Inequality and Income Segregation in Brazilian cities: A Nationwide Analysis

José Firmino de Sousa Filho (✉ jose.sousa@ufba.br)
  Center of Data and Knowledge Integration for Health (CIDACS); School of Economics (PPGE) – Federal University of Bahia   https://orcid.org/0000-0001-5057-385X

Gervásio F. dos Santos
  Center of Data and Knowledge Integration for Health (CIDACS); School of Economics (PPGE) – Federal University of Bahia   https://orcid.org/0000-0002-3545-3590

Roberto F. Silva Andrade
  Center of Data and Knowledge Integration for Health (CIDACS); Institute of Physics – Federal University of Bahia   https://orcid.org/0000-0002-9323-1400

Aureliano S. Paiva
  Center of Data and Knowledge Integration for Health (CIDACS); 6Institute of Physics – Federal University of Bahia   https://orcid.org/0000-0002-4638-4199

Anderson Freitas
  Center of Data and Knowledge Integration for Health (CIDACS);   https://orcid.org/0000-0002-6764-9301

Caio Porto Castro
  Center of Data and Knowledge Integration for Health (CIDACS); Institute of Physics – Federal University of Bahia   https://orcid.org/0000-0001-5316-7515

Amélia A. de Lima Friche
  Observatory for Urban Health in Belo Horizonte (OSUBH) – Federal University of Minas Gerais
  https://orcid.org/0000-0002-2463-0539

Sharrelle Barber
  Department of Epidemiology and Biostatistics – Drexel University Dornsife School of Public Health
  https://orcid.org/0000-0002-4699-187X

Waleska T. Caiaffa
  Observatory for Urban Health in Belo Horizonte (OSUBH) – Federal University of Minas Gerais
  https://orcid.org/0000-0001-5043-4980

Maurício L. Barreto
  Center of Data and Knowledge Integration for Health (CIDACS); Institute of Public Health (ISC) – Federal University of Bahia   https://orcid.org/0000-0002-0215-4930

Research Article

Inequality and Income Segregation in Brazilian cities: A Nationwide Analysis

José Firmino de Sousa Filho[1,2]; Gervásio F. dos Santos[1,2]; Roberto F. Silva Andrade[1]; Aureliano S. Paiva[1]; Anderson Freitas[1]; Caio Porto Castro[1]; Amélia A. de Lima Friche[4]; Sharrelle Barber[5]; Waleska T. Caiaffa[4]; Maurício L. Barreto[1,3]

[1]Center of Data and Knowledge Integration for Health (CIDACS); [2] School of Economics (PPGE) – Federal University of Bahia; [3]Institute of Public Health (ISC) – Federal University of Bahia; [4]Observatory for Urban Health in Belo Horizonte (OSUBH) – Federal University of Minas Gerais; [5]Department of Epidemiology and Biostatistics – Drexel University Dornsife School of Public Health; [6] Institute of Physics – Federal University of Bahia.

Abstract

Urban segregation has brought significant challenges to cities worldwide and has important implications for health. This study aimed to assess income segregation in the 152 largest Brazilian cities included in the SALURBAL Project and identify specific socioeconomic characteristics related to residential segregation by income. Using the Brazilian demographic census database of the year 2010, we calculated the income dissimilarity index (IDI) at census tract level for each SALURBAL city; subsequently comparing it with Gini and other local socioeconomic variables. We evaluated our results' robustness using a bootstrap correction to the IDI to examine the consequences of using different cut-offs of income that were relevant in the context of strong urban and regional inequalities. We identified a 2 minimum wages cut-off as the most appropriate and found little evidence of upward bias in the calculation of the IDI regardless of the cut-off used. Among the 10 most segregated cities, 9 are in the Northeast region, the region with the highest income inequality and poverty in Brazil. Our results indicate that the Gini index and poverty are the main variables associated with residential segregation, measured by the IDI. Social and environmental characteristics were also associated with IDI, reinforcing the notion that access to education, water, sanitation, and better residential conditions are fundamental to improving social equity.

Keywords: Segregation. Income Dissimilarity Index. Urban Inequality. Brazil.

1. Introduction

The phenomenon of the expansion of cities and the process of transformation of urban areas has led to the creation of various forms of spatial territories. Central and peripheral areas are objects of theoretical and empirical studies worldwide. The socioeconomic configurations that have emerged since then stem from dynamic city centers with multiple and often conflicting features. In this sense, residential segregation and inequality arise as a consequence of a series of actions that can be subjective and individual or material due to a superstructure of discrimination (Schelling, 1971; White, 1983; Iceland et al., 2002; Reardon & Bischoff, 2011).

The dimensions of segregation include scale, evenness, entropy, exposure, concentration, and centralization (Massey & Danton, 1988; Acevedo-Garcia *et al*., 2003; Reardon & Bischoff, 2011; Garcia-López & Moreno-Monroy, 2018). The heterogeneity of services, infrastructure, access, or lack of resources contained in each neighborhood within cities differentially impacts the income conditions of residents and affects the process of segregation. Segregation and inequality are severe problems observed worldwide (United Nations, 2020). In Latin American countries, and Brazil, in particular, the situation is often especially stark (Ferreira & Ravallion, 2008).

Residential segregation accentuates social and economic inequalities and reveals an unequal social structure. In this sense, segregation is a dimension of social inequality but presented as fractioned spatial arrangements that impede social mobility since groups are distinguished in relation to social isolation or exclusion. It can be reflected in cities where the economically most vulnerable people have great barriers to getting to work, young people going to school, and other dimensions of social interactions and engagement.

Santos *et al*. (2021) calculated income segregation for Brazilian cities and highlighted how it affects homicide patterns in large urban centers. According to the authors, segregation is further aggravated by high regional inequalities in Brazil, and under these circumstances, there is a discriminatory social structure that involves the spatialization of inequality and poverty. Thereby affecting the collective health of individuals and households that find themselves in a situation of socioeconomic vulnerability.

This study examined the association between income segregation and inequality as major factors for promoting equity and social development in large Brazilian urban centers, using the dissimilarity index. The Income Dissimilarity Index (IDI) originally proposed by Duncan and Duncan (1955) allows a spatial view of the population that would need to be "moved" within a city to make it homogeneous in terms of income. It is a measure of average segregation that indicates how far the poorest are from the city's average income (Massey & Danton, 1988; Iceland et al., 2002). To define an income cut-off at the national level, we initially tested different low-income minority groups: households that earn 0 to ½ wage, 0 to 1, 0 to 2, and 0 until 3 wages[1]. Subsequently, we obtained the IDI based on the income data of all 164,109 census tracts of 152 Brazilian cities as defined by SALURBAL project (Quistberg *et al.* 2019; Roux *et al*. 2019).

---

[1] In 2010, the official monthly nominal minimum wage in Brazil was equal to R$ 510.00 or approximately US$ 289.00.

This paper seeks to answer three important research questions: (i) What is the best s income cut-off for shares of low-income households to calculate IDI?; (ii) Is there a bias in the IDI to the point of making the indicator inconsistent?; (iii) how is the IDI associated with other socioeconomic indicators in a context of high regional inequality?

We found that the cut-off of 2 minimum wages is adequate to define a low-income minority group at the national level in Brazil since its represents 33.8% of the sample. This mean that the percentage of the low-income group is a major factor in the analysis of residential segregation measured by income. Through sensitivity analyses (Allen *et al.* 2015; Tivadar, 2019) using bootstrap method, we observed that there is no relevant bias embedded in the income-based dissimilarity index specifically for Brazil. Therefore, in the context of great regional income inequality, the spatial distribution of the low-income minority is essential to measure the heterogeneity within a city and efficiently capture the level of segregation.

The heterogeneity we refer to means reaching a low-income minority group, capable of adequately representing intra-urban and regional inequality patterns. Finally, using linear regression models we find that the Gini index and poverty are the main variables associated with residential segregation, measured by the IDI. Other social and environmental characteristics were also associated with IDI, including education, access to water, sanitation, and better residential conditions.

2. Data and Methods

2.1 Data

We used the sample of the 152 Brazilian cities included in the SALURBAL project comprising urban agglomerations with more than 100,000 inhabitants in 2010. The urban agglomerations (referred to as L1ADs) comprise 422 sub-city units (municipalities) (Quistberg et al. 2019).

Income data by census tract were collected by the Brazilian Bureau of statistics (IBGE, Portuguese acronym) for the year 2010 and was harmonized and consolidated by the SALURBAL project. In Brazil, data at the census tract level are collected every 10 years. The last demographic census was carried out in 2010.

The socioeconomic variables examined in relation to IDI are shown in Table 1. SEI is an indicator that takes into account the level of education (proportion of the population aged 25 or older who completed primary education or above), access to water, sanitation, and housing conditions based on the number of households overcrowding. The overcrowding

indicator was reverse coded so that higher values all measures indicate a better social environment. Finally, the four measures were summed and divided by 4, assuming equal weights for each one (Bilal, 2021). The SEI index and other covariates were transformed into a Z-score.

Three models were estimated taking the IDI as the exposure variable. The first, in model 1, the Gini index was used as outcome and in the second, the SEI. In the third model we included all other covariates.

Table 1. Variables used in the regression, all data for 2010.

Variables	Definition
Population	Projected population.
Gini index	Income inequality based on the household total income.
GDP per capita	Nominal GDP (UU$ dollars)/Population.
Unemployment	The unemployment rate among the total population 15 years or above in the labor force.
Poverty rate	The proportion of the population living in households with household income below the national income poverty line.
Social Environment Index (SEI)	Education / water access / sanitation / overcrowding (reverse coded). Indices summed and divided by 4 assuming equal weights for all four measures.

Source: SALURBAL Project.

2.2 Definition of the dissimilarity index

The allocation of individuals within a space can be purely random, or it can reflect the influence of economic, social, and environmental determinants. In general, the systematic allocation process is influenced by the preferences or restrictions of individuals and occupied areas.

Allen, Burgess and Windmeijer (2009), Allen *et al*. (2015), Mazza and Punzo (2015), Mazza (2017) offer statistical tools capable of measuring the distribution of individual allocations in space through the dissimilarity index. All specifications described here are

based on the methodology developed by these authors. Hereafter, we assume that there are units $j = 1, ..., J$, nested within a given urban area. Individuals may or may not have measurable characteristics on a dichotomous scale, $c = 0, 1$. The number of individuals with status c in the urban area is denoted n^c. Individuals are allocated, n_j^c individuals in unit j having status c, and the total number of individuals in unit j is represented by $n_j = n_j^1 + n_j^0$.

Then, we can use the dissimilarity index D for a given region, as follows:

$$D = \frac{1}{2} \sum_{j=1}^{J} \left| \frac{n_j^1}{n^1} - \frac{n_j^0}{n^0} \right| \tag{2.2.1}$$

The allocation is given by a set of probabilities p_j^c that assigns individual i for unit j according to individual status c:

$$p_j^a \equiv P(unit = j | c = a), j = 1, ..., J; a = 0,1. \tag{2.2.2}$$

Systematic segregation is present when there is j such that $p_j^1 \neq p_j^0$. The relationship between D and the underlying allocation probabilities is denoted by the fraction $\frac{n_j^c}{n^c}$, $c = 0,1$. Since $\hat{p}_j^c = \frac{n_j^c}{n^c}$, therefore, D is only half of $\sum_{j=1}^{J} |\hat{p}_j^1 - \hat{p}_j^0|$. The objective is to recognize the dissimilarity index as an estimator for the population quantity to allocate individuals independently. Thus, the population of a region with an individual number n, proportionally $p = n^1/n$ with status $c = 1$ is allocated in J units according to the p_j^c probability. The probability function captures all allocations and the results are the allocations $n_j = n_j^1 + n_j^0$ determined by a stochastic allocation, and the sizes of the units are given by:

$$E[n_j] = n^1 p_j^1 + n^0 p_j^0 \tag{2.2.3}$$

Therefore, we can rewrite D as:

$$D_{pop} = \frac{1}{2} \sum_{j=1}^{J} |p_j^1 - p_j^0| \tag{2.2.4}$$

$D_{pop} = 0$ if and only if $p_j^1 = p_j^0 \; \forall \, j$.

5

Maximum likelihood functions can estimate the conditional probabilities as independent allocations of multinomial distributions, and the function's log-likelihood can be described as:

$$\log L = log\left(\frac{n^1!}{n_1^1! \dots n_J^1!}\right) + log\left(\frac{n^0!}{n_1^0! \dots n_J^0!}\right) + \sum_{j=1}^{J} n_j^1 \, log(p_j^1) + \sum_{j=1}^{J} n_j^0 \, log(p_j^0) \quad (2.2.5)$$

The maximum likelihood estimator is given by $\hat{p}_j^1 = \frac{n_j^1}{n^1}$ e $\hat{p}_j^0 = \frac{n_j^0}{n^0}$, $j = 1, \dots, J$. expressed by the product of two independent multinomial distributions, one for $c = 0$ and one for $c = 1$:

$$P\left(n_1^0, \dots, n_j^0, n_1^1, \dots, n_j^1; p_1^0, \dots, p_j^0, p_1^1, \dots, p_j^1, n^0, n^1\right) = \prod_{j=1}^{J} \prod_{c=0}^{1} n^c! \, \frac{(p_j^c)^{n_j^c}}{n_j^c!} \quad (2.2.6)$$

When we assume that the size of units n_j is fixed, we apply another model and additionally assume that the size of population n and the minority proportion p are also fixed. The allocation can be given by conditioned probabilities and D_{pop} is written as:

$$D_{pop} = \frac{1}{2} \sum_{j=1}^{J} p \, (unit = j) \left| \frac{P(c = 1|unit = j)}{P(c = 1)} - \frac{1 - P(c = 1|unit = j)}{1 - P(C = 1)} \right|$$

$$= \frac{1}{2} \sum_{j=1}^{J} \frac{n_j}{n} \left| \frac{P(c = 1|unit = j)}{P(c = 1)} - \frac{1 - P(c = 1|unit = j)}{1 - P(C = 1)} \right| \quad (2.2.7)$$

Therefore, with a complete population or with a random sample, D will remain an estimator of D_{pop}, both in cases of random effects or units of fixed sizes. This statistical distribution allows us to demonstrate whether or not there is bias in the income dissimilarity index, as described below.

2.3 Measuring the presence of bias in the index

The bias can arise when the area population is small or the minority group proportion is very low. Consequently, the index is affected by differences in the proportion of the minority in the population and by the size of the areal unit of analysis, making it difficult to compare the indices across cities. Fosset (2017) presents a series of "rule-of-thumb" practices to minimize the problem of bias, among which we highlight: (i) the use of census tracts and (ii)

focus on the comparison between broader minority and the rest of the population versus more finely categorized groups.

We followed these recommendations by working with census tract level data and explored the bias generated by different income cut-offs. In order to evaluate bias embedded in the IDI we use randomization models with bootstrap. For that, we take as reference Efron (1979) in which for an estimator $\hat{\theta} = s(x)$ the bootstrap bias estimator is defined as $bias_{\hat{F}}$:
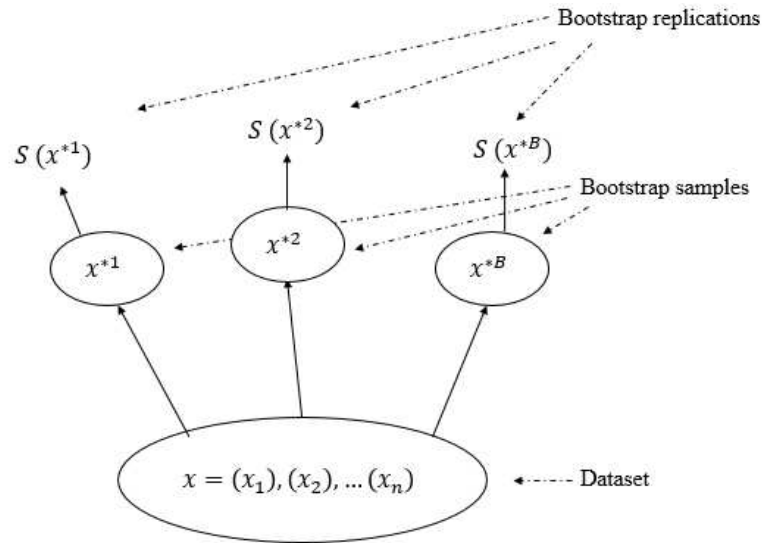
$$bias_F = bias_F\left(\hat{\theta}, \theta\right) = E_F\left[s(x)\right] - t(F) \qquad (2.3.1)$$

$$bias_{\hat{F}} = E_{\hat{F}}\left[s(x^*)\right] - t\left(\hat{F}\right) \qquad (2.3.2)$$

Where $t\left(\hat{F}\right)$ is the estimator of θ is different from $\hat{\theta} = s(x^*)$. $bias_{\hat{F}}$ is the plug-in estimator of $bias_F$, and $\hat{\theta}$ may or may not be the θ. Efron (1979) also demonstrates the ideal bootstrap estimator through simulations by Monte Carlo, in which independent bootstrap samples $(x^{*1}), (x^{*2}), \dots s(x^{*B})$, are generated, as in Figure 1, calculating the bootstrap replications $\hat{\theta}^*(b) = s(x^{*b})$ and approximating the expected value $E_{\hat{F}}\left[s(x^*)\right]$ by the mean

$$\hat{\theta}^*(.) = \sum_{b=1}^{B} \frac{\hat{\theta}^*(b)}{B} = \sum_{b=1}^{B} \frac{s(x^{*b})}{B} \qquad (2.3.3)$$

Figure 1. Illustration of bootstrap replications

Source: Efron, 1979.

Bootstrap samples are generated from the original data set. Each bootstrap sample has n elements, generated by sampling with replacement n times from the original data set. Bootstrap replicates $s(x^{*1}), (x^{*2}), \dots s(x^{*B})$ are obtained by calculating the value of the statistic $s(x)$ on each bootstrap sample. Finally, the standard deviation of the values $(x^{*1}), (x^{*2}), \dots s(x^{*B})$ is our estimate of the standard error of $s(x)$.

The bootstrap bias estimator based on B replications $\widehat{bias}_B$, is (3.3.2) with $\hat{\theta}^*(.)$ replaced by $E_{\hat{F}}[s(x^*)]$:

$$\widehat{bias}_B = \hat{\theta}^*(.) - t(\hat{F}) \tag{2.3.4}$$

Concerns about the bias embedded in the dissimilarity index are not recent. Carrington and Troske (1997) developed indexes of systematic segregation that measure the distance of randomness samples instead of the distance of uniformity. The use of randomness in the allocation implies a substantial unevenness caused by small minority shares in relation to the whole or census tracts that are not very representative in terms of income heterogeneity between households. This reinforces the need to check for bias in the indicator when examining associations with other socioeconomic, health, environmental, built factors, etc.

2.4 Associations between the dissimilarity index and other socioeconomic variables

The linear regression method is a classic technique of statistical mathematics and an essential part of modern econometrics theory (Rao & Toutenburg, 1995; Wooldridge, 2019). We take a generalized and multivariable linear equation and ran three models to look for associations between the dissimilarity index and socioeconomic variables. In the first model, the outcome of interest is the Gini index traditionally used to measure income inequality between households, cities or regions. In the second model we use the social environment index, which aggregates important variables for social development in a city. Finally, in the third model, we introduce the entire set of covariates.

3. Results and Discussion

Table 2 provides an overview of our data. We stratified the sample of studied cities by population quartiles to observe the pattern of the different indicators according to the size of the city population. Both the dissimilarity and Gini indexes increase with the population size. Cities in the second quartile of the population have a higher GDP *per capita* due to an

accentuated growth in medium-sized cities in Brazil (Henderson, 1997; Mata et al., 2005; Bolay, 2020). This finding can also be explained by the high cost of living in the country's major metropolitan centers. This is reflected in monetary matters and is also related to the quality of life and leisure of households in large cities.

Unemployment and poverty did not differ much between cities according to population size. However, the social environment index points out that larger cities can provide more services to improve the well-being of households.

Table 2. Descriptive statistics by city population quartiles

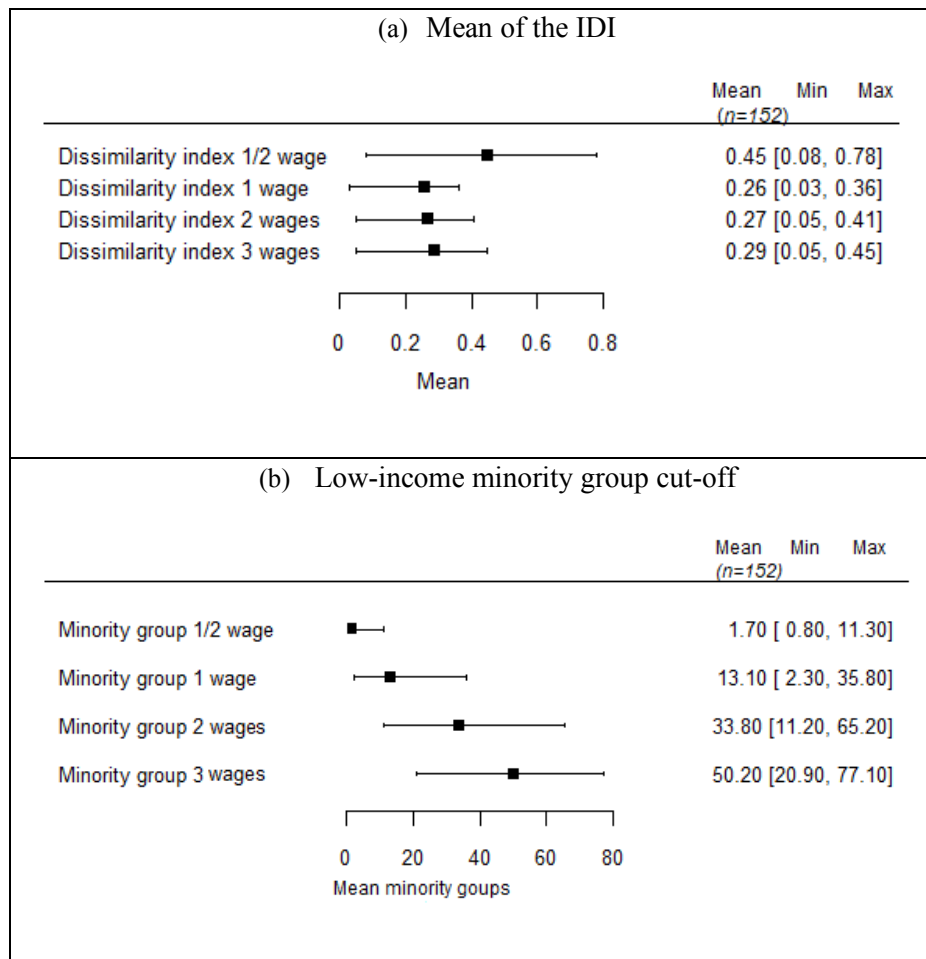| Variables	**Q1**(N = 38) Mean (SD)	**Q2** (N = 38) Mean (SD)	**Q2** (N = 38) Mean (SD)	**Q4** (N = 38) Mean (SD)
Dissimilarity index	0.25 (0.04)	0.25 (0.04)	0.26 (0.04)	0.30 (0.04)
Gini	0.53 (0.03)	0.53 (0.02)	0.54 (0.03)	0.60 (0.04)
GDP *per capita* (US$)	14.2 (9.1)	18.5 (12.8)	16.5 (9.6)	17.3 (7.0)
Unemployment	0.09 (0.03)	0.08 (0.03)	0.08 (0.02)	0.09 (0.02)
Poverty rate	0.26 (0.15)	0.24 (0.12)	0.25 (0.13)	0.24 (0.11)
SEI	0.07 (0.39)	0.02 (0.34)	0.09 (0.44)	0.19 (0.36)

As shown in Figure 2, the cut-offs analyzed for the 152 cities in the SALURBAL project presented different average, maximum, and minimum levels. The mean income segregation of the ½ minimum wage cut-off was 45%. However, the proportion of households in the minority low-income group is very small, with only 1.7% of Brazilian households (Figure 2(b). Therefore, it is not reasonable to work with this cut-off.
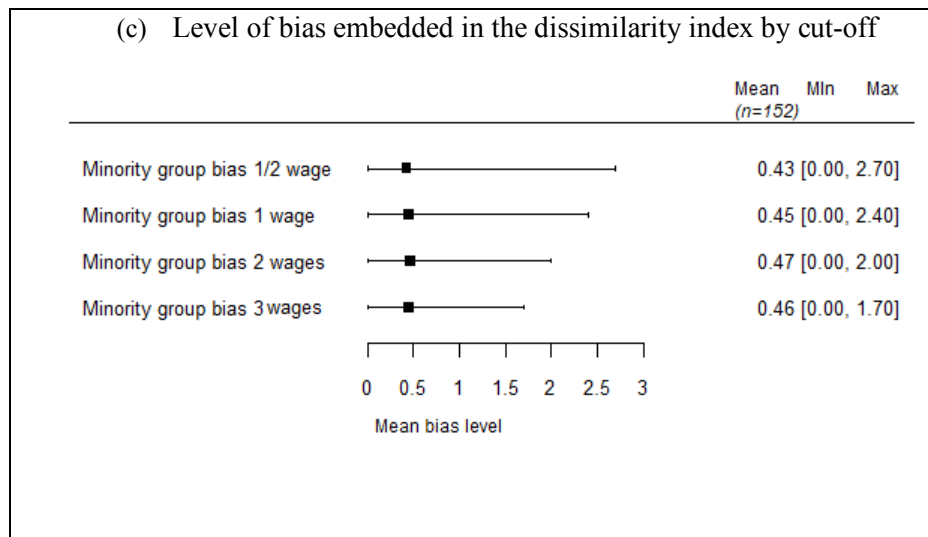
Based on this same criterion, the indicator of 1 minimum wage also represents a small proportion of households, only 13.1% are in the low-income minority group. Therefore, cut-offs of 2 and 3 minimum wages present better distributed samples and have an average of

income segregation, 27%, and 29%, respectively. However, the low-income cut-off related to households earning up to 3 minimum wages exceeds 50% of the sample, so this group cannot be used as a minority group.

The level of bias found in the IDI was low, and did not appear to be dependent on the cut-off level used, as shown in Figure 2(c). The Brazilian bureau of statistics defines census tracts as a continuous cadastral control unit. Every sector must be fully contained in an urban or rural area and its dimension in numbers of households and establishments must allow the enumerator to carry out the work within a specified period. The number of households and households in the census tracts is similar regardless of the city (IBGE, 2020), which also explains the low level of bias found. For all income groups analyzed, on average, the bias remained below 0.5%.
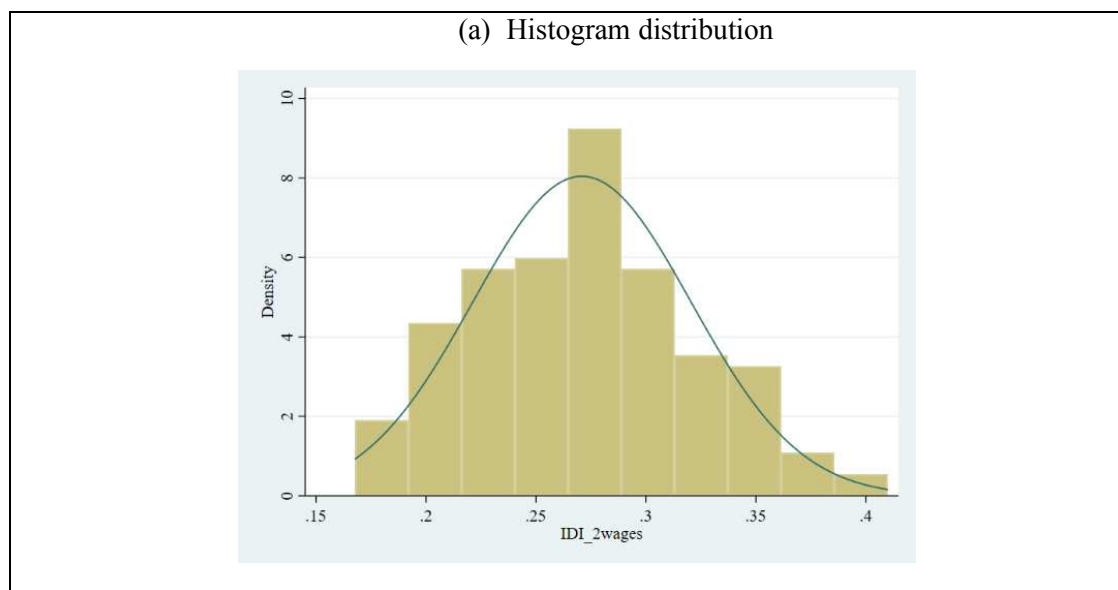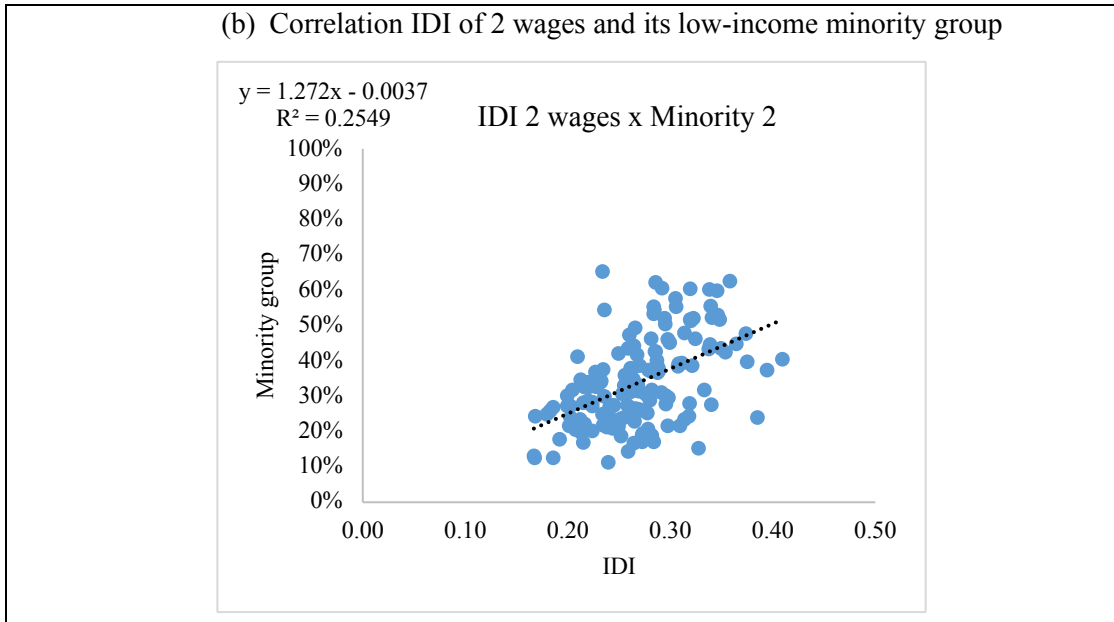
Figure 2. Descriptive statistics for the indices

(a) Mean of the IDI

	Mean Min Max (n=152)
Dissimilarity index 1/2 wage	0.45 [0.08, 0.78]
Dissimilarity index 1 wage	0.26 [0.03, 0.36]
Dissimilarity index 2 wages	0.27 [0.05, 0.41]
Dissimilarity index 3 wages	0.29 [0.05, 0.45]

(b) Low-income minority group cut-off

	Mean Min Max (n=152)
Minority group 1/2 wage	1.70 [0.80, 11.30]
Minority group 1 wage	13.10 [2.30, 35.80]
Minority group 2 wages	33.80 [11.20, 65.20]
Minority group 3 wages	50.20 [20.90, 77.10]

(c) Level of bias embedded in the dissimilarity index by cut-off

	Mean (n=152)	MIn	Max
Minority group bias 1/2 wage	0.43	[0.00,	2.70]
Minority group bias 1 wage	0.45	[0.00,	2.40]
Minority group bias 2 wages	0.47	[0.00,	2.00]
Minority group bias 3 wages	0.46	[0.00,	1.70]

Source: Research results.

Figure 3(a) shows the histogram for the distribution of the percent of households earning up 2 minimum wages across cities. As the level of bias, specifically for the Brazilian case, was not dependent on the sample mean in any of the cases, we can use the indicator with or without correction. Therefore, we use the indicator originally calculated by the SALURBAL Project. Figure 3(b) presents the distribution of 2 minimum wage IDI and the percent minority low-income group for cities. The correlation between both measures was approximately 0.25.

Figure 3. Statistical distribution for the income dissimilarity index

(a) Histogram distribution

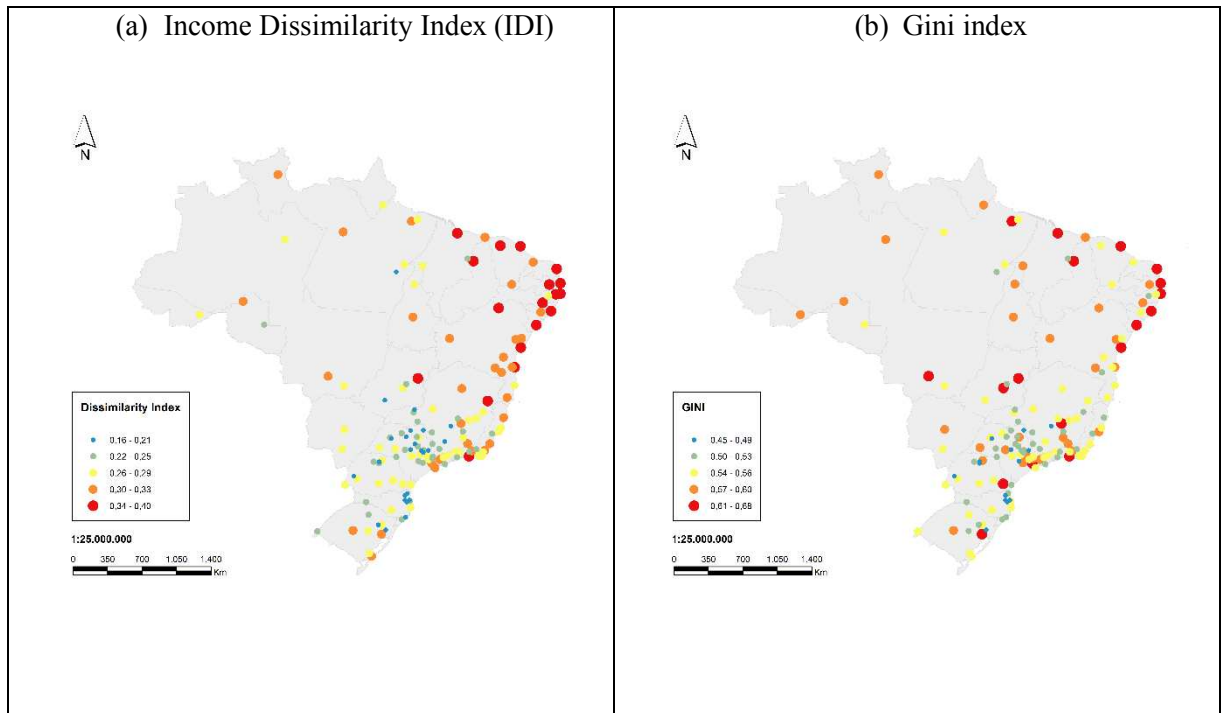(b) Correlation IDI of 2 wages and its low-income minority group

Source: Research results.

In subsequent analyses we used the 2 wages cut-off to look for associations between residential segregation indicators and other socioeconomic variables.

3.1 Spatial distribution of indicators

Figure 4 shows the spatial distribution of the IDI for the 152 Brazilian cities (or urban agglomerations) of the SALURBAL project. Spatial patterns in income segregation are similar to spatial patterns in the Gini coefficient revealing increased income vulnerabilities in Brazil's North and Northeast regions (Reis, 2014). The North and Northeast regions are also those with the highest percentage of the indigenous and black and population in the country, which points to a possible correlation between income segregation and the high levels of racial inequalities in Brazil (Bailey, Loveman & Muniz, 2013).

Figure 4. Income dissimilarity and Gini indexes for 152 Brazilian cities, 2010.

(a) Income Dissimilarity Index (IDI)	(b) Gini index

Source: SALURBAL Project.

In the field of the regional economy, studies such as Akita (2003), Elbers et al. (2003), Trendle (2005), Tarozzi and Deaton (2009), Miranti et al. (2015) used statistical and spatial methods to estimate income inequalities, poverty, demographics, Gini, market composition workplace, educational issues and, racial segregation. The literature converges in the direction that policies aimed at reducing regional inequalities, income, and greater social welfare are beneficial as a whole.

Tarozzi and Deaton (2009) also estimated levels of inequality and poverty for small areas of Mexico's cities using census data from the 2000s. According to the authors, regional inequality patterns could be observed according to predictors such as education, family spending, housing, family composition, employment level, etc. Therefore, the spatialization of inequality can be relevant in constructing policies aimed at reducing segregation and poverty.

Torres (2006) calculated the dissimilarity index for the city of São Paulo, the largest city in Brazil, and highlighted the importance of the indicator in providing consistent socioeconomic evidence on residential segregation. For the author, segregation is a phenomenon that has consequences in a broad sense, but it can be mitigated with income distribution and housing policies. However, the rapid urbanization of cities can implement policies that alleviate segregation complex, given the lack of funding for public works, increased violence, and social degradation of marginalized individuals.

13

3.2 Associations between IDI and socioeconomic variables

The relationship among the evenness indexes that measure segregation and other socioeconomic indicators are widely analyzed, especially in developed countries (Darroch, 1971; White, 1983; Reardon & Bischoff, 2011; Musterd, *et al.* 2017). However, in developing economies, data on income, race, education is more difficult to find at the census tract level which is needed to estimate residential segregation indicators.

Figure 5 shows the correlations between the variables. The Gini, poverty rate and unemployment have the greatest correlations with the dissimilarity index. The population, GDP *per capita,* and the SEI have lower correlations.

Figure 5. Correlation matrix among the indices

Source: Research results.

Table 4 shows linear regression coefficients reflecting the associations between the dissimilarity index and the set of socioeconomic variables. We can see a strong association between most of the indicators with the IDI. In the first model, only the coefficient for the Gini index is presented. For this indicator, we find that with each change of 1 standard deviation in the Gini index, there is an expected increase of 0.038 (CI 95%: 0.033, 0.043) in the level of segregation.

An important result was found in the association between the dissimilarity index and the social environment, represented by the SEI index in model 2. As expected, the relationship

between residential segregation and the indicator was negative. In other words, for each change in standard deviation in the SEI, there is a -0.036 (CI 95%: -0.056, -0.017) decrease in income segregation. Thus, the social environment proved to be relevant to the decrease in income segregation. Notably, better educational indicators, access to water, sanitation, and housing improvements in neighborhoods far from central metropolitan regions have been previously suggested to reduce segregation (Moody 2001; Watson 2009; Fiel 2015).

Finally, in the last model, we included the entire set of variables and observed that the Gini index continues to have a relevant association with residential income segregation (0.025 CI 95%: 0.018, 0.032) even after other variables are accounted for. The poverty index becomes also significant (0.012 CI95%: 0.000, 0.025). In addition to economic and social issues, inequality poses growing challenges given the urban changes occurring across the planet. Furthermore, the effects of inequalities are directly felt in the life expectancy of the most vulnerable populations in Latin America (Bilal et al., 2019).

We also estimated the models using the IDI corrected with the bootstrap method to provide more robustness to the results. As expected, the results were maintained. Appendix A shows the estimated coefficients with marginal changes, but the effects and significance remained unchanged.

Table 4. Linear regression coefficients for Brazilian cities in 2010

	Model 1 (95% CI)	Model 2 (95% CI)	Model 3 (95% CI)
Gini	0.038 (0.033, 0.043)		0.025 (0.018, 0.032)
SEI		-0.036 (-0.056, -0.017)	0.002 (-0.012, 0.017)
Poverty rate			0.012 (0.000, 0.025)
GPD *per capita*			0.000 (-0.004, 0.004)
Unemployment			0.009 (-0.001, 0.020)

It is important to highlight that segregation is among the biggest urban problems in the context of rapid urbanization in developing economies. Haddad (2020) states that urbanization and urban spatial structures affect poverty and generate inequalities in social and environmental fields, further contributing to the stratification of social groups. Spatial stratification of social groups accentuates inequalities in poorer countries and are an obstacle to the development of societies in the medium and long term (ECLAC, 2019). Understanding how social and urban changes affect population growth, income inequality, and segregation is essential for designing policies to protect the most vulnerable people and establish viable socioeconomic development conditions for everyone.

4. Final remarks

Discussing methods of measuring segregation is not a simple matter. Segregation quantifies how heterogeneous population groups are distributed in the urban space, providing an indicator of how far an income-minority group is from the city's average income. If we take as base small percentage of low-income minority groups or large spatial areas, the dissimilarity index can be affected by deviations from evenness and the indicator could be strongly biased.

We present a comprehensive description of the segregation by income in the Brazilian cities selected by the SALURBAL project as the largest urban agglomerations in the country. According to Massey and Denton (1988), a highly centralized group, spatially concentrated, unevenly distributed, clustered, and minimally exposed to the majority group is said to be "residentially segregated".

To show which cut-off to define low-income household's best deals with urban income segregation within large cities and the pattern of regional inequality, we use statistical analyses of correlation and distribution between the calculated IDI and the respective low-income minority group. Statistics point out that the cut-off of households earning up to 2 minimum wages yields an average proportion in the low-income minority group of 33.8% of households, seeming reasonable to work with this cut-off. Bias levels calculated on IDI using bootstrap methods were not relevant to the study since they did not change the regressions' estimates.

The fact that the Brazilian census tracts are built taking into account only technical criteria means that each sector's delimited area is spatially balanced, leading to each having approximately the same number of households. However, checking the bias for the indicator is important for the advancement of future research and constitutes a new statistical framework to be taken into account when dealing with other types of residential segregation.

As expected, inequality and the socioeconomic environment were strongly associated with segregation. Despite income gains in the first decade of the XXI century, economic growth in Brazil was not reflected in better income distribution (IBGE, 2017). We found that the SEI, based on education, access to water, sanitation, and housing conditions, was strongly associated with the IDI. Therefore, improvements in social areas based on these elements were found to be relevant for reducing segregation. Notably, the challenges related to improving people's lives span a range of factors, and structural issues must be taken into account when dealing with segregation, poverty, and inequality. Long-term urban planning focused on equity is essential to mitigate the severe effects of these factors.

Therefore, we highlight that the study of income residential segregation plays a significant role in understanding the social relationships within large cities. Future research should focus on other determinants, in addition to income. Understanding segregation patterns by gender, race, and education is essential for developing effective social and health policies that focus on the long term. Therefore, building a consolidated research agenda that can provide adequate scientific support to guide policies to reduce segregation is essential.

Declaration of interests

We declare no competing interests.

Ethics Approval statement

The SALURBAL study protocol was approved by the Drexel University Institutional

Review Board with ID #1612005035 and by appropriate site-specific IRBs.

References

Acevedo-Garcia, Dolores et al. 2003. Future Directions in Residential Segregation and Health Research: a multilevel approach. *American Journal of Public Health*. 93, 215-221. DOI: 10.2105/ajph.93.2.215

Akita, Takahiro. 2003. Decomposing Regional Income Inequality in China and Indonesia Using Two-stage Nested Theil Decomposition Method. *Annals of Regional Science*. 37, 55-77. https://doi.org/10.1007/s001680200107

Allen, Rebecca et al. 2015. More Reliable Inference for the Dissimilarity Index of Segregation. *The Econometrics Journal*. 18, 40-66. https://doi.org/10.1111/ectj.12039

Allen, Rebecca; Burgess, Simon and Windmeijer, Frank. 2009. More Reliable Inference for Segregation Indices. The Centre for Market and Public Organization. *Working Paper, nº 09/216*.

Bailey, Stanley R.; Loveman, Mara & Muniz, Jeronimo. 2013. Measures of "Race" and the analysis of Racial Inequality in Brazil. *Social Science Research*. 42, 106-119. https://doi.org/10.1016/j.ssresearch.2012.06.006

Bilal, Usama et al. 2019. Inequalities in Life Expectancy in Six Latin American Cities From the SALURBAL Study: an Ecological Analysis. *Lancet Planet Health*. 3, e503-10. https://doi.org/10.1016/ S2542-5196(19)30235-9

Bilal, Usama et al. 2021. Life expectancy and mortality in 363 cities of Latin America. *Nature Medicine*, 27, 463-470. https://doi.org/10.1038/s41591-020-01214-4

Bolay, Jean-Claude. 2020. An Intermediate City in Brazil: Between Inequalities and Growth". In: Urban Planning Against Poverty. *Future City*, 14. 121-166. Springer, Cham. https://doi.org/10.1007/978-3-030-28419-0_5

Carrington, William J. and Troske, Kenneth R. 1997. On Measuring Segregation in Samples with Small Units. *Journal of Business & Economic Statistics*. 15, 402-409. https://doi.org/10.1080/07350015.1997.10524718

Comisión Econnómica para América Latina y Caribe. 2017. *Panorama Multidimensional del Desarrollo Urbano en América Latina y el Caribe*. Editorial: CEPAL, 113 p.

Comisión Econnómica para América Latina y Caribe. 2019. Panorama de América Latina.

Darroch, A. Gordon ¨& Marston, Wilfred G. 1971. The Social Class Basis of the Ethinc Residential Segregation: The Canadian Case. *American Journal of Sociology,* 77, 491-510. https://doi.org/10.1086/225160

Duncan, Otis Dudley & Duncan, Beverly. 1955. A Methodological Analysis of Segregation Indexes. *American Sociological Review*. 20, 210-217. https://www.jstor.org/stable/2088328

Efron, Bradley. 1979. Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*. 7, 1-26.

Elbers, Chris et al. 2003. Are Neighbours Equal? Estimating Local Inequality in Three Developing Countries. *World Institute for Development Economics Research*. Discussion Paper nº 2003/52.

Ferreira, Francisco H. G. & Ravallion, Martin. 2008. Global poverty and inequality: A review of the evidence. *World Bank Policy Research Working Paper 4623*.

Fiel, Jeremy. 2015. Closing Ranks: Closure, Status Competition, and School Segregation. *American Journal of Sociology*. 121, 126-170. https://doi.org/10.1086/682027

Fosset, Mark. 2017. New Methods for Measuring and Analyzing Segregation. *The Springer Series on Demographic Methods and Population Analysis*. DOI: 10.1007/978-3-319-41304-4. DOI: 10.1007/978-3-319-41304-4

Garcia-López, Miquel-Àngel & Moreno-Monroy, Ana I. 2018. Income Segregation in Monocentric and Polycentric cities: does urban form really metter? *Regional Science and Urban Economics*. 71, 62-79. https://doi.org/10.1016/j.regsciurbeco.2018.05.003

Haddad, Mônica A. 2020. Residential Income Segregation and Commuting in a Latin American City. *Applied Geography*. 117, 1-11. https://doi.org/10.1016/j.apgeog.2020.102186

Henderson, Vernon. 1997. Medium size cities. *Regional Science and Urban Economics*, 27, 583-612. https://doi.org/10.1016/S0166-0462(96)02169-2

Iceland, John, Weinberg, Daniel H., & Steinmetz, Erika. 2002. U.S. *Census Bureau, Series CENSR-Racial and Ethnic Residential Segregation in the United States: 1980–2000.* Washington, DC:U.S. Government Printing Office.

Instituto Brasileiro de Geografia e Estatística. 2017. *Síntese de Indicadores Sociais: uma análise das condições de vida da população brasileira*. Available in: < https://biblioteca.ibge.gov.br/>.

Instituto Brasileiro de Geografia e Estatística. 2020. *Censo Demográfico Brasileiro 2010*. Available in: <https://censo2010.ibge.gov.br/>.

Massey, Douglas S. & Denton, Nancy A. 1988. Suburbanization and Segregation in U.S. Metropolitan Areas. *American Journal of Sociology*. 94, 592-626. https://doi.org/10.1086/229031

Mata, Daniel et al. 2005. Examining the Growth Patterns of Brazilian Cities. *World Bank Policy Research Working Paper 3724*, 1-39.

Mazza, Angelo & Punzo, Antonio. 2015. On the Upward Bias of the Dissimilarity Index and Its Corrections. *Sociological Methods & Research*. 44, 80-107. DOI: 10.1177/0049124114543242

Mazza, Angelo. 2017. Dealing with Bias of the Dissimilarity Index of Segregation. *Rivista Italiana di Economia Demografia e Statistica*. LXXI, 1-10.

Miranti, Riyana et al. 2015. Measuring Small Area Inequality Using Spatial Microsimulation: Lessons Learned from Australia. *International Journal of Microsimulation*. 8, 152-175. DOI: 10.34196/IJM.00118

Moody, James. 2001. Race, School Integration, and Friendship Segregation in America. *American Journal of Sociology*. 107, 679-716. http://www.jstor.org/stable/3081149

Quistberg, D. Alex et al. 2019. Building a Data Platform for Cross-Country Urban Health Studies: the SALURBAL study. *Journal Urban Health*. 96, 311-337. https://doi.org/10.1007/s11524-018-00326-0

Rao, C. R.; Toutenburg, H. 1995. *Linear Models. In: Linear Models*. Springer Series in Statistics. Springer, New York, NY.

Reardon, Sean F. & Bischoff, Kendra. 2011. Income Inequality and Income Segregation. *American Journal of Sociology*. 116, nº 4, 1092-1153. https://doi.org/10.1086/657114

Reis, Eustáquio. 2014. Spatial Income Inequality in Brazil, 1872-2000. *Economia*. 15, 119-140. https://doi.org/10.1016/j.econ.2014.06.006

Roux, Ana V. Diez et al. 2019. A Novel International Partnership for Actionable Evidence on Urban Health in Latin America: LAC-Urban Health and SALURBAL. *Global Challenges*, 3, 1-11. https://doi.org/10.1002/gch2.201800013

Santos, Maria I. et al. 2021. Urban income segregation and homicides: an analysis using Brazilian cities selected by the Salurbal Project. *SSM-Population Health*, 14, 1-9. https://doi.org/10.1016/j.ssmph.2021.100819

Schelling, Thomas C. 1971. Dynamics models of segregation. *Journal of Mathematical Sociology*. 1, 143-186. https://doi.org/10.1080/0022250X.1971.9989794

Tarozzi, Alessandro & Deaton, Angus. 2009. Using Census and Survey Data to Estimate Poverty and Inequality for Small Areas. *The Review of Economics and Statistics*. 91, 773-779. DOI: 10.2139/ssrn.997829

Tivadar, Mihai. 2019. OasisR: An R Package to Bring Some Order to the World of Segregation Measurement. *Journal of Statistical Software*. 89, 1-39. DOI: 10.18637/jss.v089.i07

Torres, Haroldo G. 2006. Residential segregation and public policies: Sao Paulo in the 1990's. *Rev. Bras. Ciencias Sociais*, 2, 1-14.

Trendle, Bernard. 2005. Sources of Regional Income Inequality: An Examination of Small Regions in Queensland. *The Applied Regional Science Conference*. 17, 35-50. DOI: 10.1111/j.1467-940X.2005.00095.x

United Nations. 2020. *Inequality in a Rapidly Changing World*. World Social Report 2020.

Watson, Tara. 2009. Inequality and the Measurement of Residential Segregation by Income in American Neighborhoods. *The Review of Income and Wealth*. 55, 820-844. https://doi.org/10.1111/j.1475-4991.2009.00346.x

White, Michael J. 1983. The Measurement of Spatial Segregation. *American Journal of Sociology*. 88, 1008-1018. https://doi.org/10.1086/227768

Wooldridge, Jeffrey M. *Introductory Econometrics: a modern approach*. Cengage Learning; 7th edition, 816 p.

Appendix A. Results of the sensitivity analysis with the IDI corrected by bootstrap

	Model 1 (95% CI)	Model 2 (95% CI)	Model 3 (95% CI)
Gini	0.037 (0.031, 0.043)		0.026 (0.018, 0.033)
SEI		-0.039 (-0.060, -0.019)	0.007 (-0.009, 0.024)
Poverty rate			0.017 (0.003, 0.031)
GPD *per capita*			0.000 (-0.005, 0.005)
Unemployment			0.008 (-0.002, 0.020)
Population (*log*)			0.000 (-0.005, 0.007)