

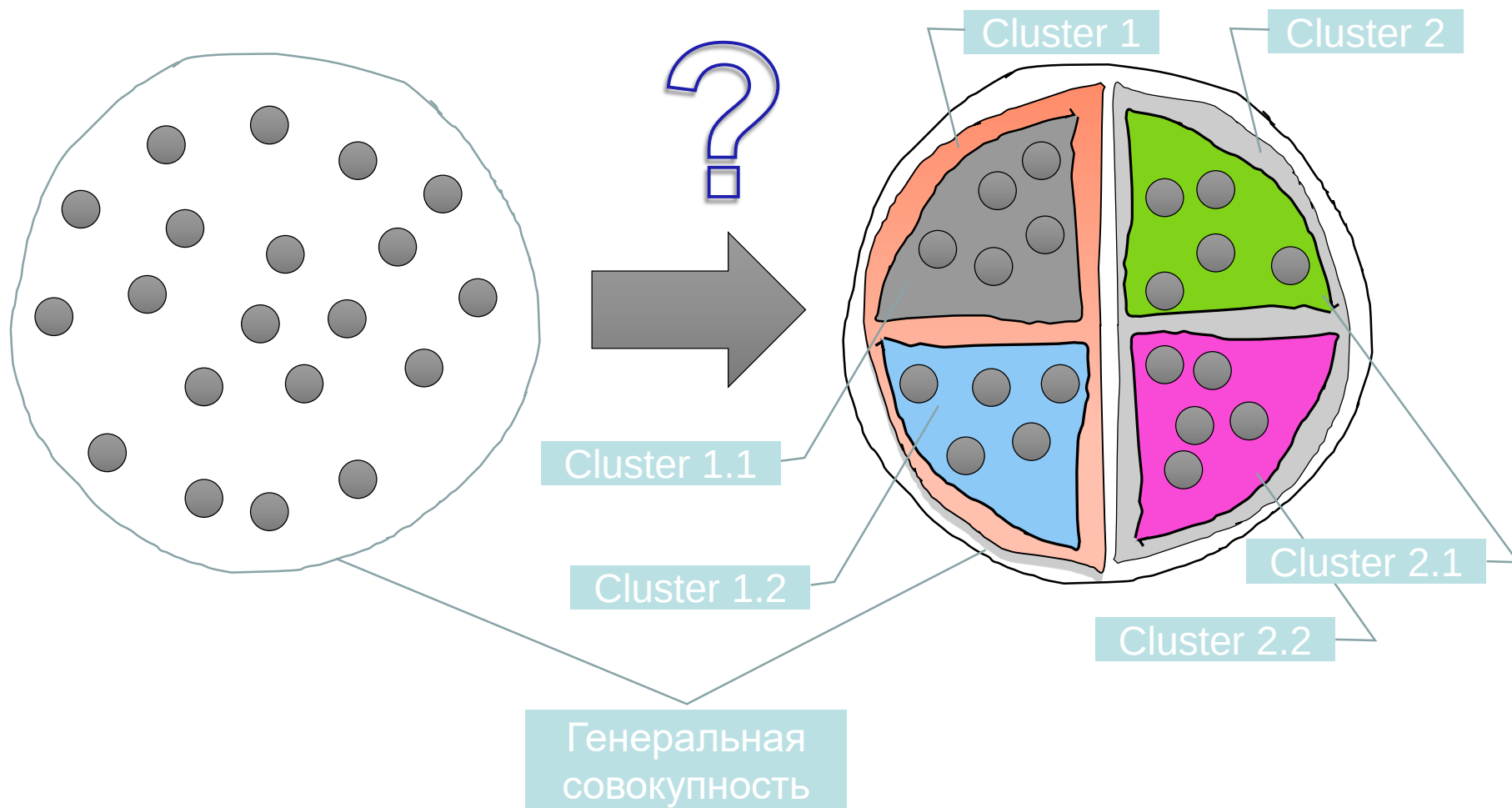
2. История возникновения и применения кластерного анализа

2.1. Сущность и задачи кластерного анализа

2.2. Примеры исследований и данных

2.3. Предостережения и рекомендации

2.1.



Кластерный анализ –

это многомерная статистическая процедура, выполняющая сбор данных, содержащих информацию о выборке объектов ($X_1, X_2, \dots, X_m$), и затем упорядочивающая объекты в сравнительно однородные группы.

Cluster (сгусток, пучок, группа) – кластер, класс, таксон, сгущение.

Комбинационная группировка

Монотетический подход

Кластерный анализ

Политетический подход

Биология

«Начала численной таксономии», Робертом Сокэл и Петер Снит (1963 г.).

Появление высокоскоростных компьютеров

Фундаментальное значение классификации как научного метода

Антропология

Определение однородных культурных областей, используя матричные методы (Czekanowski, 1911; Driver, 1965; Johnston, 1972)

Психология

«Факторный анализ бедняка», (Tryon, 1939)

Задачи кластерного анализа:

- **разработка типологии или классификации;**
- **исследование полезных концептуальных схем группирования объектов;**
- **порождение гипотез на основе исследования данных;**
- **проверка гипотез или исследования для определения, действительно ли типы (группы), выделенные тем или иным способом, присутствуют в имеющихся данных.**

Примеры ранних исследований

1. Голдстейн и Линден (1969) – классификация больных алкоголизмом с помощью психологического теста MMPI (миннесотское многофазовое обследование личности – Minnesota Multiphasic Personality Inventory).

Данные о 513 больных алкоголизмом, проходивших лечение в госпитале в Индианаполисе, штат Индиана.

Тест содержал 566 вопросов (типа да/нет), которые суммировались по 13 диагностическим шкалам (например, шкала шизофрении, шкала истерии).

Кластеры:

- 1) эмоционально-неустойчивые личности;**
- 2) психоневротики с беспокойством/депрессией;**
- 3) психопатические личности;**
- 4) больные алкоголизмом, употребляющие наркотики и обладающие параноидальными чертами.**

Примеры ранних исследований

2. Бертон и Ромни (1975). Антропологическое исследование терминов, соответствующих статусу и роли индивидов в системе человеческих отношений.

58 наиболее общих терминов, среди которых типичными были: «художник», «босс», «друг», «человек», «владелец», «поэт» и «шпион».

Были выделены три шкалы: оценочная шкала, шкала иерархии, шкала профессиональной принадлежности.

Кластеры:

- 1) кластер, включающий семь терминов родства;**
- 2) кластер дружеских отношений;**
- 3) кластер терминов принадлежности к социальным группам;**
- 4) кластер управленческих ролей.**

Примеры ранних исследований

3. Филсингер, Фолкнер и Уорленд (1969). Социологическое исследование для создания классификации верующих.

Данные были собраны с помощью шкалы религиозности в форме вопросника (37 вопросов), выборка из 220 студентов.

Для анализа было выбрано решение, состоявшее из семи кластеров, соответствующих семи типам верующих:

- тип I – неверующие;**
- тип II – консерваторы;**
- тип III – нигилисты;**
- тип IV – умеренно религиозные;**
- тип V – крайне религиозные;**
- тип VI – ортодоксы;**
- тип VII – поклонники культа и обрядности.**

Этапы кластерного анализа:

- отбор выборки для кластеризации;**
- определение множества признаков, по которым будут оцениваться объекты в выборке;**
- вычисление значений той или иной меры сходства между объектами;**
- применение метода кластерного анализа для создания групп сходных объектов;**
- проверка достоверности результатов кластерного решения.**

Таблица 1 – Данные о захоронениях

N	Возраст	Пол	Статус	Местная керамика	Наконечники стрел	Обломки браслетов	Обработанные камни	Костяные иглы	Костяные шилья	Привозная керамика	Металлические изделия
1	С	М	Н	1	0	0	1	0	0	0	0
2	С	М	Н	0	0	0	1	0	0	0	0
3	С	М	Е	1	0	0	1	0	0	1	1
4	С	Ф	Н	1	0	1	0	0	0	0	0
5	С	Ф	Е	0	0	1	0	0	0	1	0
6	Т	Ф	Е	1	0	1	0	0	0	1	0
7	Т	М	Н	1	1	0	1	0	0	0	0
8	Т	М	Н	0	1	0	1	1	0	0	0
9	Т	М	Н	1	0	0	1	1	0	0	0
10	Т	М	Н	1	1	0	1	1	0	0	0
11	Т	М	Е	1	1	0	1	1	0	1	1
12	Т	Ф	Н	0	0	0	0	1	0	0	0
13	Т	Ф	Н	1	0	0	0	1	0	0	0
14	Т	Ф	Е	1	0	0	0	1	0	1	0
15	А	М	Н	1	1	0	1	1	0	0	0
16	А	М	Н	0	1	0	1	1	0	0	0
17	А	М	Н	1	1	0	1	0	0	0	0
18	А	М	Е	1	1	0	1	1	0	1	1
19	А	М	Е	1	0	0	1	0	0	1	0
20	А	Ф	Н	0	0	0	0	0	1	0	0
21	А	Ф	Н	1	0	0	0	0	1	0	0
22	А	Ф	Н	0	0	0	0	1	1	0	0
23	А	Ф	Н	1	0	0	0	0	0	0	0
24	А	Ф	Е	1	0	0	0	1	1	1	0 ⁹
25	А	Ф	Е	1	0	0	0	1	1	1	1

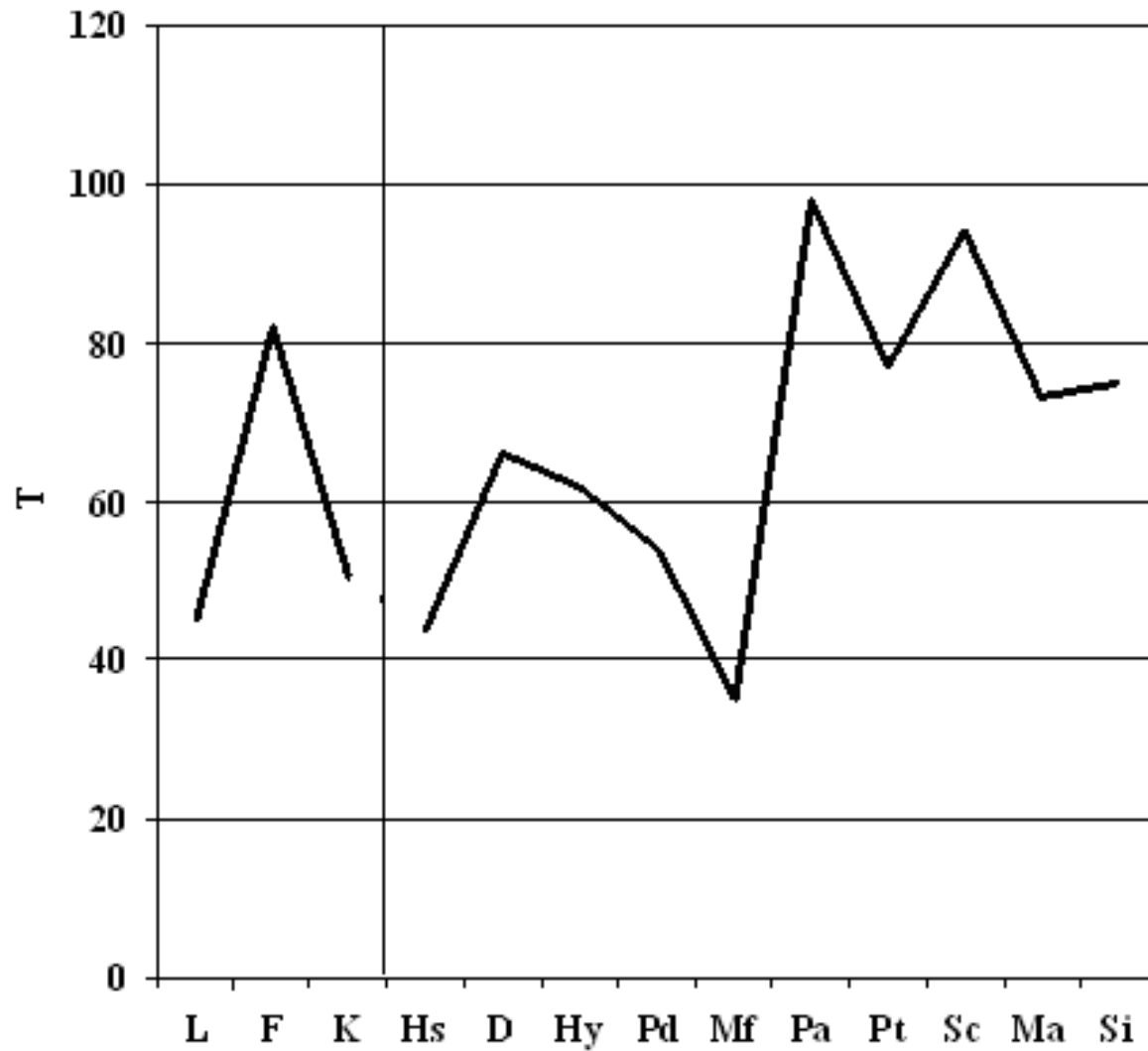


Рисунок 1 – Пример профиля данных ММРІ-теста

2.3.

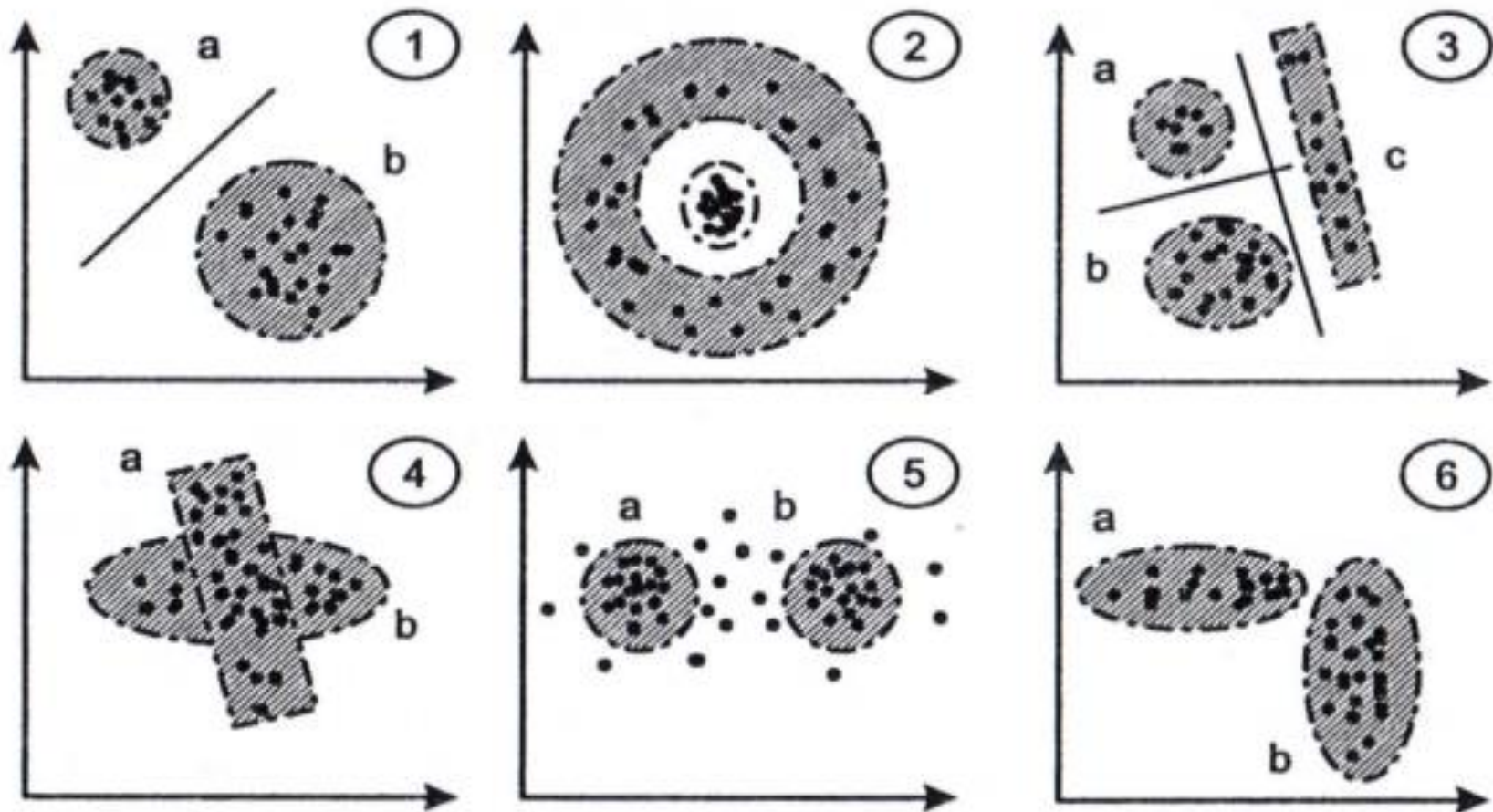


Рисунок 2 – Примеры кластеризации в двумерном пространстве

Предостережения общего характера

- 1. Многие методы кластерного анализа – довольно простые процедуры, которые, как правило, не имеют достаточного статистического обоснования.**
- 2. Методы кластерного анализа разрабатывались для многих научных 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.**

Рекомендации по составлению отчетов об исследованиях, использующих кластерный анализ

- 1) необходимо давать четкое описание метода кластеризации (это поможет избавиться от жаргона в публикациях);**
- 2) необходимо четко указывать, какая мера сходства была выбрана (или статистический критерий, если используется итеративный метод);**
- 3) необходимо указывать, какой программой пользовался исследователь;**
- 4) необходимо указывать процедуры, которые применялись при определении числа кластеров;**
- 5) необходимо привести убедительные свидетельства обоснованности решения кластерного анализа.**