

BOLETÍN CRIMINOLÓGICO

Nº 100

Instituto andaluz interuniversitario de Criminología

SEPTIEMBRE - OCTUBRE 2007

La aplicación de técnicas epidemiológicas en el estudio de determinados delitos, especialmente aquellos contra la propiedad, ha demostrado que éstos se aglomeran espacio-temporalmente. Es decir, una vez que un determinado delito tiene lugar es más probable que otro ocurra a poca distancia y poco después del inicial. La identificación de estas pautas espacio-temporales nos permite trazar mapas predictivos, muy útiles para contribuir a la prevención del delito y a la detección y detención de los delincuentes. Este trabajo metodológico, realizado por una profesora de UCL Hill Dando Institute of Crime Science de Londres, describe dos de las técnicas epidemiológicas más utilizadas en el estudio de las pautas espacio-temporales del delito: la tabla de contingencia de Knox y la simulación de Monte Carlo.

Palabras clave: pautas espacio-temporales, mapas predictivos, técnicas epidemiológicas, prevención del delito, detección del delito.

LA MEDICIÓN ESTADÍSTICA DE LAS PAUTAS ESPACIO-TEMPORALES DEL DELITO

Introducción

La identificación de las pautas espacio-temporales del delito es importante para el trazado de mapas predictivos que nos ayuden a predecir cuándo y dónde el delito ocurrirá en el futuro y así poder mejorar su prevención y detección. Hasta ahora, los mapas de puntos calientes ('hotspot maps') utilizados por la policía han sido mapas históricos, basados en información sobre delitos ocurridos en un período inmediatamente anterior al que queremos predecir. El problema es que los puntos calientes son en muchos casos fluidos (i.e. no estables con el paso del tiempo), lo que disminuye la fiabilidad de estos mapas. Es por esto que recientemente determinados estudios empíricos han empezado a estudiar las pautas espacio-temporales del delito, con el objeto de observar la fluidez de los puntos calientes y determinar si ésta se ajusta a ciertas regularidades.

Para el estudio de las pautas espacio-temporales se han empleado técnicas originalmente desarrolladas para el estudio de enfermedades cuya etiología no estaba establecida. Estas técnicas determinan si las incidencias son independientes o están relacionadas entre sí. Utilizando estas técnicas se ha podido ver que existen aglomeraciones espacio-temporales en el caso de robos de domicilio y sustracciones de objetos en vehículos. Por

ejemplo, se ha demostrado empíricamente cómo, tras la comisión de un robo en un domicilio, el riesgo de que otro robo se produzca aumenta de forma significativa en un radio de 400 metros durante el siguiente mes. Quizás lo más interesante es que estas pautas son predecibles, lo que permite el trazado de mapas predictivos que superan en fiabilidad a los mapas históricos utilizados actualmente. Este artículo describe dos de las técnicas epidemiológicas utilizadas en el estudio de las pautas espacio-temporales del delito: la tabla de contingencia de Knox y la simulación de Monte Carlo.

Ilustración

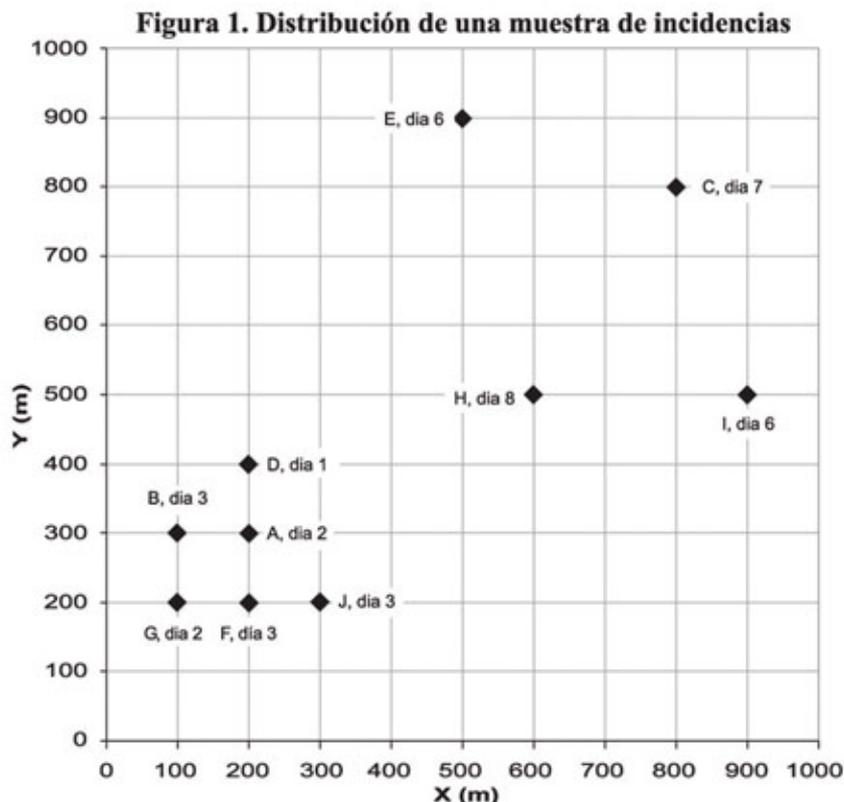
Una muestra simplificada de la medición estadística de las pautas espacio-temporales del delito se presenta en la figura. Las diez incidencias que se contemplan en la figura se denominan con las letras del alfabeto, de la A a la J. La posición de estas incidencias en el gráfico indica su posición geográfica a partir del punto de referencia (0,0). Cada punto en el gráfico está etiquetado con la letra asignada a la incidencia que representa y el día en que la incidencia ocurrió (a partir de un punto de referencia $t=0$). Por ejemplo, la incidencia D ocurrió en el punto de referencia geográfico (200,400) el primer día. Una inspección visual de la



figura indica que hay una aglomeración espacial en el espacio cuadrangular limitado por las coordenadas (100,200), (100,400), (300,200) y (300,400). Las seis incidencias que tienen lugar en este espacio también parecen estar aglomeradas temporalmente, habiendo ocurrido en un espacio de sólo tres días. Esto parece indicar que nuestra muestra posee una aglomeración espacio-temporal.

Técnicas utilizadas para la medición estadística de las pautas espacio-temporales de delitos

Aunque la anterior figura nos aporta bastante información sobre la distribución espacio-temporal de un grupo de incidencias, normalmente es necesario medir esta aparente aglomeración estadísticamente para determinar la magnitud de la misma y establecer su significado estadístico. Las técnicas disponibles para la medición de pautas espacio-temporales suponen la comparación de las distribuciones observadas y esperadas de las distancias espacio-temporales entre todos los posibles pares de incidencias. Para un total de n incidencias,



habría $(n(n-1))/2!$ pares de incidencias o combinaciones. En nuestro caso, con $n=10$, el número total de posibles parejas de incidencias sería 45.

El cálculo de las distancias observadas es fácil. Lo único que hay que hacer es restar las distancias entre las incidencias que componen cada pareja. Por ejemplo, la distancia temporal entre las incidencias A y C es de

6 días (i.e. la diferencia más uno, ya que es posible que la incidencia A haya ocurrido al empezar el día 2 y la C al terminar el día 7; este método es más conservador que utilizar simplemente la diferencia entre los dos días) y la distancia espacial es de 781.02 metros (i.e.). Las distancias temporales y espaciales entre todos los posibles pares de incidencias se muestran en la tabla 1.

Tabla 1. Distancias espaciales (d; en metros) y temporales (t; en días) entre todos los posibles pares de incidencias.

| d(m) t(días) | d(m) t(días) | d(m) t(días) |
|---------------|---------------|---------------|
| A - B 100.0 2 | B - I 824.6 4 | E - F 761.6 4 |
| A - C 781.0 6 | B - J 223.6 1 | E - G 806.2 5 |
| A - D 100.0 2 | C - D 721.1 7 | E - H 412.3 3 |
| A - E 670.8 5 | C - E 316.2 2 | E - I 565.7 1 |
| A - F 100.0 2 | C - F 848.5 5 | E - J 728.0 4 |
| A - G 141.4 1 | C - G 922.0 6 | F - G 100.0 2 |
| A - H 447.2 7 | C - H 360.6 2 | F - H 500.0 6 |
| A - I 728.0 5 | C - I 316.2 2 | F - I 761.6 4 |
| A - J 141.4 2 | C - J 781.0 5 | F - J 100.0 1 |
| B - C 860.2 5 | D - E 583.1 6 | G - H 583.1 7 |
| B - D 141.4 3 | D - F 200.0 3 | G - I 854.4 5 |
| B - E 721.1 4 | D - G 223.6 2 | G - J 200.0 2 |
| B - F 141.4 1 | D - H 412.3 8 | H - I 300.0 3 |
| B - G 100.0 2 | D - I 707.1 6 | H - J 424.3 6 |
| B - H 538.5 6 | D - J 223.6 3 | I - J 670.8 4 |

La tabla de contingencia de Knox

El método más sencillo para determinar si hay una aglomeración espacio-temporal estadísticamente significativa consistiría en producir una tabla de contingencia de Knox. Knox publicó un artículo en 1964 en el que aplicaba estos métodos al estudio de la leucemia infantil, ya que en aquellos tiempos la etiología de esta enfermedad era poco conocida. Desarrolló un método por el cual se calculan las distancias espacio-temporales de todos los posibles pares de incidencias (como se indicó anteriormente, para un total de n

incidencias, habría $(n(n-1))/2!$ parejas de incidencias o combinaciones). Después de categorizar las distancias geográficas y temporales en intervalos fijos, se crea una tabla de contingencia para contar el número de combinaciones en cada categoría. En nuestro ejemplo, las distancias espaciales se podrían categorizar en intervalos de 100 metros y las temporales en intervalos de un día. De este modo, podemos observar que hay seis pares de incidencias a 100 metros o menos de distancia, de los cuales uno ocurrió con un día de diferencia y los otros cinco con dos (ver tabla 2).

El significado estadístico de estas frecuencias se puede

Tabla 2. Tabla de contingencia de Knox, en la que se muestran las frecuencias observadas.

| | t (días) | | | | | | | | Total |
|--------------------|----------|-----------|----------|----------|----------|----------|----------|----------|-----------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | |
| 0 < d -'3d 100m | 1 | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 6 |
| 100 < d -'3d 200m | 2 | 2 | 2 | 0 | 0 | 0 | 0 | 0 | 6 |
| 200 < d -'3d 300m | 1 | 1 | 2 | 0 | 0 | 0 | 0 | 0 | 4 |
| 300 < d -'3d 400m | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 3 |
| 400 < d -'3d 500m | 0 | 0 | 1 | 0 | 0 | 2 | 1 | 1 | 5 |
| 500 < d -'3d 600m | 1 | 0 | 0 | 0 | 0 | 2 | 1 | 0 | 4 |
| 600 < d -'3d 700m | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 2 |
| 700 < d -'3d 800m | 0 | 0 | 0 | 4 | 2 | 2 | 1 | 0 | 9 |
| 800 < d -'3d 900m | 0 | 0 | 0 | 1 | 4 | 0 | 0 | 0 | 5 |
| 900 < d -'3d 1000m | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| Total | 5 | 11 | 5 | 6 | 7 | 7 | 3 | 1 | 45 |

positiva entre las distancias espaciales y temporales (i.e. mientras mayor sea la distancia espacial, mayor será la temporal,

chi cuadrado nos proporciona un valor de probabilidad para la tabla

Tabla 5. Tabla de contingencia de Knox con las categorías agregadas en la que se muestran las frecuencias observadas (o) y las esperadas (e).

| | t (días) | | | |
|-----------------|----------|------|-------|-----|
| | 1 - 4 | | 5 - 8 | |
| | o | e | o | e |
| 0 < d ≤ 500m | 20 | 14.4 | 4 | 9.6 |
| 500 < d ≤ 1000m | 7 | 12.6 | 14 | 8.4 |

Tabla 3. Tabla de contingencia de Knox, en la que se muestran las frecuencias observadas (o) y las esperadas (e).

| | t (Días) | | | | | | | | | | | | | | | |
|-----------------|----------|-----|---|-----|---|-----|---|-----|---|-----|---|-----|---|-----|---|-----|
| | 1 | | 2 | | 3 | | 4 | | 5 | | 6 | | 7 | | 8 | |
| | o | e | o | e | o | e | o | e | o | e | o | e | o | e | o | e |
| 0 < d ≤ 100m | 1 | 0.7 | 5 | 1.5 | 0 | 0.7 | 0 | 0.8 | 0 | 0.9 | 0 | 0.9 | 0 | 0.4 | 0 | 0.1 |
| 100 < d ≤ 200m | 2 | 0.7 | 2 | 1.5 | 2 | 0.7 | 0 | 0.8 | 0 | 0.9 | 0 | 0.9 | 0 | 0.4 | 0 | 0.1 |
| 200 < d ≤ 300m | 1 | 0.4 | 1 | 1.0 | 2 | 0.4 | 0 | 0.5 | 0 | 0.6 | 0 | 0.6 | 0 | 0.3 | 0 | 0.1 |
| 300 < d ≤ 400m | 0 | 0.3 | 3 | 0.7 | 0 | 0.3 | 0 | 0.4 | 0 | 0.5 | 0 | 0.5 | 0 | 0.2 | 0 | 0.1 |
| 400 < d ≤ 500m | 0 | 0.6 | 0 | 1.2 | 1 | 0.6 | 0 | 0.7 | 0 | 0.8 | 2 | 0.8 | 1 | 0.3 | 1 | 0.1 |
| 500 < d ≤ 600m | 1 | 0.4 | 0 | 1.0 | 0 | 0.4 | 0 | 0.5 | 0 | 0.6 | 2 | 0.6 | 1 | 0.3 | 0 | 0.1 |
| 600 < d ≤ 700m | 0 | 0.2 | 0 | 0.5 | 0 | 0.2 | 1 | 0.3 | 1 | 0.3 | 0 | 0.3 | 0 | 0.1 | 0 | 0.0 |
| 700 < d ≤ 800m | 0 | 1.0 | 0 | 2.2 | 0 | 1.0 | 4 | 1.2 | 2 | 1.4 | 2 | 1.4 | 1 | 0.6 | 0 | 0.2 |
| 800 < d ≤ 900m | 0 | 0.6 | 0 | 1.2 | 0 | 0.6 | 1 | 0.7 | 4 | 0.8 | 0 | 0.8 | 0 | 0.3 | 0 | 0.1 |
| 900 < d ≤ 1000m | 0 | 0.1 | 0 | 0.2 | 0 | 0.1 | 0 | 0.1 | 0 | 0.2 | 1 | 0.2 | 0 | 0.1 | 0 | 0.0 |

entonces determinar al comparar éstas con las frecuencias esperadas, calculadas sobre los totales marginales (test de chi cuadrado). En nuestro ejemplo (ver tabla 3) se observa que las frecuencias observadas en el caso de pares de incidencias que ocurren cerca, espacial y temporalmente, es a menudo mayor que lo que se esperaría en una distribución aleatoria (p. e. el número de pares de incidencias que ocurrieron a 100 metros o menos y de a uno a dos días de distancia es cinco, mientras que la frecuencia esperada estadísticamente es 1.5). Los cocientes entre las frecuencias observadas y esperadas muestran más claramente una asociación

y viceversa; ver tabla 4). El test de chi cuadrado nos confirma que estas diferencias son significativas estadísticamente ($X^2(63)=109.12, p<.01$). Sin embargo, este método presenta al menos dos problemas. El primero es un problema de interpretación. Como el test de

de contingencia total, no es posible saber con exactitud en qué categorías específicas (p. e. a 100 metros o menos y de a uno a dos días de distancia) las diferencias entre las frecuencias observadas y esperadas son significativas. El segundo problema es más grave. El test de chi cuadrado requiere: a) que no haya ninguna categoría en la tabla de contingencia con una frecuencia esperada menor que uno; b) que no haya más del 25% de las categorías con una frecuencia esperada menor que cinco; y c) que las observaciones

Tabla 4. Tabla de contingencia de Knox, en la que se muestran los cocientes entre las frecuencias observadas y las esperadas (el sombreado más oscuro indica mayores cocientes)

| | t (Días) | | | | | | | |
|-----------------|----------|-----|-----|-----|-----|-----|-----|-----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 0 < d ≤ 100m | 1.5 | 3.4 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 100 < d ≤ 200m | 3.0 | 1.4 | 3.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 200 < d ≤ 300m | 2.3 | 1.0 | 4.5 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 300 < d ≤ 400m | 0.0 | 4.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 400 < d ≤ 500m | 0.0 | 0.0 | 1.8 | 0.0 | 0.0 | 2.6 | 3.0 | 9.0 |
| 500 < d ≤ 600m | 2.3 | 0.0 | 0.0 | 0.0 | 0.0 | 3.2 | 3.8 | 0.0 |
| 600 < d ≤ 700m | 0.0 | 0.0 | 0.0 | 3.8 | 3.2 | 0.0 | 0.0 | 0.0 |
| 700 < d ≤ 800m | 0.0 | 0.0 | 0.0 | 3.3 | 1.4 | 1.4 | 1.7 | 0.0 |
| 800 < d ≤ 900m | 0.0 | 0.0 | 0.0 | 1.5 | 5.1 | 0.0 | 0.0 | 0.0 |
| 900 < d ≤ 1000m | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 6.4 | 0.0 | 0.0 |

sean independientes. En nuestro caso, ninguna de estas condiciones se cumple, lo que invalida los resultados del test.

Las dos primeras condiciones se pueden solucionar al agregar las categorías (ver tabla 5; $X^2(1)=11.67$, $p<.001$), aunque esto supone la pérdida de detalle. La tercera condición del test es más difícil de cumplir. Ésta se

que luego se asocian con los puntos geográficos, cuyo orden se mantiene constante. Esta simulación se puede realizar por medio de una simple programación informática.

Se utiliza una muestra porque el número de permutaciones sería igual al factorial del número de incidencias, el cual tiende al infinito al crecer el número de incidencias (ver tabla 6).

De este modo, podemos crear una tabla de contingencia como la anterior, donde se comparen las frecuencias observadas con las medias de las frecuencias generadas en las simulaciones (ver tabla 7). Si la frecuencia observada es mayor que la generada en el

99% de las permutaciones, nuestro resultado para esa casilla será significativo al nivel .01. Los cocientes entre las frecuencias observadas y esperadas se muestran en la tabla 8.

Los resultados obtenidos por este método son así más fiables, por no infringir ninguna condición del test. Además, nos permiten saber con exactitud cuáles son las categorías que resultan significativas estadísticamente, y así poder interpretar los resultados más fácilmente. También es importante señalar que el hecho de sólo utilizar una muestra de las permutaciones no afecta considerablemente las frecuencias esperadas.

Tabla 6. Factorial de n (n!) para varios valores de n.

| n | n! |
|-------|--------------------------------------|
| 0 | 1 |
| 1 | 1 |
| 2 | 2 |
| 3 | 6 |
| 4 | 24 |
| 5 | 120 |
| 6 | 720 |
| 7 | 5040 |
| 8 | 40320 |
| 9 | 362880 |
| 10 | 3628800 |
| 15 | 1307674368000 |
| 20 | 2432902008176640000 |
| 25 | 15511210043330985984000000 |
| 50 | $3.04140932... \times 10^{64}$ |
| 70 | $1.19785717... \times 10^{100}$ |
| 450 | $1.73368733... \times 10^{1,000}$ |
| 3249 | $6.41233768... \times 10^{10,000}$ |
| 25206 | $1.205703438... \times 10^{100,000}$ |

infringe al emparejar todas las incidencias entre ellas, lo que hace que los datos sean dependientes estadísticamente.

Simulación de Monte Carlo

Una alternativa sería utilizar simulaciones de Monte Carlo para crear las distribuciones esperadas, las cuales no requieren esta última condición. La simulación de Monte Carlo selecciona una muestra aleatoria de entre todas las permutaciones entre todos los puntos temporales disponibles,

Tabla 7. Tabla de contingencia en la que se muestran las frecuencias observadas (o) y las medias de las frecuencias esperadas calculadas por medio de simulaciones de Monte Carlo (e). (Los números en negrita expresan resultados significativos al nivel 0.01)

| | t (Días) | | | | | | | | | | | | | | | |
|-----------------|----------|-----|----------|------------|---|-----|----------|------------|----------|------------|---|-----|---|-----|---|-----|
| | 1 | | 2 | | 3 | | 4 | | 5 | | 6 | | 7 | | 8 | |
| | o | e | o | e | o | e | o | e | o | e | o | e | o | e | o | e |
| 0 < d ≤ 100m | 1 | 0.7 | 5 | 1.5 | 0 | 0.7 | 0 | 0.8 | 0 | 0.9 | 0 | 0.9 | 0 | 0.4 | 0 | 0.1 |
| 100 < d ≤ 200m | 2 | 0.8 | 2 | 1.4 | 2 | 0.7 | 0 | 0.9 | 0 | 0.9 | 0 | 0.9 | 0 | 0.3 | 0 | 0.1 |
| 200 < d ≤ 300m | 1 | 0.6 | 1 | 0.9 | 2 | 0.5 | 0 | 0.7 | 0 | 0.6 | 0 | 0.5 | 0 | 0.2 | 0 | 0.1 |
| 300 < d ≤ 400m | 0 | 0.2 | 3 | 0.7 | 0 | 0.4 | 0 | 0.3 | 0 | 0.5 | 0 | 0.5 | 0 | 0.2 | 0 | 0.1 |
| 400 < d ≤ 500m | 0 | 0.5 | 0 | 1.2 | 1 | 0.5 | 0 | 0.7 | 0 | 0.8 | 2 | 0.8 | 1 | 0.4 | 1 | 0.1 |
| 500 < d ≤ 600m | 1 | 0.3 | 0 | 1 | 0 | 0.5 | 0 | 0.4 | 0 | 0.7 | 2 | 0.7 | 1 | 0.3 | 0 | 0.1 |
| 600 < d ≤ 700m | 0 | 0.3 | 0 | 0.5 | 0 | 0.2 | 1 | 0.3 | 1 | 0.3 | 0 | 0.3 | 0 | 0.1 | 0 | 0.0 |
| 700 < d ≤ 800m | 0 | 1.1 | 0 | 2.3 | 0 | 1 | 4 | 1.2 | 2 | 1.3 | 2 | 1.4 | 1 | 0.6 | 0 | 0.2 |
| 800 < d ≤ 900m | 0 | 0.4 | 0 | 1.2 | 0 | 0.6 | 1 | 0.6 | 4 | 0.8 | 0 | 0.9 | 0 | 0.4 | 0 | 0.1 |
| 900 < d ≤ 1000m | 0 | 0.1 | 0 | 1.2 | 0 | 0.1 | 0 | 0.1 | 0 | 0.2 | 1 | 0.2 | 0 | 0.1 | 0 | 0.0 |

Tabla 8. Tabla de contingencia en la que se muestran los cocientes entre las frecuencias observadas y las esperadas calculadas por medio de simulaciones de Monte Carlo (el sombreado más oscuro indica mayores cocientes; los números en negrita expresan resultados significativos al nivel 0.01).

| | t (Días) | | | | | | | |
|-----------------|----------|------------|-----|-----|------------|-----|-----|------------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 0 < d ≤ 100m | 1.5 | 3.4 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 100 < d ≤ 200m | 2.6 | 1.4 | 3.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 200 < d ≤ 300m | 1.7 | 1.1 | 4.3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 300 < d ≤ 400m | 0.0 | 4.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 400 < d ≤ 500m | 0.0 | 0.0 | 1.8 | 0.0 | 0.0 | 2.5 | 2.6 | 9.4 |
| 500 < d ≤ 600m | 3.2 | 0.0 | 0.0 | 0.0 | 0.0 | 3.0 | 3.0 | 0.0 |
| 600 < d ≤ 700m | 0.0 | 0.0 | 0.0 | 3.0 | 3.4 | 0.0 | 0.0 | 0.0 |
| 700 < d ≤ 800m | 0.0 | 0.0 | 0.0 | 3.3 | 1.5 | 1.5 | 1.8 | 0.0 |
| 800 < d ≤ 900m | 0.0 | 0.0 | 0.0 | 1.7 | 5.1 | 0.0 | 0.0 | 0.0 |
| 900 < d ≤ 1000m | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 6.0 | 0.0 | 0.0 |

CONCLUSIÓN

En este artículo se ha demostrado cómo la simulación de Monte Carlo es una técnica más apropiada para el estudio de las pautas espacio-temporales del delito. No sólo soluciona el problema de la condición de observaciones independientes, sino que además nos proporciona información sobre el significado estadístico de cada categoría de distancias espacio-temporales, facilitando así la interpretación de los resultados y el trazado de mapas predictivos más precisos y fiables.