскости, образованной точками двухмерного пространства x и y , используется следующая формула евклидовой меры:

$$R = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2},$$

где x_1 и y_1 - координаты первой точки;
 x_2 и y_2 - координаты второй точки.

При наличии трех и более переменных определение расстояния между двумя точками x и y в любом n -мерном пространстве может быть определено, соответственно, по следующей формуле Евклида:

$$R = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}.$$

В общем случае экономико-математическая постановка задачи состоит в следующем. Пусть множество $I = \{I_1, I_2, \dots, I_n\}$ обозначает n объектов (районов или областей), принадлежащих некоторой искомой однородной зоне в изучаемом регионе P_r . Предположим также, что существует некоторое множество наблюдаемых показателей или характеристик $C = \{C_1, C_2, \dots, C_r\}$. Наблюдаемые характеристики могут быть как количественными, так и качественными. Основная часть нашего исследования будет посвящена количественным данным. Для множества объектов I мы располагаем множеством векторов измерений $X = \{X_1, X_2, \dots, X_n\}$, которые описывают множество I . Отметим, что множество X может быть представлено как n точек в c -мерном евклидовом пространстве E_c .

Задача кластерного анализа заключается в том, чтобы на основании данных, содержащихся в множестве X , разбить множество объектов I на m кластеры (однородных экономических зон) P_1, P_2, \dots, P_m так, чтобы каждый объект I_j принадлежал одному и только одному подмножеству разбиения и чтобы объекты, принадлежащие одному и тому же кластеру, были сходными, в то время как объекты, принадлежащие разным кластерам, были разноданными (несходными).

Приведем постановку региональной задачи кластерного анализа. В качестве региона I рассмотрим n образующих его административно-территориальных образований, каждое из которых характеризуется следующими основными социально-экономическими показателями, исчисляемыми в системе Министерства статистики и анализа Республики Беларусь: среднемесячная зарплата работников (C_1), денежные доходы в расчете на душу населения (C_2), объем платных услуг на человека (C_3) и т. п. [2, 3, 6, 7,

8]. Тогда X_i (вектор измерений) представляет собой набор указанных характеристик для первого административно-территориального района, X_2 - для второго административно-территориального района и т.д. Задача заключается в том, чтобы разбить данные районы по уровню развития человеческого потенциала.

Решением задачи кластерного анализа является разбиение, удовлетворяющее некоторому критерию оптимальности. Этот критерий может представлять собой некоторый функционал, выражющийший уровень желательности различных разбиений и группировок. Этот функционал можно назвать целевой функцией. Например, в качестве целевой функции может быть взята внутригрупповая сумма квадратов отклонений. В качестве примера рассмотрим восемь объектов ($n = 8$) обладающих одной характеристикой (то есть $c = 1$); результаты измерения пусть представляют собой множество $X = \{3, 4, 7, 4, 3, 3, 4, 4\}$. Сумма квадратов отклонений вычисляется по формуле

$$\sum_{i=1}^8 x_i^2 - \frac{1}{8} \left(\sum_{i=1}^8 x_i \right)^2 = 140 - 128 = 12.$$

Если множество X разбить на три группы: $G_1 = \{3, 3, 3\}$, $G_2 = \{4, 4, 4, 4\}$ и $G_3 = \{7\}$, то все внутригрупповые суммы квадратов отклонений должны быть равны нулю:

$$W_1 + W_2 + W_3 = 0 + 0 + 0,$$

где W_i обозначает сумму квадратов, соответствующую зоне G_i .

Оптимальное значение для этого примера равно нулю при условии, что осуществляется разбиение на три группы. В общем случае следует рассматривать значение целевой функции в сочетании с желаемым числом групп.

Последовательный процесс кластеризации начинается с рассмотрения n районов, затем два наиболее сходных района объединяются в один кластер и число кластеров становится равным $n-1$. Процесс повторяется до тех пор, пока все n районов не попадут в один кластер, содержащий все районы.

Для представления результатов решения задачи регионального кластерного анализа будем использовать наиболее известный метод представления матрицы расстояний (разнородности) или сходства, основанный на идеи графического изображения результатов последовательной кластеризации, которое обычно называют «дендограммой».

Дендограмму следует показывать как графическое изображение результатов процесса последовательной кластеризации, который осуществляется в терминах матрицы расстояний или сходства. В дальнейшем процесс такой кластеризации будет рассматриваться как процедуру с матрицей расстояний или сходства. Таким образом, с помощью дендограммы можно графически или геометрически изображать результат кластеризации при условии, что эта процедура оперирует только с элементами матрицы

расстояний или сходства.

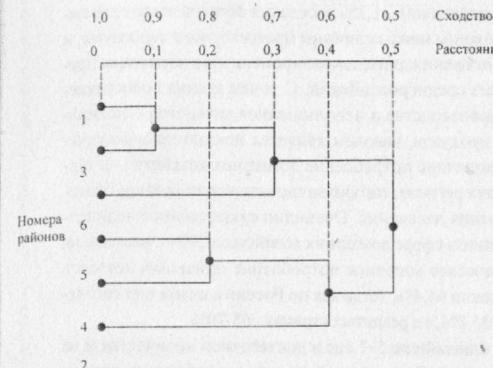


Рисунок. Дендограмма районирования в изучаемом регионе

Результаты регионального кластерного анализа.

Численное решение выделенных задач регионального кластерного анализа по Республике Беларусь и Гомельской области выполнялось с использованием пакета SPSS Version 10 (анализ статистических данных и восстановление скрытых закономерностей) [1, с. 384-417].

Структура полученных результатов по каждой экономической зоне характеризуется следующей графической и количественной информацией:

- графическая информация, отражающая динамику объединения экономических районов по мере увеличения диапазона изменения изучаемого параметра (евклидово расстояние), приведена на рисунке;

- количественная информация представляет собой:

- условия прекращения решения задачи;
- матрица близости экономических параметров по исследуемым регионам;
- последовательность выделения кластеров (районов);
- возможные варианты объединения кластеров (районов) по близости экономических показателей.

Графическая информация по результатам решения каждой из исследуемых задач по региональному анализу представлена графиком взаимосвязей изучаемых районов по близости социально-экономических показателей (дендограмма минимального дерева). При этом получаемый график развернут по оси абсцисс, показывающей направление снижения сходства технико-экономических показателей по изучаемым районам. На оси ординат показываются номера изучаемых регионов (областей или административно-экономических районов). Обратим внимание на то, что в решаемых задачах расположение номеров районов по оси абсцисс не является постоянным, что связано с особенностями алгоритма реализации задачи кластерного анализа. Представляет практический интерес и тот факт, что расположение номеров изучаемых регионов на

оси ординат также соответствует направлению снижения схожести социально-экономических показателей, что может быть использовано при построении вариантов прогноза по экономическому зонированию.

Экономическое зонирование в Республике Беларусь по выделенным локальным социально-экономическим задачам

выполнено также с использованием пакета SPSS Version 10 [1, с. 385-409] регионального кластерного анализа. Для регионального исследования по ранжированию областей Республики Беларусь использованы следующие критерии схожести социально-экономических показателей: рентабельность активов предприятий и организаций; численность населения, занятого в экономике; численность населения - всего; экономически активное население [в том числе: занятое население, безработное; уровень экономической активности населения (в процентах к численности населения в трудоспособном возрасте); уровень безработицы (в процентах к численности экономически активного населения)]; распределение численности работников по уровню образования; соотношение на-численной номинальной среднемесячной заработной платы работников и минимального потребительского бюджета; реальные денежные доходы населения; структура затрат на производство продукции, работ, услуг предприятий и организаций; структура чистой прибыли предприятий и организаций; структура формирования прибыли предприятий и организаций; ожидаемая продолжительность жизни при рождении [2, 3, 6, 7, 8]. На основании рассмотрения данных, приведенных в таблице 1, определяем номера позиций решаемых задач как ранги и соответствующие приоритеты группировки районов по сходству социально-экономических показателей. Например, получилось выделение следующих приоритетов: 1 - Гомельская область; 2 - Брестская область; 3 - Витебская область и т. д.

Количественная информация, характеризующая по каждой задаче динамику снижения схожести социально-экономических показателей по исследуемым регионам с использованием кластерного анализа, определена нами также с помощью пакета SPSS Version 10 [1, с. 385-409].

Экономическое зонирование в Гомельской области по выделенным локальным социально-экономическим задачам

выполнено также с использованием пакета SPSS Version 10 [1, с. 385-409] регионального кластерного анализа.

Для регионального исследования по ранжированию районов и городов Гомельской области использованы следующие основные элементы по критериям схожести социально-экономических показателей: рентабельность реализованной продукции, товаров, работ, услуг, в том числе: затраты на производство и реализацию, млн. рублей; прибыль (убыток) от реализации, млн. рублей; уровень рентабельности, в %; территория, кв. км; все население, тыс. человек (в том числе: городское, тыс. человек; сельское, в %); плотность населения, чел. / кв. км; уровень радиоактивного загрязнения

Таблица 1
Приоритеты группировки регионов по результатам кластерного анализа в Республике Беларусь за 2001 г.

№ п/п	Название района	Номера задач и присвоенные им ранги															Сумма рангов	Приори- теты группиро- вки регионов		
		1.1	1.2A	1.2B	1.2B	1.3	1.4	1.5A	1.5B	1.6	1.7	1.8	1.9	1.10	1.11	1.12A	1.12B			
1	Гомельская область	3	1	1	6	1	1	3	4	5	4	3	1	7	3	1	3	4	51	1
2	Брестская область	1	3	4	3	3	3	1	1	4	1	6	4	1	5	6	6	6	58	2
3	Витебская область	2	4	3	4	7	6	4	5	1	2	1	5	6	6	3	1	1	61	3
4	Гродненская область	4	5	5	1	5	4	6	6	2	3	4	2	2	4	4	4	5	66	4
5	Минская область	6	2	2	5	2	2	2	3	6	6	5	7	5	2	5	5	3	68	5
6	Могилевская область	7	6	6	2	6	5	5	2	3	5	2	3	3	7	2	2	2	68	5
7	г. Минск	5	7	7	7	4	7	7	7	7	7	7	6	4	1	7	7	7	104	6
	Сумма рангов	28	28	28	28	28	28	28	28	28	28	28	28	28	28	28	28	28	476	

ния (плотность загрязнения Km^2 по населенным пунктам); среднесписочная численность рабочих и служащих, тыс. человек; среднесписочная численность колхозников, тыс. человек; среднемесячная зарплата работников, тыс. рублей; прибыль (убыток) от реализации товаров, работ, услуг на одного работника, тыс. рублей; коэффициенты смертности; обеспеченность медицинским персоналом на 1000 населения; интегральный уровень образования, в том числе городского населения и сельского населения; ввод в действие жилых домов общей площади, m^2 / человек; общая жилая площадь, m^2 / жителя (в том числе: в городских поселениях, в сельской местности и всего); платные услуги населению (в сопоставимых ценах, в % к 1990 г.); объем платных услуг, тыс. руб. / чел.; зарплата, тыс. руб. / месяц;

задолженность на одного жителя, тыс. руб. / чел. (в том числе: дебиторская, кредиторская и сальдо) [2, 3, 6, 7, 8].

Полученные в ходе исследования результаты по ранжированию районов (городов) в Гомельской области приведены в таблице 2. По аналогии с предыдущей задачей экономического зонирования в Республике Беларусь определяем группировки регионов Гомельской области. Например, Брагинский район имеет первый номер приоритета, Наровлянский район - второй номер, Петриковский район - третий номер и т. д. Результаты решения данной задачи с использованием кластерного анализа представляют интерес для принятия управленческих решений на уровне Гомельского облисполкома.

Таблица 2

Распределение районов и городов Гомельской области по приоритетам кластерного анализа за 2001 г.

№ п/п	Название района (района)	Номера задач и присвоенные им ранги									Сумма рангов	Приоритеты группировки регионов
		2.1-2.3	2.4	2.5	2.6	2.7	2.8	2.9A	2.9B	2.10		
1	Брагинский	1	8	5	7	18	6	9	5	14	73	1
2	Наровлянский	3	3	2	3	19	8	12	14	17	81	2
3	Петриковский	9	10	9	4	17	21	1	10	6	87	3
4	Речицкий	18	13	16	1	10	15	2	2	12	89	4
5	Любовский	4	2	4	11	21	7	18	9	15	91	5
6	Кормянский	5	1	1	21	23	25	10	4	3	93	6
7	Ельский	2	5	8	13	20	22	15	8	5	98	7
8	Чечерский	7	17	6	2	12	17	14	13	10	98	7
9	Лельчицкий	11	20	18	5	6	23	7	1	8	99	8
10	Жлобинский	16	7	12	8	25	10	5	17	1	101	9
11	Светлогорский	13	12	10	6	14	19	16	12	2	104	10
12	Хойникский	10	9	7	17	11	12	6	15	20	107	11

Окончание таблицы 2

№ п/п	Название района (района)	Номера задач и присвоенные им ранги									Сумма рангов	Приоритеты группировки регионов
		2.1-2.3	2.4	2.5	2.6	2.7	2.8	2.9A	2.9B	2.10		
13	Добрушский	15	19	15	14	13	9	3	6	16	110	12
14	Октябрьский	6	16	11	16	2	24	13	7	18	113	13
15	Житковичский	12	6	17	15	24	18	4	11	7	114	14
16	Ветковский	8	18	3	10	22	16	17	3	19	116	15
17	Буда-Кошелевский	17	11	14	12	16	11	8	16	23	128	16
18	Мозырский	14	22	13	9	26	14	11	18	4	131	17
19	г. Речица	21	4	21	25	3	1	24	23	24	146	18
20	Рогачевский	20	14	20	19	4	13	23	26	9	148	19
21	Калинковичский	22	21	19	18	1	20	22	25	13	161	20
22	г. Светлогорск	23	23	22	23	8	2	20	20	21	162	21
23	Гомельский	19	15	23	20	15	26	21	24	11	174	22
24	г. Мозырь	25	26	25	24	5	5	19	22	25	176	23
25	г. Жлобин	24	24	22	7	4	25	21	26	177	24	
	Сумма рангов	351	351	351	351	351	351	351	351	351	3159	

Интегрированное экономическое зонирование в

Республике Беларусь. В задаче по интегрированному эко-

номическому зонированию, объединяющей все исследуемые факторные признаки по локальным задачам (1.1-1.12в) по Республике Беларусь, рассматриваются возможные варианты объединения районов (городов) республики в более крупные территориальные образования с целью углубления социально-экономического анализа. Сформулированная задача решена также с использованием пакета SPSS Version 10 [1, с. 385-409]. Рассмотрение промежуточных результатов вычислений в данном случае можно опустить и сосредоточить все внимание на таблице 3 (Cluster Membership) кластерного анализа.

Таблица 3

Варианты интегрированного зонирования в Республике Беларусь по факторам развития человеческого потенциала (Cluster Membership)

№ п/п	Название района	Интегрированные экономические зоны			Приоритеты группировки регионов
		4 зоны	3 зоны	2 зоны	
1	Брестская область	1	1	1	
2	Витебская область	2	1	1	
3	Гомельская область	1	1	1	
4	Гродненская область	3	2	2	
5	г. Минск	4	3	1	
6	Минская область	1	1	1	
7	Могилевская область	3	2	2	

экономической зоне - Гродненскую и Могилевскую областей;

- при выделении трех интегрированных зон следует отнести к первой экономической зоне: Брестскую, Витебскую, Гомельскую и Минскую области; ко второй зоне - Гродненскую и Могилевскую области; к третьей зоне - г. Минск;

- при выделении четырех интегрированных зон следует отнести к первой экономической зоне: Брестскую, Гомельскую и Минскую области; ко второй зоне - Витебскую; к третьей зоне - Гродненскую и Могилевскую области; к четвертой зоне - г. Минск.

Интегрированное экономическое зонирование в Гомельской области. В задаче по Гомельской области рассматриваются возможные варианты объединения регионов исследуемой области в более крупные территориальные образования также с целью более глубокого социально-экономического анализа для повышения обоснованности принимаемых региональных управленческих решений. Решаемая задача объединяет всю входную информацию, использованную рассмотренных выше девяти локальных социально-экономических задачах. Задача интегрированного экономического зонирования также решена с использованием пакета SPSS Version 10 [1, с. 385-409].

Полученный количественный результат решения задачи по интегрированному зонированию приведен в таблице 4, в которой отражены три наиболее предпочтительных варианта объединения кластеров по условию сходства социально-экономических показателей. Например, при выделении в Гомельской области четырех регионов (4 Clusters), с точки зрения интересов решаемой задачи, целесообразно следующее объединение административных образований в каждом из них: первый регион - районы Брагинский, Буда-Кошелевский, Ветковский, Гомельский, Добрушский, Ельский, Житковичский, Жлобинский, Ка-

Окончание таблицы 4

№ п/п	Районы (регионы) области	Интегрированные экономические зоны		
		4 зоны	3 зоны	2 зоны
21	Чечерский	1	1	1
22	Гомельский горсовет	2	2	2
23	г. Жлобин	3	3	1
24	г. Мозырь	3	3	1
25	г. Речица	1	1	1
26	г. Светлогорск	4	1	1

Молинковичский, Кормянский, Лельчицкий, Лоевский, Мозырский, Наровлянский, Октябрьский, Петриковский, Речицкий, Рогачевский, Светлогорский, Хойникский, Чечерский и г. Речица; второй регион - Гомельский горсовет; третий регион - гг. Мозырь и Жлобин; четвертый регион - г. Светлогорск.

Полученные результаты вычислений показывают возможность выделения следующих приоритетных вариантов интегрированных экономических зон в Гомельской области:

- при выделении двух интегрированных зон к первой экономической зоне следует отнести все административные образования Гомельской области за исключением Гомельского горсовета;

- при выделении трех интегрированных зон к первой экономической зоне следует отнести все административные образования области за исключением Гомельского горсовета, гг. Жлобина и Мозыря; ко второй экономической зоне - Гомельский горсовет; к третьей экономической зоне - гг. Жлобин и Мозырь;

- при выделении четырех интегрированных зон к первой экономической зоне следует отнести все административные образования за исключением Гомельского горсовета, гг. Жлобина, Мозыря и Светлогорска; ко второй экономической зоне - Гомельский горсовет; к третьей экономической зоне - гг. Жлобин и Мозырь; к четвертой экономической зоне - г. Светлогорск.

Таблица 4

Варианты интегрированного зонирования в Гомельской области по факторам развития человеческого потенциала (Cluster Membership)

№ п/п	Районы (регионы) области	Интегрированные экономические зоны		
		4 зоны	3 зоны	2 зоны
1	Брагинский	1	1	1
2	Буда-Кошелевский	1	1	1
3	Ветковский	1	1	1
4	Гомельский	1	1	1
5	Добрушский	1	1	1
6	Ельский	1	1	1
7	Житковичский	1	1	1
8	Жлобинский	1	1	1
9	Калинковичский	1	1	1
10	Кормянский	1	1	1
11	Лельчицкий	1	1	1
12	Лоевский	1	1	1
13	Мозырский	1	1	1
14	Наровлянский	1	1	1
15	Октябрьский	1	1	1
16	Петриковский	1	1	1
17	Речицкий	1	1	1
18	Рогачевский	1	1	1
19	Светлогорский	1	1	1
20	Хойникский	1	1	1

Выходы. 1. Показана возможность использования кластерного анализа для решения задач по экономическому зонированию. В качестве изучаемых регионов могут быть приняты Республика Беларусь в целом или более мелкие территориальные образования (области, административно-территориальные районы и т. д.).

2. Выделены приоритеты объединения регионов в более крупные экономические зоны по критерию близости учитываемых факторов человеческого потенциала на районном (городском) уровне.

3. Получены конкретные рациональные варианты выделения двух, трех и четырех экономических зон как по Республике Беларусь, так и по Гомельской области.

Литература

1. **Бюоль Ахим, Цефель Петер.** SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей: Пер. с нем. / Ахим Бюоль, Петер Цефель. - СПб.: ООО «ДиаСофтЮП», 2001.

2. Денежные доходы и расходы населения Республики Беларусь. - Мин.: Министерство статистики и анализа Республики Беларусь, 2002.

3. Доклад о развитии человека за 2002 год. Издано по заказу Программы развития Организации Объединенных Наций (ПРООН) - Нью-Йорк: Оксфорд, Оксфордуниверсити пресс, 2002.

4. **Дюран Б., Оделл П.** Кластерный анализ. Пер. с англ. Е. З. Демиденко. Под ред. А. Я. Боярского. Предисловие А. Я. Боярского. - М.: «Статистика», 1977.

5. **Кендалл М., Стьюарт А.** Многомерный статистический анализ и временные ряды. - М.: Наука, 1976.

6. Труд и занятость в Республике Беларусь. - Мин.: Министерство статистики и анализа Республики Беларусь, 2002.

7. Финансовые результаты предприятий и организаций Гомельской области за январь-декабрь 2001 года. - Гомель: Гомельское областное управление статистики, 2002.

8. Финансы Республики Беларусь: Статистический сборник. - Мин.: Министерство статистики и анализа Республики Беларусь, 2002.

9. **Юл Дж. Э., Кендалл М. Дж.** Теория статистики. Четырнадцатое издание, пересмотренное и расширенное. Перевод с английского под редакцией Ф.Д. Лившица. - М.: Госстатиздат ЦСУ СССР, 1960.

10. **Ястребинский Б. С.** Связь между элементами крестьянского хозяйства в 1917 и 1919 годах // Вестник статистики, 1920. С. 48-69.