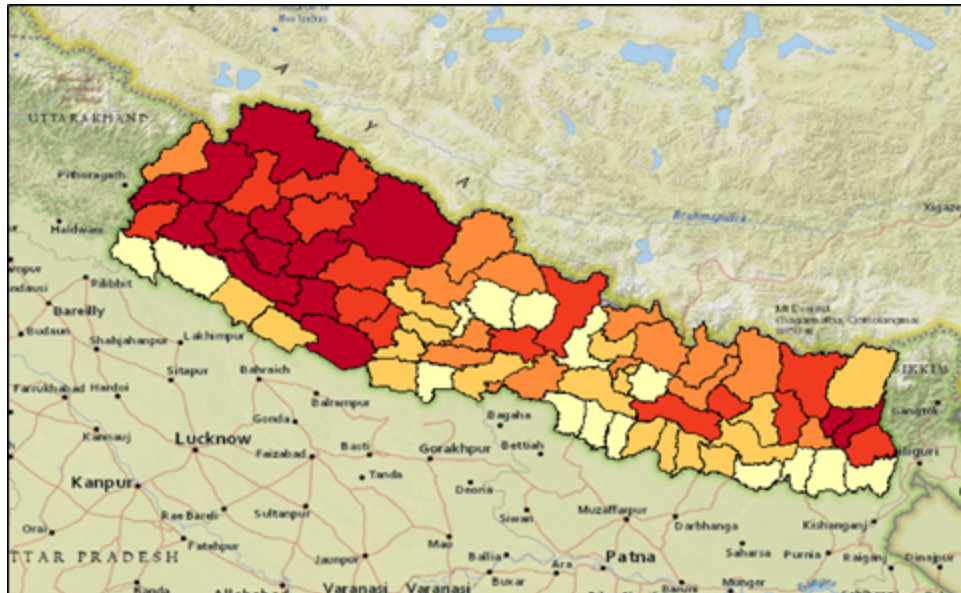


Análisis de Autocorrelación Espacial Privación de Abastecimiento Económico en Nepal 1997-2014



Daniela Marín López - 20142025027

Ana María Rojas Cantor - 20142025036

Carol Tatiana Chicuaзуque Gutiérrez - 20142025105

**Análisis de Autocorrelación Espacial Privación de Abastecimiento Económico en Nepal
1997-2014**

Presentado Por.

Daniela Marín - 20142025027

Ana María Rojas Cantor - 20142025036

Carol Tatiana Chicuazuque Gutiérrez - 20142025105

Presentado A.

Carlos Eduardo Melo Martínez

Universidad Distrital Francisco José de Caldas

Facultad de Ingeniería

Ingeniería Catastral y Geodesia

Geoestadística

Bogotá, D. C.

05 de Junio de 2018

Lista de Ilustraciones

<i>Ilustración 1.</i> Mapa de Hong Kong	11
<i>Ilustración 2.</i> Mapa del estadístico LISA.	11
<i>Ilustración 3.</i> a) Resultado del índice de privación basado en AHP. b) Escenarios OWA del índice de privación.	14
<i>Ilustración 4.</i> Análisis local de dependencia espacial	16
<i>Ilustración 5.</i> Análisis dependencia espacial local	17
<i>Ilustración 6.</i> Contigüidad en un entramado regular.	19
<i>Ilustración 7.</i> Matriz de interacciones para nueve datos situados en una cuadrícula.	20
<i>Ilustración 8.</i> Mapa división política de Nepal. Software ArcGis 10.5.	26
<i>Ilustración 9.</i> Histogramas Variables Independientes y Dependiente.	27
<i>Ilustración 10.</i> Q-Q plot variable respuesta	28
<i>Ilustración 11.</i> Diagramas de Dispersión PCINC vs PCINCM y PCINC vs PCINCP	28
<i>Ilustración 12.</i> Diagramas de Dispersión entre las variables BOYG1_5 , KIDS1_5	28
<i>Ilustración 13.</i> Criterios basados en distancia	29
<i>Ilustración 14.</i> Correlograma espacial a partir de la matriz de orden 5.	30
<i>Ilustración 15.</i> Histogramas de las matrices de contigüidad. a. Orden 1. b. Orden 2. c. Orden 3. d. Orden 4. e. Orden 5.	30
<i>Ilustración 16.</i> Distribución tasa de Privación de Aprovechamiento Económico	31
<i>Ilustración 17.</i> a) Test de Moran. b) Dispersograma de Moran	32
<i>Ilustración 18.</i> Dispersogramas Bivariados de Moran con variables Estandarizadas.	33
<i>Ilustración 19.</i> I de Moran Bivariado.	33
<i>Ilustración 20.</i> Correlogramas Bivariados de Moran.	34
<i>Ilustración 21.</i> Mapas de significancia LISA.	35
<i>Ilustración 22.</i> I de Moran Local, clusters locales.	36
<i>Ilustración 23.</i> Gráficos de detección de normalidad de los residuos	41
<i>Ilustración 24.</i> Gráficos de detección de homocedasticidad Residuales vs Predicciones.	42
<i>Ilustración 25.</i> Efectos de un cambio en el distrito de Palpa	43

Lista de Tablas

<i>Tabla 1.</i> Resultados de la Estimación de Modelos Espaciales.	9
<i>Tabla 2.</i> Criterios para la construcción del índice de privación.	12
<i>Tabla 3.</i> GWR para todas de las regresiones realizadas.	14
<i>Tabla 4.</i> Estadísticas I Morán para todos los residuales de las regresiones realizadas.	15
<i>Tabla 5.</i> Diferencias entre estadísticas globales o locales.	21
<i>Tabla 6.</i> Descripción base de datos.	25
<i>Tabla 7.</i> Estadísticas Descriptivas Variables Independientes y Dependiente.	27
<i>Tabla 8.</i> Conexiones presentadas en los criterios de contigüidad evaluados.	29
<i>Tabla 9.</i> Correlograma espacial para el DEPECPROV a partir de la matriz de orden 5 efecto reina	30
<i>Tabla 10.</i> I de Moran Global para las variables independientes.	32
<i>Tabla 11.</i> I de Moran Local para la variable dependiente	37
<i>Tabla 12.</i> G de Getis y Ord Local variable dependiente.	37
<i>Tabla 13.</i> Regresión lineal clásica obtenida. Software R	38
<i>Tabla 14.</i> Pruebas para verificación de supuestos	38
<i>Tabla 15.</i> Pruebas de los Multiplicadores de Lagrange	38
<i>Tabla 16.</i> Mejor Modelo Spatial Error. Software R	39
<i>Tabla 17.</i> Modelo Spatial Lag. Software R	40
<i>Tabla 18.</i> Test de normalidad para los residuos del modelo Spatial Lag	41
<i>Tabla 19.</i> Prueba de Heterocedasticidad para los residuos del modelo Spatial Lag	41
<i>Tabla 20.</i> Prueba de autocorrelación para los residuos del modelo Spatial Lag	42
<i>Tabla 21.</i> Diferencia entre las predicciones originales y las obtenidas	43
<i>Tabla 22.</i> Medidas de Impactos	45

Tabla de Contenido

Lista de Ilustraciones	3
Lista de Tablas	4
1. Introducción	6
2. Justificación	6
3. Objetivos	7
3.1. Objetivo General	7
3.2. Objetivos Específicos	7
	4

4.	Estado del arte	8
5.	Marco Teórico	18
	Conceptos de Econometría Espacial	18
	Análisis exploratorio de datos	18
	Contigüidad	19
	Matriz de Interacciones Espaciales	19
	Matriz de pesos espaciales	20
	Estadísticos globales y locales de autocorrelación espacial	21
	Análisis Confirmatorio de Datos Espaciales	22
6.	Descripción de la base de datos	25
7.	Análisis exploratorio	26
7.1.	Datos espaciales	26
	Área de estudio	26
	Estadísticas Descriptivas	26
	Matriz de Pesos Espaciales	29
	Análisis Exploratorio de los Datos Espaciales	31
	“Variable Dependiente: Tasa de Privación de Aprovechamiento Económico ”	31
	Análisis Exploratorio Global para la Variable Dependiente	31
	I de Moran Global para las Variables Independientes	32
	I de Moran Bivariado	32
	Correlogramas Bivariados	34
	Análisis Exploratorio Local	35
8.	Análisis Confirmatorio	37
8.1.	Modelo Clásico	37
8.2.	Pruebas de Multiplicadores de Lagrange	38
8.3.	Modelo Spatial Error (Espacial Autorregresivo en el Error)	39
8.4.	Modelo Spatial Lag (Autorregresivo de rezago espacial)	40
	8.4.1. Análisis de Supuestos	40
	8.4.2. Análisis de Impactos	43
9.	Conclusiones	45
10.	Referencias	46

1. Introducción

El análisis de autocorrelación espacial como instrumento de la estadística espacial nos permite determinar si un fenómeno particular, en este caso social, tiene un comportamiento dependiente de las características poblacionales y de localización de cada región de una zona en particular. De acuerdo al Informe sobre Desarrollo Humano del PNUD 2016 de la UNESCO, el país de Nepal es uno de los países con menor nivel de desarrollo y aproximadamente el 40% de la población vive en pobreza; lo cual genera a su vez la presencia de inequidad y baja calidad de vida. El presente informe busca determinar la existencia de una relación directa entre la privación de abastecimiento económico que sufre esta nación y las condiciones mencionadas anteriormente, y con ello determinar si su manifestación en cada distrito está relacionada con las características de los distritos vecinos, seleccionados mediante la matriz de pesos espaciales efecto reina, orden 2 (cuyo número de links no nulos resultó ser mayor en comparación con otras matrices generadas).

A continuación, se presenta la aplicación de diferentes estadísticos tanto globales como locales, su visualización gráfica e interpretación para la detección de ausencia o presencia de autocorrelación espacial y de asociaciones de valores, junto al modelamiento espacial de los datos y el respectivo análisis de impactos del modelo.

2. Justificación

La privación del abastecimiento económico genera un alto impacto en la calidad de vida de una sociedad donde a consecuencia de la misma, se presenta la falta de acceso a recursos básicos generando pobreza en una región. De acuerdo al reporte de desarrollo humano de 2001 existe una relación directa entre el índice de pobreza humana, la esperanza de vida, el analfabetismo en adultos y el abastecimiento económico; la aplicación de métodos estadísticos nos permite contrastar dichas relaciones, además de estimar el comportamiento estadístico del desabastecimiento económico con respecto a la ubicación geográfica en este caso para el país de Nepal.

3. Objetivos

3.1. Objetivo General

Presentar los resultados obtenidos en el análisis exploratorio y confirmatorio de los datos espaciales registrados en Nepal respecto a variables relacionadas con el desabastecimiento económico del país y las características de su población.

3.2. Objetivos Específicos

- Verificar la ausencia o presencia de autocorrelación espacial en las tasas de desabastecimiento económico de los distritos de Nepal, haciendo uso de pruebas univariadas y bivariadas del I de Moran.
- Interpretar los resultados de las medidas de tendencia central de las variables *índice de pobreza humana, ingreso per cápita, porcentaje de niños menores a 5 años con desnutrición, personas que no se espera vivan más de 40 años*, entre otras, para establecer su comportamiento estadístico.
- Mostrar los resultados de autocorrelación espacial registrados para la tasa de desabastecimiento económico y variables similares en diferentes regiones del mundo por medio artículos publicados en revistas especializadas.
- Estimar un modelo espacial para determinar la tasa de privación en el abastecimiento económico en Nepal para los años 1994 a 2014 y analizar los impactos de cada una de las variables en el modelo seleccionado.

4. Estado del arte

La concentración geográfica de pobreza, ligada a la desigualdad en muchos otros aspectos (económico, social, salud, entre otros), son hoy el foco de estudio en muchas zonas metropolitanas. Es por ello, que se aplican modelos espaciales acompañados previamente de pruebas de análisis exploratorio entre las más comunes están el I de Morán y la regresión geográficamente ponderada (GWR), que permiten la determinación de patrones espaciales, observar relaciones o asociaciones espaciales, así como los grados de asociación entre las variables de estudio. A continuación, se presenta el resumen de algunos artículos publicados sobre el tema a tratar.

“Desarrollo, pobreza e inequidad: Un análisis espacial de las provincias sudafricanas”[3]

La pobreza continúa siendo predominante en África, indicando que el aumento del ingreso no es suficiente para disminuirla. Este artículo investiga la relación entre el crecimiento, la pobreza y la inequidad en Sudáfrica a nivel regional, utilizando información de la Regional Explorer Database (REX), periodos anuales entre 1996 y 2013. Se adopta un modelo econométrico espacial motivado por dos hechos: primero la alta tasa de pobreza en Sudáfrica y segundo los altos niveles de pobreza e inequidad en las regiones estudiadas. Adicionalmente se tiene en cuenta que la autocorrelación espacial entre las regiones puede afectar la relación entre el desarrollo y la pobreza.

Sudáfrica está dividida en 9 provincias: Limpopo, Mpumalanga, North West, Gauteng, Northern Cape, Free State, Eastern Cape, Western Cape y KwaZulu Natal.

La tasa de pobreza más alta se encuentra en la Provincia Eastern Cape y la tasa más baja en Gauteng; Northern Cape oscila alrededor de la pobreza media y casi todas las provincias están por encima del promedio. El PIB per cápita revela un perfil distinto con en la provincia de Gauteng con el valor más bajo; el coeficiente Gini (medida de desigualdad) más bajo se encuentra en Free State; mientras el crecimiento del PIB es similar a lo largo de todas las provincias.

El modelo de regresión espacial usado permite tener en cuenta la dependencia entre las observaciones ubicadas en el espacio. El modelo de pobreza a ser estimado es el siguiente:

$$\log Poverty_{it} = \alpha_{it} + \beta_1 Trend_{it} + \beta_2 SqTrend_{it} + \beta_3 \log percapita income_{it} + \beta_4 \log GDP growth_{it} + \beta_5 \log Employment_{it} + \beta_6 \log Gini_{it} + \mu_{it} \quad (1)$$

Donde:

- *logPoverty*: Logaritmo del número total de personas pobres
- *Trend*: Variable tendencia que es rango desde 1 en 1996 hasta 18 en 2013
- *SqTrend*: Cuadrado de la tendencia
- *log per capita income*: Logaritmo del PIB per cápita
- *logGDPGrowth*: Logaritmo del crecimiento del PIB
- *logemployment*: Logaritmo de la población empleada

- *Gini*: Indicador Gini
- Se utilizaron las coordenadas de las ciudades capitales de cada una de las nueve provincias.

Inicialmente se investigó acerca de la existencia de autocorrelación espacial con el estadístico I de Moran Global, el valor que arrojó fue de 3.212 y una significancia estadística de 0.000. Es decir, existe autocorrelación espacial en los datos, justificando el uso de modelos espaciales. En la *Tabla 1* se enuncian los resultados de la estimación de los 5 modelos realizados.

Tabla 1. Resultados de la Estimación de Modelos Espaciales. Fuente: Barros et al. (2017)[3]

	SDM Modelo Espacial con aleatorización	SAC Modelo Espacial ajustado	SAR Modelo espacial con aleatorización	Han- Phillips SDM Modelo Espacial Dinámico	Modelo Espacial Arellano- Bond
<i>Intercepto</i>	12.493	-	0.158	-13.09	3.725
<i>Logpoor_{t-1}</i>	-	-	1.001	2.350	0.289
<i>W logpoor_{t-1}</i>	-	-	-	0.007	0.012
<i>Trend</i>	0.0009*	0.020	0.0166*	0.025	0.040*
<i>SqTrend</i>	-0.002*	-0.002	-0.0005*	-0.0004	-0.0015*
<i>Log per capita income</i>	-0.551*	-0.332	-0.0205	-0.501*	-0.389*
<i>LogGDPGrowth</i>	0.444	-0.047	-0.1294	-0.002	-0.313*
<i>LogEmployement</i>	-0.304**	-0.520*	-0.0383	-0.016	-0.061
<i>Gini</i>	6.227*	6.183*	0.0760	1.916*	1.170
<i>R²</i>	0.685	0.558	0.938	0.862	0.947
<i>ρ</i>	7.971	2.053*	0.442*	0.007*	0.012*
*, ** Significancia al 1% y 5% respectivamente					

El modelo elegido es el espacial de Arellano- Bond. Interpretando los resultados, la tendencia es positiva, por lo tanto, la pobreza está incrementando en el tiempo. Entre mayor sea el PIB per cápita, el crecimiento del PIB y el empleo menor será la pobreza. Al contrario, entre mayor sea el indicador Gini, mayor será la pobreza.

Se observa que el parámetro autorregresivo rezagado *W logpoor_{t-1}* es positivo y significativo estadísticamente en todos los casos, es decir, que la pobreza está incrementando en Sudáfrica.

Aunque las variables económicas hacen que disminuya la tasa de pobreza, se concluye que no es suficiente, por lo tanto, es necesario que el gobierno tome medidas para atacar este problema persistente.

“La concentración de la pobreza en una ciudad próspera: Variación geográfica y correlaciones de las tasas de pobreza en vecindarios en Hong Kong” [5]

La concentración geográfica de pobreza urbana se ha convertido en una de los temas sociales más importantes en muchas áreas metropolitanas contemporáneas. Los objetivos de este estudio fueron, investigar los patrones espaciales de pobreza de los barrios; identificar áreas con pobreza concentrada; explorar localmente y a lo largo de toda la ciudad la correlación espacial de la pobreza; y revelar las diferentes formas de concentración de la pobreza en Hong Kong.

Se utilizaron datos del censo realizado en el año 2011 en Hong Kong, de la Geo-Community Database, y la Geo-Reference Database, que contienen la información necesaria para calcular la tasa de pobreza en los barrios y otros correlatos socioeconómicos. La unidad geográfica usada para indicar un vecindario fue el gran bloque de la calle (LSB - *Large Street Block*). La totalidad del territorio de Hong Kong se dividió en 1620 LSB. La tasa de pobreza de cada LSB se calculó como la variable dependiente. Se construyeron variables independientes, según los siguientes conceptos:

- Segregación basada en clases
 1. *Tasa de Desempleo*
 2. *Proporción de personas con empleos no profesionales*
 3. *Proporción de personas con educación secundaria o menor.*
- Grado de segregación basada en la inmigración
 4. *Proporción de nuevos inmigrantes chinos*
- Estructura demográfica
 5. *Proporción de hogares de padres solteros*
 6. *Proporción de hogares de adultos mayores solteros*
 7. *Radio de dependencia (número de personas entre 0 y 14 años, y mayores de 65, dividido por el número de personas entre 15 -64 años)*
 8. *Densidad Poblacional*
- Características del vecindario influenciadas por políticas públicas
 9. *Proporción de hogares viviendo en casas públicas*
 10. *Densidad vial*
 11. *Acceso a hospitales*
 12. *Acceso a servicios de actividades físicas*

El análisis exploratorio de datos fue usado para investigar los patrones espaciales de la pobreza en los vecindarios. Para examinar la autocorrelación espacial, se utilizó el estadístico I de Moran. Los que comparten un límite fueron definidos como vecinos.

Para modelar la relación entre tasas de pobreza en los barrios y sus correlativos, se llevaron a cabo dos tipos de modelos: Regresión mediante mínimos cuadrados ordinarios (OLS) y Regresión Geográficamente Ponderada (GWR). En el modelo OLS se asume que los procesos espaciales que representan los niveles de pobreza del vecindario son los mismos en toda el área de estudio (la relación entre la variable dependiente y las variables independientes es espacialmente estacionaria); en contraste el

modelo GWR ofrece una oportunidad para evaluar el grado en que la asociación entre la variable dependiente y las variables independientes varía a lo largo del área de estudio. Este método realiza un modelo local para cada unidad de área. La elección del número óptimo de vecinos más cercanos se basó en el que hiciera menor el criterio de información Akaike (AIC). El análisis se llevó a cabo en ArcGis 10.2; el número de vecinos óptimos determinado fue 465.

La Ilustración 1 representa el mapa de Hong Kong, con tres regiones principales Hong Kong Island, Kowloon, y the New Territories. Se identifican mucho mayores índices de pobreza en New Territories y Kowloon. El I de Moran fue $0.11 (p < 0.001)$, indicando un bajo grado de autocorrelación espacial general en las tasa de pobreza a lo largo de Hong Kong. La Ilustración 2 muestra cuatro diferentes tipos de cluster; las áreas resaltadas en rojo tienen una tasa de pobreza relativamente alta y sus áreas vecinas también muestran alta concentración de pobreza; se observan aparentemente siete regiones con valores agrupados.

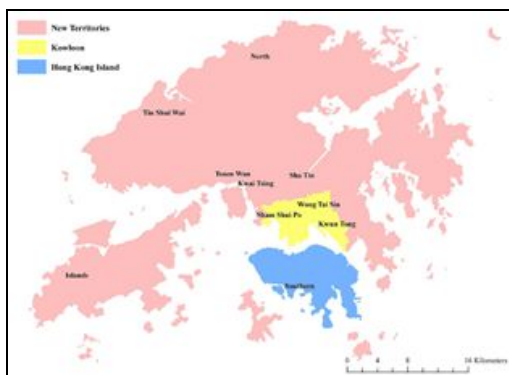


Ilustración 1. Mapa de Hong Kong.
Fuente: Guo et al. (2018)[5]

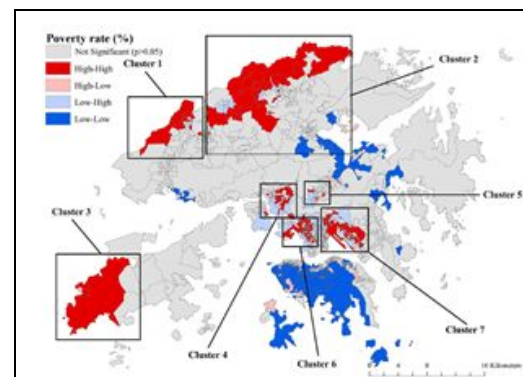


Ilustración 2. Mapa del estadístico LISA
Fuente: Guo et al. (2018)[5]

El modelo OLS mostró que, con excepción de la densidad de población, todas las demás variables de condición socioeconómica, a saber, la segregación basada en la inmigración y la estructura demográfica mostraron una asociación positiva con la tasa de pobreza del vecindario. En términos de variables relacionadas con las políticas públicas, la proporción de hogares que viven en viviendas públicas se asoció positivamente con las tasas de pobreza; por el contrario, no hubo evidencia estadística de una asociación entre la tasa de pobreza del vecindario con el acceso a hospitales o actividades físicas. La densidad de carreteras se asoció positivamente con la tasa de pobreza, contrariamente a nuestra expectativa de una asociación negativa. El R^2 ajustado fue de 0.656.

Los resultados del modelo GWR muestran la heterogeneidad de la asociación espacial con la tasa de pobreza para algunas variables. Por ejemplo, la variación espacial en la asociación con la tasa de pobreza fue particularmente aparente para acceder a hospitales, con un 11.30% de LSB con asociación negativa (principalmente en el noroccidente de New Territories) y un 18.23% de LSB mostrando una asociación positiva (principalmente en Kowloon). Los resultados de ambos modelos fueron más consistentes para la tasa de desempleo y la proporción de población con baja educación que para otras variables - asociación

positiva significativa con la tasa de pobreza fue encontrada en las del 85% de LSB. El R^2 ajustado del modelo GWR fue de 0.76, lo que indica un mejoramiento del 7% sobre el modelo OLS.

“Índice de privación espacial multicriterio para apoyar el análisis de desigualdad en salud”[4]

El estudio de caso se lleva a cabo en el área urbana del Distrito Metropolitano de Quito y sus 34 distritos urbanos en los cuales se distribuyen las 1.5 millones de personas que habitan en Quito. En este artículo se muestra la aplicación y desarrollo de un índice de privación haciendo uso de técnicas de toma de decisiones multicriterio. Esta metodología muestra cómo se aplica el Proceso de Jerarquía Analítica (AHP) la cual permite obtener los pesos para cada uno de los indicadores que estructuran el índice de privación. Luego es necesario emplear el promedio ponderado ordenado (OWA) para identificar los diferentes escenarios de privación. A partir de los bloques censales suministrados correspondientes al año 2010 se les aplicará un análisis exploratorio espacial utilizando I de Moran y una Regresión Geográficamente Ponderada (GWR) al índice de privación y sus escenarios propuestos para evaluar cómo es su relación espacial con factores de la salud, tales como: distancia a los servicios de salud y el porcentaje de personas que nunca nacieron.

A continuación, se muestra los diferentes indicadores considerados para la construcción del índice de privación. Cada indicador se considera como un criterio para el Proceso de Jerarquía Analítica:

Tabla 2. Criterios para la construcción del índice de privación. Fuente: Cabrera et al. (2015) [4]

Dimensión	Variable	Indicador/Criterio
Educación	Nivel de Educación	A: % de personas sin ningún nivel de educación.
Salud	Seguridad Social	B: % de personas sin seguro social.
	Personas con Discapacidad	C: % de personas que presentan discapacidades por más de un año.
Empleo	Trabajadores sin pago	D: % de personas que trabajan sin pago.
Vivienda	Estructura Viviendas	E: % de casas con techos dañados.
	Sobrepoblación	F: % de casas con más de cuatro personas por dormitorio.
	Servicios Públicos	G: % de viviendas sin sistema de alcantarillado.
		H: % de casas sin conexión a la red eléctrica.
		I: % de casas sin conexión al sistema de suministro de agua potable.
		J: % de casas sin caminos de acceso adoquinados o pavimentados.
		K: % de casas sin servicio de recolección de basura.
	Condición general de la vivienda	L: % de viviendas en choza

Los criterios seleccionados deben cumplir los siguientes requisitos:

- Considerar un enfoque en derechos humanos.
- Estar relacionado con la salud y tener una afinidad con las dimensiones sociales de privación.
- Poder ser representado en el nivel del bloque de censo.

Por otra parte, las variaciones de los factores de inflación (VIF) fueron calculadas para cada uno de los indicadores utilizados a fin de identificar colinealidades múltiples. El VIF muestra cuánto se incrementa la varianza de un coeficiente de regresión estimado como resultado de las colinealidades entre dos variables.

Modelo:

- Todos los VIF obtenidos fueron menos de 5, lo que significa que todos los indicadores seleccionados pueden ser usados para la construcción del índice.
- La regresión geográficamente ponderada (GWR) se aplicó considerando las medidas de privación y la distancia a los servicios de salud como las variables explicativas.
- El indicador del porcentaje de personas que nunca nacieron se consideró como la variable dependiente.
- Se realizó una regresión ponderada geográficamente (GWR) diferente para cada escenario de privación de promedio ponderado ordenada (OWA) y para el índice de privación basado en el Proceso de Jerarquía Analítica.
- La regresión geográficamente ponderada es una extensión de las técnicas de regresión estándar que permite que los parámetros β_k varíen espacialmente; también evalúa las variaciones de las relaciones entre los modelos de regresión espacial.

El modelo GWR se puede escribir como:

$$Y(s_i) = \beta_0(s) + \sum_{k=1}^M \beta_k(s) X_k(s_i) + \varepsilon(s_i) \quad (2)$$

- En cada ubicación s , todos los coeficientes β_k deben estimarse, y $\varepsilon(s_i)$ es un error aleatorio con una media de cero y una varianza constante.
- Las estimaciones de los coeficientes β_k requieren la ponderación de todas las observaciones

Resultados

En la Ilustración 3. los resultados del Índice de deprivación basado en AHP muestran la presencia de niveles medios y altos de privación (0,511 a 1,000) en zonas específicas del área de estudio, aunque la mayor parte de la ciudad de Quito tiene bajos valores de privación. Los mayores niveles de privación aparecen en los bordes del área de estudio, y representan barrios asentados relativamente recientemente creados por las personas socio-económicamente más desfavorecidas. Por otro lado, los niveles más bajos de privación (0,000 a 0,146) están comúnmente presentes en el lado norte de la Ciudad, una parte de Quito generalmente habitada por personas con mejores condiciones socioeconómicas. Las áreas moderadamente desfavorecidas se encuentran en el sur.

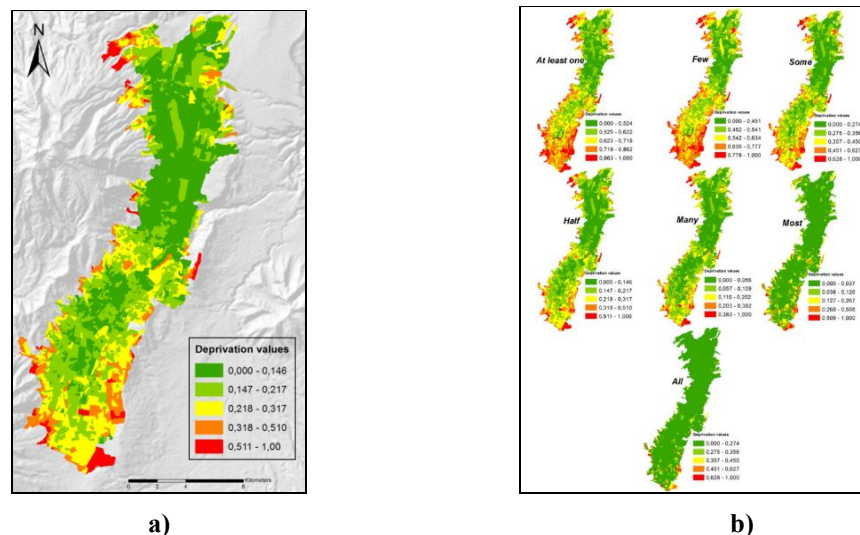


Ilustración 3. a) Resultado del índice de privación basado en AHP. **b)** Escenarios OWA del índice de privación.
Fuente: Cabrera et al (2015)[4]

En la Ilustración 4. se observan los siete escenarios propuestos por la regresión geográficamente ponderada (GWR), y se resumen así:

- En el escenario de privación "Al menos uno" representa los valores de privación más altos posibles para cada bloque del censo, con "resultados positivos" que significan "valores más altos de criterios de privación", es decir, los puntajes de privación más altos que los basados en AHP.
- El escenario "Todos" representa los valores de privación más bajos posibles para cada bloque censal.
- El escenario de privación "Mitad" es el equivalente al índice de privación basado en AHP, porque los pesos de orden son iguales y se aplican para todos los indicadores.
- Los escenarios de privación "Pocos" y "Algunos" se asignan pesos de orden mayores a valores de criterio más altos y se asignan pesos menores a los valores de criterio más bajos.
- Los escenarios de privación "Muchos" y "La mayoría" asignan mayores ponderaciones de orden a un criterio inferior de valores y pesos más pequeños se asignan a los valores de criterio más altos.

A continuación, se presenta el resultado de las GWR y las estadísticas del I Moran para todas las regresiones:

Tabla 3. GWR para todas de las regresiones realizadas.

	Basado en AHP	Al Menos Uno	Pocos	Algunos	La Mitad	Muchos	La Mayoría	Todos
AIC	18470,02	18743,45	18634,98	18483,08	18470,02	18505,63	18684,81	19653,30
R ²	0,59	0,50	0,53	0,58	0,59	0,59	0,57	0,36

- Se utilizó un kernel gaussiano para resolver cada regresión local y la extensión de la medida de la calidad de un modelo estadístico que tiene en cuenta la bondad estadística del ajuste y la compensación de los parámetros utilizados en el modelo.

- El mejor modelo es el que tiene el valor AIC más bajo.
- Los modelos GWR que tienen la mejor bondad de ajuste son el modelo "Basado en AHP" y el modelo "Mitad".
- Otros modelos con valores bajos de AIC son los modelos "Algunos" y "Muchos", que muestran la importancia de utilizar escenarios de OWA como intercambios entre un escenario neutral y escenarios extremos al describir las interacciones de privación y salud.
- Los modelos mencionados, "Basado en AHP", "Mitad", "Algunos" y "Muchos", también son modelos que representan proporciones similares de la varianza de la variable dependiente.
- R^2 entre 58% y 59%, esto significa que estas regresiones no producen una predicción de variable dependiente óptima en todas las ubicaciones.

Tabla 4. Estadísticas I Morán para todos los residuales de las regresiones realizadas.

	Basado en AHP	Al Menos Uno	Pocos	Algunos	La Mitad	Muchos	La Mayoría	Todos
Índice Moran	-0,005	0,051	0,035	-0,000	-0,006	-0,008	0,006	0,215
Puntaje z	-1,083 (Aleatorio)	10,198 (Cluster)	7,130 (Cluster)	-0,049 (Aleatorio)	-1,076 (Aleatorio)	-1,478 (Aleatorio)	1,175 (Aleatorio)	42,908 (Cluster)
p-valor	0,279	0,000	0,000	0,961	0,282	0,139	0,239	0,0000

Según la *Tabla 4* las estadísticas I de Moran:

- La hipótesis nula es la distribución aleatoria de los valores
- Identificaron agrupamientos en los valores residuales de todos los GWR realizados.
- La agrupación con altos niveles de significancia indica que faltan variables explicativas.
- Muestran una distribución aleatoria en los modelos que mostraron la mejor bondad de ajuste ("basada en AHP", "Mitad", "Algunos" y "Muchos"), así como en el modelo "La mayoría".
- Los anteriores escenarios de privación, en combinación con el factor de distancia de los servicios de salud, podría ser una variable explicativa para predecir el porcentaje de personas que nunca nacieron.
- Los modelos con la presencia de conglomerados (clusters) residuales con altos niveles de significancia ("Al menos uno", "Pocos", "Todos") son modelos que no explican completamente la variable dependiente de la salud

Finalmente, es necesaria más investigación sobre este tema y las investigaciones futuras podrían incluir la incorporación de más variables de salud explicativas que podrían interactuar con el índice de privación basado en Proceso de Jerarquía Analítica (AHP) y la privación basada en el promedio ponderado ordenada (OWA).

“Dimensión espacial de la pobreza en Colombia”[7]

Este estudio hace parte del libro Geografía económica y análisis espacial en Colombia, y tiene como finalidad mostrar que en el caso colombiano las características relacionadas con la ubicación geográfica son factores determinantes en el comportamiento de la pobreza a nivel nacional. Los datos a utilizar fueron escogidos teniendo en cuenta la disponibilidad de los mismos por lo cual se utilizaron el índice de necesidades básicas insatisfechas (NBI) y el índice de condiciones de vida (ICV) disponibles a nivel municipal para 1985 y 1993, la información faltante a nivel municipal se obtuvo por medio de interpolación espacial mediante la técnica de distancia inversa ponderada (IDW). La primera variable contempla no sólo condiciones de dependencia económica sino también de educación (asistencia escolar) y condiciones de infraestructura, mientras que la segunda considera características físicas del hogar, tomando valores de 0 a 100 donde valores más grandes indican mejores condiciones de vivienda.

Para el análisis espacial se utilizó el software Exploratory Spatial Data Analysis (ESDA), realizando un análisis local a través de indicadores de asociación espacial local (LISA) para la evaluación de hipótesis de aleatoriedad espacial de las variables en cada municipio respecto a los vecinos y el I de Moran y su gráfico para la determinación del tipo de asociación espacial. Adicionalmente se analizó la dinámica presentada en el lapso de tiempo del cual se tiene datos por medio del efecto de difusión que se puede clasificar en contagiosa y difusión espacial jerárquica. Los resultados se presentan tanto a nivel departamental como a nivel local.

A nivel departamental la aplicación del I de Moran, se hizo para primer y segundo orden con el criterio de efecto reina, para el primer orden se encontró presencia de autocorrelación espacial para los niveles de NBI mientras que para segundo orden no se evidencia ningún tipo de dependencia espacial. También hizo uso del estadístico LISA con el fin de identificar la presencia de clusters, este test arrojó como resultado la existencia de tres clusters departamentales el primero Chocó, mostrando un comportamiento alto rodeados de bajos que se puede entender como pobre rodeado de ricos, el segundo mostrando una estructura de bajo rodeados de bajos para los departamentos de Cundinamarca, Caldas, Risaralda y Tolima y el tercero conformado por los departamentos considerados nuevos donde se presenta un alto rodeado de altos, o pobres rodeado de pobres. En lo referente al ICV se encontró un comportamiento similar, en los departamentos de Guainía y Vaupés se evidencia difusión por contagio pues pasaron de tener altos niveles de calidad de vida en 1985 a tener bajos niveles de calidad de vida en 1993.

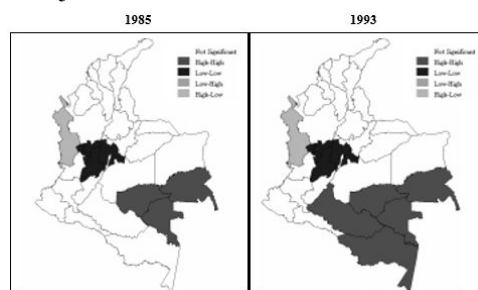


Ilustración 4. Análisis local de dependencia espacial
Fuente: (Pérez, 2005)[7]

En el análisis a nivel municipal, fue posible evidenciar comportamientos que no se veían a nivel departamental por ejemplo que los municipios más pobres hacen parte de los departamentos Bolívar, Sucre, Córdoba, Magdalena, Chocó, Casanare y Boyacá, caso contrario a Caldas, Risaralda, Quindío, Antioquia, Valle del Cauca y Cundinamarca. En el caso puntual de Cundinamarca se evidenció la relación directa en la distancia de los municipios a Bogotá, donde a mayor distancia de Bogotá mayor es la pobreza registrada; a nivel nacional el estadístico del I de Moran tanto en primer como segundo orden hay presencia de autocorrelación espacial, además se observó que los coeficientes del mismo se reducen a medida que aumenta el orden indicando que la importancia de los vecinos depende de la distancia. Adicionalmente se realizaron pruebas LISA para determinar zonas del país donde las asociaciones espaciales son significativas.

En un primer orden la parte nororiental del país (Costa Caribe y Chocó) y Nariño se caracterizan por comportarse como clúster donde se presentan altas necesidades insatisfechas al igual de sus vecinos, mientras que en la zona central del país (Sur de Antioquia, Risaralda, Quindío, Caldas y Valle de Cauca) donde bajos niveles de necesidades insatisfechas están rodeados de municipios con la misma condición. Adicionalmente se llevó a cabo el mismo ejercicio para cada una de las regiones del país, obteniendo los siguientes resultados:

- Regiones Caribe y Pacífica: El estadístico del I de Moran muestra la existencia de una dependencia espacial positiva, lo que significa que la mayoría de municipios se relacionan con vecinos de valores similares donde los clúster son significativos específicamente en la Guajira y la parte sur de Bolívar; adicionalmente se hace un ejercicio agregando una variable dummy donde se separan los municipios del norte y del sur, su aplicación permitió determinar la dicotomía entre los municipios según su ubicación. También fue posible observar que son los municipios del centro y de la periferia resultan significativos en la formación de clúster.

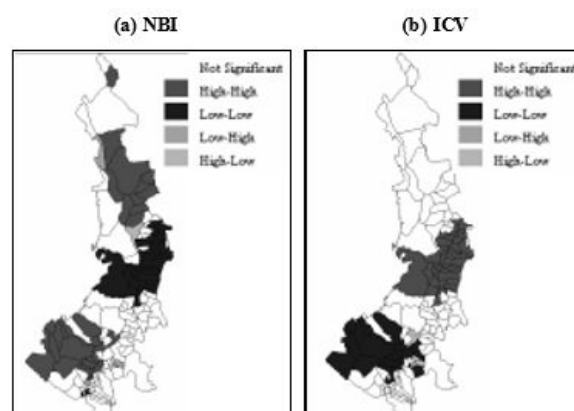


Ilustración 5. Análisis dependencia espacial local
Fuente: (Pérez, 2005)[7]

- Regiones centro - occidente, centro - sur y centro - norte: Los resultados de las pruebas de dependencia espacial mostraron presencia de autocorrelación espacial, aunque cada una de las regiones presenta una dinámica espacial diferente, pero sin embargo sobresale el hecho de la presencia de clúster espaciales.

5. Marco Teórico

Conceptos de Econometría Espacial

La econometría espacial puede ser definida como la parte de la econometría que se ocupa del tratamiento de la autocorrelación espacial y de la heterogeneidad espacial en modelos de regresión cuyo corte en transversal y datos de panel. (Anselin, 2001 [2]; citado por Yrigoyen, 2003, p. 8) [14].

- Dependencia espacial: Se refiere a la falta de independencia producida en las observaciones de corte transversal, de acuerdo a Tobler (1979) en su conocida “primera ley de la geografía” este efecto espacial se debe a que “todo tiene que ver con todo, pero las cosas cercanas están más relacionadas entre sí que las cosas lejanas”. Razón por la que autocorrelación espacial se puede definir como la presencia de una relación funcional entre los eventos ocurridos en un punto del espacio y otro lugar. (Moreno & Vayá, 2000 [6]; citado por Yrigoyen, 2003, p. 9) [14]

La dependencia espacial se produce cuando la variable dependiente en una unidad espacial se comporta como una función del valor tomado por la misma variable, pero en unidades vecinas. (Vilalta, 2005, p. 3) [13]

- Heterogeneidad espacial: Hace referencia a la variación de relaciones entre las variables en el espacio (Vilalta, 2005, p. 4) [13]. Se encuentra relacionada con la ausencia de estabilidad en el espacio del comportamiento evaluado, el problema causado por este efecto puede ser resuelto haciendo uso de la econometría estándar (Yrigoyen, 2003, pp 9-10) [14]. Este efecto espacial de acuerdo a Vilalta puede presentarse por:
 - Un problema estadístico como consecuencia de la heterocedasticidad en el modelo de regresión.
 - La existencia de una variación espacial sustantiva de la variable estudiada, donde también estará presente la heterocedasticidad.

Análisis exploratorio de datos

El AED es utilizado para la identificación de relaciones de carácter sistemático entre variables cuando no hay un a priori de la naturaleza de dichas relaciones y consiste en el tratamiento y comparación de un grupo de variables que identifican entre ellas formas estables (Yrigoyen, 2003)[14]. De acuerdo a Turkey este se puede definir como “el conjunto de herramientas gráficas y descriptivas utilizadas para el descubrimiento de patrones de comportamiento en los datos y el establecimiento de hipótesis con la menor estructura posible” (Turkey, 1997 [12]; citado por Yrigoyen, 2003, p. 28) [14].

Contigüidad

La definición de contigüidad requiere de la existencia de un mapa, a partir del cual se puedan obtener las fronteras entre unidades espaciales, cuando dichas unidades se posicionan de forma irregular, puede ser fácil conocer las fronteras entre las distintas unidades geográficas. Sin embargo, cuando las unidades pertenecen a una cuadrícula regular la determinación de la contigüidad, no es única.

Cuando las unidades espaciales son puntos regular o irregularmente distribuidos sobre el sistema, como el caso de las ciudades en una jerarquía urbana, el concepto de contigüidad se define sobre la noción de camino más corto en una red (Yrigoyen, 2003, pp. 53-54) [14].

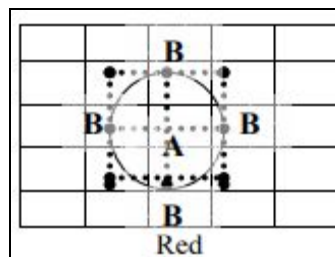


Ilustración 6. Contigüidad en un entramado regular.

Fuente: (Anselin 1988 [1]; citado por Yrigoyen, 2003, p. 54) [14].

Los nudos de la red son considerados como vecinos si se encuentran dentro de una distancia máxima (camino más corto) unos de otros. En la *Ilustración 6* los nudos B se encuentran en un radio “d” del centroide A, y pueden ser considerados como contiguos para esta concreta distancia crítica.

Matriz de Interacciones Espaciales

Es una matriz cuadrada binaria, donde cada unidad espacial se representa a la vez mediante una fila y una columna. En cada fila, los elementos no nulos de las columnas se corresponden con las unidades espaciales contiguas, es decir, una situación de vecindad o contigüidad entre dos unidades espaciales se podría expresar mediante valores de tipo 0-1. Entonces “si dos unidades espaciales tienen una frontera común de longitud no nula, se considera que son contiguas y se les asigna el valor 1” (Moreno y Vayá, 2000 [6]; citado por Yrigoyen, 2003, p. 53) [14].

Para las 9 celdas situadas en la *Ilustración 7*, es posible construir una matriz (9x9) correspondiente a estos datos (cuyas celdas están numeradas de izquierda a derecha y de arriba abajo) utilizando el criterio de contigüidad de la torre.

	1	2	3	
	4	5	6	
	7	8	9	

	1	2	3	4	5	6	7	8	9
1	0	1	0	1	0	0	0	0	0
2	1	0	1	0	1	0	0	0	0
3	0	1	0	0	0	1	0	0	0
4	1	0	0	0	1	0	1	0	0
5	0	1	0	1	0	1	0	1	0
6	0	0	1	0	1	0	0	0	1
7	0	0	0	1	0	0	0	1	0
8	0	0	0	0	1	0	1	0	1
9	0	0	0	0	0	1	0	1	0

Ilustración 7. Matriz de interacciones para nueve datos situados en una cuadrícula.
Fuente: (Anselin 1988 [1]; citado por Yrigoyen, 2003, p. 55)[14].

Matriz de pesos espaciales

Es una generalización de la matriz de interacciones espaciales. Esta matriz de pesos generalizada (W), es una matriz cuadrada no estocástica cuyos elementos w_{ij} reflejan la intensidad de la interdependencia existente entre cada par de regiones i y j . La matriz de pesos se remonta al concepto de contigüidad de primer orden donde w_{ij} es igual a 1 si las regiones i y j son físicamente adyacentes o a 0 en caso contrario (se asume por definición que $w_{ii} = 0$). (Moreno & Vayá, 2004, pp 86-87) [6].

Siguiendo el criterio de la proximidad física, diversos autores han propuesto definiciones de W basadas en la utilización de la distancia entre regiones (Moreno & Vayá, 2004) [6]. Por lo que se propone:

- Utilización de una matriz inversa de distancias al cuadrado, de manera que claramente la intensidad de la interdependencia entre dos regiones disminuye con la distancia que separa sus respectivos centros.
- Estandarizar la matriz W , dividiendo cada elemento w_{ij} por la suma total de la fila a la que pertenece, de forma que la suma de cada fila de la matriz estandarizada sea igual a la unidad.

$$\frac{W_{ij}}{\sum_{j \in J} W_{ij}} \quad (3)$$

siendo J el grupo de regiones relacionadas con i .

- A partir de la matriz de pesos espaciales, es posible obtener el denominado retardo espacial de una variable x sin más que premultiplicar la matriz W por la citada variable (Wx). Así, cada elemento de dicha variable retardada espacialmente se muestra como un promedio ponderado de los valores de la variable en el subgrupo de observaciones vecinas S_i , dado que $w_{ij} = 0$ para $j \notin S_i$

La interpretación de estos elementos es la siguiente (Yrigoyen, 2003, p. 57) [14]:

- $w_{ij} = 0$ indica ausencia de autocorrelación espacial entre las observaciones i, j (los elementos de la diagonal principal de la matriz de pesos serán igual a cero como en la matriz de interacciones).
- $w_{ij} \neq 0$ indica existencia de una interacción espacial entre las observaciones i, j , expresada como contigüidad de distancias, a través de centroides, dentro de una banda de distancia mínima

(considerándose contiguos dos puntos situados a una distancia inferior a la pre-fijada como mínima).

Estadísticos globales y locales de autocorrelación espacial

Por medio de estos estadísticos es posible contrastar la presencia o ausencia de un comportamiento de dependencia espacial a nivel univariante, para determinar el cumplimiento de la hipótesis de que la variable tiene una distribución aleatoria en el espacio o si por el contrario se evidencia una asociación significativa de valores similares o disimilares entre las regiones vecinas (Moreno & Vayá, 2004) [6]. Estos métodos pueden ser aplicados tanto global como localmente.

Tabla 5. Diferencias entre estadísticas globales o locales. (Stewart, Brunsdon & Charlton, 2002) [11]

Global	Local
Resumen de datos para toda la región	Desagregaciones a nivel local de estadísticas globales
Estadística para un solo valor	Estadísticas para múltiples valores
No mapeable	Mapeable
SIG - No amigable	SIG - Amigable
No espacial o espacial limitado	Espacial
Se enfatiza en similitudes en el espacio	Se enfatiza en diferencias en el espacio
Busca regularidades o leyes	Busca excepciones o puntos calientes locales
<i>Ejemplo</i>	<i>Ejemplo</i>
Regresión Clásica	Regresión geográficamente ponderada (GWR)

- I de Moran: Es uno de los estadísticos más comunes y usados para la medición del grado de autocorrelación espacial en una región, su forma de cálculo está dado por la siguiente ecuación (Moran, 1948- 1950; citado por Rogerson, 2001) [10]:

$$I = \frac{n \sum_i \sum_j W_{ij} (y_i - \bar{y})(y_j - \bar{y})}{(\sum_i \sum_j W_{ij}) \sum_i (y_i - \bar{y})^2} \quad (4)$$

Donde se tienen n regiones y W_{ij} es una medida de la proximidad espacial entre las regiones i y j . Este estadístico es interpretado como un coeficiente de correlación, donde los valores cercanos a +1 indican un fuerte patrón espacial donde existe una concentración de valores similares y valores cercanos a -1 que indican una fuerte correlación espacial negativa donde valores altos se encuentran localizados cerca de valores bajos. Finalmente, valores cercanos a 0 indican la ausencia de un patrón espacial. (Rogerson, 2001) [10].

- I de Moran Local: Es utilizado principalmente en la detección de clusters ya sea cuando la localización ya está especificada o cuando no se tiene una idea de la localización del cluster. Cuando el test global encuentre no significativa la desviación aleatoria, el test local permitirá descubrir puntos aislados de la incidencia de crecimiento; en el caso donde el test global indique un cluster de grado significativo, el estadístico local podrá ser utilizado para tomar una decisión:

a) el área de estudio es relativamente homogénea ó b) existen outliers espaciales que contribuyen a la significancia del estadístico global. (Rogerson, 2001) [10].

$$I_i = \frac{Z_i}{\sum_i^n Z_i^2 / n} \sum_{j \in J_i} W_{ij} Z_j \quad (5)$$

Donde se tiene que; Z_i es el valor correspondiente de la variable normalizada en la región i y J_i es el conjunto de regiones vecinas a i . La interpretación luego de su estandarización es $Z(I_i) > 0$ y significativo evidencia cluster de valores similares alrededor de la región i , mientras que $Z(I_i) < 0$ y significativo evidencia un cluster de valores disímiles alrededor de la región i . (Moreno & Vayá, 2004) [6]

- G(d) de Getis y Ord Local: Es utilizado para probar si una región particular i y sus vecinos conforman un cluster ya sea de valores altos o bajos, Ord & Getis (1995) hicieron uso del siguiente estadístico (Citado en Rogerson, 2001) [10]:

$$G_i^* = \frac{\sum_j^n W_{ij} - W_i^* \bar{x}}{s\{(n S_i^* W_i^*) / (n-1)\}^{1/2}} \quad (6)$$

Donde se tiene que:

$$W_i^* = W_i + w_{ii}; S_i^* = \sum_j W_{ij}^2; s^2 = \frac{1}{n-1} \sum_j (x_j - \bar{x})^2 \quad (7)$$

La interpretación de este estadístico de acuerdo a Moreno & Vayá [6] es la siguiente: $G_i^* > 0$ y significativo indica la presencia de valores similares y altos alrededor de la región i , mientras que $G_i^* < 0$ y significativo indica la presencia de valores similares pero bajos alrededor de la región i .

Análisis Confirmatorio de Datos Espaciales

Modelo econométrico espacial autorregresivo de primer orden (SAR)

Es una estructura general (simultánea) que se propone en forma matricial, derivada de la estructura del modelo lineal $y = X\beta + \epsilon$, que plantea la siguiente forma:

$$y = \rho W_y + u \quad (8)$$

Donde y y u son vectores de variables y el término de error, respectivamente, y W la matriz de pesos espaciales (Anselin, cap. 4, 1988 [1]; citado por Pérez, 2006, p. 145)[8]. El término ρW_y es la estructura autorregresiva en la variable dependiente que intenta explicar el modelo (el rezago espacial en la variable dependiente). Dentro de esta especificación se puede derivar una amplia gama de modelos, que pueden incluir la combinación de estructuras autorregresivas y procesos de medias móviles. La estructura que requiere la dependencia espacial para estimar en un caso de este tipo viene dada por la matriz de pesos W ,

que reduce el número de parámetros desconocidos a uno (el coeficiente de asociación espacial del proceso espacial autorregresivo o en el proceso espacial de medias móviles) (Pérez, 2006, pp. 145-146). [8]

Las consecuencias de ignorar autocorrelación espacial en un modelo de regresión cuando está presente, dependen de la forma de la hipótesis alterna. Existen dos casos:

Modelo mixto regresivo-autorregresivo espacial de primer orden (Spatial Lag Model)

También conocido como modelo de retardo espacial que tendrá la siguiente forma (rezago más otras variables explicativas):

$$y = \rho W_y + X\beta + u \rightarrow y = (1 - \rho)^{-1} X\beta + u \quad (9)$$

$$u \sim N(0, \sigma^2 I)$$

Donde y es un vector $N \times 1$, W_y es el retardo espacial de la variable dependiente y X una matriz de K variables exógenas, u es un término de perturbación ruido blanco, y ρ el parámetro espacial autorregresivo, con $H_0: \rho = 0$. Si tal forma de autocorrelación espacial se ignorará, los estimadores de MCO (mínimos cuadrados ordinarios) serían insesgados y la inferencia basada en el modelo de regresión clásica (un modelo sin W_y) sería incorrecta. (Pérez, 2006, p. 146). [8]

Modelo mixto regresivo espacial con perturbaciones espaciales autorregresivas de primer orden (Spatial Error Model)

Este modelo está relacionado con la autocorrelación espacial vinculada al término de error o autocorrelación residual (error espacial), cuya estructura plantea un proceso espacial para los términos de error:

$$y_{it} = X_{it}\beta_{it} + \varepsilon_{it}$$

$$\varepsilon_{it} = \lambda W_\varepsilon + u$$

$$y_{it} = X_{it}\beta_{it} + (1 - \lambda W)^{-1} u_{it} \quad (10)$$

$$u \sim N(0, \sigma^2 I)$$

Donde X_{it} es un vector renglón de observaciones para la unidad espacial i en el periodo t , β_{it} es un vector de parámetros específicos tiempo-espacio y ε_{it} es un término de error con las siguientes condiciones:

$$E[\varepsilon_{it}] = 0 \quad E[\varepsilon_{it}, \varepsilon_{js}] \neq 0 \quad (11)$$

La segunda condición posibilita la existencia de varias opciones de modelar las dependencias tiempo-espacio y patrones de heterogeneidad espacial si la varianza se considera constante, o la existencia de algún efecto espacial (Anselin, 1988. pp. 36-39 [1]; citado por Pérez, 2006, p. 147). [8]

Con $W\epsilon$ como el rezago espacial en el término de error, λ como el coeficiente autorregresivo y u como un término de error comportado (homocedástico e incorrelacionado). En este caso la hipótesis nula sería $H_0=\lambda=0$, y las consecuencias de ignorar dependencia espacial en el error son las mismas que para la heteroscedasticidad, los estimadores de MCO (mínimos cuadrados ordinarios) permanecen insesgados, pero no eficientes, al ignorar la correlación entre términos de error (Pérez, 2006, p. 147). [8]

Modelo Econométrico Espacial de Media Móvil SMA

Presenta una estructura de autocorrelación espacial tipo local, haciendo que las dependencias de una región se redujeron a las regiones más próximas a la misma en función de su estructura de conectividad definida inicialmente en la matriz W . Este tipo de modelo se desea modelizar el término de perturbación de forma diferente (autocorrelación residual), por lo que (Cliff y Ord, 1981; citado por Ramírez, 2012, pp. 21-22) [9], sugieren utilizar un proceso de medias móviles espaciales de orden 1:

$$\begin{aligned}y &= X\beta + \epsilon \\ \epsilon &= \theta W_1\mu + \mu \\ \mu &\sim N(0, \sigma^2 I)\end{aligned}\tag{12}$$

Donde θ es el coeficiente de medias móviles espacial y μ es el término de perturbación incorrelacionado. El esquema anterior se expande a órdenes superiores. La esperanza estimada por este modelo, es constante, haciendo que, si aumenta el tamaño de la muestra, está no cambiará su valor. Sin embargo, cuando se tratan este tipo de modelos, se tiene por naturaleza heteroscedasticidad en la ilustración (Ramírez, 2012, p. 22) [9].

Modelo Durbin espacial

El modelo robusto, debido a que los modelos espaciales autorregresivos (SAR) y modelos espaciales de error (SMA), son modelos particulares de este. Es un modelo que garantiza estimadores consistentes en el primer paso y es mucho más simple, es decir, proporciona estimaciones a los datos espacialmente correlacionados.

$$\begin{aligned}y &= \rho W_y + \alpha i_n + X\beta + WX\gamma + \epsilon \\ \epsilon &\sim N(0, \sigma^2 I)\end{aligned}\tag{13}$$

Donde y es un vector que guarda los valores de la variable dependiente, W es una matriz booleana de interacción espacial, ρ es un coeficiente que mide la dependencia de la variable dependiente y γ y α son vectores de parámetros asociado a las variables independientes y sus rezagos (Ramírez, 2012, p. 21)[9].

6. Descripción de la base de datos

La base de datos, pertenece a la plataforma de datos abiertos para el desarrollo internacional, AidData, cuya información es multitemporal entre los años 1997 - 2014 y corresponde a Nepal (Asia). Las observaciones corresponden a los 75 distritos del país, los cuales se encuentran agrupados en 14 zonas administrativas, guardados en un archivo de formas (shapefile). En la *Tabla 6*. se observan las variables que allí se consideran se describen a continuación:

Tabla 6. Descripción base de datos.

Identificador	Variable	Tipo	Descripción
ID_1	Identificador de la zona de desarrollo	Nominal	Número que identifica la zona de desarrollo a la que pertenece el distrito.
NAME_1	Nombre de la zona de desarrollo	Nominal	Nombre de la zona de desarrollo a la que pertenece el distrito.
ID_2	Identificador de la zona administrativa	Nominal	Número que identifica la zona administrativa a la que pertenece el distrito.
NAME_2	Nombre de la zona administrativa	Nominal	Nombre de la zona de desarrollo a la que pertenece el distrito.
ID_3	Identificador del distrito	Nominal	Número que identifica a cada distrito.
DISTRICT	Nombre del distrito	Nominal	Nombre correspondiente a cada distrito.
DEPECPROV	Privación en el abastecimiento económico.	Razón	Tasa de privación del abastecimiento económico por distrito.
POVINDEX	Índice de pobreza humana	Razón	Tasa de pobreza humana por distrito.
PCINC	Ingreso per cápita	Intervalo	Ingreso per cápita por distrito.
PCINCPPP	Ingreso per cápita PPP	Intervalo	Ingreso per cápita en base a la paridad del poder adquisitivo por distrito.
PCINCMPI	Ingreso per cápita, Rs. a precio del mercado	Intervalo	Ingreso per cápita en rupias a precio del mercado por distrito.
MALKIDS	Porcentaje de niños menores a 5 años con desnutrición	Razón	Porcentaje de niños menores a 5 años con desnutrición
LIF40	Personas que no se espera sobrevivan a los 40 años	Razón	Porcentaje de personas que no se espera sobrevivan a los 40 años
NOSAFH20	Población sin agua potable	Razón	Porcentaje de población sin agua potable en cada distrito.
POPULATION	Población	Intervalo	Cantidad de personas en cada distrito.
BOYG1_5	Número de niños en grados 1ro a 5to	Intervalo	Cantidad de niños inscritos en grados 1ro a 5to de primaria en cada distrito.
GIRLG1_5	Número de niñas en grados 1ro a 5to	Intervalo	Cantidad de niñas inscritas en grados de 1ro a 5to de primaria en cada distrito.
SCHOOLCNT	Número de escuelas	Intervalo	Número de escuelas en cada distrito.
AD_ILLIT	Analfabetismo en adultos	Razón	Tasa de analfabetismo en adultos medida en cada distrito.
LON	Longitud	Razón	Coordenada geográfica oeste del Distrito.
LAT	Latitud	Razón	Coordenada geográfica norte del Distrito.

7. Análisis exploratorio

7.1. Datos espaciales

Área de estudio

La región en la cual se determinará la presencia o ausencia de autocorrelación espacial respecto a la tasa de desabastecimiento económico, corresponde a Nepal (Asia); la ilustración 8 permite apreciar los 75 distritos del país agrupados en sus 14 zonas administrativas.

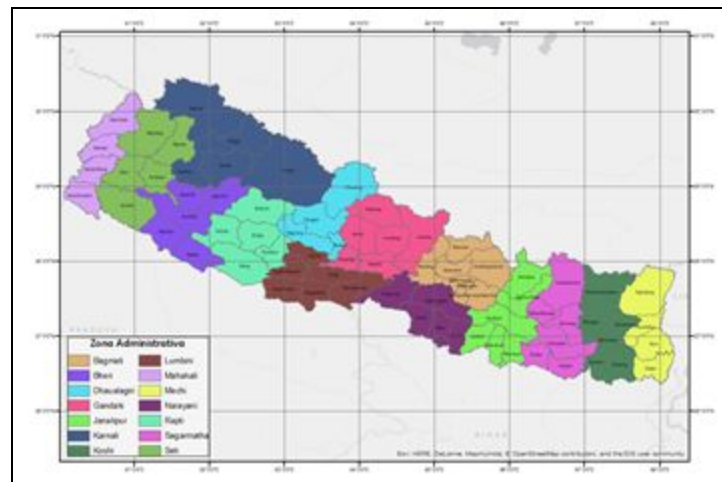
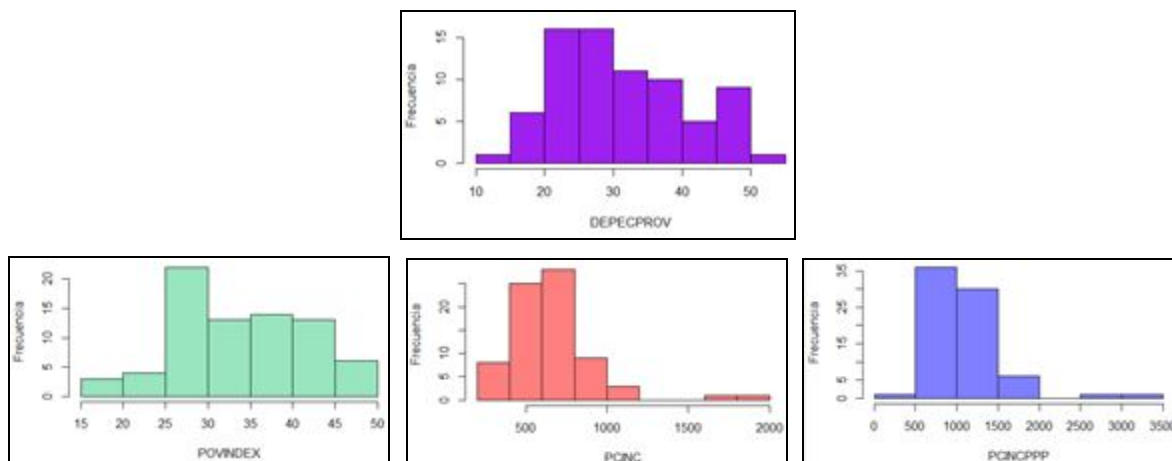


Ilustración 8. Mapa división política de Nepal. Software ArcGis 10.5.

Estadísticas Descriptivas



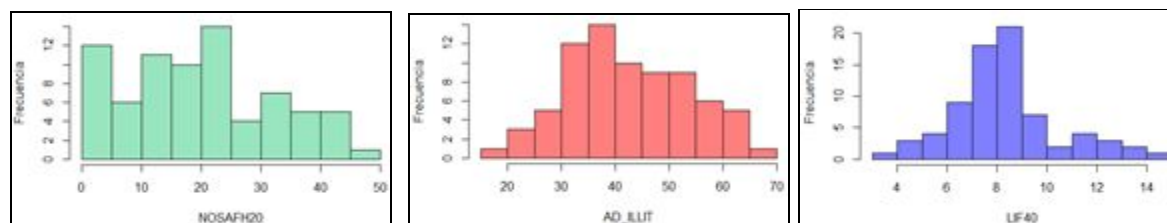


Ilustración 9. Histogramas Variables Independientes y Dependiente.

Observando la *Ilustración 9*, el histograma de la variable dependiente *DEPEC PROV* indica que la tasa de privación en el abastecimiento económico que más se presenta a lo largo de los 75 distritos se encuentra entre 20% y 30%, dato que confirman las estadísticas descriptivas de la variable (*Tabla 7*), con un promedio de 31.32%. Los datos están ligeramente sesgados hacia la derecha y presentan baja variabilidad. Aproximadamente 10 distritos exhiben una tasa del 52%, siendo ésta la más alta en todo el país.

El comportamiento del Ingreso per cápita *PCINC* y del Ingreso per cápita en base a la paridad del poder adquisitivo por distrito *PCINCP PP* es bastante similar, en su histograma ambos presentan sesgamiento hacia la derecha y valores atípicos altos, en promedio cada persona recibe un ingreso de 1079.83 rupias para subsistir, muy pocos habitantes tienen un ingreso mayor de 1500 rupias a lo largo del territorio.

Con respecto al índice de pobreza *POVINDEX*, los datos presentan una ligera asimetría hacia la izquierda, aproximadamente un 33.8% de población se encuentra en situación de pobreza en cada distrito, la variabilidad de este índice es baja, la mayoría de distritos presenta altos valores de pobreza.

El histograma del porcentaje de población sin agua potable en cada distrito *NOSAFH20* sugiere que los datos no siguen una tendencia específica, no hay una geometría definida en su forma y estructura, aproximadamente un 20% de la población de cada distrito no cuenta con agua potable.

El porcentaje promedio por distrito de personas que no se espera sobrevivan a los 40 años es del 8%, el valor más alto es del 14.48%, su desviación estándar sugiere que hay baja variabilidad y en el histograma se puede observar que aunque no hay un ajuste adecuado, podría seguir una distribución normal.

Tabla 7. Estadísticas Descriptivas Variables Independientes y Dependiente.

	Mediana	Media	Mínimo	Máximo	Desv. Estándar
DEPEC PROV	28.88	31.32	14.84	51.76	9.58
POVINDEX	33.58	33.80	16.50	49.26	7.66
PCINCP PP	1007.00	1079.83	487.00	3166.00	443.05
PCINCP	45022.00	48293.19	21772.00	141588.00	19814.94
MALKIDS	44.00	42.81	16.20	65.70	10.97
LIF40	8.14	8.28	3.31	14.48	2.18
NOSAFH20	19.26	19.83	2.14	48.12	12.27
AD_ILLIT	41.83	42.52	15.96	66.11	11.39

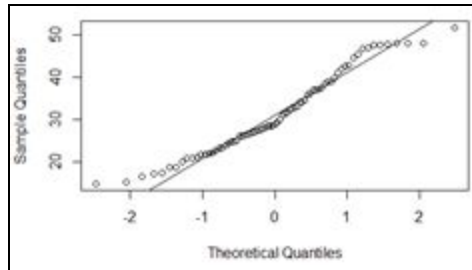


Ilustración 10. Q-Q plot variable respuesta

Corroborando la sugerencia del histograma, la variable dependiente no sigue una distribución normal, aunque se acerca a una, esto indicado por el hecho de que no todos los puntos caen sobre la recta en la *Ilustración 10*. También se puede observar presencia de valores extremos, tanto bajos como altos.

En búsqueda de tener una aproximación a las variables a usar en el modelo, se realizaron los siguientes diagramas de dispersión, ya que hay algunas variables que se calculan a partir de otras, y no sería adecuado la presencia de todas ellas dentro del modelado, puesto que generan problemas de multicolinealidad.

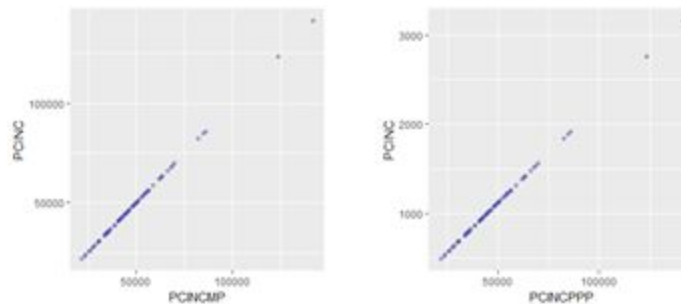


Ilustración 11. Diagramas de Dispersión PCINC vs PCINCMP y PCINC vs PCINCP

En este caso el *PCINCMP* y el *PCINCP* se calculan con respecto al *PCINC*, por lo tanto solo se usará la variable *PCINC*.

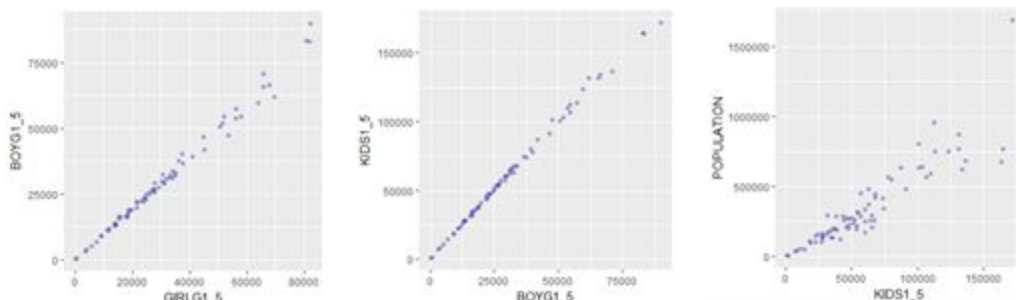


Ilustración 12. Diagramas de Dispersión entre las variables BOYG1_5, KIDS1_5, GIRLG1_5 y POPULATION

De la misma forma las variables *BOYG1.5*, *GIRLG1.5*, *KIDS1.5*, y *POPULATION* se encuentran relacionadas, por lo tanto, únicamente se tendrá en cuenta la variable *KIDS1.5* o la variable *POPULATION* dentro de la misma regresión.

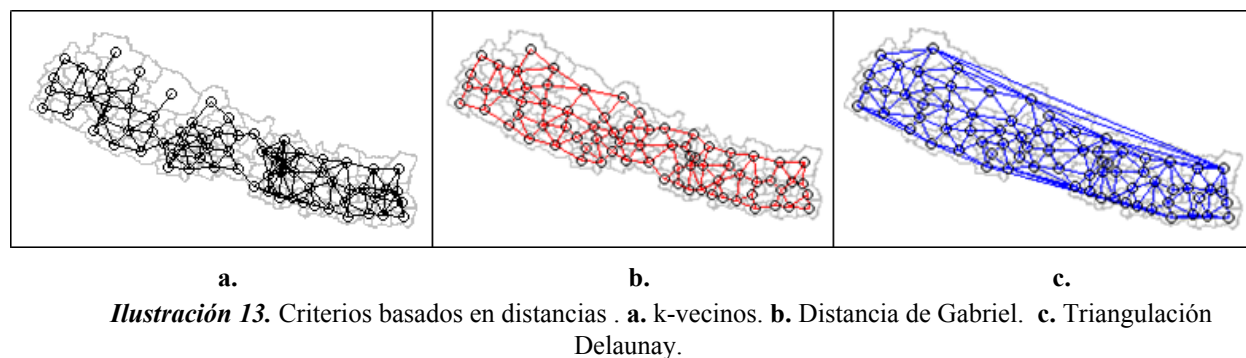
Matriz *W* de Pesos Espaciales

La elección de la matriz de pesos espaciales, se llevó a cabo mediante una revisión de las conexiones presentadas en las matrices generadas por los efectos reina, torre, k- vecinos más cercanos, distancia de Gabriel y Triangulación Delaunay . Las matrices para los efectos reina y torre fueron generadas mediante el software GeoDa, para 5 órdenes, ambos efectos arrojaron los mismos resultados, como se observa en la *Tabla 8*.

Tabla 8. Conexiones presentadas en los criterios de contigüidad evaluados. Software R.

Criterio		Links no nulos	%W no nulos	Promedio de links
Reina	Orden 1	366	6.51	4.88
	Orden 2	666	11.84	8.88
	Orden 3	794	14.11	10.58
	Orden 4	732	13.01	9.76
	Orden 5	616	19.95	8.21
Torre	Orden 1	366	6.51	4.88
	Orden 2	666	11.84	8.88
	Orden 3	794	14.11	10.58
	Orden 4	732	13.01	9.76
	Orden 5	616	19.95	8.21
K-vecinos	k=1	75	0	1
	k=2	150	2.66	2
	k=3	255	4	3
	k=4	300	5.33	4
	k=5	75	6.66	5
Distancia Gabriel		296	5.26	3.95
Triangulación Delaunay		424	7.54	5.65

Con respecto a los criterios basados en distancias, la triangulación Delaunay presentó mayor número de links no nulos, sin embargo, los órdenes de los efectos reina tienen un valor superior.



En general la matriz que presenta más conexiones es la de orden tres, efecto reina con un promedio de 10.58 links no nulos; sin embargo, al analizar el correlograma de contigüidad espacial (*Ilustración 14* y

Tabla 9), hay presencia de autocorrelación espacial para la variable DEPECPROV, en los órdenes del 2 al 5 (a un nivel de significancia del 5%).

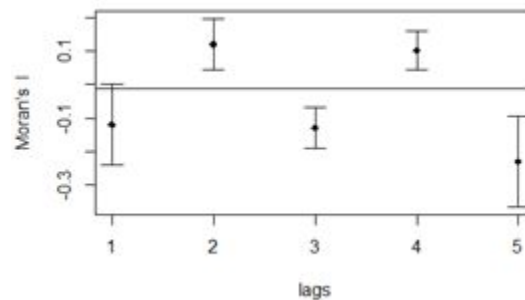


Ilustración 14. Correlograma espacial a partir de la matriz de orden 5 efecto reina.

Tabla 9. Correlograma espacial para el DEPECPROV a partir de la matriz de orden 5 efecto reina.

Orden	Estimado	E I	Varianza	Desv. Estándar	Pr(I) two sided
1	-0.12038033	-0.01351351	0.00363400	-1.7728	0.0762683
2	0.12004473	-0.01351351	0.00143478	3.5260	0.0004219
3	-0.12963367	-0.01351351	0.00091686	-3.8349	0.0001256
4	0.10169032	-0.01351351	0.00082292	4.0160	5.92e-05
5	-0.22924458	-0.01818182	0.00459917	-3.1122	0.0018568

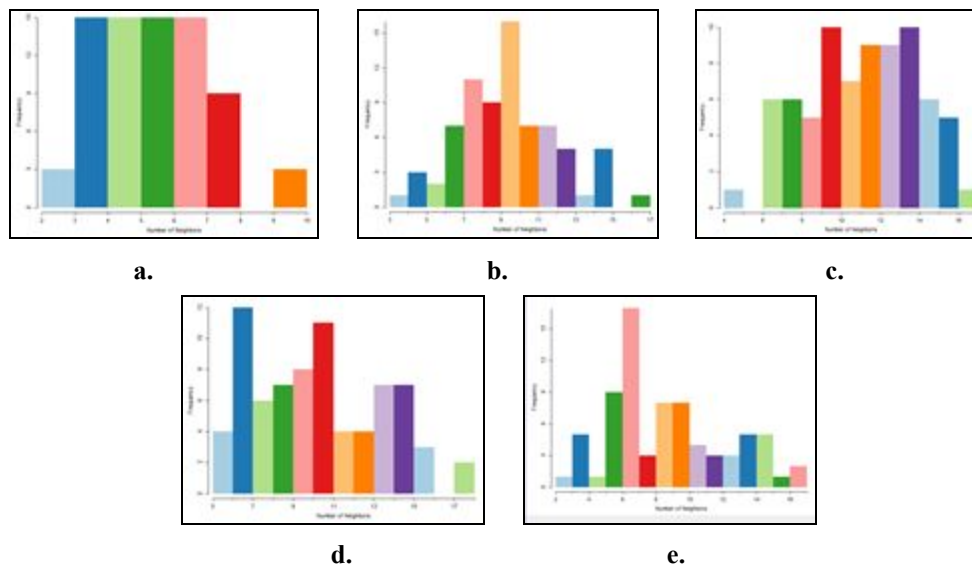


Ilustración 15. Histogramas de las matrices de contigüidad. **a.** Orden 1. **b.** Orden 2. **c.** Orden 3. **d.** Orden 4. **e.** Orden 5.

Unido al análisis de conexiones válidas entre distritos que se presenta en la Tabla 8 se generaron los histogramas de las diferentes matrices de pesos de orden 1 a 5 con el fin de tener otro criterio con el cual identificar la matriz de pesos más adecuada para el análisis de relación espacial. La Ilustración 15 nos permite ver con mayor claridad en términos de la normalidad de la variable que la matriz de contigüidad de orden 2, que hace referencia a las relaciones con los vecinos de aquellas regiones que son vecinos de primer orden, es la más óptima en el estudio a realizar. Considerando la cantidad de conexiones no se ve

afectada por el efecto con el cual se genere la matriz de pesos, y por ello finalmente se seleccionó la matriz de pesos de orden 2 y efecto reina.

Análisis Exploratorio de los Datos Espaciales

Inicialmente, se hace una primera observación del comportamiento que presenta nuestra variable endógena en la región de estudio mapeando la tasa de privación de aprovisionamiento económico registrada en cada distrito, el resultado se presenta a continuación:

“Variable Dependiente: Tasa de Privación de Aprovisionamiento Económico *DEPEC PROV*”

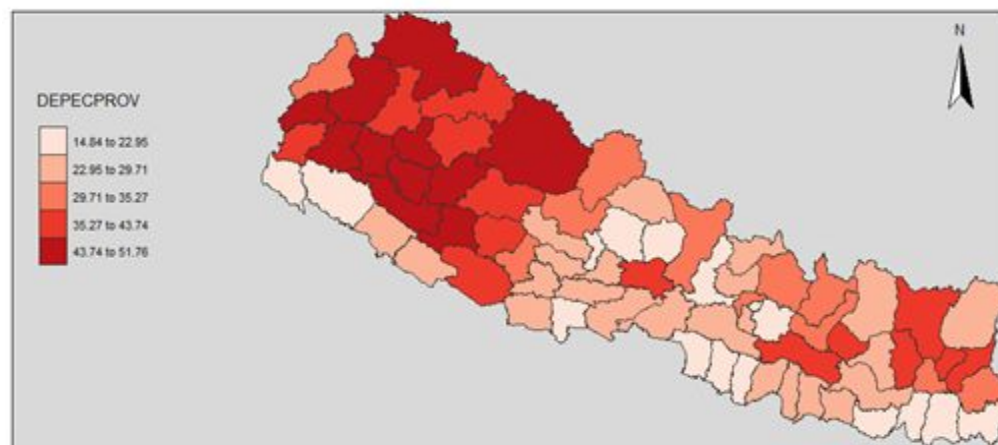


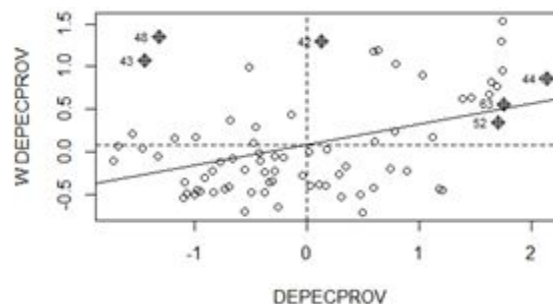
Ilustración 16. Distribución tasa de Privación de Aprovisionamiento Económico.
Software R.

La *Ilustración 16* nos permite hacer una primera identificación de posibles asociaciones zonales de valores similares en cuanto a la tasa de privación económica o clusters espaciales; en este primer acercamiento es posible inferir la existencia de un clúster de valores altos en la región del lejano y medio oeste, caso contrario la zona este y centro del país donde se presenta una distribución menos homogénea de los valores tomados por la variable de interés. Esta primera aproximación permite inferir que en aquellas regiones donde la tasa de privación económica es alta hay presencia de altos índices de pobreza y desigualdad; estos indicios serán corroborados por medio del análisis exploratorio que se presenta en este documento.

Análisis Exploratorio Global para la Variable Dependiente

Inicialmente se buscará determinar la presencia de autocorrelación a nivel global para la Tasa de privación económica y posteriormente se pasará a un análisis local; para ello se hará uso de estadísticos tales como el I de Moran y G de Getis y Ord junto con los gráficos generados en su aplicación.

Test I de Moran DEPECPROV	
Z(I) de Moran	4.7909
p-value	1.661e-06
I de Moran	0.24120
E[I]	-0.01351
V[I]	0.002826
Alternative Hypothesis: two.sided	



a)

b)

Ilustración 17. a) Test de Moran. b) Dispersograma de Moran.
Software R.

La aplicación del Test I de Moran de la variable dependiente (*DEPECPROV*) es significativa estadísticamente a un nivel de confianza del 95%, indicándonos la presencia de autocorrelación en la región respecto a la variable de interés. La *Ilustración 17* al mostrar una pendiente positiva en el dispersograma de Moran nos confirma que la autocorrelación existente es positiva lo cual nos permite inferir que hay asociaciones de valores similares; donde las regiones con tasas altas de privación del abastecimiento económico respecto al promedio se rodean de regiones con tasas altas de la misma variable y aquellas con tasas bajas respecto al promedio tienen a su vez vecinos con tasas bajas.

I de Moran Global para las Variables Independientes

En la *Tabla 10* se presenta un compilado de los resultados de la prueba global del I de Moran para las variables exógenas; en este caso únicamente las variables *SCHOOLCNT* (Cantidad de escuelas) y *NOSAFH20* (Población sin agua potable) no presenta autocorrelación espacial, en cuanto a las demás variables consideradas en relevante resaltar que todas evidencian por medio del test una autocorrelación positiva indicando la asociación de valores similares en la región de Nepal.

Tabla 10. I de Moran Global para las variables independientes. Software R.

	POVINDE	AD_ILLIT	NOSAFH20	PCINC	SCHOOLCNT	GIRLG1.5	BOYG1.5	MALKIDS
Z(I) de Moran	4.3779	2.6556	1.3707	2.9306	1.1641	2.9259	3.1794	4.2599
p-value	1.198e-05	0.007916	0.1705	0.003383	0.2444	0.003435	0.001476	2.045e-05
I de Moran	0.21926	0.127496483	0.05929	0.13303	0.04556	0.14089	0.15371	0.21252

I de Moran Bivariado

Para el análisis bivariado del I de Moran, se tomaron aquellas variables que se consideraba podrían tener una relación directa con la privación del abastecimiento económico por su naturaleza, estas variables fueron: *POVINDE* (Índice de Pobreza), *PCINC* (Ingreso per cápita), *NOSAFH20* (Población sin agua potable) y *AD_LLIT* (Analfabetismo en adultos).

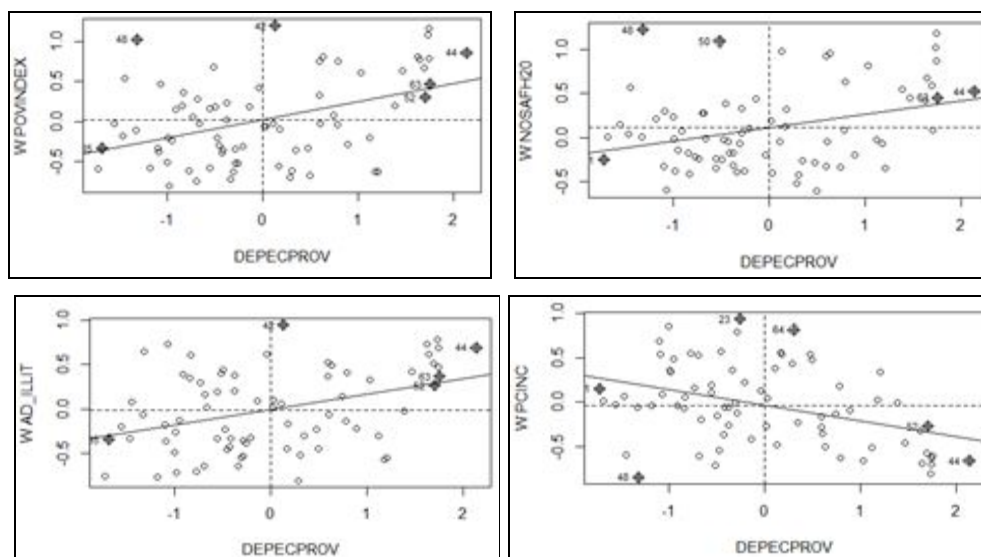


Ilustración 18. Dispersogramas Bivariados de Moran con variables Estandarizadas.

Test Bivariado I de Moran DEPEC PROV vs POVINDEX		Test Bivariado I de Moran DEPEC PROV vs PCINC	
Z(I) de Moran	4.4927	Z(I) de Moran	-3.0872
p-value	7.033e-06	p-value	0.00202
I de Moran	0.2253487	I de Moran	-0.17765
E[I]	-0.013513	E[I]	-0.01351
V[I]	0.002826	V[I]	0.002826
Alternative Hypothesis: two.sided		Alternative Hypothesis: two.sided	
Test Bivariado I de Moran DEPEC PROV vs NOSAFH20		Test Bivariado I de Moran DEPEC PROV vs AD _I LLIT	
Z(I) de Moran	3.1279	Z(I) de Moran	3.5165
p-value	0.001761	p-value	0.0004373
I de Moran	0.152787	I de Moran	0.1734480
E[I]	-0.01351	E[I]	-0.01351
V[I]	0.002826	V[I]	0.002826
Alternative Hypothesis: two.sided		Alternative Hypothesis: two.sided	

Ilustración 19. I de Moran Bivariado. Software R.

Los resultados de la aplicación del test bivariado del I de Moran se muestran de manera compilada en la *Ilustración 19*, tales resultados fueron confirmados por medio de la generación de los diferentes dispersogramas de Morán que a su vez son presentados en la *Ilustración 18*. A nivel general se puede decir que todas las relaciones analizadas presentan autocorrelación espacial en la región de estudio y su comportamiento estadístico representa lo siguiente:

- La privación de abastecimiento económico y el índice de pobreza, presentan una autocorrelación positiva lo cual de acuerdo al dispersograma de Moran generado (*Ilustración 18*) permite inferir que aquellas regiones con baja tasa de privación de abastecimiento económico se encuentran rodeadas de regiones con bajos índices de pobreza (ricos), mientras que aquellas regiones que

presentan alta tasa de privación de desabastecimiento económico se encuentran rodeadas de regiones con altos índices de pobreza (pobres).

- b. La privación del abastecimiento económico y el ingreso per cápita, presentan una autocorrelación negativa que de acuerdo al dispersograma de Moran permite inferir que aquellas regiones con baja tasa de privación de abastecimiento económico tienen vecinos con un alto ingreso per cápita, mientras que aquellas regiones con alta tasa de privación económica tienen vecinos con bajo ingreso per cápita.
- c. La privación del abastecimiento económico, la población sin agua potable y el analfabetismo en adultos, presentan una autocorrelación positiva lo cual de acuerdo los Dispersogramas de Moran generados permiten inferir que aquellas regiones con baja tasa de privación de abastecimiento económico tienen vecinos con bajos índices de población sin agua potable y de analfabetismo en adultos, mientras que aquellas regiones con alta tasa de privación de abastecimiento económico tienen vecinos con altos índices de población sin agua potable y analfabetismo en adultos. Estas dos relaciones en conjunto nos indican que, a mayor desabastecimiento económico, mayor es la desigualdad en la región.

Correlogramas Bivariados

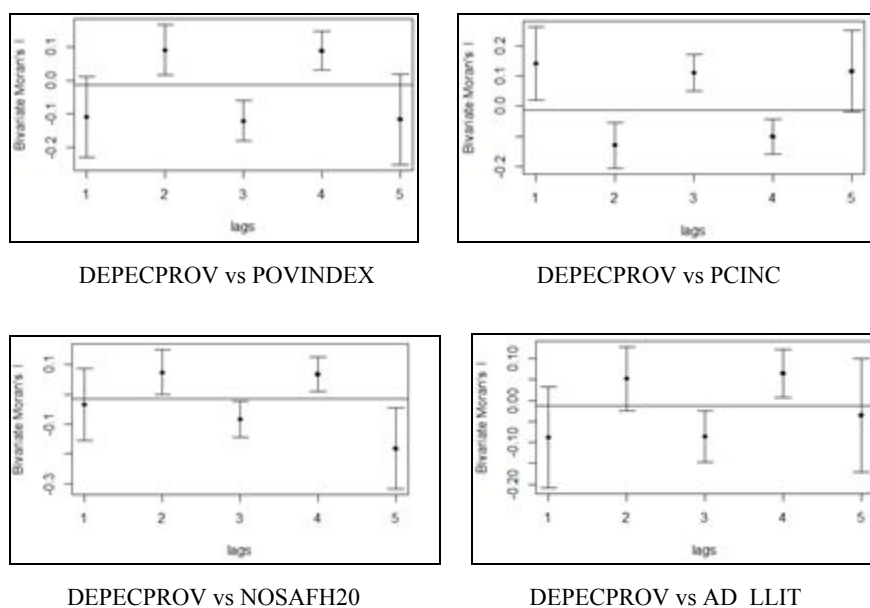


Ilustración 20. Correlogramas Bivariados de Moran.

Los correlogramas correspondientes al análisis anterior se generan para la determinación de los órdenes en los cuales las inferencias realizadas a partir de los resultados del test son válidas, como se puede observar en la *Ilustración 20* para los órdenes 1 y 5 es casi nula la autocorrelación espacial entre las variables caso contrario a los órdenes 2,3 y 4 donde es posible notar la presencia de autocorrelación espacial desde un análisis bivariado para la región.

Análisis Exploratorio Local

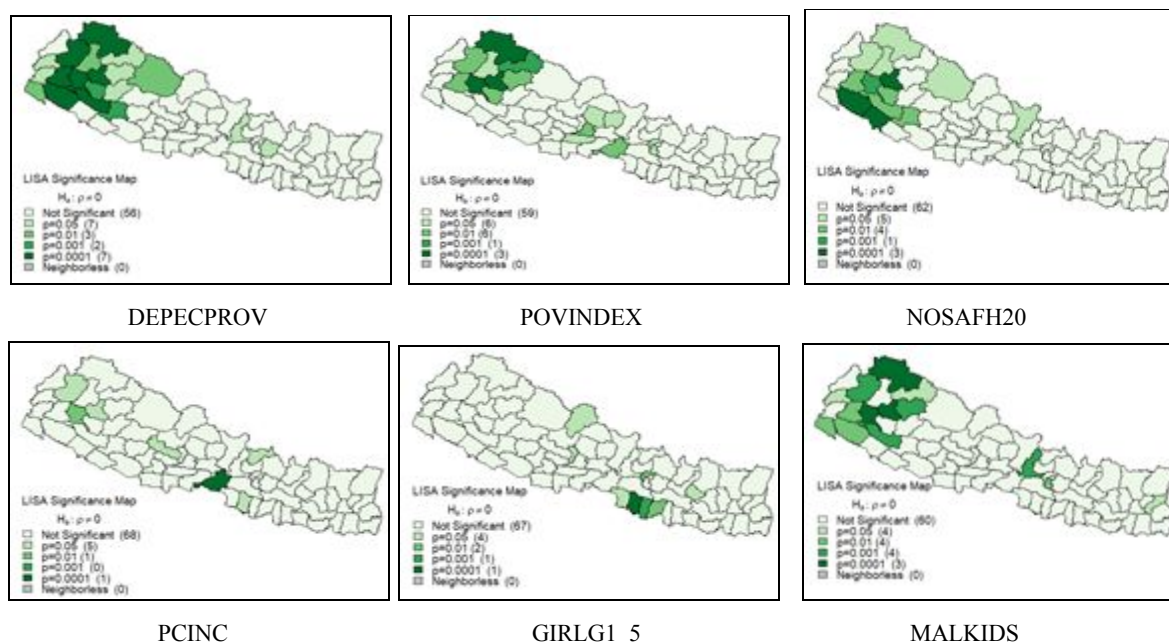


Ilustración 21. Mapas de significancia LISA.

El mapa de significancia del test LISA, nos permite identificar aquellas zonas de la región donde hay presencia de un clúster ya sea de valores similares o disimilares; la *Ilustración 21* nos permite confirmar la apreciación hecha a partir del mapeo de las tasas de privación del abastecimiento económico (*Ilustración 16*) pues efectivamente la existencia de clústeres se da principalmente en la zona del lejano y medio oeste de Nepal, sin embargo es importante resaltar que esta condición solo se presenta en la variable endógena (*DEPECPROV*) y las variables *POVINDEX* (índice de pobreza), *NOSAFH20* (población sin agua potable) y *MALKIDS* (porcentaje de niños menores a 5 años con desnutrición); por su parte las variables *PCINC* (ingreso per cápita) y *GIRLG1_5* (número de niñas en grados 1ro a 5to) presentan clusters significativos en la zona central de Nepal.

En cuanto a la variable endógena (privación del abastecimiento económico) el clúster más significativo es de valores similares (alto rodeado de altos) este se presenta en la zona del lejano y medio oeste, en dos distritos de la zona central y dos del lejano oeste se evidencia una asociación de bajo rodeado de bajos y bajo rodeado de altos respectivamente. En las variables exógenas o independientes de acuerdo a la *Ilustración 22* se evidencian los siguientes comportamientos:

- Índice de pobreza (*POVINDEX*): El clúster más significativo es de alto rodeado de altos y se presenta en distritos del lejano y medio oeste, otros clústeres de más baja significancia se presentan en el centro de Nepal y corresponde a valores bajos rodeados de bajos.
- Población sin agua potable (*NOSAFH20*): El clúster más significativo es de bajo rodeado de altos y se presenta en dos distritos uno del lejano y otro del medio oeste de Nepal, adicionalmente

se presentan clústeres con menor significancia de alto rodeado de altos y alto rodeado de bajos en las zonas del lejano - medio oeste y central respectivamente.

- c. Ingreso per cápita (*PCINC*): No presenta una gran cantidad de asociaciones, el clúster más significativo se presenta en la zona centro de Nepal y es de alto rodeado de altos, hay presencia de clústeres no tan significativos y dispersos entre la zona centro y del lejano oeste donde se presentan alto rodeado de altos y bajo rodeado de bajos respectivamente.
- d. Número de niñas en grados 1ro a 5to (*GIRLG1_5*): No tiene gran presencia de asociaciones en la región de Nepal, sin embargo, se presenta un clúster con gran significancia en la zona central este sigue un comportamiento de alto rodeado de altos.
- e. Porcentaje de niños menores a 5 años con desnutrición (*MALKIDS*): El clúster más significativo sigue un comportamiento de alto rodeado de altos y se presenta en provincias del lejano y medio oeste, en esta misma zona se evidencia otro clúster de menor significancia de bajo rodeado de altos y finalmente un tercer clúster de alto rodeado de bajos en la zona este del país.

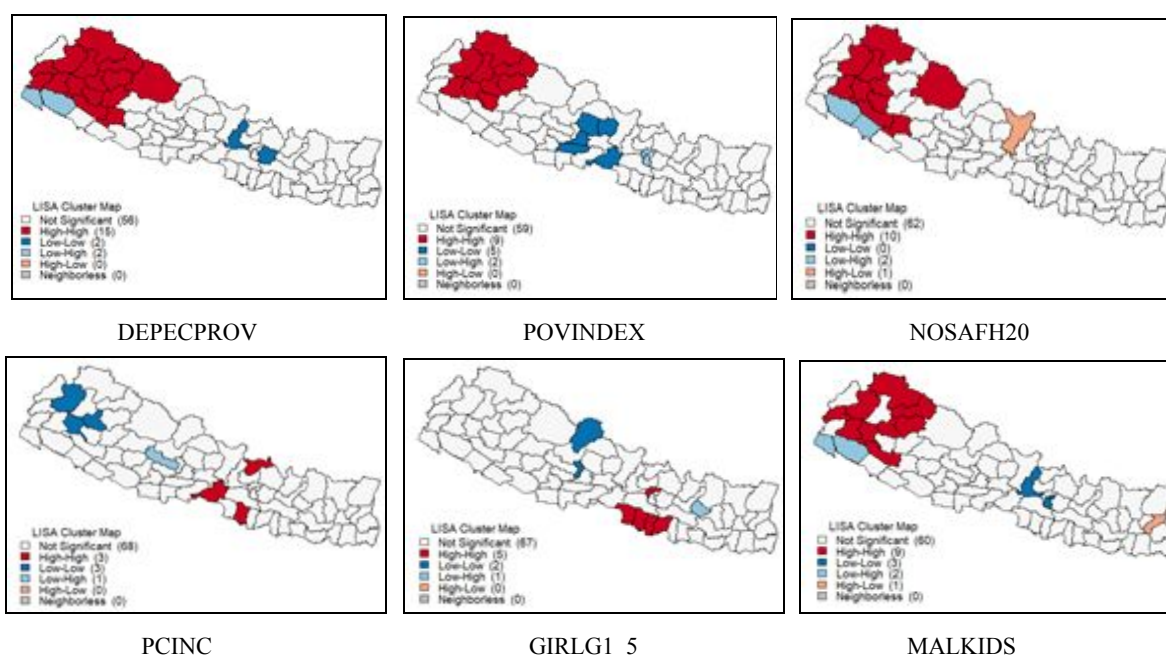


Ilustración 22. I de Moran Local, clusters locales.

El contraste local I de Moran, indica la presencia o ausencia de autocorrelación espacial alrededor de cada uno de los distritos. En la *Tabla 11* se exhiben los distritos en donde hay presencia de autocorrelación espacial, en éstos, hay concentración de valores similares significativos de la tasa de privación en el abastecimiento económico *DEPECPROV*, es decir hay presencia de autocorrelación positiva. Dichos distritos son en el orden respectivo de la tabla: Dhading, Kavrepalanchok, Rautahat, Baitadi, Dadeldhura, Achham, Bajhang, Bajura, Doti, Dailekh, Jajarkot, Surkhet, Dolpa, Humla, Jumla, Kalikot, Mugu y Salyan.

Tabla 11. I de Moran Local para la variable dependiente.

ID	I_i	$E[I_i]$	$Var[I_i]$	$Z[I_i]$	$Pr(z > 0)$
6	0.60	-0.01	0.07	2.33	0.01
8	0.53	-0.01	0.06	2.27	0.01
22	0.44	-0.01	0.08	1.63	0.05
40	0.95	-0.01	0.15	2.47	0.01
41	0.83	-0.01	0.15	2.17	0.02
44	1.86	-0.01	0.09	6.37	0.00
45	1.68	-0.01	0.15	4.34	0.00
46	0.94	-0.01	0.10	3.04	0.00
47	1.36	-0.01	0.10	4.40	0.00
51	1.10	-0.01	0.10	3.56	0.00
52	0.59	-0.01	0.07	2.28	0.01
53	1.31	-0.01	0.07	5.02	0.00
54	0.87	-0.01	0.08	3.17	0.00
55	2.28	-0.01	0.13	6.38	0.00
56	0.71	-0.01	0.10	2.33	0.01
57	2.72	-0.01	0.13	7.60	0.00
58	0.77	-0.01	0.11	2.35	0.01
63	0.99	-0.01	0.07	3.80	0.00

Tabla 12. G de Getis y Ord Local variable dependiente.

ID	G_i^*
6	-2.11429
8	-2.11659
22	-1.70403
40	1.69859
41	2.67240
44	3.10087
45	2.52386
46	2.92361
47	2.72468
51	2.23290
52	1.39791
53	3.01696
54	2.29943
55	3.73599
56	3.79467
57	4.39859
58	3.57457
63	2.22878

El contraste local G de Getis y Ord, sugiere que hay presencia de asociación espacial alrededor de cada distrito. En los distritos de Dhading, Kavrepalanchok y Rautahat se presentan valores bajos de tasa de privación económica rodeados de bajos valores de la misma variable; en los distritos Dadeldhura, Achham, Bajhang, Bajura, Doti, Dailekh, Surkhet, Dolpa, Humla, Jumla, Kalikot, Mugu y Salyan, se presentan valores altos de la tasa de privación en el abastecimiento rodeados de altos valores de la misma variable.

8. Análisis Confirmatorio

Luego de confirmar la presencia de autocorrelación espacial, se realiza el modelado de la variable privación en el abastecimiento económico mediante modelos espaciales.

8.1. Modelo Clásico

Inicialmente se realiza la estimación del modelo de corte transversal por Mínimos Cuadrados Ordinarios, de acuerdo al Análisis Exploratorio de Datos Espaciales, las variables seleccionadas para explicar la privación en el abastecimiento económico son el índice de pobreza humana, el ingreso per cápita, las personas que no se espera que sobrevivan a los 40 años, la población sin agua potable, la población y el analfabetismo en adultos.

El mejor modelo clásico obtenido se presenta en la *Tabla 13*, todas sus variables son significativas estadísticamente al 95% de confiabilidad, excepto el intercepto; su R^2 ajustado indica que aproximadamente el 95% de la variabilidad de la privación de abastecimiento económico se encuentra explicada por las variables independientes después de castigar el ingreso de las nuevas variables; según el $p - value$ de la prueba F, el modelo ajusta los datos al 1% de significancia.

Tabla 13. Regresión lineal clásica obtenida. Software R

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.3690	1.7263	0.79	0.4305
POVINDE	2.1197	0.1424	14.89	0.0000
PCINC	-0.0022	0.0011	-2.06	0.0431
LIF40	0.7933	0.1274	6.23	0.0000
NOSAFH20	0.1852	0.0341	5.42	0.0000
AD_ILLIT	-1.1870	0.0898	-13.21	0.0000
Observations	75			
R ²	0.958			
Adjusted R ²	0.955			
Residual Std. Error	2.026 (df = 69)			
F Statistic	316.909*** (df = 5; 69)			
p-value	<2.2e-16			

Con respecto a los supuestos, según la prueba de Kolmogorov- Smirnov, los residuos del modelo son normales y no se encuentran correlacionados (Test de Durbin- Watson), y hay presencia de homocedasticidad (según el test de Breusch- Pagan).

Tabla 14. Pruebas para verificación de supuestos.

Test	Estadístico	p-value
Shapiro - Wilk	0.9886	0.7434
Kolmogorov - Smirnov	0.0643	0.8960
Breusch - Pagan	0.00642	0.93611
Durbin-Watson	2.1175	0.6876

A pesar de los resultados obtenidos, se procede a utilizar modelos espaciales, con los cuales, se puede modelar adecuadamente la dependencia espacial de la variable respuesta.

8.2. Pruebas de Multiplicadores de Lagrange

Para tener una primera aproximación en el tipo de modelo espacial a elegir, se realizaron las pruebas de multiplicadores de Lagrange.

El estadístico del LMlag (prueba de multiplicadores de Lagrange donde la estructura espacial está rezagada en las variables exógenas) fue el único que resultó ser significativo estadísticamente a un nivel de significancia del 5%; lo que sugiere que un modelo Spatial Lag es el más adecuado.

Tabla 15. Pruebas de los Multiplicadores de Lagrange

Prueba	LMlerr	LMlag	RLMerr	RLMlag	SARMA
Valor	0.68843	4.0693	0.10826	3.4891	4.1776
P-value	0.4067	0.04367	0.7421	0.06177	0.1238

A pesar de lo anterior se realizó un modelo Spatial Error, uno Spatial Lag, uno SARAR (Autorregresivo en el error combinado) y un modelo Durbin - Spatial, para corroborar los resultados anteriores.

El modelo seleccionado es de tipo Spatial Lag, debido a que fue el único cuyo parámetro autorregresivo fue significativo estadísticamente y se ajustó adecuadamente a las observaciones.

A continuación se presentan los resultados obtenidos en su construcción mediante la librería *spdep* del software R.

8.3. Modelo Spatial Error (Espacial Autorregresivo en el Error)

Al realizar un modelo Spatial Error, el parámetro autorregresivo λ no es significativo estadísticamente y este modelo quedaría descartado, ya que no ajusta correctamente la realidad de los datos. Esto corroborado por el estadístico de Wald, cuya probabilidad es mayor de 0.05, indicando que el λ no influye a la hora de determinar la variable dependiente.

Tabla 16. Mejor Modelo Spatial Error. Software R

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-13.5867	6.3827	-2.13	0.0333
POVINDEX	2.0442	0.1118	18.28	0.0000
PCINC	-0.0024	0.0008	-3.02	0.0025
LIF40	0.4164	0.1148	3.63	0.0003
NOSAFH20	0.1869	0.0276	6.77	0.0000
SCHOOLCNT	-0.0055	0.0012	-4.61	0.0000
AD_ILLIT	-1.1628	0.0689	-16.87	0.0000
lat	0.7944	0.2280	3.48	0.0005

```
Lambda: -0.3268, LR test value: 1.3404, p-value: 0.24696
Asymptotic standard error: 0.24421
z-value: -1.3382, p-value: 0.18084
Wald statistic: 1.7907, p-value: 0.18084

Log likelihood: -140.5189 for error model
ML residual variance (sigma squared): 2.4537, (sigma: 1.5664)
Nagelkerke pseudo-R-squared: 0.97256
Number of observations: 75
Number of parameters estimated: 10
AIC: 301.04, (AIC for lm: 300.38)
```

A pesar de que el modelo no es el indicado, su pseudo-R-cuadrado es bastante alto y todas las variables son significativas estadísticamente al 5% de significancia, lo que puede sugerir que son adecuadas para los siguientes modelos a realizar.

8.4. Modelo Spatial Lag (Autorregresivo de rezago espacial)

Utilizando las variables que presentan autocorrelación espacial, se corrió el siguiente modelo:

Tabla 17. Modelo Spatial Lag. Software R

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-102.9939	28.0541	-3.67	0.0002
POVINDEX	2.1684	0.1164	18.63	0.0000
PCINC	-0.0035	0.0010	-3.68	0.0002
LIF40	0.5627	0.1228	4.58	0.0000
NOSAFH20	0.1377	0.0292	4.72	0.0000
AD_ILLIT	-1.2202	0.0735	-16.59	0.0000
lat	2.6875	0.5341	5.03	0.0000
lon	0.4358	0.1909	2.28	0.0225

```
Rho: -0.12802, LR test value: 4.4321, p-value: 0.035269
Approximate (numerical Hessian) standard error: 0.059858
z-value: -2.1387, p-value: 0.032459
Wald statistic: 4.5741, p-value: 0.032459

Log likelihood: -143.9344 for lag model
ML residual variance (sigma squared): 2.7143, (sigma: 1.6475)
Nagelkerke pseudo-R-squared: 0.96994
Number of observations: 75
Number of parameters estimated: 10
AIC: 307.87, (AIC for lm: 310.3)
```

Como se observa, todas las variables son significativas estadísticamente a un nivel de confianza del 95%, el parámetro autorregresivo ρ es significativo, lo que implica que esta puede ser la forma apropiada para el ajuste del modelo, ya que se tiene en cuenta la autocorrelación espacial; siendo este aspecto apoyado por el estadístico de Wald, indicando que ρ influye a la hora de determinar la variable respuesta. Su valor de pseudo-R- cuadrado es de 96% aproximadamente, lo que muestra que el modelo tiene un buen ajuste.

8.4.1. Análisis de Supuestos

Normalidad

Para la evaluación del supuesto de normalidad, se usará la siguiente prueba de hipótesis:

$$H_0 : \text{Los residuos siguen una distribución normal}$$

$$H_1 : \text{Los residuos no siguen una distribución normal}$$

Los resultados de las pruebas de normalidad, arrojaron que a un nivel de confiabilidad del 95%, no se rechaza H_0 y por lo tanto los residuos siguen una distribución normal.

Tabla 18. Test de normalidad para los residuos del modelo Spatial Lag

Test	Estadístico	p-value
Shapiro - Wilk	0.9808	0.3149
Kolmogorov - Smirnov	0.0898	0.5505

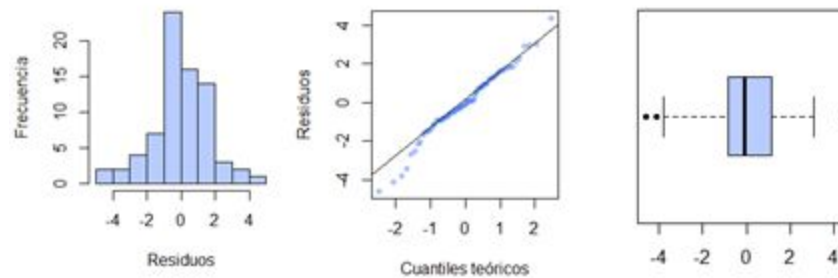


Ilustración 23. Gráficos de detección de normalidad de los residuos

Según los gráficos qq-plot y el boxplot de los residuos se confirma que los residuos siguen una distribución normal, la media en el boxplot es cero, y se señalan observaciones outliers, que debido a que se trabaja con datos espaciales, no es posible removerlas. Del mismo modo, la forma del histograma tiende a ser similar a una campana de Gauss, indicando normalidad.

Heterocedasticidad

Para la evaluación del supuesto de heteroscedasticidad, se usará el siguiente criterio:

H_0 = Hay presencia de homocedasticidad

H_1 = Hay presencia de heterocedasticidad

Tabla 19. Prueba de Heterocedasticidad para los residuos del modelo Spatial Lag

Test	Estadístico	p-value
Breusch - Pagan	8.5716	0.2849

Según los resultados de los test aplicados, el $p - value > \alpha$, ($\alpha = 0.05$), no se rechaza H_0 , es decir, los residuos son homocedásticos.

La *Ilustración 24* corrobora la afirmación acerca de que los residuos son homocedásticos, ya que éstos forman una banda horizontal aproximada alrededor de la línea 0 que indica la homogeneidad de la varianza del error.

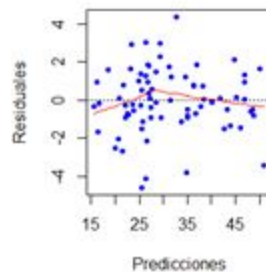


Ilustración 24. Gráficos de detección de homocedasticidad Residuales vs Predicciones.

Autocorrelación Espacial

Para la evaluación del supuesto de autocorrelación, se usará el siguiente criterio:

$$H_0 = \text{Ausencia de autocorrelación espacial en los residuos}$$

$$H_1 = \text{Presencia de autocorrelación espacial en los residuos}$$

Tabla 20. Prueba de autocorrelación para los residuos del modelo Spatial Lag

Test I de Moran	
Residuos	
Z(I) de Moran	-0.65346
p-value	0.5135
I de Moran	-0.04789
E[I]	-0.01351
V[I]	0.00276
Alternative Hypothesis: two.sided	

Según el test aplicado, el $p - value > \alpha$, ($\alpha = 0.05$), no se rechaza H_0 , es decir, hay ausencia de autocorrelación espacial en los residuos.

En general el modelo cumple con los supuestos especificados, y aunque el valor del AIC no es el menor (comparado con el modelo Spatial Error), se eligió como el que mejor modela la realidad, teniendo en cuenta el análisis anterior.

El modelo Spatial Lag estimado de la Privación en el Abastecimiento Económico, es el siguiente:

$$DEPECPROV = -102.9939 - 0.12802W DEPECPROV + 2.1684 POVINDE X - 0.0035 PCINC + 0.5627 LIF40 + 0.1377 NOSAFH20 - 1.2202ADILLIT + 2.6875lat + 0.4358lon + \mu$$

$$\mu \sim N(0, \sigma^2)$$

Donde $-0.12802WDEPECPROV$ es el valor de la Privación en el Abastecimiento Económico espacialmente retardada a partir de una matriz W de interacciones espaciales obtenida como contigüidad efecto reina de segundo orden entre los distritos de Nepal.

8.4.2. Análisis de Impactos

Para el análisis de impactos, se altera una variable independiente (predictor), en un distrito y luego se usará el modelo espacial ajustado para predecir los valores (*DEPECPROV*) en otro lugar del mapa.

¿Qué pasaría con la Privación en el Abastecimiento Económico en Nepal si el Índice de Pobreza aumenta del 25.23% al 51% en el distrito de Palpa ?

Cambiando el valor de *POVINDEX* de 25.23% a 51% y haciendo uso del modelo tipo Spatial Lag seleccionado (*Tabla 17*), se obtuvieron unas nuevas predicciones. Calculando las diferencias entre las predicciones originales y la mencionadas anteriormente, y ordenando su valor absoluto de mayor a menor, los 10 distritos en donde el impacto fue mayor, se presentan a continuación:

Tabla 21. Diferencia entre las predicciones originales y las obtenidas cambiando el valor de *POVINDEX*

ID	Distrito	Dif. Predicciones
74	Palpa	56.17
73	Nawalparasi	-2.12
29	Taplejung	-1.51
47	Dailekh	-1.44
39	Kanchanpur	1.38
24	Sunsari	1.29
48	Jajarkot	-1.25
68	Syangja	-1.19
69	Tanahu	-1.15
44	Kailali	1.05

El efecto directo del cambio realizado fue en el distrito de Palpa, como era de esperarse (fue el distrito en donde se modificó el valor de *POVINDEX*). Sin embargo, cuando se observan los 10 distritos más afectados, también se aprecia que distritos tan alejados, como Taplejung experimentaron un cambio sustancial en su tasa de privación en el abastecimiento económico; es importante apreciar que estos cambios son tanto positivos como negativos, como se presenta en la *Ilustración 25*, en algunos distritos disminuyó la tasa de *DEPECPROV* (color rojo), en algunos aumentó (color verde) y en algunos no hubo efecto.

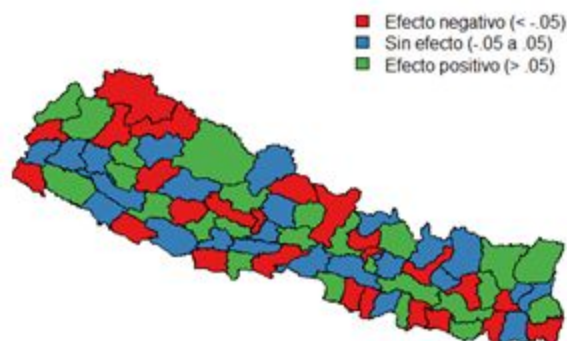


Ilustración 25. Efectos de un cambio en el distrito de Palpa (establecer *POVINDEX* = 51%) en los valores pronosticados en un modelo de rezago espacial.

Por último, mediante la función *impacts()*, se calculan los impactos, teniendo en cuenta que ésta dice el efecto directo (local), indirecto (spill-over) y total de un cambio de unidad en cada una de las variables predictoras. Los cambios informados por los impactos son el impacto promedio global

Tabla 22. Medidas de Impactos

	Directo	Indirecto	Total
POVINDEX	2.17	-0.25	1.92227
PCINC	-0.00	0.00	-0.00314
LIF40	0.56	-0.06	0.49884
NOSAFH20	0.14	-0.02	0.12209
AD_ILLIT	-1.22	0.14	-1.08169

Según los resultados arrojados en el cálculo de los impactos (mediante la función *impacts()*), un aumento del 1% en el índice de pobreza (*POVINDEX*) conduce a un aumento del 1.92% (en promedio) en la tasa de privación de abastecimiento económico (*DEPECPROV*).

Por otro lado, el aumento de 100 rupias nepalí en el ingreso per cápita (*PCINC*) disminuye la tasa de privación en el abastecimiento económico en -0.315% (en promedio) aproximadamente.

Con respecto a las personas que no se espera sobrevivan a los 40 años (*LIF40*), un aumento del 1% en esta variable, conduce a un incremento en el *DEPECPROV* del 0.49% (en promedio), de modo similar que con el porcentaje de población sin agua potable en cada distrito (*NOSAFH20*), cuyo aumento del 1% incrementa el *DEPECPROV* en 0.12% (en promedio).

Un incremento del 1% en la tasa de analfabetismo en adultos medida en cada distrito (*AD_ILLIT*), disminuye el *DEPECPROV* en 1.08% (en promedio).

9. Conclusiones

Finalizado el análisis exploratorio de los datos espaciales, se puede establecer la presencia de autocorrelación espacial positiva de la tasa de privación de abastecimiento económico en Nepal y además se evidencian relaciones con las demás variables de la base de datos; por ejemplo, la presencia de una tasa alta de privación de abastecimiento económico implica altos índices de pobreza, un alto número de población infantil sin acceso a educación y con desnutrición lo que se podría resumir en “desigualdad social” y una baja calidad de vida en la población.

Por medio de la aplicación de los test de Moran globales para cada una de las variables regresoras, es posible afirmar que las únicas dos variables que no presentaba un autocorrelación espacial en Nepal son *SCHOOLCNT* (Cantidad de escuelas) y *NOSAFH20* (Población sin agua potable); esto para una matriz de pesos espaciales efecto reina, orden 2 lo que hace referencia a la relación de una región *i* y los vecinos de sus vecinos por frontera o de primer orden.

En cuanto a la determinación de la mejor forma funcional para la definición del modelo, se seleccionó uno tipo Spatial Lag, cuyo parámetro autorregresivo espacial, resultó ser significativo. Los residuos del modelo, cumplen los supuestos y según la interpretación de los impactos en el modelo, se sugiere que las variables regresoras explican de manera adecuada la realidad.

10. Referencias

- [1] Anselin, L. (1988). *Spatial econometrics: methods and models*. Kluwer Academic Publishers.
- [2] Anselin, L. (2001). *Spatial econometrics. 2001 Summer course*. University of Illinois, Urbana-Champaign. Recuperado de <http://www.spacestat.com>.
- [3] Barros, C., Gupta, R. (2017). *Development, poverty and inequality: a spatial analysis of South African provinces*. The Journal of Developing Areas. Vol 51. No. 1. Recuperado de <https://muse.jhu.edu/article/654392>.
- [4] Cabrera, P. Murphy, T. Kienberger, S. Blaschke, T. (2015). *A multi-criteria spatial deprivation index to support health inequality analyses*. International Journal of Health Geographics. Recuperado de <https://ij-healthgeographics.biomedcentral.com/articles/10.1186/s12942-015-0004-x>.
- [5] Guo, Y., Chang, S., Sha, F., Yip, P. (2018). *Poverty concentration in an affluent city: Geographic variation and correlates of neighborhood poverty rates in Hong Kong*. PLoS ONE. 13(2) Recuperado de <https://doi.org/10.1371/journal.pone.0190566>.
- [6] Moreno, R., Vayá, E. (2004). *Econometría Espacial: nuevas técnicas para el análisis regional. Una aplicación a las regiones europeas*. En *Efectos Espaciales "Autocorrelación espacial"* (pp. 86-87). En. Investigaciones Regionales. Recuperado de <https://dialnet.unirioja.es/descarga/articulo/2124394.pdf>
- [7] Pérez, G. (2008). *Dimensión espacial de la pobreza en Colombia*. Digitalizado por la Biblioteca Luis Ángel Arango del Banco de la República. Geografía económica y análisis espacial en Colombia. (pág. 175 - 222) Recuperado de <http://repositorio.banrep.gov.co/handle/20.500.12134/466>.
- [8] Pérez, J. (2006). *Econometría espacial y ciencia regional*. En *Tipología de Modelos Espaciales* (pp. 145-147). Facultad de Economía, México. Recuperado de <http://www.redalyc.org/pdf/601/60125804.pdf>
- [9] Ramírez, Y. (2012). *Análisis de autocorrelación espacial en el desarrollo sostenible respecto al componente social, ambiental, institucional y económico de los municipios de Corpochivor para el año 2010*. En 5.2.8 *Análisis Confirmatorio de Datos Espaciales* (pp. 21-22). Bogotá, Universidad Distrital Francisco José de Caldas, Facultad de Ingeniería, Tesis de Grado.
- [10] Rogerson, P (2001). *Statistical Methods Geography*. En *Spatial Patterns*, (pp. 167, 173-175). London: Sage Publications LTD. Recuperado de <https://uk.sagepub.com/en-gb/eur/statistical-methods-for-geography/book243249>
- [11] Stewart, A., Brunsdon, C., Charlton, M. (2002). *Geographically Weighted Regression: the analysis of spatially varying relationships*. En *Local Statistics and Local Models for Spatial Data*, (p. 6). University of Newcastle, UK: John Wiley & Sons, LTD. Recuperado de <http://www.madrid.org/bvirtual/BVCM005618>.

- [12] Turkey, J.W. (1977). *Exploratory Data Analysis*. Reading: Addison-Wesley.
- [13] Vilalta, J. (2005). Cómo enseñar autocorrelación Espacial. *En definición de conceptos: la autocorrelación, la dependencia y la heterogeneidad espaciales*, (pp. 3-4). Tecnológico de Monterrey, Campus Ciudad de México Recuperado de <http://www.madrid.org/bvirtual/BVCM005618>.
- [14] Yrigoyen, C. (2003). Econometría Espacial Aplicada a la Predicción-Extrapolación de Datos Microterritoriales. En *Capítulo II. Econometría de Datos Espaciales*, (pp. 8-10). En *Capítulo III. Análisis exploratorio de Datos Espaciales* (pp. 28, 53-58). Madrid, España: Consejería de Economía e Innovación Tecnológica. Tesis Doctoral. Recuperado de <http://www.madrid.org/bvirtual/BVCM005618>.