

空間の統計学(1)：事始め

慶應義塾大学総合政策学部准教授

古谷 知之 (Furutani Tomoyuki)

■兵庫県生まれ。2001年東京大学大学院工学系研究科博士課程修了。博士(工学)。東京大学大学院助手、慶應義塾大学環境情報学部専任講師を経て、07年4月より現職。専門分野：空間統計学、都市交通計画、観光政策。



今回から、「空間の統計学」をテーマに、1年間程度にわたり、本連載を執筆することになりました。どうぞよろしくお願ひします。本連載では、以前、「Rと空間統計」に関するシリーズがありましたが(2007年9月号から2008年2月号)、今回は、空間データの統計解析を主眼に執筆する予定です。

1. はじめに

本稿では、Rを使って空間統計学や空間計量経済学を学ぶ上での基本的な情報を提供したいと考えています。第1回は、空間統計学や空間計量経済学と呼ばれる分野で使われているRのパッケージ、地図表示と主題図の作成、空間データベースの操作(属性テーブルの結合)について紹介します。できるだけWindowsとMacの両方に対応できるように心がけますが、気づかない点もあると思いますので、適宜、ご指摘いただければ幸いです。

2. Rのパッケージ

Rでは、GISデータの基本操作(インポート、エクスポート、座標系の定義、地図表示など)や空間データの統計解析、計量経済分析に加え、空間疫学や環境学など、特定分野で有用なパッケージも用意されています。

最近では、70あまりの空間統計関連パッケージが提供されるようになり(表1、[2])、ベイズ法など最新の推定方法が適用可能なパッケージも利用可能です。また、この数年、Rを使った空間データ分析に関する文献も出版されています(例えば、[1])。

空間統計学、空間計量経済学関連のパッケージは、以下の方法により一括してインストールすることができます(ただし、LANの接続速度が遅い場合は、インストールに時間がかかるので、必ずしもお勧めしません)。

```
install.packages("ctv")
library(ctv)
install.views("Spatial")
```

パッケージがインストールできたら、早速、

表1 Rの空間統計学、空間計量経済学関連パッケージ一覧

パッケージ名	概要	パッケージ名	概要
ade4	生態学、環境学関連データの分析	PBSmodelling	Pacific Biological Stationデータのモデリング
adehabitat	動物の生息環境分析	ramps	RAMPS法を使ったベイズ空間統計モデリング
ads	RepleyのK関数をはじめとする空間点過程分析	RandomFields	ランダム場のシミュレーションと分析
akima	不規則グリッドデータのスプライン補完	RArcInfo	Arc/Info V7.xデータのインポート
ash	Davit ScottのASHルーチン	RColorBrewer	主題図作成のためのカラーパレット
aspace	セントログラフィ法と計算幾何学	regress	線形共分散構造を持つガウス線型モデル
automap	ヴァリオグラムとクリギング補完の自動生成	rgdal	空間データ抽出ライブラリのバインディング (windowsのみ)
classInt	主題図作成時の一変量階級区分選択	RgoogleMaps	Google Mapsへのオーバーレイ
clustTool	クラスター分析	RPyGeo	ArcGISジオプロセッシング
Dcluster	疫学データの空間クラスター検出	RSAGA	SAGAジオプロセッシングと地形解析
diseasemapping	疫学データの標準死亡率比計算	RSurvey	空間分布データの解析
ecespa	空間点過程分析	sgeostat	S+の空間統計モデリングのためのオブジェクト指向フレームワーク
fields	空間補完をはじめとする空間分析用ツール	shapefiles	ESRIシェープファイルの読み込みと書き出し
GEOmap	地図投影	sp	空間データ用クラスとメソッド
geomapdata	GEOmapに使う地図データ	spatclus	2Dおよび3D点データの空間クラスティング
geonames	www.geonames.org 用インターフェース	spatgraphs	点データのグラフ
geoR	尤度関数やベイズ法を用いた空間統計分析	spatialCovariance	空間データの共分散行列計算
geoRglm	一般化線形空間回帰モデル (ベイズ法もカバー)	SpatialExtremes	空間極値モデリング
GeoXp	対話型の探索的空間データ分析	spatialkernel	多変量空間点過程のノンパラメトリック推定
glimmBUGS	WinBUGSを使った一般化線形混合モデル、空間モデル分析	spatstat	空間点過程分析、モデル適合、シミュレーション
gmaps	グリッドデータ用関数	spBayes	MCMCによる一変量および多変量モデリング
gmt	GMT地図作成用ソフトとRとのインターフェース	spdep	空間重み付け行列、隣接行列など空間統計指標の計算
grasp	一般化回帰モデルと空間予測	spgrass6	GRASS 6.0とRとのインターフェース
GRASS	GRASS 5.0とRとのインターフェース	spgwr	地理的加重回帰モデル
gstat	空間統計モデリング、予測、シミュレーション	spnlans	時空間点過程分析
hdeco	カテゴリカルマップ比較	spsurvey	合衆国環境保護庁の環境モニタリング・評価プログラムのための調査設計と分析
mapdata	大規模/高解像度空間データ	SQLiteMap	SQLiteを使ったグラフィカルマップの操作
mapproj	緯度経度の座標系変換	tgp	ベイズガウス過程モデル
maps	地図表示	tossm	空間構造の検出
maptools	空間オブジェクトの読み込みと操作	trip	動物移動軌跡データの空間解析
MBA	マルチレベルB-スプライン近似	tripack	二次元のドロネー三角網作成
ModelMap	学習データによる検証とクロスバリデーション	tripEstimation	動物移動軌跡のメトロポリスサンプラー
ncdf	netCDFデータファイルへのインターフェース	vegan	植物生態やコミュニティ生態の多様性分析
ncf	空間ノンパラメトリック共分散関数	VR	Venable & Ripleyの"Modern Applied Statistics with S"で用いられた関数とデータ
pastecs	生態学の時空間系列データ分析	WeedMap	雑草強度マップ
PBSmapping	Pacific Biological Stationの調査データ		

出典) CRANのHP[2]

空間データを操作してみましょう。

3. 地図データの表示

本稿で用いるデータは、筆者のHP[3]で公開しています。上記HPから「ESTRELA連載」にアクセスして、各回のデータをダウンロードして活用してください。

今回は、ESRIジャパンが無償配布している日本の全国市町村境界シェープファイル[4]を、都道府県でマージし、属性データを一部加工

した、都道府県境界シェープファイルを使うことにします。

データをダウンロードしたら、ファイルのあるディレクトリをRの作業ディレクトリに指定しておくといいでしょう。Windowsユーザは「ファイル」→「ディレクトリの変更」で、Macユーザは「その他」→「作業領域の変更」で、それぞれ作業ディレクトリを指定することができます。

インストールしたパッケージのうち、

mapproolsパッケージを、**library()**関数を使って呼び出します。

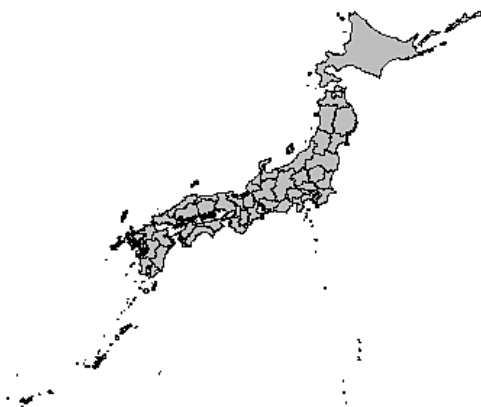
```
library(mapprools)
```

mapproolsパッケージでは、spパッケージのオブジェクトとmapsパッケージのオブジェクトを扱うことができます。ここでは、前者のオブジェクトを扱うことにします。spパッケージでは、SpatialPoints、SpatialLines、SpatialPolygonsなどのSpatial*オブジェクトを扱います。

次に、**readShapePoly()**関数を使って、シェープファイル jpn_pref.shp をインポートし、地図を表示します (図1)。

```
jpn_pref <- readShapePoly("jpn_pref.shp",  
IDvar="PREF_CODE")  
plot(jpn_pref, col="grey")  
summary(jpn_pref)
```

図1 都道府県境界



資料) 全国市区町村境界データ[4]

4. 属性テーブルのマッチング

jpn_pref.shpファイルの属性テーブルには、夜間人口、可住地面積、人口密度などのデータが含まれています。さらに、「社会生活統計

指標—都道府県の指標—」[5]の死因別死亡者数のデータのうち、2005年のデータを[6]からダウンロードし、jpn_pref.shpの属性テーブルと結合してみましょう。

ここでは、死因別死亡者数（人口10万人あたり死亡者数）のうち、生活習慣病による死亡者数 (geriatric)、悪性新生物による死亡者数 (malignant)、糖尿病による死亡者数 (diabetes)、高血圧性疾患による死亡者数 (hypertensive) のデータのみをCOD.csvファイルに格納しています。

```
jpn_COD <- read.table("COD.csv", sep=";", header=T)  
summary(jpn_COD)
```

次に、jpn_prefの属性テーブルとjpn_CODとを結合します。2つのテーブルに共通なIDを使ってマッチングするには、**match()**関数を使い、テーブル結合には**spCbind()**関数を使います。

```
ID.match <- match(jpn_pref$PREF_CODE,  
jpn_COD$PREF_CODE)  
ID.match  
jpn_COD1 <- jpn_COD[ID.match,]  
summary(jpn_COD1)  
jpn_pref_COD <- spCbind(jpn_pref, jpn_COD1)  
names(jpn_pref_COD)  
summary(jpn_pref_COD)
```

5. 主題図の作成

データの準備ができたところで、いよいよ主題図を作成していきましょう。主題図において階級区分を検討することは、空間データの分類の第一歩といえます。主題図は、(1)カラーパレットを作成し、(2)階級区分を定義し、(3)主題図を描く、という順番で作成します。

主なカラーパレットには、以下のようなものがあります。それぞれ、nは階級数、startとendは色または色の濃さ、alphaは不透明度を意味します。

```
pal0 <- c("grey","grey9")
pal1 <- gray.colors(
n=5,start=0.9,end=0.3)
pal2 <- rainbow(
n=5,start=0.6,end=0.1)
pal3 <- heat.colors(
n=5,alpha=1)
pal4 <- topo.colors(
n=5,alpha=1)
pal5 <- terrain.colors(
n=5,alpha=1)
pal6 <- cm.colors(n=5,alpha=1)
```

階級区分方法は、パッケージClassIntのclassInt()関数によって指定します。

```
library(classInt)
```

この関数は一変数データの階級区分を行う際に用いられますが、関数中のstyleに階級区分方法を指定することにより、等量分類(quantile)、等間隔分類(equal)、標準偏差分類(sd)、自然階級分類(jenksまたはfisher)、階級数に依存しない分類(pretty)、非階層クラスタリング(k-means法)による分類(kmeans)、階層クラスタリングによる分類(hclustまたはbclust)を行うことができます。

① 等量分類

等量分類は、各階級区分に等しい数の統計値(または地区)が含まれるようにする分類方法です。4つの階級に区分した場合、4分位(25%、50%、75%)でデータが区分されます。

```
q_pref <- classIntervals(
jpn_pref_COD$Pop_Dens, n=5,
style="quantile")
plot(q_pref, pal=pal0)
q_pref_Col <- findColours(
q_pref,pal0)
plot(jpn_pref_COD,col=q_pref_Col)
title("Population Density
(quantile)")
legend("topleft",fill=attr(q_pref_Col,"palette"),
legend=names(attr(q_pref_Col,"table")),bty="n")
```

② 等間隔分類

等間隔分類は、データの最大値と最小値の差(=データの範囲)を階級数で割って、等間隔で区分する方法です。

```
eq_pref <- classIntervals(
jpn_pref_COD$geriatric, n=5,
style="equal")
```

③ 標準偏差分類

標準偏差分類は、データの平均値からの乖離度について、標準偏差を±して示す方法です。

```
sd_pref <- classIntervals(
jpn_pref_COD$malignant,
style="sd")
```

④ 自然階級分類(Jenksの最適化法)

データの変化点が比較的大きいところに閾値が設定されます。styleにfisherを指定すると、Fisher-Jenksの方法で分類され、jenksを指定すると、単にJenksの方法で分類されます。

```
fj_pref <- classIntervals(
jpn_pref_COD$malignant,
style="fisher")
```

⑤ 階級数に依存しない分類

styleにprettyを指定すると、指定した階級数にかかわらず、視覚的にわかりやすい階級区分で地図を描くことができます。

```
pr_pref <- classIntervals(  
jpn_pref_COD$malignant,  
style="pretty")
```

⑥ 非階層クラスタリングによる分類

非階層クラスタリングの代表的な方法であるk-means法により分類します。この場合、指定した階級数がk-means法のクラスタ数となります。

```
km_pref <- classIntervals(  
jpn_pref_COD$Pop_Dens, n=5,  
style="kmeans")
```

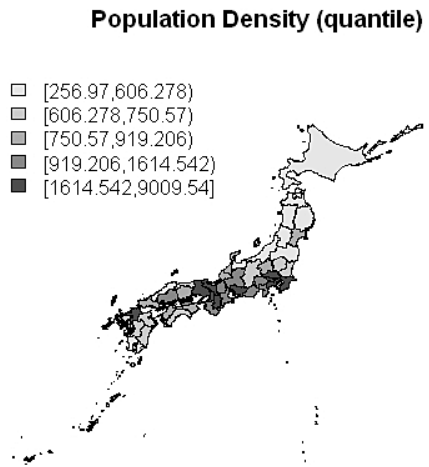
⑦ 階層クラスタリングによる分類

階層クラスタリングにより分類する方法です。階層クラスタリングの方法は、**hclust()**関数を参照してください (`help (hclust)` というコマンドを実行すると、ヘルプファイルを閲覧できます)。

```
hc_pref <- classIntervals(  
jpn_pref_COD$Pop_Dens, n=5,  
style="hclust",  
method="complete")
```

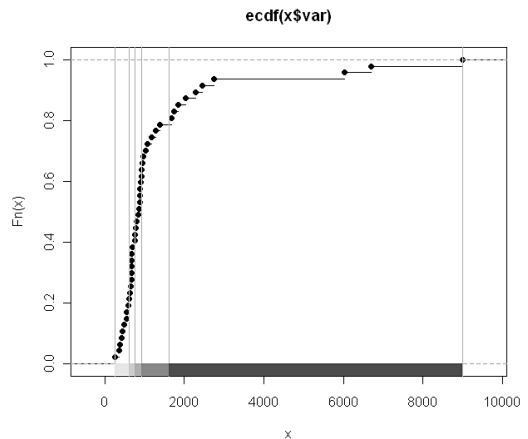
四分位分類で人口密度を表示した地図と階級区分を、それぞれ図2と図3に示します。また、誌面の都合上、生活習慣病による死亡者数、悪性新生物による死亡者数、糖尿病による死亡者数、高血圧性疾患による死亡者数

図2 人口密度の主題図 (四分位分類)



資料) 全国市区町村境界データ[4]、社会生活統計指標—都道府県の指標—2005[6]

図3 人口密度の階級区分 (四分位分類)



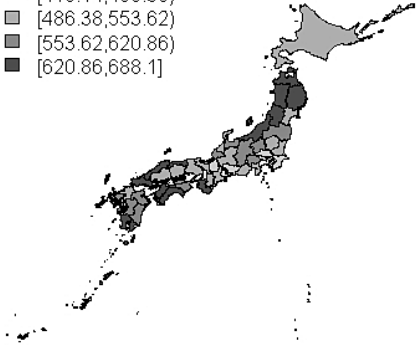
資料) 図2同様

を、それぞれ、等間隔分類、標準偏差分類、自然階級分類 (Jenksの最適化法)、階級数に依存しない分類を用いて分類した結果を、図4～7に示します。

図4 生活習慣病による死亡者数（等間隔分類）

Geriatric diseases (equal)

- [351.9,419.14)
- [419.14,486.38)
- [486.38,553.62)
- [553.62,620.86)
- [620.86,688.1]

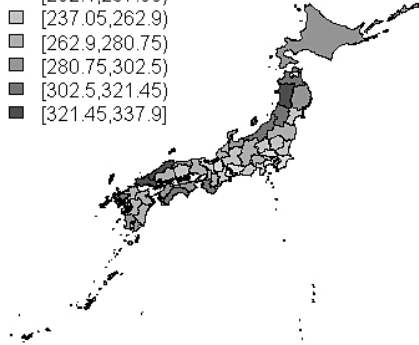


資料) 図2同様

図6 糖尿病による死亡者数（自然階級分類）

Diabetes mellitus (Fisher-Jenks)

- [185.8,202.1)
- [202.1,237.05)
- [237.05,262.9)
- [262.9,280.75)
- [280.75,302.5)
- [302.5,321.45)
- [321.45,337.9]

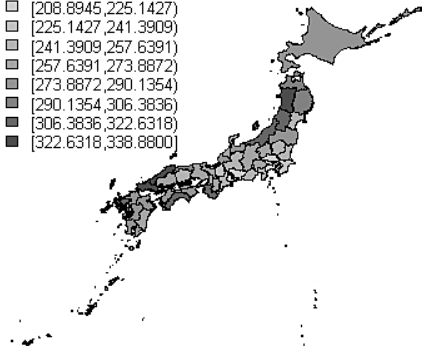


資料) 図2同様

図5 悪性新生物による死亡者数（標準偏差分類）

Malignant neoplasms (standard deviations)

- [176.3981,192.6463)
- [192.6463,208.8945)
- [208.8945,225.1427)
- [225.1427,241.3909)
- [241.3909,257.6391)
- [257.6391,273.8872)
- [273.8872,290.1354)
- [290.1354,306.3836)
- [306.3836,322.6318)
- [322.6318,338.8800]

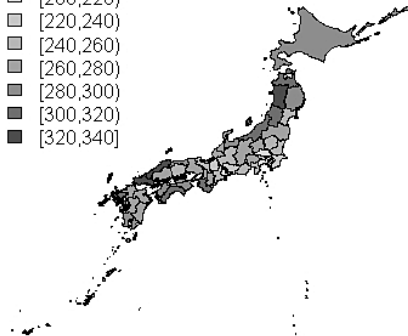


資料) 図2同様

図7 高血圧性疾患による死亡者数（階級数に依存しない分類）

Hypertensive diseases (pretty)

- [180,200)
- [200,220)
- [220,240)
- [240,260)
- [260,280)
- [280,300)
- [300,320)
- [320,340]



資料) 図2同様

* 参考文献・URL

- [1] R. S. Bivand, E. J. Pebesma and V. Gomez-Rubio (2008) : "Applied Spatial Data Analysis with R" : Springer.
- [2] CRANのHP (<http://cran.r-project.org/web/views/Spatial.html>)
- [3] 筆者のHP (<http://web.sfc.keio.ac.jp/~maunz/wiki/>)

- [4] ESRIジャパンのHP・全国市区町村界データ (http://www.esri.com/products/gis_data/japanshp/japanshp.html)
- [5] 総務省統計局：社会生活統計指標—都道府県の指標—2009 (<http://www.stat.go.jp/data/ssds/5.htm>)
- [6] e-StatのHP (<http://www.e-stat.go.jp/>)