

REVISTA

YACHAY

ISSN: 2412-2963
e-ISSN: 2520-9051

Envío: 26 mayo 2020

Aceptado: 29 Mayo 2021

Autor corresponsalLuis Eduardo Calderón Canto
lecalderoncanto@gmail.com

DOI: yachay.v10i01.268

Distribuido bajo

Creative Commons CC-BY-NC-SA 4.0

ORCIDLuis Eduardo Calderón Canto
<https://orcid.org/0000-0003-2350-401X>**OPEN ACCESS**

Análisis de patrones de puntos espaciales: Una aplicación al Covid-19 en el Perú

Luis Eduardo Calderón Canto¹¹ Universidad Nacional de San Antonio Abad del Cusco

RESUMEN

En diciembre del 2019 la enfermedad del Covid-19 emergió de la provincia de Wuhan, para luego extenderse por todo el mundo, el Perú no ha sido la excepción. Este trabajo tiene como objetivo principal estimar los patrones espaciales del Covid-19 en el Perú, además de establecer *clusters* a nivel regional. Mediante técnicas de estadística espacial se encontró asociación espacial mediante el índice de Moran (I Moran) global de 0.056 ($p < 0.05$). Por otro lado, se estableció cuatro *clusters*: en primer lugar, regiones al norte del país donde hay más casos, el sur donde hay menores cifras y *clusters* formados por outliers como Lima y Callao.

Palabras clave: Estadística espacial, patrones espaciales, asociación espacial, Covid-19.

Analysis of spatial point patterns: An application to Covid-19 in Perú

ABSTRACT

In December 2019, Covid 19 disease emerged from the Wuhan province, and then spread throughout the world, Peru has not been the exception. The main objective of this work is to estimate the spatial patterns of Covid 19 in Peru, in addition to establishing clusters at the regional level. Spatial association was found using spatial statistical techniques using the global I Moran of 0.056 ($p < 0.05$). On the other hand, four clusters were established: first, regions in the north of the country where there are more cases, the south where there are fewer numbers and clusters made up of outliers such as Lima and Callao.

Key words: Spatial statistics, Spatial patterns, Spatial association, Covid-19.

INTRODUCCION

En diciembre de 2019, la enfermedad del Covid 19 emergió de la provincia de Wuhan, para luego extenderse en toda China. A pesar de las medidas de contención del gobierno chino, la enfermedad se extendió por todo el mundo, teniendo como primeros epicentros Italia, Irán y España (Fannelli y Piazza, 2020). En Latinoamérica el primer caso se registró el 26 de febrero en Brasil y rápidamente por toda la región. En el proceso global actual, los procesos de migración y la dinámica de flujos de viaje son parte de la explicación de la rápida propagación (Murat Yüceşahin y Sirkeci, 2020).

En Perú, el primer caso se identificó el 6 de marzo en la ciudad de Lima, para luego extenderse en todo el país. La gravedad de la enfermedad y ya calificada como pandemia por la Organización Mundial de la Salud (OMS), obligó al gobierno a declarar el estado de emergencia para contener la propagación del virus a nivel nacional.

El gobierno peruano actualiza cada día la información nacional acerca de la enfermedad de manera sostenida. Los datos mostrados evidencian el comportamiento del virus a nivel regional, evidenciando el número de casos diarios, recuperados, hospitalizados, en cuidados intensivos, fallecidos y número de pruebas. Una gran ausencia en el análisis de la pandemia en el Perú ha sido la dinámica espacial, las relaciones de dependencia espacial, así como las implicancias de la heterogeneidad de la propagación en las distintas regiones.

La comunidad académica internacional ha abordado la pandemia de un modo sin precedentes. Siendo los campos de la medicina y la biología los más privilegiados. Por otro lado, la gran producción de

datos relacionados al Covid-19 ha permitido construir modelos predictivos. El análisis espacial no ha sido ajeno a este problema.

Bajo la primera ley de Tobler (1969) que indica que todo está relacionado, sin embargo, las cosas que se encuentran más cercanas, están más relacionadas que las que se distancian, lo que significó el vínculo entre la geografía y la estadística. La estadística espacial se entiende como el modelado de datos espaciales que muestran estructuras de dependencia, así como del análisis descriptivo e inferencial de los datos geográficos (Ebdon, 1977). Crassie (1992) distingue tres tipos de datos espaciales: datos geoespaciales, datos lattice y patrones de puntos. En particular el análisis de puntos espaciales ha sido útil en la investigación de la estadística epidemiológica para evaluar la distribución espacial y la concentración de determinada enfermedad (Bivand, Pedesma y Gómez-Rubio, 2013).

Por ejemplo, Han, Choi, Kim y Choi (2020) hacen un análisis espacial de la propagación del Covid-19 en 31 provincias en China. Mediante la estimación del I de Moran encuentran una significativa asociación espacial, encontrando máximos de 0.45 y 0.08 del índice en determinados días. Para el caso sudafricano se emplea el mismo indicador para encontrar hot spots del Covid-19 (Arashi et al. 2020).

Huang, Liu, Ding (2020) estudian la dependencia espacial y la propagación del Covid-19 en China central y sus vecinos durante 17 fechas. Además, evalúan las tendencias de propagación en función de los datos disponibles, mediante la I de Moran, encuentran dependencia espacial significativa a lo largo del periodo de estudio.

Giuliani et al. (2020) mediante el método de mínimos cuadrados generalizados, evalúan la propagación del Covid-19 en las provincias de Italia. Confirman el fuerte potencial de propagación de la enfermedad entre las áreas estudiadas, pero también su heterogeneidad. Evidencian que un grupo de provincias del norte italiano tendrían una transmisión endógena, en el centro por una transmisión de vecinos cercanos y en el sur una tendencia endémica.

Zhang et al. (2020) comparan el Covid-19 con el SARS, el estudio no encuentra asociación espacial significativa global mediante el I de Moran, -0.02 para el Covid y 0.06 para el SARS ($p > 0.05$). Sin embargo, mediante indicadores de asociación espacial local encuentra valores significativos. Identifica cuatro clusters de alto riesgo para el Covid 19 y dos clusters secundarios; mientras que, para el caso del SARS encuentra 3 clusters principales y 2 secundarios.

Para el caso latinoamericano y peruano hay muy poca evidencia espacial. Algunos intentos buscan hacer proyecciones en torno a las defunciones a partir de diversos modelos estadísticos (Bayes, Sal y Rosas & Valdivieso, 2020) y otros se centran en evidenciar las brechas de salud en el contexto de la pandemia (Gozzer, Canchihuamán y Espinoza, 2020).

Este trabajo tiene como objetivo principal estimar los patrones de puntos espaciales del Covid-19 en el Perú. En específico se busca cuantificar el nivel de asociación espacial entre los casos de

Covid-19 a nivel regional. También se busca establecer clusters regionales que permitan observar el comportamiento espacial de los casos de Covid 19 con sus vecinos. Para ello, nos valemos de técnicas provenientes de la estadística espacial.

MATERIAL Y METODO

Los datos provienen de la información del Ministerio de Salud para el día viernes 18 de abril de las 25 regiones del país proporcionado por el Ministerio de Salud. Por otro lado, se incluye información geo referencial de cada región considerando su centroide. Desafortunadamente no se cuenta con información provincial lo que impide tener un mayor panorama. Anselin y Forax (1995) precisan que el número de observaciones mayores a 50 se puede aproximar a una distribución normal; siendo nuestras observaciones 25, urge hacer la interpretación con más cautela (Herrera, 2012).

Con el objetivo de evidenciar si existe asociación espacial, empleamos técnicas provenientes de la estadística espacial. En este trabajo planteamos el índice de Moran (1950) y un indicador espacial de asociación local (Anselin, 1995). Además de un scatterplot desarrollado por Anselin (1996). Numerosos paquetes permiten computar información geográfica como ArcGis o ArcView. En este trabajo se emplea el software libre R en específico, se emplea la librería `spdep::moran.test()` Bivand (1992).

Matriz de pesos espaciales

Dada la matriz no negativa, $W=(w_{ij}; i,j=1, \dots, n)$ es una potencial matriz de pesos espaciales que vincula n unidades espaciales i y j . w_{ij} se conoce como peso espacial, representa el grado de influencia de la unidad j en la unidad i . El peso espacial toma diversos valores en función del criterio de relación espacial que se establezca y 0 cuando no existe relación ($w_{ii}=0$), dado que una unidad espacial no tiene relación consigo misma. En este trabajo emplearemos el criterio de distancia.

El índice de Moran global

Para poder estimar la presencia de autocorrelación espacial existen varias aproximaciones como el test de Matel (1967) o el estadístico de Geary (1954). Moran (1950) propuso el siguiente índice a partir de coeficiente de Pearson (Chen, 2013), la representación estándar de acuerdo a Cliff and Ord (1981) es:

$$I = \frac{(X_i - \bar{X})}{\sum_i \sum_i w_{ij}} \frac{N \sum_i \sum_i w_{ij} (X_i - \bar{X})(X_j - \bar{X})}{\sum_i (X_i - \bar{X})^2}$$

Donde N es el número de unidades espaciales i y j , en este caso, 25 puntos espaciales (regiones). X es una variable observable dada por el número de casos de Covid-19 a nivel regional en el Perú y $\sum_i \sum_i w_{ij}$ es la suma de los elementos de una matriz pesos espaciales. Donde la hipótesis alterna establece la existencia de dependencia espacial.

Cuando el valor del índice es cercano a 1, indica que los datos están agrupados y por lo tanto existe asociación espacial. Si el valor es cercano a -1, la distribución es aleatoria. Finalmente, si es 0 no

existe asociación espacial.

El índice de Moran local

La asociación espacial se puede dar en determinadas unidades espaciales, el indicador de asociación espacial local (LISA) permite estimar en cuánto contribuye una unidad espacial a la construcción de un valor global de asociación espacial. Anselin (1995) en el marco de mostrar de manera gráfica del I de Moran desarrollo el indicador, siendo Getis y Ord (1996) quienes contribuyeron a la parte inferencial:

$$I_i = \frac{(x_i - \bar{x})}{m_o} \sum_{j=i}^n c_{ij} (x_j - \bar{x})$$

La I_i puede identificar clusters y detectar valores atípicos. Donde x_i corresponde a una característica de la unidad espacial (casos), (\bar{x}) es el valor medio de los casos y c_{ij} corresponde al peso espacial. Por otro lado, se pueden identificar unidades espaciales bajo las siguientes formas: de alta presencia, rodeadas de unidades de alta presencia (high-high), unidades con alta presencia rodeadas de unidades de baja presencia (high-low), unidades con baja presencia rodeadas de unidades con alta presencia (low-high) de y unidades de baja presencia rodeadas de unidades de baja presencia también (low-low). Con presencia nos referimos al número de casos de Covid-19.

RESULTADOS

La Figura 1 muestra de manera preliminar el patrón de casos de Covid-19 en las regiones, donde resalta su distribución heterogénea. La costa norte concentra los mayores casos en contraste con la sierra sur del país. Se pueden observar también regiones con número de casos atípicos.

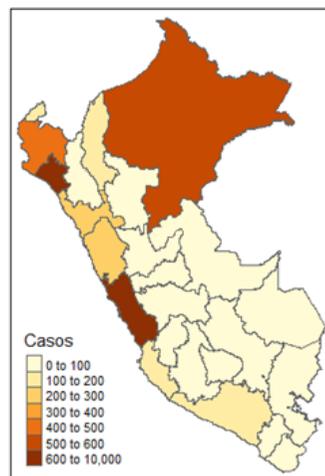


Figura 1. Número de casos de Covid-19

El valor observado de la I Moran global es 0.056 ($p < 0.05$) por lo que rechazamos la hipótesis nula y aceptamos la hipótesis alterna. A partir de lo cual se detecta una significativa asociación espacial, donde los

valores de los casos de Covid-19 presentan valores próximos a la de sus vecinos.

En cuanto a la estimación del LISA, en la Figura 2 se puede observar un primer cluster que agrupa observaciones menores a la media de casos y que cuentan con vecinos con valores altos, ubicado en las regiones del centro-norte del país en el primer cuadrante. Sucede lo contrario con el tercer cuadrante que agrupa a regiones con casos menores a la media y con vecinos con valores bajos también (cold spot). El cuadrante cuatro muestra regiones con casos bajos rodeados de regiones donde los casos son altos. Finalmente, el cuadrante 2 muestra a las regiones donde los casos son altos y de sus vecinos altos como el caso del Callao (hot spot). El cuadrante cuatro muestra el valor atípico que corresponde a Lima rodeado de regiones con casos bajos.

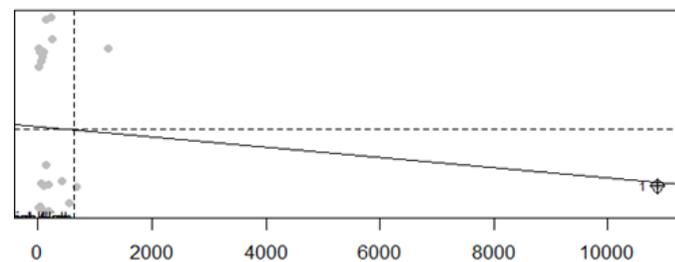


Figura 2. Clusters de casos de Covid-19

CONCLUSIONES

El presente trabajo tuvo como objetivo estimar el nivel de asociación espacial y establecer cluster a partir de los casos de Covid-19 en el Perú, en las 25 regiones. Para ello se emplearon técnicas de estadística espacial como la I de Moran y LISA, así se llegó a los siguientes hallazgos:

- Se encontró una significativa asociación espacial global 0.056, donde los casos de las regiones son parecidas a sus vecinos en determinadas áreas geográficas.
- Se pudo establecer cuatro clusters; por un lado, regiones del norte con una propagación del Covid-19 por encima del promedio nacional. Por otro lado, regiones del sur con menores casos al promedio nacional. Finalmente, outliers como las regiones de Lima y Callao.
- La divulgación de la información a nivel provincial y distrital brindaría mejores insumos para posteriores trabajos para la caracterización de la enfermedad en el Perú.

REFERENCIAS BIBLIOGRAFICAS

- Anselin, L. (1995). Local Indicators of Spatial Association—LISA. *Geographical Analysis*, 93-115.
- Anselin, L. (1996).) The Moran scatterplot as an ESDA tool to assess local instability in spatial association. En M. Fischer, H. Scholten, & D. Unwin, *Spatial Analytical perspectives on GIS* (pág. 111.125). London: Taylor and Francis.
- Anselin, L., & Forax, R. (1995). *New Directions in Spatial Econometrics*. Verlag: Springer.
- Arashi, M., Bekker, A., Salehi, M., Millard, S., Erasmus, B., Cronje, T., & Golpaygani, M. (2020). Spatial analysis and prediction of COVID-19 spread in South Africa after lockdown. <https://arxiv.org/abs/2005.09596>.
- Bayes, C., Sal y Rosas, V., & Valdivieso, L. (2020). Modelling death rates due to COVID-19: A Bayesian approach. <https://arxiv.org/abs/2004.02386>.
- Bivand, R. (1992). SYSTAT-compatible software for modeling spatial dependence among observations. *Comput Geosci*, 951–963.
- Bivand, R., Pebesma, E., & Gómez-Rubio, V. (2013). *Applied Spatial Data Analysis with R*. New York: Springer.
- Chen, Y. (2013). New Approaches for Calculating Moran's Index of Spatial Autocorrelation. *PLoS ONE* 8(7): e68336. <https://doi.org/10.1371/journal.pone.0068336>.
- Cliff, A., & Ord, K. (1981). *Spatial Processes: Models & Applications*. London: Pion.
- Cressie, N. (1992). *Statistics for spatial data*. New York: Wiley interscience publication.
- Ebdon, D. (1977). *Statistics in geography: A practical approach*. Oxford: Wiley-Blackwell.
- Fanelli, D., & Piazza, F. (2020). Analysis and forecast of COVID-19 spreading in China, Italy and France. *Chaos, Solitons and Fractals*, 1-5.
- Geary, R. (1954). he contiguity ratio and statistical mapping. *The Incorporated Statistician*, 115–145.
- Getis, A., & Ord, J. (1996). Local spatial statistics: An overview. En P. Longley, & M. Batty, *Spatial Analysis: Modeling in A GIS Environment* (págs. 261-277). New York: John Wiley & Sons.
- Giuliani, D., Dickson, M. M., Espa, G., & Santi, F. (2020). COVID-19, Italy, epidemiology, diffusion model, spatio-temporal diffusion models. Disponible en: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3559569.
- Gozzer, E., Canchihuamán, F., & Espinoza, R. (2020). COVID-19 y la necesidad de actuar para mejorar las capacidades del Perú frente a las pandemias. *Revista Peruana de Medicina Experimental y Salud Pública*.
- Herrera, M., Paz, J., & Cid, J. (2012). Introducción a la Econometría Espacial. Una Aplicación al Estudio de la Fecundidad en la Argentina usando R. *MPRA*, 1-30.
- Huang, R., Liu, M., & Ding, Y. (2020). Spatial-temporal distribution of COVID-19 in China and its prediction: A data-driven modeling analysis. *The Journal of infection in Developing Countries*.
- Kim, J.-H., Choi, J., Kang, D., & Choi, H. (2020). Spatial epidemic dynamics of the COVID-19 outbreak in China. *International Journal of Infectious Diseases*.
- Matel, N. (1967). he Detection of Disease Clustering and a Generalized Regression Approach. *Cancer Research*., 209-220.
- MINSA. (2020). Situación actual “Covid-19” al 18 de Abril 2020. Lima: Ministerio de Salud.
- Moran, P. (1950). Notes on Continuous Stochastic Phenomena. *Biometrika*, 17-23.
- Murat Yüceşahin, M., & Sirkeci, I. (2020). Coronavirus and Migration: Analysis of Human Mobility and the Spread of COVID-19. *Migration Letters*, 379-398.
- Topler, W. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region in the Detroit Region. *Economic Geography*, 234-240.
- Zhang, X., Rao, H.-X., Wu, Y., Huang, Y., & Dai, H. (2020). Comparison of the spatiotemporal characteristics of the COVID-19 and SARS outbreaks in mainland China. *medRxiv*.