

Przestrzenne metody ilościowe w R : statystyka, ekonometria, uczenie maszynowe, analiza danych / Katarzyna Kopczewska (red.). – Wydanie I. – Warszawa, 2020

Spis treści

Wstęp	13
Wykorzystanie w książce oświadczenia Amerykańskiego Stowarzyszenia Statystycznego w sprawie istotności statystycznej i <i>p-value</i>	17
Rozdział 1	
Podstawowe operacje w programie R (<i>Mateusz Kopyt</i>)	19
1.1. O oprogramowaniu R	19
1.2. Interfejs programu R	20
1.2.1. R Commander	21
1.2.2. RStudio	22
1.3. Korzystanie z pomocy	24
1.4. Pakiety dodatkowe	28
1.5. Język programu R - podstawowe cechy	31
1.6. Definiowanie i wczytywanie danych	32
1.7. Podstawowe operacje na obiektach	36
1.8. Podstawowe statystyki zbioru danych	47
1.9. Wizualizacja - podstawy	55
1.9.1. Wykres punktowy oraz liniowy	56
1.9.2. Wykres kolumnowy	59
1.9.3. Wykres kołowy	62
1.9.4. Wykres pudełkowy (<i>boxplot</i>)	63
1.10. Regrgsjã.w przykładach	65
Rozdział 2	
Dane i klasy przestrzenne oraz podstawowa grafika <i>(Katarzyna Kopczewska)</i>	73
2.1. Wczytywanie i podstawowe operacje na danych przestrzennych wektorowych	74
2.2. Tworzenie, sprawdzanie i konwersja klas przestrzennych	89
2.3. Wybrane palety kolorystyczne	95
2.4. Podstawowe mapy konturowe z warstwą kolorystyczną	100
Schemat 1 - z <code>colorRampPalette()</code> z pakietu <code>grDevices::</code>	100
Schemat 2 - z <code>choropleth()</code> z pakietu <code>GISTools::</code>	102
Schemat 3 - z <code>findInterval()</code> z pakietu <code>base::</code>	103
Schemat 4 - z <code>findColours()</code> z pakietu <code>classInt::</code>	104
Schemat 5 - z <code>spplot()</code> z pakietu <code>sp::</code>	106
2.5. Podstawowe operacje i wykresy dla danych punktowych	107
Schemat 1 - z <code>points()</code> z pakietu <code>graphics::</code> - wyłącznie lokalizacje	108

Schemat 2 - z <code>spplot()</code> z pakietu <code>sp::</code> - lokalizacje i wartości	108
Schemat 3 - z <code>hndInterval()</code> z pakietu <code>base::</code> - lokalizacje, wartości, różna wielkość symboli	110
2.6. Podstawowe operacje na rastrach	114
2.7. Podstawowe operacje na gridach	123
2.8. Geometrie przestrzenne	133

Rozdział 3

Dane przestrzenne z Web API (<i>Mateusz Kopyt, Katarzyna Kopczewska</i>)	139
3.1. Czym jest API	139
3.2. Tworzenie map kontekstowych - wykorzystanie podkładów z serwerów API	141
3.3. Sposoby wizualizacji danych przestrzennych - mapy dla danych punktowych i regionalnych	157
Schemat 1 - z <code>bubbleMap()</code> z pakietu <code>RgoogleMaps::</code>	157
Schemat 2 - z <code>ggmap()</code> z pakietu <code>ggmap::</code>	158
Schemat 3 - z <code>PlotOnStaticMap()</code> z pakietu <code>RgoogleMaps::</code>	162
Schemat 4 - z <code>GetMap()</code> z pakietu <code>RGoogleMaps::</code> i przekształceniem <code>staticMap</code> na raster	163
3.4. Dane przestrzenne w formacie wektorowym na przykładzie bazy OSM	164
3.5. Dostęp do baz i zasobów danych nieprzestrzennych w Internecie przez API w przykładach	174
3.6. Geokodowanie danych	193

Rozdział 4

Macierze wag przestrzennych, pomiar odległości, teselacja, statystyka przestrzenna (<i>Katarzyna Kopczewska, Maria Kubara</i>)	215
4.1. Wprowadzenie do analizy danych przestrzennych	216
4.2. Macierze wag przestrzennych	219
4.2.1. Ogólne ramy tworzenia macierzy wag przestrzennych	219
4.2.2. Wybór macierzy sąsiedztwa	223
4.2.3. Macierze sąsiedztwa według kryterium wspólnej granicy	224
4.2.4. Macierz k najbliższych sąsiadów (<i>k nearest neighbours, knn</i>)	229
4.2.5. Macierz oparta o kryterium odległości (sąsiadów w promieniu d km)	232
4.2.6. Macierz odwrotnej odległości	235
4.2.7. Podsumowania i edycja macierzy wag	237
4.2.8. Opóźnienia przestrzenne i sąsiedztwo wyższego rzędu	243
4.2.9. Tworzenie macierzy wag w oparciu o przynależność grupową	245
4.3. Pomiar odległości i agregacja przestrzenna	250
4.4. Teselacja	262
4.5. Statystyki przestrzenne	265
4.5.1. Statystyki globalne	269
4.5.1.1. Statystyka globalna I Morana	269
4.5.1.2. Statystyka globalna C Geary'ego (Geary C)	277
4.5.1.3. Statystyki join-count	279
4.5.2. Statystyki autokorelacji przestrzennej lokalnej	283
4.5.2.1. Statystyka lokalna I Morana (Local Moran, LISA)	284
4.5.2.2. Statystyka lokalna C Geary'ego (Local Geary)	287

4.5.2.3. Lokalna statystyka G (Getis-Ord G)	288
4.5.2.4. Lokalna heteroskedastyczność przestrzenna (LOSH)	290
4.6. Przestrzenne korelacje krzyżowe dla dwóch zmiennych	293
4.7. Korelogram	296

Rozdział 5

Stosowana ekonometria przestrzenna (<i>Katarzyna Kopczewska</i>)	303
5.1. Wartość dodana z modelowania przestrzennego i klasy modeli	304
5.2. Podstawowe modele przekrojowe	309
5.2.1. Estymacja	309
5.2.2. Ocena jakości modeli przestrzennych	330
5.2.2.1. Kryteria informacyjne i pseudo R^2 w ocenie dopasowania modelu	331
5.2.2.2. Test na heteroskedastyczność reszt modelu	333
5.2.2.3. Testy na autokorelację przestrzenną reszt	336
5.2.2.4. Testy LM na wybór typu modelu	340
5.2.2.5. Testy LR i Walda na ograniczenia w modelach	343
5.2.3. Dobór macierzy wag przestrzennych i modelowanie siły dyfuzji	345
5.2.4. Prognozy w modelach przestrzennych	348
5.2.5. Przyczynowość	351
5.3. Wybrane specyfikacje przekrojowych modeli przestrzennych	353
5.3.1. Modele interakcji przestrzennych jednokierunkowych	353
5.3.2. Modele kumulatywne	366
5.3.3. Modele bootstrapowane dla <i>big data</i>	373
5.3.4. Modele dla danych grid	385
5.4. Przestrzenne modele panelowe	392

Rozdział 6

Modelowanie heterogeniczności przestrzennej (<i>Piotr Ćwiakowski</i>)	413
6.1. Regresja geograficznie ważona	413
6.2. Estymacja podstawowego modelu GWR	416
6.2.1. Oszacowanie modelu referencyjnego (MNK)	417
6.2.2. Wybór optymalnej dla zbioru szerokości pasma (<i>bandwidth</i>)	418
6.2.3. Lokalne statystyki ważone geograficznie	423
6.2.4. Estymacja regresji ważonej geograficznie	425
6.2.5. Podstawowe testy diagnostyczne modelu GWR	428
6.2.6. Testowanie istotności parametrów w GWR	437
6.2.7. Wybór optymalnej formy funkcyjnej modelu	438
6.2.8. GWR z heteroskedastycznym błędem losowym	441
6.3. Problem współliniowości w modelach GWR	442
6.4. Mieszany model GWR	456
6.5. Regresja odporna w modelu GWR	458
6.6. Regresja ważona czasowo i geograficznie (<i>Geographically and Temporally Weighted Regression, GTWR</i>)	460

Rozdział 7

Przestrzenne uczenie nienadzorowane (<i>Katarzyna Kopczewska</i>)	465
7.1. Klastrowanie punktów przestrzennych algorytmami k-średnich, PAM i CLARA	466

7.2. Klastrowanie algorytmem DBSCAN	483
7.3. Przestrzenna Analiza Głównych Składowych (<i>Spatial Principal Component Analysis</i>)	496
7.4. Dryf przestrzenny (<i>Spatial Drift</i>)	503
7.5. Przestrzenne klastrowanie hierarchiczne (<i>spatial hierarchical clustering</i>)	513
7.6. Przestrzenne skośne drzewa decyzyjne (<i>spatial oblique decision tree</i>)	524

Rozdział 8

Analiza przestrzennych rozkładów punktowych i interpolacja

przestrzenna (<i>Kateryna Zabarina</i>)	529
8.1. Wprowadzenie i główne definicje	532
8.1.1. Zbiór danych	533
8.1.2. Tworzenie okna i rozkładu punktowego	534
8.1.3. Znaki	537
8.1.4. Zmienne towarzyszące (<i>covariates</i>)	543
8.1.5. Duplikowanie punktów	546
8.1.6. Projekcja i skalowanie	548
8.2. Analiza intensywności w rozkładach punktowych nieoznakowanych	552
8.2.1. Kwadrat test (test zliczania w kwadratach)	554
8.2.2. Testy z przestrzennymi zmiennymi towarzyszącymi (<i>spatial covariates</i>)	555
8.3. Analiza rozkładu punktowego nieoznakowanego w oparciu o odległość	559
8.3.1. Miary oparte na odległości	560
8.3.1.1. Funkcja K Ripleya	560
8.3.1.2. Funkcja F	563
8.3.1.3. Funkcja G	563
8.3.1.4. Funkcja J	564
8.3.1.5. Testy CSR oparte na odległości	564
8.3.2. Testy Monte-Carlo	565
8.3.3. Obwiednie	565
8.3.4. Testy niegraficzne	567
8.4. Wybór i oszacowanie właściwego modelu dla rozkładu punktowego nieoznakowanego (<i>unmarked point pattern</i>)	568
8.4.1. Nota teoretyczna	569
8.4.2. Wybór parametrów	571
8.4.3. Oszacowania i wyniki	573
8.4.4. Wnioski	576
8.5. Analiza intensywności w rozkładach punktowych oznakowanych	577
8.5.1. Test segregacji (<i>Segregation test</i>)	578
8.6. Analiza korelacji i rozrzutu w rozkładzie punktowym z wartościami	579
8.6.1. Analiza przy założeniu stacjonarności	579
8.6.1.1. Warianty funkcji K dla rozkładu wielorakiego	579
8.6.1.2. Funkcja powiązań wartości (<i>mark connection function</i>)	580
8.6.1.3. Analiza zależności wewnątrz typów i pomiędzy typami	581
8.6.1.4. Test randomizacyjny niezależności składników	583
8.6.2. Analiza przy założeniu niestacjonarności	584

8.6.2.1. Warianty niejednorodnej funkcji K dla wielorodzajowych wzorców	584
8.7. Wybór i oszacowanie właściwego modelu dla rozkładu punktowego oznakowanego	586
8.7.1. Uwagi teoretyczne	586
8.7.2. Wybór optymalnego promienia	587
8.7.3. Promień oddziaływania wewnątrzsektorowego	588
8.7.4. Promień interakcji między sektorami	590
8.7.5. Oszacowania i wyniki	591
8.7.6. Model bez interakcji pomiędzy sektorami	591
8.7.7. Model ze wszystkimi możliwymi interakcjami	594
8.8. Metody interpolacji przestrzennej - kriging	598
8.8.1. Podstawowe definicje	599
8.8.2. Opis wybranych metod krigingu	600
8.8.3. Przygotowanie danych do badań	602
8.8.4. Oszacowanie i dyskusja	603

Rozdział 9

Przestrzenne próbkowanie i bootstrap

(Katarzyna Kopczewska, Piotr Cwiakowski)

613

9.1. Przestrzenne dane punktowe - klasy obiektów i agregacja przestrzenna	616
9.2. <i>Spatial sampling</i> - losowanie/generowanie nowych punktów na powierzchni	619
9.3. <i>Spatial sampling</i> - losowanie próby z istniejących punktów	622
9.3.1. Losowanie proste	625
9.3.2. Możliwości pakietu <code>sperrest::</code>	627
9.3.3. Losowanie punktów z obszarów wyznaczonych algorytmem k-średnich - <i>block bootstrap</i>	634
9.3.4. Losowanie punktów z bloków ruchomych (<i>moving block bootstrap</i>)	644
9.4. Wykorzystanie próbkowania przestrzennego i bootstrap w walidacji krzyżowej modeli	652

Rozdział 10

Przestrzenne *big data* (Piotr Wójcik)

671

10.1. Przykłady zastosowania <i>big data</i>	672
10.2. Przestrzenne <i>big data</i>	674
10.2.1. Typy danych przestrzennych	674
10.2.2. Wyzwania związane z wykorzystaniem przestrzennych <i>big data</i>	676
10.2.2.1. Przetwarzanie dużych zbiorów danych	676
10.2.2.2. Mapowanie i redukcja	677
10.2.2.3. Przestrzenne indeksowanie danych	678
10.3. Pakiet <code>sf</code> - <i>simple features</i>	679
10.3.1. Klasa <code>sf</code> - specjalna ramka danych	680
10.3.2. Dane z geometrią typu POLYGON	681
10.3.3. Dane z geometrią typu POINT	690
10.3.4. Wizualizacja z wykorzystaniem pakietu <code>ggplot2::</code>	691
10.3.5. Wybrane funkcje do analiz przestrzennych	693

10.4. Wykorzystanie funkcji pakietu dplyr::	698
10.5. Przykładowa analiza dużych danych rastrowych	714
10.5.1. Pomiar nierówności ekonomicznych z kosmosu	714
10.5.2. Analiza z wykorzystaniem funkcji pakietu raster::	716
10.5.3. Inne funkcje pakietu raster::	726
10.5.4. Potencjalna alternatywa - pakiet stars::	727

Rozdział 11

Przestrzenne reguły asocjacyjne w geomarketingu

<i>(Alessandro Festi)</i>	729
11.1. Wprowadzenie do analizy koszykowej i geomarketingu	730
11.2. Dane potrzebne w analizie koszyków rynku przestrzennego	732
11.3. Symulacja danych	735
11.4. Technika analizy koszyków rynkowych stosowana do danych geolokalizacyjnych	741
11.5. Przestrzenne reguły asocjacyjne	747
11.6. Aplikacje do geomarketingu	752
11.6.1. Znalezienie najlepszej lokalizacji dla firmy	753
11.6.2. Targetowanie	756
11.6.3. Badanie konkurencji	757
11.7. Wnioski i dalsze badania	757

Aneks A: Dane wykorzystane w przykładach	759
---	------------

Aneks B: Powiązania pakietów	775
-------------------------------------	------------

Aneks C: Dane przestrzenne w pakietach R	779
---	------------

Bibliografia	787
---------------------	------------

Spis rysunków	809
----------------------	------------

Spis tabel	817
-------------------	------------

Indeks pojęć	819
---------------------	------------

Indeks pakietów	833
------------------------	------------

Indeks komend	837
----------------------	------------