



UNIVERSIDAD NACIONAL DE ROSARIO
FACULTAD DE CIENCIAS ECONÓMICAS Y ESTADÍSTICA
SECRETARIA DE CIENCIA Y TECNOLOGIA E INSTITUTOS DE INVESTIGACIONES

Resumen Ampliado

Jornadas Anuales

***“Investigaciones en la Facultad” de
Ciencias Económicas y Estadística***



Borra, Virginia Laura
Castellanos, Valentina Celeste
Pagura, José Alberto

Instituto de investigaciones Teóricas y Aplicadas, Escuela de Estadística.

MEDIDAS DE BALANCE ESPACIAL. ANÁLISIS DE DIFERENTES PROPUESTAS¹

Resumen:

En estudios donde las unidades poblacionales se distribuyen en el espacio, la presencia de autocorrelación espacial positiva, fenómeno muy frecuente, hace que la selección de unidades cercanas genere información redundante y disminuya la precisión de los estimadores. En este contexto, el balance espacial de una muestra, entendido como una adecuada dispersión territorial de las unidades seleccionadas, resulta una propiedad clave para mejorar la calidad de las estimaciones. Este trabajo compara distintas medidas propuestas en la literatura para cuantificar el balance espacial de una muestra: *Spatial Balance Index* (SBI), *Pielou's Evenness Index* (PEI), *Local Spatial Balance Measure* (LB) e Índice de Moran normalizado (IB). El análisis se realiza a partir de la selección de muestras de la población de radios censales de la ciudad de Rosario, considerando distintos métodos de muestreo espacial y el muestreo aleatorio simple como referencia. Los resultados muestran que todos los índices permiten discriminar adecuadamente entre diseños espaciales y no espaciales, destacándose el índice de Moran normalizado por su fácil interpretación y su rango acotado. Estos atributos lo convierten en una herramienta particularmente útil para evaluar y comparar diseños de muestreo espacial.

Palabras claves: Balance espacial; Muestreo espacial; Correlación espacial.

Abstract:

In studies where population units are spatially distributed, the presence of positive spatial autocorrelation, a very frequent phenomenon, causes the selection of nearby units to generate redundant information and reduces the precision of estimators. In this context, the spatial balance of a sample, understood as an adequate territorial dispersion of the selected units, is a key property for improving the quality of estimates. This paper compares different measures proposed in the literature to quantify the spatial balance of a sample: the *Spatial Balance Index* (SBI), *Pielou's Evenness Index* (PEI), the *Local Spatial Balance Measure* (LB), and the normalized Moran Index (IB). The analysis is based on samples drawn from the population of census areas of the city of Rosario, considering different spatial sampling methods, with simple random sampling used as a reference. The results show that all indices adequately discriminate between spatial and non-spatial designs, with the normalized Moran Index standing out for its ease of interpretation and bounded range. These attributes make it a particularly useful tool for evaluating and comparing spatial sampling designs.

Keywords: Spatial balance; Spatial sampling; Spatial correlation.

¹ Trabajo elaborado en el marco del Proyecto (80020220700061UR), titulado: "Muestreo espacial y su aplicación en estudios económicos y sociales", dirigido por José A. Pagura

Introducción

En muchos estudios donde se busca inferir características de una población a partir de una muestra, las unidades se encuentran distribuidas en un espacio geográfico. En este contexto, las variables analizadas suelen exhibir algún grado de dependencia o correlación espacial.

La situación más frecuente en los estudios socioeconómicos es la existencia de autocorrelación espacial positiva, lo que significa que las unidades que se encuentran cercanas en el espacio tendrán un valor de la variable de interés parecido. Es por esto que la selección de unidades vecinas brinda información redundante, teniendo como consecuencia la obtención de estimaciones poco precisas. Cuando puede asumirse la existencia de dicha clase de correlación, los métodos de muestreo espacial resultan muy beneficiosos.

Para esta situación, diferentes autores han propuesto métodos de selección que se orientan a la inclusión en la muestra de unidades dispersas. Entre ellos pueden mencionarse SCPS (*Spatial Correlated Poisson Sampling*), LPM2 (*Local Pivotal Method*), GRTS (*Generalized Random Tessellation Stratified*) y HPWD (*Heuristic Product Within Distance*), que son los considerados en este estudio.

La calidad de las estimaciones proporcionadas por una muestra se puede evaluar con el error cuadrático medio del estimador, el que puede estimarse con los datos de una muestra en particular. Sin embargo, con este mismo propósito, se puede considerar alguna otra medida que refleje la dispersión de la muestra en la región delimitada para el estudio, asumiendo que frente a la existencia de autocorrelación positiva, unidades de la muestra bien distribuidas en el espacio producen estimaciones más precisas. En este sentido, en la literatura más reciente, se han propuesto diferentes índices para medir el "balance espacial" de una muestra.

Se dice que una muestra está espacialmente balanceada si está bien repartida en el territorio poblacional. Esta característica mejora la precisión de los estimadores de totales o medias, ya que disminuye la redundancia en la información de la muestra.

A partir de este concepto, se han definido diferentes índices, como ser: *SBI* (Grafström, Lundström, Schelin, 2012), *PEI* (Dumelle, Kincaid, Olsen, Weber, 2023), *LB* (Prentius, Grafström, 2024) e *IB* (Tillé, Dickson, Espa, Giuliani, 2018). La descripción se presenta en la sección correspondiente.

Objetivos

Presentar los fundamentos de cada uno de los índices propuestos y realizar un estudio comparativo para evaluar el balance espacial de muestras de radios censales de la ciudad de Rosario.

Metodología

El balance espacial se considera una de las propiedades más importantes y útiles de una muestra seleccionada de una población espacial. Si bien no hay una definición formal única sobre este concepto, se entiende que una muestra está espacialmente balanceada si está "bien" distribuida en el territorio estudiado.

UNR

A partir de la comprensión de este concepto, se proponen diferentes procedimientos para cuantificar el balance espacial.

Originalmente, Stevens y Olsen (2004) propusieron medir el equilibrio espacial subdividiendo el territorio en estudio en polígonos conocidos como "de Voronoi" de acuerdo a las siguientes definiciones.

Sea una población espacial finita U de N unidades, y de acuerdo a un plan de muestreo establecido, cada unidad i tiene una probabilidad de inclusión π_i , tal que $\sum_{i=1}^N \pi_i = n$, donde n es el tamaño de la muestra.

Para una muestra $S = (s_1, s_2, \dots, s_n)$ seleccionada de U , se define el polígono de Voronoi asociado a la unidad s_i incluida en la muestra S como el conjunto de todas las unidades de la población U que están más cerca de s_i que de cualquier otra unidad s_j de la muestra. Además, sea v_i la suma de las probabilidades de inclusión de las unidades que pertenecen al i -ésimo polígono de Voronoi, cantidad aleatoria ya que depende de la muestra seleccionada, con $E(v_i) = 1$. A partir de este resultado, Grafström et. al. (2012), establecen que una muestra está perfectamente balanceada espacialmente si todos los v_i son iguales a 1.

Índice SBI (Spatial Balanced Index)

Grafström et. al. (2012) definieron el índice SBI para medir el balance espacial de una muestra como:

$$SBI = \frac{1}{n} \sum_{i=1}^n (v_i - 1)^2.$$

Este índice tendrá un valor mayor o igual que 0, $SBI \geq 0$, significando el valor 0, un balance espacial perfecto.

El rango posible de SBI depende del patrón espacial de la población U y, por lo tanto, difiere entre distintas poblaciones, lo que implica que SBI puede utilizarse con éxito sólo para comparar el nivel de balance espacial de diferentes muestras de la misma población. Además, como el índice no varía dentro de un rango finito fijo, no está caracterizado por un valor de referencia específico que discrimine entre ausencia y presencia de balance espacial. Los paquetes *BalancedSampling* y *Spsampling* de R, permiten el cálculo del índice SBI mediante las funciones $sb()$ y $sbi()$ respectivamente.

Índice PEI (Pielou's Evenness Index)

Otro índice recientemente presentado, que también toma en consideración los polígonos de Voronoi, es el índice PEI (Dumelle et. Al., 2023). El mismo se define como:

$$PEI = 1 + \sum_{i=1}^n \frac{v_i}{n} \frac{\ln\left(\frac{v_i}{n}\right)}{\ln(n)}.$$

Este índice se basa en el índice de equidad de Pielou, que es una métrica de pérdida (Pielou, 1966).

El índice PEI varía entre 0 y 1, $0 \leq PEI \leq 1$, indicando 0 un balance espacial perfecto y 1 la falta de balance espacial, es decir que, a medida que el PEI aumenta, el equilibrio

espacial empeora. El paquete *spsurvey* de R, tiene la función *sp_balance()* que permite su cálculo.

Índice LB (Local Spatial Balance Measure)

El *índice LB* (Prentius, Grafström, 2024) está basado en los polígonos de Voronoi y en una ecuación de equilibrio espacial que tiene en cuenta información auxiliar de las unidades. Este índice toma valores iguales o mayores que 0, $LB \geq 0$, indicando el valor 0 un balance espacial perfecto. Esta medida puede utilizarse para comparar el balance espacial de diferentes muestras tomadas de la misma población.

Los autores señalan como una de las potenciales aplicaciones más relevantes, el tratamiento del problema de la no respuesta, orientando los esfuerzos hacia las unidades que provean más información.

Índice de Moran normalizado IB

Por último, se presenta el *Índice de Moran normalizado IB* (Tillé et. al., 2018) que se basa en el conocido índice de Moran utilizado para cuantificar la correlación espacial. La diferencia radica en que, en lugar de utilizar la variable en estudio y_i , se emplea una variable indicadora de inclusión en la muestra:

$$\delta_i = \begin{cases} 1 & \text{si } i \in S \\ 0 & \text{en otro caso} \end{cases}$$

Cuando la muestra está bien distribuida en el espacio, las unidades muestrales tienden a ubicarse relativamente alejadas entre sí y, por lo tanto, la variable δ_i está negativamente autocorrelacionada. Por otro lado, si la muestra está espacialmente agrupada, las unidades muestrales tienden a ubicarse relativamente cerca unas de otras y, en consecuencia, la variable δ_i está autocorrelacionada positivamente.

Para lograr la medida del balance espacial, se propone una normalización del índice de Moran. Su expresión matemática es:

$$IB = \frac{(\delta - \bar{\delta}_w)^T W (\delta - \bar{\delta}_w)}{\sqrt{(\delta - \bar{\delta}_w)^T D (\delta - \bar{\delta}_w) (\delta - \bar{\delta}_w)^T B (\delta - \bar{\delta}_w)}}$$

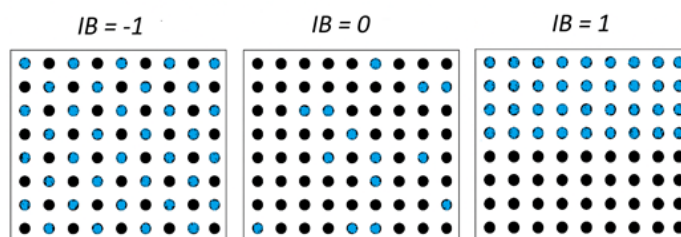
donde:

- W es la matriz de pesos espaciales,
- D y B matrices construidas a partir de los pesos espaciales.

Este índice puede tomar valores entre -1 y 1 , $-1 \leq IB \leq 1$. El valor $IB = 1$, corresponde a una máxima concentración en la muestra seleccionada; mientras que $IB = -1$ indica un balance espacial perfecto. El valor 0 del índice está asociado a una distribución aleatoria de las unidades de la muestra (Figura 1).



Figura 1: Ejemplos de situaciones hipotéticas en las que el IB alcanza sus valores extremos.



Nota: Las ubicaciones de las unidades muestrales se representan con círculos celestes y las ubicaciones de las unidades no muestrales se representan con círculos negros.

El rango de variación permite que el índice refleje de manera clara el grado de dispersión espacial de una muestra y se aprecie si la muestra está espacialmente balanceada o agrupada. Por lo tanto, esta medida puede tener una aplicabilidad más general que simplemente evaluar qué diseño de muestreo espacial es mejor que otro para lograr la dispersión deseada en una población dada.

El cálculo de este índice puede realizarse recurriendo a la función $IB()$ del paquete *WaveSampling* de R.

Resultados

Para evaluar el comportamiento de los diferentes índices considerados, se extrajeron 10.000 muestras de tamaño $n = 100$ de la población de radios censales de la ciudad de Rosario ($N = 1043$), por cada uno de los cinco métodos de selección considerados: HPWD, LPM2, SCPS, GRTS y MSA. Los primeros cuatro métodos tienen en cuenta que la población se distribuye espacialmente en el territorio, mientras que el MSA no incorpora este aspecto espacial en el diseño. Para cada muestra se calcularon los índices de balance espacial propuestos.

Como medidas de resumen, se calcularon la media, el desvío estándar y el coeficiente de variación de estos índices para las 10.000 muestras en los diferentes métodos de selección (Tabla 1).



Tabla 1: Medidas resumen de los índices de balance espacial para cada método de muestreo

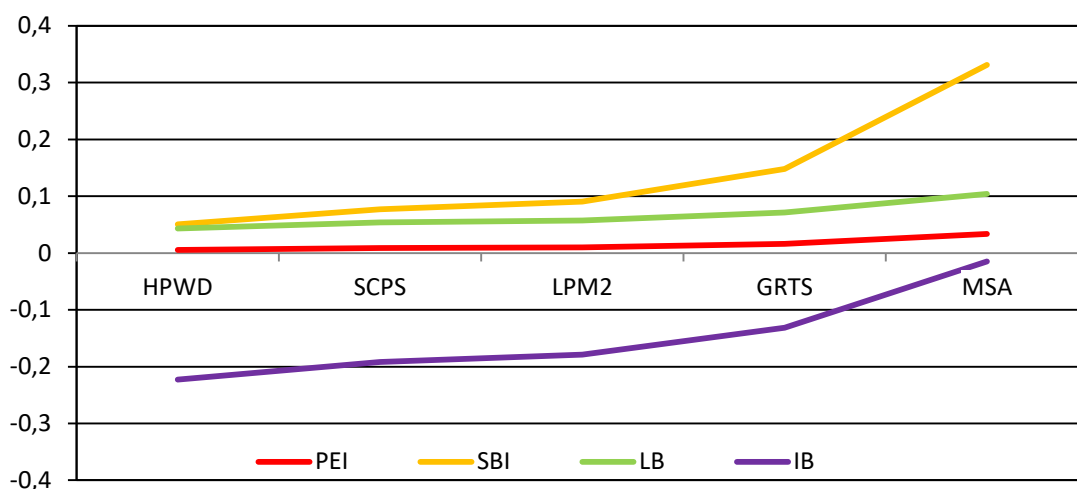
Índice	Método	Promedio	Desvío estándar	CV
<i>PEI</i>	HPWD	0,006	0,001	0,147
	SCPS	0,009	0,001	0,166
	LPM2	0,010	0,002	0,156
	GRTS	0,016	0,002	0,154
	MSA	0,034	0,006	0,171
<i>SBI</i>	HPWD	0,050	0,007	0,144
	SCPS	0,077	0,012	0,160
	LPM2	0,090	0,014	0,153
	GRTS	0,148	0,024	0,160
	MSA	0,331	0,069	0,207
<i>IB</i>	HPWD	-0,223	0,025	0,114
	SCPS	-0,192	0,026	0,134
	LPM2	-0,179	0,026	0,144
	GRTS	-0,131	0,029	0,224
	MSA	-0,015	0,035	2,354
<i>LB</i>	HPWD	0,043	0,004	0,090
	SCPS	0,054	0,006	0,105
	LPM2	0,057	0,006	0,098
	GRTS	0,071	0,007	0,094
	MSA	0,104	0,012	0,113

El campo de variación de cada uno de los índices es diferente. Sin embargo, todos los índices coinciden en que el mejor resultado es para el método HPWD y el peor es para MSA.

Los índices *IB* y *SBI* distinguen muy bien entre un método "no espacial" y otro que tiene en cuenta la dispersión de las unidades. El índice *IB* toma un valor cercano a 0 para el MSA, indicando una distribución aleatoria en la localización de las unidades de la muestra y *SBI* toma el valor más grande para el MSA (Figura 2).

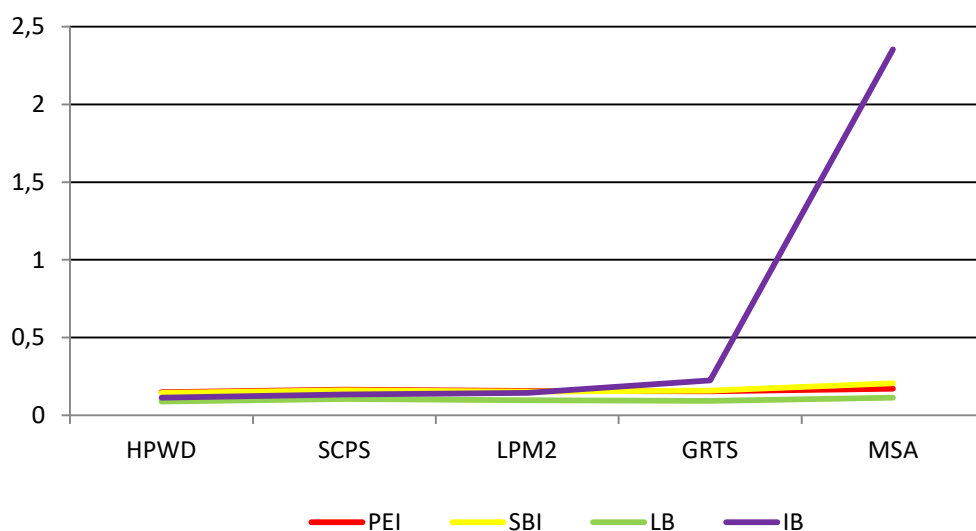


Figura 2: Promedio de los índices de balance espacial según métodos de muestreo



Sería deseable utilizar un índice que presente menor variabilidad considerando aquellos métodos que tienen en cuenta la dispersión de la muestra. Para poder apreciar esta condición se calculó el coeficiente de variación, representando su comportamiento en la Figura 3. El índice que mejor se ajusta a esta condición es el índice de Moran normalizado, *IB*.

Figura 3: Coeficiente de variación de los índice de balance espacial según métodos de muestreo



Comentarios finales

La eficiencia de un plan de muestreo se evalúa mediante el error cuadrático medio del estimador que se emplea. En el caso de los métodos de selección que tienen en cuenta la localización de las unidades, muestreo espacial, el uso de índices de balance espacial presenta una visión complementaria en cuanto a que se calculan



UNR

sobre una muestra particular y revela si la misma está bien distribuida en el espacio, característica deseable cuando existe autocorrelación espacial positiva.

Diferentes autores han definido indicadores de balance espacial con puntos de partida diferentes y el interés del presente trabajo estuvo en evaluar como se comportan esos indicadores, a través de las diferentes muestras posibles de seleccionar por cada uno de cinco planes de muestreo, cuatro de ellos que corresponden a la categoría de muestreo espacial y utilizando un quinto método, el muestreo aleatorio simple, como punto de referencia. Para realizar este estudio se extrajeron 10000 muestras de la población de radios censales para cada método de selección, es decir un total de 50000 muestras. Se calcularon los cuatro índices que fueron considerados y se obtuvieron medidas de resumen de los resultados encontrados con cada método de selección: el valor medio, la desviación estándar y el coeficiente de variación. Al momento de realizar las comparaciones, de debió poner especial cuidado a que los indicadores presentan diferentes rangos de variación y los valores asociados al balance espacial no son siempre los mismos.

Del análisis de estos resultados, se encuentra que los indicadores estudiados presentan un buen desempeño en cuanto a qué para los métodos de muestreo espacial, sus valores son siempre más cercanos a la situación de balance espacial perfecto que en el caso del muestreo aleatorio simple, es decir, reconocen los métodos de selección que han tenido en cuenta la localización de las unidades. Otro detalle de interés se observa en que los coeficientes de variación son mucho menores en los índices de balance espacial, para los métodos de muestreo espacial que para el muestreo aleatorio simple.

En cuanto a la facilidad de interpretación, el índice de Moran normalizado, IB , presenta la ventaja de tomar valores dentro de un rango establecido, $-1 \leq IB \leq 1$ y tanto sus valores extremos como su valor central se encuentran asociados a situaciones perfectamente identificables. El índice PEI presenta la misma ventaja en cuanto a que varía en un rango acotado. Otra ventaja del índice de Moran normalizado es que permite comparar planes de muestreo en poblaciones diferentes.

En cuanto al índice LB sus autores señalan que la posibilidad de incorporación de información auxiliar en su composición sería de gran utilidad en diferentes situaciones, entre ellas el tratamiento de la falta de respuesta.

Es de interés continuar estudiando las propiedades de los índices revisados, abarcando poblaciones con diferentes estructuras de correlación espacial y evaluando su comportamiento con diferentes tamaños de muestra.

También, se debe profundizar en el estudio de SBI , PEI y LB cuando se incorpora información auxiliar ya sea en las probabilidades de inclusión y, en el caso particular del índice LB su inclusión en las ecuaciones de balance que se utilizan como condición de partida.

Referencias Bibliográficas

- Benedetti R, Piersimoni F (2017). "Fast Selection of Spatially Balanced Samples." arXiv:1710.09116 [stat.ME], URL <https://arxiv.org/abs/1710.09116>.

- Dumelle, M., Kincaid, T., Olsen, A. R., & Weber, M. (2023). *spsurvey: Spatial Sampling Design and Analysis in R* (versión 5.4.1) [Software]. *Journal of Statistical Software*, 105(3), 1–29. <https://doi.org/10.18637/jss.v105.i03>
- Grafström A., Lundström N., 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>
- 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>
- 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>
- R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- Stevens D. L., & 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>
- Tillé, Y., Dickson, M. M., Espa, G., & Giuliani, D. (2018). Measuring the spatial balance of a sample: A new measure based on the Moran's I index. *Spatial Statistics*. <https://doi.org/10.1016/j.spasta.2018.02.001>