

УДК 332.1: 311: 327.8      DOI: 10.14451/1.256.368

# Многомерная группировка постсоветских стран инструментами среды программирования R

© 2026 Цыпин Александр Павлович

Кандидат экономических наук, доцент, доцент кафедры бизнес-аналитики, Факультет налогов, аудита и бизнес-анализа. Финансовый университет при Правительстве РФ, Москва, Россия.

E-mail: aptsypin@fa.ru

**Ключевые слова:** постсоветские страны, структура экономики, группировка, кластеризация, кластерный анализ.

В статье рассматриваются вопросы выбора методики кластеризации экономических систем постсоветских стран. При этом используются три конкурирующих метода – k-средних, k-методы, k-медиан. В результате формируется вывод об оптимальном методе разбиения 15 республик на группы со схожими структурами экономик.

## Введение

Группировка данных является одной из основных задач статистической науки. Наиболее простым способом считается одномерная группировка, когда объекты (домохозяйства, предприятия, регионы, страны) распределяются на отдельные группы на основе одного показателя (признака). Существенным недостатком данного подхода является игнорирование многомерности признаков пространства. Данная проблема решается при рассмотрении методов многомерной группировки. Одним из таких подходов является кластерный анализ, который позволяет формировать группы объектов с учетом веса множества признаков.

Обращение к российскому сектору научной литературы позволило нам выявить ряд публикаций, посвященных применению методов мно-

гомерной группировки к совокупности субъектов Российской Федерации. В частности, можно указать на работы таких авторов, как Айвазян С. А. [1], Боженко С. В. [2], Гичиев Н. С. [3], Клименко А. В. [4], Шаталова О. М. [6]. Тем не менее исследований, посвященных кластеризации постсоветских стран, а тем более сопоставления разных алгоритмов разбиения не было выявлено. Поэтому в рамках проводимого исследования обратимся к данным по 15 странам – экс-членам СССР и сопоставим результаты работы методов секционирования и кластеризации.

## Материалы и методы

В практике статистических исследований используется несколько подходов к кластеризации объектов. Согласно публикациям [7–9], можно выделить четыре наиболее часто используемых подходов (табл. 1).

**Таблица 1.** Методы и алгоритмы кластеризации.

Наименование	Сущность	Алгоритмы
Иерархические алгоритмы	Предназначены для формирования иерархии между группируемыми объектами	Агломеративная кластеризация – от частного (объекты) к общему (группы). Девизивная кластеризация – от общего (одной группы) к частному (множеству кластеров)
Методы секционирования и кластеризации	Алгоритмы разделяют объекты на основе нескольких центроидов (средних, метойдов, медиан) в наборе данных.	k-средних, k-метойды, k-медиан
Методы кластеризации на основе распределения	Предполагается, что данные получены из смеси распределений вероятностей, и оценивает параметры этих распределений для выявления кластеров.	Метод максимального правдоподобия (EM), модели гауссовских смесей (GMM)
Кластеризация на основе плотности	Кластеризация на основе плотности работает путём обнаружения областей, в которых сосредоточены факторы и в которых они разделены с помощью областей, которые могут быть пустыми или разреженными. Точки, которые не являются частью кластера, классифицируются как шум. Выбросы не относятся к кластерам и поэтому игнорируются в этих алгоритмах.	DBSCAN (пространственная кластеризация приложений на основе плотности с учётом шума). OPTICS (упорядочивание точек для определения структуры кластеризации).

Источник: составлено на основе данных публикаций в информационно-телекоммуникационной сети Интернет [7–9]

Из приведенных в таблице 1 подходов в проводимом исследовании были рассмотрены методы секционирования и кластеризации, при этом во всех трех алгоритмах кластеризации было применено манхэттенское расстояние.

В качестве объекта изучения были выбраны постсоветские страны, то есть 15 республик, входивших в состав СССР до 1991 года. Предметом изучения является оценка схожести экономических структур постсоветских стран.

Источником информации является Статистический комитет ООН. В частности, были взяты удельные веса видов экономической деятельности в ВВП (%), за 2023 год по 15 постсоветским странам: A\_B – доля сельского хозяйства; C\_E – доля промышленности; F – доля строительства; G\_H – доля торговли, гостиничного и ресторанного бизнеса; I – доля отрасли транспорта

и связи; J\_P – доля сферы услуг (финансы, образование, здравоохранение и т. д.). При этом стоит отметить, что обозначение отраслей (видов экономической деятельности) осуществлялось на основе ОКВЭД1.

Сокращенные названия постсоветских стран, приводимые в исследовании, были взяты из международного стандарта ISO 3166-1.

#### Результаты исследования

Прежде чем приступить к непосредственной группировке методами секционирования и кластеризации, обратимся к иерархическому методу и построим дендограмму [5]. Это позволит определиться с количеством групп, на которые необходимо разбивать изучаемую совокупность постсоветских стран. Результаты кластеризации приведены на рисунке 1.

Опираясь на приведенный рисунок 1, можно сделать вывод о выделении трех групп объектов. Так, в первую группу входят страны: Туркменистан, Таджикистан, Узбекистан, Азербайджан. Во вторую группу – Беларусь, Литва, Казахстан



**Рис. 1.** Круговая дендограмма постсоветских стран (Примечание: для построения диаграммы были использованы библиотеки `cluster` и `ape` среды программирования R).

и Российская Федерация. В третью группу – Латвия, Эстония, Грузия, Кыргызстан, Молдавия, Армения и Украина.

Проведя «разведочный» анализ на основе иерархического метода и получив информацию о числе групп, перейдем к классификации методом  $k$ -средних. При этом число кластеров  $k$  задается пользователем (в нашем случае  $k = 3$ ).

Суть процедуры заключается в следующем: на первом шаге определяют  $k$  кластеров-эталонов. Далее каждый объект присоединяется к ближайшему эталону. В качестве критерия используется минимальное расстояние внутри кластера относительно среднего.

Как только объект включается в кластер, среднее пересчитывается. После пересчета эталона объекты снова распределяются по ближайшим кластерам и т. д. Процедура заканчивается при стабилизации процесса, то есть при стабилизации центров тяжести.

В результате реализации данной процедуры с применением функции `kmeans {amar}` среды программирования R, получаем диаграмму с выделенными кластерами (рис. 2).

Согласно информации, приведенной на рисунке 2, можно сделать вывод о формировании процедурой  $k$ -средних трех кластеров. Первый

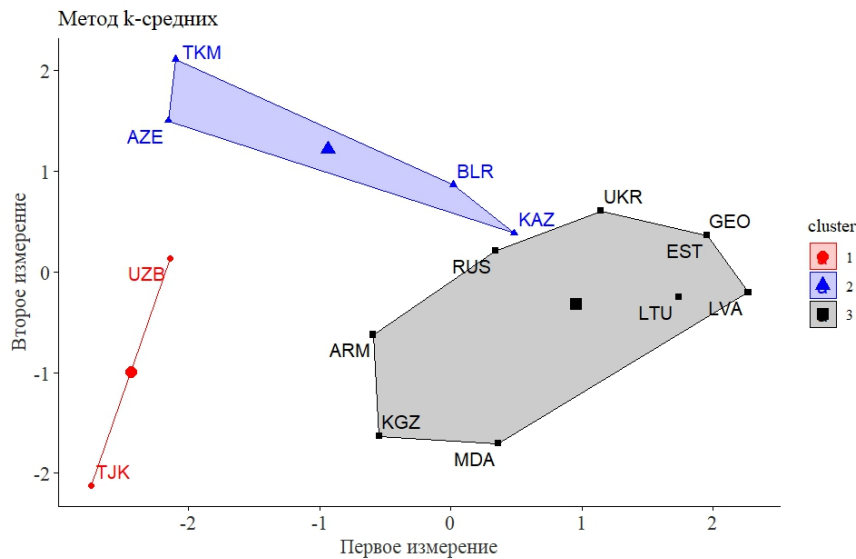
содержит 2 объекта, второй 4 и третий 9 постсоветских стран.

Использование средних значений подразумевает, что кластеризация с использованием  $k$ -средних значений очень чувствительна к выбросам. Это может серьезно повлиять на распределение наблюдений по кластерам. Более надежный алгоритм обеспечивается алгоритмом PAM – (Partitioning Around Medoids – Разделение вокруг Медоидов), который относится к группе методов  $k$ -метойды.

Стоит отметить, что медоид (medoid) – это объект, принадлежащий набору данных или кластеру, различие которого с другими объектами в наборе данных или кластере минимально. Медоиды близки по смыслу центроидам, но в отличие от них, являются объектом, принадлежащим кластеру, и как правило используются в тех случаях, когда невозможно вычислить средние координаты или центр масс кластера.

Алгоритм PAM подразумевает выполнение следующей последовательности действий:

1. Выбираем  $k$  объектов, чтобы они стали медоидами;
2. Вычисляем матрицу различий;
3. Назначаем каждому объекту ближайший медоид;



**Рис. 2.** Диаграмма кластеров постсоветских стран методом k-средних (Примечание: для построения диаграммы была использована функция `factoextra{fviz_cluster}` среды программирования R).

4. Для каждого кластера оцениваем – не уменьшает ли какой-либо объект кластера средний коэффициент непохожести; если это так, выбираем объект, который уменьшает этот коэффициент больше всего, в качестве медоида для этого кластера;
5. Если хотя бы один медоид изменился, переходим к пункту № 3, в противном случае завершаем алгоритм.

Для осуществления РАМ алгоритма используем библиотеку `ram{cluster}` [10] среды программирования R. Результаты разбиения совокупности на 3 группы приведены на рисунке 3.

Согласно данным, приведенным на рисунке 3, совокупность 15 постсоветских стран была разбита на 3 группы, но их состав в значительной степени отличается от результатов полученных на основе k-средних.

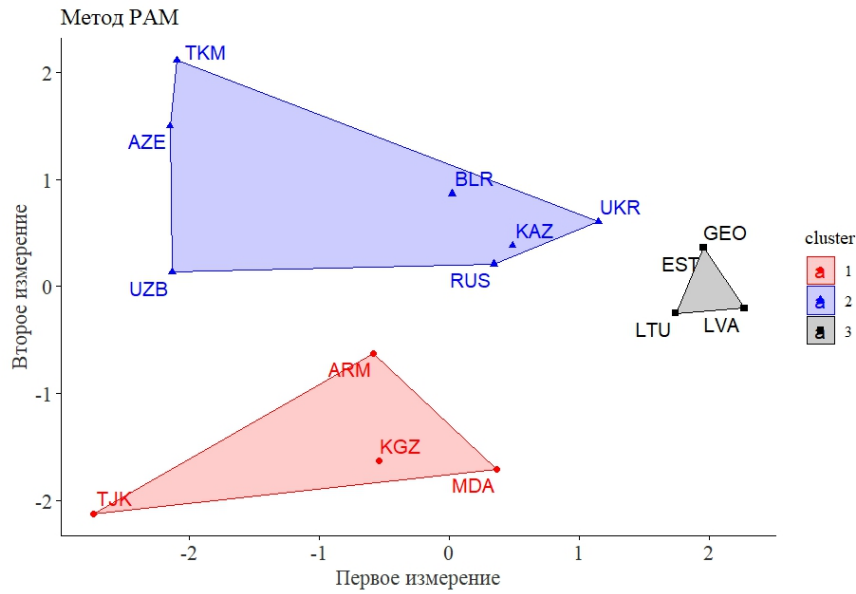
Рассмотрим еще один подход к осуществлению кластерного анализа, относящийся к группе методов секционирования и кластеризации, это k-медиан. Его особенностью является невосприимчивость к выбросам, которые часто наблюдаются в совокупности макроэкономических объектов. Для реализации алгоритма обратимся

к функции `kcca{flexclust}`, в которой укажем аргумент `kmedians`. Результаты кластеризации приведем на рисунке 4.

Приведенные на рисунке результаты кластеризации методом k-медиан идентичны результатам, полученным в ходе применения метода k-средних. Итоговый результат группировки тремя методами представлен в таблице 2.

Как видим из приведенной таблицы 2, алгоритмы k-средних и k-медиан выдают схожие результаты, то есть 13 субъектов из 15 вошли в аналогичные группы. Тогда как результаты группировки на основе k-метойдов в значительной степени отличаются от двух конкурирующих. При этом стоит отметить, что данные результаты (k-метойдов) легко интерпретируются. Так, страны, вошедшие в 1 группу, обладают слабой экономикой и как следствие ее структура далека от среднемировой. Тогда как объекты, вошедшие в 3 группу это экономики, приближенные по структуре к среднеевропейским.

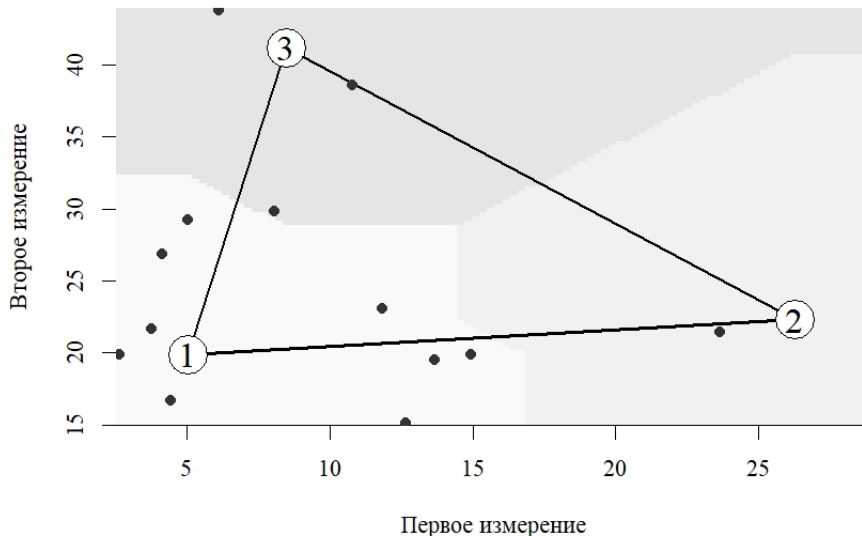
Используем результаты кластеризации, полученные на основе подхода k-метойдов и рассчитаем внутригрупповые средние по всем 6 показателям, участвующим в анализе (табл. 3).



**Рис. 3.** Диаграмма кластеров постсоветских стран методом  $k$ -метойды (Примечание: для построения диаграммы была использована функция  `pam{cluster}` среды программирования R).

**Таблица 2.** Результаты многомерной группировки постсоветских стран в шестимерном признаковом пространстве.

Страны	Номер кластера		
	$k$ -средних	$k$ -метойдов	$k$ -медиан
ARM	3	1	3
AZE	2	2	2
BLR	2	2	3
EST	3	3	3
GEO	3	3	3
KAZ	2	2	3
KGZ	3	1	3
LVA	3	3	3
LTU	3	3	3
MDA	3	1	3
RUS	3	2	3
TKM	2	2	2
TJK	1	1	1
UKR	3	2	3
UZB	1	2	1



**Рис. 4.** Диаграмма кластеров постсоветских стран методом k-медиан (Примечание: для построения диаграммы были использованы функции `image` и `points` из стандартной библиотеки `graphics` среды программирования R).

**Таблица 3.** Внутригрупповые средние для трех кластеров постсоветских стран, полученных на основе k-метойдов.

№	Состав кластера	A_B	C_E	F	G_H	I	J_P
1	ARM, KGZ, MDA, TJK	16,23	19,01	9,97	16,87	8,56	28,82
2	AZE, BLR, KAZ, RUS, TKM, UKR, UZB	10,69	30,65	6,54	12,90	9,66	29,02
3	EST, GEO, LVA, LTU	3,37	19,54	6,57	16,03	14,25	40,14

Приведенные в таблице 3 внутригрупповые средние показывают, что в третьей группе доминируют такие отрасли, как «транспорт и связь» и сфера услуг. Во второй группе высокая доля промышленности (C\_E). Во второй группе наблюдаются максимальные значения по удельному весу сельского хозяйства, строительства и сферы гостеприимства. Таким образом, с помощью метода k-метойдов получены три группы с ярко выраженной отраслевой структурой экономики.

#### Выводы

Проведенное исследование, направленное на сопоставление трех методов секционирования и кластеризации, позволяет сделать ряд выводов:

- два (k-средних и k-медиан) из трех методов секционирования и кластеризации показывают схожие результаты разбиения совокупности 15 постсоветских стран по набору показателей, характеризующих структуру экономических систем;
- по-нашему мнению наилучшее разбиение получено при использовании подхода, основанного на k-метойдов, так как в полной мере проявляются и раскрываются особенности каждой группы стран;
- результаты кластеризации методом k-метойдов позволяют утверждать, что в результате трансформационного перехода от плановой к рыночной экономике наблюдается дифференциация постсоветских стран по структуре экономической системы.

**Библиографический список**

1. Айвазян С. А., Афанасьев М. Ю., Кудров А. В. Метод кластеризации регионов РФ с учетом отраслевой структуры ВРП // Прикладная эконометрика. – 2016. – 1 (41). – С. 24–46.
2. Боженко С. В., Брут-Бруляко А. А. Кластерный анализ в экономических исследованиях: направления и опыт использования // Вестник Костромского государственного технологического университета. – 2008. – № 18. – С. 35–38. – EDN KVSITN.
3. Гичиев Н. С. Кластерный анализ в экономике: теоретический аспект // Региональные проблемы преобразования экономики. – 2020. – 8(118). – С. 176–186. – DOI: [10.26726/1812-7096-2020-8-176-186](https://doi.org/10.26726/1812-7096-2020-8-176-186).
4. Клименко А. В., Слащев И. С. Кластерный анализ данных // Вестник науки. – 2019. – Т. 1, 1 (10). – С. 159–163.
5. Цыпин А. П. Эконометрическое моделирование влияния мезопоказателей на объемы ипотечного жилищного кредитования в регионах России // Экономические науки. – 2025. – № 243. – С. 346–352. – DOI: [10.14451/1.243.346](https://doi.org/10.14451/1.243.346). – EDN UEOSEP.
6. Шаталова О. М., Касаткина Е. В., Лившиц В. Н. Кластерный анализ и классификация промышленно ориентированных регионов РФ по экономической специализации // Экономика и математические методы. – 2022. – Т. 58, № 1. – С. 80–91. – DOI: [10.31857/S042473880018971-7](https://doi.org/10.31857/S042473880018971-7).
7. 6 Different Types of Clustering: All You Need To Know! – URL: <https://datarundown.com/types-of-clustering/> (visited on 03/02/2025).
8. Cluster Analysis – Types, Methods and Examples. – URL: <https://researchmethod.net/cluster-analysis> (visited on 03/02/2025).
9. Cluster Analysis in R – Complete Guide on Clustering in R. – URL: <https://techvidvan.com/tutorials/cluster-analysis-in-r/> (visited on 03/02/2025).
10. Partitioning Around Medoids / The R Project for Statistical Computing. – URL: <https://search.r-project.org/CRAN/refmans/cluster/html/pam.html> (visited on 03/02/2025).