



UNIVERSIDAD DE SANTIAGO DE CHILE

FACULTAD DE CIENCIAS

DEPARTAMENTO DE MATEMÁTICAS Y CIENCIA DE LA
COMPUTACIÓN

PROYECTO TESIS

Análisis Espacial para la determinación de factores que influyen en el reporte de delitos

Katherine Carrasco Vargas

Profesor Guía:

Felipe Elorrieta López

Comisión:

Lucía Dammert

Luis Riquelme

Resumen

Hoy en día, uno de los temas más preocupante, principalmente en la Región Metropolitana, es la delincuencia. Cada vez es más común ser víctimas o conocer a personas que son víctimas de diferentes delitos, algunos de mayor magnitud que otros. ¿Por qué ocurren estos actos con tanta frecuencia? ¿Las víctimas denuncian cuando esto les ocurre? Muchas veces es normal escuchar que se ha sufrido algún acto delictivo y sin embargo, se queda en un susto y no se tomas acciones al respecto. De aquí, la importancia de indagar en posibles factores que permitan relacionar la ocurrencia y denuncias de diferentes hechos delictuosos, enfocándose principalmente en un análisis a nivel comunal para intentar encontrar algún patrón común o una relación al respecto de acuerdo a la localidad de las personas, mediante herramientas de la estadística espacial.

Los modelos utilizados, dado las características de los datos y la selección de su estudio, fueron modelos de datos de área, específicamente Modelo autoregresivo simultáneo (SAR) y modelo autoregresivo condicional (CAR) cuando se presentara una dependencia global, de lo contrario, se generaron modelos de regresión lineales.

De acuerdo a los resultados , a nivel general, independiente del mejor modelo encontrado para cada variable respuesta estudiada, se logra comprender que existe una gran relación a nivel socio-económico respecto a las ocurrencias y disposición de denunciar delitos de diferente carácter, pues como algunos estudios indican, personas de mejor situación tienden a tomar más acciones frente a esta problemática.

Palabras Claves: *Estadística Espacial, Delincuencia, Modelos Autoregrsivos, Test de Moran, Test de Geary.*

Índice general

índice de tablas	4
índice de figuras	4
1. Introducción	7
1.1. Planteamiento del Problema	9
1.2. Objetivos	11
1.2.1. Objetivo General	11
1.2.1.1. Objetivos Específicos	11
1.3. Metodología	12
2. Marco Teórico	13
2.1. Estadística Espacial	13
2.2. Datos de área	14
2.2.1. Dependencia Espacial	14
2.2.2. Modelos Espaciales Autoregresivos	19
2.3. Regresión Lineal	25
2.3.1. Análisis de Varianza (ANOVA)	26
2.3.1.1. Test Global	26
2.3.1.2. Test Individual	27
2.4. Selección de Variables	27
2.4.1. Eliminación hacia atrás (backward elimination)	27
2.4.2. Eliminación hacia adelante (forward selection)	28
2.4.3. Regresión Stepwise	28
2.5. Validación de los Modelos	28
2.5.1. Bondad de Ajuste	28
2.5.1.1. Criterio de Información de Akaike (AIC):	28
2.5.2. Validación de supuestos	29
2.5.2.1. Normalidad de los residuos	29
2.5.2.2. Homocedasticidad	30

2.5.2.3. Errores No Correlacionados Espacialmente	31
3. Análisis de los datos	32
3.1. Encuesta Nacional Urbana de Victimización (ENUSC)	32
3.2. Análisis de Preguntas	33
3.2.1. Módulo de Víctimización	33
3.2.2. Modulo de Inseguridad y reacción frente al delito	35
3.2.3. Caracterización socio-demográfica	36
3.3. Construcción de Variables	37
3.3.1. Variable Respuesta	37
3.3.2. Variables Predictoras	37
3.4. Análisis de datos	38
3.4.1. Análisis por delito	38
3.4.2. Análisis de delitos por comuna	40
3.4.2.1. Hurto	40
3.4.2.2. Robo con violencia e intimidación	46
3.4.2.3. Robo por sorpresa	52
3.4.2.4. Robo en vivienda	58
3.4.3. Análisis por clasificación de delitos	63
3.4.3.1. Análisis de Delitos Violentos	65
3.4.3.1.1. Tasa de Denuncias	65
3.4.3.1.2. Tasa de Ocurrencias	67
3.4.3.2. Análisis Delitos no Violentos	69
3.4.3.2.1. Tasa de denuncias	69
3.4.3.2.2. Tasas de ocurrencias	71
4. Conclusión	74
5. Referencias	76
6. Anexos	78
6.1. Matriz de Comunas	78
6.2. Funciones Espaciales	78
6.3. Modelos opcionales para buscar el mejor ajuste:	80

Índice de cuadros

2.1. Valores para la distribución Normal	17
2.2. Comparación modelo SAR y CAR	25
2.3. Comparación varianza SAR y CAR	25
2.4. Tabla ANOVA para regresión lineal múltiple	26
3.1. Tabla Resumen por delitos	39
3.2. Test de dependencia para Tasas de ocurrencias de Hurto	42
3.3. Modelos Lineales para tasa de ocurrencias del hurto	42
3.4. Test de dependencia espacial para tasas de denuncias de Hurto	44
3.5. Comparación SAR Y CAR para tasa de denuncias por hurto	45
3.6. Test de dependencia para Tasas de ocurrencias de Hurto	48
3.7. Comparación SAR Y CAR para tasa de denuncias por hurto	49
3.8. Modelo Final Tasa de ocurrencias Hurto	49
3.9. Test de dependencia espacial para tasas de denuncias de Hurto	51
3.10. Test de dependencia para Tasas de ocurrencias de Robo con Violencia	53
3.11. Comparación SAR Y CAR para tasa de denuncias por robo por sorpresa	55
3.12. Test de dependencia espacial para tasas de denuncias de Robo con Violencia	56
3.13. Test de dependencia para Tasas de ocurrencias de robo en vivienda	59
3.14. Comparación SAR Y CAR para tasa de denuncias por robo en vivienda	60
3.15. Test de dependencia espacial para tasas de denuncias de robo en vivienda	61
3.16. Tabla resumen delitos Violentos y No violentos	63
3.17. Test de dependencia para Tasas de denuncias por delitos violentos	65
3.18. Test de dependencia para Tasas de ocurrencias por delitos violentos	67
3.19. Test de dependencia para Tasas de denuncias por delitos no violentos	69
3.20. Test de dependencia para Tasas de denuncias por delitos no violentos	72

Índice de figuras

1.1. Esquema de la Metodología	12
2.1. Gráfico Q-Q Plot	30
2.2. Histograma	30
3.1. Gráfico de Ocurrencias v/s Denuncias por delito	40
3.2. Gráfico Hurto : casos v/s denuncias	41
3.3. Mapa tasa ocurrencia hurto: Distribución Espacial	41
3.4. Test Local Moral para tasa de ocurrencia de delitos	42
3.5. Gráfico de Dependencia espacial a k distancias para la tasa de ocurrencia por hurto	43
3.6. Mapa Tasa denuncia Hurto: Distribución espacial	44
3.7. Test Local de Moran para tasa de denuncias de hurto	44
3.8. Gráfico de Dependencia espacial a k distancias para la tasa de denuncias por hurto	45
3.9. Gráfico de dependencia a k distancias para tasa de denuncias de hurto (Modelo)	46
3.10. Gráfico Robo con violencia e intimidación: casos v/s denuncias	47
3.11. Mapa de distribución espacial de la tasa de ocurrencias de robo con violencia	47
3.12. Test Local Moral para tasa de ocurrencia de robo con violencia	48
3.13. Gráfico de dependencia a k distancias	49
3.14. Gráfico de Dependencia espacial a k distancias para tasa de ocurrencia por robo con violencia e intimidación (Modelo)	50
3.15. Mapa tasa de denuncias robo violencia e intimidación: Distribución Espacial	50
3.16. Test Local de Moran para tasa de denuncias de robo con violencia	51
3.17. Gráfico de dependencia a k distancias para tasa de denuncias de robo con violencia e intimidación (Modelo)	52
3.18. Gráfico Robo por sorpresa : casos v/s denuncias	53
3.19. Mapa de distribución espacial de la tasa de ocurrencias de robo con sorpresa	53
3.20. Test Local Moral para tasa de ocurrencia de robo con violencia	54
3.21. Gráfico de dependencia a k distancias para tasas de robo sorpresa	54

3.22. Gráfico de Dependencia espacial a k distancias para tasa de ocurrencia por robo por sorpresa (Modelo)	55
3.23. Mapa tasa denuncia robo sorpresa: Distribución Espacial	56
3.24. Test Local de Moran para tasa de denuncias de robo con violencia	56
3.25. Gráfico de dependencia a k distancias para tasa de denuncias de robo por sorpresa (Modelo)	57
3.26. Gráfico Robo en vivienda: casos v/s denuncias	58
3.27. Mapa de distribución espacial de la tasa de ocurrencias de robo en vivienda	59
3.28. Test Local Moral para tasa de ocurrencia de robo con violencia	59
3.29. Gráfico de Dependencia espacial a k distancias para tasa de ocurrencia por robo en vivienda (Modelo)	60
3.30. Mapa tasa denuncia robo vivienda: Distribución Espacial	61
3.31. Test Local de Moran para tasa de denuncias de robo en vivienda	61
3.32. Gráfico de dependencia a k distancias para tasa de denuncias de robo en vivienda (Modelo)	62
3.33. Gráfico delitos Violentos	63
3.34. Gráfico delitos no violentos	64
3.35. Mapa distribución espacial delitos violentos	65
3.36. Test local Moran para tasa de denuncias de delitos violentos	66
3.37. Gráfico de dependencias a k distancias para tasas de denuncias por delitos violentos (Modelo)	66
3.38. Mapa tasa ocurrencia delitos violentos: Distribución Espacial	67
3.39. Test local Moran para tasa de ocurrencias de delitos violentos	68
3.40. Gráfico de dependencias a k distancias para tasas de ocurrencias por delitos violentos (Modelo)	68
3.41. Mapa distribución espacial delitos no violentos	69
3.42. Test local Moran para tasa de denuncias de delitos no violentos	70
3.43. Gráfico de dependencias a k distancias para tasas de denuncias por delitos no violentos (Modelo)	71
3.44. Mapa distribución espacial delitos no violentos	71
3.45. Test local Moran para tasa de denuncias de delitos no violentos	72
3.46. Gráfico de dependencias a k distancias para tasas de ocurrencias por delitos no violentos (Modelo)	73

Capítulo 1

Introducción

Actualmente, la delincuencia es una realidad inevitable que se enfrenta fuertemente en nuestro país, y de acuerdo a la gran frecuencia que ha alcanzado en los últimos años, se ha transformado en una de las temáticas que genera mayor preocupación para la ciudadanía.

En materia de victimización, bajo los resultados y alcances de la XIV Encuesta Nacional Urbana de Seguridad Ciudadana (ENUSC) 2017, es posible constatar un aumento del 5,2% en los últimos 5 años, si en el año 2013 la victimización era del 22,8% , en el año 2017 pasó a ser del 28% , teniendo alzas sostenidas durante todos estos años [1].

En la actualidad, para las personas no es novedad enterarse o presenciar distintos delitos de diferente magnitud, desde hurtos de un móvil en plena vía pública hasta asaltos terminados en crímenes de mayor nivel, lo que ha generado inseguridad y miedo en los habitantes. Lo antes descrito, repercute en la tranquilidad de las personas y en las distintas instituciones a cargo de la seguridad de los ciudadanos.

Bajo el artículo 1 del Código Penal, es delito *“toda acción u omisión voluntaria penada por la ley”*¹ , de aquí, lo que hace característico al delito es la existencia de una norma jurídica dictada con anterioridad que presente una pena o sanción estipulada por la ley para quien realiza aquel acto delictivo [2].

Para hacer frente a esta situación, resulta importante dimensionar mediante diferentes herramientas la realidad de este fenómeno, lo que actualmente se realiza a través de información recolectada de acuerdo a las denuncias emitidas a Carabineros de Chile, Policía de Investigaciones o Ministerio Público [3].

Con el objetivo de analizar de manera más profunda la no denuncia de actos delictivos y conocer los distintos factores que intervienen en esta decisión, se dispone de la información anual entregada por la ya mencionada ENUSC, la cual en su último estudio del año 2017, describe que la tasa de reporte de denuncias para ese año corresponde a un 38,5%, sobresaliendo con un 18% la percepción

¹La ley es una declaración de la voluntad soberana que, manifestada en la forma prescrita por la Constitución, manda, prohíbe o permite.(Artículo 1º del Código Civil)

de que la pérdida no fue lo suficientemente seria, seguida por alrededor del 15% que se basa en la justificación de que las entidades correspondientes no habrían hecho nada [4] y de los datos semestrales del índice de Paz Ciudadana-Adimark, donde en el Segundo semestre del 2016 se obtuvo que el porcentaje de personas que denuncia los delitos del que son víctimas es de un 52,6% donde predominan los mismos factores de desmotivación a reportar lo ocurrido. [5].

El reporte de delitos, juega un rol importante al momento de conocer la verdadera cantidad de delitos que ocurren realmente.

No es menor tener en cuenta que existe un número importante de factores que afecta a la decisión de denunciar o no, luego de vivir un hecho delictivo, entre los que se pueden encontrar la edad, gravedad del hecho, victimización previa, situación socioeconómica, relación entre víctima y victimario, miedo a represarías, entre otros [6].

Por lo tanto, el siguiente estudio estadístico, permitirá analizar mediante el estudio del comportamiento de delitos, tener una panorámica real de estos, permitiendo con lo anterior prevenir y/o tomar medidas de seguridad ante determinadas situaciones delictuales en diferentes localidades.

1.1. Planteamiento del Problema

Como se mencionó, la presencia de la actividad delictiva, es algo que se lleva a cabo todos los días en diferentes lugares y dependencias de la Región Metropolitana, sin embargo, pese a los diferentes datos que se disponen, continúa existiendo una encrucijada que describe la baja posibilidad de dominar las causas de la delincuencia, imposibilitando la opción de prever de manera exacta dónde y cuándo se cometerá un delito. Es por esto, que la motivación o desmotivación de emitir una denuncia afecta conjuntamente a la hora de actuar y detener todos los hechos delictivos que efectivamente están sucediendo, dado que hay una alta tasa de delitos desconocidos por no ser reportados, dejando escapar victimarios que podrían cometer estos mismos delitos con distintas víctimas u otros delitos de mayor magnitud.

Sería, por tanto, muy importante encontrar aquellos factores que afectan en que los ciudadanos denuncien los delitos a las autoridades, tanto los que sufran en carne propia o perciban a su alrededor, porque es la única forma de combatir la actividad de los criminales y de conseguir que las autoridades trabajen sobre cifras fiables, con el fin de adoptar las medidas pertinentes en la lucha contra la delincuencia [6].

Si bien la no denuncia es un problema que ha dejado sin reportar diferentes tipos de delitos, existe algunos que pasan de ser reportados con mayor frecuencia que otros, y las razones cambian según sea el acto delictivo. Entre el delito menos denunciado se encuentra el hurto y robo por sorpresa, lo que se ve afectado por la confianza que tiene la víctima sobre las instituciones que intervienen en el control social del delito y su persecución. Así mismo, existen delitos donde la predisposición de reportar se ve afectada por miedo a represarías, comentarios que se podrían generar, o cercanía con el victimario como lo es con los fraudes, violaciones o violencia intrafamiliar[3].

Por otro lado, bajo una investigación realizada por S. Avdija et al (2012) [7], es posible indicar que existen antecedentes los cuales indican que entre los factores destacados que influyen en el comportamiento de las personas en denunciar delitos se encuentra el Nivel Socioeconómico. Aquí, se menciona que bajo diferentes resultados se indican cosas tales como que personas que viven en barrios desfavorecidos económicamente están menos dispuestas a cooperar con la policía, personas que se encuentran en situaciones financieras difíciles (por ejemplo, desempleados) son menos propensas a reportar crímenes a la policía, actitudes hacia la policía son más negativas entre las personas con ingresos inferiores o personas que no son propietarias de viviendas, entre otros resultados. Así, el nivel socioeconómico muestra ser un importante causa que explica la cantidad de denuncias que se realizan.

Por lo anterior, sería de interés aplicar técnicas que evalúen no tan solo los delitos que ocurren sino también dónde están sucediendo qué hechos delictivos exactamente, y es aquí donde aparece el concepto de **Estadística Espacial**.

Cuando se habla de Estadística Espacial, se hace referencia a la reunión de un conjunto de metodologías apropiadas para el análisis de datos que corresponden a la medición de variables en diversos sitios o puntos de espacio de una región determinada.

Un caso particular de la estadística espacial son los datos de áreas, lo que matemáticamente se caracteriza mediante un proceso estocástico de la forma $\{Z(w, s) : w \in \omega, s \in D \subset \mathbb{R}^p\}$, donde s p -dimensional representa la ubicación en el espacio euclidiano y $Z(s)$ es una variable aleatoria que representa la distribución de todas las posibles realizaciones $z(s)$ en la localización s [8]. Así, suponiendo una posible existencia de dependencia entre comunas cercanas, el uso de esta metodología podría ayudar a encontrar factores similares que estarían influyendo en el reporte de delitos de los habitantes.

Para poner en práctica un análisis de esta índole, existen modelos de datos de áreas adecuados para la realización de este estudio, como lo son el Modelo Autoregresivo Simultáneo SAR y el Modelo Autoregresivo Condicional CAR [9], los cuales se aplicarán a los resultados de la última encuesta realizada por la ENUSC, puesto que estos datos generan información de gran interés para apoyar en la solución y explicación del problema en cuestión, permitiendo obtener variables que afectan en el comportamiento del reporte de delitos, las razones de la no denuncia y por consecuencia, factores que afectan en los delitos en sí de cada localidad.

1.2. Objetivos

1.2.1. Objetivo General

Identificar factores que afectan en el comportamiento de reporte de delitos a nivel comunal en la Región Metropolitana utilizando herramientas de estadística espacial.

1.2.1.1. Objetivos Específicos

- Efectuar un estudio de diferentes bibliografías disponibles, referentes a estadística espacial y a modelos de datos de área SAR y CAR.
- Realizar un análisis de la base de datos (ENUSC 2017) para conocer profundamente la información disponible para el estudio y encontrar diferentes variables de interés.
- Realizar un análisis descriptivo de las observaciones mediante el Software estadístico R.
- Ajustar los modelos a los datos seleccionados mediante el programa antes mencionado.
- Analizar el ajuste y calidad de los modelos.
- Estudiar los resultados de la aplicación de los modelos e identificar si mediante el uso de la estadística espacial se ha logrado encontrar posibles factores propios socioeconómicos en localidades cercanas, que influyen en la voluntad de reportar delitos.

1.3. Metodología

Para comenzar, será esencial analizar documentos que abarquen y expliquen la correspondiente estadística espacial [10], que será la herramienta que nos permitirá dar solución al problema en cuestión, para luego continuar con un estudio de datos de área, específicamente de los modelos SAR y CAR ya mencionados [11]. Para conocer su aplicación e implementación, será conveniente indagar en estudios que hayan dado uso a estos modelos para una mayor comprensión de su funcionamiento.

Luego, será necesario estudiar y familiarizarse con la base de datos correspondiente a los resultados de la encuesta ENUSC 2017, estudiando detalladamente cada ítem y respuesta que conforman los diferentes módulos presentes, para poder así calcular variables de interés que puedan ser útiles para el estudio de acuerdo a la literatura sobre el tema.

A continuación, ya conociendo la fundamentación y aplicación de los modelos a utilizar, se realizará un análisis de la dependencia espacial y posteriormente se ajustarán los modelos a los datos para distintas comunas de la Región Metropolitana, separando el análisis por tipo de delito. Esta implementación, se realizará mediante el software estadístico R, bajo el paquete “spdep” [12]. Si no se presenta autocorrelación espacial, el modelo adecuado será una regresión Lineal.

Una vez que ya se tengan los modelos, se evaluará el ajuste de los datos y la calidad de los modelos, mediante el análisis de los supuestos y una bondad de ajuste².

Finalmente, se analizarán los resultados obtenidos y será posible tomar diferentes conclusiones respecto al comportamiento de reporte de delitos, para poder dar una posible solución a la problemática mencionada.

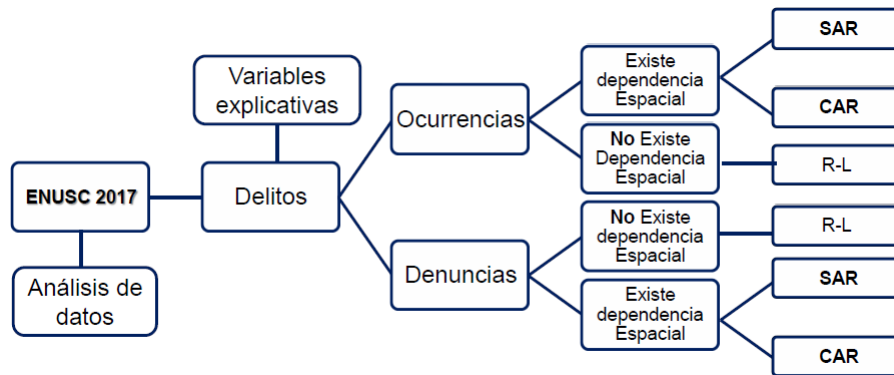


Figura 1.1: Esquema de la Metodología

²La bondad de ajuste de un modelo estadístico describe qué tan bien se ajusta a este un conjunto de observaciones

Capítulo 2

Marco Teórico

2.1. Estadística Espacial

La estadística espacial, es una rama de la estadística que estudia diferentes fenómenos aleatorios teniendo en cuenta la ubicación o distancia entre objetos, entregando un conjunto de metodologías que permiten realizar un análisis de datos que incluye diversas técnicas que hacen posible visualizar y modelar relaciones espaciales.

De manera más formal, se puede decir que la estadística espacial trata con el análisis de realizaciones de un proceso estocástico (o también denominado proceso aleatorio o campo aleatorio) definido de la siguiente forma [10]:

$$\{Z(s) : s \in D \subset \mathbb{R}^p\}$$

Donde:

- s : ubicación en el espacio Euclidiano P -dimensional.
- $Z(s)$ variable aleatoria en la ubicación s y s varía sobre un conjunto de índices o espacio de parámetros $D \subset \mathbb{R}^p$.

Dentro de la estadística espacial, se encuentran los siguientes tres grandes áreas:

- **Geoestadística:** Estudia datos de procesos estocásticos en los que el espacio de parámetros es continuo, es decir, existen infinitos lugares en donde medir, sin embargo, el investigador puede hacer selección de puntos del espacio a conveniencia o puede seleccionar los sitios bajo algún esquema de muestreo probabilístico. Por ejemplo:

1. Contenido de nitrógeno en sitios de un parcela experimental.
2. Precipitación en sitios de Colombia.

- **Datos de área:** En este caso el proceso estocástico tiene espacio de parámetros discreto y la selección de los sitios de medición depende del investigador. Por ejemplo:
 1. Tasa de mortalidad
 2. Producción cafetera en kilogramos de todas las fincas productoras de café del país.
- **Patrones puntuales:** A diferencia de los dos anteriores, el espacio de parámetros es aleatorio, es decir que la decisión respecto a dónde se hace la medición no depende del investigador. El conjunto puede ser discreto o continuo, pero la ubicación de los sitios donde ocurre el hecho a estudiar es dada por la naturaleza. Por ejemplo:
 1. Ubicaciónn de nidos de pájaros en una regiónn dada.
 2. Sitios de terremotos en Colombia.

De acuerdo a las definiciones anteriores, la gran diferencia entre cada área radica en el conjunto de índices del proceso estocástico de interés, y dado el problema en cuestión, los datos y la forma en que se trabajaran (análisis comunal) , se dará uso a **datos de áreas** (también denominados datos regionales o lattices).

2.2. Datos de área

Como se mencionó anteriormente, de manera más formal, se tienen las dos importantes siguientes características [13]:

1. El proceso estocástico tiene espacio de parámetros discreto $D \subset \mathbb{R}^p$, es decir, $Z(s)$ puede ser observado en una serie de ubicaciones fijas que pueden enumerarse, las cuales pueden ser regular o irregularmente espaciadas.
2. Los sitios de medición dependen del investigar, esto es, que D es fijo y las ubicaciones en este no son estocásticas.

Luego, para realizar el análisis comunal correspondiente y abordar el problma en cuestión bajo herramientas de la estadística espacial, específicamente datos de área, será primordial realizar un análisis de **dependencia espacial** de los datos.

2.2.1. Dependencia Espacial

Como es conocido, las estadísticas clásicas se basan en la independencia de los valores observados, los cuales se consideran realizaciones independientes de la misma variable aleatoria. Sin embargo, cuando los valores observados están basados en el espacio, la hipótesis de independencia ya no es válida. En el contexto espacial, esta falta de independencia recibe el nombre de dependencia o autocorrelación espacial.

La autocorrelación espacial, tiene que ver tanto con la localización geográfica como con los valores hallados de la variable que se esté estudiando, pues una variable posee esta característica espacial cuando el valor observado en un lugar determinado depende de los valores observados en lugares próximos. Luego, la existencia de una autocorrelación espacial, prueba lo que se establece en la Primera Ley de Geografía: "Todo está relacionado con todo lo demás, pero las cosas cercanas están más relacionadas que las cosas lejanas" (Tobler, 1970)[13].

Para verificar la existencia de dependencia, se utilizan diferentes estadísticos de correlación espacial, los cuales pueden ser medidas **globales o locales**.

Las medidas de dependencia **global**, utilizan la información completa del conjunto de datos con el fin de obtener un valor promedio para todo el espacio geográfico, pero al resumir en un único valor toda la información, no permiten detectar la variabilidad de la dependencia ni la localización de estos patrones.

Por otra parte, Anselin (1995) propone un conjunto de indicadores **locales** de asociación espacial denotados LISA ("Local Indicators of Spatial Association"), capaces de detectar la contribución de cada región a un indicador de dependencia espacial global. Estos, realizan el análisis en un subconjunto de datos y para cada localización brindan un valor particular que puede ser contrastado estadísticamente clasificándolo en diferentes grupos o cluster [14].

Para la detección de este fenómeno, será importante determinar un criterio que estipule cuándo dos localidades serán consideradas cercanas o vecinas, dando así uso a la denominada **Matriz de proximidad o Matriz de vecindades**, cuyos valores se obtienen bajo la siguiente definición:

$$W_{ij} = \begin{cases} 1 & \text{si las localizaciones } i, j \text{ son vecinas} \\ 0 & \text{si no lo son} \end{cases}$$

Es importante destacar, que en esta ocasión, se considerará que dos comunas son vecinas si estas tienen frontera en común, lo que a la vez implica que tienen un **lag o rezago** igual a 1.

Entre los test existentes, es posible encontrar los siguientes:

■ Test Globales

● Test de Moran (I)

Este test, entrega un indicador que refleja una forma espacialmente ponderada del coeficiente de correlación de Pearson.

El índice I de Moran (Moran, 1950), considerando $Z(s_1), \dots, Z(s_n)$ como las variables medidas en n áreas, sigue la forma básica para los índices globales de autocorrelación espacial con similitud entre las regiones i y j definidas como el producto de la diferencia respectiva entre Y_i e Y_j con la media global [15]:

$$sim_{ij} = (Y_i - \bar{Y})(Y_j - \bar{Y})$$

Donde:

$$\bar{Y} = \frac{\sum_{i=1}^N Y_i}{N}$$

Luego, se divide de forma básica por la varianza observada, obteniendo finalmente el indicador I de la siguiente manera:

$$I = \frac{1}{S^2} \frac{\sum_{i=1}^N \sum_{j=1}^N \omega_{ij} sim_{ij}}{\sum_{i=1}^N \sum_{j=1}^N \omega_{ij}}$$

con:

- $S^2 = \frac{1}{N} \sum_{i=1}^N (Y_i - \bar{Y})^2$
- ω_{ij} : Elemento de la matriz de vecindades espaciales del par i, j .

Se debe obtener el valor de I observado insertando las observaciones en la ecuación anteriormente definida. Así, cuando las regiones vecinas tienden a tener valores similares, su valor será **positivo**. Si las regiones vecinas tienden a tener valores diferentes, el valor será **negativo**. Finalmente, cuando no hay correlación entre los valores vecinos, el valor esperado de I se aproxima a **cero** a medida que aumenta N. Lo anterior, en términos del indicador, se expresa de la forma:

- $I = 0$: Ausencia de correlación espacial.
- $I > 1$: Correlación espacial positiva.
- $I < 1$: Correlación espacial negativa.

Bajo **normalidad**, se contrasta la siguiente hipótesis:

$$H_o : La AE es nula \quad v/s \quad H_1 : La AE es no nula$$

Donde **AE**: Autocorrelación Espacial.

En términos del indicador, la hipótesis anterior queda definida de la siguiente manera:

$$H_o : I = 0 \quad v/s \quad H_1 : I \neq 0$$

Luego, el valor esperado y varianza bajo la hipótesis nula, están dados por:

- $E(I) = -\frac{1}{N-1}$
- $Var(I) = \frac{N^2 S_1 - N S_2 + 3 S_0^2}{(N-1)(N+1) S_0^2}$

Donde:

- $S_0 = \sum_{i=1}^N \sum_{j=1}^N \omega_{ij}$
- $S_1 = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (\omega_{ij} + \omega_{ji})^2$

- $S_2 = \sum_{i=1}^N (\omega_{i+} + \omega_{+j})^2$
- $\omega_{i+} = \sum_{j=1}^N \omega_{ij} \quad ; \quad \omega_{+i} = \sum_{j=1}^N \omega_{ji}$

Su distribución asintótica normal es:

$$z = \frac{[I - E(I)]}{\sqrt{Var(I)}}$$

Cada valor z, se encuentra asociado a un pvalue para poder así fijar los niveles de significancia, donde los criterios más utilizados son los de la siguiente tabla:

Puntuación z (Desviación Estándar)	Valor P (Probabilidad)	Nivel de Confianza (Porcentaje)
< -1,65 o > +1,65	< 0,10	90 %
< -1,96 o > +1,96	< 0,05	95 %
< -2,58 o > +2,58	< 0,01	99 %

Cuadro 2.1: Valores para la distribución Normal

En esta ocasión se utilizará $z = 1,65$ para $\alpha = 0,10$, lo que indicará que:

- Si $p_value \leq 0,10 \rightarrow$ Se rechaza H_0
- Si $p_value > 0,10 \rightarrow$ No se rechaza H_0

Así mismo, otra manera de realizar el cálculo de la desviación típica del estadístico I (denotada $SD(I)$ o $Var(I)$) es bajo el método o proceso de **randomization o aleatorización**. Aquí, los datos de las unidades espaciales se intercambian al azar, es decir, una aleatorización por permutación que permite obtener diferentes valores de autocorrelación que se comparan con el valor obtenido.

Así, la hipótesis se define de la siguiente forma:

$$H_0 : \text{El proceso es aleatorio} \quad \text{v/s} \quad H_1 : \text{El proceso no es aleatorio}$$

Luego, al igual que en el caso de normalización, se especifica el nivel de significancia que se utilizará y que va a indicar la probabilidad de rechazar o no la hipótesis nula planteada.

Observación:

A diferencia de un coeficiente de correlación tradicional, los valores para el índice de Moran no necesitan estar restringido al intervalo $[-1, 1]$, aunque normalmente $|I| < 1$.

• **Test de Geary (c)**

En este indicador c (Geary, 1954), las similitudes se miden mediante la siguiente expresión [15]:

$$sim_{ij} = (Y_i - Y_j)^2$$

De acuerdo a esto, se construye un índice global de autocorrelación espacial basado en un valor ponderado de media de los valores de similitud observados para todos los pares de regiones que asignan ponderaciones por proximidad espacial, esto es, asignando pesos más altos para pares de observaciones más cercanos.

Así, se da resultado a denominada c de Geary, la cual es definida de la forma:

$$c = \frac{N-1}{2 \sum_{i=1}^N (Y_i - \bar{Y})^2} \frac{\sum_{i=1}^N \sum_{j=1}^N \omega_{ij} sim_{ij}}{\sum_{i=1}^N \sum_{j=1}^N \omega_{ij}}$$

El valor de índice c de Geary oscila entre los valores 0 y un valor positivo que solo en situaciones concretas se excede de 2. Así, de forma contraria al estadístico I de Moran, se establece lo siguiente:

- $c = 0$: Perfecta correlación espacial positiva.
- $c > 1$ (c tiende a 2): Perfecta autocorrelación espacial negativa.
- $c = 1$: Ausencia de correlación espacial.

Al igual que en el test global de Moran, la hipótesis es:

$$H_0 : La AE es nula \quad v/s \quad H_1 : La AE es no nula$$

Lo que en esta ocasión, bajo los valores que toma el indicador c , puede ser definida como:

$$H_0 : c = 0 \quad v/s \quad H_1 : c \neq 1$$

Observación:

Este indicador, corresponde a la estadística de Durbin-Watson, utilizada para probar la autocorrelación regresiones y series temporales.

Al analizar los posibles resultados de ambos indicadores globales, es posible ver que para la interpretación de los valores c de **Geary** es opuesta a la de los valores de I de **Moran**, pues altos valores de c equivalen a bajos valores de I y viceversa. Griffith (1987), señala que este tipo de relación inversa entre los índices es básicamente de naturaleza lineal. Las desviaciones de la linealidad se atribuyen a las diferencias en lo que mide cada uno de ellos, es decir, para c de Geary se ocupa de las comparaciones pareadas y por otro lado I de Moran de las covariaciones.

Es importante destacar que Cliff y Ord (1981) señalan que I parece estar menos afectado por la distribución de los datos que c y adjudican una ligera ventaja estadística de I respecto a c . Sin embargo, el índice I de Moran parece ser más sensible a los valores extremos (Legendre y Fortin, 1989).

- **Test Locales**

- **Test Local de Moran (I_i)**

Otro indicador que se emplea y es fundamental para cuando el test de Moran detecte la presencia de autocorrelación, es el test de Moran **local**. Así, se define un indicador de dependencia local basado test I de Moran, I_i , del modo siguiente:

$$I_i = z_i \sum_{j=1}^{J_i} \omega_{ij} z_j$$

Donde:

- z_i y z_j : Variables Y_i estandarizadas.
- $\sum_{j=1}^{J_i}$: Sumatoria únicamente de los valores vecino de i , $j \in J_i$

Con este índice, será posible evaluar la influencia de ubicaciones individuales en la magnitud del estadístico de autocorrelación espacial global correspondiente.

Es posible comprobar que la suma de los I_i para todas las observaciones i es proporcional al estadístico I de Moran dado por la ecuación:

$$\sum_{i=1}^N I_i = \sum_{i=1}^N z_i \sum_{j \in J_i} \omega_{ij} z_j$$

Al igual que con el estadístico I global, se puede calcular la desviación de las estadísticas locales de los valores esperados bajo suposiciones de normalidad y aleatorización.

2.2.2. Modelos Espaciales Autoregresivos

Los modelos autorregresivos espaciales son útiles para analizar, caracterizar e interpretar diferentes fenómenos espaciales de valor real $X = Z(s)$, $s \in S$ definidos en redes espaciales discretas S que tienen un vecindario geométrico[16]. Paralelamente a los modelos autoregresivos (AR) de series temporales, los cuales indican la existencia de una relación lineal entre lo ocurrido con la variable en un periodo de tiempo t y lo ocurrido en periodos anteriores, cuando se refiere al plano espacial, dicho modelo relaciona un suceso $Z(s_i)$ con lo ocurrido en lugares vecinos denotados como $Z(s_{-i})$.

Dentro de los enfoques existentes para modelar la autocorrelación en los modelos de regresión espaciales, destacan el modelo autorregresivo simultáneo (SAR) y el modelo autorregresivo condicional (CAR). Estos modelos permiten incorporar la estructura de vecindad utilizada cuando se trabaja con este tipo de datos.

La principal diferencia que presentan estos dos modelos es que los **CAR** se utilizan cuando existe dependencia entre datos cercanos, mientras que los modelos **SAR** toman en cuenta la dependencia espacial existente con mayor alcance.

1. Modelo SAR:

En la especificación para este modelo, los términos de error e están modelados de manera que dependen unos de otros de la siguiente manera[17]:

$$e_i = \sum_{i=1}^m b_{ij} e_i + \epsilon_i \quad (2.1)$$

Donde:

- b_{ij} : dependencia espacial entre áreas, b_{ii} debe ajustarse a cero para que cada área no recaiga sobre sí misma.
- ϵ_i : error aleatorio, distribuye de forma independiente según una distribución Normal con una matriz de media cero y matriz de varianza - covarianza diagonal Σ_ϵ con los elementos $\sigma_{\epsilon_i}^2$, $i = 1, \dots, n$.

Luego, considerando lo anterior y partiendo de la idea de un modelo lineal con datos normales, este modelo de regresión se define como:

$$Z(s) = X(s)\beta + e_i \quad (2.2)$$

Con:

- $Z(s)$: Variable respuesta.
- $X(s)$: vector de variables explicativas
- β : matriz simétrica de orden n que contiene los parámetros de dependencia b_{ij} a estimar. Si todos los b_{ij} son igual a cero, no hay autoregresión y el modelo se reduce al modelo tradicional de regresión lineal con errores independientes.

De las dos ecuaciones anteriores, es posible expresar el modelo de la forma: [18]

$$Z(s) = X(s)\beta + e_i \quad (2.3)$$

Lo que indica también que:

$$e_i = Z(s) - X(s)\beta \quad (2.4)$$

Luego, bajo la expresión (2.1):

$$Z(s) = X(s)\beta + Be_i + \epsilon_i \quad (2.5)$$

Sustituyendo ahora lo de (2.4), se tiene:

$$Z(s) = X(s)\beta + B(Z(s) - X(s)\beta) + \epsilon_i \quad (2.6)$$

$$Z(s) = X(s)\beta + BZ(s) - BX(s)\beta + \epsilon_i \quad (2.7)$$

$$Z(s) - BZ(s) = X(s)\beta - BX(s)\beta + \epsilon_i \quad (2.8)$$

$$Z(s)(I - B) = X(s)\beta(I - B) + \epsilon_i \quad (2.9)$$

Así, los términos del error pueden ser expresados de la siguiente manera:

$$\epsilon_i = Z(s)(I - B) - X(s)\beta(I - B) \quad (2.10)$$

$$\epsilon_i = (I - B)(Z(s) - X(s)\beta) \quad (2.11)$$

Donde:

- B: matriz que contiene los parámetros de dependencia b_{ij} .
- I: matriz de identidad de dimensión requerida.
- (I - B) Matriz invertible.

Luego, considerando la expresión (2.1), se tiene:

$$e_i = Be_i + \epsilon_i \quad (2.12)$$

De aquí se obtiene:

$$e_i(I - B) = \epsilon_i \implies e_i = \epsilon_i(I - B)^{-1} \quad (2.13)$$

Bajo la definición anterior del modelo, $Z(s)$ distribuye de acuerdo a una distribución normal multivariante con la siguiente media y matriz de covarianza:

$$\mathbb{E}(Z(s)) = X(s)\beta + \mathbb{E}(e_i) \quad (2.14)$$

Posteriormente, bajo (2.13) se tiene que:

$$\mathbb{E}(Z(s)) = X(s)\beta + \mathbb{E}(I - B)\mathbb{E}(\epsilon_i) \quad (2.15)$$

Y como se mencionó al comienzo, ϵ_i tiene media igual a cero, por lo tanto, la media de $Z(s)$ finalmente queda de la siguiente manera:

$$\mathbb{E}(Z(s)) = X(s)\beta + \mathbb{E}(I - B) \cdot 0 \implies \mathbb{E}(Z(s)) = X(s)\beta \quad (2.16)$$

Así, la varianza queda expresada de la forma:

$$\mathbb{V}(Z(s)) = \mathbb{V}(e_i) \quad (2.17)$$

$$\mathbb{V}(Z(s)) = \mathbb{V}((\epsilon_i)(I - B)^{-1}) \quad (2.18)$$

$$\mathbb{V}(Z(s)) = (I - B)^{-1} \sigma^2 I ((I - B)^{-1})^T \quad (2.19)$$

$$\mathbb{V}(Z(s)) = (I - B)^{-1} \Sigma_S (I - B^T)^{-1} \quad (2.20)$$

Lo anterior, se da bajo el supuesto de que $(I - B)^{-1}$ exista.

Finalmente, se puede obtener una re-parametrización útil de este modelo definiendo:

$$B = \lambda W$$

Donde:

- λ : parámetro de autocorrelación espacial.
- W : Matriz de proximidad espacial (por lo general simétrica).

Luego, bajo esta especificación, y considerando que Σ_ϵ bajo un único parámetro σ^2 tal que $\Sigma_\epsilon = \sigma^2 I$, la expresión de la varianza de $Z(s)$ se convierte en:

$$\Sigma_{SAR} = \sigma^2 (I - \lambda W)^{-1} (I - \lambda W^T)^{-1}$$

Para que el modelo esté definido de manera correcta, se requiere que $(I - \lambda W)^{-1}$ sea no singular (invertible).

A este modelo, se le conoce como modelo autorregresivo simultáneo (SAR) y el término autorregresivo describe las sucesivas autorregresiones que ocurren al mismo tiempo en cada región.

2. Modelo CAR:

Puede resultar más intuitivo trabajar con un enfoque análogo al análisis de series de tiempo y especificar modelos para las distribuciones de probabilidad de cada observación $Z(s)$ condicionada a los valores observados de las otras observaciones. Es decir, modelar $f(Z(s_i)|Z(s_{-i}))$, donde $Z(s_i)$ corresponde al vector de todas las observaciones menos la correspondiente a la región $i \forall i$.

En el contexto de las series de tiempo, las variables aleatorias cumplen la propiedad de Markov cuando $f(Y_{t+1}|Y_t, \dots, Y_1) = f(Y_{t+1}|Y_t)$, lo que significa que el valor en el tiempo $t+1$ depende tan solo de lo ocurrido en el tiempo inmediatamente anterior t . Si esto se abarca a datos espaciales, el valor Y_i dependerá únicamente de lo que ocurra en sus vecinos, de esta manera, $Z(s_i)$ depende de $Z(s_j)$ solo si la localización j pertenece al conjunto de vecinos de i , N_i . Cuando esto se cumple se dice que el proceso $Z(S)$ es un campo aleatorio de Markov.

Considerando una distribución gaussiana, el modelo puede ser expresado de la forma:

$$Z(s) = X(s)\beta + C(Z(s) - X(s)\beta) + \epsilon_i \quad (2.21)$$

Así, con el enfoque autorregresivo condicional, la esperanza y varianza condicional de la variable respuesta dado el conjunto de vecinos queda de la siguiente manera:

- $\mathbb{E}(Z(s_i)|Z(s_{-i})) = X(s)\beta + C(Z(s) - X(s)\beta)$
- $\mathbb{V}(Z(s_i)|Z(s_{-i})) = \sigma_i^2; i = 1, \dots, n$

Por otro lado, paralelamente al modelo SAR, esta distribución condicional permite construir una distribución conjunta multivariada con:

- $E[Z(s)] = X(s)\beta$
- $Var[Z(s)] = (I - C)^{-1}\Sigma_C$

Con Σ_c definido como la diagonal de la matriz simétrica de varianzas covarianzas, constituida por los elementos $\sigma_1^2 \dots \sigma_n^2$.

Bajo lo anterior, los modelos CAR de primer orden, presentan la siguiente varianza:

$$Var[Z(s)] = \sigma_c^2(I - \lambda W)^{-1}$$

Donde: los componentes fueron definidos anteriormente en el modelo SAR.

Modelo	Planteamiento de la Fórmula
SAR	$Z(s) = X(s)\beta + Be_i + \epsilon$
CAR	$Z(s) = X(s)\beta + C[Z(s) - X(s)\beta + \epsilon]$

Cuadro 2.2: Comparación modelo SAR y CAR

Considerando ambos modelos autoregresivos, tanto el simultáneo como el condicional, es posible notar que en su notación general ambos modelos se componen por los mismos elementos, sin embargo, su diferencia radica en la componente que pondera la matriz de proximidad B o C, respectivamente.

Por otro lado, en términos de su varianza, se observa:

Modelo	Planteamiento de la Fórmula
SAR	$\Sigma_{SAR} = (I - B)^{-1}\Sigma_S(I - B^T)^{-1}$
CAR	$\Sigma_{CAR} = (I - \lambda C)^{-1}\Sigma_C$

Cuadro 2.3: Comparación varianza SAR y CAR

Donde tanto B como C, pueden ser expresadas por la reparametrización ya definida λW .

Finalmente, la **estimación de parámetros** para ambos modelos es posible realizarla mediante mínimos cuadrados generalizados y por máxima verosimilitud .

Observación:

Como se indicó en la metodología, cuando no exista presencia de autocorrelación espacial, el modelo adecuado más que un modelo autoregresivo, será una regresión lineal, cuyas características se mencionarán a continuación.

2.3. Regresión Lineal

El modelo de regresión lineal [20], considera una variable respuesta o de interés, la cual será explicada a partir de una variable explicativa (Regresión simple) o más variables explicativas (Regresión múltiple). Así, el objetivo es estimar o conocer el valor final de la variable de interés dado las variables que lo explican, es decir, estimar los parámetros.

Dentro de este modelo de regresión se debe realizar un análisis exhaustivo, de tal manera que se pueda determinar si las variables que se acondicionaron al modelo están asociadas y de ser así, en qué sentido lo están.

De forma general, para regresión múltiple se plantea el modelo:

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi} + \epsilon_i \quad ; \quad i = 1, \dots, n$$

Donde:

- Y_i : variable que representa la i-ésima observación de la variable respuesta.
- β_0 : intercepto, no tendrá una interpretación directa cuando el valor de $x = 0$ no pertenece al recorrido de la variable.
- β_1 : variación media de la variable respuesta cuando X aumenta en una unidad.
- X_i : i-ésima observacion de la variable independiente.
- ϵ_i : i-ésimo error aleatorio.

2.3.1. Análisis de Varianza (ANOVA)

La tabla ANOVA, se compone de la siguiente forma:

F.V.	g.l.	Suma cuadrados	Cuadrado medio	Estadístico F
Modelo	p	$SCM = \hat{\beta}'X'Y - n\bar{Y}^2$	CMM=SCM/p	CMM/CME
Error	n-(p+1)	$SCE = Y'Y - \hat{\beta}'X'Y$	CME= S^2 =SCE/n-(p+1)	
Total	n-1	$SCT = Y'Y - n\bar{Y}^2$		

Cuadro 2.4: Tabla ANOVA para regresión lineal múltiple

Luego, una vez ajustado el modelo, resulta importante evaluar de manera individual y global el aporte realizado por las variables independientes.

2.3.1.1. Test Global

El objetivo del test global, es evaluar si el conjunto de todas las variables independientes explican o no la variabilidad del modelo múltiple.

La dócima proviene de la tabla ANOVA:

$$H_0 : \beta_0 = \beta_1 = \dots = \beta_p = 0 \text{ versus } H_1 : \exists j \text{ tal que } \beta_j \neq 0 \quad ; \quad j = 1, \dots, p$$

El estadístico de prueba es:

$$F_{obs} = \frac{CMModelo}{CMError} \sim F_{(p, n-p-1)}$$

Luego, se considera:

- Si $F_{obs} > F_{(p, n-p-1)}$ se rechaza H_0 , por lo tanto, al menos una variable independiente es significativa en el modelo.
- Si $(F_{obs} \leq F_{(p, n-p-1)})$ el modelo no estaría ajustado.

2.3.1.2. Test Individual

Si en el teste global la hipótesis nula es rechazada, esto significa que al menos una de las variables independientes es significativa para el modelo. Luego, es de interés saber cuál o cuáles son. Para esto, se debe realizar dósimas individuales y analizar cada una de las p variables del modelo en cuestión.

Así, la dócima será:

$$H_0 : \beta_j = 0 \text{ versus } H_1 : \beta_j \neq 0 \quad ; \quad j = 1, 2, \dots, p$$

El estadístico utilizado es:

$$t_{obs} = \frac{\hat{\beta}_j}{\sqrt{\text{var}(\hat{\beta}_j)}} \sim t_{1-\frac{\alpha}{2}}(n-2)$$

Luego, si $t_{obs} > t_{1-\frac{\alpha}{2}}(n-2)$, existe evidencia suficiente para rechazar H_0 , es decir, la variable independiente x_j es significativa para el modelo en cuestión. De lo contrario, la variable no es significativa.

2.4. Selección de Variables

Sea Y la variable respuesta y X_i las p variables predictoras, el objetivo de la selección de variables es encontrar un subconjunto de las p variables independientes que presenten un buen ajuste de modelo de regresión para explicar la variable dependiente Y .

El objetivo del mejor modelo ajustado, será que entregue buenas predicciones y que las estimaciones de los coeficientes de interés sean significativas.

Para encontrar aquel subconjunto de variables predictivas necesarias para lograr el objetivo anterior, existen tres estrategias diferentes de selección. La idea de cada uno de estos, es elegir el mejor modelo en forma secuencial pero incluyendo (o excluyendo) una sola variable predictora en cada paso de acuerdo a ciertos criterios. El proceso secuencial termina cuando una regla de parada se satisfice.

Entre procedimientos existentes más utilizados, se encuentran los mencionados a continuación.

2.4.1. Eliminación hacia atrás (backward elimination)

En este caso, se comienza con el modelo completo y en cada paso se va eliminando una variable. Si todas las variables predictoras son significativas, es decir, tienen p-value pequeños para la prueba t establecida, entonces el mejor modelo es el que tiene todas las variables predictoras disponibles. En caso contrario, en cada paso se van eliminando variables independientes hasta que todas en el modelo sean significativas de acuerdo a algún criterio escogido. En este caso, el criterio será AIC, el cual se explicará más adelante.

El método **Backward** padece un efecto de anidamiento, esto es ya que toda variable que es eliminada del modelo, ya no vuelve a entrar a él.

2.4.2. Eliminación hacia adelante (forward selection)

Aquí, contrario a la estrategia anterior, el procedimiento consiste en ir agregando variables en cada fase hasta que ninguna de las variables que no han ingresado al modelo justifiquen que su inclusión sea un aporte significativo.

En esta ocasión, también está presente el efecto de anidamiento, puesto que toda variable que es añadida al modelo, ya no puede salir del mismo.

2.4.3. Regresión Stepwise

Este proceso fue introducido por Efron y Tibshirani en el año 1960 para corregir el problema de anidamiento de los dos métodos anteriores. De cierta manera, puede ser considerado como una modificación del método **Forward**, ya que se empieza con un modelo de regresión simple y en cada paso se puede añadir una variable en forma similar a este método, pero se evalúa si alguna de las variables que ya están presentes en el modelo puede ser eliminada.

El proceso es finalizado cuando ninguna de las variables, que no han entrado aún, tienen importancia suficiente como para entrar al modelo.

2.5. Validación de los Modelos

2.5.1. Bondad de Ajuste

Cuando se tiene más de un modelo, resulta importante saber cuál es el que mejor se ajusta a los datos. Para ello, se utilizan diferentes criterios de información basados en la verosimilitud. El más utilizado es el **Criterio de Información de Akaike (AIC)** (Akaike 1974), el cual será utilizado para esta comparación y se define a continuación.

2.5.1.1. Criterio de Información de Akaike (AIC):

Este estadístico se basa en la información y como se mencionó, mide la calidad que tiene un modelo. Su cálculo se da mediante la siguiente fórmula:

- **Regresión Lineal Simple**

$$AIC = 2l + 2(p + 1)$$

- **Resgresión Lineal Múltiple**

$$AIC = n + n \log(2\pi) + n \log\left(\frac{SSR}{n}\right) + 2((p + 1) + 1)$$

Donde:

- **p**: número de variables predictoras.
- **l**: logaritmo de la función de similitud.
- **SSR**: suma de cuadrados residual.

Una vez calculado el resultado para cada modelo, se debe elegir aquel cuyo AIC es mínimo.

2.5.2. Validación de supuestos

Una vez que el modelo está ajustado, el siguiente paso es realizar un diagnóstico del mismo con el objetivo de comprobar si los supuestos se cumplen. De esta manera, se podrá tener una noción respecto a si el modelo es apropiado o no para los datos existentes.

Algunos de estos supuestos se basan en el análisis de los residuos, los cuales corresponden a la diferencia entre el valor observado Y_i y el valor estimado \hat{Y}_i :

$$\epsilon_i = Y_i - \hat{Y}_i \quad ; \quad i = 1, \dots, n$$

Estos, presentan las siguientes propiedades:

- La media muestral es cero, $\bar{\epsilon} = \frac{\sum_{i=1}^n \epsilon_i}{n} = 0$
- El estimador de la varianza poblacional está basado en ellos.
- Los residuos, como variables aleatorias, no son independientes entre sí.

Los supuestos que se deben cumplir, son:

2.5.2.1. Normalidad de los residuos

Este supuesto es relevante para hacer inferencia en regresiones lineales múltiples y puede ser verificada de manera gráfica o numérica, no obstante, este supuesto puede no ser tan relevante cuando se tienen muestras de gran tamaño.

Respecto a las formas gráficas, se tienen las siguientes dos opciones:

1. Gráfico Cuantil - Cuantil (Q-Q Plot):

Este gráfico, compara cuantiles de una variable respecto a cuantiles de una distribución normal. Se debe observar si los puntos se aproximan al comportamiento lineal señalado en el gráfico; de ser así, se podría considerar que la distribución empírica es similar a la teórica. Por otro lado, si los puntos se alejan de la recta señalada, significaría la distribución empírica se aleja de una distribución normal.

2. Histograma

Este tipo de gráfico, muestra la forma de la distribución empírica y traspone una línea de la distribución normal teórica. De esta forma, se puede realizar una comparación y observar a primera vista si los datos se ajustan o no a una distribución normal.

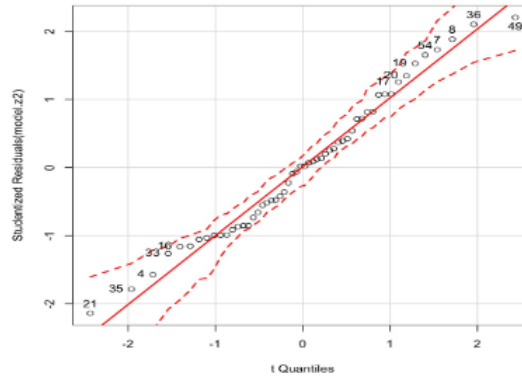


Figura 2.1: Gráfico Q-Q Plot

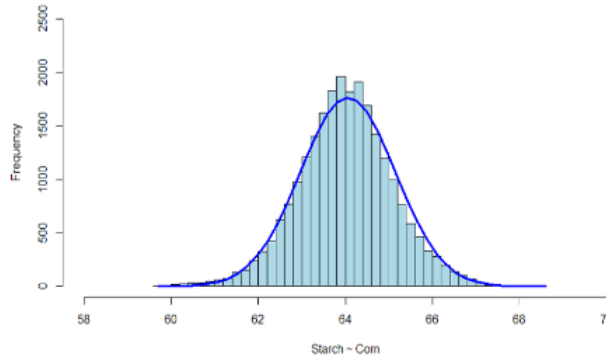


Figura 2.2: Histograma

Por otra parte, es posible encontrar Test que verifican la existencia de normalidad, como lo es el test **Shapiro - Wilks**.

En esta prueba, la hipótesis a probar es la siguiente:

$$H_0 : \text{Errores distribuyen normal versus } H_1 : \text{Errores no distribuyen normal}$$

El estadístico de prueba es:

$$W = \frac{(\sum_{i=1}^n a_i x_{(i)})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Donde $x_{(i)}$ es el número que ocupa la i -ésima posición en la muestra.

La región crítica del test está dada por $W/W_{obs} < W_{n,\alpha}$.

2.5.2.2. Homocedasticidad

Un modelo de regresión lineal exige que la varianza de las perturbaciones aleatorias sea constante, esto es, que exista homocedasticidad.

Si este supuesto no se cumple, esto es, si existe heterocedasticidad, se estaría indicando la varianza de los estimadores estaría sesgada, siendo los estadísticos t o F poco confiables.

Para detectar si se cumple o no este supuesto, se puede utilizar el **Contraste de Breusch Pagan**, el cual analiza si la varianza estimada de los residuos de una regresión depende de los valores de las variables regresoras.

La hipótesis a contrastar es:

H_0 : Los errores son homocedásticos versus H_1 : Los errores son heterocedásticos

Se utiliza el siguiente Estadístico de prueba:

$$BP = nR_{e^2}^2 \sim X_{k-1}^2$$

Donde k es el número de variables exógenas.

2.5.2.3. Errores No Correlacionados Espacialmente

Si la variable respuesta presenta dependencia espacial, se espera que luego de generar el modelo, los errores sean independientes entre sí, ya que la autocorrelación espacial ya estaría presente en el modelo.

Para verificar si se cumple este supuesto, es posible utilizar los test I de Moran y C de Geary ya definidos en el punto 2.2.1 (Dependencia Espacial).

Capítulo 3

Análisis de los datos

3.1. Encuesta Nacional Urbana de Victimización (ENUSC)

La ENUSC, es una encuesta realizada desde el año 2003 por la Subsecretaría de Prevención del Delito del Ministerio del Interior y Seguridad Pública en conjunto con el apoyo técnico del Instituto Nacional de Estadísticas (INE), con el objetivo de medir la victimización e inseguridad para la sociedad chilena más relevante en el país, sus regiones y las principales comunas urbanas.

Este instrumento de medida social, se realiza de manera anual y ha logrado constituirse como una de las encuestas de victimización más relevantes en América Latina.

Por lo anterior, los datos generados por esta encuesta podrán ser de gran utilidad para la realización del estudio mencionado y el cumplimiento de los objetivos planteados.

Se tomará en cuenta, como se plante inicialmente en el objetivo, los datos correspondientes a la Región Metropolitana, ya que es la región con mayor concentración de presencia de delitos a lo largo de Chile. A la vez, se utilizará la recolección de datos del año 2017 (publicadaos en el año 2018), dado que corresponden a la información más reciente disponible.

La base de datos consta de 88.237 casos muestrales y 671 variables. Luego, como se mencionó, el objetivo es analizar el comportamiento de delitos a nivel comunal, específicamente en la **Región Metropolitana**, se ha generado un filtro que mantenga los datos de solo aquellas personas cuya residencia coincida con la indicada mediante el uso de la siguiente variable y nivel especificado:

Variable	Nombre Variable	Código	Etiqueta
Producto enc_región	Región	13	Resgión Metropolitana

Luego de este proceso, la muestra se reduce a 22.961 casos en donde se encuentran 42 comunas diferentes, las cuales son:

Buín	Cerrillos	Cerro Navia	Colina
Conchalí	El Bosque	Estación Central	Huechuraba
Independencia	La Cisterna	La Florida	La Granja
La Pintana	La Reina	Lampa	Las Condes
Lo Barnechea	Lo Espejo	Lo Prado	Macul
Maipú	Melipilla	Ñuñoa	PAC
Padre Hurtado	Paine	Peñaflor	Peñalolen
Providencia	Pudahuel	Puente Alto	Quilicura
Quinta Normal	Recoleta	Renca	San Bernardo
San Joaquín	San Miguel	San Ramón	Santiago
Talagante	Vitacura		

Una vez ya seleccionados los casos de interés, será posible en base a los datos, analizar variables y preguntas seleccionar y construir diferentes variables.

3.2. Análisis de Preguntas

En primer lugar, para dar una correcta lectura a los datos, es importante mencionar que en el Manual de usuario entregado por la ENUSC, se menciona el uso del factor expansión a cada unidad de análisis según sea el tipo de pregunta. El factor de expansión a utilizar en esta ocasión, corresponde al **factor hogar** y se interpreta como la **cantidad de hogares representados en la población**. Este, es calculado como el inverso de la probabilidad de selección de la unidad muestral.

Aquí, posterior a su aplicación, se obtiene que existe un total de 2.079.435 hogares respresentados con la muestra.

Para tener un primer conocimiento del comportamiento de los datos frente a la problemática, será útil indagar en ciertas preguntas realizadas en algunos módulos de interés.[4]

3.2.1. Módulo de Víctimización

Este módulo está compuesto por la medición de los siete delitos de mayor connotación social que mide la ENUSC, los cuales son [19]:

- **Robo con violencia e intimidación (A):** Este delito consiste en la apropiación de algún objeto ajeno, efectuada sin la voluntad de su dueño y con ánimo de lucro, ejerciendo el autor fuerza física y moral sobre la víctima, pues utiliza violencia o coacción en las personas.
- **Robo por sorpresa (B):** Es aquel que se produce cuando el autor del delito se aprovecha de un agolpamiento de personas o lo simula para arrebatar las pertenencias de manera agresiva, imprevista y fugaz, pero sin propinarle algún golpe o amenazas a la víctima .

- **Robo con fuerza en vivienda (C):** El robo con fuerza en las cosas se caracteriza por el empleo de energía para vencer la especial protección de que la cosa está revestida, en esta ocasión, el uso de fuerza para invadir dentro de la vivienda.
- **Hurto (D):** El delito de hurto consiste en la apropiación no violenta de una cosa mueble ajena, efectuada sin la voluntad de su dueño y con ánimo de lucro.
- **Lesiones (E):** Se refiere a causar daño o menoscabo a la integridad corporal o salud de una personas.
- **Robo de vehículos (G):** Realizar la apropiación de un vehículo ajeno sin la voluntad de su dueño.
- **Robo o hurto desde vehículo (H):** Realizar la acción de robar o hurtar objetos del interior de un vehículo.

En el año 2015, se agregan a esta lista también:

- **Delitos de connotación económica (I):** Son todas aquellas conductas ilícitas cometidas por personas naturales, personalmente o a través de personas jurídicas, que afectan el patrimonio de una o más víctimas, el sistema financiero o el mercado en general.
- **Amenazas (J):** Consistente en amenazar o anunciar la realización de causar un mal futuro verosímil a una persona, a su familia, honra o propiedad con el fin de causar inquietud o miedo en el amenazado.
- **Daños o Vandalismo (J):** Este delito incluye cualquier conducta intencional destinada a destruir, alterar o profanar los bienes que pertenece a un tercero.
- **Delitos CibernéticoS (J):** Corresponden a delitos mediante el uso de computadora, que generalmente pueden dividirse en específicos del uso de computadoras y delitos que se facilitan gracias a las estas pero que ya existen.

Para la construcción de diferentes variables, se dará uso a ciertas preguntas de interés establecidas en este modulo de la encuesta, las cuales son:

1. **Durante los últimos doce meses, ¿usted o algún miembro de su hogar ha sufrido este delito? (A1_1_1)**

- | | |
|------------|----------------|
| 1 Si | 99 No responde |
| 2 No | 96 Sin dato |
| 88 No Sabe | |

2. **¿Usted o alguien denunció el o los delitos? (A2_1_1)**

- | | |
|------------|----------------|
| 1 Si | 99 No responde |
| 2 No | 96 Sin dato |
| 88 No Sabe | |

3. La última vez, ¿dónde le sucedió esto a usted o a algún miembro de su hogar? (A4_1.1)

- | | |
|---|---------------------------|
| 1 Dentro de la vivienda | 6 En otra región del país |
| 2 Frente de la propia vivienda o edificio | 7 En el extranjero |
| 3 En su barrio | 88 No sabe |
| 4 En otra parte de la comuna donde vive | 99 No responde |
| 5 En otra comuna de la región donde vive | 96 Sin dato |

4. ¿Cuál fue el principal motivo para no denunciar el incidente? (A18_1.1)

- | | |
|---|----------------|
| 1 La pérdida no fue lo suficientemente seria | 77 Otro |
| 2 Porque el problema se solucionó | 88 No sabe |
| 3 Porque no tenía testigos | 99 No responde |
| 4 Por temor a amenazas/represalias | 96 Sin dato |
| 5 Porque no tenía seguro | |
| 6 Porque conoce a los responsables | |
| 7 Por temor a encarar a los responsables en el juicio | |
| 8 La policía no podría haber hecho nada | |
| 9 Porque los responsables lo/a amenazaron | |
| 10 La justicia (tribunales) no hubiera hecho nada | |
| 11 Porque el trámite demanda mucho tiempo | |
| 12 Porque tiene algún parentesco con los responsables | |
| 13 La policía no era necesaria | |
| 14 La policía no hubiera hecho nada | |
| 15 La policía me recomendó no registrar la denuncia | |

3.2.2. Modulo de Inseguridad y reacción frente al delito

Luego, será de gran interés y utilidad, analizar las preguntas estipuladas en este modulo para conocer las percepciones existentes respecto a la seguridad y los actos delictivos, y por ende, factores que estarían afectando en la decisión de reportar o no un delito.

En esta ocasión, se utilizaron las siguientes interrogantes:

1. De acuerdo con su percepción ¿Cuál diría usted que es la principal causa de los niveles de delincuencia que actualmente existen en su barrio? En primer lugar (P8_1.1).

1	La falta de presencia de carabineros en su barrio	88	No sabe
2	La falta de vigilancia municipal	99	No responde
3	La falta de efectividad de las policías en su barrio	96	Sin dato
4	Falta de intervención del municipio en problemas sociales del barrio		
5	Falta de organización de los vecinos del barrio		
6	La existencia de problemas familiares en los hogares del barrio		
7	La ocupación de lugares del barrio por pandillas y grupos peligrosos		
8	La existencia de comercio en el barrio		
9	La falta de espacios públicos para el uso de vecinos en el barrio		
10	La existencia de lugares sucios y abandonados en el barrio		
11	La falta de iluminación de calles y casas del barrio		
12	Venta de drogas en el barrio		
13	La distribución de las casas en el barrio		
14	Lo atractivo del barrio para los delincuentes		
15	Concentración de locales de diversión nocturna en que se consume mucho alcohol		
16	Consumo de drogas o alcohol en plazas y espacios públicos		
77	Otra		
85	No cree que exista delincuencia en su barrio		

2. Pensando en la delincuencia, usted diría que durante los últimos doce meses la delincuencia: En su comuna (P3_2.1):

1	Aumentó	2	Se mantuvo
3	Disminuyó	88	No sabe
99	No responde	96	Sin dato

3. Pensando en la delincuencia, usted diría que durante los últimos doce meses la delincuencia: En su barrio (P3_3.1).

1	Aumentó	2	Se mantuvo
3	Disminuyó	88	No sabe
99	No responde	96	Sin dato

3.2.3. Caracterización socio-demográfica

Pensando en el objetivo establecido, será de interés tomar en cuenta preguntas que apunten al Nivel Socioeconómico de las personas. Para esto, se seleccionó la siguiente pregunta:

1. **R6: Nivel más alto alcanzado (rph_nivel).**

0 Nunca asistió	10 Profesional
1 Sala cuna / Jardín infantil	11 Postítulo
2 Prekinder/Kinder	12 Magíster
3 Básica	13 Doctorado/Postdoctorado
4 Primaria o Preparatoria	90 Educación diferencial
5 Media científico-humanista	96 Sin dato
6 Humanidades (sistema antiguo)	99 Nivel Ignorado
7 Media Técnico Profesional	
8 Técnica Comercial, Industrial o Normalista (sistema antiguo)	
9 Técnico de Nivel Superior	

3.3. Construcción de Variables

Para construir variables de interés, se tomarán en cuenta ciertas respuestas por cada pregunta especificada anteriormente, según lo que se desee calcular y conocer.

A continuación, se menciona lo que indica cada variable construida y se explica por cada una lo considerado del ítem correspondiente para su formulación, indicando finalmente entre paréntesis la pregunta utilizada en cada caso (Denotadas en el punto 3.2).

3.3.1. Variable Respuesta

1. **V1:Tasa de ocurrencias:** calculada mediante el cociente de la suma de los casos cuya respuesta fue la opción **1.sí** como respuesta al haber sufrido el delito en cuestión, sobre el total de casos (**A1_1_1**).
2. **V2: Tasa de denuncias:** calculada mediante el cociente de la suma de los **1.sí** como respuesta al haber denunciado el delito en cuestión, sobre el total de casos (**A2_1_1**).

3.3.2. Variables Predictoras

1. **V3: Tasa de delitos que ocurren en la misma comuna de residencia:** calculada mediante el cociente entre la suma de los casos en que la respuesta respecto al lugar de ocurrencia del delito, indican un sitio dentro de su comuna. Las opciones contempladas son: 1. Dentro de la vivienda, 2. Frente de la propia vivienda o edificio, 3. En su barrio y 4. En otra parte de la comuna donde vive (**A4_1_1**).
2. **V4: Tasa de personas que no denuncian delitos por tener una percepción negativa respecto a la efectividad de la justicia:** calculada mediante el cociente entre la suma de casos cuya respuesta es una de las dos siguientes opciones: 10. La justicia (tribunales) no hubiera hecho nada y 14. La policía no hubiera hecho nada, sobre el total de casos (**A18_1_1**).

3. **V5: Tasa de personas que piensa que la efectividad de la policía en su barrio es negativa y que se relaciona directamente con los niveles de delincuencia en sus barrios:** calculada mediante el cociente entre la suma de los casos en que la respuesta seleccionada fue: 3. La falta de efectividad de las policías en su barrio, sobre el total de casos (**P8_1_1**).
4. **V6: Tasa de personas que considera que su barrio seguro:** calculada mediante el cociente entre la suma de casos cuya respuesta fue la opción: 85. No cree que exista delincuencia en su barrio, sobre el total de casos (**P8_1_1**).
5. **V7: Tasa de personas que considera que en el último año la delincuencia ha tenido un aumento a nivel comunal :** calculada mediante el cociente entre la suma de casos en que la respuestas fue: 1. Aumentó, sobre el total de casos (**P3_2_1**).
6. **V8: Tasa de personas que considera que en el último año la delincuencia ha tenido un aumento en el barrio:** calculad mediante el cociente entre la suma de casos en que la respuesta fue: 1. Aumentó, sobre el total de casos (**P3_3_1**).
7. **V9: Tasa de personas que posee nivel educacional profesional:** calculada mediante el cociente de la suma de personas que responde a su nivel más alto alcanzado como: 10. Profesional, sobre el total de casos (rph_nivel).
8. **V10: Tasa de personas cuyo máximo nivel es Enseñanza Media:** calculada mediante el cociente entre la suma de personas cuya respuesta a su máximo nivel alcanzado fue algunas de las opciones: 7. Media Técnico Profesional (o aquellas inferior a este), sobre el total de casos (rph_nivel).

Con la construcción de las variables anteriormente descritas, será posible generar una base de datos que contemple aquellas variables mencionadas tanto para cada tipo de delito como a nivel general, para posteriormente, de acuerdo al objetivo y problemática en cuestión, realizar un análisis descriptivo y espacial para la búsqueda de resultados de interés.

3.4. Análisis de datos

3.4.1. Análisis por delito

En esta ocasión, los datos se han agrupado por delito, obteniendo la cantidad de casos a nivel general en la Región Metropolitana, tanto de ocurrencia como de denuncias, para así poder analizar un comportamiento global de estos hechos delictuosos.

Se ha realizado la siguiente tabla resumen:

De lo anterior, se obtienen las siguientes observaciones:

Delito	Casos	Denuncias	Tasa de Denuncias
Robo con intimidación	163.495	78.424	47.97 %
Robo por sorpresa	131.100	34.152	26.05 %
Robo en vivienda	88.324	38.655	43.77 %
Hurto	217.139	41.121	18.94 %
Lesión	36.913	14.787	40.06 %
Robo de vehículo	14.812	14.528	98.08 %
Robo desde vehículo	174.533	57.227	32.79 %
Económico	201.892	78.795	39.09 %
Amenaza	99.671	41.550	41.69 %
Vandalismo	122.474	3.674	3.00 %
Vandalismo de vehículo	100.001	7.612	7.61 %
Cibernético	161.973	37.657	23.35 %
Total	1.512.327	448.182	29.64 %

Cuadro 3.1: Tabla Resumen por delitos

1. Existen 1.512.327 casos en donde al menos un integrante del hogar es víctima de algún tipo de delito en la Región Metropolitana, los cuales son denunciados en un 29.64 % de las ocasiones.
2. El delito con mayor ocurrencia en la Región Metropolitana (RM) en el año, es el hurto con un total de 217.139 casos. Le siguen los delitos económicos, robo desde autos y delitos cibernéticos.
3. El delito que ocurre con menor frecuencia corresponde al robo de vehículos, sin embargo, al mismo tiempo es el más denunciado (se realiza la denuncia en el 98.08 % de los casos, lo cual se podría deber a que por lo general tienen asociado un seguro).
4. El delito menos denunciado es el vandalismo ya sea en hogares o vehículos, pues se denuncian un 3.00 % y 7.61 % de las veces respectivamente.

Lo anterior, puede ser atambién apreciado en el siguiente gráfico:

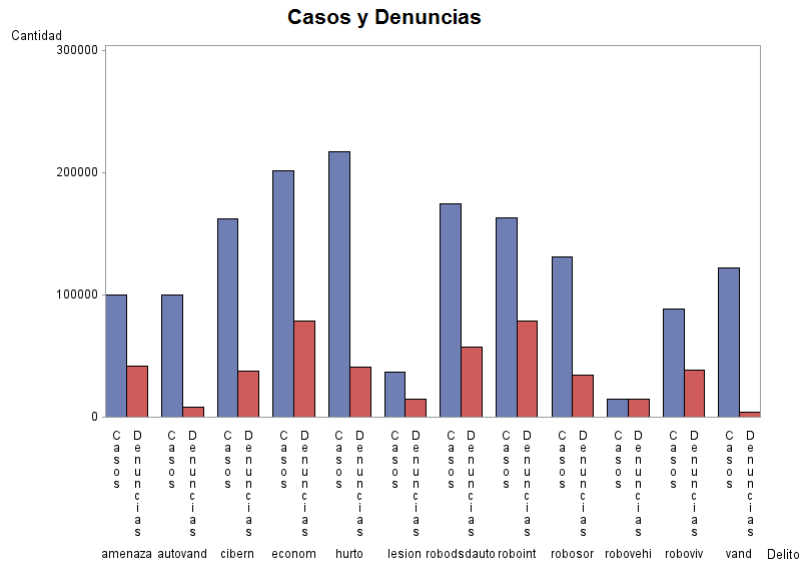


Figura 3.1: Gráfico de Ocurrencias v/s Denuncias por delito

3.4.2. Análisis de delitos por comuna

Considerando cada uno de los delitos de mayor connotación social en conjunto con los resultados analizados anteriormente, se ha seleccionado de acuerdo a sus ocurrencias el análisis para cada uno de los siguientes delitos agrupados por comuna.

3.4.2.1. Hurto

- Análisis general

1. La gente cuya residencia es en la comuna de Maipú, es la que más ha sido víctima de hurto en los últimos 12 meses (29.305 casos), seguida por quienes viven en Las Condes y Puente Alto.
2. La gente que vive en la comuna de Melipilla, es la que menos ha sido víctima de hurto en los últimos 12 meses (423 casos), seguidas por quienes residen en Huechuraba y Paine.
3. La gente que más tiende a denunciar este delito, es la que vive en la comuna de Huechuraba (se reporta aproximadamente un 67.00% de las veces), seguida por quienes viven en Las Condes y la Reina.
4. Las personas que menos denuncian el delito de hurto, son las que viven en las comunas Padre Hurtado (se denuncian el 2.5% de las veces), Conchalí y Lo Espejo.

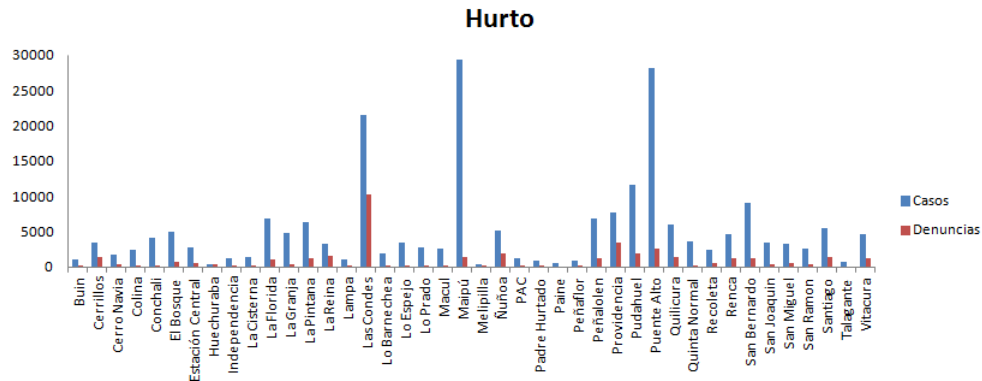


Figura 3.2: Gráfico Hurto : casos v/s denuncias

■ **Análisis de dependencia espacial**

1. **Tasa de ocurrencias**

• **Mapa de Distribución Espacial**

Con este tipo de mapa es posible identificar a primera vista la distribución de comunas en la Resgión Metropolitana, junto a la posible existencia o no de dependencia espacial, la cual se representa en la tonalidad de los colores. Tonos más claros apuntan a tasas menores y colores más oscuros a tasas mayores. La secuencia de una misma tonalidad podría indicar autocorrelación entre aquellas comunas, logrando hacer una idea de lo que podría estar sucediendo al respecto y en qué sectores. En esta oportunidad, no se identifica alguna distribución muy clara que pudiera indicar existencia de dependencia espacial:

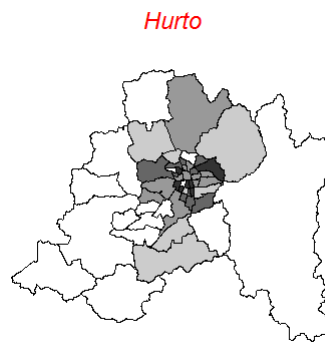


Figura 3.3: Mapa tasa ocurrencia hurto: Distribución Espacial

• Test de dependencia espacial

Test	Indicador	p_value (Normal)	p_value (Randomization)
I Moran	0.1082	0.1582	0.1610
c de Geary	0.8631	0.1587	0.1551

Cuadro 3.2: Test de dependencia para Tasas de ocurrencias de Hurto

De acuerdo a los p_valores obtenidos, bajo un nivel del 10% de significancia, no se rechaza en todos los casos la hipótesis nula y por lo tanto, no existe dependencia espacial global. Luego, se verifica si existe de manera local mediante el siguiente gráfico:

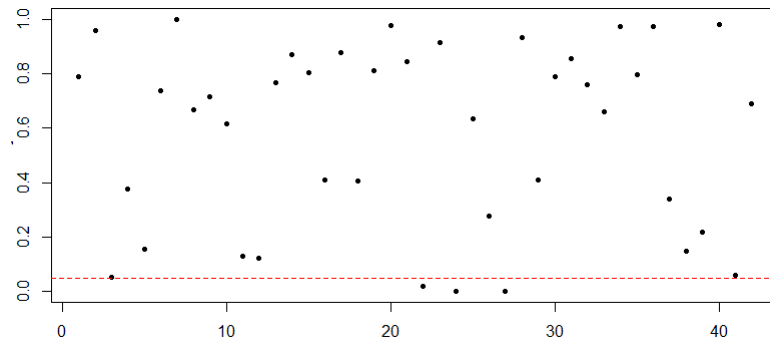


Figura 3.4: Test Local Moral para tasa de ocurrencia de delitos

Es posible verificar que existen 3 comunas en la zona de rechazo (bajo la línea roja), las cuales son: **Peñaflor, Pedro Aguirre Cerda y Melipilla** y que poseen dependencia espacial local, lo que significa que las tasas de ocurrencias de hurto son similares en estas comunas respecto a sus comunas vecinas.

• Aplicación de los modelos

Dado que no existe presencia de autocorrelación espacial, el modelo que se deberá generar corresponde a un modelo de Regresión Lineal. Se obtienen los dos siguientes modelos candidatos:

Modelo	N variables	AIC
Mod 1	1	-142.4
Mod 2	2	-144.2

Cuadro 3.3: Modelos Lineales para tasa de ocurrencias del hurto

Ambos modelos poseen un AIC muy similar, luego, bajo principio de parsimonia ¹ el modelo final será el primero (denotado Mod 1):

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.0888	0.0113	7.8352	0.0000
V9	0.0633	0.0429	1.4741	0.1083

Como se ve, a un nivel de 10% de significancia, la variable del modelo es significativa y afecta en dirección positiva. Esto significa que mientras mayor es la tasa de profesionales, mayor es la tasa de ocurrencia de hurtos.

- **Errores no correlacionados espacialmente**

Gráficamente es posible verificar que los residuos del modelo, no poseen dependencia espacial, validándose así este supuesto:

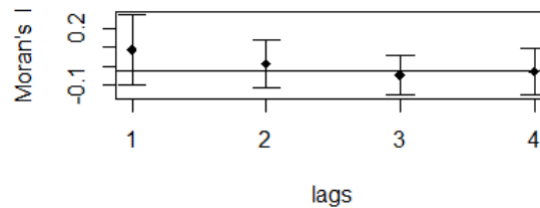


Figura 3.5: Gráfico de Dependencia espacial a k distancias para la tasa de ocurrencia por hurto

Como muestra la figura, a un lag 1, el intervalo de confianza no se encuentra posicionado en ninguno de los dos sectores de rechazo, no dejando estipulado de esta forma la existencia del fenómeno de autocorrelación espacial ni positiva ni negativamente.

2. Tasa de denuncias

- **Mapa de distribución espacial**

En el siguiente mapa, se logra identificar una posible presencia de dependencia espacial, ya que los tonos más oscuros del mapa están concentrados al centro de este y se agrupan tonos más leves hacia el costado norte de la Región (lo que indicaría tasas menores de la variable en cuestión):

¹Este principio se utiliza generalmente para elegir una de dos alternativas en situación de aparente igualdad y hace referencia a lograr explicar la variable respuesta de buena forma mediante menos variables explicativas, es decir, un modelo más simple pero de igual modo que se ajusta de buena manera.

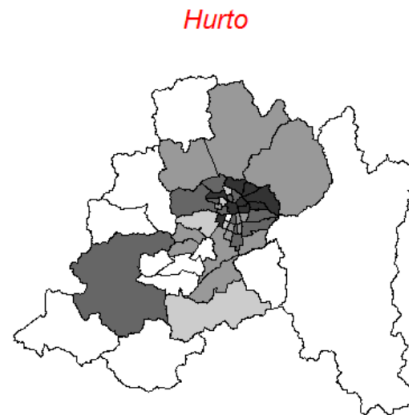


Figura 3.6: Mapa Tasa denuncia Hurto: Distribución espacial

- Test de dependencia espacial

Test	Indicador	p_value (Normal)	p_value (Randomization)
I Moran	0.1688	0.0399	0.0351
c de Geary	0.8095	0.0498	0.0569

Cuadro 3.4: Test de dependencia espacial para tasas de denuncias de Hurto

Anlizando los p-valores obtenidos, es posible confirmar para todos los casos, que con un nivel del 10% de significancia, se rechaza H_0 y por lo tanto, existe presencia de autocorrelación espacial global.

Por otro lado, es posible analizar mediante un Test Local de Moran el siguiente gráfico y resultado:

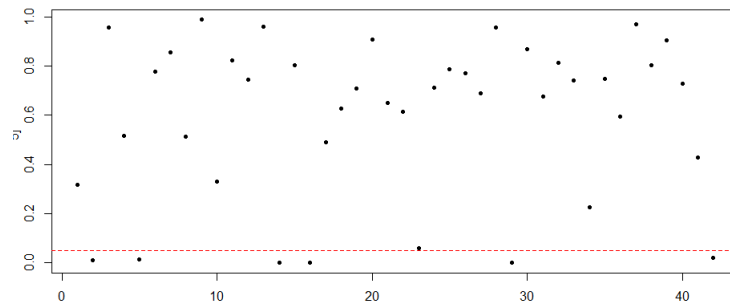


Figura 3.7: Test Local de Moran para tasa de denuncias de hurto

Se distinguen 6 puntos en la zona de rechazo, los cuales representan a las siguientes 6 comunas:

Cerrillos	Las Condes
Conchalí	Providencia
La Reina	Vitacura

Luego, un nivel del 10% de significancia, estas comunas poseen dependencia espacial de forma local, esto es, que las tasas de denuncias tienen un comportamiento similar en cada de estas 6 comunas respecto a sus comunas vecinas.

Finalmente, es posible observar también en la siguiente figura la existencia de este fenómeno en sus residuos para un lag=1 y lag=2 (pues el intervalo indica que el valor se encuentra por sobre la línea, es decir, en la zona de rechazo, indicando en esta ocasión que existe autocorrelación positiva).

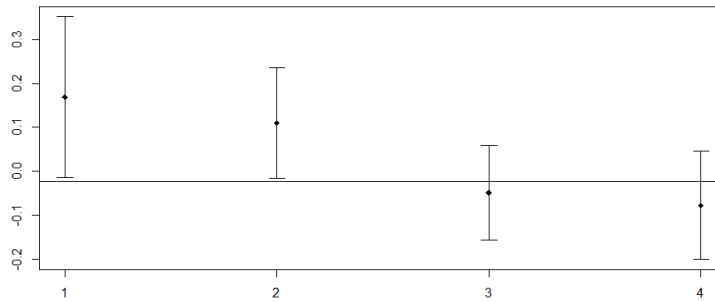


Figura 3.8: Gráfico de Dependencia espacial a k distancias para la tasa de denuncias por hurto

• Aplicación de los modelos

Modelo	N Variables	AIC	Significancia parámetro espacial
SAR	3	-56.49	No
CAR	2	-55.72	No

Cuadro 3.5: Comparación SAR Y CAR para tasa de denuncias por hurto

Tomando en cuenta el criterio AIC, el modelo que se ajusta de mejor forma es el SAR. Sin embargo, su parámetro espacial no resulta ser necesario incluirlo para poder explicar este fenómeno. Por lo tanto, se realiza un modelo de regresión lineal con las variables que tenía este modelo más no considerando , que corresponde al parametro espacial. Los resultados son:

Las tres variables del modelo muestran ser significativas y V9 influye directamente en la tasa de denuncias por hurto, mientras que las otras eafectan en sentido contrario.

Para V5, mientras mayor es la tasa de personas que piensa que falta vigilancia policial , se da una menor tasa de denuncias.

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.1854	0.0547	3.3871	0.0017
V9	0.5905	0.1208	4.8864	0.0000
V5	-1.7645	0.7796	-2.2635	0.0294
V6	-1.2875	0.8054	-1.5986	0.1082

Para V6, se tendrá que mientras más se piensa que el barrio es seguro, se da una menor tasa de denuncias de este delito o viceversa.

■ Errores no correlacionados espacialmente

Una vez verificada la existencia de dependencia espacial y habiendo generado un modelo bajo estos términos, se procede a a verificar el cumplimiento de el supuesto de no correlación de los residuos, puesto que con el modelo aplicado, este fenómeno debería estar en él, más no en los residuos.

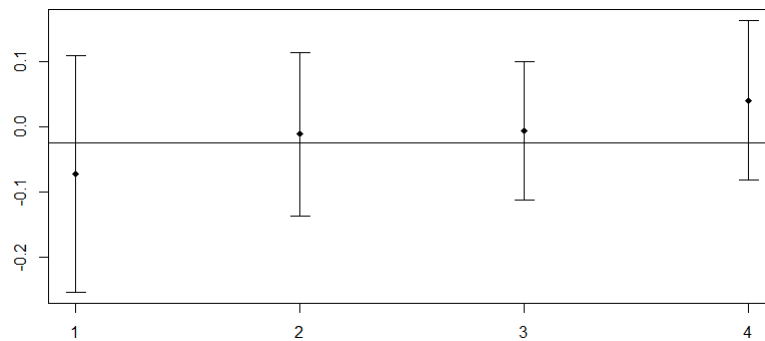


Figura 3.9: Gráfico de dependencia a k distancias para tasa de denuncias de hurto (Modelo)

En efecto, tanto a lag = 1 como para lag = 2, no existe dependencia espacial en los residuos, pues ya no se posicionan en la zona de rechazo como en la figura 3.7, cumpliendo así el supuesto de no correlación.

3.4.2.2. Robo con violencia e intimidación

1. La mayoría de las víctimas de este delito en los últimos 12 meses, corresponde a gente de la comuna de Puente Alto (27.313 casos), seguida por quienes residen en Maipú y Pudahuel.
2. La menos cantidad de víctimas vive en la comuna de Paine (415 casos), Buin y Padre Hurtado.
3. La gente que más denunció este delito tiene su hogar en Pedro Aguirre Cerda (se denuncia el 85.00 % de las veces), seguido por Lo Barnechea e Independencia.
4. La gente que menos tiende a denunciar este delito, es la que vive en la comuna de Melipilla (24.00 % de la veces),Peñalolen y Lo Espejo.

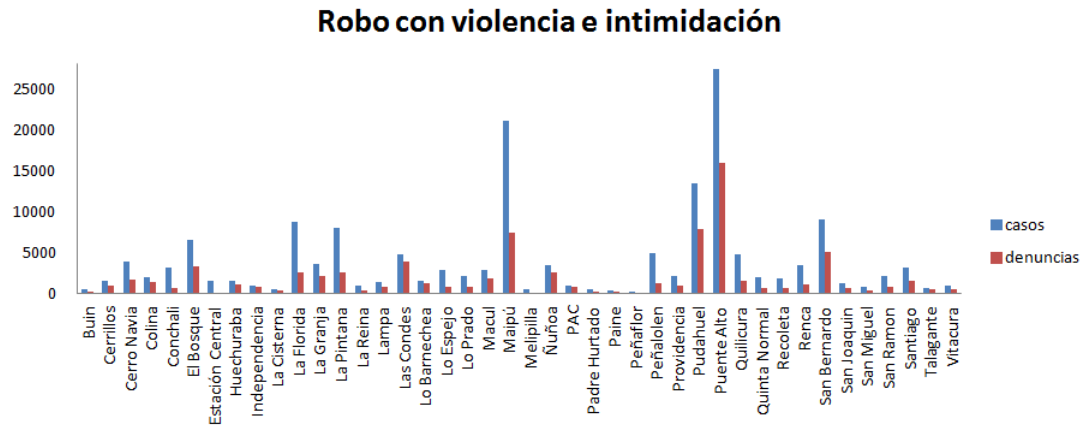


Figura 3.10: Gráfico Robo con violencia e intimidación: casos v/s denuncias

Luego, se realiza un análisis de dependencia espacial :

1. Tasa de ocurrencias

■ Mapa de Distribución Espacial

Robo con violencia e intimidación

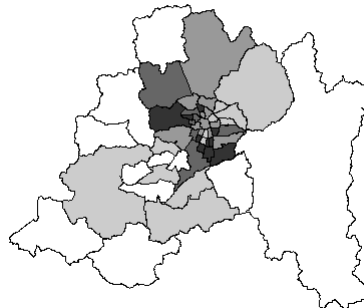


Figura 3.11: Mapa de distribución espacial de la tasa de ocurrencias de robo con violencia

Es posible observar una posible presencia de dependencia espacial, como se ve, hacia el sector suroeste del mapa se da una continuidad de comunas vecinas con baja tasa de ocurrencia de robo con violencia (ya que poseen tonos claros) respecto al centro del mapa que posee tonos más oscuros indicando una alta tasa de ocurrencia de este delito.

■ Test de dependencia espacial

De acuerdo a los p_ valores obtenidos, bajo un nivel del 10% de significancia, se rechaza en todos los casos la hipótesis nula y por lo tanto, existe dependencia espacial global.

Test	Indicador	p_value (Normal)	p_value (Randomization)
I Moran	0.3366	0.0001	0.0001
c de Geary	0.6398	0.0002	0.0002

Cuadro 3.6: Test de dependencia para Tasas de ocurrencias de Hurto

Luego, localmente se obtiene:

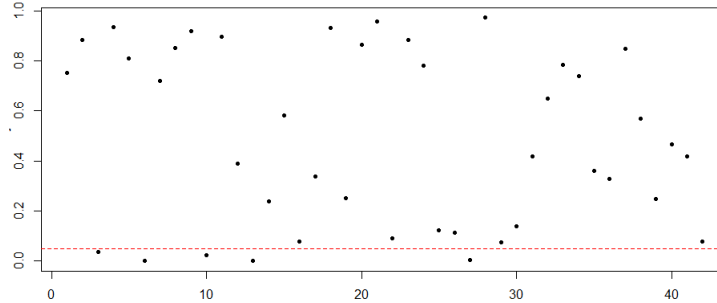


Figura 3.12: Test Local Moral para tasa de ocurrencia de robo con violencia

Existen 5 comunas en la zona de rechazo (bajo la línea roja), donde las tasas de ocurrencias de robo con violencia son similares en ella con respecto a sus comunas vecinas. Estas comunas son:

Cerro Navia	La Pintana
El Bosque	Peñaflor
La Cisterna	

Luego, para los residuos se observa el siguiente comportamiento:

En la figura, a un lag=1 y lag=3 se aprecia una presencia de autocorrelación espacial positiva y negativa respectivamente.

■ **Aplicación de los modelos**

Considerando que en ambos modelos solo existe una variable, se toma en cuenta el criterio AIC. Así, el modelo que se ajusta de mejor forma es el SAR por una leve ventaja. El parámetro espacial resulta significativo tomando un valor =0.5268. Los resultados se muestran en cuadro 3.8.

Así, con un nivel del 10% de significancia, tanto V10 como ρ , son significativos para la contribución del modelo.

Luego, que V10 tenga un valor estimado positivo, va a indicar que en lugares donde la tasa de personas cuyo máximo estudio es la enseñanza media, existe una mayor tasa de ocurrencia de robos con violencia.

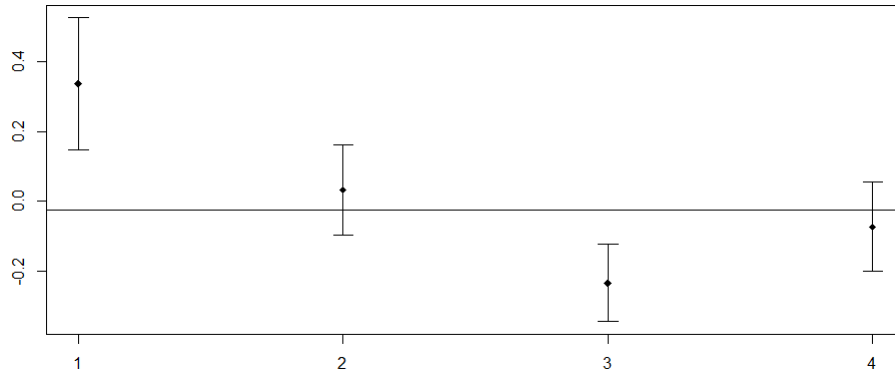


Figura 3.13: Gráfico de dependencia a k distancias

Modelo	N Variables	AIC	Significancia parámetro espacial
SAR	1	-167.86	Si
CAR	1	-167.26	Si

Cuadro 3.7: Comparación SAR Y CAR para tasa de denuncias por hurto

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.0267	0.0187	1.43	0.1530
V10	0.0714	0.0251	2.85	0.0044

Cuadro 3.8: Modelo Final Tasa de ocurrencias Hurto

- **Errores no correlacionados espacialmente**

Gráficamente es posible verificar que los residuos del modelo, no poseen dependencia espacial, pues se encuentran fuera de las zona de rechazo, validándose así este supuesto:

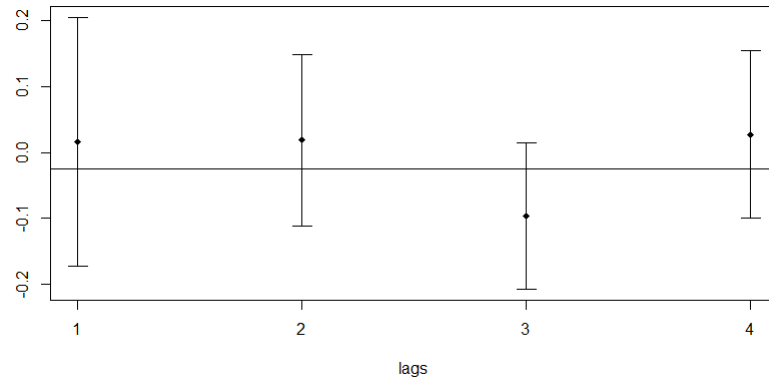


Figura 3.14: Gráfico de Dependencia espacial a k distancias para tasa de ocurrencia por robo con violencia e intimidación (Modelo)

2. Tasa de denuncias

- **Mapa de Distribución Espacial**

Robo con violencia e intimidación

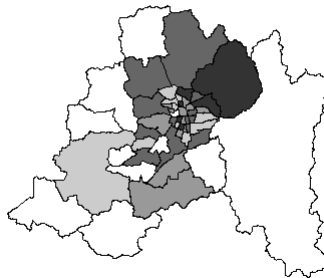


Figura 3.15: Mapa tasa de denuncias robo violencia e intimidación: Distribución Espacial

No se observa una gran dependencia espacial a nivel general en el mapa, sin embargo si en ciertos sectores puntuales, donde dos o 3 comunas vecinas muestran misma tonalidad.

- **Test de dependencia espacial**

Anlizando los p-valores obtenidos, es posible confirmar para todos los casos, que con un nivel del 10 % de significancia, no se rechaza H_0 y por lo tanto, no existe dependencia espacial global.

Test	Indicador	p_value (Normal)	p_value (Randomization)
I Moran	-0.0271	0.9764	0.9766
c de Geary	1.003	0.9750	0.9747

Cuadro 3.9: Test de dependencia espacial para tasas de denuncias de Hurto

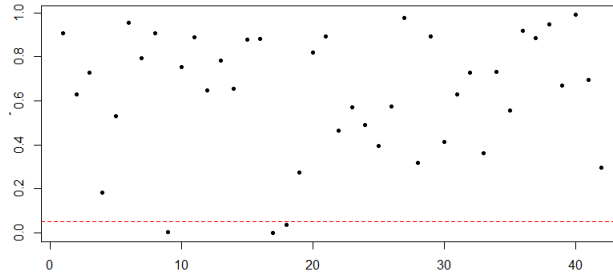


Figura 3.16: Test Local de Moran para tasa de denuncias de robo con violencia

Por otro lado, con el test local se obtiene:

Se distinguen 3 puntos en la zona de rechazo, los cuales representan a las comunas: **Independencia, Lo Barneche y Lo Espejo**, lo que indica pese a no existir dependencia global, estas 3 comunas poseen dependencia espacial de carácter local.

■ **Aplicación de los modelos**

Dado que no existe presencia de autocorrelación espacial global, el modelo que se deberá generar corresponde a un modelo de Regresión Lineal. El mejor modelo es el siguiente:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.8083	0.1071	7.55	0.0000
V4	-0.3315	0.1325	-2.50	0.0166
V10	-0.3251	0.1360	-2.39	0.0218

Este modelo, posee un AIC=19.32, y con un nivel del 10% de significancia las variables V4 Y V10 son significativas y se encuentran en dirección negativa. Esto es que una alta presencia de ambas variables implican una mayor tasa de denuncia de robos con violencia e intimidación.

Por ejemplo, para V4, Mientras mayor es la percepción negativa de la efectividad de la justicia, menor es la tasa de denuncias de este hecho delictivo.

■ **Errores no correlacionados espacialmente**

Resulta necesario a verificar el cumplimiento de el supuesto de no correlación de los residuos. Se obtiene:

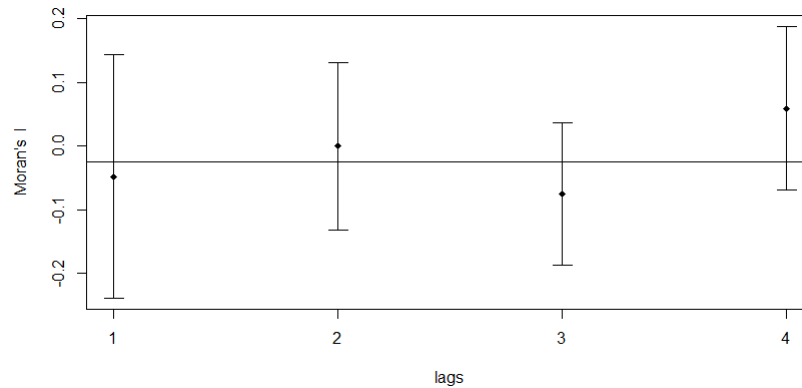


Figura 3.17: Gráfico de dependencia a k distancias para tasa de denuncias de robo con violencia e intimidación (Modelo)

En efecto, el supuesto se cumple, pues el intervalo no se establece en ninguna de las dos zonas de rechazo.

3.4.2.3. Robo por sorpresa

1. La gente cuya residencia es Maipú, es la que más ha sido víctima de robo por sorpresa en los últimos 12 meses (21.367 casos), seguida por quienes viven en Puente Alto y Las Condes.
2. A quienes menos les ha ocurrido este delito, es a quienes residen en Buin (423 casos), Paine y Melipilla.
3. Las personas que denuncia mayormente este delito, son quienes viven en la comuna de La Reina (100% de las veces), seguido por Vitacura y Paine.
4. Las personas que viven en Conchalí, El Bosque y Lo Barnechea no han denunciado ningún robo por sorpresa en el último año.

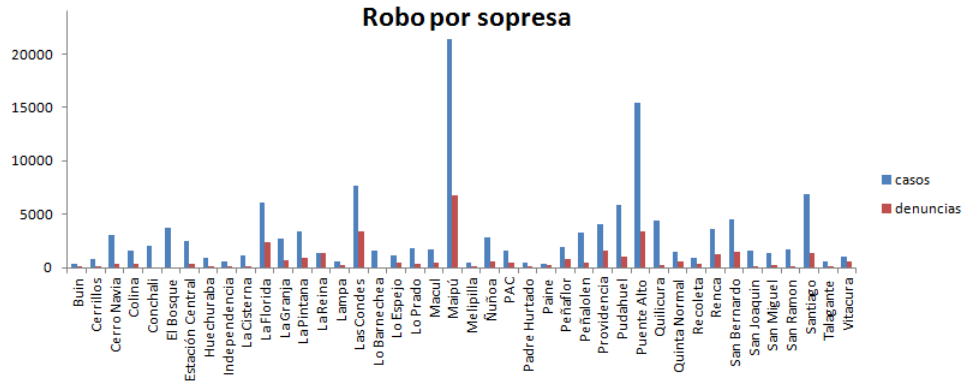


Figura 3.18: Gráfico Robo por sorpresa : casos v/s denuncias

Análisis de dependencia espacial

1. Tasa de ocurrencias

■ **Mapa de Distribución Espacial**

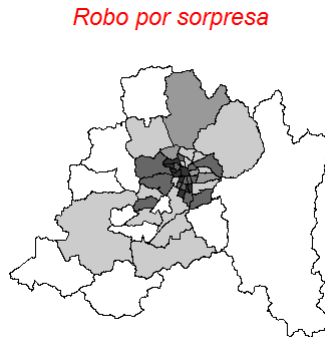


Figura 3.19: Mapa de distribución espacial de la tasa de ocurrencias de robo con sorpresa

Se logra observar una posible presencia de dependencia espacial, en el sector suroeste del mapa y en el centro de el con tasas de mayor valor (colores más oscuros). La distribución de este mapa es muy similar al que posee el robo con violencia e intimidación.

■ **Test de dependencia espacial**

Test	Indicador	p_value (Normal)	p_value (Randomization)
I Moran	0.2679	0.0019	0.0017
c de Geary	0.7318	0.0058	0.0061

Cuadro 3.10: Test de dependencia para Tasas de ocurrencias de Robo con Violencia

De acuerdo a los p_ valores obtenidos, bajo un nivel del 10 % de significancia, se rechaza en todos los casos la hipótesis nula y por lo tanto, se indica la existencia de autocorrelación espacial global.

Luego, localmente se obtiene:

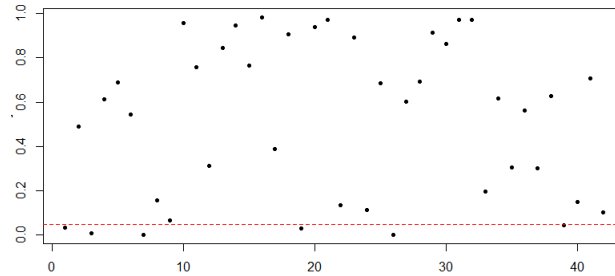


Figura 3.20: Test Local Moral para tasa de ocurrencia de robo con violencia

Existen 5 comunas en la zona de rechazo, donde las tasas de ocurrencias de robo por sorpresa son similares en ella con respecto a sus comunas vecinas. Estas comunas son:

Buin	Lo Prado
Cerro Navia	Paine
Estación Central	

Luego, para los residuos se observa el siguiente comportamiento:

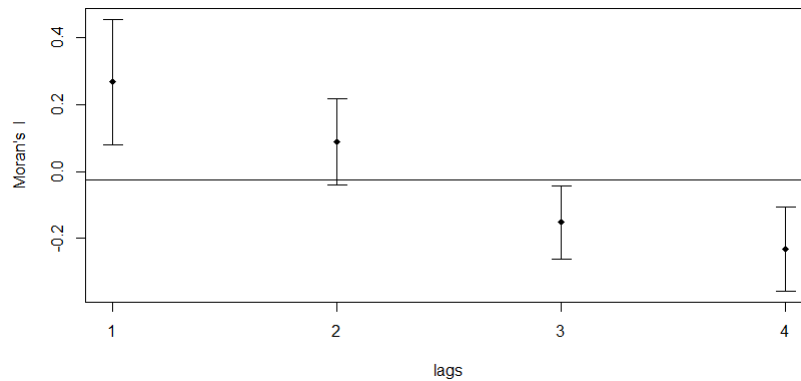


Figura 3.21: Gráfico de dependencia a k distancias para tasas de robo sorpresa

En la figura, a un lag=1 y lag=3 se ve una presencia de autocorrelación espacial positiva y negativa respectivamente.

■ **Aplicación de los modelos**

Modelo	N Variables	AIC	Significancia parámetro espacial
SAR	-	-	-
CAR	1	-203.5	Si

Cuadro 3.11: Comparación SAR Y CAR para tasa de denuncias por robo por sorpresa

Considerando que no se logro ajustar un modelo adecuado para el tipo SAR, el mejor modelo es el CAR. Aquí, el parámetro espacial resulta significativo tomando un valor =0.6543 . Los resultados son:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.0491	0.0066	7.47	0.0000
V4	0.0288	0.0128	2.26	0.0239

Así, con un nivel del 10% de significancia, tanto V4 como , son significativos para el modelo.

Luego, que un mayor aumento de esta variable implica una disminución de tasa de ocurrencia de robos por sorpresa.

■ **Errores no correlacionados espacialmente**

Gráficamente es posible verificar que luego de aplicar el modelo, los residuos no poseen dependencia espacial, pues se encuentran fuera de las zona de rechazo, validándose así este supuesto:

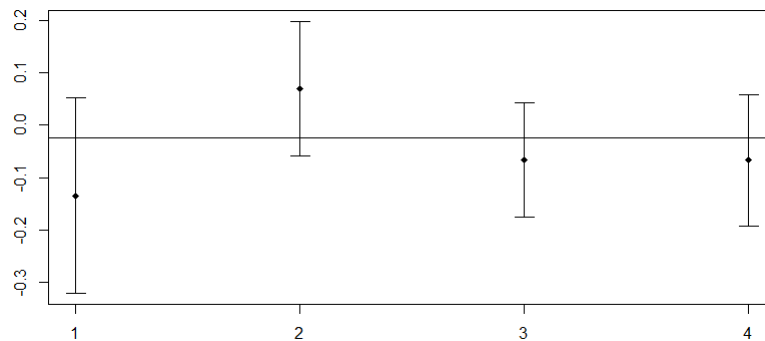


Figura 3.22: Gráfico de Dependencia espacial a k distancias para tasa de ocurrencia por robo por sorpresa (Modelo)

2. Tasa de denuncias

■ Mapa de Distribución Espacial

Robo por sorpresa



Figura 3.23: Mapa tasa denuncia robo sorpresa: Distribución Espacial

En el mapa no es posible confirmar a nivel general una clara dependencia espacial, sin embargo, algunos sectores hacia el centro y sector suroeste existen un par de sectores con una tonalidad común indicando una posible similar tasa de este fenómeno en ellas.

■ Test de dependencia espacial

Test	Indicador	p_value (Normal)	p_value (Randomization)
I Moran	-0.0167	0.9345	0.9301
c de Geary	0.92482	0.4387	0.4710

Cuadro 3.12: Test de dependencia espacial para tasas de denuncias de Robo con Violencia

Analizando los p_valores obtenidos, es posible confirmar para todos los casos, que con un nivel del 10 % de significancia, no se rechaza H_0 y por lo tanto, no existe dependencia espacial global.

Por otro lado, con el test local se obtiene:

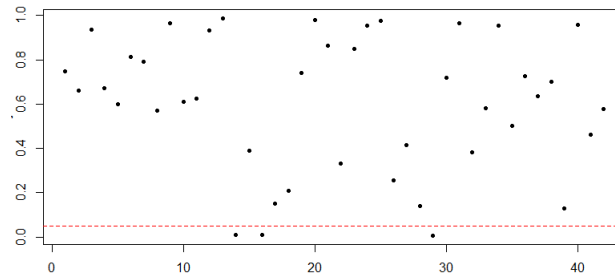


Figura 3.24: Test Local de Moran para tasa de denuncias de robo con violencia

Se distinguen 3 puntos en la zona de rechazo, los cuales representan a las comunas:

La Reina, Las Condes y Providencia, lo que indica pese a no existir dependencia global, estas 3 comunas poseen dependencia espacial de carácter local.

■ Aplicación de los modelos

Dado que no existe presencia de autocorrelación espacial global, el modelo que se deberá generar corresponde a un modelo de Regresión Lineal. El mejor modelo generado fue el siguiente:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.1558	0.0460	3.39	0.0016
V9	0.4692	0.1740	2.70	0.0102

Este modelo, posee un AIC=-24.83, y con un nivel del 10% de significancia la variable v9 es significativa y se encuentra en dirección positiva. Esto es que una alta presencia de esta variable implican una mayor tasa de denuncia de robos por sorpresa.

Mientras mayor es la tasa de perofesionales, mayor es la tasa de denuncias de este hecho delictivo.

■ Errores no correlacionados espacialmente

Luego, resulta necesario a verificar el cumplimiento de el supuesto de no correlación de los residuos. Se obtiene:

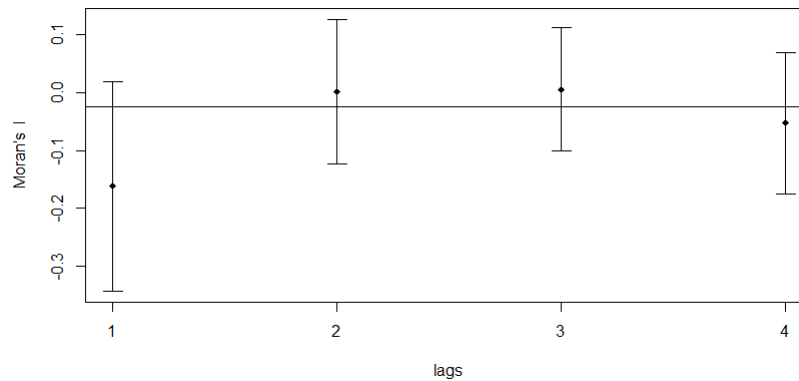


Figura 3.25: Gráfico de dependencia a k distancias para tasa de denuncias de robo por sorpresa (Modelo)

En efecto, el supuesto se cumple, pues el intervalo no se establece en ninguna de las zonas de rechazo.

3.4.2.4. Robo en vivienda

1. La comuna donde ocurre más este hecho delictivo es en Puente Alto (13.624 casos), seguida por Pudahuel y La Florida.
2. En las comunas donde se da el menor número de robo en vivienda es Independencia (135 casos), seguidas por Cerrillos y San Joaquín.
3. La mayor cantidad de denuncias de este hecho se da en la comuna de El Bosque (77.00 % de las veces), Las Condes y Providencia.
4. Las comunas con menor tasa de denuncias de este delito son Independencia, Lo Barnechea y Huechuraba.

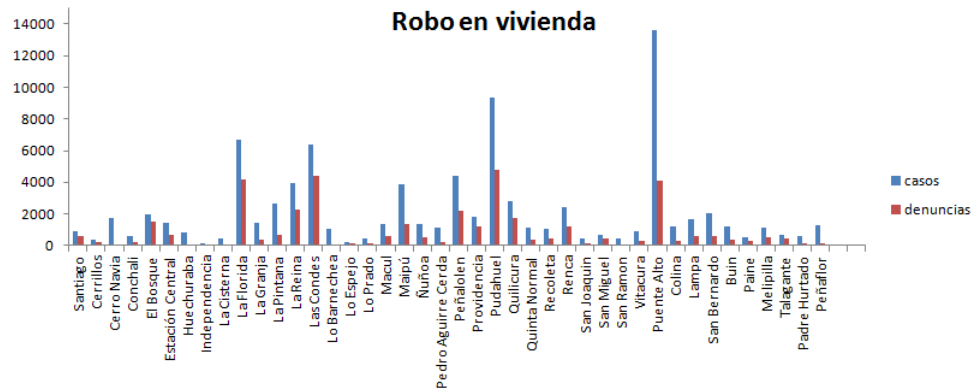


Figura 3.26: Gráfico Robo en vivienda: casos v/s denuncias

Análisis de dependencia espacial

1. Tasa de ocurrencias

■ **Mapa de Distribución Espacial**

Se logra observar una posible presencia de dependencia espacial, principalmente en el sector sur y hacia el este en el centro:

Robo en vivienda

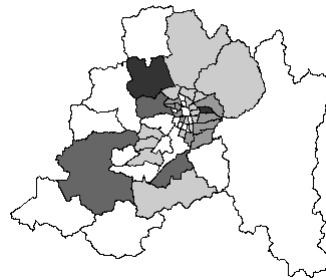


Figura 3.27: Mapa de distribución espacial de la tasa de ocurrencias de robo en vivienda

■ Test de dependencia espacial

Test	Indicador	p_value (Normal)	p_value (Randomization)
I Moran	0.1150	0.1379	0.1287
c de Geary	0.8402	0.09973	0.1096

Cuadro 3.13: Test de dependencia para Tasas de ocurrencias de robo en vivienda

De acuerdo a los p_ valores obtenidos, en el Teste de Geary, bajo un nivel del 10% de significancia, se rechaza en ambos casos la hipótesis nula y por lo tanto, se indica la existencia de autocorrelación espacial global bajo este indicador.

Luego, en terminos de Moran, localmente se obtiene:

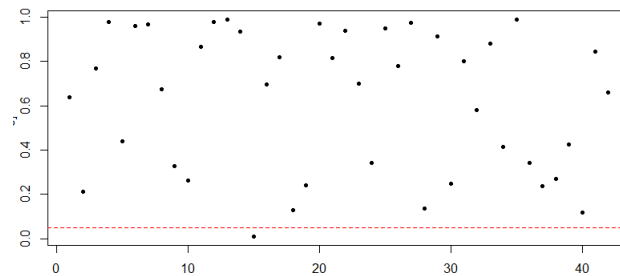


Figura 3.28: Test Local Moral para tasa de ocurrencia de robo con violencia

Existen 1 comuna en la zona de rechazo, donde las tasas de ocurrencias de robo en vivienda son similares en ella con respecto a sus comunas vecinas. Estaa es la comuna de : **Lampa**.

■ Aplicación de los modelos

Ambos modelos tienen 1 variable y un AIC muy similar, sin embargo, el menor es el del modelo CAR, pero considerando que el parámetro espacial no es significativo para estar

Modelo	N Variables	AIC	Significancia parámetro espacial
SAR	1	-186.32	No
CAR	1	-186.34	No

Cuadro 3.14: Comparación SAR Y CAR para tasa de denuncias por robo en vivienda

en el modelo, se realiza una regresión lineal simple con la variable que este posee. . Los resultados son:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.0352	0.0052	6.71	0.0000
V4	0.0352	0.0130	2.71	0.0100

El modelo posee un AIC=-187.532 (mejora).Luego, con un nivel del 10 % de significancia, V4 es una variable significativa para el modelo.

Luego, que un mayor aumento de esta variable implica una disminución de tasa de ocurrencia de robos en vivienda.

■ Errores no correlacionados espacialmente

Gráficamente es posible verificar que luego de aplicar el modelo, los residuos no poseen dependencia espacial, pues se encuentran fuera de las zona de rechazo, validándose así este supuesto.

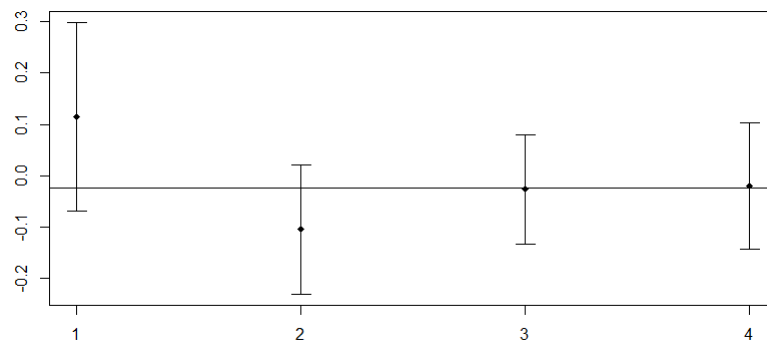


Figura 3.29: Gráfico de Dependencia espacial a k distancias para tasa de ocurrencia por robo en vivienda (Modelo)

2. Tasa de denuncias

■ Mapa de Distribución Espacial

Robo en vivienda

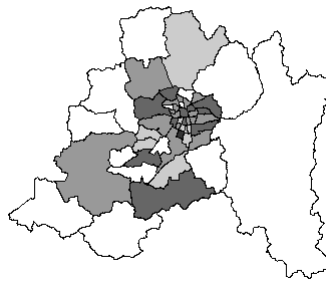


Figura 3.30: Mapa tasa denuncia robo vivienda: Distribución Espacial

Bajo la distribución del mapa, no es posible notar que exista una dependencia espacial para la tasa de denuncias de robos en viviendas.

■ Test de dependencia espacial

Test	Indicador	p_value (Normal)	p_value (Randomization)
I Moran	-0.0333	0.9246	0.9250
c de Geary	0.9887	0.9074	0.9066

Cuadro 3.15: Test de dependencia espacial para tasas de denuncias de robo en vivienda

Analizando los p_valores obtenidos, es posible confirmar para todos los casos, que con un nivel del 10% de significancia, no se rechaza H_0 y por lo tanto, no existe dependencia espacial global.

Por otro lado, con el test local se obtiene:

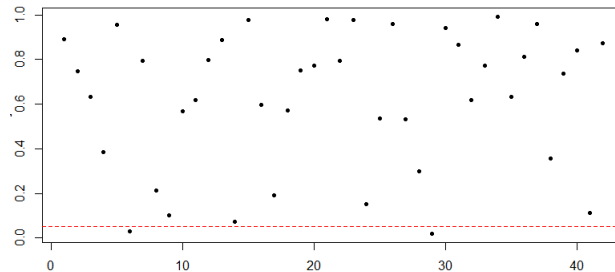


Figura 3.31: Test Local de Moran para tasa de denuncias de robo en vivienda

Se distinguen 2 puntos en la zona de rechazo, los cuales representan a las comunas: **Colina y Ñuñoa**, lo que indica pese a no existir dependencia global, estas comunas poseen dependencia espacial de carácter local.

■ Aplicación de los modelos

Dado que no existe presencia de autocorrelación espacial global, el modelo que se genera un modelo de Regresión Lineal. El mejor modelo generado fue el siguiente:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.5873	0.0945	6.22	0.0000
V10	-0.3092	0.1407	-2.20	0.0338

Este modelo, posee un AIC=-15.52, y con un nivel del 10% de significancia la variable V10 es significativa y se encuentra relacionada de modo inverso. Esto es, que una alta presencia de V10 (mayor tasa de personas con nivel educacional máximo enseñanza media) se tiene una menor tasa de denuncia de robos en vivienda y viceversa.

■ Errores no correlacionados espacialmente

Luego, resulta necesario a verificar el cumplimiento de el supuesto de no correlación de los residuos. Se tiene el siguiente gráfico:

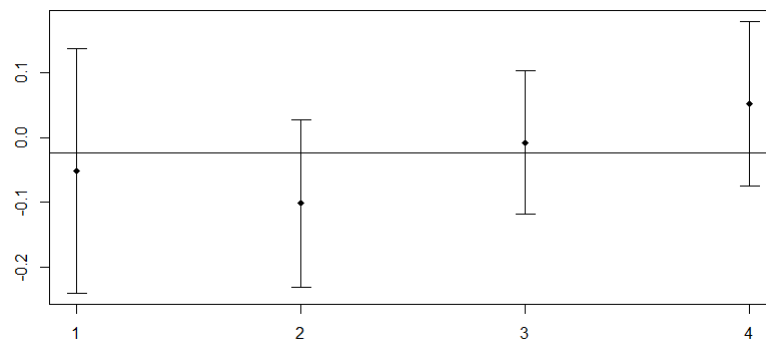


Figura 3.32: Gráfico de dependencia a k distancias para tasa de denuncias de robo en vivienda (Modelo)

En efecto, el supuesto se cumple, pues el intervalo a lag=1, no se establece en ninguna de las dos zonas de rechazo.

3.4.3. Análisis por clasificación de delitos

De acuerdo a los tipos de delitos definidos anteriormente, es posible clasificarlos como delitos **violentos o no violentos**. Así, los hechos delictuosos analizados fueron agrupados de la siguiente manera:

1. Delitos Violentos:

- Robo con violencia e intimidación
- Lesiones
- Robo en vivienda

2. Delitos no Violentos:

- Robo por sorpresa
- Hurto

Para clasificar cada delito, se tomó en cuenta la definición y contexto de las preguntas estipulado en la encuesta.

Luego, es posible realizar un análisis general de comportamiento para cada uno de los dos tipos de delito, obteniendo los siguientes dos gráficos y tabla resumen:

Tipo de delito	Casos Totales	Mín. casos	Máx. casos	Deuncias	Tasa de Denuncias
Violentos	288.732	1.005	45.672	131.866	46.00 %
No Violentos	348.239	856	50.672	75.273	22.00 %

Cuadro 3.16: Tabla resumen delitos Violentos y No violentos

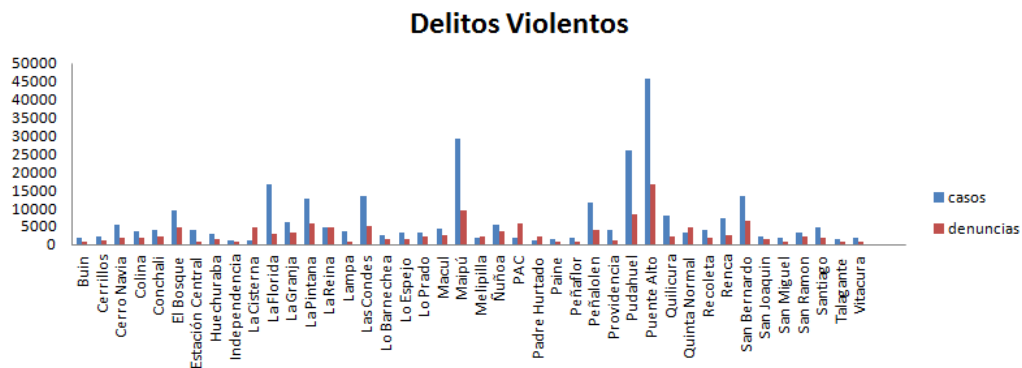


Figura 3.33: Gráfico delitos Violentos

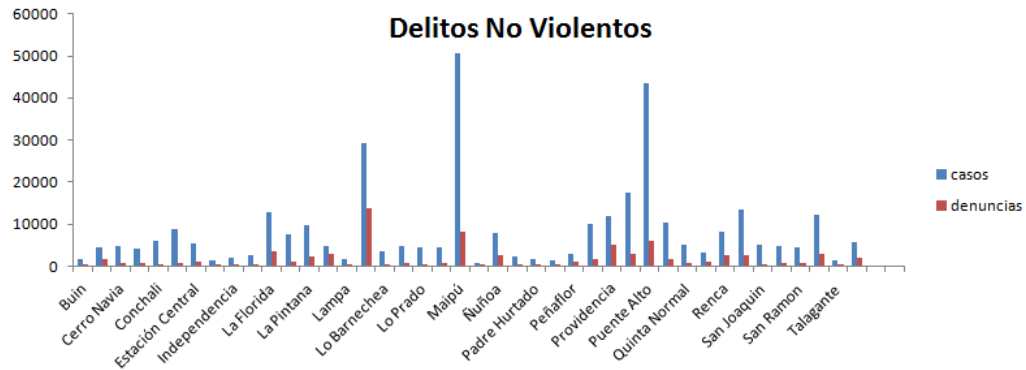


Figura 3.34: Gráfico delitos no violentos

Observaciones:

1. A nivel general, ocurren mayormente delitos no violentos que delitos violentos.
2. El menor número de casos de delitos violentos le ha ocurrido a las personas de la comuna de La Cisterna (1.005 hechos).
3. El menor número de casos de delitos no violentos se da en quienes viven en la comuna de Melipilla (856 hechos).
4. La mayor cantidad de delitos violentos ocurridos, le ocurrieron a personas que viven en la comuna de Puente Alto (45.672 casos).
5. La mayor cantidad de delitos no violentos, le ocurrieron a gente que reside en la comuna de Maipú. (45.672 casos).
6. Los delitos violentos son más denunciados que los no violentos.

3.4.3.1. Análisis de Delitos Violentos

3.4.3.1.1. Tasa de Denuncias

1. Mapa de distribución Espacial

En la siguiente figura se puede apreciar que existe a primera vista una dependencia espacial ,en varios pequeños sectores, principalmente en la parte alta del mapa desde el Oeste, pues se mantiene mismos tonos indicando una tasa similar de esta variable respuesta.

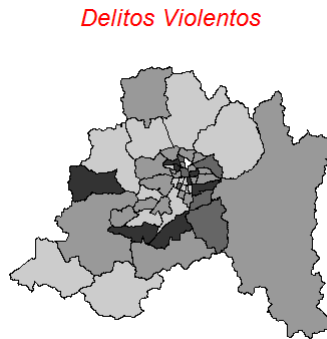


Figura 3.35: Mapa distribución espacial delitos violentos

2. Test de dependencia espacial

Test	Indicador	p_value (Normal)	p_value (Randomization)
I Moran	0.0918	0.2164	0.2160
c de Geary	0.8767	0.2040	0.2044

Cuadro 3.17: Test de dependencia para Tasas de denuncias por delitos violentos

Con un nivel del 10% de significancia, no existe evidencia para rechazar la hipótesis nula de ausencia de autocorrelación espacial, por lo tanto, no existe dependencia de este tipom globalmente.

En términos de dependencia local, en la figura 3.32 es posible ver que existen 7 comunas ubicadas en el área de rechazo, indicando la existencia de dependencia local, estas comunas son:

La Reina	Santiago
Las Condes	Talagante
Peñaflor	Vitacura
Providencia	

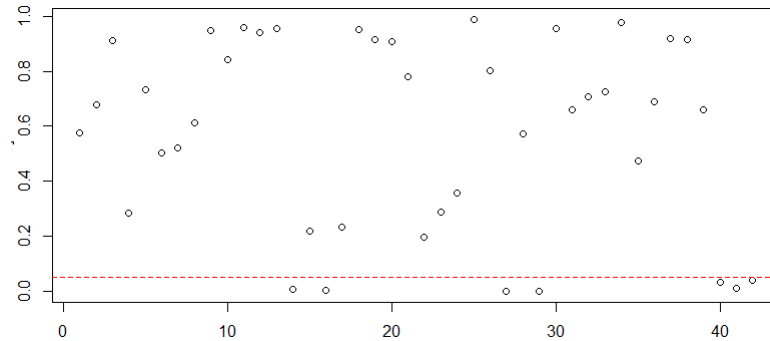


Figura 3.36: Test local Moran para tasa de denuncias de delitos violentos

3. Aplicación de los Modelos

Dado que no existe autocorrelación espacial global, se realiza la selección del mejor modelo de regresión lineal. El mejor modelo fue:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.0275	0.0906	-0.30	0.7631
V1	0.9385	0.2300	4.08	0.0002
V9	0.3204	0.1291	2.48	0.0178
V4	-0.2003	0.1079	-1.86	0.0713
V5	-1.6204	0.7565	-2.14	0.0389

El AIC del modelo es de -62.5111, y se ve que las variables V1 y V9 influyen directamente en la tasa de denuncias de delitos violentos, esto es, que mientras mayor es la tasa de estas variables explicativas también mayor es la tasa de denuncias emitidas o viceversa. Por el contrario, para V4 y V5 sucede al revés, pues afectan de forma inversa, es decir, una alta precencia de estas hace que la tasa de denuncias se vuelva menor.

4. Errores no correlacionados espacialmente

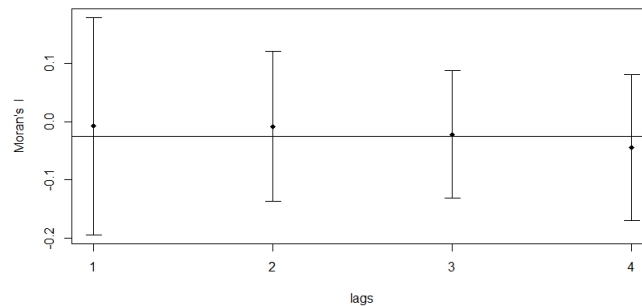


Figura 3.37: Gráfico de dependencias a k distancias para tasas de denuncias por delitos violentos (Modelo)

Como es posible apreciar, los residuos cumplen el supuesto pues los intervalos a ningún lag se posicionan en alguna de las zonas de rechazo.

3.4.3.1.2. Tasa de Ocurrencias

1. Mapa de distribución Espacial

En la siguiente figura se puede apreciar que existe a primera vista una dependencia espacial, principalmente en la parte baja de Oeste a Este, pues se mantienen mismos tonos indicando una tasa similar de esta variable respuesta.



Figura 3.38: Mapa tasa ocurrencia delitos violentos: Distribución Espacial

2. Test de dependencia espacial

Test	Indicador	p_value (Normal)	p_value (Randomization)
I Moran	0.0728	0.3009	0.3035
c de Geary	0.9292	0.4660	0.4628

Cuadro 3.18: Test de dependencia para Tasas de ocurrencias por delitos violentos

Con un nivel del 10% de significancia, no existe evidencia para rechazar la hipótesis nula de ausencia de autocorrelación espacial, por lo tanto, no existe dependencia espacial global.

En términos de dependencia local, es posible ver que existen 4 comunas ubicadas en el área de rechazo, indicando la existencia de dependencia local, estas comunas son:

Pedro Aguirre Cerda	Providencia
San Miguel	Santiago

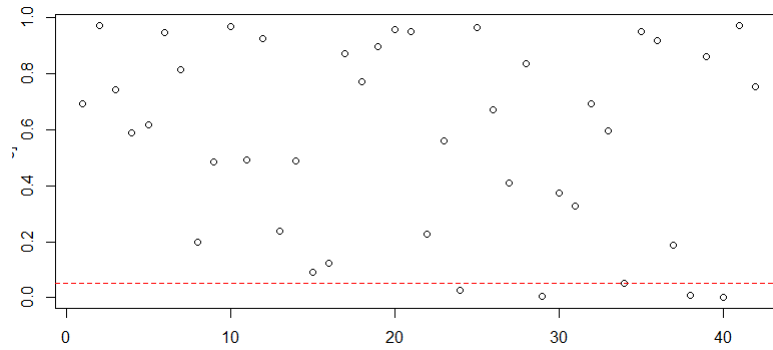


Figura 3.39: Test local Moran para tasa de ocurrencias de delitos violentos

3. Aplicación de los Modelos

Dado que no existe autocorrelación espacial global, se realiza la selección del mejor modelo de regresión lineal. El mejor modelo fue:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.1913	0.0336	5.69	0.0000
V2	0.2119	0.0663	3.20	0.0028
V3	0.4030	0.0911	4.43	0.0001
V4	0.1746	0.0485	3.60	0.0009

El AIC del modelo es de -117.3791, y se ve que las 3 variables ingresadas al modelo influyen directamente en la tasa de ocurrencias de delitos violentos, esto es, que mientras mayor es la tasa de estas variables explicativas también mayor es la tasa de denuncias emitidas o viceversa.

4. Errores no correlacionados espacialmente

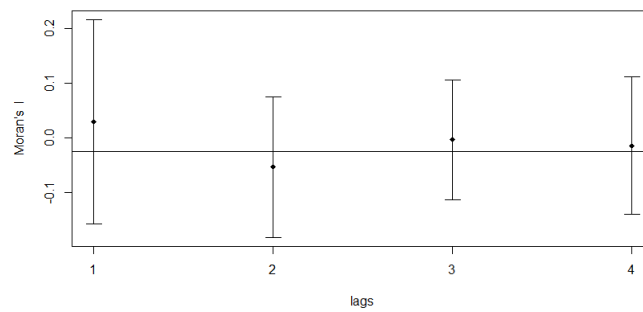


Figura 3.40: Gráfico de dependencias a k distancias para tasas de ocurrencias por delitos violentos (Modelo)

Como es posible apreciar, los residuos cumplen el supuesto estipulado, pues los intervalos a ningún lag se posicionan en alguna de las zonas de rechazo.

3.4.3.2. Análisis Delitos no Violentos

3.4.3.2.1. Tasa de denuncias

1. Mapa de distribución Espacial

Delitos No Violentos

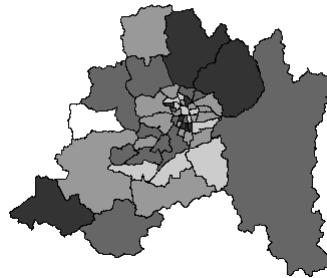


Figura 3.41: Mapa distribución espacial delitos no violentos

Como muestra la figura, no se ve una fuerte dependencia espacial, aunque hay ocasiones donde hay dos comunas vecinas con igual tonalidad, como es en la parte Norte del mapa y sector Oeste.

2. Test de dependencia espacial

Test	Indicador	p_value (Normal)	p_value (Randomization)
I Moran	0.05098	0.4225	0.0.4226
c de Geary	0.91465	0.3794	0.3795

Cuadro 3.19: Test de dependencia para Tasas de denuncias por delitos no violentos

Luego, con un nivel del 10% de significancia, no existe evidencia para rechazar la hipótesis nula de ausencia de autocorrelación espacial, por lo tanto, no existe dependencia a nivel global.

En términos de dependencia local, se obtiene el siguiente gráfico:

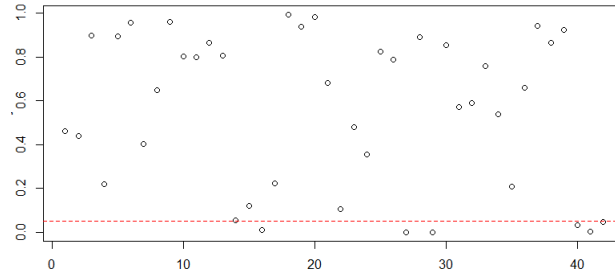


Figura 3.42: Test local Moran para tasa de denuncias de delitos no violentos

Es posible ver que existen 5 comunas ubicadas en el área de rechazo, indicando la existencia de dependencia local. Estas son:

Peñaflor	Talagante
Providencia	Vitacura
Santiago	

Estas mismas 5 comunas, también son parte de las comunas que poseen esta característica para delitos violentos. Esto, indicaría que ambos tipos de delitos o mas bien, en general, estas comunas tienen con sus vecinos un comportamiento similar en las tasas de denuncias.

3. Aplicación de los Modelos

Dado que no existe autocorrelación espacial global, se realiza la selección del mejor modelo de regresión lineal. El mejor modelo fue:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.1775	0.1156	1.54	0.1330
V1	1.1737	0.2340	5.02	0.0000
V4	-0.2804	0.1140	-2.46	0.0187
V5	1.8537	0.8139	2.28	0.0286
V9	-0.2148	0.1319	-1.63	0.1118

El AIC del modelo es de -56.4016, indicando que el modelo para delitos violentos tiene un mejor ajuste.

Las variables V1 y V5 influyen directamente en la tasa de denuncias de delitos no violentos, esto es, que mientras mayor es la tasa de estas variables explicativas también mayor es la tasa de denuncias emitidas. Por el contrario, para V4 y V9 sucede al revés, pues afecta de forma inversa.

Las variables para la tasa de denuncia de delitos tanto violentos como no, son las mismas. Sin embargo, cambia la dirección en que afectan a la variable de la respuesta.

4. Errores no correlacionados espacialmente

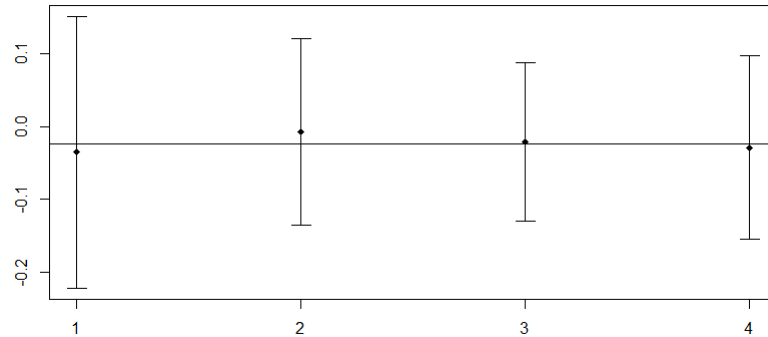


Figura 3.43: Gráfico de dependencias a k distancias para tasas de denuncias por delitos no violentos (Modelo)

Como se logra observar, los residuos cumplen el supuesto pues los intervalos a ningún lag se posicionan en alguna de las zonas de rechazo.

3.4.3.2.2. Tasas de ocurrencias

1. Mapa de distribución Espacial



Figura 3.44: Mapa distribución espacial delitos no violentos

Como muestra la figura, no se ve una fuerte dependencia espacial, pues las toniladidas se hallan muy dispersas, aunque hay ocasiones donde hay dos comunas vecinas con igual tonalidad, como es en la parte Norte y Suroeste del mapa.

2. Test de dependencia espacial

Test	Indicador	p_value (Normal)	p_value (Randomization)
I Moran	0.1029	0.1756	0.1781
c de Geary	0.8971	0.2893	0.2857

Cuadro 3.20: Test de dependencia para Tasas de denuncias por delitos no violentos

Luego, con un nivel del 10% de significancia, no existe evidencia para rechazar la hipótesis nula de ausencia de autocorrelación espacial, por lo tanto, no existe dependencia a nivel global.

En términos de dependencia local, se obtiene el siguiente gráfico:

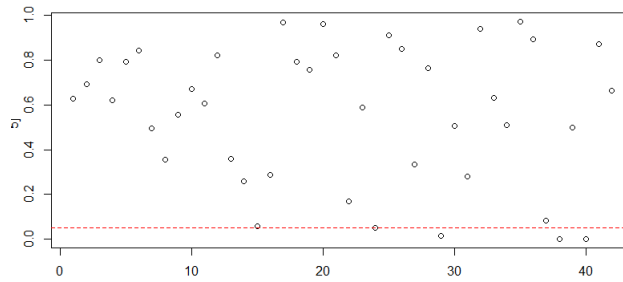


Figura 3.45: Test local Moran para tasa de denuncias de delitos no violentos

Es posible ver que existen 3 comunas ubicadas en el área de rechazo, indicando la existencia de dependencia local en ellas. Estas son:

Providencia	San Miguel
Santiago	

3. Aplicación de los Modelos

Dado que no existe autocorrelación espacial global, se realiza la selección del mejor modelo de regresión lineal. El mejor modelo fue:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.0126	0.0390	0.32	0.7478
V1	0.2133	0.0605	3.53	0.0011
V3	0.3957	0.0848	4.67	0.0000
V4	0.1700	0.0476	3.57	0.0010

El AIC del modelo es de -119.9851, indicando que el modelo para delitos no violentos tiene un mejor ajuste.

Las variables del mdelo influyen directamente en la tasa de ocurrencias de delitos no violentos, esto es, que mientras mayor es la tasa de estas variables explicativas también mayor es la tasa de ocurrencias .

Las variables para la tasa de pcurrencia de delitos tanto violentos como no, al igual que en el caso de las denuncias, son las mismas. Sin embargo, cambia la dirección en que afectan a la variable de la respuesta.

4. Errores no correlacionados espacialmente

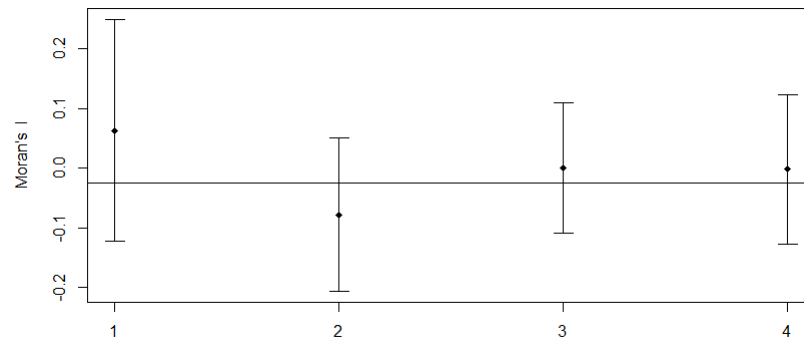


Figura 3.46: Gráfico de dependencias a k distancias para tasas de ocurrencias por delitos no violentos (Modelo)

Como se logra observar, los residuos cumplen el supuesto pues los intervalos a ningún lag se posicionan en alguna de las zonas de rechazo.

Capítulo 4

Conclusión

Entre las distintas situaciones reflejadas, resulta importante destacar que aquellos delitos de menor dificultad o más bien, de menor pérdida económica son los menos denunciados. Por el contrario, cuando las pérdidas son mayores, la denuncia es más común a ocurrir, lo cual es bastante esperado, ya que por lo general estos bienes poseen seguros y reportar el delito tendrá a corto o largo plazo una respuesta segura. Este comportamiento, afecta en la ocurrencia de ciertos delitos, por ejemplo, el hurto es el delito que registra mayor frecuencia, lo que se podría ver afectado por la no denuncia de este, haciendo que el delincuente no dimensione este acto delictivo, ya que como se mencionó, el hurto no genera daño a la persona y la pérdida no tiende a ser de gran valor económico. Así mismo, en el último estudio dividido entre delitos violentos y no violentos, se logra identificar que la proporción de los delitos no violentos es mayor frente a los violentos, lo cual confirma lo mencionado anteriormente a que en gran parte el valor de la pérdida afecta mucho en cómo llevar la situación, ya que pese a ser los delitos no violentos los de mayor frecuencia, no son los más denunciados.

Entre los factores que más se repiten tanto en la ocurrencia como denuncia de hechos delictivos, se encuentran la percepción de las personas respecto a la efectividad y confianza en la justicia, indicando que afecta principalmente en aquellas personas que viven en comunas que podrían ser consideradas de nivel socioeconómico menor e indicando la gran importancia que tiene esto a la hora de tomar decisiones frente a hechos delictuosos. De aquí, resultaría interesante e importante poder indagar de qué manera podría mejorarse esta percepción ya que esto traería en efecto que las personas tengan una mayor disposición a reportar delitos.

Como se mencionó en el primer capítulo, un estudio de S. Avdija et al (2012), en general, indica que si se tiene un mejor nivel socioeconómico se tiende a tener una mejor disposición a denunciar diferentes delitos, por lo que se dio paso a crear dos variables socioeconómicas (V9: tasa de profesionales y V10: tasa de personas cuyo mayor nivel alcanzado es la enseñanza media), las cuales se repitieron en los modelos finales en más de una ocasión, y de acuerdo al modo en que afectaban (positiva o negativamente) verifican lo estipulado en aquel estudio, como es en el caso del robo en vivienda y robo con violencia e intimidación, pues personas cuyo mayor nivel educacional

no es enseñanza media (lo que podría indicar a comunas de mayor nivel socioeconómico), tienden a reportar más lo ocurrido.

En algunas ocasiones, pese a no darse una fuerte autocorrelación espacial global o en situaciones en las que solo se dio dependencia espacial manera local, de igual modo se puede continuar identificando el comportamiento anterior. Si bien, la comuna en esta ocasión por la modalidad de trabajo de la encuesta no indicaba la comuna de ocurrencia del hecho, si indicaba el lugar de vivienda de estas personas, aportando de igual modo a poder encontrar factores y comportamientos comunes entre gente que reside en los mismos sectores, tal como se da en el caso de ocurrencias de hurto, denuncias de robo con violencia e intimidación y tasa de denuncias del robo por sorpresa. En este último caso, por ejemplo, no se presenta dependencia global, sin embargo, ciertas comunas apuntan a un comportamiento similar con sus comunas vecinas y de acuerdo al modelo final, se podría confirmar que en comunas de mayor nivel socioeconómico se da un mayor reporte de este tipo de delito, ya que como es en por ejemplo en el caso de denuncias de robo por sorpresa, la variable explicativa del modelo corresponde a la variable tasa de profesionales, indicando que mientras mayor es esta tasa, mayor es también la tasa de denuncias.

El hecho de no actuar frente a todos los delitos, dejando pasar por alto más de uno, puede causar a un victimario la confianza de poder seguir comitiendo hecho delictivos al saber que su probabilidad de denuncia es baja, por lo que a partir de esto, sería de gran ayuda poder tomar aquellos factores claves que puedan equilibrar y dar interés de evitar cualquier delito por mínimo que sea, ya que todos son actos indebidos.

El presente estudio, logró identificar factores que se dan e intervienen en la ocurrencia y denuncia de hechos delictuosos considerando su lugar de residencia. Sin embargo, sería de gran utilidad e interés poder realizar paralelamente un estudio considerando aquellas las comuna en donde ocurren los hechos, para poder tener en cuenta no solo, a quienes les ocurre más o quienes denuncian más de acuerdo a características y percepciones, si no también, que comunas necesitan de un mayor cuidado. De esta manera, se podría conocer mayormente el verdadero comportamiento e influencia que está teniendo la delincuencia en la actualidad.

Otra forma de complementar este trabajo podría ser considerar otros tipos de modelos de carácter espacial o bien, considerando estos mismos ya utilizados ampliando o reduciendo las áreas, ya que siempre hubo al menos dependencia local, por lo que tal vez, por ejemplo, por comunas podrían encontrarse los sectores con mayor tendencia a ocurrencia de este tipo de actos.

Capítulo 5

Referencias

- [1] R. Retting (2018): Seguridad pública: análisis y alcances de la Enusc 2017
<https://www.biobiochile.cl/noticias/nacional/chile/2018/05/14/indice-de-victimizacion-llega-al-28-a-nivel-nacional-y-alcanza-punto-mas-alto-en-4-anos.shtml>
- [2] Diccionario Jurídico Chileno (2001).
En línea: <http://www.juicios.cl/dic300/DELITO.htm>
- [3] D. Quinteros (2014): Delitos del espacio público y el problema de la “cifra negra”: una aproximación a la no-denuncia en Chile, Vol. 9, No. 18,2014, pp. 691-712.
- [4] Instituto Nacional de Estadísticas (2018): XIV Encuesta nacional Urbana de Seguridad Ciudadana
<http://www.ine.cl/estadisticas/sociales/encuesta-de-seguridad-ciudadana>
- [5] Índice Paz Ciudadana- GfK Adimark (2016): Resultados del Estudio año 2016 Versión conferencia
<http://www.pazciudadana.cl/wp-content/uploads/2017/01/ipc-2016-version-conferencia-120117.pdf>
- [6] El Heraldo S.A. (2013): La importancia de denunciar los delitos, Colombia.
- [7] Avdija, S. et al (2012): The impact of Prior Victimization and Socio-Economic Status on Peoples Crime-Reporting Behavior, USA.
- [8] R. Borda (2011): Comparación de las metodologías: Modelo Lineal Generalizado mixto marginal espacial con varianza CAR bajo respuesta Poisson y Modelo Lineal Generalizado Poisson Log-lineal con distribución subyacente gaussiana en el estudio de datos de área, Universidad Nacional de Colombia, Medellín-Antoquia.
- [9] R. Bivand (2007): Analysing Spatial Data in R Worked example: spatial autocorrelation Department of Economics Norwegian School of Economics and Business Administration Bergen, Norway.
- [10] R. Girardo (2011): Estadística Espacial, Colombia.
- [11] A. Waller, A. Gotway (2004): Applied Spatial Statistics for Public Health Data, New Jersey, Canada.

- [12] R. Bivand (2018): The Problem of Spatial Autocorrelation: forty years on.
<https://cran.r-project.org/web/packages/spdep/vignettes/CO69.pdf>
- [13] M. Montero, G. Fernandez, J. Mateu (2015): Spatial and spatio-temporal Geostatistical Modeling and Kriging, Chennai, India.
- [14] M. Herrera (2015): *Econometría Espacial usando STATA: Guía Teórico - Aplicada*, Chile.
- [15] A. Waller, A. Gotway (2004): *Applied Spatial Statistics for Public Health Data*, Canada.
- [16] C. Geaten, X. Guyon (2010) : *Springer Series in Statistics*, London.
- [17] R. Bivand, E. Pebesma (2008): *Applied Spatial Data Analysis with R,USA*.
- [18] c. Martinez (2018): *Análisis de datos espaciales de devoluciones*, Coruña.
- [19] Derecho Chile (2012): <http://www.derecho-chile.cl/robo-con-violencia-o-intimidacion-en-las-personas-i-parte/>
- [20] A. Edgar (2011): *Análisis de Regresión*, Puerto Rico.

Capítulo 6

Anexos

6.1. Matriz de Comunas

```
attr(“rm.nb, com.id”) = c(“CERRILLOS”, “CERRO NAVIA”, “COLINA”, “CONCHALI”, “.EL BOS-  
QUE”, “.ESTACION CENTRAL”, “HUECHURABA”, “INDEPENDENCIA”, “LA CISTERNA”, “LA  
FLORIDA”, “LA GRANJA”, “LA PINTANA”, “LA REINA”, “LAMPA”, “LAS CONDES”, “LO  
BARNECHEA”, “LO ESPEJO”, “LO PRADO”, “MACUL”, “MAIPU”, “MELIPILLA”, “NUNOA”, “PAC”,  
“PADRE HURTADO”, “PAINE”, “PENAFLOL”, “PENALOLEN”, “PROVIDENCIA”, “PUDAHUEL”, “PUENTE  
ALTO”, “QUILICURA”, “QUINTA NORMAL”, “RECOLETA”, “RENCA”, “SAN BERNARDO”, “SAN  
JOAQUIN”, “SAN MIGUEL”, “SAN RAMON”, “SANTIAGO”, “TALAGANTE”, “VITACURA”)
```

6.2. Funciones Espaciales

- **Test de Moran:**

- **Función:**

```
moran.test(x, listw, randomisation=TRUE, alternative=“greater”)
```

Donde:

- **x:** vector numérico que contiene en este caso, los valores de las tasas de ocurrencias.
- **listw:** matriz creada de vecindades.
- **randomisation:** FALSE o TRUE, según si se quiere realizar el cálculo e hipótesis bajo normalidad o aleatoriedad respectivamente.
- **alternative:** dependerá de la hipótesis alternativa, en esta ocasión se quiere evaluar directamente la existencia o no existencia de AE, sin importar su dirección. De aquí, se da uso a la opción “two.side” (distinto, ambos lados).

- **Test de Geary**

```
geary.test(x, listw, randomisation=TRUE, alternative="greater")
```

Aquí, los argumentos se definen de la misma forma que en el test de *I* Moran.

6.3. Modelos opcionales para buscar el mejor ajuste:

- Hurto

1. SAR:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.7514	0.1893	3.97	0.0001
p5	-0.1928	0.4926	-0.39	0.6956
pc5	-0.0092	0.1320	-0.07	0.9446
ppoli1	-2.5286	0.8875	-2.85	0.0044
ps1	-1.6971	1.0461	-1.62	0.1047
pcd1	-0.0668	0.5321	-0.13	0.9001
pcdc1	-0.4619	0.3883	-1.19	0.2343
pnop5	-0.1421	0.1304	-1.09	0.2758

2. CAR:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.7414	0.1895	3.91	0.0001
p5	-0.0962	0.4927	-0.20	0.8452
pc5	-0.0089	0.1319	-0.07	0.9463
ppoli1	-2.4378	0.8880	-2.75	0.0060
ps1	-1.8430	1.0447	-1.76	0.0777
pcd1	-0.2022	0.5319	-0.38	0.7038
pcdc1	-0.3741	0.3880	-0.96	0.3350
pnop5	-0.1186	0.1305	-0.91	0.3635

- Robo por sorpresa

1. SAR:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.0635	0.0264	2.41	0.0161
pd4	-0.0115	0.0149	-0.77	0.4424
pc4	0.0137	0.0149	0.92	0.3565
ppoli1	0.0399	0.1178	0.34	0.7351
ps1	0.1393	0.1492	0.93	0.3506
pcd1	-0.0001	0.0660	-0.00	0.9986
pcdc1	-0.0367	0.0468	-0.78	0.4332
pnop4	0.0292	0.0120	2.44	0.0148

2. CAR:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.7414	0.1895	3.91	0.0001
p5	-0.0962	0.4927	-0.20	0.8452
pc5	-0.0089	0.1319	-0.07	0.9463
ppoli1	-2.4378	0.8880	-2.75	0.0060
ps1	-1.8430	1.0447	-1.76	0.0777
pcd1	-0.2022	0.5319	-0.38	0.7038
pcdc1	-0.3741	0.3880	-0.96	0.3350
pnop5	-0.1186	0.1305	-0.91	0.3635

■ **Robo con Violencia e Intimidación**

1. **SAR:**

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.0347	0.0472	0.74	0.4621
pd1	-0.0163	0.0248	-0.66	0.5114
pc1	-0.0210	0.0287	-0.73	0.4639
ppoli1	-0.0711	0.2082	-0.34	0.7325
ps1	-0.0848	0.2660	-0.32	0.7499
pcd1	0.0859	0.1170	0.73	0.4629
pcdc1	0.0311	0.0906	0.34	0.7312
pnop1	0.0094	0.0237	0.40	0.6924

2. **CAR:**

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.0400	0.0495	0.81	0.4189
pd1	-0.0209	0.0266	-0.79	0.4316
pc1	-0.0033	0.0311	-0.11	0.9148
ppoli1	-0.0401	0.2242	-0.18	0.8580
ps1	-0.3404	0.2845	-1.20	0.2314
pcd1	0.0288	0.1237	0.23	0.8156
pcdc1	0.0630	0.0919	0.69	0.4931
pnop1	0.0090	0.0248	0.36	0.7157

- Robo con violencia e intimidación:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.3178	0.0565	5.63	0.0000
ppoli1	-2.6634	0.8410	-3.17	0.0015

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.2950	0.0589	5.00	0.0000
ppoli1	-2.6696	0.8338	-3.20	0.0014

- Robo por sorpresa:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.0499	0.0076	6.54	0.0000
pnop4	0.0282	0.0120	2.35	0.0190

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.0694	0.0200	3.47	0.0005
pc4	0.0294	0.0146	2.02	0.0439
pcd1	-0.0742	0.0385	-1.93	0.0536
pnop4	0.0304	0.0123	2.48	0.0132

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.0919	0.0126	7.32	0.0000
pf1	-0.0904	0.0373	-2.42	0.0153

- Robo en vivienda:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.0370	0.0064	5.81	0.0000
pnop3	0.0308	0.0117	2.62	0.0087