**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 a describir los problemas de aplicación en los que se utilizarán los índices.

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 básicos.**

En muchos problemas estadísticos, los datos con los que se debe tratar se observan a unidades que se encuentran situadas en el espacio y cuya 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 que puede presentar esta clase de datos es la dependencia espacial, fenómeno conocido como autocorrelación espacial.

Para el tratamiento de esta clase de datos se ha desarrollado un conjunto de métodos que se conocen bajo el título de Estadística Espacial, y tienen por objeto la exploración, descripción, visualización, y análisis de datos considerando las características particulares debidas a su distribución en el espacio. Los mismos se basan en considerar a los datos espaciales como la realización de un proceso estocástico en la región que se estudia.

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 que es el más ampliamente difundido, puede encontrarse en muchos estudios publicados y se dispone para el cálculo, en cualquier software de Estadística Espacial.

Este índice puede presentar inconvenientes cuando las unidades son de tamaños diferentes y por ese motivo se han propuesto otras alternativas que consideran dichas diferencias, las cuales que se tratan en la presente tesina, como lo son el índice de Oden y el Empirical Bayes Index (EBI).

**Tipos de datos espaciales**

En función de la manera en que se considere el espacio 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 áreas 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 estar afectado por 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 que aquellos que se encuentran en sitios más alejados, 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, se puede 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.

**3.2.1. Criterios 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.

**Criterios de vecindad**

Los más utilizados y divulgados en 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. Una forma usual de referirse a la distancia entre dos áreas es hacer que ella sea 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 k 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.

**Pesos espaciales**

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 la unidad 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 de vecindad, 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.

Existen otros criterios más específicos para determinar los pesos espaciales de la matriz de vecindad que han sido estudiados, pero en los párrafos anteriores se han mencionados los más utilizados.

**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 como se comenta más adelante.

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 las razones *pi*

*wji* es el elemento i,j de la matriz de conectividad que corresponde al peso 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. Se obtiene calculando los productos cruzados de las diferencias entre las razones y su media para cada par (*i,j*) de unidades, ponderados por el peso *wji* correspondiente. 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 puede utilizar 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.

Un instrumento gráfico habitual en el análisis de autocorrelacián espacial es el diagrama de dispersión de Moran, el cual representa en la abscisa la variable de interés estandarizada y en la ordenada los retardos espaciales de dicha variable estandarizada, el cual es igual al promedio ponderado de la variable estandarizada en las áreas vecinas; las ponderaciones son los pesos asignados a cada unidad vecina. Este gráfico permite identificar los diferentes tipos de asociación: positiva, negativa o nula. Las categorías de asociación espacial positiva se corresponden a los cuadrantes I y III, mientras que la negativa se encuentra en los cuadrantes II y IV. Si se ajusta una recta de regresión lineal sobre los puntos del gráfico, el índice de Moran será igual a su pendiente. Si la nube de puntos está dispersa en los cuatro cuadrantes es indicio de ausencia de autocorrelación espacial.

**Efectos de unidades de diferentes tamaños**

Si se considera una región dividida en áreas con tamaños diferentes, como ocurre 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.

Assunção y otros (1999) estudiaron los efectos de tener diferentes tamaños de unidades sobre el error de tipo I (α) en el test de significación del índice de Moran. Si se elige la prueba normal, se debe suponer que las razones están distribuidas normalmente idénticamente y son independientes. Si las unidades tienen diferente tamaño y la variable aleatoria que se considera para cada unidad es una tasa o una razón, la variancia será diferente ya que dependerá del tamaño de la unidad y la distribución posiblemente sea Poisson o Binomial. Por otra parte, las medias podrían ser diferentes ya que la inexistencia de autocorrelación espacial no descarta la heterogeneidad espacial.

Si se aplica un test permutacional y la situación es la descripta en el párrafo anterior, las distribuciones no son “intercambiables”, condición requerida por esa clase de pruebas.

**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 tendrán valores iguales a 2, a diferencia de las matrices de vecindad ya vistas donde la diagonal está compuesta por elementos iguales a 0. Oden propone esta modificación, con el objeto de diferenciar la situación 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 el índice de Moran, la prueba de significación estadística de la existencia de autocorrelación espacial con el índice de Oden se realiza bajo el supuesto de normalidad o mediante un test permutacional.

Bajo la hipótesis nula de inexistencia de autocorrelación espacial, 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 =

Bajo la hipótesis nula: E[] = , y si m es suficientemente grande, la estadística Z = se distribuye como una variable aleatoria normal estándar.

Oden (1995) muestra, mediante estudios por simulación, que el test que utiliza a es más potente que la prueba realizada con la estadística *I* de Moran, cuando los tamaños de las unidades espaciales consideradas son distintos. Esta capacidad de capturar la variabilidad se debe al efecto del 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 se debe poner atención a la observación realizada por Assunção y otros (1999), quien advierte sobre 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 subyacentes 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 y otros, 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)**

Assunção y otros (1999) proponen un indicador alternativo para probar la existencia de autocorrelación espacial que llaman Empirical Bayes Index (EBI). En el artículo citado, realizan una reseña metodológica de los índices de Moran, Oden y Waldhör y una comparación de los tests destinados a comprobar la existencia de autocorrelación espacial mediante simulaciones en diferentes escenarios. Consideran la ciudad de Belo Horizonte, Minas Gerais, Brasil, la variable que se estudia es tasa de homicidios en 1994, calculada en cada uno de los 81 distritos como el cociente entre el número de homicidios ocurridos y la población del distrito. La población de los distritos varía entre 31 y 70870 habitantes. Simulan para cada distrito el número de casos de acuerdo a una distribución de Poisson y se asumen tres opciones de poblaciones en cada distrito, una de ellas población constante y otras dos con poblaciones heterogéneas. Realizan 1000 simulaciones en cada escenario, llevando a cabo los tests correspondientes con un nivel de significación α del 5% y observan la proporción de veces que rechazan la H0 (no hay autocorrelación espacial), encontrando que el menos afectado por la heterogeneidad en el tamaño de la población es el índice de Moran. Luego de proponer el EBI, comparan mediante un estudio por simulación en escenarios similares, su desempeño con respecto al del índice de Moran, evaluando la potencia de los tests y encontrando diferencias a favor del EBI en los casos de heterogeneidad en el tamaño de las poblaciones. El artículo finaliza con la aplicación de los índices al problema de la distribución espacial de homicidios en Belo Horizonte mostrando los resultados que se obtienen y destacando la importancia del EBI.

En los siguientes párrafos se describen brevemente los aspectos teóricos del EBI.

Sean θ1, θ2, …, θi, ..., θm, las razones subyacentes en las m áreas en estudio. Se realiza el supuesto de que el número de eventos observados ni 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.

Considerando el enfoque bayesiano, se realiza el supuesto de que las razones θi tienen a priori 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 y otros, 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).

Es posible construir un gráfico de dispersión del EBI, el cual se construye de igual manera que para el índice de Moran, pero utilizando los en lugar de la variable de interés estandarizada, la forma de interpretar este diagrama es equivalente a Moran.

Assunção y otros (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.

**3.3 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 del Instituto Nacional de Estadística y Censos.

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, problema similar al utilizado en Assunção y otros (1999).

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 son reticulares irregulares, donde cada una de las áreas o agregaciones espaciales corresponden a los radios censales de ciudad de Rosario.

A la hora de determinar la matriz de conectividad, se adopta 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 emplea el criterio de “estandarización por filas” 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 destaca que debió desarrollarse un programa R para el cálculo del índice de Oden, ya que no se encontró en ningún paquete de R ni en otro software.

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)