**3. Material y Método**

Este capítulo está dedicado a presentar los fundamentos metodológicos de diferentes alternativas para el cálculo de indicadores de autocorrelación espacial, así como de los problemas en los que se utilizarán para su comparación.

En primer lugar, se introducen los conceptos sobre Estadística Espacial necesarios para la construcción de indicadores de autocorrelación espacial.

Luego, se realizan consideraciones a tener en cuenta a la hora de trabajar con índices de autocorrelación espacial, presentando los fundamentos teóricos de los índices de Moran, Oden y el índice Empírico de Bayes.

En tercer lugar, se presentan dos problemas en los que se utilizan los tres índices mencionados, con las particularidades que condujeron a su elección: el estudio de la distribución espacial de los Hogares con Necesidades Básicas Insatisfechas (NBI) en la ciudad de Rosario en el año 2010 y Heridos de armas de fuego en la ciudad de Rosario en un determinado período. Se agrega la descripción de los conjuntos de datos empleados para las mencionadas aplicaciones.

**3.1 Estadística Espacial. Conceptos necesarios para el presente trabajo**

En muchos problemas estadísticos, los datos con los que se debe tratar se observan unidades que se encuentran situadas en el espacio y su localización puede especificarse (datos georreferenciados). Si la ubicación de las unidades es relevante para la descripción y análisis del fenómeno que se estudia, estos datos se denominan “datos espaciales”.

Una particularidad de los datos espaciales es la existencia de autocorrelación espacial. Se denomina Estadística Espacial al conjunto de métodos que tienen por objeto la exploración, descripción, visualización, y análisis de datos considerando sus características de distribución en el espacio.

En un análisis de datos espaciales pueden distinguirse tres etapas:

* Análisis exploratorio: se aplican métodos estadísticos convencionales para realizar un reconocimiento del comportamiento de las variables que se estudian y evidenciar la existencia de autocorrelación espacial. Sin la realización de esta etapa no se puede avanzar sobre las siguientes. Los métodos descriptivos se complementan con el cálculo de índices de autocorrelación espacial; el más utilizado es el índice de Moran, cuyos fundamentos se presentan más adelante.
* Análisis estructural: constituye el objetivo principal de la Estadística Espacial y consiste en la construcción de modelos que describen la autocorrelación espacial.
* Predicción: a partir del modelo construido predecir valores de la variable aleatoria que se estudia en sitios donde no se han realizado observaciones.

El interés del presente trabajo está en el estudio de los índices disponibles para la evaluación de la autocorrelación espacial. El ya mencionado índice de Moran, el más ampliamente difundido, puede encontrarse en muchos estudios publicados y se dispone en cualquier software de Estadística Espacial.

Este índice presenta inconvenientes en situaciones particulares y por ese motivo se han propuesto otras alternativas, menos difundidas, que se tratan en la presente tesina como el índice de Oden y el Empirical Index Bayes (EBI).

**Tipos de datos espaciales**

En función de la manera en que se considere el espacio en donde se observa la variable en estudio, los datos espaciales pueden dividirse en tres subtipos:

* Geoestadísticos o espacialmente continuos: estos datos se pueden observar en cualquier posición y corresponden a un fenómeno que se desarrolla en forma continua en la región que se considera. Se puede seleccionar cualquier localización del espacio en estudio para realizar una observación de la variable de interés.
* Reticulares o Lattice (látices): en esta situación cada observación se corresponde con agregaciones espaciales, es decir se observa una variable aleatoria sobre cada una de diferentes áreas en las que se divide la región que se estudia. Estas áreas son polígonos definidos por vértices y lados (fronteras). Según la forma que presenten estas superficies serán regulares o irregulares, las primeras dividen al espacio total de estudio en subáreas idénticas, mientras que las segundas presentan distintas formas y tamaños. La definición de las áreas no resulta trivial, ya que el resultado final del estudio podría depender de la delimitación elegida. La naturaleza del problema determinará la mejor manera de definir la división territorial en distintas áreas.
* Puntuales: la población en estudio es un conjunto de objetos distribuidos en el espacio y de tamaño pequeño en relación a la distancia que los separa. Cada objeto se encuentra en una posición aleatoria en el espacio.

**3.2 Indicadores de Autocorrelación Espacial**

Los métodos estadísticos tradicionales, asumen que las observaciones de una variable se toman bajo condiciones idénticas y de manera independiente. Ellos consideran que los datos son una muestra aleatoria simple, es decir, son independientes e idénticamente distribuidos. Bajo esta suposición se construye la mayoría de la teoría estadística.

Considerar dependencia en los datos es un gran inconveniente a la hora de trabajar con los modelos usuales. Sin embargo, en muchos casos los modelos que incluyen dependencia son más realistas que los que no lo hacen. La idea de que datos cercanos, en el espacio o en el tiempo, son más parecidos, es natural.

En el contexto espacial, esta falta de independencia recibe el nombre de dependencia o autocorrelación espacial, la cual se define mediante una relación funcional entre lo que ocurre en una unidad determinada del espacio y en sus unidades vecinas. En otras palabras, existirá autocorrelación espacial cuando el valor observado de una variable en una unidad o área determinada dependa, en cierta manera, de los valores observados en unidades o áreas vecinas.

La autocorrelación espacial puede ser:

* Negativa: Se presenta una relación inversa entre las unidades vecinas. Áreas con valores altos de la variable serán vecinas de áreas con valores bajos. A modo de ejemplo, considerar la competencia entre plantas por la luz, donde zonas de plantas fuertes pueden estar rodeadas de otras con plantas menos fuertes.
* Positiva: la variable asumirá valores similares en regiones cercanas. Esta situación representa el efecto contagio, lo que ocurre en una unidad se “contagia” a áreas vecinas. Un área con un valor alto de la variable estará rodeada de unidades donde la variable también asuma valores altos.
* Nula: en esta situación no existe autocorrelación espacial, en otras palabras, la variable se distribuye de manera aleatoria en el espacio.

Quienes permiten verificar si se cumple la hipótesis de que una variable se encuentra distribuida en forma aleatoria en el espacio o si, por el contrario, existe asociación significativa entre unidades vecinas son los índices de autocorrelación espacial. Para su cálculo se necesita proporcionar criterios de cercanía entre unidades y pesos que reflejen la fuerza de la influencia en la relación entre las mismas.

**3.2.1Criterios de Vecindad y Pesos Espaciales**

Como dice la primera ley de la geografía, o principio de autocorrelación espacial “Todo está relacionado con todo lo demás, pero las cosas cercanas están más relacionadas que las cosas distantes” (Tobler, 1970). Pero ¿qué se considera cercano? Para poder responder a esta pregunta nace el concepto de vecindad.

Bajo la perspectiva de datos reticulares se considera que no todas las áreas influyen sobre el valor que asume la variable en una determinada área, sino, que solo influirán aquellas que sean vecinas.

Definir qué características deben poseer dos áreas para que sean consideradas vecinas es una cuestión de suma relevancia. Existen varios criterios de vecindad que pueden utilizarse y se debe escoger el más apropiado para el conjunto de datos y según la naturaleza del problema.

Todos los criterios de vecindad deben cumplir que, al seleccionar una unidad o área, el resto de ellas queden particionadas en dos conjuntos mutuamente excluyentes, uno compuesto por unidades vecinas y otro por las que no lo son.

Una característica importante a considerar en los criterios de vecindad, es la simetría. Sea A un área determinada y según el criterio que se está utilizando la unidad B es vecina suya. Si el criterio utilizado es simétrico, entonces B también tendrá como vecina a A. Si el criterio no fuera simétrico, B no necesariamente tendrá a A entre el conjunto de sus unidades vecinas.

Los criterios de vecindad más utilizados y aceptados por la bibliografía son:

* Vecinos por contigüidad. Se define como áreas vecinas a aquellas en las que para ir de una a otra no haya que pasar por una tercera, es decir, que estén contiguas en el mapa. Existen tres criterios de vecindad basados en contigüidades: Reina, Torre y Alfil, reciben estos nombres porque las unidades vecinas son aquellas a las que se accede según el movimiento de las piezas en el tablero de ajedrez.
  + Reina: en el ajedrez puede moverse a lo largo de la fila, la columna y las diagonales de la casilla en que se encuentre. Extrapolando esos movimientos a la situación de interés, dos áreas serán vecinas si tienen al menos un punto común.
  + Torre: solo puede moverse a lo largo de la fila y la columna en que se encuentre, no puede moverse en diagonal. Análogamente se considera que dos áreas son vecinas si tienen más de un punto en común.
  + Alfil: solo puede moverse a lo largo de la diagonal de la casilla en la que se encuentre. De esta manera, dos unidades en el espacio serán vecinas si y solo si tienen tan solo un punto en común.

Este criterio cumple la condición de simetría ya que, si un área tiene un punto o más en común con una segunda, esta segunda también tendrá un punto o más en común con la primera.

* Vecinos basados en la distancia euclídea. Este método considera vecinas dos áreas si cumplen cierta condición referente a la distancia que las separa. De aquí en adelante se refiere a la distancia entre dos áreas como la distancia entre sus centroides. Existen dos variantes:
  + Los k vecinos más cercanos. Se calcula la distancia de la unidad considerada a todas las demás: serán vecinas las k unidades cuyas distancias sean las menores. El número k se determina en base a la naturaleza de cada problema. Será una relación asimétrica en la que todas las áreas tendrán el mismo número de vecinos.
  + Dos áreas serán vecinas si y solo si la distancia entre ellas es menor a una magnitud fijada a priori. Este método es adecuado cuando las áreas tienen una distancia similar entre ellas, ya que si hay una distancia mucho mayor a las otras se presentará el problema de dejar esta unidad sin vecinos, o considerar un número de vecinos demasiado alto en el resto de áreas. Esta relación es simétrica.

Una vez definido el criterio de vecindad a utilizar, resulta de interés cuantificar la fuerza de cada relación, esto es lo que se conoce como pesos espaciales. Por ejemplo, se sabe que A tiene dos áreas vecinas B y C, pero lo que se desconoce es si ambas influyen de la misma manera. En esta sección se explicarán los distintos criterios que pueden adoptarse a la hora de definir los pesos espaciales.

Los pesos se representan de forma matricial mediante la matriz cuadrada W. Cada elemento wij representa el peso de la relación de vecindad entre las áreas i y j. Cuando wij = 0 las unidades no son vecinas. La diagonal de la matriz de conectividad W será 0 ya que, por convención, una unidad no se considera vecina de ella misma Por lo que W es una matriz cuadrada con todos sus elementos mayores o iguales a 0.

Esta matriz será simétrica si el criterio utilizado para definir los vecinos y los pesos lo son. Los pesos resultan simétricos cuando un área A ejerce sobre B la misma influencia que B sobre A. Un ejemplo de una situación en la que tiene sentido utilizar una relación asimétrica es considerar la influencia de las ciudades grandes sobre los pueblos de alrededor. Las grandes ciudades influyen más en las características de los pueblos que a la inversa.

Dentro de todos los posibles criterios para asignar los pesos, se distinguen dos grandes grupos; aquellos donde por el simple hecho de ser vecinos cada unión tendrá un peso común y aquellos en la que la importancia de las uniones variará en base a ciertas características.

* Binario: Es el criterio más sencillo, asume que wij = 1 cuando i y j son unidades vecinas y wij = 0 cuando no lo son. Este método es el más utilizado cuando existe poca información del proceso espacial. Bajo este criterio la suma de los pesos de un área es el número de vecinos que tiene.
* Estandarización por filas: Este método se basa en que los pesos de cada fila de la matriz sumen 1. Para ello se divide la unidad entre el número de áreas vecinas que posee el área considerada. Según este método los pesos de áreas con pocos vecinos serán mayores que los de áreas con un número de vecinos mayor. Es decir, cada vecino de un área con pocos vecinos ejerce gran influencia sobre ella, mientras que los vecinos de áreas con muchos vecinos ejercen una influencia menor.
* Estandarización Global: considera el mismo peso para todas las relaciones, definiendo el peso como el cociente entre la unidad y el número total de vecinos. De esta manera, la suma de todos los pesos será igual a uno.
* Estabilización de variancia: Los pesos de áreas con muchos vecinos varían mucho al utilizar un criterio u otro, de los mencionados anteriormente. Lo que busca este método es reducir esta variación. Puede pensarse como un híbrido del método de estandarización por filas y el de estandarización global. Será siempre asimétrico, pero si el criterio de vecindad es simétrico la matriz W estará bastante cerca de ser simétrica.

**3.2.2 El índice de Moran (I)**

Este es el índice más usual. Fue desarrollado por Moran en 1950 y en la casi totalidad de los programas geoestadísticos se incluye su cálculo. Su interpretación es sencilla ya que es similar a la del coeficiente de correlación de Pearson. Se conocen sus propiedades estadísticas y pueden hacerse pruebas de hipótesis sobre su significación estadística. Sin embargo, cuando los tamaños de las unidades son diferentes y la variable para la cual se quiere evaluar la correlación espacial es una razón, este índice puede conducir a resultados erróneos.

Se considera una región R dividida en m áreas ri, i=1…, m. Por ejemplo, si la región es la ciudad de Rosario, las áreas o unidades podrían ser los radios censales.

Sea xi el valor de una variable que refleja el tamaño del área i, por ejemplo, el total de hogares en el radio censal i.

Sea ni el valor de la variable de interés en el área i, en el primer problema que se estudia es el número de hogares con NBI en el radio censal i.

La razón observada en el área i se define como pi=

El índice de Moran para la razón pi, se define como (Moran, 1950):

I = ∀ *i≠j*, donde

*pi* es el valor de la razón en la unidad ia la que se le asocia el conjunto de coordenadas **s***i*, vector cuyas componentes son las coordenadas espaciales,

= es la media de la razón *pi*,

*wji* es el elemento i,j de la matriz de conectividad que recoge la relación de vecindad entre las unidades *i* y *j* definida de la siguiente manera:

=

El índice de asociación de Moran resume la intensidad y dirección de la dependencia entre los valores de una variable observados en distintas unidades del espacio, calculando los productos cruzados entre los valores de pares de unidades y ponderando por una medida de la relación de vecindad entre las unidades de cada par. Por lo tanto, *I* puede considerarse como una medida de correlación de cada *pi* con el resto de las áreas con las que se encuentra vinculada. Al igual que el índice de correlación de Pearson, varía entre -1 y 1, y E[I]= bajo la hipótesis de aleatoriedad espacial.

Un coeficiente *I* mayor que su valor esperado indica autocorrelación espacial positiva, mientras que un valor inferior a E[I] pone de manifiesto la existencia de autocorrelación espacial negativa. Un valor cercano a E[I] (la cual tiende a 0 cuando *n* crece) indica ausencia de autocorrelación.

Para probar la significación estadística del índice *I* y así comprobar la hipótesis de no autocorrelación espacial se puede utilizar un test de hipótesis basado en supuestos de normalidad.

Bajo la hipótesis nula de que no existe autocorrelación espacial y si p*i* ∼ *N*(*µ, σ*2) o si mes suficientemente grande, la estadística Z = sigue una distribución normal estándar donde:

* E[I] =
* Var[I] = -

Cuando no se cumple el supuesto de normalidad de pi se utiliza un test permutacional: se encuentran las m! posibles configuraciones de las unidades asumiendo que sus valores son aleatorios y sobre cada una de ellas se calcula el valor de I, para luego calcular la probabilidad asociada a la hipótesis de aleatoriedad.

Si m es grande, el trabajo computacional puede volverse dificultoso. Un recurso utilizado en estos casos, es un test basado en el Método de Montecarlo, que consiste en la realización de un test permutacional, pero sólo considerando un subconjunto de configuraciones. Los softwares suelen utilizar 999 permutaciones, y considerando la muestra observada, resultan 1000 valores de I para construir la distribución de aleatorización.

**Efectos de unidades de diferentes tamaños**

Si se considera una región dividida en áreas con tamaños diferentes, como lo son en los problemas que se estudian en esta tesina, al calcular la razón observada en cada unidad se estarían utilizando cantidades con diferentes denominadores en el cálculo de los índices.

Por ejemplo, si se desea estudiar la autocorrelación espacial para la variable razón de hogares con necesidades básicas insatisfechas (NBI) observada en los radios censales de la ciudad de Rosario, puede ocurrir que un radio censal con 100 hogares tenga 10 con necesidades básicas insatisfechas, y otro radio censal con 10000 hogares tenga 1000 con NBI; en ambos casos, la razón de hogares con NBI es 0,10 pero evidentemente la situación es diferente.

Si se utiliza el índice de Moran con las razones, no hay distinción alguna entre estas dos situaciones al momento de hacer los cálculos, es decir no se tiene en cuenta el “tamaño” de las correspondientes áreas.

Teniendo en cuenta el supuesto de normalidad cuando las unidades son de diferentes tamaños, las áreas de menor “tamaño” tienen razones más variables y, por lo tanto, es más probable que asuman un valor extremo.

Se han estudiado los efectos de las desviaciones del supuesto de normalidad en las pruebas de autocorrelación espacial. La P(e1) en estos casos se incrementa cuando los “tamaños” en las áreas son heterogéneos y la razón es constante (Walter, 1992).

Assunção et al. (1999) estudia el efecto de “tamaños” de unidades heterogéneos sobre la potencia del test, es decir evalúa el impacto de la variación de los “tamaños” de las áreas cuando existe una correlación espacial entre las razones.

**3.2.3 El índice de Oden ()**

Cuando las unidades son de diferente tamaño Oden, (1995) propuso un ajuste al índice de Moran. El conocido como Índice de Oden es:

=

Donde:

n = total de la variable en estudio en la región,. Por ejemplo, el total de hogares con NBI en la ciudad de Rosario.

x = es el total en la región de la variable utilizada como denominador, siguiendo con el ejemplo: total de hogares en la ciudad de Rosario.

, razón de interés en la región.

, razón con respecto al total en la región de la variable en estudio en la unidad i.

razón con respecto al total en la región de la variable utilizada como denominador en la unidad i.

Se llama al peso espacial definido por Oden de la siguiente manera:

Mij =

Como puede verse los elementos de la diagonal principal de la matriz de vecindad tendrá como valor 2, a diferencia de las matrices de vecindad ya vistas donde la diagonal está compuesta por elementos iguales a 0. Oden propone esta modificación, de alguna manera intentando diferenciar las situaciones de dos áreas que no son vecinas a aquella cuando se compara un área consigo misma.

La cantidad incluida en el índice de Oden es:

Al igual que Moran, el índice de Oden puede obtenerse bajo el supuesto de normalidad o mediante un test permutacional.

Bajo la hipótesis nula de inexistencia de autocorrelación espacial o si m es suficientemente grande, la estadística Z = se distribuye como una variable aleatoria normal estándar con E[] = y la variancia de puede obtenerse mediante la igualdad Var[] = E[] – E2[], siendo:

E[] =, donde los valores que se encuentran en el exponente entre paréntesis denotan factoriales descendientes nk = n (n - 1) (n – 2)…(n – k + 1)

El cálculo de E[] es complejo pero puede obtenerse mediante los términos mencionados anteriormente, definidos por Oden de la siguiente manera:

* S0 = x2A – xB
* S1 = x2C/2 – 2xD
* S2 = x3 E – 4x2F + 4Xd, donde:
* A =
* B =
* C =
* D =
* E =
* F =

Oden (1995) muestra que el test que utiliza a es más potente que la prueba asociada a I cuando los tamaños de las unidades espaciales consideradas son distintos. Esta capacidad de capturar la variabilidad se debe al primer término en el numerador, que es una versión espacial de la prueba chi-cuadrado convencional para la heterogeneidad de proporciones.

**Las hipótesis estadísticas para los índices de Moran y Oden**

A pesar de la mayor potencia obtenida por el índice de Oden con respecto al índice de Moran, Assunção et al. (1999) advierte sobre ciertas particularidades que posee este índice, las cuales se deben considerar en caso de decidir su utilización.

La consideración más importante es lo inapropiado de la comparación entre las pruebas de hipótesis que se realizan con I e .

Para realizar la comparación entre los test se definen tres estados con respecto a la configuración espacial de las razones de las áreas:

1. Razones espaciales homogéneas o constantes
2. Razones heterogéneas sin correlación espacial
3. Razones heterogéneas correlacionados espacialmente

La prueba de existencia de correlación espacial que propone Moran considera las siguientes hipótesis nula y alternativa:

H0) A U B H1) C

Mientras que la prueba propuesta por Oden plantea las siguientes hipótesis:

H0) A H1) B U C

Como se muestra en la Figura 1.

**Tabla n°1**: Hipótesis probadas por Moran y Oden.

|  |  |  |
| --- | --- | --- |
| Índices/Hipótesis | H0 | H1 |
| I | A U B | C |
|  | A | B U C |

Puede apreciarse que el estado B, razones heterogéneas sin correlación espacial, en el índice de Moran se acepta como parte de la H0 y en el índice de Oden cuando ella se rechaza. El test basado en el índice de Oden conduce a aceptar la existencia de correlación espacial a una situación en la que no hay (B).

En consecuencia no sorprende que el test basado en tenga mayor potencia, especialmente en estados como B, frente a los cuales el índice de Moran debería tener como máxima potencia la probabilidad de error de tipo I (Assunção et al., 1999).

A medida que se fueron desarrollando las investigaciones correspondientes que desencadenaron en las consideraciones mencionadas anteriormente con respecto al índice de Oden, este mismo dejó de utilizarse.

De esta manera, resurgió la necesidad de encontrar un índice que tenga en cuenta el tamaño de las distintas áreas consideradas de una región a la hora de determinar si existe autocorrelación espacial de una variable aleatoria.

**3.2.4 Empirical Bayes Index (EBI)**

Una nueva alternativa que se considera para probar la autocorrelación espacial, es el Empirical Bayes Index (EBI). Este indicador fue desarrollado por Assunção et al. y fue presentado en el año 1999. Estos autores, en el artículo en el que presentan la nueva propuesta, realizan comparaciones con el índice de Moran empleando simulaciones por Monte Carlo y estimando la potencia de las pruebas de independencia espacial en diferentes escenarios llegando a mostrar el mejor desempeño del EBI.

Sean θ1, θ2, …, θi, ..., θm, donde i=1…, m, las razones de interés desconocidas de las áreas en estudio. Se realiza el supuesto de que el número de eventos observados ni (durante un período de referencia) tiene una distribución Poisson[[1]](#footnote-1) con media condicional E(ni|θi) = Var(ni|θi) = xi θi siendo xi el “tamaño” poblacional del área i. De esta forma, la media condicional de la razón estimada pi es E(pi|θi) = θi y su variancia condicional es igual a Var(pi|θi) = , por lo tanto las razones estimadas poseen distintas medias y variancias condicionales.

Adoptando un enfoque bayesiano, se realiza el supuesto de que las razones θi tienen a priori una esperanza y variancia igual a β y α respectivamente. De esta manera, la esperanza marginal de pi es β y su variancia marginal es α + (Assunção et al., 1999). Dicho de otra manera, las áreas poseen la misma esperanza marginal y sólo las variancias difieren entre las unidades, las cuales se incrementan a medida que los **“**tamaños” de las áreas disminuyen.

Para estimar los parámetros α y β desconocidos, Marshall (1991) propone utilizar el método de los Momentos, el cual conduce a los siguientes resultados:

= a = s2 - = b = ,

dónde s2 = , x y n son el “tamaño” de la región y el total de la variable de interés en la región respectivamente.

Por lo tanto, la esperanza y variancia marginales de pi son estimadas por b y = a + , respectivamente. Por convención, si < 0, se define = .

En lugar de utilizar las razones estimadas pi (como se emplean en el índice de Moran), se propone un nuevo índice que toma las razones estandarizadas, utilizando las estimaciones presentadas anteriormente.

=

El índice Empírico de Bayes (EBI) se define de la siguiente manera:

EBI =

Al igual que el índice de Moran, el EBI será significativamente distinto de cero si las razones están correlacionadas espacialmente. La prueba de independencia espacial depende de la distribución bajo la hipótesis nula del EBI, la cual se puede obtener mediante permutaciones.

Por lo tanto, se permuta independientemente el vector y se asignan aleatoriamente a las áreas una determinada cantidad de veces (en general, se utilizan 999 permutaciones al igual que en Moran). Para cada una de las combinaciones obtenidas se calcula el EBI. El valor de la probabilidad asociada al test de hipótesis está dado por el cociente entre la cantidad de veces que el EBI permutado excede el EBI observado (numerador) y la cantidad de permutaciones utilizadas (denominador).

**3.3 Problemas de aplicación**

Para poder observar los valores resultantes de los índices estudiados se consideran dos problemas de aplicación: el estudio de la distribución espacial de los Hogares con NBI en la ciudad de Rosario en el año 2010 y Heridos de armas de fuego en la ciudad de Rosario durante un determinado período.

Es oportuno que las comparaciones entre los índices de Moran, Oden y el EBI, se realicen sobre una región formada por unidades de “tamaños” distintos, tal como lo son los radios censales de la ciudad de Rosario, motivo principal por lo que se escogieron dichas aplicaciones.

Para el primer problema, se dispone de un archivo de datos georreferenciado que contiene entre otras variables el número de hogares con NBI y el total de hogares para cada radio censal de la ciudad de Rosario en el año 2010, obtenido del sitio web de la Sociedad Argentina de Estadística (SAE).

En cuanto al segundo problema a considerar, se cuenta con un conjunto de datos ficticios que corresponden al número de Heridos de armas de fuego en la ciudad de Rosario en un determinado año. Este problema fue tratado debido a que, en el artículo principal referenciado en la presentación del EBI, se aplica este indicador al mismo problema en la ciudad de Belo Horizonte, Minas Gerais, Brasil.

En base a los distintos tipos de datos espaciales que fueron estudiados en la presente tesina, es importante destacar que los conjuntos de datos considerados se corresponden con datos reticulares irregulares, donde cada una de las áreas o agregaciones espaciales se corresponden con los radios censales y la región es la ciudad de Rosario.

A la hora de determinar la matriz de conectividad, se adoptará el criterio de vecindad por contigüidad, y más específicamente el de tipo Reina, ya que parecería ajustarse de una manera más adecuada a la naturaleza de los problemas que se consideran. Por otro lado, se empleará el criterio de Estabilización de variancia para determinar el peso de cada una de las relaciones entre vecinos.

El software que se utiliza para realizar la presente tesina es R y en particular los paquetes sp y spdep. Se desarrolla un programa para el cálculo del índice de Oden, ya que no se encontró en ningún paquete de R ni en otro software.

1. **Aplicación y Resultados**

En este capítulo se presentan los resultados hallados, como consecuencia de aplicar los índices desarrollados en el capítulo Materiales y Métodos sobre los problemas considerados: Hogares con Necesidades Básicas Insatisfechas (NBI) en el año 2010 y Heridos de armas de fuego, ambos en la ciudad de Rosario.

El segundo caso de aplicación se trata de un escenario ficticio, cuya motivación se debe a que el en el artículo referenciado en la presentación del EBI (Asuunção, 1999) se aplica este indicador a un problema equivalente en la ciudad de Belo Horizonte, Minas Gerais, Brasil.

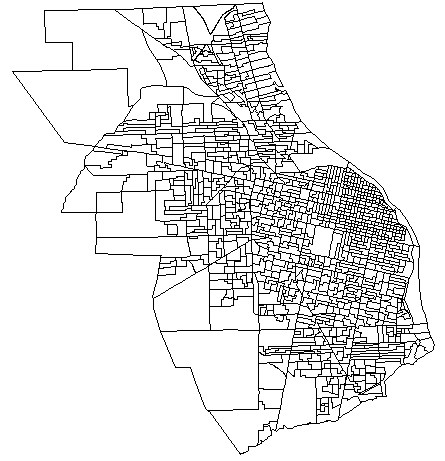
Se analizará el comportamiento de los índices de Moran, Oden y el EBI sobre los problemas mencionados, y en base a los resultados observados se escogerá cuál de los tres índices es el más acorde.

* 1. **Vecindad**

Una decisión fundamental previa a calcular algún índice de autocorrelación espacial es la elección de un criterio para saber cuándo dos unidades serán vecinas, y otro método para determinar el peso de esas relaciones.

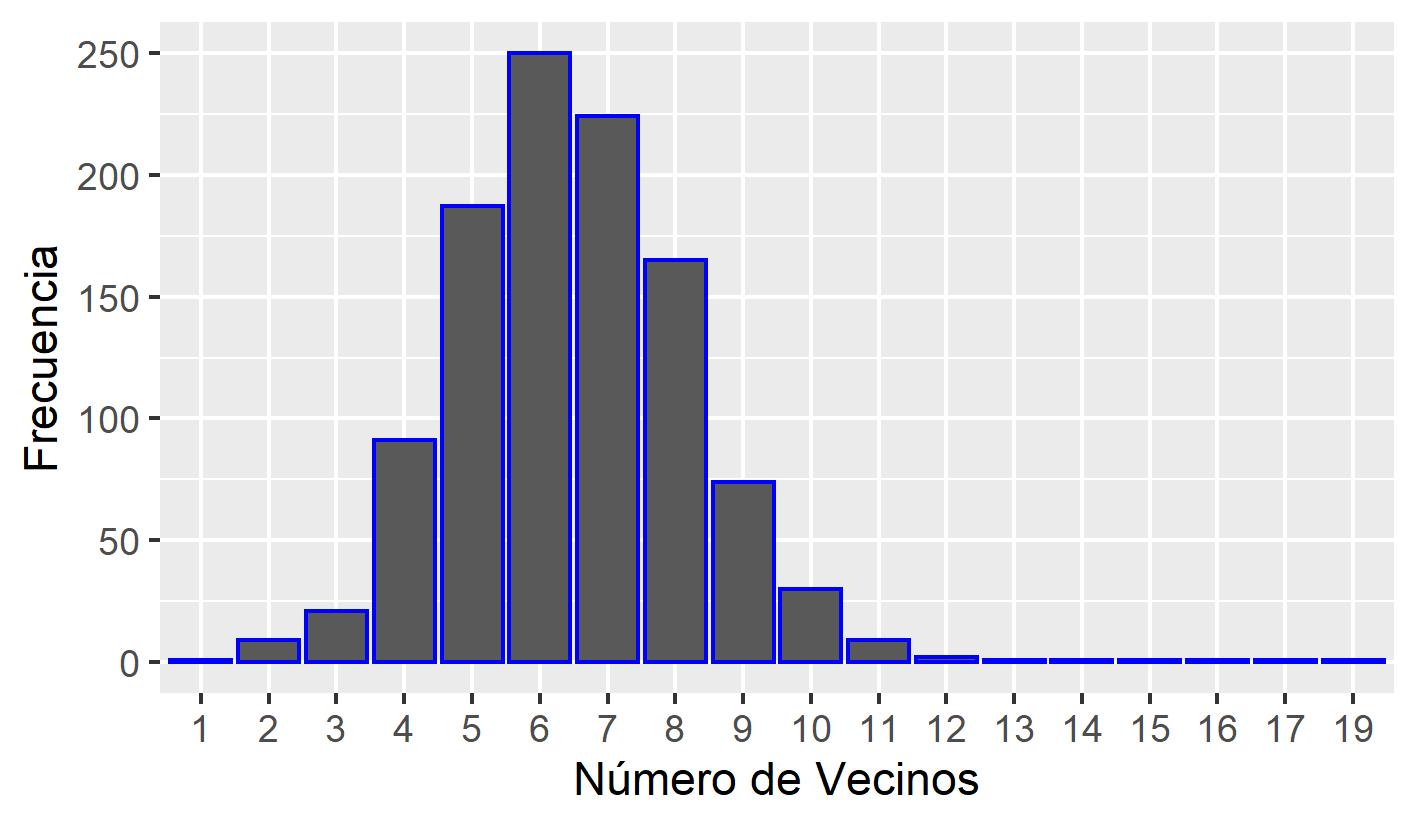
Se utiliza un criterio de vecindad por contigüidad de tipo Reina para determinar las vecindades entre radios censales y un criterio de Estabilización de variancias para determinar la fuerza de las relaciones entre vecinos, de esta manera se construye la matriz de vecindad, la cual es la misma en ambas aplicaciones, ya que las áreas consideradas para ambos problemas son los radios censales de la ciudad de Rosario.

**Figura 4.1**: Mapa de los radios censales de la ciudad de Rosario.

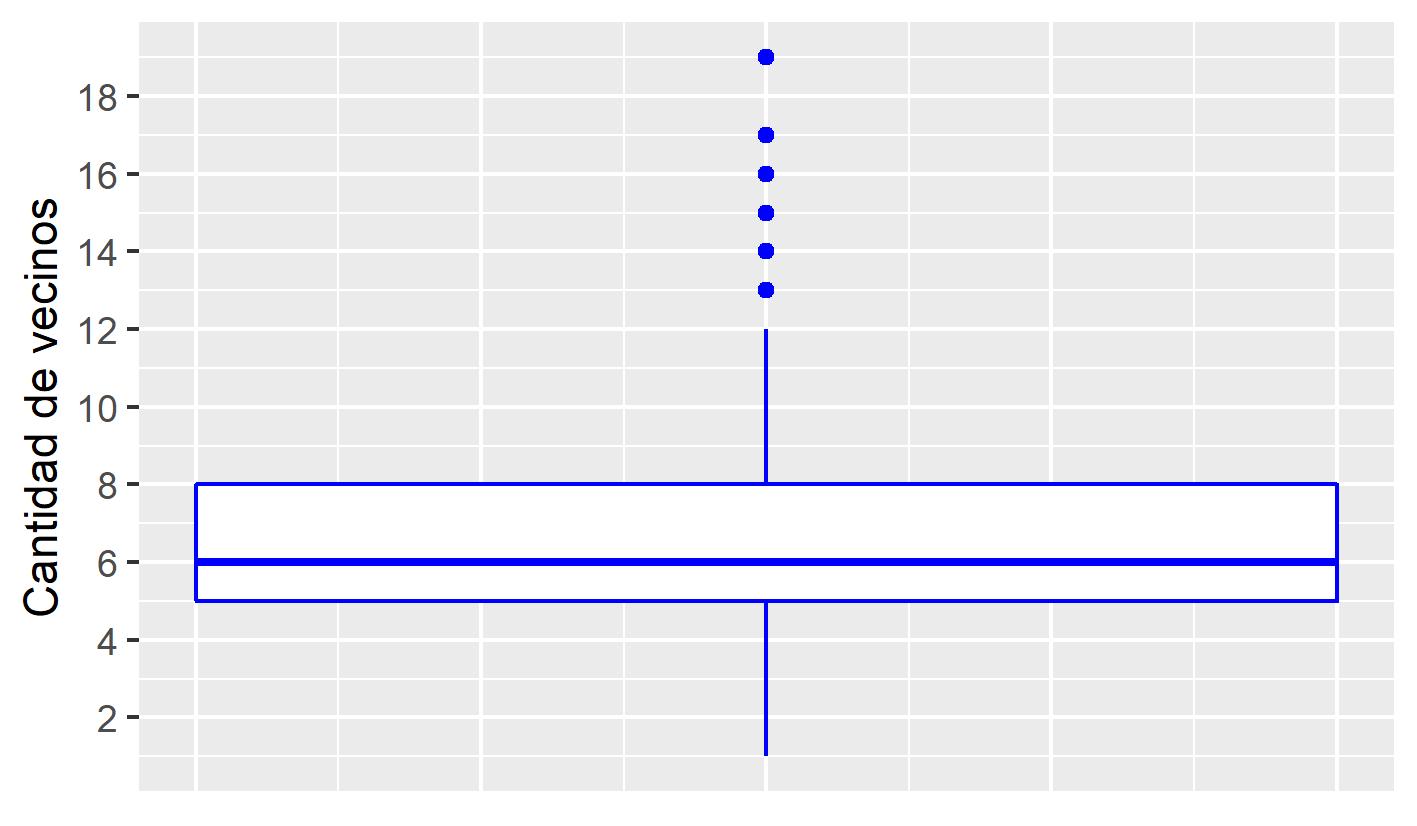


La figura 4.1 permite observar los 1069 radios censales de la ciudad de Rosario, cada par de unidades que compartan al menos un punto en el espacio serán vecinos, ya que así lo establece el criterio de vecindad por contigüidad del tipo Reina.

**Figura 4.2.a**: Distribución del número de vecinos de los radios censales de la ciudad de Rosario.



**Figura 4.2.b**: Boxplot de la cantidad de vecinos de los radios censales de la ciudad de Rosario.



La figura 4.2.a permite apreciar la cantidad de vecinos que posee cada radio censal de la ciudad de Rosario, con una distribución aproximadamente simétrica, donde gran parte de las áreas poseen entre 5 y 8 vecinos. La media observada es cercana a 6,5 y la desviación estándar es aproximadamente igual a 1,8.

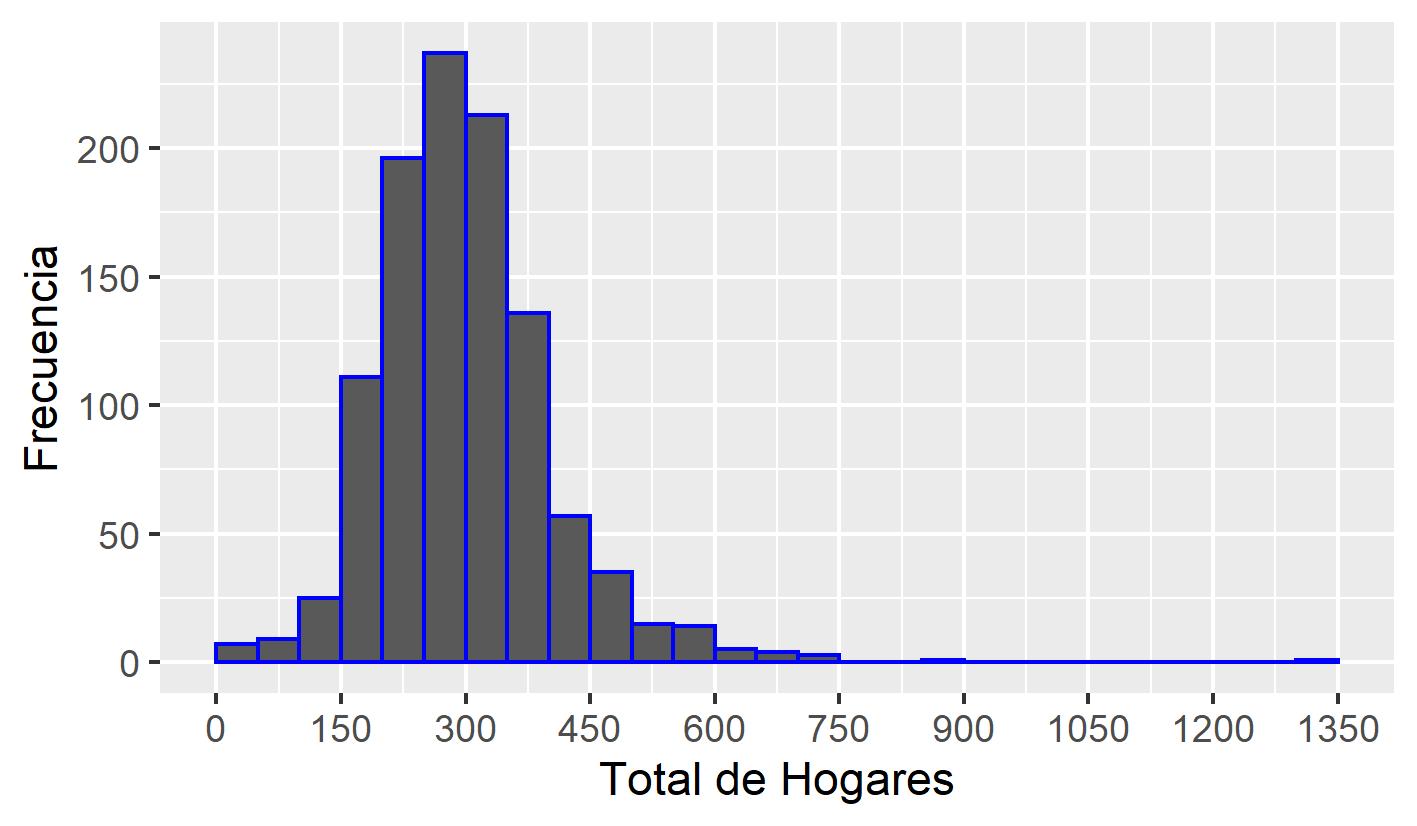
El boxplot contenido en la figura 4.2.b muestra que los radios censales poseen una cantidad mediana de vecinos igual a 6 y los valores que acumulan el 25% y 75% de la distribución de los datos son 5 y 8 respectivamente.

* 1. **Resultados en Hogares con NBI**

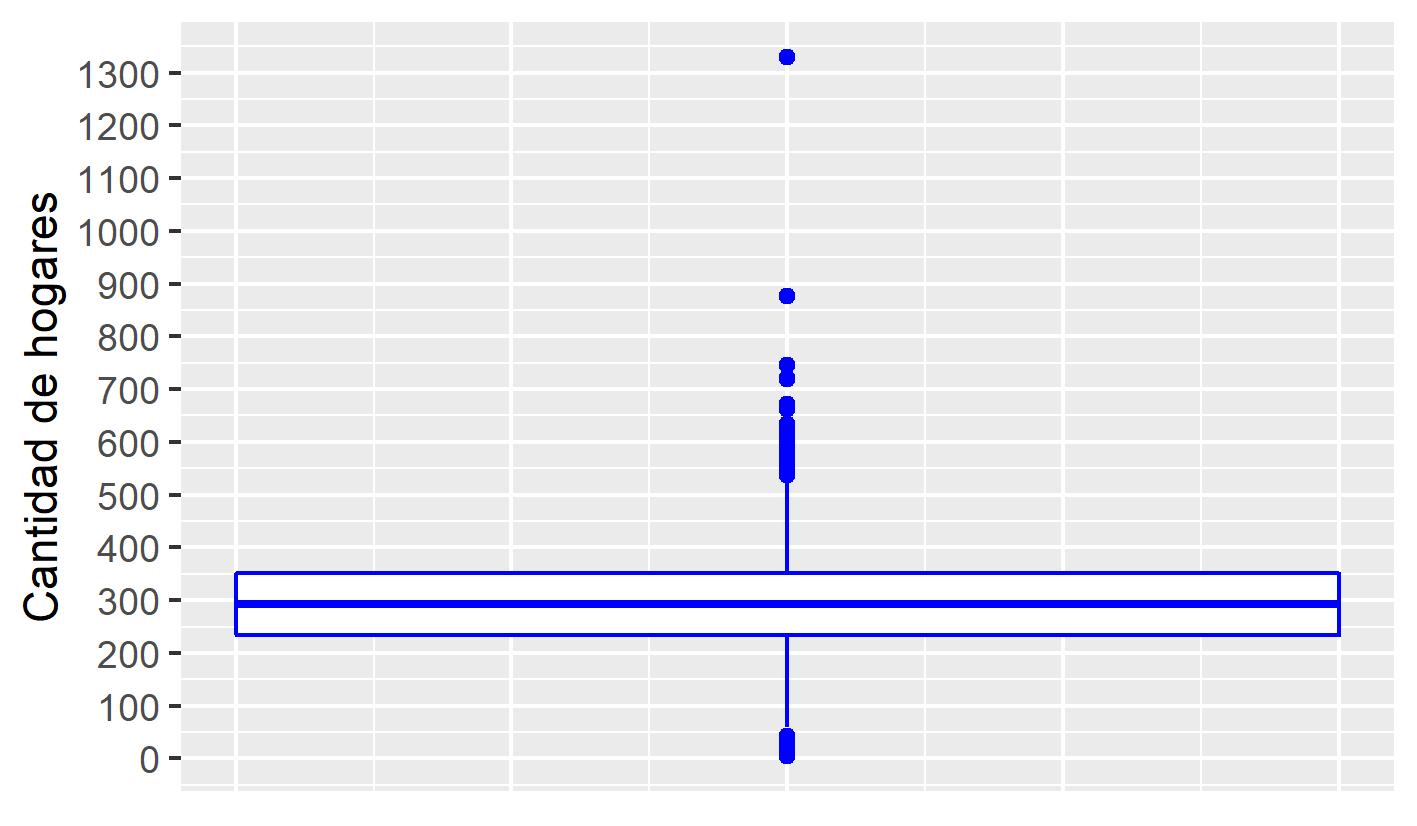
Para el cálculo de los índices propuestos se define la proporción de Hogares con NBI como el cociente compuesto por:

* Numerador: Cantidad de Hogares con NBI en un determinado radio censal de la ciudad de Rosario
* Denominador: Cantidad de Hogares en un determinado radio censal de la ciudad de Rosario.

**Figura 4.3.a**: Distribución de la cantidad de hogares en los radios censales de la ciudad de Rosario.



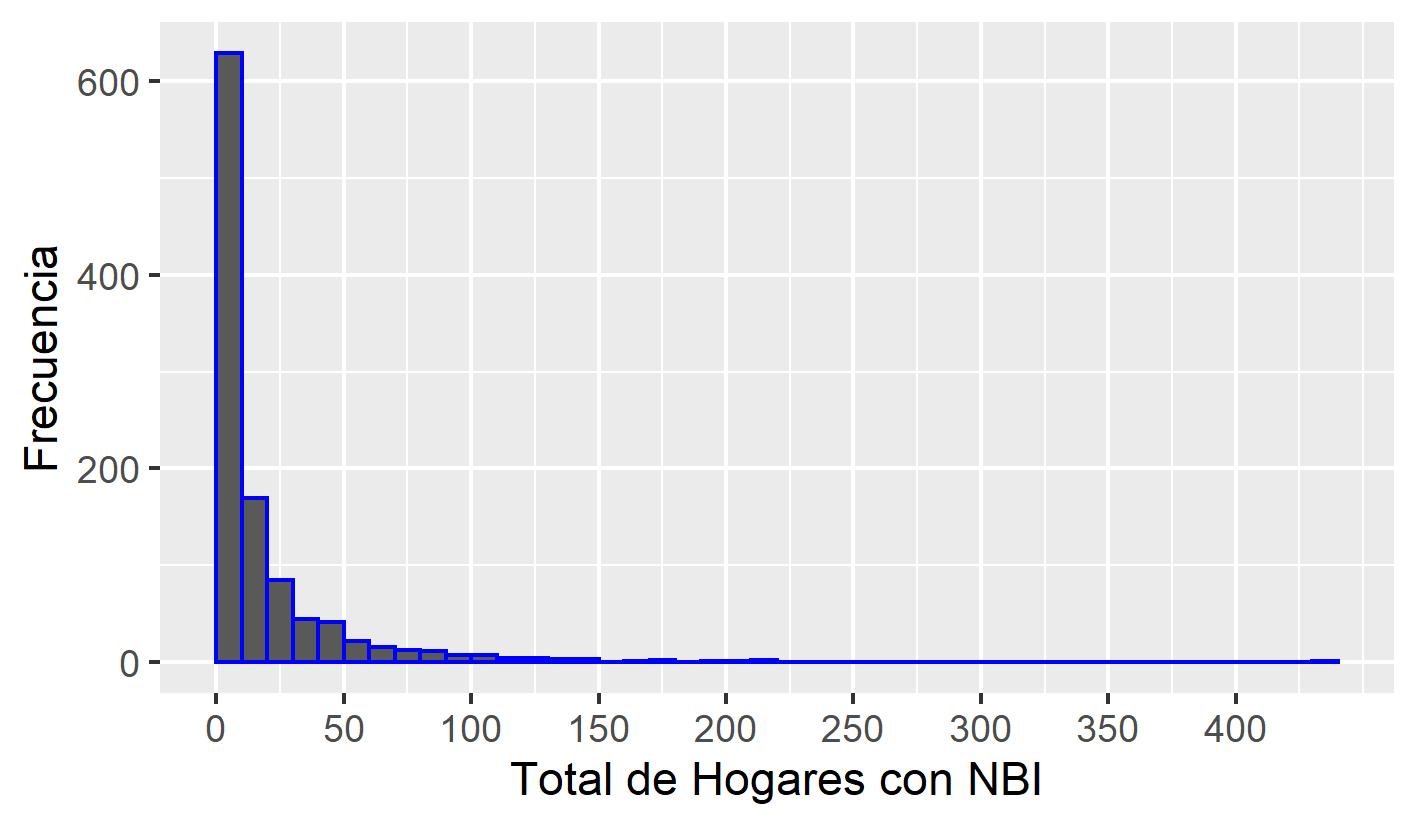
**Figura 4.3.b**: Boxplot de la variable total de hogares en los radios censales de la ciudad de Rosario.



La figura 4.3.a permite apreciar que la cantidad de hogares en los radios censales de la ciudad de Rosario se comporta de una manera aproximadamente simétrica, con una leve cola por derecha. Su valor medio y desviación estándar son aproximadamente iguales a 300 y 107 respectivamente.

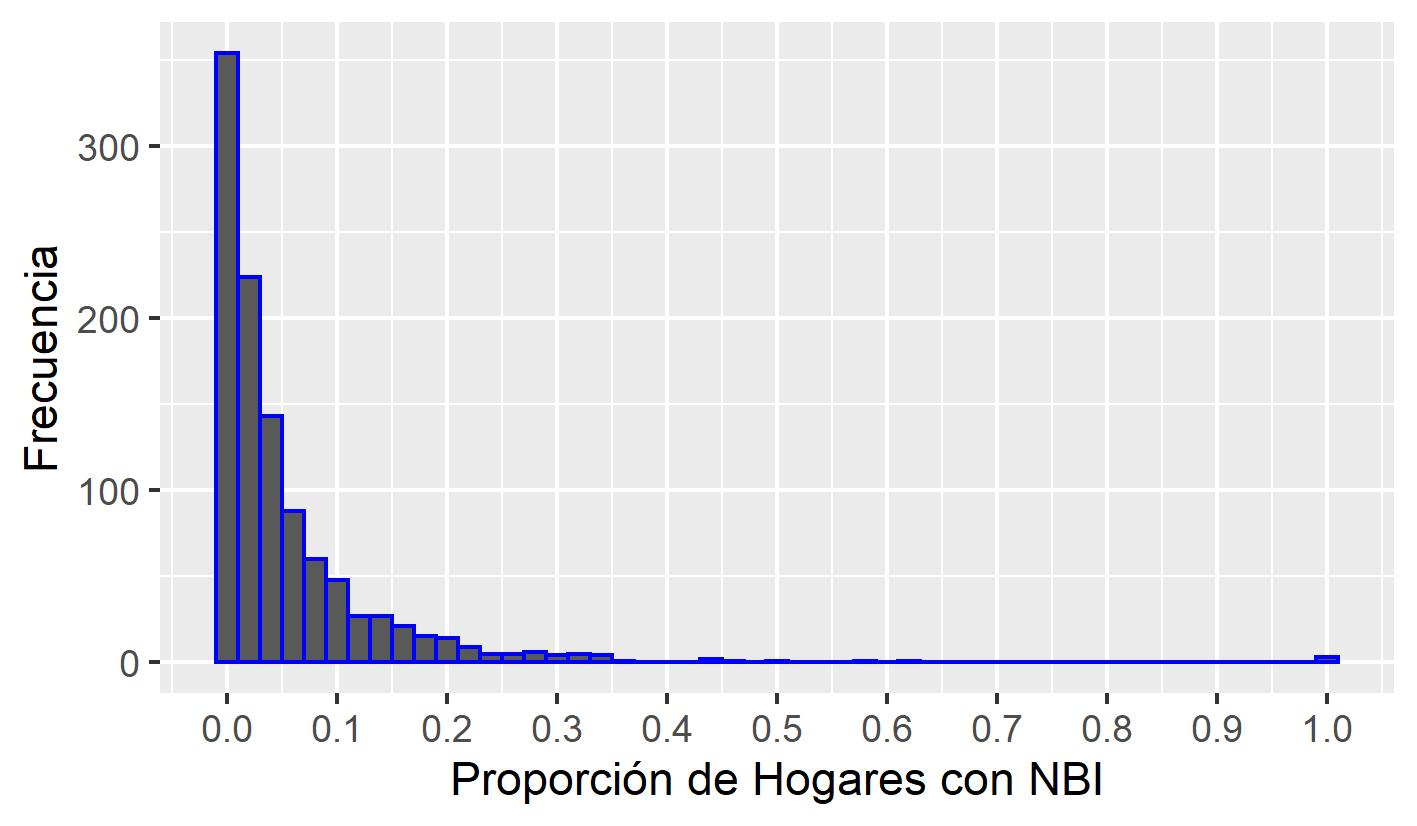
El diagrama de cajas y bigotes de la figura 4.3.b muestra que la mediana de la cantidad de hogares que tienen los radios censales de la ciudad de Rosario es 292, también puede identificarse que el 50% de los radios censales poseen entre 233 y 352 hogares.

**Figura 4.4**: Distribución de la cantidad de hogares con NBI en los radios censales de la ciudad de Rosario.



La figura 4.4 permite ver que la cantidad de hogares con NBI en los radios censales de la ciudad de Rosario posee una pronunciada cola derecha, con media cercana a 7 y un desvío estándar aproximadamente igual a 31.

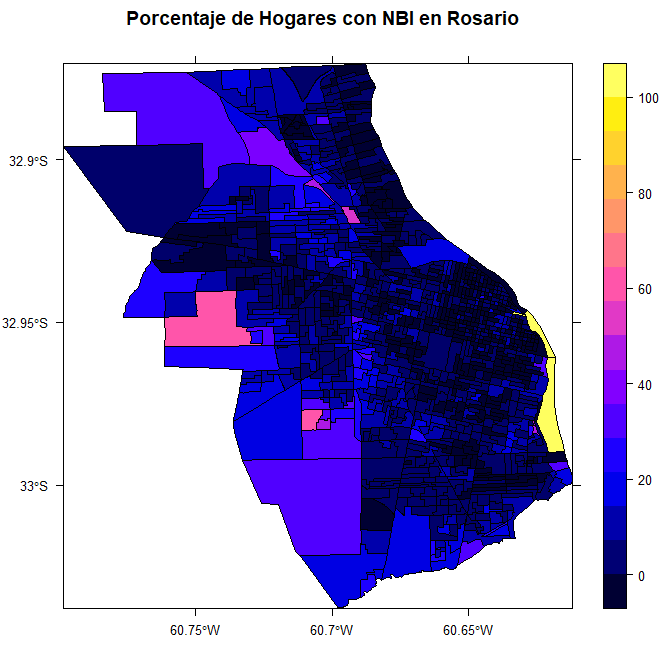
**Figura 4.5**: Distribución de la proporción de hogares con NBI en los radios censales de la ciudad de Rosario.



Al igual que la cantidad de hogares con NBI, se aprecia que la proporción de hogares con NBI en los radios censales de la ciudad de Rosario posee una notable asimetría por derecha, con media cercana a 0,025 y un desvío estándar aproximadamente igual a 0,088 (figura 5.5).

Resulta de interés observar el comportamiento espacial de la proporción de hogares con NBI, ya que es la variable de interés principal sobre la cual se aplicarán los indicadores de autocorrelación espacial estudiados a lo largo de esta tesina.

**Figura 4.6**: Mapa de calor del porcentaje de hogares con NBI a lo largo de los radios censales de la ciudad de Rosario.



La figura 4.6 permite detectar cuales son los radios censales con mayor proporción de hogares con NBI, los radios censales que poseen los peores valores de este indicador se encuentra en la zona sureste (color amarillo).

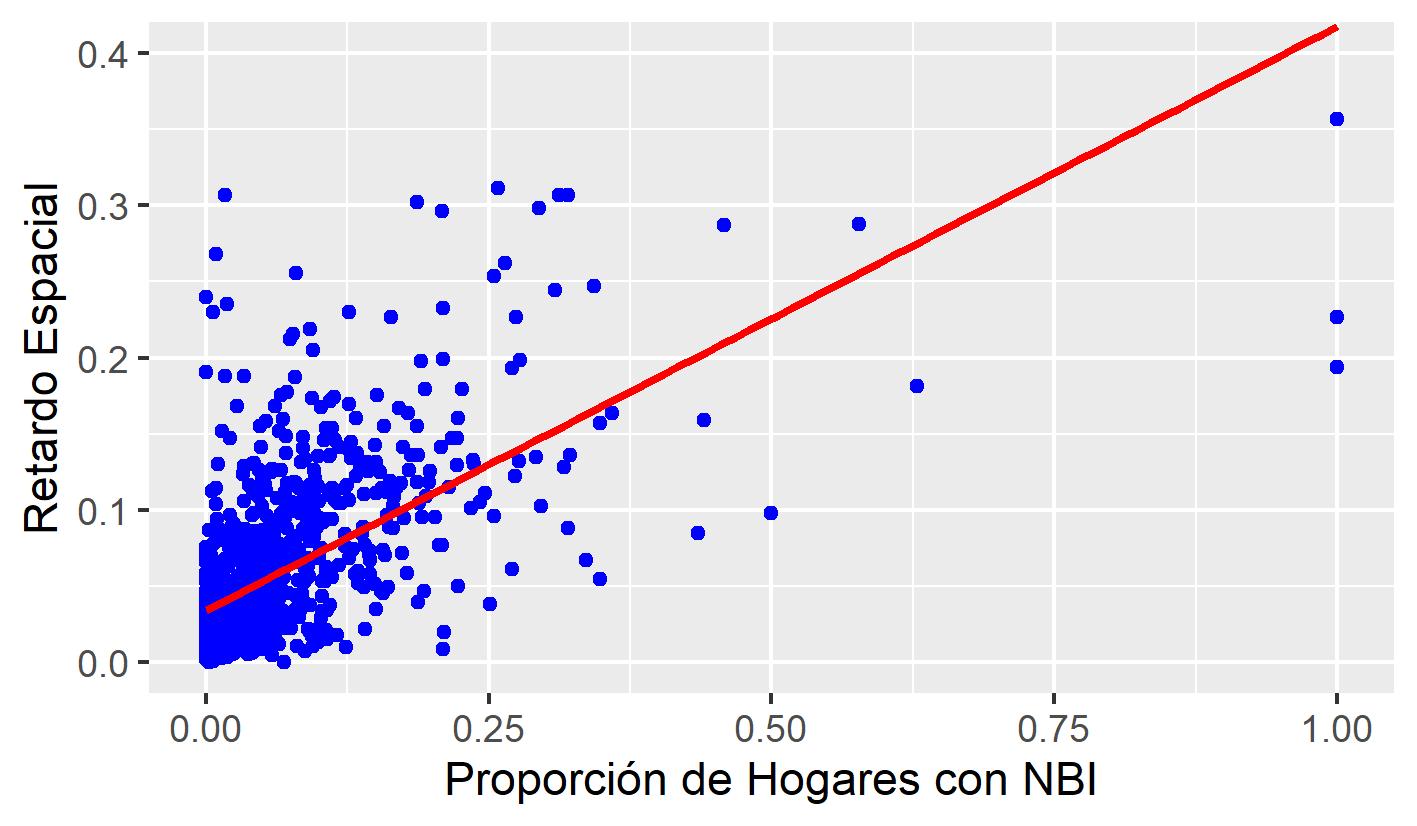
A priori, al observar el mapa de calor, se podría pensar que existe una dependencia espacial de los radios censales con respecto a la variable analizada, pero para comprobarlo científicamente y obtener una medida que lo cuantifique se procede a calcular los índices de autocorrelación espacial estudiados en la presente tesina.

**Tabla 4.1**: índices de autocorrelación espacial calculados para la proporción de hogares con NBI.

|  |  |  |
| --- | --- | --- |
| **Índice** | **Estadístico** | **P-Valor** |
| Moran (I) | 0,38527 | 0,001 |
| Oden () | 0,08890 | <0,001 |
| EBI | 0,41943 | 0,001 |

La tabla 4.1 contiene la magnitud del índice de Moran, Oden y el EBI junto con sus correspondientes probabilidades asociadas a la prueba de hipótesis de ausencia de autocorrelación espacial.

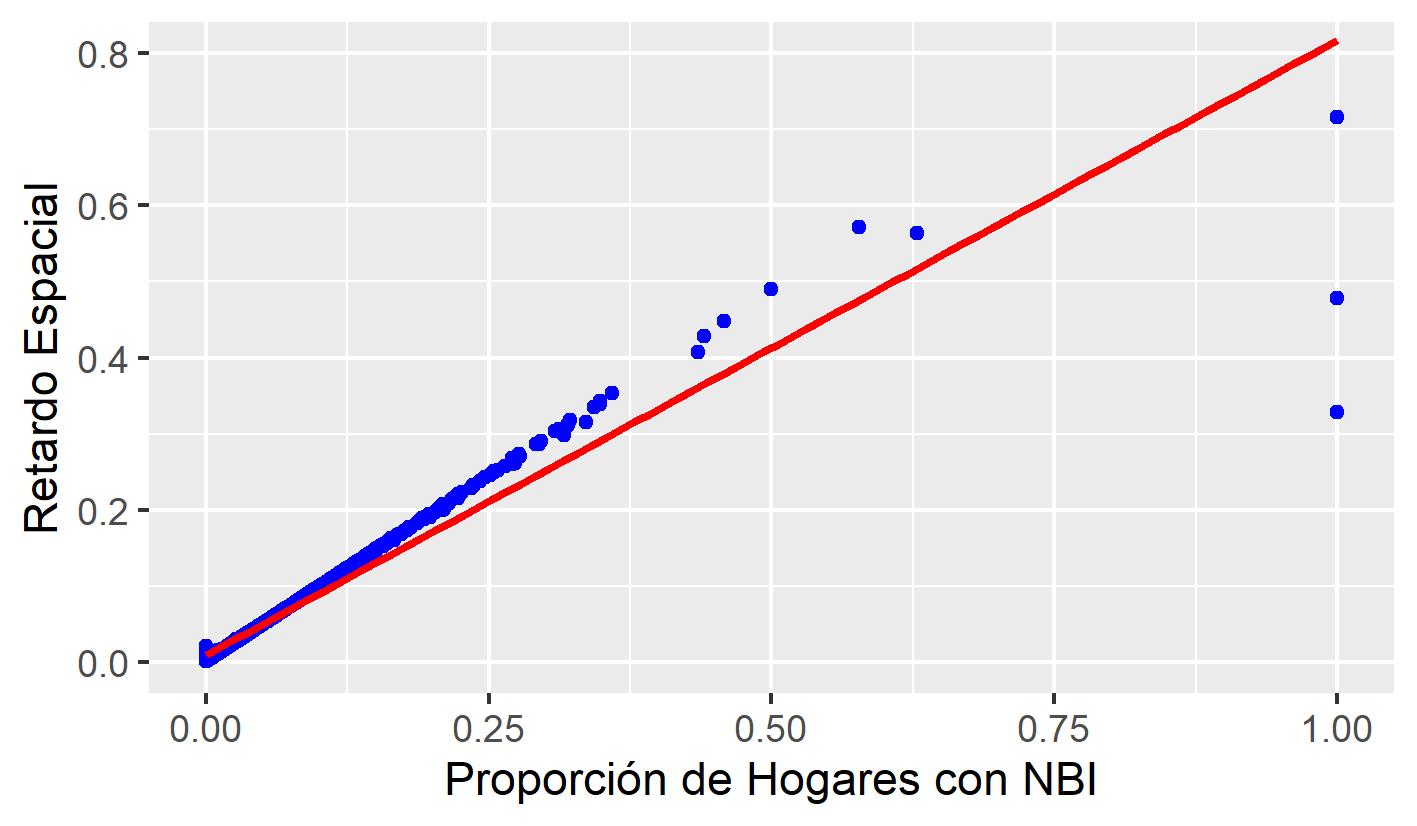
**Figura 4.7**: Gráfico de dispersión de Moran para la proporción de hogares con NBI



La figura 4.7 muestra que los puntos no se presentan de manera al azar, sino que, por el contrario, indican una autocorrelación espacial positiva de la proporción de hogares con NBI . La pendiente de la recta de mínimos cuadrados ajustada sobre la nube de puntos coincide con el estadístico de Moran (0,38527, tabla 4.1).

Las magnitudes obtenidas para el Índice de Moran resultan iguales ya sea que se utilice el supuesto de normalidad o un test permutacional.

**Figura 4.8**: Gráfico de dispersión del EBI para la proporción de hogares con NBI



Al observar la figura 4.8 se evidencia la existencia de autocorrelación espacial positiva mediante el uso del índice empírico de Bayes, ya que los puntos están lejos de presentarse de manera aleatoria. En este caso, la pendiente de la recta que minimiza la suma de las distancias al cuadrado de cada uno de los puntos hacia ella no coincide con el EBI, ya que este índice diferencia cada una de las proporciones que se encuentran en el eje de las abscisas por los tamaños poblacionales de cada una de las áreas asociadas.

Se aprecia que ambos índices arrojan resultados significativos, aunque el EBI muestra una mayor autocorrelación de la proporción de hogares con NBI, la cual es aproximadamente igual a 0,42 mientras que el estadístico de Moran es cercano a 0,39.

Como el EBI sólo puede ser calculado mediante un test permutacional, es importante mencionar que se utilizaron 999 simulaciones para la obtención del mismo.

A pesar de incluir el índice de Oden en la tabla y notar que es el que arroja la menor probabilidad asociada, hay que recordar que su comparación con Moran y con el EBI no es adecuada ya que su par de hipótesis nula y alternativa no concuerdan.

La pequeña magnitud de la estadística de Oden junto con su probabilidad asociada casi nula puede explicarse por el hecho de que la prueba es muy potente, por lo tanto, pequeños alejamientos del valor esperado serán detectados, aunque sin poder diferenciar si el rechazo de la hipótesis de ausencia de autocorrelación espacial se debe a los diferentes tamaños poblacionales de las radios censales o si, realmente, existe autocorrelación espacial.

* 1. **Resultados en Heridos de Armas de Fuego**

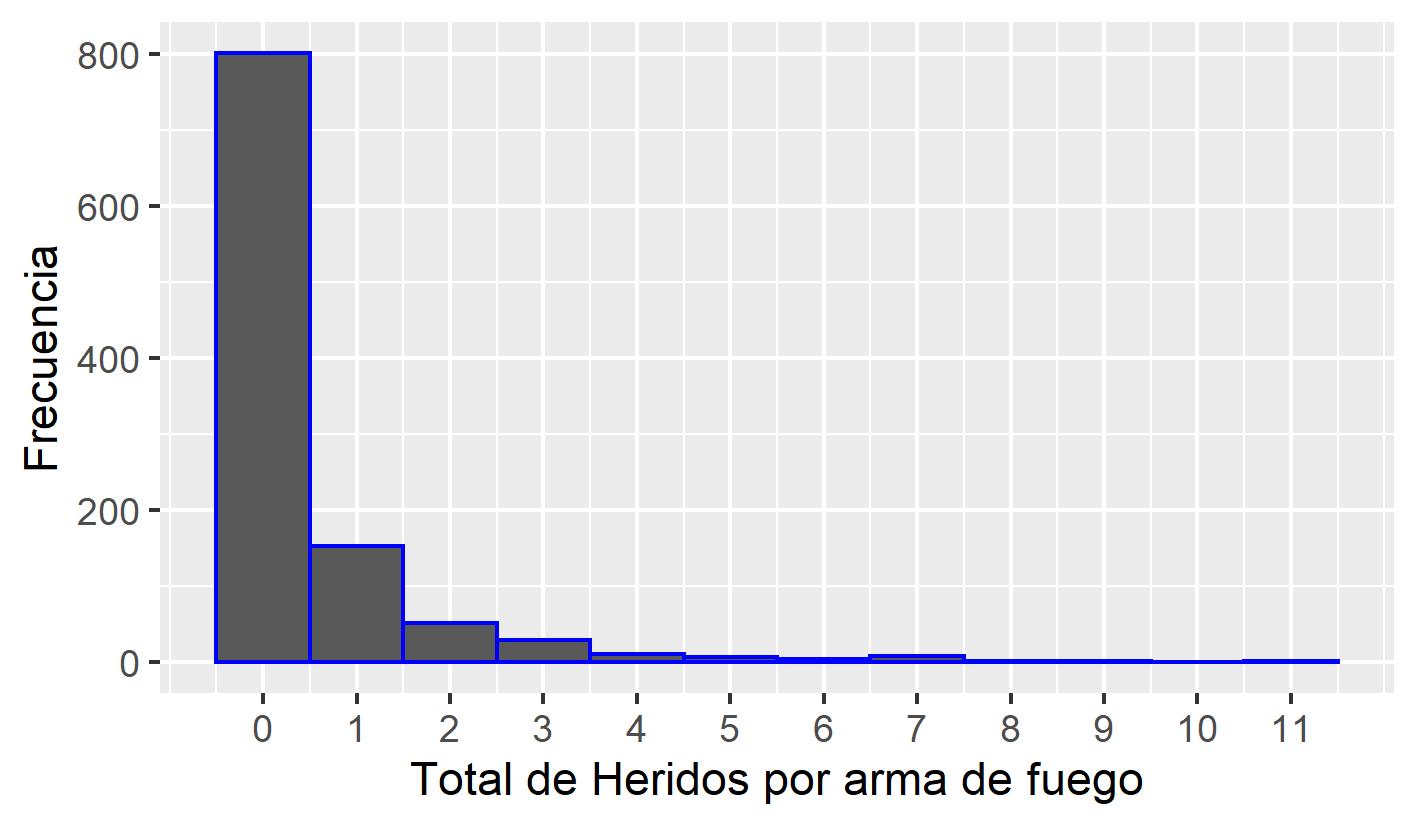
De manera similar a la sección 5.1, se aplicarán los distintos índices sobre el conjunto de datos compuesto por los Heridos de Arma de Fuego en la ciudad de Rosario.

Nuevamente, se menciona que este conjunto de datos es ficticio. La motivación de realizar esto es que en el paper referenciado en la presentación del EBI (Assunção et al., 1999), se aplica este índice al mismo problema en la ciudad de Belo Horizonte, Minas Gerais, Brasil.

La proporción de heridos de arma de fuego por radio censal se define como el cociente entre:

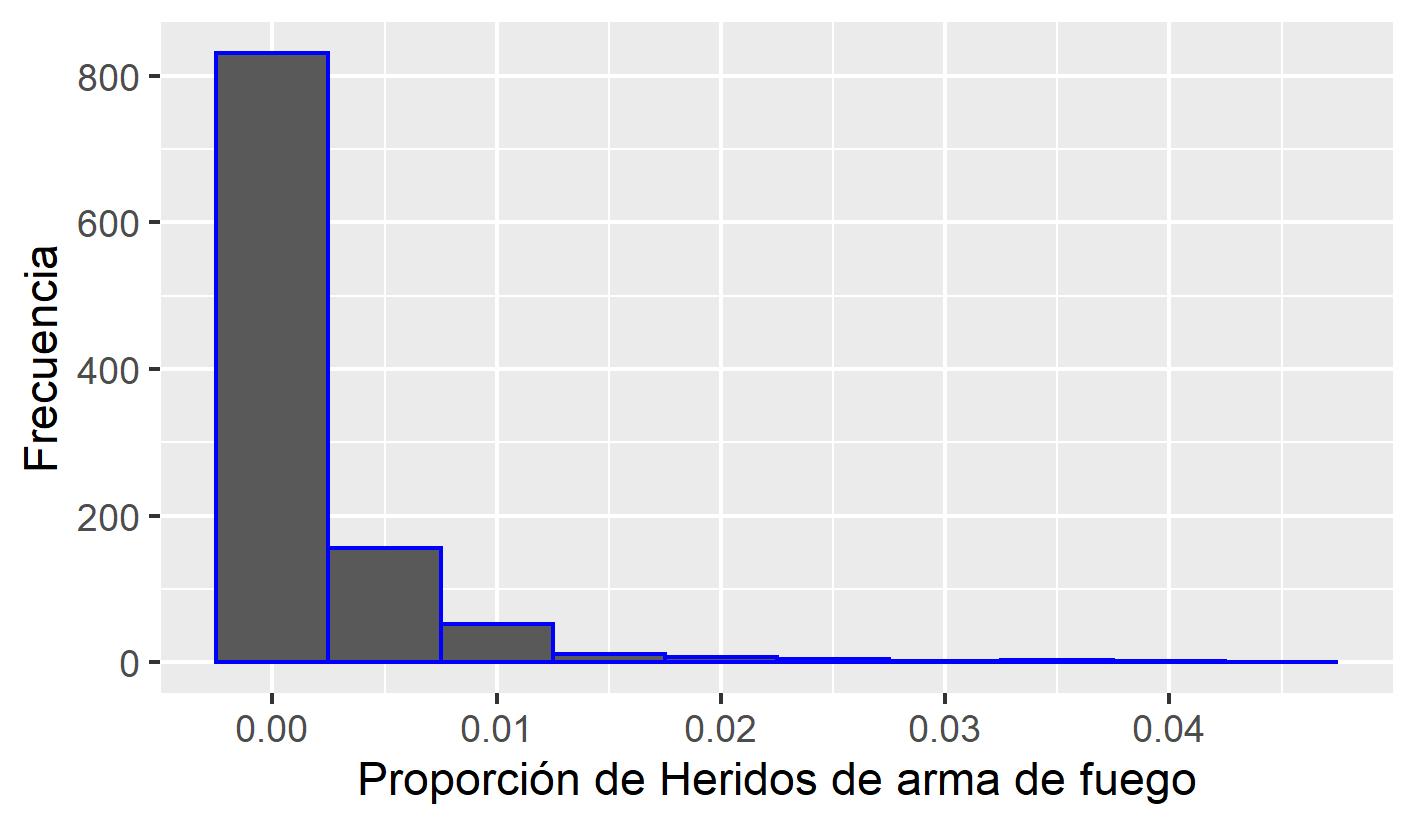
* Numerador: Cantidad de Heridos de Arma de Fuego en cada radio censal de la ciudad de Rosario
* Denominador: Cantidad de Hogares en cada radio censal de la ciudad de Rosario. Siendo esto una aproximación de la población en riesgo, resulta totalmente lógico pensar que en aquellos radios de la ciudad donde existen más hogares, existirán más personas en riesgo.

**Figura 4.9**: Distribución de la cantidad de heridos de arma de fuego en los radios censales de la ciudad de Rosario.



La figura 4.9 permite ver que la cantidad de heridos de arma de fuego en los radios censales de la ciudad de Rosario posee una gran asimetría por derecha, con media y desvío estándar aproximadamente iguales a 0,32 y 1,22 respectivamente.

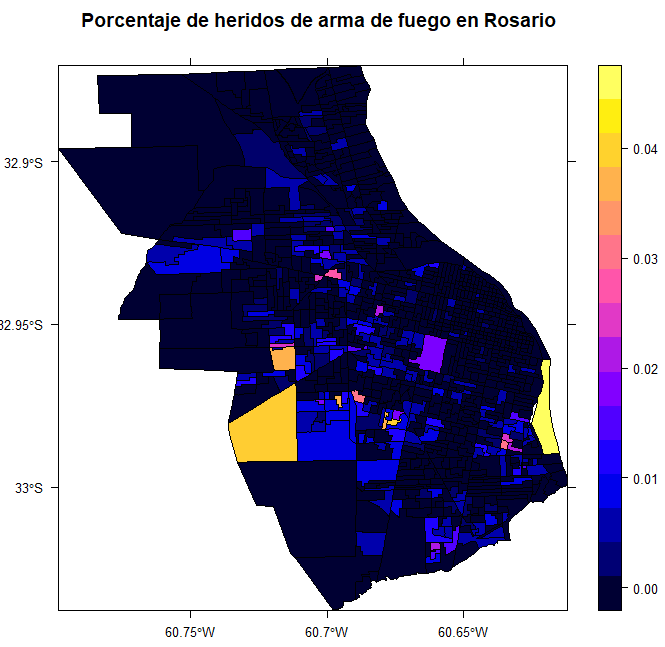
**Figura 4.10**: Distribución de la proporción de heridos de arma de fuego en los radios censales de la ciudad de Rosario.



En la figura 4.10 se puede observar que la distribución de la proporción de heridos de arma de fuego en la ciudad de Rosario posee una distribución con una pronunciada cola derecha, con media cercana a 0,0016 y desvío estándar aproximadamente igual a 0,0044.

Resulta pertinente observar la distribución espacial de la proporción de hogares con NBI, ya que es la variable bajo estudio sobre la cual se aplicarán los indicadores de autocorrelación espacial mencionados a lo largo de la presente tesina.

**Figura 4.11**: Mapa de calor del porcentaje de heridos de arma de fuego a lo largo de los radios censales de la ciudad de Rosario.



La figura 4.11 permite detectar cuales son las zonas con mayor proporción de heridos de arma de fuego, los radios censales que parecerían ser más propensos a la ocurrencia de este suceso delictivo se encuentran en las zonas pintadas de amarillo.

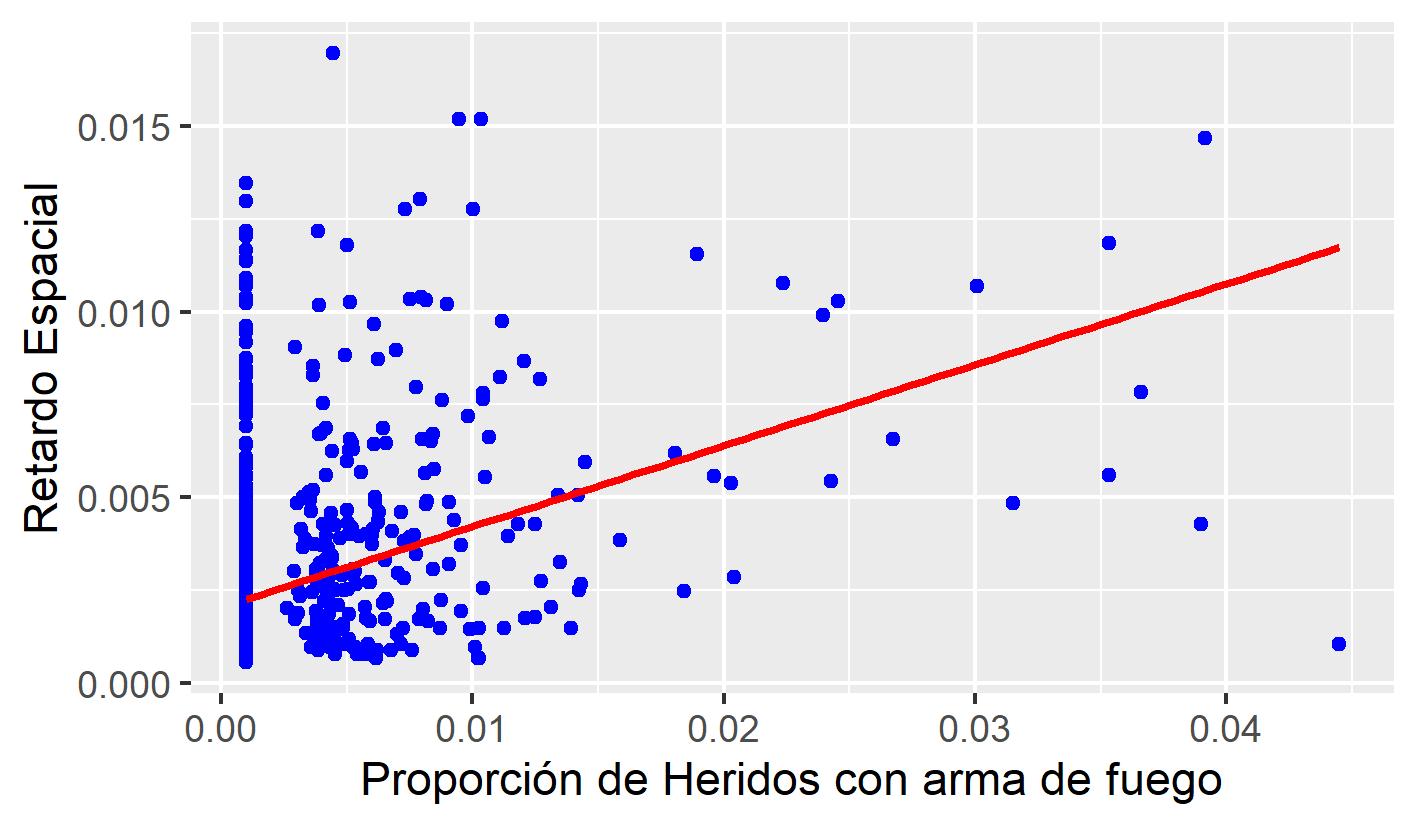
Al igual que lo ocurrido con el mapa de calor para los hogares con NBI, parecería existir un patrón espacial a lo largo de los radios censales de la ciudad, pero para comprobarlo estadísticamente y obtener una medida que lo cuantifique es importante calcular algún índice de autocorrelación espacial.

**Tabla 4.2**: índices de autocorrelación espacial calculados para la proporción de heridos de arma de fuego.

|  |  |  |
| --- | --- | --- |
| **Índice** | **Estadístico** | **P-Valor** |
| Moran (I) | 0,21796 | 0,001 |
| Oden () | 0,00549 | <0,001 |
| EBI | 0,24375 | 0,001 |

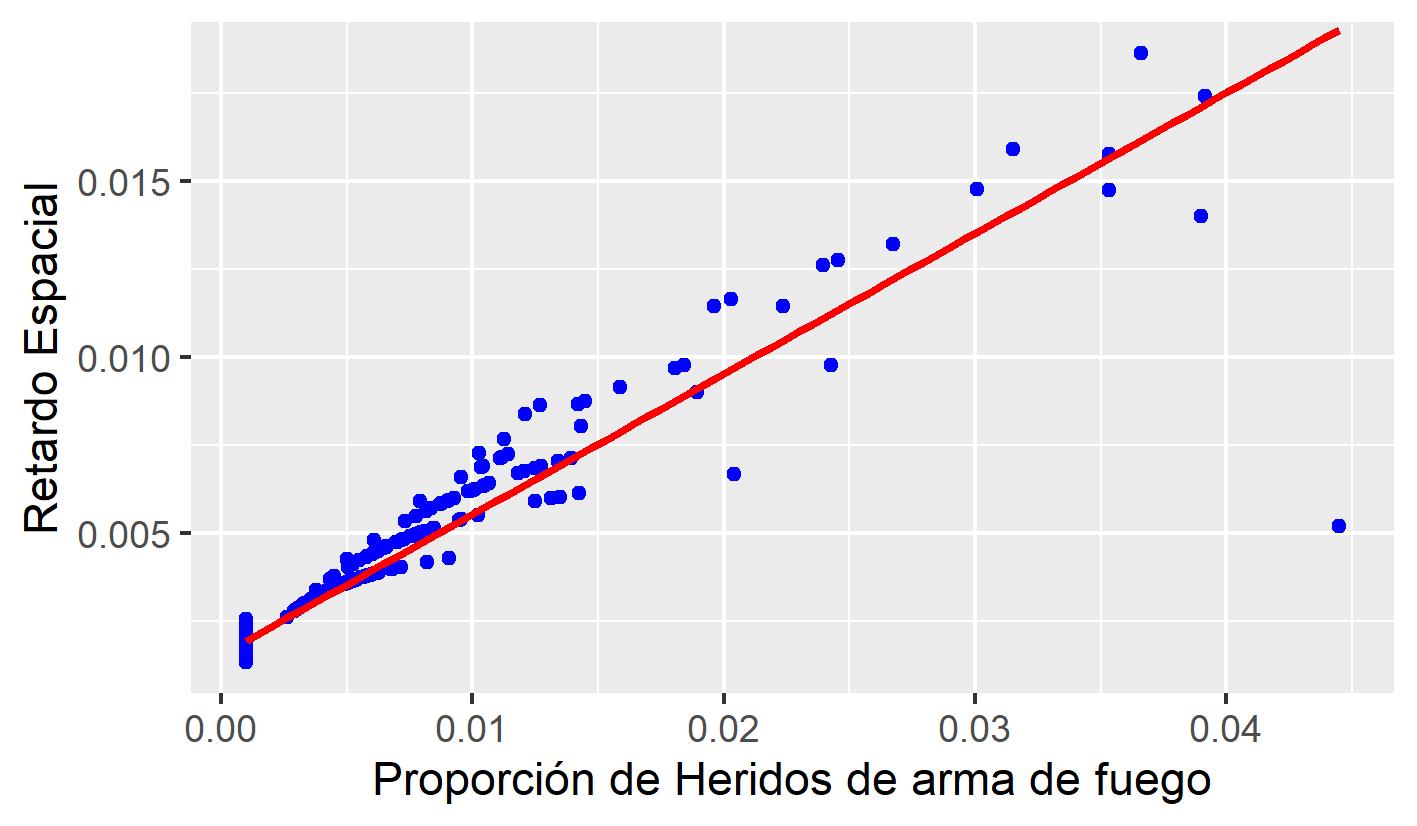
La tabla 4.2 muestra los resultados obtenidos para el índice de Moran, Oden y el EBI junto con sus correspondientes probabilidades asociadas a la prueba de hipótesis de ausencia de autocorrelación espacial.

**Figura 4.12**: Gráfico de dispersión de Moran para la proporción de heridos de arma de fuego.



La figura 4.12 muestra que los puntos no se presentan de manera aleatoria, sino que la pendiente de la recta de regresión ajustada sobre los puntos es significativamente mayor a 0, indicando una autocorrelación espacial positiva, ya que coincide con el estadístico de Moran (0,21796, tabla 5.2).

**Figura 4.13**: Gráfico de dispersión del EBI para la proporción de heridos de arma de fuego.



La estadística observada correspondiente al EBI es cercana a 0.24, resultando significativa y de una magnitud levemente superior al índice de Moran. Vale destacar, nuevamente, que la coincidencia entre la pendiente de la recta de regresión ajustada sobre los puntos y el valor del índice no es válida para el EBI, esta igualdad solo se presenta para el índice de Moran.

Al igual que ocurría con los hogares con NBI, el índice de Oden arroja la menor probabilidad asociada, a pesar de poseer una menor magnitud en cuanto a la estadística observada, el motivo por el cual ocurre esto, es que la prueba es más potente, como se mencionó en el problema analizado anteriormente.

En términos generales, los resultados de los tres índices muestran una tendencia similar a la aplicación en hogares con NBI.

En base a los resultados observados en los dos problemas considerados, puede concluirse que se halló un índice de autocorrelación espacial que considera los distintos tamaños de las áreas de una región (EBI) y que parecería tener un funcionamiento similar al índice de Moran, aunque en el siguiente capítulo se mencionarán algunas propiedades que dejaran en evidencia que el mismo es igual o superior, dependiendo las características del problema de interés.

1. **Consideraciones finales**

Es común que, en situaciones en la que se trabaje con datos espaciales, los tamaños poblacionales de las áreas consideradas para una determinada región sean distintos. Lo que lleva a considerar la necesidad de tener en cuenta el tamaño de las distintas unidades contempladas a la hora de construir indicadores que expresen la autocorrelación espacial. En otras palabras, resulta de interés evaluar alternativas distintas al índice de Moran.

En la presente tesina se han estudiado dos índices que tienen en cuenta esta cuestión mencionada anteriormente, el índice de Oden y el EBI. La utilización del primero de ellos queda descartada por los fundamentos mencionados en la sección Materiales y Métodos.

Por otro lado, el EBI mostró un comportamiento satisfactorio en los dos problemas analizados y posee algunas propiedades superadoras con respecto al índice de Moran, las más importantes se mencionarán en los siguientes párrafos.

El EBI y Moran poseen una potencia similar ante escenarios de tamaños poblacionales parecidos de las distintas áreas consideradas, pero a medida que se alejan de esta situación, el EBI incrementa su potencia de manera considerable con respecto al índice de Moran (Assunção et al., 1999).

También el EBI posee interesantes cualidades de robustez, ya que no se ve afectado por valores atípicos, a diferencia de Moran que da el mismo peso a cada unidad contemplada (Assunção et al., 1999). Una forma sencilla de entender esta propiedad es considerar el gráfico de dispersión de Moran, donde la pendiente de la recta de mínimos cuadrados coincide con el estadístico de Moran, y esto no ocurre con el EBI.

La P(e1) del EBI no se ve alterada cuando las poblaciones en las áreas son heterogéneas (Assunção et al., 1999), mientras que La P(e1) se incrementa para el índice de Moran cuando los tamaños poblaciones de las áreas son distintos y la proporción es constante (Walter, 1992).

1. **Referencias bibliográficas**

**[1] Anselin, Luc, Ibnu Syabri and Younggihn Kho (2006).** GeoDa: An Introduction to Spatial Data Analysis. Geographical Analysis 38 (1), 5-22.

**[2] Assunção, R.M.; Reis, E. A. (1999)** *“*A new proposal to adjust Moran´s I for population density”. *Statist. Med. 18, 2147-2162.*

**[3] Borra V. (2015)** “Estadística Espacial. Muestreo y modelización para la aplicación en estudios socioeconómicos”.

**[4] Marshall, R. J. (1991)** “Mapping disease and mortality rates using empirical Bayes estimators”, Applied Statistics, **40**, 283–294.

**[5] Moran, P. A. P. (1950)** “Notes on continuous stochastic phenomena”, Biometrika, **37**, 17–23.

**[6] Oden, N. (1995)** “Adjusting Moran's I for population density”. *Statistics in Medicine, 14, 17-26*.

**[7] Tiefelsdorf M, Griffith DA, Boots B (1999)** “A variance-stabilizing coding scheme for spatial link matrices”. Environment and Planning A 31:165–180.

**[8] Tobler W (1970)** “A Computer Movie Simulation Urban Growth in the Detroit Region”. Economic Geography 46(2):234-240.

**[9] Walter, S.D. (1992)** “The analysis of regional patterns in health data. I. Distributional considerations”. *American Journal of Epidemiology, 136, 730-741.*

1. Por las características de las poblaciones con las que se trabajan en la presente tesina se considera la distribución de Poisson. En otras situaciones puede emplearse la distribución Binomial. [↑](#footnote-ref-1)