

Residential Segregation in Brazil: An Inference Framework Based study for Selected Cities

Renan Xavier Cortes
Department of Statistics
Federal University of Rio Grande do Sul
Porto Alegre, Brazil
renan.cortes@ufrgs.br

February 17, 2026

Abstract

This is a simple abstract. It briefly explains the objective, methods, main results, and conclusions of the study.

Keywords: Keyword1; Keyword2; Keyword3

1 Introduction

The measurement of residential segregation has been a central concern in urban sociology and demography for nearly a century, tracing its intellectual lineage to the pioneering work of the Chicago School and the quantitative innovations of the mid-twentieth century. Despite this long history, the field has experienced persistent debates about how best to capture the multifaceted nature of spatial separation between social groups. At its core, segregation measurement grapples with a fundamental challenge: how to reduce complex spatial patterns of population distribution into summary statistics that can meaningfully inform our understanding of social inequality, urban structure, and demographic change.

The canonical contribution to this literature remains Massey and Denton [1988] seminal work, which brought conceptual clarity to a field that had become fragmented by competing definitions and measures. In their influential paper, they argued persuasively that residential segregation is not a unidimensional phenomenon but rather encompasses five conceptually distinct dimensions of spatial variation: evenness, exposure, concentration, centralization, and clustering. *Evenness* refers to the differential distribution of groups across areal units, with the dissimilarity index (D) serving as the most widely used measure. *Exposure* captures the degree of potential contact between groups, typically measured through interaction and isolation indices (P^*). *Concentration* reflects the relative amount of physical space occupied by a group, while *centralization* describes the extent to which a group resides near the urban core. Finally, *clustering* measures the degree to which areas inhabited by minority members adjoin one another in space. Through a comprehensive factor analysis of twenty segregation indices across sixty U.S. metropolitan areas, Massey and Denton [1988] empirically validated this five-dimensional

structure, demonstrating that while the dimensions are correlated in practice, they capture distinct aspects of spatial organization that cannot be reduced to a single measure.

Despite the conceptual advances of Massey and Denton's framework, a critical limitation persisted throughout the segregation literature well into the 1990s and beyond: the almost complete absence of statistical inference. Researchers routinely calculated segregation indices as descriptive statistics, treating observed values as fixed population parameters rather than estimates subject to sampling variation. This practice implicitly assumed that the data represented the complete enumeration of a population, ignoring the stochastic nature of the allocation processes that generate observed spatial patterns. As Ransom [2000] forcefully articulated, even when working with census data that purportedly represent the entire population, there are compelling reasons to adopt a statistical perspective. Individuals are allocated to residential locations through processes that contain random elements, and the observed distribution on any given census day represents but one realization of a stochastic allocation process. Moreover, census data themselves are subject to measurement error and temporal sampling variation, as populations are in constant flux due to migration and mobility.

Ransom [2000] provided a rigorous foundation for statistical inference in segregation research by developing the asymptotic sampling distributions of the dissimilarity index (D) and the Gini index (G) under a multinomial sampling model. In this framework, the observed counts of group members across spatial units are treated as draws from a multinomial distribution with fixed underlying probabilities. Ransom derived the asymptotic normality of D and G under conditions where the group proportions differ across units, providing a computationally tractable method for constructing confidence intervals and hypothesis tests. The delta method was employed to obtain variance estimators, enabling researchers to test whether observed differences in segregation across time or place were statistically significant. Ransom's Monte Carlo simulations demonstrated that tests based on these asymptotic distributions maintained appropriate significance levels with moderately large samples, offering a practical tool for applied researchers.

However, the multinomial model proposed by Ransom [2000] rests on assumptions that may be problematic in many empirical contexts. Specifically, it assumes independence across spatial units and treats the population counts as fixed at the time of measurement. These assumptions become particularly troublesome when unit sizes are small, a common situation in studies of school segregation, occupational segregation, or residential segregation in sparsely populated areas. When unit sizes are small, the dissimilarity index exhibits substantial upward bias, as even random allocation produces nonzero index values due to sampling variation. This insight, while noted by earlier researchers including Cortese et al. [1976], was systematically addressed by Carrington and Troske [1997], who proposed adjustments to segregation indices that account for the expected value under random allocation.

Allen et al. [2015] advanced this line of inquiry by developing a comprehensive inferential framework for the dissimilarity index that explicitly addresses the bias induced by small unit sizes and small minority proportions. Their approach conceptualizes segregation as the outcome of a stochastic allocation process governed by conditional probabilities that may differ between groups. In this framework, systematic segregation exists precisely when these conditional probabilities differ across groups. Allen et al. [2015] derived a likelihood ratio test for the presence of any systematic segregation, as well as bias-adjusted estimators based on bootstrap methods and asymptotic normal approximations. Their density-corrected estimator (D_{DC}) substantially reduces the upward bias of the dissimilarity index.

ity index when unit sizes are small, while their Monte Carlo simulations demonstrate that bootstrap-based inference procedures outperform asymptotic approximations in terms of size and power properties. The authors further developed tests for comparing segregation across areas or over time, providing a coherent framework for addressing the comparative questions that have long motivated segregation research.

A parallel development in the statistical literature on segregation has been the incorporation of spatial autocorrelation into inferential frameworks. Lee et al. [2015] made a crucial observation: the areal units used to compute segregation indices exhibit strong spatial patterning, with neighboring units typically displaying similar proportions of minority populations. This spatial autocorrelation violates the independence assumptions underlying conventional inference procedures and, if ignored, leads to misleading uncertainty quantification. citelee2015bayesianinference proposed a Bayesian hierarchical modeling approach that explicitly accounts for spatial dependence through conditional autoregressive (CAR) priors. Their framework estimates the underlying probability surface across areal units, borrowing strength from neighboring observations to improve estimation precision, particularly when unit populations are small. The posterior predictive distribution of the dissimilarity index then provides point estimates and credible intervals that appropriately reflect both sampling variation and spatial structure. Through an empirical application to religious segregation in Northern Ireland, Lee et al. [2015] demonstrated that conventional bootstrap confidence intervals are excessively wide and fail to achieve nominal coverage, whereas the Bayesian spatial model yields intervals with appropriate properties.

More recently, Rey et al. [2021] have addressed perhaps the most challenging question in comparative segregation analysis: when we observe a difference in segregation between two cities or time periods, to what extent is this difference attributable to variations in population composition versus differences in spatial structure? This decomposition question has profound implications for both theoretical understanding and policy design, as interventions targeting segregation must be informed by its underlying drivers. Rey et al. [2021] developed a novel framework combining counterfactual distributions with Shapley decomposition to partition observed differences in segregation indices into spatial and compositional components. The approach generates counterfactual populations by imposing the tract-level composition distribution of one city onto the spatial structure of another, then computes segregation indices for these counterfactual scenarios. Through an empirical analysis of 50 U.S. metropolitan areas, they demonstrated that for evenness and isolation measures, differences between cities are typically dominated by compositional variation, while for clustering, concentration, and centralization, spatial structure plays a more significant role. This decomposition framework represents a significant advance in enabling researchers to move beyond simple comparisons of segregation levels toward understanding the mechanisms generating observed differences.

Building on these methodological developments, Cortes et al. [2020] introduced the PySAL segregation module, an open-source Python package that implements a comprehensive suite of spatial and aspatial segregation measures within a unified computational framework. The module provides point estimation for 25 distinct segregation indices spanning all five dimensions identified by Massey and Denton, including bias-corrected versions of the dissimilarity index following Allen et al. [2015]. Critically, the PySAL segregation module incorporates inferential functionality that has been largely absent from previous software implementations. For single-value inference, the module offers multiple approaches to generating distributions under the null hypothesis, including systematic

allocation (assuming equal conditional probabilities across groups), evenness (assuming independent binomial distributions with constant global probability), spatial permutation (randomizing units across space while preserving marginal totals), and combinations thereof. For comparative inference, the module implements two approaches: random labeling, which randomly reassigned observations between comparison groups, and counterfactual composition, which swaps composition distributions between areas while preserving spatial structure.

However, the inferential framework implemented in the PySAL segregation module has attracted important methodological critiques that warrant careful consideration. A reviewer of the Cortes et al. [2020] manuscript raised fundamental questions about the conceptual foundations of simulation-based inference for segregation measures. The central critique concerns the definition of the null hypothesis: the module offers multiple options (systematic, evenness, permutation, and their combinations) without providing clear guidance about which null is appropriate for which research question. This ambiguity is not merely a technical oversight but reflects deeper conceptual issues about what constitutes "no segregation" in different contexts. For evenness measures like the dissimilarity index, the condition of no segregation is conventionally defined as equal distributions of groups across all units, yielding an index value of zero. Yet if one conceptualizes no segregation as random allocation, the expected value of D under the null is not zero but some positive quantity that depends on unit sizes and minority proportions. The PySAL module's "systematic" approach generates simulations under the assumption of equal conditional probabilities, which corresponds to the former conceptualization, while the "evenness" approach generates independent binomial draws with constant global probability, which approximates the latter. These different null hypotheses can lead to markedly different inferences, particularly when unit sizes are small or minority proportions are extreme.

The reviewer further noted that the randomization framework underlying the PySAL inference procedures may generate unrealistic spatial distributions by ignoring constraints on unit population sizes. When simulating under the null, the module expands the sampling space to include configurations that could not plausibly occur in real populations, potentially leading to incorrect inferences. This critique echoes longstanding debates in the spatial analysis literature about the appropriate reference distribution for tests of spatial pattern. Moreover, the review pointed to recent developments in empirical Bayes approaches (Lee et al. [2015]) that offer alternative frameworks for uncertainty quantification, noting that the PySAL module's simulation-based approach has not been systematically compared to these Bayesian alternatives. These limitations highlight that while the PySAL segregation module represents a significant step forward in making inferential tools accessible to applied researchers, the conceptual foundations of segregation inference remain an active area of methodological development, and users must exercise careful judgment in selecting and interpreting null hypotheses.

The present chapter contributes to this evolving methodological literature by applying these inferential frameworks to the measurement of segregation in Brazil. The Brazilian context presents unique challenges and opportunities for segregation research. Brazilian cities are characterized by extreme socioeconomic inequality, rapid urbanization, and complex patterns of racial and ethnic diversity that differ substantially from the North American contexts that have dominated methodological development. Brazilian census data are available at fine spatial scales but exhibit the small unit sizes and variable minority proportions that make bias correction essential. Moreover, Brazilian metropolitan areas display diverse spatial structures, from the dense, consolidated urban fabric of São

Paulo to the more dispersed patterns of newer cities, providing fertile ground for decomposition analyses that disentangle compositional and spatial contributions to observed segregation differences. By applying the methods developed by Allen et al. [2015], Lee et al. [2015], Rey et al. [2021], and implemented in the PySAL segregation module (Cortes et al. [2020]), this chapter advances our understanding of Brazilian segregation patterns while contributing to the broader project of developing statistically rigorous approaches to segregation measurement in diverse international contexts. In doing so, we remain mindful of the caveats identified in methodological critiques, and we interpret our results with appropriate caution regarding the sensitivity of inferences to the choice of null hypothesis and simulation framework.

2 Related work in Brazil

Residential racial segregation has long occupied a central position in international debates on urban inequality, yet its treatment in Brazil has historically been shaped by the country's particular racial formation and the enduring myth of "racial democracy." While the absence of formal segregationist laws distinguished Brazil from contexts such as the United States, a growing body of demographic and urban research demonstrates that racialized spatial inequalities have been persistent and structurally embedded since abolition in 1888. Contemporary scholarship has decisively challenged narratives of harmonious race relations, revealing instead patterned spatial separations between white, black (*preto*), and brown (*pardo*) populations across Brazilian cities.

Classic analyses by Telles [2004], particularly in *Race in Another America*, were foundational in demonstrating that racial and income segregation in Brazil, although lower than in many U.S. metropolitan areas, was nonetheless substantial and socially consequential. Using census data from the 1980s onward, Telles identified systematic spatial distances between white and black populations, followed by brown and black groups, revealing how racial hierarchies were inscribed in metropolitan space. Importantly, he highlighted how the absence of explicit segregationist legislation did not prevent the consolidation of racially stratified urban patterns.

Recent nationwide evidence has expanded and refined this perspective. Sousa Filho et al. [2023] provide one of the most comprehensive demographic assessments of racial and economic segregation using 2010 census tract data. Applying the dissimilarity index, the authors demonstrate that racial segregation is particularly pronounced in cities in the South and Southeast regions and disproportionately affects the self-declared black population. Notably, they show that income plays a mediating but not exhaustive role: while per capita household income is strongly associated with segregation among the poorest families—especially those identifying as black—racial segregation cannot be reduced solely to economic stratification. Their findings reinforce the need to disentangle race and class dimensions within spatial-demographic analyses.

City-level studies further nuance the Brazilian panorama. Barros and Feitosa [2018] examine the uneven geography of racial and socioeconomic segregation in Brazil, emphasizing how spatial inequality is structured across metropolitan areas. Specifically, in the Metropolitan Region of São Paulo, Barros and Feitosa [2024] suggests that São Paulo exhibits segregation patterns comparable to, and in some dimensions exceeding, those of London. These comparative analyses contribute to repositioning Brazilian metropolises within global debates on urban polarization, gated communities, and fragmented urbanism. The proliferation of enclosed condominiums and high-income enclaves reflects

broader Latin American trends of socio-spatial distancing, intensifying both economic and racial separation.

França [2022] reveals how coastal valorização imobiliária and uneven urban expansion have reinforced racialized and income-based spatial divisions. Similarly, longitudinal analyses of the Metropolitan Region of Belo Horizonte (2010–2022), as discussed in Gonçalves and Mercedes Strauch [2026], indicate the persistence—and in some cases, reconfiguration—of segregation patterns over the last decade, despite social policies aimed at poverty reduction and housing expansion. These temporal perspectives are crucial for understanding how macroeconomic shifts, federal housing programs, and demographic transitions interact with entrenched racial hierarchies.

Beyond its spatial and economic implications, residential segregation has increasingly been examined through a public health lens. Building on international scholarship linking segregation to differential exposure to environmental risks, violence, and limited access to health services, Brazilian research has begun to articulate how racialized spatial concentration shapes health trajectories. The study presented in Barber et al. [2018], drawing on data from the Brazilian Longitudinal Study of Adult Health (ELSA-Brasil), demonstrates associations between segregation measures and health outcomes, suggesting that spatial isolation may influence cardiovascular risk, mental health, and other chronic conditions. These findings resonate with frameworks developed by scholars such as Massey and Denton [1988], who conceptualized segregation as a “structural linchpin” of racial inequality affecting multiple life domains, including health.

In the Brazilian context, the intersection between segregation and public health is further mediated by the structure of the Unified Health System (SUS) and by persistent inequalities in access to preventive and curative services. Spatial concentration of black and brown populations in peripheral neighborhoods often coincides with precarious sanitation, longer commuting times, environmental hazards, and reduced healthcare accessibility, reinforcing cumulative disadvantage across the life course. Thus, segregation emerges not merely as a demographic pattern but as a determinant of social vulnerability and health inequity.

From a methodological standpoint, the Brazilian literature has predominantly relied on classical indices—such as the dissimilarity and isolation indices—calculated at the census tract level. While these measures remain central for comparability, recent advances in Spatial Data Science open new analytical possibilities. High-resolution geocoded census data, spatial autocorrelation statistics, local indicators of spatial association (LISA), and multiscale segregation metrics enable more refined assessments of how racial and income groups cluster, interact, and experience urban space. Moreover, integrating demographic microdata with administrative, environmental, and health datasets—an approach exemplified by data infrastructures such as CIDACS/Fiocruz—allows for multidimensional analyses linking segregation to socioeconomic and epidemiological outcomes.

In sum, contemporary research demonstrates that residential racial segregation in Brazil is neither negligible nor reducible to income inequality alone. It is spatially heterogeneous across regions, historically rooted, and dynamically reshaped by urban development, housing markets, and public policy. For black and brown populations, segregation structures differential access to opportunities, services, and well-being. Anchoring demographic data within spatial analytical frameworks is therefore essential to advancing both scholarly understanding and evidence-based policymaking. This volume situates itself within this evolving field, mobilizing tools from Spatial Data Science to deepen the empirical and conceptual examination of racialized urban inequality in Brazil.

The Brazilian literature on residential racial segregation has been methodologically consistent, but also relatively narrow in scope. Most empirical studies—whether national or metropolitan—rely predominantly on the Dissimilarity Index (D) as the principal metric of segregation. This emphasis reflects a strong focus on the evenness dimension, measuring how evenly black, brown, and white populations are distributed across census tracts. While the dissimilarity index ensures comparability and interpretability, its dominance constrains analytical depth: other dimensions of segregation—such as exposure, clustering, centralization, and spatial interaction—are rarely examined. As a result, segregation is frequently reduced to a single summary statistic derived from administratively defined units, with limited attention to spatial dependence or multiscalar processes. Equally important is the relative absence of robust inferential frameworks. Much of the literature remains descriptive, comparing index values across cities or over time without formally modeling the mechanisms that produce segregation.

3 Framework

Here, I discuss the hypotheses present the null hypothesis and alternative hypothesis and limitations of the approach recovering the Reviewer comments.

Discuss the dimensions of Massey...

Queen matrix

Why we chose those cities

Segregation measures chosen

4 Results

Hardware,

4.1 Point Estimation of segregation measures

Here I put the maps Here I put all point measurements and discuss

4.2 Inference results

Computational challenges

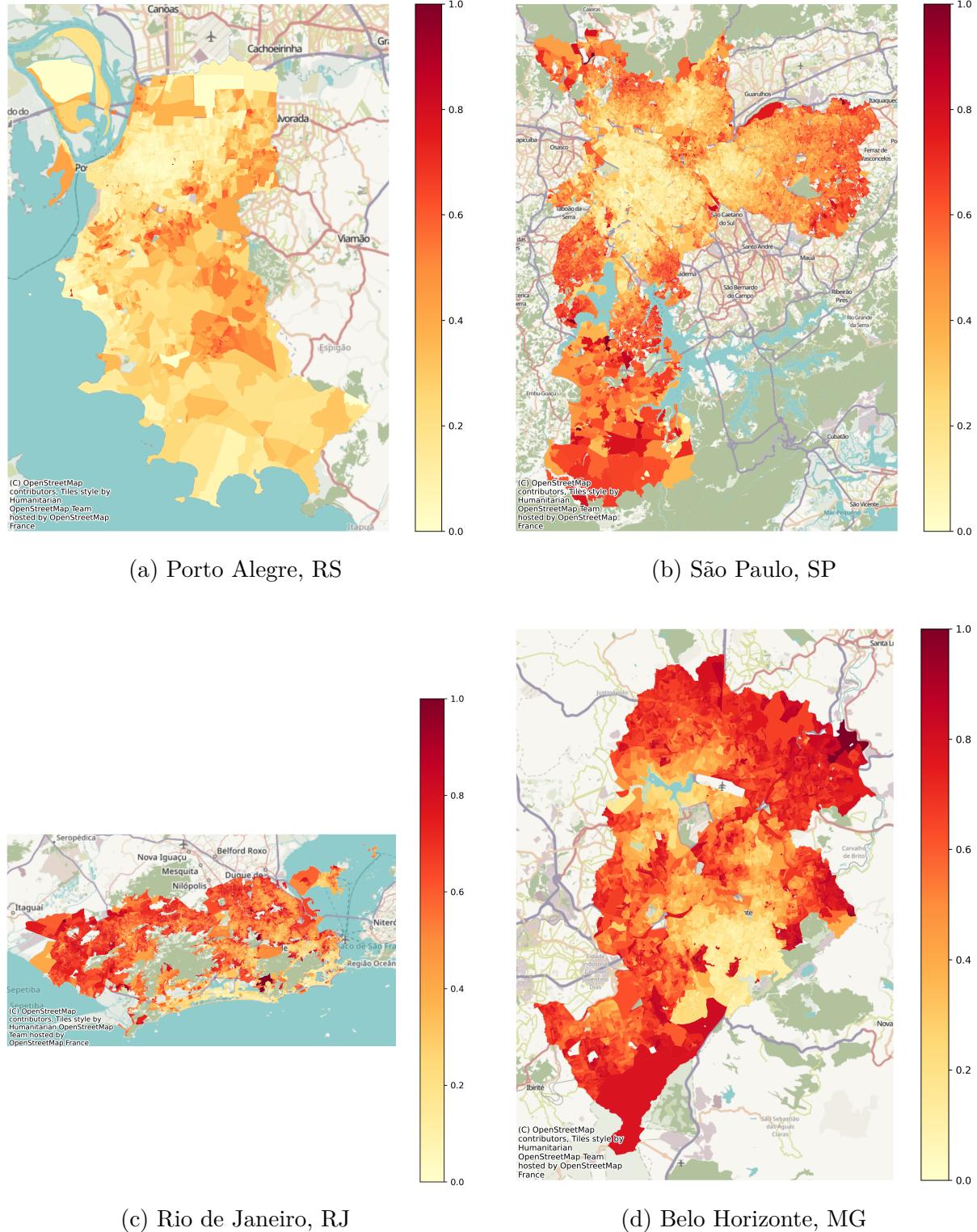


Figure 1: Composition of all selected cities

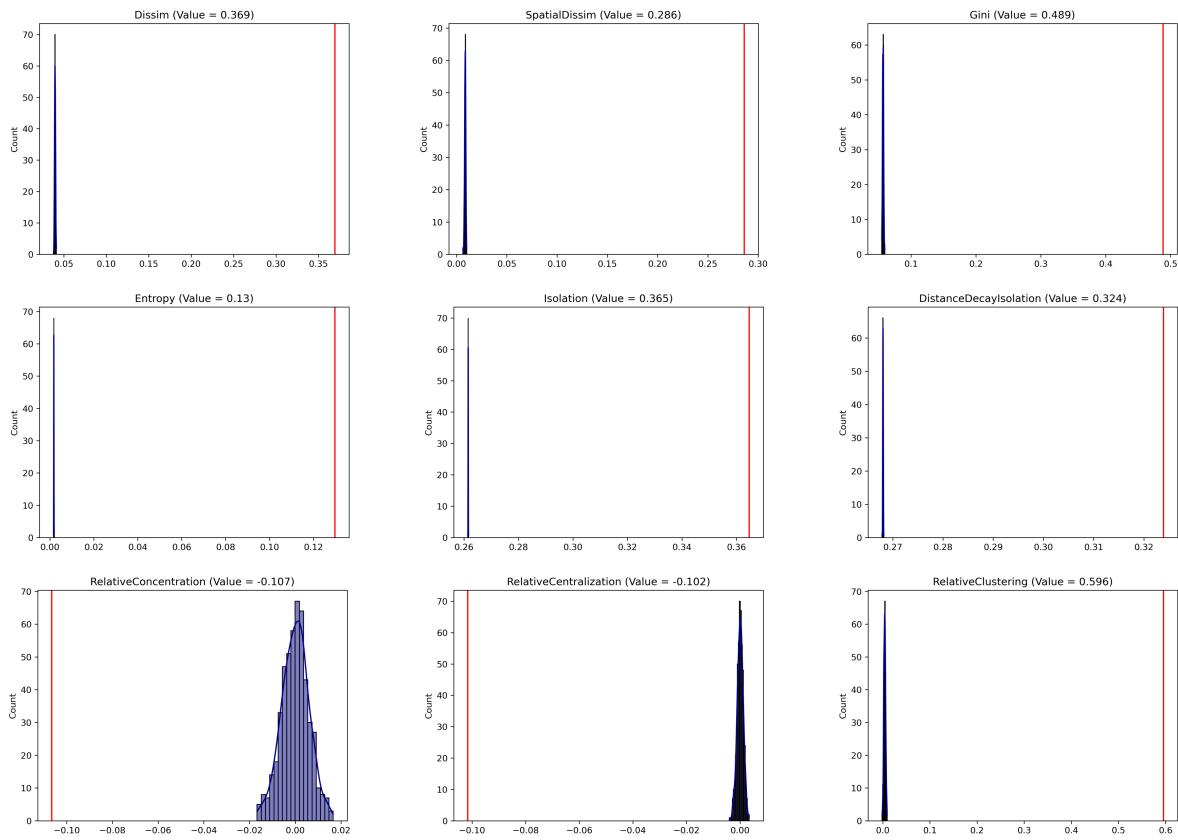


Figure 2: Point Estimation with systematic approach

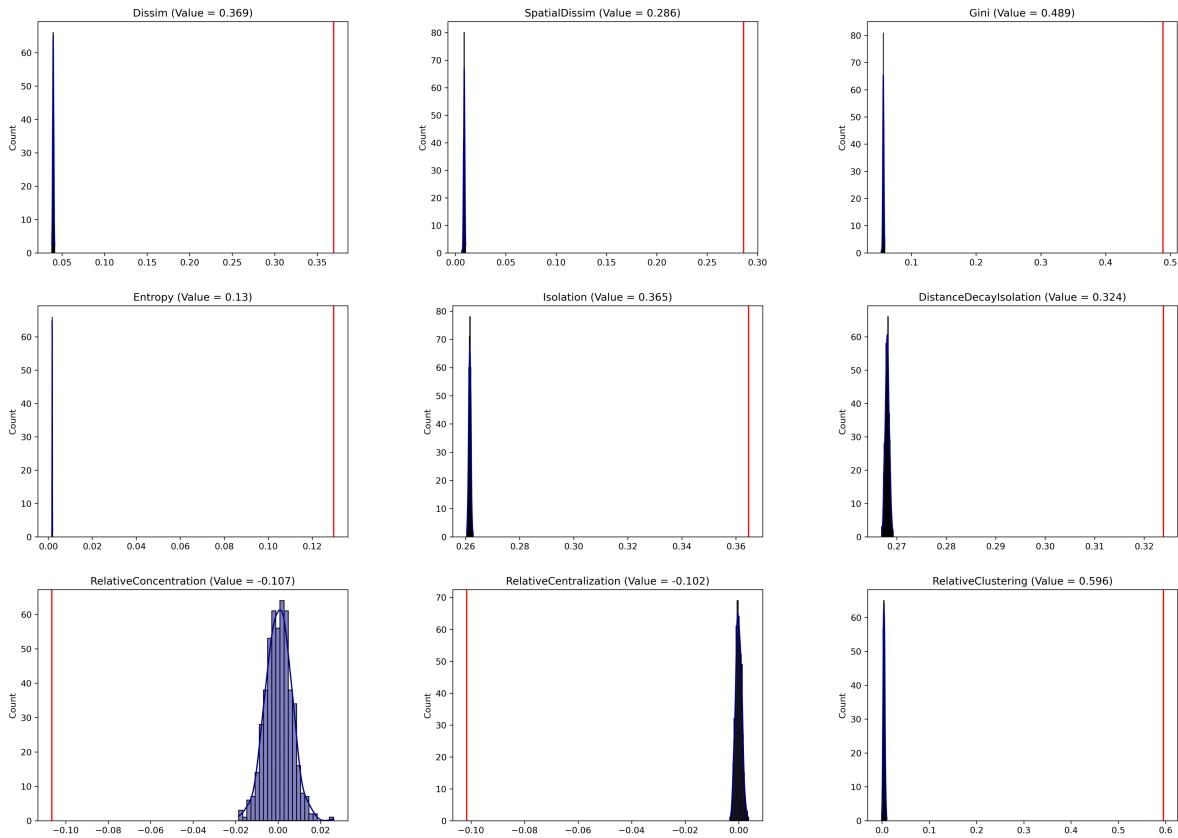


Figure 3: Point Estimation with evenness approach

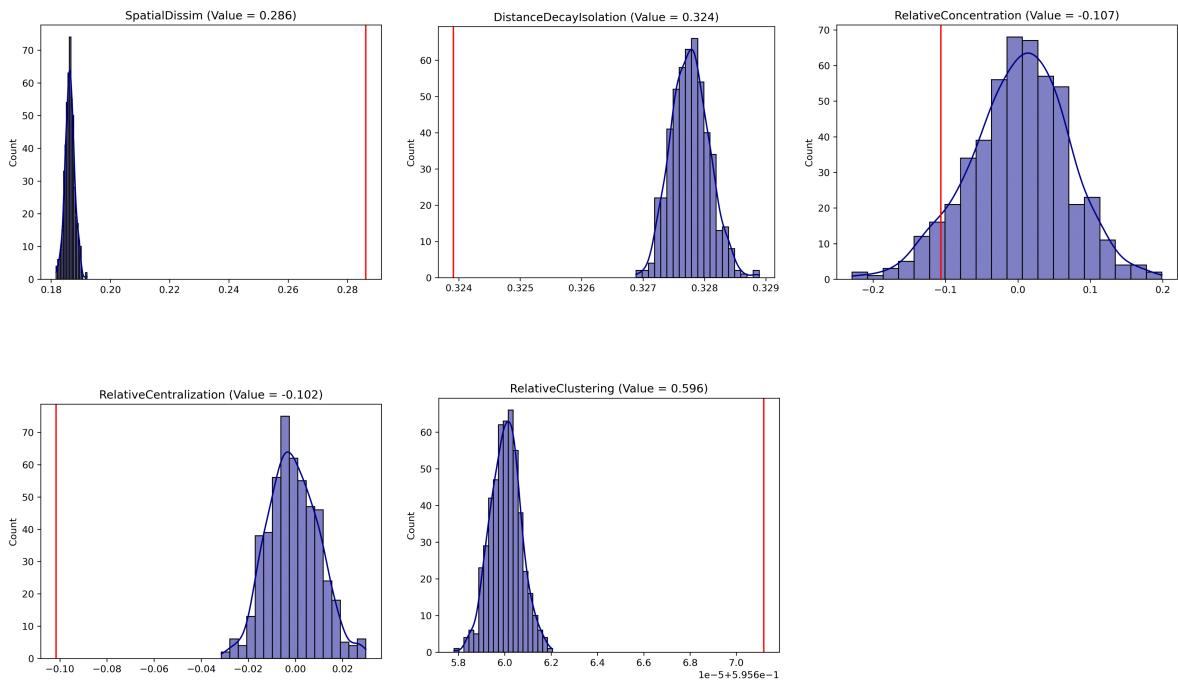


Figure 4: Point Estimation with Geographic Permutation approach

5 Discussion and Future Work

More cities Comparative framework detangling with shapley Street network based.

References

- Rebecca Allen, Simon Burgess, Russell Davidson, and Frank Windmeijer. More reliable inference for the dissimilarity index of segregation. *The econometrics journal*, 18(1):40–66, 2015.
- Sharrelle Barber, Ana V. Diez Roux, Letícia Cardoso, Simone Santos, Verônica Toste, Sherman James, Sandhi Barreto, Maria Schmidt, Luana Giatti, and Dora Chor. At the intersection of place, race, and health in brazil: Residential segregation and cardio-metabolic risk factors in the brazilian longitudinal study of adult health (elsa-brasil). *Social Science & Medicine*, 199:67–76, 2018. ISSN 0277-9536. doi: <https://doi.org/10.1016/j.socscimed.2017.05.047>. URL <https://www.sciencedirect.com/science/article/pii/S0277953617303489>. The role of Racism in Health Inequalities: Integrating Approaches from Across Disciplines.
- Joana Barros and Flavia da Fonseca Feitosa. A segregação étnico-racial nas regiões metropolitanas de são paulo e londres: reflexões sobre um estudo comparativo. *Revista Brasileira de Estudos de População*, 41:1–28, 8 2024. doi: 10.20947/S0102-3098a0262. URL <https://rebep.org.br/revista/article/view/2362>.
- Joana Barros and Flavia F Feitosa. Uneven geographies: Exploring the sensitivity of spatial indices of residential segregation. *Environment and Planning B: Urban Analytics and City Science*, 45(6):1073–1089, 2018. doi: 10.1177/2399808318760572. URL <https://doi.org/10.1177/2399808318760572>.
- William J Carrington and Kenneth R Troske. On measuring segregation in samples with small units. *Journal of Business & Economic Statistics*, 15(4):402–409, 1997.
- Renan Xavier Cortes, Sergio Rey, Elijah Knaap, and Levi John Wolf. An open-source framework for non-spatial and spatial segregation measures: the pysal segregation module. *Journal of Computational Social Science*, 3(1):135–166, 2020. doi: 10.1007/s42001-019-00059-3.
- Charles F Cortese, R Frank Falk, and Jack K Cohen. Further considerations on the methodological analysis of segregation indices. *American sociological review*, pages 630–637, 1976.
- Danilo França. Segregação residencial por raça e classe em fortaleza, salvador e são paulo. *Caderno CRH*, 35:e022045, 12 2022. doi: 10.9771/ccrh.v35i0.42018. URL <https://periodicos.ufba.br/index.php/crh/article/view/42018>.
- Ana Paula Vasconcelos Gonçalves and Julia Celia Mercedes Strauch. Segregação racial em belo horizonte: o que mudou em mais de uma década? *Cadernos Metrópole*, 28(65), 2 2026. URL <https://revistas.pucsp.br/index.php/metropole/article/view/71128>.

Duncan Lee, Jon Minton, and Gwilym Pryce. Bayesian inference for the dissimilarity index in the presence of spatial autocorrelation. *Spatial Statistics*, 11:81–95, 2015. ISSN 2211-6753. doi: <https://doi.org/10.1016/j.spasta.2014.12.001>. URL <https://www.sciencedirect.com/science/article/pii/S2211675314000633>.

Douglas S Massey and Nancy A Denton. The dimensions of residential segregation. *Social forces*, 67(2):281–315, 1988.

Michael R Ransom. Sampling distributions of segregation indexes. *Sociological Methods & Research*, 28(4):454–475, 2000.

Sergio Rey, Renan Cortes, and Elijah Knaap. Comparative spatial segregation analytics. *Spatial Demography*, 9:1–26, 01 2021. doi: 10.1007/s40980-021-00075-w.

José Firmino de Sousa Filho, Sara Costa Pedeira, Gervásio F. dos Santos, Joanna M. N. Guimarães, Andréa J. F. Ferreira, Flávia Jôse O. Alves, Gabrielle R. Nascimento, Aureliano S. S. Paiva, Roberto F. S. Andrade, Emanuelle F. Góes, Maurício L. Barreto, and Estela M. L. Aquino. Racial and economic segregation in brazil: a nationwide analysis of socioeconomic and socio-spatial inequalities. *Brazilian Journal of Population Studies*, 40:1–24, 8 2023. doi: 10.20947/S0102-3098a0247. URL <https://rebep.org.br/revista/article/view/2166>.

Edward E Telles. *Race in Another America: The Significance of Skin Color in Brazil*. Princeton University Press, Princeton, NJ, 2004. ISBN 978-0-691-11866-6.