

Análisis de los hurtos en Colombia durante el año 2017 mediante los modelos de regresión lineal múltiple y la regresión ponderada geográficamente

Analysis of thefts in Colombia during 2017 using multiple linear regression models and geographically weighted regression

Análise dos furtos na Colômbia durante o ano 2017 mediante os modelos de regressão linear múltipla e a regressão ponderada geograficamente

Fecha de recepción: 2018/10/23 | Fecha de evaluación: 2019/05/13 | Fecha de aprobación: 2019/05/24

Nelson Ricardo López Herrera

MSc Geographical Information Science & Systems
Especialista en Estadística - Universidad Industrial de Santander
Bucaramanga, Colombia
nr.lopez@gmail.com

Marlon Augusto Aceros Bueno

Economista
Especialista en Estadística - Universidad Industrial de Santander
Bucaramanga, Colombia
marlonaceros@gmail.com

Mariana Luzardo Briceño

Doctora en Estadística
Docente de tiempo completo
Universidad Pontificia Bolivariana Seccional Bucaramanga
Bucaramanga, Colombia
mariana.luzardo@upb.edu.co

Para citar este artículo / To reference this article / Para citar este artigo: López-Herrera, N., Aceros, M. & Luzardo, M. (2019). Análisis de los hurtos en Colombia durante el año 2017 mediante los modelos de regresión lineal múltiple y la regresión ponderada geográficamente. *Revista Criminalidad*, 61(3): 141-163

Resumen

Según información procedente del observatorio del delito de la Policía Nacional de Colombia, los hurtos a personas y de celulares han presentado una tendencia al alza desde el año 2003 (Norza, Peñalosa y Rodríguez, 2017). Esta tendencia motivó la realización del presente estudio para analizar la relación entre los factores socioeconómicos y el hurto en los diferentes municipios de Colombia durante el año 2017, mediante el uso de modelos de regresión lineal múltiple y regresión geográficamente ponderada utilizando fuentes de información secundaria segregada a nivel municipal. Se constató que las variables matriculados en instituciones de

educación superior por cada mil personas, presupuesto per cápita asignado por el sistema general de participaciones y la categoría del municipio explican el 69,5% de la variabilidad del logaritmo del hurto a personas y de celulares en 532 municipios mediante un modelo de regresión lineal múltiple estimado globalmente y el 50,16% utilizando el modelo de regresión ponderada geográficamente omitiendo en este último la categoría del municipio. En este modelo hubo ligeras variaciones en los coeficientes a nivel municipal, lo que refleja que la heterogeneidad social y económica tiene efectos en los indicadores de hurto a nivel nacional.

Palabras clave

Factores de la criminalidad, hurto, estadísticas criminales, medición de la criminalidad (fuente: Tesauro de Política Criminal Latinoamericana - ILANUD) Estadística, GWR, OLS (autor).

Abstract

According to information from the crime observatory of the National Police of Colombia, thefts from people and of cell phones have shown an upward trend since 2003 (Norza, Peñalosa and Rodríguez, 2017). This trend motivated the carrying out of the current study to analyze the relationship between socioeconomic factors and theft in the different municipalities of Colombia during 2017, through the use of multiple linear regression models and geographically weighted regression using secondary information sources segregated to municipal level. It was validated that variables enrolled in higher education institutions per thousand

people, budget per capita allocated by the general system of participations and the category of the municipality account for 69,5% of the variability of the logarithm of theft from individuals and of cellphones in 532 municipalities using a globally estimated multiple linear regression model and 50,16% using the geographically weighted regression model omitting in the latter the category of the municipality. In this model there were slight variations in the coefficients at the municipal level, reflecting that the social and economic heterogeneity has effects on indicators of theft nationwide.

Keywords

Factors of crime, theft, criminal statistics, crime measuring (source: Tesouro de Política Criminal Latinoamericana [Thesaurus of Latin American Criminal Policy] - ILANUD) Statistics, GWR, OLS (author).

Resumo

Segundo informação proveniente do observatório do delito da Polícia Nacional da Colômbia, os furtos a pessoas e de celulares têm apresentado uma tendência de aumento desde o ano 2003 (Norza, Peñalosa y Rodríguez, 2017). Esta tendência motivou a realização do presente estudo para analisar a relação entre os fatores socioeconômicos e o furto em os diferentes municípios da Colômbia durante o ano 2017, mediante o uso de modelos de regressão linear múltipla e regressão geograficamente ponderada utilizando fontes de informação secundária segregada a nível municipal. Constatou-se que as variáveis matriculadas em instituições

de ensino superior por cada mil pessoas, orçamento per capita atribuído pelo sistema geral de participações e a categoria do município explicam o 69,5% da variabilidade do logaritmo do furto a pessoas e de celulares em 532 municípios mediante um modelo de regressão linear múltipla estimado globalmente e o 50,16% utilizando o modelo de regressão ponderada geograficamente omitindo neste último a categoria do município. Neste modelo houve ligeiras variações nos coeficientes a nível municipal, o que reflete que a heterogeneidade social e econômica tem efeitos nos indicadores de furto a nível nacional.

Palavras-chave

Fatores da criminalidade, furto, estatísticas criminais, medição da criminalidade (fonte: Tesouro de Política Criminal Latinoamericana - ILANUD) Estatística, GWR, OLS (autor).

Introducción

Al ser el hurto un delito de tipo monetario es razonable pensar que puede estar relacionado con indicadores económicos que reflejen las condiciones de inversión social y calidad de vida en los municipios. Este hecho ha sido estudiado por Sandoval y Barón (2008), quienes argumentan que el crimen depende de las oportunidades a las que se tenga acceso en la sociedad y que incentiven el desarrollo personal y económico de los individuos que les permita satisfacer sus necesidades básicas.

Según los datos reportados por el observatorio del delito de la Policía Nacional de Colombia, desde el año 2003, en Colombia se ha presentado una tendencia al incremento en el hurto a personas (Norza et al., 2017). Dicho comportamiento impulsa la realización de investigaciones que ayuden a entender el contexto en el que ocurren los delitos y los principales factores socioeconómico-demográficos asociados con el crimen para poder entender las causas directas e indirectas que propician la actividad delictiva.

Teniendo en cuenta la heterogeneidad del territorio nacional (Castillo, 2007) —diferencias de tipo cultural, social y económico—, se espera que existan variaciones significativas en las estadísticas criminales entre municipios que no puedan ser explicadas adecuadamente por un modelo de regresión lineal múltiple global. Por esta razón en este trabajo se utilizará la técnica de regresión ponderada geográficamente (GWR por sus siglas en inglés) (Brunsdon, Fotheringham y Charlton, 1996), en la que se pueden identificar variaciones espaciales entre el hurto y las variables explicativas a partir de la variación de los coeficientes del modelo. Dichos modelos pueden contribuir a la planificación de políticas públicas, adecuadas al contexto de cada municipio, y que pueden ser promovidas por las diferentes entidades gubernamentales que podrían tener impacto en el aumento o disminución de los índices de criminalidad.

En este trabajo se compara un modelo de regresión lineal múltiple y un modelo GWR mediante el uso tanto del coeficiente de determinación ajustado (R_a^2) como de los Criterios de Información de Akaike (AIC) y el criterio de información bayesiano (BIC).

Marco teórico

Relación entre el crimen y las condiciones socioeconómicas

Sandoval y Barón (2008, p. 105) analizaron la situación actual de los diferentes temas relacionados con la criminalidad, en su estudio afirman que los delitos

... tienen origen en aspectos de los individuos, los cuales dependiendo de las posibilidades que tengan en la sociedad pueden llegar a ser productivos para ésta o pueden tener incentivos a ser criminales y delinquentes dadas las limitaciones que encuentren a progresar personal y laboralmente; así mismo, a partir de las características propias de las zonas geográficas la delincuencia tiende a ser una forma de vida para las personas que aprovechan los recursos que brindan las regiones, lo cual puede llegar a propagarse hacia las zonas aledañas y convertirse en aros de criminalidad que a su vez pueden generar violencia producto del aprovechamiento que desean tener las redes delictivas de los dividendos generados por las actividades ilícitas.

Si bien la conducta delictiva puede asociarse a aspectos individuales como la capacidad de autocontrol (Gottfredson y Hirschi, 1990), el consumo de drogas y alcohol (Baron, 2003; Eisner, 2002; Harrison, 1992), traumas en la infancia asociados a violencia intrafamiliar, abuso sexual (Chen, Thrane, Whitbeck, Johnson y Hoyt, 2007), entre otros, tanto como a aspectos sociales del entorno en el cual las personas viven, es razonable pensar que distintos tipos de delitos se asocian a distintos tipos de factores que pueden promover la conducta criminal. Los delitos violentos podrían estar relacionados con aspectos emocionales, como una conducta insegura que lleve a celos —en el caso del feminicidio o la violencia de género—, o sociales, como una idiosincrasia machista o intolerante (Chávez, 2018), aunque en el caso del sicariato el móvil es la búsqueda de lucro por parte del autor material.

Por su parte, en los delitos contra la propiedad es de esperar que la principal motivación del delincuente sea obtener una recompensa monetaria por su actuar con el objetivo de satisfacer sus necesidades económicas. Diferentes estudios realizados mediante métodos estadísticos de tipo espacial, no espacial y temporal han sido realizados con la finalidad de entender las relaciones existentes entre los hurtos y aspectos sociales y económicos. Fajnzylber, Lederman y Loayza (2002) han hallado correlaciones positivas entre las tasas de hurtos y variables que indican pobreza y desigualdad (e. g., índice de Gini, índice de calidad de vida, entre otros) al analizar los datos de 37 países; este mismo estudio evidenció que existen relaciones inversas entre variables relacionadas con el desarrollo económico (e. g., porcentaje de variación del PIB en términos reales). Además, se encontró que el nivel de ingresos promedio, el nivel educativo promedio de la población adulta y el grado de urbanización del país no presentan relaciones significativas con este delito.

Otro indicador de calidad de vida es el índice de necesidades básicas insatisfechas (NBI), el cual ha sido utilizado para identificar y caracterizar la población según las carencias críticas (Mancero y Feres, 2001); por lo tanto, se considera un indicador útil para estimar las condiciones de vida de las personas en la sociedad y por ende se presume que en una población con alto NBI los individuos sean más propensos a cometer hurto en el afán de mejorar sus condiciones de vida. A su vez, el presupuesto destinado a los municipios por parte del Sistema General de Participaciones puede ser una variable relacionada con el número de robos por cuanto se supone que a mayor sea el presupuesto del municipio, mejor será la calidad de vida de las personas y por ende estarán motivadas a la delincuencia. Este mismo efecto atenuante se espera de variables como la cantidad de personas con acceso a la educación.

Regresión lineal múltiple

En Charlton y Fotheringham (2009) la regresión es descrita como un conjunto amplio de métodos para modelar las relaciones entre una variable dependiente (Y) y un conjunto de variables independientes (X). En la forma más sencilla, un modelo de regresión lineal múltiple poblacional puede expresarse de la siguiente forma:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_m x_{mi} + \varepsilon_i \quad \text{para } i = 1 \dots n \quad (1)$$

En la ecuación anterior y_i y x_i son las variables respuesta e independientes (predictoras), respectivamente, medidas en la localidad i ; ε_i representa el error aleatorio; β_0 y β_{mi} son los coeficientes estimados por el método de Mínimos Cuadrados Ordinarios (MCO) para las m variables de forma que se minimice el valor de $\sum_{i=1}^n (y_i - \hat{y}_i)^2$ en las n observaciones. \hat{y}_i es el valor predicho por el modelo para la i -ésima observación dado el i -ésimo valor de X .

Las predicciones son obtenidas mediante una combinación lineal de las variables independientes. El término $(y_i - \hat{y}_i)$ es denominado residual para la i -ésima observación. Para que el modelo sea válido estadísticamente se debe garantizar que los residuales no presenten autocorrelación serial, se distribuyan de manera normal con media cero y varianza constante, la no existencia de multicolinealidad entre las variables predictoras y que cambios en las variables predictoras tengan efecto en las predicciones, es decir, los parámetros del modelo $\hat{\beta}$ deben ser significativamente diferentes de cero (Charlton, Fotheringham y Brunson, 2006).

Según Brunson *et al.* (1996), el estimador MCO se expresa como

$$\hat{\beta} = (X^T X)^{-1} X^T y \quad (2)$$

Siendo $\hat{\beta}$ el vector de los coeficientes estimados, X es la matriz de diseño, la cual contiene los valores de las variables independientes y una columna con valores 1; y es el vector de los valores observados y $(X^T X)^{-1}$ es la inversa de la matriz de varianza-covarianza (Charlton *et al.*, 2006).

Para evaluar la autocorrelación de los residuales se utilizó la prueba de Durbin-Watson (1950, 1951, 1971), que tiene como hipótesis nula la ausencia de autocorrelación.

En los modelos de regresión lineal múltiple se asume que las relaciones entre las variables son las mismas en la zona de estudio, es decir, hay homogeneidad espacial. Sin embargo, dicha suposición puede no cumplirse ya que los procesos que generan los datos pueden variar con el espacio –heterogeneidad espacial (Charlton y Fotheringham, 2009)–. Estas relaciones pueden ser el resultado de diferencias intrínsecas en el espacio debido a diferencias culturales o a contextos políticos distintos (Charlton *et al.*, 2006).

Cuando la regresión lineal es utilizada con datos espaciales, la localización en la que fueron tomados los datos no se tiene en cuenta en el proceso de modelado (Brunson, Fotheringham y Charlton, 1998) ya que se asume que el proceso examinado es constante en el espacio y por lo tanto se utiliza un solo modelo (Charlton *et al.*, 2006).

Regresión geográficamente ponderada

La ley de la geografía indica que “*Todo se relaciona con todo, pero las cosas cercanas se relacionan más que las lejanas*” (Tobler, 1970, p. 236), lo que implica que en procesos espaciales puede no cumplirse la suposición de independencia entre las observaciones asumidas en los modelos de regresión tradicional. Esto es conocido como correlación espacial y puede estar presente en diferentes fenómenos e influir en las estimaciones de los parámetros en un modelo de regresión lineal tradicional ya que, si los residuales del modelo tienen una estructura espacial, la estimación de los parámetros será ineficiente, por lo cual los errores estándar de los parámetros serán muy grandes, produciendo que parámetros potencialmente significativos parezcan no serlo (Gutiérrez-Puebla, García-Palomares y Daniel-Cardozo, 2012).

La regresión geográficamente ponderada fue desarrollada por Brunson *et al.* (1996) como un método para explorar la no estacionariedad espacial,

es decir, la condición en la cual un modelo “global” no puede explicar adecuadamente las relaciones entre los conjuntos de variables. Esta situación puede darse en modelos con información georreferenciada en los cuales los coeficientes de la regresión varían con el espacio.

La GWR extiende el método de regresión tradicional, permitiendo estimar parámetros locales en vez de globales (Charlton *et al.*, 2006). De este modo, el modelo puede ser expresado con la siguiente ecuación:

$$y_i = a_0 + (u_i, v_i) + \sum_k \beta_k(u_i, v_i) X_{ik} + \varepsilon_i \quad k = 1, 2, \dots, n \quad (3)$$

Donde (u_i, v_i) denota las coordenadas del i -ésimo punto en el espacio y $\beta_k(u_i, v_i)$ son los coeficientes en el punto i (Brunsdon *et al.*, 1996). Por lo tanto, se permite que exista una superficie continua de valores de parámetros, y las mediciones de esta superficie se toman en ciertos puntos para indicar la variabilidad espacial de la superficie (Charlton *et al.*, 2006). De este modo, el modelo de regresión lineal múltiple es un caso particular de la GWR cuando se asume que los parámetros son constantes.

En la GWR los valores próximos al punto son ponderados de acuerdo con su proximidad, es decir, los más próximos influyen más que los lejanos. Algebraicamente, el estimador se puede expresar como

$$a(u_i, v_i) = (X'W(u_i, v_i)X)^{-1}X'W(u_i, v_i)y \quad (4)$$

Donde $W(u_i, v_i)$ es una matriz $n \times n$ compuesta por ceros a excepción de los elementos en la diagonal principal, los cuales representan las ponderaciones geográficas de los datos observados en el punto i (Charlton *et al.*, 2006).

$$w(u_i, v_i) = \begin{pmatrix} w_{i1} & 0 & 0 & \dots & 0 \\ 0 & w_{i2} & 0 & \dots & 0 \\ 0 & 0 & w_{i3} & \dots & 0 \\ \cdot & \cdot & \cdot & \dots & \cdot \\ 0 & 0 & 0 & \dots & w_{in} \end{pmatrix}$$

Donde w_{in} corresponde al peso del dato en el punto n en la calibración del modelo alrededor del punto i . Las ponderaciones cambian con i , lo cual establece la diferencia con respecto a la técnica de mínimos cuadrados ponderados, donde la matriz de ponderación es constante (Charlton *et al.*, 2006).

En el caso de un modelo de regresión tradicional OLS (i. e. global), la matriz de ponderaciones está implícitamente definida como:

$$w(u_i, v_i) = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \cdot & \cdot & \cdot & \dots & \cdot \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix}$$

lo cual quiere decir que cada dato tiene un peso equivalente a 1, que se interpreta como que no hay variaciones espaciales en los parámetros estimados.

Con la técnica GWR no solo se obtienen coeficientes locales, también se obtienen versiones locales de los diagnósticos utilizados en la regresión tradicional como la bondad de ajuste (R^2).

Para evaluar la presencia de autocorrelación espacial de datos medidos en áreas teniendo en cuenta la localización y su valor, se utiliza comúnmente el Estadístico I de Moran (1950) definido como (Environmental Systems Research Institute [ESRI], 2016):

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n W_{ij} Z_i Z_j}{\sum_{i=1}^n \sum_{j=1}^n W_{ij} \sum_{i=1}^n Z_i^2} \quad (5)$$

Donde Z_i y Z_j son las desviaciones del valor en la posición i y en la posición j , respectivamente, con respecto al valor de la media (i.e. $Z_i = x_i - \bar{X}$). W_{ij} es la matriz de vecindad, la cual define si las posiciones i, j son vecinas ($W_{ij}=1$) o no ($W_{ij}=0$), n es el número total de datos.

Si los datos tienden a un *cluster* espacial, el índice de Morán será positivo; por el contrario, si los valores altos tienden a estar cerca de valores bajos, el índice será negativo. Además, si los productos positivos entre las desviaciones igualan a los negativos, el índice tenderá a cero. Por lo tanto, el índice de Morán tendrá valores comprendidos entre -1 y 1. Un valor positivo del índice -1 de Morán- indica una tendencia hacia la acumulación, mientras que un valor negativo indica una tendencia hacia la dispersión (ESRI, 2016).

El I de Morán también viene acompañado de un Z-Score y un p-valor que responde a la hipótesis nula de aleatoriedad en los datos, según su resultado se puede concluir lo expresado en la Tabla 1.

Tabla 1.
Interpretación del I de Morán según el p-valor y el Z-score

p-valor no significativo	Hay evidencia estadísticamente significativa de que los valores se distribuyen debido a un proceso aleatorio.
p-valor significativo	Hay evidencia significativa de que los valores no se distribuyen debido a un proceso aleatorio. Cuando el Z-score es positivo, la distribución espacial de los datos está más agrupada que lo esperado en un proceso aleatorio. En caso contrario, indica que los valores están más dispersos que lo esperado en un proceso aleatorio.

La presencia de correlación espacial en los residuales del modelo GWR puede ser señal de un modelo mal definido por la ausencia de una variable clave en el modelo (ESRI, 2016).

Antecedentes

La formulación del modelo GWR ha sido ampliamente utilizada en investigaciones relacionadas con variables georreferenciadas, donde pueden existir cambios considerables en el comportamiento en los coeficientes de las variables analizadas en la región de estudio, como salubridad pública (Fotheringham, Charlton y Brunson, 1998; Wheeler y Tiefelsdorf, 2005), mercado inmobiliario (Brunson, Fotheringham y Charlton, 1999), en exploración geoquímica (Zhao, Wang y Cheng, 2014), en estudios de movilidad (Gutiérrez-Puebla et al., 2012), en estudios del rendimiento académico (Fotheringham, Charlton y Brunson, 2001), estudios biológico-ecológicos (Ortiz-Yusty, Páez y Zapata, 2013), entre otros.

Así mismo, se han realizado estudios que evalúan la relación entre el hurto a motocicletas y factores geográficos en el Valle del Cauca, Colombia, mediante técnicas de econometría espacial, los cuales demuestran que este delito presenta una distribución regional no aleatoria en el espacio (Ramírez, López, Castro y Ocampo, 2017).

Los aspectos relacionados con la criminalidad también han sido analizados mediante el uso de la GWR, tal es el caso del estudio del efecto del terrorismo en el crecimiento económico de Turquía (Öcal y Yildirim, 2010) y el estudio de los crímenes violentos en la ciudad de Portland, Estados Unidos, realizado por Cahill y Mulligan (2007), que reportan la utilidad de

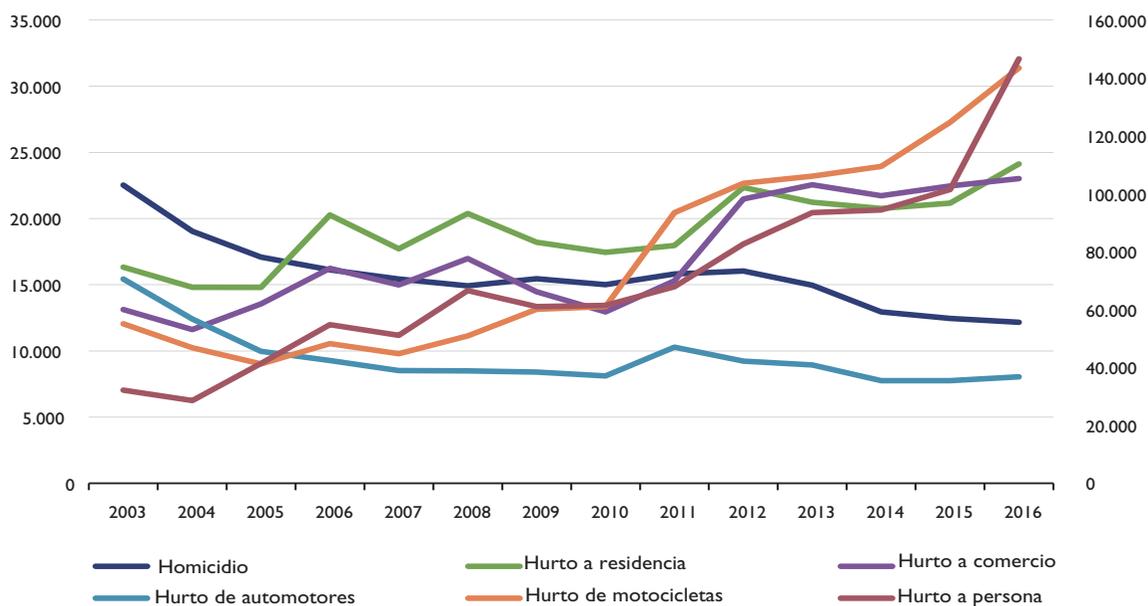


Figura 1. Historial de los principales crímenes en Colombia desde el 2003 (Norza et al., 2017, p. 13)

la técnica para modelar aspectos sociales como el crimen y reconocen un mejor ajuste en comparación con los modelos de regresión lineal. Así mismo, Rincón-Ruiz, Pascual y Flantua (2013) analizaron los factores asociados con los cultivos de coca y comprobaron que dichas relaciones varían espacialmente. Fowler (2013) analizó algunas variables socioeconómicas y su relación con las tasas municipales de homicidios en Colombia durante el año 2005. Finalmente, Norza *et al.* (2017) analizaron y compararon en un estudio descriptivo el comportamiento de los principales delitos de alto impacto en la sociedad colombiana entre 2015 y 2016. Se registró un aumento en el número de hurtos a residencias, locales comerciales, motocicletas y a personas, como se observa en la Figura 1.

Metodología

Este estudio es de tipo cuantitativo-correlacional (Dankhe, 1986) en el que se hace uso de información secundaria emitida por diferentes instituciones de orden nacional segregada por municipios; los datos utilizados (Tabla 2) provienen de bases de datos generadas por entidades gubernamentales, entre las que se encuentran el Departamento Administrativo Nacional de Estadística (DANE), el Departamento Nacional de Planeación (DPN), el Ministerio de Educación Nacional, Migración Colombia y el grupo de información de la Policía Nacional de Colombia. La información fue agrupada a nivel municipal y georreferenciada según el Marco Geoestadístico Nacional del año 2017 suministrado por el DANE.

Tabla 2.
VARIABLES EVALUADAS

Nombre	Variables	Tipo	Transformación	Descripción	Fuente
hurtos	loghurtos	Cuantitativa continua	Logarítmica ¹	Corresponde al logaritmo de la cantidad de hurtos a personas y celulares reportados durante el año 2017	Base de datos del grupo Observatorio del Delito de la Policía Nacional ²
sgppercapita	logsgppercapita	Cuantitativa continua	Logarítmica ¹	Corresponde al presupuesto en millones asignado por el Sistema general de Participaciones.	Departamento Nacional de Planeación ³
matri1000p	logmatri1000p ¹	Cuantitativa-continua	Logarítmica ¹	Corresponde al logaritmo de la cantidad de personas matriculadas en instituciones de educación superior	Datos del Ministerio de Educación Nacional -Sistema Nacional de Información de la Educación Superior ⁴
Categoría*	categoría	Cualitativa-nominal		Categoría del municipio según la Contaduría General de la Nación	Contaduría General de la Nación ⁵

* La variable categoría corresponde a la clasificación otorgada por la Contaduría General de la República teniendo en cuenta la normatividad vigente definida por el artículo primero de la Ley 617 de 2000 y el artículo séptimo de la Ley 1551 de 2012.

1 Transformación utilizada para cumplir los supuestos de normalidad y homocedasticidad, y hacer que datos medidos en escalas diferentes sean más comparables entre sí.

2 <https://www.policia.gov.co/grupo-informacion/C3%B3n-criminalidad/estadistica-delictiva>

3 https://sicodis.dnp.gov.co/ReportesSGP/SGP_ReportePorEntidadyFuente.aspx

4 <https://www.mineduacion.gov.co/sistemasinfo/Informacion-a-la-mano/212400:Estadisticas>

5 http://www.contaduria.gov.co/wps/wcm/connect/759e1fc3-bd8b-4d63-a95b-f7c37d509140/CT01+-+Categorizacion+2018+-+DNP.xlsx?MOD=AJPERES&CONVERT_TO=url&CACHEID=759e1fc3-bd8b-4d63-a95b-f7c37d509140

Nombre	Variables	Tipo	Transformación	Descripción	Fuente
Extranjeros	Totalextran	Cuantitativa-discreta		Corresponde a la cantidad de extranjeros hospedados en los municipios	Registro de ingreso de extranjeros según ciudad de hospedaje, Ministerio de Relaciones Exteriores de Colombia ⁶
Cobertura neta total	cober	Cuantitativa-continua	Logarítmica I	Porcentaje de estudiantes matriculados en el sistema educativo, sin contar los que están en extraedad (por encima de la edad correspondiente para cada grado)	Ministerio de Educación ⁷
Cobertura neta secundaria	Cobertura_neta	Cuantitativa-continua		Porcentaje de estudiantes matriculados en la educación secundaria, sin contar los que están en extraedad (por encima de la edad correspondiente para cada grado)	Ministerio de Educación ⁸
Cobertura neta educación básica	CobNeta	Cuantitativa-continua		Corresponde al porcentaje de cobertura neta en educación preescolar, básica, media y secundaria	Ministerio de Educación ⁹
Población total	PobTotal	Cuantitativa tipo razón		Corresponde a la población total del municipio	DANE. Proyecciones de la población total por municipio para el año 2017 ¹⁰
Población rural	PobRural	Cuantitativa tipo razón		Corresponde a la cantidad de residentes en el sector rural del municipio	DANE. Proyecciones de la población total por municipio para el año 2017
Departamento	Depto	Cualitativa nominal		Corresponde al departamento en el cual se encuentra el municipio	Marco Geoestadístico Nacional del 2017 ¹¹
Región	Región	Cualitativa nominal		Corresponde a la región geográfica en la cual se encuentra el municipio	Marco Geoestadístico Nacional del 2017 ¹²

6 http://migracioncolombia.gov.co/index.php/es/?option=com_content&view=article&id=718

7 <https://www.datos.gov.co/Educaci-n/ESTADISTICAS-EN-EDUCACION-BASICA-POR-MUNICIPIO/nudc-7mev>

8 <https://www.datos.gov.co/Educaci-n/ESTADISTICAS-EN-EDUCACION-BASICA-POR-MUNICIPIO/nudc-7mev>

9 <https://www.datos.gov.co/Educaci-n/ESTADISTICAS-EN-EDUCACION-BASICA-POR-MUNICIPIO/nudc-7mev>

10 https://www.dane.gov.co/files/investigaciones/poblacion/proyepobla06_20/Municipal_area_1985-2020.xls

11 <https://geoportal.dane.gov.co/v2/?page=elementoDescargaMGN>

12 <https://geoportal.dane.gov.co/v2/?page=elementoDescargaMGN>

La categorización de los municipios se realiza bajo los siguientes parámetros: los límites definidos de población en ingresos corrientes de libre destinación (en adelante ICLD) y el valor máximo definido de

los gastos de funcionamiento como proporción a los ICLD para las diferentes entidades territoriales, importancia económica y situación geográfica según lo estipulado en la Tabla 3.

Tabla 3.
Criterios para categorizar a los municipios según el artículo séptimo de la Ley 1551 de 2012

Categoría	Población	ICLD (en SMLMV)*	Importancia económica
Especial (0)	≥500.001	>400.000	Grado 1
Primera (1)	100.001 - 500.000	100.000 - 400.000	Grado 2
Segunda (2)	50.001 - 100.000	50.000 - 100.000	Grado 3
Tercera (3)	30.001 - 50.000	30.000 - 50.000	Grado 4
Cuarta (4)	20.001 - 30.000	25.000 - 30.000	Grado 5
Quinta (5)	10.001 - 20.000	15.000 - 25.000	Grado 6
Sexta (6)	≤10.000	hasta 15.000	Grado 7

* Salarios mínimos legales mensuales vigentes
Fuente: Contaduría General de la República.

En primer lugar, se realizó un análisis exploratorio de los datos, a partir de la evaluación de la correlación entre las variables, con el objetivo de identificar la posible presencia de relaciones lineales entre la variable predicha y las predictoras, al igual que la posible multicolinealidad entre predictoras. Se modelaron además diferentes regresiones lineales múltiples combinando las distintas variables mediante la función *step* de la librería *stats* del programa *R* (R Core Team, 2018). Adicionalmente, se utilizó un proceso empírico de prueba y error, con el cual se definió el modelo de regresión lineal múltiple que satisface los supuestos de normalidad, homocedasticidad y no multicolinealidad entre variables predictoras. Para complementar, se identificaron los valores atípicos en los residuales del modelo final y se comprobó que corresponden a información real característica de los municipios, por lo cual no fueron eliminados. Es importante mencionar que los modelos fueron elaborados con los datos de 532 municipios, los cuales contaban con la información completa en todas las variables.

Para satisfacer la condición de normalidad y homocedasticidad, las variables fueron recalculadas tomando en cuenta la población total de cada municipio (*i*) mediante la ecuación (6), con el propósito de obtener una tasa que facilite la comparación entre municipios.

$$x'_i = \frac{x_i}{Población_i} \times 1000 \quad (6)$$

Adicionalmente, se estandarizaron las variables usadas en los modelos mediante la ecuación (7).

$$x'_i = \frac{x_i - \mu}{\sigma} \quad (7)$$

Se aplicó una transformación logarítmica a los datos con la intención de mitigar el sesgo a la derecha presente en las variables. Esta transformación permite acotar el rango de las variables y reduce la sensibilidad de las estimaciones a los valores extremos, característica que resulta útil para comparar datos absolutos en inversión, crímenes, población, número de delitos y número de extranjeros en ciudades capitales contra pequeñas poblaciones. De la misma forma, esta transformación permitió que los modelos lineales cumplieran con los supuestos requeridos.

Para la construcción del modelo GWR se utilizó el programa estadístico de uso libre GWR4 (Nakaya, 2016), que permite la manipulación, procesamiento de datos georreferenciados y la aplicación de pruebas estadísticas, entre las que están la prueba de variación geográfica que determina la presencia de cambios significativos de los coeficientes en el espacio. Este modelo fue construido mediante un Kernel tipo “adaptativo *bi square*” que utiliza la siguiente ecuación para establecer las ponderaciones:

$$W_{ij} = \begin{cases} \left[1 - \frac{d_{ij}^2}{d^2}\right]^2 & \text{si } d_{ij} \leq d \\ 0 & \text{si es lo contrario} \end{cases} \quad (8)$$

Resultados

Análisis exploratorio de los datos

Los modelos fueron validados mediante varias pruebas: el de Shapiro-Wilk (Shapiro y Wilk, 1965) para evaluar la normalidad en los residuos, el de Breusch-Pagan (Breusch y Pagan, 1979) para garantizar la igualdad de varianza, el test de Durbin-Watson (1950, 1951, 1971) que tiene como hipótesis nula la ausencia de autocorrelación y el criterio de Factor de Inflación de Varianza (VIF¹³) para identificar la posible existencia de multicolinealidad entre variables regresoras.

Una vez obtenidos los modelos finales (GWR y regresión lineal múltiple), se visualizaron geográficamente los residuales en ArcMap y se calculó el I de Morán (Moran, 1950) para determinar la presencia de correlación espacial. Finalmente, la comparación entre modelos se realizó mediante los criterios de información de Akaike, Bayesiano y el coeficiente de determinación ajustado (R_a^2).

Mediante un análisis de los coeficientes de correlación de Pearson, entre las diferentes variables estudiadas se logró determinar las asociadas al hurto, proceso que permitió descartar algunas variables como predictoras e identificar redundancia de información aportada por algunas variables. En este sentido, la Figura 1 muestra que la única variable altamente correlacionada con el hurto a personas y de celulares es la población total (PobTotal). Las demás variables presentan correlaciones por debajo de 0,5. Tampoco existen correlaciones importantes entre las variables predictoras. La relación entre el presupuesto per cápita asignado en el Sistema General de Participaciones (sgp_percapita) y el hurto, aunque es muy baja, presenta el comportamiento esperado (a mayor presupuesto por persona, menor el número de hurtos).

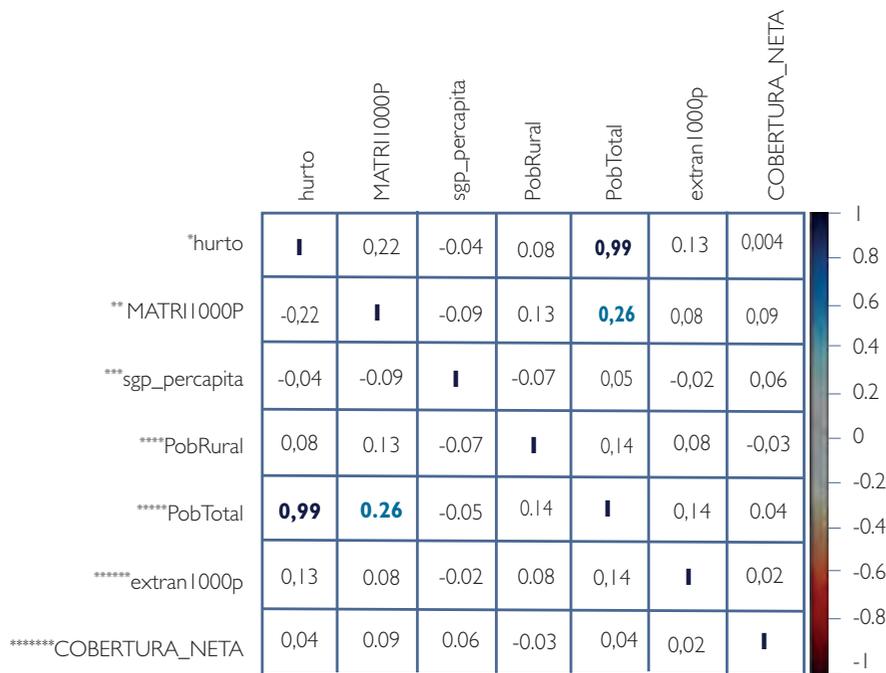


Figura 2. Matriz de correlación de las variables

* hurto corresponde al número de hurto a personas y celulares en 2017.

** MATRI1000P corresponde al número de matriculados en instituciones superiores por cada 1.000 personas.

*** sgp_percapita corresponde al presupuesto per cápita asignado por el Ministerio de Hacienda.

**** PobRural corresponde a la población rural.

***** PobTotal corresponde a la población total.

***** extran1000p corresponde al número de extranjeros por cada 1.000 personas.

***** COBERTURA_NETA corresponde a la cobertura neta de educación.

13 Valores mayores a 10 indican presencia de multicolinealidad grave Citar.

En la Figura 3 se observa una disminución del número de hurtos (logaritmo de los hurtos) al aumentar la categoría de los municipios. Es decir, la mayor cantidad de delitos registrados se localiza principalmente en las ciudades capitales, en municipios de categoría especial (categoría cero) y

de categoría uno. La mayor dispersión se presenta en los municipios de categoría 6. Es importante resaltar que existen municipios con baja población que presentan alto número de hurtos registrados, como el caso de algunos localizados en la región de los Llanos Orientales (Figura 4).

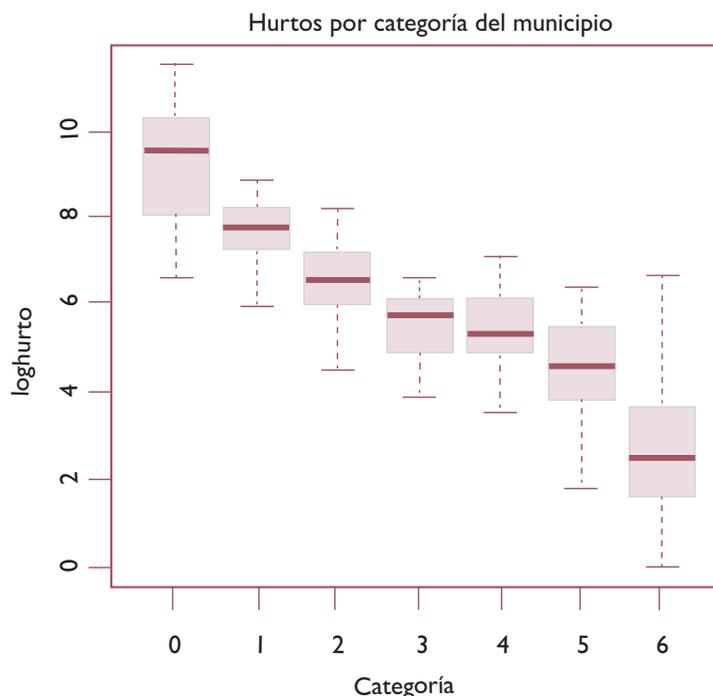


Figura 3. Diagramas de caja de la variable log hurto vs. categoría

Al realizar el análisis descriptivo de la variable hurto, se observa que en promedio se registraron en el 2017 en los 532 municipios 2,1 hurtos por cada mil habitantes, con una desviación estándar de 2,52. El valor mínimo fue de 0,03 en Moñitos, en el departamento de Córdoba, y el máximo, 15,6, en Pasto, departamento de Nariño.

En la Figura 5 se aprecia que existe una dispersión grande en los recursos económicos per cápita asignados por el Sistema General de Participaciones. En promedio, el presupuesto per cápita asignado por el SGP es de 530.090 pesos, con una desviación estándar de ± 232.439 ; el valor mínimo es de 136.461 pesos, asignado al municipio de Sopó, en el departamento de Cundinamarca, y el valor máximo es de 2'059.951 pesos, asignado al municipio de San Miguel, en el departamento de Santander.

La Figura 6 muestra que las principales ciudades del país (Bogotá, Cali, Barranquilla, Medellín, Bucaramanga y Cartagena) concentran el mayor número de estudiantes de educación superior (IES). El valor promedio de matriculados en IES es de 28,1 por cada mil habitantes; sin embargo, al hacer comparables los municipios por medio de la tasa de matriculados por cada mil personas, en el municipio de Marinilla, en el departamento de Antioquia, se presenta la menor tasa, con 0,018 estudiantes por cada mil personas. De forma opuesta, en Nilo, Cundinamarca, y Pamplona, Norte de Santander, más del 51% de sus habitantes se encuentran realizando estudios de educación superior, seguidos por Bucaramanga, Santander, con una proporción de más del 40%.

El número de municipios en cada categoría y su localización se presenta en la Figura 7.

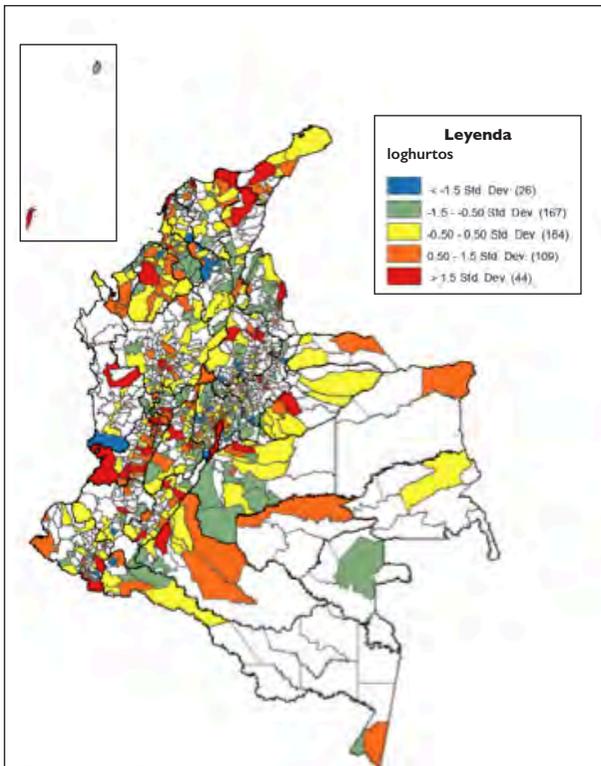


Figura 4. Hurto a personas y celulares (loghurto)

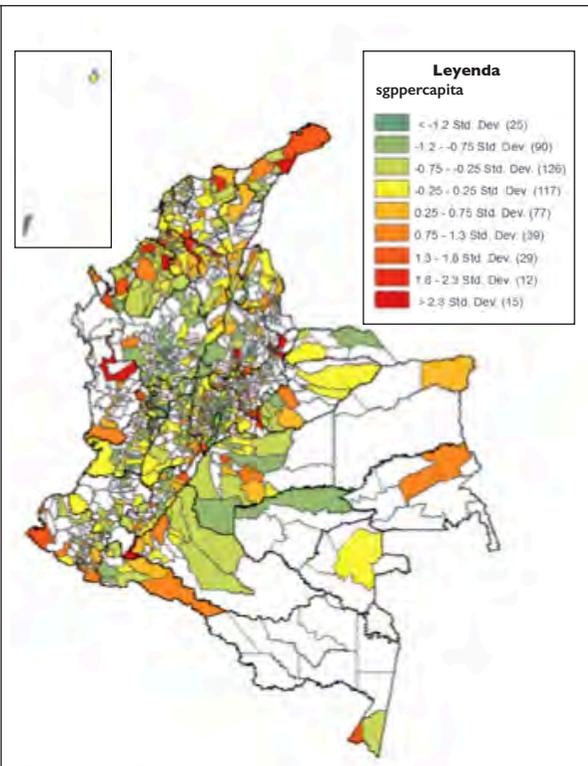


Figura 5. Presupuesto per cápita a partir del Sistema General de Participaciones

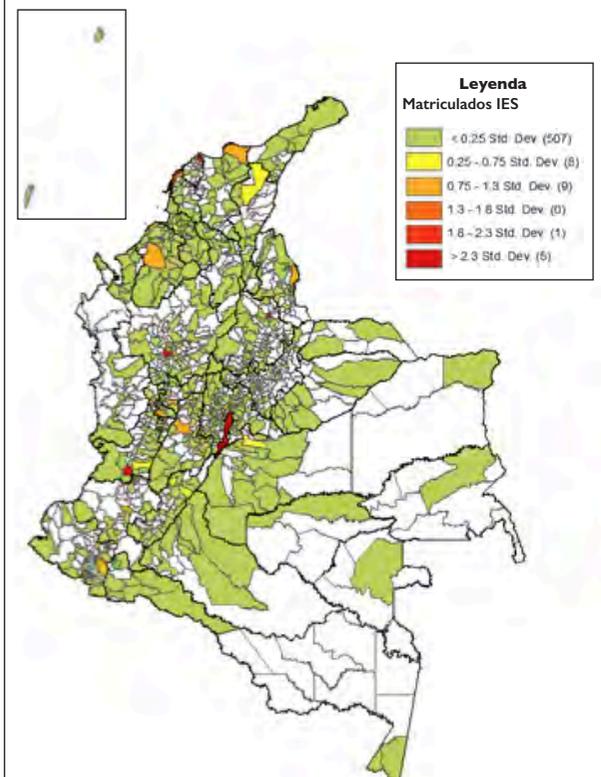


Figura 6. Matriculados en instituciones de educación superior

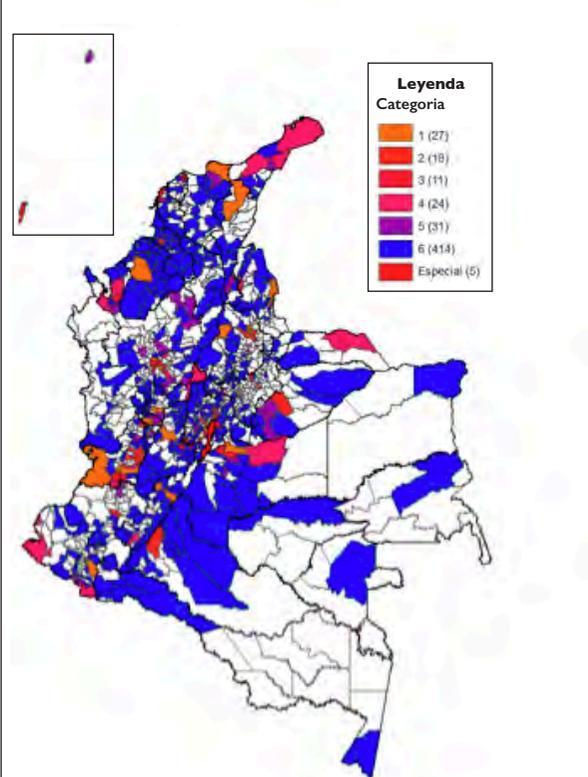


Figura 7. Categoría municipal

En este apartado se presentan los resultados para los modelos de regresión lineal múltiple y el espacial (GWR).

Modelo de regresión lineal múltiple

En la Tabla 4 se presentan los modelos lineales generados con el resultado de los valores AIC,

BIC el R_a^2 en porcentaje y las pruebas de normalidad, homocedasticidad y multicolinealidad. Nótese que entre los modelos evaluados, solo los “Modelo 14”, “Modelo 15” y “Modelo 16” cumplen con los supuestos de normalidad y homocedasticidad.

Tabla 4.
Modelos lineales evaluados y diagnóstico

Modelo lineal evaluado		Diagnóstico					
Modelo	Fórmula	AIC	BIC	R_a^2	*Normalidad residual	**Varianza constante	Multicolinealidad (VIF)
1	Loghurto ~ PobTotal	2.202,5	2.215,4	16,09	No	No	
2	loghurto ~ sgppercapitamillones + MATRI1000P + extran1000p + Cobertura_neta + NBI_Prop + Categoría + PobTotal + PobRural+ Depto + Region	1.581	1.782,8	75,85	No	No	
3	Loghurto ~ sgppercapitamillones + MATRI1000P + extran1000p + Cobertura_neta + NBI_Prop + Categoría + PobTotal + PobRural + Depto	1.581,8	1.782,81	75,85	No	No	Sí
4	Loghurto ~ sgppercapitamillones + MATRI1000P + extran1000p + Cobertura_neta + NBI_Prop + Categoría + PobTotal + PobRural	1.684,8	1.749	68,99	Sí	No	No
5	Loghurto ~ sgppercapitamillones + MATRI1000P + Cobertura_neta + NBI_Prop + Categoría + PobTotal + PobRural	1.683,71	1.743,5	69	Sí	No	Sí
6	Loghurto ~ log(PobTotal)	1.558,17	1.571	75	No	Si	
7	loghurto ~ logsgppercapita + logmatri1000p + extran1000p + logcober + log(NBI_Prop) + Categoría + log(PobTotal) + log(PobRural) + Depto + Region	1.275,8	1.455,5	86,3	No	No	Sí
8	loghurto ~ logmatri1000p + extran1000p + logcober + log(NBI_Prop) + log(PobTotal) + log(PobRural) + Depto	1.272,1	1.443	86,3	No	No	Sí
9	loghurto ~ logmatri1000p + extran1000p + log(NBI_Prop) + log(PobTotal) + log(PobRural) + Depto	1.273,4	1.440,2	86,29	No	No	Sí
10	loghurto ~ logmatri1000p + extran1000p + log(NBI_Prop) + log(PobTotal) + log(PobRural)	1.330,4	1.360,3	83,83	No	No	No
11	loghurto ~ logmatri1000p + extran1000p + log(NBI_Prop) + log(PobRural)	1.826	1.852	58,86	No	No	No

(Pasa)

(Viene)

Modelo lineal evaluado		Diagnóstico					
Modelo	Fórmula	AIC	BIC	R ² _a	*Normalidad residual	**Varianza constante	Multicolinealidad (VIF)
12	loghurto ~ logmatri1000p + extran1000p + log(NBI_Prop) + log(PobRural)+categoria	1.688,6	1.739,9	68,6	Sí	No	No
13	loghurto ~ logmatri1000p + log(NBI_Prop) + log(PobRural) + Categoría	1.687	1.734	68,63	Sí	No	No
14	loghurto ~ logmatri1000p + log(NBI_Prop) + log(PobRural) + Categoría + logsgpper capita	1.648,77	1.700,1	70,86	Sí	Sí	No
15	Loghurto ~ logmatri1000p + log(PobRural) + Categoría + logsgpper capita	1.647,8	1.710	70,8	Sí	Sí	No
16	Loghurto ~ logsgpper capita + logmatri1000p + Categoría	1.671	1.713,9	69,5	Sí	Sí	No
17	hurtoestan ~ coberestan + nbiestan + sgpperestan + matriestan + extranestan	242,49	272,43	90,9	No	No	No
18	hurtoestan ~ nbiestan + matriestan + extranestan	238,7	260,12	90,95	No	No	No
19	hurtoestan ~ nbiestan + matriestan + extranestan + Categoría	201,15	248,19	91,66	No	No	No
20	hurtoestan ~ nbiestan + coberestan+matriestan + Categoría	1.262,78	1.309,82	38,6	No	No	No
21	hurtoestan ~ extranestan +Categoría	199,58	238	91,65	No	No	No
22	Hurtoestan ~ nbiestan + coberestan + matriestan + sgpper capitamillones	1.491	1.516,6	4	No	No	No
23	Hurtoestan ~ nbiestan + coberestan + matriestan + sgpper capitamillones + Categoría	1.264,6	1.315,98	38	No	No	No
24	hurtoestan ~ nbiestan + coberestan + matriestan	1.489	1.510	5	No	No	No

* Se cumple al 5% de significancia.

** Se cumple al 5% de significancia.

$$\begin{aligned}
 \text{loghurto} = & 24.94971 - 1.30122\text{logsgpper capita} + (0.21779)\text{logmatri1000p} \\
 & - (1.30239)\text{Categoría1} - (2.30343)\text{Categoría2} \\
 & - (3.63049)\text{Categoría3} - (3.19047)\text{Categoría4} \\
 & - (4.26926)\text{Categoría5} - (5.25923)\text{Categoría6}
 \end{aligned}
 \tag{9}$$

La anterior fórmula expresa el modelo final (modelo 16) que tiene como grupo de referencia a los municipios con categoría especial.

El factor de inflación de varianza del modelo (Tabla 5) indica que no existen relaciones lineales fuertes entre variables predictoras. Por tanto, el modelo no presente multicolinealidad.

Tabla 5. Factor de inflación de varianza del modelo final (VIF)

Variable	VIF
Logmatri1000p	1.3594
Logsgpper capita	1.1697
Categoría	1.5522

Analizando cada una de las variables involucradas en el modelo anterior y manteniendo constantes las variables no mencionadas, se tiene que

- Para un aumento del 1% en la tasa de matriculados por cada mil personas, en promedio se espera un aumento del 0,21% en la cantidad de hurtos.
- Para un incremento del 1% en la inversión del SGP por habitante, se espera una reducción del 1,3% en promedio en la cantidad de hurtos.

Así mismo, el modelo indica que los municipios clasificados en las categorías uno, dos, tres, cuatro, cinco y seis presentan respectivamente en promedio 1,3%, 2,3%, 3,6%, 3,19%, 4,2% y 5,25% de hurtos menos en comparación con la categoría especial.

De los resultados obtenidos, este modelo (modelo 16 –ecuación 9–) fue escogido por principio de parsimonia entre aquellos que cumplieron con los supuestos de normalidad, ausencia de multicolinealidad e igualdad de varianza y significancia estadística (p -valor $< 0,05$). Sin embargo, el modelo no cumple

con el supuesto de autocorrelación entre los residuos, determinado a partir de test de Durbin Watson y partiendo de ello se aplicará a los datos un test de autocorrelación espacial. En cuanto a los análisis por variable, se confirma la significancia de cada una de ellas. Finalmente, se puede afirmar que el modelo explica 65,9% de la variación del logaritmo del hurto.

Modelo GWR

A partir del resultado del test de autocorrelación de Durbin-Watson aplicado a los residuales del modelo, se evaluó la autocorrelación espacial mediante el índice de Morán, el cual con un valor de 0,091625, y un p -valor $< 0,001$ apoya concluir que existe agrupamiento espacial en los residuales del modelo.

En el mapa de residuales del modelo lineal (Figura 8) se destacan, con circunferencias negras, los *clusters* de altos residuales presentes en la región Caribe y en el departamento de Caquetá.

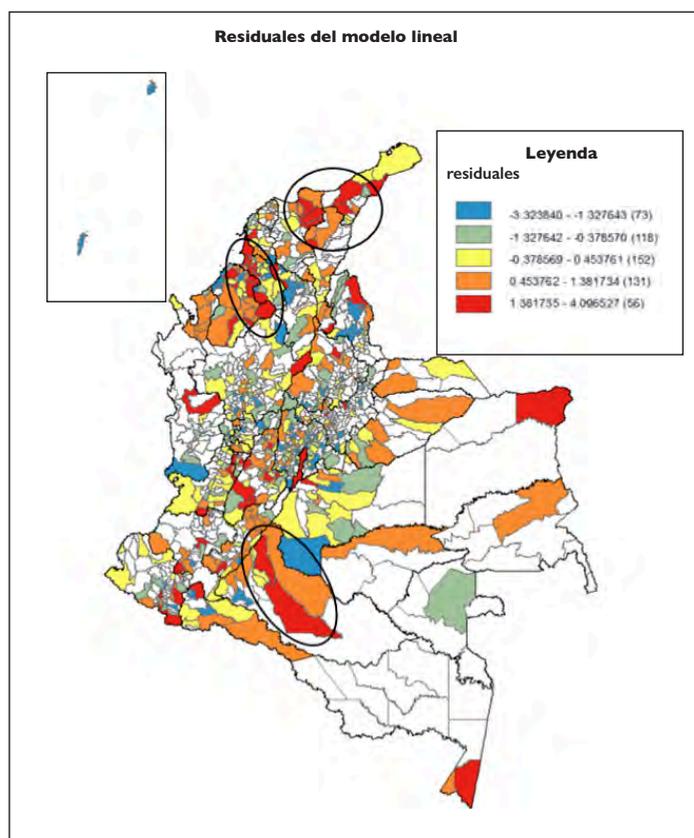


Figura 8. Mapa de distribución de los residuales del modelo lineal múltiple.

Al comprobarse la existencia de autocorrelación espacial en los residuales del modelo de regresión lineal múltiple final –ecuación 9–, se construyó un modelo GWR utilizando las mismas variables, con excepción de la variable categórica, debido al alto riesgo de presentar problemas de multicolinealidad local producto del alto agrupamiento espacial que exhibe esta variable y de que los modelos GWR se centran en variables que siguen distribuciones gaussianas (Dong, Nakaya y Brunson, 2018).

Intercepto del modelo GWR

Los valores más bajos se encuentran agrupados hacia el sur y suroeste del país, en los municipios de Puerto Nariño y Leticia, en el Amazonas, y Puerto Leguízamo, en el Putumayo.

Por otro lado, los valores más altos se concentran en el nororiente del país, en los municipios de Sardinata y Ocaña, en el departamento de Norte de Santander, y San Alberto, en el Cesar (Figura 9).

Coefficiente variable logsgppercapita

En la Figura 10 se observa que el coeficiente de la variable presupuesto per cápita asignado por el Sistema General de Participaciones, transformada logarítmicamente, presenta una correlación negativa con el hurto en todo el país, por lo cual tiene un

comportamiento atenuante en relación con el hurto; sin embargo, esta relación es variable espacialmente. Los municipios con mayor valor se concentran hacia el sur y suroeste del país, en Puerto Nariño y Leticia, en el Amazonas, y Puerto Leguízamo, en el Putumayo.

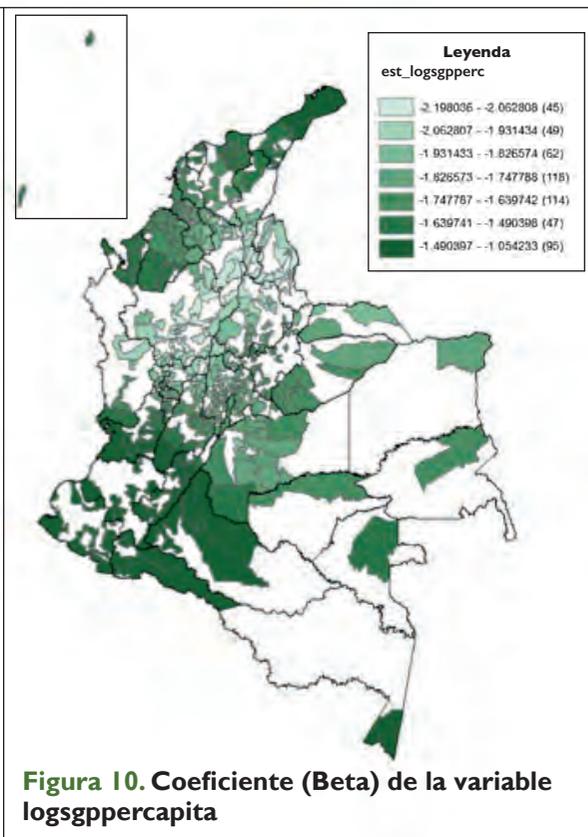
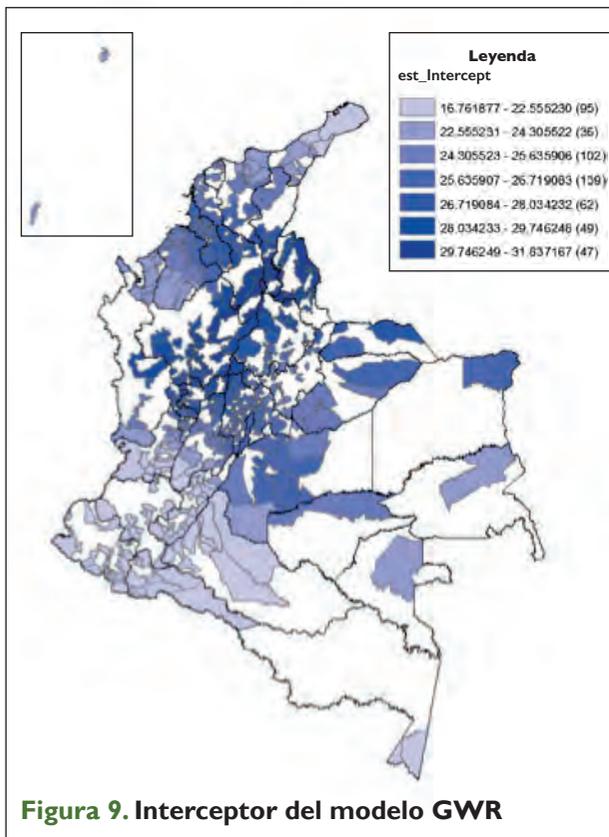
Los valores más bajos se concentran hacia la zona del Magdalena medio: Norte de Santander –Sardinata, Gramalote– y Cesar –San Alberto–. Esto indica que en estas regiones esta variable influye en menor medida sobre los hurtos.

Coefficientes variables logmatri1000p

En la Figura 11 se observa que la tasa de matriculados en instituciones de educación superior por cada mil personas presenta una relación directa con el hurto; sin embargo, esta relación varía espacialmente en menor medida en comparación con la variable logsgppercapita y el intercepto del modelo.

Los valores más altos se concentran en el norte y suroeste del país, siendo los municipios de Tumaco y El Charco, en el departamento de Nariño, y Guapi, en el Cauca, los que presentan los mayores coeficientes.

Por otra parte, los menores valores se concentran principalmente en los departamentos de Antioquia y los Santanderes; los municipios de Amalfi, Yolombó y Yondó, en Antioquia, presentan los menores coeficientes.



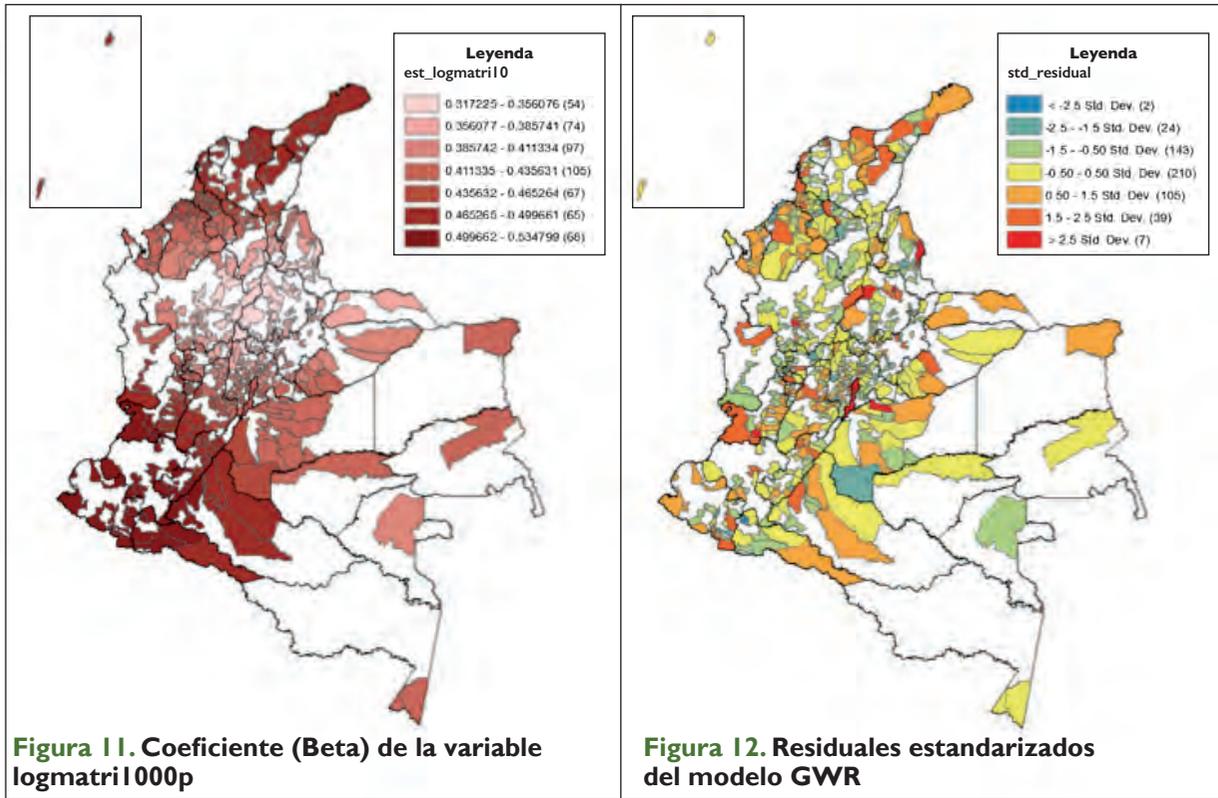


Figura 11. Coeficiente (Beta) de la variable logmatrilo00p

Figura 12. Residuales estandarizados del modelo GWR

Diagnóstico del modelo GWR

Tal como los señalan Estrada y Durán (2015, p. 10), en los procesos espaciales no se cumplen los supuestos de normalidad y homocedasticidad

... debido a lo irregular de las unidades geográficas (los municipios) por diferencias significativas de áreas, y fundamentalmente por el papel del espacio, que no es sólo contenedor de información, sino factor explicativo de los procesos que en él se presentan.

```

*****
GWR (Geographically weighted regression) result
*****
Bandwidth and geographic ranges
Bandwidth size:                248.374222
Coordinate                    Min            Max            Range
-----
X-coord                       167428.000000  1658641.000000  1491213.000000
Y-coord                       89436.000000   1980257.000000  1890821.000000

Diagnostic information
Residual sum of squares:      1151.156315
Effective number of parameters (model: trace(S)): 16.758168
Effective number of parameters (variance: trace(S'S)): 10.734758
Degree of freedom (model: n - trace(S)): 515.241832
Degree of freedom (residual: n - 2trace(S) + trace(S'S)): 509.218422
ML based sigma estimate:      1.470995
Unbiased sigma estimate:      1.503540
-2 log-likelihood:            1920.390077
Classic AIC:                   1955.906413
AICc:                           1957.204478
BIC/MDL:                       2031.851766
CV:                             2.328926
R square:                       0.501662
Adjusted R square:             0.479323
    
```

Figura 13. Resumen y medidas de diagnóstico del modelo GWR.

Fuente: Salida gráfica programa GWR4

Por lo tanto, los procesos espaciales tienden a ser no estacionarios, es decir, que presentan variación regional o local.

El modelo resultante (Figura 13) indica que aproximadamente el 50% de la variabilidad del logaritmo del hurto a personas y celulares es explicado por las variables predictoras.

A partir del modelo GWR se obtienen valores de los coeficientes para las variables en cada municipio.

Los resultados mostrados en la Tabla 6 indican que la variable logsgppercapita es la que presenta mayor variación geográfica, con una desviación estándar de 0.2120, lo que podría indicar que diferencias en la inversión per cápita en los 532 municipios incluidos en el modelo inciden en los indicadores de hurto. Por otro lado, la tasa de matriculados por cada mil personas muestra una baja dispersión, con un coeficiente de variación del 12,5%.

Tabla 6. Mínimo, máximo, rango y valor promedio de los estimadores de las variables predictoras del modelo GWR

Variable	Mínimo	Máximo	Rango	Media	Desviación
Intercepto	16.761877	31.837167	15.075289	25.750007	2.6803
logmatri1000p	0.317225	0.534799	0.217574	0.426423	0.0535
logsgppercapita	-2.198036	-1.054233	1.143803	-1.737462	0.2121

La prueba de variabilidad geográfica tiene como objetivo evaluar si las variables predictoras presentan variabilidad significativa en su coeficiente Beta a nivel espacial. El resultado mostrado en la Figura 14 indica que existe variación significativa en el intercepto, y en la variable logsgppercapita (logsgpperc), por el contrario, la variable logmatri1000p (logmatri10)

no varía significativamente. Esto se deduce a partir del valor *DIFF of Criterion* (AICc) menor a 2, que es el resultado de la diferencia entre los AICc de los modelos con el parámetro global o local. Lo anterior concuerda con la poca dispersión observada en el coeficiente de la variable logmatri1000p presentado en Tabla 5.

```

*****
Geographical variability tests of local coefficients
*****
Variable           F           DOF for F test  DIFF of Criterion
-----
Intercept          994.233346    5.440  515.242    -1287.633180
logmatri10         2.035058     4.589  515.242     0.199089
logsgpperc         119.371867    4.763  515.242    -385.477611
    
```

Figura 14. Resultado del test de variabilidad espacial.

Fuente: Resultados obtenidos a partir del programa GWR4.

En la Figura 12 se observa el mapa de distribución de los residuales estandarizados del modelo GWR, los cuales, a partir del resultado del índice de Morán mostrado en la Figura 15, no se encuentran distribuidos

aleatoriamente ya que existe agrupamiento espacial. Este comportamiento puede ser debido a la ausencia de variables clave en el modelo.

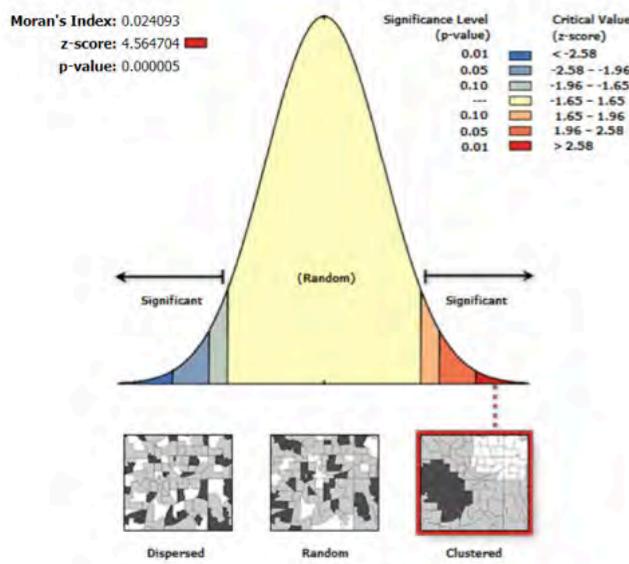


Figura 15. Resultado del índice de Morán aplicado a los residuos estandarizados del modelo GWR.

Finalmente, el coeficiente de determinación R^2 del modelo GWR mostrado en la Figura 16 indica que existen diferencias en el ajuste del modelo a nivel espacial. El menor ajuste se observa en el departamento de Córdoba y la región del Urabá, donde las variables predictoras logran explicar

tan solo entre 39 y el 43% aproximadamente de la variabilidad del hurto. Por otro lado, el mejor ajuste del modelo GWR se presenta hacia el centro del país en los departamentos de Cundinamarca, Boyacá y el sur de Santander principalmente, donde se alcanzan cifras entre el 53 y 57% de variabilidad explicada.

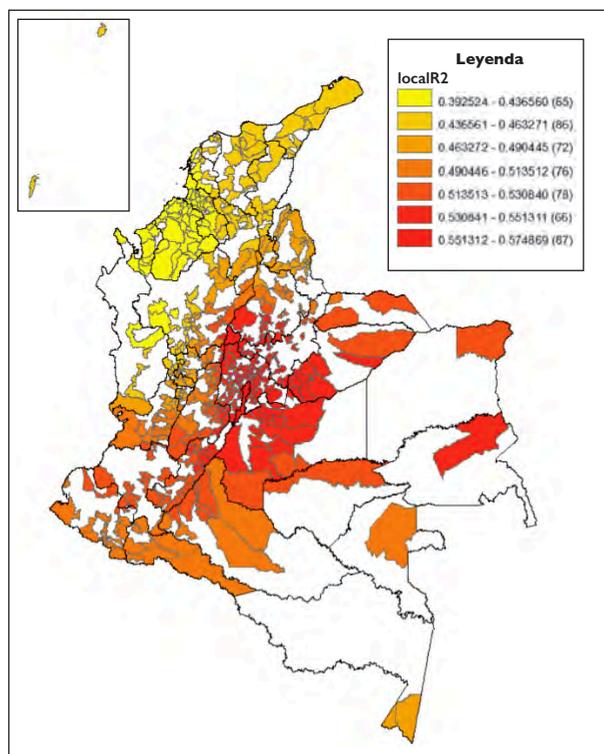


Figura 16. Coeficiente de determinación R^2 local según el modelo GWR.

Comparación entre los modelos de regresión lineal múltiple y el modelo GWR

Para comparar modelos equivalentes, se realizó un modelo de regresión lineal múltiple sin tomar en cuenta la categoría del municipio (MCO global) ya que este con las variables $\log \text{mari}1000p$, $\log \text{sgppercapita}$ no cumple los supuestos de normalidad y homogeneidad de varianza.

Al comparar el modelo OLS global con su equivalente GWR, se observa que el modelo espacial explica en aproximadamente 4% más la variabilidad del hurto a personas y celulares. De igual forma, presenta un menor AIC.

Al cotejar el modelo de regresión lineal múltiple final (OMC Modelo 16), que incluye la variable Categoría, con el resto, se logra un aumento considerable en el R^2 , una disminución del AIC y un modelo que cumple con los supuestos de normalidad y homocedasticidad en los residuales (Tabla 7).

Tabla 7.
Principales medidas de bondad de ajuste para la comparación entre modelos

Modelo	Predictoras	R ² (%)	R ² ajustado	AIC
OLS global	$\log \text{mari}1000p$, $\log \text{sgppercapita}$	46,13	0,4582	1.969,9
OLS Modelo 16	$\log \text{sgppercapita}$, $\log \text{mari}1000p$, Categoría	69,96	0,695	1.672,6
GWR (local)	$\log \text{mari}1000p$, $\log \text{sgppercapita}$	50,16	0,4793	1.955,9*

Conclusiones

El modelo de regresión lineal múltiple estimado por mínimos cuadrados ordinarios que tiene como variables predictoras el presupuesto per cápita asignado por el Sistema General de Participaciones y la tasa de matriculados en instituciones de educación superior por cada mil habitantes requirió una transformación logarítmica y la inclusión de la variable nominal categoría del municipio. De esta forma, se cumple con los supuestos de normalidad y homocedasticidad de los residuales y no multicolinealidad en las variables predictoras. Este modelo logra explicar el 69,5% de la variabilidad del logaritmo de los hurtos a personas y celulares registrados por la Policía Nacional durante el año 2017 en los 532 municipios incluidos en el análisis. Sin embargo, este tipo de modelos realiza una valoración global sin considerar la localización de los datos; por lo tanto, no tiene en cuenta las posibles heterogeneidades propias de los territorios, que se reflejan en la presencia de autocorrelación espacial de los residuales.

El modelo de regresión ponderada geográficamente estimado mediante el uso de un *kernel* de tipo adaptativo *bi-square* según el criterio de selección AICc, con las variables predictoras: presupuesto per cápita asignado por el Sistema General de Participaciones y la tasa de matriculados en instituciones de educación

superior por cada mil habitantes, ambas transformadas logarítmicamente, sin incluir la variable categoría, explica el 50,16% de la variación del logaritmo del hurto con un AIC de 1955,9. De esta forma, se obtiene un mejor ajuste en comparación con el mismo modelo de regresión lineal múltiple estimado mediante mínimos cuadrados ordinarios, el cual presenta un coeficiente de determinación de 46,1 y un AIC de 1969,8. Por lo tanto, se puede concluir que al aplicar la técnica GWR a datos que presentan autocorrelación espacial, se obtienen mejores resultados.

A pesar del mejor ajuste logrado mediante la técnica GWR en comparación con el mismo modelo estimado por mínimos cuadrados ordinarios, el método de regresión espacial está limitado al uso de variables numéricas. Por tanto, al no permitir el uso de variables categóricas, se restringe el uso de posible información clave para el modelo.

La inclusión de la categoría del municipio en el modelo aumentó el coeficiente de determinación a 69,96%, lo cual representa una mejora de 19,44% con respecto al modelo GWR (que presentó un R^2 de 50,16%) y del 23,83% respecto al modelo de regresión múltiple sin la variable categórica.

Del análisis realizado y los modelos evaluados se puede deducir que, para los 532 municipios incluidos en el modelo, las variables cantidad de extranjeros por cada mil habitantes, población rural y cobertura neta

en educación preescolar, básica, media y secundaria no presentan relación significativa con el hurto a personas y celulares durante el año 2017.

En los tres modelos finales se observa una relación inversa entre el hurto a personas y celulares y el presupuesto per cápita asignado a través del Sistema General de Participaciones, con lo que se concluye que a mayor inversión en los municipios, sin importar su localización, se tiende a reducir las cifras de hurtos.

Los matriculados en instituciones de educación superior por cada mil habitantes presentan una relación directa con el hurto a personas y celulares, relación que posiblemente sea el reflejo de una tendencia en este tipo de hurtos en zonas de alto flujo de personas, como lo son los centros universitarios. Sin embargo, este resultado amerita estudios posteriores sobre la conducta criminal y ocurrencia de delitos que permitan aclarar el porqué de esta relación.

Los resultados obtenidos indican que cuanto mejor es la categoría de los municipios, mayor es la ocurrencia de hurtos a personas y celulares.

Recomendaciones

Los estudios sobre criminalidad, especialmente los enfocados en atentados contra el patrimonio económico, requieren de la inclusión de variables e información relacionada con la calidad de vida y pobreza, ya que es razonable pensar que este tipo de delitos está motivado por el insuficiente acceso a condiciones de vida dignas asociadas a bajos ingresos. Se recomienda para futuros estudios la inclusión de las variables distintos indicadores de pobreza y tasas de desempleo, que no fueron incluidas en este estudio debido a la ausencia de datos actualizados y segregados a nivel municipal.

Los modelos obtenidos pueden ser utilizados como insumo en posteriores estudios de política pública cuyo objetivo sea mitigar el hurto a nivel municipal teniendo en cuenta la heterogeneidad del territorio.

Para futuros estudios sobre criminalidad, se recomienda tener en cuenta variables disuasivas que dificultan la ejecución de los delitos, como el pie de fuerza policial, militar y judicial presente de los municipios. Esta información no fue tenida en cuenta en este estudio debido a la falta de información de fácil acceso público.

La regresión geográfica permite mejorar los ajustes, neutralizar la dependencia espacial en los residuos y conocer la distribución espacial tanto de las elasticidades en las variables explicativas como en la significación local del modelo, todo ello sin renunciar

al carácter directo y de respuesta rápida (Gutiérrez-Puebla et al., 2012, p. 1). Además, la combinación de precisión, bajo coste y sencillez hace que los modelos estadísticos se conviertan en una alternativa razonable para municipios de tamaño pequeño o de recursos limitados que necesitan evaluar de forma rápida y precisa el número de hurtos.

Referencias

- Baron, S.W. (2003). Self-Control, Social Consequences, and Criminal Behavior: Street Youth and the General Theory of Crime. *Journal of Research in Crime and Delinquency*, 40(4): 403-425. <https://doi.org/10.1177/0022427803256071>.
- Breusch, T. S. & Pagan, A. R. (1979). A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica*, 47(5): 1287. <https://doi.org/10.2307/1911963>.
- Brunsdon, C., Fotheringham, A. S. & Charlton, M. (1996). Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity. *Geographical Analysis*, 28(4), 281-298. <https://doi.org/10.1111/j.1538-4632.1996.tb00936.x>.
- Brunsdon, C., Fotheringham, A. S. & Charlton, M. (1998). Geographically Weighted Regression. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 47(3): 431-443. <https://doi.org/10.1111/1467-9884.00145>.
- Brunsdon, C., Fotheringham, A. S. & Charlton, M. (1999). Some Notes on Parametric Significance Tests for Geographically Weighted Regression. *Journal of Regional Science*, 39(3): 497-524. <https://doi.org/10.1111/0022-4146.00146>.
- Castillo G, L. C. (2007). Etnicidad y nación: El desafío de la diversidad en Colombia (1. ed.). Colección Libros de investigación. Cali Colombia: Programa Editorial Universidad del Valle. Recuperado de <https://books.google.es/books?id=HyYPkuRP891C>
- Cahill, M. & Mulligan, G. (2007). Using Geographically Weighted Regression to Explore Local Crime Patterns. *Social Science Computer Review*, 25(2), 174-193. <https://doi.org/10.1177/0894439307298925>.
- Charlton, M. & Fotheringham, A. S. (2009). *Geographically weighted regression (White Paper)*. Recuperado de https://www.geos.ed.ac.uk/~gistea/fspat/gwr/gwr_arcgis/GWR_WhitePaper.pdf.
- Charlton, M., Fotheringham, A. S. & Brunsdon, C. (2006). *Geographically weighted regression: NCRM Methods Review Papers/NCRM/006*. Recuperado de <http://eprints.ncrm.ac.uk/90/>.

- Chávez, S. M. (2018). Causas y factores que conllevan a cometer el delito de feminicidio en la provincia de Huánuco 2016-2017. Universidad de Huánuco, Huánuco - Perú.
- Chen, X., Thrane, L., Whitbeck, L. B., Johnson, K. D. & Hoyt, D. R. (2007). Onset of conduct disorder, use of delinquent subsistence strategies, and street victimization among homeless and runaway adolescents in the Midwest. *Journal of interpersonal violence*, 22(9): 1156-1183. <https://doi.org/10.1177/0886260507303731>.
- Dankhe, G. L. (1986). Investigación y comunicación. En C. Fernández-Collado y G. L. Dankhe (Eds.), *La comunicación humana: ciencia social* (pp. 385-454). México: McGraw-Hill.
- Dong, G., Nakaya, T. & Brunson, C. (2018). Geographically weighted regression models for ordinal categorical response variables: An application to geo-referenced life satisfaction data. *Computers, Environment and Urban Systems*, 70, 35-42. <https://doi.org/10.1016/j.compenurbsys.2018.01.012>.
- Durbin, J. & Watson, G. S. (1950). Testing for Serial Correlation in Least Squares Regression: I. *Biometrika*, 37(3/4): 409. <https://doi.org/10.2307/2332391>.
- Durbin, J. & Watson, G. S. (1951). Testing for Serial Correlation in Least Squares Regression. II. *Biometrika*, 38(1/2): 159. <https://doi.org/10.2307/2332325>.
- Durbin, J. & Watson, G. S. (1971). Testing for Serial Correlation in Least Squares Regression. III. *Biometrika*, 58(1): 1. <https://doi.org/10.2307/2334313>.
- Eisner, M. (2002). Crime, Problem Drinking, and Drug Use: Patterns of Problem Behavior in Cross-National Perspective. *The ANNALS of the American Academy of Political and Social Science*, 580(1): 201-225. <https://doi.org/10.1177/000271620258000109>.
- Environmental Systems Research Institute (2016). *Regresión ponderada geográficamente (GWR)-Ayuda*. Recuperado de <http://desktop.arcgis.com/es/arcmap/10.3/tools/spatial-statistics-toolbox/geographically-weighted-regression.htm>.
- Estrada, L. & Durán, C. (2015). *Estudio sobre las relaciones espaciales locales entre la pobreza multidimensional, la ruralidad y la capacidad institucional*. Recuperado de https://www.researchgate.net/publication/317721631_Estudio_sobre_las_relaciones_espaciales_locales_entre_la_pobreza_multidimensional_la_ruralidad_y_la_capacidad_institucional.
- Fajnzylber, P., Lederman, D. & Loayza, N. (2002). Inequality and violent crime. *The Journal of Law and Economics*, 45(1): 1-39.
- Fotheringham, A. S., Charlton, M. & Brunson, C. (1998). Geographically Weighted Regression: A Natural Evolution of the Expansion Method for Spatial Data Analysis. *Environment and Planning A*, 30(11), 1905-1927. <https://doi.org/10.1068/a301905>.
- Fotheringham, A. S., Charlton, M. & Brunson, C. (2001). Spatial Variations in School Performance: A Local Analysis Using Geographically Weighted Regression. *Geographical and Environmental Modelling*, 5(1): 43-66. <https://doi.org/10.1080/13615930120032617>.
- Fowler, B. (2013). Understanding Colombian Violence Through Geographic Information Systems and Statistical Approaches (Master Thesis). Western Kentucky University, Bowling Green, Kentucky.
- Gottfredson, M. R. & Hirschi, T. (1990). *A general theory of crime*. Stanford Calif.: Stanford University Press.
- Gutiérrez-Puebla, J., García-Palomares, J. C. & Daniel-Cardozo, O. (2012). Regresión Geográficamente Ponderada (GWR) y estimación de la demanda de las estaciones del Metro de Madrid. En *XV Congreso Nacional de Tecnologías de la Información Geográfica*. Madrid.
- Harrison, L. D. (1992). The Drug-Crime Nexus in the USA. *Contemporary Drug Problems*, 19, 203.
- Mancero, X. & Feres, J. C. (2001). *El método de las necesidades básicas insatisfechas (NBI) y sus aplicaciones en América Latina. Serie estudios estadísticos y perspectivas: Vol. 7*. Santiago: Naciones Unidas, CEPAL. Recuperado de <http://hdl.handle.net/11362/4784>.
- Moran, P. A. P. (1950). Notes on Continuous Stochastic Phenomena. *Biometrika*, 37(1/2): 17-23. <https://doi.org/10.2307/2332142>.
- Nakaya, T. (2016). *GWR4.09 User Manual: GWR4 Windows Application for Geographically Weighted Regression Modelling*. Recuperado de https://raw.githubusercontent.com/gwrtools/gwr4/master/GWR4manual_409.pdf.
- Norza, E., Peñalosa, M. J. & Rodríguez, J. D. (2017). Exégesis de los registros de criminalidad y actividad operativa de la Policía Nacional en Colombia, año 2016. *Revista Criminalidad*, 59(3): 9-40.

- Öcal, N. & Yildirim, J. (2010). Regional effects of terrorism on economic growth in Turkey: A geographically weighted regression approach. *Journal of Peace Research*, 47(4): 477-489. <https://doi.org/10.1177/0022343310364576>
- Ortiz-Yusty, C. E., Páez, V. & Zapata, F. A. (2013). Temperature and precipitation as predictors of species richness in northern Andean amphibians from Colombia. *Caldasia*, 35(1), 65-80.
- R Core Team. (2018). Vienna, Austria: R Foundation for Statistical Computing. Recuperado de <https://www.R-project.org/>.
- Ramírez, K.V., López, Y., Castro, J. & Ocampo, J. F. (2017). Factores geográficos que influyeron en la concentración del hurto de motocicletas en diez estaciones policiales del Valle del Cauca, entre el 2010 y el 2015. *Revista Criminalidad*, 59(2): 9-31.
- Rincón-Ruiz, A., Pascual, U. y Flantua, S. (2013). Examining spatially varying relationships between coca crops and associated factors in Colombia, using geographically weight regression. *Applied Geography*, 37, 23-33. <https://doi.org/10.1016/j.apgeog.2012.10.009>
- Sandoval, L. E. & Barón, D. M. (2008). Una revisión al estudio de la delincuencia y criminalidad. *Revista Facultad de Ciencias Económicas: Investigación y Reflexión*, 16(1): 105-117. Recuperado de <https://dialnet.unirioja.es/descarga/articulo/4237385.pdf>.
- Shapiro, S. S. & Wilk, M. B. (1965). An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, 52(3/4): 591. <https://doi.org/10.2307/2333709>.
- Tobler, W. R. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, 46(sup1): 234-240.
- Wheeler, D. & Tiefelsdorf, M. (2005). Multicollinearity and correlation among local regression coefficients in geographically weighted regression. *Journal of Geographical Systems*, 7(2), 161-187. <https://doi.org/10.1007/s10109-005-0155-6>.
- Zhao, J., Wang, W. & Cheng, Q. (2014). Application of geographically weighted regression to identify spatially non-stationary relationships between Fe mineralization and its controlling factors in eastern Tianshan, China. *Ore Geology Reviews*, 57, 628-638. <https://doi.org/10.1016/j.oregeorev.2013.08.005>.