

**Лабораторная работа** □ 1

ГРУППИРОВКА ГОРОДОВ С ПРИМЕНЕНИЕМ

КЛАСТЕРНОГО АНАЛИЗА

(курс «Статистика»)

**1. Цель работы**

Ознакомление с методами кластерного анализа на примере классификации населенных пунктов.

**2. Общие положения**

**Кластерный анализ** - многомерная статистическая процедура, которая упорядочивает объекты в сравнительно однородные группы (кластеры или таксоны), методы построения кластеров - это и есть кластерный анализ. Кластерный анализ применяется в экономике, маркетинге, социологии, медицине, биологии.

### Многомерные группировки

**Группировка** – это распределение единиц по группам в соответствии со следующим принципом: различия между единицами, отнесенными к одной группе, должны быть меньше, чем между единицами, отнесенными к разным группам.

Группировка производится по варьирующим характеристикам (признакам), по которым будет проводиться группировка так называемых *группировочных* признаков и их значений, отделяющих одну группу от другой (*интервал группировки*).

Если группированный признак один, то группировка называется простой или монистической. Если же группировочных признаков несколько, то различают иерархическую группу и многомерную группировку. Признаки в многомерных группировках равноправны. Например: предприятия группируются по эффективности и финансовому положению. Финансовое положение характеризует набор равноправных признаков. Методом многомерной группировки является кластерный анализ, по которому элементы совокупности объединяются в группы или классы двумя основными способами: либо по расстоянию между ними в производстве признаков, либо по тесноте корреляций связи.

### Основы кластерного анализа

Классификация объектов по осмысленным группам – кластеризация – является важной процедурой в области социально-экономических исследований. «Кластерный анализ» - это общее название множества вычислительных процедур, используемых при создании классификации. В результате работы с процедурами образуются «кластеры» или группы очень похожих объектов. Более точно кластерный метод – это многомерная статистическая процедура, выполняющая сбор данных, содержащих информацию о выборке объектов, и затем упорядочивающая объекты в сравнительно однородные группы.

Различные приложения кластерного анализа можно свести к четырём основным задачам:

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

Главная цель кластерного анализа - нахождение групп схожих объектов в выборке данных. Эти группы удобно называть кластерами. Не существует общепринятого или просто полезного определения термина «кластер». Несмотря на отсутствие определения, ясно, что кластеры обладают теми или иными свойствами, наиболее важными из которых являются плотность, дисперсия, форма и отделимость.

Каждая единица совокупности в кластерном анализе рассматривается как точка в заданном признаковом пространстве. Значение каждого из признаков у данной единицы служит ее координатой в этом «пространстве» по аналогии с координатами точки в нашем реальном трехмерном пространстве. Таким образом, *признаковое пространство* - это область варьирования всех признаков совокупности изучаемых явлений. Если мы

уподобим это пространство обычному пространству, имеющему евклидову метрику, то тем самым мы получим возможность измерять «расстояния» между точками признакового пространства. Эти расстояния называют *евклидовыми*.

Их вычисляют по тем же правилам, что и в обычной евклидовой геометрии. На плоскости, т.е. в двумерном пространстве, расстояние между точками А и В равно корню квадратному из суммы квадратов разностей координат этих точек по оси абсцисс и по оси ординат - на основании теоремы Пифагора (рис.1). Евклидово расстояние вычисляется по формуле:

$$. \quad (1)$$

Плотность – это свойство, которое позволяет определить кластер как скопление точек в пространстве данных, относительно плотное по сравнению с другими областями пространства, содержащими либо мало точек, либо не содержащими их вовсе.

Дисперсия характеризует степень рассеяния точек в пространстве относительно центра кластера. Кластер можно называть плотным, если все точки находятся от центра тяжести, и неплотным, если они разбросаны вокруг центра.

Форма кластер – это расположение точек в пространстве.

Разработанные кластерные методы образуют семь основных семейств:

- 1) иерархические агломеративные методы;
- 2) иерархические димизивные методы;
- 3) итеративные методы группировки;

- 4) методы поиска модальных значений плотности;
- 5) факторные методы;
- 6) методы сгущений;
- 7) методы, использующие теорию графов.

Наиболее известными семействами кластерных методов, используемыми в социальных науках, являются иерархические агломеративные, иерархические дивизивные и факторные.

К иерархическим агломеративным методам относятся: метод одиночной связи, метод полной связи, метод Уорда, метод средней связи.

Рассмотрим алгоритм средней связи более подробно.

### **Алгоритм «средней связи»**

Представим каждый  $i$ -ый объект выборки точкой в  $m$ -мерном пространстве признаков. При классификации объектов в качестве меры сходства обычно используются показатели расстояния или меры близости (как правило, евклидово расстояние). При классификации признаков в качестве меры сходства обычно выступают коэффициенты первой корреляции между признаками.

Для определения евклидова расстояния между  $k$ -й и  $l$ -й точками  $d_{kl}$  исходные данные  $x_{ij}$  нормируются так, как признаки имеют различную размерность.

(2)

(3)

где  $i=1, \dots, n$  - номер объекта;  $j=1, \dots, m$  - номер признака;

- среднее  $\bar{x}_j$  по всем точкам выборки;

- стандартное отклонение  $s_j$ .

Евклидово расстояние определить по формуле:

. (4)

Объединение объектов в класс, или объединение кластеров  $q$  и  $s$ , осуществляется по формуле  $d_{qs} = \min d_{kl}$ .

После объединения новый кластер выступает как отдельный объект. Матрица евклидовых расстояний пересчитывается по формуле:

$$D_{l,q+s} = m_q d_{kl} / (m_q + m_s) + m_s d_{ls} / (m_q + m_s), \quad (5)$$


где  $m_q$  – число объектов в классе  $q$ ;

$m_s$  – число объектов в кластере  $s$ .

Подобный алгоритм называется алгоритмом средней связи. Процесс объединения продолжается до получения одного кластера, при этом по результатам объединения можно выделить отдельные классы объектов и выявить объекты, не вошедшие ни в один из классов.

Результат оформления в виде дендрограммы (рис. 2), на которой по оси абсцисс откладываются номера объектов, а по оси ординат – евклидовы расстояния ( $D$ ).

При объединении объектов в классы координаты нового объекта получаются как координаты центра тяжести.

 ма евклидовых расстояний

Процесс объединения продолжается до получения одного кластера, при этом по результатам объединения можно выделить отдельные классы объектов и выявить объекты, не вошедшие ни в один из классов.

### **КЛАССИФИКАЦИЯ ГОРОДОВ**



Город - интеграл человеческой деятельности, средоточие разнообразных видов выполняемых им функций, которые определяют "лицо" города. Обычно функции подразделяются на экономические (промышленные, транспортные и т.д.) и неэкономические (административно-политические, культурные, научные, организационные).

Среди функций города выделяют также градообразующие и градообслуживающие. Деятельность города по обслуживанию внегородских связей - экономических, культурных, административных и других, направленных на выполнение основных задач данного города в районном или общегосударственном масштабе, называют градообразующей, а удовлетворяющую потребности самого города - градообслуживающей.

Города делятся на моно- и полифункциональные. Первые, особенно в России, - это преимущественно небольшие поселения курортного, научного профиля и т.д., а также города - районные центры.

Среди полифункциональных городов наиболее полным набором функций и мощным их развитием выделяются, помимо Москвы и Санкт-Петербурга, также межрегиональные (например, Самара, Новосибирск) и некоторые областные и краевые центры.

Функциональная структура влияет на многие черты города, в первую очередь на численность и динамику его населения. Как правило, многофункциональные города отличаются большой людностью и более быстрым ростом населения. Функциональная структура города влияет на половозрастной и профессиональный состав населения, его образовательный уровень, образ жизни и т.д.

Важными разделами изучения городов являются их классификация и типология. Под классификацией обычно понимают распределение городов по какому-либо признаку или их сочетанию. Типология - более высокий уровень обобщения, позволяющий дать комплексную характеристику городов и на этой основе наметить стратегию их развития; выделение совокупности городов по наиболее важным признакам.

Наиболее часто используется классификация городских поселений:

- 1) по величине (численности населения);
- 2) функциям;
- 3) степени участия в территориальном разделении труда;
- 4) генетическим признакам;
- 5) по экономико-географическому положению.

Классификация городов *по величине* имеет широкое применение. В зависимости от численности населения в официальных статистических источниках и градостроительной практике выделяют следующие группы городов:

малые - до 20 тыс. жителей;

средние – 20 - 100 тыс. жителей;

крупные – 100 - 500 тыс. жителей;

крупнейшие - 500 тыс. жителей и выше;

города-миллионеры.

При группировке городов по величине классификация в известной мере совпадает с типологией: можно говорить о малых, больших, городах-миллионерах как об определенных типах поселений.

Классификации и типологии городов *по функциям* отражают их наиболее важные особенности. По преобладанию и сочетанию различных функций можно выделить пять основных типов городов.

1. Многофункциональные, сочетающие административно-политические, культурные, экономические функции градообразующего значения с развитой промышленностью и транспортом. Обычно это крупные города, являющиеся важными районообразующими центрами с широкими и разнообразными связями.

2. Города с резким преобладанием промышленных и транспортных функций межрайонного значения. Схематично этот тип городов можно разделить на промышленные, транспортные, промышленно-транспортные.

3. Города с преобладанием административных, культурных и обслуживающих функций. Обычно это небольшие поселения - местные центры административных районов с развитыми организационно-хозяйственными видами деятельности.

4. Города-курорты, составляющие особую группу.

5. Быстро развивающиеся в последнее время города - научно-экспериментальные центры, или города науки.

Для объективного выделения функциональных типов городов по составу их функций пользуются обычно данными о распределении занятого населения по отраслям и видам деятельности, разделяя при этом работающих в градообразующей и градообслуживающей сферах деятельности.

*По степени участия в территориальном разделении труда* выделяют города, участвующие преимущественно в местных связях (они обслуживают небольшие территории и являются местными центрами), во внутрирайонном, межрайонном, международном разделении труда. Эти различия отражают масштабность градообразующих функций, выполняемых городом.

Для *генетической классификации* городов их разделяют по времени и причинам возникновения, а также учитывают степень сохранения различных исторических черт в

современной планировке и облике города. Такая классификация важна, например, при изучении внутренней географии города, для решения его планировочных проблем.

И наконец, наиболее сложная и пока наименее разработанная типология городов *по их экономико-географическому положению (ЭГП)*, которая позволяет определить направление будущего развития города путем оценки потенциальных возможностей района или какой-то его фокусной точки. В зависимости от ЭГП выделяют различные типы городов, например:

города, расположенные в узлах транспортных путей (Нижний Новгород, Новосибирск, Красноярск);

города, расположенные в крупных добывающих районах (Донецк, Рудный);

города в районах с развитой обрабатывающей промышленностью (Ярославль, Иваново, Серпухов);

города в районах интенсивного сельского хозяйства (Краснодар, Ставрополь).

Изучение функциональной структуры городов, их ЭГП, места в территориальном разделении труда, связей позволяет понять роль городов в качестве опорных элементов систем расселения разного типа и ранга, которые формируются в ходе эволюции урбанизации и развития различных видов деятельности.

Многомерная квалификация позволяет разбить города не по одному а по нескольким равноправным признакам.

### **3. Содержание домашней подготовки**

3.1. Ознакомиться с целью работы.

3.2. Ознакомиться с методами кластерного анализа и существующей классификацией городов.

3.3. Подготовить ответы на контрольные вопросы.

### **4. Порядок выполнения работы**

4.1. С помощью программы stat.exe, табл. 1 и табл. 2 исходных данных построить дендрограмму.

4.2. Используя дендрограмму, разбить объекты на выбранное количество кластеров.

4.3. Повторить п.2 и п.3 , выбрав другие информативные признаки.

4.4. Описать полученные кластеры и дать им наименование.

### **5. Содержание отчета о работе**

5.1. Теоретическая часть.

5.2. Дендрограммы по различным признакам.

5.3. Описание полученных кластеров.

5.4. Выводы.

## **6. Контрольные вопросы**

6.1. Основные этапы кластерного анализа.

6.2. Методы выбора информативных признаков.

6.3. Методы оценки сходства кластеров.

6.4. Семейства методов кластерного анализа для создания кластеров.

6.5. Понятие и функции города.

6.6. Классификация городских поселений.

---

## **Библиографический список**

1. Экономическая статистика. 2-ое изд. доп.: Учебник / Под ред. Ю.Н. Иванова. М.: ИФРА- М, 2000.

2. Терёхин В.И., Федотов Н.И. Применение регрессионных моделей в планировании и управлении производством: Учеб. пособие. Рязань: РРТИ, 1962.

3. Мандель А.Д., Миркин Б.Г. Кластерный анализ и смежные вопросы. Краткий обзор направлений // Автоматика. 1987. № 2. С. 72-82.

4. Елисеева И.И., Юзбашев М.М. Общая теория статистики. М.: Финансы и статистика, 2000.

Таблица 1

Список городов

( нумерация по вертикали таблицы 1)

№ города

Город

№ города

Город

№ города

Город

1

2

3

4

5

6



7

8

9

10

11

12

13

14

15

16

17

Владимир

Калуга

Кострома

Москва

Орел

Рязань

Ярославль

Белгород

Воронеж

Тамбов

Новгород

Псков

Санкт-Петербург

Архангельск

Вологда

Сыктывкар

Н.Новгород

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

Чебоксары

Астрахань

Волгоград

Казань

Пенза

Самара

Ульяновск

Элиста

Владикавказ

Грозный

Майкоп

Махачкала

Нальчик

Ростов-на-Дону

Черкеск

Екатеринбург

Ижевск

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

Оренбург

Пермь

Уфа

Челябинск

Барнаул

Кемерово

Новосибирск

Омск

Тюмень

Иркутск

Красноярск

Кызыл

Улан-Уде

Чита

Владивосток

Магадан

Список признаков



( нумерация по горизонтали таблицы 2 )

1 - численность населения на конец года, тыс. чел.,

2 - число промышленных предприятий,

3 - жилищный фонд города, на душу населения,

4 - число вузов,

5 - число больничных учреждений,

6 - число театров,

7 - число гостиниц и домов для приезжих,

8 - протяженность транспортной линий (автобусных), км.

В каждый вариант включается по 3 признака (номер первого признака соответствует номеру варианта).

## **Таблица 2**

### **Характеристика городов по признакам**

№ города

Признаки

1

2

3

4

5

6

7

8

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

339

316

282

8801

345

527

638

311

900

310

234

208

4467

420

289

224

1445

438

512

1007

1105

551

1244

648

93

306

401

58

72

87

880

76

101

110

65

134

62

55

50

464

86

78

51

162

78

101



157

177

108

154

90

34

80

73

17

17,1

16

16

15,1

15,5

16,7

15,5

15,2

15,2

15,6

16,3

17,7

15,9

17,6

16,7

15,7

14,8

15,8

16,9

15

14,7

15,4

15,9

15,1

15,2

15,2

2

1

2

76

2

4

5

3

9

3

9

1

41

3

9

2

10

3

4

8

11

4

10

3

1

4

3

17

16

13

220

15

20

31

17

39

18

14

16

130

25

14

16

70

23

32

48

60

24

44

26

12

15

13

2

1

2

54

3

3

3



2

4

2

1

2

17

2

3

2

5

5

3

4

6

2

4

2

2

4

4

5

3

4

172

4

8

5

5

14

9

4

4

62

9

6

11

31

4

8

20

17

9

15

14

6

2

7

197

260

250

5946

313

258

574

447

681

257

300

152

1890

286

235

364

1540

274

455

965

647

437

1150

701

140

376

631

Окончание табл.2

№ города

Признаки

1

2

3

4

5

6

7

8

28

29

30

31

32

33

34

35

36



37

38

39

40

41

42

43

44

45

46

47

48

49

50

153

334

241

1028

117

1374

647

557

1100

1097

1148

607

521

1446

1167

494

641

924

87

362

376

648

155

40

72

57

154

32

154

43

91

149

221

144

111

95

189

151

93

125

132

29

64

70

71

37

16,2

15,3

14

15,4

14,5

15,8

14,8

14,4

14,2

13,6

15,8

14,8

15,5

14,2

15

15,9

14,6

14,8

13,9

13,8

13

15,1

14,7

1

5

2

10

1

15

4

4

7

7

7



6

5

15

10

5

8

12

1

4

3

8

1

10

27

13

53

12

64

37

30

50

42

47

40

26

60

68

22

33

49

12

25

20

46

18

2

5

5

4

2

5

4

4

4

5

4

4

3

6

4

2

4

5

1

4

2

3

2

2

7

8

20

5

15

10

8

20

11

17

9

11

23

14

16

13

26

4

12

11

12

5

85

161

250

743

132

690

311

559

1095

747

627



270

703

1000

1636

698

775

825

166

437

349

445

226

**Лабораторная работа** □ 2

КЛАССИФИКАЦИЯ МУНИЦИПАЛЬНЫХ ОБРАЗОВАНИЙ

С ПРИМЕНЕНИЕМ КЛАСТЕРНОГО АНАЛИЗА

( курс «Экономика города»)

**1. Цель работы**

Проведение классификации муниципальных образований (МО) по социально-экономическим признакам.

**2. Общие положения**

Муниципальные образования, как и любые территориальные образования вообще, характеризуются достаточно большим объемом различных показателей. Наиболее часто

в социально-экономическом анализе используются следующие группы показателей.

*Показатели демографические:*

- численности проживающего населения;
- коэффициент рождаемости;
- коэффициент естественного прироста (убыли) населения и др.

*Уровень жизни населения:*

- доходы населения;
- жилищные условия населения;
- благоустройство жилищного фонда;
- средний размер месячной пенсии и др.

*Уровень социального обеспечения:*

- число дошкольных учреждений;

- численность учащихся государственных дневных общеобразовательных учреждений;
- численность врачей;
- число библиотек и др.

*Уровень экономического развития:*

- объем промышленной продукции;
- инвестиции в основной капитал;
- себестоимость производимой продукции и др.

**Уровень развития сельского хозяйства:**

- **распределение общей земельной площади;**
- **объем продукции растениеводства;**
- **объем продукции животноводства и др.**

**Финансовые показатели муниципалитета:**

- **поступление налогов и сборов в бюджетную систему России;**
- объем бюджета;
- удельный вес убыточных предприятий и др.

Муниципалитеты весьма существенно различаются между собой по перечисленным

показателям. Однако это не означает, что отдельные муниципалитеты не могут образовывать однородные группы (классы), в рамках которых некоторые показатели этих муниципалитетов достаточно близки по своим значениям.

Если удастся выделить такие однородные классы, то это в свою очередь означает, что к ним могут быть применены одинаковые методы управления. Например, класс депрессивных МО или класс МО с высоко развитым промышленным производством. Понятно, что к этим двум классам могут быть применены совершенно различные методы управления. Классификация же в данном случае может рассматриваться как логическая операция деления объема понятия, объекта или явления на некоторые однородные группы (классы).

### **3. Содержание домашней подготовки**

3.1. Ознакомиться с целью и общими положениями работы.

3.2. Ознакомиться с материалами областного комитета статистики, характеризующими социально-экономическое положение МО Рязанской области.

3.3. Ознакомиться с методами кластерного анализа, изложенными в первой части настоящей работы.

3.3. Подготовить ответы на контрольные вопросы.

### **4. Порядок выполнения работы**

4.1. На основании материалов областного комитета статистики, характеризующих социально-экономическое положение МО Рязанской области, самостоятельно выбрать от 3 до 8 показателей, по которым будет осуществляться классификация МО.

4.2. С помощью программы stat.exe построить дендрограмму.

4.3. Используя дендрограмму, разбить объекты на выбранное количество кластеров.

4.5. Описать полученные классы и дать им наименование.

### **5. Содержание отчета о работе**

5.1. Теоретическая часть.

5.2. Дендрограммы по различным признакам.

5.3. Описание полученных классов МО.

5.4. Выводы.

### **6. Контрольные вопросы**

6.1. Социально-экономические показатели, характеризующие МО.

6.2. Обоснование выбора информативных признаков.

6.3. Методы кластерного анализа.

6.4. Влияние выбранных показателей на формирование классов.

6.5. Классификация МО Рязанской области.

#### **Библиографический список.**

1. Социально-экономическое положение районов Рязанской области в 2001 году / Ряз. Обл. комитет государственной статистики, 2002.

2. Занодворов В.С., Занодворова А.В. Экономика города: Учебное пособие. М.: ИЧП «Издательство Магистр», 1998. 272 с.

3. Райзин Д.В. Классификация и кластер. М.: Мир, 1980.

4. Перфильев С.В., Терехин В.И. Анализ территориального неравенства и бюджетного регулирования территорий. М.: Экономика, 2000. 215 с.