An R package for more reliable estimates of Residential Segregation

*Un paccetto R per stime piú attendibili sul livello di segregazione residenziale*

Angelo Mazza and Antonio Punzo

**Abstract** The dissimilarity index is widely used to evaluate residential segregation, although there is a widespread awareness that this index is inherently subject to an upward bias and, under certain conditions, can be highly misleading. Common strategies used in literature to deal with the index bias rely on the use of informal rules of thumb, which at least have the side effect of restricting the scope of segregation studies.

Bias correction methods have been proposed, but they require the computation of the sampling distribution of the index through computation-intensive techniques and the lack of user-friendly computer programs has affected their adoption. Hence, we introduce an R package that allows for the computation of bias corrections based on bootstrap, iterated bootstrap, grouped jackknife, and on a technique proposed by the authors. Confidence intervals and tests for absence of segregation are implemented.

**Abstract** *L'indice di dissimilarità è ampiamente utilizzato per valutare la segregazione residenziale, sebbene sia noto da tempo che questo tenda a sovrastimare il reale livello di segregazione. Le strategie comunemente seguite in letteratura per limitare la distorsione dell’indice si basano sull'impiego di regole empiriche, malgrado da tempo siano stati proposti metodi formali per la riduzione del bias. Tali metodi, tuttavia, richiedono il calcolo della distribuzione campionaria dell'indice attraverso tecniche computazionali, e la mancanza di programmi informatici user-friendly ne ha ristretto l'adozione. Per ovviare a questa mancanza, noi proponiamo un pacchetto R che implementa diverse tecniche di riduzione del bias dell’indice di dissimilarità, basate su bootstrap, iterated bootstrap, grouped jackknife raggruppato e su una tecnica proposta dagli stessi autori. Vengono implementati anche intervalli di confidenza e test per l'assenza di segregazione.*

**Keywords:** residential segregation, bias correction, dissimilarity index

Introduction

The segregation of demographic groups, often associated with ethnicity, age or gender, is an important area of research among sociologists, demographers, and other social scientists. The evaluation of segregation within a population is typically based on the proportions of demographic groups belonging to some kind of allocation units, such as residential areas, workplaces, or schools (Mazza and Punzo, 2015).

Many segregation indexes have been suggested, with different formulations denoting different definitions of segregation (see Massey and Denton, 1988 for an overview). Among these, the dissimilarity index $D$, proposed by Duncan and Duncan (1955), is widely used to assess the differential distribution of two groups among allocation units.

The observed allocation pattern may be conceived as one of many possible outcomes of a stochastic allocation process. Usually, researchers are interested in understanding the “systematic” characteristics of the allocation process, apart from random fluctuations that may affect a single observed pattern (Mazza and Punzo, 2015). In this view, the observed dissimilarity $\hat{D}$ is merely an estimator of an unknown level of dissimilarity in the population.

A problem with the use of the dissimilarity index is that $\hat{D}$ appears to be an upward biased estimator of $D$. Within a multinomial framework, Allen et al. (2009) demonstrate, using simulations, that random allocation generates substantial unevenness, and hence an upward bias, especially when dealing with small units, a small minority proportion, and a low level of segregation. Accordingly, different correction approaches have been proposed in literature (see, e.g., Allen et al., 2009, and Altavilla, Mazza and Punzo, 2010 for two examples of bootstrap-based bias correction, and Mazza and Punzo 2015 for an analytical computation of bias and a newer bias correction which outperforms previous correction attempts).

Although several bias correction methods have been proposed, they require the computation of the sampling distribution of the index through computation-intensive techniques and the lack of user-friendly computer programs has affected their adoption. We introduce an R package (R Core Team, 2017 ) that allows for the computation of bias corrections based on bootstrap, iterated bootstrap, grouped jackknife, and on the technique recently proposed in Mazza and Punzo (2015). Confidence intervals and tests for absence of segregation are implemented.

Estimators

In this section, we introduce four alternative bias correction techniques.

Bootstrap based estimator

Allen et al. (2009) adopt a bootstrap-based bias correction, based on

|  |  |
| --- | --- |
| $$D-\hat{D}\_{obs}≈\hat{D}\_{obs}-E\left(\hat{D}|\hat{p}\_{1}^{0},…,\hat{p}\_{k}^{0},\hat{p}\_{1}^{1},…,\hat{p}\_{k}^{1},n^{0},n^{1}\right),$$ |  |

where $\hat{D}\_{obs}$ denotes the observed counterpart of $\hat{D}$. The observed conditional probabilities $\hat{p}\_{j}^{0}$ and $\hat{p}\_{j}^{1}$, $j=1,…,k$, are used to generate, by multinomial sampling, $B$ bootstrap allocations with the same group sizes $n^{0}$ and $n^{1}$. Then, a measure of $Bias\left(\hat{D}\right)$ is given by $\overline{D}\_{Boot}-\hat{D}$, and the bootstrap bias corrected estimate of $D$ can be obtained as

|  |  |
| --- | --- |
| $$\hat{D}\_{Boot}=\hat{D}\_{obs}-\left(\overline{D}\_{Boot}-\hat{D}\_{obs}\right)=2\hat{D}\_{obs}-\overline{D}\_{Boot}.$$ |  |

This bias correction would work well if the bias were constant for different values of $D$. This is not the case here, and this bias correction is therefore not expected to “eliminate”, but only to “reduce”, the existing bias. Instead of bootstrapping $E\left(\hat{D}|\hat{p}\_{1}^{0},…,\hat{p}\_{k}^{0},\hat{p}\_{1}^{1},…,\hat{p}\_{k}^{1},n^{0},n^{1}\right)$, Mazza and Punzo (2015) show that this expectation may be computed analytically, using a binomial based formulation for a small number of units with small sizes or with a folded normal approximation when , , is sufficiently large.

Grouped jackknife and iterative bootstrap estimators

Alternative to the bootstrap, a standard practice for bias correction is the Jackknife. Hence, we added to the package a grouped jackknife estimator $\hat{D}\_{JK}$; this estimator was implemented following Efron (1982, Section. 2.2). Finally, a double bootstrap estimator $\hat{D}\_{DB}$, based on the approach documented in Davison and Hinkley (1997, Section. 3.9) has also been implemented.

A recently introduced estimator

Mazza and Punzo (2015) introduce an estimator of $D$ that reduces the bias with respect to $\hat{D}\_{Boot}$. Its rationale consists in choosing a value $\tilde{D}$ which minimizes

$$E\left(\hat{D}|\tilde{p}\_{1}^{0},…,\tilde{p}\_{k}^{0},\tilde{p}\_{1}^{1},…,\tilde{p}\_{k}^{1},n^{0},n^{1}\right)-\hat{D}\_{obs}$$

with $\tilde{D}=\frac{1}{2}\sum\_{j=1}^{k}‍\left|\tilde{p}\_{j}^{1}-\tilde{p}\_{j}^{0}\right|.$ There may be different criteria for choosing $\tilde{D}$. On way is to require the sequence of differences $\left|\tilde{p}\_{j}^{0}-\tilde{p}\_{j}^{1}\right|$ to be a flattened variant of its observed counterpart. Flattening is obtained by spreading the difference $Δ=\hat{D}\_{obs}-\tilde{D}\geq 0$, among the $k$ differences $\left|\tilde{p}\_{j}^{0}-\tilde{p}\_{j}^{1}\right|$, proportionally to the residuals $\hat{d}\_{j}=\left|\hat{p}\_{j}^{0}-\hat{p}\_{j}^{1}\right|$. An optimization procedure, which adopts a combination of golden section search and successive parabolic interpolation is described in Mazza and Punzo (2015).

Conclusions

It has long been recognized that the sensitivity of the dissimilarity index of Duncan and Duncan (1955) to random allocation implies an upward bias, particularly evident with smaller unit sizes, small minority proportions and lower levels of segregation. Although several bias correction methods have been proposed, they typically require computation-intensive techniques and the lack of user-friendly computer programs has limited their adoption. Hence, we introduce an R package that allows for the computation of bias corrections based on bootstrap, iterated bootstrap, grouped jackknife, and on the technique proposed in Mazza and Punzo (2015).

References

1. Allen, R., S. Burgess, and F. Windmeijer: More Reliable Inference for Segregation Indices. Technical Report 216, The Centre for Market and Public Organisation, University of Bristol, Bristol, UK (2009)
2. Altavilla A.M., Mazza A., Punzo A.: Sull’impiego di un indice di dissimilarità nello studio della disposizione di popolazioni straniere su un territorio urbano. Rivista Italiana di Economia, Demografia e Statistica, vol. LXIV, p. 7-14 (2010)
3. Duncan, O. D. and B. Duncan: A Methodological Analysis of Segregation Indexes. American Sociological Review 20:210-217 (1955)
4. Massey D. S., Denton, N. A.: The dimensions of residential segregation. Social Forces, 67(2), 281–315 (1988)
5. Mazza, A., Punzo, A.: On the upward bias of the dissimilarity index and its corrections. Sociological Methods & Research 44(1): 80–107 (2015).
6. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/> (2017).