

Оглавление

| | |
|--|-----------|
| Предисловие от издательства..... | 11 |
| 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 |
| Практическая работа 1. Пространственный анализ инвестиций на рынке недвижимости | 54 |
| А. ArcGIS | 60 |
| В. GeoDa..... | 67 |
| Упражнение 1.1. Знакомство с данными и областью исследования | 67 |
| 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..... | 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 | |
| 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 |
| Вопросы и ответы..... | 128 |
| Практическая работа 2. Исследовательский анализ пространственных данных (ESDA): анализ и отображение данных..... | 132 |
| А. ArcGIS..... | 133 |
| Упражнение 2.1. Инструменты ESDA: отображение и анализ распределения доходов..... | 133 |
| Упражнение 2.2. Двумерный анализ: анализ расходов по уровню образования..... | 147 |
| В. GeoDa..... | 152 |
| Упражнение 2.1. Инструменты ESDA: отображение и анализ распределения доходов..... | 152 |
| Упражнение 2.2. Двумерный анализ: анализ расходов по уровню образования..... | 158 |

| | |
|--|------------|
| 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 | 166 |
| 3.1.3. Центральный объект | 167 |
| 3.1.4. Стандартное расстояние | 169 |
| 3.1.5. Эллипс стандартного отклонения | 171 |
| 3.1.6. Географические и пространственные выбросы | 173 |
| 3.2. Анализ пространственных закономерностей: анализ структуры точечных закономерностей | 178 |
| 3.2.1. Определения: пространственный процесс, полная пространственная случайность, эффекты первого и второго порядка | 180 |
| 3.2.2. Пространственный процесс | 182 |
| 3.3. Методы анализа структуры точечных закономерностей | 185 |
| 3.3.1. Анализ ближайших соседей | 186 |
| 3.3.2. K -функция Рипли и L -функция ее преобразования | 187 |
| 3.3.3. Ядерная функция плотности | 191 |
| 3.4. Заключительные замечания к главе | 195 |
| Вопросы и ответы | 196 |
| Практическая работа 3. Пространственная статистика: оценка географических распределений | 199 |
| Упражнение 3.1. Оценка географических распределений | 200 |
| Упражнение 3.2. Анализ точечных закономерностей | 207 |
| Упражнение 3.3. Ядерная оценка плотности | 214 |
| Упражнение 3.4. Географические выбросы | 219 |
| Глава 4. Пространственная автокорреляция | 222 |
| Теория | 222 |
| 4.1. Пространственная автокорреляция | 223 |
| 4.2. Глобальная пространственная автокорреляция | 226 |
| 4.2.1. Индекс I Морана и диаграмма рассеяния | 226 |
| 4.2.2. Индекс S Гири | 232 |
| 4.2.3. Общая G -статистика | 233 |
| 4.3. Инкрементальная пространственная автокорреляция | 235 |
| 4.4. Локальная пространственная автокорреляция | 238 |
| 4.4.1. Локальный индекс I Морана (анализ кластеров и выбросов) | 238 |
| 4.4.2. Оптимизированный анализ выбросов | 242 |
| 4.4.3. Индексы Гетиса–Орда G_i^* и G_i^* (анализ горячих точек) | 243 |
| 4.4.4. Оптимизированный анализ горячих точек | 245 |
| 4.5. Пространственно-временной корреляционный анализ | 246 |
| 4.5.1. Двумерный индекс I Морана пространственно-временной корреляции | 246 |
| 4.5.2. Дифференциальный индекс I Морана | 248 |
| 4.5.3. Анализ возникновения горячих точек | 250 |
| 4.6. Проблема множественного сравнения и пространственная зависимость | 251 |
| 4.7. Заключительные замечания к главе | 253 |

| | |
|---|-----|
| Вопросы и ответы | 254 |
| Практическая работа 4. Пространственная автокорреляция | 258 |
| А. ArcGIS | 259 |
| Упражнение 4.1. Глобальная пространственная автокорреляция..... | 259 |
| Упражнение 4.1. Глобальная пространственная автокорреляция..... | 260 |
| Упражнение 4.2. Инкрементальная пространственная автокорреляция и матрица пространственных весов..... | 264 |
| Упражнение 4.3. Анализ кластеров и выбросов (локальный индекс I Морана) | 270 |
| Упражнение 4.4. Анализ горячих точек (индекс G_i^* Гетиса–Орда) и оптимизированный анализ горячих точек..... | 273 |
| Упражнение 4.5. Оптимизированный анализ горячих точек криминальных событий | 279 |
| В. GeoDa..... | 282 |
| Упражнения 4.1 и 4.2. Глобальная пространственная автокорреляция и матрица пространственных весов..... | 282 |
| Упражнение 4.3. Анализ кластеров и выбросов (локальный индекс I Морана) | 287 |
| Упражнение 4.4. Анализ горячих точек (индекс G_i^* Гетиса–Орда) | 290 |

Глава 5. Многомерные данные в географии: сокращение

| | |
|---|------------|
| размерности данных и пространственная кластеризация | 292 |
| Теория..... | 292 |
| 5.1. Анализ многомерных данных..... | 293 |
| 5.2. Метод главных компонент | 297 |
| 5.3. Факторный анализ | 307 |
| 5.4. Многомерное масштабирование | 308 |
| 5.5. Кластерный анализ | 310 |
| 5.5.1 Иерархическая кластеризация..... | 311 |
| 5.5.2. Алгоритм k средних (разделяющая кластеризация) | 316 |
| 5.6. Регионализация | 322 |
| 5.6.1. Метод SKATER..... | 323 |
| 5.6.2. Метод REDCAP..... | 326 |
| 5.7. Кластеризация на основе плотности: DBSCAN, HDBSCAN, OPTICS..... | 327 |
| 5.8. Анализ сходства: косинусное сходство | 329 |
| 5.9. Заключительные замечания к главе | 331 |
| Вопросы и ответы | 332 |
| Практическая работа 5. Многомерная статистика: кластеризация | 335 |
| А. ArcGIS | 336 |
| Упражнение 5.1. Кластеризация методом k средних | 336 |
| Сводные данные по переменным | 345 |
| Упражнение 5.2. Пространственная кластеризация (Regionalization)..... | 347 |
| Упражнение 5.3. Анализ сходства..... | 350 |
| Упражнение 5.4. Обобщение..... | 354 |
| В. GeoDa..... | 363 |
| Упражнение 5.1. Кластеризация методом k средних | 363 |
| Упражнение 5.2. Пространственная кластеризация | 366 |

| | |
|--|------------|
| Глава 6. Моделирование отношений: | |
| регрессия и географически взвешенная регрессия..... | 369 |
| Теория..... | 369 |
| 6.1. Простая линейная регрессия..... | 370 |
| 6.1.1. Предположения в основе простой линейной регрессии..... | 373 |
| 6.1.2. Обычный метод наименьших квадратов (для определения точки пересечения и наклона)..... | 373 |
| 6.2. Множественная линейная регрессия..... | 374 |
| 6.2.1. Основы множественной регрессии..... | 374 |
| 6.2.2. Переобучение модели: выбор количества переменных путем определения функциональной взаимосвязи..... | 377 |
| 6.2.3. Отсутствующие значения..... | 378 |
| 6.2.4. Выбросы и точки с большим плечом..... | 378 |
| 6.2.5. Фиктивные переменные..... | 379 |
| 6.2.6. Методы включения переменных в модель множественной линейной регрессии: объяснительный анализ; выявление причин и следствий..... | 381 |
| 6.3. Оценка результатов линейной регрессии: метрики, критерии и диаграммы..... | 383 |
| 6.3.1. Множественный R -квадрат..... | 383 |
| 6.3.2. Дисперсия и коэффициент детерминации R -квадрат..... | 384 |
| 6.3.3. Скорректированный R -квадрат..... | 388 |
| 6.3.4. Прогнозный R -квадрат..... | 389 |
| 6.3.5. Стандартная ошибка (отклонение) регрессии (или стандартная ошибка оценки)..... | 390 |
| 6.3.6. F -критерий общей значимости..... | 391 |
| 6.3.7. t -статистика (критерий коэффициентов)..... | 393 |
| 6.3.8. Критерий Вальда (критерий коэффициентов)..... | 394 |
| 6.3.9. Стандартизированные коэффициенты (бета)..... | 394 |
| 6.3.10. Остатки, диаграммы остатков и стандартизированных остатков..... | 397 |
| 6.3.11. Факторы влияния: выбросы и наблюдения с высоким плечом.... | 400 |
| 6.4. Предположения в основе множественной линейной регрессии: диагностика и исправление..... | 402 |
| 6.5. Мультиколлинеарность..... | 408 |
| 6.6. Практический пример: простая и множественная линейная регрессия..... | 410 |
| 6.7. Исследовательская регрессия..... | 418 |
| 6.8. Географически взвешенная регрессия..... | 422 |
| 6.8.1. Типы пространственных ядер..... | 423 |
| 6.8.2. Ширина полосы..... | 424 |
| 6.8.3. Интерпретация результатов GWR и практические рекомендации..... | 425 |
| 6.9. Заключительные замечания к главе..... | 429 |
| Вопросы и ответы..... | 430 |

| | |
|---|------------|
| Практическая работа 6. Обычный метод наименьших квадратов (OLS), исследовательская регрессия, географически взвешенная регрессия (GWR)..... | 433 |
| Упражнение 6.1. Исследовательская регрессия | 434 |
| Упражнение 6.2. Регрессия обычным методом наименьших квадратов | 448 |
| Упражнение 6.3. Географически взвешенная регрессия..... | 459 |
| Глава 7. Пространственная эконометрика..... | 471 |
| Теория..... | 471 |
| 7.1. Пространственная эконометрика | 472 |
| 7.2. Пространственная зависимость: модели и критерии пространственной регрессии | 473 |
| 7.2.1. Критерии пространственной зависимости | 474 |
| 7.2.2. Выбор между моделями пространственного лага и пространственной ошибки | 477 |
| 7.2.3. Методы оценки..... | 479 |
| 7.3. Модель пространственного лага | 479 |
| 7.3.1. Пространственный двухшаговый метод наименьших квадратов (S2SLS) | 482 |
| 7.3.2. Максимальное правдоподобие..... | 486 |
| 7.4. Модель пространственной ошибки..... | 487 |
| 7.5. Пространственная фильтрация | 489 |
| 7.6. Пространственная неоднородность: модели пространственной регрессии..... | 490 |
| 7.7. Пространственные режимы | 491 |
| 7.8. Заключительные замечания к главе | 494 |
| Вопросы и ответы | 494 |
| Практическая работа 7. Пространственная эконометрика..... | 497 |
| Общий прогресс | 497 |
| Сфера анализа | 497 |
| Упражнение 7.1. Регрессия OLS..... | 498 |
| Упражнение 7.2. Модель пространственной ошибки | 508 |
| Упражнение 7.3. Регрессия OLS с пространственными режимами | 511 |
| Упражнение 7.4. Модель пространственной ошибки с режимами..... | 519 |
| Список использованной литературы | 522 |
| Предметный указатель | 532 |

Предисловие от издательства

Отзывы и пожелания

Мы всегда рады отзывам наших читателей. Расскажите нам, что вы думаете об этой книге – что понравилось или, может быть, не понравилось. Отзывы важны для нас, чтобы выпускать книги, которые будут для вас максимально полезны.

Вы можете написать отзыв на нашем сайте www.dmkpress.com, зайдя на страницу книги и оставив комментарий в разделе «Отзывы и рецензии». Также можно послать письмо главному редактору по адресу dmkpress@gmail.com; при этом укажите название книги в теме письма.

Если вы являетесь экспертом в какой-либо области и заинтересованы в написании новой книги, заполните форму на нашем сайте по адресу http://dmkpress.com/authors/publish_book/ или напишите в издательство по адресу dmkpress@gmail.com.

Список опечаток

Хотя мы приняли все возможные меры для того, чтобы обеспечить высокое качество наших текстов, ошибки все равно случаются. Если вы найдете ошибку в одной из наших книг – возможно, ошибку в основном тексте или программном коде, – мы будем очень благодарны, если вы сообщите нам о ней. Сделав это, вы избавите других читателей от недопонимания и поможете нам улучшить последующие издания этой книги.

Если вы найдете какие-либо ошибки в коде, пожалуйста, сообщите о них главному редактору по адресу dmkpress@gmail.com, и мы исправим это в следующих тиражах.

Нарушение авторских прав

Пиратство в интернете по-прежнему остается насущной проблемой. Издательство «ДМК Пресс» очень серьезно относится к вопросам защиты авторских прав и лицензирования. Если вы столкнетесь в интернете с незаконной публикацией какой-либо из наших книг, пожалуйста, пришлите нам ссылку на интернет-ресурс, чтобы мы могли применить санкции.

Ссылку на подозрительные материалы можно прислать по адресу dmkpress@gmail.com.

Мы высоко ценим любую помощь по защите наших авторов, благодаря которой мы можем предоставлять вам качественные материалы.

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 связкой (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.

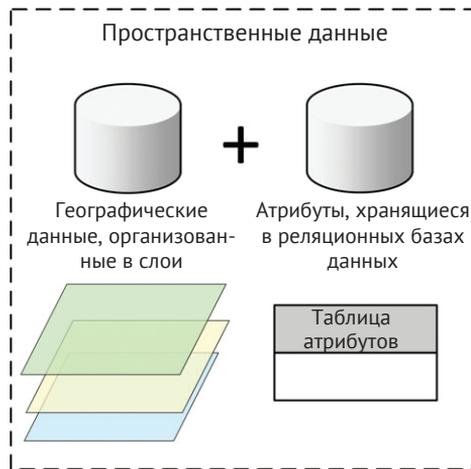


Рис. 1.2. Пространственные объекты в совокупности с атрибутами образуют пространственные данные

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 |

Конец ознакомительного фрагмента.

Приобрести книгу можно

в интернет-магазине

«Электронный универс»

e-Univers.ru