Spatial Indices of Residential Segregation

Maurizio Pisati

Department of Sociology and Social Research
University of Milano-Bicocca (Italy)
maurizio.pisati@unimib.it

6th Italian Stata Users Group meeting
Grand Hotel Baglioni - Firenze, November 19-20, 2009

Outline

1. Overview
   - Introduction
   - Analysis of residential segregation
2. Description
   - Program
   - Notation
   - Formulas
3. Examples
4. Conclusion
5. References

Introduction

- The measurement of residential segregation is generally based on indices that do not appropriately take into account the spatial patterning of the phenomena of interest

Introduction

- The measurement of residential segregation is generally based on indices that do not appropriately take into account the spatial patterning of the phenomena of interest
- To overcome such limitation, new indices have been proposed that adequately account for the relationships of spatial contiguity among residential locations

Introduction

- The measurement of residential segregation is generally based on indices that do not appropriately take into account the spatial patterning of the phenomena of interest
- To overcome such limitation, new indices have been proposed that adequately account for the relationships of spatial contiguity among residential locations
- The purpose of this talk is to present a preliminary version of `spseg`, a novel user-written Stata program that computes three indices of this kind: two multigroup spatial clustering indices – the spatial dissimilarity index $\widetilde{D}$ and the spatial information theory index $\widetilde{H}$ – and the spatial exposure index $\widetilde{P}^*_{k\ell}$

What is residential segregation

- Residential segregation can be defined as the extent to which, within a given geographical area (e.g., a city), individuals belonging to different social groups live in neighborhoods characterized by different social compositions (Reardon and O'Sullivan 2004)

What is residential segregation

- Residential segregation can be defined as the extent to which, within a given geographical area (e.g., a city), individuals belonging to different social groups live in neighborhoods characterized by different social compositions (Reardon and O'Sullivan 2004)
- For example, suppose the inhabitants of Florence are classified into two social classes: white collars and blue collars. If the majority of white collars live in neighborhoods populated primarily by white collars, while most blue collars reside in neighborhoods whose population consists mainly of blue collars, then we can conclude that Florence is characterized by a certain degree of residential segregation by social class

Measuring residential segregation

- Measuring residential segregation within a given geographical area entails two basic operations: (a) defining the neighborhoods within which individuals live, and (b) quantifying the extent to which the distribution of the social attribute of interest (e.g., social class) varies across neighborhoods (Reardon and O'Sullivan 2004)

Measuring residential segregation

- Measuring residential segregation within a given geographical area entails two basic operations: (a) defining the neighborhoods within which individuals live, and (b) quantifying the extent to which the distribution of the social attribute of interest (e.g., social class) varies across neighborhoods (Reardon and O'Sullivan 2004)
- The indices of residential segregation most commonly used in the literature focus on the second operation, but completely elude the first one, thus failing to take into account the spatial nature of the phenomenon

Measuring residential segregation

- Expressly, the traditional indices of residential segregation equate the neighborhoods within which individuals live with the administrative units (e.g., the census tracts) into which the geographical area of interest is divided

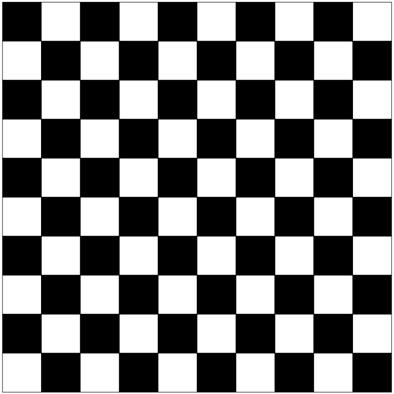
Measuring residential segregation

- Expressly, the traditional indices of residential segregation equate the neighborhoods within which individuals live with the administrative units (e.g., the census tracts) into which the geographical area of interest is divided
- The establishment of such equivalence between administrative units and neighborhoods, however, raises two issues potentially relevant to the measurement of residential segregation: (a) the checkerboard problem, and (b) the comparability problem

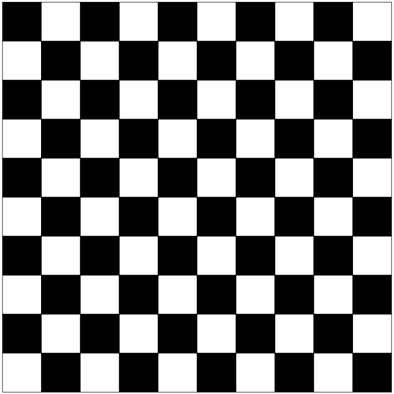
The checkerboard problem

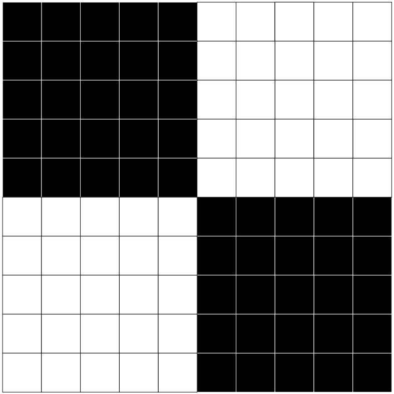
- The *checkerboard* problem stems from considering each administrative unit in isolation from the others, thus neglecting the overall social composition of its surrounding space

The checkerboard problem



The checkerboard problem





The comparability problem

- The issue of *comparability* is also very important and comes from the fact that different geographical areas – or the same geographical area at different times – are often divided into administrative units according to different criteria

The comparability problem

- The issue of *comparability* is also very important and comes from the fact that different geographical areas – or the same geographical area at different times – are often divided into administrative units according to different criteria
- This means that, whenever we equate neighborhoods with administrative units and are interested in carrying out a comparative analysis of residential segregation, different areas – or the same area at different times – will likely correspond to different definitions of neighborhoods, thus making any comparison of segregation levels unreliable

Some solutions

- To tackle the checkerboard and the comparability problems, in the last few years several scholars have devised new indices of residential segregation that take into account the spatial dimension of the phenomenon (e.g., Morgan 1983; White 1983; Morrill 1991; Wong 1993, 1998, 2003; Grannis 2002; Feitosa *et al.* 2004; Reardon e O'Sullivan 2004; O'Sullivan e Wong 2007)

Some solutions

- To tackle the checkerboard and the comparability problems, in the last few years several scholars have devised new indices of residential segregation that take into account the spatial dimension of the phenomenon (e.g., Morgan 1983; White 1983; Morrill 1991; Wong 1993, 1998, 2003; Grannis 2002; Feitosa *et al.* 2004; Reardon e O'Sullivan 2004; O'Sullivan e Wong 2007)
- These indices are based on definitions of neighborhoods that are less sensitive to the nature of pre-existing administrative units

Neighborhood definition

- Following Reardon e O'Sullivan (2004), the neighborhood N_r of each individual r living in the geographical area of interest is defined as the union of two sets of sub-areas (henceforth *tracts*):

Neighborhood definition

- Following Reardon e O'Sullivan (2004), the neighborhood N_r of each individual r living in the geographical area of interest is defined as the union of two sets of sub-areas (henceforth *tracts*):
  - The tract R where r lives

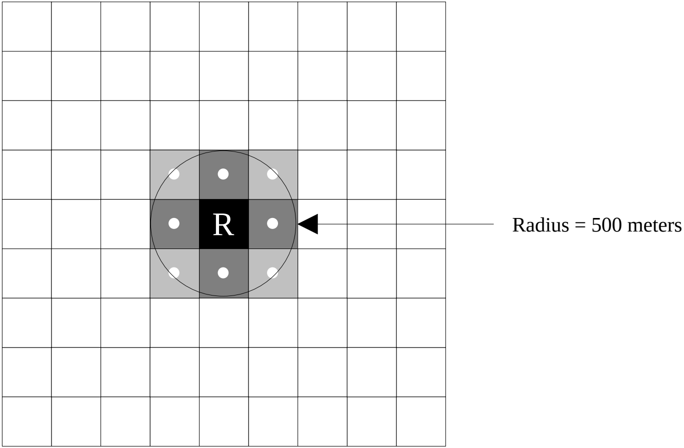
Neighborhood definition

- Following Reardon e O'Sullivan (2004), the neighborhood N_r of each individual r living in the geographical area of interest is defined as the union of two sets of sub-areas (henceforth *tracts*):
  - The tract R where r lives
  - The tracts whose centroids are located within a given radius from the centroid of tract R

Neighborhood definition

- Following Reardon e O'Sullivan (2004), the neighborhood N_r of each individual r living in the geographical area of interest is defined as the union of two sets of sub-areas (henceforth *tracts*):
  - The tract R where r lives
  - The tracts whose centroids are located within a given radius from the centroid of tract R
- The social composition of each neighborhood N_r is then computed by weighting every other individual q residing in N_r by a quantity proportional to the inverse of the distance between the centroid of tract R and the centroid of the tract where q lives, so as to give greater importance to the nearest neighbors of r, i.e., to those individuals with whom r is more likely to interact

Neighborhood definition



spseg

- spseg is a new user-written Stata program that implements the approach to measuring residential segregation outlined above

spseg

- spseg is a new user-written Stata program that implements the approach to measuring residential segregation outlined above
- Specifically, spseg allows the user to compute three spatial indices of residential segregation:

spseg

- spseg is a new user-written Stata program that implements the approach to measuring residential segregation outlined above
- Specifically, spseg allows the user to compute three spatial indices of residential segregation:
  - the spatial dissimilarity index $\widetilde{D}$

spseg

- spseg is a new user-written Stata program that implements the approach to measuring residential segregation outlined above
- Specifically, spseg allows the user to compute three spatial indices of residential segregation:
  - the spatial dissimilarity index $\widetilde{D}$
  - the spatial information theory index $\widetilde{H}$

spseg

- spseg is a new user-written Stata program that implements the approach to measuring residential segregation outlined above
- Specifically, spseg allows the user to compute three spatial indices of residential segregation:
  - the spatial dissimilarity index $\widetilde{D}$
  - the spatial information theory index $\widetilde{H}$
  - the spatial exposure index $\widetilde{P}^{*}_{k\ell}$

spseg

- spseg is a new user-written Stata program that implements the approach to measuring residential segregation outlined above
- Specifically, spseg allows the user to compute three spatial indices of residential segregation:
  - the spatial dissimilarity index $\widetilde{D}$
  - the spatial information theory index $\widetilde{H}$
  - the spatial exposure index $\widetilde{P}^*_{k\ell}$
- The definition of these three indices follows closely that given in Reardon e O'Sullivan (2004)

Basic quantities (1)

Symbol	Description
T	Number of tracts ($i, j = 1, \ldots, T$)
K	Number of groups ($k = 1, \ldots, K$)
n_i	Number of subjects living in tract i
n_k	Number of subjects belonging to group k
n_{ik}	Number of subjects living in tract i and belonging to group k
n	Total number of subjects living in the study area
p_i	n_i/n
p_k	n_k/n
I	$\sum_{k=1}^{K} p_k(1 - p_k)$

Basic quantities (2)

Symbol	Description
$\tilde{n}_i$	$\sum_{j=1}^{T} w_{ij} n_j$
$\tilde{n}_{ik}$	$\sum_{j=1}^{T} w_{ij} n_{jk}$
$\tilde{p}_{k\|i}$	$\tilde{n}_{ik} / \tilde{n}_i$
w_{ij}	$f(d_{ij}, r)$
d_{ij}	Euclidean distance between the centroids of tracts i and j
r	Radius of the neighborhood

Spatial dissimilarity index $\widetilde{D}$

The spatial dissimilarity index $\widetilde{D}$ is defined as follows:

$$\widetilde{D} = \sum_{i=1}^{T} p_i \widetilde{D}_i$$

where

$$\widetilde{D}_i = \sum_{k=1}^{K} \frac{1}{2I} \left| \tilde{p}_{k|i} - p_k \right|$$

Spatial dissimilarity index $\widetilde{D}$

- $\widetilde{D}$ can be interpreted as a measure of how different the social composition of neighborhoods is, on average, from the social composition of the study area

Spatial dissimilarity index $\widetilde{D}$

- $\widetilde{D}$ can be interpreted as a measure of how different the social composition of neighborhoods is, on average, from the social composition of the study area
- $\widetilde{D}$ can take values ranging from 0 (no segregation) to 1 (maximum possible segregation)

Spatial information theory index $\widetilde{H}$

The spatial information theory index $\widetilde{H}$ is defined as follows:

$$\widetilde{H} = \sum_{i=1}^{T} p_i \left(\frac{E - \widetilde{E}_i}{E} \right)$$

where

$$E = -\sum_{k=1}^{K} (p_k) \log_K (p_k)$$

$$\widetilde{E}_i = -\sum_{k=1}^{K} (\tilde{p}_{k|i}) \log_K (\tilde{p}_{k|i})$$

Spatial information theory index $\widetilde{H}$

- $\widetilde{H}$ can be interpreted as a measure of how much less socially diverse neighborhoods are, on average, than is the study area

Spatial information theory index $\widetilde{H}$

- $\widetilde{H}$ can be interpreted as a measure of how much less socially diverse neighborhoods are, on average, than is the study area
- $\widetilde{H}$ can take values ranging from 0 to 1, where 0 corresponds to the case in which each neighborhood has the same social composition as the study area (no segregation), and 1 corresponds to the case in which each neighborhood is populated by subjects belonging to one group only (maximum possible segregation)

Spatial exposure index $\widetilde{P}^*_{k\ell}$

The spatial exposure index $\widetilde{P}^*_{k\ell}$ is defined as follows:

$$\widetilde{P}^*_{k\ell} = \sum_{i=1}^{T} \frac{n_{ik}}{n_k} \tilde{p}_{\ell|i}$$

Spatial exposure index $\widetilde{P}^*_{k\ell}$

- $\widetilde{P}^*_{k\ell}$ expresses the degree of spatial exposure of group k to group ℓ, and corresponds to the average proportion of group ℓ in the neighbordhood of each member of group k

Spatial exposure index $\widetilde{P}^*_{k\ell}$

- $\widetilde{P}^*_{k\ell}$ expresses the degree of spatial exposure of group k to group ℓ, and corresponds to the average proportion of group ℓ in the neighbordhood of each member of group k
- When $\ell \equiv k$, then $\widetilde{P}^*_{kk}$ expresses the degree of spatial *isolation* of group k, i.e., the spatial exposure of group k to itself

Spatial exposure index $\widetilde{P}^*_{k\ell}$

- $\widetilde{P}^*_{k\ell}$ expresses the degree of spatial exposure of group k to group ℓ, and corresponds to the average proportion of group ℓ in the neighbordhood of each member of group k
- When $\ell \equiv k$, then $\widetilde{P}^*_{kk}$ expresses the degree of spatial *isolation* of group k, i.e., the spatial exposure of group k to itself
- In absence of segregation, $\tilde{P}^*_{k\ell}$ corresponds exactly to p_ℓ, i.e., to the proportion of group ℓ in the study area

Spatial exposure index $\widetilde{P}^*_{k\ell}$

- $\widetilde{P}^*_{k\ell}$ expresses the degree of spatial exposure of group k to group ℓ, and corresponds to the average proportion of group ℓ in the neighbordhood of each member of group k
- When $\ell \equiv k$, then $\widetilde{P}^*_{kk}$ expresses the degree of spatial *isolation* of group k, i.e., the spatial exposure of group k to itself
- In absence of segregation, $\tilde{P}^*_{k\ell}$ corresponds exactly to p_ℓ, i.e., to the proportion of group ℓ in the study area
- On the contrary, when there is segregation $\tilde{P}^*_{k\ell} \neq p_\ell$

The data

- Study area: City of Florence
- Administrative units: Census tracts
- Year: 2005
- Social attribute of interest: citizenship
- Social groups: (1) Italians, (2) Foreigners from "rich" countries, (3) Foreigners from "poor" countries

Dissimilarity index and Information theory index

Aspatial versions computed by user-written programs seg and spseg – Selected output

```
. use "K:\Sug09it\Florence-Attributes.dta", clear

. seg italian forrich forpoor, d h
Segregation Measures
```

--------	Dissim.	Inf. Theory
Total	0.2684	0.0640

```
. spseg italian forrich forpoor, x(xcoord) y(ycoord) band(1) kernel(quartic) d h
Spatial indices of segregation
```

Index	Estimate
Information Theory	0.064
Dissimilarity	0.268

Dissimilarity index and Information theory index

Spatial versions computed by spseg – Selected output

```
. spseg italian forrich forpoor, x(xcoord) y(ycoord) band(500) kernel(quartic) d h
Spatial indices of segregation
```

Index	Estimate
Information Theory	0.049
Dissimilarity	0.237

```
. spseg italian forrich forpoor, x(xcoord) y(ycoord) band(1000) kernel(quartic) d h
Spatial indices of segregation
```

Index	Estimate
Information Theory	0.041
Dissimilarity	0.211

```
. spseg italian forrich forpoor, x(xcoord) y(ycoord) band(2000) kernel(quartic) d h
Spatial indices of segregation
```

Index	Estimate
Information Theory	0.029
Dissimilarity	0.173

```
. spseg italian forrich forpoor, x(xcoord) y(ycoord) band(100000) kernel(quartic) d h
Spatial indices of segregation
```

Index	Estimate
Information Theory	0.000
Dissimilarity	0.000

Spatial exposure indices

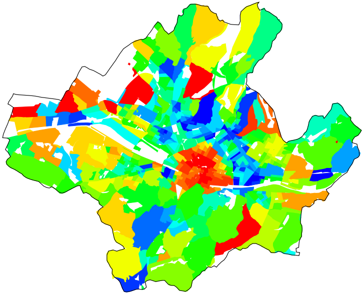
```
. spseg italian forrich forpoor, x(xcoord) y(ycoord) band(500) kernel(quartic) p

Spatial indices of segregation

Number of tracts   =         481
Number of groups   =           3
Bandwidth          =         500
Kernel function    =     quartic
```

Index	Estimate
Exposure of group 1 to 1	0.915
Exposure of group 1 to 2	0.011
Exposure of group 1 to 3	0.075
Exposure of group 2 to 1	0.885
Exposure of group 2 to 2	0.022
Exposure of group 2 to 3	0.093
Exposure of group 3 to 1	0.884
Exposure of group 3 to 2	0.013
Exposure of group 3 to 3	0.103

Marginal distributions

Groups	Absolute	Relative
Italians	333668	0.912
Foreigners - Rich countries	3988	0.011
Foreigners - Poor countries	28262	0.077

Spatial dissimilarity index – Local components $\widetilde{D}_i$



```
. use "K:\Sug09it\Florence-Attributes.dta", clear

. spseg italian forrich forpoor,      ///
    x(xcoord) y(ycoord) band(500)    ///
    kernel(quartic) d                ///
    ligen(i) id(id)

. spmap d_i using                             ///
    "Florence-Coordinates.dta",               ///
    id(id) cln(20) fcolor(Rainbow)            ///
    ocolor(none ..) ndocolor(none)            ///
    legend(off)                               ///
    polygon(data("Florence-Outline.dta"))
```

Conclusion

- `spseg` complements the existing Stata programs for the analysis of segregation, namely `seg` (by Sean F. Reardon), `duncan` (by Ben Jann) and `hutchens` (by Stephen P. Jenkins)

Conclusion

- `spseg` complements the existing Stata programs for the analysis of segregation, namely `seg` (by Sean F. Reardon), `duncan` (by Ben Jann) and `hutchens` (by Stephen P. Jenkins)
- spseg is still at a very preliminary stage and it will take some time before it reaches a publishable form

References

- Feitosa, F.F., G. Câmara, A.M.V. Monteiro, T. Koschitzki, and M.P.S. Silva. 2004. Spatial Measurement of Residential Segregation, in *Proceedings of GeoInfo 2004*, pp. 59–73. Geneve: IFIP.
- Grannis, R. 2002. Discussion: Segregation Indices and Their Functional Inputs. *Sociological Methodology* 32: 69–84.
- Morgan, B.S. 1983. An Alternate Approach to the Development of a Distance-Based Measure of Racial Segregation. *American Journal of Sociology* 88: 1237–1249.
- Morrill, R.L. 1991. On the Measure of Spatial Segregation. *Geography Research Forum* 11: 25–36.
- O'Sullivan, D. and D.W.S. Wong. 2007. A Surface-Based Approach to Measuring Spatial Segregation. *Geographical Analysis* 39: 147–168.

References

- Reardon, S.F. and D. O'Sullivan. 2004. Measures of Spatial Segregation. *Sociological Methodology* 34: 121–162.
- White, M.J. 1983. The Measurement of Spatial Segregation. *American Journal of Sociology* 88: 1008–1018.
- Wong, D.W.S. 1993. Spatial Indices of Segregation. *Urban Studies* 30: 559–572.
- Wong, D.W.S. 1998. Measuring Multiethnic Spatial Segregation. *Urban Geography* 19: 77–87.
- Wong, D.W.S. 2003. Implementing Measures of Spatial Segregation in GIS. *Computers, Environment and Urban Systems* 27: 53–70.