

# КЛАСТЕРНЫЙ АНАЛИЗ

План:

## 1. Общие сведения о кластеризации:

*1.1. Понятие о кластеризации. Примеры.*

*1.2. Меры сходства объектов и кластеров.*

*1.3. Качество классификации.*

*1.4. Классификация методов кластерного анализа.*

## 2. Кластеризация в SPSS:

*2.1. Иерархическая кластеризация. Анализ результатов.*

*2.2. Кластеризация k-средними. Интерпретация результатов*

*2.3. Двухэтапный кластерный анализ.*

## 3. Практикум





# Сегментация потребителей

| Категория  | Сегменты |    |     |     |
|--|----------|----|-----|-----|
|  | 1        | 2  | 3   | 4   |
|  | 18%      | 7% | 60% | 15% |
| Моющая способность (результаты отстирывания и замачивания) | 3        | 8  | 2   | 7   |
| Отдушка (придание белью запаха после стирки)               | 5        | 5  | 7   | 1   |
| Цена   | 8        | 7  | 1   | 2   |
| Безвредность для здоровья                                  | 1        | 4  | 8   | 3   |
| Эффект отбеливания   | 2        | 6  | 3   | 6   |
| Подсинивание (для белого белья)                            | 4        | 1  | 6   | 8   |
| Способность быстро растворяться                            | 7        | 1  | 4   | 5   |
| Отсутствие пыления при пересыпании                         | 6        | 2  | 5   | 4   |

# Меры расстояния между объектами

Евклидово расстояние:

$$d_{ij} = \sqrt{\sum_{k=1}^m (x_{ik} - x_{jk})^2}$$

Взвешенное евклидово расстояние:

$$d_{ij} = \sqrt{\sum_{k=1}^m w_k (x_{ik} - x_{jk})^2}$$

Расстояние Минковского:

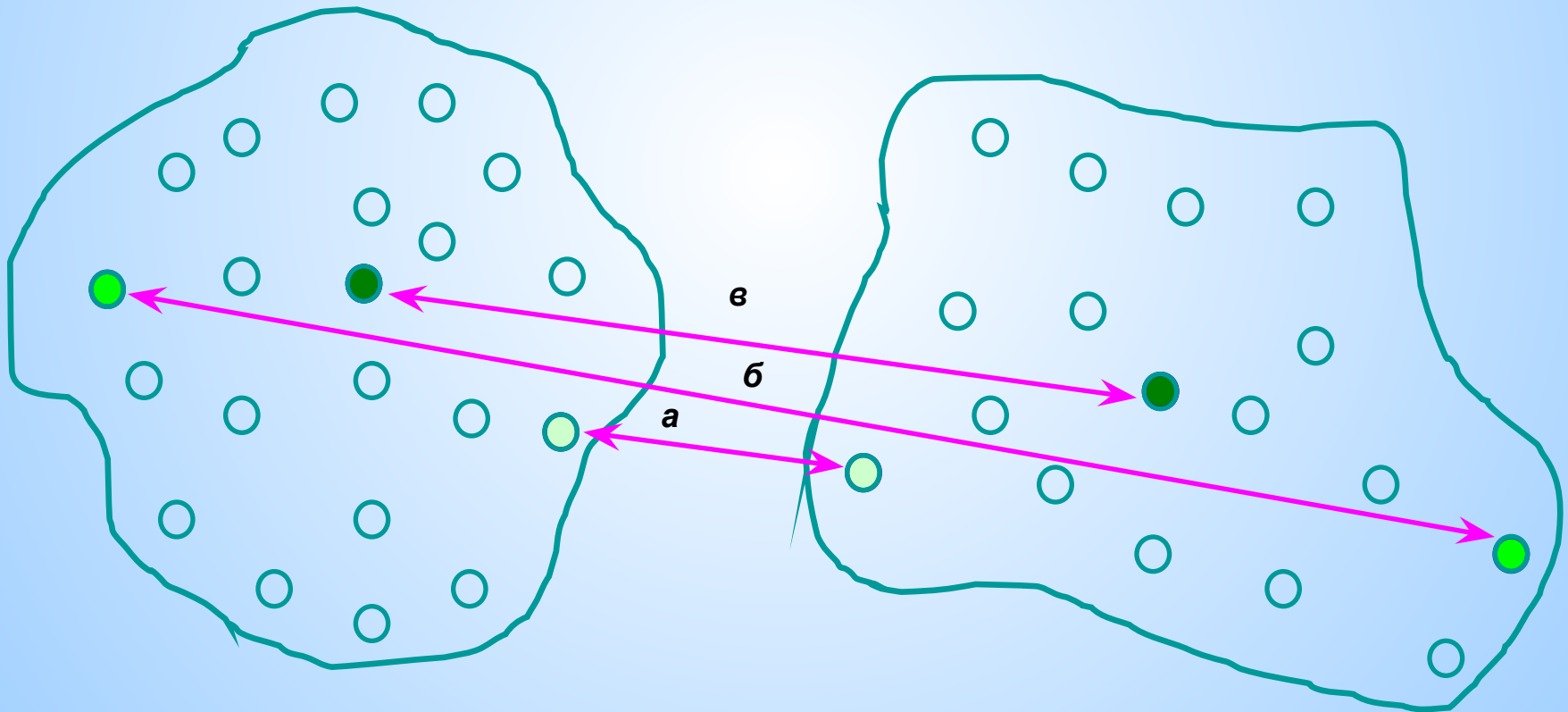
$$d_{ij} = \left( \sum_{k=1}^m |x_{ik} - x_{jk}|^p \right)^{1/p}$$

Расстояние city-block:

$$d_{ij} = \sum_{k=1}^m |x_{ik} - x_{jk}|$$

# Расстояние между кластерами

- а) метод «ближнего соседа»
- б) метод «дальнего соседа»
- в) центроидный метод
- г) метод межгрупповых связей



# Качество классификации

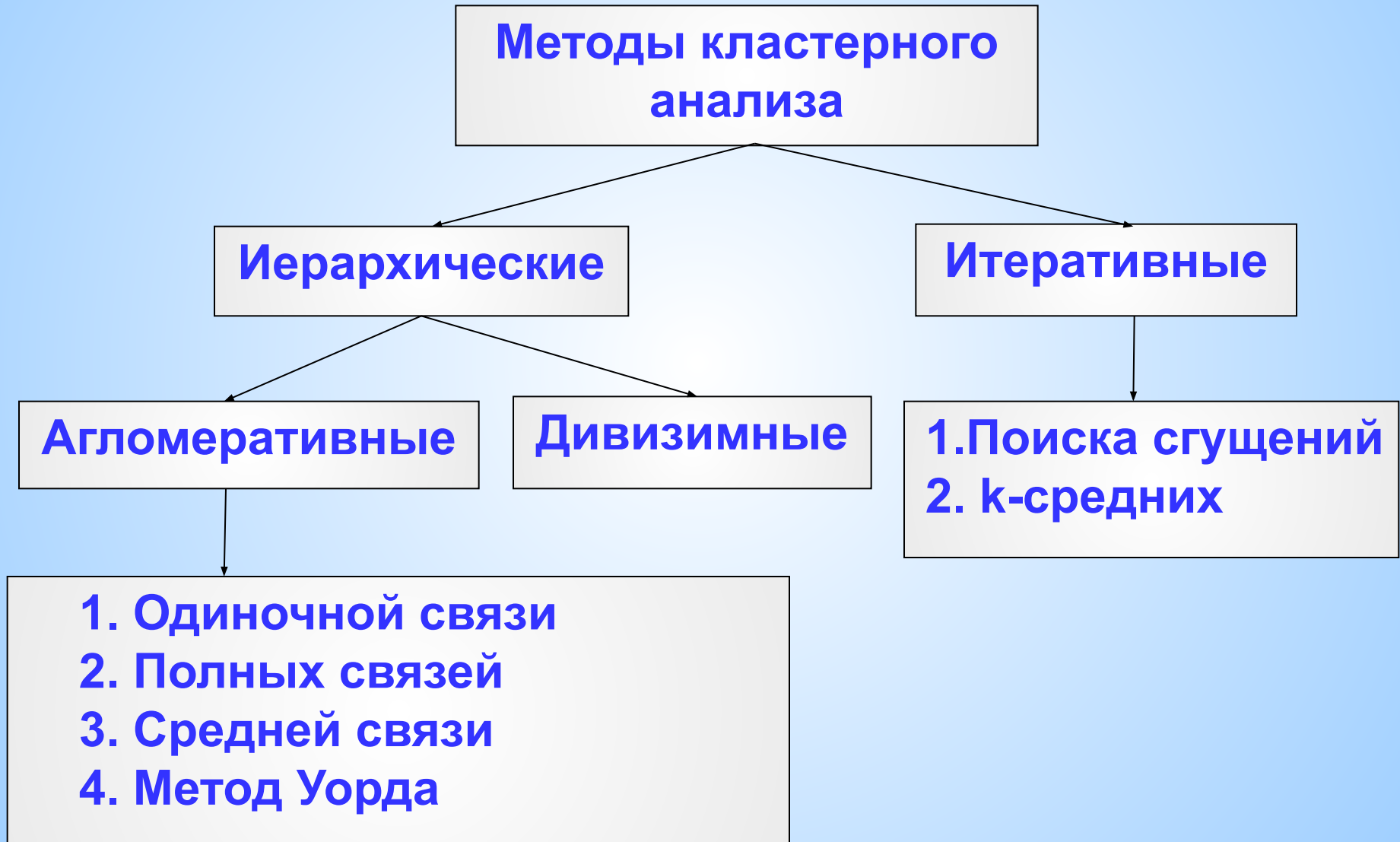
## Функционалы качества

- *сумма внутриклассовых дисперсий*  $Q_1(S) = \sum_{\mathbb{K}=1}^p \sum_{x_i \in S_{\mathbb{K}}} d^2(x_i, \bar{x}_{\mathbb{K}})$
- *сумма попарных внутриклассовых расстояний между элементами*  $Q_2(S) = \sum_{\mathbb{K}=1}^p \sum_{x_i, x_j \in S_{\mathbb{K}}} d^2(x_i, x_j)$

## Кластерные профили

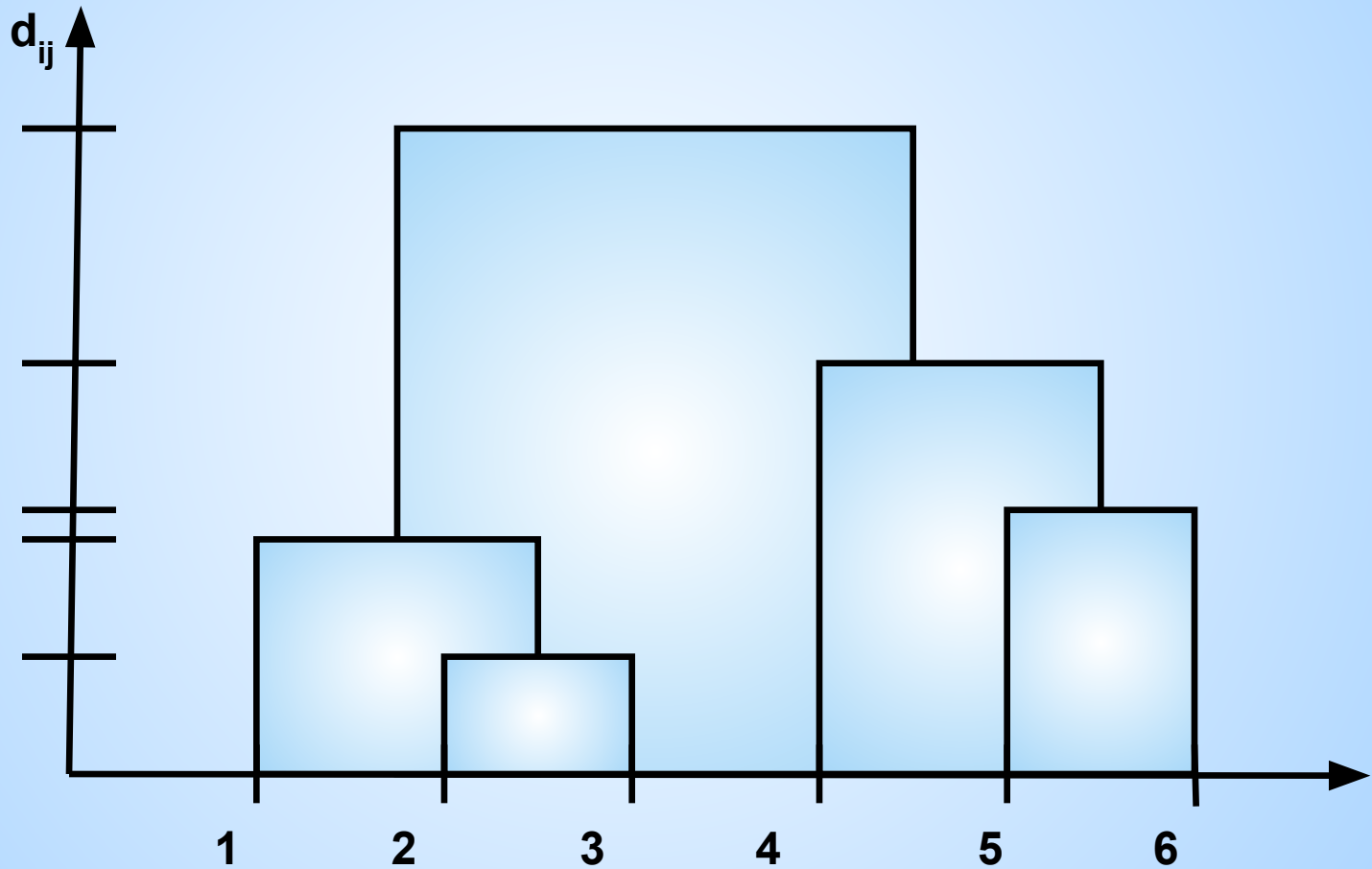
| Кластер | Муж. | 30-50 лет | > 50 лет | Рук. | Мед. | Льготы | З/п | Стаж | Образов. |
|---------|------|-----------|----------|------|------|--------|-----|------|----------|
| 1       | 80   | 90        | 5        | 70   | 10   | 12     | 95  | 30   | 30       |
| 2       | 50   | 70        | 45       | 13   | 60   | 70     | 60  | 40   | 20       |
| 3       | 90   | 35        | 10       | 5    | 30   | 20     | 70  | 20   | 50       |

# Классификация методов



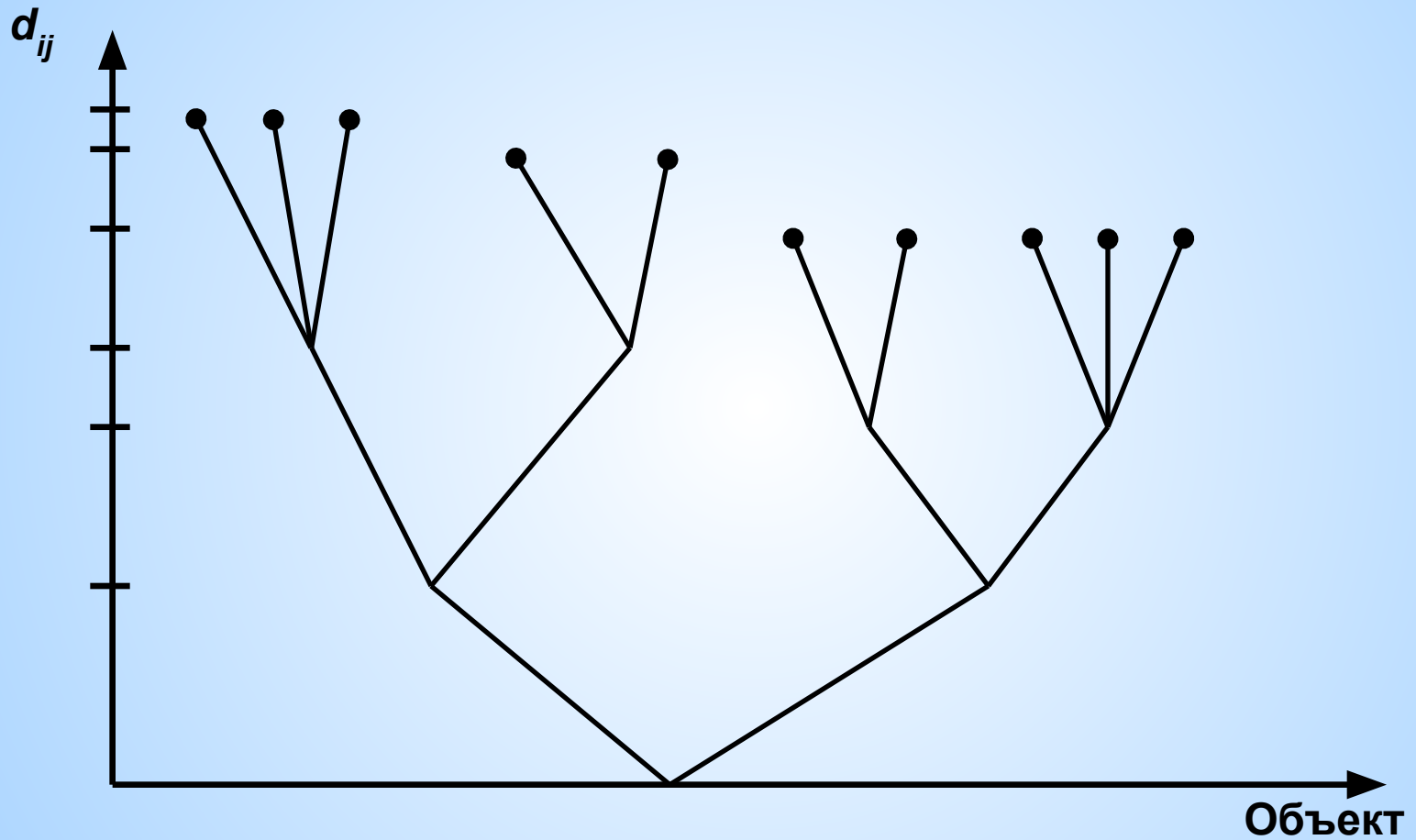


# Иерархические агломеративные методы





# Иерархические дивизимные методы



# Итеративный метод k-средних

