

UNIVERSIDAD
ICESI

LA GEOGRAFÍA DE LA EDUCACIÓN EN CALI

AUTORES:

JOHN JAIRO LOPEZ VALENCIA

DARLY MARCELA REVELO IBARRA

DIRECTOR DEL PROYECTO:

JUAN TOMÁS SAYAGO GÓMEZ

UNIVERSIDAD ICESI

FACULTAD DE CIENCIAS ADMINISTRATIVAS Y ECONÓMICAS

ECONOMÍA Y NEGOCIOS INTERNACIONALES

SANTIAGO DE CALI

2019

Tabla de contenido

Introducción	8
Objetivos	10
Objetivo general.....	10
Objetivos específicos	10
Marco teórico	11
Resultados	23
Conclusiones	30
Referencias bibliográficas.....	32

Lista de tablas

Tabla 1 ÍNDICE DE MORAN TOTAL 2013.....	23
Tabla 2 ÍNDICE DE MORAN INGLÉS.....	28
Tabla 3 REGRESIÓN (AÑO BASE 2013)	29

Lista de ilustraciones

Ilustración 1 GRÁFICO ÍNDICE DE MORAN	18
Ilustración 2 GRÁFICO ÍNDICE DE MORAN TOTAL 2013	24
Ilustración 3 GRÁFICO ÍNDICE DE MORAN TOTAL 2017	24
Ilustración 4 LISA TOTAL 2013	25
Ilustración 5 LISA TOTAL 2017	27
Ilustración 6 LISA INGLÉS 2013	28

Resumen

La Geografía de la Educación en Cali es un proyecto de investigación de tipo exploratorio que se encuentra dentro de la categoría de educación y que usa como instrumento las Pruebas Saber 11. Este proyecto pretende evaluar el efecto que tiene la ubicación geográfica de las instituciones educativas (en la ciudad de Cali) en los resultados de las pruebas, empleando la metodología de aglomeración y la correlación espacial. Los resultados obtenidos muestran que en la ciudad de Cali se observan clusters educativos que condicionan los resultados de las Pruebas Saber 11, de esta manera también fue posible determinar que hay una autocorrelación positiva, es decir que los colegios con rendimiento académico alto/bajo tienen vecinos con rendimiento alto/bajo, agrupándose en clusters educativos.

Palabras clave

Pruebas Saber 11, Correlación Espacial, Análisis de Clúster, Aglomeraciones.

Abstract

The Geography of Education in Cali is an explanatory research project that falls within the category of education and uses the Pruebas Saber 11 as an instrument. This project aims to evaluate the effect of the geographic location of the educational institutions (in the city of Cali) in the results of the tests, using the methodology of agglomeration and spatial correlation. The results obtained show that in the city of Cali there are educational clusters that condition the results of the Pruebas Saber 11, in this way it was also possible to determine that there is a positive autocorrelation, that is to say that the schools with high/low academic performance have neighbors with high/low performance, grouped in educational clusters.

Key words

Pruebas Saber 11, Spatial Correlation, Cluster Analysis, Agglomerations.

Introducción

La educación trae consigo beneficios tanto para los individuos, como para la sociedad. En cuanto a los beneficios individuales se tiene que existe una correlación positiva entre el nivel educativo y las oportunidades de empleo y obtención de rentas. Además, se puede percibir mejoras en la salud, pautas de consumo más apropiadas y mayor estatus social, entre otros (Salas, 2008).

Por otro lado, los beneficios para la sociedad incluyen los cambios en las actitudes y valoraciones de la sociedad, el aumento de la participación social, la mejora de la eficiencia y el descenso de la delincuencia (Salas, 2008). Lo que quiere decir que la educación es una variable que toma especial importancia en la sociedad, no solo como determinante individual o social, sino también como determinante del desarrollo económico de un país.

“Según Gary Becker (1983) cuando el capital humano aumenta, gracias a la alta inversión en los sectores educativos, el retorno en la inversión de capital humano aumenta hasta que éste crece lo suficiente y encuentra un equilibrio, es decir, se eleva el ingreso per cápita de la sociedad debido a la relación directa con el crecimiento económico y el stock de capital humano”, (Becker, 1983, citado en Apache, Vinasco & Rojas, 2018, p.4).

Para el caso de Colombia esto no es ajeno, el Departamento Nacional de Planeación (2019, párr.1) afirma que “la educación cumple un papel estratégico en el crecimiento y desarrollo económico y social del país, dado que aporta al progreso de las personas y la sociedad en conjunto” Por lo tanto, es importante la información que se genera a partir de estudios que permiten obtener indicadores de la situación de la educación, aportando al cierre

de brechas de desinformación y permitiendo generar planes específicos de desarrollo en pro de la calidad de la educación para formar capital humano cada vez más eficiente, lo que a largo plazo generará el acotamiento de brechas de desigualdad y oportunidades para los menos favorecidos.

Específicamente en la ciudad de Cali se han estado desarrollando proyectos que mejoran la calidad de la educación, optimizando redes eléctricas, cocinas, baños, comedores, techos y zonas verdes, además, de la instalación de filtros de agua potable y arreglos de infraestructura, con lo cual se han beneficiado 6.846 estudiantes y 257 docentes, señalaron fuentes del Grupo energético Celsia (Celsia, 2018, citado en Ortegón R., 2018). En este sentido, se puede apreciar que se vienen adelantando obras enfocadas en la infraestructura de los planteles educativos, que buscan generar espacios que contribuyan a un ambiente estudiantil adecuado que permita mejorar los procesos educativos. Pero, aún es necesario saber cuáles exactamente son los lugares donde se debería hacer las inversiones más significativas.

Es así como la presente investigación busca hacer un bosquejo de cómo se encuentran distribuida la educación en la ciudad de Cali, mostrando los clusters educativos y el nivel de autocorrelación entre los colegios, dado el lugar geográfico donde se encuentran. Además de los efectos que podrían traer las instituciones vecinas en el desempeño de una institución.

Este proyecto se encuentra dividido en cuatro secciones incluyendo esta que es la introducción. La sección que continua es el marco teórico donde se mostrará las bases de tipo conceptuales usadas como eje en la investigación. La siguiente sección será la de resultados donde se describirán los hallazgos obtenidos en el proceso investigativo. Y la última sección contendrá las respectivas conclusiones del proyecto.

Objetivos

Objetivo general

Evaluar el efecto que tiene la ubicación geográfica de las instituciones educativas de la ciudad de Cali en los resultados de las pruebas saber 11 para los años 2013 a 2017 empleando la metodología de aglomeración y la correlación espacial.

Objetivos específicos

1. Determinar la existencia de clusters educativos en la ciudad de Cali, que contribuyan a que los resultados de ciertas delimitaciones geográficas sean homogéneos.
2. Determinar la correlación entre la pertenencia de las instituciones educativas a clusters educativos y los resultados de las pruebas saber 11.

Marco teórico

La necesidad de dar explicación a ciertos efectos o condiciones en un espacio determinado ha dado lugar a estudios que respondan incógnitas de fenómenos de organización. Dentro de estos estudios se encuentra la variable geografía, debido a que la localización de un lugar puede afectar los rendimientos productivos de una comunidad. De esta manera, algunos métodos gráficos, parámetros muestrales e indicadores, ayudan a identificar estos patrones permitiendo dar explicación a estos fenómenos.

La autocorrelación espacial es la asociación determinada entre las observaciones de una sola variable (auto significada) estrictamente atribuible a la proximidad de esas observaciones en el espacio geográfico. Esta noción se resume mejor en la primera ley de Tobler que establece que "todo está relacionado con todo lo demás, pero las cosas cercanas están más relacionadas que las cosas distantes" (Tobler, 1970, citado en Waller & Gotway, 2004, p.3). Ahora bien, Tobler fue uno de los primeros en introducir el término de "autocorrelación espacial" y es mediante esta idea que se parte para explicar el fenómeno que se presenta en la ciudad de Cali (Colombia) con respecto a la educación, puesto que la localización de los colegios puede estar condicionada por factores de esparcimiento de conocimiento, afectando de esta manera, los rendimientos educativos.

Cabe resaltar que el tema de autocorrelación espacial es ampliamente usado para diferentes campos de estudio. Dentro del campo de los estudios económicos, se encuentra el análisis exploratorio de datos espaciales (AEDE) que agrupa técnicas que permiten visualizar distribuciones espaciales, localizar esquemas de asociación y sugerir heterogeneidad espacial como estructuras en el espacio (Hoef, 1993 citado en Corso, Pinilla, & Gallego, 2018). Estas

técnicas relacionan conceptos de interacción que, dado un grado de concentración o dispersión, permite obtener herramientas visuales y que en conjunto a un análisis estadístico se puede analizar el comportamiento de las observaciones, tendencias, aglomeraciones, puntos atípicos y estructuras espaciales mediante modelos de regresión introducidos por Anselin en 1999.

De acuerdo a Waller & Gotway (2004) existen tres componentes que deben tener los datos espaciales: características, soportes y atributos. Para este caso, se hará uso de la característica “punto” que se refiere a “una ubicación precisa en el espacio” (Waller & Gotway, 2004, p.38), que para efectos de la investigación será cada una de las instituciones educativas de la ciudad de Cali seleccionadas para el estudio. Cada característica tiene una orientación espacial específica, los cuales, tomados en conjunto, forman el soporte de los datos (Waller & Gotway, 2004). Es decir que el conjunto de “puntos”, en este caso el conjunto de colegios es el soporte de los datos. En cuanto a los atributos, “son observaciones o valores medidos asociados con características” (Waller & Gotway, 2004, p.39). Para esta investigación, los atributos principales serán los resultados del ICFES para los años 2013 a 2017, estos resultados permiten comprobar el grado de desarrollo de las competencias de los estudiantes que están por finalizar el grado 11 de la educación media (ICFES, 2019).

En el estudio de la autocorrelación espacial se puede hacer uso de diferentes indicadores con el fin de realizar el análisis exploratorio de los datos, el cual nos permite además de describir distribuciones espaciales a través de la formulación de modelos de regresión, también obtener estimaciones de indicadores, entre ellos, el indicador de Moran que fue introducido en 1950 y puede ser medido de manera local o global. “La base para las pruebas locales y las medidas de autocorrelación espacial proviene de la estadística de producto cruzado que permite

comparaciones de autocorrelación espacial para una observación dada” (Fischer & Wang, 2011, p.23).

Por su parte, el Índice de Moran Global permite establecer el grado de autocorrelación (positiva, negativa o neutra (ausencia de correlación)) de acuerdo a la distribución y concentración de los datos. La autocorrelación positiva “se observa cuando un evento que tiene lugar en una región condiciona o favorece que se produzca en otra región” (Barros & Aroza, 2014, p.3). En cuanto a la autocorrelación negativa “se produce cuando el evento de una región limita su aparición en otra” (Barros & Aroca, 2014, p.3).

Ahora bien, en términos de la presente investigación se espera que, si los puntajes son altos/bajos en una localización, sus vecinos próximos también obtengan puntajes altos/bajos lo que daría como resultado una autocorrelación positiva. Por el contrario, si los puntajes son altos/bajos en un lugar y sus vecinos presentan puntajes bajos/altos, la autocorrelación es negativa.

“En general se acepta que existe autocorrelación espacial (AE) siempre que haya una variación espacial sistemática en los valores de una variable a través de un mapa, es decir un patrón en el comportamiento de la variable según la ubicación geográfica del dato” (Celemín, 2009, p.13)

Este patrón de comportamiento de variación puede ser mejor observado agrupando las variables en un indicador global, como lo es el LISA (*Indicador local de asociación espacial*). Este indicador fue derivado por Anselin (1995), con la motivación de descomponer las estadísticas de autocorrelación espacial global, como el I de Moran y la C de Geary, en la

contribución de cada observación individual (Celemín, 2009). “El mapa de LISA permite representar aquellas localizaciones con valores significativos, en indicadores estadísticos de asociación espacial” (Getis y Ord, 1992; Anselin, 1995; Ord y Getis, 1995; Unwin, 1996, citado en Palacios & Moreno, 2015, p.46).

Adicionalmente, para el análisis que se desea hacer en la presente investigación, es necesario tener en cuenta algunos conceptos importantes, evidentemente tomando como referencia el campo de estudio de la economía. El primero de ellos es el termino cluster, el cual está relacionado con el análisis de aglomeraciones “que busca agrupar elementos o variables tratando de lograr la máxima homogeneidad en cada grupo y la mayor diferencia entre los grupos” (De la Fuente, 2011, p.1).

Además, la formación de grupos generaría el mismo tipo de economías de aglomeración que observamos en las ciudades modernas. “Los diferentes grupos podrían compartir los proveedores de insumos intermedios, y en él también facilitaría la innovación y el aprendizaje” (Arnott & McMillen, 2006, p.53). Esto es importante puesto que, al facilitar el aprendizaje, podría crear el efecto de contagio o desbordamiento (spillover), lo que implica que su presencia en una región podría causar extensión a regiones vecinas, favoreciendo la concentración del fenómeno en la zona (Palacios & Moreno, 2015).

Ahondando en el tema de la presente investigación, se debe decir que en Colombia la educación no resulta ser un tema especialmente importante, en esta medida, existe recurrente desinformación por parte de los planteles educativos hacia el gobierno, lo que a su vez causa que no se puedan generar indicadores que permitan que se tomen decisiones correctas para

implementar programas de calidad. Patrinos (2000, p.62) evidencia que “las inversiones en personas son críticas para el desarrollo económico y social de un país”.

Las personas invierten en su capital humano por varias razones, una de estas (quizás la más importante) es el beneficio monetario. Esto se debe a que las inversiones personales en educación y capacitación aumentan la productividad, de manera que la expectativa es a que exista una relación directa entre productividad y salario. Por ende, la educación toma un rol importante en la sociedad.

En ese sentido, como se mencionó anteriormente, las Pruebas Saber 11 juegan un papel importante en Colombia debido a que no solo son usadas para medir el rendimiento de los estudiantes, sino que también sirven para que las instituciones puedan tener una retroalimentación de su nivel de preparación de los estudiantes, y con ello su desempeño como entidad prestadora del servicio educativo. Las Pruebas Saber 11 históricamente han contado con diferentes cambios, pero el cambio más importante en los últimos años ha sido el que ocurrió en el periodo 2014-2. Para conocer mejor los cambios que se realizaron a las pruebas es necesario recurrir a algunos estudios que se realizaron sobre este tema.

“Entre las modificaciones que se hicieron a la Prueba Saber 11, se tiene que: 1) se introdujo la prueba de competencias ciudadanas, 2) en el área de matemáticas, se distinguió entre lo que es genérico y lo que no lo es, y 3) se fusionaron diferentes pruebas en torno a las competencias genéricas que evalúan en común; lenguaje y filosofía se unieron en una prueba de lectura crítica; física, química y biología en ciencias naturales...de esta manera, el número de competencias por evaluar pasaron de ocho a cinco” (Díaz & Tobar, 2016, p.13).

En términos de la presente investigación se tomó como base el año 2013 donde todavía no se había realizado los cambios. Además de los años posteriores, para analizar el efecto que se dio en los planteles educativos con relación a su georreferenciación.

Metodología

Para la evaluación de la influencia de la ubicación geográfica en los resultados de las pruebas saber 11 en la ciudad de Cali se hará uso de la econometría espacial. Puntualmente se usarán las herramientas del Índice de Moran Global y el mapa de Lisa (Local Indicator of Spatial Association) a los cuales se hizo referencia anteriormente. Además de un modelo de retardo espacial.

De acuerdo a Barros & Aroca (2014), el Índice de Moran usa la siguiente formula:

$$I = \frac{N \sum_{(2)} W_{ij} (y_i - \bar{y})(y_j - \bar{y})}{S_0 \sum_{i=1}^N (y_i - \bar{y})^2}$$

Donde:

W_{ij} : Es el elemento de la matriz de pesos espaciales correspondiente al par (i, j)

y : Es el valor medio o esperado de la variable y .

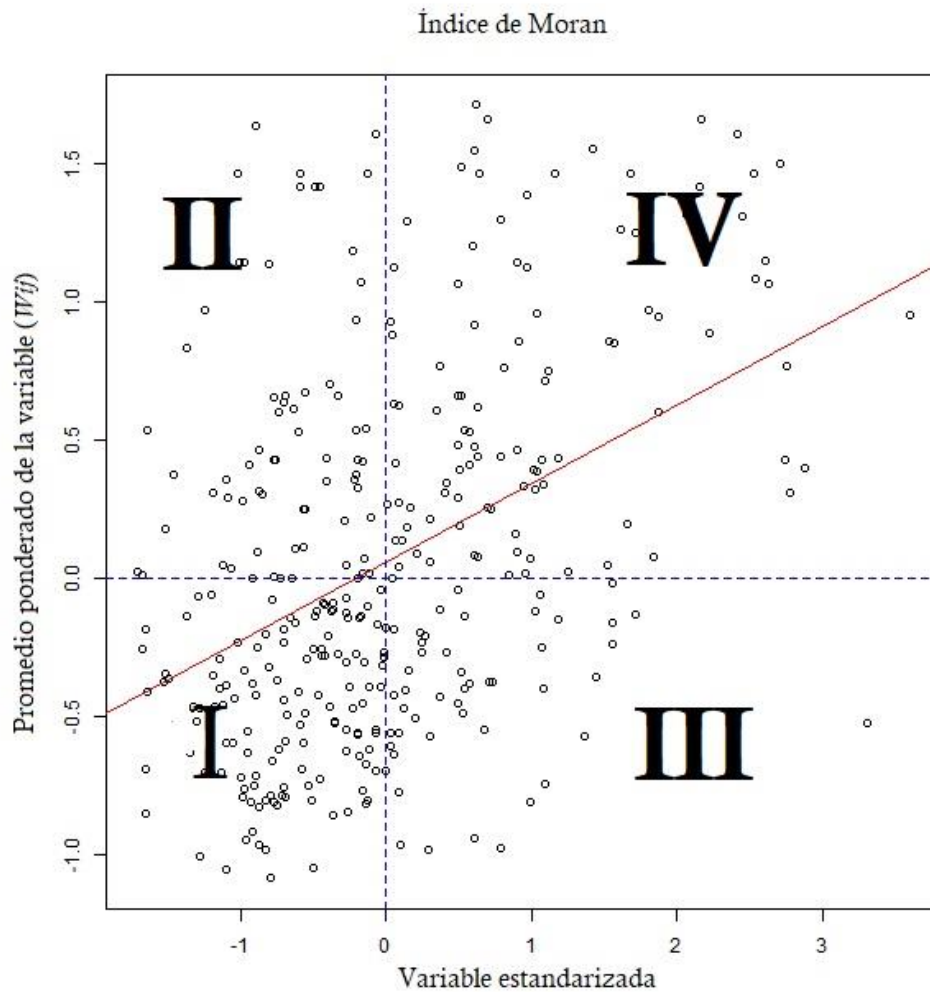
N : Es el Número de observaciones o tamaño muestral.

$$S_0 = \sum_i \sum_j W_{ij} = \sum_{(2)} W_{ij}$$

El test Global de Moran se basa en la pendiente de la recta. De esta manera, “mientras mayor sea el ángulo que forma con el eje de las abscisas, ..., mayor es el grado de autocorrelación” (Barros & Aroca, 2014, p.3).

Además, es necesario aclarar la interpretación del gráfico del Índice de Moran Global, que se presenta a continuación:

Ilustración 1 GRÁFICO ÍNDICE DE MORAN



Fuente: elaboración propia

El primer y el cuarto cuadrante muestran la presencia de autocorrelación positiva, lo que “indica la existencia de valores altos rodeados de valores altos, o valores bajos rodeados de valores bajos” (Barros & Aroca, 2014, p.4). Por su parte, el segundo y tercer cuadrante indican autocorrelación espacial negativa, lo que “significa la presencia de valores bajos de la variable rodeadas de valores altos o viceversa”. (Barros & Aroca, 2014, p.4).

Por otro lado, el mapa LISA “se basa en el estadístico I de Moran de asociación local y adquiere un valor diferente para todas y cada una de las observaciones..., es así como este estadístico mide el grado de concentración de valores alto/bajos de una variable en el entorno geográfico de cada una de las observaciones de la muestra..., para cada valor es posible realizar una inferencia para evaluar el nivel de significatividad estadística” (Chasco, 2001, p.17). Lo que quiere decir que es posible determinar la concentración de valores con ciertas características y ubicarlo en un punto en el espacio.

El cálculo de este indicador y del mapa de LISA se hace a través del código de programación de R-Studio, que básicamente es un lenguaje de computadora que le permite al usuario programar algoritmos y usar herramientas programadas por otros para realizar cálculos y técnicas estadísticas, además de gráficos (Zuur, Ieno, & Meesters, 2009). Previo a esto es necesario contar con la base de datos que se requiere para el análisis. En ese sentido, se procederá a describir la manera en la cual fue obtenida la base de datos, ya que toma un papel especialmente relevante en la investigación.

En primera instancia, fue necesario la recolección de los datos dado que no existía un documento que tuviese consolidado los resultados de todos los años, por lo cual se debió hacer la unión de los datos en una misma base. Para la construcción de la base de datos se obtuvieron los resultados históricos de las Pruebas Saber 11, que se encuentran disponibles en la página web del ICFES¹. Cabe aclarar, que solo se pudieron obtener los resultados de los años 2011 a 2018-1, debido a que la página web no suministraba datos anteriores, ni posteriores a estos hasta ese momento.

¹ Los datos obtenidos de las pruebas saber 11 se encuentran disponibles en la página web: www.icfes.gov.co

La construcción de la base de datos requirió que para los años que van hasta el 2014-1 se realizara un re-escalamiento en los puntajes obtenidos por las instituciones debido a que para ese entonces los resultados de las Pruebas Saber 11 contaban con un formato en el cual se evaluaban 8 componentes, en cambio, desde el 2014-2 hasta la actualidad se evalúan solo 5 de estos. A pesar de que para los años 2014-2 y 2015 los componentes iban acompañados de 2 subpruebas, no se tenía en cuenta el resultado de estas últimas para obtener el puntaje global.

Es así como se obtuvo una base de datos de aproximadamente 500 observaciones en las cuales había gran cantidad de datos faltantes por muchos factores, entre ellos: algunos colegios que presentaban información para un año no presentaban información para los siguientes dos y reaparecían con información para el tercer año; o se daban casos en los que algunos colegios fueron creados después del año que se había tomado como referencia, de manera que no se encontraba la información completa.

Dada la disponibilidad de los datos y que los principales faltantes se encontraban antes del año 2013, se decidió delimitar la base de datos entre los años 2013 y 2017, además, se decidió eliminar de la base de datos aquellos colegios que no presentaran información para alguno de los años en cuestión. Al final se obtuvo una base de datos con 347 observaciones.

Posteriormente, fue necesario obtener las direcciones y coordenadas geográficas de cada una de las instituciones del municipio que contaba con resultados en las Pruebas Saber 11 para la franja de años anteriormente establecida y así imputar esta información a cada uno de los colegios. Cabe resaltar, que las coordenadas geográficas se obtuvieron bajo el sistema de coordenadas World Geodetic System - 1984 (WGS-84) que es “un Sistema Terrestre

Convencional realizado mediante la modificación del Sistema de Satélites de Navegación Marina (NNSS), ... girándolo para que su meridiano de referencia coincida con el meridiano cero definidos por la Oficina Internacional de Salud (BIH)” (ICAO, 2002, p.21). Una vez obtenida la base de datos, se procedió a usar el lenguaje de programación de R-Studio para realizar el correspondiente análisis de datos.

Por último, otro de los métodos empleados para realizar el análisis de auto correlación espacial fue el modelo de regresión. Para empezar, es fundamental describir un modelo de regresión lineal (múltiple) debido a que este es el punto de partida para explicar el modelo de regresión de retardo espacial. De acuerdo a Borrego (2018), la forma funcional es la siguiente:

$$Y = \sum_{q=1}^Q X_q B_q + \varepsilon$$

Y su versión muestral:

$$y_i = \sum_{q=1}^Q x_{iq} \beta_q + \varepsilon_i \quad i = 1, \dots, n$$

Donde y_i es una observación de la variable dependiente o de interés, que en este caso es el resultado total para los años 2014, 2015, 2016 y 2017 y x_{iq} es cada una de las variables explicativas, con $q = 1, 2 \dots Q$ (Borrego, 2018) el cual tendrá como base el resultado total del año 2013, que es el año donde aún no se había realizado el cambio más significativo en las Pruebas Saber 11, además del factor tipo vec que fue calculado mediante el indicador de moran de cada uno de los años y es dependiente de la ubicación, es decir, del cuadrante en el cual se encuentra la observación. Por último, se tiene que “ B_q es el coeficiente que mide la influencia

por sí sola de la q-variable explicativa en la variable dependiente, es decir, mide el cambio en Y por cada cambio unitario en X_q manteniendo las demás variables constantes. El término ε_i es el error aleatorio, que incluye variables no controladas y la variabilidad muestral” (Borrego, 2018).

El modelo de regresión de retardo espacial presenta correlación (dependencia) en la variable dependiente Y . La notación matricial se muestra a continuación (Borrego, 2018):

$$Y = \rho WY + X\beta + \varepsilon$$

“ WY se conoce como una variable dependiente espacialmente retardada” (Borrego, 2018, p.37). Para la investigación se usó una representación dispersa de una matriz de ponderaciones espaciales lo que permitió calcular el vector de retardo, esto se hizo para cada una de las variables por año. Al final de la regresión se sumó el factor tipo vec, y finalmente se obtuvo el resultado de la regresión.

Resultados

Mediante el Índice de Moran aplicado al promedio total de los resultados de las Pruebas Saber 11 por plantel educativo para los años 2013 a 2017, se pudo observar que los resultados del test fueron consistentes, significativos y positivos para todos los años evaluados y se comprueba la existencia de autocorrelación espacial global puesto que se rechaza la hipótesis nula de inexistencia de autocorrelación en el espacio para todos los casos. En la *tabla 1* se puede observar que para el año 2013 el indicador es $I=0,2656$ con un P.value: 0,00. Esto indica que se rechaza la ausencia de autocorrelación.

Tabla 1 ÍNDICE DE MORAN TOTAL 2013

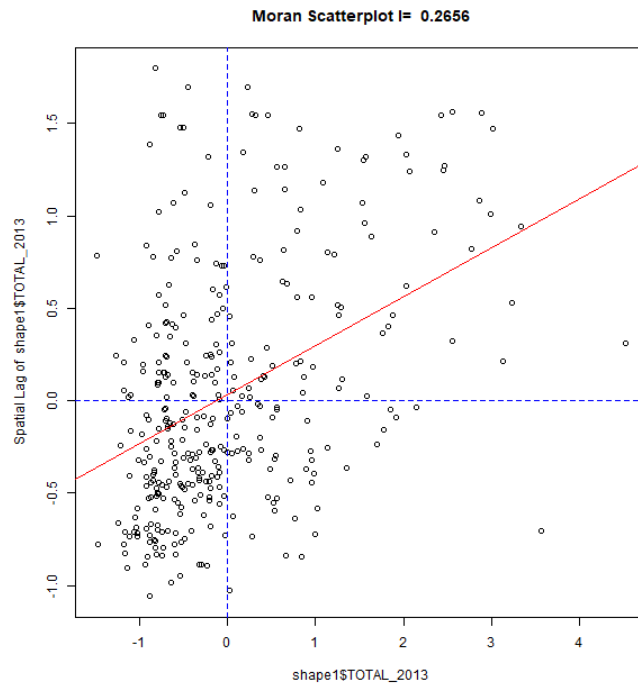
	Mora n I 2013	Mora n I 2014	Mora n I 2015	Mora n I 2016	Mora n I 2017
Index	0,265 6	0,235 5	0,308 7	0,314 8	0,288 4
P.valu e	0.00	0.00	0.00	0.00	0.00
Mora n I statistic standard deviate	8.589	7.626	9.953	10.15 1	9.157

Fuente: elaboración propia

Otra consideración relevante es que el mayor índice se da en el año 2016 con un valor de 0,3148, una posible explicación de este resultado es que se intuye la presencia de convergencia en los resultados, dado que para un primer momento la dispersión de los datos no fue tan alta, pero posteriormente los colegios con mejores resultados reaccionaron ante los cambios realizados por el ICFES y volvieron a alejarse con respecto a los que no son tan buenos. En la *Ilustración 1 y 2* se muestra el gráfico de autocorrelación de Moran respecto a la matriz de pesos espaciales del criterio de contigüidad de vecinos relativos para el 2013 y 2017, dando como resultado una dependencia espacial positiva que indica que los colegios con

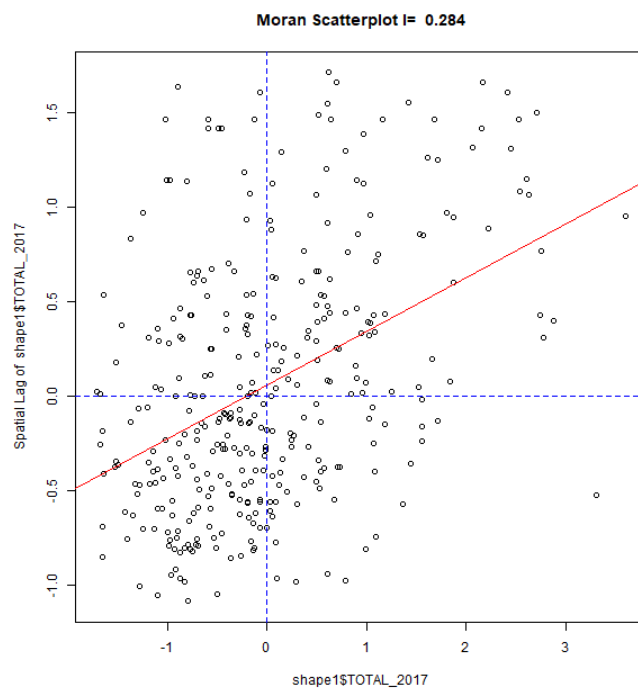
bajo/alto rendimiento académico en las Pruebas Saber 11 están rodeados de otros colegios con la misma condición.

Ilustración 2 GRÁFICO ÍNDICE DE MORAN TOTAL 2013



Fuente: elaboración propia

Ilustración 3 GRÁFICO ÍNDICE DE MORAN TOTAL 2017



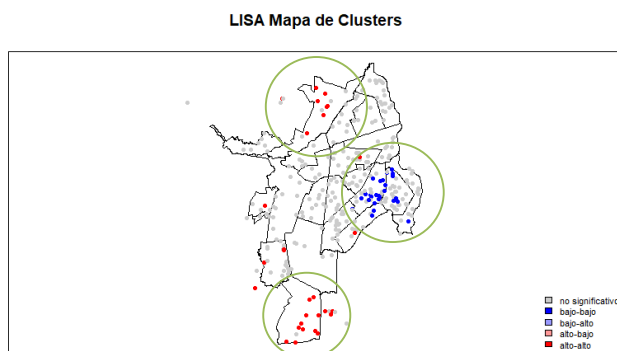
Fuente: elaboración propia

Además, para el año 2013 se tiene que el 43,22% de las observaciones se encuentran en el cuadrante I (bajo-bajo), siendo este el mayor porcentaje en todos los años de análisis. Por otro lado, se obtuvo que para el 2017 hubo un aumento en el número de colegios en el cuadrante IV (alto-alto) con un porcentaje de 26,5%, que también el mayor con respecto a todos los años, lo que refuerza la idea de que los colegios que habían tenido bajas en sus rendimientos académicos dado el cambio en la estructura de la prueba presentado en el 2014-2, volvieron a reforzar su aprendizaje y de nuevo se separaron de los colegios con bajo rendimiento.

Por otro lado, después de la detección de autocorrelación espacial se empleó el test de Moran local, el cual se obtiene del test global. A partir de este indicador se pudo obtener los mapas locales, donde “se representan las localizaciones con valores significativos, por lo que se pudo detectar la existencia de puntos atípicos y formación de clusters”. (Barros & Aroca, 2014, p.4).

En ese sentido, se puede evidenciar la existencia de tres clusters educativos principales en la ciudad de Cali para el año 2013 y los cuales se mantienen aproximadamente constantes durante los años subsiguientes, estas aglomeraciones serán representados por medio del mapa LISA (Ver *Ilustración 4*) y estarán clasificados de la siguiente manera:

Ilustración 4 LISA TOTAL 2013



Fuente: elaboración propia

Clúster 1: Este clúster se encuentra en la parte Nor-occidental de la ciudad de Cali y comprende la comuna 2, se caracteriza por que sus valores más altos (color rojo) pertenecen al grupo de alto-alto (A-A) lo que quiere decir que en este sector los colegios con alto rendimiento en las Pruebas Saber 11 están rodeados de otros colegios con alto rendimiento.

Clúster 2: Se ubica en la parte del distrito de Agua Blanca y comprende las comunas 13,14,21 y una pequeña parte de la comuna 15, se caracteriza por sus valores bajos (color azul) y pertenecen al grupo bajo-bajo (B-B) lo que quiere decir que en esta zona existen colegios con bajo rendimiento en las Pruebas Saber 11 que están rodeados por otros colegios con rendimiento similares. Como característica especial en este sector se puede tomar en cuenta la variable socioeconómica de los barrios que conforman las comunas, de acuerdo a un estudio realizado por el Observatorio de Políticas Públicas de la universidad Icesi, las personas perciben que la seguridad en estas comunas es baja, lo que podría estar asociados a variables socioeconómicas inestables (Mera, 2018).

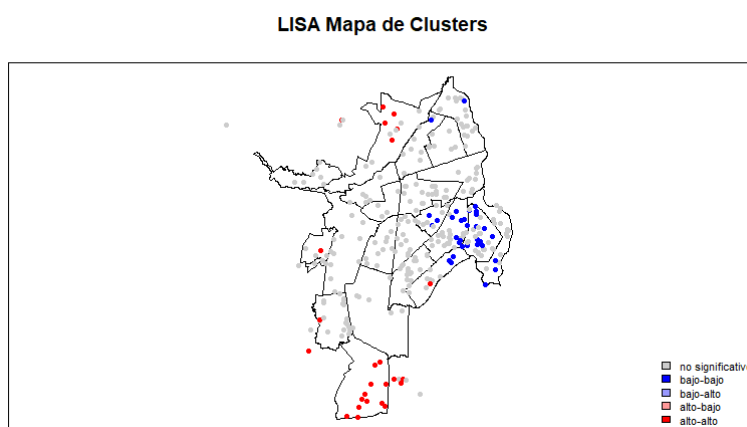
Clúster 3: Esta aglomeración se ubica en el Sur de la ciudad, principalmente en la comuna 22 y se caracteriza de igual forma que el primer clúster anteriormente descrito. De este clúster se puede decir, que en esta comuna se encuentran barrios de estratos altos, lo que podría ser una variable por considerar, dadas las condiciones académicas de los planteles educativos, sin embargo, en la presente investigación no es significativo puesto que esta variable no está considerada dentro del modelo.

Cabe mencionar la existen de algunos puntos atípicos representados en el mapa, estos puntos se encuentran en las comunas 8, 18 y 20 y comprenden barrios como Siloé, Villacolombia y el sector de Meléndez respectivamente. En el marco de esta investigación no

es posible identificar las causas de estos atípicos debido a que, para poder hacerlo, se debe tener en cuenta otras variables no consideradas en el estudio.

Adicionalmente, en la *Ilustración 5* se muestra el mapa LISA para el año 2017, en el cual se mantienen los 3 clusters descritos anteriormente y adicional se observan diferencias: en primera instancia, se observan nuevos puntos con valores A-A en las comunas 4 y 6, además, el punto presente en el 2013 en comuna 8 ya no es significativo y el clúster del distrito de Agua Blanca se extiende hasta la comuna 12.

Ilustración 5 LISA TOTAL 2017



Fuente: elaboración propia

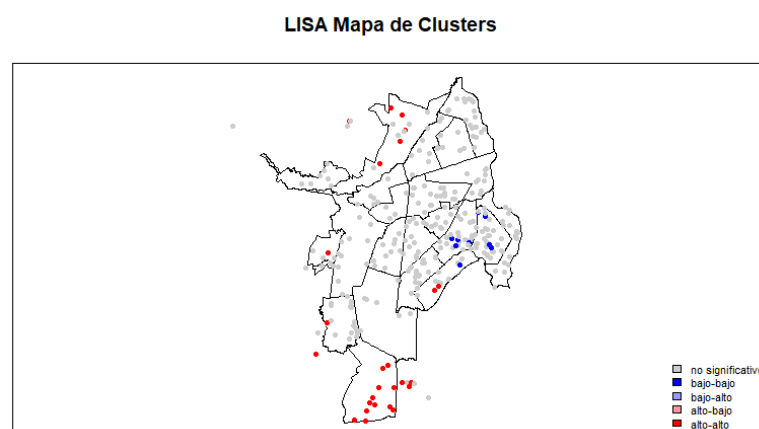
Por otro lado, utilizando el Índice de Moran para analizar los resultados del componente de inglés se obtuvo que para todos los años los resultados son significativos, positivos y sus valores oscilan en un rango de 0,23 y 0,33 aproximadamente (Ver *tabla 2*) Para el 2014 se obtuvo el indicador más bajo, lo que podría estar asociado a efectos provocados por el cambio en las pruebas. En cuanto al mapa Lisa (Ver *Ilustración 6*) indica la conformación de clusters similares a los obtenidos para los resultados promedios totales de la prueba.

Tabla 2 ÍNDICE DE MORAN INGLÉS 2013

	Mora n I 2013	Mora n I 2014	Mora n I 2015	Mora n I 2016	Mora n I 2017
Index	0,318 5	0,236 3	0,323 6	0,339 5	0,327 6
p.valu e	0.00	0.00	0.00	0.00	0.00
Mora n I statistic standard deviate	10.30 0	7.677	10.45 7	10.95 3	10.55 8

Fuente: elaboración propia

Ilustración 6 LISA INGLÉS 2013



Fuente: elaboración propia

Por último, se hizo una regresión de retardo espacial en la cual se determinó la covariación de los resultados globales con respecto al año 2013, tomándose esta como variable X y como variable Y los demás años (2014, 2015, 2016 y 2017), además de tomar como referencia el cuadrante del Índice de Moran donde se encuentra. Lo anterior para mostrar el grado de reacción de los colegios ante el cambio para el año donde se realizó y los posteriores a este (Ver *tabla 3*) Se obtuvo que para el 2014 los coeficientes son negativos puesto que fue

el año donde se realizaron los cambios y es muy probable que los colegios no hayan tenido oportunidad de reaccionar, mientras que para el resto de años es positivo, lo que quiere decir que los colegios probablemente tomaron medidas de preparación para las pruebas.

Tabla 3 REGRESIÓN (AÑO BASE 2013)

	<i>Dependent variable:</i>			
	TOTAL_2014 (1)	TOTAL_2015 (2)	TOTAL_2016 (3)	TOTAL_2017 (4)
TOTAL_2013	0.954*** (0.036)	1.036*** (0.033)	0.704*** (0.026)	0.661*** (0.027)
TipoVec)2	-2.011 (2.632)	4.778* (2.462)	3.041 (1.886)	2.524 (1.988)
TipoVec)3	-6.521* (3.701)	3.570 (3.462)	1.240 (2.652)	3.783 (2.796)
TipoVec)4	-2.543 (4.124)	6.937* (3.858)	7.416** (2.955)	6.262** (3.116)
Constant	17.788 (12.133)	-16.085 (11.351)	10.057 (8.695)	19.717** (9.167)
Observations	347	347	347	347
R ²	0.852	0.898	0.878	0.851
Adjusted R ²	0.851	0.897	0.877	0.850
Residual Std. Error (df = 342)	18.342	17.160	13.144	13.858

Note:

* p<0.1; ** p<0.05; *** p<0.01

Fuente: elaboración propia

Conclusiones

Teniendo en cuenta los análisis realizados en el presente proyecto de investigación, se logró determinar con suficiente evidencia estadística que en la ciudad de Cali se presentan clusters educativos que afectan los resultados de las Pruebas Saber 11. Lo que indica que en diferentes sectores de la ciudad se encuentran conglomeraciones geográficas donde los resultados demuestran que la educación parece ser homogénea entre los integrantes de los clusters y heterogénea entre los mismos.

Esto se pudo evidenciar gracias a la existencia de autocorrelación positiva, es decir que los colegios con rendimiento académico alto/bajo tienen vecinos con rendimientos alto/bajo. Entre los diferentes clusters que se encuentran en la ciudad de Cali, se pudo tres principales, que comparten la característica de que el comportamiento de los resultados de las pruebas en algunos colegios está siendo favorecido por el comportamiento de los resultados de otros colegios próximos geográficamente. Además, los clusters se mantiene aproximadamente constantes para los años de estudio, lo que quiere decir que no se da de manera estocástica.

El primer y tercer cluster se encuentran respectivamente en la comuna 2 y 22 del municipio. Ambos se caracterizan porque los colegios pertenecientes tienen resultados altos y se encuentran rodeados de más colegios con resultados altos. El segundo cluster se encuentra entre las comunas 13, 14 y parte de la comuna 15 y 21, y se caracteriza porque los resultados de los colegios son bajos y se encuentran rodeados de otros colegios con resultados bajos.

Por otro lado, la evaluación del Índice de Moran permite evidenciar la existencia de una convergencia en los resultados, de manera que se detecta una curva de aprendizaje que va desde

el año 2014-2 que es el año en el cual se realiza el cambio más significativo en la evaluación, hasta la actualidad. Lo que quiere decir que, pese a que inmediatamente después del cambio la diferencia entre los colegios no resultaba ser tan amplia, paulatinamente los colegios con mejores rendimientos históricos en las pruebas se han dispersado de los otros.

La presente investigación puede ser utilizada como herramienta para tomar decisiones importantes acerca del gasto en educación en el municipio, ya que permite conocer los lugares en los cuales es especialmente necesaria la intervención estatal. Evidentemente este estudio no toma en cuenta todas las variables que podrían llegar a tenerse en cuenta, por lo tanto, solo presenta un análisis de muchos posibles.

Referencias bibliográficas

- Apache Roa, C., Vinasco, M., & Rojas, Á. (2018). *Eficiencia del gasto público en educación en Colombia durante el 2017 desde las Pruebas Saber 3, 5 y 9* (tesis de pregrado) Universidad Icesi, Cali, Colombia.
- Arnott, R. J., & McMillen, D. P. (2006). A Companion to Urban Economics. En *A Companion to Urban Economics*. Chichester, United Kingdom:Blackwell Publishing
- Barros Díaz, O., & Aroca González, P. (2014). Econometría espacial y el análisis sociodemográfico. Aplicación en la formación de agrupaciones espaciales de envejecimiento en Cuba, período 2003-2009. *Novedades en población*, (10), p.10.
- Borrego Sánchez, J. (2018). *Modelos de Regresión para Datos Espaciales* (tesis de pregrado). Universidad de Sevilla, Sevilla, España.
- Celemín, J. P. (2009). Autocorrelación espacial e indicadores locales de asociación espacial. Importancia, estructura y aplicación. *Revista universitaria de geografía*, (18), pp.11-31.
- Chasco, C. I. L. R. K. (2001). Análisis exploratorio de datos espaciales al servicio del geomarketing. En *III Seminario sobre Nuevas Tecnologías en la Investigación, el Marketing y la Comunicación. e-Participación*.
- Corso, G., Pinilla Rivera, M., & Gallego, J. (2018). Métodos gráficos de análisis exploratorios de datos espaciales con variables espacialmente distribuidas. *Cuaderno Latinoamericano de Administración*, (XIII), p.p.92-104.
- De la Fuente Fernández, S. (2011). Análisis De Conglomerados. Universidad Autónoma de Madrid, 1(1), 55. Obtenido de: <http://www.fuenterrebollo.com/Economicas/ECONOMETRIA/SEGMENTACION/CONGLOMERADOS/conglomerados.pdf>

- Departamento Nacional de Planeación. (2019). *DNP*. Obtenido de Departamento Nacional de Planeación: <https://www.dnp.gov.co/programas/desarrollo-social/subdireccion-de-educacion/Paginas/subdireccion-de-educacion.aspx>
- Díaz Rosero, M., & Tobar Bedoya, J. (2016). *Causas de las diferencias en desempeño escolar entre colegios públicos y privados: Colombia en las Pruebas Saber 11 2014* (tesis de pregrado) Pontificia Universidad Javeriana, Bogotá, Colombia.
- Fischer, M., & Wang, J. (2011). *Spatial Data Analysis. Models, Methods and Techniques*. New York, EE.UU.: Springer.
- ICAO (2002). *World Geodetic System -1984 (WGS-84) Manual*. International Civil Aviation Organization.
- ICFES. (2019). Estructura general del examen Icfes Saber Pro - Portal Icfes. Obtenido de: <http://www.icfes.gov.co/web/guest/acerca-examen-saber-once>
- Mera, A. (19 de diciembre de 2018). Inseguridad en Cali, ¿es una realidad o es solo percepción? *El Tiempo*. Obtenido de: <https://www.elpais.com.co/judicial/inseguridad-en-cali-una-realidad-o-es-solo-percepcion.html>
- Ortegón R., F. (30 de 07 de 2018). Una buena nota en educación superior. *El Tiempo*. Obtenido de: <https://www.eltiempo.com/colombia/cali/una-buena-nota-en-educacion-superior-248768>
- Palacios, Y., & Moreno, N., (2015). *Análisis de autocorrelación espacial de datos de área del virus del chikunguya en la provincia del Alto Magdalena departamento de Cundinamarca* (tesis de pregrado) Universidad de Cundinamarca, Fusagasugá, Cundinamarca, Colombia.
- Patrinos, H., (2000). Market Forces in Education. *European Journal of Education*. (35), p.p.61-80.
- Salas, M (2008). *Economía de la educación. Aspectos teóricos y actividades prácticas*. Madrid, España: Pearson.
- Waller, L., & Gotway, C. (2004). *Applied Spatial Statistics for Public Health Data*. Hoboken, New Jersey, EE.UU.: John Wiley sons, inc.

Zuur, A., Ieno, E., & Meesters, E. (2009). *A Beginner's Guide to R*. London, New York: Springer.