

Patrones espaciales del cambio demográfico en las escuelas públicas de Costa Rica: 2010-2018

Katherine Barquero Mejías¹

M^aestría en Estadística, Universidad de Costa Rica. 2060 San José, Costa Rica.

Resumen

La investigación analiza si existen patrones espaciales en el comportamiento de la disminución de la matrícula que ha venido experimentando la educación primaria durante la última década. Se hace uso de dos técnicas de estadística geoespacial: Procesos puntuales y Estadística de Áreas. Los resultados obtenidos a partir de la primera técnica sugieren que los eventos, es decir las escuelas que ha sufridos disminuciones en la matrícula, no tienen un comportamiento aleatorio en el espacio y que la intensidad de los eventos se concentra principalmente en el Gran Área Metropolitana. En el segundo caso, el análisis se realiza a partir de la disminución de la matrícula promedio en los distritos de Costa Rica, donde se corrobora la existencia de autocorrelación espacial y se identifican además los distritos de influencia alta como en el caso de San Carlos.

Palabras clave: Patrones espaciales, autocorrelación espacial, matrícula, cambio demográfico, primaria

Introducción

En Costa Rica, según datos del Departamento de Análisis Estadístico del Ministerio de Educación Pública (MEP) durante el periodo 2010-2018 el 65 % de las escuelas públicas del país experimentó disminuciones en la matrícula. Del total de 3.692 escuelas, una tercera parte tuvo reducciones superiores al 25 %.

Aunado a este fenómeno, según el Centro Centroamericano de Población (CCP), en 2018 la población con edad de asistir a la educación primaria, es decir cuyo rango de edad se encuentra entre 6 a 11 años, registró 57.000 niños (as) menos que en el 2000. Las proyecciones indican que el descenso prolongado se empezará a experimentar a partir del 2024, lo que podrá ocasionar cambios importantes en la estructura actual de las escuelas del país: cierre de alguna de ellas, traslados, variaciones en el personal, etc.

Este panorama lejos de ser una amenaza debe ser visto como una oportunidad para mejorar la calidad de la enseñanza que actualmente se brinda en este nivel. Actualmente, el 93 % de las escuelas no ofrece el currículo completo y funciona bajo la modalidad de horario alterno, se espera que la universalización se alcance hasta el 2034, si se continúa con el mismo ritmo de crecimiento y si se supone que no se crean otras escuelas. Por otra parte, cerca del 40 % del total de centros de primaria en el país funcionan bajo una modalidad unidocente, y además, el aprendizaje de los estudiantes al culminar la primaria es deficiente, según la prueba Terce de la Unesco, el 50 % de los estudiantes se ubicaron en los niveles de desempeño más bajos en materia de comprensión lectora. Por esta razón, es necesario analizar cuál es el comportamiento especial de la matrícula inicial en las escuelas públicas del país, a fin de analizar si es posible realizar cambios en su estructura que ofrezcan la posibilidad de mejorar la calidad de la educación primaria.

Dado lo anterior, esta investigación hace uso de técnicas en estadística geoespacial para identificar si existen patrones espaciales en la disminución de la matrícula de las escuelas públicas de Costa Rica que permiten elaborar estrategias de intervención de política educativa diferenciada para mejorar la enseñanza de la calidad educativa de la primaria.

Metodología

Este estudio utiliza la base de datos georreferenciada de las escuelas públicas de Costa Rica, específicamente para el periodo 2010-2018 del Programa Estado de la Nación la cual incluye información de la matrícula inicial, es decir, la matrícula reportada por todas las

escuelas públicas del país cuando inicia el ciclo lectivo (regularmente en febrero de cada año). Se utiliza como principal variable la disminución de la matrícula estimada como la diferencia entre la matrícula inicial del 2019 con respecto a la matrícula inicial del 2010.

Para desarrollar el objetivo de investigación se aplicarán dos técnicas de análisis de datos geoespaciales: procesos puntuales y estadísticas de áreas. El método de procesos puntuales permite determinar si los puntos, o unidades geoestadísticas de análisis, presentan o no un patrón sistemático como agrupaciones o un patrón regular. Las variables puntuales con las que se cuenta son la matrícula al inicio de cada ciclo lectivo y las coordenadas geográficas de cada escuela vigente en el periodo de análisis.

Además, en este caso se entiende el proceso puntual como aquel en donde se registró una disminución de la matrícula superior al 15 % entre el año 2010 y 2018. Este análisis se desarrollará la siguiente forma: se preparan y manipulan los datos, se estiman las estadísticas descriptivas de los eventos a partir de las propiedades de según orden pues están relacionadas están relacionadas con la dependencia espacial e implican relacionar pares de eventos con respecto al número de eventos que suceden entre ellos. Los métodos que se emplearán son Vecinos más cercanos y la función K y se identificará cuál es la distribución que presentan los eventos de análisis a partir de la función G, y se determinará si pueden haber agrupaciones en los datos o si responde a un comportamiento aleatorio. Además, se examinará si se presentan puntos de calor que representen mayores disminuciones en la matrícula.

Una vez que se han analizado las escuelas y sus locaciones como unidades geoestadísticas, se procede a emplear el método de estadística de áreas, el cual pretende probar si el fenómeno de estudio presenta alguna relación de dependencia considerando regiones y vecinos. Para el empleo de este método se utiliza la cantidad de escuelas agrupadas por distrito y que presentaron de escuelas agrupadas por distrito y que presentaron una disminución en la matrícula en el periodo de análisis definido para esta investigación. En este caso, se evalúan distintos métodos para estimar los vecinos y pesos agregados a nivel distrito, posteriormente se utiliza la I-Moran para

detectar si hay existe autocorrelación espacial y finalmente, a partir de test locales se detecta si hay agrupaciones del fenómeno de análisis.

El análisis de datos se realizó con el software estadístico R[1], y utilizando las librerías *dplyr*[2], *ggplot2*[3], *haven*[4], *spatstat*[5–7], *raster*[8], *rgdal*[9], *spdep*[10, 11] y *RColorBrewer*[12].

Discusión de resultados

Durante el periodo comprendido entre el 2010 y el 2018, el 65 % de las escuelas públicas del país reportó disminuciones en la matrícula. Los estadísticas descriptivas de este fenómeno revelan que no sucede de manera homogénea en el territorio nacional, según se muestra en la figura 1, si bien en prácticamente todas las direcciones regionales educativas se presentaron disminuciones de la matrícula (a excepción de Penninsular, San Carlos, Aguirre y Santa Cruz que si reportaron algún incremento), en Desamparrados, San José, Cartago, Limón y Heredia se registraron disminuciones superiores a 3 mil estudiantes. [13]. De acuerdo con las proyecciones de población estimadas por el Centro Centroamericano de Población (CCP) entre el 2020 y el 2025 la población con edad de asistir a la educación primaria disminuirá a un ritmo anual promedio del -0.2 %, sin embargo durante el 2020 y el 2022 en promedio crecerá un 0.4 % pero en adelante presentará una tasa de decrecimiento más acelerada que ronda el -1.5 % en promedio. En el caso de la población entre 10 y 14 años más bien se espera un crecimiento anual promedio del 0.1 %, sin embargo en este grupo es importante considerar que la edad de asistir a la primaria es hasta los 12 años, debido a la agregación de las estadísticas queda pendiente mayor investigación para saber si es el rango de 10 a 12 años quien o de 12 a 14 el que está influenciando a una tasa positiva de crecimiento.

Un hecho que si llama la atención es las proyecciones para el grupo entre los 0 y 4 años de edad, es decir, la población que se espera que ingresa a la primaria en los próximos años, contrario a los grupos anteriormente descritos, en particular se proyecta un decrecimiento todos los años que ronda entre el -0.6 % y el -0.9 % desde el 2020 y hasta el 2025, esto plantea retos sobre la matrícula esperada pa-

ra los próximos años y llama la atención en profundizar donde se están experimentando para tomar medidas de acción que ante menor población en el sistema educativo permita mejorar la calidad de la enseñanza y sentar bases educativas sólidas para enfrentar los ciclos siguientes.

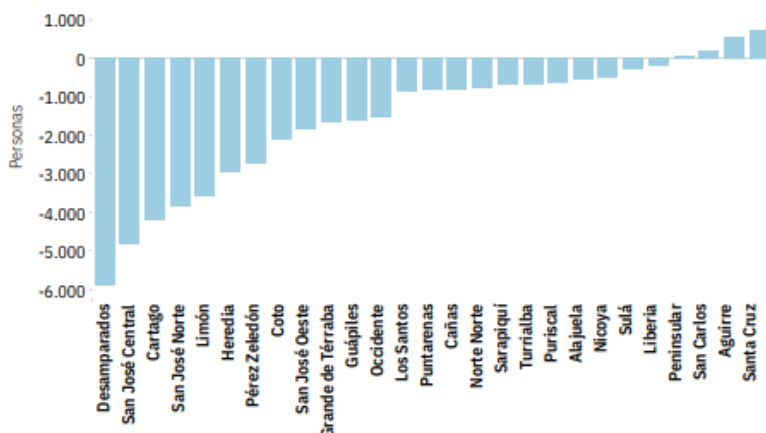


FIGURA 1: *Diferencia en la matrícula inicial reportada en 2018 y 2010, según Dirección Regional Educativa.*

A partir del análisis de procesos puntuales se procede a estimar la intensidad de los eventos, es decir, escuelas que presentan disminuciones en su matrícula en el territorio nacional, a fin de determinar si el proceso de estudio es uniforme o homogéneo. Los resultados obtenidos se presentan en la figura 2, el color rojo indica las provincias donde se presentan las intensidades más altas, estas corresponden a San José, Puntarenas y Alajuela. Es importante notar que estos resultados son consistentes a los mostrados en la figura 1, sin embargo en esta se presentaban los datos por Dirección Regional Educativa no por provincia, y esta división administrativa que emplea el MEP desagrega la provincia de Alajuela en otras direcciones como la de Occidente que también pertenece a esta provincia, por eso de manera agregada la intensidad es mayor que los valores absolutos presentados en ella. Los resultados obtenidos por lo tanto reflejan que el proceso no es homogéneo, en la provincia de San José

los cantones con mayor intensidad son Montes de Oca, particularmente el distrito de San Pedro, en Desamparados el distrito de San Antonio y en Pérez Zeledón el distrito de San Isidro de El General, En el caso de Puntarenas la mayoría se presenta en el cantón de Buenos Aires, y en el caso de Alajuela en el cantón de San Carlos.

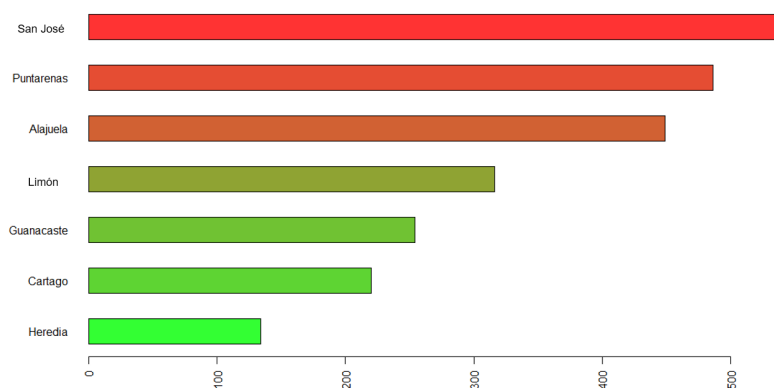


FIGURA 2: *Intensidad de los eventos según provincia.*

Por su parte la densidad se estimó utilizando la técnica de suavizamiento de Kernel que calcula la intensidad de forma continua en toda el área de estudio. Para esto se define el ancho de banda de la estimación de la densidad que determina el área de influencia, los resultados se muestrab en la figura 3, según se observa si se utiliza un anchos de banda de 11291.92 se identifica una mayor densidad en el zona del Gran Área Metropolitana (GAM) y en el cantón de Limón particularmente en el área de Talamanca. Es importante mencionar que se utilizaron varios anchos de bandas pero todos los resultados fueron consistentes.

Finalmente, para completar el análisis de procesos puntuales se estimó la función G para determinar si el patrón de los eventos presenta un comportamiento aleatorio, es decir, si los eventos implicados en este proceso tienen la misma probabilidad de ocurrencia en todas las ventanas de estudio. De acuerdo los resultados obtenidos de acuerdo con las estimaciones realizadas la función en particular

crece de forma rápida al inicio lo que podría sugerir que los eventos pueden estar presentando agrupaciones y que no se distribuyen de manera aleatoria en el espacio.

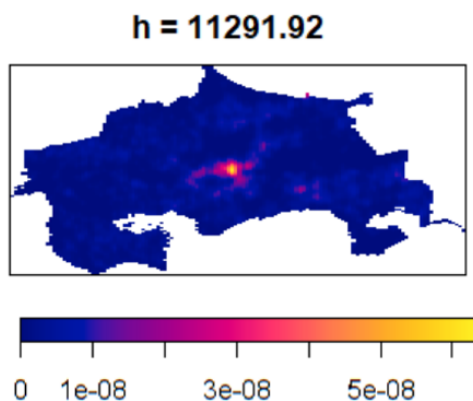


FIGURA 3: *Intensidad de los eventos según provincia.*

En cuanto al método de estadística de áreas lo primero que se realizó fue crear la variable agregada a nivel distrital que se utilizará en el análisis, esta corresponde a la magnitud de la disminución promedio reportada en los centros educativos que se ubican en un mismo distrito. Seguidamente se procedió a definir e identificar los vecinos y su matriz de pesos, para esto se utilizó el método de estimación con base en la distancia y utilizando dos vecinos para la identificación. Una vez realizados estos procedimientos se evaluó si existe dependencia espacial en los datos estimados a nivel distrital a partir del estadístico I Moran donde se parte de la hipótesis nula que indica que los datos están distribuidos aleatoriamente o bien que no hay autocorrelación espacial. El valor del estadístico obtenido fue de 0.24 con una probabilidad muy baja del $8.4601e-06$ por lo que hay evidencia estadística para establecer que si hay autocorrelación espacial y que por lo tanto no están distribuidos aleatoriamente.

en el espacio, lo cual es consistente con los resultados obtenidos a partir del análisis de procesos puntuales.

A partir del cálculo y estimación de test locales se procedió a identificar si habían agrupaciones o clusters que permitiesen explicar el fenómeno de estudio y a la identificación de los distritos de influencia. En el primer caso, los resultados obtenidos se presentan en la figura 4. La estimación se realizó a partir del cálculo del test de Moran Local bajo el supuesto estándar y con el supuesto de riesgos constantes, el cual corrige el modelo considerando las diferencias que de la población que hay en cada área. Dado que la variable de análisis está estrechamente relacionada con la población se sugiere como método más permite el que suma el supuesto de riesgos constantes, según se observa las mayores magnitudes de la disminución de la matrícula se concentran en las regiones del Gran Área Metropolitana, lo cual es consistente con las estimaciones de la intensidad de los eventos obtenidos a partir del método de procesos puntuales. Asimismo se identifica en el distrito de Liberia como una de las principales donde han habido mayores reducciones promedio en la matrícula de primaria.

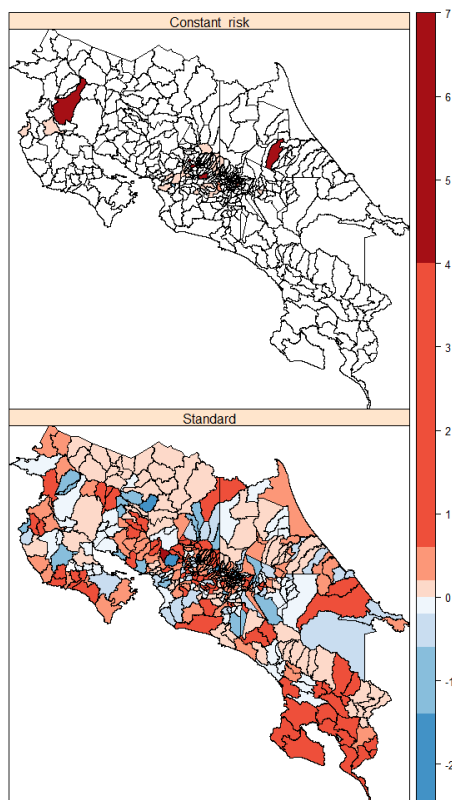


FIGURA 4: *Clusters estimados a partir del Test de Moran local.*

Finalmente se procedió a identificar los distritos de influencia, las estimaciones obtenidas se muestran en la figura 5, en esta figura la etiqueta HL(High-Low) significa que un distrito con alta disminución está rodeado por otros con menores magnitudes en la disminución de la matrícula, LH(Low-High) corresponde a distritos con bajas disminuciones pero que están alrededor de distritos con altas disminuciones de la matrícula y HH(High-High) corresponde a distritos rodeado por otros con altas disminuciones de la matrícula. Del análisis destacan los distritos de Monterrey y Venado del cantón de San Carlos con la etiqueta High-High, y para el caso del Low-High el distrito de Vásquez de Coronado del cantón de San José, la identificación de estos distritos es muy importante para poder

realizar intervenciones de política educativa más que se identificó que existe dependencia espacial entre los datos.

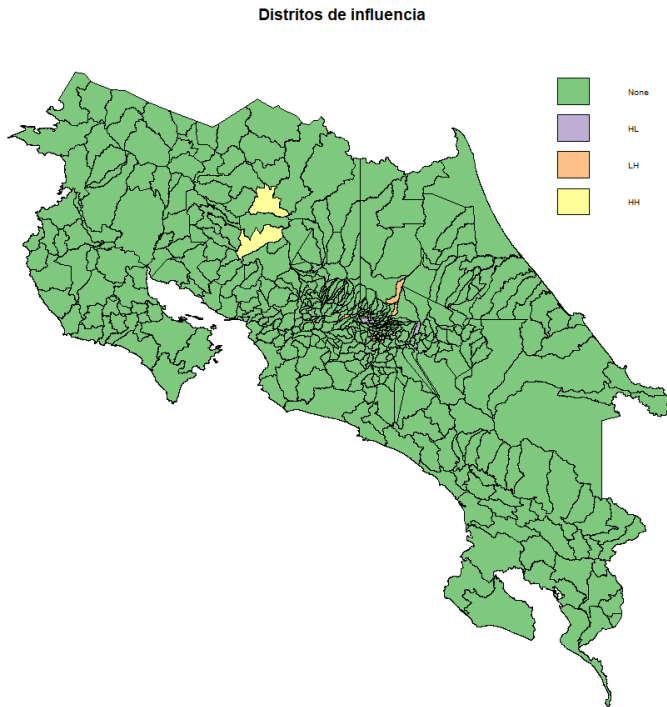


FIGURA 5: *Distritos de influencia.*

Conclusiones

- El cambio demográfico plantea una oportunidad para generar políticas educativas diferenciadas que permitan mejorar la calidad de la enseñanza, el empleo de técnicas geoestadísticas representan una importante herramienta de análisis.
- El análisis de procesos puntuales permite determinar que no existe una distribución aleatoria de los eventos definidos como disminuciones de matrícula en las escuelas en el país. Las

disminuciones se concentran en el Gran Área Metropolitana, en la provincia de San José y Alajuela, específicamente en el cantón de desamparados para el primer caso y en la zona de occidente para el segundo.

- Cuando se emplea la técnica de estadística de áreas que analiza los datos agrupados a nivel distrital se identifica la presencia de autocorrelación espacial, lo que significa que la locación espacial importa.
- La técnica de estadística de áreas detecta cuáles son distritos de influencia en los que habría que realizar un seguimiento más exhaustivo del comportamiento de las proyecciones de población y de la matrícula inicial a fin de analizar la aplicación de estrategias de política educativa diferenciada principalmente en los del cantón de San Carlos

Referencias

- [1] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria (2020).
- [2] H. Wickham, R. François, L. Henry, and K. Müller, *dplyr: A Grammar of Data Manipulation* (2020), r package version 1.0.2.
- [3] H. Wickham, *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag New York, 2016).
- [4] H. Wickham and E. Miller, *haven: Import and Export 'SPSS', 'Stata' and 'SAS' Files* (2020), r package version 2.3.1.
- [5] A. Baddeley, E. Rubak, and R. Turner, *Spatial Point Patterns: Methodology and Applications with R* (Chapman and Hall/CRC Press, London, 2015).
- [6] A. Baddeley, R. Turner, J. Mateu, and A. Bevan, *Journal of Statistical Software* **55**, 1 (2013).
- [7] A. Baddeley and R. Turner, *Journal of Statistical Software* **12**, 1 (2005).

-
- [8] R. J. Hijmans, *raster: Geographic Data Analysis and Modeling* (2020), r package version 3.3-13.
 - [9] R. Bivand, T. Keitt, and B. Rowlingson, *rgdal: Bindings for the 'Geospatial' Data Abstraction Library* (2020), r package version 1.5-12.
 - [10] R. S. Bivand, E. Pebesma, and V. Gomez-Rubio, *Applied spatial data analysis with R, Second edition* (Springer, NY, 2013).
 - [11] R. Bivand and D. W. S. Wong, TEST **27**, 716 (2018).
 - [12] E. Neuwirth, *RColorBrewer: ColorBrewer Palettes* (2014), r package version 1.1-2.
 - [13] P. E. de la Nación, .