



МАТЕМАТИЧЕСКИЕ МЕТОДЫ В ГЕОГРАФИИ

Карпиченко Александр Александрович

***доцент кафедры почвоведения и
земельных информационных систем***

Литература

- elib.bsu.by
- Математические методы в географии: учебно-методическое пособие / Н. К. Чертко, А. А. Карпиченко. – Минск: БГУ, 2009.

3. КЛАСТЕРНЫЙ АНАЛИЗ

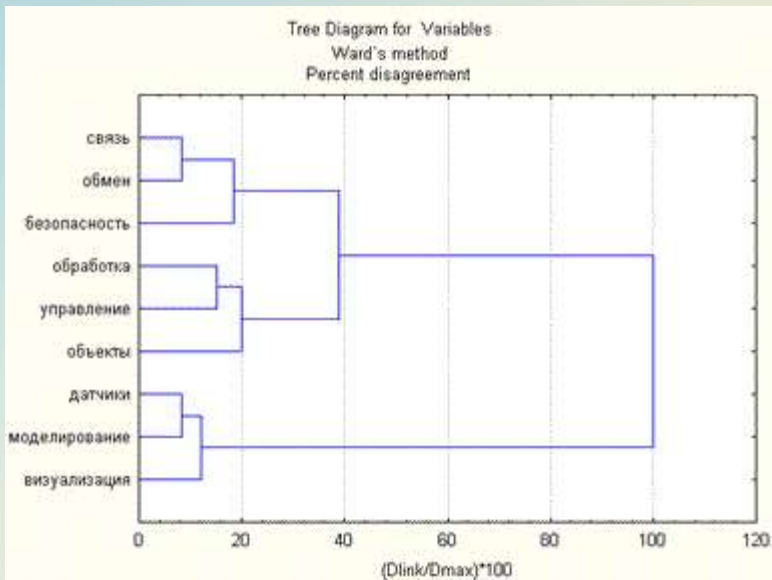
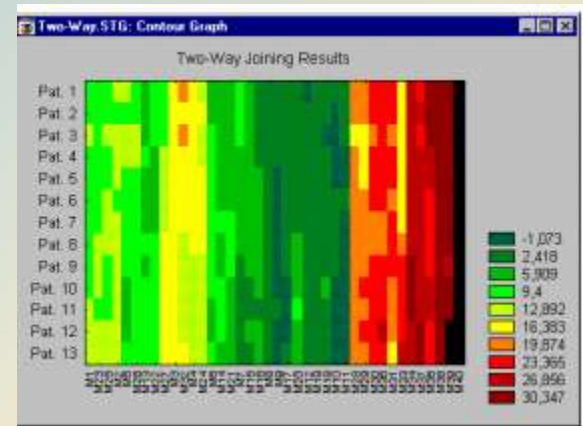
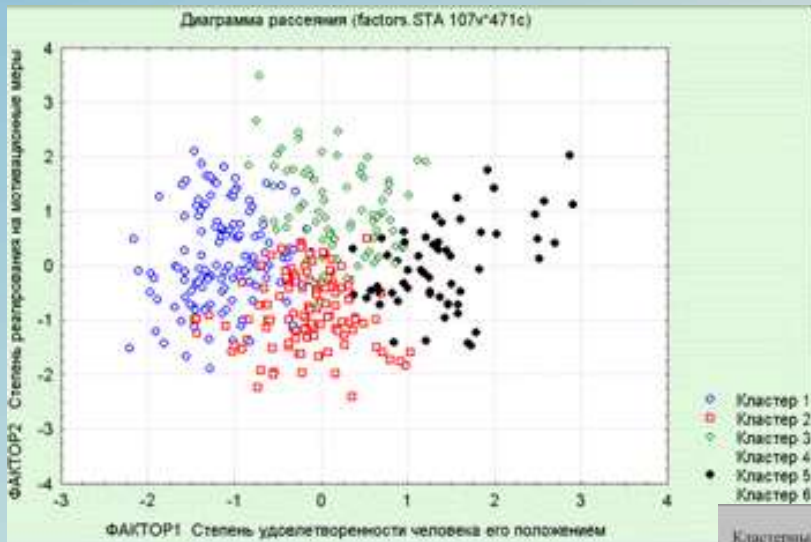
При проведении географических исследований, как правило, возникает проблема *объединения по сходству (кластеризация)* объектов, которые характеризуются множеством признаков, выраженных в разных единицах измерения. Для этой цели используется *кластерный анализ*. Поскольку кластерный анализ занимается классификацией объектов, а факторный исследует связи между ними, то оба метода дополняют друг друга и между ними иногда трудно провести четкие границы.

Методологические особенности кластерного анализа сводятся к выявлению единой меры, охватывающей ряд исследуемых признаков. Эти признаки объединяются с помощью метрики (расстояния) в один кластер сходства группируемых объектов.

3. КЛАСТЕРНЫЙ АНАЛИЗ

Кластерный анализ – это общее название множества вычислительных процедур, используемых при создании классификации. Главная цель кластерного анализа – *нахождение групп схожих объектов в выборке данных.* Эти группы удобно называть кластерами. Не существует общепринятого определения термина «*кластер*», однако считается, что кластеры обладают некоторыми свойствами, наиболее важными из которых являются плотность, дисперсия, размеры, форма и отделимость. Исходя из их свойств кластеры – это непрерывные области некоторого пространства с относительно высокой плотностью точек, отделенные от других таких же областей областями с относительно низкой плотностью точек.

3. КЛАСТЕРНЫЙ АНАЛИЗ



3. КЛАСТЕРНЫЙ АНАЛИЗ

Применяя процедуры кластерного анализа, **всегда следует помнить**, что:

1. Многие методы кластерного анализа – довольно простые процедуры, которые, как правило, не имеют достаточного статистического обоснования (то есть большинство методов являются эвристическими).
2. Методы кластерного анализа разрабатывались для многих дисциплин, а потому несут на себе отпечатки специфики этих дисциплин.
3. Разные кластерные методы могут порождать и порождают различные решения для одних и тех же данных.
4. Цель кластерного анализа заключается в поиске существующих структур. В то же время его действие состоит в привнесении структуры в анализируемые данные, и эта структура может не совпадать с искомой «реальной».

3. КЛАСТЕРНЫЙ АНАЛИЗ

При относительной формализации методов кластерного анализа они носят эвристический (теоретический) характер, реализуют принцип здравого смысла. Для оценки сходства объектов по ряду признаков используют три типа мер:

- *коэффициент подобия* – для группировки объектов и признаков, если уровни показателей являются действительно целыми числами;
- *коэффициенты связи* – чаще применяются для группировки признаков с использованием коэффициента корреляции;
- *показатели расстояния* – характеризуют степень взаимной удаленности признаков и применяются в основном для кластеризации объектов; признаки объектов должны быть независимыми, что предварительно можно уточнить с помощью корреляционного анализа.

3. КЛАСТЕРНЫЙ АНАЛИЗ

Многомерное наблюдение может быть интерпретировано геометрически в виде точки в многомерном пространстве. *Геометрическая близость точек в пространстве означает близость физических состояний объектов, их однородность.* Решающим в интерпретации остается выбор масштаба метрики, т.е. задание расстояния между объектами, которые объединяют или разъединяют объекты. В результате разбиения объектов на группы по сходству признаков образуются *кластеры (таксоны, образы)*. Необходимость разбиений совокупности объектов на однородные группы возникает при проведении социально-экономических, землеустроительных, географических исследований и т. д.

3. КЛАСТЕРНЫЙ АНАЛИЗ

Выбор метрики (меры близости) является важнейшим моментом исследования, который определяет окончательный вариант разбиения объектов на группы. Это зависит от цели исследования, физической и статистической природы вектора наблюдений (x), полноты априорных сведений о характере вероятностного распределения x .

В задачах кластер-анализа широко используются следующие метрики: **Эвклида, манхэттенское расстояние, Чебышева, Минковского, Хемминга, меры близости, задаваемые потенциальной функцией.** **Эвклидова метрика наиболее употребительна.**

3. КЛАСТЕРНЫЙ АНАЛИЗ

Обычно расстояние по формулам: среднее Эвклидовое рассчитывается по

$$d_{kl} = \sqrt{\frac{1}{m} \sum_{j=1}^m (Z_{kj} - Z_{lj})^2} \quad (3.1)$$

где m – число признаков x ; Z_{kj} , Z_{lj} – стандартизированные значения признака j для k и l объектов соответственно, или:

$$d_{kl} = \sqrt{\frac{(Z_{kj_1} - Z_{lj_1})^2 + (Z_{kj_2} - Z_{lj_2})^2 + \dots + (Z_{kj_m} - Z_{lj_m})^2}{m}}$$

3. КЛАСТЕРНЫЙ АНАЛИЗ

Метод дендритов. Исследуемые объекты, разделенные на кластеры, можно изобразить в виде дендрограммы, которая представляет собой графическое изображение матрицы расстояний или сходства. Такой анализ объектов исследования носит название метода дендритов. Имея n объектов, можно построить большое количество дендрограмм, которые соответствуют избранной процедуре кластеризации. Для конкретной матрицы расстояний или сходства существует только одна дендрограмма.

3. КЛАСТЕРНЫЙ АНАЛИЗ

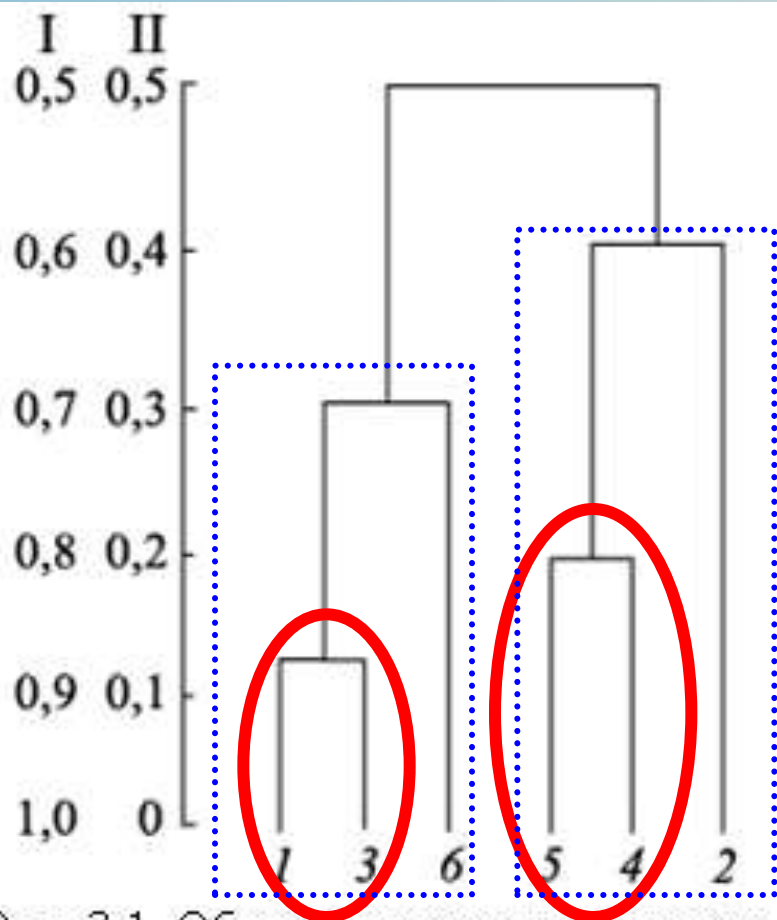


Рис. 3.1. Общий вид дендрограммы:
I – сходство, II – расстояние

Представим дендрограмму с шестью объектами ($n = 6$) (рис. 3.1). Объекты 1 и 3 наиболее близки, т. е. наименее удалены друг от друга, поэтому объединяются в один кластер на уровне сходства, равном 0,9 (образуют 1-й шаг). Объекты 4 и 5 объединяются при уровне сходства 0,8 (2-й шаг). На 3-м и 4-м шагах процесса образуются кластеры 1, 3, 6 и 5, 4, 2, соответствующие уровню сходства соответственно 0,7 и 0,6. Окончательно все объекты группируются в один кластер при уровне сходства 0,5.