

Maestría Profesional en Estadística
Curso: SP-1649 Tópicos de Estadística Espacial Aplicada
Profesora: Dra. Marcela Alfaro Córdoba

ANÁLISIS DE ESTADÍSTICA ESPACIAL: COMPARACIÓN DE UN MODELO ESPACIAL Y NO ESPACIAL EN LA ESTIMACIÓN DEL ÍNDICE DE NECESIDADES BÁSICAS INSATISFECHAS (NBI) POR DISTRITO EN COSTA RICA, AÑO 2011

Elaborado por:

María Elena Zúñiga Brenes
Alejandro Zúñiga Madrigal

- 12 de diciembre de 2019 -

CONTENIDO

1	INTRODUCCIÓN	3
2	ANTECEDENTES Y JUSTIFICACIÓN	3
2.1	Objetivos de la investigación	4
3	ASPECTOS METODOLÓGICOS	5
3.1	Población de estudio	5
3.2	Fórmulas utilizadas	7
4	RESULTADOS Y DISCUSIÓN	7
5	CONCLUSIÓN	19
6	BIBLIOGRAFÍA	20

1 INTRODUCCIÓN

En este trabajo se realiza un análisis estadístico espacial, tomando como base de datos de distritos a nivel nacional y la distribución del índice de necesidades básicas insatisfechas (NBI). Estos registros los recolectó el Instituto Nacional de Estadística y Censos (INEC) en el 2011.

La intención del trabajo es explorar y realizar la estimación planteando las siguiente pregunta de interés:

- ¿Existen conglomerados que se puedan localizar geográficamente por distritos para el NBI?
- ¿Al modelar el NBI, existe algún aporte adicional de un modelo espacial en comparación a uno no espacial?

Este trabajo establece como propuesta las preguntas de investigación para lograr con ello, identificar la distribución del índice de necesidades básicas insatisfechas (NBI); y de esta manera, generar criterios para su estudio contemplando la dimensión espacial.

2 ANTECEDENTES Y JUSTIFICACIÓN

El término “pobreza” tiene muchos significados y abarca una infinidad de situaciones, es aceptable definirlo como “la situación de aquellos hogares que no logran reunir, en forma relativamente estable, los recursos necesarios para satisfacer las necesidades básicas de sus miembros (...)” (CEPAL / DGEC, 1988a). A ello puede agregarse que “la pobreza (...) es un síndrome situacional en el que se asocian el infraconsumo, la desnutrición, las precarias condiciones de vivienda, los bajos niveles educacionales, las malas condiciones sanitarias, una inserción inestable en el aparato productivo o dentro de los estratos primitivos del mismo, actitudes de desaliento y anomía, poca participación en los mecanismos de integración social, y quizás la adscripción a una escala particular de valores, diferenciada en alguna manera de la del resto de la sociedad” (Altimir, 1979).

En América Latina, el método de las Necesidades Básicas Insatisfechas ha significado un importante aporte para la identificación de ciertas carencias críticas de la población y la caracterización de la pobreza. La siguiente figura resume la construcción del NBI.

Figura 1. Composición del Índice de Necesidades Básicas Insatisfechas

NECESIDADES BÁSICAS, DIMENSIONES Y VARIABLES CENSALES		
Necesidades Básicas	Dimensiones	Variables Censales
Acceso a vivienda	a) Calidad de la vivienda	Materiales de construcción utilizados en piso, paredes y techo
	b) Hacinamiento	i) Número de personas en el hogar ii) Número de cuartos de la vivienda
Acceso a servicios sanitarios	a) Disponibilidad de agua potable	Fuente de abastecimiento de agua en la vivienda
	b) Tipo de sistema de eliminación de excretas	i) Disponibilidad de servicio sanitario ii) Sistema de eliminación de excretas
Acceso a educación	Asistencia de los niños en edad escolar a un establecimiento educativo	i) Edad de los miembros del hogar
		ii) Asistencia a un establecimiento educativo
Capacidad económica	Probabilidad de insuficiencia de ingresos del hogar	i) Edad de los miembros del hogar
		ii) Último nivel educativo aprobado
		iii) Número de personas en el hogar
		iv) Condición de actividad

Fuente: CEPAL/PNUD (1989)

Como se puede observar, las variables para la construcción del NBI son obtenidas a través de los Censos, lo que implica que la construcción de este, depende de los momentos censales de cada país. Por ello, este trabajo, busca de manera exploratoria modelar el NBI a través de otras variables, como lo son la relación de dependencia demográfica, la relación del sector primario con respecto al terciario, y el indicador de seguro social de la jefatura del hogar (ver anexos); todo esto para los distritos del año 2011.

También se busca conocer si existe alguna asociación espacial en la distribución del NBI para los distritos, y si esta puede influir sobre los resultados obtenidos en el modelaje del indicador en cuestión.

2.1 OBJETIVOS DE LA INVESTIGACIÓN

- Determinar si existe autocorrelación espacial en la distribución del NBI para los distritos de Costa Rica del año 2011.
- Comparar modelos con estructura espacial con un modelo no espacial, y determinar cuál presenta el mejor ajuste.

3 ASPECTOS METODOLÓGICOS

El presente trabajo intenta modelar el índice de Necesidades Básicas Insatisfechas a partir modelos lineales, tanto no espaciales como espaciales.

3.1 FÓRMULAS UTILIZADAS

Modelo de regresión lineal Gaussiano

Como modelo no espacial se emplea un modelo de regresión lineal Gaussiano, del siguiente tipo:

$$\hat{y} = \alpha + \beta x + e$$

En cuanto a los modelos espaciales, se trabajan los siguientes:

Modelo de Error Espacial (SEM)

Se define de la siguiente manera:

$$\hat{y} = \alpha + \beta x + e, \text{ en donde } e = W e + u$$

Modelo SAR y CAR

- Sea $\{Z(A_i): A_i \in (A_1 \dots A_n)\}$ donde $\{A_1 \dots A_n\}$ forman una grilla en D.
- $\{A_1 \dots A_n\}$ forman una grilla en D si $\{A_1 \dots A_n\}$ es una partición simple de D, por ejemplo: $A_1 \cup A_2 \cup \dots \cup A_n = D$ y $A_i \cap A_j = \emptyset$ para todo $i \neq j$
- Entonces se puede modelar el proceso por medio de modelos autorregresivos SAR y CAR.

Modelo SAR

$$Z(A_i) = \mu_i + \sum_{j=1}^n b_{ij}(Z(A_j) - \mu_j) + \varepsilon_i$$

donde

$$\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)' \sim N(\mathbf{0}, \mathbf{A}) \quad E(Z(A_i)) = \mu_i$$

- b_{ij} son constantes conocidas o desconocidas y
- $b_{ii} = 0, i = 1, \dots, n$
- Este modelo se denomina “simultáneo” porque en general los términos del error están correlacionados con $\{Z(A_j): j \neq i\}$
- Si n es finito, $\mathbf{B} = (b_{ij})$ es la matriz que contiene b_{ij}

La función de enlace es:

$$\mathbf{Z} \sim N(\boldsymbol{\mu}, (\mathbf{I}_n - \mathbf{B})^{-1} \mathbf{A} (\mathbf{I}_n - \mathbf{B})^{-1'})$$

Donde:

$$\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_n)$$

\mathbf{I}_n : matriz identidad

Modelo CAR

$$Z(A_i) | Z(A_{(-i)}) \sim N \left(\mu_i + \sum_{j=1}^n c_{ij} (Z(A_j) - \mu_j), \tau_i^2 \right)$$

- Donde: $Z(A_{(-i)}) = \{Z(A_j): j \neq i\}$, $E(Z(A_i)) = \mu_i$, τ_i^2 es la varianza condicional c_{ij} son constantes conocidas o desconocidas, en particular $c_{ii} = 0, i = 1, \dots, n$
- Si n es finito, se forman las matrices $\mathbf{C} = (c_{ij})$ y $\mathbf{T} = \text{diag}\{\tau_1^2, \tau_2^2, \dots, \tau_n^2\}$ y entonces:

$$\mathbf{Z} \sim N(\boldsymbol{\mu}, (\mathbf{I}_n - \mathbf{C})^{-1} \mathbf{T})$$

- La estructura de \mathbf{B} y \mathbf{C} está especificada por la forma de la grilla.

- Una forma común de construir B y C es con un parámetro que escala una matriz W para indicar si las regiones son vecinas o no.
- Donde $W = \{w_{ij}\}$
 - $w_{ij} = 1$ si la región A_i comparte frontera con la región A_j
 - $w_{ij} = 0$ si $i=j$
 - $w_{ij} = 0$ para cualquier otro caso.
- Para el modelo SAR, $B = \rho sW$ y para el modelo CAR, $C = \rho cW$. Donde ρs y ρc se refieren a parámetros de correlación espacial o dependencia espacial y son estimados.
- Existen otras formas de construir la matriz W, como:
- Restringiendo las filas de W a sumar 1
- Crear funciones de peso que dependen del largo de las fronteras entre regiones.
- Otros autores recomiendan matrices ponderadas como:

$$w_{ij}^* = w_{ij}/w_i$$

3.2 POBLACIÓN DE ESTUDIO

Se estudian 472 distritos de Costa Rica. Para cada uno se analiza el índice de necesidades básicas insatisfechas; las cuales fueron estimadas por el Censo del año 2011 por parte del Instituto Nacional de Estadística y Censos. Es importante destacar que para este estudio se excluye la Isla del Coco; con ello, tanto las estimaciones del NBI, así como de las variables del modelo, y los shapes son facilitados por el INEC, y corresponde a los datos y capas del año 2011.

Software

Para el procesamiento de la información se utiliza el software estadístico R Studio versión 3.6. Además se utilizan las librerías “spdep”, “maptools”, “rgdal”.

Manejo de datos

Las capas de información de los distritos están en formato shape (.shp). Se agrega la información del NBI para cada uno de estos así como las otras variables en cuestión. Finalmente se lleva a cabo el modelaje de la información.

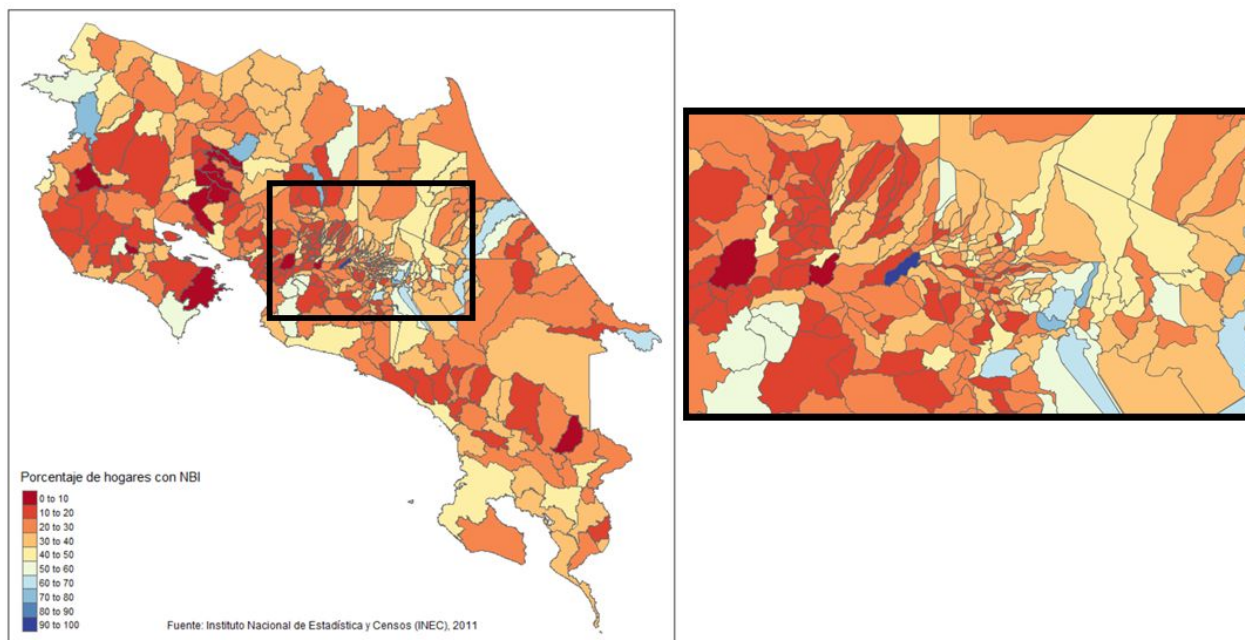
4 RESULTADOS Y DISCUSIÓN

La siguiente figura muestra la distribución del índice de necesidades básicas insatisfechas para los 472 distritos del año 2011. Se aprecia que en la zona norte del país se tienen niveles de entre 20%

y 40% de los hogares con NBI, y un comportamiento similar se puede apreciar en la Zona Sur, en la Región Atlántica y parte de la Región Pacífica.

Se aprecian algunos pocos distritos con alto porcentaje de hogares con NBI, entre 60% y 80% distribuidos por todo el país, teniendo a sus alrededores distritos con niveles medios de este indicador. En general se observa un comportamiento similar del NBI para grupos de distritos; es decir, salvo algunos valores altos de este indicador, se ven similitudes entre distritos vecinos; lo cual, parece indicar que existe relación espacial de esta variable entre distritos.

Figura 2. Índice de necesidades básicas insatisfechas por distrito, Costa Rica, 2011.



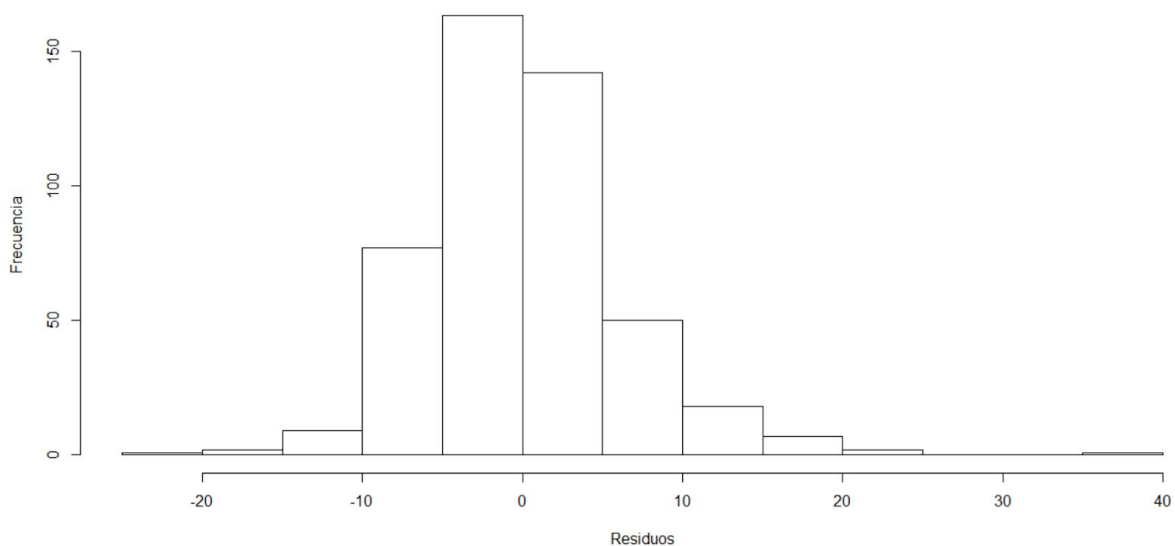
Como primer paso se lleva a cabo un modelo de regresión lineal no espacial del porcentaje de hogares con NBI y la relación del sector primario con el terciario (Rel_Sec12), indicador del seguro social de la persona jefa del hogar (IND_SSJe), y la relación de dependencia demográfica (Rel_DepDem). La siguiente tabla muestra un resumen de los resultados de este modelo, se aprecia que todos los coeficientes de regresión son estadísticamente significativos, y que el modelo permite explicar un 77% de la variabilidad del fenómeno en estudio.

Tabla 1: Modelo de regresión lineal

Coeficientes	Estimado	Error Estandar	Valor t	Pr(> t)
Intercepto	-23,4614	2,5193	-9,312	< 2e-16 ***
Relación del sector primario con el terciario	0,0416	0,0031	13,4780	< 2e-16 ***
Indicador del seguro social de la persona jefa del hogar	0,2696	0,0407	6,6290	9,35e-11***
Relación de dependencia demográfica	0,9053	0,0618	14,6440	< 2e-16 ***
Error estándar residual: 6,301 on 468 gl.				
R cuadrado: 0,7765, R cuadrado ajustado: 0,7751				
F-statistic: 542,1 con 3 y 468 gl, valor p: <2,2e-16				

En cuanto al análisis de los supuestos, no se cumple el de normalidad, pese a que estos sí están centrados en cero, se tienen valores extremos que rechazan la hipótesis nula que asume normalidad, tal y como se puede apreciar en el siguiente histograma de residuos.

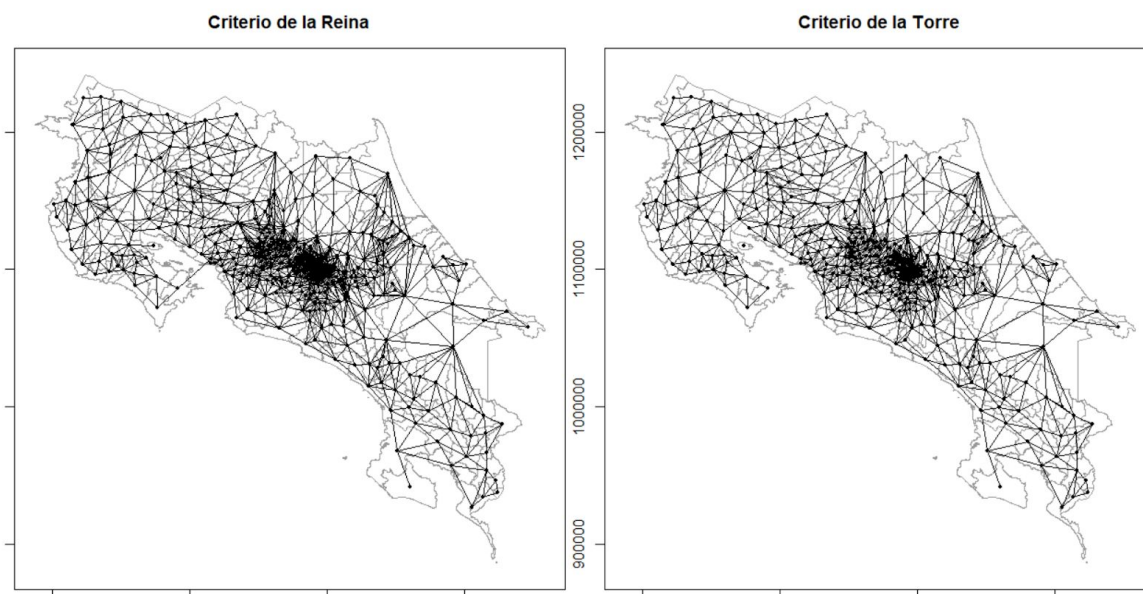
Figura 3: Histograma de residuos del modelo lineal no espacial



Al analizar el supuesto de homocedasticidad mediante el test de Breusch Pagan, se tiene un p-value asociado menor a 0.0001; por lo que se tiene suficiente evidencia estadística para asumir la hipótesis alternativa de que el modelo presenta heterocedasticidad en sus residuos. De manera paralela se lleva a cabo el test de Durbin Watson, y el p-value asociado es de 0.001089; por lo que de igual manera, se tiene evidencia estadística de que existe autocorrelación de los residuos.

Para llevar a cabo los modelos espaciales, primero se debe definir el criterio de vecinos para los distritos. La figura 4 muestra el resultado de agrupar a los distritos bajo el criterio de la reina y de la torre. Se aprecia que existen similitudes entre ambos criterios, y que bajo el criterio de la reina, los distritos tienen una mayor cantidad de vecinos y en todas direcciones. Para este trabajo, los modelos espaciales resultantes se llevan a cabo con el criterio de la reina, pues teóricamente resulta adecuado debido al comportamiento del NBI a nivel distrital.

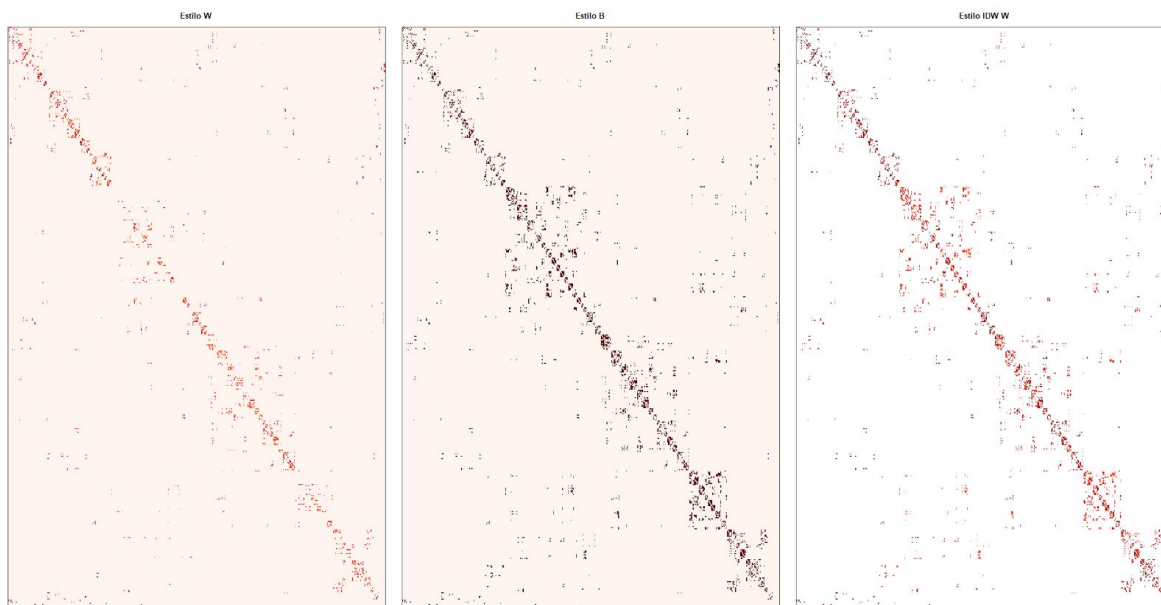
Figura 4: Agrupamiento de distritos vecinos bajo el criterio de la reina y de la torre



Posteriormente se debe seleccionar el estilo de la matriz de vecindades, en donde se evalúa el estilo w, este mismo estilo ponderado por el inverso de la distancia, y el estilo B. La figura 5 muestra que el estilo B los valores son de 1; mientras que valores más claros hacen referencia a su cercanía con el valor 0. Para este trabajo se decide utilizar la matriz

W ponderada por el inverso de la distancia; pues presenta mayor variabilidad en los valores de distancias; lo que enriquece el análisis espacial.

Figura 5: Distancias para cada estilo de matriz de vecindades



Para ver si existe una estructura espacial se hace las pruebas de autocorrelación espacial. Primero se lleva a cabo el test de Moran y luego el test del Multiplicador de Lagrange. Ambas pruebas permiten determinar la existencia de autocorrelación espacial, y la segunda además sugiere el tipo de modelo espacial a emplear.

Para la primera prueba, el Test de Moran, se tiene un valor observado de este indicador de 0.06, mientras que el valor esperado es de -0.003 y una varianza de 0.0007; lo que da como resultado un p-value de 0.0137. Por lo tanto, a partir de esta prueba, existe evidencia estadística para asumir que existe autocorrelación espacial.

**Tabla 2: Prueba Global de Moran para Regresión Residual
(peso: W1)**

Indicador	Valor
Observado	0,0611
Expectativa	-0,0031
Variancia	0,0007

Por su parte, la prueba del Multiplicador de Lagrange coincide con la prueba de Moran, ya que brinda evidencia de la existencia de dependencia espacial. Además sugiere que el modelo del error espacial es el más probable para el tratamiento de los datos. Pese a esta sugerencia, en este trabajo se estiman además modelos CAR y SAR, con el fin de encontrar el modelo que mejor ajuste.

**Tabla 3: Prueba Multiplicador de Lagrange para Dependencia Espacial
(peso: IWD Estilo W)**

Modelos	Valor	df	valor p
LMerr	5,3879	1	0,0203
LMLag	2,1189	1	0,1455
RLMerr	3,3715	1	0,0663
RLMlag	0,1025	1	0,7489
SARMA	5,4904	2	0,0642

Modelo de error espacial

La siguiente tabla muestra los resultados obtenidos por el modelo de error espacial estimado a través del método de máxima verosimilitud. Se aprecia que todos los coeficientes estimados son estadísticamente diferentes de cero; pero también se debe prestar atención a su similitud con los coeficientes de regresión estimados por el modelo lineal no espacial.

También se tiene un valor de lambda de 0.16 el cual además es estadísticamente significativo; lo que este tiene una influencia relevante en la estructura de los errores, y con este la relevancia de contemplar la estructura espacial.

Tabla 4: Modelo del error máxima verosimilitud

Coeficientes	Estimado	Error Estandar	Valor t	Pr(> Z)
Intercepto	-23,6441	2,5723	-9,312	< 2,2e-16 ***
Relación del sector primario con el terciario	0,0407	0,0031	13,4780	< 2,2e-16 ***
Indicador del seguro social de la persona jefa del hogar	0,2757	0,0407	6,6290	1,337e-11
Relación de dependencia demográfica	0,9080	0,0624	14,6440	< 2,2e-16 ***

Lambda: 0,16085, Valor test LR: 4,7089, valor p: 0,0300

Error estándar asimétrico: 0,0750, valor z: 2,1443, valor p: 0,0320

Estadístico de Wald: 4,5979, valor p: 0,03201

Log Verosimilitud: -1534,207 para modelo de error

ML variancia residual (sigma cuadrado): 38,813, (sigma: 6,23)

Número de observaciones: 472

Número de parámetros estimados: 6

AIC: 3080,4 , (AIC para Lm: 3083,1)

Modelo SAR Auto Regresivos

La tabla 5 muestra los resultados obtenidos por el modelo espacial autoregresivo Sar. Al igual que los modelos anteriores, todos los coeficientes de regresión estimados son estadísticamente significativos; y además son bastante similares a los obtenidos por los dos modelos anteriores. Por otra parte, se tiene un valor estimado de rho de 0.05, el cual no estadísticamente significativo.

Tabla 5: Modelo SAR Auto Regresivos

Coeficientes	Estimado	Error Estandar	Valor t	Pr(> Z)
Intercepto	-24,0356	2,5293	-9,5026	< 2,2e-16 ***
Relación del sector primario con el terciario	0,0406	0,0031	13,5638	< 2,2e-16 ***
Indicador del seguro social de la persona jefa del hogar	0,2644	0,0407	6,5043	7,803e-11
Relación de dependencia demográfica	0,8862	0,0632	14,0189	< 2,2e-16 ***

Rho: 0,054687, Valor test LR: 1,9517, valor p: 0,1624

Error estándar asimétrico: 0,04008, valor z: 1,3645, valor p: 0,17242

Estadístico de Wald: 1,8617, valor p: 0,17242

Log Verosimilitud: -1535,585 para modelo de resago

ML variancia residual (sigma cuadrado): 39,188, (sigma: 6,26)

Número de observaciones: 472

Número de parámetros estimados: 6

AIC: 3083,2 , (AIC para Lm: 3083,1)

Valor del test: 2,2965, valor p: 0,12966

Los siguientes modelos intentan maximizar la función de máxima verosimilitud a través del paquete spdep mediante la función spautolm del software estadístico R. La tabla 6 muestra los resultados del modelo Sar ajustado por esta función. Nuevamente se tienen coeficientes de regresión bastante similares a los modelos anteriores; pero para este caso, se tiene un Rho estimado de 0.16 el cual es estadísticamente significativo, lo cual refleja la importancia de la estructura espacial en el modelo.

Tabla 6: Modelo Auto SAR

Coeficientes	Estimado	Error Estandar	Valor t	Pr(> Z)
Intercepto	-23,6440	2,5723	-9,1919	< 2,2e-16 ***
Relación del sector primario con el terciario	0,0407	0,0031	13,2021	< 2,2e-16 ***
Indicador del seguro social de la persona jefa del hogar	0,2757	0,0408	6,7646	1,337e-11
Relación de dependencia demográfica	0,9079	0,0624	14,5601	< 2,2e-16 ***

Lambda: 0,16085, Valor test LR: 4,7089, valor p: 0,0300

Hessian error estandar de lambda: 0,072839

Log Verosimilitud: -1534,207

ML variancia residual (sigma cuadrado): 38,813, (sigma: 6,23)

Número de observaciones: 472

Número de parámetros estimados: 6

AIC: 3080,4

La tabla 7 muestra los resultados obtenidos al ajustar un modelo Car mediante la función anterior. Nuevamente se aprecian valores similares en los coeficientes estimados en comparación a los modelos anteriores; todos estadísticamente significativos. Este modelo estima un valor de lambda de 0.33, el cual además es estadísticamente significativo.

Tabla 7: Modelo Auto Car

Coeficientes	Estimado	Error Estandar	Valor t	Pr(> Z)
Intercepto	-23,3200	2,5786	-9,0436	< 2,2e-16 ***
Relación del sector primario con el terciario	0,0399	0,0031	12,9271	< 2,2e-16 ***
Indicador del seguro social de la persona jefa del hogar	0,2861	0,0408	7,0130	2,332e-11 ***
Relación de dependencia demográfica	0,8990	0,0624	14,4015	< 2,2e-16 ***

Lambda: 0,33283, Valor test LR: 5,111, valor p: 0,023774

Hessian error estandar de lambda: 0,13847

Log Verosimilitud: -1534,006

ML variancia residual (sigma cuadrado): 38,57, (sigma: 6,2105)

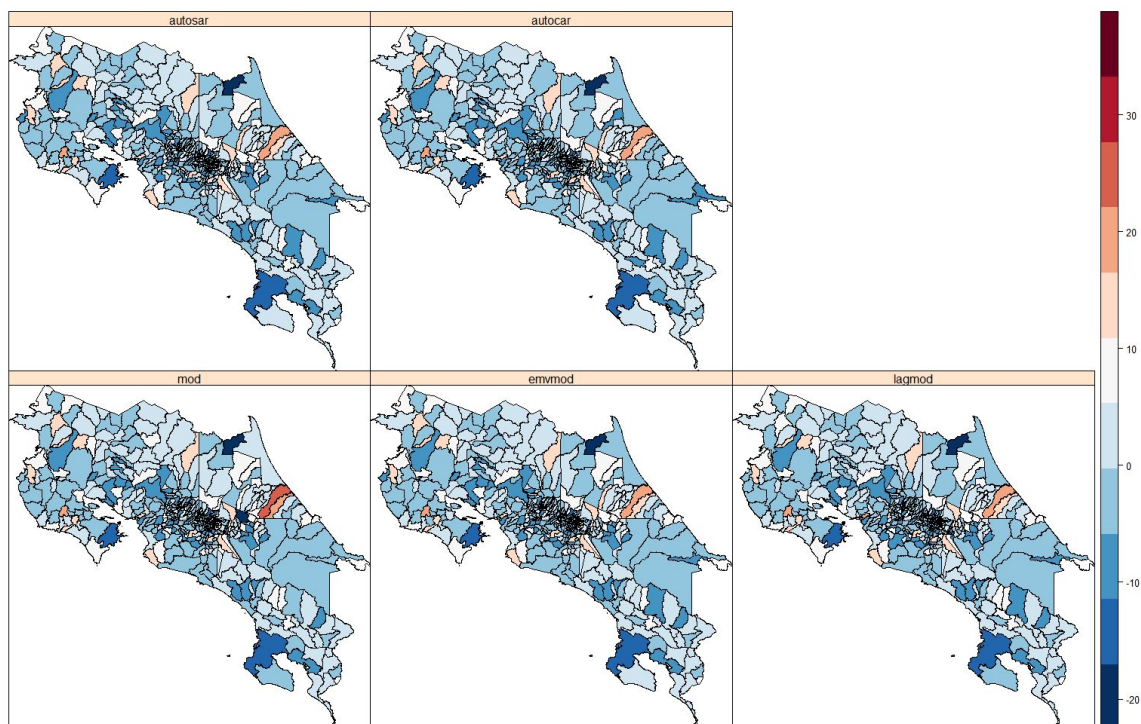
Número de observaciones: 472

Número de parámetros estimados: 6

AIC: 3080

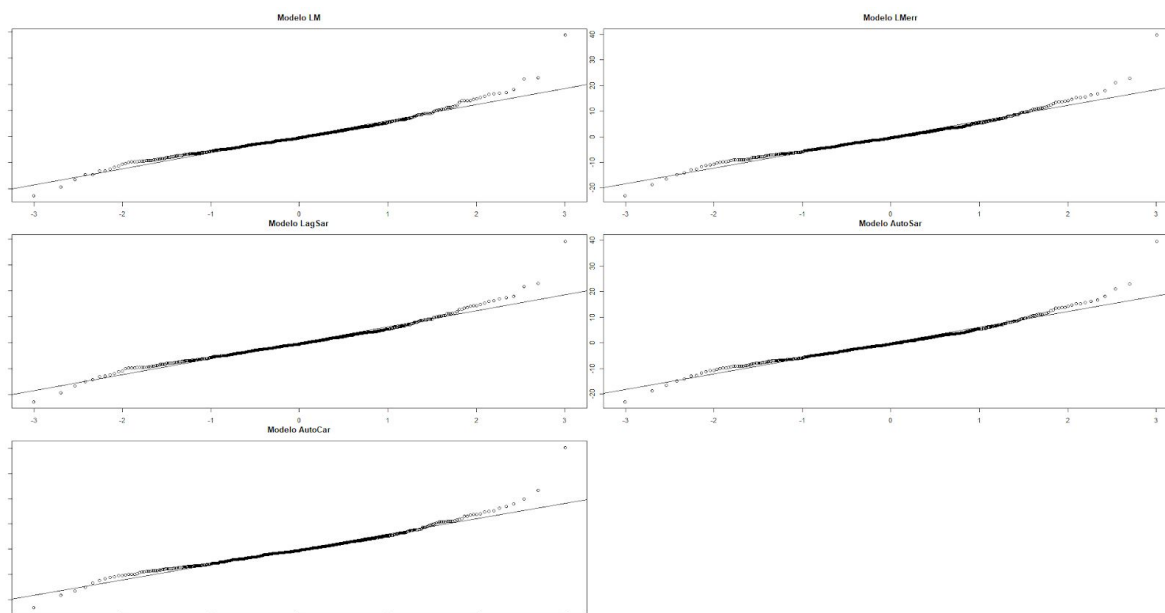
A continuación los gráficos con los residuos de los diferentes modelos para cada uno de los distritos. Se puede apreciar la alta similitud entre todos los residuos, pero la figura presenta limitaciones para analizar los residuos de la gran área metropolitana. Pese a esto, es de esperar que dada la similitud de los coeficientes de regresión de todos los modelos estimados, los residuos tienden a ser muy similares entre modelos.

Figura 6: Gráfico de los residuos de los diferentes modelos realizados para cada distrito



Siempre analizando los residuos, la figura 7 muestra la relación de cuantiles residuales observados y los cuantiles teóricos de una distribución normal, para cada uno de los modelos estimados en el trabajo.

Figura 7: Gráficos cuantil-cuantil de residuos de los diferentes modelos.



Tal y como es de esperar, al realizar la prueba formal de Shapiro wilk de normalidad de los residuos de los diferentes modelos, esta se rechaza en todos los casos, lo que coincide con las gráficas, las cuales muestran un buen ajuste a la distribución normal, pero con valores en las colas que influyen en el cumplimiento de este supuesto.

Finalmente, para seleccionar el modelo con mejor ajuste se lleva a cabo la comparación de estos a través del criterio de información de Akaike. La tabla 8 muestra que para este caso, el modelo autoregresivo Car estimados a través de la maximización de la verosimilitud, pues su valor es el más pequeño de todos.

Tabla 8: Criterio de información de Akaike (AIC)

Modelos	AIC
Gaussiano (mod)	3083,1220
Error estándar (Emvmod)	3080,4130
Sar con rezago (Lagmod)	3083,1700
Auto sar (autosar)	3080,4130
Auto car (autocar)	3080,0110

5 CONCLUSIÓN

El problema de ignorar la estructura espacial de los datos implica que las estimaciones de modelos de regresión sin estructura espacial pueden ser sesgadas, inconsistentes o ineficientes, esto si la dependencia subyacente es de esta naturaleza. Por ello, resulta siempre necesario probar si existe dependencia espacial en este tipo de fenómenos.

Para el porcentaje de hogares con NBI sin embargo, se puede apreciar que los modelos con estructura espacial y el modelos sin esta, estiman coeficientes de regresión bastante similares; y se observan comportamientos semejantes entre los residuos de los diferentes modelos.

Esto se puede deber a que el valor del índice de Moran es de 0.06; es decir, el indicador de correlación espacial es muy cercano a cero. Dado esto, es posible que su significancia estadística se deba al bajo valor esperado y su baja varianza. Por esto se recomienda analizar y considerar las limitaciones del test de Moran, pues puede arrojar información confusa.

De mano a lo anterior, se debe tomar en cuenta que la definición del tipo de vecindades, así como el estilo de sus matrices influye directamente en los resultados obtenidos; por ello, se debe hacer un esfuerzo de ligar aspectos teóricos para fundamentar el uso de una u otra manera para trabajar el modelaje espacial.

Pese a estas implicaciones, como resultado se tiene que un modelo CAR autoregresivo es el que presenta el mejor ajuste de los datos, y que pese a que los coeficientes de regresión son bastante similares a los obtenidos por el modelo no espacial, se asume que las estimaciones hechas por este modelo espacial son insesgadas, consistentes y eficientes.

6 BIBLIOGRAFÍA

- Alfaro M. (2019). Notas clase del curso Estadística Espacial.
<https://mediacionvirtual.ucr.ac.cr>
- CEPAL (2011). Hacia una nueva definición de “rural” con fines estadísticos en América Latina. Naciones Unidas.
- Food and Agriculture Organization of the United Nations (FAO). (2015). Rural Households’ Livelihood and Well-Being. United Nations Economic Commission for Europe (UNECE)
- Sancho J. y Reinoso D. (2012). Construcción de una tipología rural/urbana para los municipios españoles. Estudios Geográficos.
- Goerlich F., Reig E., Cantarino I. (2016). Goerlich F., Reig E., Cantarino I. (2016). Construcción de una tipología rural/urbana para los municipios españoles.
- Sáez C. (1975). La definición de lo urbano. ISSN 0014-1496, Vol. 36, Nº 138-139. págs. 265-302.

ANEXO: Ficha Técnica de las Variables Utilizadas

Dimensión	Nombre del indicador	Definición conceptual	Fórmula de cálculo	Componentes de la fórmula de cálculo	Unidad de medida del indicador	Interpretación	Desagregación	Línea base	Meta
Sociodemográfica	Relación de dependencia demográfica (RDD)	Se expresa como la cantidad de personas de 0 a 14 años y de 65 años y más (edades de dependencia), por cada cien personas de 15 años a 64 años de edad (edades productivas)	$RDD = \left(\frac{\text{Población menor de 15 años} + \text{Población de 65 años y más}}{\text{Población de 15 a 64 años}} \right) * 100$	<ul style="list-style-type: none"> * Población menor de 15 años * Población de 65 años y más * Población de 15 a 64 años * Población total 	Relación	Cantidad de personas en edad dependiente por cada 100 personas en edad de trabajar	<ul style="list-style-type: none"> * País * Región * Provincia * Cantón * Distrito * UGM 	47,2	Obtener un indicador que permita clasificar áreas geográficas, viviendas y población en la variable de zona.
Sociodemográfica	Índice de tipo de seguro social del jefe	Es la relación porcentual de los jefes de hogar con seguro social por el estado o sin seguro, respecto al total de hogares	Tipo de seguro social= (Jefes de hogar con tipo de seguro por el estado o sin seguro)/(total de hogares)*100	<ul style="list-style-type: none"> * Jefes sin seguro o con tipo de seguro por el estado * Total de hogares 	Porcentaje	Porcentaje de hogares con condición de aseguramiento por el estado o sin seguro.	<ul style="list-style-type: none"> * País * Región * Provincia * Cantón * Distrito * UGM 		Obtener un indicador que permita clasificar áreas geográficas, viviendas y población en la variable de zona.
Sociodemográfica	Relación sector primario y sector terciario	Corresponde a la cantidad de personas del sector primario de la economía por cada 100 personas que trabajan en el sector terciario.	$RSTP = \left(\frac{\text{Población que trabaja en el sector primario}}{\text{Población que trabaja en el sector terciario}} \right) * 100$	<ul style="list-style-type: none"> * Población de personas que trabajan en el sector primario de la economía * Población de personas que trabajan en el sector terciario de la economía 	Relación	Cantidad de personas que trabajan dentro del sector terciario por cada 100 personas que trabajan en el sector primario	<ul style="list-style-type: none"> * País * Región * Provincia * Cantón * Distrito * UGM 	20,2	Obtener un indicador que permita clasificar áreas geográficas, viviendas y población en la variable de zona.