

Análisis exploratorio de datos espaciales de pacientes con rinofaringitis, asma y bronquitis en consulta externa del Hospital Santa Rosa - 2006

Carla Patricia Zuñiga Vilca¹ y Emma Norma Cambillo Moyano²

Resumen: El presente trabajo tiene como objetivo evaluar el tipo de patrón espacial que siguen los casos de pacientes con rinofaringitis, asma y bronquitis del año 2006 en la ciudad de Lima que fueron atendidos por el Hospital Santa Rosa. Se utilizó un análisis exploratorio de datos espaciales (AEDE) obteniendo como resultados que los casos de asma atendidos por el Hospital Santa Rosa no presentan un patrón espacial a lo largo de Lima Metropolitana (distribución no homogénea), siendo en mayor intensidad en aquellos distritos más cercanos a dicho hospital y específicamente en el distrito de Pueblo Libre. Además, la presencia de clústeres significativos se dio en los distritos de Pueblo Libre, Breña, Jesús María, Magdalena del Mar, San Miguel y Lima.

Palabras clave: análisis exploratorio de datos espaciales; infecciones respiratorias agudas; índice de Morán.

Exploratory analysis of spatial data of patients with rhino pharyngitis, asthma and bronchitis in the outpatient clinic of Santa Rosa Hospital - 2006

Abstract: The objective of this work is to evaluate the type of spatial pattern followed by the cases of patients with Rhino pharyngitis, asthma and bronchitis the year 2006 in the city of Lima, who were treated by the Santa Rosa hospital. An exploratory analysis of spatial data (AEDE) was used, obtaining as a result that the cases of asthma treated by the hospital Santa Rosa do not present a spatial pattern throughout Metropolitan Lima (inhomogeneous distribution), being more intense in those districts more close to said hospital and specifically in the district of Pueblo Libre. In addition, the presence of significant clusters occurred in the districts of Pueblo Libre, Breña, Jesús María, Magdalena del Mar, San Miguel and Lima.

Keywords: exploratory analysis of spatial data; acute respiratory infections; Moran's index.

Recibido: 08/01/2019. *Aceptado:* 26/04/2019. *Publicado online:* 13/05/2019.

¹UNMSM, Facultad de Ciencias Matemáticas, e-mail: czunigav@unmsm.edu.pe

²UNMSM, Facultad de Ciencias Matemáticas, e-mail: ecambillom@unmsm.edu.pe

1. Introducción

El presente artículo es desarrollado a partir de la tesis titulada “Análisis exploratorio de datos espaciales de las enfermedades respiratorias en consulta externa del hospital Santa Rosa - 2006” para la obtención de grado de Magister en Bioestadística

Las enfermedades respiratorias en Lima Metropolitana siguen en aumento debido a los problemas de contaminación del medio ambiente. Se sabe que las principales causas de atenciones por consulta externa en el Ministerio de Salud - MINSA, para el período 2002 - 2005 son las infecciones agudas de las vías respiratorias como asma, bronquitis, enfisemas, entre otras.

Todos los años en la época de invierno se observa el incremento de las infecciones respiratorias observando de este modo un patrón que es familiar para cada centro de salud, de este modo cada servicio de salud puede prevenir la presencia de estas enfermedades respiratorias (Ministerio de Salud [MINSA], 2002, p.21).

La Oficina Ejecutiva de Planeamiento Estratégico del Hospital Santa Rosa, publicó el documento El Plan Operativo 2008 en el cual indica que el distrito que demanda el mayor porcentaje de atenciones es Pueblo Libre (24.86 %), le siguen en orden correlativo los distritos del Cercado de Lima, San Miguel, San Martín de Porres, Jesús María y Magdalena. Así mismo, la principal causa de atenciones que brinda el hospital en consulta externa son las enfermedades del sistema respiratorio (18 %) para los pacientes que proceden de los diferentes distritos de Lima Metropolitana.

Por esta razón; esta investigación tiene por objetivo determinar el patrón espacial de la morbilidad del asma en consulta externa en el Hospital Santa Rosa en el 2006.

2. Metodología

El tipo de estudio es exploratorio porque se va investigar por primera vez el patrón espacial del asma en los pacientes que acuden al Hospital Santa Rosa por consulta externa.

El diseño de investigación es observacional - transversal debido a que observan tal cual se presenta las enfermedades del sistema respiratorio para luego analizar el patrón espacial de estas enfermedades y transversal porque se mide la enfermedad en un momento dado del tiempo.

La recolección de los datos se realizó en el Hospital Santa Rosa; durante el 2006 a todos pacientes diagnosticados con rinofaringitis, asma y bronquitis en consulta externa, contando con un total de 8 196 pacientes de los 43 distritos de Lima Metropolitana

Para el análisis de los resultados se utilizó los softwares estadísticos SPSS v21 y R Project 3.2.5 para el caso del análisis exploratorio de datos (AED) y para el AEDE se usó QGIS y Geoda 1.8.

2.1. Análisis exploratorio de datos espaciales

El análisis exploratorio de datos espaciales (AEDE) es una técnica que describe y visualiza la distribución espacial ayudando a descubrir la presencia de patrones de asociación espacial y agrupamientos (puntos calientes) así como identificar observaciones atípicas espaciales (spatial

outliers) y sugerir formas de heterogeneidad espacial (Anselin, 1988).

El AEDE se presenta de manera univarida cuando se utiliza una sola variable, bivariada cuando se busca la relación de dos variables y multivariada cuando se utiliza más de dos variables.

Cuando no existe una teoría previa acerca del fenómeno que se estudia se recomienda el uso de las técnicas del AEDE, debido a que no se conoce la distribución de datos geográficos a priori (grandes bases de datos geográficos).

2.2. Medidas de asociación espacial global

La autocorrelación espacial (AE) es definida como la agrupación de los valores de una variable en un mapa; es decir expresa el grado en que objetos en una unidad geográfica son similares a otros objetos en unidades geográficas próximas (Goodchild, 1987).

Baronio, Vianco y Rabanal (2012) afirman:

El análisis de autocorrelación espacial global realiza un examen conjunto de todas las unidades que componen la muestra para determinar si las unidades espaciales se encuentran distribuidas aleatoriamente o si, por el contrario, lo hacen conforme a un patrón determinado. (p.13)

A continuación, se describen una medida de asociación espacial:

I de Moran: permite averiguar la existencia de *dependencia espacial* de una variable dada en diferentes lugares o, caso contrario, indica una distribución espacialmente aleatoria.

Para determinar la autocorrelación espacial usando el *Índice de Moran* se usa la siguiente fórmula:

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n \sum_{j=1}^n w_{ij} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (1)$$

donde:

- n : número de áreas
- x_i : valor de la variable considerada en el área i
- \bar{x} : valor medio de la variable en la región de estudio
- w_{ij} : pesos de la matriz W
- x_j : valor de la variable considerada en el área j

A través de este indicador, podemos averiguar el grado de autocorrelación espacial de una variable x en diferentes áreas i y j , (x_i, x_j) , ponderada por la proximidad geográfica medida por w_{ij} , donde el numerador muestra el promedio de los productos de las desviaciones de las zonas i y j en la media global, y el denominador es una medida de la variabilidad de las desviaciones con respecto a la media global.

El valor del índice I de Moran varía entre -1 y $+1$, donde los valores negativos indican un conglomerado espacial de unidades territoriales con valores de análisis distintos y los valores positivos indican un conglomerado espacial de unidades territoriales con valores de análisis similares, sean estos altos o bajos.

Si el valor de este índice 0 o cercano a 0 indica que las variables son espacialmente independientes, lo que significa que no hay autocorrelación o no hay dependencia espacial.

2.3. Mapa de LISA

Las letras LISA significan “Local Indicator of Spatial Association” (Indicador local de asociación espacial). En el mapa LISA muestra la presencia de puntos calientes (“hot spots”) o atípicos espaciales, cuya mayor o menor intensidad dependerá de las medidas de asociación espacial global. En esta investigación se utilizó el estadístico de asociación local I de Morán.

2.4. Pruebas de aleatoriedad espacial completa

Si se sospecha que la intensidad (número de casos por unidad de área) no es homogénea a lo largo del área de estudio, se realizan las pruebas de Aleatoriedad Espacial Completa (CSR, por sus siglas en inglés).

Para realizar la prueba estadística para la homogeneidad se asume que las áreas no son iguales basado en los conteos de los cuadrantes y asumiendo que el proceso puntual espacial es Poisson, se considera las siguientes hipótesis:

De forma general, dejando de asumir que las áreas son iguales, una prueba estadística para homogeneidad, basada en los conteos de los cuadrantes, asumiendo que el proceso puntual espacial es Poisson, consiste en considerar, las siguientes hipótesis:

H_0 : La intensidad es homogénea (CSR) y como H_1 : La intensidad es no homogénea (Es un proceso Poisson no homogéneo). Si la hipótesis nula es verdadera, entonces los n_j son realizaciones de variables aleatorias Poisson independientes, con valores esperado $\mu_j = \rho a_j$.

Dado que el número de puntos $n = \sum_j n_j$ y el área total de la ventana $a = \sum_j a_j$, entonces la intensidad estimada es $\hat{\rho} = n/a$ y el conteo esperado en el cuadrante B_j es $e_j = \hat{\rho} a_j = n a_j / a$.

El estadístico ji - cuadrado es:

$$\chi^2 = \sum_j \frac{(n_j - e_j)^2}{e_j} = \sum_j \frac{(n_j - \hat{\rho} a_j)^2}{\hat{\rho} a_j} \sim \chi^2_{(m-1)} \quad (2)$$

Si los cuadrantes tienen igual área, entonces los n_j son independientes con igual valor esperado bajo la hipótesis nula. La prueba estadística se reduce a:

$$\chi^2 = \sum_j \frac{(n_j - n/m)^2}{n/m} \sim \chi^2_{(m-1)} \quad (3)$$

Bajo la hipótesis nula, la distribución del test estadístico es aproximadamente una χ^2 con $m - 1$ grados de libertad y los residuales Pearson son $r_j = (n_j - e_j) \sqrt{e_j}$. La aproximación es

tradicionalmente aceptable cuando los conteos esperados e_j son mayores que 5 para todos los cuadrantes.

3. Resultados y Discusión

El análisis exploratorio de datos espaciales en muchos casos inicia con la representación de las variables en estudio a través de un mapa coloreada de acuerdo a un criterio de generación de intervalos, con el fin de tener una idea vaga acerca del tipo de distribución espacial de la variable en estudio.

En la figura 1, en la cual los distritos con mayor número de casos son coloreados en azul más intenso. En forma de punto y de color rojo se representa a la ubicación del hospital Santa Rosa. Para los pacientes con asma se observa que existen determinados distritos con un mayor número de casos, en esencia se observa que gran parte de la demanda de pacientes se encuentran en los distritos cercanos a la ubicación del hospital, suponemos que los pacientes prefieren atenderse en un hospital más cercano a estos. Esto conlleva a sospechar que la distribución espacial de los pacientes del hospital Santa Rosa están distribuidos en forma no homogénea a lo largo de la ciudad de Lima y más cercanas a estas.

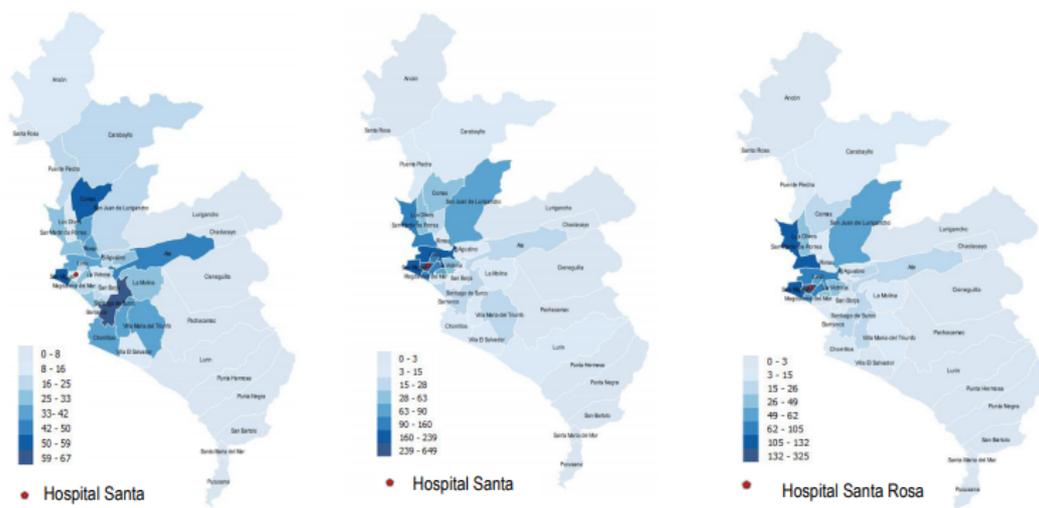


Figura 1. Número de casos de pacientes con rinofaringitis (lado izquierdo), asma (lado central) y bronquitis aguda (lado derecho).

Para corroborar que los pacientes prefieren atenderse en un hospital más cercano al distrito en que vive, se elaboró un diagrama de dispersión, en la cual la variable dependiente corresponde a la tasa de IRAS atendidas por el hospital Santa Rosa en el 2006 versus la distancia en kilómetros a dicho hospital (DIST KM). La tasa de IRAS se calcula con el total de casos con enfermedades respiratorias de rinofaringitis, asma y bronquitis que acudieron al hospital Santa Rosa por consulta externa dividido por la población expuesta al 2006.

La figura 2, en la cual se observa que los distritos que se encuentran más cercanos al hospital Santa Rosa son los mismos que demandan una mayor atención por consulta externa al hospital Santa Rosa, mientras que, en el caso contrario, son más reducidas la proporción de pacientes

que acuden desde distritos lejanos. En general los datos se ajustan de mejor forma a una curva no lineal del tipo exponencial.

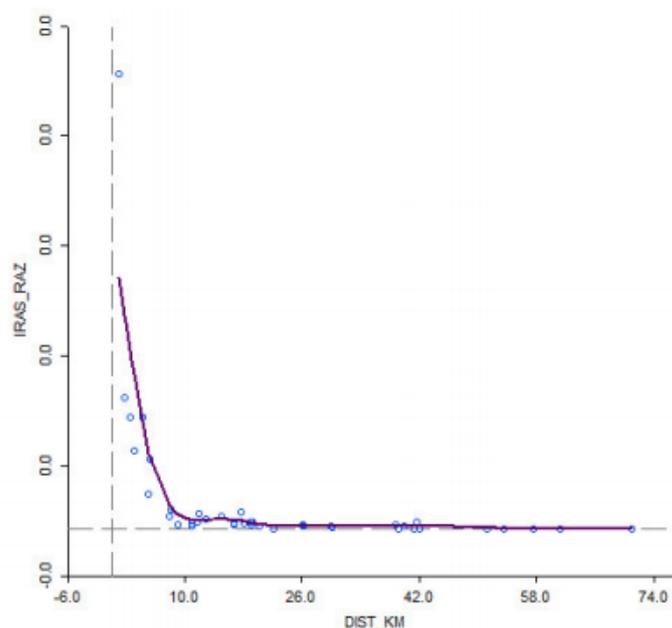


Figura 2. Diagrama de dispersión de la distancia en kilómetros versus la tasa de IRAS atendidas por el Hospital Santa Rosa 2006.

Otra alternativa para estudiar los casos de pacientes con infecciones respiratorias corresponde a analizar el número de casos por unidad de área (km²), denominada intensidad de casos. Esta puede considerarse homogénea u no homogénea a lo largo de la ciudad de Lima.

Los resultados se muestran en la Tabla 1, del cual a un nivel de confianza de (99 %), se rechaza la hipótesis nula, con lo cual se concluye que los casos de pacientes con infecciones respiratorias que acudieron al hospital Santa Rosa en el 2006, estuvieron distribuidos de forma no homogénea a lo largo de la ciudad de Lima siguiendo un proceso espacial Poisson no homogéneo.

Tabla 1. Prueba de aleatoriedad espacial completa, para los casos de IRAS atendidos por el Hospital Santa Rosa 2006 y total de casos de IRAS en Lima Metropolitana 2015.

Tipo de cuadrantes	Nro. De poligonos	Área total (Km ²)	Nro. De casos.	χ^2	Df	p-value
IRAS – Hospital Santa Rosa, 2006	43	2641.4	8196	466878.8	42	0.000

Finalmente, en base a los indicadores locales de asociación espacial, se identificó la presencia de clústeres altamente significativo (altas tasas de pacientes con infecciones respiratorias) para los distritos Pueblo Libre, Breña, Jesús María, Magdalena del Mar, San Miguel y Lima tal cual se muestra en la figura 3.

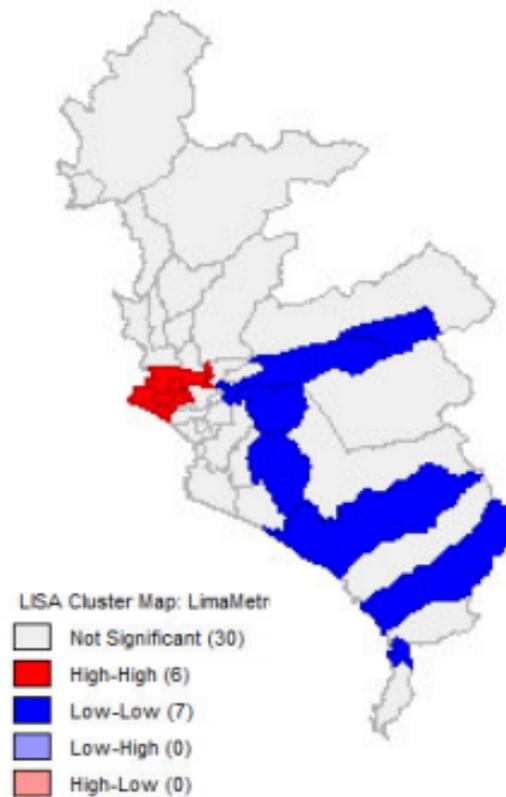


Figura 3. Detección de clúster espaciales para la tasa de incidencia de infecciones respiratorias de las consultas externas del Hospital Santa Rosa 2006.

4. Conclusión

La morbilidad de la rinofaringitis, asma y bronquitis en consulta externa del Hospital Santa Rosa - 2006 están distribuidas de forma no homogénea a lo largo de ciudad de Lima en la cual los distritos más cercanos al hospital requieren de mayores consultas externas de dicho establecimiento de Salud.

A partir del índice de LISA y el índice de Morán se encontraron la presencia de clústeres significativos de distritos con alta demanda las tres enfermedades más prevalentes del sistema respiratorio en consulta externa del Hospital Santa Rosa. Estos distritos son: Pueblo Libre, Breña, Jesús María, Magdalena del Mar, San Miguel y Lima.

Referencias bibliográficas

- [1] Anselin, L. (1988). *Spatial Econometric: Methods and Models*. Dordrecht, Holanda: Kluwer Academic Publisher.
- [2] Baddeley, A., Rubak, E., y Turner, R. (2016). *Spatial Point Patterns Methodology and Applications with R*. United States of America: CRC Press Chapman & Hall BOOK.

- [3] Baronio, A., Vianco A. y Rabanal C. (2012). *Una Introducción a la Econometría Espacial. Dependencia y Heterogeneidad*. Recuperado el 11 de noviembre del 2018 de <http://www.econometricos.com.ar/wp-content/uploads/2012/11/Espacial.pdf>.
- [4] Goodchild, M. (1987). Spatial Analytical Perspective on Geographical Information Systems. *International Journal of Geographical Information Systems*, 1(4), 327-334.
- [5] Haining, R., Wise, S. y Signoretta, P. (2000). Providing scientific visualization for spatial data analysis: Criteria and an assessment of SAGE. *Journal of Geographical Systems*, 2(2), 121-140.
- [6] Ministerio de Salud. (2002). *Análisis de la Situación de Salud - Guía para el análisis del proceso salud - enfermedad*. Lima, Perú
- [7] Oficina Ejecutiva del Planeamiento Estratégico - Hospital Santa Rosa. (2008). *El Plan Operativo Anual del Hospital Santa Rosa*. Lima, Perú.
- [8] Zuñiga, C. (2018). *Análisis exploratorio de datos espaciales de las enfermedades respiratorias en consulta externa del Hospital Santa Rosa - 2006* (Tesis de Maestría). Universidad Nacional Mayor de San Marcos, Perú.