

ТЕХНИЧЕСКИЕ НАУКИ

Моденова Анна Александровна

магистрант

Якимов Игорь Максимович

доцент, профессор

ФГБОУ ВПО «Казанский национальный исследовательский
технический университет им. А.Н. Туполева»
г. Казань, Республика Татарстан

КЛАСТЕРНЫЙ АНАЛИЗ РЕГИОНОВ РОССИИ ПО НАУЧНОЙ И ИННОВАЦИОННОЙ АКТИВНОСТИ

Аннотация: статья посвящена проблеме кластерного анализа регионов России. Как отмечают авторы, федеральные округи в значительной степени различаются по уровню развития научной и инновационной активности. Для принятия управленческих решений в условиях, когда существует ряд объектов управления, всегда требуется классификация по ряду важных признаков. Выделение общих признаков и критериев уровня развития регионов в научной и инновационной деятельности позволит определить и основные факторы развития регионов округа, а также сделать в дальнейшем прогнозы. По мнению авторов, для статистических исследований целесообразно разделение регионов на кластеры с однородными значениями признаков.

Ключевые слова: научная активность, инновационная активность, кластерный анализ.

При проведении кластерного анализа было выделено 47 признаков инновационной активности регионов из статистического сборника Росстата [3]. Для примера назовём 5, наиболее характерных, из них: объём инновационных товаров и услуг; количество выданных патентов на изобретения; численность персонала, занятого научными исследованиями и разработками; численность аспирантов; затраты на научные исследования и разработки.

Годовые значения исходных статистических данных за 2013 год, взяты из [3].

Проведем кластеризацию иерархическим методом Уорда, по которому в пакете прикладных программ Statistica 6.0 [1] имеется стандартная процедура. Принимается допущение об одинаковой значимости всех признаков регионов. В качестве расстояния между кластерами берется прирост суммы квадратов расстояний объектов до центров кластеров, получаемый в результате их объединения. На каждом шаге алгоритма объединяются такие два кластера, которые приводят к минимальному увеличению целевой функции, т.е. внутригрупповой суммы квадратов (1). Этот метод направлен на объединение близко расположенных кластеров и «стремится» создавать кластеры малого размера. Дендрограмма кластеров приведена на рис. 1.

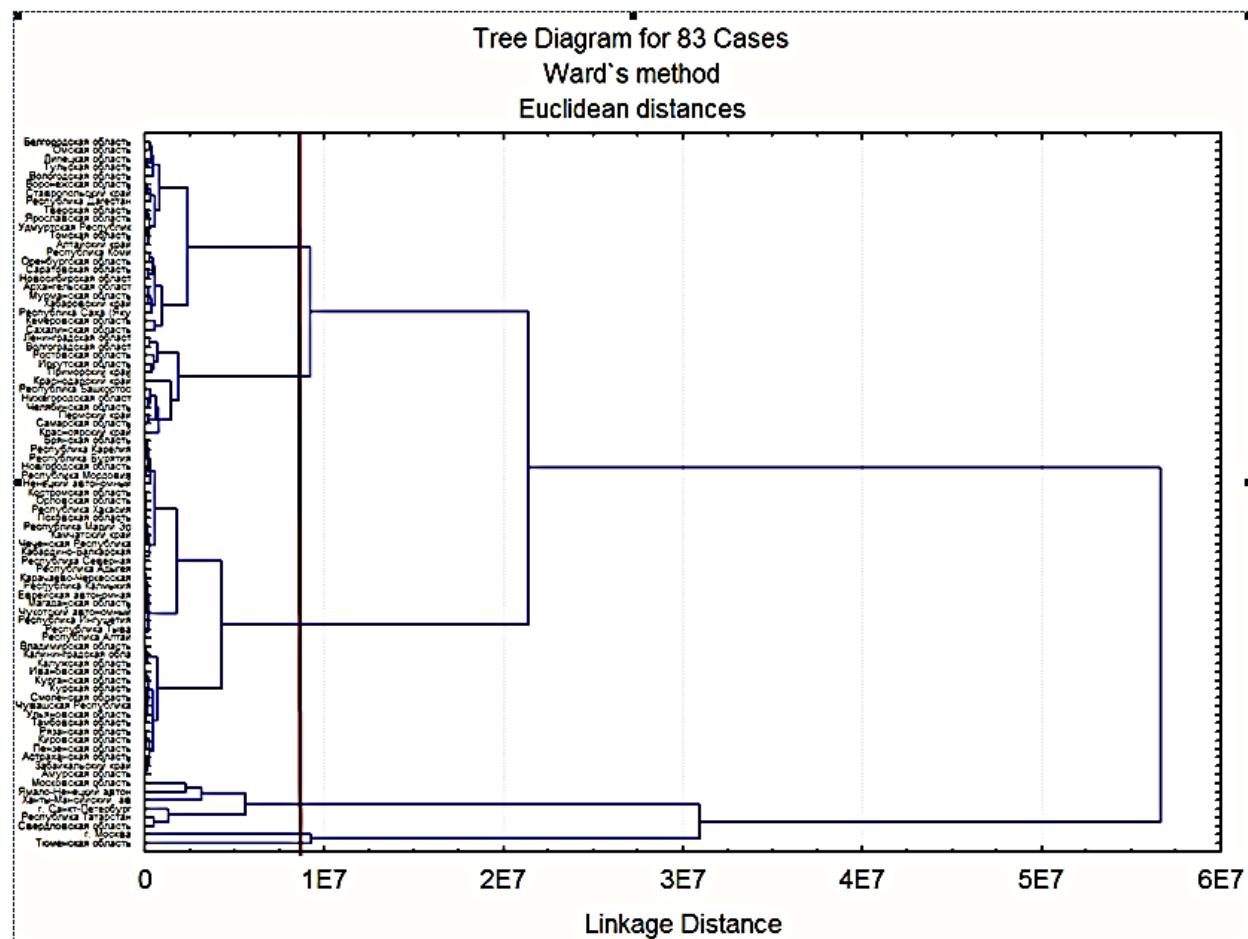


Рис. 1. Дендрограмма

Поставим задачу выделения кластеров по показателям расстояния между признаками в группируемых объектах исследования с выполнением следующих условий.

$$\sum_{i=1}^k \sum_{j=1}^k \delta_{ij} d_{ij} \leq D_{\text{порог}} , \quad (1)$$

$$\sum_{i=1}^k \sum_{j=1}^k (1 - \delta_{ij}) d_{ij} > D_{\text{порог}} , \quad (2)$$

где k – количество объектов (регионов);

d_{ijg} – расстояние между g -ым признаком i -го и j -го объектов;

δ_{ij} – символ Кронекера, принимающий значение 1, если i -ый и j -ый объекты входят в один и тот же кластер; и значение 0, если не входят;

$D_{\text{порог}}$ – пороговое значение расстояния между объектами для выделения кластеров.

Для определения количества кластеров, на которые целесообразно разбить все регионы РФ, нужно выбрать пороговое расстояние $D_{\text{порог}}$ – то есть такое расстояние, при превышении которого объединяться будут уже слишком далекие с точки зрения исследователя объекты. После выбора порогового расстояния проводится перпендикуляр через точку, соответствующую выбранному расстоянию, и подсчитывается количество его пересечений с «ветвями» дендрограммы (рис. 1). Количество пересечений и будет определять количество выделяемых кластеров, а объекты, оказавшиеся на «отсечённой» ветке – состав кластеров. При выбранном пороговом расстоянии – красная вертикаль на рис.1 – количество кластеров будет равно 6. Сдвигом вертикали вправо количество выделяемых кластеров уменьшается и увеличивается суммарное расстояние между признаками в кластере. Сдвигом вертикали влево количество выделяемых кластеров увеличивается и уменьшается суммарное расстояние между признаками в кластере.

Шестой кластер, наиболее значимый по значениям признаков, образует единственный регион – Тюменская область. Пятый кластер – город Москва. Республика Татарстан (РТ) вошла в четвёртый кластер, в который вошли ещё пять следующих регионов: город Санкт-Петербург, Свердловская область, Московская область, Ямало-Ненецкий автономный округ и Ханты-Мансийский автономный округ –

ЮГра. В третий кластер вошло 40 регионов. Во второй кластер вошло 12 регионов. В первый кластер вошло 23 региона.

Проведённая кластеризация другими иерархическими методами: методом наиболее удаленных соседей (полной связи), методом одиночной связи, методом невзвешенного попарного среднего и методом взвешенного попарного среднего, показала, что за счёт выбора порогового значения – $D_{порог}$ можно каждым методом добиться одинаковых результатов.

Кластеризация методом К-средних позволяет формировать кластеры, задаваясь количеством кластеров. Поставим задачу выделения кластеров по показателям расстояния между признаками в группируемых объектах исследования с выполнением следующих условий.

$$\sum_{i=1}^k \sum_{j=1}^k \delta_{ij} d_{ij} \rightarrow \min , \quad (3)$$

$$\sum_{i=1}^k \sum_{j=1}^k (1 - \delta_{ij}) d_{ij} \rightarrow \max , \quad (4)$$

где k – количество объектов;

d_{ijg} – расстояние между g -ым признаком i -го и j -го объектов;

δ_{ij} – символ Кронекера, принимающий значение 1, если i -ый и j -ый объекты входят в один и тот же кластер; и значение 0, если не входят.

Условие (3) обеспечивает минимум расстояний между признаками объектов, вошедших в один и тот же кластер; а (4) максимум этих расстояний между объектами, вошедшими в разные кластеры.

Была проведена кластеризация при двух различных значениях $k=4$ и $k=6$.

В заключении отметим следующие результаты кластерного анализа.

1. Проведённая кластеризация не противоречит здравому смыслу, в частности регионы, вошедшие в четвёртый кластер, являются весьма состоятельными в экономическом отношении и могут себе позволить выделять средства на научные и инновационные работы.

2. Наибольшее различие, сыгравшее самое существенное значение при выделении кластеров, наблюдается в следующих признаках: x_{36} – валовый региональный

продукт, x_{37} – объём основных фондов, x_{38} – инвестиции в основной капитал, x_{45} – затраты на технологические инновации, x_{46} – объём инвестиционных товаров и услуг.

3. Количество регионов, вошедших в кластер с РТ – 6, можно считать вполне приемлемым, что в дальнейшем позволит генерировать управленческие решения по повышению эффективности научной и инновационной деятельности РТ на основании его предыдущих результатов и ещё пяти регионов, выделенных вместе с ним в один кластер с однородными значениями признаков.

Список литературы

1. Боровиков В.П. Statistica – Статистический анализ и обработка данных в среде Windows / В.П. Боровиков, И.П. Боровиков. – М.: Информационно-издательский дом «Филин», 1998. – 608 с.
2. Елисеева И.И. Общая теория статистики / И.И. Елисеева, М.М. Юзбашев. – М.: Финансы и статистика, 1998. – 500 с.
3. Регионы России. Социально-экономические показатели. Стат. сб. / Росстат. – М., 2012. – 990 с.