**Análisis espacial de factores asociados al crimen y la vulnerabilidad social en Boston**

**Esteban Vargas y Miguel Coto**

**Resumen**

Se tomaron diferentes variables asociadas al fenómeno de vulnerabilidad social en Boston, para a partir de ellas, modelar de forma espacial el nivel de criminalidad en la zona. Mediante la prueba de Moran, se obtuvo evidencia de autocorrelación espacial lo cual permitió evaluar distintos modelos, de los cuales el Durbin fue que realizó un mejor ajuste. Las variables de cantidad de población de color y cantidad de personas con enfermedades médicas fueron las de mayor aporte al modelaje espacial.

**Palabras clave**

Estadística de áreas, prueba de Moran, autocorrelación espacial.

**Introducción**

La vulnerabilidad social es un tipo de susceptibilidad desproporcionada de algunos grupos sociales a los impactos de los peligros, como, por ejemplo, la muerte, las lesiones, la pérdida o la interrupción de los medios de vida. La noción del término tiene su origen en estudios sobre desastres naturales y en diversos análisis acerca de cómo las comunidades rurales pobres enfrentan las consecuencias de eventos socioeconómicos traumáticos. El Centro Latinoamericano para el Desarrollo CEPAL (2002) distingue cuatro usos que se le suelen dar al concepto de vulnerabilidad social, descrita a continuación.

1. Se utiliza el concepto de vulnerabilidad con referencia a grupos específicos de la población en el análisis y en las políticas sociales:
2. Los grupos en situación de riesgo social están conformados por individuos que son más propensos a presentar conductas anómicas (por factores propios de su ambiente doméstico o comunitario), a sufrir diversos daños por acción u omisión de terceros, o a tener desempeños deficientes en esferas claves para la inserción social
3. Para delimitar segmentos de la población que, por el hecho de actuar según un patrón de conducta común, tienen mayor probabilidad de ser objeto de algún evento dañino.
4. Para identificar colectivos que comparten algún atributo básico común que se supone que genera problemas similares.

También se asocia a expresiones sobre daño, fragilidad, indefensión, desamparo, debilidad y se traslapa en varios aspectos como el de la exclusión, pero hace referencia a las condiciones de los individuos, hogares o comunidades que tienen mayor propensión de ser afectados de forma negativa y con menor capacidad de respuesta ante cambios del entorno, los cuales los puede conducir a situaciones de pobreza o afectar los pilares que definen la integración social en la actualidad, es decir el empleo, el acceso a los servicios de salud, de educación, de protección social y de vivienda (Mosquera, N Rengifo, N, et al. 2013). (Mosquera, et al, 2013) destacan situaciones familiares y sociales asociadas a vulnerabilidad, relacionadas con bajos niveles educativos y falta de empleo de los padres o acudientes, falta de provisión de servicios públicos y básicos de salud, hacinamiento, desprotección social y de vivienda, situaciones que constituyen factores que vulneran el desarrollo humano de las familias.

La vulnerabilidad social incluye a las personas que tienen más probabilidades de sufrir desproporcionadamente debido a sus circunstancias sociales existentes, como las relacionadas con la edad, el género, la raza, las enfermedades médicas, la discapacidad, la alfabetización y el dominio del inglés, y el aislamiento social. Sus circunstancias aumentan la probabilidad de aislamiento social, que es la falta de compromiso en los lazos sociales, las conexiones institucionales o la participación comunitaria. El aislamiento social en la vida cotidiana o posterior al desastre se correlaciona directamente con una mayor mortalidad. (Martin, S. 2014).

A nivel latinoamericano, para comprenderlo desde un punto de vista mas cercano a nuestra cotidianeidad, un estudio realizado por Kaztman (2000) destaca que, en la mayoría de los países de la zona, las oficinas de estadística revelan en forma regular encuestas de hogares, para intentar detectar el grado de vulnerabilidad social y familiar. Los gobiernos hacen uso de estas mediciones para diseñar políticas públicas con el fin de mejorar la calidad de vida de las poblaciones más vulnerables en cada uno de estos países.

El objetivo del presente escrito se centra en establecer relacione entre diferentes factores asociados a la vulnerabilidad social. A partir de datos espaciales de la corporación “Analyze Boston”, se obtuvo un registro de una serie de indicadores sobre variables asociadas a este fenómeno para el cual se realizó un análisis espacial tomando como división las zonas delimitadas de la ciudad de Boston. Se hizo el supuesto de que los datos que contienen pocos o ningún error, es decir, que los resultados son fiables. A pesar de esto se toma en consideración que casi todos los datos tienen errores que pueden surgir al medir tanto la ubicación (puntos, líneas, áreas) como las propiedades de atributo de los objetos espaciales.

Según la base de datos de Climate Ready Boston, los grupos identificados como más vulnerables son adultos mayores, niños, personas de color, personas con dominio limitado del inglés, personas con bajos ingresos o sin ingresos, personas con discapacidades y personas con enfermedades médicas. Estos indicadores se contrastaron con los crímenes en la región para obtener a partir de modelos espaciales, la mejor estructura que describiera la correlación espacial entre el fenómeno y los indicadores de vulnerabilidad.

**Definiciones de términos asociados a vulnerabilidad:**

**Crímenes:**

Corresponden a registros del nuevo sistema de informe de incidentes delictivos, que incluye un conjunto de datos sobre el tipo de incidente, así como cuándo y dónde ocurrió.

**Adultos mayores:**

Este tipo de población es más vulnerable en un evento climático; sufren tasas más altas de enfermedad médica que el resto de la población y pueden tener algunas limitaciones funcionales en un escenario de evacuación, así como también cuando se preparan y se recuperan de un desastre. Fuente de datos: datos de estimaciones de 5 años (ACS) de la Encuesta de la Comunidad Estadounidense 2008-2012 por sección censal para la población mayor de 65 años.

**Niños:**

Las familias con menores requieren recursos adicionales en un evento climático, por ejemplo, cuando se cancela la escuela, los padres necesitan opciones alternativas de cuidado de niños, lo que puede significar la falta de trabajo. Fuente de datos: 2010 Estimaciones de 5 años de la Encuesta sobre la Comunidad Estadounidense (ACS) por sección censal para la población menor de 5 años.

**Población de color:**

Este tipo de población abarca más de la mitad de la población de Boston. Estadísticamente, las personas de color tienen niveles más bajos de ingresos y niveles más altos de pobreza que la población en general. Las personas de color, muchas de las cuales también tienen un dominio limitado del inglés, además de ser mas propensas a ser rechazadas y discriminadas. Fuente de datos: Datos de estimaciones de 5 años de la Encuesta de la Comunidad Estadounidense (ACS) 2008-2012 por tramo censal: hispanos negros, nativos americanos, asiáticos, isleños, otros, multi, no blancos.

**Dominio limitado del inglés:**

Las personas con más limitaciones pueden ser más susceptibles a perder información crucial sobre cómo prepararse por peligros. En caso de inundación, por ejemplo, los residentes también pueden enfrentar desafíos para comunicarse con el personal de respuesta a emergencias. Fuente de datos: datos de estimaciones de 5 años (ACS) de la Encuesta de la Comunidad Estadounidense 2008-2012 por sector censal, definidos como habla inglés solamente o habla inglés "muy bien".

**Ingreso bajo o nulo:**

La falta de recursos financieros afecta la capacidad de un hogar para prepararse para un evento de desastre y para apoyar a amigos y vecindarios. Por ejemplo, los residentes sin televisores, computadoras o teléfonos móviles basados ​​en datos pueden enfrentar desafíos para recibir noticias sobre peligros o recursos de recuperación. Los inquilinos pueden tener problemas para encontrar y pagar depósitos para viviendas de reemplazo si su residencia se ve afectada por las inundaciones. Fuente de datos: datos de estimaciones de 5 años (ACS) de la Encuesta de la Comunidad Estadounidense 2008-2012 por sección censal para poblaciones de ingresos bajos o nulos. Los datos representan un campo calculado que combina personas que estaban 100% por debajo del nivel de pobreza y aquellas que estaban entre 100 y 149% del nivel de pobreza.

**Personas con discapacidades:**

Este tipo de sector se encuentra entre los más vulnerables en una emergencia; sostienen tasas desproporcionadas de enfermedades, lesiones y muertes en eventos de desastre. Las personas con discapacidad pueden tener dificultades para prepararse adecuadamente para un evento de desastre, incluido el traslado a un lugar más seguro. Es más probable que se queden atrás o abandonados durante las evacuaciones. Fuente de datos: datos de estimaciones de 5 años (ACS) de la Encuesta de la Comunidad Estadounidense 2008-2012 por tramo censal para la población civil no institucionalizada total, que incluyen: dificultad auditiva, dificultad para la visión, dificultad cognitiva, dificultad ambulatoria, dificultad para el autocuidado y vida independiente dificultad.

**Enfermedad médica:**

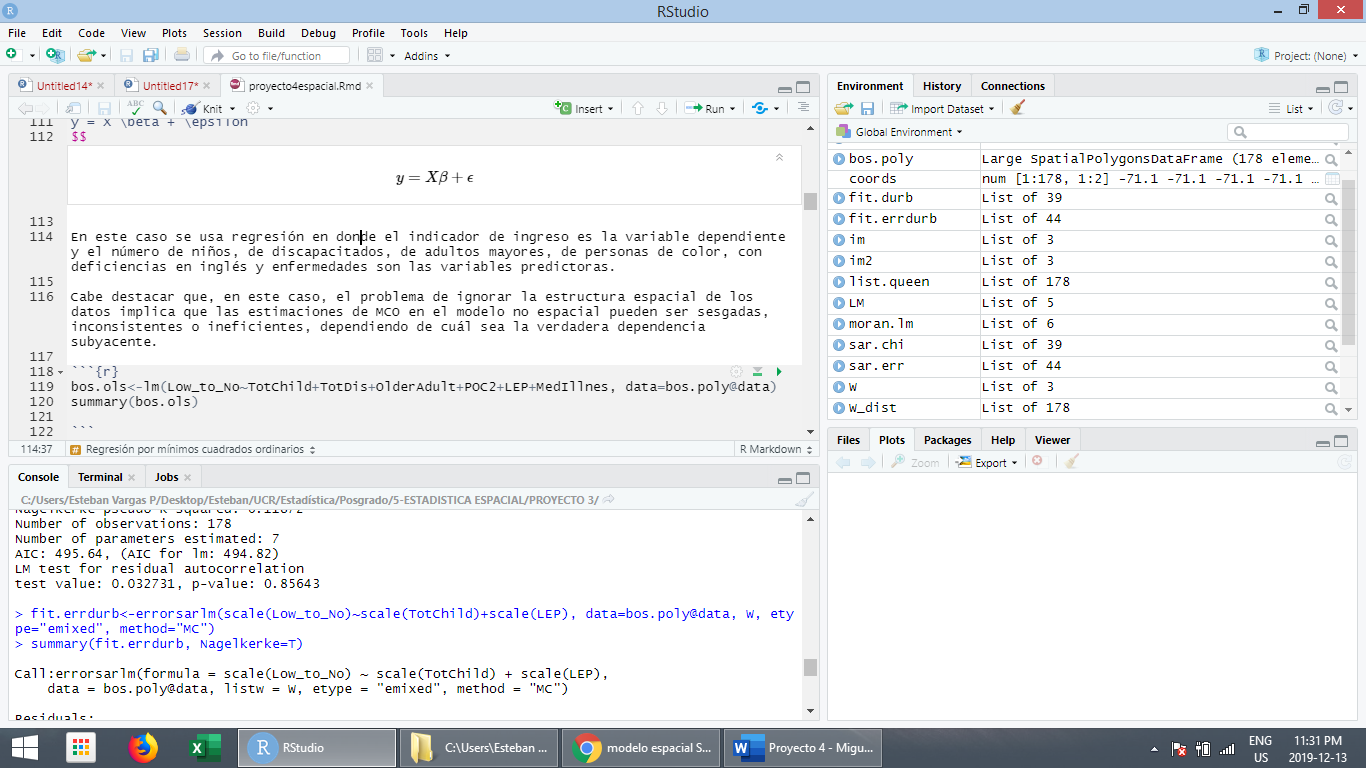
Los síntomas de enfermedades médicas existentes a menudo se ven exacerbados por las altas temperaturas. Por ejemplo, el calor puede desencadenar ataques de asma o aumentar la presión arterial ya alta debido al estrés de las altas temperaturas que ejerce el cuerpo. Los eventos climáticos pueden interrumpir el acceso a fuentes normales de atención médica e incluso medicamentos para mantener la vida. Fuente de datos: La enfermedad médica es una medida indirecta que se basa en datos EASI a los que se accede a través de Simply Map. Los datos de salud a nivel local en Massachusetts no están disponibles más allá de los códigos postales. EASI modeló las estadísticas de salud para la población de EE. UU. En función de las probabilidades de edad, sexo y raza utilizando los datos de la Oficina del Censo de EE. UU.

**Metodología**

Pata llevar a cabo el análisis se tomaron como referencia los siguientes modelos espaciales:

**Regresión por mínimos cuadrados ordinarios (OLS)**

Tradicionalmente, estos problemas se modelan dejando de lado el elemento espacial usando técnicas como la regresión lineal:



En este caso se usa regresión en donde el indicador de ingreso es la variable dependiente y el número de niños, de discapacitados, de adultos mayores, de personas de color, con deficiencias en inglés y enfermedades son las variables predictoras.

Cabe destacar que, en este caso, el problema de ignorar la estructura espacial de los datos implica que las estimaciones de MCO en el modelo no espacial pueden ser sesgadas, inconsistentes o ineficientes, dependiendo de cuál sea la verdadera dependencia subyacente.

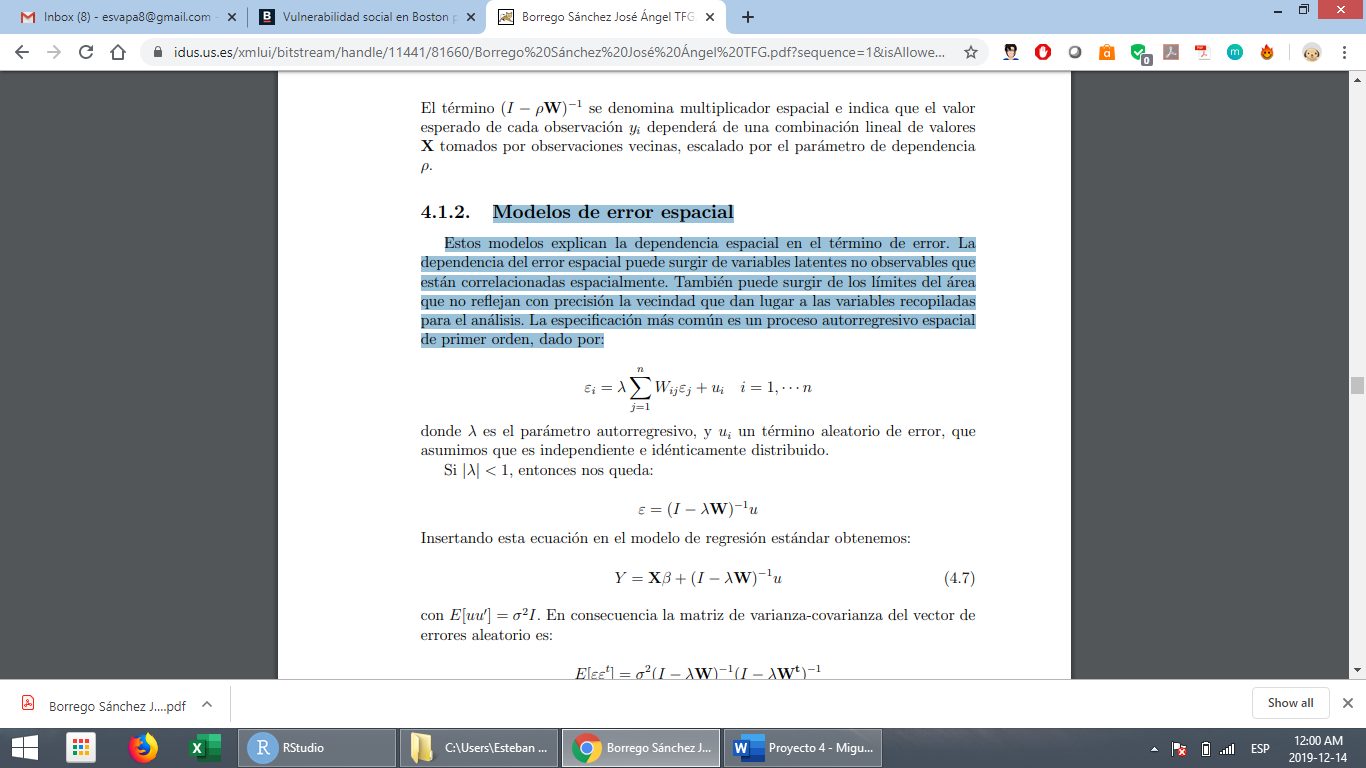
**Modelo SAR**

La estimación del modelo SAR se puede abordar de dos maneras. Una forma es asumir la normalidad del término de error y utilizar máxima verosimilitud.

Una de las especificaciones de un SAR (Modelo espacial autorregresivo), donde la variable dependiente yi está asociada con la zona i y se modela como una función de las variables explicativas xij y la matriz de adyacencia, donde las "i" son variables aleatorias gaussianas independientes con media cero y varianza σ2 o en notación vectorial Y = Xβ + λWY + ε, donde ε es un vector gaussiano con media cero y matriz de varianza-covarianza. El modelo de regresión estándar, cuenta con un término adicional (λWY) que permite que la variable predictora en zonas adyacentes a una zona dada tenga algún impacto. En este caso, como modelamos los crímenes, esto supone que, además de las características de una zona que afecta el nivel de ceiminalidad en dicha zona, estos también se verían influenciados por el crimen de los sujetos en áreas cercanas. También se puede estimar el modelo de error espacial.

**Modelos de error espacial**

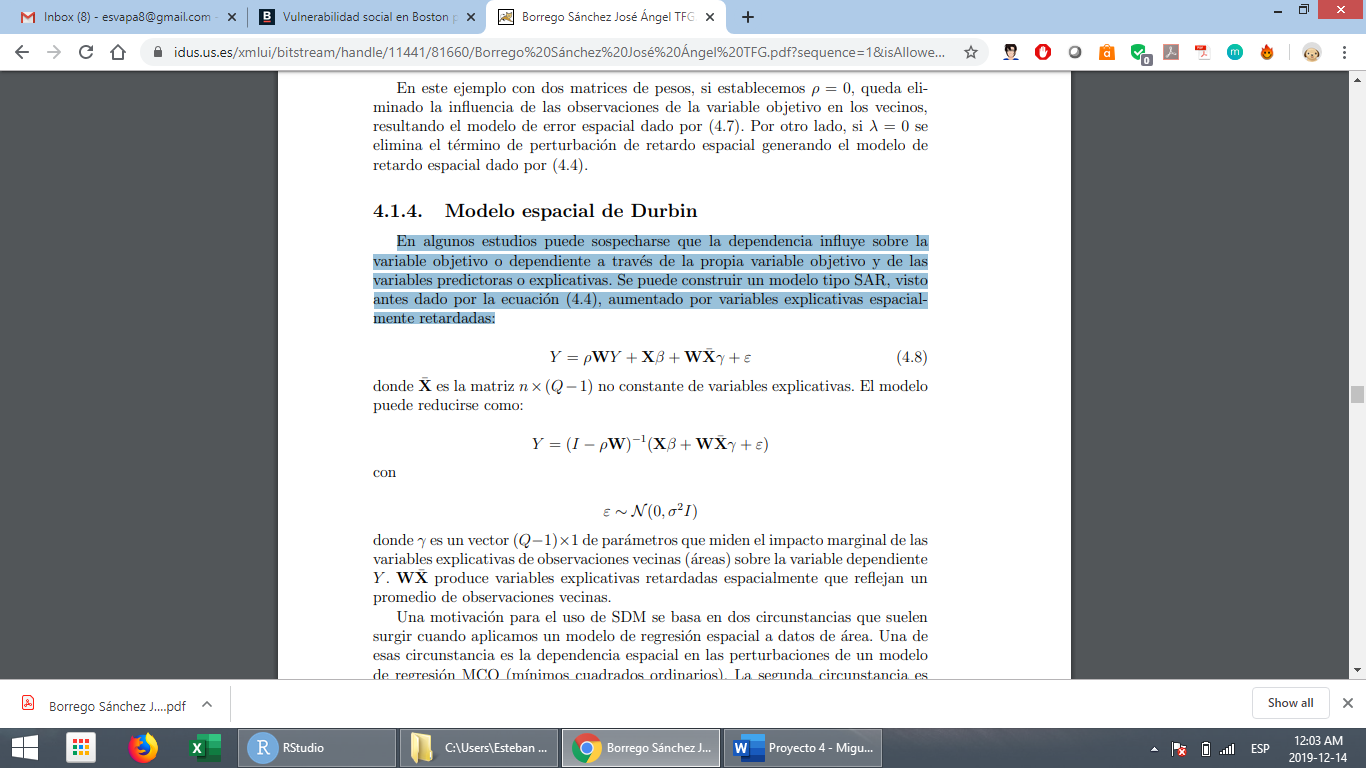
Estos modelos explican la dependencia espacial en el término de error. La dependencia del error espacial puede surgir de variables latentes no observables que están correlacionadas espacialmente. También puede surgir de los límites del área que no reflejan con precisión la vecindad que dan lugar a las variables recopiladas para el análisis. La especificación más común es un proceso autorregresivo espacial de primer orden, dado por:



donde λ es el parámetro autorregresivo, y ui un término aleatorio de error, que asumimos que es independiente e idénticamente distribuido.

**Modelo Durbin**

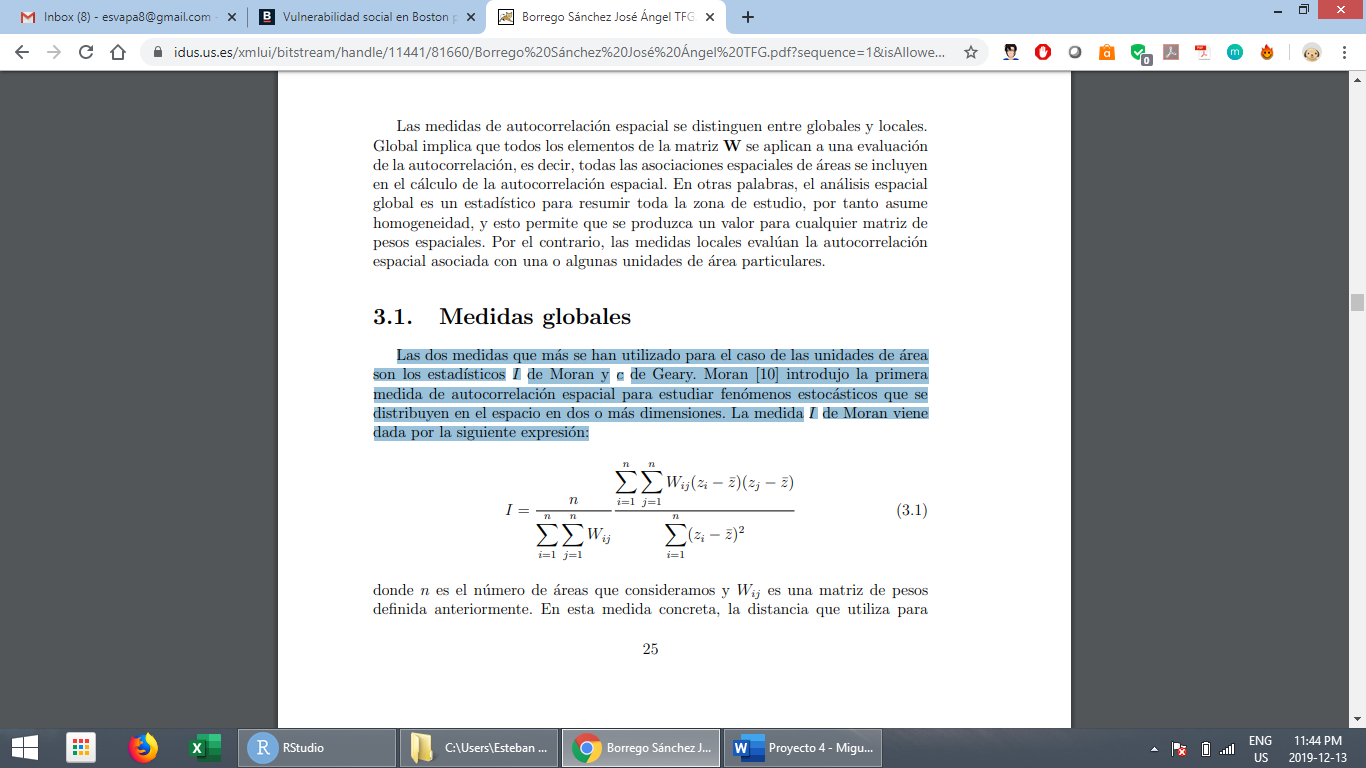
Estos modelos usan múltiples matrices de pesos proporcionando una generalización sencilla de los modelos SAR y SEM. En algunos estudios puede sospecharse que la dependencia influye sobre la variable objetivo o dependiente a través de la propia variable objetivo y de las variables predictoras o explicativas. Se puede construir un modelo tipo SAR, aumentado por variables explicativas espacialmente retardadas:



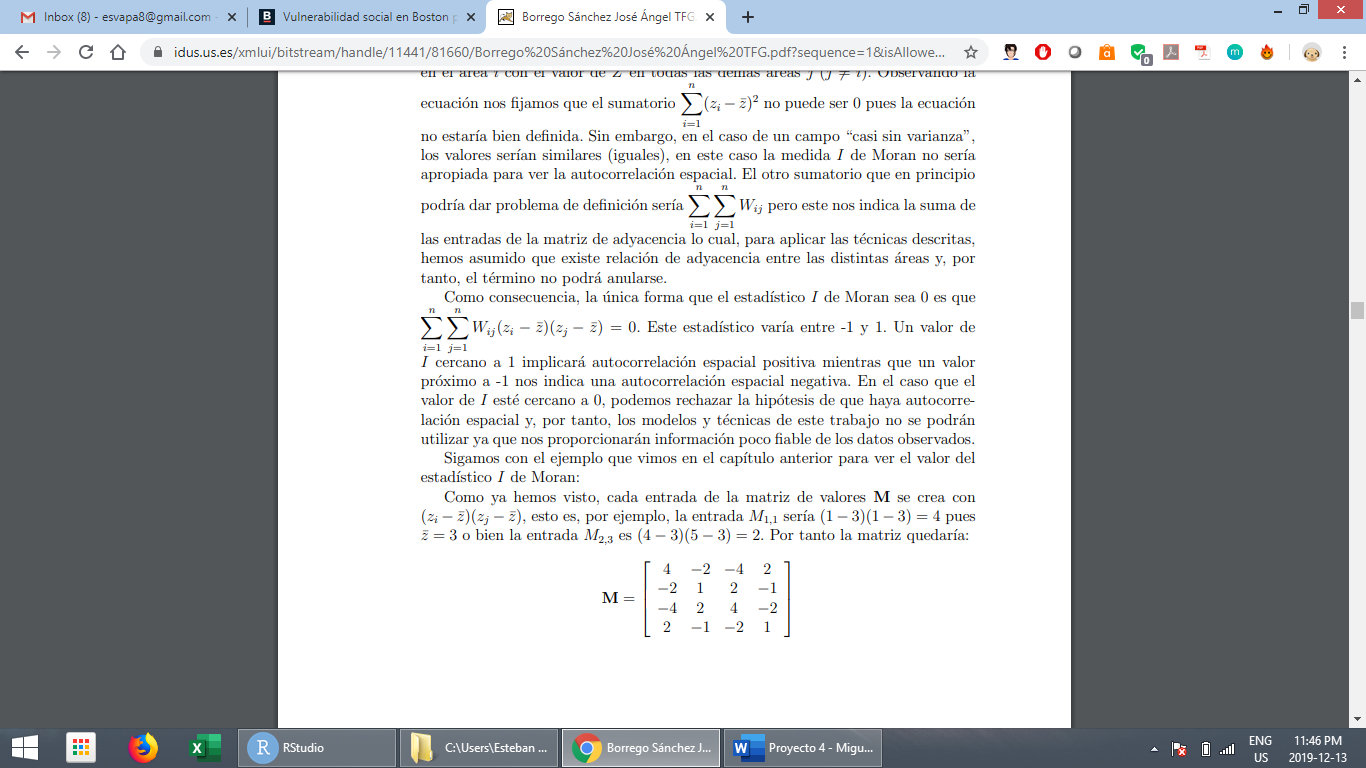
En dicho modelo los erroes se distribuyen normal con media cero y varianza conocida.

**Medidas de autocorrelación espacial: Prueba de Moran**

Las dos medidas que más se han utilizado para el caso de las unidades de área son los estadísticos I de Moran y c de Geary. Moran introdujo la primera medida de autocorrelación espacial para estudiar fenómenos estocásticos que se distribuyen en el espacio en dos o más dimensiones. La medida I de Moran viene dada por la siguiente expresión:

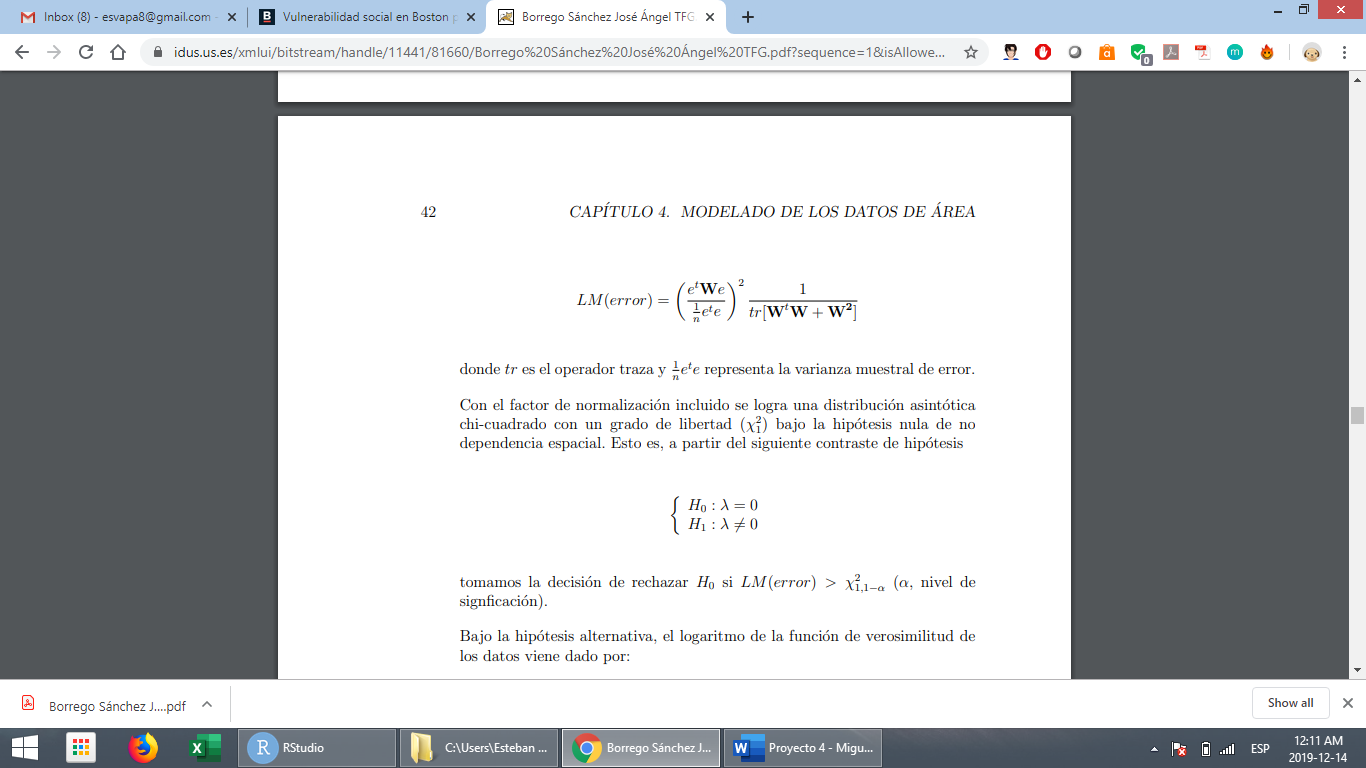


El estadístico de Moran puede tomar varios valores, como consecuencia, la única forma que el estadístico I de Moran sea 0 es que:



Este estadístico varía entre -1 y 1. Un valor de I cercano a 1 implicaría autocorrelación espacial positiva mientras que un valor próximo a -1 nos indica una autocorrelación espacial negativa. En el caso que el valor de I este cercano a 0, no podemos rechazar la hipótesis de que no haya autocorrelación espacial y, por tanto, los modelos y técnicas de este trabajo no se podrán utilizar ya que nos proporcionarán información poco fiable de los datos observados.

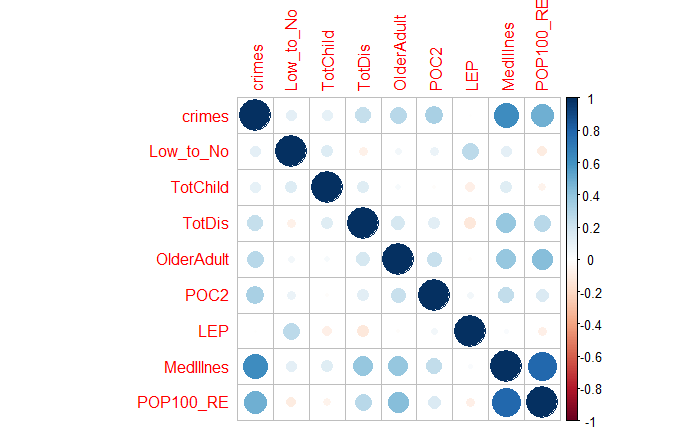
En la práctica, la inferencia mediante la prueba I de Moran, se basa en una aproximación normal utilizando un valor estandarizado obtenido restando la media, bajo la hipótesis nula de no dependencia espacial, y dividiendo por la raíz cuadrada de la varianza. La causa de la dependencia espacial bajo la hipótesis alternativa no está especificada, por lo tanto, la prueba de Moran es una prueba general para detectar autocorrelación espacial. Una prueba alternativa se basa en el principio del multiplicador de Lagrange (LM), sugerido por Burridge, que se calcula a través de los residuos MCO y que es diferente según qué tipo de dependencia espacial queramos probar. Para una dependencia de tipo error dado por la ecuación (4.7), el estadístico LM (error) viene dado por:



**Resultados**

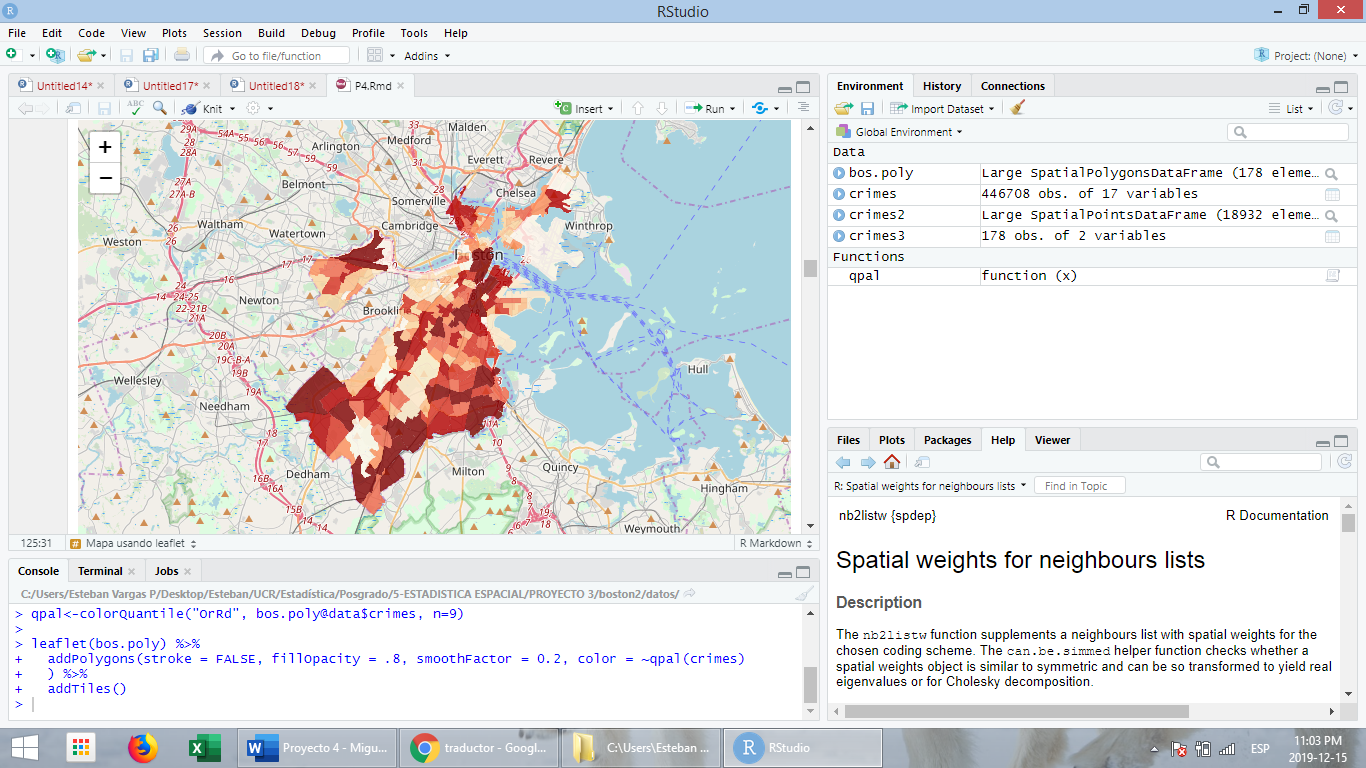
Como primer aspecto de análisis se obtuvieron las correlaciones entre las variables en cuestión. La figura 1 muestra que todas ellas tienen una asociación positiva, principalmente entre crímenes y enfermedades medicas o cantidad población. Otras variables como el ingreso y la tenencia de niños presentan correlación baja.

**Figura 1.** Correlaciones entre variables sobre vulnerabilidad social



Una vez realizada una primera inspección sobre la asociación entre las variables del conjunto de datos, se observó la distribución en el espacio geográfico de la variable dependiente: crímenes.

**Mapa 1.** Crímenes en la ciudad de Boston



El registro de crímenes en la ciudad de Boston parece ser más frecuente en la región sur de la ciudad, sin embargo, en la mayoría de regiones se observan niveles altos de criminalidad. En la región noroeste parecen concentrarse los niveles más bajos de crímenes.

**Modelo de regresión lineal**

El modelo de regresión lineal inicial tomó en cuenta todas las variables, sin embargo, no todas resultaron ser significativas por lo cual se eliminaron algunas para obtener el modelo más parsimonioso. El modelo final se muestra a continuación:

**Cuadro 1.** Modelo de regresión lineal para el numero de crímenes

|  |  |  |
| --- | --- | --- |
| **Predictor** | **Valor F** | **Valo p asociado** |
| Población de color | 32,2 | <0.01 |
| Niños | 3,9 | 0.05 |
| Discapacitados | 9,4 | 0.003 |
| Adultos mayores | 9,4 | 0.003 |
| Enfermedades médicas | 71,4 | <0.01 |

De acuerdo con los resultados del cuadro 1, se puede observar que los predictores que resultaron predecir de forma eficiente el nivel de criminalidad en la zona fueron: población de color, cantidad de niños, discapacitados, adultos mayores y enfermedades médicas. En R2 fue del 42%.

**Prueba de Moran**

La autocorrelación espacial mide el grado en que el nivel de criminalidad se correlaciona consigo mismo en el espacio. En otras palabras, los valores similares aparecen cerca uno del otro, o agrupados, en el espacio (autocorrelación espacial positiva) o los valores vecinos son diferentes (autocorrelación espacial negativa).

En este caso la autocorrelación fue positiva (0,33), por lo cual el nivel de criminalidad entre zonas cercanas fue similar. Para la prueba se utilizó la matriz de pesos (torres) la cual establece que dos unidades están cerca si comparten un lado o un borde. Tomando un 95% de confianza, el valor p asociado fue menor a 0,05. Dado estos resultados, se puede rechazar la hipótesis nula, es decir, la distribución espacial de los valores altos y los valores bajos de crímenes está más agrupada espacialmente de lo que se esperaría si los procesos espaciales subyacentes fueran aleatorios. Se puede asumir autocorrelación espacial. La prueba de Lagrange también sugirió autocorrelación espacial en la mayoría de modelos espaciales. Los modelos que permitieron modelar la distribución espacial se tomaron en cuenta para el modelaje. Estos se muestran a continuación:

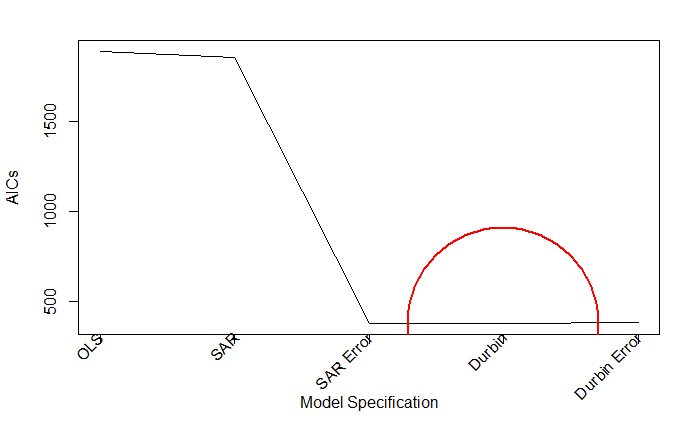
**Cuadro 2.** Modelo de espaciales para el nivel de criminalidad en Boston

|  |  |
| --- | --- |
| **Modelo** | **Variables** |
| MCO | Población de color, cantidad de niños, discapacitados y enfermedades médicas |
| SAR | Población de color y enfermedades médicas |
| SAR Error | Población de color y enfermedades médicas |
| Durbin | Población de color y enfermedades médicas |
| Durbin Error | Población de color y enfermedades médicas |

De Acuerdo al cuadro 2, los indicadores que resultaron ser los que describían de mejor manera los niveles altos y bajos de criminalidad, fueron la población de color de la región y las enfermedades médicas.

Para elegir el modelo que mejor describía la estructura espacial se tomó en cuenta el criterio del AIC, en donde aquellos modelos con un nivel más bajo fueron los que mejor ajustaron.

**Figura 2**. AIC observado para cada modelo espacial



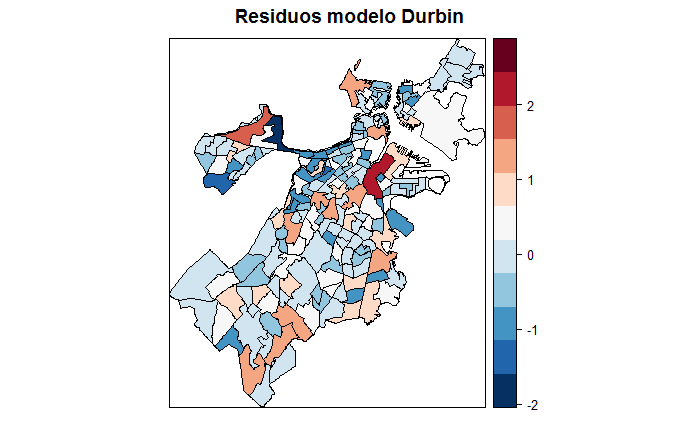
La medida de ajuste AIC permitió obtener el modelo que realizó una mejor predicción de la estructura espacial del nivel de criminalidad en Boston. El modelo de Durbin fue el que presentó un mejor ajuste. A pesar de que el gráfico muestra una gran similitud entre el modelo SAR error, Durbin y Durbin error, después de realizarse una prueba de razón de verosimilitudes entre el SAR error y los dos Durbin, el modelo de Durbin fue el elegido. Además, no se observó diferencia estadística entre la calidad del modelo SAR error y el modelo Durbin error.

**Evaluación del mejor modelo**

**Impacto**

El análisis del nivel de impacto total del mejor modelo (Durbin) mostró que la variable con mayor influencia fue la variable de enfermedades médicas, la cual mostró un valor de 0.64. En el caso de la población negra, el valor fue de 0.52.

**Figura 3.** Residuos de modelo Durbin



También se obtuvo el gráfico de los residuos de modelo Durbin, el cual mostró que la en general los erroes fueron bastante bajos o cercanos a 0. Las zonas con mayor cantidad de crimenes fueron estimadas de manera satisfactoria y las regiones con mayores errores fueron en el centro de la ciduad de boston y en el area oeste.

**Discusión y Conclusión**

El análisis obtenido indicó que la distribución espacial de los niveles de criminalidad presenta correlación espacial significativa. Los mejores modelos espaciales para predecir en fenómeno en cuestión, fueron el SAR Error y los Durbin, sin embargo, el Durbin fue el mejor de todos.

A partir del análisis realizado fue posible obtener un modelo satisfactorio para para ajustar el nivel de criminalidad. Es importante resaltar que el cantidad de personas con enfermedades médicas y la cantidad de población de color son los principales predictores en prácticamente todos los modelos analizados. Esto, evidencia el problema altamente investigado en Estados Unidos que asocia las poblaciones no blancas con las tasas de criminalidad en muchos estados. Además, aunque intuitivamente la cantidad de personas con enfermedades no parece tener mucho sentido en el contexto de la criminalidad debe verse como una variable proxy de los niveles de salud publica y de la ejecución de los planes de salud en los distintos distritos de la ciudad de Boston.

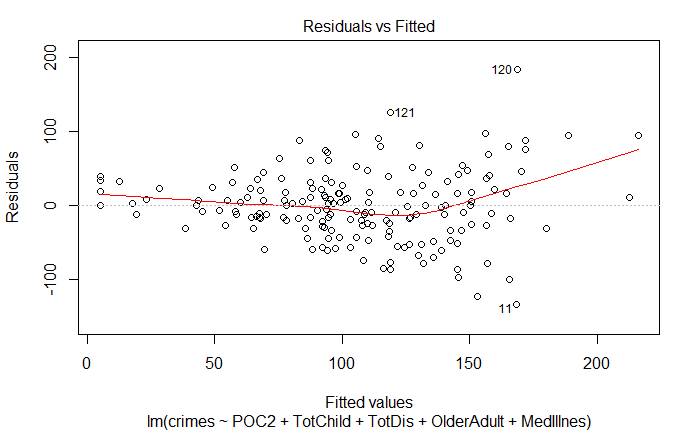
Por supuesto, el presente análisis muestra solo la asociación entre estas variables y como se asocian espacialmente con el fenómenos de la criminalidad en la ciudad de Boston, pero faltan otras variables para poder determinar y probar las hipótesis de las relaciones entre cantidad de personas de color y cantidad de personas con enfermedades con la cantidad de crímenes.

**Referencias**

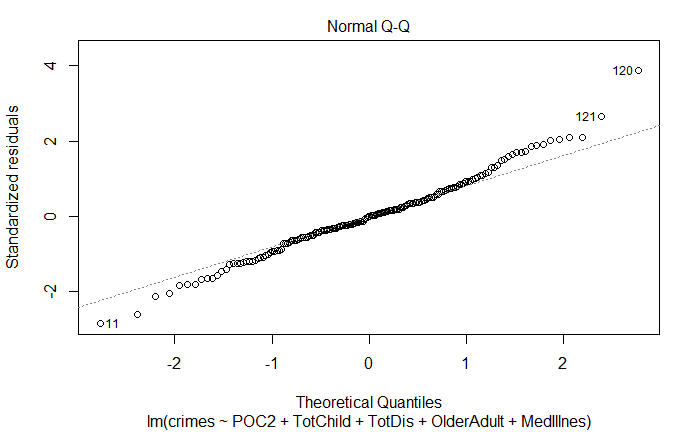
1. Burridge, P. (1980). On the Cliff-Ord Test for Spatial Correlation. Journal of the Royal Statistical Society. Series B (Methodological), Vol. 42, No. 1.
2. CEPAL (2002). Comisión Económica para América Latina y el Caribe. México. Naciones Unidas. México, D.F.
3. CEPAL / CELADE (2002), Vulnerabilidad socio demográfica; viejos y nuevos riesgos para comunidades, hogares y personas. Separata.
4. Kaztman, Rubén (2002). Convergencias y divergencia: exploración sobre los efectos de las nuevas modalidades de crecimiento sobre la estructura social de cuatro áreas metropolitanas en América Latina. Trabajo y ciudadanía. Los cambiantes rostros de la integración y exclusión social en cuatro áreas metropolitanas de América Latina. P 23 – 60.
5. Martin, S.. (2014). A framework to understand the relationship between social factors that reduce resilience in cities: Application to the City of Boston. International Journal of Disaster Risk Reduction.
6. Moran, P.A.P. (1950). Notes on Continuous Stochastic Phenomena.Biometrika, Vol. 37, No. 1/2, pp. 17-23.
7. Mosquera, N Rengifo, N, et al. (2013). Análisis de factores familiares y sociales asociados a vulnerabilidad. Dialnet, Vol. 12, Nº. 2 (diciembre), 2013, págs. 136-161

**Anexos**

**Anexo 1. Residuales estimados y predichos del modelo lineal**



**Anexo 2. QQ-plot para residuales estandarizados**



**Anexo 3. Distancia de Cook**

