Stallion Journal for Multidisciplinary Associated Research Studies
ISSN (Online): 2583-3340
Volume-4 Issue-5 || October 2025 || PP. 58-64 https://doi.org/10.55544/sjmars.4.5.8

Development and Introduction of a Novel Method for Measuring Spatial Balance

Pohanyar Ahmad Saboor Ahmadzay
Teaching Assistant, Lecturer of Math Department, Education Faculty, Faryab University, AFGHANISTAN.

Corresponding Author: aszahmadzay@gmail.com

www.sjmars.com || Vol. 4 No. 5 (2025): October Issue

Date of Submission: 27-09-2025	**Date of Acceptance:** 04-10-2025	**Date of Publication:** 26-10-2025

ABSTRACT

Spatial balance of samples in the space of auxiliary variables—especially when these variables are correlated with the target variables—plays a crucial role in improving accuracy and reducing the variance of statistical estimates. A uniform and dispersed distribution of sample units in this space, without the need to collect additional information, can provide a better representation of the population structure, since auxiliary data are usually already available. In recent years, various methods have been introduced for designing spatially balanced sampling, including the cube method, local pivotal sampling, and distance- or spatial correlation-based designs. The common feature of these methods is achieving samples with a balanced distribution in the auxiliary space. This paper develops and introduces a novel index for assessing the degree of spatial balance of samples. The proposed index is based on the Voronoi structure and provides a quantitative measure for evaluating spatial balance by analyzing the spatial relationships of each sample unit with its neighbors. Empirical results indicate that using this index can be effective in comparing and optimizing sampling designs and ultimately in reducing the variance of estimates.

Keywords- Spatial balance; balance indices (Moran, Voronoi, balanced Voronoi); local pivotal sampling; inclusion probability; auxiliary variable; spatial dispersion; statistical population.

I. INTRODUCTION

In recent years, the use of auxiliary information in the design of sampling schemes has attracted significant attention due to its potential to increase estimation accuracy and reduce variance (Deville & Tille, 2004; Stevens & Olsen, 2004). In particular, sampling designs that consider the spatial balance of samples within the space of auxiliary variables have shown better performance in representing the population structure and reducing estimator variance (Grafstrom et al., 2012; Robertson et al., 2013). One of the main advantages of using auxiliary information is its availability prior to data collection, which reduces the cost and time of implementing the design. For example, when variables such as gender or age are known and correlated with the target variables, stratified sampling designs can be employed to enhance estimation accuracy (Grafstrom & Schelin, 2014).

In this regard, various methods have been proposed to utilize auxiliary information in sampling, including the cube method (Deville & Tille, 2004), generalized random tessellation stratified (GRTS) sampling (Stevens & Olsen, 2004), local pivotal method (Grafstrom et al., 2012), balanced sampling (Robertson et al., 2013), distance-based sampling «Benedetti & Piersimoni 2017», and weakly associated vectors method (Haziza & Tille, 2020). A common feature among many of these methods is the emphasis on appropriate dispersion of sample units in the auxiliary space, which leads to better representation of the population structure compared to simple random sampling (Grafstrom & Schelin, 2014).

A sample is considered spatially balanced when the distribution of auxiliary variables within it reflects their distribution in the entire population. Such balance brings the Horvitz–Thompson estimator closer to the true values of the auxiliary variables (Grafstrom & Lundström, 2013), and also reduces the likelihood of large gaps in the auxiliary space,

Stallion Journal for Multidisciplinary Associated Research Studies
ISSN (Online): 2583-3340
Volume-4 Issue-5 || October 2025 || PP. 58-64 https://doi.org/10.55544/sjmars.4.5.8

providing more comprehensive coverage of that space. In this article, a novel index for measuring the spatial balance of samples is introduced and analyzed. This index is based on the Voronoi structure and, by evaluating the relative position of each sample unit with respect to its neighbors in the auxiliary variable space, provides an effective metric for assessing the spatial quality of the sample. This index can play a significant role in reducing the variance of population estimates, especially in the presence of strong auxiliary variables.

Grafstrom and Schelin (2014) define spatial balance as a characteristic of a sample in which, for every coherent subset $U^* \subset U$ of the finite population U, the following relationship is approximately satisfied:

$$\sum_{i \epsilon U^*} \delta_i \approx \sum_{i \epsilon U^*} \pi_i,$$

where δ_i is the selection indicator for unit i, and π_i is its inclusion probability. A *coherent subset* refers to a convex region in the space of auxiliary variables. From their perspective, a balanced sample is essentially a scaled-down version of the population that reflects its overall structure. Based on this idea, the weighted distribution of auxiliary variables for the population and the sample can be defined, respectively, as $Gx(U^*)$and $\hat{G}x(U^*)$. When $\boldsymbol{x} = \pi$, these distributions can be rewritten as follows:

$$\hat{G}_\pi(U^*) = \frac{\sum_{i \in U^*} \pi_i}{\sum_{i \in U} \pi_i}, \qquad G_\pi(U^*) = \frac{\sum_{i \in U^*} \delta_i}{\sum_{i \in U} \pi_i}.$$

Spatial balance refers to the closeness of these two distributions:

$$\hat{G}_\pi(U^*) \approx G_\pi(U^*).$$

To assess the overall balance in a sampling design, the expected value of the spatial balance index B is defined as follows:

$$E(B) = \sum_s p(s)\ .B(s),$$

which is typically approximated through Monte Carlo simulation.

$$E^*(B) = \frac{1}{R}\sum_{r=1}^{R} B(s_r).$$

II. OBJECTIVE

The aim of this study is to develop and introduce a novel index for measuring the spatial balance of samples within the space of auxiliary variables. The proposed index is designed based on the Voronoi structure and provides a quantitative measure for evaluating spatial balance by analyzing the spatial relationships among sample units.

III. NEW INDEX FOR SPATIAL BALANCE

One of the well-known methods for measuring spatial balance is the Voronoi index, which was first introduced by Stevens Jr. and Olsen (2004). This index is based on partitioning the population U into Voronoi regions $\{U_i\}_{i \in S}$ according to proximity to the sample units S, such that each U_i contains the population units closest to unit iii.
A version of the Voronoi spatial balance index is defined as follows:

$$B_{VO}(s) = n \sum_{i \in S} \left(\hat{G}_\pi(U_i) - G_\pi(U_i)\right).^2$$

where $\hat{G}_\pi(U_i)$represents the actual sample proportion and $G_\pi(U_i)$ denotes the expected proportion (based on π_i) in the region U_i. If these two values are equal, the sample is considered perfectly balanced according to the Voronoi index.
Ideally, balance in the entire population is established as follows:

$$\hat{G}_\pi(U) = G_\pi(U).$$

However, for a more detailed analysis of local balance in the space of auxiliary variables, the following vector difference can be defined for each region U^*:

Stallion Journal for Multidisciplinary Associated Research Studies
ISSN (Online): 2583-3340
Volume-4 Issue-5 || October 2025 || PP. 58-64 https://doi.org/10.55544/sjmars.4.5.8

$$d_{\boldsymbol{x}}(U^*) = t_{\boldsymbol{x}}(U) \cdot \left(\hat{G}_{\boldsymbol{x}}(U^*) - G_{\boldsymbol{x}}(U^*)\right).$$

which represents the deviation of the sample from balance in the auxiliary variables xxx within the region U^*.
Based on this, the local spatial balance index is defined using the Mahalanobis distance as follows:

$$B_{LB}(s) = \sqrt{\frac{1}{\mid U \mid} \sum_{i \in S} d(U_i)^T \, Q^{-1} d(U_i)},$$

where $Q = \boldsymbol{x}^T\boldsymbol{x}$ is the design matrix. The choice of Q instead of the covariance matrix not only accounts for the statistical orientation of the variables but also incorporates balance in terms of the size of the region.
Let me know if you'd like to translate and format the matrix or equation that comes next.

$$\boldsymbol{x} = [1, \boldsymbol{x}_1, \dots, \boldsymbol{x}_p],$$

The first component (the unit vector) enables the assessment of balance based on the number of units in each region, as is also considered in the Voronoi index. Although indices such as the Balanced Voronoi Index (BVO) and the Local Spatial Balance index (BLB) are powerful tools for evaluating spatial balance in samples, they share a common limitation: these indices lack a normalized scale and their values can vary within the range $[0, \infty)$. For this reason, it is not possible to directly interpret their values to determine whether a sampling design is "good" or "bad." In practice, to compare two designs in terms of spatial balance, the distribution of the index across repeated sampling iterations must be analyzed—something that is only feasible for small populations where all possible samples can be enumerated.

To empirically examine these indices, a point population with a defined structure (Figure 1) has been considered. In Figure 1, samples from this population are displayed. Assuming a sample size of n=3n = 3n=3, the values of three spatial balance indices, including:

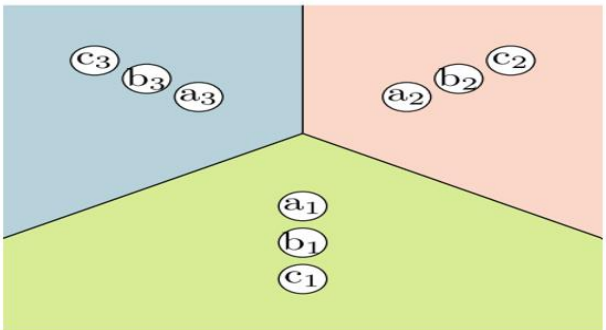


Figure 1: depicts a population consisting of three clusters (1, 2, and 3), each containing three units: a, b, and c.

The distance between units within each cluster is smaller than the distance between any two units from different clusters have been calculated for the samples selected from the target population. The results of this comparison are presented in Table 1.

Table 1: The spatial balance measures for samples of size 3 from the population in Figure 1.

Sample	#	VO	MI	LB
(a) Stratified samples				
aaa	1	0	−1	0.26
aab	3	0	−1	0.21
aac	3	0	−1	0.26
abb	3	0	−1	0.15
abc	6	0	−1	0.21
acc	3	0	−1	0.26
bbb	1	0	−1	0.00
bbc	3	0	−1	0.15
bcc	3	0	−1	0.21
ccc	1	0	−1	0.26
Mean	27	0	−1	0.20
(b) Clustered samples				
$a_ib_ic_i$	3	0.89	1	0.99
$a_ib_ia_j$	6	0.52	0	0.62
$a_ib_ib_j$	6	0.07	0	0.61
$a_ib_ic_j$	6	0.07	0	0.63
$a_ic_ia_j$	6	0.50	0	0.67
$a_ic_ib_j$	6	0.17	0	0.67
$a_ic_ic_j$	6	0.17	0	0.68
$b_ic_ia_j$	6	0.52	0	0.73
$b_ic_ib_j$	6	0.52	0	0.76
$b_ic_ic_j$	6	0.30	0	0.79
Mean	84	0.23	−0.29	0.54

This comparison clearly demonstrates how different indices may or may not be sensitive to subtle differences in spatial dispersion. It also highlights the growing need for developing normalized versions of these indices to enable better analysis and comparison in larger populations. As observed, neither the Moran index nor the Voronoi index made meaningful distinctions between stratified and clustered sample cases. In this analysis, two different sampling methods were considered:

Stallion Journal for Multidisciplinary Associated Research Studies
ISSN (Online): 2583-3340
Volume-4 Issue-5 || October 2025 || PP. 58-64 https://doi.org/10.55544/sjmars.4.5.8

cluster sampling and stratified sampling. In the cluster sampling method, one unit was selected from each cluster, and the different cluster members were combined to form the final sample.

The results indicate that the Moran and Voronoi indices did not differentiate between the stratified samples, whereas the local balance index was able to identify this distinction. According to this index, all stratified samples were considered perfectly balanced by the Moran and Voronoi indices; however, the local balance index, by taking into account the position of each sample within its respective cluster, calculated the degree of balance more precisely. Additionally, in cluster sampling, the Moran index largely failed to distinguish among most samples, treating them similarly to stratified random samples. In contrast, the Voronoi index showed some degree of differentiation, and the local balance index based on Voronoi calculated and demonstrated these differences with greater accuracy.

IV. SIMULATION

To evaluate the performance of spatial balance indices under various conditions, a series of simulations were conducted on populations with different spatial distribution patterns and two types of inclusion probabilities (equal and unequal) (Figure 2). These populations consist of three distinct spatial structures: regular, random, and clustered. For each structure, two scenarios for inclusion probabilities were considered, where in the unequal scenario, the size of the points varies proportionally with the selection probability. To examine the effect of different sampling designs on spatial balance, the following three methods were employed:

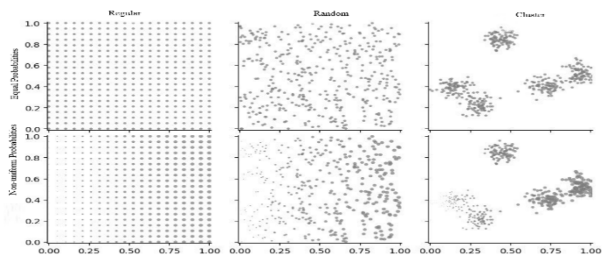


Figure 2 illustrates six simulated populations, representing three different spatial dispersion patterns from left to right: regular, random, and clustered. The upper row shows populations with equal inclusion probabilities, while the lower row depicts populations with unequal inclusion probabilities.

These designs were implemented across six combined scenarios arising from the different spatial dispersion patterns and types of inclusion probabilities. To analyze performance, the Moran, Voronoi, and Balanced Voronoi indices were calculated for each sampling. The simulation process was conducted using the Monte Carlo method with 1,000 iterations, and the results are presented in boxplot diagrams.

V. ANALYSIS OF THE MORAN INDEX

The analysis of the Moran index (Figure 3) shows that the local pivotal sampling design performed significantly better than other methods in most scenarios. Specifically, in clustered and regular populations, this design produced more negative Moran index values, indicating higher dispersion and better balance in the samples. For example, in the clustered scenario with equal inclusion probabilities, the Moran index for the local pivotal design was mostly within the range [-0.2,-0.1], while the other two methods yielded values close to zero. Across all population structures, even under unequal inclusion probabilities, the local pivotal design managed to generate a more uniform distribution with stronger negative spatial autocorrelation among sample units. On the other hand, simple random sampling and maximum entropy methods exhibited more variable behavior, producing weak positive autocorrelation in some cases and weak negative autocorrelation in others. Given the definition of the Moran index, where more negative values indicate greater dispersion and reduced clustering, it can be concluded that the local pivotal design has a clear advantage in terms of spatial balance over the other two methods. Next, the results related to the Voronoi and Balanced Voronoi indices will be examined for a more detailed analysis of local balance.

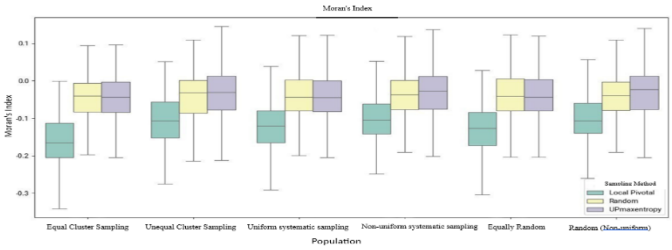


Figure 3 illustrates the simulation results using the Moran index over 1,000 iterations, presented as boxplots.

VI. ANALYSIS OF THE VORONOI SPATIAL BALANCE INDEX

Similar to the Moran index, the Voronoi index was able to reveal significant differences in the performance of various sampling designs under different conditions (Figure 4). The analysis shows that the local pivotal design generally produced lower values of this index, indicating better spatial balance in the samples. Across all population structures (regular, random, clustered) and under both types of inclusion probabilities (equal and unequal), the index values for the local pivotal method mostly fell within a narrow and consistent range of [0.15,0.20]. In contrast, the simple random sampling and maximum entropy methods frequently produced larger and more variable index values, sometimes reaching up to 1. For example:

- In clustered populations, the local pivotal method maintained a range between [0.15,0.20] under both inclusion probability scenarios, whereas the maximum entropy method produced index values up to 1.00 and simple random sampling up to 0.50.
- In the regular structure, the performance of the local pivotal method remained stable within the same range, while the other methods covered broader ranges from [0.20,0.75].
- In the random structure as well, the superiority of the local pivotal method in maintaining stability and minimizing the index was preserved.

Since the Voronoi index measures the deviation between the expected and actual distribution of sample units within neighboring regions, lower values indicate a more accurate representation of the spatial structure of the population. Therefore, it can be concluded that the local pivotal method has succeeded in producing samples with a more balanced and uniform structure relative to the population. The median values of the indices obtained from this method consistently remained within the range [0.15,0.20], while the medians for other methods were considerably higher and more variable. This pattern confirms that simple random sampling and maximum entropy methods have a lower capability in reproducing the population's structure in a spatially balanced manner. Overall, as observed in the analysis of the Moran index, the Voronoi index also clearly demonstrates the relative superiority of the local pivotal design in generating more spatially balanced and dispersed samples.

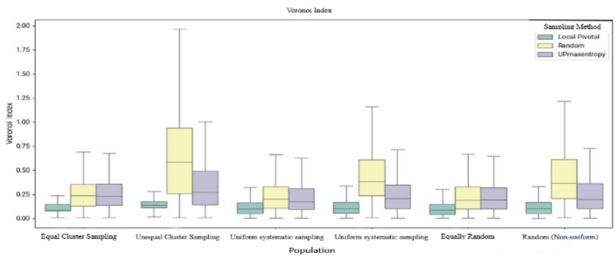


Figure 4 presents the simulation results using the Voronoi index over 1,000 iterations, illustrated with boxplots.

VII. ANALYSIS OF THE BALANCED VORONOI INDEX

The Balanced Voronoi Index (Figure 5), which evaluates local balance not only in terms of geographic position but also based on auxiliary variables, confirms the relative superiority of the local pivotal design in most scenarios. By utilizing the Mahalanobis distance, this index is able to more accurately assess spatial balance in the feature space. The results show that:

- In clustered populations (under both types of inclusion probabilities), the local pivotal method produced consistently low index values within the range [0.2,0.5], whereas the simple random and maximum entropy methods registered higher values, up to 1.5.
- In regular populations, the index for the local pivotal method again remained within the [0.2,0.5] range, while the other methods produced values as high as approximately 1.0.

ISSN (Online): 2583-3340
Volume-4 Issue-5 || October 2025 || PP. 58-64 https://doi.org/10.55544/sjmars.4.5.8

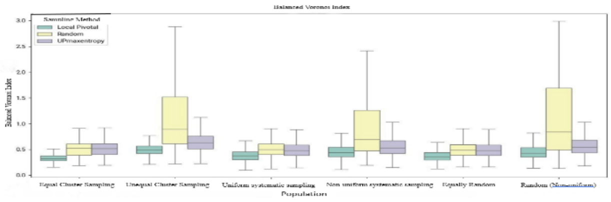
- In the random structure, the local pivotal method once more demonstrated more stable performance. Its index values were confined to the range [0.3,0.5], whereas the other methods occasionally yielded broader ranges, up to [0.3,1.5].

Lower values of the Balanced Voronoi Index—especially when approaching one-quarter—indicate a desirable level of balance in sampling. In most of the scenarios studied, the median of the index for the local pivotal method fell within the range [0.2,0.5], which, compared to the higher values (up to 1.5) seen in other methods, reflects this method's greater accuracy in representing the population structure.

Another notable point is the stability of the index values for the local pivotal method. This trait demonstrates that even when confronted with diverse spatial structures and varying inclusion probability distributions, this method was able to maintain spatial balance in the samples—whereas other methods showed significant fluctuations.

In summary, the analysis of the Balanced Voronoi Index clearly demonstrates that:

- The local pivotal method is capable of producing more uniformly and spatially balanced samples.
- It exhibits high stability and adaptability in response to changes in population structure and probability distributions.
- Overall, like the Moran and Voronoi indices, this index also confirms the superior performance of the local pivotal design compared to the other sampling strategies examined.



The simulation, conducted using the Balanced Voronoi Index over 1,000 iterations, produced results that are presented as boxplots in Figure 5.

VIII. CONCLUSION

A comprehensive analysis of the three indices—Moran, Voronoi, and Balanced Voronoi—demonstrated that the local pivotal sampling method consistently outperforms the two other common methods across all scenarios and conditions. By producing stable values close to the optimal range (i.e., more negative values for the Moran index and lower values for the other indices), this method has achieved a more desirable dispersion of samples within the target population.

In contrast, the simple random sampling method showed lower efficiency in terms of spatial balance, due to greater fluctuations and a wider range of index values. While the maximum entropy method offers higher statistical stability from a theoretical perspective and enhances entropy in sampling, its performance in achieving spatially balanced sample dispersion was inferior to the local pivotal method.

Overall, the findings indicate that the local pivotal method—by simultaneously maintaining spatial balance and result stability—serves as a reliable and efficient option for sampling in populations with complex and variable spatial structures. This method is particularly suitable for studies aiming to obtain optimally dispersed and accurately representative samples of the population.

REFERENCES

[1] Benedetti, R., & Piersimoni, F. (2017). A spatially balanced design with probability function proportional to the within sample distance. *Biometrical Journal, 59*(5), 1067–1084. https://doi.org/10.1002/bimj.201600194

[2] Bivand, R. S., Pebesma, E., & Gómez-Rubio, V. (2013). *Applied spatial data analysis with R* (2nd ed.). Springer. https://asdar-book.org/

[3] Deville, J.-C., & Tillé, Y. (2004). Efficient balanced sampling: The cube method. *Biometrika, 91*, 893–912.

[4] Deville, J.-C., & Tillé, Y. (2005). Variance approximation under balanced sampling. *Journal of Statistical Planning and Inference, 128*(2), 569–591. https://doi.org/10.1016/j.jspi.2003.11.011

[5] Graf, M. (2011). Use of survey weights for the analysis of compositional data. In V. Pawlowsky-Glahn & A. Buccianti (Eds.), *Compositional data analysis: Theory and applications* (pp. 114–127). Wiley.

[6] Grafström, A. (2012). Spatially correlated Poisson sampling. *Journal of Statistical Planning and Inference, 142*(1), 139–147. https://doi.org/10.1016/j.jspi.2011.07.003

[7] Grafström, A., Lisic, J., & Prentius, W. (2024). *BalancedSampling: Balanced and Spatially Balanced Sampling* (R package version 2.0.6). https://CRAN.R-project.org/package=BalancedSampling

[8] Grafström, A., Lundström, N. L., & Schelin, L. (2012). Spatially balanced sampling through the pivotal method. *Biometrics, 68*(2), 514–520. https://doi.org/10.1111/j.1541-0420.2011.01699.x

[9] Grafström, A., & Lundström, N. L. P. (2013). Why well spread probability samples are balanced. *Open Journal of Statistics, 3*(1), 36–41. https://doi.org/10.4236/ojs.2013.31005
[10] Grafström, A., & Schelin, L. (2014). How to select representative samples. *Scandinavian Journal of Statistics, 41*(2), 277–290. https://doi.org/10.1111/sjos.12016
[11] Grafström, A., & Tillé, Y. (2013). Doubly balanced spatial sampling with spreading and restitution of auxiliary totals. *Environmetrics, 24*(2), 120–131. https://doi.org/10.1002/env.2194
[12] Jauslin, R., & Tillé, Y. (2020). Spatial spread sampling using weakly associated vectors. *Journal of Agricultural, Biological and Environmental Statistics, 25*(3), 431–451. https://doi.org/10.1007/s13253-020-00407-1
[13] Jauslin, R., Panahbehagh, B., & Tillé, Y. (2022). Sequential spatially balanced sampling. *Environmetrics, 33*(8), e2776. https://doi.org/10.1002/env.2776
[14] Kruskal, W., & Mosteller, F. (1979). Representative sampling, I–III. *International Statistical Review, 47*(1, 2, 3), 13–265. https://doi.org/10.2307/1403202
[15] Panahbehagh, B., Jauslin, R., & Tillé, Y. (2023). A general stream sampling design. *Computational Statistics, 39*(6), 2899–2924. https://doi.org/10.1007/s00180-023-01408-7
[16] Prentius, W., & Grafström, A. (2024). How to find the best sampling design: A new measure of spatial balance. *Environmetrics, 35*(7), e2878. https://doi.org/10.1002/env.2878
[17] Robertson, B. L., Brown, J. A., McDonald, T. L., & Jaksons, P. (2013). BAS: Balanced acceptance sampling of natural resources. *Biometrics, 69*(3), 776–784. https://doi.org/10.1111/biom.12059
[18] Särndal, C. E. (2011). The 2010 Morris Hansen lecture: Dealing with survey nonresponse in data collection and in estimation. *Journal of Official Statistics, 27*(1), 1.
[19] Stevens, D. L., Jr., & Olsen, A. R. (2004). Spatially balanced sampling of natural resources. *Journal of the American Statistical Association, 99*(465), 262–278. https://doi.org/10.1198/016214504000000250
[20] Y., Dickson, M. M., Espa, G., & Giuliani, D. (2018). Measuring the spatial balance of a sample: A new measure based on Moran's I index. *Spatial Statistics, 23*, 182–192. https://doi.org/10.1016/j.spasta.2018.02.001