

КЛАСТЕРНЫЙ ПОДХОД К АНАЛИЗУ ПЕНСИОННОЙ ОБЕСПЕЧЕННОСТИ РЕГИОНОВ РОССИИ

Дмитриев Павел Андреевич

студент Северо-Западного института управления Российской Академии Народного Хозяйства и Государственной Службы при Президенте Российской Федерации, РФ, Санкт-Петербург

На конец 2016 года на учете в системе Пенсионного фонда Российской Федерации числилось почти 43 миллиона пенсионеров.

Обеспечение достойного уровня жизни этих граждан – одна из основных задач государства [3].

Российская Федерация – это 85 регионов, каждый из которых характеризуется разным уровнем пенсионного обеспечения. Кластерный подход в исследовании уровня пенсионного обеспечения призван выявить группы похожих регионов, обеспечить понимание выборки, найти сильно отличающиеся от других субъекты федерации.

Для анализа был сформирован набор данных за 2015 года, содержащий следующие показатели уровня жизни в 83 регионах России (без учета Республики Крым и г. Севастополь):

· Ожидаемая средняя продолжительность жизни

· Средняя пенсия

· Процент населения пенсионного возраста

· Среднегодовая стоимость минимального набора продуктов питания

Для наглядности величины взяты в сравнении со средними по стране (таблица 1).

Таблица 1.

Отношение показателей уровня жизни в регионах к средним по стране [3]

Регион	Отношение средней ожидаемой продолжительности жизни регионе к стране	Отношение средней пенсии регионе к стране	Отношение процента пенсионеров в регионе к стране	Отношение стоимости набора продуктов питания в регионе к стране
Алтайский край	0,999	0,923	1,054	
Амурская область	0,954	1,002	0,904	
Архангельская область	1,003	1,243	1,033	

Астраханская область	1,012	0,889	0,946	
Белгородская область	1,030	0,961	1,088	
...	...	...	...	...
Ярославская область	1,007	1,008	1,150	

Перед тем, как приступить к решению задачи кластеризации, проведем предобработку данных – стандартизируем выборку: среднее равно 0, стандартное отклонение равно 1 (таблица 2).

Таблица 2.

Стандартизированный набор данных

Регион	Отношение средней ожидаемой продолжительности жизни в регионе к стране	Отношение средней пенсии в регионе к стране	Отношение процента пенсионеров в регионе к стране	Отношение стоимости набора продуктов к стране
Алтайский край	-0,083	-0,514	0,479	
Амурская область	-1,367	-0,121	-0,272	
Архангельская область	0,026	1,085	0,375	
Астраханская область	0,290	-0,684	-0,063	
Белгородская область	0,796	-0,326	0,646	
...	...	...	...	...
Ярославская область	0,136	-0,090	0,958	

Следующим шагом найдем статистику Хопкинса. Статистика Хопкинса используется для оценки тенденции кластеризации набора данных.

Если статистический показатель близок к 0 (сильно меньше 0,5), то можно сделать вывод, что данные значимо кластеризуемы. Статистика Хопкинса для исследуемой выборки равна 0,192, следовательно, набор данных хорошо подходит для решения задачи кластеризации.

Экспертным путем было установлено, что для данного типа задачи оптимальным методом кластеризации является метод Варда (Ward's method). Этот метод отличается от всех других методов, поскольку для оценки расстояний между кластерами он использует методы дисперсионного анализа.

Метод минимизирует сумму квадратов для любых двух (гипотетических) кластеров, которые могут быть сформированы на каждом шаге. [2]. Рассчитаем индексы для выбора оптимального числа кластеров (таблица 3.)

Таблица 3.

Индексы оптимального числа кластеров

Индекс	Оптимальное число кластеров	Индекс	Оптимальное число кластеров
KL	4	Duda	2
CH	8	Pseudot2	2
Hartigan	4	Beale	2
CCC	9	Ratkowsky	4
Scott	3	Ball	3
Marriot	3	Ptbiserial	4
Trcovw	4	Frey	1
Tracew	4	McClain	4
Friedman	3	Dunn	4
Rubin	8	Hubert	0
Cindex	2	Sdindex	4
DB	10	Dindex	0
Silhouette	2	DSbw	9

Согласно правилу большинства, оптимальное число кластеров для исследуемой выборки равно 4.

Далее построим дендрограмму (произведем иерархическую кластеризацию) методом Варда (рис. 1). Дендрограмма показывает степень близости отдельных объектов и кластеров, а также наглядно демонстрирует в графическом виде последовательность их объединения или разделения [1].

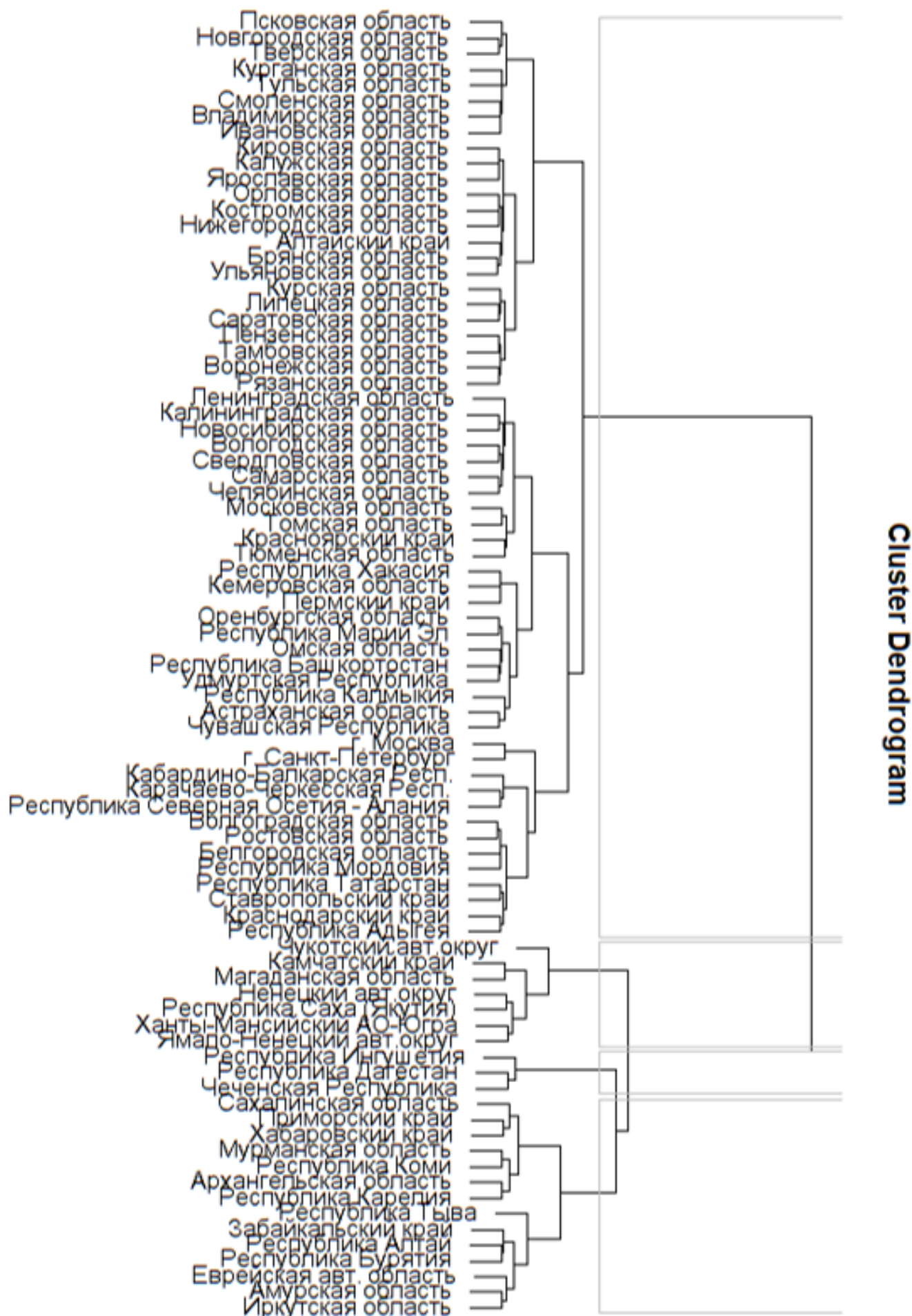


Рисунок 1. Результат иерархической кластеризации

Выведем кластеры в таблицу (таблица 4).

Таблица 4.

Результаты кластеризации

Кластер 1	Кластер 2	Кластер 3	Кластер 4
Алтайский край	Амурская обл.	Камчатский край	Респ. Дагестан
Астраханская обл.	Архангельская обл.	Магаданская обл.	Респ. Ингушетия
Белгородская обл.	Еврейская авт. обл.	Ненецкий авт.округ	Чеченская Респ.
Брянская обл.	Забайкальский край	Респ. Саха (Якутия)	
Владимирская обл.	Иркутская обл.	Ханты-Мансийский АО-Югра	
Волгоградская обл.	Мурманская обл.	Чукотский авт.округ	
Вологодская обл.	Приморский край	Ямало-Ненецкий авт.округ	
Воронежская обл.	Респ. Алтай		
г. Москва	Респ. Бурятия		
г. Санкт-Петербург	Респ. Карелия		
Ивановская обл.	Респ. Коми		
Кабардино-Балкарская Респ.	Респ. Тыва		
Калининградская обл.	Сахалинская обл.		
Калужская обл.	Хабаровский край		
...	...		
Ярославская обл.			

Интерпретируем результаты. Первый, самый большой кластер, представлен группой достаточно «ровных» регионов, где все показатели не сильно разнятся с единицей. Второй кластер отличается от первого относительно высокой стоимостью (выше среднего по стране) минимального набора продуктов питания.

Третий кластер характеризует северные регионы с самыми высокими пенсиями (в 1,5 раза больше средней по стране), невысоким процентом пенсионеров (вдвое меньше среднего по стране) и очень высокой стоимостью минимального набора продуктов питания (1,6 раза больше среднего по стране). Четвертый, самый маленький кластер, представлен республиками Северного Кавказа, с низким процентом жителей пенсионного возраста и невысокой пенсией. Как известно, эти регионы выделяются молодым населением и высокой рождаемостью.

Исследование показало, что при решении задачи повышения уровня пенсионного обеспечения регионов, реформирования региональных пенсионных систем т. д., можно сократить размерность задачи, рассматривая не каждый регион в отдельности, а 4 кластера.

Список литературы:

1. Дендрограмма [Электронный ресурс]. URL: <https://basegroup.ru/community/glossary/dendrogram> (дата обращения: 21.01.2018)
2. Кластерный анализ [Электронный ресурс]. URL: <http://statsoft.ru/home/textbook/modules/stcluan.html> (дата обращения: 19.01.2018)
3. Официальный сайт Федеральной службы государственной статистики URL: [Электронный ресурс]. <http://www.gks.ru> (дата обращения: 21.01.2018)