



UNIVERSIDAD DE LA REPÚBLICA
Facultad de Ciencias Económicas y de Administración
Instituto de Estadística

Determinación de regiones de imputación para datos espaciales utilizando el algoritmo PCNM: un ejemplo de aplicación a los datos del Censo 2011

Riaño, Eugenia

Noviembre, 2016

Documentos de Trabajo

Serie DT (16 / 2) - ISSN : 1688-6453

Forma de citación sugerida para este documento:

Riaño Eugenia. 'Determinación de regiones de imputación para datos espaciales utilizando el algoritmo PCNM: un ejemplo de aplicación a los datos del Censo 2011'
Serie Documentos de Trabajo, DT 11/2016. Instituto de Estadística, Facultad de Ciencias Económicas y Administración, Universidad de la República, Uruguay.

Determinación de regiones de imputación para datos espaciales utilizando el algoritmo PCNM: un ejemplo de aplicación a los datos del Censo 2011

Riaño, Eugenia ¹

RESUMEN

En general, la calidad y cobertura de los Censos de Población y Vivienda del año 2011 fue calificada como positiva, cumpliendo con los estándares exigidos internacionalmente. Sin embargo, su implementación no estuvo exenta de inconvenientes. No se cuenta con información de determinados hogares cuyo domicilio fue relevado, y para algunos se cuenta con sólo información parcial relativa a la composición del hogar. La omisión censal se concentra en zonas socioeconómicamente vulnerables. Esto afectaría la construcción del mecanismo utilizado por el Ministerio de Desarrollo Social para seleccionar a la población beneficiaria de los programas de transferencia monetaria. Este mecanismo se basa en la Encuesta Continua de Hogares cuyo marco muestral es el del Censo, y refleja los problemas de omisión. Para el caso de Montevideo, el patrón espacial de la población objetivo y de la propia omisión indican que es necesario definir regiones de imputación, dado que la distribución espacial no es continua en el mapa. La selección de los modelos a utilizar para la imputación es muy sensible a la escala del mapa, por lo que la definición de las regiones condiciona la selección del modelo final a utilizarse para realizar la imputación. El objetivo de este trabajo es construir las regiones a partir del algoritmo PCNM (Principal Coordinates of Neighbour Matrices). El método consiste en aplicar Análisis de Componentes principales a una matriz de distancias truncada entre las observaciones. Así se obtiene un conjunto de variables explicativas que captan la variabilidad espacial en diferentes escalas. A partir de ellas pueden construirse las posibles regiones de imputación, y mediante un modelo de regresión, analizar cuáles son las escalas que se encuentran más asociadas con la variable de respuesta.

Palabras clave Análisis de Componentes Principales, Estadística Espacial, Índice de Moran, Modelos de Regresión Poisson,

Códigos JEL: C21, C31, C38, C88

Clasificación MSC2010: 62H11, 62H25, 62P12, 62M40

¹ *Instituto de Estadística - Facultad de Ciencias Económicas y de Administración - UdelaR.*

1. Introducción

En general, la calidad y cobertura de los Censos de Población y Vivienda del año 2011 fue calificada como positiva, cumpliendo con los estándares exigidos internacionalmente. Sin embargo, su implementación no estuvo exenta de inconvenientes. No se cuenta con información de determinados hogares cuyo domicilio fue relevado, y para algunos se cuenta sólo con información parcial relativa a la composición del hogar.

La omisión censal se concentra en zonas socioeconómicamente vulnerables. Esto afectaría la construcción del mecanismo utilizado por el Ministerio de Desarrollo Social (de aquí en más MIDES) para seleccionar a la población beneficiaria de los programas de transferencia monetaria. Este mecanismo se basa en la Encuesta Continua de Hogares cuyo marco muestral es el del Censo, y refleja los problemas de omisión.

En el presente documento se presentan los resultados de una etapa intermedia de investigación, de un trabajo cuyo objetivo final es la imputación de población elegible en las zonas omisas del Censo. El marco geográfico es el del departamento de Montevideo. Se observa que el patrón espacial de la omisión es diferente por regiones, por lo que se hace necesario definir regiones de imputación.

Para la construcción de las regiones se utiliza el algoritmo PCNM (Principal Coordinates of Neighbour Matrices). El método consiste en aplicar Análisis de Componentes principales a una matriz de distancias truncada entre las observaciones. Así se obtiene un conjunto de variables explicativas que captan la variabilidad espacial en diferentes escalas. A partir de ellas pueden construirse las posibles regiones de imputación, y mediante un modelo de regresión, analizar cuáles son las escalas que se encuentran más asociadas con la variable de respuesta.

En las siguientes secciones se presenta primero una descripción del problema, el marco teórico y la metodología a utilizar, se presentan los resultados obtenidos y finalmente se realiza una discusión en donde se plantean las direcciones futuras a seguir en el tema de investigación.

2. Descripción del problema

En el año 2008, en el marco de un convenio entre el MIDES y la Universidad de la República, se construye un índice denominado Índice de Carencias Críticas (de aquí en más ICC), con el fin de definir inicialmente la población elegible del programa de Asignaciones Familiares del Plan de Equidad (AFAM -PE). Luego también es utilizado para definir la población elegible de los programas de transferencias Tarjeta Uruguay Social (TUS) (Ministerio de Desarrollo Social 2013).

El ICC es un modelo *probit* que predice la probabilidad de que un hogar pertenezca al primer quintil de ingresos, en base a variables que reflejan su situación en términos de educación, vivienda, confort y composición del hogar. Junto con la estimación del ICC, se determina también un punto de corte por región (Montevideo - Interior) que permite distinguir entre los hogares que según la predicción del modelo pertenecen a la población elegible, y los que no. Es decir, se fija un valor del ICC a partir del cual un hogar sería población elegible para el programa por región.

Dichos umbrales se fijan de modo de permitir la elegibilidad de la cantidad de hogares que la ley define deben recibir el beneficio, tomando como referencia la Encuesta Continua de Hogares (ECH) que se considera representativa del total poblacional.

Los hogares de la ECH cuentan con sus correspondientes ponderadores muestrales anuales, que permiten realizar las estimaciones pertinentes al caso. Para fijar los umbrales, se estima el ICC para los hogares de la ECH y se calcula la probabilidad de pertenecer al primer quintil de ingresos. Luego se ordenan en forma decreciente según la probabilidad calculada por el modelo, y dada la cantidad de hogares que pretende alcanzar el programa se estima la distribución de los mismos entre Montevideo - Interior. Se realiza una suma acumulada de los ponderadores muestrales que permite identificar cuáles son los hogares de la muestra que representarían a los beneficiarios del programa y luego tomando esos hogares se estima la proporción Montevideo - Interior.

Una vez determinada la cantidad de posibles beneficiarios del programa por región, el umbral del ICC se fija en el lugar donde (ordenado en forma decreciente) la estimación de la población beneficiaria se iguale a la cantidad de beneficiarios que pretende alcanzar el programa.

El procedimiento descrito en los párrafos anteriores corresponde al programa AFAM. Para los programas TUS se modifica levemente, tomando como criterio para ordenar los hogares la capacidad de compra de la canasta básica alimentaria (CBA) en lugar de los ingresos. Esto tiene como consecuencia que la cantidad de partidas de beneficios aumente para Montevideo, ya que la CBA es más cara.

Dado que el marco de la ECH es el del Censo 2011, cuya omisión se concentra en zonas socioeconómicamente vulnerables, se cuestiona cual sería el efecto de la omisión en los umbrales del ICC. Es decir, si los hogares que no se incluyen en la ECH por no estar en el marco son más vulnerables que los demás, los umbrales se verían modificados al alza. La omisión del censo tendría como consecuencia que la distribución de los programas entre los hogares beneficiarios no fuera la correcta, así como también podría verse modificada la distribución de las partidas de beneficios entre Montevideo e Interior.

Surge así la necesidad de realizar una estimación de la cantidad de población elegible en las zonas omisas del Censo, con el fin de estimar el impacto en los umbrales principalmente de los programas TUS y en la distribución por región de las partidas de beneficios.

2.1 Definiciones

En este apartado se describe principalmente la operativa del Censo 2011, necesaria para entender las causas y los tipos de omisión.

El operativo del Censo comienza con un relevamiento previo al propiamente denominado como Censo, en donde los Jefes de Campo realizaron un empadronamiento de domicilios por observación que luego sirvió como punto de partida para los censistas que efectivamente realizaron el Censo.

Se tiene entonces que por un domicilio por lo menos deberían haber pasado dos personas al momento del Censo. Como el primer empadronador sólo incluye domicilios en su listado por observación, puede suceder que se omitan viviendas que no se encontraran a la vista, como puede darse el caso de varios domicilios en un mismo padrón. Este error se subsanaría cuando el domicilio fuera visitado por el censista, que una vez que ingresa a una vivienda, indaga por la existencia de otras viviendas en el mismo padrón. El censista tiene las herramientas para agregar o eliminar domicilios, es decir, para corregir errores u omisiones que haya cometido el primer empadronador.

En la realidad surgieron varios problemas que dan lugar a los diferentes tipos de omisión, con los que se trabajarán en el documento.

El primero es el caso en que el empadronador releva el domicilio, pero el censista no. Se tiene un registro de la existencia de la vivienda, pero no se tiene información de quienes residen en ella. A este caso se lo denomina Morador Ausente. Al estar registrada en el marco censal, la vivienda tiene posibilidades de ser seleccionada en la ECH. Si bien no se tienen datos al momento del censo sobre la cantidad de población elegible residente en la vivienda, al no quedar excluida del sorteo de la ECH, no genera un problema de marco.

El segundo es consecuencia de una operativa de emergencia al final del relevamiento del Censo. Dado que se aproximaba fin de año, y en enero no sería posible continuar con el operativo del Censo, se decidió relevar solamente los datos referentes a la composición del hogar, es decir, sexo, edad y relación de parentesco, en un cuestionario papel. Para estos hogares no es posible calcular el ICC, pero sí se tiene algún tipo de información que puede ser de importancia para determinar si el hogar es población elegible o no, como ser la cantidad de menores de 15 años. A estos casos se los denomina "Formato Papel". Como en el caso anterior, al no excluirse la vivienda del marco, no tiene consecuencias graves sobre la ECH, sí sobre la estimación de población elegible al momento del censo.

El tercer caso es en el que en donde la vivienda no fue relevada ni por el empadronador, ni por el censista. Para poder corroborar esta situación es necesaria una fuente externa de información, ya que en el marco del INE no hay registros de estos domicilios. El MIDES realiza un relevamiento permanente en zonas socioeconómicamente vulnerables que permitió identificar algunos de estos casos en zonas censales que se encontraban vacías según el censo. El trabajo que el MIDES realiza en campo no permite identificar a todas las zonas que fueron omitidas completamente por el censo, ya que la selección de zonas a visitar se basa en informantes calificados y no necesariamente coinciden con las omisas. Sí se obtiene una cota inferior para esta cantidad, no siendo posible identificar el resto por no contar con fuentes externas oficiales que lo permitan.

El tercer caso sí tiene consecuencias importantes sobre las estimaciones de la ECH ya que estas zonas no son incluidas en el marco. El error que implique esta omisión se arrastra a todas las ECH futuras hasta una nueva actualización del marco. Por este motivo es que este trabajo se enfoca en la imputación de los datos en estas zonas, en donde actualmente no se tiene ningún tipo de información.

La unidad de análisis es la zona censal, y se imputarán aquellas sin datos en forma completa, ya sea por moradores ausentes, formato papel o por omisión propiamente dicha. Si bien existen zonas en donde parte de las viviendas fueron omitidas, en una primera instancia no se realizará la imputación de las mismas.

2.2 Los datos

Las bases cartográficas con las que se trabajan corresponden a las actualizadas por el Censo 2011. Las zonas censales son unidades de muestreo en la selección de la muestra de la ECH. En el país existen un total de 69.752 zonas de las cuales 16.486 corresponden a “zonas verdes”. Ejemplos de zonas verdes son los parques, canteros, reservas naturales, etc. Estas zonas no se censan y no se toman en cuenta para la imputación.

A continuación se presenta el total de zonas por departamento, clasificando las zonas según las tres categorías mencionadas en el apartado anterior:

Tabla 1: Zonas censales por tipo según departamento

Departamento	Total de zonas	Total sin zonas verdes	Morador Ausente	Formato Papel	Zonas Mixtas	Omisas
Montevideo	13621	10559	20	182	290	140
Artigas	1806	1424	0	1	0	21
Canelones	12614	9894	19	6	2	52
Cerro Largo	2415	1885	9	0	0	2
Colonia	3954	3024	4	2	6	8
Durazno	1900	1515	0	0	0	20
Flores	819	681	0	0	0	0
Florida	2313	1775	0	0	0	11
Lavalleja	1658	1365	0	0	0	5
Maldonado	6838	4267	49	3	1	8
Paysandú	3296	2493	0	0	0	31
Río Negro	1665	1337	0	0	0	4
Rivera	2149	1762	0	0	0	16
Rocha	3291	2468	6	1	17	9
Salto	2550	2015	0	0	0	11
San José	2597	1984	1	0	0	11
Soriano	2120	1686	4	0	0	12
Tacuarembó	2151	1603	0	0	0	13
Treinta y Tres	1529	466	0	0	0	8
Total	69752	53266	112	195	316	382

Se agrega una categoría de “zonas mixtas” para aquellas en donde todos sus hogares son Moradores Ausentes o Formato Papel.

A simple vista se aprecia que los grandes problemas de omisión se concentran en el departamento de Montevideo, completando un total de 632 zonas para imputar de un total de 1005 zonas en todo el país. La categoría Formato Papel casi se concentra con exclusividad en Montevideo.

En el presente trabajo sólo se considera el departamento de Montevideo, ya que se considera que es el más importante en problemas de omisión, y es el que puede tener mayores consecuencias en la distribución de las partidas de beneficios.

La distribución geográfica de las zonas sin información se presenta en la Figura 1. Los problemas de omisión claramente se concentraron en zonas socioeconómicamente más vulnerables. Los centros comunales zonales 9 y 17 concentran el 77.2% de las zonas sin información. También con una participación menor se encuentran los comunales 10, 11 y 12 que concentran un 15% de los casos (ver Anexo 1).

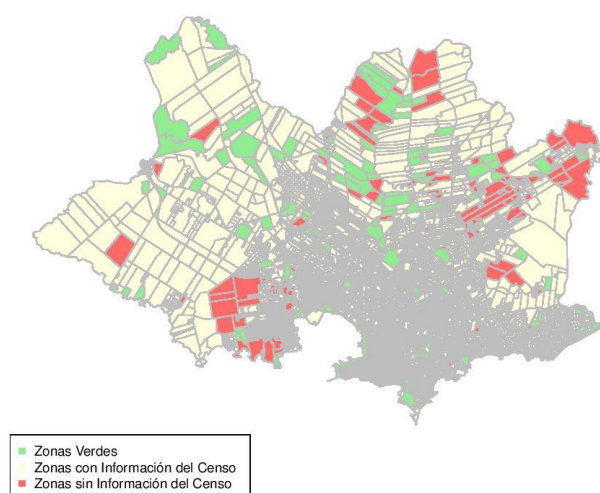


Figura 1: Distribución Geográfica de las zonas sin información del Censo

Los comunales se toman como referencia descriptiva, en el sentido que la omisión presenta cierta continuidad en el espacio, y no se delimita estrictamente por los límites geográficos de los centros comunales. Del análisis anterior surge que existen unas dos o tres regiones de análisis: la del Cerro y la de Camino Maldonado, y una zona Norte más dispersa que podría plegarse a la de Camino Maldonado o que podría ser tratada por separado.

Definir y elegir los límites de las regiones de análisis es el objetivo de este trabajo, ya que se entiende que el patrón espacial trasciende los límites administrativos, y que éstos deben definirse de acuerdo a un mejor ajuste de los modelos a estimar o a la detección de procesos espaciales diversos en las distintas zonas del departamento.

2.3 Distribución espacial de la población elegible

Como se mencionó anteriormente, el ICC es calculado en base a los datos de la ECH. Los cuestionarios de la ECH y del censo son diferentes, por lo que el cálculo del ICC

debe ser adaptado para poder aplicarse a los datos del Censo. Este trabajo estuvo a cargo del Instituto de Economía de la Facultad de Ciencias Económicas y de Administración (Bai, Carrasco y Colafranceschi 2014).

Al reproducir el ICC en los datos del censo se obtiene como resultado un mapa con la distribución territorial de los potenciales beneficiarios de los programas prioritarios del MIDES (AFAM y TUS). Los programas son acumulativos, es decir, el beneficiario de la TUS Doble, también lo es de AFAM, así como los beneficiarios de la TUS Simple.

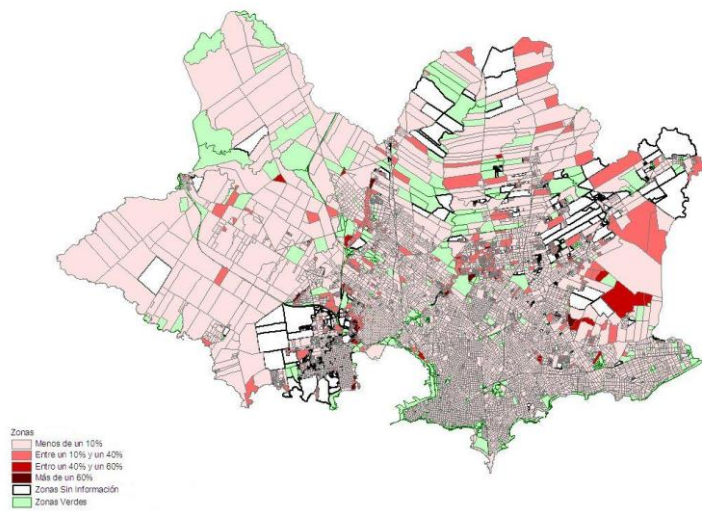


Figura 2: Distribución espacial de la población elegible para el programa TUS Doble

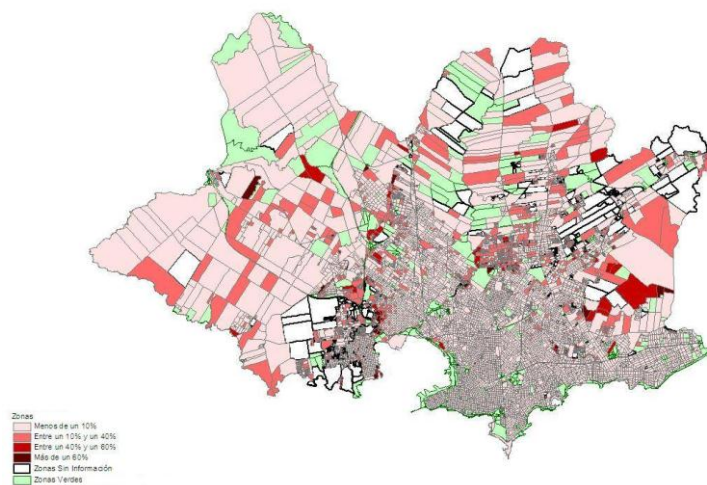


Figura 3: Distribución espacial de la población elegible para el programa TUS

Sin ser el caso del programa AFAM (Figura 4) que tiene una distribución más “continua” en la superficie, se observa que las zonas omisas son próximas a las que

tienen una mayor concentración de población elegible para los programas TUS y TUS Doble.

La regionalización de las zonas dependerá del patrón espacial de la variable a imputar, es decir, para los diferentes programas podrían obtenerse diferentes regiones de imputación. En este documento se presentará los resultados obtenidos para el programa TUS Doble.

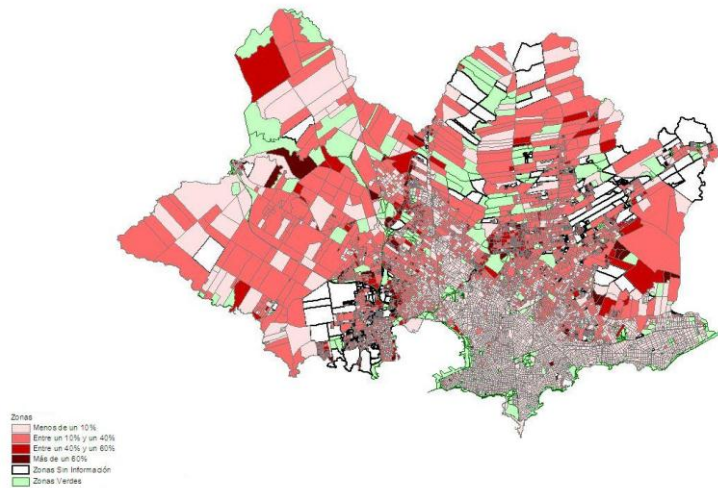


Figura 4: Distribución espacial de la población elegible para el programa AFAM

3. Marco Teórico

3.1 Introducción

Dentro de la Estadística espacial según (Guyon y Gaetan 2011), existen tres grandes áreas de estudio:

- Análisis de Patrones de Puntos
- Geoestadística,
- Datos de Área.

En Análisis de Patrones de Puntos (*Spatial Point Pattern*) el interés se centra en el lugar en donde ocurrirán los eventos. Se utiliza por ejemplo en epidemiología, y se concentra en la modelización de procesos estocásticos en tiempo y espacio.

En el segundo caso, los datos pueden medirse en un principio en cualquier punto del espacio (datos continuos). El interés no es el patrón de puntos observados en sí mismo, si no en la predicción sobre un espacio continuo de una variable de interés medida en los sitios observados.

Cuando los datos espaciales son observados en polígonos, se los definen como Datos de Área. En la mayoría de los casos los polígonos corresponden a unidades administrativas, como ser una zona censal, un departamento, país, etc. Los datos

observados son frecuentemente agregados dentro de los límites del polígono, como por ejemplo totales o promedios dentro de las áreas administrativas. Los datos de área son datos discretos.

En el caso de Geoestadística, las distancias sobre una superficie continua son la base para determinar la estructura de la autocorrelación espacial. En el caso de Datos de Área no existe una noción de distancia euclídea, si no que se deben definir estructuras de vecindad - cercanía entre los polígonos. En una superficie continua, todos los puntos son vecinos uno a uno, aunque a algunos puede dársele un peso menor, por encontrarse muy apartados. En una superficie de polígonos, se define una partición del conjunto (excluyendo una observación i) separando los polígonos que son vecinos de la observación i del resto de las observaciones. Así se conforma un grafo dirigido que describe la dependencia espacial entre los datos. A cada polígono en la vecindad a su vez se le pueden asignar diferentes pesos, que reflejen la intensidad de la dependencia espacial. Esta diferencia entre Geoestadística y Datos de Área es fundamental, ya que en el segundo caso es el investigador que introduce la estructura de vecinos.

El presente trabajo corresponde a Datos de Área y las unidades de observación son las zonas censales del departamento de Montevideo. A continuación se presentan una serie de conceptos básicos para el análisis de este tipo de datos, comenzando por la definición de vecindad espacial.

3.2 Definición de Vecindad Espacial

El primer paso es definir el criterio para elegir los vecinos. El segundo es asignar un peso a los vínculos entre observaciones definidos en el paso anterior.

Existen varias formas de definir los vecinos (Bivand, Pebesma y Gómez-Rubio 2013):

- Por contigüidad: se definen como vecinos a aquellos polígonos que compartan bordes o vértices con la observación. También se puede elegir el orden, en términos de series temporales, rezagos.
- Vecinos basados en grafos: triangulación de Delaunay, Esferas de Influencia, Gabriel graph, entre otros.
- Vecinos basados en distancias: vecinos más cercanos, radios de distancia.

Los métodos mencionados anteriormente pueden ser implementados tanto en datos con polígonos irregulares (el caso de estudio por ejemplo) como regulares (una grilla regular como pueden ser los píxeles de una imagen). Existen otros métodos exclusivos para grillas regulares que no se consideran en este caso por no ser aplicables a los datos a analizar.

3.3 Definición de la matriz de Pesos Espaciales

Una vez establecida la matriz de conectividad que define a los vecinos se debe elegir la forma de ponderar las relaciones establecidas en la matriz de conectividad. A esta matriz de pesos se la denota como W y juega un rol importante en el cálculo de la autocorrelación, en la estimación de modelos y es la base del algoritmo PCNM.

Existen varias formas de definir la matriz W (Bivand, Pebesma y Gómez-Rubio 2013). La primera forma es definir la matriz como binaria, indicando con unos y ceros cuáles son vecinos y cuáles no. Otra forma es mantener la distancia euclídea entre los vecinos, asignando un cero en el resto de los vínculos. No se realizará una descripción exhaustiva de la forma de asignar los pesos, pero existe una variedad de métodos basados en distancia, que penalizan de distinta forma los vecinos más alejados.

Se dice que la matriz W está estandarizada cuando la suma de sus filas es uno. Cuando esto se cumple la interpretación de los parámetros de los modelos es más sencilla.

3.4 Autocorrelación espacial: Índice de Moran

Aunque no es el único, el Índice de Moran es el estadístico más utilizado para testear la autocorrelación espacial. Se define como:

$$I = \frac{N}{\sum_{i=1}^N \sum_{j=1}^N w_{ij}} \frac{\sum_{i=1}^N \sum_{j=1}^N w_{ij} (X_i - \bar{X})(X_j - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

Donde N es el número de observaciones, X es la variable de interés y w_{ij} son los pesos de la matriz W . El valor esperado bajo la hipótesis nula de no autocorrelación es

$$E(I) = \frac{-1}{N-1}$$

Los valores de I van de -1 a 1. Cuando se acerca a cero indica un patrón espacial aleatorio (no hay autocorrelación).

Existen dos formas de interpretar la hipótesis nula de no existencia de autocorrelación espacial:

1. Las observaciones son mutuamente independientes, o
2. Cada permutación de las observaciones x_i es igualmente probable.

Se puede demostrar que la distribución asintótica del Índice de Moran escalado por una constante es $N(0,1)$ (Guyon y Gaetan 2011).

Una propiedad interesante del Índice de Moran son sus valores extremos. El máximo y el mínimo de I se encuentran asociados a los valores propios de la matriz W independientemente de los valores observados en los puntos de la variable X . Esta propiedad es la base del algoritmo denominado *Moran's Eigenvector Map* (MEM), siendo el PCNM un caso particular de éste, que se describe a continuación.

3.5 Moran's Eigenvector Map

El Índice de Moran puede ser reescrito en forma matricial de la siguiente manera:

$$I = \frac{N}{1^T W 1} \frac{X^T (I - 11^T/N) W (I - 11^T/N) X}{X^T (I - 11^T/N) X}$$

Para la matriz espacial W , de Jong et al (1984) demostraron que los valores máximos y mínimos del Índice de Moran están dados por $(N/1^T W 1)\lambda_{max}$ y $(N/1^T W 1)\lambda_{min}$ donde λ_{max} y λ_{min} son los valores extremos de la matriz $\Omega = (I - 11^T/N)W(I - 11^T/N)$. Por lo tanto, los vectores propios de Ω son vectores de norma uno que maximizan el I de Moran bajo la restricción de ortogonalidad.

Los valores propios de esta matriz son iguales a los I de Moran post multiplicados por una constante, y similarmente pueden ser positivos o negativos. Los vectores propios asociados con los valores propios positivos (o negativos) más grandes tienen una autocorrelación positiva (o negativa) alta, y describen estructuras globales. A medida que los valores propios decrecen, se asocian a estructuras espaciales más locales, es decir, capturan la variabilidad espacial en escalas más pequeñas. Los vectores se obtienen realizando Análisis de Componentes Principales sobre la matriz Ω .

Los vectores propios encontrados son utilizados para lo que se denomina *filtrado espacial*. Los vectores asociados a los mayores valores propios al capturar la tendencia global sirven para quitar lo que sería el componente de tendencia de los datos (la metodología es análoga a la utilizada en series de tiempo). Así por ejemplo, los vectores encontrados son agregados como variables explicativas a los modelos que se desean estimar, eliminando así del ruido estos componentes.

En este contexto, el algoritmo PCNM (Dray, S., Legendre, P. y Peres Neto 2006) es un caso particular del método MEM, en donde la matriz W tiene la siguiente forma:

$$W = \begin{cases} d_{ij} & \text{si } d_{ij} \leq t \\ 4t & \text{si } d_{ij} > t \end{cases}$$

Donde d_{ij} es la distancia euclídea entre los puntos i y j , y t es un umbral definido por el investigador, es el límite que define qué polígonos son considerados como vecinos.

Una matriz de distancias es simétrica y con todos los elementos positivos a excepción de la diagonal en donde valen cero. Al estar truncada en este caso, pueden obtenerse valores propios negativos, por lo que la matriz W no puede ser representada enteramente en el espacio euclídeo. En cualquier caso, los valores propios positivos representan los componentes euclidianos de los vínculos de vecindad de la matriz truncada.

En el problema presentado en este trabajo se utilizarán los factores resultantes de este algoritmo para analizar la tendencia global en la superficie, utilizándolos como base para la construcción de las regiones de imputación.

4. Metodología

Previo a la implementación del algoritmo PCNM, se realiza un análisis exploratorio de los datos calculando los I de Moran para varias estructuras de vecinos, y se presenta un variograma con el fin de justificar la regionalización previa a la imputación.

Luego, el primer paso es definir la matriz de conectividad, es decir, qué criterio tomar para establecer la vecindad o cercanía entre observaciones. Se opta por definir los

vecinos basados en la distancia, y en un principio se toma un umbral de 1000 metros. La matriz W es la asociada con el algoritmo PCNM.

Se implementa el algoritmo PCNM y luego se estima un modelo de Poisson, que tiene como variable de respuesta al total de población elegible del programa TUS Doble y como variables explicativas a los factores encontrados con el algoritmo PCNM. Se analiza el patrón espacial de los factores que son significativos en el modelo.

Por último se presentan posibles regionalizaciones considerando diferentes factores con diferentes distancias, y se discute la pertinencia de considerar diferentes umbrales para regionalizar, dadas algunas particularidades de los datos en algunas partes del mapa. Queda abierto a discusión qué criterio utilizar para elegir la mejor regionalización, problema que al momento no se ha resuelto en la presente investigación.

5. Resultados

5.1 Análisis exploratorio de los Datos

En esta sección se presenta el variograma obtenido para la variable TUS Doble tomando como marco de referencia la mapa completo de Montevideo, y los correlogramas utilizando el Índice de Moran.

El variograma es una herramienta de Geoestadística que permite analizar el comportamiento espacial de una variable sobre un área definida (Bohling, G. 2005), obteniendo como resultado un variograma experimental que refleja la distancia máxima y la forma en que un punto tiene influencia sobre otro punto a diferentes distancias. Si bien los datos a tratar son datos de área, se utiliza en este caso como ejemplo para mostrar el problema de la sensibilidad de las técnicas de estadística espacial al área definida como marco de análisis.

El variograma empírico se define como:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(u_i + h) - z(u_i)]^2$$

Donde $N(h)$ es la cantidad de pares de puntos separados por la distancia h , y u_i es el punto en donde se observa la variable z . Es decir, es la media de los cuadrados de las diferencias entre pares de muestras separados por una distancia h .

El variograma se analiza observando tres características que se definen a continuación:

- El umbral (*sill*): es el valor del variograma en donde la varianza se vuelve constante.
- Rango: distancia en la cual el variograma alcanza el umbral.
- *Nugget*: en teoría el variograma en cero debería ser cero. Si es diferente de cero, al valor del variograma en la distancia cero se le denomina “*nugget*”, y representa la variabilidad en distancias menores a las tomadas en los puntos muestrales, incluyendo los errores de medida.

Existe dependencia espacial para aquellas distancias menores al rango. Una vez que se alcanza este valor, el umbral, la varianza se hace constante.

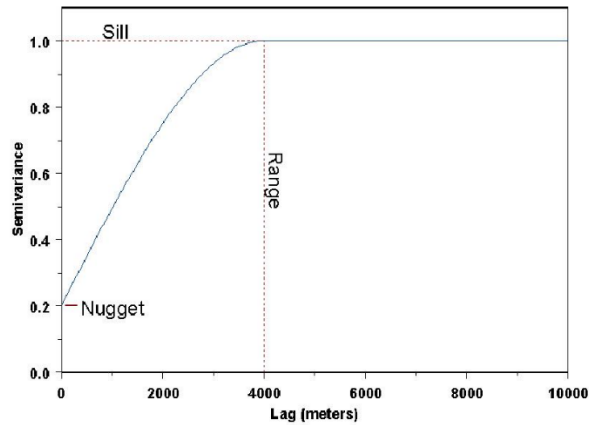


Figura 5: Interpretación gráfica del Variograma

Para poder calcular el variograma se deben tomar los centroides de los polígonos. Para la variable TUS Doble, tomando como marco el mapa de la Figura 2, el variograma obtenido es el siguiente:

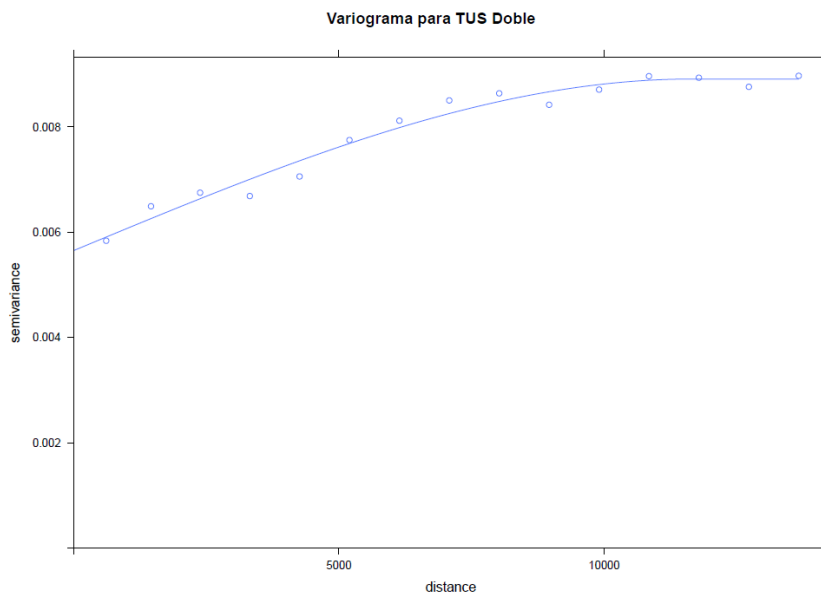


Figura 6: Variograma Empírico para TUS Doble

Se observa un efecto *nugget* importante, que puede ser explicado por la irregularidad de los puntos (no son equidistantes en el plano). Pero lo más importante es que marca una dependencia espacial de aproximadamente 10.000 metros. Esto concuerda con las distancias en el mapa entre los puntos donde se concentran casos de población elegible para el programa TUS Doble, y muestra la sensibilidad del método a la escala del mapa elegido.

Para la imputación final, sería más adecuado un modelo local. Al observarse distintos patrones de omisión en diferentes zonas, los modelos utilizados para la imputación y

para la propia descripción de los datos podrían ser diferentes. De aquí la importancia de la construcción de las regiones de imputación.

Por otra parte, si se observa el correlograma calculado con el Índice de Moran, tomando como rezago el criterio de contigüidad (Figura 7) se llega a una conclusión similar. Los valores cercanos a cero son valores del Índice que no son significativos, es decir, que no se puede rechazar la hipótesis de no autocorrelación espacial. A medida que se incrementa el rezago, el índice baja, hasta acercarse a cero en el rezago 18, para luego volver a subir en el rezago 30 aproximadamente. Al considerar como marco el mapa completo, el índice vuelve a ser significativo en rezagos en donde encuentra correlación, aunque se encuentren alejados en distancia.

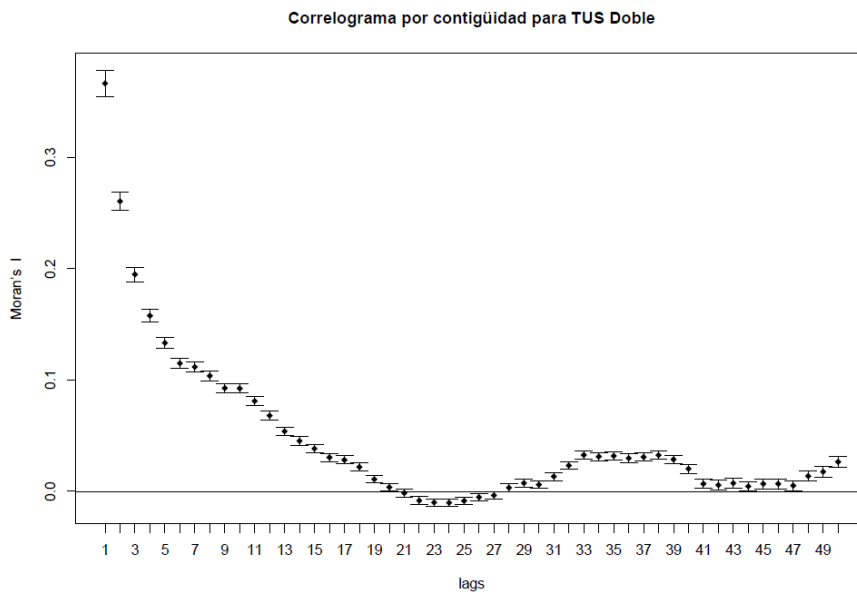


Figura 7: Correlograma basado en el Índice de Moran para TUS Doble

5.2 Construcción de la matriz W

Los pesos de la matriz W se encuentran determinados por la matriz de distancias que utiliza el algoritmo PCNM. En este apartado se presentan algunos ejemplos gráficos, para una zona en particular del mapa, considerando como umbrales de 100, 200 y 300 metros.

Grafo de vecinos con umbral 100 metros

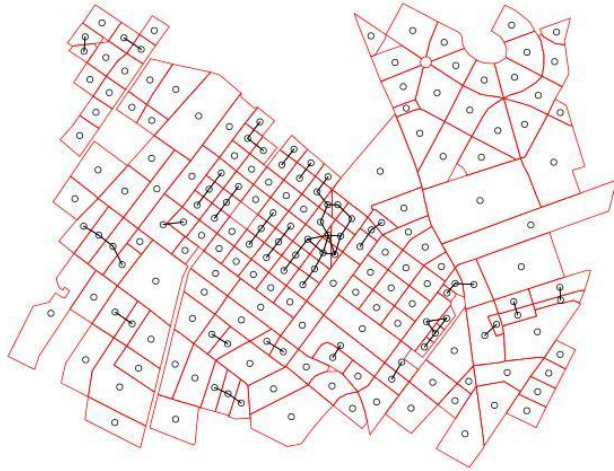


Figura 8: Grafo de vecinos para una distancia de 100 metros

Grafo de vecinos con umbral 200 metros

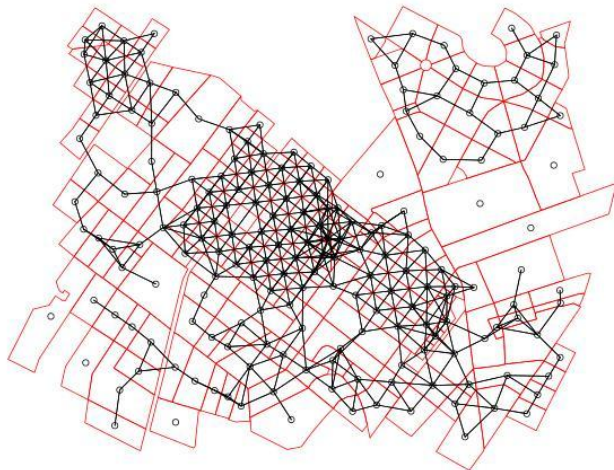


Figura 9: Grafo de vecinos para una distancia de 200 metros

Grafo de vecinos con umbral 300 metros

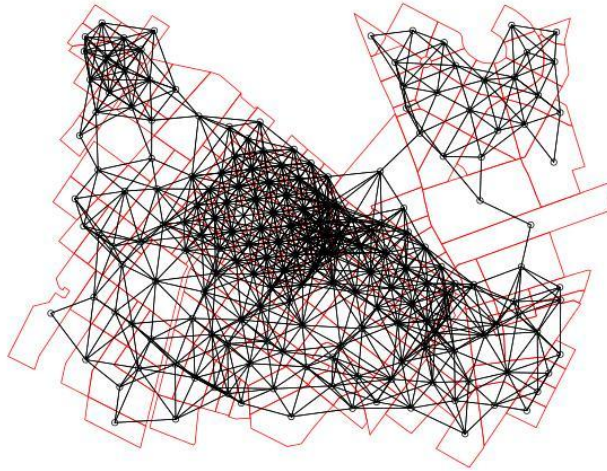


Figura 10: Grafo de vecinos para una distancia de 300 metros

5.3 Implementación del algoritmo PCNM

Una vez construida la matriz W se implementa el algoritmo PCNM. Para ello se utiliza el paquete `codep` (Guenard, Legendre y Pages 2015) del software estadístico R (R Core Team 2016). Como resultado, se obtienen los vectores propios que, según sus valores propios, se encuentran asociados a diferentes escalas de variabilidad en el mapa. Cuanto mayor es el valor propio, más asociado se encuentra a tendencias globales. Cabe recordar que este análisis es independiente de la variable de interés, solo depende de la matriz W .

La interpretación de los factores no es sencilla. Son los vectores propios de la matriz $\Omega = (I - 11^T/N)W(I - 11^T/N)$ que maximizan el Índice de Moran independientemente de la variable a la que se le esté midiendo la autocorrelación. La ausencia de una variable de respuesta dificulta la interpretación, que debe basarse en la matriz de conectividad. El grafo que se genera al definir la matriz W determina los posibles valores del Índice de Moran, y los resultados del algoritmo PCNM. La bibliografía se concentra en la utilización de los factores para filtrado de los datos, pero no en su interpretación en sí misma. En el siguiente apartado, se intentará dar una interpretación a los factores que sean significativos en el modelo de Poisson, para explicar el comportamiento de la variable TUS Doble, considerando que en este caso la interpretación puede ser más simple.

5.4 Construcción del Modelo Poisson y caracterización de factores significativos.

La regresión de Poisson es un modelo lineal generalizado en el que la variable de respuesta tiene una distribución de Poisson y el logaritmo de su valor esperado puede ser modelado por una combinación lineal de parámetros desconocidos. Se utiliza para modelizar datos de conteo.

Se asume que $Y|X \sim \text{Poisson}(e^{\alpha x + \beta})$, por lo tanto $E(Y|X) = e^{\alpha x + \beta}$. Realizando una transformación logarítmica se llega a que

$$\log(E(Y|X)) = \alpha x + \beta$$

En nuestro caso la variable de respuesta Y es la cantidad de hogares que son población elegible del programa TUS Doble por zona.

El modelo especificado toma en cuenta como variables explicativas sólo a los factores obtenidos con el algoritmo PCNM. Del total de factores incluidos (9927) sólo 27 son significativos en el modelo (ver Anexo 2).

De los factores significativos se toman aquellos de menor orden, ya que son los que captan una estructura más global. A continuación se presentan tres ejemplos para los vectores 26, 15 y 12, en el orden que aparecen en el modelo. Se grafican las coordenadas positivas y negativas de cada vector propio, que en un contexto de análisis factorial, serían las características discriminantes del factor.

Si se observa la Figura 11, el vector 26 con coordenadas negativas, a excepción de la zona de Carrasco, separa una región en donde se encuentran la mayor parte de los casos de TUS Doble. En la zona de Carrasco Norte, se encuentra un punto, no visible en la escala de los mapas actuales, con un valor alto de población elegible para TUS Doble. Al considerar un umbral de 1000 metros, no es posible separar este fenómeno de las zonas dentro de ese radio, por eso queda incluida con coordenadas negativas. Por este motivo luego se presentarán los mapas resultantes de PCNM con un umbral de 500 metros.

Vector 26 PCNM 1000 metros

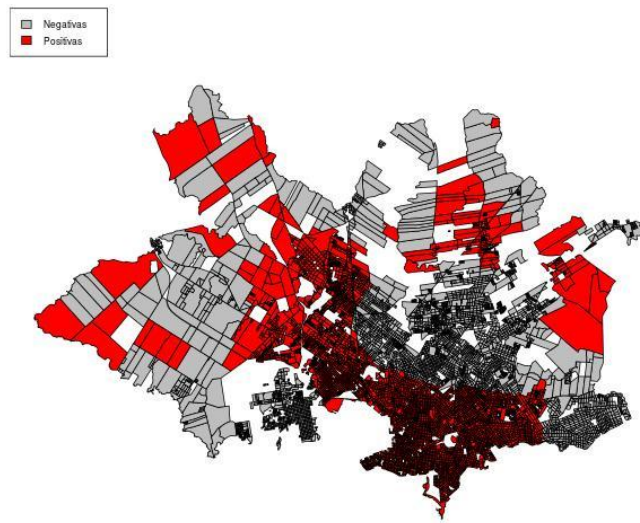


Figura 11: Vector Propio 26 para umbral de 1000 metros

Vector 15 PCNM 1000 metros

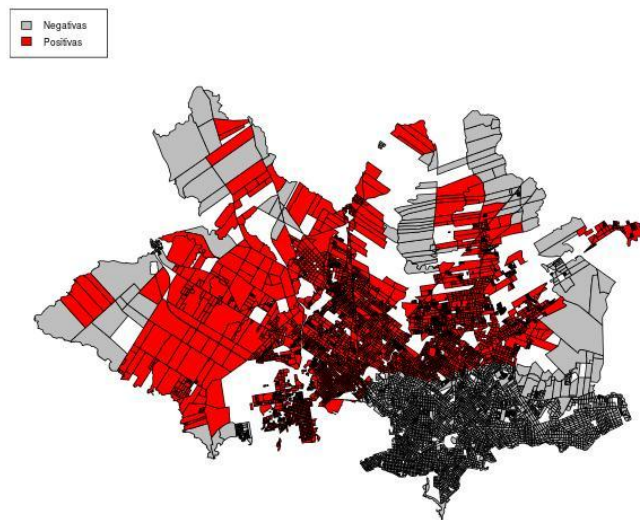


Figura 12: Vector Propio 15 para umbral de 1000 metros

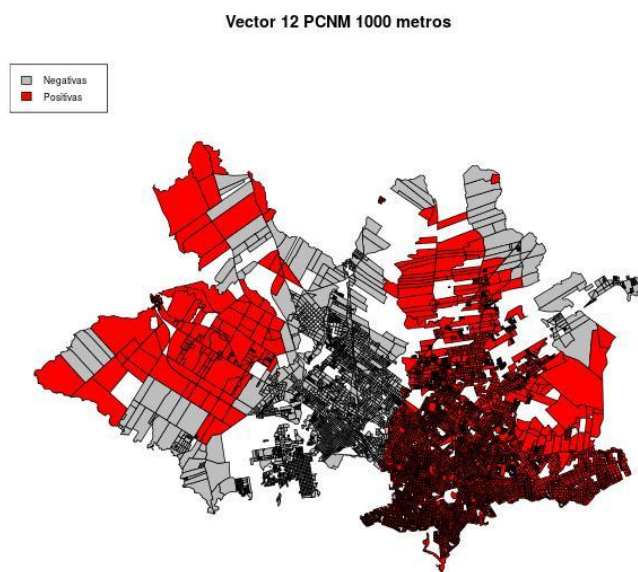


Figura 13: Vector Propio 12 para umbral de 1000 metros

El vector 15 (Figura 12) delimita una región más precisa con coordenadas positivas, en donde también se concentra una densidad importante de población elegible para TUS Doble. En la zona del Cerro Norte, al haber un área omisa de gran tamaño, es muy difícil como veremos a continuación con otros ejemplos además del presente, que se obtenga una región continua en la superficie.

El vector 12 (Figura 13) separa con coordenadas negativas a una región central, con algunas excepciones.

De combinar los mapas anteriores surgen diferentes regionalizaciones, de las cuales se presenta la obtenida de combinar los vectores 15 y 26 (Figura 14).

Al considerar el vector 26 en la zona Este del departamento persiste el problema mencionado anteriormente, pero la zona centro presenta dos regiones bien delimitadas, que podrían considerarse como regiones de imputación. El problema es con qué criterio elegir estas regiones, y a su vez, notando que el método puede ser sensible a datos atípicos, ¿se deben construir las regiones con un único criterio de distancia o pueden combinarse varios umbrales para una mejor representación de los datos? Esta es una pregunta que al momento no tiene una respuesta, pero que plantea las direcciones a seguir en la presente investigación.

Reginalización a partir de vectores 15 y 26 de PCNM 1000 metros

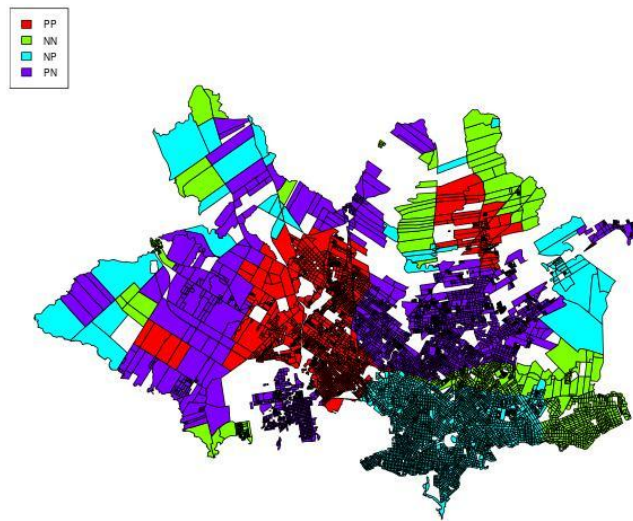


Figura 14: Regiones obtenidas al combinar los vectores 15 y 26 para un umbral de 1000 metros

Para un umbral de 500 metros se obtienen 68 vectores significativos, apareciendo primero en el modelo factores de orden alto (por ejemplo 300) que captan una variabilidad a nivel más local que no es de interés en este caso para construir las posibles regiones de imputación. De los de menor orden, aparecen como significativos los vectores propios 22, 23, 27 y 30, y se presentan a continuación en las figuras 15 a 18.

Los vectores 22 (figura 15) y 23 (figura 16) son similares, y ambos separan, uno con coordenadas positivas y el otro con negativas, zonas clave en donde puede haber una alta densidad de población elegible del programa TUS Doble, en la zona Rural de Montevideo, a excepción de algunos casos.

El vector 27 (figura 17) es similar al vector 26 (figura 11) obtenido con un umbral de 1000 metros, pero se contrae excluyendo por ejemplo la zona de Carrasco incluida con el umbral de 1000 metros. Esta contracción se da en todo el mapa, por lo que el área de coordenadas negativas es menor a la del umbral de 1000 metros.

El mapa del factor 30 (figura 18) es similar que al factor 15 (figura 12) con un umbral de 1000 metros, y al igual que en el caso anterior, el área gris se achica debido a que se está utilizando un umbral de 500 metros.

Vector 22 PCNM 500 metros

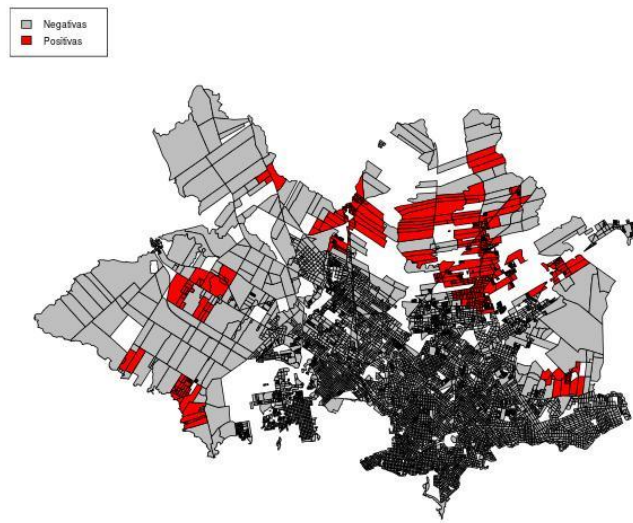


Figura 15: Vector Propio 22 para umbral de 500 metros

Vector 23 PCNM 500 metros



Figura 16: Vector Propio 23 para umbral de 500 metros

Vector 27 PCNM 500 metros

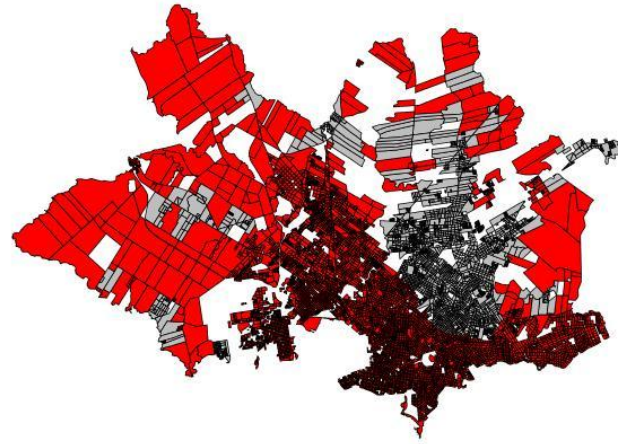
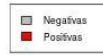


Figura 17: Vector Propio 27 para umbral de 500 metros

Vector 30 PCNM 500 metros

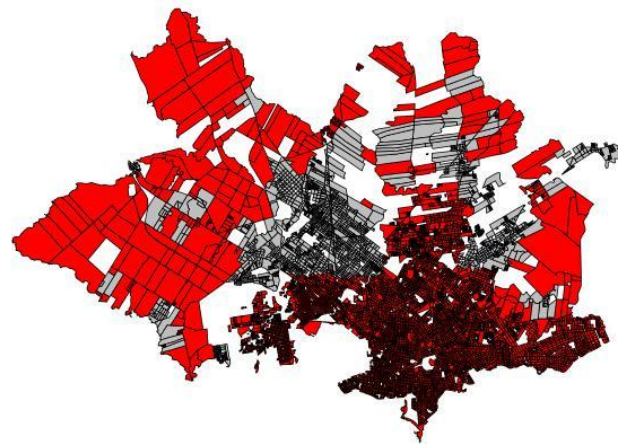
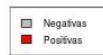


Figura 18: Vector Propio 30 para umbral de 500 metros

De combinar los factores 27 y 30, se obtienen las siguientes regiones (Figura 19), en donde también se observan dos regiones centrales bien definidas, de forma similar que en el caso del umbral de 1000 metros con los vectores 15 y 26 (ver Figura 14).

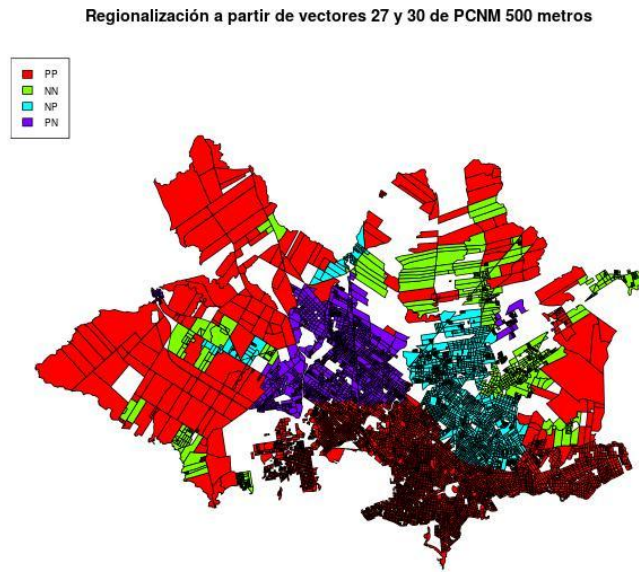


Figura 19: Regiones obtenidas al combinar los vectores 27 y 30 para un umbral de 500 metros

Por último se presentan los resultados para un umbral intermedio, de 700 metros. En este caso, el total de vectores propios significativos en el modelo de Poisson es de 29. De los de menor orden, son significativos los vectores propios 16, 20, 24, 25, 28 y 29. A continuación se presentan algunos de los casos de los vectores significativos (figuras 20, 21 y 22), y se observará que al igual que en los casos anteriores, que las regiones con coordenadas negativas, se contraen respecto al umbral de 1000 metros o se expanden respecto al umbral de 500 metros.

Las regiones producto de combinar algunos de los factores anteriores presentan al igual que en los casos anteriores dos regiones bien definidas en el centro del departamento (ver figura 23), que pueden variar en tamaño respecto a los umbrales anteriores.

Vector 16 PCNM 700 metros

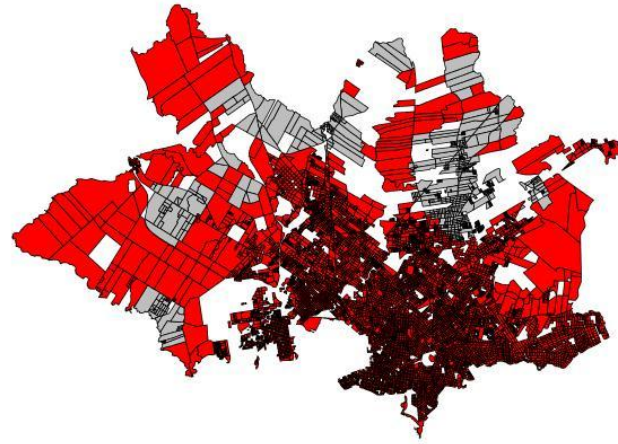
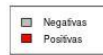


Figura 20: Vector Propio 16 para umbral de 700 metros

Vector 20 PCNM 700 metros

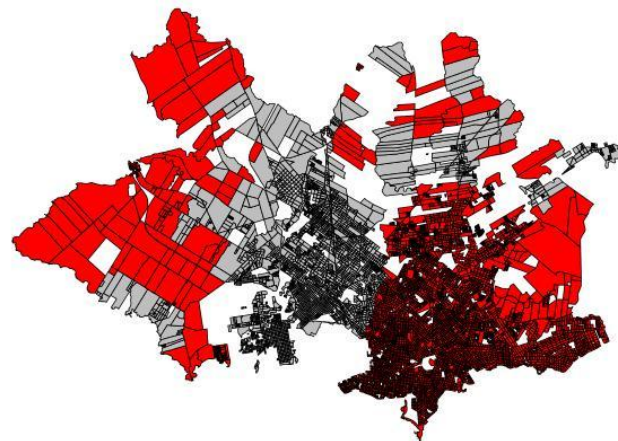
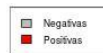


Figura 21: Vector Propio 20 para umbral de 700 metros

Vector 24 PCNM 700 metros

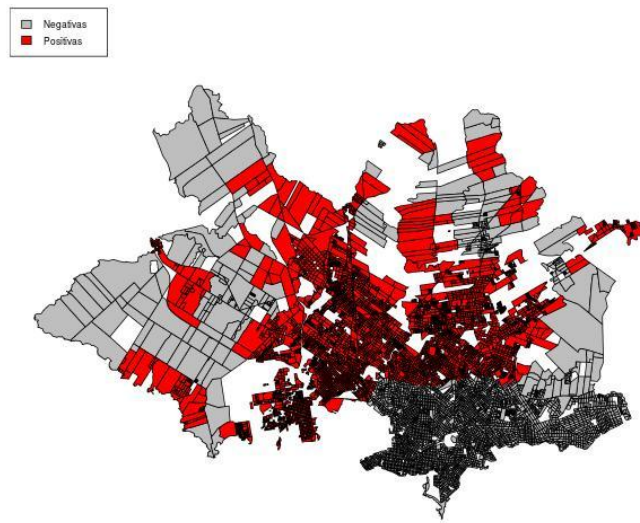


Figura 22: Vector Propio 24 para umbral de 700 metros

Reginización 1 a partir de vectores 20 y 24 de PCNM 700 metros

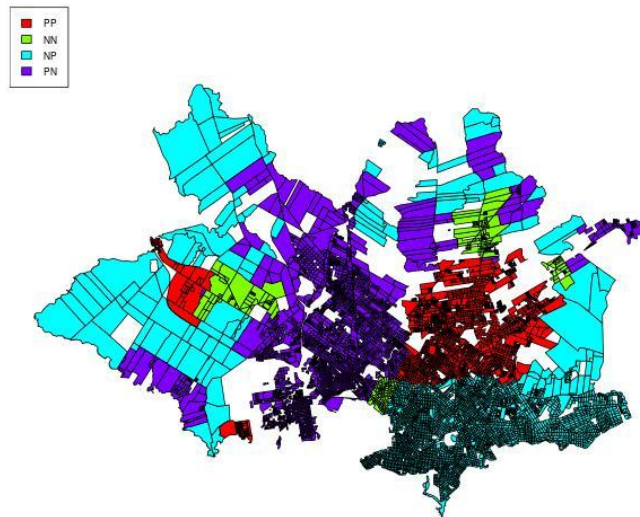


Figura 23: Regiones obtenidas al combinar los vectores 20 y 24 para un umbral de 700 metros

6. Discusión y líneas futuras de investigación

A grandes rasgos, se puede decir que dentro de los vectores correspondientes a escalas globales, se encuentran tres tipos:

- Los que captan las zonas más periféricas del departamento, zonas rurales, que son las más irregulares en términos de polígonos, y las más grandes en área.
- Los que fraccionan el mapa en una región “Norte - Sur”(figuras 12, 18 y 22) y por último
- Los que separan una zona central (figuras 13 y 21).

De estas tres grandes tendencias podrían construirse regiones bien definidas para la imputación. El problema a discutir a futuro es encontrar un criterio que defina automáticamente estas regiones en base a los vectores propios encontrados utilizando el algoritmo PCNM.

Dada la sensibilidad a los umbrales en algunas partes del mapa, debido a la existencia de puntos atípicos, se debería considerar la posibilidad de combinar vectores propios de distintos umbrales como una posibilidad a resolver este problema.

También podría incorporarse un modelo sencillo para la imputación de los datos, con el fin de evaluar las regiones. Es decir, la delimitación de las zonas de imputación podría depender del modelo final a ajustar, y realizar la imputación y regionalización en forma simultánea, utilizando criterios estándar para la evaluación de modelos, como ser el AIC, BIC, entre otros.

Otra posibilidad sería implementar análisis de cluster, con una restricción de contigüidad espacial entre las observaciones, utilizando como variables los vectores propios obtenidos con el algoritmo PCNM.

Cabe destacar la importancia del investigador en la definición de la matriz W . Las conclusiones finales dependen enteramente de esta definición, y probar con varias matrices es necesario para una decisión final. También, el algoritmo PCNM no es el único caso, podría analizarse qué sucede con otras matrices de pesos como las definidas por contigüidad o las binarias.

Las estrategias mencionadas anteriormente no son las únicas, quedando abierta la posibilidad (y a discusión) de aplicar otras técnicas o criterios para lograr el objetivo de la construcción de regiones de imputación.

Bibliografía

- [1] Guyon, X y Gaetan, C. (2011) *Spatial Statistics and Modeling*. Springer.
- [2] Dray S., Legendre P. Peres - Neto, P. (2006). “*Spatial Modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM)*”, *Ecological Modelling*, Vol. 196, pp 483-493.
- [3] De Jong, P., Sprenger, C. y van Veen, F. “*On Extreme Values of Moran’s I and Geary’s c*”, *Geographical Analysis*, Volume 16, Issue 1.
- [4] Ministerio de Desarrollo Social (2013). *¿Qué es el Índice de Carencias Críticas?* Serie de documentos “Aportes a la conceptualización de la pobreza y la focalización de las políticas sociales en Uruguay”.
- [5] Bai, H., Carrasco, P. y Colafranceschi, M. (2014) “*Aplicación del Índice de Carencias Críticas al CENSO y análisis del vínculo con NBI*”. IECON -MIDES.
- [6] Bohling, G. (2005). *Introduction to Geostatistics and Variograma Analysis*. Notas de clase, Kansas University.
- [7] Bivand, R., Pebesma, E. y Gómez- Rubio, E. (2013). *Applied Spatial Data Analysis with R*. Springer.
- [8] Guenard, G., Legendre, P. y Pages, B. (2015). codep: Multiscale Codependence Analysis. R package version 0.5-1. <https://CRAN.R-project.org/package=codep>
- [9] R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

Anexo 1

CCZ	Total de zonas sin Información	%	Morador ausente	%	Formato Papel	%	Zonas Mixtas	%	Omisas	%
1	1	0,2	1	5	0	0	0	0	0	0
2	1	0,2	1	5	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0
5	2	0,3	2	10	0	0	0	0	0	0
6	9	1,4	0	0	1	0,5	0	0	8	5,7
7	1	0,2	0	0	0	0	0	0	1	0,7
8	4	0,6	2	10	1	0,5	0	0	1	0,7
9	212	33,5	5	25	53	29	118	40,7	36	26
10	55	8,7	2	10	26	14	15	5,2	12	8,6
11	16	2,5	1	5	0	0	0	0	15	11
12	24	3,8	2	10	12	6,6	5	1,7	5	3,6
13	10	1,6	0	0	1	0,5	1	0,35	8	5,7
14	7	1,1	1	5	0	0	0	0	6	4,3
15	2	0,3	0	0	1	0,5	0	0	1	0,7
16	1	0,2	0	0	1	0,5	0	0	0	0
17	276	43,7	2	10	86	47	150	51,7	38	27
18	11	1,7	1	5	0	0	1	0,35	9	6,4
Total	632	100	20	100	182	100	290	100	140	100

En el comunal 9 se encuentran los barrios de Curva de Maroñas, Flor de Maroñas, Jardines del Hipódromo, Ideal, Bella Italia, Málaga, Ituzaingó, Industrial, Punta de Rieles, Km. 16 Cno. Maldonado y Villa García.

Los barrios del comunal 17 son Casco del Cerro, Casabó, Pajas Blancas, Santa Catalina, Cerro Norte, La Boyada, Cerro Oeste y parte de la zona rural de Montevideo.

Anexo 2

Modelo con umbral de 1000 metros

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-0.40050	0.01431	-27.987	< 2e-16	***
fitted(f1)vec26	-58.80149	1.03314	-56.915	< 2e-16	***
fitted(f1)vec15	40.33419	1.06787	37.771	< 2e-16	***
fitted(f1)vec47	-3.69190	0.37788	-9.770	< 2e-16	***
fitted(f1)vec32	-11.58145	0.98903	-11.710	< 2e-16	***
fitted(f1)vec161	10.39429	0.59909	17.350	< 2e-16	***
fitted(f1)vec61	15.62651	1.14431	13.656	< 2e-16	***
fitted(f1)vec12	-23.49607	1.03504	-22.701	< 2e-16	***
fitted(f1)vec57	14.21107	0.97822	14.527	< 2e-16	***
fitted(f1)vec91	-9.25815	0.90364	-10.245	< 2e-16	***
fitted(f1)vec147	-11.43596	0.98263	-11.638	< 2e-16	***
fitted(f1)vec62	15.22277	2.19380	6.939	3.95e-12	***
fitted(f1)vec25	-11.46230	0.88651	-12.930	< 2e-16	***
fitted(f1)vec17	23.63240	3.32034	7.117	1.10e-12	***

fitted(f1)vec34	-11.96077	1.14140	-10.479	< 2e-16	***
fitted(f1)vec178	10.40485	0.72325	14.386	< 2e-16	***
fitted(f1)vec267	9.43465	0.92806	10.166	< 2e-16	***
fitted(f1)vec292	-5.15380	0.96896	-5.319	1.04e-07	***
fitted(f1)vec9532	5.42474	1.06928	5.073	3.91e-07	***
fitted(f1)vec123	-13.80441	1.09535	-12.603	< 2e-16	***
fitted(f1)vec66	-9.25591	1.22643	-7.547	4.45e-14	***
fitted(f1)vec9668	-6.07721	1.00883	-6.024	1.70e-09	***
fitted(f1)vec81	-10.77122	1.26672	-8.503	< 2e-16	***
fitted(f1)vec269	-5.03877	0.93104	-5.412	6.23e-08	***
fitted(f1)vec9597	-9.64748	1.11012	-8.690	< 2e-16	***
fitted(f1)vec9519	-6.77517	1.02706	-6.597	4.21e-11	***
fitted(f1)vec139	-5.72062	1.19852	-4.773	1.81e-06	***
fitted(f1)vec40	-17.90215	1.34872	-13.273	< 2e-16	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Modelo con umbral de 500 metros

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-0.72107	0.01982	-36.375	< 2e-16	***
fitted(f2)vec36	65.61882	1.56205	42.008	< 2e-16	***
fitted(f2)vec39	-83.09832	2.12252	-39.151	< 2e-16	***
fitted(f2)vec336	1.56912	0.34398	4.562	5.07e-06	***
fitted(f2)vec27	-22.99226	0.88152	-26.083	< 2e-16	***
fitted(f2)vec140	-8.83392	0.64543	-13.687	< 2e-16	***
fitted(f2)vec127	-10.43479	0.56635	-18.425	< 2e-16	***
fitted(f2)vec30	-26.97552	1.03640	-26.028	< 2e-16	***
fitted(f2)vec40	-20.19869	0.86878	-23.250	< 2e-16	***
fitted(f2)vec34	-24.28693	1.78829	-13.581	< 2e-16	***
fitted(f2)vec78	-28.70702	1.13923	-25.199	< 2e-16	***
fitted(f2)vec77	-25.26727	1.11151	-22.732	< 2e-16	***
fitted(f2)vec79	18.49002	1.14690	16.122	< 2e-16	***
fitted(f2)vec142	9.83386	0.92753	10.602	< 2e-16	***
fitted(f2)vec244	-12.21693	0.83909	-14.560	< 2e-16	***
fitted(f2)vec31	33.00978	1.24772	26.456	< 2e-16	***
fitted(f2)vec128	12.55065	1.13286	11.079	< 2e-16	***
fitted(f2)vec111	-2.77423	0.54942	-5.049	4.43e-07	***
fitted(f2)vec365	-5.22457	0.83405	-6.264	3.75e-10	***
fitted(f2)vec423	10.16462	0.97077	10.471	< 2e-16	***
fitted(f2)vec253	10.23254	1.18218	8.656	< 2e-16	***
fitted(f2)vec165	12.97042	1.19952	10.813	< 2e-16	***
fitted(f2)vec186	2.37626	1.35990	1.747	0.0806	.
fitted(f2)vec87	-7.39269	0.92553	-7.988	1.38e-15	***
fitted(f2)vec204	11.63086	1.76574	6.587	4.49e-11	***
fitted(f2)vec182	-9.31056	0.88121	-10.566	< 2e-16	***
fitted(f2)vec440	6.35566	1.12950	5.627	1.83e-08	***
fitted(f2)vec308	7.66739	1.38704	5.528	3.24e-08	***
fitted(f2)vec116	-10.12311	1.13881	-8.889	< 2e-16	***
fitted(f2)vec101	-7.59856	0.81345	-9.341	< 2e-16	***
fitted(f2)vec356	-5.87251	0.63930	-9.186	< 2e-16	***
fitted(f2)vec22	12.13037	0.81866	14.817	< 2e-16	***
fitted(f2)vec23	-10.53653	0.83838	-12.568	< 2e-16	***
fitted(f2)vec18	-6.13038	0.57110	-10.734	< 2e-16	***
fitted(f2)vec346	11.71111	1.23623	9.473	< 2e-16	***

fitted(f2)vec598	-8.83347	0.90274	-9.785	< 2e-16	***
fitted(f2)vec304	-5.95899	1.16625	-5.110	3.23e-07	***
fitted(f2)vec66	-5.59186	0.65615	-8.522	< 2e-16	***
fitted(f2)vec457	7.51308	0.99311	7.565	3.87e-14	***
fitted(f2)vec94	11.80652	1.20309	9.814	< 2e-16	***
fitted(f2)vec583	-8.24677	0.97796	-8.433	< 2e-16	***
fitted(f2)vec158	-14.57940	1.65533	-8.808	< 2e-16	***
fitted(f2)vec378	9.29381	0.96095	9.672	< 2e-16	***
fitted(f2)vec250	8.91524	1.09143	8.168	3.12e-16	***
fitted(f2)vec285	-11.64915	1.37911	-8.447	< 2e-16	***
fitted(f2)vec26	-7.82945	0.83104	-9.421	< 2e-16	***
fitted(f2)vec84	3.72720	1.20590	3.091	0.0020	**
fitted(f2)vec416	-9.04326	1.11450	-8.114	4.89e-16	***
fitted(f2)vec436	-8.82175	0.82264	-10.724	< 2e-16	***
fitted(f2)vec512	5.51421	0.79291	6.954	3.54e-12	***
fitted(f2)vec744	6.46995	0.81067	7.981	1.45e-15	***
fitted(f2)vec218	5.35629	0.94191	5.687	1.30e-08	***
fitted(f2)vec97	3.86297	0.68416	5.646	1.64e-08	***
fitted(f2)vec368	-10.52503	1.37319	-7.665	1.79e-14	***
fitted(f2)vec103	3.47644	0.69437	5.007	5.54e-07	***
fitted(f2)vec511	-5.92005	0.89460	-6.618	3.65e-11	***
fitted(f2)vec371	-3.97047	0.64289	-6.176	6.58e-10	***
fitted(f2)vec81	8.18956	1.19130	6.874	6.22e-12	***
fitted(f2)vec63	-21.55191	1.35918	-15.857	< 2e-16	***
fitted(f2)vec62	7.56962	0.92320	8.199	2.42e-16	***
fitted(f2)vec96	7.74083	1.07708	7.187	6.63e-13	***
fitted(f2)vec187	4.72013	0.81722	5.776	7.66e-09	***
fitted(f2)vec86	-5.76933	1.07805	-5.352	8.72e-08	***
fitted(f2)vec595	7.74003	0.96147	8.050	8.27e-16	***
fitted(f2)vec41	10.73201	0.96853	11.081	< 2e-16	***
fitted(f2)vec758	6.07010	0.83631	7.258	3.92e-13	***
fitted(f2)vec73	17.72241	1.81559	9.761	< 2e-16	***
fitted(f2)vec725	-8.21634	1.03409	-7.945	1.93e-15	***
fitted(f2)vec9847	5.96270	0.53740	11.096	< 2e-16	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Modelo con umbral de 700 metros

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-0.56851	0.01686	-33.715	< 2e-16	***
fitted(f3)vec24	91.89161	1.60341	57.310	< 2e-16	***
fitted(f3)vec28	28.70211	0.75955	37.788	< 2e-16	***
fitted(f3)vec41	9.32346	0.35671	26.137	< 2e-16	***
fitted(f3)vec136	-9.45558	0.59934	-15.777	< 2e-16	***
fitted(f3)vec143	-10.70717	0.61329	-17.459	< 2e-16	***
fitted(f3)vec25	-19.40930	0.78667	-24.673	< 2e-16	***
fitted(f3)vec90	-9.11738	0.69163	-13.183	< 2e-16	***
fitted(f3)vec152	6.95483	0.78957	8.808	< 2e-16	***
fitted(f3)vec72	17.62149	1.21029	14.560	< 2e-16	***
fitted(f3)vec46	11.41951	0.78576	14.533	< 2e-16	***
fitted(f3)vec55	-25.17346	1.24526	-20.215	< 2e-16	***
fitted(f3)vec70	18.44098	1.09770	16.800	< 2e-16	***
fitted(f3)vec73	-19.14056	1.37127	-13.958	< 2e-16	***
fitted(f3)vec252	-9.86242	1.00641	-9.800	< 2e-16	***


```

fitted(f3)vec134  16.32362    1.34379   12.147 < 2e-16 ***
fitted(f3)vec29   -8.80406    0.70573  -12.475 < 2e-16 ***
fitted(f3)vec16  -13.80225    0.77922  -17.713 < 2e-16 ***
fitted(f3)vec63  -19.27753    1.09864  -17.547 < 2e-16 ***
fitted(f3)vec62  -39.79534    1.77563  -22.412 < 2e-16 ***
fitted(f3)vec20  -28.42629    2.88792   -9.843 < 2e-16 ***
fitted(f3)vec192  -6.89919    0.97088   -7.106 1.19e-12 ***
fitted(f3)vec127  -7.78207    0.84677   -9.190 < 2e-16 ***
fitted(f3)vec99   7.47033    0.98943    7.550 4.35e-14 ***
fitted(f3)vec160 -10.61599    0.95048  -11.169 < 2e-16 ***
fitted(f3)vec392  9.20491    0.88153   10.442 < 2e-16 ***
fitted(f3)vec275  -7.22835    0.82366   -8.776 < 2e-16 ***
fitted(f3)vec145  10.89155    1.10394    9.866 < 2e-16 ***
fitted(f3)vec175  -7.48780    0.92901   -8.060 7.63e-16 ***
fitted(f3)vec48  -28.47439    1.48279  -19.203 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Instituto de Estadística

Documentos de Trabajo



Eduardo Acevedo 1139. CP 11200 Montevideo, Uruguay

Teléfonos y fax: (598) 2410 2564 - 2418 7381

Correo: ddt@iesta.edu.uy

www.iesta.edu.uy

Área Publicaciones

Noviembre, 2016

DT-2016/2