

Modelos estadísticos en procesos puntuales espaciales poisson para evaluar la distribución espacial de los hechos delictivos en Lima, Perú

Statistical models in poisson spatial point processes to evaluate the spatial distribution of criminal acts in Lima, Perú

Braulio Quispe Quispe^{1,2}; Carlos López de Castilla Vasquez^{3*}

*Autor de correspondencia

Resumen

Se plantea una aplicación de los modelos estadísticos de procesos puntuales espaciales Poisson para evaluar la distribución espacial de hechos delictivos y su relación con algunas covariables espaciales para el área comprendida por los distritos de Lima Centro y Residencial. La información utilizada fue las ubicaciones georreferenciada de los hechos delictivos reportadas por las víctimas durante los años 2013 y 2014 (patrón puntual de hechos delictivos). A través de las estadísticas de resumen se identificaron las zonas con mayor incidencia de hechos delictivos como: Lince, Trébol de Javier Prado y Lima Cercado, y el tipo de distribución espacial no homogéneo (existen conglomerados o agregación de puntos) que siguen los hechos delictivos. El modelamiento estadístico se realizó a través de la intensidad de puntos usando los modelos log-lineales para representar su relación con un conjunto de covariables espaciales, determinándose que el número de hechos delictivos por unidad de área (intensidad) guarda relación con la ubicación de los límites distritales, la inversión destinada al orden interno y la densidad poblacional.

Palabras claves: Procesos puntuales espaciales; covariables espaciales; intensidad; distribución espacial; hechos delictivos; imagen satelital.

Abstract

An application of the statistical models of Poisson spatial point processes is proposed to evaluate the spatial distribution of criminal acts and its relation with some spatial covariables for the area comprised by the districts of Lima Centro and Residencial. The information used was the georeferenced location of the criminal acts reported by the victims at the end of 2013 until the beginning of 2014 (spatial point pattern of criminal acts). Through the summary statistics were identified the areas with the highest incidence of criminal acts: Lince, Trébol de Javier Prado and Lima Cercado, and the type of spatial distribution in homogeneous (there are clustering or aggregation of points) that follow the criminal acts. Statistical modeling was performed through the intensity of points using log-linear models to represent their relationship with a set of spatial covariables, determined that the number of criminal acts per unit area (intensity) is related to the location of the limits districts, the investment destined to the internal order and the population density.

Keywords: Spatial point processes; spatial covariables; intensity; spatial distribution.

1. Introducción

El análisis de los eventos observados en forma de puntos expresados por sus coordenadas de ubicación (longitud, latitud) dentro de un espacio bidimensional, son tratados en el análisis espacial de Patrones Puntuales Espaciales, basado en la teoría de procesos puntuales espaciales, donde el interés radica principalmente en determinar el tipo de distribución espacial y/o las causas que generan la aparición de estos puntos. El modelo estadístico más conocido para el análisis de patrones puntuales espaciales es el proceso puntual espacial Poisson. A partir de este se construyen modelos más complejos (Baddeley, Moller y Waagepetersen, 2000). El modelamiento se plantea en el

término intensidad (los más usados se expresan en la forma log-lineal) con el fin de analizar el efecto de las covariables espaciales. El modelo nulo por lo general es el proceso puntual Poisson homogéneo, conocido también como Aleatoriedad Espacial Completa (CSR, por sus siglas en inglés), en la cual la intensidad (número de puntos por unidad de área) se asume homogénea a lo largo de todo el espacio de estudio.

Por otra parte, la inseguridad ciudadana es uno de los mayores problemas que afronta el país. Una de las herramientas más utilizadas por las municipalidades distritales de Lima, son los mapas del delito (MLM, 2016), donde los hechos delictivos se representan en forma de puntos. Analizar la

¹Especialista Estadístico, Ministerio de Educación, Lima, Perú. E-mail: bquispeq@hotmail.com

²Universidad Nacional Agraria La Molina, Lima, Perú. E-mail: bquispeq@hotmail.com

³Universidad Nacional Agraria La Molina, Lima, Perú. Email: clopez@lamolina.edu.pe

distribución espacial de los hechos delictivos (eventos), involucra considerar su localización o ubicación espacial y a partir de estas evaluar su distribución en el espacio, pudiendo estar distribuidos homogéneamente en todo el espacio de estudio o de lo contrario pueden existir zonas con mayor incidencia e inclusive formando agregados espaciales. Las técnicas estadísticas clásicas suponen que al estudiar un fenómeno se toman observaciones bajo circunstancias idénticas e independientes entre sí y por ello no son convenientes para analizar fenómenos que varían en el espacio. Así también, la información observada, i.e., el patrón puntual no solo consiste en un conjunto de ubicaciones de puntos, sino también la ausencia en otras ubicaciones es informativo (Baddeley *et al.*, 2005). Esto es útil para entender las relaciones de la ausencia de puntos con las covariables espaciales.

Los objetivos planteados en esta investigación fueron:

Estimar el mejor modelo estadístico de procesos puntuales espaciales que se ajuste al patrón puntual de hechos delictivos y específicamente, determinar:

El tipo de patrón puntual espacial que caracteriza a la distribución los hechos delictivos en la ciudad de Lima. Es decir, determinar si los hechos delictivos se encuentran distribuidos de forma aleatoria, o regular, o se encuentran formando agregados a lo largo de la ciudad de Lima.

Si en los lugares con mayor densidad poblacional, existe mayor ocurrencia de hechos delictivos.

Si la inversión en orden interno efectuado por los Municipios, favorece a la reducción de la delincuencia.

Si los límites jurisdiccionales (distritales) favorecen al incremento de los hechos delictivos.

Si la distribución espacial de las comisarías y serenazgos, tiene efecto reductor en los hechos delictivos.

2. Materiales y métodos

La presente investigación es de tipo exploratoria, descriptiva y correlacional. El diseño de la investigación es no experimental de corte transversal, dado que los datos corresponden a un determinado espacio y tiempo.

Formulación de hipótesis

La hipótesis general considerada es que el patrón puntual de hechos delictivos, se ajusta de mejor forma a un modelo de un proceso puntual espacial Poisson no homogéneo, mientras las específicas son: Los hechos delictivos se encuentran distribuidos de forma no homogénea a lo largo de la ciudad de Lima; en los lugares con mayor densidad poblacional existe mayor ocurrencia de hechos delictivos; la inversión en orden interno efectuado por los Municipios favorece a la reducción de la delincuencia; los límites jurisdiccionales (distritales) favorecen al incremento de los hechos delictivos; la distribución espacial de las comisarías y serenazgos tiene efecto reductor en los hechos delictivos.

Población y muestra

La población bajo estudio está constituida por el conjunto de todos los puntos (ubicaciones) en donde se registró un hecho delictivo (robos en la modalidad de: arrebato, bujiazos, cogoteo, raqueto, etc) durante los años 2013 y 2014, en la ciudad de Lima Centro y Residencial. Específicamente, el área de estudio o ventana de observación W se muestra en la Figura 1 y constituye la región conformada por los distritos de Lima Centro (La Victoria, Lima, Lince, Breña, Rímac) y Lima residencial (Barranco, Jesús María, La Molina, Magdalena, Miraflores, Pueblo Libre, San Borja, San Isidro, San Luis, San Miguel, Santiago de Surco, Surquillo).

El lugar de ocurrencia de un hecho delictivo en W , corresponde a un punto y constituye un vector aleatorio representado por sus coordenadas geográficas (Longitud, Latitud). Se asume que el tamaño de la población N , es grande y desconocido por distintos motivos, por ejemplo, no todas las víctimas se atreven a denunciarlo y/o reportarlo. Por tanto, si se toman todos los $n=1,082$ puntos recolectados en los años 2013 y 2014, se puede considerar a estos como una muestra aleatoria de toda la población. Este mismo criterio es utilizado en la aplicación de Minería de Datos. Así, la unidad de análisis son los puntos en los cuales se registró un hecho delictivo durante los años 2013 y 2014 y que fueron reportados y ubicados por las víctimas a través del aplicativo web Datea (desarrollada por Giancarlo Díaz Pardo, policía en actividad) al interior de W . Cabe recalcar que esto no debe confundirse con la metodología de procesos puntuales espaciales, para el cual este conjunto de puntos en si representa una única realización del proceso puntual espacial bidimensional, lo cual difiere notablemente de los estudios clásicos donde cada dato se considera una observación y el conjunto la muestra.

Metodología aplicada

La estimación de los parámetros del modelo de procesos puntuales espaciales Poisson, se realizó a través del método de máxima verosimilitud en el programa R Project utilizando el paquete spatstat. El procedimiento se desarrolló en el siguiente orden: Mapeo de puntos, Estimación de la intensidad de hechos delictivos, Pruebas de Aleatoriedad Espacial Completa, Estimación de las estadísticas de resumen de segundo orden, Estimación de Modelos y la Selección de modelos.

Procesos Puntuales en R^d

Moller y Waagepetersen (2004), definen un proceso puntual espacial X , como un subconjunto aleatorio numerable de un espacio $S \subseteq R^d$ (el caso más simple es cuando $d=2$). Una realización de tal proceso es un patrón puntual espacial $x = \{x_1, \dots, x_n\}$ de $n \geq 0$ puntos contenidos en S . Para especificar la distribución de X , puede definirse la distribución del número de puntos $n(X)$ y para cada $n \geq 1$ condicionado a $n(X)=n$, la distribución conjunta de los n

puntos en X . En la práctica se observa solo los puntos contenidos en una ventana de observación fija y conocida $W \subseteq S$. Así, dado que $X \setminus W$ no es observado, genera el problema de efectos de borde (Baddeley, Rubak y Turner, 2016).

Procesos Puntuales espaciales Poisson

Un proceso puntual X en S , con función de intensidad $\rho: S \rightarrow [0, \infty)$ integrable localmente y medida de intensidad $\mu(B) = \int_B \rho(u) du$ localmente finita, es de Poisson si para cada región acotada $B \subseteq S$ con $\mu(B) > 0$:

- a) $N(B) \sim Po(\mu(B))$, es decir, el número de puntos en B sigue una distribución Poisson con media $\mu(B)$.
 - b) Teniendo en cuenta $N(B)$, los puntos en $X_B = X \cap B$ son i.i.d. con densidad proporcional a $\rho(u)$, es decir $f(u) = \rho(u) / \mu(B)$, donde $u \in B$.
- y se escribe $X \sim Poisson(S, \rho)$.

donde, μ determina el número esperado de puntos en B , es decir, $E[N(B)] = \mu(B)$.

De los postulados a) y b) la densidad conjunta de n y (x_1, \dots, x_n) está dado por:

$$f\{(x_1, \dots, x_n), n\} = \left\{ \frac{e^{-\mu(B)} \prod_{i=1}^n \rho(x_i)}{n!} \right\} \tag{1}$$

Si ρ es constante, el proceso *Poisson* (S, ρ) se denomina proceso *Poisson* homogéneo en S , conocido también como Aleatoriedad Espacial Completa (CSR, por sus siglas en inglés) con intensidad ρ y de lo contrario se dice un proceso de Poisson no homogéneo sobre S Para más detalle consulte Cressie (1991).

Estadísticas de resumen

Se basan principalmente en las características que describen las propiedades de primer orden (la intensidad) en analogía a la media comúnmente conocida en el caso no espacial, ya que estos describen la densidad o intensidad en términos del número promedio de puntos por unidad de área. Mientras que las de segundo orden tales como: las funciones de correlación por pares (g), las medidas de momentos de segundo orden reducidas (como las llamadas funciones L y K) y las estadísticas de resumen basadas en las distancias entre puntos (como las funciones F , G y J) se relacionan con la posición relativa de los puntos o la interacción entre estos. Para más detalle consulte Stoyan y Stoyan (1994), Moller y Waagepetersen (2004) y las referencias en estas.

La medida de la intensidad

Representa el número esperado o promedio de puntos del proceso puntual situado en alguna región del espacio. Así, la medida de intensidad μ en R^d está dada por:

$$\mu(B) = E[N(B)] = \int_B \rho(u) du, B \subseteq \mathbb{R}^2 \tag{2}$$

donde ρ es la función de intensidad no negativa A menudo se usa la estimación de la intensidad en lugar de su correspondiente medida. Así, supóngase que ha sido posible observar un único patrón puntual $X_W = x$ en una ventana acotada $W \subseteq R^d$ con área $|W| > 0$; en el caso homogéneo el estimador de máxima verosimilitud MV es:

$$\hat{\rho} = n(x) / |W| \tag{3}$$

y para el caso no homogéneo, un estimador núcleo (no paramétrico) de la función de intensidad es:

$$\hat{\rho}_b(u) = \sum_{v \in x} k_b(u-v) / c_{W, b(v)}, u \in W \tag{4}$$

donde k_b es un núcleo (Silverman, 1986) con un ancho de banda $b > 0$, es decir $k_b(u) = k(u/b) / b^d$ donde k es una función de densidad dada, y $c_{W, b(v)} = \int_W k_b(u-v) du$ es un factor de corrección de borde.

Pruebas de conteo por cuadrantes para CSR

Este método consiste en dividir W en subregiones disjuntas denominadas cuadrantes de área a_1, \dots, a_m . Si los conteos de puntos n_j para $j = 1, \dots, m$ son divididos por su correspondiente área, obtenemos un simple estimado de la función de intensidad. Así, una prueba estadística para homogeneidad, asumiendo que el proceso puntual es Poisson, se basa en los conteos de los cuadrantes considerando como H_0 : La intensidad es homogénea (CSR) y como H_1 : Es un proceso Poisson no homogéneo. Si H_0 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 j -ésimo cuadrante es $e_j = \hat{\rho} a_j = (n a_j) / a$. La prueba estadística Ji cuadrado, es:

$$\chi^2 = \sum_j \frac{(n_j - e_j)^2}{e_j} = \sum_j \frac{(n_j - \hat{\rho} a_j)^2}{\hat{\rho} a_j} \sim \chi^2_{(m-1)} \tag{5}$$

La aproximación es tradicionalmente aceptable cuando los conteos esperados e_j son mayores que 5 para todos los cuadrantes. Para más detalle consulte Schabenberger y Gotway (2005), Lloyd (2007), Baddeley (2008) y Diggle (2014).

Estadísticas de resumen de segundo orden

Permiten determinar si los puntos, tienden a formar agregaciones o presentar dispersión (regularidad) en un rango de distancias.

En el caso de la *Función – g* de correlación por pares, si los puntos son independientes unos de otros como es el caso del proceso Poisson $g(r)=1$, si $g(u,v)>1$ indica la presencia de atracción entre puntos (agregaciones) hasta una distancia r , mientras que si $g(r) < 1$ indica repulsión entre puntos (regularidad). En el caso de la *Función – K*, $\rho K(r)$ representa el número esperado puntos en una bola de radio r cuyo centro es un punto en X no considerado en el conteo. La función $-K$ y la función $-L$ están en correspondencia uno a uno, siendo la función $-L$ la más usada por representar la identidad de un proceso de Poisson.

Las estadísticas de resumen basada en las distancias entre puntos tales como: La función $-F$ de espacio vacío, es la función de distribución de la distancia desde un origen (u otro punto fijo en R^2) hasta el punto más cercano en X . La función $-G$, conocida también como “Función al vecino más cercano”, se interpreta como la función de distribución de la distancia desde un punto típico en hasta el vecino más cercano en X y la función $-J$, propuesta por Lieshout y Baddeley (1996), se define por $J(r) = (1 - G(r)) / (1 - F(r))$ para $F(r) < 1$. Para un proceso Poisson homogéneo en R^2 , se cumple $J(r) < 1$. En general, para valores pequeños de $r>0$, $F(r)<G(r)$ (o $J(r)<1$) indica agregación o clustering, y $F(r)>G(r)$ (o $J(r)> 1$) regularidad (Moller y Waagepetersen, 2004).

Máxima verosimilitud

Considere $x=\{x_1, \dots, x_n\}$ una realización del proceso puntual espacial Poisson X observada sobre la región W y la intensidad $\rho_\theta(u)$ que depende del vector de parámetros $\theta = (\theta_1, \dots, \theta_p)$, $\Theta \subseteq R^p$, para reflejar dependencia sobre una covariable espacial, entonces la función de log-verosimilitud ignorando la constante $n!$, es:

$$\log L(\theta; x; W) = \sum_{i=1}^n \log \rho_\theta(x_i) - \int_W \rho_\theta(u) du \tag{6}$$

El estimador de MV de θ se calcula utilizando algoritmos numéricos y la integral en (6) es aproximada a través de una suma ponderada $\int_W \rho_\theta(u) du \approx \sum_{(j=1)}^m \rho_\theta(u_j) a_j$, donde W es dividida en pequeños pixeles de igual área, conocido como dispositivo de Berman-Turner (*Berman-Turner device*), propuesto por Baddeley y Turner (2000). u_j es el centro del j -ésimo pixel. De esta forma, la sumatoria en (6) puede escribirse como $\sum_{(i=1)}^n \log \rho_\theta(x_i) = \sum_{(j=1)}^m n_j \log \rho_\theta(u_j)$, donde n_j es el número de puntos que caen en el j -ésimo pixel. Así la función log-verosimilitud (6) queda expresada como:

$$\log L(\theta) = \sum_{j=1}^m \{n_j \log \rho_j - \rho_j a_j\} \tag{7}$$

donde $\rho_j = \rho_\theta(u_j)$ y la expresión $\log L(\theta)$ tiene la misma

forma como el log-verosimilitud de una variable aleatoria n_j con distribución de Poisson y media $a_j \rho_\theta(u_j)$.

La intensidad es modelada a través de un modelo log-lineal de la forma, $\rho_j = \exp\{B(u_j) + \theta^T Z(u_j)\}$ ya que $\rho_j \geq 0$, ya que, lo cual se cumple en este modelo. Por otra parte considerando $c = \log a + B(u_j)$, el número esperado de puntos en un pixel de área a , es: $u_j = \exp\{c + \theta^T Z(u_j)\}$. Esta expresión corresponde a una regresión Poisson log-lineal con covariables $Z(u_j)$ y una variable de exposición, también denominada “offset”. Así, los coeficientes estimados θ para la regresión Poisson log-lineal son los estimadores aproximados de θ para el modelo de un proceso puntual espacial Poisson log-lineal.

Resultados y discusión

Mapeo de puntos

La representación gráfica del patrón puntual de hechos delictivos consistente en un mapa de puntos (en color azul) se muestra en la parte derecha de la Figura 1, donde es todo el espacio territorial ocupado por los distritos de Lima Centro y Residencial, delimitado por una línea de color gris.

A simple vista, se observan zonas con mayor concentración de puntos; por ejemplo, en los distritos de Lima Cercado y Lince. En el caso opuesto, es decir, las zonas con menor concentración de puntos son los distritos de La Molina y Santiago de Surco. Estos hechos dan una impresión de que el Proceso Puntual Espacial de hechos delictivos no sería completamente aleatorio, es decir, no existe homogeneidad en la distribución espacial de puntos.

Estimación de la intensidad

Asumiendo que el proceso puntual espacial es Completamente Aleatorio, la estimación de la intensidad según la expresión (3), es de 6.56 hechos delictivos por km^2 ; esto se detalla en la Tabla 1:

Tabla 1. Estadísticas de resumen del patrón puntual bidimensional homogéneo

Tipo	Frecuencia de puntos	Área de W (km ²)	Intensidad (puntos por km ²)
Hechos delictivos	1082	164.98	6.56

Pruebas de Aleatoriedad espacial completa (CSR)

Dado que los hechos delictivos parecieran no estar distribuidos homogéneamente en el espacio de estudio, se aplicó la prueba de aleatoriedad espacial completa, a partir de tres configuraciones de cuadrantes, basada en sus respectivos conteos. La hipótesis nula es: La intensidad es homogénea (CSR), mientras que la alterna corresponde a un proceso Poisson no homogéneo. Los resultados se muestran en la Tabla 2.

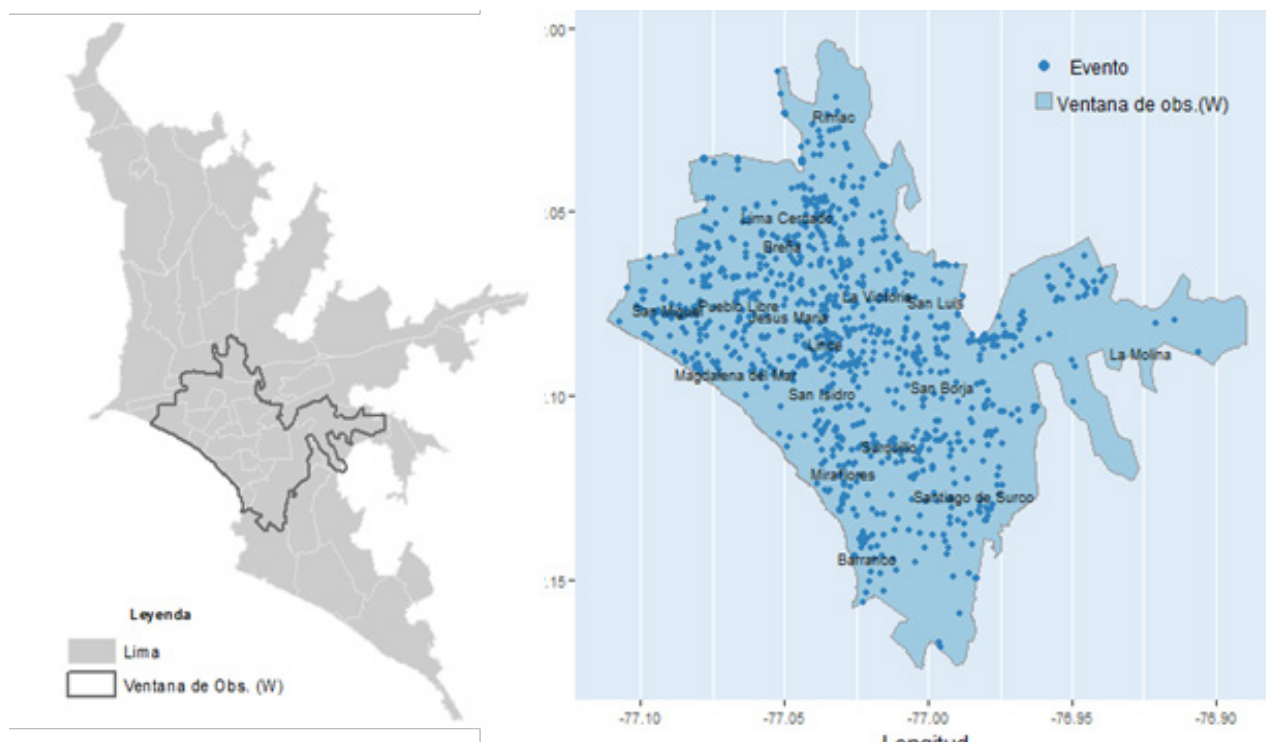


Figura 1. Patrón puntual bidimensional ($d=2$) de hechos delictivos, donde la ventana de observación W representa las zonas de Lima Centro y Residencial coloreadas en color celeste

Tabla 2. Pruebas de aleatoriedad espacial completa, basado en conteo de cuadrantes

Tipo	Nro. De polígonos	χ^2	Df	p-value
Polígonos dist.	17	333.27	16	2.2e-16
Cuadrados irreg.	12	232.85	11	2.2e-16
Dirichlet	9	194.62	8	2.2e-16

Dado que $p\text{-valor} < 0.01$, en cada uno de los tipos de cuadrantes, se rechaza la hipótesis nula. Es decir, existe suficiente evidencia estadística para afirmar que el patrón puntual de hechos delictivos es una realización de un proceso Poisson no homogéneo subyacente en la región de estudio; es decir, la intensidad es no homogénea. Gráficamente, la falta de homogeneidad puede observarse para cada tipo de cuadrante, en la Figura 2. Asimismo se muestran las etiquetas de los valores observados (izquierda), esperados y los residuales correspondientes. Los polígonos son representados en color gris de acuerdo a la intensidad de hechos delictivos, lo cual permite identificar los cuadrantes con mayor incidencia delictiva. Así, en el caso de polígonos distritales, estas se encuentran representadas en mayor grado en los distritos de Lima Cercado, Breña, Lince y Surquillo.

Otra prueba importante de bondad de ajuste, es la prueba de Kolmogorov – Smirnov, espacial (Baddeley *et al.*, 2016), las hipótesis son las mismas de CSR. Esta se realizó utilizando como covariable espacial a la coordenada de longitud, para el cual el estadístico calculado $D=0.157$ y un $p\text{-valor} < 0.01$, con lo cual se concluye que existe suficiente

evidencia estadística para mencionar que existe tendencia espacial a lo largo de la coordenada de longitud, y por tanto no existe CSR.

Después de haberse rechazado la hipótesis de falta de homogeneidad, se realizó la estimación de la intensidad de forma no paramétrica a través de un núcleo Gaussiano estándar (para una ancho de ventana, $h=400$ m). Los resultados se muestran en la Figura 3, en la cual la intensidad se superpuso por encima de la imagen satelital de Google. Se observa que las zonas reales con mayor incidencia de hechos delictivos, fueron los distritos de Lima Cercado y Lince. Por otra parte, en la barra de colores se observa que la intensidad fluctúa desde cero (en color azul) hasta un poco más de 30 hechos delictivos/km² (en color rojo), siendo estas las zonas con mayor incidencia delictiva. Estas fueron etiquetadas con las letras A, B y C. En la zona A, se observa que los hechos delictivos principalmente han ocurrido a lo largo de la Av. Arequipa hasta llegar al cruce con la Av. Juan de Zela, cuya intensidad fluctúa en el rango de 25 a 30 hechos delictivos/km², aproximadamente. En la zona B, se observa que los hechos delictivos principalmente han ocurrido en la zona conocida como el Trebol de Javier Prado con intensidades similares al caso A y finalmente la zona C corresponde a Lima Cercado. De esta forma, estimando la intensidad de un proceso puntual por el método no paramétrico descrito anteriormente y con la ayuda de las imágenes satelitales se logró identificar las zonas específicas con mayores incidencias de hechos delictivos. Por otro lado, el valor que toma esta nos ayuda a tener una idea de la magnitud del problema en cada una de las zonas identificadas.

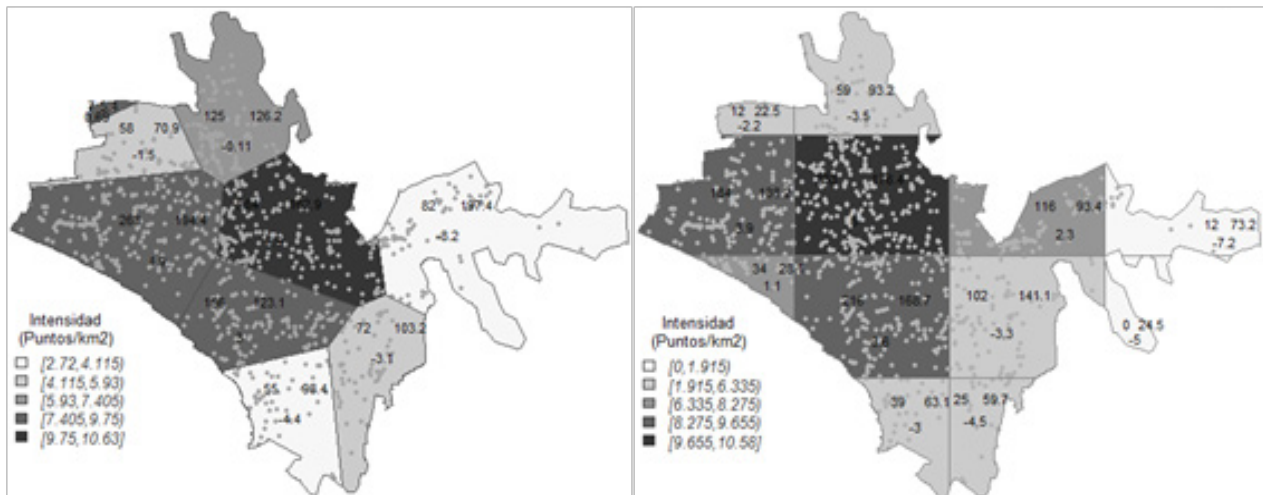
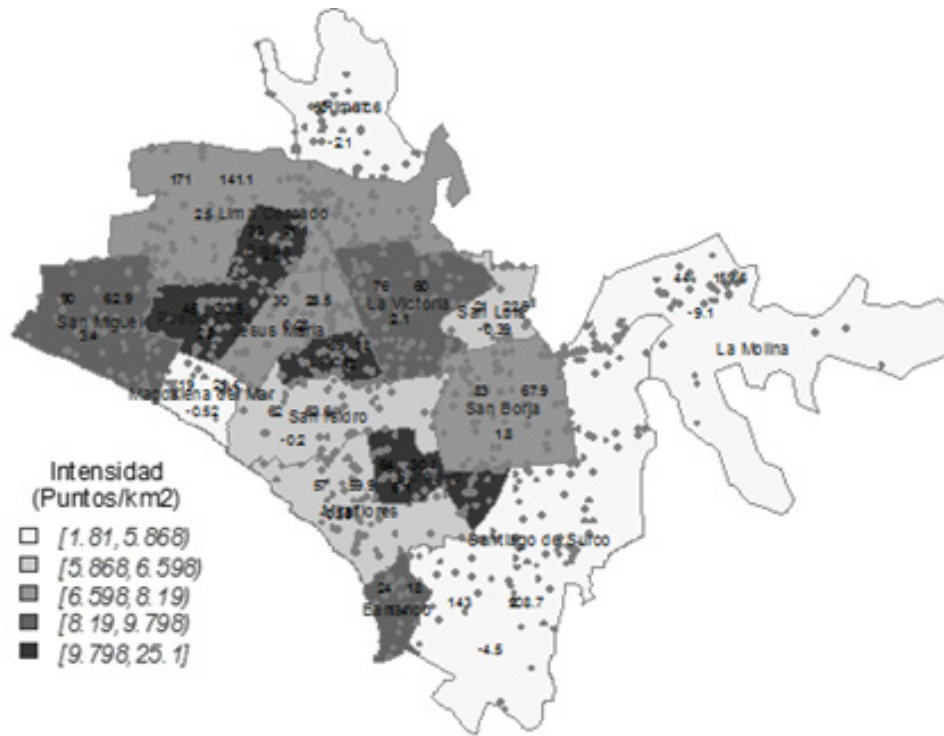


Figura 2. Mapa de la intensidad basado en cuadrantes. En la parte superior los cuadrantes están de acuerdo a la configuración distrital, en la parte inferior - izquierda para el caso de teselación de Dirichlet y al lado derecho para el caso de cuadrantes (tiles)

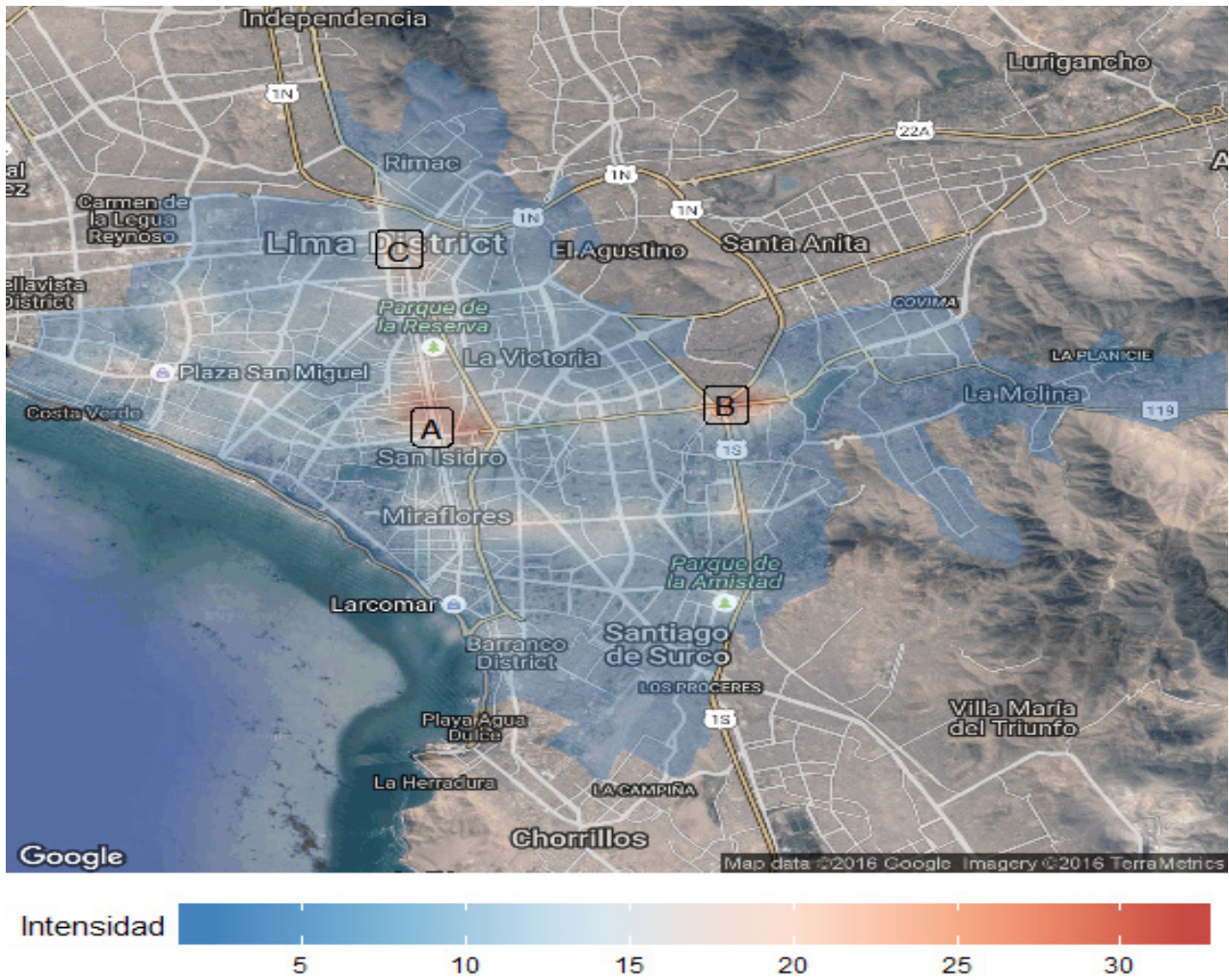


Figura 3. Mapa de la Intensidad de hechos delictivos (Puntos/km²) para para $h=0.4$ km, con fondo de imagen satélite Google Maps

Estadísticas de resumen de segundo orden

Las estadísticas de resumen de segundo orden tales como: la función- g de correlación por pares y la función $-L$ asumiendo homogeneidad y no homogeneidad, se muestran en la Figura 4, respectivamente. Los valores estimados de resultaron superiores a los valores teóricos $L(r)-L_{pois}>0$, lo cual indica la existencia de agregación o agrupamiento de hechos delictivos hasta una distancia menor a 1.5 km; sin embargo este agrupamiento es más notorio en distancias cercanas a 400 m. La existencia de agregación de puntos en ciertos lugares, indica que la intensidad es no homogénea y a su vez que existen zonas específicas que son las más preferidas por los delincuentes para delinquir.

De la misma forma, para el caso de la función $-g$ de correlación por pares (asumiendo homogeneidad y no homogeneidad), se observa que para valores menores a 0.4 km (es decir 400 metros), $g(r=0.4)>1$, lo cual confirma la existencia de agrupamientos de hechos delictivos y a su vez la falta de homogeneidad.

Las estadísticas de resumen basadas en las distancias

entre puntos se muestran en la Figura 5. La función $-G$ de distancia al vecino más cercano, se muestra en la parte superior-derecha; en esta se observa que la función teórica G_{pois} , se encuentra alejada de los valores estimados de G , lo cual indica que no existe CSR en el patrón puntual de hechos delictivos. Por otra parte, indica también que a menos de 250 metros, en el 80% de los casos ocurrió otro hecho delictivo (los hechos delictivos se dan en distancias cercanas unas de otras).

La función $-F$, representa la distribución de un punto o lugar cualesquiera de la región de estudio hasta el lugar más cercano donde se registró un hecho delictivo. Así, a partir de la Figura 4, se observa que para un punto arbitrario dentro de la ventana de observación y hasta una distancia de 300 metros, existía una probabilidad del 60% que se registre un hecho delictivo en él, durante el periodo de estudio. En general, los valores de las funciones $-F$ y G son altos a distancias cortas y además dado que $J(r)<1$, esto indica y corrobora la existencia de agregaciones en el patrón puntual de hechos delictivos.

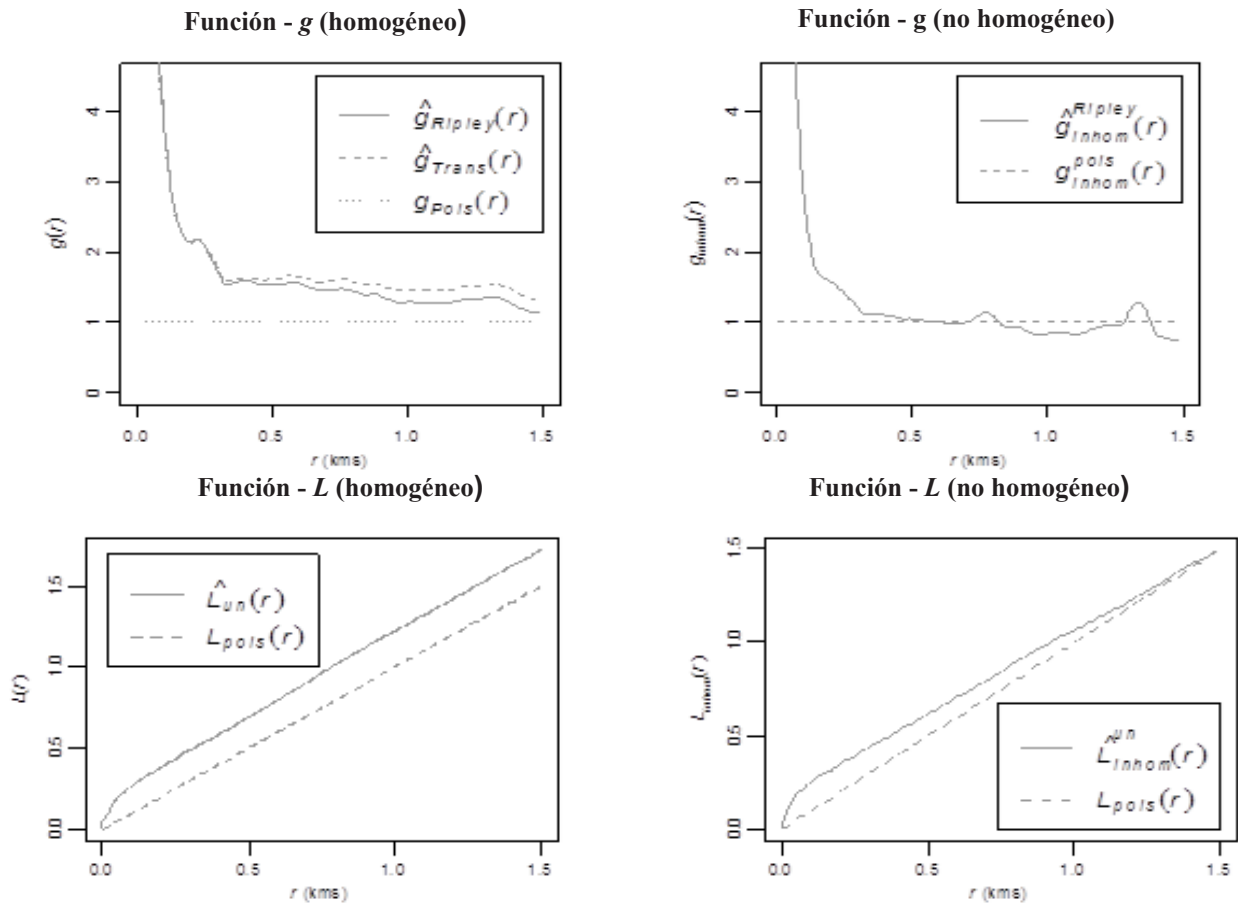


Figura 4. Función - g y L, para los datos de hechos delictivos. Al lado izquierdo para el caso homogéneo y el lado derecho para el caso no homogéneo.

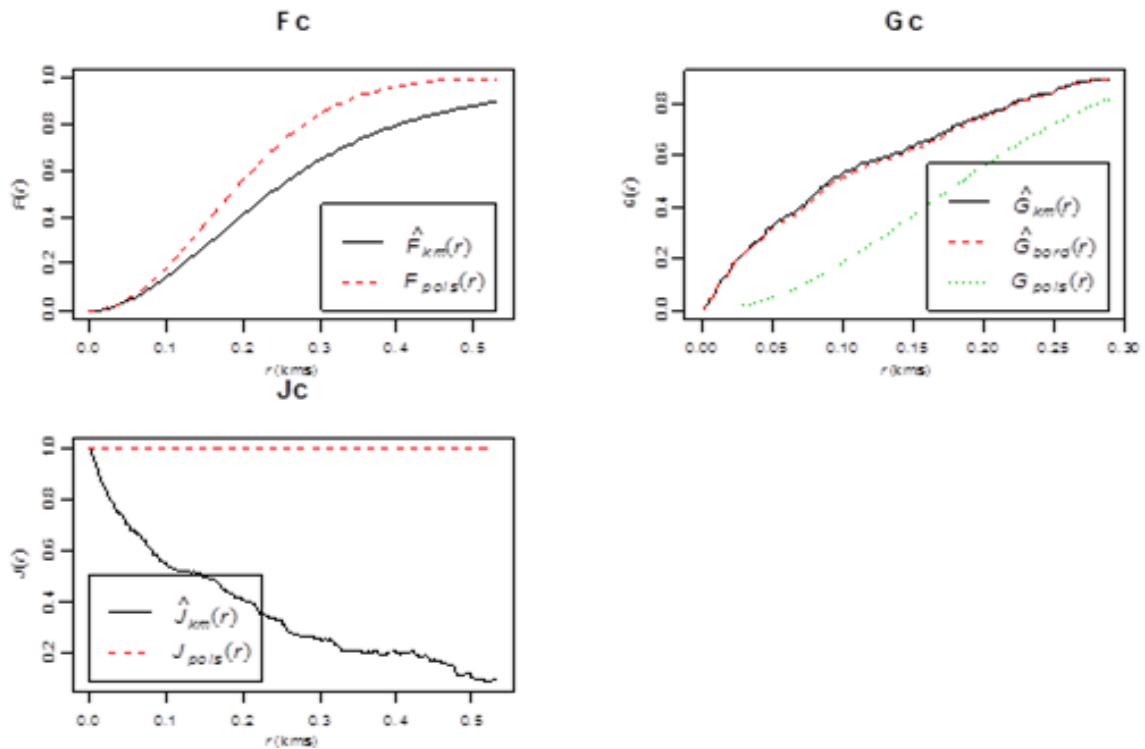


Figura 5. La función - F, G y J, basado en distancia entre puntos

Estimación de modelos

La variable dependiente ρ_i representa la intensidad de hechos delictivos; es decir, el número de delitos por km^2 registrado en un punto cualesquiera de W , que se asume son influenciados por un conjunto de covariables espaciales registradas en cada pixel de 20 metros de lado en el área de estudio W . Estas se presentan en la Figura 6, donde: Z_1 y Z_2 son la distancia (km) de un punto cualquiera a un puesto de seguridad más cercano y a un límite distrital, respectivamente; Z_3 es el Presupuesto Institucional per cápita (en nuevos soles) destinado al orden interno y Z_4 representa la densidad poblacional existente en un punto cualquiera.

Estas llevan superpuestos los puntos del patrón puntual de hechos delictivos, con el fin de verificar su dependencia con respecto a las covariables espaciales. En el caso de las distancias hacia una comisaría, se observa que los hechos delictivos se encuentran en su mayoría alejados de esta, actuando estas como puntos de repulsión. De forma inversa ocurre con límites distritales en donde los puntos tienden a situarse a distancias reducidas. Por otra parte, se visualiza que los puntos tienden a situarse en zonas donde el presupuesto destinado al orden interno es menor y donde se registra una densidad poblacional alta.

Para evaluar la dependencia espacial de los hechos delictivos con respecto a las covariables espaciales se

formularon diversos modelos estadísticos cuyos parámetros fueron estimados por Máxima Verosimilitud. Para el caso de la dependencia espacial de una covariable a la vez, estas se describen en el Tabla 3, mientras para el caso múltiple en el Tabla 4. En el caso del proceso Poisson homogéneo (CSR), la intensidad es de la forma $p(u,x)=e^{(1.880741)}=6.558363$ y representa el número de hechos delictivos por para la región de estudio. Por definición de este modelo, se asume que la intensidad es homogénea y por tanto no depende de la ubicación. Con respecto a los modelos (1) - (4) con dependencia de una covariable espacial, el modelo de proceso Poisson no homogéneo tiene la siguiente forma $\rho(u,x)=e^{(\beta_0+\beta_1 Z_1)}$; así, para el modelo (2) se tiene:

$$\rho(u,x)=e^{2.366-1.057Z_1}$$

Respecto a la significancia de los coeficientes de los modelos, del Tabla 3, estos resultaron altamente significativos para cada una de las covariables (a excepción de la distancia hacia una centro de seguridad), lo cual indica que la intensidad de hechos delictivos, dependen de estas covariables. Con respecto a los signos de la covariables espaciales que influyen significativamente al fenómeno de estudio, se concluye que cuanto más alejado este un punto de los límites distritales, la intensidad de hechos

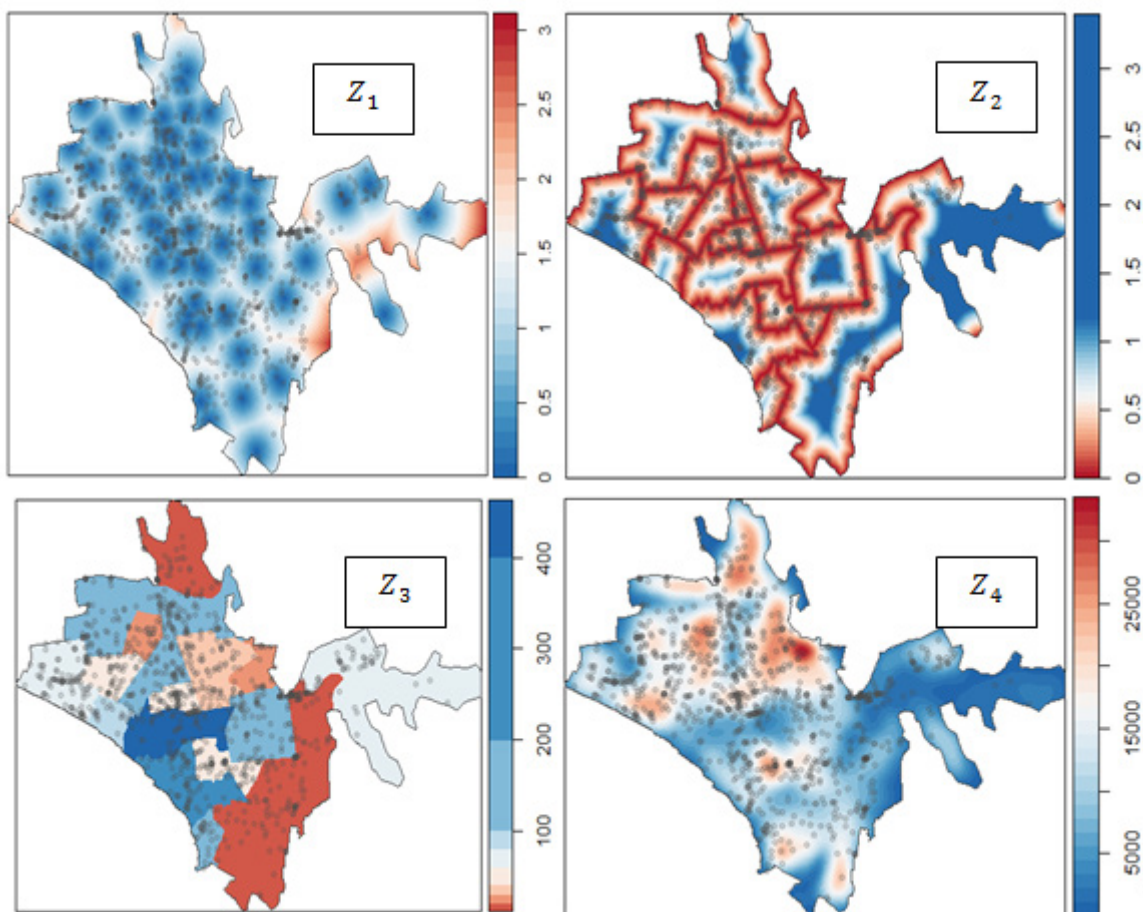


Figura 6. Mapas de covariables espaciales donde se superpone el patrón puntual de hechos delictivos (Fuente: Elaboración propia)

delictivos son menores; es decir, los hechos delictivos ocurren en zonas cercanas a los límites distritales. Esto presumimos ocurre porque las comisarías de los distritos velan solo por la seguridad de sus ciudadanos, la cual resulta beneficioso para los delincuentes al lograr evadirse de los arrestos.

Para el caso de la dependencia espacial múltiple en un proceso Poisson no homogéneo, se formularon cuatro modelos que van de (5) – (8), y se muestran en el Tabla 4.

El modelo que incluye todas las covariables espaciales fue especificado en el modelo (5), en el cual resultaron no significativos los coeficientes asociados a las covariables Z_1 y Z_3 , por lo cual se procedió a estimar los modelos (6) y (7) en la cual se elimina una covariable en cada caso.

En ambos modelos, estas covariables resultaron no ser significativas por lo que se procedió a estimar el modelo (8) que no incluyen a dichas covariables. Así en el modelo (8) todos sus coeficientes resultaron altamente significativo tanto para Z_2 y Z_4 . Esta serie de modelos se plantearon con el fin de estimar el modelo más adecuado que se ajuste a los datos del patrón puntual.

En el caso del modelo (8), se estimó un modelo clúster de Neyman – Scott (Waagepetersen, 2007), descrito en el Tabla 5, donde la forma de la intensidad para este modelo es la misma que del modelo (8). La diferencia la hacen los parámetros del clúster, “kappa” y “scale”, estimadas a partir de la función – K paramétrica y a través del método de *Mínimo Contraste*.

Tabla 3. Coeficientes del modelo log - lineal univariado para el caso de un proceso espacial Poisson no homogéneo

Modelo	Variabes	Estimate	S.E.	CI95.lo	CI95.hi	Ztest	Zval
Poisson Homogéneo (0)	(Intercept)	1.880741	0.030401	1.821157	1.940326	***	61.86467
Poisson No homogéneo (1)	(Intercept)	1.880806	0.030628	1.820776	1.940837	***	61.40768
	$Z_1 (<0.11)$	-0.001073	0.251869	-0.49473	0.492582		-0.00426
Poisson No homogéneo (2)	(Intercept)	2.366108	0.043981	2.279907	2.452309	***	53.79854
	Z_2	-1.051617	0.085483	-1.219161	-0.884073	***	-12.30203
Poisson No homogéneo (3)	(Intercept)	1.981458	0.044368	1.894498	2.068417	***	44.65979
	$Z_3 (>60)$	-0.182021	0.060915	-0.30141	-0.06263	**	-2.98810
Poisson No homogéneo (4)	(Intercept)	1.423270	0.064411	1.297026	1.549514	***	22.09657
	Z_4	0.000038	0.000004	0.000029	0.000047	***	8.65450

Tabla 4. Coeficientes del modelo log - lineal múltiple para el caso de un proceso espacial Poisson no homogéneo

Modelo	Variabes	Estimate	S.E.	CI95.lo	CI95.hi	Ztest	Zval
Poisson No homogéneo (5)	(Intercept)	2.024665	0.091209	1.845898	2.203431	***	22.19806
	$Z_1 (<0.11)$	-0.092417	0.252170	-5.87E-01	4.02E-01		-0.36649
	Z_2	-1.004802	0.087974	-1.177227	-0.832376	***	-11.4216
	$Z_3 (>60)$	-0.011861	0.063282	-1.36E-01	1.12E-01		-0.18743
	Z_4	0.000027	0.000005	0.000018	0.000036	***	5.672683
Poisson No homogéneo (6)	(Intercept)	2.024694	0.091203	1.845941	2.203448	***	22.19998
	Z_2	-1.004844	0.087987	-1.177295	-0.832393	***	-11.4204
	$Z_3 (>60)$	-0.012677	0.063244	-1.37E-01	1.11E-01		-0.20045
	Z_4	0.000027	0.000005	0.000018	0.000036	***	5.662029
Poisson No homogéneo (7)	(Intercept)	2.015458	0.076331	1.865853	2.165064	***	26.404309
	Z_1	-0.093998	0.252026	-0.587960	0.399965		-0.372968
	Z_2	-1.004982	0.087974	-1.177408	-0.832556	***	-11.423610
	$Z_4 (>60)$	0.000027	0.000005	0.000018	0.000036	***	5.937187
Poisson No homogéneo (8)	(Intercept)	2.014646	0.076305	1.865091	2.164201	***	26.40252
	Z_2	-1.004815	0.087978	-1.177248	-0.832381	***	-11.4212
	Z_4	0.000027	0.000005	0.000018	0.000036	***	5.929459

Tabla 5. Coeficientes del modelo log – lineal, para el caso de un proceso espacial clúster Neyman - Scott (no homogéneo)

Modelo	Variabes	Estimate	S.E.	CI95.lo	CI95.hi	Ztest	Zval
Modelo clúster	(Intercept)	2.014646	0.233457	1.557080	2.472213	***	8.629639
	Z_2	-1.004815	0.205377	-1.407347	-0.602283	***	-4.892532
	Z_4	0.000027	0.000015	-0.000003	0.000057		1.753346
	Kappa	0.399605					
	Scale	0.203248					

Selección del modelo

Se realizó con la prueba de razón de verosimilitud, a través de comparaciones entre pares de modelos adecuados. Estos corresponden a los modelos univariados (2) y (4) versus el modelo (8). Los resultados se muestran en el Tabla 6.

Tabla 6: Pruebas de razón de verosimilitud

Comparación de modelos	Df	Deviance	Pr(>Chi)
(2) vs (8)	1	34.71200	0.00000 ***
(4) vs (8)	1	165.42000	0.00000 ***

0 **** 0.001 *** 0.01 * 0.05 . 0.1 ' ' 1

En ambos casos el modelo (8) presentó una diferencia significativa alta respecto a los modelos univariados (2) y (4), determinándose así el modelo (8) como el mejor ya que presenta un mejor ajuste a los datos y corresponde al modelo de proceso puntual espacial Poisson no homogéneo, de la siguiente forma:

$$\hat{\rho}(u) = e^{2.014646 - 1.004815Z_2 + 0.000027Z_4}$$

donde Z_2 y Z_4 representan: la distancia (km) a un límite distrital y la densidad poblacional (población/ km²), respectivamente.

4. Conclusiones

De acuerdo a las pruebas de CSR, se determinó que el patrón puntual de hechos delictivos, se ajusta de mejor forma a un modelo de proceso puntual espacial Poisson no homogéneo. Así, el modelo elegido teniendo en cuenta la prueba de razón de verosimilitud fue el modelo log-lineal, que incluye como covariables espaciales a: la distancia al límite distrital y la densidad poblacional.

Los hechos delictivos se encuentran distribuidos de forma no homogénea a lo largo de la ciudad de Lima Centro y Residencial, donde existen zonas con mayor incidencia de hechos delictivos, tales como los distritos de Lima Cercado y Lince. Particularmente también se identificaron puntos específicos donde la incidencia de hechos delictivos es muy alta tales como lo ocurrido en la Av. Arequipa hasta llegar al cruce con la Av. Juan de Zela, cuya intensidad fluctúa en el rango de 25 a 30 hechos delictivos / km², aproximadamente. Otra zona corresponde a la zona conocida como el Trebol de Javier Prado con intensidades similares al primer caso.

A partir de los modelos Poisson univariado, se concluyó: que en los lugares con mayor densidad poblacional existe mayor ocurrencia de hechos delictivos, una mayor inversión en orden interno efectuado por los Municipios, favorece a la reducción de la delincuencia; los límites jurisdiccionales (distritales) favorecen al incremento de los hechos delictivos. Esto corrobora muchas afirmaciones periodísticas que muestran como las barreras

jurisdiccionales existen y resultan beneficiosos para los delincuentes en el momento de evasión y perjudican a las víctimas al momento de reportar el hecho delictivo y su denuncia, dado que deben realizarla en una comisaría de su jurisdicción.

No se encontró evidencia estadística, para afirmar que la distribución espacial de las comisarías y serenazgos, tiene efecto reductor en los hechos delictivos.

En general, las ocurrencias de hechos delictivos no solo obedecen a un factor sino a una conjunto de estos (considerándose también aquellos que no fueron incluidos en el modelo) y que involucran la participación de diversos sectores desde los distintos niveles de gobiernos tanto locales, nacional y de los ciudadanos en general.

Desde un punto de vista metodológico, el análisis estadístico de patrones puntuales espaciales se constituye en una herramienta valiosa para en la lucha contra la delincuencia, ya que no solo permite identificar los lugares donde estos ocurren, sino también porque detecta las zonas de mayor incidencia delictivas a través de la intensidad y usa las covariables espaciales relacionadas a este problema a través de los modelos estadísticos de procesos puntuales espaciales.

5. Recomendaciones

Para futuros trabajos, se recomienda emplear modelos de procesos puntuales espaciales marcados multivariados para modelar los hechos delictivos considerando sus distintas modalidades (arrebato, bujiazco, cogoteo, raqueteo), dado que se podría esperar resultados distintos en cada una de sus modalidades.

Incorporar la dimensión temporal, empleando modelos de procesos puntuales espacio-temporales, como una alternativa para predecir los patrones de hechos delictivos en un determinado espacio.

6. Literatura citada

- Baddeley, A. J; Moller, J; Waagepetersen, R. 2000. Non-an semi-parametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica* (2000), 54(3): 329-350.
- Baddeley, A. J; Turner, R. 2000. Practical maximum pseudolikelihood for spatial point patterns (with discussion). *Australian and New Zealand Journal of Statistics*, 42(3):283-322.
- Baddeley, A; Turner, R; Moller J; Hazelton, M. 2005. Score, Pseudo-Score and Residual Diagnostics for Spatial Point Process Models.
- Baddeley, A. 2008. *Analysing spatial point patterns* in R, CSIRO and University of Western Australia.
- Baddeley, A; Rubak, E; Turner, R. 2016. *Spatial Point Patterns Methodology and Applications with R*. CRC Press Chapman & Hall CRC BOOK.
- Cressie, N. A. C. 1991. *Statistics for Spatial Data*, 1 edn, John Wiley & Sons.

- Diggle, P. J. 2014. *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns*, 2 edn. CRC Press A Chapman and Hall, Boca Raton.
- Lloyd, C. 2007. *Local Models for Spatial Analysis*, 1 edn, CRC Press.
- MLM-Municipalidad de Lima Metropolitana, PE. 2016. *Plan Regional de Seguridad Ciudadana de Lima Metropolitana 2016*. GSGC Gerencia de Seguridad Ciudadana.
- Moller, J; Waagepetersen, R. P. 2004. *Statistical Inference and Simulation for Spatial Point Processes*, 1 edn, Chapman & Hall CRC.
- Schabenberger, O; Gotway, C. 2005. *Statistical Methods for Spatial Data Analysis*, 1 edn, Chapman & Hall CRC. Boca Raton London NewYork Washington, D.C.
- Stoyan, D; Stoyan, H. 1994. *Fractals, Random Shapes and Point Fields*, Wiley, Chichester.
- Waagepetersen, R. P. 2007. An Estimating Function Approach to Inference for Inhomogeneous Neyman-Scott Processes. *Biometrics*, 63(1): 252-258.