

УДК
ББК

Джордж Грекусис

Методы и практика пространственного анализа. Описание, исследование и объяснение с использованием ГИС / пер. с англ. А. Н. Киселева. – М.: ДМК Пресс, 2021. – 500 с.: ил.

ISBN 978-5-97060-912-5

Книга знакомит читателей с приемами и методами пространственного анализа и пространственной статистики с помощью ГИС. Каждая глава представляет отдельную группу методов и метрик, объясняет, как интерпретировать результаты, и содержит практические примеры.

В числе рассматриваемых тем описание и отображение данных посредством исследовательского пространственного анализа данных, анализ географического распределения и точечных закономерностей, пространственная автокорреляция, пространственная кластеризация, географически взвешенная регрессия и регрессия методом наименьших квадратов, пространственная эконометрика. Примеры, объединяющие теорию и практику, связаны в единое исследование с использованием программного обеспечения и сопровождаются пояснительными иллюстрациями. Упражнения решаются дважды: сначала с помощью ArcGIS, а затем – GeoDa.

Книга будет полезна аспирантам и исследователям, занимающимся анализом геопрограммных данных через призму пространственного анализа, в том числе всем, кто использует геоинформационные системы в экологии, географии и социальных науках.

УДК
ББК

Все права защищены. Любая часть этой книги не может быть воспроизведена в какой бы то ни было форме и какими бы то ни было средствами без письменного разрешения владельцев авторских прав.

ISBN (англ.) 978-1-108-71293-4
ISBN (рус.) 978-5-97060-912-5

© 2020 George Grekousis
© Оформление, издание, перевод, ДМК Пресс, 2021

Оглавление

Вступление	12
------------------	----

Глава 1. Пространственное мышление: основные понятия пространственного анализа и концептуализация пространства	15
---	-----------

Теория.....	15
1.1. Введение: пространственный анализ.....	16
1.2. Основные определения	20
1.3. Пространственные данные: что в них особенного?	25
1.4. Концептуализация пространственных отношений	30
1.5. Измерение расстояния	32
1.5.1. Фиксированное расстояние (сфера влияния)	33
1.5.2. Расстояние затухания	35
1.6. Близость: матрица смежности	39
1.6.1. Близость полигонов	39
1.6.2. Матрица смежности.....	41
1.7. Взаимодействие.....	42
1.8. Окрестности и соседи	43
1.8.1. Метод k ближайших соседей	43
1.8.2. Метод пространственно-временного окна	45
1.8.3. Метод полигонов близости	46
1.8.4. Триангуляция Делоне и нерегулярные триангуляционные сети ..	47
1.9. Пространственные веса и стандартизация строк.....	48
1.10. Заключительные замечания к главе	50
Вопросы и ответы	51
Практика	54

Глава 2. Инструменты и статистики исследовательского анализа пространственных данных	73
---	-----------

Теория.....	73
2.1. Введение в исследовательский анализ пространственных данных, описательную статистику, статистику выводов и пространственную статистику	74
2.2. Простые инструменты ESDA и описательные статистики для визуализации пространственных данных (одномерные данные).....	78
2.2.1. Фоновые картограммы	78
2.2.2. Распределение частот и гистограммы.....	80
2.2.3. Оценка центра.....	84
2.2.4. Оценки формы	86

2.2.5. Оценки рассеяния/изменчивости – вариативность	87
2.2.6. Процентили, квартили и квантили.....	90
2.2.7. Выбросы	91
2.2.8. Коробчатая диаграмма	93
2.2.9. Нормальный график КК.....	96
2.3. Инструменты ESDA и описательные статистики для анализа двух и более переменных (двумерный анализ).....	97
2.3.1. Диаграмма рассеяния	97
2.3.2. Матрица диаграмм рассеяния.....	99
2.3.3. Ковариационная и дисперсионно-ковариационная матрицы....	100
2.3.4. Коэффициент корреляции.....	102
2.3.5. Парная корреляция.....	105
2.3.6. Нормальный график КК.....	106
2.4. Изменение масштаба данных	106
2.5. Статистика выводов и ее роль в пространственном анализе	110
2.5.1. Параметрические методы	111
2.5.2. Непараметрические методы	115
2.5.3. Доверительный интервал.....	116
2.5.4. Стандартная ошибка, стандартная ошибка среднего, стандартная ошибка пропорции и выборочного распределения.....	117
2.5.5. Проверки значимости, гипотезы, p -значение и z -оценка	119
2.6. Использование нормального распределения в географическом анализе	125
2.7. Заключительные замечания к главе	127
Практическая работа 2. Исследовательский анализ пространственных данных (ESDA): анализ и отображение данных	132
А. ArcGIS	133
В. GeoDa.....	152

Глава 3. Анализ географического распределения и структуры точечных закономерностей..... 162

Теория.....	162
3.1. Анализ географического распределения: центрография	163
3.1.1. Средний центр	163
3.1.2. Медианный центр.....	165
3.1.3. Центральный объект.....	166
3.1.4. Стандартное расстояние.....	168
3.1.5. Эллипс стандартного отклонения.....	169
3.1.6. Географические и пространственные выбросы.....	171
3.2. Анализ пространственных закономерностей: анализ структуры точечных закономерностей	175
3.2.1. Определения: пространственный процесс, полная пространственная случайность, эффекты первого и второго порядка	176
3.2.2. Пространственный процесс	178

3.3. Методы анализа структуры точечных закономерностей.....	181
3.3.1. Анализ ближайших соседей	182
3.3.2. К-функция Рипли и L-функция ее преобразования	184
3.4. Заключительные замечания к главе	190
Практическая работа 3. Пространственная статистика: оценка географических распределений	194
Глава 4. Пространственная автокорреляция.....	206
Теория.....	206
4.1. Пространственная автокорреляция.....	207
4.2. Глобальная пространственная автокорреляция	210
4.2.1. Индекс <i>I</i> Морана и диаграмма рассеяния	210
4.2.2. Индекс <i>C</i> Гири	216
4.2.3. Общая G-статистика	216
4.3. Инкрементальная пространственная автокорреляция.....	219
4.4. Локальная пространственная автокорреляция	222
4.4.1. Локальный индекс <i>I</i> Морана (анализ кластеров и выбросов).....	222
4.4.2. Оптимизированный анализ выбросов	226
4.4.4. Оптимизированный анализ горячих точек	230
4.5. Пространственно-временной корреляционный анализ.....	231
4.5.1. Двумерный индекс <i>I</i> Морана пространственно-временной корреляции	231
4.5.2. Дифференциальный индекс <i>I</i> Морана.....	232
4.5.3. Анализ возникновения горячих точек	233
4.6. Проблема множественного сравнения и пространственная зависимость.....	235
4.7. Заключительные замечания к главе	237
Практическая работа 4. Пространственная автокорреляция	241
Глава 5.....	258
Многомерные данные в географии: сокращение размерности данных и пространственная кластеризация.....	258
Теория.....	258
5.1. Анализ многомерных данных.....	259
Определения	259
Назначение	261
Обсуждение и практические рекомендации.....	261
Определение	263
Назначение	264
Интерпретация	264
Обсуждение и практические рекомендации (рабочий процесс и пространственные данные)	267
5.3. Факторный анализ	272

5.4. Многомерное масштабирование	273
5.5. Кластерный анализ	275
5.5.1 Иерархическая кластеризация.....	275
5.5.2. Алгоритм k средних (разделяющая кластеризация)	279
5.6. Регионализация	284
5.6.1. Метод SKATER.....	286
5.6.2. Метод REDCAP	289
5.7. Кластеризация на основе плотности: DBSCAN, HDBSCAN, OPTICS.....	290
5.8. Анализ сходства: косинусное сходство	291
5.9. Заключительные замечания к главе	293
Практическая работа 5. Многомерная статистика: кластеризация	297

Глава 6. Моделирование отношений:

регрессия и географически взвешенная регрессия..... 314

Теория.....	314
6.1. Простая линейная регрессия.....	315
6.1.1. Предположения в основе простой линейной регрессии.....	317
6.1.2. Обычный метод наименьших квадратов (для определения точки пересечения и наклона).....	318
6.2. Множественная линейная регрессия.....	319
6.2.1. Основы множественной регрессии	319
6.2.2. Переобучение модели: выбор количества переменных путем определения функциональной взаимосвязи	322
6.2.4. Выбросы и точки с большим плечом	323
6.2.5. Фиктивные переменные	324
6.2.6. Методы включения переменных в модель множественной линейной регрессии: объяснительный анализ; выявление причин и следствий	326
6.3. Оценка результатов линейной регрессии: метрики, критерии и диаграммы.....	328
6.3.1. Множественный R -квадрат.....	328
6.3.2. Дисперсия и коэффициент детерминации R -квадрат	329
6.3.3. Скорректированный R -квадрат	332
6.3.4. Прогнозный R -квадрат	334
6.3.5. Стандартная ошибка (отклонение) регрессии (или стандартная ошибка оценки).....	335
6.3.6. F -критерий общей значимости.....	336
6.3.7. t -статистика (критерий коэффициентов).....	337
6.3.8. Критерий Вальда (критерий коэффициентов).....	339
6.3.9. Стандартизированные коэффициенты (бета)	339
6.3.10. Остатки, диаграммы остатков и стандартизированных остатков	341
6.3.11. Факторы влияния: выбросы и наблюдения с высоким плечом.....	344

6.4. Предположения в основе множественной линейной регрессии: диагностика и исправление	346
6.6. Практический пример: простая и множественная линейная регрессия	354
6.7. Исследовательская регрессия	360
6.8. Географически взвешенная регрессия	364
6.8.1. Типы пространственных ядер	365
6.8.2. Ширина полосы	366
6.8.3. Интерпретация результатов GWR и практические рекомендации	367
6.9. Заключительные замечания к главе	371
Практическая работа 6. Обычный метод наименьших квадратов (OLS), исследовательская регрессия, географически взвешенная регрессия (GWR)	374
Глава 7. Пространственная эконометрика	400
Теория	400
7.1. Пространственная эконометрика	401
7.2. Пространственная зависимость: модели и критерии пространственной регрессии	402
7.2.1. Критерии пространственной зависимости	403
7.2.2. Выбор между моделями пространственного лага и пространственной ошибки	406
7.2.3. Методы оценки	407
7.3. Модель пространственного лага	408
7.3.1. Пространственный двухшаговый метод наименьших квадратов (S2SLS)	411
7.3.2. Максимальное правдоподобие	415
7.4. Модель пространственной ошибки	416
7.5. Пространственная фильтрация	418
7.6. Пространственная неоднородность: модели пространственной регрессии	420
7.7. Пространственные режимы	420
7.8. Заключительные замечания к главе	423
Практическая работа 7. Пространственная эконометрика	426
Список использованной литературы	447
Предметный указатель	457

Вступление

По мере того как пространственные данные становятся все более доступными, а службы, основанные на информации о местоположении (от приложений для смартфонов до систем мониторинга умных городов), становятся стандартом повседневного общения и взаимодействия людей, все больше исследователей, ученых и специалистов, далеких от географических дисциплин, осознают необходимость анализа данных с привязкой к географическому местоположению. Географические информационные системы позволяют присваивать местоположениям разнообразные атрибуты, однако пространственные данные скрывают в себе гораздо больше, чем гляцевое картографическое представление. Для извлечения этой информации необходим пространственный анализ, предлагающий методы и инструменты преобразования пространственных данных в знания и помогающий принимать более обоснованные решения. Таким образом, в широком спектре дисциплин существует огромная потребность в точном анализе данных с привязкой к географическим координатам (включая большие данные).

Для удовлетворения этой потребности и была написана данная вводная книга «Методы и практика пространственного анализа», описывающая приемы пространственного анализа и вычисления статистик с помощью ГИС. В книге представлены методы анализа пространственных и геоинформационных данных для решения различных задач с использованием подхода «описание – исследование – объяснение». Каждая глава посвящена одной теме, знакомит с соответствующей теоретической базой, объясняет, как интерпретировать результаты, и, наконец, предоставляет практические примеры.

Вот краткий список обсуждаемых тем:

- глава 1 «Пространственное мышление: основные понятия пространственного анализа и представления пространства» (приводит примеры, решаемые с помощью ArcGIS, GeoDa);
- глава 2 «Инструменты и статистики исследовательского анализа пространственных данных» (приводит примеры, решаемые с помощью ArcGIS, GeoDa);
- глава 3 «Анализ географического распределения и структуры точечных закономерностей» (приводит примеры, решаемые с помощью ArcGIS);
- глава 4 «Пространственная автокорреляция» (приводит примеры, решаемые с помощью ArcGIS, GeoDa);
- глава 5 «Многомерные данные в географии: сокращение размерности данных и пространственная кластеризация» (приводит примеры, решаемые с помощью ArcGIS, GeoDa, Matlab);
- глава 6 «Моделирование отношений: регрессия и географически взвешенная регрессия» (приводит примеры, решаемые с помощью ArcGIS, Matlab);
- глава 7 «Пространственная эконометрика» (приводит примеры, решаемые с помощью GeoDa Space).

Каждая глава в книге имеет теоретический («Теория») и практический («Практика») разделы. Такая организация принята с целью «обучения на практике». В разделах «Теория» подробно описываются идеи, методы и метрики, а в разделах «Практика» эти знания применяются для пошагового решения примеров с помощью ArcGIS и GeoDA. В двух главах также предлагаются сценарии Matlab.

ТЕОРИЯ

Методы и приемы пространственного анализа подробно и последовательно описаны в подразделах:

- «Определение»: каждый подраздел с этим названием начинается с определений методов, которые будут представлены далее. Это позволяет запоминать, где даются новые определения, идеи или метрики;
- «Назначение»: в этих подразделах дается начальное объяснение важности метода или метрики, а также описываются типы задач, для решения которых эти методы и метрики подходят лучше всего;
- «Интерпретация»: эти подразделы объясняют, как следует интерпретировать результаты, полученные с применением описываемых методов и метрик пространственного анализа, и идет чуть дальше простого представления данных или карт без дальнейшего критического обсуждения;
- «Обсуждение и практические рекомендации»: в этих подразделах обсуждаются плюсы и минусы каждого метода и каждой метрики. Здесь также даются ценные советы по практической реализации. Например, даются рекомендации по выбору подходящих значений параметров (статистик/метрик/инструментов), помогающие избежать необоснованного использования значений по умолчанию, предлагаемых программным обеспечением. Экспериментирование со значениями и настройками различных параметров позволяет лучше понять влияние каждого из них на конечный результат. Также в этих разделах упоминаются интересные тематические исследования;
- «Заключительные замечания»: в конце каждой главы представлен список важных замечаний и рекомендаций, в которых обобщаются основные теоретические положения;
- «Вопросы и ответы»: предлагают набор из 10 вопросов и ответов для самооценки читателем своих знаний.

ПРАКТИКА

Разделы «Практика» нацелены на получение практического опыта через решение хорошо спроектированных примеров. Все основные метрики, используемые в решении, описываются в разделе «Теория» этой же главы. Это позволяет читателям учиться выполнять пространственный анализ и составлять отчеты о результатах в ArcGIS или GeoDa, выполняя пошаговые инструкции. В данных разделах также подчеркивается, насколько важно критически подходить к интерпретации результатов, чтобы пространственный анализ приводил к извлечению знаний, помогающих принимать решения и осуществлять пространственное планирование.

На протяжении всей книги исследуется единственный рабочий пример. Благодаря этому читатели могут глубже исследовать разные подходы в пространственном анализе. Переходя от главы к главе, читатели будут получать все более полное представление об изучаемом регионе, и, таким образом, интерпретация результатов будет становиться для них все более простой и содержательной.

Вот общая структура разделов «Практика»:

- «Общий прогресс»: в начале каждого практического раздела читателю напоминает, на какой стадии проекта он находится. Также в начале каждого практического раздела приводится графическая схема, представляющая упражнения, используемые инструменты и ожидаемые результаты;
- «Сфера анализа»: описывает решаемую задачу;
- «Шаги»: приводит пошаговое руководство, описывающее порядок решения задачи, и сообщает результаты;
- «Интерпретация результатов»: описывает интерпретацию результатов с точки зрения пространственного анализа и применительно к рассматриваемой проблеме.

Эта книга является ценным ресурсом для широкого круга читателей и адресована не только географам. Она может принести пользу аналитикам, учителям, студентам различных специальностей и исследователям, работающим на стыке дисциплин, которые занимаются или собираются заняться анализом геопространственных данных. Здесь описываются все понятия, методы, метрики и приемы применения инструментов геопространственного анализа (ArcGIS, GeoDa и GeoDa Space), которые пригодятся при исследовании широкого спектра реальных задач социально-экономической сферы, анализе местоположений и планирования, оценке состояния городов и для эффективного содействия принятию решений в государственной политике. Никаких предварительных знаний в области пространственного анализа не требуется.

Я благодарен за помощь и советы многим ученым, и поскольку упущения и ошибки неизбежны, я буду очень признателен, если вы пришлете свои исправления или предложения, которые помогут сделать эту книгу еще лучше. Исправления будут опубликованы на сайте книги.

Глава 1

Пространственное мышление: основные понятия пространственного анализа и концептуализация пространства

ТЕОРИЯ

Цели обучения

Эта глава описывает:

- основные понятия, термины и определения, относящиеся к пространственному анализу;
- процедуру пространственного анализа в стиле «описание – исследование – объяснение»;
- причины, по которым пространственные данные считаются особенными, а именно с пространственной автокорреляцией, масштабом, проблемой изменяющегося масштаба, пространственной неоднородностью, пограничными эффектами и экологическими ошибками;
- почему концептуализация пространственных отношений чрезвычайно важна в пространственном анализе;
- подходы к концептуализации пространственных отношений;
- как при концептуализации пространства используются расстояние, близость/смежность, соседство, полигоны близости и пространственно-временное окно;
- матрицу пространственных весов, которая используется почти в каждой пространственной статистике/методике;
- представляет реальный проект с соответствующим набором данных, над которым мы будем работать на протяжении всей книги.

Тщательно изучив теоретический и практический разделы, вы узнаете:

- как организовать комплексную процедуру проведения пространственного анализа;
- как различать пространственные и непространственные данные;
- почему пространственные данные следует обрабатывать с применением новых методов (таких как пространственная статистика);
- почему важно применять методы концептуализации, соответствующие рассматриваемой задаче;
- об основных понятиях пространственного анализа, таких как расстояние, близость/смежность, соседство, полигоны близости и пространство-время;
- как описать процедуру пространственного анализа для проекта, рассматриваемого в этой книге;
- как получить визуальное представление данных проекта в ArcGIS и GeoDa.

1.1. ВВЕДЕНИЕ: ПРОСТРАНСТВЕННЫЙ АНАЛИЗ

«Мы верим только Богу. Все остальные должны предоставлять данные», – сказал У. Эдвардс Деминг (W. Edwards Deming, американский статистик и профессор, 1900–1993 гг.), потому что без данных почти ничего нельзя сделать. Подсчет объектов или людей и измерение их характеристик – это основа почти любого исследования. С *географическими информационными системами* (ГИС) появилась простая возможность связывать непространственные данные (например, уровень дохода, безработица, образование, пол) с пространственными (страны, города, районы, дома) и создавать большие *базы геоданных*. Фактически связь данных с местоположением делает их анализ более интригующим, и *пространственный анализ* с наукой о географии приобретает особое значение, потому что необработанные данные имеют небольшую ценность. Анализ данных с помощью методов и приемов пространственного анализа позволяет повысить их ценность, создавая информацию, а затем знания. В этом контексте пространственный анализ можно определить по-разному:

- **пространственный анализ** – это совокупность методов, статистик и приемов, объединяющих такие понятия, как местоположение, площадь, расстояние и взаимодействие для анализа, исследования и объяснения в географическом контексте закономерностей или особенностей в наблюдениях с пространственной привязкой, возникающих в результате процессов, действующих в пространстве;
- **пространственный анализ** – это количественное исследование явлений, возникающих в пространстве (Anselin 1989, стр. 2);
- **пространственный анализ** изучает, «как физическая среда и деятельность человека меняются в пространстве – иными словами, как эти действия меняются с расстоянием от исходных местоположений или интересующих объектов» (Wang 2014, стр. 27);
- **пространственный анализ** – это «процесс, с помощью которого мы превращаем исходные данные в полезную информацию в погоне за научными открытиями или более эффективными решениями» (Longley et al. 2011);

- **пространственный анализ (данных)** – это «совокупность методов для поиска закономерностей, выявления аномалий или проверки гипотез и теорий на основе пространственных данных» (Goodchild 2008, стр. 200);
- **пространственный анализ** – это широкий термин, включающий (а) манипулирование пространственными данными с помощью географических информационных систем (ГИС), (б) анализ пространственных данных описательным и исследовательским способами, (в) пространственную статистику, использующую статистические процедуры для исследования возможности делать выводы, и (д) пространственное моделирование, включая построение моделей для определения взаимосвязей и прогнозирования результатов в пространственном контексте (O’Sullivan & Unwin 2010, стр. 2).

Назначение пространственного анализа?

Понятия, методы и теория пространственного анализа вносят ценный вклад в изучение:

- **социальных систем:** методы пространственного анализа можно применять для изучения взаимодействий между людьми в социальном, экономическом и политическом контекстах, потому что пространство является основным слоем для всех действий и *взаимосвязей* между людьми;
- **среды:** методы пространственного анализа можно применять при исследовании природных явлений и опасностей, связанных с изменениями климата, для управления природными ресурсами, защиты окружающей среды и устойчивого развития;
- **экономики:** методы пространственного анализа можно применять для анализа, картирования и моделирования взаимосвязей между людьми и различными параметрами экономической жизни.

Главное преимущество пространственного анализа – способность выявлять закономерности в данных, которые прежде не только не определялись, но даже не наблюдались. Например, с помощью методов пространственного анализа можно идентифицировать *кластеризацию* эпидемии и разработать механизмы для предотвращения ее распространения или даже устранения (Bivand et al. 2008). В этом отношении пространственный анализ позволяет вырабатывать лучшие *решения* и успешнее осуществлять *пространственное планирование* (Grekousis 2019).

Всего существует четыре типа пространственного анализа:

- **пространственный анализ структуры точечных закономерностей:** совокупность точек данных анализируется с целью определить ее состояние: *сгруппированное, рассредоточенное, случайное*. Рассмотрим, например, пространственное распределение приступов инсульта в исследуемой области. Сгруппированы ли они в определенном регионе или случайно распределены по всей области? Затем пространственный анализ продолжается выявлением, например, движущих факторов, объясняющих эту группировку (наличие поблизости промышленных зон и связанного с ними загрязнения). Анализ структуры точечных законо-

мерностей также включает центрографию – определение совокупности пространственных статистик, используемых для определения центра, разброса и направленности тренда точечных закономерностей. В этом типе анализа данные обычно относятся ко всей генеральной совокупности, а не к выборке;

- **пространственный анализ площадных данных:** данные объединяются в заранее определенные зоны (например, участки переписи, участки отделений связи и т. д.) и затем анализируются связи и взаимодействия между зонами. Например, группируются ли люди с высоким или низким доходом вокруг определенных регионов или они расселяются случайно? Центральными понятиями в этом типе анализа являются *пространственная зависимость, пространственная неоднородность, пространственная автокорреляция, концептуализация пространства* (через матрицу пространственных весов) и *регионализация* (пространственная группировка, или кластеризация);
- **геостатистический анализ данных (анализ непрерывных данных):** геостатистический анализ – это раздел статистики, анализирующий и моделирующий *переменные непрерывного поля* (O’Sullivan & Unwin 2010, стр. 115). В этом отношении геостатистические данные представляют собой совокупность выборочных наблюдений за непрерывным явлением. Используя различные геостатистические подходы (например, интерполяцию), можно рассчитать значения для всей поверхности. Например, загрязнение контролируется ограниченной сетью пунктов наблюдения. Чтобы оценить загрязнение в каждой отдельной точке, можно применить методы интерполяции. Геостатистический анализ в этой книге не рассматривается;
- **пространственное моделирование:** пространственное моделирование в основном занимается проблемой возможности моделирования пространственной зависимости, пространственной автокорреляции и пространственной неоднородности для получения надежных пространственных предсказаний. Пространственное моделирование можно использовать, например, для моделирования связи стоимости дома с его местоположением. Ключевыми методами пространственного моделирования являются пространственная регрессия и пространственная эконометрика.

Последовательность пространственного анализа

Пространственный анализ – обширная дисциплина с большим разнообразием методов, подходов и приемов, поэтому необходимо четко понимать, как проводить такой анализ. В этой книге представлена уникальная последовательность пространственного анализа, следующая структуре описание – исследование – объяснение, которая помогает отвечать на вопросы что, где и почему, соответственно (см. рис. 1.1).

Шаг А: описание (что). Это первый шаг в процессе пространственного анализа. Он описывает набор данных с помощью *описательной статистики*. Описательная статистика используется для обобщения характеристик данных и помогает понять особенности распределения значений, их диапазон и наличие выбросов.

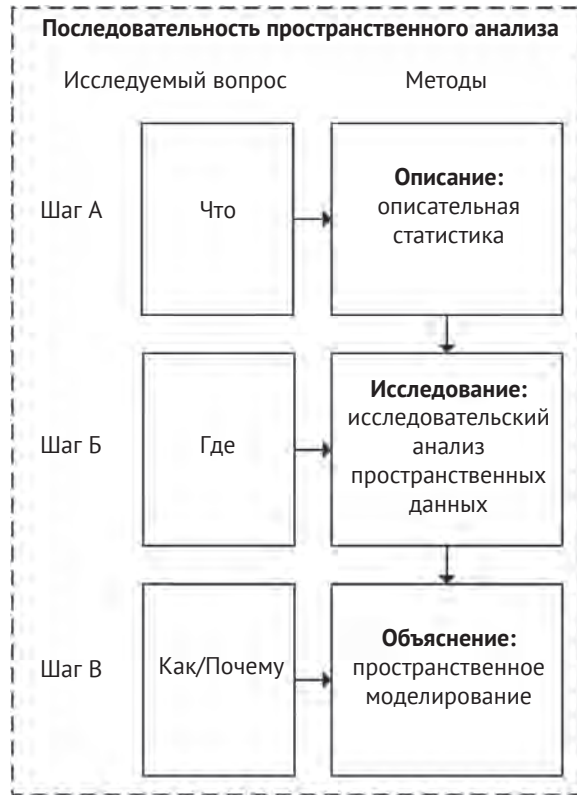


Рис. 1.1. Последовательность пространственного анализа

Обычно этот шаг отвечает на вопрос «что?», например каков средний доход в районе или какова доля населения, живущего за чертой бедности. На этом шаге достигается начальное понимание набора данных и его конкретных характеристик. Однако если данные были собраны неаккуратно, никакой анализ не сможет дать точных и полезных результатов. Поэтому любая совокупность данных должна проверяться на согласованность и точность, прежде чем углубиться в анализ. Не используйте совокупности данных, которые не сопровождаются подробным описанием методов их сбора и оценкой точности (всегда в своих исследованиях указывайте, какая база данных использовалась, какие методы выборки данных применялись и какие средства контроля качества были задействованы).

Шаг Б: исследование (где). На втором этапе применяется *исследовательский анализ пространственных данных* (Exploratory Spatial Data Analysis, ESDA) для изучения данных, обнаружения выбросов, проверки основных предположений и выявления в них тенденций и ассоциаций, таких как наличие пространственной автокорреляции или пространственной кластеризации. На этом этапе мы в основном отвечаем на вопрос «где?», например: какие районы характеризуются низкими и высокими доходами, существует ли какая-либо пространственная группировка районов по распределению дохода на душу населения, где они располагаются и где в городе находятся горячие точки преступности?

Шаг В: объяснение (почему/как). На последнем этапе применяется статистический анализ, объясняющий причины и следствия с помощью моделей. На этом этапе мы пытаемся ответить на вопросы «почему?» и «как?». Этот анализ не просто идентифицирует ассоциации, но также пытается выявить (а) отношения, объясняющие причины происходящего, и (б) движущие силы изменений. В число типичных вопросов в географическом контексте входят: почему криминальные события сосредоточены в определенном районе? Есть ли какая-то связь с конкретными социально-экономическими характеристиками этого района? Почему доход на душу населения связан с местоположением, и как доход связан с размером домовладения? Каковы основные причины повышения уровня моря и как рост населения влияет на урбанизацию? В анализе этого типа и в контексте этой книги мы рассматриваем как независимые, так и зависимые переменные. Зависимая переменная (эффект) – это явление/состояние/переменная, которые мы пытаемся объяснить. Например, если в результате анализа делается вывод, что прирост населения (независимая переменная, фактор) объясняет x % изменения уровня урбанизации (зависимая переменная, эффект), то устанавливается связь (отношение), которая объясняет степень влияния фактора на эффект. После построения модели, объясняющей причины происходящего, ее можно использовать для прогнозов. Этот шаг идет намного дальше, чем шаги А и Б, которые в основном касаются того, «что происходит» или «где это происходит». На этом этапе анализа используются пространственная регрессия и пространственная эконометрика, которые будут описаны далее в данной книге. С точки зрения пространственного анализа на этом этапе также можно попробовать найти ответы на некоторые дополнительные вопросы: можно ли чему-то научиться, применяя используемую методологию к данному набору данных? Были ли созданы новые знания в ходе анализа? Каким должен быть следующий шаг? В каком направлении следует продолжить исследования? Знания, полученные по завершении пространственного анализа, упрощают процессы принятия решений и пространственного планирования.

1.2. ОСНОВНЫЕ ОПРЕДЕЛЕНИЯ

Врезка 1.1. В предыдущем разделе было упомянуто более 20 терминов (выделены курсивом), связанных с пространственным анализом, пространственной статистикой и *пространственным мышлением* (еще один термин). Некоторые термины просты и понятны, другие – совершенно новые, а третьи – довольно расплывчатые. В этом разделе мы сформируем общий словарь и рассмотрим некоторые ключевые определения. В разных книгах определения, термины и формулы обычно различаются, что сбивает с толку не только неспециалистов, но и ученых, и вызывает множество недоразумений. Эта путаница также затрудняет использование программного обеспечения для статистического, геоинформационного и пространственного анализа, особенно это относится к уравнениям и формулам. В данной книге представлены наиболее часто используемые названия и символы для обозначения терминов и статистик.

Определения

Пространственная статистика использует статистические методы для анализа пространственных данных, количественной оценки пространственного процесса, выявления скрытых закономерностей или неожиданных тенденций и моделирования данных в географическом контексте. Пространственная статистика может рассматриваться как часть исследовательского анализа пространственных данных (ESDA), пространственной эконометрики и анализа дистанционного зондирования (Fischer & Getis 2010, стр. 4). Она в значительной степени основана на статистических выводах и проверке гипотез в отношении географических закономерностей, помогая улучшить модель явления, изменяющегося в пространстве (Fischer & Getis 2010, стр. 4). Пространственная статистика количественно определяет и представляет то, что интуитивно видит человеческий глаз и разум при чтении карты, изображающей пространственное расположение, распределение или тенденции (Scott & Janikas 2010, стр. 27; см. также главу 2).

Пространственное моделирование занимается созданием моделей, объясняющих или предсказывающих пространственные результаты (O'Sullivan & Unwin 2010, стр. 3).

Геопространственный анализ – это совокупность методов, приемов и моделей пространственного анализа, объединенных в географические информационные системы (ГИС; de Smith et al. 2018). Геопространственный анализ подкрепляется богатыми возможностями ГИС и используется для разработки новых моделей или интеграции существующих в среду ГИС. Этот термин также часто употребляется как синоним «пространственного анализа» (de Smith et al. 2018); однако, строго говоря, пространственный анализ является частью геопространственного анализа (см. врезку 1.2).

Врезка 1.2. Иногда сложно различить термины «географический», «пространственный» и «геопространственный». Эти термины были определены многими экспертами в разных научных контекстах. Приведенные здесь определения не являются исчерпывающими; они лишь служат основой для общей терминологии. Даже в географической науке эти термины могут пересекаться и различаться с большим трудом. Термин «географический» относится к местоположению на поверхности Земли в сочетании с представлениями некоторого типа. Термин «пространственный» относится не только к поверхности Земли; он охватывает местоположение в сочетании с дополнительными атрибутивными данными. Термин «геопространственный» больше ориентирован на вычисления и относится к информации, основанной на пространственных данных и моделях, и сочетает географический анализ с пространственным анализом и моделированием.

Пространственные данные относятся к пространственным объектам, характеризующимся геометрическими параметрами и пространственной привязкой (координаты и система координат), которые также имеют другие непространственные атрибуты (см. рис. 1.2; Vivand et al. 2008, стр. 7). Например, город можно описать численностью его населения, уровнем безработицы, доходом на душу населения или среднемесячной температурой. Связав эти данные

с местоположением через пространственные объекты (например, городские округа), мы получаем пространственные данные. Диапазон атрибутов, которые присоединяются к пространственным объектам, зависит от изучаемой проблемы и доступности данных (например, результатов переписи). Изображения с географической привязкой также считаются пространственными данными.

Существует два основных способа представления географических объектов и мира как такового в цифровом виде: объектный и полевой (Haining 2010, стр. 199).

Представление объекта – это представление, описывающее мир с использованием уникальных пространственных объектов, которые привязаны к определенному месту посредством координат. Пространственные объекты в объектном представлении моделируются как точки, линии или полигоны (также называемые признаками). Эта модель данных называется векторной моделью данных (O’Sullivan & Unwin 2010, стр. 6). Объектное представление и векторная модель могут использоваться для отображения, например, демографических или социально-экономических данных. При изменении масштаба анализа пространственные объекты могут быть представлены иначе.

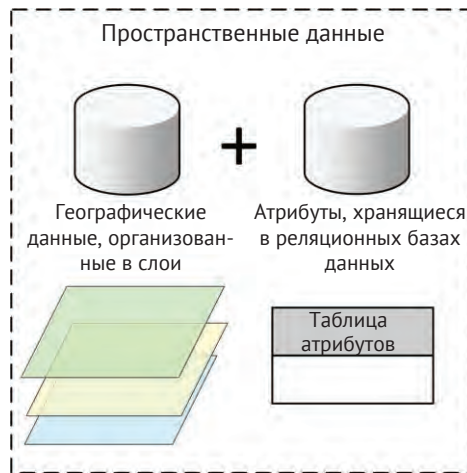


Рис. 1.2. Пространственные объекты в совокупности с атрибутами образуют пространственные данные

Например, при исследовании в масштабе страны город может быть представлен точкой, а на более локальном уровне – как полигон.

Полевое представление – это представление, описывающее мир как поверхность с постоянно меняющимися свойствами (O’Sullivan & Unwin 2010, стр. 7). Полевое представление мира больше подходит для изображения непрерывных явлений и свойств (например, температуры, загрязненности, типа почвенного покрова, высоты). Полевое представление записывается с использованием растровой модели данных. В этой модели прямоугольные ячейки (называемые пикселями), организованные в виде регулярной решетки, отображают географическое изменение исследуемого свойства. Другой способ моделирования полей – использование нерегулярных триангуляционных сетей (Triangulated Irregular Networks, TIN). Пространственные данные можно преобразовывать

из одной модели в другую (например, векторную модель можно преобразовать в растровую, и наоборот) в соответствии с потребностями исследования.

Переменная – это любая характеристика объекта или человека. Например, возраст, рост и вес – все это переменные, характеризующие людей, животных или объекты.

Атрибуты – это информация, которую несут пространственные данные. Они хранятся в виде столбцов в таблице ГИС. Атрибут эквивалентен переменной в классической статистике и считается более предпочтительным термином в ГИС-анализе, хотя эти термины могут использоваться как взаимозаменяемые. Атрибутом может быть численность населения городского округа или годовой доход на душу населения в округе.

Данные образуются в процессе измерения характеристик объектов или людей.

Значение – это результат измерения (или отклик) характеристики. В статистике для описания значения переменной также используется термин **оценка**.

Выброс – это необычное, очень маленькое или очень большое значение по сравнению с остальными значениями.

Набор данных – это коллекция переменных объектов любого типа. Обычно набор пространственных данных имеет табличный формат, в котором столбцы соответствуют атрибутам, а строки – пространственным объектам.

Генеральная совокупность – это полная коллекция наблюдений/объектов/измерений, для которой запрашивается информация.

Выборка – это часть генеральной совокупности.

Уровень измерения переменной описывает, как ее значения расположены по отношению друг к другу (de Vaus 2002, стр. 40). Переменные/атрибуты группируются по трем уровням измерения: номинальному, порядковому и интервальному или относительному (Haining 2010, стр. 201; см. табл. 1.1).

- **Номинальные переменные** – это переменные, значения которых нельзя упорядочить. Например, переменная, отражающая расу, может иметь значения белая = 1, азиатская = 2, латиноамериканская = 3. Это номинальная переменная, потому что значения «1, 2, 3» не определяют порядок, а используются только для обозначения различных категорий. Нельзя сложить значения двух разных объектов, например «1 + 3 = 4», потому что «4» не отражает какого-либо значимого значения. Другой пример номинальных переменных: «название города» (например, Афины, Пекин, Нью-Йорк) или «название земного покрова» (например, лес, город, водоем). Атрибуты этого типа представляют описательную информацию и могут использоваться для обозначения полигонов на карте. К ним применимы такие операторы, как «равно» или «не равно» (=, ≠).
- **Порядковые переменные** – это переменные, которые можно упорядочивать, но численные различия которых не имеют смысла и не могут быть вычислены. Например, переменная «студент» может принимать следующие значения: «отличник» = 1, «хорошист» = 2, «посредственный» = 3. Категории можно упорядочить сверху вниз (или наоборот), но в вычитании категорий нет смысла («отличник» – «хорошист» = –1). К ним можно применять операторы «равно», «не равно», «больше чем» и «меньше чем» (=, ≠, >, <). Атрибуты пространственного объекта, измеренные на номинальном или порядковом уровне, также называются «категориальными».

- **Интервальные и относительные переменные** (также называемые «числовыми») – это переменные, для которых каждое наблюдение можно значимо выразить в числовом виде. Числа используются не только как метки, но и для вычисления статистик (например, среднего). Если числовые значения переменной ограничены определенными категориями, такая переменная называется дискретной, или интервальной. Интервальный уровень – это класс относительного уровня. Категории в интервальных измерениях определяются фиксированными расстояниями. Интервальные данные позволяют выполнять операции сложения и вычитания (Haining 2010, стр. 201). Однако интервальные переменные не сохраняют отношения (O’Sullivan & Unwin 2003, стр. 13). Дихотомические переменные (например, «пол», которая может иметь только два значения «мужской = 1» и «женский = 0» или наоборот) тоже могут рассматриваться как дискретные интервального уровня. В этом случае ноль означает отсутствие чего-либо. Если набор возможных значений не ограничен определенными категориями между низкими и высокими значениями, тогда переменная является непрерывной числовой (или «относительной»). Относительные переменные имеют значащий ноль. С относительными переменными можно использовать все операторы (=, ≠, >, <, +, -, ×, /).

Таблица 1.1. Уровни измерений для моделей с разной структурой данных (векторные/растровые) и примеры для каждого типа данных. В скобках указаны применимые логические и арифметические операции. Многие статистические процедуры и методы можно использовать не на всех уровнях измерений, потому что на разных уровнях применяются разные логические и арифметические операции. Например, бинарная логистическая регрессия может применяться к дихотомическим зависимым переменным и не может к относительным переменным. Уровень измерения определяет набор статистических процедур, которые можно будет использовать в дальнейшем анализе. Со статистической точки зрения, для анализа относительных переменных можно использовать больше методов, чем для номинальных и порядковых переменных; поэтому относительные переменные выглядят предпочтительнее (de Vaus 2002, стр. 43)

Уровень измерения	Векторная модель данных (объектное представление)			Растровая модель данных (полевое представление)
	Точка	Линия	Полигон	Пиксель
Номинальный (=, ≠)	Название населенного пункта	Название дороги	Округ (населенный пункт)	Тип земного покрова
Порядковый (=, ≠, >, <)	Ранг города по привлекательности для проживания	Тип дороги (шоссе, автострада)	Классификация округов (населенных пунктов) по уровню образования	Разновидности лесов
Интервальный (=, ≠, >, <, +, -)	Уровень бедности	Ширина дороги	Уровень бедности в населенном пункте	Температура на поверхности земли
Относительный (=, ≠, >, <, +, -)	Численность населения	Объем перевозок	Данные о населенном пункте: количество жителей, доход на душу населения	Загрязненность воздуха частицами PM2.5

Для анализа переменных/атрибутов можно использовать статистические методы. Существует три основных раздела классической статистики: **описательная статистика**, **статистика выводов** и **пояснительная статистика** (Linneman 2011 стр. 20). Далее в книге мы рассмотрим все эти разделы.

Описательная статистика – это набор статистических процедур, обобщающих основные характеристики данного распределения. Описательная статистика обычно обобщает конкретную выборку и не подходит для выводов относительно всей совокупности (если у нас под рукой нет всей совокупности). В результате они разрабатываются не на основе теории вероятностей, как статистика выводов. В этом смысле результаты описательной статистики применимы только к конкретному набору данных, для которого они были получены. Описательная статистика использует таблицы, графики и простые статистические процедуры (Linneman 2011, стр. 21).

Статистика выводов – это раздел статистики, в котором производится анализ выборок, чтобы сделать выводы для всей генеральной совокупности. К числу типичных подходов, используемых в статистике выводов, относятся: проверка значимости (проверка гипотез), определение доверительного интервала и байесовский вывод.

Пояснительная статистика – это раздел статистики, в котором используются методы и приемы, помогающие выявить взаимосвязи между переменными и возможные «объяснения» причинно-следственных связей. В статистике этого типа переменные рассматриваются как зависимые или независимые (Linneman 2011, стр. 21). Зависимая переменная – это то, что мы пытаемся объяснить с помощью набора независимых переменных. Обычно в пояснительной статистике используется регрессионный анализ.

1.3. ПРОСТРАНСТВЕННЫЕ ДАННЫЕ: ЧТО В НИХ ОСОБЕННОГО?

Представьте, что риелтор хранит в своем мобильном телефоне контактные данные своих клиентов. Эти данные хранятся в памяти телефона бессистемно. Если эти данные связать с местоположениями (координатами или адресами), то их можно будет преобразовать в пространственные данные. То есть каждый контакт будет привязан к одному пространственному объекту (например, точке, обозначающей домашний адрес каждого клиента, несущей дополнительную информацию – атрибуты, – такую как номер телефона, имя, дата рождения и адрес электронной почты). Преобразование данных в пространственные данные дает гораздо больше, чем просто гляцевую визуализацию. Оно позволяет выполнять углубленный геопропространственный анализ и выполнять расширенные пространственные запросы. Например: какой клиент ближе к определенной точке, где проживает большинство клиентов, где проживают клиенты, которые платят больше всего, как лучше проехать к клиенту, какой процент клиентов живет в радиусе 1 или 2 километров от определенного местоположения, группируются ли клиенты вокруг определенных районов, и каков социально-экономический профиль этих районов? На некоторые вопросы невозможно быстро ответить, не связав контакты с местоположениями (представьте, что у вас есть тысячи клиентов), а на другие вообще не получится ответить без этого.

Применение методов пространственного анализа к пространственным данным помогает выявить скрытую информацию. Чтобы найти эти информационные сокровища, пространственные данные анализируются с использованием различных приемов и методов описательной, исследовательской и пояснительной статистики. Следующие главы посвящены этим обширным классам пространственных методов. Однако отличия между пространственными и непостоянными данными не всегда очевидны, и не всегда ясно, почему мы должны применять новые методы для их анализа. Давайте посмотрим, что особенного в пространственных данных (см. также врезку 1.3).

Во-первых, пространственный анализ подразумевает концентрацию внимания на множестве географических параметров, таких как местоположение, расстояние, форма, площадь, соседство, смежность и взаимосвязь. Включение географических параметров в пространственный анализ отличает методы пространственного анализа от методов классического (непространственного) анализа (Anselin 1989, стр. 2).

Во-вторых, многие из традиционных статистических подходов, используемых для анализа непостоянных данных, неприменимы напрямую к пространственным данным, потому что исследование пространственных данных сопряжено со следующими проблемами.

Наличие **пространственной автокорреляции/зависимости**. Многие статистические критерии, касающиеся непостоянных данных, основаны на гипотезе о случайном формировании выборок и независимости наблюдений (O'Sullivan & Unwin 2010, стр. 34). Однако в отношении выборок пространственных данных эта гипотеза обычно нарушается. Это явление описывается как «пространственная зависимость». Первый закон географии Тоблера объясняет это основное свойство пространственных данных, утверждая, что «любой географический объект связан со всеми остальными географическими объектами, но близкие объекты связаны сильнее, чем далекие» (Tobler 1979). Пространственная зависимость тесно связана с пространственной автокорреляцией. Пространственная автокорреляция – это степень пространственной зависимости, ассоциации или корреляции между наблюдаемыми значениями атрибутов соседствующих пространственных объектов (Dall'era 2009). Наличие пространственной автокорреляции в наборе пространственных данных не обязательно является проблемой. Если бы между объектами не было пространственной зависимости, то не было бы никакого смысла в пространственном анализе. Иначе говоря, если бы пространственные характеристики не имели значения, то и географический анализ не представлял бы интереса. Пространственная автокорреляция существует во многих географических задачах, поэтому для их решения необходимо использовать специальные инструменты. Более подробно о пространственной автокорреляции мы поговорим в главе 4.

Концептуализация пространства. Области в пространстве не изолированы друг от друга. Между соседними или удаленными областями существуют социальные, экономические и демографические связи. Эти связи/отношения имеют пространственное измерение, поскольку они разворачиваются в пространстве, поэтому местоположение и расстояние имеют значение. Как уже упоминалось, согласно первому закону географии Тоблера «близкие объекты связаны сильнее, чем далекие». Чтобы расшифровать это правило, нам нужно

определить, на каком расстоянии друг от друга должны находиться объекты, чтобы называться «близкими» и «далекими», и как объект должен изображаться для вычисления расстояния (например, как точка, линия или полигон). Приступая к анализу пространственных данных, мы должны математически определить, насколько близко находится «близкий» объект, как определяется «смежность», каков размер близлежащих окрестностей и как интегрировать «пространственно-временной» анализ. Это означает, что мы должны установить ряд географических параметров, чтобы с их помощью определить пространственные отношения между объектами. Это называется **концептуализацией пространственных отношений**, и это основное отличие между методами анализа пространственных и непространственных данных. Подробнее об этом мы поговорим в следующем разделе данной главы.

Выбор географического масштаба. До начала географического/пространственного анализа следует выбрать подходящий географический масштаб, потому что он напрямую влияет на порядок сбора данных, выбор модели данных, методов их анализа и способов формирования выводов. Например, в масштабе страны город может быть представлен как точка, в масштабе региона – как полигон, а в более локальном масштабе – как набор полигонов (например, городских округов). Для более детального анализа можно опуститься до уровня округа (обычно это территория с населением 1000 человек).

Выбор масштаба анализа данных. Масштаб анализа тесно связан с географическим масштабом, но это разные понятия (хотя часто они используются как синонимы). Масштаб анализа определяет размер и форму области, по которой вычисляются пространственные статистики после установки географического масштаба. По сути, это уровень представления пространственных явлений, тесно связанный с рассматриваемой задачей. Масштаб анализа также можно представить как географическую протяженность, на которой анализируется пространственная и временная изменчивость посредством построения соответствующей матрицы пространственных весов. Географический масштаб же определяет масштаб данных (например, масштаб страны [1:100 000] или масштаб города [1:10 000]).

Масштаб анализа может отличаться при исследовании разных переменных в одном и том же наборе данных. То есть масштаб анализа зависит от пространственного распределения значений атрибутов, тогда как географический масштаб часто остается неизменным для всего набора данных. Как правило, большие расстояния отражают более общие тенденции (например, с востока на запад), а малые расстояния – более локальные тенденции (например, между районами). Поэтому, применяя крупномасштабный анализ на локальном уровне, мы можем обобщить и потерять скрытую пространственную неоднородность.

В гипотетическом примере кластеризация безработицы может быть очевидна в масштабе анализа 100 м и 1000 м, отражая закономерности на уровне городских округов. На практике масштаб анализа определяется расстоянием, на котором будут анализироваться пространственные объекты (например, городские округа) для вычисления пространственных статистик (см. также раздел 4.3).

Проблема изменяющейся пространственной единицы (Modifiable Areal Unit Problem, MAUP). Атрибутивные данные, такие как данные переписи или социально-экономические данные, часто агрегируются по соображениям

конфиденциальности или простоты в заранее определенные зоны. Проблема MAUP относится к влиянию дизайна зоны на результаты анализа. Другой способ группировки, вероятно, привел бы к другим результатам. Основным недостатком заключается в том, что зоны (их границы и протяженности) определяются произвольно по отношению к конкретной задаче географического анализа. Например, во многих городах округа в центре меньше, чем на окраинах. Насколько иными были бы статистические результаты, если бы все округа имели одинаковую площадь? Типичный и хорошо изученный пример MAUP – президентские выборы в США 2000 года. Эл Гор (Al Gore) получил больше голосов, чем Джордж Буш (George Bush), но проиграл выборы из-за особенностей разграничения округов внутри каждого штата (O’Sullivan & Unwin 2010, стр. 38). Другое разграничение округов могло привести к другому исходу. Проблема MAUP оказывает влияние как на масштаб анализа, так и на агрегирование данных. В общем случае, когда имеются более крупные территориальные единицы, мы склонны агрегировать данные на более высоком уровне, чтобы обобщение было более очевидным. Проще говоря, в результате обобщения агрегированные значения имеют тенденцию быть более похожими на общее среднее (глобальное среднее), а отклонения сглаживаются. В таком случае можно потерять ценную информацию.

Гетерогенность (неоднородность) пространства. Пространство неоднородно. Это главный фактор, отличающий пространственные данные от непространственных. Наш повседневный жизненный опыт предлагает различные примеры негладких, прерывистых и неізотропных пространственных эффектов. Например, естественные и искусственные преграды, такие как реки, дороги и парки, нарушают целостность пространства. Местоположения в неоднородном пространстве имеют разные вероятности определенного значения/действия или процесса. Например, стоимость земли на одном берегу реки может быть значительно выше, чем на другом. Пространственная неоднородность – это вариации между значениями в наборе наблюдений внутри исследуемой области (Dall’erba 2009). Когда имеет место пространственная неоднородность, могут наблюдаться тенденции в определенных направлениях (например, с востока на запад) или большая разница значений в соседних наблюдениях. Например, соседние районы могут иметь резкие социально-экономические различия, что свидетельствует о пространственной неоднородности (например, во многих мегаполисах трущобы расположены рядом с богатыми пригородами).

При наличии пространственной неоднородности сбор данных может быть сопряжен с некоторыми проблемами. Представьте, что население неравномерно распределено по территории города (высокая неоднородность населения). Если собрать и нанести на карту события приступов инсульта по всему городу, то мы обнаружим области, в которых эти события образуют кластеры. Это не обязательно означает, что риск инсульта в этих областях выше. Концентрация приступов инсульта может быть обусловлена большей численностью людей, проживающих в этих районах (высокая неоднородность населения). В менее населенных районах ожидается меньшее количество приступов. В подобных случаях следует учитывать распределение населения в пространстве, чтобы понять, есть ли связь между местоположением и неожиданно высоким количе-

ством приступов инсульта. Например, можно вычислить такой показатель, как плотность инсультов на душу населения, для каждого района в городе и внести поправку на неоднородность населения.

Пространственная неоднородность не означает отсутствия пространственной автокорреляции. Внутри исследуемой области могут быть подобласти с высокой пространственной неоднородностью или с отрицательной либо положительной пространственной автокорреляцией. На самом деле пространственную зависимость нелегко отличить от пространственной неоднородности. В литературе это называется «обратной задачей» (Anselin 2010). Когда имеет место пространственная зависимость, корреляция или ковариация между переменными в разных местах определяется расположением объектов в географическом пространстве (Anselin 2010). Но, несмотря на то что кластеры и закономерности могут обнаруживаться с помощью разных процедур, таких как анализ пространственной автокорреляции, нет никакой возможности определить, являются ли эти кластеры результатом структурной неоднородности или истинного процесса, создающего кластеры независимо от неоднородности пространства.

Проблема краевых эффектов. В проблеме краевых эффектов пространственные единицы, расположенные в центре изучаемой области, как правило, имеют соседей во всех направлениях, тогда как пространственные единицы на краях изучаемой области имеют соседей только в некоторых определенных направлениях. Для учета этой асимметрии в подсчете соседей обычно используется метод стандартизации (нормализации) строк (см. раздел о пространственных весах далее в этой главе).

Экологическая ошибка. Эта проблема возникает, когда предполагается, что связь, статистически значимая на одном уровне анализа, сохраняется на более детальном уровне. Эта типичная ошибка имеет место при использовании агрегированных данных для описания поведения людей. Например, если на уровне округа переменная «высокий доход» сильно коррелирует с переменной «высшее образование», то это не обязательно означает, что каждый, имеющий высшее образование, будет иметь высокий доход. Это проблема заблуждения – убеждение, что если что-то верно на уровне агрегирования, то это также будет верно на более низком, более детальном уровне (например, на уровне индивидуума). Данный результат правильнее интерпретировать так: «на участках, где проживает больше людей с высшим образованием, доход на душу населения, как правило, выше», но никак не «каждый человек с высшим образованием будет иметь высокий доход». Чтобы сделать вывод на уровне индивидуумов о том, как образование связано с доходом, мы должны провести исследование на этом уровне анализа (использовать данные на уровне индивидуумов, а не агрегированные на каком-то другом уровне).

Врезка 1.3. Данные, которые следовало бы рассматривать как пространственные, часто анализируются без учета их пространственных характеристик. Это происходит, потому что аналитикам не хватает пространственного мышления или они не знают, как использовать инструменты пространственного анализа. В этом случае географическое пространство исключается из анализа, и данные,

Врезка 1.3 (окончание)

которые следовало бы изучать в различных географических и аналитических масштабах, изучаются только в том масштабе, в каком они доступны. Поскольку пространственное измерение не участвует в анализе, возникает одна или несколько из вышеупомянутых причин, которые делают пространственные данные особенными. Например, если существует пространственная автокорреляция, то применение классической статистики часто нарушает допущения, необходимые для получения статистически значимых результатов, и результаты могут получиться смещенными (Lee & Wong 2001). Таким образом, пространственные данные являются особенными, и для их анализа следует использовать новые методы и приемы, учитывающие пространственные отношения и пространственную концептуализацию.

1.4. КОНЦЕПТУАЛИЗАЦИЯ ПРОСТРАНСТВЕННЫХ ОТНОШЕНИЙ

Определение

Концептуализация пространственных отношений – это моделирование отношений и взаимосвязей между объектами в пространстве. Проще говоря, это математическое определение таких понятий, как «близко», «далеко», «смежный», «непрерывный», «соседний», «окрестность» и «расстояние», для набора пространственных объектов в виде определенных значений или функций.

Назначение

Сошлемся еще раз на первый закон географии Тоблера: объекты, расположенные по соседству или близко друг от друга, обладают общими характеристиками и, вероятно, будут взаимодействовать друг с другом чаще, чем объекты, находящиеся далеко друг от друга. Концептуализация пространственных отношений помогает определить, какие объекты можно считать близкими, далекими или соседствующими, и имеет большое значение для любого географического анализа и применения инструментов пространственной статистики. На этом этапе важно:

- выбрать подходящую меру расстояния (например, евклидово, манхэттенское, время в пути или обратное расстояние);
- определить, что значит смежность (например, по наличию общих границ или по вхождению в заранее определенную область);
- определить количество соседей для вычислений пространственных статистик (например, 10, 100 или 1);
- определить метод выбора ближайшего соседа (например, по расстоянию или по смежности; если по расстоянию, то выбрать меру расстояния – евклидово или манхэттенское; если по смежности, то следует уточнить – выбирать ли по наличию общих границ или по вхождению в предопределенную зону).

При выборе разных пространственных отношений пространственный анализ одного и того же набора данных приводит к разным результатам. По этой причине очень важно определить пространственные отношения после тщательного исследования, избегая значений по умолчанию, которые применяются в большинстве программных пакетов ГИС.

Интерпретация

Чем точнее концептуализация пространственных отношений для набора пространственных объектов, тем точнее будут результаты статистических исследований и моделей. Если концептуализация не отражает внутреннюю структуру отношений пространственных объектов, результаты анализа будут вводить в заблуждение. Представьте, что нам нужно определить область охвата обслуживанием (окрестность) для кафе, чтобы в дальнейшем проанализировать социально-экономические характеристики людей, живущих или работающих в этой области. Для целей моделирования пространственных отношений между людьми и кафе сделаем следующее предположение: люди в радиусе 1 км от кафе с большей вероятностью будут посещать его, чем люди за пределами этой области. Эта простая концептуализация пространственных отношений определяет форму окрестности (круг) и ее размер (1 км). Также можно использовать другие формы (например, квадрат или шестиугольник) и размеры (например, 2 км). Очевидно, что допущение о более обширной области обслуживания, например 10 км, приведет к неточным результатам. Разумнее предположить, что лучше подойдет меньшее расстояние, потому что основная целевая категория потребителей для заведений этого типа – это люди, живущие или работающие в пределах шаговой доступности. Разные концептуальные представления об области обслуживания будут приводить к совершенно разным результатам.

Обсуждение и практические рекомендации

Существуют разные методы концептуализации пространственных отношений, каждый из которых имеет целью лучше смоделировать внутреннюю структуру набора пространственных данных с точки зрения пространственной смежности и формирования окрестностей. Например:

- А. **Расстояние.** Определить пороговое значение или функцию расстояния (см. раздел 1.6):
- мера расстояния (например, евклидово, манхэттенское, Минковского, сетевое);
 - фиксированное расстояние (порог);
 - расстояние затухания (функция расстояния).
- Б. **Смежность.** Определить, какие объекты считаются смежными (см. разделы 1.6, 1.7):
- только имеющие общие границы (правило ладьи);
 - имеющие общие углы (правило ферзя);
 - смежность высшего порядка;
 - взаимодействующие (включая расстояние и близость).

В. Соседство. Определите окрестности (см. раздел 1.8):

- k ближайших соседей;
- полигоны близости;
- триангуляция Делоне.

Г. Пространство-время. Определить расстояния и временные окна для пространственно-временного анализа (см. раздел 1.8.2):

Чтобы выбрать подход для использования, сначала нужно изучить проблему с концептуальной точки зрения и проследить потенциальные пространственные отношения. Например, при исследовании дорожного движения вместо евклидова расстояния рациональнее использовать сетевое или манхэттенское. При исследовании плотности населения в городской агломерации целесообразнее использовать функцию затухания (например, обратное расстояние), потому что районы, расположенные дальше от центра, вероятно, будут менее населенными.

Многие инструменты пространственной статистики требуют перед любым анализом установить метод концептуализации пространственных отношений. Вот некоторые из статистических методов, подробно представленных в этой книге:

- глобальная пространственная автокорреляция (общий индекс I Морана, общая G -статистика);
- анализ кластеров и выбросов (локальный индекс I Морана);
- анализ горячих точек (Getis-Ord G_i^*);
- пространственно-временные автокорреляции (двумерный индекс I Морана, дифференциальный индекс I Морана);
- создание матрицы пространственных весов;
- географически взвешенная регрессия.

1.5. ИЗМЕРЕНИЕ РАССТОЯНИЯ

В географическом анализе чаще всего используются следующие меры расстояния: евклидово расстояние, манхэттенское расстояние, расстояние Минковского, корреляционное расстояние Пирсона, корреляционное расстояние Спирмена, сетевое расстояние и геодезическое расстояние. В пространственной статистике чаще всего используются евклидово и манхэттенское расстояния.

Определения и формулы

Расстояние между двумя точками A и B с координатами X_1, Y_1 и X_2, Y_2 соответственно в системе координат проекции (плоская поверхность с декартовой системой координат) можно определить следующим образом:

- **евклидово** расстояние – расстояние между двумя точками A и B , соединенными прямой линией, рассчитывается по формуле (1.1):

$$S = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2}; \quad (1.1)$$

- **манхэттенское** расстояние – сумма разностей вертикальных и горизонтальных координат точек A и B рассчитывается по формуле (1.2):

$$S = |X_2 - X_1| + |Y_2 - Y_1|. \quad (1.2)$$

Манхэттенское расстояние можно использовать для моделирования расстояния между точками в городской среде, когда уличная сеть недоступна. Здания создают препятствия, и обычно нет возможности переместиться из точки А в точку В по прямой. Манхэттенское расстояние имитирует такой вид перемещений, поэтому его следует использовать при изучении атрибутов, связанных с поездками (например, расстояние до школы, зона обслуживания магазина) в городской среде. При анализе социально-экономических характеристик, таких как численность населения, уровень дохода или образование, предпочтительнее использовать евклидово расстояние;

- **расстояние Минковского** – это обобщение евклидова и манхэттенского расстояний и рассчитывается по формуле (1.3):

$$S = \left((X_2 - X_1)^p + (Y_2 - Y_1)^p \right)^{1/p}. \quad (1.3)$$

При $p = 1$ мы получаем манхэттенское расстояние, а при $p = 2$ – евклидово. Расстояние Минковского используется также в анализе главных компонент и в кластерном анализе (см. главу 5);

- **корреляционное расстояние Пирсона** считает два вектора близкими (похожими), если они сильно коррелированы. В многомерном наборе данных для двух переменных с корреляцией r корреляционное расстояние Пирсона рассчитывается по формуле (1.4) (определение и уравнение коэффициента корреляции r см. в разделе 2.3.4):

$$S = 1 - r. \quad (1.4)$$

В процессе концептуализации пространственных отношений по расстоянию, кроме меры расстояния, следует также выбрать диапазон расстояний, в пределах которого пространственные объекты рассматриваются как имеющие пространственные отношения. Диапазон расстояний тесно связан с масштабом анализа. На практике наиболее широко используются два метода определения диапазона расстояний – фиксированное расстояние и расстояние затухания.

1.5.1. Фиксированное расстояние (сфера влияния)

Определение

Фиксированный диапазон расстояний – это значение, выражающее размер сферы влияния, окружающей пространственный объект. Все пространственные объекты внутри сферы влияния получают одинаковый вес, а пространственным объектам за ее пределами присваивается нулевой вес.

Назначение

Фиксированный диапазон расстояний используется для формирования окрестностей, содержащих одинаково интерпретируемые пространственные объекты.

Интерпретация

Сфера влияния определяется выбором фиксированного расстояния (порогового значения). Внутри этой сферы все пространственные объекты получают одинаковый вес, а объекты вне сферы не учитываются в расчетах (например, при вычислении весовой матрицы или пространственной статистики), потому что предполагается, что из-за удаленности они не взаимодействуют с объектами внутри сферы (т. е. их вес равен нулю).

Выбор подходящего значения для диапазона расстояний имеет большое значение, потому что задает масштаб анализа и определяет реакцию пространственной статистики. Например, для вычисления большинства пространственных статистик требуется наличие, по крайней мере, одного соседа у каждого пространственного объекта. Когда количество соседей невелико, статистические результаты (например, z-оценки) ненадежны. Выбор маленького расстояния может привести к ситуации, когда у объектов не будет соседей. Выбор большого значения может создать чрезмерно большие сферы влияния и, как результат, нежелательные агрегаты. Кроме того, при выборе большого расстояния увеличиваются вычислительные затраты, особенно когда в вычислениях участвуют тысячи соседей. Диапазон расстояний следует выбирать так, чтобы, с одной стороны, он гарантировал наличие хотя бы одного соседа у каждого пространственного объекта, а с другой – чтобы число соседей было не слишком велико. Когда распределения искажены, для получения надежных результатов необходимо иметь около восьми соседей. Как поясняется ниже, для определения диапазона расстояний, подходящего для конкретного набора данных или области исследования, можно использовать различные инструменты и процедуры.

Инструменты для определения диапазона расстояний

Не существует какого-то оптимального диапазона расстояний для решения конкретной задачи. При наличии неоднородности пространства могут применяться разные масштабы анализа. Определение подходящего диапазона расстояний зависит от решаемой задачи и масштабов анализа. Задача оценки распределения школ в центре города по отношению к численности населения требует относительно небольшого масштаба анализа, от нескольких сотен метров до 1–2 км, потому что было бы неразумно заниматься ежедневной доставкой детей в более удаленные школы. С другой стороны, для оценки распределения больниц в том же исследовании диапазон расстояний можно увеличить, потому что больницы имеют большую зону обслуживания. Вот некоторые типичные подходы к выбору диапазона расстояний:

- выбор на основе предыдущего опыта. Если исследователь обладает глубокими знаниями изучаемого пространственного явления и опытом изучения конкретного показателя или существуют публикации с результатами большого количества исследований, подтверждающих эффективность использования определенного диапазона расстояний для данной задачи, то расстояние можно выбрать на основе своего опыта или опыта других исследователей;
- выбор диапазона расстояний, гарантирующего наличие у каждого объекта хотя бы одного соседа. Следует отметить, что при наличии географических выбросов (см. главу 3) этот подход неуместен, так как такое

расстояние будет чересчур большим. Гарантируя наличие хотя бы одного соседа у некоторых объектов, оно может создавать окрестности вокруг других объектов, содержащие сотни соседей. Как следствие масштаб анализа может приводить к неверным выводам. Чтобы решить эту проблему, нужно исключить географические выбросы и рассчитать расстояние, при котором каждый объект имеет хотя бы одного соседа. Также следует установить минимальное количество соседей для тех объектов, у которых нет соседей в пределах выбранного диапазона расстояний. Такой подход помогает определить наиболее реалистичный диапазон расстояний, гарантируя, что выбросы также будут включены, путем вычисления их весов на основе количества ближайших соседей, а не расстояния (подробнее об этом рассказывается в разделе 4.3);

- выбор расстояния, на котором пространственная автокорреляция наиболее выражена (см. главу 4). Когда имеет место пространственная кластеризация, для определения подходящего диапазона расстояний могут использоваться методы пространственной автокорреляции, такие как инкрементная пространственная автокорреляция, оптимизированный анализ горячих точек и k -функция Рипли.

Обсуждение и практические рекомендации

Прежде чем определить подходящий диапазон расстояний, необходимо проверить наличие географических выбросов (см. главу 3). Если данные искажены, то расстояние не должно быть слишком маленьким, обеспечивающим включение только одного или двух соседей, или слишком большим, чтобы площади окрестностей не оказались сопоставимыми с площадью изучаемой области, потому что z -оценки, вычисленные различными методами пространственной статистики, будут ненадежными. С искаженными данными приемлемая надежность z -оценок достигается, когда количество соседей близко к восьми.

Большинство программных пакетов вычисляют диапазон расстояний по умолчанию, при котором каждый объект имеет не меньше одного соседа. Однако это не всегда верный выбор, особенно когда имеются географические выбросы (см. предыдущий раздел «Инструменты для определения диапазона расстояний»). Фиксированное расстояние уместно использовать, когда есть хорошая теоретическая база для решения поставленной задачи или хорошее знание области исследования. Фиксированное расстояние также хорошо подходит для точечных данных и данных в форме полигонов при анализе горячих точек и в других методах анализа локальной пространственной автокорреляции.

1.5.2. Расстояние затухания

Определение

Расстояние затухания – это любая функция, которая подразумевает плавное и непрерывное ослабление влияния некоторого фактора на значения атрибутов соседних пространственных объектов с увеличением расстояния (Longley et al. 2011, стр. 108; см. рис. 1.3).

Назначение

Расстояние затухания используется для выражения пространственных взаимодействий между разными местоположениями путем применения весов, меняющихся в зависимости от расстояния, когда более близкие объекты имеют больший вес (оказывают более сильное влияние), чем более удаленные. Это расстояние используется как практическое воплощение первого закона географии Тоблера.

Интерпретация

Идея затухания предполагает ослабление физических или социально-экономических взаимодействий между двумя точками в пространстве с увеличением расстояния между ними. Для представления затухания могут использоваться различные типы функций, в том числе обратная функция (или обратное расстояние; см. рис. 1.3А), отрицательная степенная функция (или обратный квадрат расстояния; см. рис. 1.3В), отрицательная экспоненциальная или обратная линейная функция (используются довольно редко). Обратное расстояние имеет менее крутой график, чем обратный квадрат расстояния. Типичное соотношение расстояния затухания (обратное расстояние или обратный квадрат расстояния) можно использовать для описания высоты зданий в городе: чем дальше от центра, тем ниже здания. Эти функции можно использовать для описания плотности населения: чем дальше от центра города, тем ниже плотность.

Обсуждение и практические рекомендации

Фиксированное расстояние обычно рекомендуется использовать в сочетании с функцией затухания (в качестве точки отсечки или в качестве области равнозначности):

- фиксированное расстояние, применяемое в качестве точки отсечки (см. рис. 1.3С). За этой точкой расстояние между объектами не учитывается, потому что взаимодействия между ними перестают быть значимыми; никакой анализ с использованием расстояний, превышающих это значение, не дает ценной информации;
- фиксированное расстояние перед функцией затухания (область равнозначности, см. рис. 1.3D). В этом случае определяется расстояние, на котором все пространственные объекты интерпретируются как взаимодействующие одинаково сильно. За пределами этого расстояния применяется функция затухания, и веса рассчитываются в соответствии с ней. Такой подход применим, когда есть область, в пределах которой все объекты взаимодействуют одинаково сильно. Чтобы не создавать резкую границу после этой области, применяется функция затухания, плавно изменяющая веса объектов, в зависимости от удаленности.

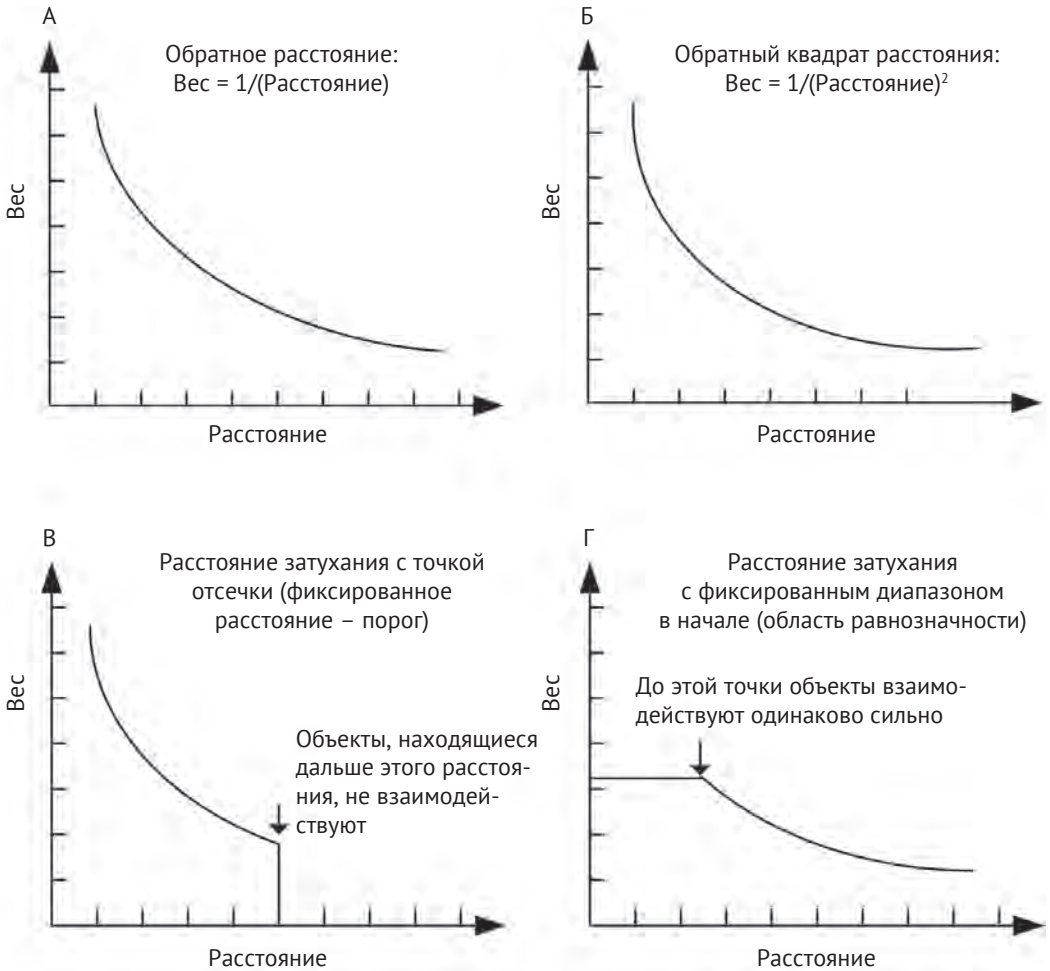


Рис. 1.3. Графики функций затухания. Чем больше расстояние между двумя точками, тем меньше они влияют друг на друга. (А) График функции обратное расстояние похож на график обратного квадрата расстояния, но менее крутой. (Б) Функцию обратного квадрата расстояния можно использовать, когда есть доказательства, что взаимовлияние объектов ослабляется с расстоянием в большей степени, чем предполагает функция обратного расстояния. (В) Пороговое значение (или фиксированное расстояние) можно использовать для завершения функции затухания. (Г) Фиксированное расстояние можно использовать до функции затухания. Это расстояние также называют «областью равнозначности», потому что все объекты внутри нее интерпретируются одинаково. Объекты, удаленные дальше этого расстояния, взвешиваются в соответствии с функцией затухания

Комбинирование функции затухания с фиксированным расстоянием сокращает время вычислений и лучше отражает реальность. Поскольку на очень больших расстояниях объекты практически не взаимодействуют и их веса почти равны нулю, вместо вычисления их влияния можно прервать процесс вычислений на расстоянии с заранее заданным значением. Функции

затухания могут использоваться для непрерывных данных. Например, обратное евклидово расстояние можно использовать для моделирования температуры или загрязненности. Обратное манхэттенское расстояние, с другой стороны, больше подходит для точечного анализа внутри городской среды (например, для анализа клиентов), когда информация о дорожной сети недоступна. При анализе горячих точек следует избегать использования обратного расстояния в роли функции затухания, так как оно имеет тенденцию создавать небольшие изолированные горячие или холодные точки (см. главу 4).

Пример

Предположим, вы решили выбрать кафе, куда можно приходить каждый день, чтобы попить кофе, и расстояние от дома является одним из наиболее важных критериев (в реальной жизни выбор определяется также рядом дополнительных параметров, но мы будем использовать только расстояние; см. рис. 1.4). Не важно, находится ли кафе от вашего дома на расстоянии 200 м или 250 м. В действительности есть два кафе – одно в 200 м, а другое в 250 м от вашего дома, и вы, скорее, пройдете лишние 50 м, если второе кафе окажется лучше. Однако если расстояние составляет 1 км, то вы дважды подумаете, стоит ли идти так далеко. Таким образом, мы можем использовать фиксированный диапазон расстояний от дома, например 300 м, когда вы одинаково отнесетесь к идее пройти его пешком, чтобы добраться до кафе. Для дистанций больше этого расстояния мы можем применить функцию затухания; например, идея посетить кафе, отстоящее всего на 30 м дальше точки отсечения (300 м), не кажется иррациональной, однако чем дальше находится кафе, тем меньший вес оно получает при выборе, по сравнению с кафе, находящимися в зоне влияния. С другой стороны, если вы хотите как можно быстрее сходить в аптеку, то наверняка отдадите предпочтение ближайшей. В этом случае для оценки лучше подойдет обратный квадрат расстояния, чтобы особенно подчеркнуть вес/важность близости аптеки. Наконец, если вы собираетесь ездить на велосипеде, то можете купить еду в более удаленном местном продуктовом магазине. Тогда уместно использовать функцию обратного расстояния.

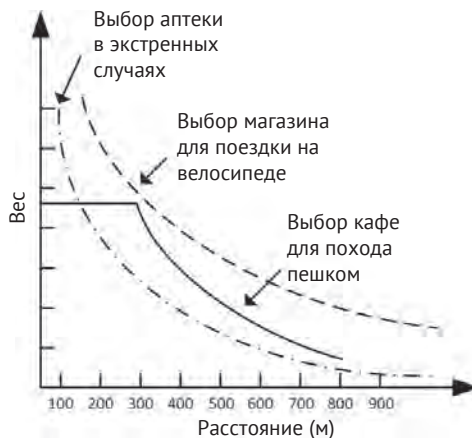


Рис. 1.4. Пример использования разных функций затухания для анализа

1.6. Близость: МАТРИЦА СМЕЖНОСТИ

1.6.1. Близость полигонов

Определение

Близость – это пространственное свойство, описывающее, находятся ли целевой объект и один или несколько других объектов в непосредственной близости друг от друга. На практике под близостью понимается, какие полигоны считать соседями целевого объекта. Вот наиболее распространенные методы представления близости для объектов-полигонов:

- **только имеющие общие границы (правило ладьи).** Соседями считаются только полигоны, имеющие общую границу (край), и только они включаются в вычисления, связанные с целевым полигоном (см. рис. 1.5А);
- **имеющие общие углы (правило ферзя).** Соседями считаются полигоны, имеющие общие границы, а также общие углы (узлы), и все они включаются в вычисления, связанные с целевым полигоном (см. рис. 1.5В).

Назначение

Понятие близости используется для определения окрестностей, а также при вычислении матрицы пространственных весов и различных пространственных статистик.

Обсуждение и практические рекомендации

Предыдущие методы представления близости используются для определения окрестностей или силы взаимодействий при моделировании данных или процессов, представленных полигонами. Степень близости также важна при определении соседей. Близость первого порядка имеет место, когда взаимодействия и соседство учитываются только для объектов, располагающихся в непосредственной близости, а близость второго порядка отражает влияние на целевой объект соседей его соседей (Anselin 2016). При вычислении весов близости высшего порядка следует решить, необходимо ли учитывать соседей более низкого порядка (Anselin 2018; см. рис. 1.5В). Когда при вычислении весов высшего порядка не учитываются веса более низкого порядка, это можно назвать дизъюнктивной (исключающей, или чистой) близостью высшего порядка (Anselin 2018). Когда близость высшего порядка вычисляется с учетом более низкого порядка, она называется конъюнктивной (или включающей) близостью высшего порядка (Anselin 2018). Например, окрестность второго порядка, включающая также окрестность первого порядка, называется конъюнктивной (включающей) окрестностью второго порядка. Чистая близость высшего порядка (не включающая соседей более низкого порядка) используется при изучении влияния пространственных лагов на пространственную автокорреляцию и на модели пространственной авторегрессии, не учитывающих избыточные и циклы полигональных объектов (Anselin 2018).

Определение окрестности с применением правила ладьи или ферзя хорошо зарекомендовало себя в случаях, когда пространственное взаимодействие

увеличивается, если два полигона имеют общее ребро, узел или и то, и другое. В общем случае, если полигоны имеют одинаковый размер, близость полигонов является подходящим методом. Для полигонов разного размера (например, небольших в центре города и больших на окраинах) при их исследовании используются разные масштабы анализа. По этой причине для вычисления матрицы пространственных весов следует использовать прием стандартизации (нормализации) строк в совокупности с методом определения близости полигонов (см. раздел 1.9).

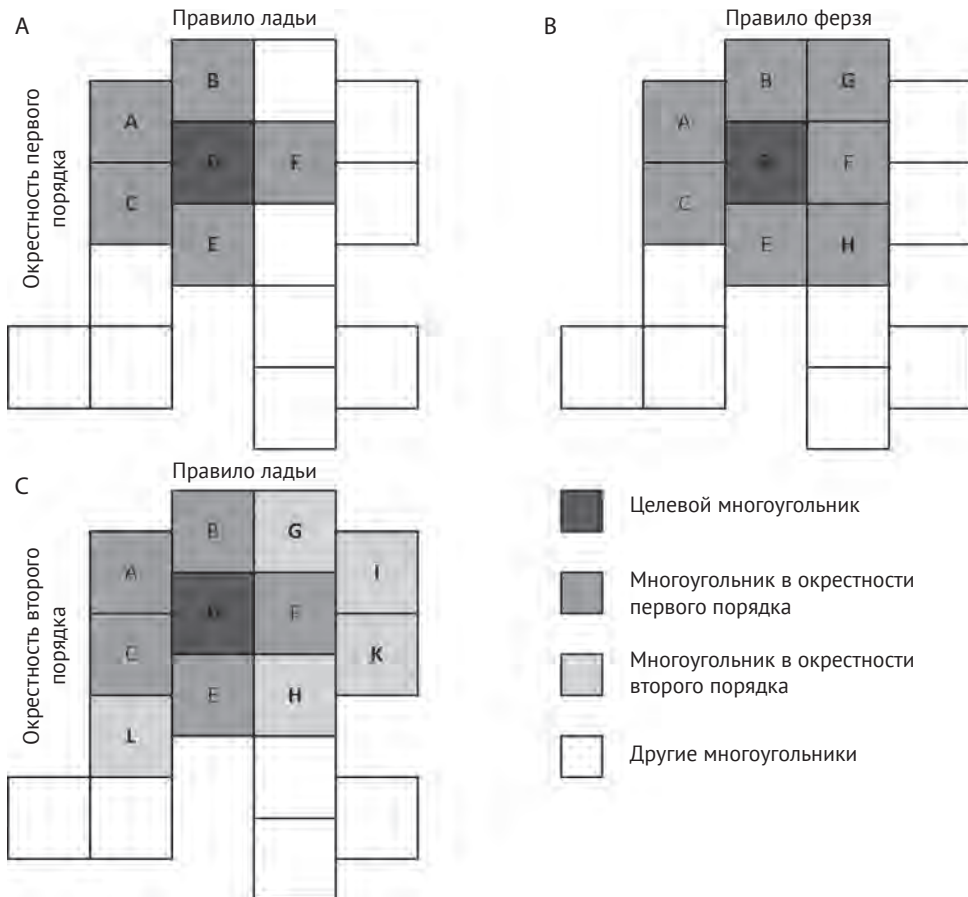


Рис. 1.5. (А) Определение окрестности с применением правила ладьи (имеется общая граница). Окрестность составляют полигоны, непосредственно примыкающие к целевому объекту D, – это A, B, C, F и E (окрестность первого порядка). (В) Правило ферзя (имеется общая граница или угол). В соответствии с этим правилом в окрестность D дополнительно включены полигоны G и H, и теперь ее составляют полигоны A, B, C, E, F, G и H (окрестность первого порядка). (С) Дизъюнктивная (исключающая, или чистая) окрестность второго порядка с применением правила ладьи для D включает G, H, I, K и L, потому что они являются соседями соседей первого порядка (A, B, C, E и F). Конъюнктивная (включающая) окрестность второго порядка для D состоит из непосредственных соседей и из соседей этих соседей (A, B, C, E, F, G, H, I и K)

1.6.2. Матрица смежности

Определение

Для представления полигонов в контексте пространственного анализа используется **матрица смежности** – квадратная матрица, элементы которой указывают, являются пары полигонов смежными или нет.

Назначение

Матрица смежности применяется для представления разных форм смежности и определения окрестностей для дальнейшего пространственного анализа.

Интерпретация

Матрица смежности – симметричная матрица, недиагональные элементы которой принимают значение 0 или 1, а диагональные не имеют значений. Если два пространственных объекта являются смежными (первого, или высшего, порядка), то соответствующему элементу матрицы присваивается 1, иначе – 0. Полигоны со значением смежности 1 в матрице для заданного целевого полигона (в строке или в столбце, соответствующем целевому полигону) определяют его окрестность (набор объектов, которые необходимо учитывать при вычислении пространственной статистики для целевого объекта).

Обсуждение и практические рекомендации

Для случая, изображенного на рис. 1.5А, когда смежность определяется правилом ладьи (смежными считаются только полигоны, имеющие общие границы, или ребра), матрица смежности имеет вид (1.5):

$$Adj = \begin{bmatrix} & A & B & C & D & E & F & SUM \\ A & * & 1 & 1 & 1 & 0 & 0 & 3 \\ B & 1 & * & 0 & 1 & 0 & 0 & 2 \\ C & 1 & 0 & * & 1 & 1 & 0 & 3 \\ D & 1 & 1 & 1 & * & 1 & 1 & 5 \\ E & 0 & 0 & 1 & 1 & * & 0 & 2 \\ F & 0 & 0 & 0 & 1 & 0 & * & 1 \\ SUM & 3 & 2 & 3 & 5 & 2 & 1 & * \end{bmatrix}. \quad (1.5)$$

Для обозначения смежности полигонов с самими собой мы используем звездочку (*). Сумма (SUM) в каждой строке или столбце равна общему количеству полигонов, смежных с данным. Если соседство с целевым объектом (полигоном) определяется фактом смежности (близостью первого порядка), то полигон считается соседом целевого объекта, если значение в матрице равно 1; в противном случае они не соседи. Например, полигон А имеет трех соседей первого порядка (В, С, D).

1.7. ВЗАИМОДЕЙСТВИЕ

Определение

Взаимодействие – это степень связи между двумя местоположениями в пространстве, пунктами отправления и назначения. Она вычисляется на основе расстояния и значения смежности. В простейшем случае формула имеет вид (1.6):

$$\text{Interaction} = w_{i,j} \frac{P_i P_j}{d_{i,j}^b}, \quad (1.6)$$

где

i, j – местоположения, обозначающие пункты отправления и назначения соответственно;

w – некоторый весовой коэффициент, корректирующий величину связи между i и j . Например, w может быть равен 0, если объекты несмежные, и 1, если смежные;

P_i и P_j – соответствующие значения определенной переменной (например, численность населения);

$d_{i,j}$ – это любая функция расстояния между i, j ;

b – показатель степени, определяющий скорости уменьшения связи с увеличением расстояния.

Назначение

В пространственном статистическом анализе вычисление величины взаимодействия можно использовать для получения весов в пространственной матрице весов (см. раздел 1.9).

Интерпретация

Чем больше значение, тем сильнее взаимодействуют два местоположения.

Обсуждение и практические рекомендации

Модели пространственного взаимодействия в основном используются для изучения пространственных потоков, таких как миграция, туризм, поездки на работу, международная торговля и финансовые потоки. Пространственное взаимодействие возникает, если выполняются три независимых условия:

- имеются множества местоположений «предложения» (пунктов назначения) и «спроса» (пунктов отправления), например пассажиры [спрос], направляющиеся на работу [предложение];
- имеются альтернативные местоположения для обоих множеств пунктов отправления и назначения;
- пункты отправления и назначения связаны друг с другом.

1.8. ОКРЕСТНОСТИ И СОСЕДИ

Определение

Окрестность в контексте пространственного анализа – это географически локализованная область, к которой применяются локальный пространственный анализ и статистические методы, основывающиеся на гипотезе о том, что объекты в пределах этой области взаимодействуют сильнее, чем объекты за ее пределами.

Назначение

Определение окрестности необходимо для точного вычисления пространственных статистик. Большая часть статистических методов требует определения соседства и построения пространственной матрицы весов, которая отражает интенсивность отношений между пространственными объектами.

Интерпретация

Размер окрестности определяет способ агрегирования наблюдений. Анализ окрестностей со слишком маленьким или слишком большим количеством объектов, скорее всего, приведет к ненадежным статистическим результатам.

Обсуждение и практические рекомендации

Определение окрестности – нетривиальная задача, потому что существует масса способов ее формирования. Выше уже были представлены методы диапазона расстояний (раздел 1.5.1) и смежности полигонов (раздел 1.6.2) для определения окрестностей. В числе других методов можно назвать: метод k ближайших соседей, пространственно-временной близости, полигоны близости и триангуляции Делоне; все они описываются ниже.

1.8.1. Метод k ближайших соседей

Определение

В контексте пространственного анализа **метод k ближайших соседей** (*k -Nearest Neighbors, k NN*) – это метод, используемый для определения окрестности по k ближайшим соседям к целевому объекту.

Назначение

Этот метод гарантирует, что каждый объект будет иметь не меньше определенного количества соседей. В более широком смысле он используется для:

- определения окрестности (области), пространственные объекты в которой будут учитываться при вычислении пространственных статистик;
- определения окрестности, пространственные объекты в которой могут быть более похожими на целевой объект, чем более удаленные объекты;
- моделирования пространственных отношений.

Интерпретация

Метод k NN основан на вычислении матрицы расстояний любого типа между всеми объектами в наборе данных.

Например, матрица расстояний (основанная на евклидовом расстоянии) для всех пар полигонов на рис. 1.6 имеет вид (1.7):

$$Distance = \begin{bmatrix} & A & B & C & D & E \\ A & * & 4.0 & 2.2 & 2.9 & 4.3 \\ B & 4.0 & * & 4.1 & 2.5 & 4.2 \\ C & 2.2 & 4.1 & * & 2.4 & 2.6 \\ D & 2.9 & 2.5 & 2.4 & * & 1.9 \\ E & 4.3 & 4.2 & 2.6 & 1.9 & * \end{bmatrix}. \quad (1.7)$$

Для случая $k = 2$ матрица ближайших соседей (NN) имеет вид (1.8):

$$NN = \begin{bmatrix} & A & B & C & D & E & SUM \\ A & * & 0 & 1 & 1 & 0 & 2 \\ B & 1 & * & 0 & 1 & 0 & 2 \\ C & 1 & 0 & * & 1 & 0 & 2 \\ D & 0 & 0 & 1 & * & 1 & 2 \\ E & 0 & 0 & 1 & 1 & * & 2 \\ SUM & 2 & 0 & 3 & 4 & 1 & \end{bmatrix}. \quad (1.8)$$

Чтобы построить матрицу ближайших соседей (NN) для случая $k = 2$, в каждой строке исходной матрицы расстояний $Distance$ (1.7) мы заменяем два (k) наименьших значения единицами, а остальные – нулями. Получившаяся матрица ближайших соседей (NN) напоминает матрицу смежности, но она не симметрична. Сумма (SUM) элементов в каждой строке равна 2 (равна k), но суммы в столбцах получаются разными. Сумма в столбце показывает, сколько раз соответствующий объект (его имя указано в заголовке столбца) оказался ближайшим (в данном случае первым или вторым) к другим полигонам.

В случае $k = 2$ каждая окрестность включает три пространственных объекта (целевой объект и два его ближайших соседа). Например, ближайшими соседями полигона В (взгляните на соответствующую строку в матрице NN) являются полигоны А и D (см. рис. 1.6). Полигон D ближе к В, чем к А, поэтому он называется «первым ближайшим соседом для В», а полигон А называется «вторым ближайшим соседом для В» (см. матрицу расстояний $Distance$). Взглянув на строки А и D, можно заметить, что полигон В не является ближайшим соседом ни для одного из них. В этом нет ничего нелогичного, потому что имеются другие полигоны, расположенные к А и D ближе, чем В. Именно поэтому суммы в столбцах не равны, но суммы в строках должны быть равны двум (k).

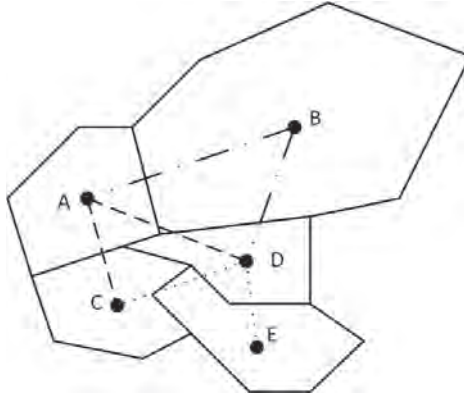


Рис. 1.6. Пример определения k ближайших соседей по матрице расстояний

Обсуждение и практические рекомендации

Выбор оптимального значения k – сложная задача. Это компромисс между созданием областей, однородных по своим характеристикам и близких по размеру. Если распределение пространственных объектов подчиняется конкурентному процессу, то есть имеется тенденция расположения объектов на равных расстояниях друг от друга по всей области исследования, то метод k ближайших соседей дает надежные результаты. Однако он очень чувствителен к размерам полигонов и плотности размещения точек. Когда объекты распределены в пространстве неоднородно или размеры полигонов сильно различаются, применение метода k ближайших соседей может потребовать значительного изменения масштаба анализа. Например, округа в деловом центре города обычно имеют небольшие зоны обслуживания, а на окраинах – значительно больше. Если установить одинаковое количество соседей для центра и окраин, то на окраинах окрестности получатся намного больше, что может привести к излишнему усреднению данных и к потере ценной информации. По этой причине иногда нежелательно применять фиксированное значение k для всей области исследования. В таких случаях рекомендуется использовать функцию затухания с фиксированным диапазоном расстояний и применять метод k ближайших соседей только для объектов, не имеющих соседей в выбранном фиксированном диапазоне расстояний. Это гарантирует, что каждый объект будет иметь хотя бы одного соседа. Если распределение значений, связанных с объектами, искажено влиянием каких-то посторонних факторов, то, как показывает практика, для получения надежных оценок необходимо, чтобы у каждого объекта имелось около восьми соседей.

1.8.2. Метод пространственно-временного окна

Определение

Пространственно-временное окно – это метод определения окрестности на основе расстояния и времени. Согласно этому методу объект считается находящимся в одной окрестности с целевым объектом, если он находится в пределах указанного расстояния и попадает в указанный интервал времени.

Назначение

Этот метод можно использовать для определения пространственно-временных горячих точек или кластеров, образованных по признакам пространственной и временной близости.

Интерпретация

Пространственные объекты, близкие друг к другу не только в пространстве, но и во времени, анализируются вместе, как пространственно-временные соседи. Если объект находится рядом с целевым в пространстве, но не во времени (или наоборот), он не включается в вычисления пространственных статистик. Иначе говоря, два пространственно близких объекта могут не входить в одну пространственно-временную окрестность, если они слишком удалены друг от друга во времени.

Обсуждение и практические рекомендации

При использовании метода пространственно-временного окна необходимо определить два параметра: диапазон расстояний и временное окно. Если, например, задать расстояние в 1 км и временной интервал в три часа, то все объекты, удаленные от целевого не далее, чем на 1 км в пространстве, и не более, чем на три часа во времени, будут проанализированы вместе, как пространственно-временные соседи. Выбор подходящего диапазона расстояний и временного окна – непростая задача, и для этого необходимо хорошо знать регион и решаемую задачу.

1.8.3. Метод полигонов близости

Определение

Полигоны близости – это полигональные области, делящие пространство так, что расстояние от любой точки в этой области до центра полигона меньше, чем до центров любых других полигонов. Полигоны близости также называются полигонами Тиссена и полигонами Вороного (O’Sullivan & Unwin 2003, стр. 43).

Назначение

Полигоны близости равномерно делят пространство по признаку близости. Полигоны Тиссена и Вороного – это методы деления плоскости на неперекрывающиеся полигоны, они могут использоваться для пространственной интерполяции или оценки областей, охватываемых общественными службами или коммерческими предприятиями (Illian et al., 2008, стр. 46).

Интерпретация

Для заданного набора точечных объектов (например, почтовых отделений, полицейских участков, кафе) с помощью метода полигонов близости можно определить области со следующими характеристиками:

- каждый объект (например, почтовое отделение) является центроидом своего полигона;
- любая точка в пространстве входит только в один полигон (полигоны не перекрываются);

- ближайшим объектом (например, почтовым отделением) для любой точки является центроид полигона, которому она принадлежит;
- точки на границах равноудалены от центроидов соответствующих полигонов.

Обсуждение и практические рекомендации

Используя полигоны близости, мы определяем окрестности или зоны влияния определенных пространственных объектов. Если взять в качестве примера почтовые отделения, то все, живущие в пределах определенного полигона близости, будут включены в окрестности почтового отделения, находящегося в центре этого полигона. В экологическом анализе полигоны Тиссена используются для оценки значений непрерывного поля (например, температуры, загрязнения). Любая точка внутри полигона будет иметь то же значение, что и точка в его центре. Однако наличие резких изменений значений в соседних точках, лежащих близко к общей границе двух смежных полигонов, может сделать такое представление проблематичным. Как объясняется далее, метод триангуляции Делоне предлагает возможность более плавной интерполяции значений.

1.8.4. Триангуляция Делоне и нерегулярные триангуляционные сети

Определение

Триангуляция Делоне делит пространство точечных объектов, играющих роль центроидов, на треугольники, полигоны близости которых имеют общие границы (O'Sullivan & Unwin 2010, стр. 51).

Назначение

Этот метод подходит в случаях, когда в наборе данных имеются изолированные полигоны (например, острова) или когда пространственное распределение объектов далеко от нормального (например, когда большие полигоны перемешаны с маленькими).

Интерпретация

Точечные объекты-центроиды, соединенные ребрами треугольника, считаются соседями. Это гарантирует наличие у каждого объекта хотя бы одного соседа.

Обсуждение и практические рекомендации

Ребра треугольников, построенных методом триангуляции Делоне, не пересекаются и создают набор треугольных фрагментов, эффективно изображающих поверхности. Эти треугольники образуют сеть, называемую нерегулярной триангуляционной сетью (Triangular Irregular Network, TIN). Что касается концептуализации пространства, то триангуляция Делоне создает окрестности, состоящие из более или менее правильных треугольников, потому что стремится максимизировать величину наименьшего внутреннего угла всех треугольников, благодаря чему предотвращается создание длинных треугольников. В общем случае вычисления основываются на высоте точек и их комбинировании с векторными данными, такими как дороги, ручьи или горные вершины, что обеспечивает более реалистичное изображение земной поверхности.

1.9. ПРОСТРАНСТВЕННЫЕ ВЕСА И СТАНДАРТИЗАЦИЯ СТРОК

Определения

Пространственные веса – это числовые значения, отражающие расстояние, время или стоимость между целевым и любым другим пространственным объектом в наборе данных или в выбранной окрестности. Пространственные веса количественно определяют пространственные или пространственно-временные отношения между пространственными точками в окрестности.

Матрица пространственных весов – это матрица, хранящая пространственные веса.

Стандартизация строк – это процесс масштабирования пространственных весов в диапазон от 0 до 1, чтобы избежать смещенности при выборке данных или их агрегировании из более крупных наборов.

Назначение

Матрица пространственных весов используется для отображения степени связи между объектами внутри определенной окрестности (Dall'Erba 2009).

Интерпретация

Любой метод концептуализации пространственных отношений приводит к созданию матрицы пространственных весов, количественно определяющей пространственные отношения между объектами. Она используется в качестве основной матрицы в большинстве методов пространственного анализа. В простейшем случае матрица пространственных весов содержит двоичные значения (при использовании фиксированного диапазона расстояний, k ближайших соседей и пространственных отношений смежности): значение 1 соответствует наличию пространственных отношений между двумя объектами, а 0 – их отсутствию. В случаях концептуализации взаимодействий с использованием функции затухания или определения весов пользователем элементы матрицы могут принимать любое значимое значение. Чем больше вес, тем теснее отношения.

При применении метода обратного расстояния ($1/(Distance)$) к полигонам, изображенным на рис. 1.6, элементы матрицы пространственных весов (1.9) отражают обратное расстояние до соответствующего набора полигонов (см. матрицу Distance (1.7)):

$$Spatial\ Weights = \begin{bmatrix} & A & B & C & D & E & SUM \\ A & * & 0.244 & 0.455 & 0.345 & 0.233 & 1.276 \\ B & 0.244 & * & 0.250 & 0.400 & 0.238 & 1.132 \\ C & 0.455 & 0.250 & * & 0.417 & 0.385 & 1.506 \\ D & 0.345 & 0.400 & 0.417 & * & 0.526 & 1.688 \\ E & 0.233 & 0.238 & 0.385 & 0.526 & * & 1.382 \end{bmatrix}. \quad (1.9)$$

Матрица пространственных весов почти всегда автоматически генерируется программным обеспечением, которое используется для вычисления пространственных статистик, при определении метода концептуализации.

Обсуждение и практические рекомендации

Стандартизация строк – это метод масштабирования весов в диапазон от 0 до 1. Каждый вес делится либо на сумму всех весов в его строке, либо на сумму весов соседних объектов (1.10). То есть результаты корректируются так, чтобы количество соседей не влияло на конечные результаты.

$$\text{Standartized Spatial Weights} = \begin{bmatrix} & A & B & C & D & E & \text{SUM} \\ A & * & 0.191 & 0.356 & 0.270 & 0,182 & 1 \\ B & 0.215 & * & 0.221 & 0.353 & 0.210 & 1 \\ C & 0.302 & 0.166 & * & 0.277 & 0.255 & 1 \\ D & 0.204 & 0.237 & 0.247 & * & 0.312 & 1 \\ E & 0.168 & 0.172 & 0.278 & 0.381 & * & 1 \\ \text{SUM} & 0.890 & 0.767 & 1.102 & 1.281 & 0.960 & \end{bmatrix} \cdot (1.10)$$

В стандартизированной матрице (1.10) сумма элементов каждой строки равна 1. Итоговые значения в столбцах – степень вовлеченности пространственного объекта во взаимодействия. Например, объект D имеет наибольшую сумму в столбце, что свидетельствует о самой сильной его вовлеченности во взаимодействия с остальными объектами. Это также указывает, что он занимает наиболее центральное местоположение. Объект B имеет наименьшую сумму в столбце, что свидетельствует о самой слабой его вовлеченности во взаимодействия с остальными объектами. В предыдущем примере мы вычислили пространственную матрицу для всех объектов в регионе. Однако обычно пространственные веса вычисляются только для соседствующих объектов (согласно матрице смежности).

Стандартизацию строк рекомендуется использовать, когда имеет место смещенность в распределении пространственных объектов и значений их атрибутов из-за плохо продуманных процедур выборки. Например, когда собранные образцы сгруппированы в некоторых частях изучаемой области (в результате плохо продуманного метода выборки данных), из-за чего некоторые объекты имеют больше соседей на близких расстояниях и, соответственно, значения атрибутов могут группироваться из-за пространственной автокорреляции. Стандартизация строк смягчает эти эффекты.

Стандартизацию строк следует использовать, когда объекты-полигоны отражают административные или любые другие искусственно определяемые области. Данные переписи или социально-экономических исследований часто агрегируются из более крупных наборов данных в полигоны неравных размеров. А как мы уже знаем, методы формирования полигонов могут повлиять на результаты пространственного анализа, это называется проблемой изменяющейся пространственной единицы (см. раздел 1.3). По этой причине веса лучше стандартизировать как процент от суммы весов соседей. Это позволяет обобщить результаты на более высоком уровне и смягчить проблему проектирования зон.

Стандартизацию строк также рекомендуется применять к двоичным матрицам весов, получаемым при использовании методов фиксированного диапазона расстояний и смежности полигонов, чтобы учесть разницу в количестве соседей у наблюдаемых объектов. Наконец, стандартизация строк не требуется

при анализе точечных объектов (например, дорожно-транспортных происшествий, криминальных событий), которые не являются результатом агрегирования и не зависят от процедуры выборки.

1.10. ЗАКЛЮЧИТЕЛЬНЫЕ ЗАМЕЧАНИЯ К ГЛАВЕ

- Процедура пространственного анализа должна включать этапы описания, исследования и объяснения.
- Пространственные данные могут быть точками, линиями, полигонами или полями (например, пикселями).
- Пространственные данные – особенные и требуют особого обращения.
- Пространственная автокорреляция – основная причина, почему к пространственным данным нельзя применять классические статистические подходы.
- Выбор метода концептуализации пространственных отношений во многом зависит от изучаемой проблемы.
- Неправильный выбор метода концептуализации может привести к недостоверным статистическим результатам.
- Очень большое или очень маленькое фиксированное расстояние может исказить результаты и привести к неправильной интерпретации.
- Все методы концептуализации, представленные в этой главе, предполагают непрерывное и одинаковое влияние расстояния во всех направлениях (изотропность). Однако в действительности чаще наблюдается пространственная неоднородность.
- Наличие неизотропного (прерывистого) пространства не запрещает описанных выше методов, а лишь требует выбора подходящих методов.
- Географический масштаб определяет представление данных, порядок сбора данных, методы анализа, вопросы исследования и результаты. По этой причине первой задачей любого географического анализа обычно является выбор правильного масштаба.
- Выбор подходящего масштаба анализа определяет размер и форму области, в которой вычисляются пространственные статистики после выбора географического масштаба. По сути, это уровень представления пространственных явлений, тесно связанный с рассматриваемой проблемой.
- Довольно часто меньшие расстояния лучше подходят для географического анализа в местном масштабе.
- Масштаб анализа также можно рассматривать как географическую протяженность, в которой анализируются пространственная и временная изменчивости посредством построения соответствующей матрицы пространственных весов.
- Не следует предполагать, что статистическая взаимосвязь, имеющаяся на агрегированном уровне, описывает поведение людей (экологическая ошибка).
- Влияние пространственного объекта на соседей зависит не только от расстояния. Поэтому необходимо предусмотреть параметры, отличные от расстояния. Это сделает географический анализ более полным и точным.
- Стандартизация строк почти всегда необходима, когда в анализе участвуют объекты-полигоны.

ВОПРОСЫ И ОТВЕТЫ

В этом разделе приводятся краткие ответы на вопросы. Чтобы получить более подробные ответы, обращайтесь к соответствующим разделам этой главы.

Вопрос 1. Что такое пространственный анализ и геопространственный анализ? Есть ли разница между этими терминами?

Ответ 1. Пространственный анализ – это совокупность методов, статистик и приемов, объединяющих такие понятия, как местоположение, площадь, расстояние и взаимодействие для анализа, исследования и объяснения в географическом контексте закономерностей или особенностей в наблюдениях с пространственной привязкой, возникающих в результате процессов, действующих в пространстве. Геопространственный анализ – это совокупность методов, приемов и моделей пространственного анализа, объединенных в географические информационные системы (ГИС). Геопространственный анализ подкрепляется богатыми возможностями ГИС и используется для разработки новых моделей или интеграции существующих в среду ГИС. Термин «геопространственный анализ» можно использовать как синоним термина «пространственный анализ»; однако, строго говоря, пространственный анализ является частью геопространственного анализа и в большей степени фокусируется на методах, чем на технологиях.

Вопрос 2. Какие этапы пространственного анализа представлены в этой книге?

Ответ 2. Шаг А: описание (что). Это первый шаг в процессе пространственного анализа. Он описывает набор данных с помощью *описательной статистики*. Шаг Б: исследование (где). На втором этапе применяется исследовательский анализ пространственных данных (Exploratory Spatial Data Analysis, ESDA) для изучения данных, обнаружения выбросов, проверки основных предположений и выявления в них тенденций и ассоциаций, таких как наличие пространственной автокорреляции или пространственной кластеризации. Шаг В: объяснение (почему/как). На последнем этапе применяется статистический анализ, объясняющий причины и следствия с помощью моделей. На этом этапе мы пытаемся ответить на вопросы «почему?» и «как?».

Вопрос 3. Какие ключевые вопросы пытается решить пространственный анализ, если судить по основным этапам решения данной проблемы?

Ответ 3. Пространственный анализ пытается ответить на три основных вопроса: **что, где и как/почему**. Пытаясь найти ответ на вопрос «**что?**», мы исследуем значения конкретных переменных. Например, вычисляем среднее, максимальное и минимальное значения переменной. Также, применяя методы исследовательского анализа пространственных данных и картирования, мы выясняем особенности распределения переменных в пространстве. Затем мы задаем вопрос «**где?**». Какие районы характеризуются низкими и высокими доходами? Существует ли какая-либо пространственная группировка районов по распределению дохода на душу населения? Где они располагаются? Где в городе находятся горячие точки преступности? Чтобы получить ответы на эти вопросы, требуется применение методов пространственной статистики, позволяющих количественно оценить результаты как значимые или нет и выявить интересные пространственные закономерности, которые невозможно

обнаружить иным способом. На заключительном этапе пространственного анализа мы пытаемся смоделировать отношения, чтобы ответить на вопросы «как?» и «почему?». Например, мы можем создать модель, которая показывает, как значение переменной связано с ее местоположением.

Вопрос 4. Каковы три уровня измерения переменной/атрибута? Приведите примеры каждого и назовите основные отличия.

Ответ 4. Переменные/атрибуты группируются по трем уровням измерения: номинальному, порядковому и интервальному или относительному. Номинальные переменные – это переменные, значения которых нельзя упорядочить. Например, переменная, отражающая расу, может иметь значения белая = 1, азиатская = 2, латиноамериканская = 3. Номинальные переменные нельзя складывать или вычитать. Порядковые переменные – это переменные, которые можно упорядочивать, но численные различия которых не имеют смысла и не могут быть вычислены. Например, переменная «студент» может принимать следующие значения: «отличник» = 1, «хорошист» = 2, «посредственный» = 3. Это порядковая переменная, потому что ее значения можно упорядочить сверху вниз (или наоборот), но в их вычитании нет смысла («отличник» – «хорошист» = -1). Относительные переменные – это переменные, для которых каждое наблюдение можно значимо выразить в числовом виде. Например, численность населения – это переменная отношения.

Вопрос 5. Что такое концептуализация пространственных отношений? Почему это важно?

Ответ 5. Концептуализация пространственных отношений – это моделирование отношений и взаимодействий между объектами в пространстве. Проще говоря, это математическое определение таких понятий, как «близко», «далеко», «смежный», «непрерывный», «соседний», «окрестность» и «расстояние», для набора пространственных объектов в виде определенных значений или функций. Методы и приемы пространственного анализа основаны на моделировании пространственных отношений между пространственными объектами. Концептуализация пространственных отношений важна, потому что чем ближе к реальности концептуализация пространственных отношений, тем точнее результаты статистических исследований и моделей. С другой стороны, если применяемый метод концептуализации не отражает внутреннюю структуру пространственных отношений в наборе данных, результаты анализа будут вводить в заблуждение.

Вопрос 6. Что такое сфера влияния и область равнозначности?

Ответ 6. Сфера влияния – это фиксированное расстояние, до которого все пространственные объекты внутри определяемой им области имеют одинаковый вес, а за ее пределами – нулевой вес. Фиксированное расстояние можно использовать как точку отсечки, начиная с которой применяется функции затухания. Область равнозначности – это диапазон расстояний до точки отсечки. Внутри этой области все объекты интерпретируются одинаково, а за ее пределами объекты взвешиваются в соответствии с их удаленностью (с применением выбранной функции затухания).

Вопрос 7. Что такое окрестность, и почему это понятие важно для пространственного анализа?

Ответ 7. В контексте пространственного анализа окрестность – это географически локализованная область, к которой применяются локальный пространственный анализ и статистические методы, основывающиеся на гипотезе о том, что объекты в пределах этой области взаимодействуют сильнее, чем объекты за ее пределами. Определение окрестности необходимо для точного вычисления пространственных статистик. Большая часть статистических методов требует определения соседства и построения пространственной матрицы весов, которая отражает интенсивность отношений между пространственными объектами в окрестности.

Вопрос 8. Что такое полигоны Тиссена и триангуляция Делоне? Каково их назначение?

Ответ 8. Полигоны Тиссена, также называемые «полигонами Вороного», являются полигонами близости. Триангуляция Делоне делит пространство точечных объектов, играющих роль центроидов, на треугольники, полигоны близости которых имеют общие границы. Оба метода используются для деления пространства на области с конкретными характеристиками, которые можно использовать для определения окрестностей, пространственной интерполяции или оценки областей, охватываемых общественными службами или коммерческими предприятиями.

Вопрос 9. Что такое пространственные веса и матрица пространственных весов? Каково их назначение?

Ответ 9. Пространственные веса – это числовые значения, отражающие расстояние, время или стоимость между целевым и любым другим пространственным объектом в наборе данных или в выбранной окрестности. Пространственные веса количественно определяют пространственные или пространственно-временные отношения между пространственными точками в окрестности. Матрица пространственных весов хранит пространственные веса, отражающие степени связи между объектами внутри определенной окрестности.

Вопрос 10. Что такое стандартизация строк, и когда ее следует применять?

Ответ 10. Стандартизация строк – это процесс масштабирования пространственных весов в диапазон от 0 до 1. Процедура стандартизации строк заключается в делении каждого веса либо на сумму всех весов в его строке, либо на сумму весов соседних объектов. То есть результаты корректируются так, чтобы количество соседей не влияло на конечные результаты. Стандартизацию строк рекомендуется использовать в основном в двух случаях: (а) когда есть вероятность систематической ошибки в распределении пространственных объектов и значений их атрибутов из-за плохо продуманного метода выборки данных и (б) когда данные были агрегированы из более крупных наборов данных в полигоны неравных размеров.

ПРАКТИКА

Практическая работа 1. Пространственный анализ инвестиций на рынке недвижимости

Каждая практическая работа состоит из двух разделов, а именно раздела А (ArcGIS) и раздела Б (GeoDa). В разделе А приводятся: пошаговые инструкции по выполнению практических упражнений с использованием ArcGIS, примеры интерпретации результатов и заключительные замечания. В разделе Б те же упражнения решаются с использованием возможностей GeoDa. Благодаря этому читатель получает возможность выбрать инструмент для решения практических заданий – ведущее коммерческое программное обеспечение или хорошо зарекомендовавшее себя бесплатное программное обеспечение с открытым исходным кодом. Интерпретация результатов и выводы не повторяются в разделе Б и приводятся только в абзацах «Интерпретация результатов» в разделе А. Читатель должен внимательно изучить эти абзацы, потому что они не зависят от используемого программного обеспечения. Кроме того, разделам А и Б предшествуют разделы «Общий прогресс» и «Сфера анализа», в которых приводятся дополнительные сведения, чтобы помочь лучше понять процесс пространственного анализа и практическую ценность анализа в каждой практической работе.

Общий прогресс

Пространственный анализ/практика



Рис. 1.7. Практическая работа 1 в общем прогрессе

Сфера анализа

Риелтор хочет не просто перечислить своим клиентам выставленную на продажу недвижимость, но подобрать лучшие варианты. Девиз компании: «местоположение, местоположение и еще раз местоположение», поэтому она собирает географические и социально-экономические данные (например, данные переписи населения, данные о доходах, данные о преступности) и предлагает аналитику местоположений в виде расширенной пространственной статистики и ГИС. Проводя анализ, компания стремится стать более конкурентоспособной и надежной, иметь возможность четко отвечать на вопросы клиентов с учетом их инвестиционных потребностей. Чем больше информации будет доступно о местоположении и прилегающих территориях, тем выше вероятность успешных инвестиций. Пространственный анализ предлагает все необходимые инструменты и математический аппарат для получения ответов на многие вопросы, визуализированные с помощью карт ГИС, что обычно намного лучше, чем личное мнение или общее представление о местоположении.

Эта практическая работа решает следующую задачу. Инвестор желает подобрать место для нового кафе и обращается в компанию по недвижимости за советом и консультацией.

Инвестора интересует не только арендная плата и сопутствующие расходы (см. врезку 1.4), он также хочет найти подходящий район, исходя из следующих целей (см. табл. 1.2):

Поскольку предполагается предоставлять услуги высшего качества, кафе должно располагаться в районе, жители которого (то есть потенциальные клиенты) имеют **высокие доходы**.

Кафе должно располагаться в районе с **низким уровнем преступности**.

Желательно, чтобы клиенты охотно тратили больше денег, чем в среднем, на дополнительные продукты, сопутствующие кофе. Следует выявить тех, кто проявляет **высокую покупательскую активность**.

Следует определить **движущие социально-экономические факторы**, лежащие в основе ежемесячных расходов людей (в том числе на услуги, связанные с посещением кафе).

Таблица 1.2. Цели и методы каждой практической работы

№ п/п	Цели	Методы	Практическая работа
1	Местоположение: высокий доход	Картирование, пространственная автокорреляция	2, 4, 5
2	Местоположение: низкий уровень преступности	Картирование, центрография, пространственная автокорреляция	3, 4, 5
3	Кластеризация: высокая покупательская активность	Кластеризация	5
4	Моделирование: выявление движущих факторов	Регрессия, пространственная регрессия	2, 6, 7

Четвертая цель не связана напрямую с поиском оптимального местоположения, как первые три цели. Она связана с выявлением пространственных отношений, которые можно использовать для моделирования, проникновения на рынок и ана-

лиза клиентуры. Для подобного анализа требуется наличие такой информации, как предпочтения потребителей, повседневные привычки, тип работы и суммы, потраченные на посещение кафе. По соображениям наглядности и краткости анализа мы сосредоточимся лишь на основных социально-экономических параметрах.

Врезка 1.4. Полное исследование также должно включать такие факторы, как местонахождение конкурентов, размер арендной платы и сопутствующих расходов, доступность общественного транспорта (например, метро), размер бюджета, ежедневное количество посетителей, численность постоянно проживающего населения и работающих поблизости. Ответить на все эти вопросы в книге просто невозможно, поэтому решение данной задачи под силу только тем, кто занимается пространственным анализом. Полученные результаты можно объединить с результатами анализа рынка, проводимого маркетологами или бизнес-специалистами, чтобы в конечном итоге получить надежный бизнес-план. Пространственное планирование является ключом к успеху не только в бизнесе, но и в реализации национальной, региональной и местной политики по различным вопросам, таким как образование, здравоохранение, труд, чрезвычайные ситуации и государственное управление.

Несмотря на кажущуюся простоту целей нашего исследования, анализ может оказаться бесконечным (что вполне возможно в пространственном анализе). Поэтому мы сосредоточимся на относительно небольшом множестве важных вопросов и рассмотрим дополнительные возможности моделирования. Чтобы ответить на поставленные вопросы, мы используем многие методы пространственного анализа, такие как исследовательский анализ пространственных данных, пространственная автокорреляция, кластеризация данных, пространственная кластеризация и пространственная регрессия (см. табл. 1.2 и табл. 1.3).

Пространственный анализ будет состоять из трех этапов: описание, исследование и объяснение. При этом мы дадим ответы на три основных вопроса: «что?», «где?» и «как/почему?» (см. рис. 1.1 и рис. 1.7).

- Чтобы ответить на вопрос «что?», мы изучим конкретные показатели. Например, выясним средний доход в изучаемой области, какая доля населения имеет доходы выше определенного значения, имеются ли выбросы в общей картине доходов. Это даст нам начальное представление о социально-экономическом характере изучаемой территории. Применяв исследовательский анализ пространственных данных в сочетании с методами картирования, мы получим предварительное представление о распределении переменных в пространстве.
- Затем мы попробуем ответить на вопрос «где?». Определим области с низким и высоким доходами, выясним, имеет ли место пространственная группировка областей с высокими доходами, существуют ли горячие точки преступности. Ответы на эти вопросы помогут получить надежные выводы в ходе анализа, основанного на пространственной статистике, которая количественно оценивает значимость результатов, а также выявляет интересные пространственные закономерности, которые невозможно обнаружить иными средствами.

- Наконец, мы проведем еще более глубокий анализ, чтобы ответить на вопросы «как?» и «почему?». Для этого создадим несколько регрессионных и эконометрических моделей, имитирующих ежемесячные расходы (независимые друг от друга) на основе набора зависимых переменных (например, местоположения, дохода).

Данные

Область исследования – город Афины, Греция (далее просто «город»). Пространственные данные привязаны к городским округам (полигоны; см. табл. 1.4). Социально-экономические данные относятся к переписи населения 2011 года (см. табл. 1.5). Некоторые данные были изменены по соображениям конфиденциальности.

Задачи

Основные задачи перечислены в табл. 1.3.

Таблица 1.3. Задачи, решаемые в практических работах, согласно рабочему процессу описание – исследование – объяснение

Задача	Практическая работа	Инструменты	Цели
<i>Что (описание)</i>			
Набор данных и исследуемая область. Создание и отображение карт	1	Символы	Определить задачу, область исследования и базу данных. Упростить отображение распределения переменных на карте (например, плотность населения)
Описание и отображение переменных на карте. Вычисление и отображение z-оценок	2	Фоновые картограммы, гистограммы, базовые статистики (асимметрия, эксцесс и др.), нормальный график КК, коробчатые диаграммы, отображение z-оценок	Найти области с высоким или низким доходом людей, проживающих в них. Определить, искажено ли распределение переменных. Определить, соответствует ли распределение переменных нормальному распределению. Выявить выбросы
Провести корреляционный и попарный корреляционный анализы переменных в наборе данных	2	Диаграммы рассеяния, матрица диаграммы рассеяния	Выявить наличие линейных отношений между переменными, чтобы упростить моделирование на более позднем этапе
<i>Где (исследование)</i>			
Анализ и оценка географического распределения криминальных событий в исследуемой области	3	Среднее, медиана, стандартное отклонение, эллипс стандартного отклонения	Выявление пространственных и временных тенденций совершения преступлений, чтобы выбрать более удачное место для кафе
Анализ структуры точечных закономерностей	3	Среднее ближайшее соседство, K-функция Рипли	Определение случайности пространственной структуры преступлений, ее рассредоточенности или сгруппированности, чтобы выяснить, на каком расстоянии группировка или рассредоточенность выражена наиболее ярко

Задача	Практическая работа	Инструменты	Цели
Создание карт плотности	3	Ядерная оценка плотности	Создание гладкой карты, охватывающей исследуемую область, с отображением участков с низкой и высокой плотностью совершения преступлений
Географические выбросы	3	Объект в точку, ближайший объект	Выявление географических выбросов и их удаление перед вычислением пространственных статистик
Выявление пространственной автокорреляции по величине доходов	4	Матрица пространственных весов, общий индекс / Морана, пошаговая пространственная автокорреляция, локальный индекс / Морана, общая G-статистика	Концептуализация пространства с созданием матрицы пространственных весов. Выявление группировки жителей с высоким или низким доходом в пространстве. Поиск горячих и холодных точек. Определение масштаба анализа
Выявление пространственной автокорреляции преступлений	4	Оптимизированный анализ горячих точек	Более точное определение горячих или холодных точек криминальных событий. Местоположения, совпадающие с горячими точками криминальных событий, следует исключить из возможных мест для размещения кафе
Многофакторная кластеризация данных	5	Кластеризация методом k средних	Проведение геодемографического анализа на основе множества социально-экономических переменных
Пространственно-ограниченная многофакторная кластеризация	5	Рассеяние	Пространственная кластеризация (регионализация)
Анализ подобия	5	Поиск подобия (косинусное сходство)	Выявление участков, подобных целевому, которые могут служить альтернативами для размещения кафе
Обобщение	5	Выбор по атрибутам/местоположению, экспорт, повторная переклассификация	Выбор лучшего места на основе критериев оценки
Как/почему (объяснение)			
Моделирование отношений	6	Исследовательская регрессия, метод наименьших квадратов, географически взвешенная регрессия	Моделирование расходов (независимая переменная) на основе набора зависимых переменных для определения факторов, способствующих покупательской активности
Моделирование отношений	6	Пространственный лаг, пространственная ошибка, пространственный режим	Моделирование расходов (независимая переменная) на основе набора зависимых переменных с использованием пространственной эконометрики. Выявление связей между расходами и пространственными переменными

Структура набора данных

Структура наборов данных и сопутствующих файлов в папке Booklabs показана на рис. 1.8 (см. также врезку 1.5).

Врезка 1.5. Загрузите данные для практической работы, находящиеся по адресу: www.cambridge.org/9781108498982, и сохраните их в папке I:\BookLabs\ (вы можете сохранить данные на другом диске, например C:\).

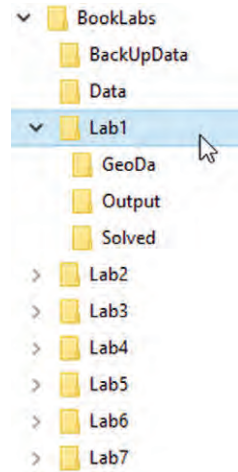


Рис. 1.8. Структура набора данных

В папке BackUpData хранится резервная копия исходных данных, хранящихся в папке Data. В случае повреждения, случайного удаления или неправильного изменения данных просто скопируйте набор данных из папки BackUpData в папку Data. Все остальные папки (например, Lab1) хранят файлы .mxd для каждой конкретной практической работы, а также подпапки Output и Solved. Папка Output используется для хранения файлов с результатами анализа, таких как шейп-файлы, графики, PDF-файлы или изображения. Это основная папка для анализа данных. Папка Solved содержит решение (например, Solved_Lab1_GettingToKnowDataSet.mxd) с окончательным набором данных после правки и применения инструментов. Используйте эту папку для сравнения с вашими результатами. Дополнительная папка GeoDa используется только в упражнениях, которые решаются с помощью программы GeoDa (см. раздел В).

Пространственные данные (в папке Data), используемые в этой книге, описаны в табл. 1.4.

Таблица 1.4. Пространственные данные

Файлы	Содержит
City.shp	90 округов (полигонов), составляющих область исследования
Downtown.shp	Внешний полигон центральной части города
Assaults.shp	Точки криминальных событий – нападений
Burglaries.shp	Точки криминальных событий – краж со взломом
Crime.shp	Точки криминальных событий – нападений и краж со взломом вместе

Данные атрибутов шейп-файла City описаны в табл. 1.5.

Таблица 1.5. Социально-экономические данные относятся к переписи 2011 года (для конфиденциальности некоторые данные изменены). Эти переменные являются атрибутами в City.shp

Атрибуты	Описание
Population	Общая численность населения
Density	Плотность населения (человек на квадратный метр)
Foreigners	Доля (%) иностранцев (за исключением коренных греков)
Owners	Доля (%) населения, владеющего домами (не платящего аренду)
SecondaryE	Доля (%) населения со средним образованием или ниже
University	Доля (%) населения со степенью бакалавра
PhD_Master	Доля (%) населения со степенью магистра или выше
Income	Средний годовой доход на душу населения в евро
Insurance	Среднемесячная плата за страховку на душу населения (в евро)
Rent	Среднемесячная арендная плата (в евро)
Expenses	Среднемесячные расходы на душу населения на ежедневные покупки, например продукты, кофе (в евро)
Area	Площадь округа в квадратных метрах
Postcode	Пятизначный почтовый индекс

Общие положения

Словом «**ДЕЙСТВИЕ:**» обозначается начало последовательности взаимодействий с программным обеспечением.

Символы, используемые для описания действий, перечислены в табл. 1.6.

Таблица 1.6. Основные символы, используемые в описаниях взаимодействий с программным обеспечением

Символы	Описание
>	Следующее действие
TOC	Содержимое (Table of Contents)
RC	Щелчок правой кнопкой мыши (Right-click)
DC	Двойной щелчок левой кнопкой мыши (Double-click)
TAB =	Выбор вкладки
=	Присвоить значение

A. ArcGIS

Упражнение 1.1. Знакомство с данными и областью исследования

Это упражнение описывает, как рассчитать и отобразить на карте плотность населения.

Совет для ArcGIS: все файлы mxd были созданы с использованием версии ArcGIS 10.4. Если у вас установлена более ранняя версия, откройте пустой файл mxd и вставьте шейп-файлы из всех упражнений в папке Data.

Перейдите в папку, где вы сохранили набор данных, сопровождающий книгу, и дважды щелкните на файле Lab1_GettingToKnowDataSet.mxd.

Например: I:\BookLabs\Lab1\Lab1_GettingToKnowDataSet.mxd

Совет: путь к файлу можно ввести непосредственно в Проводнике Windows (просто измените имя буквы диска; если вы сохранили данные на диске C, замените I на C).

Прежде всего сохраните исходный файл с новым именем:

Главное меню > File (Файл) > Save As (Сохранить как) > My_Lab1_GettingToKnowDataSet

в папку I:\BookLabs\Lab1\Output:

ТОС (Таблица содержания) > RC (щелчок правой кнопкой) на слое City > Open Attribute Table (Открыть таблицу атрибутов, см. рис. 1.9)

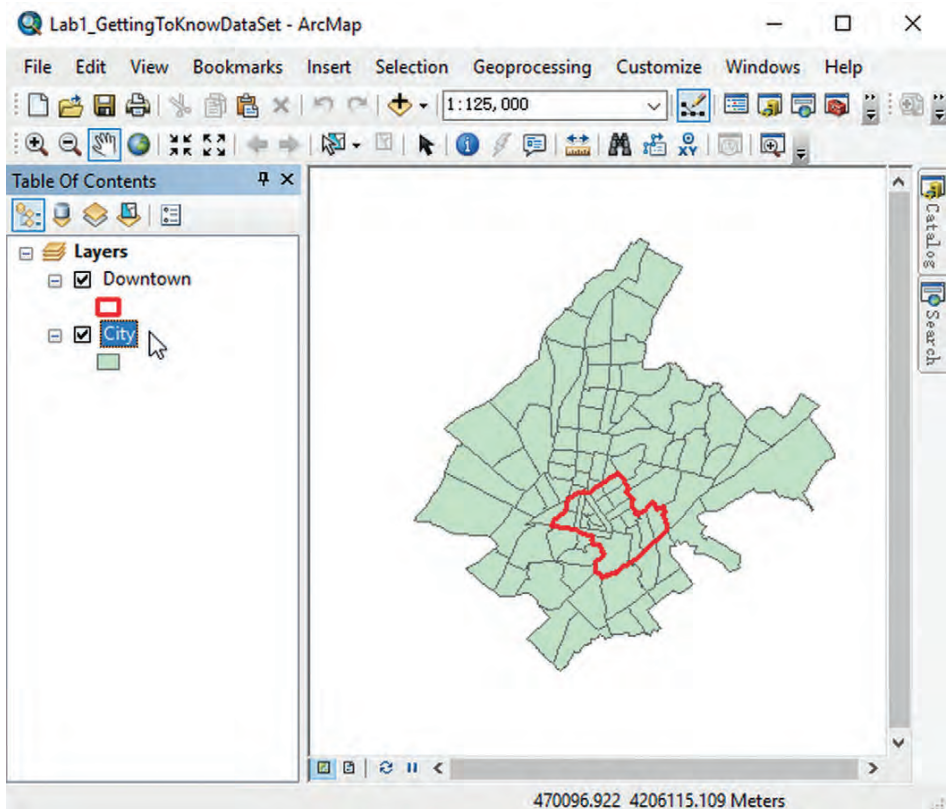


Рис. 1.9. Исследуемая область

Table

City

	PostCode	Population	Density	Foreigners	Owners	SecondaryE	University	Phd_Master	Income
▶	10434	7063	0.02126	28.67	36.28	51.28	19.82	1.81	15451.62
	10435	4900	0.017464	22.04	37.41	42.29	9	0.9	10992.84
	10436	3221	0.015488	35.7	30.22	45.76	11.27	0.56	10294.89
	10440	10919	0.045376	29.01	34.68	51.41	13.75	0.98	13460.38
	10442	8038	0.006106	12.88	49.19	46.04	8.55	0.54	12498.29
	10443	25933	0.021544	12.45	46.67	46.7	10.25	0.59	12352.07
	10444	19455	0.023705	13.19	42.69	47.2	11.5	0.72	11548.98
	10445	25094	0.03117	16.61	40.52	48.81	12.31	0.69	11981.84
	10446	22598	0.048898	27.22	36.02	49.94	13.99	0.76	11195.79
	10553	1317	0.006529	49.81	35.36	36.07	10.1	0.38	12694.28
	10554	419	0.003104	53.7	18.72	34.84	9.55	1.19	14777.12
	10557	651	0.00116	45.31	25.61	27.8	14.9	3.99	26649.54
	10560	142	0.002897	33.1	31.08	60.56	15.49	1.41	20021.22
	10562	329	0.007299	74.55	17.65	26.99	4.87	0.91	15492.32

(0 out of 90 Selected)

City

Рис. 1.10. Всего в таблице атрибутов хранятся сведения о 90 округах, по 17 полей для каждого, из которых 11 являются социально-экономическими переменными (см. табл. 1.5)

Layer Properties

General Source Selection Display **Symbology** Fields Definition Query Labels Joins & Relates Time HTML Popup

Show:

Features
Categories
Quantities
Graduated colors
Graduated symbols
Proportional symbols
Dot density
Charts
Multiple Attributes

Draw quantities using color to show values. Import...

Fields
Value: Population
Normalization: none

Classification
Natural Breaks (Jenks)
Classes: 5
Classify...

Color Ramp: [Color Ramp]

Symbol	Range	Label
[Yellow]	3.000000 - 2834.000000	3.000000 - 2834.000000
[Light Orange]	2834.000001 - 7063.000000	2834.000001 - 7063.000000
[Orange]	7063.000001 - 12378.000000	7063.000001 - 12378.000000
[Dark Orange]	12378.000001 - 17729.000000	12378.000001 - 17729.000000
[Red]	17729.000001 - 25933.000000	17729.000001 - 25933.000000

Show class ranges using feature values

Advanced

OK Cancel Apply

Рис. 1.11. Диалог Layer properties (Свойства слоя) для настройки символов

ТОС (Таблица содержания) > RC на City > Properties (Свойства) > TAB = Symbology (Символы) > Quantities (Количество) > Graduated colors (Градуированные цвета) > Value (Значение) = Population (см. рис. 1.11)

Color Ramp (Цветовая схема) = от желтого до коричневого

Щелкните на кнопке Classify (Классифицировать).

Введите следующие значения в окно Break Values (Граничные значения) справа внизу (см. рис. 1.12):

Break Values (Граничные значения) > 2000 > Enter > 5000 > Enter > 10000 > Enter > 15000 > Enter > 30000 > Enter > ОК

RC Label (Подпись) > Format Labels (Формат подписей) > выберите элемент Numeric (Числовой) > Rounding (Округление) > Number of decimal places (Число десятичных знаков) = 2 > ОК (см. рис. 1.13)

Щелкните на кнопке Apply (Применить) > ОК.

ТОС (Таблица содержания) > RC на City > Save As Layer File (Сохранить как файл слоя, см. рис. 1.14) > Name (Имя) = Population.lyr

В папку I:\BookLabs\Lab1\Output

добавьте слой в ТОС (Таблица содержания).

Сохраните.

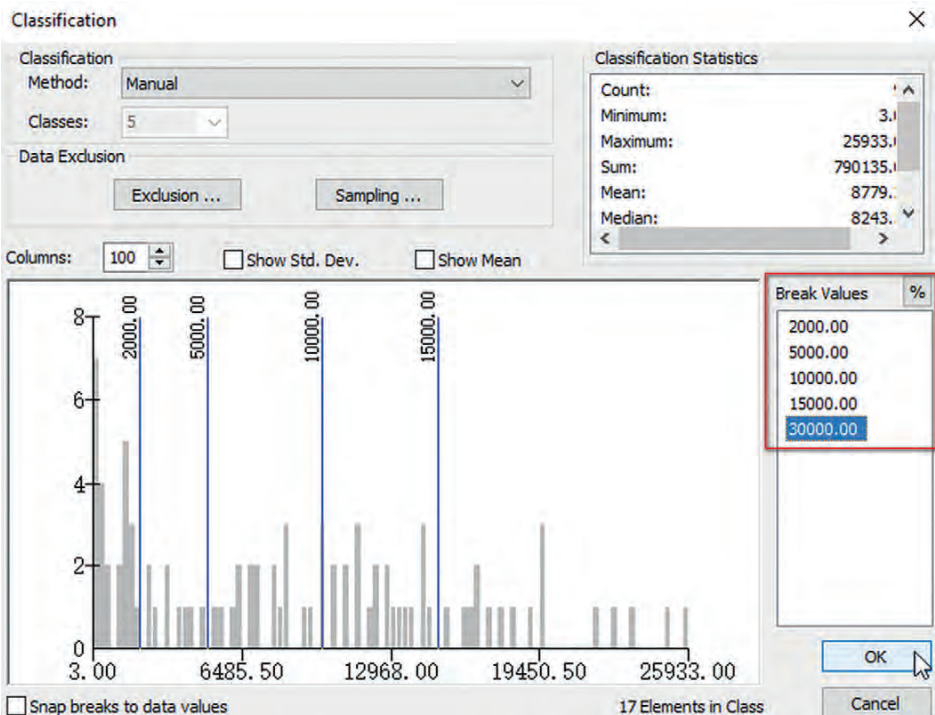


Рис. 1.12. Настройка диапазонов значений категорий

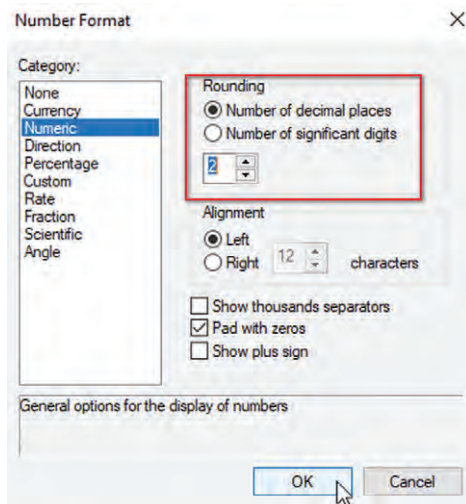


Рис. 1.13. Определение числового формата

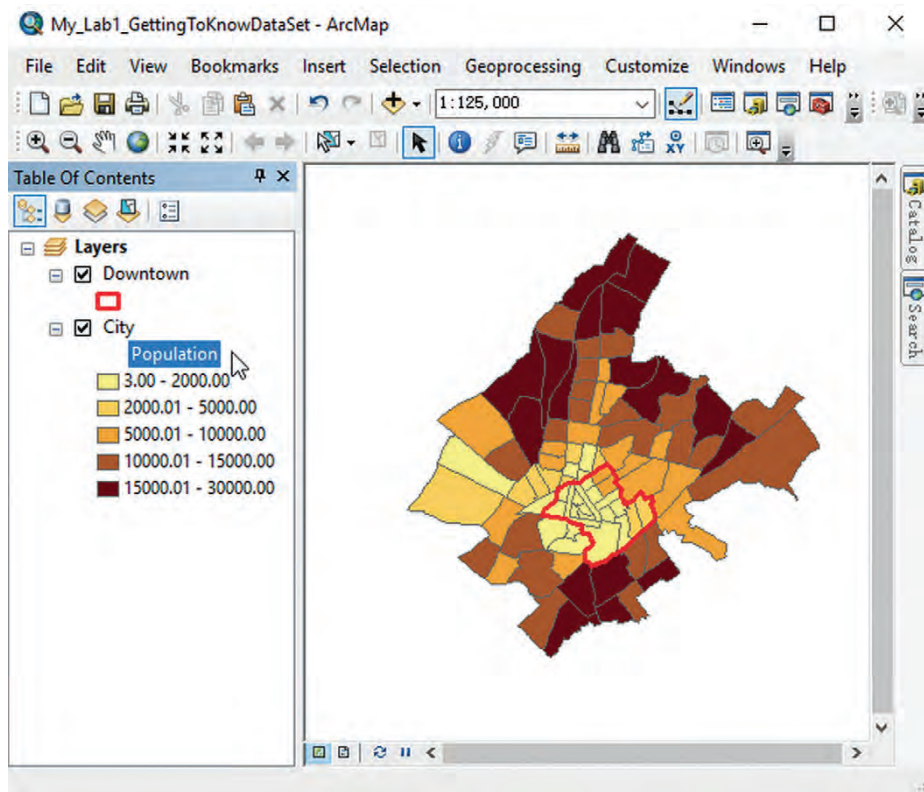


Рис. 1.14. Фоновая картограмма распределения населения

Интерпретация результатов: исследуемая область включает 90 почтовых округов (пространственные объекты; см. рис. 1.9). Открыв таблицу атрибутов City, можно исследовать значения переменных всех пространственных объектов (см. рис. 1.10). В центральных округах проживает меньше людей, чем на окраинах (см. рис. 1.14). Поскольку округа в центре меньше, рекомендуется дополнительно отобразить плотность населения, так как это поможет получить более полное представление о распределении населения в пределах исследуемой области.

ДЕЙСТВИЕ: вычисление и отображение на карте плотности населения.

RC на слое City (не Population.lyr) > Properties (Свойства) > TAB = Symbology (Символы) > Quantities (Количество) > Graduated colors (Градуированные цвета)

Value (Значение) = Population

Normalization (Нормирование) = Area

Color Ramp (Цветовая схема) = от светло- до темно-зеленого

Classes (Классов) = 4

Щелкните на кнопке Classify (Классифицировать) > Break Values (Граничные значения) > 0.0100 > Enter > 0.0200 > Enter > 0.0300 > Enter > 1.000 > OK.

Значения плотности и граничные значения измеряют численность населения на квадратный метр. То есть 0,01 означает, что на площади 1 м² живет 0,01 человека, или 1 человек на 100 м².

RC Label (Подпись) > Format Labels (Формат подписей) > Numeric (Числовой) > Number of decimal places (Число десятичных знаков) = 2 > OK > Apply (Применить) > OK

TOC (Таблица содержания) > RC на City > Save As Layer File (Сохранить как файл слоя) > Name (Имя) = PopDensity.lyr

В папку I:\BookLabs\Lab1\Output

добавьте слой в TOC (Таблица содержания).

Главное меню > File (Файл) > Save (Сохранить)

Совет: при сохранении данных о населении, нормированных по площади, в файл слоя (.lyr) сохраняется представление плотности. Когда слой добавляется в таблицу содержания, ему дается имя исходного шейп-файла (в этом примере City), а не имя, с которым он был сохранен (например, PopDensity.lyr, см. рис. 1.15).

Интерпретация результатов: на картограмме плотности населения видно, что в центре города (центр города – красный полигон) плотность населения (в большинстве округов) ниже, чем на окраинах (см. рис. 1.15). Наиболее густонаселенные округа находятся в северной части города. Кроме того, низкая плотность населения в центре города, вероятно, объясняется концентрацией в ней деловых кварталов и историческими причинами (здесь меньше количество постоянных жителей). Можно также отметить сходство с картой

распределения населения (см. рис. 1.14), но в целом карта плотности населения позволяет лучше понять, как население распределено по округам. Например, округа в центре города точнее характеризуются плотностью населения.

Совет для ArcGIS: процедура нормирования в ArcGIS основана на делении одной переменной на другую. Это дает возможность рассчитать темп изменений (прирост населения), доли (например, долю земного покрова), числовые показатели на душу населения (например, доход) и плотности (например, плотность населения). Нормирование не следует путать с процессом нормализации, который изменяет масштаб данных, приводя их в диапазон $[0,1]$ или $[-1,1]$ (см. раздел 2.4). Инструмент нормирования в ArcGIS – это операция деления двух переменных. Например, если известен совокупный доход всех, проживающих на территории каждого округа, можно вычислить и отобразить доход на душу населения. Недостаток этого инструмента в том, что он не позволяет сохранить полученные значения плотности населения в новом поле. Мы просто отображаем результаты. Однако мы можем получить это соотношение, используя вычисляемые поля. Инструмент нормирования в ArcGIS можно использовать для проверки различных комбинаций соотношений. Как только мы решим, какое из соотношений сохранить, мы можем рассчитать значения с помощью вычисляемого поля.

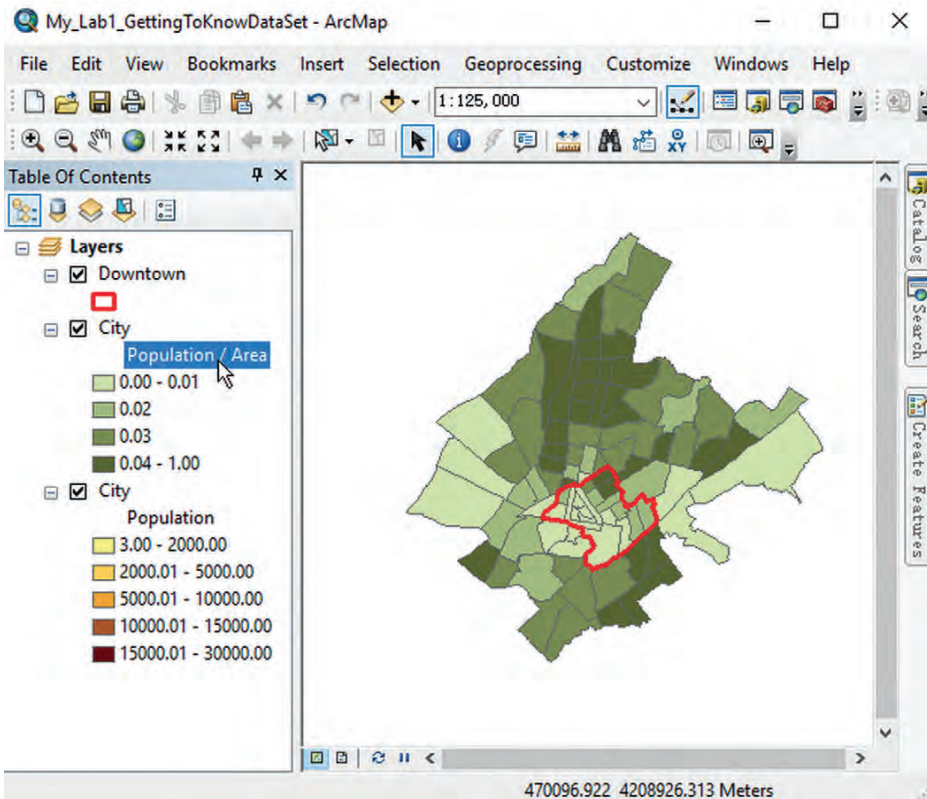


Рис. 1.15. Плотность населения

В. GeoDa

Врезка 1.6. Загрузите и установите бесплатное программное обеспечение GeoDa с открытым исходным кодом, которое можно получить по адресу <http://geodacenter.github.io>. Загляните также в раздел документации, где можно найти подробное руководство. Приложение GeoDa было разработано доктором Люком Анселином (Luc Anselin) и его командой, чей вклад в область пространственного анализа трудно переоценить.

Упражнение 1.1. Знакомство с данными и областью исследования

Это упражнение описывает, как рассчитать и отобразить на карте плотность населения.

Используемые инструменты GeoDa: Category Editor (Редактор категорий), Zoom (Масштаб), Table (Таблица).

ДЕЙСТВИЕ: открытие набора данных и отображение информации о населении.

Перейдите в папку, где вы сохранили набор данных, сопровождающий книгу, и дважды щелкните на файле `click Lab1_GettingToKnowDataSet_GeoDa.gda` в папке GeoDa.

Например:

`I:\BookLabs\Lab1\GeoDa\Lab1_GettingToKnowDataSet_GeoDa.gda`

Совет: путь к файлу можно ввести непосредственно в Проводнике Windows (просто измените имя буквы диска; если вы сохранили данные на диске C, замените I на C). Пространственные данные для упражнений GeoDa хранятся в папке GeoDa, а не в папке Data. Пространственные данные и значения атрибутов показаны в табл. 1.4 и 1.5.

Главное меню > щелкните на кнопке Table (Таблица; см. рис. 1.16 и 1.17). Затем щелкните на окне Map – CityGeoDa, чтобы активировать его.

Главное меню > Map (Карта) > Custom Breaks (Пользовательские границы) > Create New Custom Breaks (Создать новые пользовательские границы) (см. рис. 1.18)

В окне Variable Settings (Настройки переменных; см. рис. 1.19) выберите:

Population > OK

В окне New Custom Categories Title (Название новых пользовательских категорий) введите Custom Breaks (Population) и щелкните на кнопке OK.

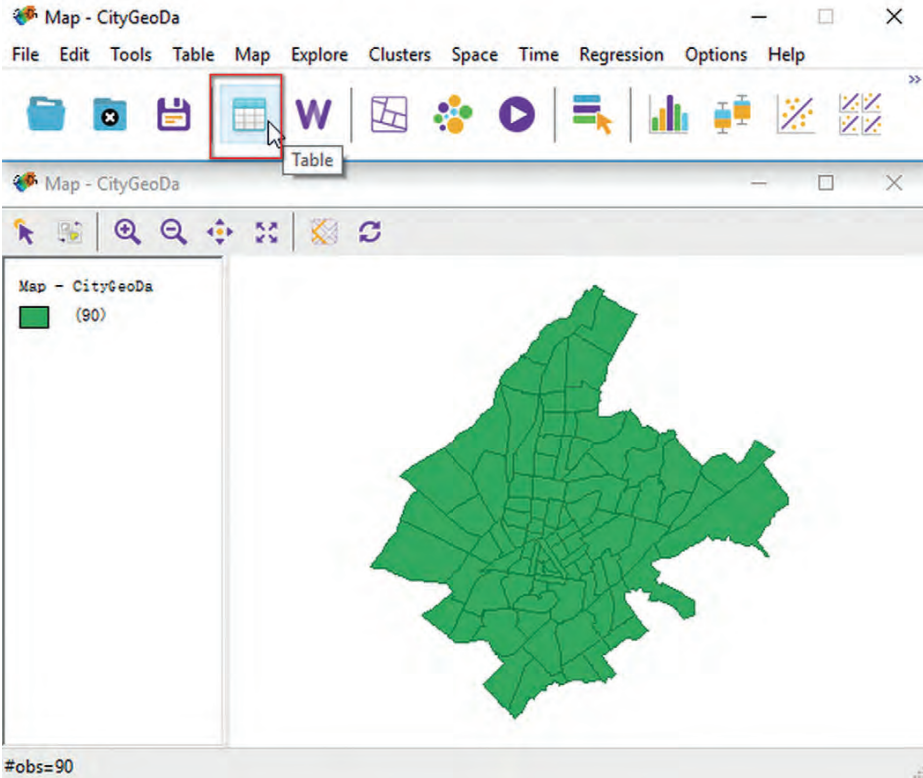


Рис. 1.16. Исследуемая область в GeoDa

	Municipal	PostCode	Population	SecondaryE	University	Phd_Master	Income	Insurance	Rent	Expenses	Area	Regimes	Density	Foreigners	Owners
1	ATHENS	10434	7063	51.260000	19.820000	1.810000	15451.630000	212.980000	632.040000	178.583919	332220.136904	2	0.020000	28.670000	36.280000
2	ATHENS	10435	4900	42.290000	9.000000	0.900000	10992.850000	118.410000	633.700000	79.890622	280381.940596	2	0.020000	22.040000	37.410000
3	ATHENS	10436	3221	45.760000	11.270000	0.560000	10294.890000	108.400000	634.930000	72.294562	207968.081786	2	0.020000	35.700000	30.220000
4	ATHENS	10440	10919	51.410000	13.750000	0.960000	13460.390000	116.000000	610.580000	98.858519	240655.086409	2	0.050000	29.010000	34.680000
5	ATHENS	10442	8036	46.040000	8.550000	0.540000	12498.300000	146.160000	492.590000	99.116531	1915988.950180	2	0.010000	12.880000	49.190000
6	ATHENS	10443	25933	46.700000	10.250000	0.590000	12352.070000	141.770000	480.810000	106.902080	1203704.016840	2	0.020000	12.450000	46.670000
7	ATHENS	10444	19455	47.200000	11.500000	0.720000	11548.990000	133.530000	377.480000	92.826920	820720.173172	2	0.020000	13.190000	42.690000
8	ATHENS	10445	25094	48.810000	12.310000	0.690000	11981.840000	138.630000	568.090000	108.538586	805075.936223	2	0.030000	16.610000	40.320000
9	ATHENS	10446	22398	49.840000	13.960000	0.760000	11195.800000	108.430000	608.680000	117.538586	462.14530779	2	0.050000	27.220000	36.020000
10	ATHENS	10553	1317	34.070000	10.100000	0.380000	12894.380000	157.320000	526.170000	64.621173	201720.185958	2	0.010000	49.810000	35.260000
11	ATHENS	10554	419	34.840000	9.550000	1.190000	14777.130000	304.580000	655.830000	178.809314	134869.127457	1	0.000000	53.700000	18.720000
12	ATHENS	10557	651	27.800000	14.900000	3.990000	26649.570000	445.730000	855.960000	442.114011	561843.171923	1	0.000000	43.310000	25.610000
13	ATHENS	10560	142	60.560000	15.490000	1.410000	20021.220000	462.310000	623.960000	195.567287	49023.275437	1	0.000000	33.100000	31.080000
14	ATHENS	10562	329	28.990000	4.870000	0.910000	19492.330000	441.640000	916.820000	164.762596	45076.450886	1	0.010000	74.590000	17.650000
15	ATHENS	10564	488	23.690000	6.350000	0.200000	21802.800000	404.940000	565.430000	176.925739	124950.134666	1	0.000000	7.450000	50.000000
16	ATHENS	10672	1406	34.780000	35.500000	11.380000	26102.290000	403.190000	777.740000	301.895511	122351.422990	1	0.010000	6.530000	39.000000
17	ATHENS	10682	1990	45.730000	24.610000	3.200000	23212.890000	348.600000	679.350000	318.759713	184824.421266	2	0.010000	18.200000	38.180000
18	ATHENS	11362	13511	50.480000	19.800000	2.050000	14097.120000	178.540000	600.270000	177.894526	298305.543340	2	0.050000	14.370000	40.050000
19	ATHENS	10683	2834	42.730000	27.200000	3.740000	17602.320000	238.930000	775.970000	200.488897	97401.934882	2	0.030000	11.970000	32.540000
20	ATHENS	11141	19642	47.990000	21.580000	1.850000	15545.310000	219.410000	607.030000	185.828585	843011.636198	2	0.020000	5.830000	44.750000
21	ATHENS	11142	18378	47.440000	14.660000	0.800000	14860.840000	204.280000	538.490000	134.450185	870428.121102	2	0.020000	8.330000	52.440000
22	ATHENS	10680	8433	41.430000	24.150000	4.070000	20829.760000	302.830000	641.110000	182.858107	148636.577176	1	0.040000	23.390000	30.800000
23	ATHENS	10681	6433	41.430000	24.150000	4.070000	17452.160000	245.370000	749.290000	176.242771	117805.016579	1	0.050000	23.390000	30.800000
24	ATHENS	11143	16526	49.070000	15.760000	1.010000	14385.850000	191.910000	387.420000	144.239849	399665.867959	2	0.020000	18.100000	46.000000

	PostCode	Population	SecondaryE	University	Phd_Master	Income
1	10434	7063	51.26	19.82	1.81	154
2	10435	4900	42.29	9.00	0.90	109
3	10436	3221	45.76	11.27	0.56	102
4	10440	10919	51.41	13.75	0.96	134
5	10442	8036	46.04	8.55	0.54	124
6	10443	25933	46.70	10.25	0.59	123
7	10444	19455	47.20	11.50	0.72	115
8	10445	25094	48.81	12.31	0.69	119
9	10446	22398	49.84	13.99	0.76	111

Рис. 1.17. Всего в таблице атрибутов хранятся сведения о 90 округах, по 17 полей для каждого, из которых 11 являются социально-экономическими переменными (см. табл. 1.5)

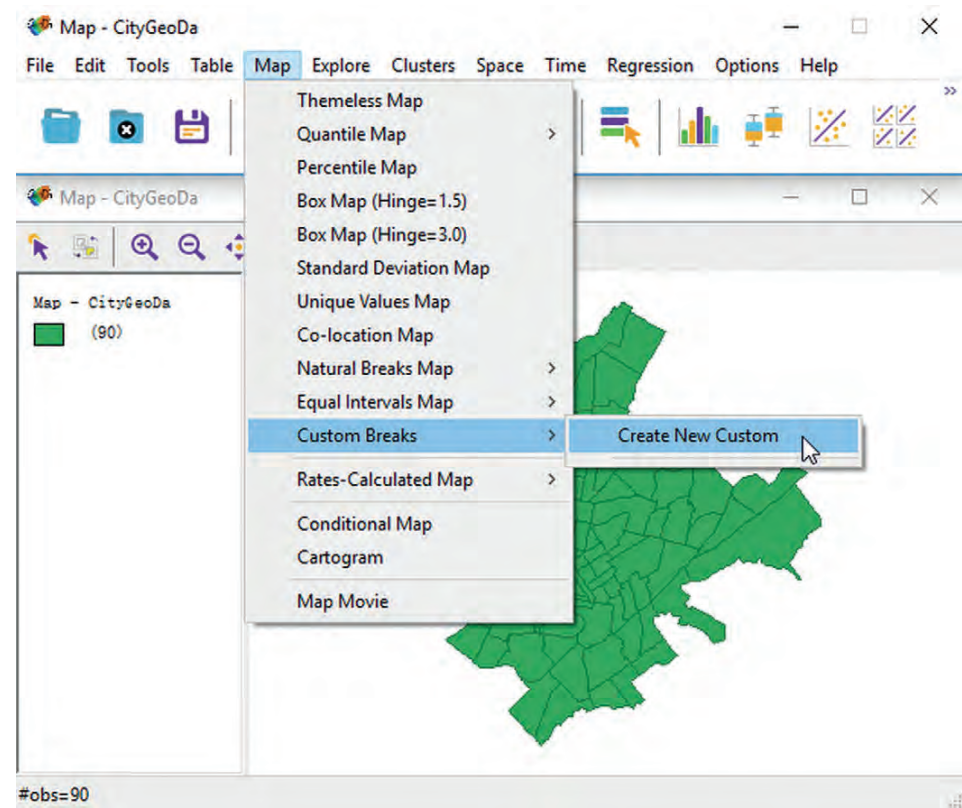


Рис. 1.18. Создание фоновой картограммы распределения населения

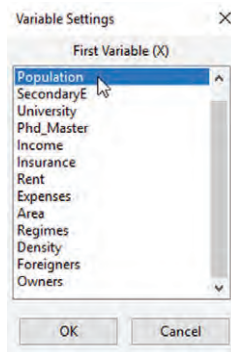


Рис. 1.19. Диалог выбора переменных

В Category Editor (Редактор категорий) измените следующие поля:
Breaks (Границы) = User Defined (Определяемые пользователем; см. рис. 1.20)

Categories (Категории) = 5

Введите следующие значения в поля break (граница):

break 1 = 2000 / break 2 = 5000 / break 3 = 10000 / break 4 = 15000 / break 5 = 20000 > закройте диалог.

Карта обновится, как показано на рис. 1.21.

Главное меню > Save (Сохранить)

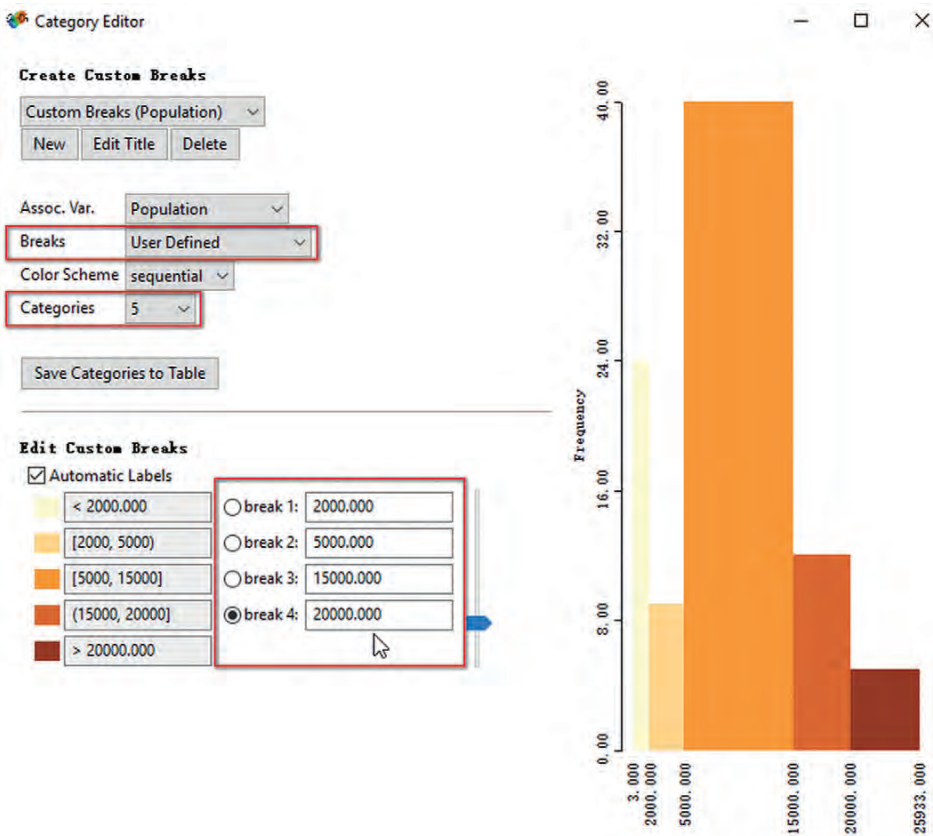


Рис. 1.20. Диалог редактора категорий

Интерпретация результатов: см. раздел А.

Совет: интерпретация результатов и выводы, полученные в ходе анализа, приводятся в разделе А, не будут повторяться в разделе В каждого упражнения. Читателю следует внимательно изучить эти абзацы, потому что они не зависят от используемого программного обеспечения.

ДЕЙСТВИЕ: вычисление и отображение на карте плотности населения.

Главное меню > Options (Параметры) > Rates (Отношения) > Raw Rate (Прямое отношение)

Event Variable (Переменная события) = Population (см. рис. 1.22)

Base Variable (Базовая переменная) = Area

Map Themes (Темы карт) = Natural Breaks (Естественные границы)

Categories (Категории) = 4 > ОК (см. рис. 1.23)

Save (Сохранить)

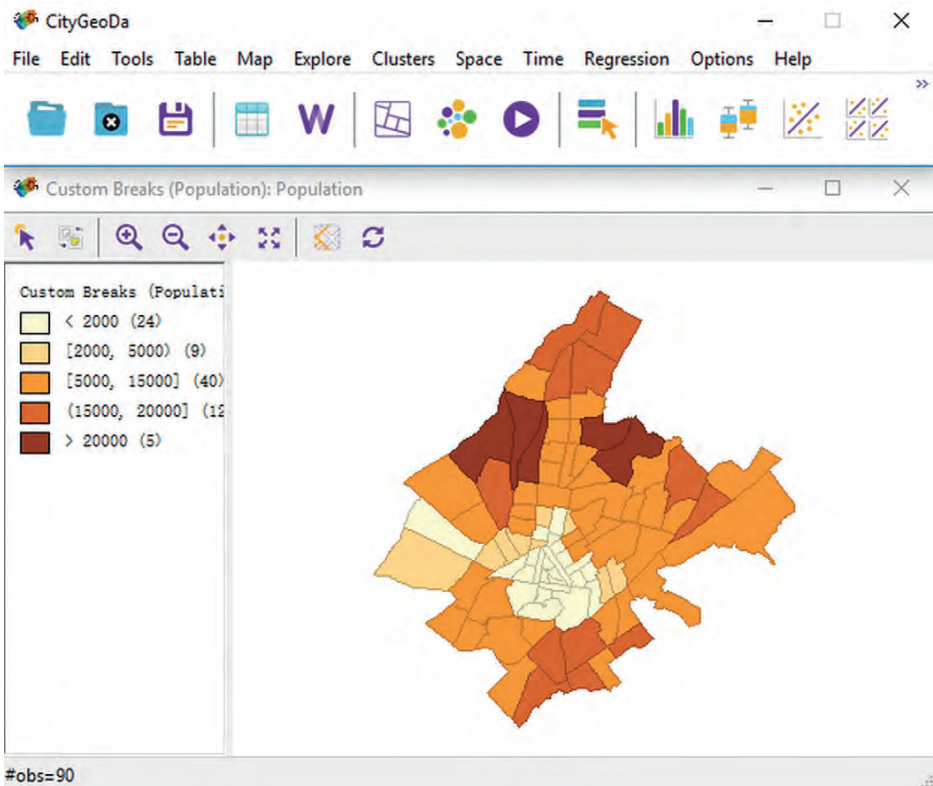


Рис. 1.21. Фоновая картограмма распределения населения

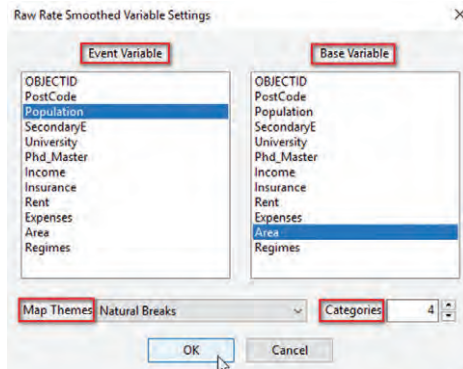


Рис. 1.22. Выбор плотности населения для построения картограммы

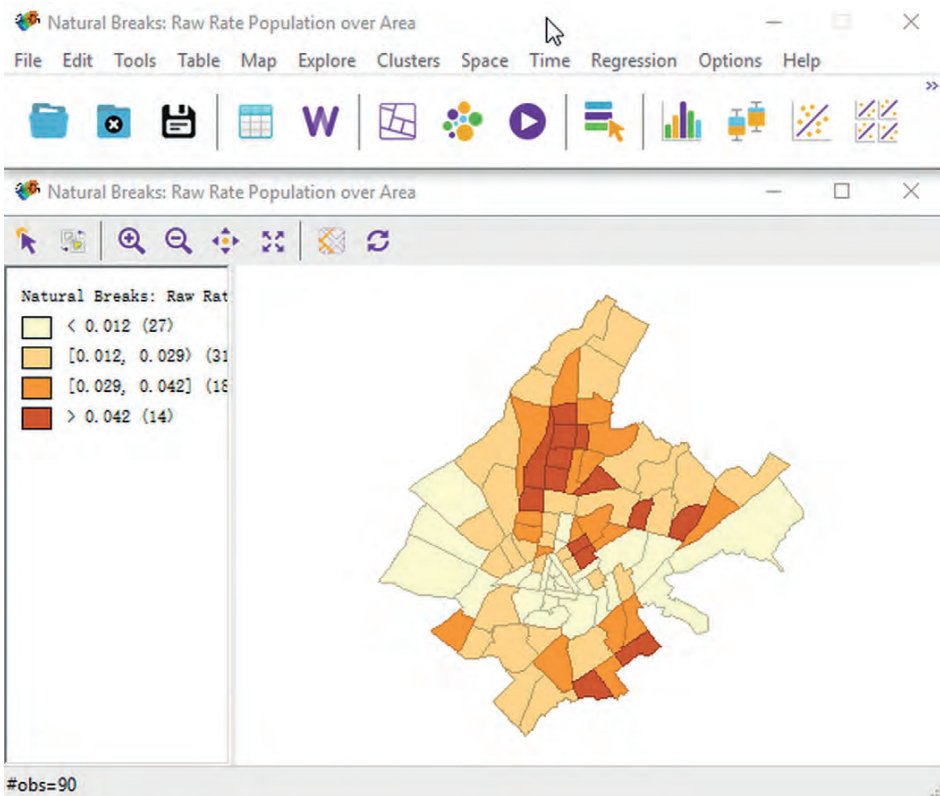


Рис. 1.23. Картограмма плотности населения

Интерпретация результатов: см. раздел А.