Geographic scales impacts on segregation index

The measurement of spatial segregation phenomena was mainly studied in the United States until the 1950's. A multitude of segregation indexes have been elaborated to analyze all the aspects of this subject. In 1988, Massey and Denton synthetized the scientist corpus by a classification among 5 dimensions: evenness, exposure, concentration, centralization and clustering.

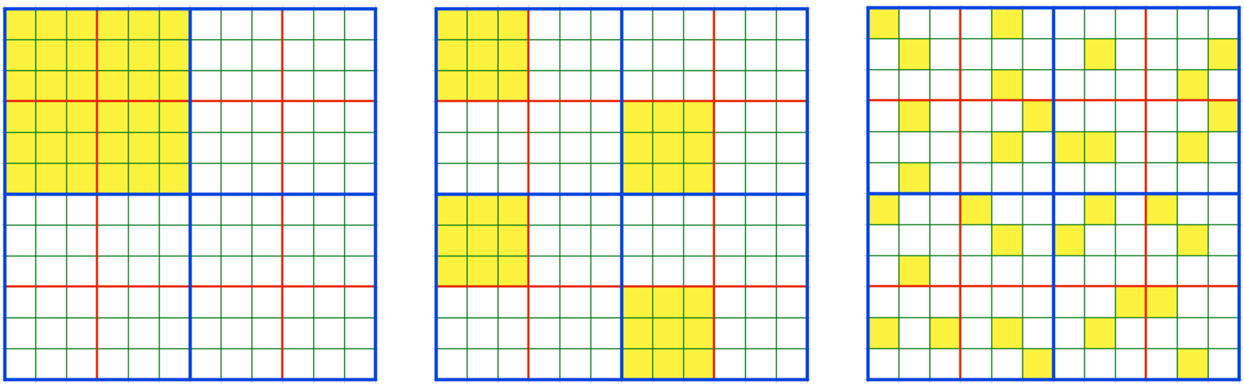
Whatever are the dimensions retained for the study, 3 choices need to be done:

- **The area of study**: it corresponds to the geographical or institutional spaces within we try to measure spatial segregation. It can involve a city, a region or a country.
- **The geographical subdivisions of the area of study**: spatial segregation indexes measure disparities of distribution in a population within the spatial units, which compose the area of study.
- **The population of study**: the choice of the population studied is made according to the objective of the study, under constraint of data's availability.

This work will focus on the usual Duncan's Segregation index. We will demonstrate in particular the geographical scale impact on this index.

To proceed, we will use a microsimulation model in order to test various hypotheses of scale level segregation. This model generates a fictive area cut at 3 different geographical scales named N1, N2, and N3. At the beginning of the simulation, the population will be randomly posted among the area whatever the geographical scales. In a second time, we will urge some people to move in specific spatial units with 3 hypotheses illustrated by the illustrations below. The yellow parts represent the units of grouping.

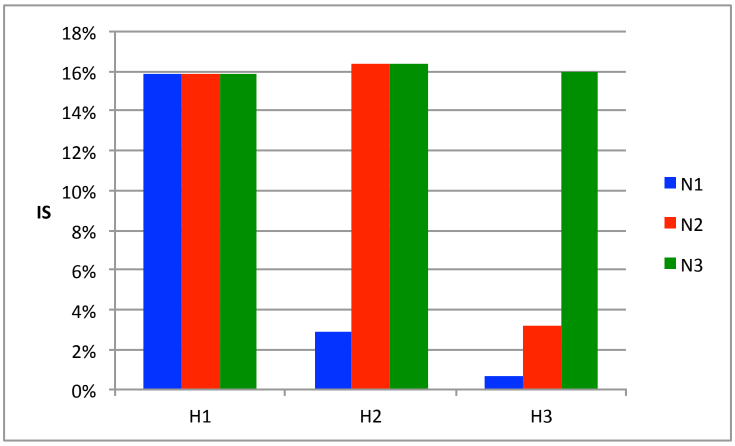
Hypotheses 1 *Hypotheses 2* *Hypotheses 3*



Source: Aurélien Dasre

The segregation can be calculated for the 3 geographical scales levels. Graphs below show the results for the 3 hypotheses at the 3 geographic scales:

Figure 1: IS calculated at the 3 geographical scales for the 3 hypotheses



Source: Aurélien Dasre

- **Hypothesis 1**: the segregation index calculated for the 3 geographical scales gives the same values.
- **Hypothesis 2**: the values calculated at N2 and at N3 scales remain stable while those that have been calculated at N1 scale decrease.
- **Hypothesis 3**: the index calculated at the N3 scale remain stable when the two others indexes give low values.

How can we explain these mechanisms? If we focus at figures schematizing the hypothesis of grouping by N1 scale, we notice that, if the people are congregated in some N1 units, it involves then inevitably that they group together into specifics N2 and N3 units. If there is no grouping at another level of aggregation, we can thus demonstrate that an index calculated at the finest level is mathematically superior or equal than indexes calculated at more included levels [1].

Within the framework of Duncan's segregation index, we can mathematically prove that there are simple relations between indexes calculated at different geographical scales.

Index calculated at the finest level is equal to the sum of the value of the index obtained at the most aggregated levels, plus the differences obtained by the index at the calculation scale.

[1] In theory, the indications calculated at an upper level of aggregation at the level of grouping put in hypothesis 2 and 3 should send back(dismiss) invalid(useless) values. In practice, during the edition of the reference municipalities, the effect of the random(unpredictable) variation makes that the distribution of these last ones between areas is not completed where from a not perfect distribution of the individuals between areas. A part of the value of the indication is also bound(connected) to the fact that the number of municipalities by area is not constant

Let's get back to the segregation index formula and see how it works:

By convention, we will say that an area contains "n" spatial units at N1 scale named Ai. In each Ai, there are "m" spatial units at N2 scale.

The basic formula calculated at the N1 level is:

$$IS_X = \frac{1}{2}\sum_{i=1}^{i=n}\left|\frac{X_{Ai}}{X} - \frac{Y_{Ai}}{Y}\right|$$

If we decompose this formula by N1, we find:

$$IS_X = \frac{1}{2}\left[\left|\frac{X_{A1}}{X} - \frac{Y_{A1}}{Y}\right| + \left|\frac{X_{A2}}{X} - \frac{Y_{A2}}{Y}\right| + \ldots + \left|\frac{X_{An}}{X} - \frac{Y_{An}}{Y}\right|\right]$$

Every elements of this sum represent the contribution of the N1 unit i to the global value of the index. We can decompose each of these elements according to the N2 units that compose every N1 unit. We find then for a N2 area made of m units the following relations:

$$\left|\frac{X_{Ai}}{X} - \frac{Y_i}{Y}\right| = \left[\left|\frac{X_{c1} + X_{c2} + \ldots + Xc_m}{X} - \frac{Y_{c1} + Y_{c2} + \ldots + Y_{cm}}{Y}\right|\right]$$

=>

$$\left|\frac{X_{Ai}}{X} - \frac{Y_i}{Y}\right| = \left[\left|\frac{X_{c1}}{X} + \frac{X_{c2}}{X} + \ldots + \frac{X_{cm}}{X} - \frac{Y_{c1}}{Y} + \frac{Y_{c2}}{Y} + \ldots + \frac{Y_{cm}}{Y}\right|\right]$$

=>

$$\left|\frac{X_{Ai}}{X} - \frac{Y_i}{Y}\right| = \left[\left|\frac{X_{c1}}{X} - \frac{Y_{c1}}{Y} + \frac{X_{c2}}{X} - \frac{Y_{c2}}{Y} + \ldots + \frac{X_{cm}}{X} - \frac{Y_{cm}}{Y}\right|\right]$$

At the same time, if we decompose the formula calculated at the N2 units level and for which we try to determine the impact of the N2 scale of an N1 area i on the total value of the index, we get the following formula:

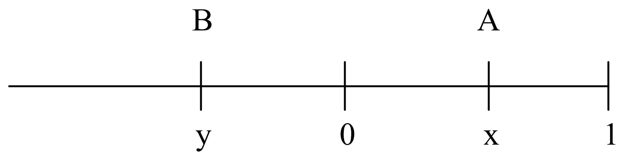
$$\left[\left|\frac{X_{c1}}{X} - \frac{Y_{c1}}{Y}\right| + \left|\frac{X_{c2}}{X} - \frac{Y_{c2}}{Y}\right| + \ldots + \left|\frac{X_{cm}}{X} - \frac{Y_{cm}}{Y}\right|\right]$$

Or, we can demonstrate that:

$$\left[\left|\frac{X_{c1}}{X} - \frac{Y_{c1}}{Y} + \frac{X_{c2}}{X} - \frac{Y_{c2}}{Y} + \ldots + \frac{X_{cm}}{X} - \frac{Y_{cm}}{Y}\right|\right] \leq \left[\left|\frac{X_{c1}}{X} - \frac{Y_{c1}}{Y}\right| + \left|\frac{X_{c2}}{X} - \frac{Y_{c2}}{Y}\right| + \ldots + \left|\frac{X_{cm}}{X} - \frac{Y_{cm}}{Y}\right|\right]$$

We can prove this relation using the demonstration of the triangular inequality:

Let a line (O; I) and points A and B of respective abscissa x and -y



We find this relation:

$$|x + y| = |x - (-y)| = d(AB)$$ [2]

With
$$|x| = d(OA)$$
$$|y| = |-y| = d(OB)$$

Or

$$d(AB) \leq d(OA)+d(OB)$$

So,

$$|A + B| \leq |A| + |B|$$

This demonstration confirms us that Duncan's index of segregation calculated at different geographical scales cannot produce the same results. The finest will be the scale, the higher will be the index. But differences between indexes calculated at different levels follow a mathematical relation, which can help us in a better comprehension of the segregation phenomena's.

We have thus the capacity to determine the relative weight of every levels of geographical aggregations into the value of the index calculated at the finest level.

If applying this methodology for the French urban area, we will be able to demonstrate that some kind of populations will be segregated in a confrontation inner city/suburbs like in Bordeaux, while in other areas like Marseille: people are segregated regarding to city blocs.

[2] With d(AB)=distance between A and B