

Михеева Татьяна Николаевна

студентка

ФГБОУ ВПО «Поволжский государственный

технологический университет»

г. Йошкар-Ола, Марий Эл Республика

КЛАСТЕРНЫЙ АНАЛИЗ РЕГИОНОВ ПФО ПО СУММАРНОЙ ПЛОЩАДИ ЖИЛЬЯ, ПРИХОДЯЩЕГОСЯ В СРЕДНЕМ НА ОДНОГО ЖИТЕЛЯ

***Аннотация:** данная статья посвящена кластерному анализу изменений величины общей площади жилых помещений, приходящихся в среднем на одного жителя в зависимости от временного интервала.*

***Ключевые слова:** кластерный анализ, иерархический агломеративный метод, метод k -средних.*

Одной из задач данного исследования является выявление произошедших изменений по величине общей площади жилых помещений, приходящихся в среднем на одного жителя, в зависимости от принадлежности их к определенным кластерам в разное время.

Целью данной работы является проведение сравнительного анализа регионов по выделенным кластерам, используя сравнения средних значений по кластерам и построением графика k -средних.

Для исследования были взяты данные по величине общей площади жилых помещений, приходящихся в среднем на одного жителя, в регионах ПФО за период 2002–2013 гг. Кластерный анализ проводился с использованием пакета STATISTICA.

Проведем анализ данных за два временных периода с использованием иерархических агломеративных методов. Для разбиения на кластеры данных от 2002–2007 гг. воспользуемся методом взвешенного попарного среднего. Результат разбиения представлен на рисунке 1.

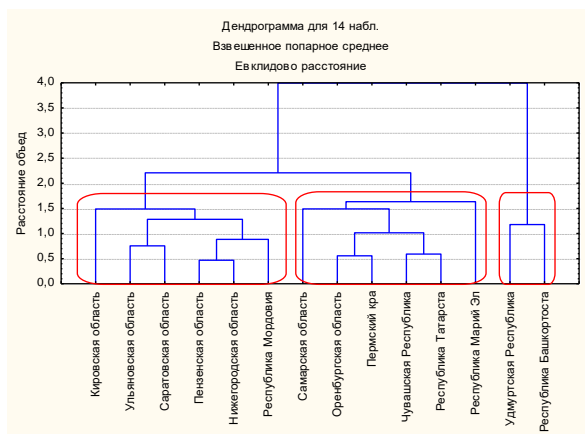


Рис. 1. Дендрограмма по методу взвешенного попарного среднего (2002–2007)

С помощью метода попарного среднего получили разбиение областей на 3 кластера (рисунок 1).

Аналогичную операцию проведем с данными за 2008–2013 гг. Результат отобразим на рисунке 2.

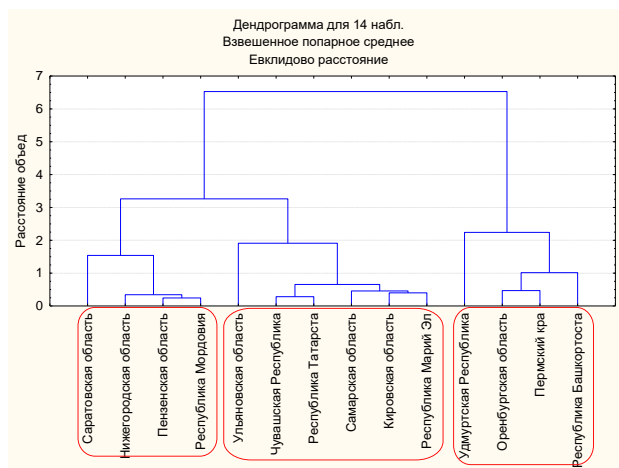


Рис. 2. Дендрограмма по методу взвешенного попарного среднего (2008–2013 гг.)

С помощью взвешенного попарного среднего метода получили разбиение областей на 3 кластера (рисунок 2).

Однако использование одного метода в процессе кластеризации данных неэффективно и не дает полной уверенности в достоверности полученных результатов. Поэтому для подтверждения полученных выводов воспользуемся методом k-средних.

Метод k-средних наиболее распространен среди неиерархических методов кластерного анализа. В отличие от иерархических методов, которые не требуют предварительных предположений относительно числа кластеров, для этого метода необходимо иметь гипотезу о наиболее вероятном количестве кластеров [2, с. 123].

На рисунке 3 представим график средних для каждого кластера данных за период 2002–2007 гг., полученный на основе результатов кластерного анализа указанным выше методом.

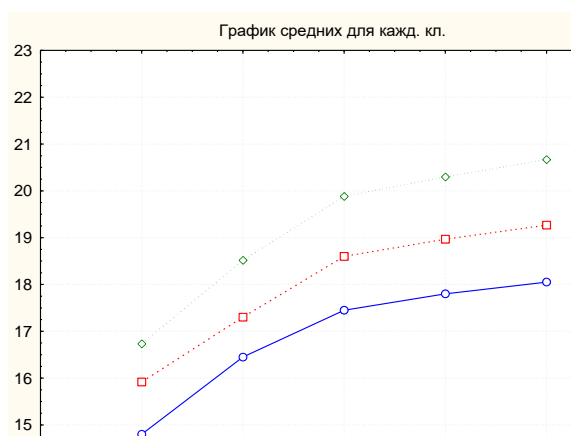


Рис. 3. График средних (2002–2007 гг.)

Сравнив данные значения и средние значения полученных кластеров, видно, что средние значения кластеров 1 и 2 сравнительно ниже, чем значения по ПФО, значение кластера 3 сравнительно выше. Также мы наблюдаем постепенный рост величины общей площади жилых помещений, приходящихся в среднем на одного жителя, по всему ПФО.

Аналогичная ситуация при построении графика средних получилась за 2008–2013 гг.

Сравнив кластеры, полученные за 2002–2007 гг. и 2008 – 2013 гг., получилось, что в целом ситуация с величиной общей площади жилых помещений, приходящихся в среднем на одного жителя, по всему ПФО изменилась в положительную сторону.

Можно заметить, что в Оренбургской области и Пермском крае ситуация с величиной общей площади жилых помещений, приходящихся в среднем на одного жителя ухудшилась, т. к. они примкнули к кластеру, где уровень величины общей площади жилых помещений ниже среднего по ПФО. Аналогичная ситуация в Кировской, Саратовской и Самарской областях.

Список литературы

1. Боровиков В.П. Популярное введение в современный анализ данных в системе STATISTICA: Учебное пособие для вузов / В.П. Боровиков. – М.: Гор. линия-Телеком, 2013. – 288 с.
2. Дубров А.М. Многомерные статистические методы: Учебник / А.М. Дубров, В.С. Мхитарян, Л.И. Трошин. – М.: Финансы и статистика, 2000.