

*КЛАСТЕРНОЕ ЗОНИРОВАНИЕ ПО
ФАКТОРАМ РАЗВИТИЯ ЧЕЛОВЕЧЕСКОГО
ПОТЕНЦИАЛА В БЕЛАРУСИ*

Постановка задачи. Территориальная дифференциация параметров, характеризующих количественную и качественную экономическую природу изучаемых объектов, явлений и т.д. может быть выявлена на основе более глубокого исследования 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 многомерного анализа. Наиболее существенные методологические черты кластерного анализа, полезные для регионального исследования, могут быть отражены по двум направлениям: образование единой меры, охватывающей ряд характеризующих признаков, и 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. Б.С. Ястремский рассматривал 34 характеристики уездов, влиявшие на эту связь [1, 48–69]. Можно привести и другие примеры группировки территориальных единиц по комплексу признаков, неизменно имевшихся в задачах районирования. Но в кластерном анализе признаки объединяются с помощью некоторой “метрики” в один количественный показатель сходства (различия) группируемых объектов [2–4].

В кластерном региональном анализе группировочные признаки могут объединяться с помощью различных функций расстояния. В данном исследовании это функция евклидова расстояния — самая распространенная мера 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 установления расстояния между двумя точками на плоскости, образованной точками двухмерного пространства x и y , используется формула евклидовой меры

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

где x_1, y_1 — координаты первой точки; x_2, y_2 — координаты второй точки.

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

ластей), принадлежащих некоторой искомой однородной зоне в изучаемом регионе P_i . Предположим также, что существует некоторое множество наблюдаемых показателей или характеристик $C = (C_1, C_2, \dots, C_p)$. Наблюдаемые характеристики могут быть количественными и качественными. Однако основная часть нашего исследования посвящена количественным данным. Для множества объектов I мы располагаем множеством векторов измерений $X = \{X_1, X_2, \dots, X_n\}$, которые описывают множество I . Множество X можно представить 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 кластеров (однородных экономических зон) P_1, P_2, \dots, P_m так, чтобы каждый объект I_i принадлежал одному и только одному подмножеству разбиения, объекты же, принадлежащие одному и тому же кластеру, были сходными, в то время как объекты, относящиеся к разным кластерам, — разнородными (несходными).

Приведем постановку региональной задачи кластерного анализа. В качестве региона I рассмотрим n -образующие 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. [5–9]. Тогда X_1 (вектор измерений) представляет собой набор указанных характеристик для первого, X_2 — для второго административно-территориального районов 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. Он может представлять собой некоторый функционал, выражающий уровни желательности различных разбиений и группировок. Его можно назвать целевой функцией. Например, в качестве последней можно взять внутригрупповую сумму квадратов отклонений. В качестве примера рассмотрим 8 объектов (n) с одной характеристикой (т. е. $c = 1$); 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:

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

W_i обозначает сумму квадратов, соответствующую зоне G_i . Оптимальное значение для этого примера равно нулю при условии, что ведется разбиение на три группы. В общем случае следует рассматривать значение целевой функции в сочетании с желаемым числом групп.

Последовательный процесс кластеризации начинается с анализа n районов; затем два наименее сходных района объединяются в один кластер и их число 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 используем наиболее известный метод представления матрицы расстояний (разнородности) или сходства, основанный на идее графического изображения результатов последовательной кластеризации, которое обычно называют дендограммой.

Ее следует показывать как графическое изображение результатов процесса последовательной кластеризации, который осуществляется в терминах матрицы расстояний или сходства. В дальнейшем процесс такой кластеризации будем рас-

смагивать 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.

Результаты регионального кластерного анализа. Задачи регионального кластерного анализа по Беларуси (Гомельской области) решались на компьютере с использованием пакета SPSS Version 10 (анализ статистических данных и восстановление скрытых закономерностей) [10, 384–417].

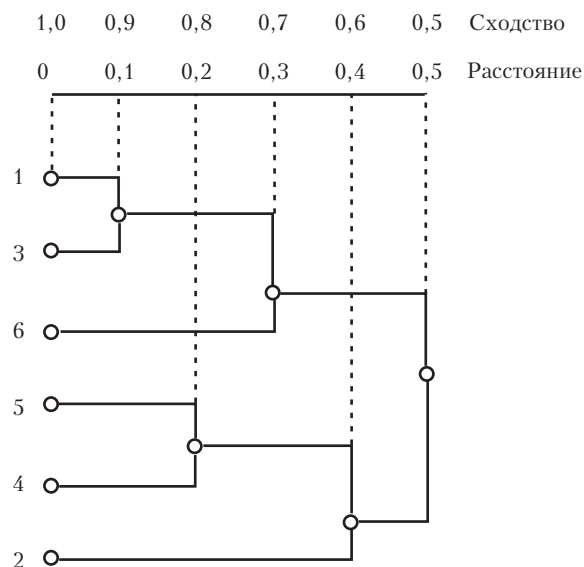
Графическая информация, отражающая динамику объединения экономических районов по мере увеличения диапазона изменения изучаемого параметра (евклидово расстояние), приведена на рисунке.

Количественная информация представлена в табл. 1–3.

Графическая информация по результатам решения каждой из исследуемых задач по региональному анализу представлена графой взаимосвязей изучаемых районов по близости социально-экономических показателей (дендограмма минимального дерева). При этом графа развернута по оси абсцисс, показывающей направления снижения сходства технико-экономических показателей по районам. На оси ординат — номера регионов

(областей или административно-экономических районов). Обратим внимание на то, что в решаемых задачах расположение номеров районов по оси абсцисс не постоянное. Это 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. Это можно использовать при построении вариантов прогноза по экономическому зонированию.

Экономическое зонирование в Беларуси по выделенным локальным социально-экономическим задачам выполнено с использованием пакета SPSS Version 10 [10, 385–409] регионального кластерного анализа. Для такого исследования использованы следующие критерии схожести социально-экономических показателей: рентабельность активов предприятий и организаций по областям Беларуси; численность населения, занятого в экономике; распределение численности населения по областям Беларуси; экономически активное население (в том числе занятые, безработное население; уровень его экономической активности (в процентах к его численности в трудоспособном возрасте); уровень безработицы (в процентах к численности экономически активного населения)); распределение численности работников по уровню образования; соотношение начисленной номинальной среднемесячной заработной платы работников и минимального потребительского бюджета по областям Беларуси; реальные денежные доходы населения Беларуси; уровень безработицы; структура затрат на производство продукции, работ, услуг; структура чистой прибыли; структура формирования прибыли предприятий и организаций по областям; ожидаемая продолжительность жизни при рождении по областям в Беларуси [5–9].



Дендограмма районирования в изучаемом регионе:
1–6 — номера районов

На основании данных табл. 1 определяем номера позиций решаемых задач как ранги и соответствующие приоритеты группировки регионов по схожести социально-экономических показателей. Например, выделились следующие приоритеты: 1 — Гомельская; 2 — Брестская; 3 — Витебская области и т. д.

Количественная информация, характеризующая по каждой задаче динамику снижения схожести социально-экономических показателей по исследуемым регионам с применением кластерного анализа определена нами на основе использования прикладного статистического пакета [10, 385–409]. Результаты решения задачи кластерного анализа, например, “Рентабельность активов предприятий и организаций по регионам Беларуси”, представлены в табл. 1–4, их содержание необходимо пояснить начинающим пользователям данным статистическим пакетом.

Решение задачи приостанавливается в случае, когда по всем изучаемым регионам будет достигнуто полное отсутствие (100 %) различия в сходстве социально-экономических показателей.

Матрица близости получается в печати треугольной и носит зеркальный характер. Например, коэффициенты близости Гомельской и Гродненской областей равны 9,500 (табл. 2).

Для каждой кластерной комбинации (табл. 3) показаны коэффициенты сходства социально-экономических показателей и шаг вычислений, на котором они получены. При этом последний определяется точками объединения регионов. Как видно, на первом шаге вычислений районы первый (Брестская область) и второй (Витебская область) имеют коэффициент взаимной близости 0,770. В табл. 3 также показано, что данная комбинация кластеров появилась на нулевом (исходном) этапе вычислений.

Приводим сводные результаты решения задачи кластерного анализа. На наш взгляд, существует три варианта наиболее рационального объединения регионов (кластеров). Для практической реализации может быть принят один из них с учетом дополнительных условий целесообразности объединения регионов (управляемость, местные, географические особенности и т.д.). Как видно, при выделении в Республике Беларусь четырех регионов целесообразны следующие их объединения: первый — Брестская, Витебская, Гродненская и Минская области, второй — Гомельская область, третий — г. Минск, четвертый — Могилевская область.

Экономическое зонирование в Гомельской области по выделенным локальным социально-экономическим задачам выполнено также с помощью регионального кластерного анализа [10, 385–409]. При этом использованы следующие основные элементы по критерию схожести социально-экономических показателей: рентабельность реализованной продукции, товаров, работ, услуг по Гомельской области, в том числе: затраты на производство и реализацию, млн р.; прибыль (убыток) от реализации, млн р.; уровень рентабельности, %; распределение населения Гомельской области на 1.01.2002 г., в том числе: территория, км²; все население, тыс. чел. (в том числе: городское, тыс. чел.; городское, %; сельское, тыс. чел.); плотность населения, чел. / км²; уровни радиоактивного загрязнения регионов Гомельской области (плотность загрязнения Бк / км² по населенным пунктам (НП)); среднесписочная численность рабочих и служащих, тыс. чел.; колхозников; среднемесячная зарплата работников, тыс. р.; прибыль (убыток) от реализации товаров, работ, услуг на одного работника, тыс. р.; коэффициенты смертности; ввод в действие жилых домов; обеспеченность медицинским персоналом на 1000 чел. населения; региональный уровень образования населения Гомельской области в 2001 г., в том числе городского и сельского (на 1000 чел.); ввод в действие жилых домов общей площади, м² / чел.; общая жилая площадь, м² / 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. р. / чел.; зарплата, тыс. р. / месяц; задолженность жителя, тыс. р. / чел. (в том числе: дебиторская, кредиторская и сальдо) [5–9].

Результаты по ранжированию районов (городов) в Гомельской области приведены в табл. 4. По аналогии с предыдущей задачей экономического зонирования в Беларуси определяем группировки регионов Гомельской области. Результаты ре-

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

Область	Номера задач и присвоенные им ранги																	Сумма рангов	Приоритеты группировки регионов
	1.1	1.2А	1.2Б	1.2В	1.3	1.4	1.5А	1.5Б	1.6	1.7	1.8	1.9	1.1	1.11	1.12А	1.12Б	1.12В		
Гомельская	3	1	1	6	1	1	3	4	5	4	3	1	7	3	1	3	4	51	1
Брестская	1	3	4	3	3	3	1	1	4	1	6	4	1	5	6	6	6	58	2
Витебская	2	4	3	4	7	6	4	5	1	2	1	5	6	6	3	1	1	61	3
Гродненская	4	5	5	1	5	4	6	6	2	3	4	2	2	4	4	4	5	66	4
Минская	6	2	2	5	2	2	2	3	6	6	5	7	5	2	5	5	3	68	5
Могилевская	7	6	6	2	6	5	5	2	3	5	2	3	3	7	2	2	2	68	5
г. Минск	5	7	7	7	4	7	7	7	7	7	7	6	4	1	7	7	7	104	6
Сумма	28	28	28	28	28	28	28	28	28	28	28	28	28	28	28	28	28	476	—

Таблица 2. Матрица близости социально-экономических параметров областей Республики Беларусь по данным за 2001 г.

Исследование по областям	Евклидово расстояние							
	Брестская	—	0,770	11,140	1,820	14,950	1,720	5,750
Витебская	0,770	—	9,970	3,650	11,980	3,030	6,520	
Гомельская	11,140	9,970	—	9,500	4,690	7,700	29,790	
Гродненская	1,820	3,650	9,500	—	14,570	3,300	9,070	
г. Минск	14,950	11,980	4,690	14,570	—	14,770	34,100	
Минская	1,720	3,030	7,700	3,300	14,770	—	12,170	
Могилевская	5,750	6,520	29,790	9,070	34,100	12,170	—	

Таблица 3. Последовательность выделения кластеров (районов)

Стадия	Сочетание кластеров		Коэффициенты	Появление первой стадии кластера		Новая стадия
	Кластер 1	Кластер 2		Кластер 1	Кластер 2	
1	1	2	0,770	0	0	2
2	1	6	2,375	1	0	3
3	1	4	2,923	2	0	5
4	3	5	4,690	0	0	6
5	1	7	8,378	3	0	6
6	1	3	15,847	5	4	0

Таблица 4. Распределение районов и городов Гомельской области по приоритетам кластерного анализа за 2001 г.

Район	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	2.9А	2.9Б	2.10		
Брагинский	1	8	5	7	18	6	9	5	14	73	1
Наровлянский	3	3	2	3	19	8	12	14	17	81	2
Петриковский	9	10	9	4	17	21	1	10	6	87	3
Речицкий	18	13	16	1	10	15	2	2	12	89	4
Лоевский	4	2	4	11	21	7	18	9	15	91	5
Кормянский	5	1	1	21	23	25	10	4	3	93	6
Ельский	2	5	8	13	20	22	15	8	5	98	7
Чечерский	7	17	6	2	12	17	14	13	10	98	7
Лельчицкий	11	20	18	5	6	23	7	1	8	99	8
Жлобинский	16	7	12	8	25	10	5	17	1	101	9
Светлогорский	13	12	10	6	14	19	16	12	2	104	10
Хойникский	10	9	7	17	11	12	6	15	20	107	11
Добрушский	15	19	15	14	13	9	3	6	16	110	12
Октябрьский	6	16	11	16	2	24	13	7	18	113	13
Житковичский	12	6	17	15	24	18	4	11	7	114	14
Ветковский	8	18	3	10	22	16	17	3	19	116	15
Буда-Кошелевский	17	11	14	12	16	11	8	16	23	128	16
Мозырский	14	22	13	9	26	14	11	18	4	131	17
г. Речица	21	4	21	25	3	1	24	23	24	146	18
Рогачевский	20	14	20	19	4	13	23	26	9	148	19
Калинковичский	22	21	19	18	1	20	22	25	13	161	20
г. Светлогорск	23	23	22	23	8	2	20	20	21	162	21
Гомельский	19	15	23	20	15	26	21	24	11	174	22
г. Мозырь	25	26	25	24	5	5	19	22	25	176	23
г. Жлобин	24	24	24	22	7	4	25	21	26	177	24
Гомельский горсовет	26	25	26	26	9	3	26	19	22	182	25
Сумма рангов	351	351	351	351	351	351	351	351	351	3159	—

шения данной задачи с использованием кластерного анализа представляют интерес для принятия управленческих решений на уровне Гомельского облисполкома.

Интегральное экономическое зонирование в Беларуси. В задаче по интегральному экономическому зонированию, объединяющей исследуемые факторные признаки по семнадцати локальным задачам (1.1–1.12в), рассматриваются возможные варианты объединения районов (городов) страны в более крупные территориальные образования в целях углубления социально-экономического анализа [10, 385–409].

В результате вычислений получены следующие приоритеты объединения областей на интегрированные экономические зоны:

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

Интегральное экономическое зонирование в Гомельской области. В задаче по Гомельской области рассматриваются возможные варианты объединения регионов в более крупные территориальные образования также в целях более глубокого социально-экономического анализа для повышения обоснованности принимаемых региональных управленческих решений. Она объединяет всю входную информацию, использованную в девяти локальных социально-экономических задачах [10, 385–409].

Количественный результат решения задачи по интегральному зонированию содержит три наиболее предпочтительных варианта объединения кластеров по условию сходства социально-экономических показателей. Например, при выделении в Гомельской области четырех регионов, с точки зрения интересов решаемой задачи, целесообразно объединение административных образований в регионы: первый — районы Брагинский, Буда-Кошелевский, Ветковский, Гомельский, Добрушский, Ельский, Житковичский, Жлобинский, Калинковичский, Кормянский, Лельчицкий, Лоевский, Мозырский, Наровлянский, Октябрьский, Петриковский, Речицкий, Рогачевский, Светлогорский, Хойникский, Чечерский и город Речица; второй — Гомельский горсовет; третий — города Мозырь и Жлобин; четвертый — г. Светлогорск.

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

Выводы. 1. Показана возможность использования кластерного анализа для решения задач по экономическому зонированию. В качестве изучаемых регионов могут быть приняты Беларусь 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 экономических показателей на районном (городском) уровне.

3. Получены конкретные рациональные варианты выделения в Гомельской области двух, трех и четырех экономических зон.

Литература

1. *Ястремский Б.С.* Связь между элементами крестьянского хозяйства в 1917 и 1919 годах // Вестн. статистики. 1920. № 9—12 (сент. — дек.).
2. *Дюран Б., Оддел П.* Кластерный анализ / Пер. с англ. Е.З. Демиденко. М., 1977.
3. *Кендалл М., Стьюарт А.* Многомерный статистический анализ и временные ряды. М., 1976.
4. *Юл Дж., Кендэл М.* Теория статистики. 14-е изд., доп. / Пер. с англ.; Под ред. Ф.Д. Лившица. М., 1960.
5. Денежные доходы и расходы населения Республики Беларусь. Мн., 2002.
6. Доклад о развитии человека за 2002 год. Нью-Йорк, 2002.
7. Труд и занятость в Республике Беларусь. Мн., 2002.
8. Финансовые результаты предприятий и организаций Гомельской области за январь — декабрь 2001 года. Гомель, 2002.
9. Финансы Республики Беларусь: Стат. сб. Мн., 2002.
10. *Бююль А., Цефель П.* SPSS: искусство обработки информации: Анализ статистических данных и восстановление скрытых закономерностей / Пер. с нем. СПб., 2001.