



**UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERIA INDUSTRIAL**

**ANÁLISIS DEL FENÓMENO DELICTUAL UTILIZANDO UN MODELO DE
REGRESIÓN LOGÍSTICA EN BASE A ATRIBUTOS**

**TESIS PARA OPTAR AL GRADO DE MAGISTER EN GESTION DE
OPERACIONES**

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

MARÍA PAZ JOFRÉ ALEGRÍA

SANTIAGO, CHILE
2011



**UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERIA INDUSTRIAL**

**ANÁLISIS DEL FENÓMENO DELICTUAL UTILIZANDO UN MODELO DE
REGRESIÓN LOGÍSTICA EN BASE A ATRIBUTOS**

**TESIS PARA OPTAR AL GRADO DE MAGISTER EN GESTION DE
OPERACIONES**

MEMORIA PARA OPTAR AL TITULO DE INGENIERO CIVIL INDUSTRIAL

MARÍA PAZ JOFRÉ ALEGRÍA

**PROFESOR GUÍA:
RICHARD WEBER HAAS**

**MIEMBROS DE LA COMISIÓN:
GASTÓN L'HUILLIER CHAPARRO
FRANCISCO MARTÍNEZ CONCHA
JAVIERA BLANCO SUÁREZ**

SANTIAGO DE CHILE

NOVIEMBRE, 2011

Este trabajo contó con el apoyo del proyecto Proyecto Anillo Act-87, CONICYT.

*A mis abuelas
Betsy y Mona (QEPD)*

Agradecimientos

Primero que todo quiero agradecer a mi familia por el amor incondicional que me ha entregado a lo largo de mi vida. Gracias a mis padres Keka y Nano por la gran oportunidad que me dieron de estudiar y por apoyarme en todos mis proyectos. Gracias a mi hermano Tato que siempre ha estado presente para ayudarme en todo. Agradecerles también a mis abuelas Betsy y Mona con quienes pasé tantos veranos completos en Algarrobo donde compartimos momentos de mucha alegría y amor, y también a mis tíos Pame y Jimmy por ser siempre tan apañadores y jugados.

Muchas gracias a mis amigos del colegio, aquellos que han estado junto a mí desde la infancia. Los considero como mis hermanos y los quiero muchísimo. Nos hemos acompañado en muchos momentos, algunos felices y otros muy duros, pero siempre nos hemos mantenido juntos como el gran ghetto que somos. Gracias Marita, Ditus, Vicho, Trini, Elisa, Drex, Peny, Vito, Vero, Nico, Monti, Make, Koke y Matitas. Los voy a querer por siempre.

También quiero agradecerles a mis amigos y partners de la universidad. A ellos que participaron en aquellas largas jornadas de estudio y carretes en las cuales se fue forjando una gran amistad, en particular con los que ahora conformamos el grupo más divertido del mundo: José, Barbi, Zeta, Cami, Dixie, Pía, Baci, Galle y Rafa. No saben lo bien que lo paso con ustedes, son los mejores. Además quiero darle las gracias a todos los que me han acompañado en estos largos años en Beaucheff y en el magíster: Vale, Gabo, Nelson, Solari y Julie Lagos, solo por nombrar a algunos. Por último, quiero agradecerle a Román Marchant, mi gran compañero de vida, quien desde segundo año me ha acompañado en todo momento. Gracias por toda la paciencia, compañía y amor que me has entregado durante estos 5 años. Sin tí, nada de esto hubiera sido posible. Es momento de empezar una nueva etapa y estoy segura de que nos va a ir muy bien, siempre y cuando nos mantengamos juntos. Te amo.

María Paz Jofré Alegría

Resumen Ejecutivo

La criminología es el estudio del crimen desde un punto de vista científico, por lo que integra múltiples disciplinas, tales como la sociología, psicología, economía y matemáticas. Todas estas materias aportan desde sus percepciones en el entendimiento del fenómeno delictual y evidencian la complejidad inherente del problema de la delincuencia y de la seguridad ciudadana.

El objetivo de este estudio es determinar los atributos que influyan en mayor medida en el nivel de actividad delictual presente en la zona delimitada por la Primera Comisaría de la comuna de Santiago. Se busca definir el sentido y magnitud de estas relaciones en base a las características del lugar del suceso. Para lograr esta meta se propone un modelo de regresión logística binaria multivariable, que pertenece a la familia de los modelos de elección discreta.

La comuna de Santiago concentra las más relevantes actividades de gobierno, servicios, comercio, finanzas, cultura, recreación y turismo. Además, dada su ubicación en la ciudad y a la gran presencia de población flotante, se concluye que esta comuna representa un territorio con alta probabilidad de delitos contra la propiedad, en especial los referentes a hurtos.

Para una aplicación correcta del modelo, es necesario dividir la zona estudiada en áreas más Pequeñas, generando así una mayor cantidad de observaciones. En este estudio se requieren subdivisiones del espacio que tengan un sentido lógico para los individuos que circulan por el sector en cuestión. Los diagramas de Voronoi conforman áreas que no está restringida a una forma fija, por lo que éste es un método que se acomoda bastante bien a las exigencias del problema.

El estudio se realiza en dos escenarios, los que se diferencian en la cantidad de atributos considerados. Los resultados de cada caso se analizan en base a los valores de los parámetros obtenidos y del nivel de significancia de éstos. Al analizar el primer escenario se observa que de las variables significativas, se tiene que las más influyentes sobre el nivel delictual son la presencia de farmacias y de paseos peatonales. Los atributos que le siguen en importancia son el comercio masivo y los bancos.

El segundo escenario corresponde a la aplicación del modelo propuesto, pero ahora considerando únicamente los atributos significativos. La selección de variables se lleva a cabo aplicando el método *backward elimination*, el cual redujo la cantidad de atributos a tres. Los resultados obtenidos en este segundo escenario son bastantes similares al anterior, pero las medidas de bondad de ajuste sugieren peores resultados. Luego, se un resultado muy interesante es que la selección de variables no mejora al modelo original de 18 atributos.

Este trabajo sugiere futuros desafíos tanto en dimensiones teóricas como prácticas. Desde el punto de vista teórico, la incorporación de herramientas tales como teoría de juegos, modelos estocásticos y simulación, permitirían robustecer la investigación. En la dimensión práctica, el modelo puede complementarse con la inclusión de un análisis dinámico tanto de las variables independientes como de los delitos. Esto servirá para observar cambios importantes en el comportamiento de los delincuentes al incluirse o eliminarse alguna variable del entorno.

Índice general

1. Introducción	1
1.1. Contexto General	1
1.2. Objetivos	3
1.2.1. Objetivo General	3
1.2.2. Objetivos Específicos	3
1.3. Metodología de Trabajo	4
1.4. Estructura de la Tesis	5
1.5. Alcances y Limitaciones	6
2. Antecedentes Generales	8
2.1. Definiciones Generales	8
2.1.1. Delito	8
2.1.2. Hurto	10
2.2. Realidad Nacional	11
2.2.1. Realidad Delictual en Chile	11
2.2.2. Realidad Delictual en Santiago	13
2.2.3. Fuentes de Información	14
2.3. Caracterización de la Zona a Estudiar	15
2.3.1. Contexto General y Características del Sector	16
2.3.2. Área de Estudio	17
2.3.3. Estadísticas y Análisis de Datos del Sector	18
2.3.4. Tendencias Espaciales de los Delitos	21
2.4. Revisión Bibliográfica	22

2.4.1. Criminología Cuantitativa	23
2.4.2. Modelos de Elección Discreta	26
3. Modelo Matemático	28
3.1. Modelos Lineales	28
3.2. Modelos de Elección Discreta	30
3.2.1. Definiciones Previas	30
3.2.2. Regresión Logística Binaria	32
3.2.3. Método de Máxima Verosimilitud	33
3.2.4. Significación Estadística de los Estimadores	34
3.2.5. Medidas de Bondad del Ajuste	34
3.2.6. Interpretación de los Estimadores	36
4. Aplicación del Modelo Propuesto	38
4.1. Definición del Problema a Resolver	38
4.2. Metodología Aplicada en la Primera Comisaría de Santiago	39
4.2.1. Recolección de los Datos	39
4.2.2. Variables del Modelo	40
4.2.3. Diagramas de Voronoi	46
4.2.4. Selección de Atributos	51
5. Resultados y Análisis	53
5.1. Análisis Bivariante	53
5.2. Resultados Todas las Variables	55
5.3. Resultados Selección de Variables	57
6. Conclusiones y Trabajos Futuros	60
6.1. Resultados Obtenidos	60
6.2. Futuros Desafíos	61
Apéndices	66
A. Método de Newton-Raphson	67

B. Función de Verosimilitud	69
C. Diagramas de Voronoi	70
D. P-Valor	71
E. Coeficientes y Significancia Método Backward Elimination	72
F. Códigos Fuentes	76

Índice de figuras

1.1. Distribución Porcentual de Hechos Delictuales Dentro del Hogar y en la Vía Pública	2
1.2. Principales Preocupaciones de la Población 2000 - 2009	3
1.3. Porcentaje de Hogares del Total del País que Denunciaron al Menos un Delito	6
1.4. Porcentaje de Delitos Denunciados del Total del País	6
2.1. Porcentaje de Hogares Victimizadas del Total del País Según Delitos 2005 - 2010	10
2.2. Evolución Denuncias de Delitos en Chile 1998 - 2007	12
2.3. Denuncias de Todos los Delitos en Chile en el año 2007	12
2.4. Denuncias de Hurtos en Chile 1998 - 2007	13
2.5. Denuncias de Hurtos por Región en el año 2007	13
2.6. Denuncias de Delitos en la Región Metropolitana (% del Total del País) 1998 - 2007	13
2.7. Denuncias de Hurtos de la Región Metropolitana Según Comuna 2007	14
2.8. Ubicación de la comuna de Santiago dentro de la ciudad de Santiago	16
2.9. Mapa Primera Comisaría de Santiago	18
2.10. Mapa Cuadrantes en la Primera Comisaría de Santiago	19
2.11. Denuncias Según Día de la Semana	20
2.12. Denuncias Según Día del Mes	20
2.13. Denuncias Según Mes	20
2.14. Denuncias Según Rango Horario	20
2.15. Denuncias Según Monto Avaluado	21
2.16. Mapas de Densidad de Delitos en la 1ª Comisaría de Santiago 2001 - 2004	22
2.17. Mapas de Densidad de Hurtos en la 1ª Comisaría de Santiago 2001 - 2004	23
2.18. Mapas de Densidad de Hurtos Según Hora en la 1ª Comisaría de Santiago 2001 - 2004	24

2.19. Mapas de Densidad de Hurtos Según Monto Avaluado 1ª Comisaría de Santiago 2001 - 2004	25
2.20. Mapas de Densidad de Hurtos Según Día de la Semana Avaluado 1ª Comisaría de Santiago 2001 - 2004	25
4.1. Distribución de los Hurtos en la Primera Comisaría	41
4.2. Dependencias Gubernamentales	43
4.3. Fuerzas Policiales: Carabineros, PDI y Casetas de Seguridad	43
4.4. Centros de Salud	44
4.5. Establecimientos Educacionales	44
4.6. Transporte Público: Salidas de Metro y Paraderos de Buses	44
4.7. Establecimientos Culturales	44
4.8. Comida y Entretención	44
4.9. Comercio Masivo	44
4.10. Iglesias	45
4.11. Alojamiento	45
4.12. Casas de Cambio	45
4.13. Farmacias	45
4.14. Bancos	45
4.15. Centros de Pago	45
4.16. Supermercados	46
4.17. Estacionamientos	46
4.18. Áreas Verdes	46
4.19. Paseos Peatonales	46
4.20. Primera Comisaría de Santiago dividida en Grillas	47
4.21. Diagramas de Voronoi en la 1ª Comisaría de Santiago 2001 - 2004	48
4.22. Nivel Delictual en la Primera Comisaría de Santiago	50
5.1. Curva ROC 18 Atributos	57
5.2. Curva ROC Atributos Seleccionados	59

Índice de tablas

2.1. Porcentaje Tipo de Delitos	19
2.2. Denuncias Según Cuadrante	19
4.1. Base de Datos de Denuncias	40
4.2. Atributos Considerados	43
4.3. Estadísticas Descriptivas Variable Dependiente Continua	49
4.4. Histograma Variable Dependiente	49
4.5. Comparación Métodos de Categorización	50
4.6. Estadísticas Descriptivas Variables Independientes Continuas	51
4.7. Estadísticas Descriptivas Variables Independientes Dicotómicas	51
5.1. Correlación Atributos con la Variable Dependiente	54
5.2. Iteraciones Bloque Inicial	55
5.3. Resumen del Modelo y Prueba de Hosmer y Lemeshow	55
5.4. Coeficientes Asociados a las Variables Predictoras	56
5.5. Tabla de Clasificación 18 Atributos	56
5.6. Área Bajo la Curva ROC 18 Atributos	57
5.7. Atributos eliminados en cada Iteración	58
5.8. Resumen del Modelo y Prueba de Hosmer y Lemeshow	58
5.9. Coeficientes Asociados a las Variables Seleccionadas	58
5.10. Tabla de Clasificación Atributos Seleccionados	58
5.11. Área Bajo la Curva ROC Atributos Seleccionados	59
E.1. Coeficientes y Significancia de los Atributos en cada Iteración Backward Elimination	75

Capítulo 1

Introducción

Contenido

1.1. Contexto General	1
1.2. Objetivos	3
1.2.1. Objetivo General	3
1.2.2. Objetivos Específicos	3
1.3. Metodología de Trabajo	4
1.4. Estructura de la Tesis	5
1.5. Alcances y Limitaciones	6

En este primer capítulo se presenta un contexto general de la delincuencia tanto en Chile como específicamente en la zona a estudiar. Luego, se plantean los objetivos que se desean lograr con esta tesis, seguido por la metodología de trabajo que se aplica a lo largo del estudio. Por último, se presenta la estructura que se seguirá para exponer la tesis y los alcances que tendrá la misma.

1.1. Contexto General

La seguridad ciudadana es un tema que ha tomado relevancia en los últimos años, lo que se ve reflejado en la última encuesta CEP¹ de diciembre de 2010 en la que se concluye que la delincuencia, los asaltos y los robos son la principal temática al cual el gobierno debería dedicarle un mayor esfuerzo en solucionar [1]. La inquietud por este tema no es algo nuevo, situación que se evidencia en la Figura 1.2 donde se observa que la preocupación por la delincuencia superan el 45% de las respuestas desde el año 2003 hasta la fecha².

En cuanto a la medición de la seguridad ciudadana, actualmente en Chile se tienen dos formas de estimar el nivel de delincuencia. La primera es a través de las encuestas de victimización y la segunda por medio de la base de datos de denuncias perteneciente a Carabineros. Es de conocimiento general que ambas cifras no son exactas dado que la primera está influenciada por diferentes factores,

¹Catálogo de Encuestas. Centro de Estudios Públicos CEP, Chile. [Fecha de consulta: 02 Agosto 2011]. Disponible en: <http://www.cepchile.cl>.

²Este 45% es el porcentaje de personas, relativas al total de los encuestados, que señalan este tema como primera prioridad o preocupación.

mientras que la segunda tiene un sesgo importante debido a la llamada *cifra negra*, término relativo a los delitos que no se denuncian.

Una de las encuestas de victimización más importante es la Encuesta Nacional Urbana de Seguridad Ciudadana (ENUSC). La ENUSC del año 2010 [2] da cuenta de que el 33% de los encuestados vive en un hogar en que algún miembro de la familia ha sido víctima de robo o intento de robo dentro o fuera del hogar en los últimos 6 meses. En cuanto al Índice de Temor, éste revela que el 67% de la población se ubica en un nivel medio de temor, mientras que el 12,2% de los encuestados presenta un nivel de alto temor.

En cuanto a las denuncias reportadas a Carabineros, se observa que el Índice Global de Denuncia presentó un aumento en la ENUSC de 2010, revelando que las notificaciones relativas a los delitos de robo crecen (56,9%), mientras que la *cifra negra* se reduce (43,1%) siendo ésta la más baja de la década. Una tendencia que merece ser advertida es la evolución de las denuncias por delitos según el lugar de ocurrencia. Los registros de Carabineros de Chile del año 2009 indican que existe una mayor concentración de delitos denunciados ocurridos en espacios públicos (42,3%) que en otros lugares, tales como domicilio (23,7%) y locales comerciales (18,1%), entre otros (15,9%).

Lo anterior coincide con lo concluido por Paz Ciudadana en un informe publicado en enero de 2011 [3] donde se revela que cerca del 49% de la población cree muy probable que será víctima de robo en la vía pública. Por otra parte, se observa (Figura 1.1) que desde el año 2004, se tiene que más del 80% de las denuncias anuales son delitos ocurridos en la vía pública.

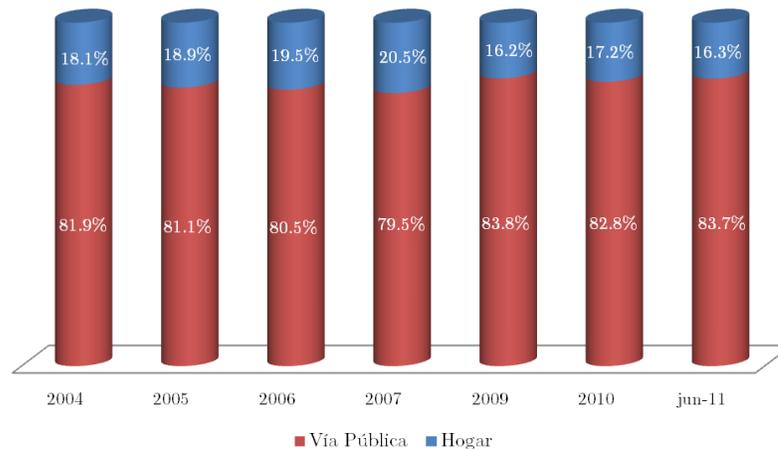


Figura 1.1: Distribución Porcentual de Hechos Delictivos Dentro del Hogar y en la Vía Pública
Fuente: Paz Ciudadana, Índice Paz Ciudadana 2011

Muy interesante es lo que ocurre en la comuna de Santiago Centro. Este sector presenta un alto nivel de denuncias con respecto al total regional (más del 12%, y la comuna que le sigue posee menos del 6%). Adicionalmente, la encuesta CEP de diciembre de 2010 revela que el 61% de los encuestados siente temor de ser asaltado o robado en el centro de la ciudad. Ambos hechos, sugieren una actividad delictual intensa en el centro de la capital.

Lo anterior revela lo importante que es para la sociedad la seguridad ciudadana, por lo que las mejoras que puedan realizarse en este tema tienen un efecto muy significativo.

En el último tiempo, la criminología ha logrado grandes avances en el ámbito de las ciencias sociales. En general, los estudios en esta área se enfocan en el comportamiento humano y las relaciones

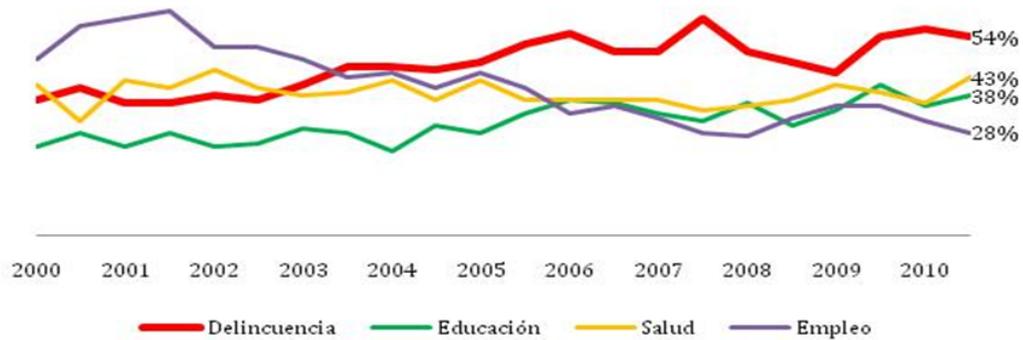


Figura 1.2: Principales Preocupaciones de la Población 2000 - 2009
Fuente: Paz Ciudadana, Balance de la Delincuencia 2009

que surgen entre los distintos agentes de la sociedad. Actualmente, los enfoques de la criminología definen diferentes áreas de estudios, las que pueden caracterizarse tanto por el desarrollo de estudios cualitativos como cuantitativos.

Los primeros estudios en el área cuantitativa son las relacionadas al ámbito de la econometría y estadística. Actualmente, gracias al desarrollo de herramientas tecnológicas y softwares computacionales avanzados, ha sido posible la expansión de estas investigaciones y el desarrollo de modelos más complejos y exactos. No obstante, estas técnicas no se destacan a la hora de considerar otros factores relevantes, tales como presencia de diferentes actores en el medio y la interacción entre ellos, o características del entorno que puedan definir una mayor o menor atractividad del lugar.

Como consecuencia de lo anterior, varios modelos alternativos se han desarrollado con el objetivo de explicar de mejor forma el fenómeno delictual. Sin embargo, hay muchos factores que no son capturados de forma satisfactoria por dichos modelos: factores psicológicos y sociológicos subyacentes a la conducta delictiva, características del entorno del lugar donde ocurre el delito o zonas alternativas con rasgos similares, entre otros.

En este trabajo se desarrolla un modelo de elección discreta donde se incluyen algunas características de la comuna de Santiago con el propósito de determinar las variables que la hacen más o menos atractiva para los delincuentes.

1.2. Objetivos

1.2.1. Objetivo General

Desarrollar un modelo que determine los atributos que influyen en mayor medida en el nivel de actividad delictual presente en una zona particular.

1.2.2. Objetivos Específicos

1. Plantear un modelo de elección discreta que describa la atractividad de un lugar en base a los atributos que presenta.

2. Identificar las variables de entorno relevantes que determinen en mayor medida el nivel delictual presente en diferentes zonas.
3. Desarrollar una metodología que permita la aplicación del modelo teórico con datos reales de crimen, utilizando división del espacio y selección de atributos.
4. Proveer sugerencias relacionadas a la planificación y organización de los recursos relacionados al área de la criminología y seguridad ciudadana usando criterios de optimalidad.

1.3. Metodología de Trabajo

1. Antecedentes Generales, Recolección, Preparación y Exploración de Datos

En esta primera parte, se presenta la realidad delictual de Chile, la Región Metropolitana y de la comuna de Santiago. Además, se realiza un análisis socio demográfico y estadístico de esta comuna, definiéndose las tendencias delictuales que la caracterizan. A continuación, se procede a una revisión bibliográfica de esta disciplina determinando así el estado del arte de los modelos utilizados en esta materia.

Una vez definido el marco teórico, se da paso a la recolección de los datos, tanto de las denuncias ocurridas en el sector estudiado como la ubicación de los diferentes atributos que conforman esta área. La base de datos de denuncias de delitos ocurridos entre Enero de 2001 y Diciembre 2004 de la Primera Comisaría de la comuna de Santiago Centro fue facilitada por Carabineros de Chile, mientras que los atributos se obtuvieron en visitas a terreno y gracias a la información entregada por diferentes instituciones y organizaciones.

En esta segunda parte de la tesis, se realiza una limpieza y preparación de los datos, definiendo así una primera mirada a lo que se tiene disponible. Además, se llevan a cabo estudios exploratorios sobre los datos anteriormente mencionados con el objetivo de disponer de información ordenada para las etapas siguientes.

2. Planteamiento del Modelo de Regresión Logística

En este punto, se exponen diversos modelos matemáticos utilizados para estudiar una variable en particular en función de otras variables relevantes. Se especifican las ventajas y desventajas de cada uno de ellos y se justifica la utilización de regresión logística binaria para el desarrollo de este estudio.

A continuación, se plantea matemáticamente el modelo elegido y se detalla el problema de optimización que se resuelve en el desarrollo del mismo. Además, se definen los elementos y herramientas estadísticas que se utilizarán para analizar el nivel de significancia de los resultados obtenidos.

3. Aplicación y Evaluación del Modelo Propuesto

La aplicación del modelo y la obtención de resultados se lleva a cabo utilizando diferentes

softwares apropiados para tales efectos, tales como GRASS GIS 6.5, SPSS 17.0, MATLAB 2009 y Microsoft Office Excel 2007.

Dada la gran cantidad de atributos considerados, se desarrollan algunos algoritmos de selección de variables que permiten la clasificación de las variables que aportan en mayor medida a obtener mejores resultados.

La evaluación del modelo se realiza en base a herramientas estadísticas y medidas de bondad propias de la regresión logística, tales como estadísticos matemáticos e índice de predicción.

4. Análisis de los Resultados y Conclusiones

Una vez aplicado el modelo de regresión logística, se analizan los resultados obtenidos en distintos escenarios. Se realiza un estudio acabado de las correlaciones y asociaciones de la variable dependiente con las variables explicativas, y se determina el nivel de significancia del modelo y otras medidas de bondad en cada una de las situaciones analizadas.

Por último, se concluye acorde a lo que proyectan los resultados obtenidos. En base a lo anterior, se realizan algunas sugerencias desde el punto de vista de las organizaciones que combaten el crimen y se detalla una propuesta de futuros desafíos relacionados tanto con los datos y supuestos considerados como con el desarrollo del modelo propuesto.

1.4. Estructura de la Tesis

En el capítulo 2 se expone el marco general de la realidad que se desea modelar. Se definen algunos conceptos claves para el desarrollo del modelo, seguido por un análisis que contextualiza la realidad delictual tanto en Chile como en Santiago. A continuación, se realiza una caracterización del área de estudio y de las tendencias espaciales de los delitos ocurridos en esta zona. Por último, se realiza una revisión bibliográfica que ayuda a determinar el estado del arte de los estudios del crimen desde la estadística, matemáticas y computación. El objetivo de esta sección es contextualizar el fenómeno de la delincuencia en Chile y revisar las distintas modelaciones del crimen que se han desarrollado en diferentes partes del mundo.

En el capítulo 3 se presentan los puntos teóricos centrales de este trabajo. Se detallan diferentes modelos matemáticos adecuados para este tipo de análisis, y los beneficios y desventajas de cada uno de ellos para el problema planteado en esta tesis. Luego, se define el modelo de regresión logística y se plantea matemáticamente el problema de optimización que se desea resolver. Además, se detallan las herramientas estadísticas que se utilizarán para analizar el nivel de significancia de los resultados y diferentes medidas empleadas para examinar la bondad de ajuste del modelo propuesto.

En el capítulo 4 se realiza una preparación acabada de los datos que se tienen, identificando las características de las distintas variables consideradas en este estudio. Una vez resuelto lo anterior, se lleva a cabo la aplicación del modelo propuesto utilizando datos reales de denuncias de delitos y de atributos del lugar. Adicionalmente, se desarrollan distintas metodologías para la selección de variables las cuales se justifican con diferentes pruebas y análisis estadísticos.

En el capítulo 5 se realiza un extenso análisis bivariado. Luego, se exponen los resultados obtenidos al considerar todas las variables independientes y los conseguidos al realizar una selección de estas variables. Asimismo, se analiza la confiabilidad y validez de los resultados encontrados.

Finalmente, en el capítulo 6 se presentan las principales conclusiones y hallazgos. Luego, se formulan

las propuestas que buscan mejorar la toma de decisiones por parte de las instituciones encargadas de combatir el crimen y se da lugar a la discusión de futuros desafíos en diferentes áreas de investigación.

1.5. Alcances y Limitaciones

Esta tesis contempla el desarrollo de un modelo que permita establecer de forma cuantitativa las variables que tienen mayor influencia en la atractividad de un lugar específico para los delincuentes. Para tal fin, se considerarán los registros de denuncias históricos ubicados en la Primera Comisaría de Santiago.

Una vez definidos los alcances generales de esta tesis, se determinan los puntos donde se encuentran los límites de la misma.

En primer lugar, el modelo propuesto se aplica exclusivamente en el terreno especificado anteriormente y no se espera una generalización para otros sectores. La comuna de Santiago presenta características particulares diferentes a otras zonas de la capital y del país, por lo que el modelo y los resultados obtenidos en este estudio no pueden hacerse extensivos a otras áreas de interés.

En cuanto a la calidad de los datos, se considerará que las denuncias de los delitos notificados en la Primera Comisaría de Santiago entre enero de 2001 y diciembre de 2004, son representativos de los crímenes ocurridos en este sector durante este periodo de tiempo. Lo anterior se justifica dado el alto porcentaje de denuncias, más del 43%, establecido en la ENUSC para delitos principalmente contra la propiedad y lesiones (Figura 1.3). Además, de la Figura 1.4 se observa que el porcentaje de denuncias en relación al total cometido es bastante alto también, por lo que los datos considerados son, efectivamente, representativos de la canasta total de delitos cometidos.

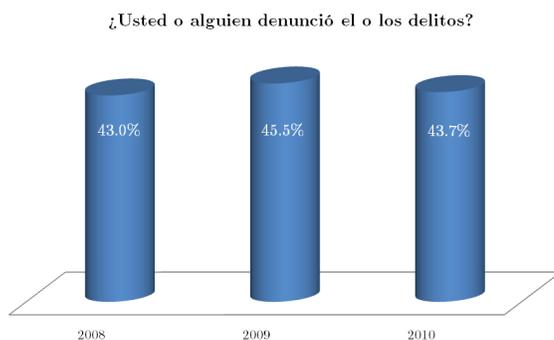


Figura 1.3: Porcentaje de Hogares del Total del País que Denunciaron al Menos un Delito

Fuente: Paz Ciudadana, Índice Paz Ciudadana 2011

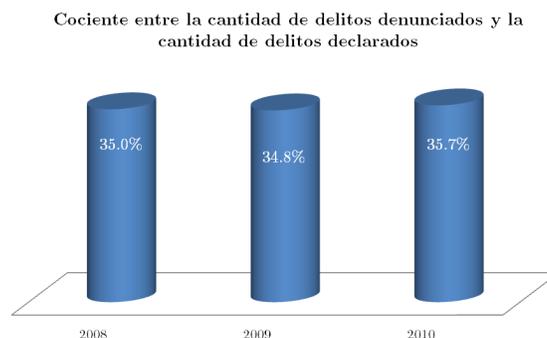


Figura 1.4: Porcentaje de Delitos Denunciados del Total del País

Fuente: Paz Ciudadana, Índice Paz Ciudadana 2011

Por otro lado, no se cuestionará el ingreso de los datos, tales como errores de tipeo y valores incorrectos y/o imprecisos, ni se argumentará acerca de la *cifra negra* de estas mismas denuncias.

Una buena forma de complementar los datos que se tienen, es la utilización de la encuesta de victimización de la ENUSC. Sin embargo, esta encuesta entrega resultados a nivel nacional, regional y comunal, y dado que no se tiene información más específica, no es posible su interpretación a nivel de zonas más pequeñas.

Como ya se mencionó, el objetivo de este estudio es desarrollar un modelo de optimización

considerando únicamente las características propias del lugar de estudio y su atractividad. Luego, la inclusión de variables propias de los agentes y la interacción entre los diversos individuos están fuera del alcance de la tesis, por lo que aplicaciones de teoría de juegos, modelos estocásticos y simulación quedan propuestas como futuros desafíos.

Los delitos pueden ser de diferente índole, lo que implica un comportamiento criminal distinto para cada uno de ellos. Luego, se restringe este trabajo a un tipo de delito definido. Por lo tanto, la modelación del comportamiento delictual considerando otros tipos de delitos queda propuesta como un estudio en el que se puede seguir ampliando la investigación.

Capítulo 2

Antecedentes Generales

Contenido

2.1. Definiciones Generales	8
2.1.1. Delito	8
2.1.2. Hurto	10
2.2. Realidad Nacional	11
2.2.1. Realidad Delictual en Chile	11
2.2.2. Realidad Delictual en Santiago	13
2.2.3. Fuentes de Información	14
2.3. Caracterización de la Zona a Estudiar	15
2.3.1. Contexto General y Características del Sector	16
2.3.2. Área de Estudio	17
2.3.3. Estadísticas y Análisis de Datos del Sector	18
2.3.4. Tendencias Espaciales de los Delitos	21
2.4. Revisión Bibliográfica	22
2.4.1. Criminología Cuantitativa	23
2.4.2. Modelos de Elección Discreta	26

2.1. Definiciones Generales

Para comprender de manera correcta el contexto delictual, es necesario familiarizarse con algunas definiciones y conceptos que se relacionan con el fenómeno y comportamiento criminal.

2.1.1. Delito

La noción de qué es un delito no es algo que esté definido de manera clara. Existen diversos tipos de crímenes y no hay una explicación que los defina a todos de manera global, ni tampoco una única forma de prevenirlos y reprimirlos. Sin embargo, la Real Academia Española¹ define un delito como

¹Diccionario de la Lengua Española. Real Academia Española RAE. [Fecha de consulta: 02 Agosto 2011]. Disponible en: <http://www.rae.es/>.

“una acción u omisión voluntaria o imprudente que se encuentra penada por la ley”. Por lo tanto, el delito supone un quebrantamiento de las normas y acarrea un castigo para el responsable.

Por otra parte, en el informe de seguridad ciudadana del año 2003 se hizo un esfuerzo por definir este concepto. En consecuencia, el Código Penal define el delito como “toda acción u omisión voluntaria penada por la Ley. Las acciones u omisiones penadas por la Ley se reputan siempre voluntarias, a no ser que se conste lo contrario” (Código Penal Art. 1).

Además, el Código Penal distingue el delito del crimen en función a la pena asignada a quien la comete. De esta forma, los ilícitos cuyas penas son superiores a cinco años y un día se denominan crímenes, mientras que para aquellos cuya pena es inferior, se denomina simple delito. Debido a que esta distinción es de orden procesal, en esta tesis se utilizarán los términos de delitos y crímenes como sinónimos.

Los criminólogos Lawrence Cohen y Marcus Felson desarrollaron en el año 1979 la teoría de la Actividad Rutinaria [4]. Se postula que los delitos ocurren cuando en el mismo tiempo y espacio están presentes tres elementos básicos: un objetivo conveniente, la ausencia de un guardián capaz de reprimir el delito y la presencia de un delincuente.

Los objetivos pueden ser una persona, un objeto o un lugar. Éstos tienen que ser llamativos para el delincuente, por lo que deben ser valiosos, visibles y accesibles. Un guardián capaz es usualmente una persona cuya presencia puede hacer que los delincuentes no actúen. Así, cuando un objetivo está desprotegido, existe la posibilidad de que el delito se cometa. El elemento final para que el crimen ocurra es la presencia de un potencial delincuente.

Por otra parte, se tiene que los delitos pueden ser de distinta índole. Debido a esta razón y para no incurrir en ambigüedades o imprecisiones, se deben identificar de manera clara los diferentes tipos de delitos que existen. Cada año, la Oficina Federal de Investigaciones (Federal Bureau of Investigation o FBI²) hace una recopilación uniforme de datos sobre el crimen en los Estados Unidos y publica estadísticas de los distintos delitos como parte de un programa llamado *Uniform Crime Reporting*. Los informes no solo brindan estadísticas útiles, sino también clasificaciones y definiciones de los delitos, los cuales se presentan a continuación.

Delitos Contra las Personas

- Violación: Tener acceso carnal con alguien en contra de su voluntad o cuando se halla privado de sentido o discernimiento.
- Homicidio: Acción de matar a una persona intencionalmente.
- Lesiones: Ataque ilegal a otra persona para causarle lesiones corporales severas o muy graves.
- Robo con Violencia o Intimidación en las Personas: Acción de quitar o intentar quitar algo de valor a una persona por la fuerza o amenazándola con usar fuerza o violencia.

Delitos Contra la Propiedad

- Hurtos: Acción de quitar o intentar quitar algo de valor a una persona por la fuerza o amenazándola con usar fuerza o violencia.

²Federal Bureau of Investigation, Estados Unidos. [Fecha de consulta: 02 Agosto 2011]. Disponible en: <http://www.fbi.gov/>.

- Robo con Fuerza en las Cosas: Sustracción de especies mediante fracturas, forzamiento, uso de llaves falsas o verdaderas sustraídas.
- Daños: Acción de destruir, estropear o inutilizar una cosa mueble o inmueble ajenos.
- Drogas: Ejecución de actos de cultivo, elaboración o tráfico, o de otro modo promuevan, favorezcan o faciliten el consumo ilegal de drogas tóxicas, estupefacientes o sustancias psicotrópicas, o las posean con aquellos fines.

2.1.2. Hurto

Como se menciona en el apartado anterior, existen diferentes tipos de delitos, lo que implica por su parte un comportamiento criminal distinto para cada uno de ellos. El *modus operandis* de un crimen en particular se caracteriza por un comportamiento delictual particular y por toda una cadena delictual propia. Por lo tanto, realizar un análisis considerando todos los tipos de delitos podría conducir a errores importantes. Luego, para evitar lo anterior, se debe definir el tipo de delito que se desea estudiar.

Como resultado de la última Encuesta Nacional Urbana de Seguridad Ciudadana del año 2010, se observa (Figura 2.1) que a nivel nacional, los hogares son victimizados principalmente por hurtos, seguido de robo con violencia o intimidación y robo por sorpresa.

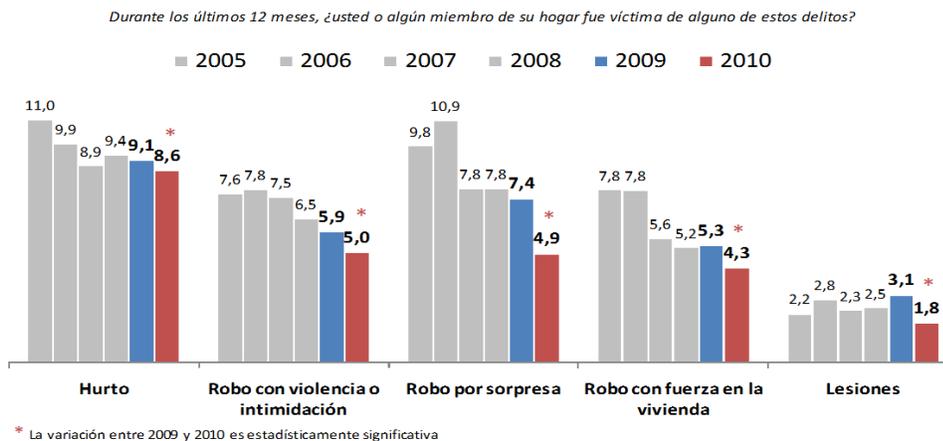


Figura 2.1: Porcentaje de Hogares Victimizados del Total del País Según Delitos 2005 - 2010

Fuente: *ENUSC 2010, Ministerio del Interior y Seguridad Pública*

Como se verá más adelante en esta misma sección, en esta tesis se consideran los atributos físicos de una zona caracterizada por una alta concentración de actividades de comercio, finanzas, servicios, gobierno, cultura, recreación y turismo. Considerando todo lo anterior, se decide trabajar únicamente con los delitos relativos a *hurtos*.

El Código Penal chileno establece el delito de *hurto* como “la apropiación de una cosa mueble ajena, con ánimo de lucrarse, sin la voluntad del dueño, no existiendo ni violencia, ni intimidación en las personas, ni fuerza en las cosas”.

Existen diferentes clasificaciones de este delito, las cuales se identifican a continuación.

- Apropiación de cables del tendido eléctrico o de comunicaciones

- Hurto simple
- Hurto de hallazgo
- Hurto agravado
- Hurto supermercado u otro comercio
- Hurto falta
- Hurto simple por un valor de media a 4 U.T.M.
- Hurto simple por un valor de 4 a 40 U.T.M.
- Hurto simple por un valor sobre 40 U.T.M.

2.2. Realidad Nacional

Como se menciona en la sección anterior, el problema de la delincuencia se encuentra entre las principales preocupaciones de la sociedad chilena. Es por esta razón que se hace urgente que los distintos organismos encargados de este tema tomen medidas que ayuden a apaciguar el crecimiento de la actividad delictual.

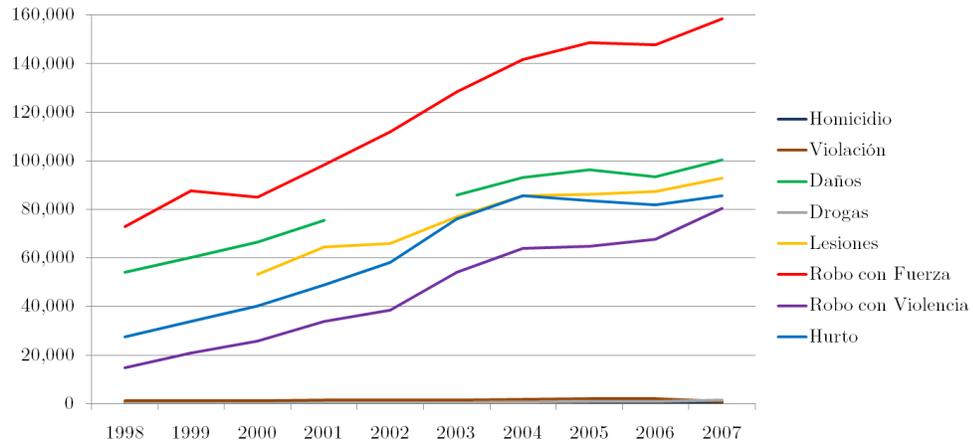
2.2.1. Realidad Delictual en Chile

Antes de comenzar a desarrollar esta materia, es fundamental mencionar que una de las dificultades más importantes es la falta de estadísticas criminales unificadas. En consecuencia de lo anterior, la Fundación Paz Ciudadana se ha preocupado de publicar un Anuario de Estadísticas Criminales, el cual permite concentrar en una sola parte la información proveniente de Carabineros, Investigaciones y el Poder Judicial. Para esta tesis se utilizarán los datos de denuncias de Carabineros publicados en este anuario.

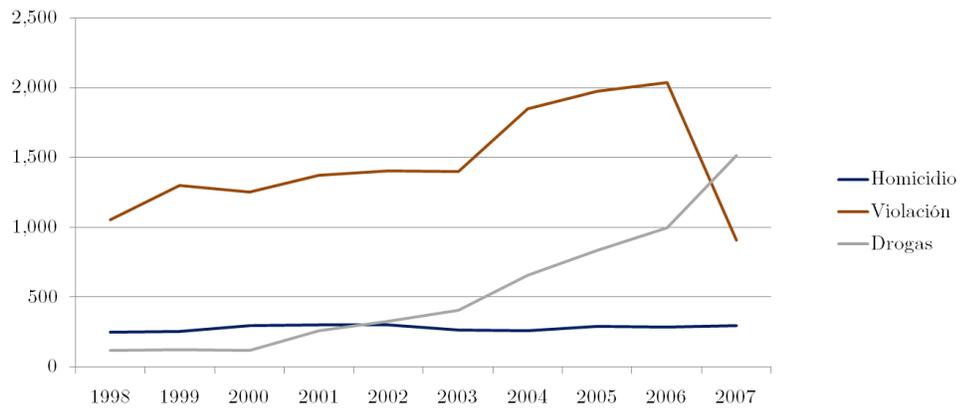
Considerando estos datos [5], se puede ver que Chile presenta un panorama muy particular en cuanto a la realidad delictual que vive. De la Figura 2.2 se puede observar que desde el año 1998, los delitos de droga y los delitos contra las personas presentan una variación importante a nivel nacional. Las violaciones se duplicaron entre 1998 y 2006, presentando luego una disminución significativa el año 2007. Los homicidios se mantuvieron constantes en la última década y los daños aumentaron casi el doble.

La tendencia que presenta el robo con fuerza, robo con violencia y hurto, es un aumento bastante uniforme a través de los años. Al considerar únicamente los datos del año 2007, se observa (Figura 2.3) que alrededor del 45 % de las denuncias hacen referencia a robo con violencia o con fuerza, lo que revela un escenario bastante violento.

En el caso propio de los hurtos, se observa en la Figura 2.4 que en la última década se ha triplicado el número de denuncias en el país, definiéndose desde el año 2004 un desacelero en la tendencia de estos delitos. Por otra parte, la Figura 2.5 muestra una clara concentración de denuncias de hurtos en la Región Metropolitana, lo que deja en manifiesto la urgencia de tomar medidas que ayuden a apaciguar este comportamiento.



(a) Denuncias de Todos los Delitos



(b) Denuncias de Delitos de Menor Magnitud

Figura 2.2: Evolución Denuncias de Delitos en Chile 1998 - 2007
Fuente: *Sistema Informático de Automatización Policial de Carabineros de Chile (AUPOL)*

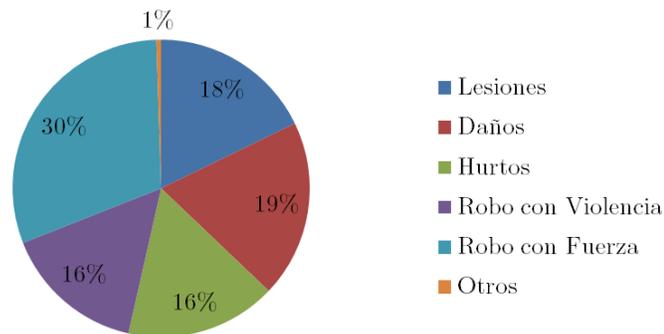


Figura 2.3: Denuncias de Todos los Delitos en Chile en el año 2007
Fuente: *Sistema Informático de Automatización Policial de Carabineros de Chile (AUPOL)*

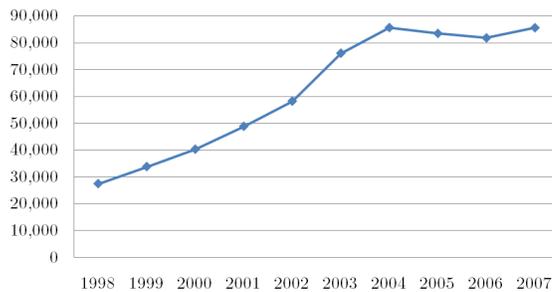


Figura 2.4: Denuncias de Hurtos en Chile 1998 - 2007

Fuente: *Sistema Informático de Automatización Policial de Carabineros de Chile (AUPOL)*

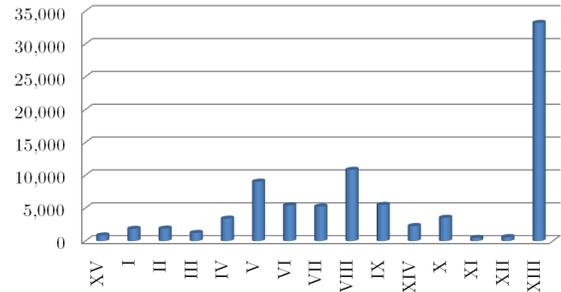


Figura 2.5: Denuncias de Hurtos por Región en el año 2007

Fuente: *Sistema Informático de Automatización Policial de Carabineros de Chile (AUPOL)*

2.2.2. Realidad Delictual en Santiago

Después de este primer análisis, se puede decir que la Región Metropolitana es una zona de alta frecuencia delictual. De la Figura 2.6 se observa que esta región presenta porcentajes muy altos, relativos al total del país, especialmente en robo con violencia, homicidio, violación y drogas.

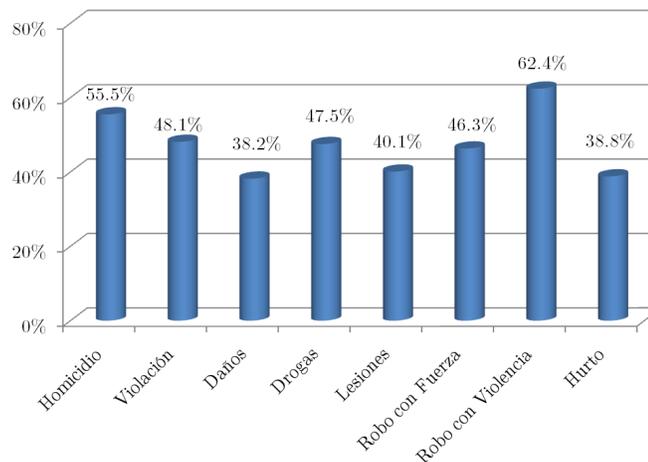


Figura 2.6: Denuncias de Delitos en la Región Metropolitana (% del Total del País) 1998 - 2007

Fuente: *Sistema Informático de Automatización Policial de Carabineros de Chile (AUPOL)*

Por otro lado, un análisis de las comunas de la Región Metropolitana permite establecer que hay zonas que se caracterizan por presentar altas tasas³ de ocurrencia de ciertos delitos. En cuanto a los crímenes relativos a homicidio, las comunas de La Granja y San Joaquín son las que presentaron los valores más altos de la capital, con un 8,7 y un 8,2 respectivamente. Por otro lado, las comunas con mayores índices de violación son La Pintana (16,4), Independencia (15,9) y La Granja (15,6). En cuanto a los delitos de daños, las comunas que presentan mayores tasas son Santiago (1903) y Providencia (1.201). Respecto a las lesiones, nuevamente Santiago muestra los índices más altos de

³Estas tasas de ocurrencia se medirán en cantidad de delitos cada 100.000 habitantes.

la capital. De la misma forma, cuando se habla de los delitos de robo, tanto con fuerza como con violencia, se puede observar que la comuna con mayor concentración de denuncias es la del centro de la capital, presentando valores superiores a los 40.00 delitos cada 100.000 habitantes en ambos casos.

En cuanto a los hurtos, se observa (Figura 2.7) que la zona con mayor concentración de estos delitos es la comuna de Santiago con un alarmante 18 %, valor muy superior a las cifras presentadas por las comunas de Providencia (9,1%) y Las Condes (9%) que son las que le siguen en magnitud. Al realizar un análisis por habitantes, se observa el mismo fenómeno, quedando en evidencia que la cantidad de hurtos por cada 100.000 habitantes en la comuna de Santiago alcanza los 3.306, mientras que en Providencia se observa una tasa de 2.310. Finalmente, queda demostrada la crítica situación que se vive en la comuna de Santiago con respecto a este tipo de delito en particular.

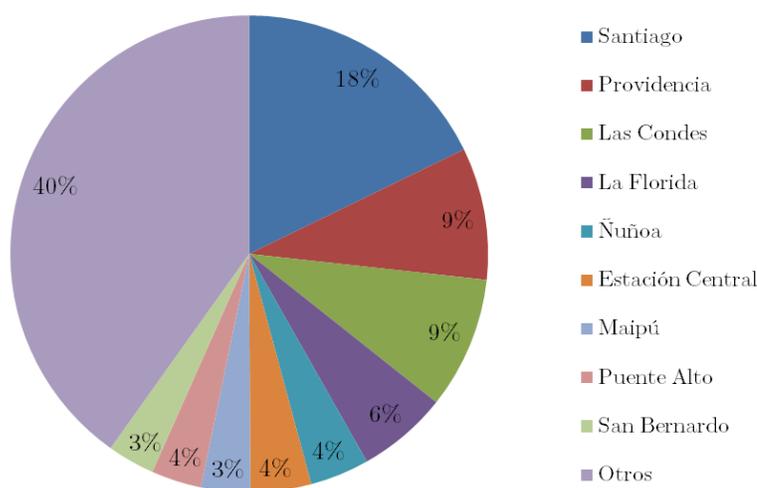


Figura 2.7: Denuncias de Hurtos de la Región Metropolitana Según Comuna 2007
Fuente: Sistema Informático de Automatización Policial de Carabineros de Chile (AUPOL)

2.2.3. Fuentes de Información

A continuación se especifican las diferentes fuentes de información que se utilizan para describir el contexto general del presente trabajo. Se detallan las diferentes encuestas nacionales y estadísticas policiales utilizadas para definir la realidad de Chile y de las distintas comunas de Santiago.

Encuesta Nacional Urbana de Seguridad Ciudadana ENUSC

La Encuesta Nacional Urbana de Seguridad Ciudadana, ENUSC, es una encuesta realizada por el Instituto Nacional de Estadísticas, INE⁴, con el objetivo de construir una fuente de información confiable para el desarrollo de políticas de seguridad ciudadana en el país.

Durante el año 2003 y a petición del Ministerio del Interior, se aplica la primera Encuesta Nacional Urbana de Seguridad Ciudadana en Chile. Luego, para actualizar esta valiosa información respecto

⁴Más información en <http://www.ine.cl/>. [Fecha de consulta: 02 Agosto 2011].

al fenómeno delictual, el año 2005 se lleva a cabo una segunda encuesta nacional. Desde entonces, se ha realizado esta encuesta todos los años siguientes hasta la actualidad.

El objetivo de este estudio es obtener información relevante sobre los niveles de victimización y de percepción de inseguridad que afectan a la población. Se busca conocer además, la reacción de la ciudadanía frente al delito y su opinión sobre el desempeño de los organismos estatales y los actores principales vinculados a la seguridad ciudadana.

La ENUSC toma como periodo de referencia los últimos 12 meses anteriores a la fecha de entrevista, excepto en preguntas de percepción. Esta encuesta no aborda todos los tipos de delitos y se desarrolla a nivel comunal regional y nacional, mas no a nivel de barrio. El número de encuestados supera los 25 mil individuos, por lo que representa uno de los estudios más representativos del fenómeno delictual del país. La población objetivo la constituyen los hogares que habitan las viviendas particulares urbanas y las personas de 15 o más años que los componen, siendo seleccionada una de ellas para responder esta encuesta. El tipo de entrevista es el denominado *cara a cara* y la cobertura geográfica del estudio es Nacional-Urbano.

Índice Paz Ciudadana-Adimark

El Índice Paz Ciudadana-Adimark es una encuesta realizada por la Fundación Paz Ciudadana⁵ desde el año 2000. Esta encuesta permite conocer la evolución de delitos contra la propiedad y donde se abarcan temas como el comportamiento de las personas victimizadas hacia la denuncia y la evaluación que ellas hacen de las autoridades responsables de la seguridad ciudadana.

Ésta es una encuesta de victimización y se realiza de forma telefónica a personas mayores de 18 años que residen en hogares con teléfono en las comunas seleccionadas de la Región Metropolitana y de otras regiones. El número de encuestados es alrededor de los 8 mil individuos y se tiene una representación general de la realidad regional y comunal.

Estadísticas Policiales

Las estadísticas policiales se registran en carabineros (sobre el 90 %) y en PDI (entre 5 % y 10 %). Se trata de aprehensiones y denuncias, por lo que el gobierno desde hace un tiempo los trata como casos policiales y corresponde a la suma de ambos. Las denuncias hacen referencia a distintos tipos de delitos, tales como Homicidio, Infanticidio y Parricidio, Violación, Daños, Drogas, Lesiones, Robo con Fuerza, Robo con Violencia y Hurto. Las aprehensiones, por su parte, son las relativas a la captura y detención de los individuos que cometieron uno o más de los delitos mencionados.

2.3. Caracterización de la Zona a Estudiar

A pesar de que la ciudad de Santiago es muy segura en comparación con otras regiones de América Latina [6], es considerada como una metrópoli con temor debido a que gran parte de sus habitantes la percibe con gran inseguridad frente al delito y la violencia.

Dadas las características que determinan esta zona, se tienen áreas que ofrecen mayores probabilidades ambientales y físicas de ser afectadas por los delitos de robos y hurtos. Lo anterior

⁵Más información en <http://www.pazciudadana.cl/>. [Fecha de consulta: 02 Agosto 2011].

permite identificar algunas tendencias espaciales que condicionan el nivel de delincuencia que afecta a este terreno en particular.

2.3.1. Contexto General y Características del Sector

La comuna de Santiago es el centro histórico de la ciudad del mismo nombre. Como se puede ver en la Figura 2.8, esta comuna está ubicada al centro de la macro región capitalina [7], limitando con diez comunas de la Región Metropolitana, lo que refleja su fácil acceso.

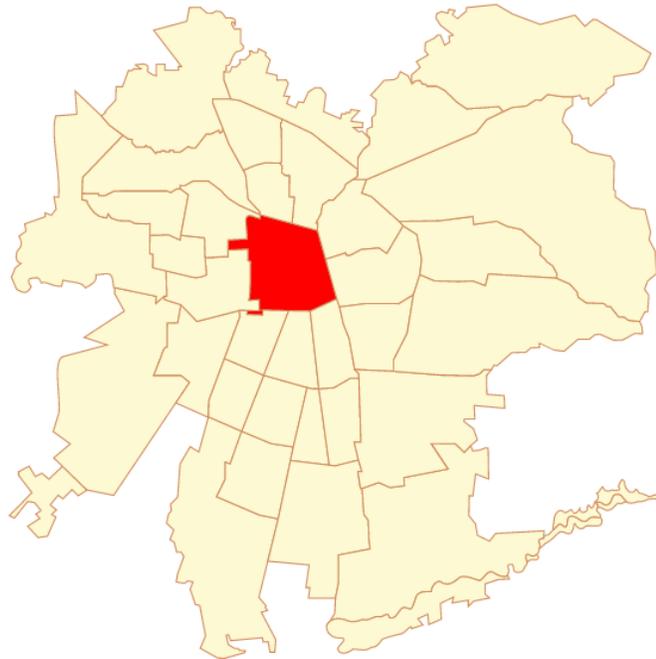


Figura 2.8: Ubicación de la comuna de Santiago dentro de la ciudad de Santiago
Fuente: *Wikipedia*

Adicionalmente y relativo al transporte, se tiene que tres líneas del Metro de Santiago y dos carreteras de la Red de Autopistas atraviesan esta comuna. Además, la Avenida Libertador General Bernardo O'Higgins es uno de los ejes viales de mayor importancia de la ciudad, al igual que la Avenida Vicuña Mackenna, calle que define el límite oriente de la comuna. En consecuencia de lo anterior, se tiene que los resultados arrojados por la última de las encuestas de origen y destino aplicada en el año 2001, confirman que el 17 % de los viajes en el área urbana de la Región Metropolitana se generan en la zona central de la capital. De estos viajes, se sabe que el 56 % de las personas lo hace en bus o taxibus y el 13,4 % en automóvil, mientras que el 12 % de los encuestados prefiere caminar.

Es importante destacar, por otro lado, que la población de la comuna aumenta notablemente durante la jornada laboral por sobre el total de la población permanente, pues existe un importante y grueso flujo de población flotante en la comuna. Ésta se ha estimado en 1.800.000 personas⁶, un millón

⁶Municipalidad de la Comuna de Santiago, Santiago, Chile. [Fecha de consulta: 02 Agosto 2011]. Disponible en: <http://www.municipalidadesantiago.cl>.

más que la comuna que le sigue en este ranking, que es el caso de Providencia. Se justifica entonces la gran cantidad de vehículos y personas que circulan por este sector.

Por otro lado, se observa que Santiago presenta una gran cantidad de edificaciones estatales y espacios públicos, tales como el Palacio de La Moneda, ubicado en la Plaza de la Constitución, el Palacio de Tribunales, la Ilustre Municipalidad de Santiago, la Catedral Metropolitana de Santiago y el Correo Central, estos tres últimos situados en torno a la Plaza de Armas. Se puede concluir entonces que en esta zona existe una alta concentración de actividades sociales y de gobierno.

En esta comuna también se ubica la Bolsa de Comercio de Santiago y una gran cantidad de casas matrices de bancos y financieras. Se caracteriza por ser el centro financiero y comercial del Gran Santiago⁷, albergando el 61 % de las colocaciones bancarias de la ciudad y el 56 % de los depósitos y captaciones bancarias del país. El Paseo Ahumada y el Paseo Estado, ubicados entre la Plaza de Armas y la Alameda, son paseos peatonales de gran actividad comercial y de alta congestión, en los que se encuentran las principales casas y locales comerciales. Sin duda entonces, queda en evidencia la gran concentración de actividades de servicios, comercio y finanzas.

Por otra parte, esta zona también se caracteriza por reunir muchos establecimientos culturales y de recreación. Existe un eje cultural conformado por la avenida Costanera Andrés Bello que bordea el Río Mapocho desde el Museo Nacional de Bellas Artes, donde se encuentra el Mercado Central y la Estación Mapocho, a unas pocas cuadras de distancia del Teatro Municipal de Santiago y de la Biblioteca Nacional de Chile, ubicada al costado del Cerro Santa Lucía o Cerro Huelén.

En consecuencia de todo lo anteriormente mencionado, queda en evidencia que la comuna de Santiago concentra las más relevantes actividades de gobierno, servicios, comercio, finanzas, cultura, recreación y turismo. Además, dada su ubicación en el mapa de la capital y a la gran afluencia vehicular y de población flotante, se concluye que esta comuna representa un territorio con alta probabilidad de delitos contra la propiedad.

2.3.2. Área de Estudio

Para efectos de esta tesis, se estudiará el terreno de la Primera Comisaría de Santiago. Como se observa en la Figura 2.9, sus límites son: al norte el Río Mapocho, al sur la Avenida Libertador Bernardo O'Higgins, al oriente la Avenida Vicuña Mackenna y al poniente la calle Amunátegui.

Esta zona corresponde a la parte más concurrida de la comuna de Santiago. Se caracteriza principalmente por la existencia de departamentos residenciales, oficinas y una marcada actividad comercial. La existencia de paseos peatonales y la alta afluencia de personas durante el día, incentiva la presencia del comercio informal y genera un escenario bastante atractivo para la ocurrencia de delitos en la vía pública.

En cuanto a los recursos policiales, Carabineros cuenta con 310 funcionarios en el terreno de la Primera Comisaría de Santiago, de los cuales 287 son operativos y 23 administrativos. Cuentan además con 12 radiopatrullas, 7 furgones, 16 motos todo terreno y 2 retenes móviles [8].

Además, en este terreno se encuentran los Cuarteles Generales de la Policía de Investigaciones, la Brigada de Investigaciones Policiales Especiales, la Brigada Investigadora de Delitos contra el Medio Ambiente y la Brigada Investigadora de Lavados de Activos. Sin embargo, como una unidad operativa asiste y cubre la demanda de varias comunas, es difícil determinar la cantidad de personal destinado a los delitos de un sector reducido como lo es el de la zona en cuestión.

⁷Asesoría Urbana Municipal. Fundación CONAMA 9. Madrid, España. [Fecha de consulta: 02 Agosto 2011]. Disponible en: <http://www.conama9.org>.

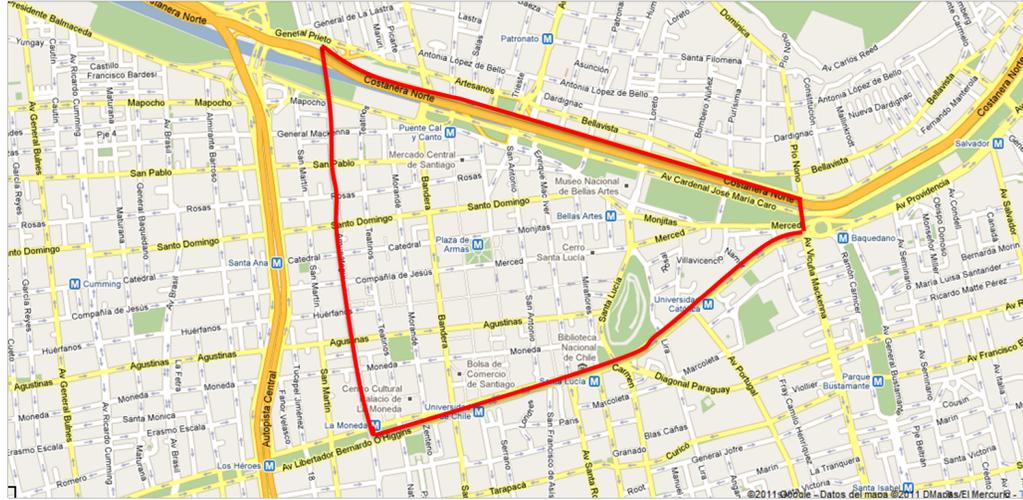


Figura 2.9: Mapa Primera Comisaría de Santiago
Fuente: *Google Maps*

Por último, el Programa de Seguridad Comunal de Santiago⁸ contempla servicios de Guardias Municipales que recorren las calles de esta comuna, concentrándose en los puntos más conflictivos, tales como el Paseo Huérfanos, entre Mac-Iver y Teatinos, y el área comprendida por Ahumada, Nueva York, La Bolsa y La Unión.

En la actualidad existen 16 Guardias Municipales, los que actúan coordinados con carabineros de la Primera Comisaría de Santiago, y asisten sus labores enfrentando el comercio ambulante ilegal, reprimiendo hurtos y robos, y recuperando especies sustraídas.

Es importante mencionar que la comuna de Santiago está compuesta por 16 cuadrantes de los cuales sólo tres pertenecen a la Primera Comisaría (Figura 2.10). Los cuadrantes [9] son regiones menores en las cuales se divide cada comisaría con el objetivo de reducir la victimización y la sensación de temor de la ciudadanía por medio de una mayor eficacia y eficiencia del actuar policial.

2.3.3. Estadísticas y Análisis de Datos del Sector

Como se mencionó anteriormente, para el desarrollo de esta tesis se cuenta con una base de datos de denuncias de delitos de la Primera Comisaría de la comuna de Santiago ocurridos entre enero de 2001 y diciembre de 2004. En esta base de datos se especifican varios detalles del delito ocurrido, tales como el cuadrante donde sucedió la transgresión, fecha, ubicación, hora, tipo de delito y valoración monetaria del crimen, entre otros.

Al analizar estos datos, se puede ver (Tabla 2.1) que el delito más denunciado en el sector de la Primera Comisaría de Santiago es el Hurto con más del 45 % de los registros. El segundo delito más denunciado es el Robo con Violencia (34 %).

Por otro lado, se observa de la Tabla 2.2 que los delitos se concentran en el segundo cuadrante, presentando más del 55 % de los registros, lo que posiciona a esta zona como la más atractiva de

⁸Más información en <http://www.municipalidaddesantiago.cl/seguridad/seguridad.php>. [Fecha de consulta: 02 Agosto 2011].

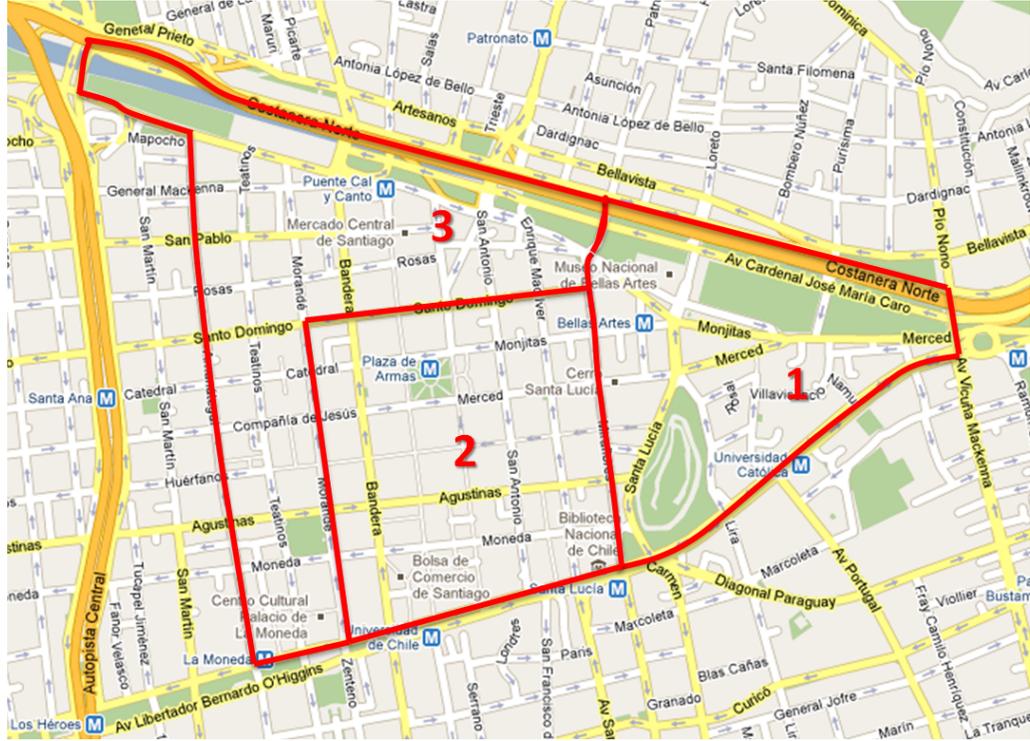


Figura 2.10: Mapa Cuadrantes en la Primera Comisaría de Santiago
Fuente: Google Maps

la Primera Comisaría para los delincuentes. Notar que el tamaño de los cuadrantes son bastante similares, por lo que se sugiere que este efecto es resultado de las características singulares de esta área.

Delito	Frecuencia	Porcentaje
Hurtos	11.529	46 %
Robo con Violencia	8.549	34 %
Robo con Fuerza	2.462	10 %
Lesiones	2.316	9 %
Otros	340	1 %
TOTAL	25.144	100 %

Tabla 2.1: Porcentaje Tipo de Delitos

	Frecuencia	Porcentaje
1er Cuadrante	5.107	20,7 %
2do Cuadrante	13.622	55,1 %
3er Cuadrante	5.977	24,2 %
TOTAL	24.706	100 %

Tabla 2.2: Denuncias Según Cuadrante

Un resultado bastante interesante es el que se obtiene al analizar los horarios, días, meses y trimestres en los que se notificaron las denuncias. De la Figura 2.11 se observa que el día de la semana en que más delitos ocurren es el viernes (19%), mientras que el día con menor frecuencia es el domingo (7%).

Otro resultado muy llamativo es el que se observa al estudiar lo ocurrido según el día del mes. De la Figura 2.12, se puede ver que el primer día del mes es el que presenta una mayor cantidad de denuncias, duplicando las notificaciones de cualquier otro día.

En cuanto a la dinámica delictual durante el año, se observa en la Figura 2.13 que la proporción de denuncias según el mes en que fueron notificadas no tiene una variación muy importante a lo

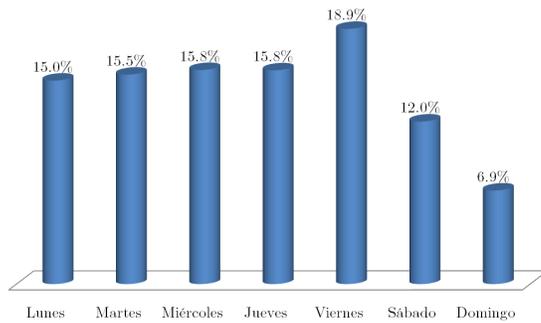


Figura 2.11: Denuncias Según Día de la Semana
Fuente: *Elaboración Propia*

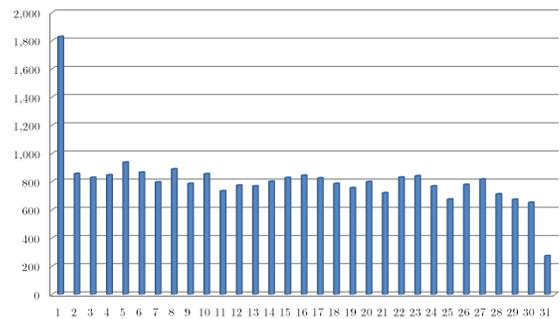


Figura 2.12: Denuncias Según Día del Mes
Fuente: *Elaboración Propia*

largo del año. Los meses con más denuncias son Octubre con el 9,2% del total y Abril con el 9%, mientras que los meses con menor cantidad de denuncias son Febrero y Septiembre, con el 6,8% y 7,3% respectivamente.

Por otra parte, la Figura 2.14 da cuenta de lo que ocurre con los delitos en los diferentes rangos horarios. Cerca del 30% de éstos ocurren entre las 12:00 y las 16:00, y el 25% entre las 16:00 y las 20:00. Luego, se puede decir que más del 55% de los crímenes suceden entre estas 8 horas. El intervalo que presenta menor cantidad de denuncias es entre las 4 y 8 a.m.

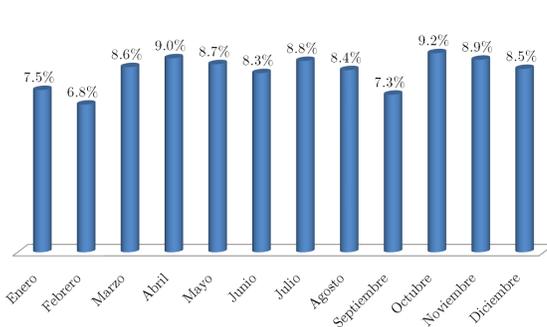


Figura 2.13: Denuncias Según Mes
Fuente: *Elaboración Propia*

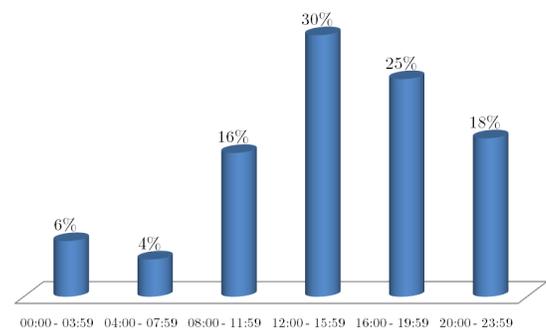


Figura 2.14: Denuncias Según Rango Horario
Fuente: *Elaboración Propia*

En cuanto a lo ocurrido en los diferentes semestres y trimestres, se percibe que los valores son bastante similares, por lo que se concluye que no hay diferencias importantes entre estos periodos de tiempo.

Por último, al analizar el monto avaluado de las denuncias, se observa que a menor valor monetario, mayor cantidad de notificaciones. De la Figura 2.15 se observa que más del 40% de las denuncias se encuentran entre los 10.000 y 50.000 pesos, 25% entre los 50.000 y los 100.000 y el 15% entre los 100.000 y los 250.000. El 19% restante se reparte en montos superiores a lo anteriormente mencionado.

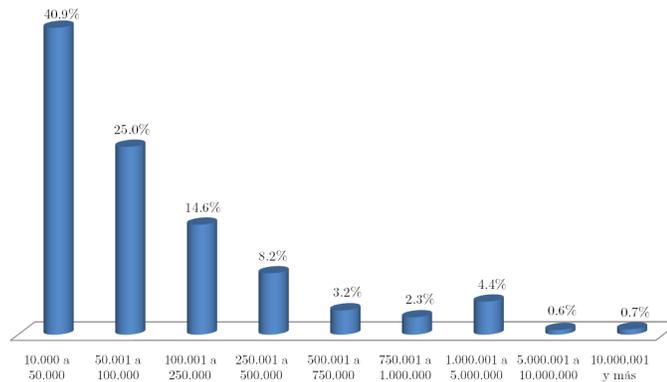


Figura 2.15: Denuncias Según Monto Avaluado
Fuente: *Elaboración Propia*

2.3.4. Tendencias Espaciales de los Delitos

Al realizar un análisis espacial de los datos estudiados en la sección anterior, se pueden advertir situaciones muy singulares que valen la pena mencionar.

La Figura 2.16 ilustra los mapas de densidad de los diferentes tipos de delitos, exceptuando los hurtos, para el periodo estudiado. Se pueden observar distintos patrones espaciales de los crímenes denunciados, concentrándose en diferentes puntos calientes y con una clara diferencia en la densidad de los sucesos.

En cuanto a la distribución espacial de los hurtos, se puede observar en la Figura 2.17 una actividad delictual bastante alta en la esquina del Paseo Ahumada con la Alameda, apreciándose otros tres puntos importantes de concentración de este tipo de delitos a lo largo de este paseo peatonal coincidiendo con las calles Moneda, Huérfanos y Rosas. Se percibe otra zona de intensidad criminal en la Alameda entre las calles San Antonio y Enrique Mac Iver. Lo anterior sugiere una relación evidente entre los *hot spots* y la afluencia de peatones.

Al analizar lo que ocurre con los hurtos en los diferentes intervalos del día, se advierten unos pocos *hot spots* que van desplazándose a lo largo de la jornada. De la Figura 2.18 se observa que en la madrugada, desde las 00:00 y las 03:59, los puntos más conflictivos son las esquinas del Paseo Ahumada con la Alameda y Enrique Mac Iver con la Alameda. Entre las 04:00 y las 07:59 los hurtos se trasladan a Plaza Italia, volviendo al Paseo Ahumada con la Alameda en el intervalo sucesivo. A continuación, entre las 12:00 y las 15:59 se produce una conglomeración de estos delitos a lo largo del Paseo Ahumada hasta la Plaza de Armas, concentrándose en las esquinas de Paseo Ahumada con la Alameda y Paseo Puente con Rosas entre las 16:00 y las 20:00. Al final del día, se observa un único punto caliente en la conflictiva esquina definida por las calles Paseo Ahumada y Alameda.

Por otro lado, al estudiar lo que sucede con los diferentes avalúos de los montos sustraídos, se observa en la Figura 2.19 que cuando este valor se encuentra en el rango definido entre los \$5.000 y los \$250.000, entonces los hurtos se concentran a lo largo del Paseo Ahumada, mientras que cuando la valoración es mayor se amplía la zona de riesgo en los alrededores directos de esta conflictiva calle.

Finalmente, se aprecia en la Figura 2.20 que no hay diferencias muy notorias entre los días Lunes a Jueves y Viernes a Domingo. Sin embargo, se observa un aumento importante en la actividad

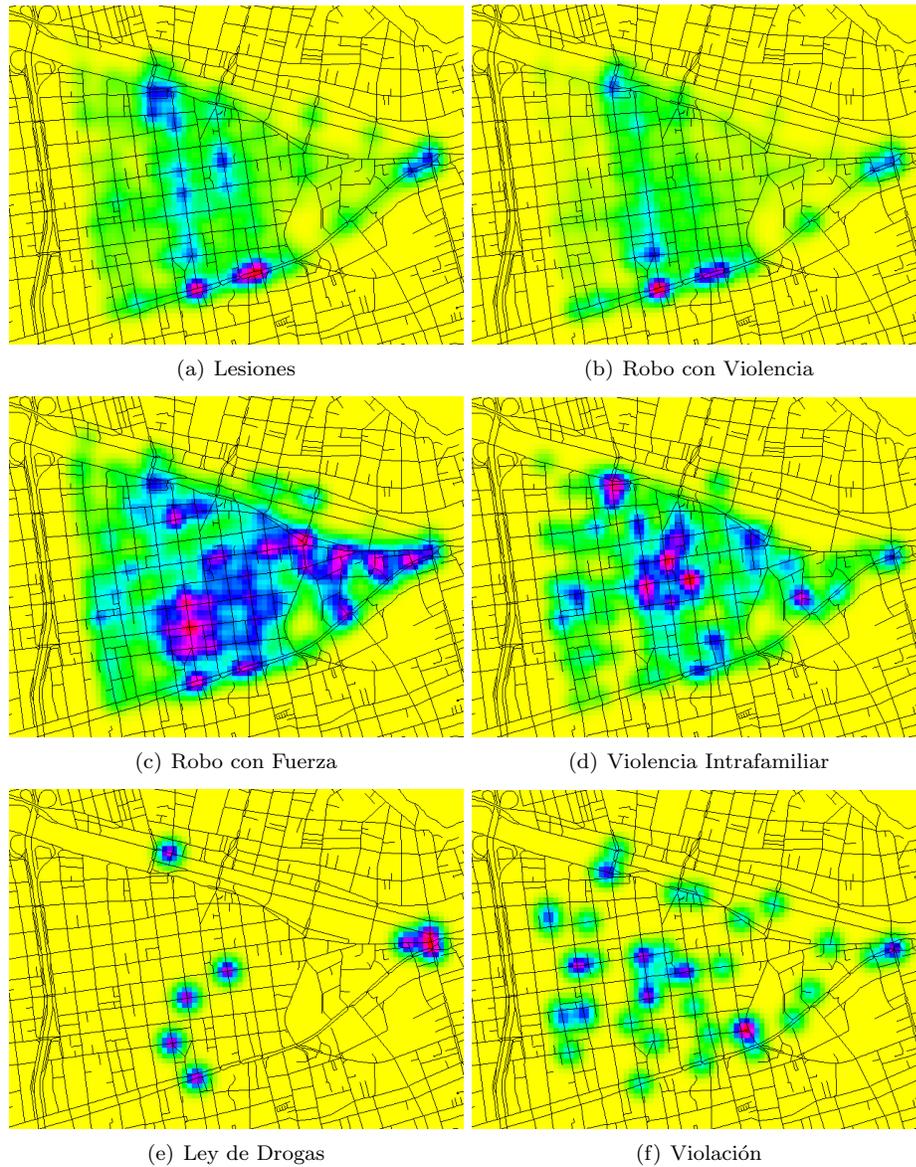


Figura 2.16: Mapas de Densidad de Delitos en la 1^a Comisaría de Santiago 2001 - 2004
Fuente: *Elaboración Propia*

delictual en la esquina de Paseo Puente y Rosas durante los últimos días de la semana.

2.4. Revisión Bibliográfica

En esta sección se revisa la bibliografía relacionada con los temas abordados en esta tesis. Primero se detalla la revisión bibliográfica relacionada con los primeros estudios en criminología cuantitativa y a los diferentes enfoques que se están investigando en la actualidad. Luego, se realiza un repaso de estudios y trabajos en los que se han desarrollado modelos de elección discreta. Finalmente, se presentan algunas publicaciones de criminología que basan sus estudios en modelos de este tipo,

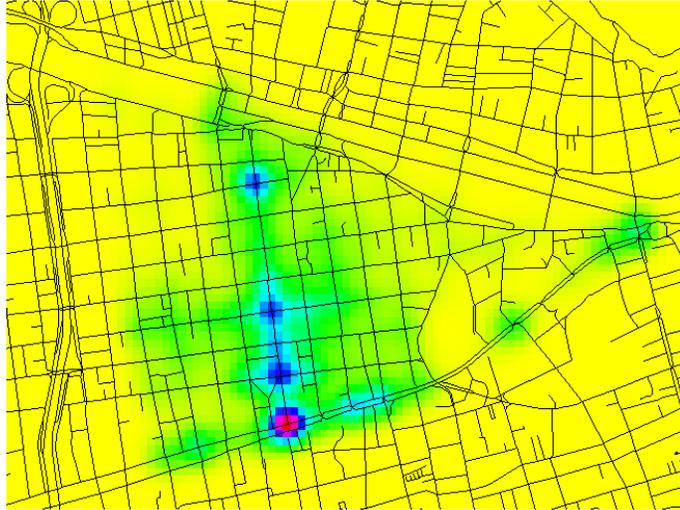


Figura 2.17: Mapas de Densidad de Hurtos en la 1ª Comisaría de Santiago 2001 - 2004
Fuente: *Elaboración Propia*

específicamente en regresiones logísticas.

2.4.1. Criminología Cuantitativa

Orígenes de la Criminología Cuantitativa

Los primeros estudios en criminología surgieron del ámbito cualitativo y carecieron de factores cuantitativos. Los primeros intereses en aplicar una metodología cuantitativa en el crimen se encargaba de recolectar estadísticas generales del suceso, lo que dio paso a un área muy desarrollada del análisis econométrico.

Uno de los pioneros en esta área fue Adolphe Quételet quien centró sus investigaciones relacionadas a la criminología en factores sociales y de entorno. Fue conocido, entre otras cosas, por su gran aporte a la Física Social. En una de sus publicaciones [10] menciona sus tres conclusiones fundamentales del acto delictual: (i) el delito es un fenómeno social que puede conocerse y determinarse estadísticamente, (ii) los delitos se cometen año a año con absoluta regularidad y precisión, y (iii) los factores que influyen como causas de la actividad delictuosa son, entre otros, el clima, la pobreza, la miseria y el analfabetismo.

La introducción de la variable espacial al área de la criminología se debe al abogado André-Michel Guerry quien basó sus principales investigaciones en los estudios espaciales de los crímenes en Francia [11]. Sus aportes en este campo inspiraron el desarrollo de la georreferenciación y la inclusión de los atributos espaciales en esta nueva área de investigación.

Criminología Cuantitativa en Tiempos Modernos

Actualmente, la gran mayoría de las publicaciones en el área de la criminología se apoyan en datos estadísticos. Además, gracias al gran desarrollo tecnológico y computacional, se tienen herramientas

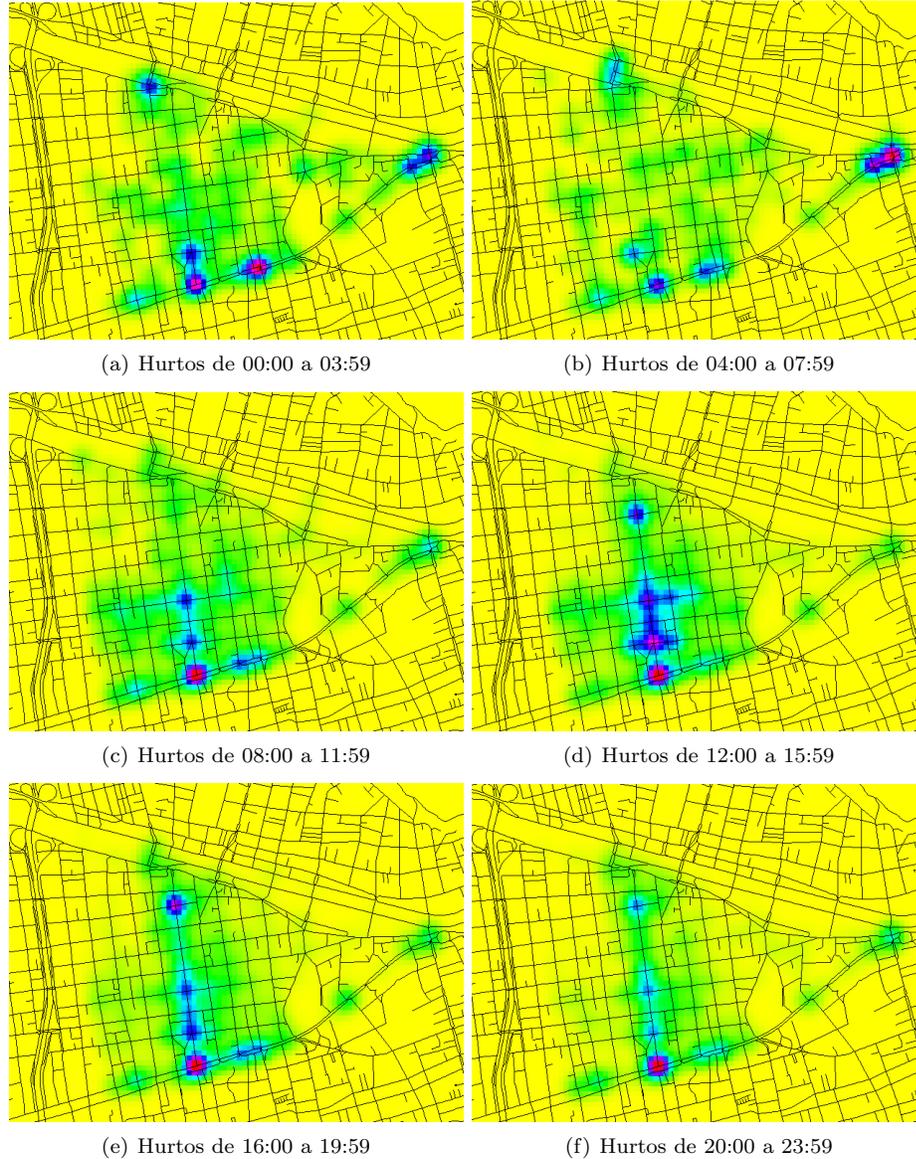


Figura 2.18: Mapas de Densidad de Hurtos Según Hora en la 1^a Comisaría de Santiago 2001 - 2004
Fuente: *Elaboración Propia*

que apoyan de manera considerable los estudios econométricos desde un punto de vista espacial y multivariado.

Existen diversos trabajos y modelos relacionados con la criminalidad. Muchos de ellos nacen desde diferentes perspectivas del problema, por lo que la forma de resolverlo puede ser bastante disímil.

- Minería de Datos

Cuando se habla del *predictive crime mapping* se refiere al estudio de la distribución de los datos, específicamente de la disposición espacial de los mismos, por lo que un enfoque matemático bastante utilizado es el de *Data Mining*, técnica que busca extraer información

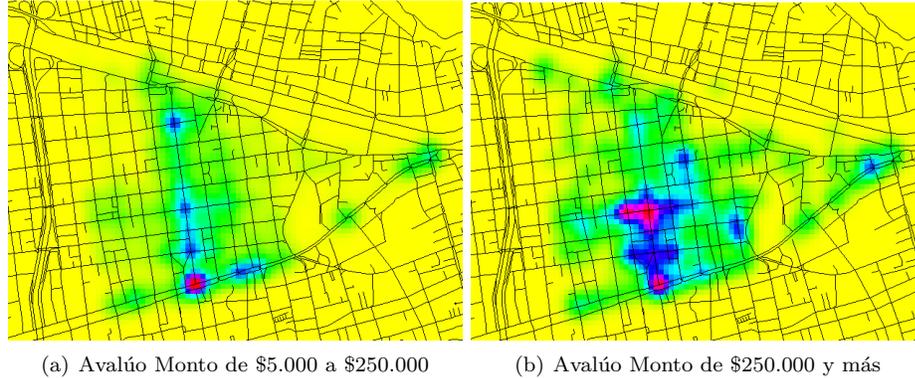


Figura 2.19: Mapas de Densidad de Hurtos Según Monto Avaluado 1^a Comisaría de Santiago 2001 - 2004

Fuente: *Elaboración Propia*

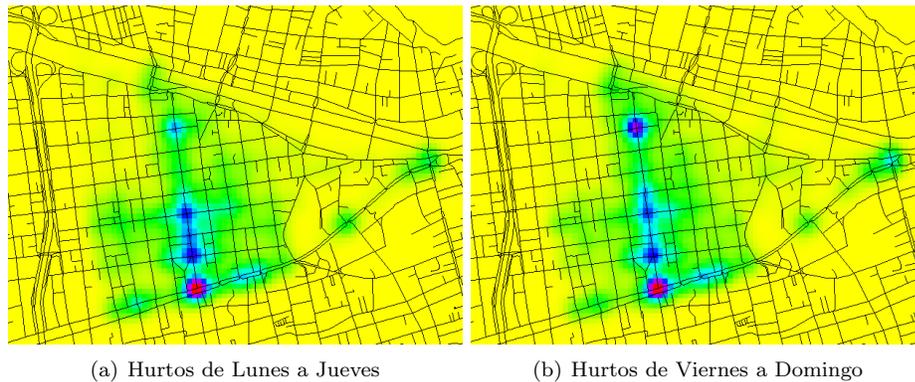


Figura 2.20: Mapas de Densidad de Hurtos Según Día de la Semana Avaluado 1^a Comisaría de Santiago 2001 - 2004

Fuente: *Elaboración Propia*

relevante de los datos que se tienen.

Múltiples técnicas de minería de datos han sido utilizadas en criminología, las cuales pueden clasificarse en relación a lo que persiguen: predicción, clustering o clasificación.

Uno de los métodos de predicción más común es el denominado *Hot Spots* en el que se asume que los puntos con mayor actividad delictual persisten durante varios periodos, por lo que utilizan datos correspondientes a series de tiempo de los delitos. Existen varios exponentes de esta rama, destacándose Spelman [12] y Adams-Fuller [13].

Una metodología predictora más actual de la minería de datos son las redes neuronales artificiales con las cuales se busca pronosticar diferentes situaciones delictuales en base a una metodología de aprendizaje y testeo [14].

En cuanto a la técnica de clustering, se tienen varios métodos para formar las agrupaciones. El primero y más conocido es el análisis de densidad de puntos para construir los clusters [15]. Otra técnica de clustering que se ha empleado con mucho éxito es el algoritmo *k*-medias [16].

Por último, las técnicas de clasificación más conocidas son las Support Vector Machines [17] y las redes bayesianas [18], entre otras.

- Teoría de Juegos

Otro método bastante usado en la literatura es la aplicación de Teoría de Juegos debido a que la lógica detrás de esta herramienta se ajusta bastante bien a la problemática en cuestión.

La teoría de juegos es un área de la matemática aplicada que utiliza modelos para estudiar la interacción entre dos o más agentes con intereses opuestos. En el caso particular de la criminología, se pueden simular diversas relaciones, tales como criminal-víctima, policía-criminal o policía-criminal-víctima, entre otros.

Muchos modelos de teoría de juegos se han desarrollado en torno al tema de la seguridad, tales como el trabajo realizado en el aeropuerto de Los Angeles USA (LAX) y en el Servicio Federal de Policía Aérea USA (FAMS) [19].

- Simulación Computacional

La tecnología y la computación han permitido el desarrollo de herramientas de simulación cruciales para el estudio del fenómeno delictual.

La simulación computacional es el proceso de diseñar un modelo virtual que emula un sistema real, para luego realizar experimentos y comprender el comportamiento del sistema estudiado.

La ventaja de utilizar esta herramienta es que se evitan situaciones en la vida real que podrían ser muy difíciles de reproducir y/o muy costosas.

Una de las técnicas de simulación que más se está usando en el área de la criminología son los modelos basados en agentes, los cuales combinan elementos de teoría de juegos, ecuaciones diferenciales y sistemas complejos. Lo anterior se ha empleado para el análisis del impacto de diferentes estrategias preventivas de robos y asaltos en Gran Bretaña [20] y en los Países Bajos [21].

2.4.2. Modelos de Elección Discreta

Los modelos de elección discreta involucran la preferencia entre dos o más alternativas. Estos modelos relacionan estadísticamente la opción elegida con las características de las alternativas a considerar.

Diversos autores han desarrollado estos modelos para el análisis de diferentes fenómenos. McFadden [22] y Hensher [23] estudiaron la elección del modo de viaje en los trayectos urbanos. Schmidt y Strauss [24] y Boskin [25] analizaron las opciones profesionales entre múltiples alternativas. Por otro lado, Terza [26] ha estudiado la asignación de calificaciones de bonos corporativos como una elección entre muchas alternativas de bonos.

Estos modelos originalmente se utilizaban en economía para estudiar el comportamiento de los individuos [27]. Otras áreas en las que se han aplicado con gran efectividad es en marketing [28], transporte [29, 30] y uso de suelos [31].

En cuanto a la utilización de estos modelos en el área de la criminología, la literatura es bastante amplia y diversa. Xue y Brown [32] desarrollaron una regresión logística con el objetivo de identificar las características socio demográficas más importantes que determinan una alta tasa de robo a casas.

Por otra parte, Bernasco [33] responde a la pregunta de cómo los criminales seleccionan las casas que van a robar en base a un modelo logit condicional que considera variables socio demográficas y de entorno, tales como el valor de la casa, rotación residencial, heterogeneidad étnica, proximidad al centro de la ciudad y unidades residenciales, entre otros.

Capítulo 3

Modelo Matemático

Contenido

3.1. Modelos Lineales	28
3.2. Modelos de Elección Discreta	30
3.2.1. Definiciones Previas	30
3.2.2. Regresión Logística Binaria	32
3.2.3. Método de Máxima Verosimilitud	33
3.2.4. Significación Estadística de los Estimadores	34
3.2.5. Medidas de Bondad del Ajuste	34
3.2.6. Interpretación de los Estimadores	36

Muchas son las situaciones en las que existe una relación inherente entre dos o más variables y resulta necesario explorar la naturaleza de esta relación. Existen diversas técnicas y herramientas estadísticas que permiten modelar esta situación, de las cuales una muy común es el análisis de regresión lineal. Esta técnica se desempeña relativamente bien cuando la variable que se desea explicar es continua. Sin embargo, cuando esta condición no se cumple pueden presentarse varios problemas, razón que motiva el uso de modelos de regresión no lineales específicamente pensados para explicar variables categóricas.

3.1. Modelos Lineales

Un modelo de regresión múltiple, que no necesariamente es lineal, permite explicar el comportamiento de una variable dependiente Y en función de una serie de variables independientes X_1, X_2, \dots, X_k y de un término de perturbación u , tal y como se muestra en la ecuación 3.1.

$$(3.1) \quad Y = f(X_1, X_2, \dots, X_k, u)$$

En el caso particular de que el modelo de regresión sea lineal, se tiene una expresión de la forma que se detalla en la ecuación 3.2.

$$(3.2) \quad Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k + u$$

El objetivo de la regresión es estimar los parámetros que acompañan a las variables predictoras de forma que el modelo resultante se ajuste lo mejor posible a las observaciones. Cuando la variable dependiente Y es continua, resulta frecuente utilizar un modelo de regresión lineal múltiple como el anterior. En tal caso, la estimación de los parámetros $\beta_0, \beta_1, \dots, \beta_k$ se lleva a cabo mediante el método de Mínimos Cuadrados.

Por otro lado, puede ocurrir que la variable dependiente Y sea una variable cualitativa, es decir, que sólo pueda tomar un conjunto reducido de valores. En tales circunstancias, el modelo de regresión lineal presenta una serie de inconvenientes serios, por lo que es necesario recurrir a otros modelos tales como la regresión logística. A diferencia del modelo lineal, en la regresión logística se emplea el método de Máxima Verosimilitud para llevar a cabo la estimación de los parámetros del modelo.

En muchas áreas de las ciencias sociales, estos modelos de regresión con variable endógena categórica suelen emplearse para explicar la decisión Y que toma un individuo a partir de un conjunto de variables explicativas exógenas. Es por este motivo que los modelos de variable endógena cualitativa son también llamados modelos de elección discreta.

En esta tesis se desea estudiar una variable dependiente binaria, es decir, que toma valores 0 o 1 según una condición dada. Luego, siempre se cumplirá que:

$$(3.3) \quad E[Y] = 0 \cdot P(Y = 0) + 1 \cdot P(Y = 1) = P(Y = 1)$$

Bajo el supuesto habitual de que los términos de perturbación se distribuyen de forma normal con media nula, es decir $E[u] = 0$, y suponiendo conocidos los valores que toman las variables explicativas, se tiene que al reemplazar la ecuación 3.2 en la expresión anterior se obtiene lo que se conoce como modelo de probabilidad lineal.

$$(3.4) \quad P(Y = 1) = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k = Y - u$$

Luego, este modelo inicial no es válido para explicar el comportamiento de variables dependientes binarias debido a que presenta varios problemas e incongruencias, algunas más serias que otras. Dos de estos problemas están relacionados con el término de perturbación, los cuales hacen referencia al no cumplimiento de una distribución normal y de la hipótesis de homocedasticidad. Sin embargo, las incongruencias más graves se encuentran al ver lo que sucede con la variable dependiente.

Como bien se menciona anteriormente, la variable dependiente sólo puede tomar los valores 0 y 1, por lo que al estimar los parámetros del modelo de probabilidad lineal se estarán ajustando diferentes puntos distribuidos en el plano a una recta definida por $Y = 0$ e $Y = 1$. El uso de dicha recta para predecir los nuevos valores de Y a partir de valores dados de X_i , puede proporcionar valores mayores que 1 o menores que 0, lo cual está en clara contradicción con la definición de probabilidad.

Un segundo problema bastante importante es el que se obtiene al derivar la ecuación 3.4. El resultado de esto, que se observa en la ecuación 3.5, revela que la variación en $P(Y = 1)$ causada por cambios en alguna de las variables explicativas es constante y, por tanto, independiente del valor actual de dicha variable explicativa, lo cual es una hipótesis muy poco realista.

$$(3.5) \quad \frac{dP(Y = 1)}{dX_i} = \beta_i, \forall i = 0, \dots, k$$

Dados los inconvenientes mencionados, se resuelve utilizar un modelo de elección discreta que se ajuste a las características del problema.

3.2. Modelos de Elección Discreta

Muchos modelos de este tipo han surgido en las últimas dos décadas como herramientas para el análisis de la localización de eventos delictuales. Dadas las características del problema a resolver, es necesario identificar un modelo matemático que considere las características físicas y de entorno de las diferentes alternativas que el delincuente enfrenta y la atracción que cada opción ejerce sobre él.

Estos modelos involucran la preferencia entre dos o más alternativas. Relacionan estadísticamente la opción elegida por cada individuo a las características de éste y a los atributos de las alternativas, por lo que son ideales para este tipo de casos.

Los modelos de elección discreta pueden tomar muchas formas, entre las cuales están los modelos de Regresión Logística Binaria, Regresión Logística Multinomial, Regresión Logística Condicional, Regresión Logística Anidada, Regresión Logística Mixta, Probit Binario, Probit Multinomial y los modelos Generalizados de Valores Extremos. Todos estos modelos tienen características comunes, las cuales se detallan a continuación.

3.2.1. Definiciones Previas

Set de Alternativas

El set de alternativas es el conjunto de opciones que están disponibles para el individuo que debe tomar la decisión, que en este caso es el delincuente. Para un modelo de elección discreta, este set de alternativas deben cumplir ciertos requisitos.

Primero, este conjunto debe ser exhaustivo, por lo que debe incluir todas las opciones posibles.

La segunda exigencia es que las alternativas que componen este set deben ser mutuamente excluyentes, es decir, si se elige una alternativa entonces no se pueden elegir otra. Este requisito implica que la persona elige sólo una alternativa de todo el conjunto de opciones consideradas.

El tercer y último requerimiento es que este conjunto debe estar compuesto por un número finito de opciones.

Probabilidad de Elección

El modelo de elección discreta especifica la probabilidad de que una persona elija una alternativa en particular en función de las variables observadas que se relacionan con estas opciones y con la persona en cuestión.

En su forma general, la probabilidad de que la persona n -ésima elija la alternativa i se expresa como se detalla a continuación.

$$P_{ni} = G(X_{ni}, X_{nj} \forall j \neq i, S_n, \beta)$$

En esta ecuación,

X_{ni} : vector de atributos de la alternativa i que enfrenta el individuo n ,

$X_{nj} \forall j \neq i$: vector de atributos de las alternativas distintas de i que enfrenta el individuo n ,

S_n : vector de características del individuo n ,

β : conjunto de parámetros que relacionan las distintas variables con las probabilidades, los cuales son estimados estadísticamente.

Como ya se mencionó, la componente S_n representa las características del individuo, y es, por lo tanto, la misma para todas las opciones. En este análisis se supondrá que este término no interactúa con los atributos y que tiene una relación lineal con la probabilidad, por lo que, dadas las condiciones anteriores, sale de la ecuación.

Se tiene entonces que la probabilidad de elegir un lugar en específico no depende de las características del individuo. Por consiguiente, la probabilidad anteriormente mencionada se puede expresar como se detalla en la ecuación 3.6.

$$(3.6) \quad P_{ni} = G(X_{ni}, X_{nj} \forall j \neq i, \beta)$$

Dependiendo del modelo de elección discreta en particular con el que se trabaje, entonces la función G será diferente.

Utilidad del Individuo

Para entender adecuadamente el comportamiento de los delincuentes, se asumirá que estos individuos son racionales, es decir, siempre buscan maximizar sus utilidades. Luego, cuando se enfrentan varias opciones, se elige la alternativa que le ofrece el mayor beneficio.

Sea U_{ni} la utilidad o beneficio neto que la persona n obtiene al elegir la alternativa i . La elección de esta persona estará representada por una variable *dummy* para cada alternativa, y_{ni} . Como se observa en la ecuación 3.7, esta variable toma el valor 1 cuando la utilidad de una alternativa es mayor que la de otra alternativa, mientras que es nula en caso contrario.

$$(3.7) \quad y_{ni} = \begin{cases} 1, & \text{si } U_{ni} > U_{nj} \forall i \neq j \\ 0, & \text{si } U_{ni} \leq U_{nj} \end{cases}$$

Luego, la probabilidad de que una persona elija una alternativa particular se determina mediante la comparación del beneficio obtenido al elegir esa opción con otras alternativas.

$$\begin{aligned} P_{ni} &= P(y_{ni} = 1) \\ &= P(U_{ni} > U_{nj} \forall i \neq j) \\ &= P(U_{ni} - U_{nj} > 0 \forall i \neq j) \end{aligned}$$

Así, la probabilidad de elección sólo depende de la diferencia de utilidades entre las alternativas, no del nivel absoluto de las mismas. Es por esta razón que al añadir una constante a las utilidades de todas las opciones no cambia las probabilidades de elección.

Como ya se ha mencionado, la elección del individuo depende de muchos factores, algunos de los cuales el investigador observa y otros que no advierte. Por lo tanto, si se asume linealidad en la función de beneficio entonces ésta puede descomponerse en dos partes, una que depende de los atributos percibidos y otra de las variables no observadas.

$$U_{ni} = \beta_n X_{ni} + \varepsilon_{ni}$$

En esta ecuación,

X_{ni} es el vector de variables observadas relativas a la alternativa i por el individuo n , β_n es el vector de coeficientes que acompañan a las variables observadas, y ε_{ni} captura el impacto de los factores no observados que influyen en la decisión.

Por consiguiente, la probabilidad de elección se puede expresar como se muestra en la siguiente ecuación.

$$\begin{aligned} P_{ni} &= P(y_{ni} = 1) \\ &= P(U_{ni} > U_{nj} \quad \forall i \neq j) \\ &= P(\beta_n X_{ni} + \varepsilon_{ni} > \beta_n X_{nj} + \varepsilon_{nj} \quad \forall i \neq j) \\ &= P(\varepsilon_{nj} - \varepsilon_{ni} > \beta_n X_{ni} - \beta_n X_{nj} \quad \forall i \neq j) \\ &= P(\varepsilon_{nj} - \varepsilon_{ni} > \beta_n (X_{ni} - X_{nj}) \quad \forall i \neq j) \end{aligned}$$

Así, una vez determinados los coeficientes β_n , se tiene que la probabilidad de elección es la probabilidad de que los términos aleatorios $\varepsilon_{nj} - \varepsilon_{ni}$ sean menores que $\beta_n (X_{ni} - X_{nj}) \quad \forall i \neq j$.

Diferentes modelos de elección, es decir, diferentes especificaciones de la función G , surgen de distintas distribuciones de ε_{ni} para todo i y diversos tratamientos de los coeficientes β_n .

3.2.2. Regresión Logística Binaria

La regresión logística es uno de los muchos modelos de elección discreta que existen. Éste es un modelo estadístico en el que se desea conocer la relación entre una variable dependiente cualitativa y una o más variables explicativas independientes.

En este caso, la variable dependiente es una variable dicotómica que toma los valores 0 ó 1 según la ausencia o presencia de actividad delictual. Por lo tanto, se desarrolla un modelo de Regresión Logística Binaria. Luego, la forma de la función G para este caso viene dado por la ecuación 3.8:

$$(3.8) \quad G(X_{ni}) = \frac{\exp(X_{ni})}{1 + \exp(X_{ni})}$$

Además, se asume que los términos no observados, ε_n , tienen una distribución logística, es decir, cumplen con la condición especificada en 3.9.

$$(3.9) \quad F(\varepsilon_n) = \exp(e^{-\varepsilon_n})$$

Luego, las ecuaciones que definen un modelo de regresión logística binaria son las presentadas a continuación.

$$(3.10) \quad P(y_i = 0|X_i) = \frac{1}{1 + \exp(\beta_0 + \sum_k \beta_k x_{ik})}$$

$$(3.11) \quad P(y_i = 1|X_i) = \frac{\exp(\beta_0 + \sum_k \beta_k x_{ik})}{1 + \exp(\beta_0 + \sum_k \beta_k x_{ik})}$$

Donde $P(y_i = 0|X_i)$ es la probabilidad de que no ocurra un delito en la alternativa i en presencia de los atributos X_i y $P(y_i = 1|X_i)$ es la probabilidad de que sí ocurra un delito en esa zona. Por otro lado, β_0 es la constante del modelo y β_k son los coeficientes que acompañan a las variables independientes.

Dado que no se tiene información relativa a los delincuentes, se asume que todos estos individuos tienen características similares, por lo que el índice n ha sido eliminado de las ecuaciones.

Para la estimación de los coeficientes del modelo y de sus errores estándar se recurre al cálculo de estimaciones de máxima verosimilitud [34], es decir, estimaciones que hagan máxima la probabilidad de obtener los valores de la variable dependiente. Estas estimaciones no son de cálculo directo, por lo que se emplean métodos iterativos, tales como el de Newton–Raphson¹. De estos métodos surgen no sólo las estimaciones de los coeficientes de regresión, sino también de sus errores estándar y de las covarianzas entre las variables del modelo.

Dado que el cálculo de estas estimaciones es bastante complejo, hay que recurrir al uso de rutinas de programación o a paquetes estadísticos. En este trabajo se utilizará el software matemático MATLAB².

3.2.3. Método de Máxima Verosimilitud

A diferencia de la regresión lineal, que hace uso de los métodos de estimación por Mínimos Cuadrados, en la regresión logística se emplean los métodos de Máxima Verosimilitud para llevar a cabo la estimación de los parámetros del modelo.

Este método consiste en hallar los valores de β que hacen más factible la probabilidad de elección de una alternativa. Estos valores se definen mediante una función de verosimilitud, la cual debe ser lo más grande posible.

Se define la función de verosimilitud que caracterizará el problema a resolver en la ecuación 3.12.

$$(3.12) \quad L = \prod_m \prod_i P_i^{d_{mi}}(\beta, X_i)$$

En la ecuación, $P_i = P(y_i = 1|X_i)$ representa la probabilidad de que la alternativa i sea escogida dadas sus características, y

¹Mayor detalle de este método en el Anexo A.

²Más información en www.mathworks.com/products/matlab/. [Fecha de consulta: 02 Agosto 2011].

$$d_{mi} = \begin{cases} 1, & \text{si el delito } m\text{-ésimo ocurrió en la alternativa } i \\ 0, & \text{en caso contrario} \end{cases}$$

Habitualmente se trabaja con el logaritmo neperiano de la verosimilitud, tal como se muestra en la ecuación 3.13, que es la función que hay que maximizar.

$$(3.13) \quad \ln[L] = \sum_m \sum_i d_{mi} \ln[P_i(\beta, X_i)]$$

Finalmente, el problema que se desea resolver es el siguiente³:

$$(3.14) \quad \text{Max}_{\beta} \ln[L] = \sum_m \sum_i d_{mi} \ln[P_i(\beta, X_i)]$$

3.2.4. Significación Estadística de los Estimadores

Como se menciona en uno de los apartados anteriores, la estimación de los coeficientes del modelo y de sus errores estándar se lleva a cabo mediante el cálculo de estimaciones de máxima verosimilitud. Para el cálculo de estas estimaciones se recurre a métodos iterativos, de los cuales surgen no sólo las estimaciones de los coeficientes de regresión, sino también de sus errores estándar y de las covarianzas entre las variables independientes del modelo.

El siguiente paso es comprobar la significación estadística de cada uno de los coeficientes obtenidos del modelo de regresión logística. Para ello se emplea el estadístico de Wald, el cual se obtiene al comparar el parámetro de la estimación de máxima verosimilitud con la estimación de su error estándar. Este *test* contrasta la hipótesis que plantea que el parámetro obtenido es no nulo y que sigue una distribución normal de media 0 y varianza 1. Luego, la obtención de significación para este estadístico indica que dicho coeficiente es no nulo y merece la pena su conservación en el modelo.

Su valor viene dado por el cociente entre el valor del coeficiente y su correspondiente error estándar, tal y como se aprecia en la ecuación 3.15.

$$(3.15) \quad W(\beta_j) = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}$$

Donde $\hat{\beta}_j$ es el coeficiente estimado que acompaña al atributo j y $SE(\hat{\beta}_j)$ es la desviación estándar de este coeficiente.

3.2.5. Medidas de Bondad del Ajuste

Una de las características técnicas que determinan la factibilidad de los resultados de un modelo es su grado de reproducibilidad. Ésta se refiere al hecho de que los resultados obtenidos en una

³Mayor detalle de la función de verosimilitud en el Anexo B.

determinada ocasión, bajo ciertas condiciones, deberían ser similares si se volviese a medir el mismo rasgo en condiciones idénticas. Este aspecto de la exactitud con que un modelo mide lo que se pretende medir es lo que se denomina la *confiabilidad* de la medida.

Específicamente, la confiabilidad estadística es la probabilidad de que un estimador se encuentre en un cierto intervalo elegido de antemano. El error estadístico es la probabilidad de que el estimador no se encuentre en este intervalo. Así, si el error deseado es, por ejemplo, del 5%, entonces la confiabilidad estadística es de 95%.

Para que los resultados de un instrumento puedan ser interpretables, es decir, que tengan significado, es necesario que los mismos sean confiables. No es posible determinar la relación entre dos o más variables si los instrumentos utilizados para medirlas son poco confiables.

A continuación se definen las medidas de confiabilidad que se utilizan para determinar la exactitud del modelo de regresión logística binomial propuesto en esta tesis.

-2 Log Likelihood

Cuanto más coincidan los resultados pronosticados con los datos reales, mejor se ajustará el modelo. Luego, uno de los primeros indicadores de importancia para apreciar el ajuste del modelo logístico es este indicador, el cual corresponde a menos dos veces logaritmo del estadístico de verosimilitud.

Se trata de un estadístico que sigue una distribución similar a χ^2 y compara los valores de la predicción con los valores observados en dos momentos: (i) en el modelo sin variables independientes, es decir, sólo con la constante y (ii) una vez introducidas las variables predictoras. Por lo tanto, el valor de la verosimilitud debiera disminuir sensiblemente entre ambas instancias e, idealmente, tender a cero cuando el modelo predice bien. Luego, cuanto más pequeño sea el valor, mejor será el ajuste.

R Cuadrado de Cox y Snell

Este indicador es un coeficiente de determinación generalizado que se utiliza para estimar la proporción de varianza de la variable dependiente explicada por las variables independientes. La R cuadrado de Cox y Snell es muy similar al estadístico anterior debido a que también se basa en la comparación de los logaritmos de verosimilitud y puede interpretarse como el porcentaje de variación de la variable dependiente que puede ser explicado por la variable independiente. Sus valores oscilan entre 0 y 1 donde valores altos determinan un buen ajuste del modelo.

R Cuadrado de Nagelkerke

Este indicador es una versión corregida de la R cuadrado de Cox y Snell. La R cuadrado de Cox y Snell tiene un valor máximo inferior a 1, incluso para un modelo *perfecto*. La R cuadrado de Nagelkerke corrige la escala del estadístico para cubrir el rango completo de 0 a 1.

Prueba de Hosmer-Lemeshov

Esta es otra prueba para evaluar la bondad del ajuste de un modelo de regresión logística. Se basa en agrupar los casos en deciles de riesgo y comparar la probabilidad observada con la probabilidad

esperada dentro de cada decil. Un modelo puede considerarse aceptable si tanto la especificidad como la sensibilidad tienen un nivel alto de al menos el 75 % [35].

Tabla de Clasificación

Si bien los coeficientes de bondad de ajuste no son del todo fiables, la tabla de clasificación es normalmente el criterio más confiable para indicar la bondad de ajuste del modelo. En esta tabla se muestran los casos bien clasificados en la diagonal principal y los casos mal clasificados en la segunda diagonal. Luego, a mayor porcentaje de aciertos, mejor es el ajuste del modelo.

Área Bajo la Curva ROC

Se entiende como Área Bajo la Curva ROC a la representación de la razón de verdaderos positivos frente a la razón de falsos positivos, según se varía el umbral de discriminación, que es el valor a partir del cual se decide que un caso es un positivo. En otras palabras, se grafica la probabilidad de clasificar correctamente la variable dependiente en dos casos, cuando toma valor 0 y cuando toma valor 1, considerando que esos dos casos son seleccionados al azar.

A los falsos positivos, o error de tipo I, se les denomina *sensibilidad*, y a los falsos negativos, o error de tipo II, se les llama *especificidad*. Luego, en la curva ROC se presenta la sensibilidad en función de los falsos positivos para distintos puntos de corte.

Los valores del ROC van entre 0,5, lo que es igual a si se hubiese calculado al azar, y el máximo que es 1. Se suele admitir como valor aceptable de discriminación cuando este índice supera el 0,7 [36].

3.2.6. Interpretación de los Estimadores

En este tipo de modelos [37], no es posible interpretar directamente las estimaciones de los parámetros β debido a que son modelos no lineales. Sin embargo, el signo de estos estimadores tienen una interpretación muy clara. Si el estimador es positivo, entonces incrementos en la variable asociada causan aumentos en $P(y_i = 1|X_i)$ de magnitud desconocida. Por el contrario, si el estimador muestra un signo negativo, ello supondrá que incrementos en la variable asociada causarían disminuciones en la probabilidad de elección.

En este modelo se suelen usar otros dos conceptos para profundizar más en la interpretación de los estimadores:

1. Odds: es el cociente de probabilidad.

$$(3.16) \quad \text{Odd}_i = \frac{P(y_i = 1|X_i)}{1 - P(y_i = 1|X_i)} = \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik})$$

Al tomar logaritmo neperiano en la expresión anterior, se obtiene una expresión lineal para este modelo.

$$(3.17) \quad \ln \left[\frac{P(y_i = 1|X_i)}{1 - P(y_i = 1|X_i)} \right] = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik}$$

A partir de lo obtenido en la ecuación 3.17, se puede interpretar el estimador del parámetro β_j como la variación en el término del logaritmo neperiano del cociente de probabilidades, causada por una variación unitaria en la variable x_{ij} , suponiendo constantes el resto de las variables explicativas.

2. Odds-Ratio: cociente entre los dos *odds* asociados. Se hace referencia al incremento unitario en una de las variables explicativas del modelo. Luego, si se asume que ha habido un incremento unitario en la variable X_i , entonces se tiene la siguiente expresión.

$$(3.18) \quad Odds - Ratio = \frac{Odds_{i2}}{Odds_{i1}} = exp(\beta_i)$$

De la expresión anterior se deduce que un coeficiente β_i cercano a cero, o un odds-ratio cercano a uno, significará que cambios en la variable explicativa X_i asociada no tendrán efecto alguno sobre la variable dependiente Y .

Capítulo 4

Aplicación del Modelo Propuesto

Contenido

4.1. Definición del Problema a Resolver	38
4.2. Metodología Aplicada en la Primera Comisaría de Santiago	39
4.2.1. Recolección de los Datos	39
4.2.2. Variables del Modelo	40
4.2.3. Diagramas de Voronoi	46
4.2.4. Selección de Atributos	51

4.1. Definición del Problema a Resolver

El objetivo de esta tesis consiste en desarrollar un modelo de regresión logística para determinar la influencia de distintos atributos sobre el nivel de actividad delictual de una zona geográfica particular, estableciendo así una herramienta que ayude a la planificación de recursos por parte de las organizaciones encargadas de combatir el crimen. El terreno contemplado para su desarrollo es la Primera Comisaría de Santiago y se considerarán exclusivamente los delitos relativos a hurtos.

La razón por la que se consideran únicamente este tipo de delitos es debido a las características inherentes de la comuna de Santiago, especialmente en el terreno de la Primera Comisaría, la cual se destaca por una alta afluencia peatonal y una importante concentración de actividad comercial. Estas características motivan en mayor medida los delitos contra la propiedad, y, dado que casi el 50% de la base de datos hace referencia a hurtos (Tabla 2.1), entonces se justifica el estudio exclusivo de estos delitos.

Al momento de decidir su ataque, el delincuente se enfrenta principalmente con dos problemas [38]. El primero es determinar el conjunto de alternativas relevantes, es decir, los lugares que más le llaman la atención, mientras que el segundo es el decidir qué lugar atacar dadas las alternativas consideradas anteriormente. En este último paso, los individuos comparan los atributos de las diferentes opciones, por lo que establecer cuáles son las variables que determinan en mayor grado el atractivo del lugar podría ayudar de manera importante en la prevención de delitos de este tipo.

La labor de asignación de recursos en la vía pública, tanto policiales como preventivos, es una problemática con la que se enfrentan diariamente los organismos encargados de combatir la

delincuencia. De ella depende gran parte de la efectividad de estas instituciones preventivas. Hoy en día se utilizan sofisticados softwares georreferenciados para visualizar el comportamiento de los delincuentes y de los *hot spots*. Sin embargo, la toma de decisiones tiene mucho de intuitivo, por lo que se propone respaldo y mejoría en base a técnicas estadísticas y matemáticas avanzadas.

Considerando todo lo anterior, se tiene que el problema a resolver consiste básicamente en determinar de qué manera, positiva o negativamente, contribuyen los distintos atributos en el nivel delictual de una región geográfica, y cuáles influyen en mayor medida.

4.2. Metodología Aplicada en la Primera Comisaría de Santiago

4.2.1. Recolección de los Datos

Para el desarrollo de este estudio se cuenta con dos tipos de datos: (i) denuncias de delitos ocurridos en la Primera Comisaría de Santiago de Chile entre el día 1 de Enero del año 2001 y el día 31 de Diciembre del año 2004, (ii) atributos presentes en el área determinada por la Primera Comisaría de la Comuna de Santiago.

Para la visualización y agregación de estos datos se utiliza GRASS GIS (Geographical Information Systems), un sofisticado software georreferenciador que permite ubicar en el plano los distintos puntos relativos tanto a hurtos como a atributos, permitiendo ver tanto las zonas calientes, o *hot spots*, como las zonas de menor actividad delictual. Su licencia es abierta para todo público y la manipulación de datos es muy sencilla.

La base de datos de las denuncias otorgada por Carabineros de Chile hacen referencia a todos los delitos ocurridos en la zona mencionada entre los años 2001 y 2004. Se especifican diferentes características del suceso detallándose el lugar de ocurrencia, permitiendo así la ubicación de los delitos en el plano geográfico. En la siguiente sección se estudian en detalle los datos relativos al delito de hurto.

Por otra parte, la información relativa a la ubicación de los atributos se obtuvo, principalmente, como resultado de salidas a terreno, datos adquiridos de MapCity¹ y Google Maps², e informes proporcionados por la Municipalidad de Santiago³ y Carabineros de Chile⁴.

Mapcity y Google Maps son servidores gratuitos que permiten la aplicación de mapas en Internet. Ofrecen imágenes de mapas desplazables de distintas ciudades del mundo en los que se puede identificar la ubicación de ciertos lugares de interés, tales como dependencias gubernamentales, alojamientos, transporte, locales comerciales y otros establecimientos relacionados a cultura, educación, turismo, deporte, salud y religión, entre otros.

Por otro lado, los archivos proporcionados por la Municipalidad permiten identificar la ubicación de casetas de seguridad municipal y patentes comerciales actualizadas a la fecha⁵, que dan cuenta de la localización de tiendas comerciales pequeñas, restaurantes, pubs y discotecas. En cuanto a los informes entregados por Carabineros, se detalla la disposición de las distintas dependencias

¹MapCity. [Fecha de consulta: 02 Agosto 2011]. Disponible en: www.mapcity.cl

²Google Maps. [Fecha de consulta: 02 Agosto 2011]. Disponible en: www.maps.google.com

³Municipalidad de la Comuna de Santiago, Santiago, Chile. [Fecha de consulta: 02 Agosto 2011]. Disponible en: www.municipalidaddesantiago.cl.

⁴Carabineros de Chile. [Fecha de consulta: 02 Agosto 2011]. Disponible en: www.carabineros.cl.

⁵Patentes comerciales otorgadas por la Municipalidad de Santiago durante el año 2008.

policiales y casetas de seguridad, además de la dotación de recursos policiales de la Primera Comisaría de Santiago.

4.2.2. Variables del Modelo

Como se explica en el capítulo 3, los modelos de regresión logística binaria son modelos estadísticos en los que se desea conocer la relación entre una variable dependiente cualitativa dicotómica y una o más variables explicativas independientes, o atributos, ya sean cualitativas o cuantitativas, siendo la ecuación inicial del modelo de tipo exponencial, si bien su transformación logarítmica permite su uso como una función lineal.

El objetivo primordial que resuelve esta técnica es el de modelar cómo influye en la probabilidad de aparición de un suceso dicotómico, la presencia o no de diversos factores y el valor o nivel de los mismos. Para lograr esta meta, es necesario definir de manera clara y objetiva tanto la variable dependiente como las independientes.

Variable Dependiente

Como se menciona anteriormente, se cuenta con una base de datos de denuncias de delitos ocurridos en la Primera Comisaría de la comuna de Santiago entre los días 1 de Enero de 2001 y 31 de Diciembre de 2004. La información que se detalla en esta base de datos hace referencia directa a las circunstancias en que ocurre el evento (Tabla 4.1), especificándose la fecha, cuadrante, hora, tipo de delito, monto avaluado de la sustracción y coordenadas geográficas.

Día	Mes	Año	Sem	Trim	Cuad	Hora	Tipo Delito	Monto	Coord X	Coord Y
1	Enero	2001	1	1	2	11:00	Lesiones	No Aplica	346402,40	6299266,00
23	Julio	2001	2	3	2	19:00	Hurto	250.001 a 500.000	346913,26	6298399,27
1	Febrero	2002	1	1	2	13:00	RCV	50.001 a 100.000	346342,59	6298746,00
8	Abril	2002	1	2	2	9:00	Hurto	10.000 a 50.000	346913,26	6298399,27
8	Noviembre	2002	2	4	3	19:00	VI	No Aplica	346419,55	6299504,66
29	Mayo	2003	1	2	2	10:00	Hurto	500.001 a 750.000	346546,00	6299161,00
19	Diciembre	2003	2	4	2	23:00	RCF	250.001 a 500.000	346986,65	6298774,52
7	Septiembre	2004	2	3	2	11:00	Hurto	50.001 a 100.000	346913,26	6298399,27

Tabla 4.1: Base de Datos de Denuncias

Tal y como se menciona en secciones anteriores, en esta tesis se consideran únicamente los delitos relativos a hurtos. En la Figura 4.1 se observa la distribución espacial de las denuncias de este tipo de delitos. Se advierten zonas con concentraciones interesantes de hurtos y otras con una menor actividad delictual, distinguiéndose una clara densidad en el centro de la zona estudiada.

La variable dependiente es aquella que deseamos modelar. Por lo tanto, para este estudio se define como variable dependiente el número de hurtos ocurridos en una región determinada. Como se puede ver, ésta es una variable continua, por lo que la aplicación de un modelo de regresión logística binaria no es posible.

Se tiene entonces que la variable dependiente toma valores enteros determinados por la cantidad de hurtos ocurridos en una zona. Luego, el valor que esta variable toma no tiene ningún significado debido a que es un simple número que no está comparado ni normalizado con ningún otro. Por lo tanto, es necesario transformar esta variable en alguna otra que adquiera sentido.

Dado que lo que se busca en esta tesis hace relación al nivel de actividad delictual de una zona en específico, entonces se resuelve dicotomizar la variable dependiente para así poder aplicar el modelo propuesto sin problemas. Puede que al transformar una variable inicialmente recogida como

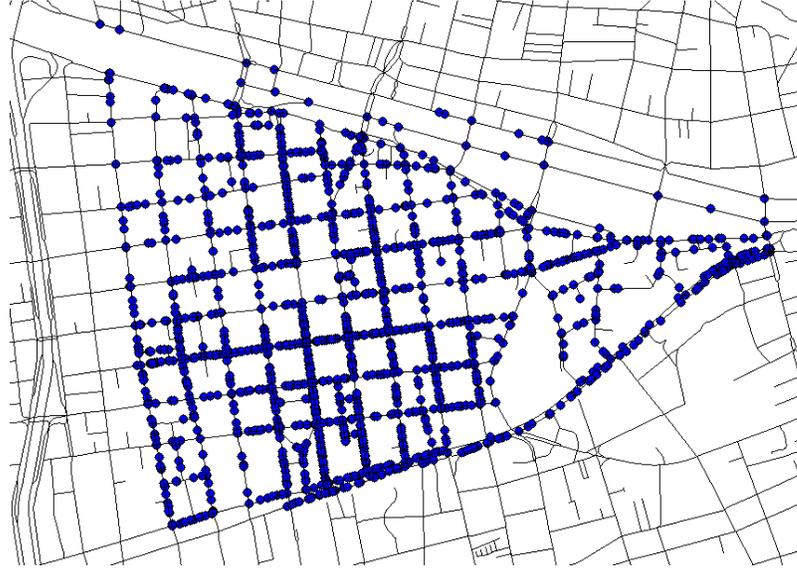


Figura 4.1: Distribución de los Hurtos en la Primera Comisaría
Fuente: *Elaboración Propia*

continua en categórica signifique perder un poco de información. No obstante, a la hora del análisis se gana en eficiencia y, sobre todo, en claridad en la interpretación.

Para poder categorizar esta o cualquier otra variable, es necesario primero dividir la zona estudiada en áreas más pequeñas y generar observaciones de esta variable en diferentes instancias. Antes de determinar que metodología se utilizará para este propósito, se deben caracterizar las variables independientes del modelo.

VARIABLES INDEPENDIENTES

El objetivo de este trabajo es estimar el nivel delictual de una región geográfica particular en base a los atributos que describan su entorno. Por esta razón, se realiza un análisis detallado de las características de la zona de estudio en el capítulo 2. Como conclusión de ello, se evidenció que la Primera Comisaría de Santiago presenta una concentración importante de diferentes actividades, tales como de gobierno, servicios, comercio, finanzas, cultura, recreación y turismo, entre otras. Se destaca particularmente la gran afluencia de población flotante en esta área, en especial por los paseos peatonales que la cruzan.

Dado lo anterior, se podrían definir muchas variables que caracterizaran esta región de manera única. Sin embargo, para que estas variables representen de manera fidedigna el entorno y el modelo se ajuste correctamente a la realidad, deben cumplir dos condiciones básicas: (i) ser observables, es decir, deben ser atributos físicos visibles y claramente distinguibles, y (ii) ser cuantificables, es decir, deben ser atributos que se puedan medir con una escala de medición clara y objetiva.

Las variables se pueden medir de diferentes maneras dependiendo de la naturaleza de las mismas. Es por esto que a continuación se definen los distintos tipos de atributos.

- **Atributos Cualitativos:** Son los atributos que expresan distintas cualidades, características o

modalidades.

1. Dicotómicos: Estos atributos sólo pueden tomar dos valores posibles.
 2. Policotómicas: Estos atributos pueden adquirir tres o más valores. Dentro de esta clasificación se pueden distinguir los atributos *ordinales*, que toman distintos valores ordenados siguiendo una escala establecida, y los atributos *nominales*, que toman valores que no pueden ser sometidos a un criterio de orden.
- Atributos Cuantitativos: Son las variables que se expresan mediante cantidades numéricas. Los atributos cuantitativos pueden ser:
 1. Discretos: Presenta separaciones o interrupciones en la escala de valores que pueden tomar. Estas separaciones indican la ausencia de valores entre los distintos valores específicos que la variable pueda asumir.
 2. Continuos: Pueden adquirir cualquier valor dentro de un intervalo especificado de valores.

La definición de posibles atributos se realizó en base a juicios de expertos y estudios similares donde se analizan variables propias del lugar en cuestión [39], [40], [41]. Entre estas variables se encuentran las relacionadas con características directas del entorno, otras relativas al flujo de peatones y población flotante, y las demás a las condiciones socio-demográficas de los residentes de la zona.

La información que se tiene de la población flotante y de las variables socio-demográficas de la Primera Comisaría de Santiago son datos agregados a nivel comunal y, dado que no se tiene información más específica, no es posible su interpretación y desagregación a nivel de zonas más pequeñas. En cuanto a los atributos del entorno, existen algunas variables que no cumplen con las condiciones mencionadas anteriormente, como es el caso del nivel de iluminación nocturna y el ancho de las calles, que claramente son variables determinantes. Sin embargo, no son atributos que se puedan distinguir de manera evidente y es muy difícil determinar una escala de medición objetiva.

Después de identificar todos los atributos observables y cuantificables, se determinaron 18 variables capaces de caracterizar la zona de estudio. En la Tabla 4.2 se resumen los diferentes atributos considerados, el tipo de clasificación según su naturaleza y la cantidad total de estas variables presentes en la zona de la Primera Comisaría de Santiago.

La distribución espacial de estos atributos en el plano se presenta a continuación. En primer lugar se tienen las dependencias gubernamentales, las cuales hacen referencia a la municipalidad de la comuna, casa de gobierno, registro civil, fiscalías, embajadas y ministerios (Figura 4.2). Se observa además, en la Figura 4.3, la distribución de las fuerzas policiales que, por su parte, están compuestas por establecimientos de carabineros, policía de investigaciones y casetas de seguridad.

En la Figura 4.4 se distinguen los centros de salud, que hacen referencia a hospitales, centros médicos, consultorios y centros dentales. En cuanto a los establecimientos educacionales (Figura 4.5), se tienen unos cuantos institutos y universidades. En la Figura 4.6 se observa la ubicación de las salidas de metro y de los paraderos de buses, correspondientes a la variable transporte público. En cuanto a los establecimientos culturales, se puede ver de la Figura 4.7 la distribución de museos, teatros, bibliotecas y centros culturales.

Además, en la Figura 4.8 se observan los locales de comida y entretenimiento, tales como restaurantes, cafés, pastelerías, heladerías, pubs, discotecas y salas de cine. También (Figura 4.9) se percibe la ubicación de los distintos locales de comercio masivo, tales como multitiendas y centros comerciales.

	Atributos	Tipo	Cantidad
X1	Dependencias Gubernamentales	Continua	27
X2	Fuerzas Policiales	Continua	10
X3	Centros de Salud	Continua	10
X4	Establecimientos Educacionales	Continua	5
X5	Transporte Público	Continua	124
X6	Establecimientos Culturales	Continua	18
X7	Comida y Entretención	Continua	111
X8	Comercio Masivo	Continua	52
X9	Iglesias	Continua	6
X10	Alojamiento	Continua	27
X11	Casas de Cambio	Continua	17
X12	Farmacias	Continua	42
X13	Bancos	Continua	57
X14	Centros de Pago	Continua	24
X15	Supermercados	Continua	9
X16	Estacionamientos	Continua	8
X17	Áreas Verdes	Dicotómica	N/A
X18	Paseos Peatonales	Dicotómica	N/A
	TOTAL		547

Tabla 4.2: Atributos Considerados

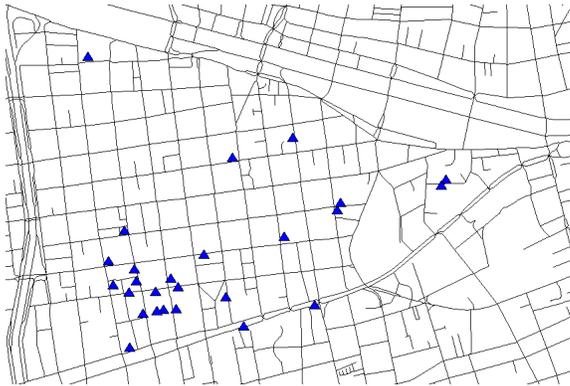


Figura 4.2: Dependencias Gubernamentales
Fuente: *Elaboración Propia*

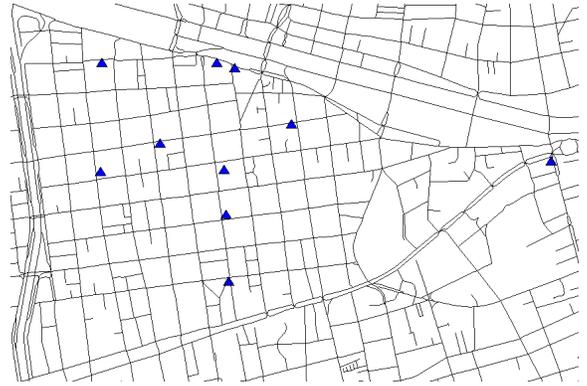


Figura 4.3: Fuerzas Policiales: Carabineros, PDI y Casetas de Seguridad
Fuente: *Elaboración Propia*

Por otro lado, la distribución de las iglesias se observa en la Figura 4.10 y el alojamiento en la Figura 4.11

Las casas de cambio se observan en la Figura 4.12, mientras que las farmacias se distribuyen como se muestra en la Figura 4.13. La ubicación de los bancos se ve en la Figura 4.14 y la de los centros de pago en la Figura 4.15. Los supermercados y estacionamientos se observan en las Figuras 4.16 y 4.17.

Como se mencionó anteriormente, se tiene que del total de atributos sólo dos son variables dicotómicas, es decir, que toman valores 1 o 0 según cumplan o no una condición, que en este caso son la presencia de áreas verdes y paseos peatonales (Figuras 4.18 y 4.19).

Cuando se tienen variables de diferente naturaleza, es decir, que unas son continuas y otras dicotómicas, y que los rangos de los valores de las variables cuantitativas son distintos, entonces se

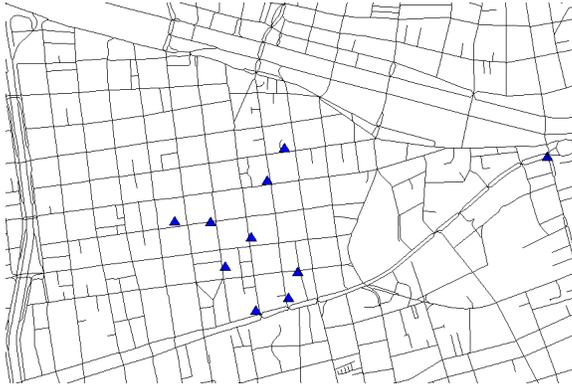


Figura 4.4: Centros de Salud
Fuente: *Elaboración Propia*

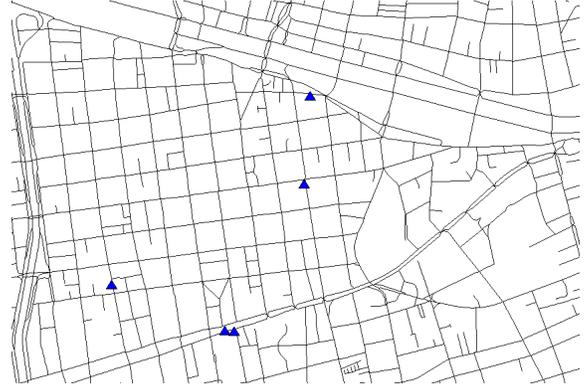


Figura 4.5: Establecimientos Educativos
Fuente: *Elaboración Propia*

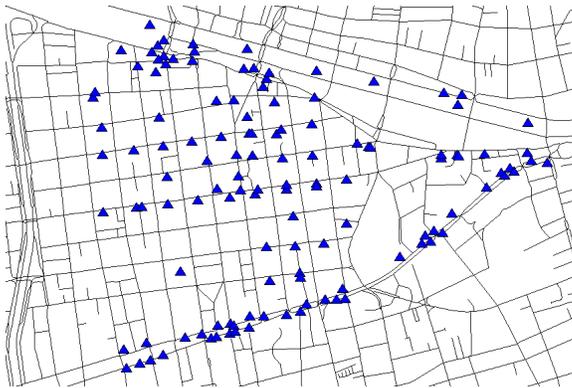


Figura 4.6: Transporte Público: Salidas de Metro y Paraderos de Buses
Fuente: *Elaboración Propia*

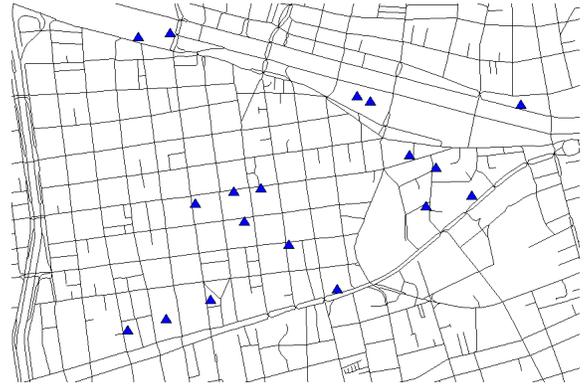


Figura 4.7: Establecimientos Culturales
Fuente: *Elaboración Propia*

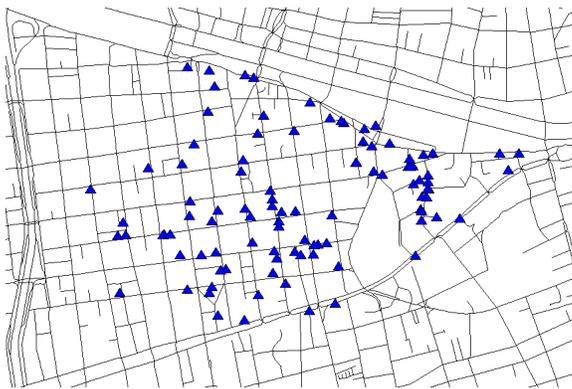


Figura 4.8: Comida y Entretenimiento
Fuente: *Elaboración Propia*

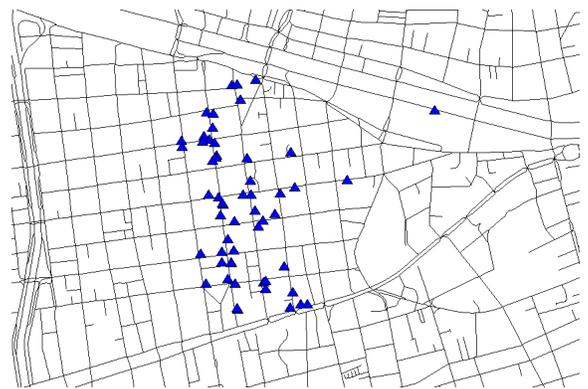


Figura 4.9: Comercio Masivo
Fuente: *Elaboración Propia*

tiene un problema grave con respecto al modelo de regresión logística binaria. Esto se debe a que se realizan las estimaciones de los coeficientes considerando una relación lineal entre las variables

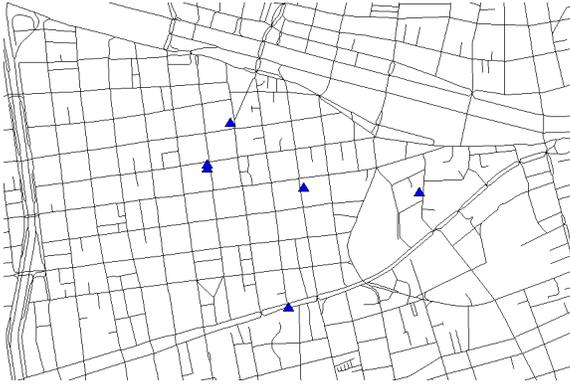


Figura 4.10: Iglesias
Fuente: *Elaboración Propia*

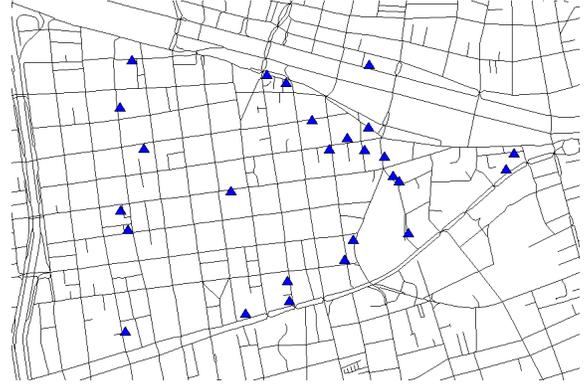


Figura 4.11: Alojamiento
Fuente: *Elaboración Propia*

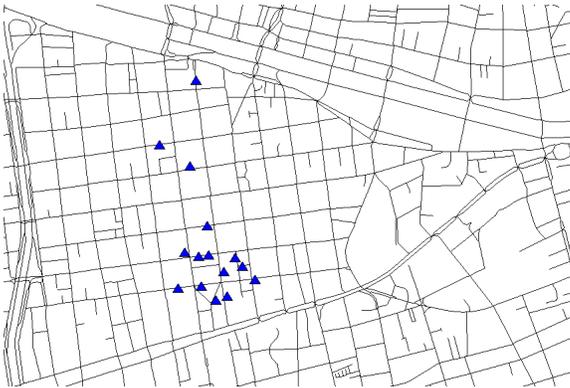


Figura 4.12: Casas de Cambio
Fuente: *Elaboración Propia*

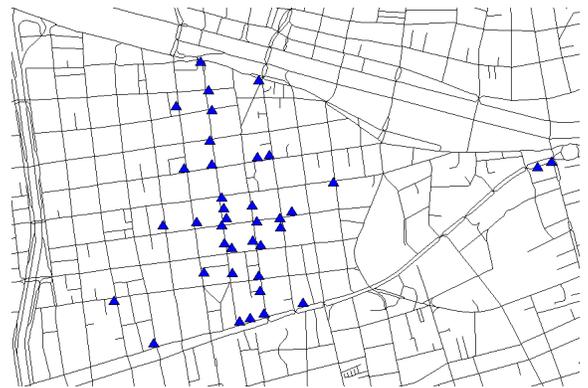


Figura 4.13: Farmacias
Fuente: *Elaboración Propia*

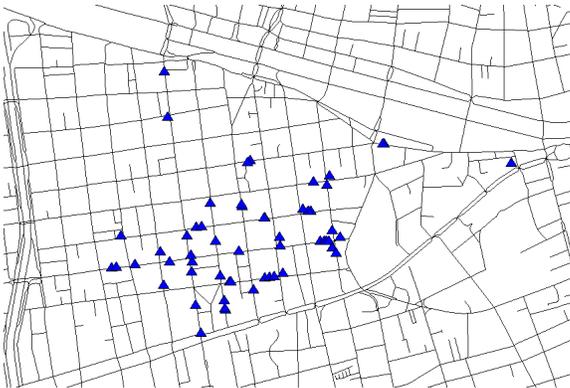


Figura 4.14: Bancos
Fuente: *Elaboración Propia*

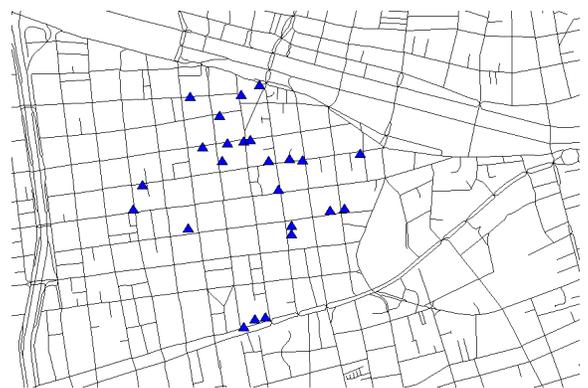


Figura 4.15: Centros de Pago
Fuente: *Elaboración Propia*

independientes (Ecuaciones 3.10 y 3.11), lo que implica entonces la necesidad de un ajuste en cuanto a la naturaleza de éstas.

Pasar de una variable dicotómica a una continua es un método un tanto arbitrario debido a que no

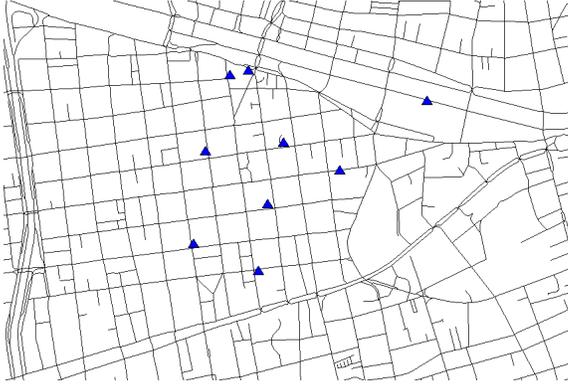


Figura 4.16: Supermercados
Fuente: *Elaboración Propia*

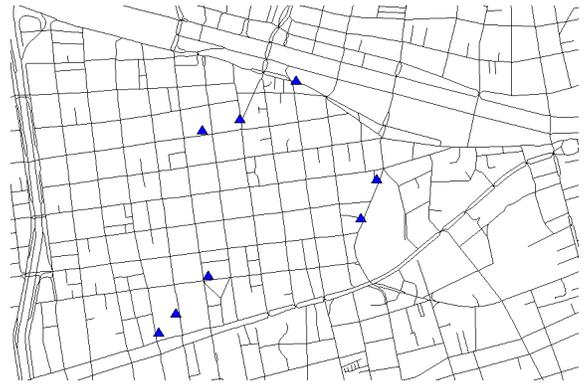


Figura 4.17: Estacionamientos
Fuente: *Elaboración Propia*

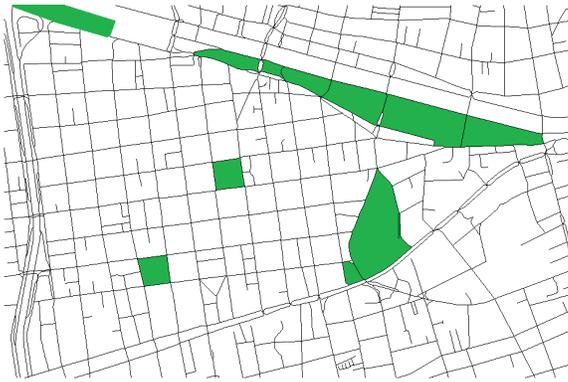


Figura 4.18: Áreas Verdes
Fuente: *Elaboración Propia*

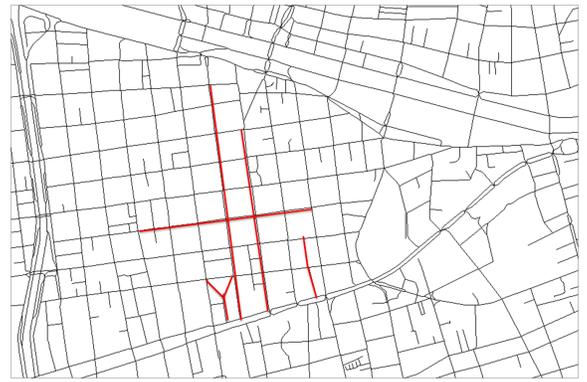


Figura 4.19: Paseos Peatonales
Fuente: *Elaboración Propia*

hay reglas de clasificación claras. Luego, se pueden realizar transformaciones basadas en diferentes reglas que podrían dar resultados muy disímiles. Así, la dicotomización de las variables continuas es una opción bastante objetiva que da solución al problema que se tiene.

Muchos sinónimos puede tener el concepto de atributo, tales como variable independiente, factor, característica o propiedad del lugar, entre otros. Por lo tanto, para evitar confusiones y ambigüedades, este término se expresará únicamente como *atributo*.

4.2.3. Diagramas de Voronoi

Para realizar un análisis certero y tener un modelo más exacto, es necesario dividir la región de estudio en áreas más pequeñas. Esta técnica de división del espacio se utiliza mucho en áreas de investigación orientadas al estudio de especies animales u otros organismos vivos debido a que en estas investigaciones el tamaño del área muestral es fundamental para la detección de patrones de comportamiento.

Una forma muy común de dividir una zona en particular es la creación de una grilla con celdas contiguas de igual tamaño en toda la región estudiada [42]. Una división de la Primera Comisaría de Santiago en base a este modelo se muestra en la Figura 4.20.

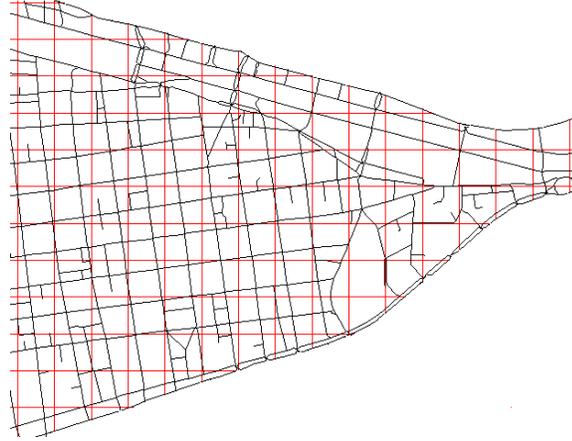


Figura 4.20: Primera Comisaría de Santiago dividida en Grillas
Fuente: *Elaboración Propia*

A pesar de ser una técnica que se usa frecuentemente en otras áreas de estudio, este procedimiento presenta algunas desventajas importantes, lo que se debe principalmente a la forma de las subdivisiones. Éstas son muy restrictivas y no permiten alteraciones del espacio. Lo anterior genera que muchas celdas no contengan características relevantes para el análisis, como es el caso de aquellas que encierran únicamente cuadras y no calles. Esto se traduce posteriormente en la no inclusión de delitos y/o atributos, en otras palabras, se tienen muchas celdas sin datos. Además, muchas de ellas no tienen significado físico, como es el caso de las celdas que cruzan calles o cuadras en direcciones aleatorias.

La desventaja más importante que presenta esta metodología para el caso particular de esta tesis, es que no se pueden unir celdas para construir *hot spots*. Luego, para conformar espacios de interés se deben hacer subdivisiones muy pequeñas, lo que puede generar los problemas mencionados anteriormente.

Para este estudio se requieren divisiones del espacio que tengan un sentido lógico para los individuos que circulan por el sector en cuestión. Luego, el procedimiento anterior no cumple con las condiciones básicas exigidas.

Un método que se acomoda de mejor manera a las exigencias del problema son los diagramas de Voronoi⁶, también denominados *polígonos*, debido a que la forma de las subdivisiones no está restringida a un área fija [43] y los límites de estas regiones pueden alterarse para la formación de áreas que sean interesantes de estudiar. Luego, ya no se tienen las dificultades de la técnica anterior.

Para poder generar los diagramas de Voronoi se deben seguir ciertos pasos explicados a continuación. En primer lugar, se deben definir los centros de estos polígonos. Esta etapa es la más importante del proceso debido a que estos puntos determinarán la forma y tamaño de estas subdivisiones. Dado que estos polígonos deben tener un significado físico que siga una lógica acorde a los objetivos propuestos, entonces se definen como centro a las esquinas, es decir, los puntos de intersección de las calles. Lo anterior debido a que esta división permite estudiar el comportamiento delictual no sólo en esos puntos sino también en sus alrededores (Figura 4.21 (a)).

Una vez definidos los centros de los polígonos, viene el paso donde se deben determinar las fronteras

⁶Mayor detalle de los Diagramas de Voronoi en el Anexo C.

de estas áreas de tal forma de maximizar la superficie de cada uno de ellos. Esta etapa queda bastante clara al observar la Figura 4.21 (b). A continuación, se delimita la zona a estudiar (Figura 4.21 (c)), lo que se realiza calculando los puntos mínimos y máximos, en cuanto a coordenadas, de todos los hurtos y atributos que forman parte del plano estudiado. Esto ayuda a acotar el problema en una región fija, lo que posteriormente tiene un efecto muy provechoso en términos computacionales.

Por último, se ubican los datos de delitos y atributos en el plano estudiado. En la Figura 4.21 (d) se pueden ver los polígonos de Voronoi aplicados en el sector de la Primera Comisaría de Santiago con los hurtos graficados en color verde y los atributos en color morado.

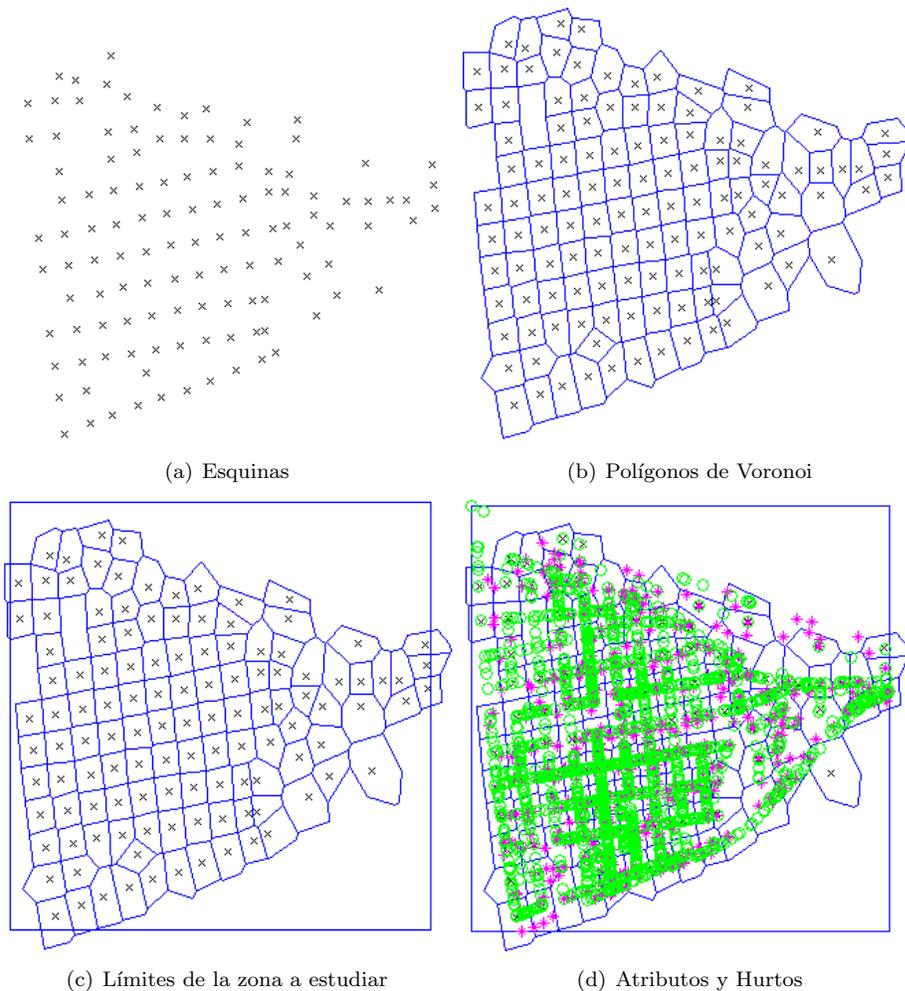


Figura 4.21: Diagramas de Voronoi en la 1ª Comisaría de Santiago 2001 - 2004
Fuente: *Elaboración Propia*

Finalmente, como resultado de la aplicación de los diagramas de Voronoi, se obtiene una partición más intuitiva y no homogénea del espacio. El sector estudiado se divide en 120 zonas más pequeñas con sus centros en las esquinas en las que se tienen datos relativos tanto a hurtos como a atributos. Notar que estos polígonos se conforman si y sólo si esta área contiene a ambas variables; en caso contrario, se excluye del análisis.

Una ventaja muy importante de la división del espacio es que ahora se puede definir el nivel delictual

de cada una de estas áreas. Además, es posible dicotomizar los atributos según el polígono que se está observando. Sin una división visible del espacio, lo anterior no es viable debido a que se tienen los datos agregados como un todo en el área de la Primera Comisaría, es decir, se tiene una sola observación. Luego, al separar esta región, se incrementa el número de observaciones, y por ende, ya se puede realizar un análisis estadístico más interesante. De ahora en adelante, el concepto de polígono se expresará con el término *observación*.

Variable Dependiente

Como se menciona en la sección anterior, el aumento de las observaciones permite generar más datos. Para el caso de la variable dependiente, se obtienen estadísticas descriptivas que permitirán la definición de un alto o bajo nivel delictual dependiendo de la cantidad de hurtos que se obtenga en cada observación. A continuación, se observan las estadísticas descriptivas en la Tabla 4.3 y el histograma de frecuencias de esta variable 4.4.

Hurtos		
Observaciones	Válidas	120
	Perdidas	0
Promedio		95,45
Mediana		55,00
Desviación Estándar		122,96
Mínimo		1
Máximo		856

Tabla 4.3: Estadísticas Descriptivas Variable Dependiente Continua

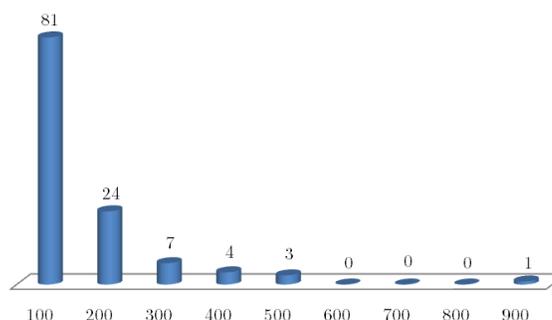


Tabla 4.4: Histograma Variable Dependiente
Fuente: *Elaboración Propia*

De lo anterior se observa que el valor promedio de número de hurtos en las distintas observaciones (96) es bastante menor que el valor máximo (856), lo que se ve reflejado, además, en un valor de la mediana muy bajo y de desviación estándar muy alto. Lo anterior sugiere una concentración de datos con valores pequeños, lo que se observa claramente en el histograma. Hay más de 80 observaciones que presentan menos de 100 hurtos, cerca el 67%, mientras que el 33% restante de las observaciones se distribuye en un rango mucho mayor.

Como se explica en las secciones anteriores, es necesario dicotomizar esta variable. Para realizar lo anterior, es necesario elegir el punto de corte que definirá si la variable dependiente toma el valor 0 o 1. Existen diversos métodos para la definición de estos puntos de corte, algunos de los cuales se definirán a continuación.

1. Intervalos de Igual Amplitud: se define el punto de corte de manera que se generen categorías con la misma distancia entre los límites del rango.
2. Percentiles Iguales: se define el punto de corte de manera que se generen intervalos con un número aproximadamente igual de casos en cada categoría.
3. Puntos de Corte en Media y Desviaciones Típicas Seleccionadas: se define el punto de corte de manera que se generen categorías basándose en los valores de la media y la desviación típica de la distribución de la variable.

En cuanto a las ventajas y desventajas de cada una de las técnicas, no existe un método mejor que otro *per se*. La decisión se debe tomar en base a los propósitos que se persiguen. Si el objetivo es definir intervalos iguales, entonces se debe elegir el primer procedimiento, mientras que si lo que se busca es dividir la muestra en partes iguales, entonces habría que inclinarse por el segundo. Estas dos primeras técnicas se recomiendan cuando se tienen datos que se comportan según una distribución normal. Por el contrario, el tercer método se sugiere cuando los datos no siguen una distribución de este tipo.

La Tabla 4.5 resume la información que resulta al categorizar la variable dependiente con las diferentes técnicas descritas. Se puede ver que el primer procedimiento clasifica casi la totalidad de las observaciones en una categoría, mientras que el segundo las separa en dos grupos con cantidades de observaciones similares. Considerando lo anterior y dado que los datos que se tienen claramente no se distribuyen de forma normal, se resuelve por aceptar la tercera técnica como método de categorización de la variable dependiente.

Categoría	Intervalos Iguales	Percentiles Iguales	Media y Desviación Estándar
$y = 0$	116	61	79
$y = 1$	4	59	41
TOTAL	120	120	120

Tabla 4.5: Comparación Métodos de Categorización

Una vez elegido el método de categorización, se determina el punto de corte que permite la dicotomización de la variable que coincide con el promedio de la variable continua. Luego, se procede a clasificar las observaciones según el umbral definido. Como resultado de lo anterior, se obtiene la Figura 4.22 donde se puede observar el nivel delictual de cada una de estas observaciones. Las zonas de color rojo son las que presentan un nivel de actividad criminal alto, es decir, que fueron categorizadas como $y = 1$, mientras que las zonas amarillas revelan un nivel bajo de delitos o $y = 0$.

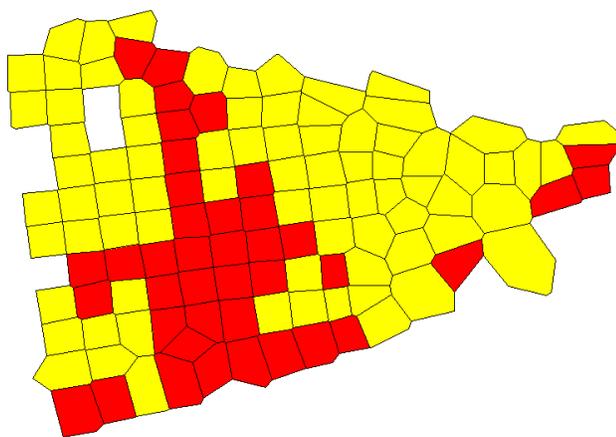


Figura 4.22: Nivel Delictual en la Primera Comisaría de Santiago
Fuente: *Elaboración Propia*

Atributos

Como se deja claro en las secciones pasadas, cuando se tienen variables explicativas continuas, se recomienda intentar transformarlas a variables categóricas, preferiblemente dicotómicas. Se

	X1	X2	X3	X4	X5	X6	X7	X8	X9
Promedio	0,23	0,08	0,08	0,04	1,03	0,14	0,94	0,44	0,05
Desviación Estándar	0,65	0,31	0,28	0,20	1,29	0,38	1,43	1,26	0,26
Mínimo	0	0	0	0	0	0	0	0	0
Máximo	4	2	1	1	6	2	10	8	2

	X10	X11	X12	X13	X14	X15	X16	X17	X18
Promedio	0,23	0,14	0,36	0,48	0,20	0,07	0,07	0,25	0,22
Desviación Estándar	0,48	0,46	0,71	1,04	0,50	0,25	0,25	0,44	0,42
Mínimo	0	0	0	0	0	0	0	0	0
Máximo	2	2	3	7	2	1	1	1	1

Tabla 4.6: Estadísticas Descriptivas Variables Independientes Continuas

	X1	X2	X3	X4	X5	X6	X7	X8	X9
Promedio	0,15	0,08	0,08	0,04	0,58	0,14	0,48	0,16	0,04
Desviación Estándar	0,36	0,27	0,28	0,20	0,49	0,34	0,50	0,37	0,20
Mínimo	0	0	0	0	0	0	0	0	0
Máximo	1	1	1	1	1	1	1	1	1

	X10	X11	X12	X13	X14	X15	X16	X17	X18
Promedio	0,20	0,10	0,25	0,26	0,16	0,07	0,07	0,25	0,22
Desviación Estándar	0,40	0,30	0,43	0,44	0,37	0,25	0,25	0,44	0,42
Mínimo	0	0	0	0	0	0	0	0	0
Máximo	1	1	1	1	1	1	1	1	1

Tabla 4.7: Estadísticas Descriptivas Variables Independientes Dicotómicas

procederá entonces con la misma metodología desarrollada con la variable predictora.

De la Tabla 4.6 se calculan las estadísticas descriptivas de las variables independientes continuas. Luego, se procede a clasificar las observaciones según el umbral definido para cada una de las variables. Como resultado de lo anterior, se obtiene la Tabla 4.7 donde se pueden ver las estadísticas descriptivas de los atributos ahora binarios.

Coincidentemente el punto de corte de todos los atributos dio 1. Esto debido a que este valor converge al promedio redondeado de la variable continua. Luego, se tiene que la no presencia de las variables continuas, es decir $X_i = 0$, en la observación define un atributo dicotómico de valor 0, mientras que la presencia de las mismas, es decir $X_i > 0$, se transforma a 1. Por lo tanto, se puede decir entonces que los atributos ahora están definidos como la presencia o ausencia de esa característica particular.

4.2.4. Selección de Atributos

Como ya se ha mencionado anteriormente, el objetivo de este análisis es predecir la variable dependiente, que en este caso es el nivel delictual, a partir de las variables predictoras o atributos, obteniendo una fórmula matemática que sirva para calcular la probabilidad del suceso estudiado en una nueva observación en razón de los valores que presenten los atributos incluidos en el modelo. Bajo esta óptica, se debe buscar, entre todos los posibles modelos, el más parsimonioso, que es el que con el menor número de variables posibles, genera una predicción más precisa y válida de la respuesta evaluada.

Introducir variables poco relevantes tiende a enmascarar el proceso de modelado y puede llevar a

estimaciones no válidas. Por otra parte, intentar construir un modelo con muchas variables puede ser un problema cuando hay pocas observaciones, ocasionando estimaciones inestables y poco precisas. Luego, una vez que se dispone de un modelo inicial con estimadores poco significativos, entonces debe procederse a su reducción hasta obtener un modelo simplificado que siga explicando los datos.

Hay que tener en cuenta que un mayor número de variables en el modelo implica mayores errores estándar, por lo que es aconsejable realizar una selección de los atributos que expliquen adecuadamente la variable dependiente. Este procedimiento estadístico es muy importante debido a tres razones fundamentales: (i) no todas las variables predictoras tienen igual importancia (variables irrelevantes), (ii) algunas variables pueden perjudicar la confiabilidad del modelo (variables redundantes), y (iii) computacionalmente es más fácil trabajar menos variables.

El paso a seguir entonces, es eliminar las variables irrelevantes y redundantes. Luego, se incluirán sólo las variables que en un análisis univariado demuestren una relación suficiente con la variable que se desea explicar, que en este caso es el nivel delictual. Existen muchas técnicas que sirven para seleccionar variables [44], siendo las más utilizadas el *Forward Selection* (o selección secuencial hacia adelante), *Backward Elimination* (o selección secuencial hacia atrás) y la selección del *Stepwise*. Muchos programas computacionales tienen incluidos estos métodos y la aplicación de éstos es bastante simple.

El *Forward Selection* consiste en elegir las variables más representativas de todo el conjunto de atributos según un método de validación predeterminado, hasta llegar a un máximo fijado inicialmente o hasta que se cumpla alguna condición. El algoritmo comienza considerando una única variable para después incluir marginalmente las variables predictoras significativas. El problema de esta metodología es que la definición del valor máximo de variables o de la condición de detención puede ser un tanto subjetivo. En consecuencia, se pueden obtener resultados muy diferentes a medida que estas condiciones cambien.

El *Backward Elimination*, por su parte, es un algoritmo *top-down* que selecciona una variable, para su eliminación, en cada paso. Este método es muy similar al anterior, pero en lugar de partir con una variable, se comienza con la totalidad de los atributos. Paso a paso se van eliminando las variables que no son significativas. El método se detiene cuando se obtiene el conjunto de los atributos más representativos.

Por último, el método de *Stepwise* es una mezcla de los dos anteriores. La idea es ejecutar primero uno y después el otro, seleccionando y deseleccionando los atributos las veces que se desee. La desventaja de esta metodología es que no se tienen muy claras las reglas de detención, por lo que, al igual que el primer caso, se pueden obtener resultados poco confiables. Considerando todo lo anterior, se concluye que el método más adecuado es el *Backward Elimination*. La aplicación y resultados de esta técnica se detallarán en el capítulo 5.

Capítulo 5

Resultados y Análisis

Contenido

5.1. Análisis Bivariante	53
5.2. Resultados Todas las Variables	55
5.3. Resultados Selección de Variables	57

En este capítulo se presentan los principales resultados obtenidos. Primero se desarrolla un análisis bivariante donde se estudian las relaciones de la variable independiente con cada una de las variables explicativas. Seguidamente, se exponen los resultados obtenidos y su nivel de significancia al considerar las 18 variables estudiadas inicialmente. Por último, se muestran los resultados conseguidos al aplicar el método *Backward Elimination* de selección de atributos.

5.1. Análisis Bivariante

Antes de entrar de lleno en el modelo de regresión logística, se debe realizar un análisis bivariante, esto es, analizar las relaciones de la variable dependiente con cada uno de los atributos. Entre las técnicas más utilizadas para realizar este tipo de análisis, caben destacar la correlación lineal de Pearson¹.

La correlación es la técnica estadística bivariante que se usa frecuentemente para resumir la fuerza de la asociación entre dos variables. Este método se utiliza específicamente para medir la intensidad de esta la relación. Luego, si se desea analizar la dependencia entre Y y X , entonces la correlación lineal entre ambas indica el grado en el que la variación de la variable Y se relaciona con la variación de la variable X .

Los valores de estos coeficientes van desde -1 a 1. El signo del coeficiente de correlación indica la dirección de la relación entre estos dos atributos, la cual puede ser positiva o negativa, mientras que el valor absoluto indica la fuerza de esta dependencia. Luego, a mayor valor absoluto se tienen relaciones más fuertes.

¹Estos coeficientes se obtienen dividiendo la covarianza de las dos variables por el producto de sus desviaciones estándar.

Por otra parte, cada coeficiente de correlación tiene asociado un nivel de significancia o *p-valor*², concepto asociado a la probabilidad de obtener resultados iguales a los observados. Si el nivel de significación es pequeño, entonces la correlación es significativa y las dos variables están relacionadas linealmente. En caso contrario, se concluye que las variables no están relacionadas. El punto de corte para determinar significancia se define en 0,05, lo que indica que el 95 % de los resultados predichos concuerdan con los resultados reales.

Atributos		Correlación con Y	P-Valor
X1	Dependencias Gubernamentales	0,095	0,309
X2	Fuerzas Policiales	0,064	0,491
X3	Centros de Salud	0,296**	0,001
X4	Establecimientos Educativos	0,027	0,771
X5	Transporte Público	0,022	0,812
X6	Establecimientos Culturales	0,082	0,375
X7	Comida y Entretención	0,203*	0,027
X8	Comercio Masivo	0,417**	0,000
X9	Iglesias	0,027	0,771
X10	Alojamiento	-0,139	0,132
X11	Casas de Cambio	0,292**	0,001
X12	Farmacias	0,548**	0,000
X13	Bancos	0,427**	0,000
X14	Centros de Pago	0,222*	0,016
X15	Supermercados	0,163	0,078
X16	Estacionamientos	0,021	0,825
X17	Áreas Verdes	-0,171	0,063
X18	Paseos Peatonales	0,570**	0,000

* Correlación significativa a nivel de 0,05 (dos colas)

** Correlación significativa a nivel de 0,01 (dos colas)

Tabla 5.1: Correlación Atributos con la Variable Dependiente

De la Tabla 5.1 se puede observar que de los 18 atributos considerados inicialmente hay sólo 8 que presentan una correlación estadísticamente significativa al 95 % con la variable dependiente. Notar que el punto de corte que separa el nivel de significancia 0,01 y 0,05 es el valor 0,250.

Se advierten que los atributos que se correlacionan significativamente a nivel de 0,01 con la variable dependiente, lo hacen en sentido positivo. Luego, la presencia de estos atributos determina un mayor nivel de delincuencia. De éstos se destacan los paseos peatonales y las farmacias debido al alto valor de sus correlaciones. Le siguen en magnitud las variables relativas a bancos y comercio masivo. Por último, se observa que la presencia de centros de salud y casas de cambio también definen un mayor nivel delictual pero en menor medida que los atributos mencionados anteriormente.

En cuanto a los atributos que influyen en la variable dependiente con un nivel de significancia de 0,05, se tiene que también lo hacen de manera positiva. Luego, la presencia de centros de pago y locales de comida y entretención definen un nivel delictual más alto.

Por otro lado, se tienen correlaciones negativas en el caso de los atributos relativos al alojamiento y a las áreas verdes. Luego, a medida que se esté en presencia de estas variables, entonces se espera un nivel delictual bajo. Estos resultados no son significativos, por lo que no se puede generalizar el efecto de estos atributos en la variable dependiente.

Se observa que las dependencias gubernamentales y las fuerzas policiales se correlacionan con la variable dependiente de manera tímida pero positiva. Esto sugiere que en presencia de instituciones

²Mayor detalle del nivel de significancia en el Anexo D.

de gobierno y policiales se podría tener un nivel delictual alto. Sin embargo, estos resultados no son significativos.

5.2. Resultados Todas las Variables

Una vez realizado un análisis bivalente, se procede a ejecutar el modelo de regresión logística. Como se menciona en secciones anteriores, se consideran 18 atributos inicialmente. La aplicación del modelo propuesto permitirá la obtención de coeficientes que determinen el sentido y magnitud de la relación entre estos atributos y la variable dependiente.

Iteración	-2 Log Likelihood	Constantes
Paso 0	1 151,140	-0,644
	2 151,125	-0,668
	3 151,125	-0,668

Tabla 5.2: Iteraciones Bloque Inicial

Paso 1	-2 Log Likelihood	Cox & Snell R^2	Nagelkerke R^2
	83,559	0,436	0,604
Paso 1	χ^2	gl	p -valor
	5,232	8	0,732

Tabla 5.3: Resumen del Modelo y Prueba de Hosmer y Lemeshow

En el bloque inicial se calcula la verosimilitud de un modelo que sólo tiene el término constante (β_0). El estadístico *-2 Log Likelihood*, mide hasta qué punto un modelo se ajusta bien a los datos. Cuanto más pequeño sea este valor, mejor será el ajuste. La Tabla 5.2 muestra un resumen del proceso iterativo de estimación del parámetro β_0 . El proceso ha necesitado de tres ciclos para estimar correctamente el término constante, determinando un valor igual a -0,668.

Después de esta primera etapa, se introducen todas las variables independientes al modelo. La Tabla 5.3 entrega tres medidas de resumen de modelo que sirve para evaluar de forma global la validez del mismo. Se observa que el término *-2LogLikelihood* disminuye considerablemente respecto al paso anterior, lo que es una señal de buen ajuste. El valor del R^2 de Nagelkerke indica que el 60,4 % de la variación de la variable dependiente es explicada por los atributos considerados. Se observa también que el p -valor de la prueba de Hosmer y Lemeshow es un poco menor que 75 %, por lo que el modelo puede considerarse aceptable.

Una vez aplicado el modelo de regresión logística a la totalidad de variables consideradas, se obtuvieron los resultados presentados en la Tabla 5.4. Se puede observar que los p -valores de los estadísticos de Wald asociados a los atributos farmacias, bancos y paseos peatonales son inferiores a 0,05. Por lo tanto, para un nivel de significación de 0,05 se rechaza la hipótesis nula de que dichos coeficientes son nulos. Luego, las variables asociadas a estos parámetros son relevantes en el modelo.

En cuanto a la magnitud de estos coeficientes, se puede ver que el atributo que influye en mayor medida sobre el nivel delictual es la presencia de farmacias. El valor de este coeficiente es positivo lo que indica que al variar este atributo, y suponiendo constante el resto de las variables, entonces el nivel de actividad delictual se altera en la misma dirección. En otras palabras, si se está en presencia de farmacias, entonces se debe esperar un nivel delictual alto.

Otras variables que presentan coeficientes importantes son los paseos peatonales, comercio masivo y bancos. Al igual que el caso anterior, se tiene que los valores de los parámetros asociados a estos atributos influyen de manera positiva en el nivel de actividad delictual.

Por otro lado, se observa que los atributos relativos a iglesias, alojamiento y supermercados tienen un efecto contrario en la variable dependiente. Es decir, la presencia de estas variables motiva una disminución del nivel delictual. Sin embargo, estos coeficientes no son significativos, por lo que se

Atributos	β	D.E.	Wald	gl	p -valor	$\text{Exp}(\beta)$
X1 Dependencias Gubernamentales	0,72	0,87	0,67	1	0,41	2,05
X2 Fuerzas Policiales	0,19	1,58	0,02	1	0,90	1,21
X3 Centros de Salud	1,20	1,15	1,10	1	0,30	3,33
X4 Establecimientos Educativos	0,23	1,30	0,03	1	0,86	1,26
X5 Transporte Público	1,03	0,73	2,02	1	0,16	2,80
X6 Establecimientos Culturales	1,30	0,96	1,86	1	0,17	3,69
X7 Comida y Entretención	0,13	0,66	0,04	1	0,85	1,14
X8 Comercio Masivo	1,57	1,01	2,42	1	0,12	4,79
X9 Iglesias	-0,89	1,48	0,36	1	0,55	0,41
X10 Alojamiento	-1,00	0,86	1,37	1	0,24	0,37
X11 Casas de Cambio	0,57	1,07	0,28	1	0,60	1,76
X12 Farmacias	1,94	0,72	7,35	1	0,01	6,99
X13 Bancos	1,49	0,69	4,67	1	0,03	4,43
X14 Centros de Pago	0,72	0,91	0,62	1	0,43	2,05
X15 Supermercados	-0,39	1,11	0,13	1	0,72	0,67
X16 Estacionamientos	0,43	1,32	0,11	1	0,75	1,54
X17 Áreas Verdes	1,10	0,72	2,28	1	0,13	2,99
X18 Paseos Peatonales	1,80	0,86	4,40	1	0,04	6,02
Constante	-3,68	0,90	16,82	1	0,00	0,03

Tabla 5.4: Coeficientes Asociados a las Variables Predictoras

espera que estos atributos no tengan un efecto importante a la hora de explicar el comportamiento de la variable dependiente.

Por último, se puede apreciar que las dependencias gubernamentales y las fuerzas policiales tienen un efecto positivo con la variable dependiente, situación similar a la que se obtuvo con el análisis bivariante. No obstante, estos resultados son muy poco significativos.

		Y		Porcentaje
		0	1	Correcto
Y	0	72	6	92,3 %
	1	13	27	67,5 %
Promedio			83,9 %	

Tabla 5.5: Tabla de Clasificación 18 Atributos

Se puede comprobar de la Tabla 5.5 que el modelo obtenido clasifica correctamente sólo 27 observaciones de nivel alto de delincuencia de un total de 40, por lo que su sensibilidad sería del 67,5%. Por el contrario, clasifica correctamente una gran parte de las observaciones de bajo nivel delictual, por lo que la especificidad del modelo es del 92,3%. Por último, se tiene que de forma global el modelo ha clasificado correctamente al 83,9% de las observaciones, lo que representa un valor relativamente alto.

Por último, se puede observar en la Figura 5.1 el área bajo la curva ROC para el modelo de regresión logística aplicado a todos los atributos considerados inicialmente. La línea diagonal del gráfico expresa la nula capacidad de discriminación. Cuanto más se acerque la curva al extremo superior izquierdo de la cuadrícula, entonces se tendrá una mejor capacidad de discriminación. De la Tabla 5.6 se observa que el área bajo la curva es de un 0,915 y presenta un p -valor nulo, por lo que se puede concluir que el nivel de discriminación del modelo se considera muy aceptable.

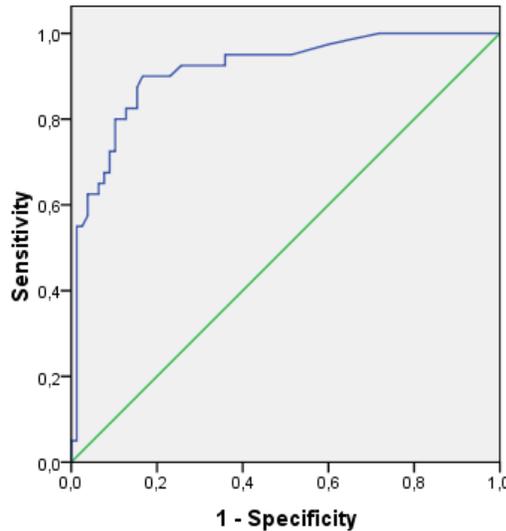


Figura 5.1: Curva ROC 18 Atributos
Fuente: *Elaboración Propia*

área	S.E	<i>p-valor</i>	Intervalo Confianza	
			L.I.	L.S.
0,915	0,028	0,000	0,860	0,969

Tabla 5.6: Área Bajo la Curva ROC 18 Atributos

5.3. Resultados Selección de Variables

Después de un completo análisis realizado en el capítulo anterior, se llega a la conclusión de que el algoritmo más adecuado para la selección de variables es el denominado *Backward Elimination* debido a que posee reglas claras y objetivas de detención. Este método considera todas las variables inicialmente y selecciona, en cada iteración, un atributo que no sea significativo. A continuación, sustrae esta variable del modelo y vuelve a estimar los parámetros. Se detiene al momento de obtener sólo atributos significativos a nivel de 0,05.

De la Tabla 5.7 se puede observar el atributo eliminado en cada una de las iteraciones³. Después de 16 etapas se logra un modelo que considera únicamente variables significativas, lo que sugiere un número un tanto excesivo de atributos contemplados inicialmente.

De la Tabla 5.8 se observa que el término *-2 Log Likelihood* aumenta a medida que hay menos variables. El valor del R^2 de Nagelkerke indica que el 51,1% de la variación de la variable dependiente es explicada por los atributos considerados. Se observa también que el *p-valor* de la prueba de Hosmer y Lemeshow es un muy bajo y menor que 75%, por lo que no se puede considerar al modelo aceptable.

En la Tabla 5.9 se puede observar que se seleccionaron sólo tres atributos. Es interesante ver también que no se presentan coeficientes negativos, por lo que se puede concluir que todas las variables seleccionadas influyen de manera positiva sobre el nivel delictual. Además, se advierte que

³Más detalle de coeficientes y significancia de los atributos en cada iteración en Anexo E.

Paso	Variable Eliminada
2	X2 Fuerzas Policiales
3	X4 Establecimientos Educativos
4	X7 Comida y Entretención
5	X16 Estacionamientos
6	X15 Supermercados
7	X11 Casas de Cambio
8	X9 Iglesias
9	X14 Centros de Pago
10	X1 Dependencias Gubernamentales
11	X5 Transporte Público
12	X17 Áreas Verde
13	X10 Alojamiento
14	X8 Comercio Masivo
15	X6 Establecimientos Culturales
16	X3 Centros de Saluds

Tabla 5.7: Atributos eliminados en cada Iteración

Paso 16	-2 Log Likelihood	Cox & Snell R^2	Nagelkerke R^2
	96,748	0,369	0,511

Paso 16	χ^2	gl	Sig
	2,660	3	0,447

Tabla 5.8: Resumen del Modelo y Prueba de Hosmer y Lemeshow

Atributos	β	D.E.	Wald	gl	p -valor	Exp(β)
X12 Farmacias	1,922	0,613	9,823	1	0,002	6,835
X13 Bancos	1,273	0,576	4,887	1	0,027	3,573
X18 Paseos Peatonales	2,001	0,684	8,567	1	0,003	7,400
Constante	-2,030	0,349	33,876	1	0,000	0,131

Tabla 5.9: Coeficientes Asociados a las Variables Seleccionadas

la presencia de paseos peatonales es la variable más influyente sobre el nivel delictual, resultado que sugiere que la existencia de paseos peatonales en un área determinada hace más atractivo este lugar para los delincuentes, por lo que se debería esperar un nivel de actividad delictual alto. El segundo atributo más influyente sobre el nivel de delitos es la presencia de farmacias, seguido finalmente por la presencia de bancos.

	Y		Porcentaje Correcto
	0	1	
Y	0	74	94,9 %
	1	17	57,5 %
Promedio			82,2 %

Tabla 5.10: Tabla de Clasificación Atributos Seleccionados

Posteriormente, se puede comprobar (Tabla 5.10) que el modelo obtenido clasifica correctamente sólo 23 observaciones de nivel alto de delincuencia de un total de 40, por lo que su sensibilidad sería del 57,5 %. Por el contrario, clasifica correctamente a la mayoría de las observaciones de bajo nivel delictual, por lo que la especificidad del modelo es del 94,9 %. Por último, se tiene que de forma

global el modelo ha clasificado correctamente al 82,2% de las observaciones, lo que representa un valor relativamente alto.

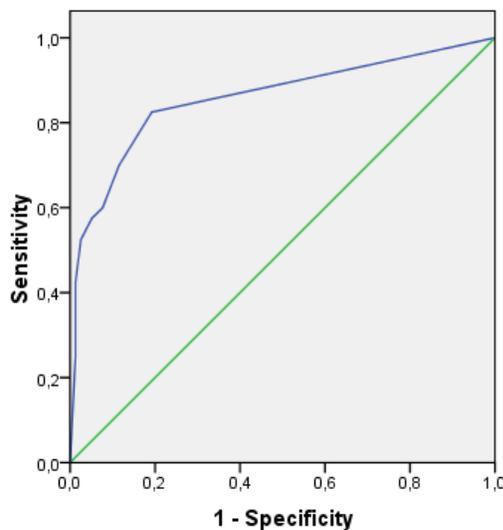


Figura 5.2: Curva ROC Atributos Seleccionados
Fuente: *Elaboración Propia*

Finalmente, se puede observar en la Figura 5.2 que el valor del área bajo la curva ROC para el modelo de regresión logística aplicado a las variables seleccionadas es superior a 0,7, por lo que el nivel de discriminación del modelo se considera un valor aceptable. Además, el $p - valor$ es nulo, lo que sugiere que este resultado sea significativo. Sin embargo, es importante señalar que el valor del área en este escenario es menor que la obtenida en el caso inicial.

Área	S.E	$p - valor$	Intervalo Confianza	
			L.I.	L.S.
0,858	0,041	0,000	0,778	0,937

Tabla 5.11: Área Bajo la Curva ROC Atributos Seleccionados

Capítulo 6

Conclusiones y Trabajos Futuros

Contenido

6.1. Resultados Obtenidos	60
6.2. Futuros Desafíos	61

La criminología es el estudio del fenómeno delictual en la cual convergen múltiples disciplinas científicas. Cada estudio entrega un aporte al entendimiento del crimen desde su propia perspectiva, lo que refleja la gran complejidad del problema de la delincuencia y la seguridad ciudadana.

Este trabajo se enfoca en el desarrollo de un modelo de regresión logística binaria que determine los atributos que influyen en mayor medida en el nivel de actividad criminal que presenta una zona geográfica particular. El objetivo de este estudio es poder determinar el sentido y magnitud de la relación entre las variables mencionadas.

Un aporte muy importante del análisis desarrollado en esta tesis es entregar información objetiva a personas e instituciones encargadas de combatir el crimen. Al conocer los atributos que más influyen en la probabilidad de ocurrencia de un delito, se pueden tomar medidas concretas que ayuden a disminuir este efecto, logrando así una mejor planificación y organización de los recursos por parte de las distintas organizaciones preocupadas de la seguridad ciudadana.

6.1. Resultados Obtenidos

Los resultados obtenidos provienen de la aplicación del modelo de regresión logística binaria a datos reales de denuncias de delitos en la vía pública. Se cuenta con más de 25.000 registros de denuncias de delitos, los cuales son caracterizados por atributos espaciales, temporales y circunstanciales (tipo de delito).

Al procesar y limpiar los datos, se establece utilizar los registros de la Primera Comisaría de Santiago concernientes al delito de hurto durante el período definido entre el 1 de Enero de 2001 y el 31 de Diciembre de 2004, lo que representa 11.529 datos. La zona estudiada corresponde a la parte más concurrida de la capital y se caracteriza principalmente por la existencia de una gran población flotante y una marcada actividad comercial.

Antes de la aplicación del modelo propuesto, se realiza un análisis bivalente. Se calculan las

correlaciones entre la variable dependiente y los atributos. Se observa que sólo 6 atributos están correlacionados de manera significativa a nivel de 0,01. La correlación más fuerte es la que presentan los paseos peatonales con la variable dependiente. Luego, se destacan las correlaciones con las farmacias, bancos y comercio masivo. Otras variables que presentan altas correlaciones con el nivel delictual son los centros de salud y las casas de cambio. Por otra parte, también se puede ver que los centros de pago y los locales de comida y entretención están correlacionados de manera significativa, pero a nivel de 0,05.

La aplicación del modelo de regresión logística se realiza en dos escenarios, los que se diferencian en la cantidad de atributos considerados. En el primer escenario se consideran 18 atributos, los que cumplen con las propiedades de ser observables y cuantificables. Estas variables hacen referencia al comercio, turismo, cultura, transporte, gobierno y de entorno, y fueron elegidas en base a juicios de expertos y estudios similares donde se analizan variables propias del lugar en cuestión.

Los resultados obtenidos en este escenario definen con un nivel de significancia de 0,05 que los atributos que más influyen en la probabilidad de ocurrencia de un delito son la presencia de farmacias, paseos peatonales y bancos. En otras palabras, si estos atributos están presentes en un lugar, entonces se debe esperar un nivel delictual alto en ese sector. Se observa además, que el comercio masivo presenta un coeficiente bastante alto, pero sin significación estadística.

En cuanto a las métricas de bondad de ajuste, se observa que la prueba de Hosmer y Lemeshow determina que el modelo se puede considerar aceptable. Además, de la tabla de clasificación se concluye que el modelo predice correctamente alrededor del 84 %, valor bastante elevado. Por último, se tiene que el área bajo la curva es superior a 0,7, por lo que el nivel de discriminación del modelo se considera muy aceptable.

El segundo escenario corresponde a la aplicación del modelo propuesto, pero ahora considerando únicamente los atributos significativos. La selección de variables se lleva a cabo aplicando el método *backward elimination*, el cual redujo la cantidad de atributos a tres. Los resultados obtenidos en este segundo escenario son bastantes similares al escenario anterior, debido a que se llega a la conclusión de que los atributos más influyentes son los paseos peatonales, farmacias y bancos. El gran aporte de la selección de variables es la significancia que acompaña a los resultados.

Al analizar lo que ocurre con las medidas de bondad de ajuste, se observa que el estadístico de la prueba Hosmer y Lemeshow disminuye, considerablemente, por lo que no se puede considerar a este modelo como aceptable. Por otro lado, se puede ver que la capacidad de predicción es cercana al 82 %, muy similar al caso anterior. Sin embargo, al estudiar lo que ocurre con la curva ROC, se observa que el área calculada es mayor que 0,7, pero inferior que la obtenida en el primer escenario. Se comprueba entonces que la selección de variables no mejora al modelo original de 18 atributos.

En conclusión, se puede decir que en ambos casos se obtuvo que las variables más influyentes en el nivel de actividad delictual son los atributos asociados a los paseos peatonales, farmacias y bancos. En otras palabras, se deduce que a medida que se esté en presencia de estos atributos, entonces se debe esperar un nivel de hurtos alto. Luego, se aconseja destinar una mayor vigilancia policial en las zonas donde estén presentes estas variables y aún con mayor énfasis cuando se encuentren presenten más de una, situación bastante común.

6.2. Futuros Desafíos

El modelo propuesto se aplica exclusivamente en el terreno de la Primera Comisaría de Santiago, por lo que se propone como trabajo futuro una generalización para otros sectores. La comuna de

Santiago presenta características bastante comunes a otras zonas de alta actividad comercial, por lo que la extensión de esta herramienta a otras áreas de interés podría ser un desafío interesante.

Una buena forma de complementar los datos empleados en este estudio, es la utilización de los resultados de distintas encuestas de victimización, por lo que se propone la inclusión de la información obtenida por otros medios.

El objetivo de este estudio era desarrollar un modelo de regresión logística binaria considerando únicamente las características propias del lugar de estudio y su atractividad. Luego, la inclusión de datos de los delincuentes y del entorno donde viven, enriquecerían de manera considerable el análisis realizado.

Esta tesis enfocó su aplicación en los delitos de tipo hurto. Luego, la modelación del comportamiento delictual considerando otros tipos de crímenes queda propuesta como un estudio en el que se puede seguir ampliando la investigación.

Por otra parte, la sinergia de estos modelos con otras herramientas, tanto matemáticas como computacionales, sería un gran desafío futuro. Los resultados obtenidos podrían ser implementados en modelos donde se aplique la teoría de juegos, ayudando a definir el nivel de atractividad de una zona y así, poder identificar de mejor manera el comportamiento de los delincuentes.

Otra área donde se pueden aplicar estos modelos es la simulación computacional. Al igual que en el caso anterior, al comprender los atributos que motivan a los criminales se podrían definir reglas de comportamiento y simular experimentos difíciles o imposibles de realizar en la vida real.

Por otra parte, se motiva a incluir otros atributos que no fueron considerados en este estudio debido a la dificultad de medirlos. Hay variables que claramente hacen más atractiva una zona para un delincuente, así como una pobre iluminación, calles angostas o existencia de árboles o estructuras donde sea fácil esconderse. Luego, la inclusión de estos atributos podría ser un desafío interesante.

Por último, se propone incluir un análisis dinámico tanto de las variables independientes como de los delitos. Esto puede lograrse realizando observaciones de estos atributos cada cierto tiempo y entre periodos regulares. Lo anterior serviría para observar cambios importantes en el comportamiento de los delincuentes al incluirse o eliminarse alguna variable del entorno.

Bibliografía

- [1] “Estudio Nacional de Opinión Pública N°63,” *Centro de Estudios Públicos*, Noviembre-Diciembre 2010.
- [2] “VII Encuesta Nacional Urbana de Seguridad Ciudadana,” *Instituto Nacional de Estadísticas de Chile (INE)*, 2010.
- [3] “Índice Paz Ciudadana-Adimark,” *Fundación Paz Ciudadana & Adimark-GfK*, Abril 2011.
- [4] L. Cohen, M. Felson, “Social Change and Crime Rate Trends: A Routine Activity Approach,” *American Sociological Review*, vol. 44, no. 4, pp. 588–608, 1979.
- [5] “Anuario de Estadísticas Criminales,” *Fundación Paz Ciudadana*, 2008.
- [6] E. Oviedo, “Violencia Delictual y Ciudad: Evolución histórica de hurtos y robos en siete ciudades chilenas y análisis de su distribución intercomunal en el Gran Santiago,” *Revista Proposiciones*, vol. 22, pp. 173–207, 1993.
- [7] “Plan Estratégico de Desarrollo Comunal de Santiago,” *Municipalidad de Santiago*, 2010.
- [8] “Cuenta Pública 2009: Santiago,” *Carabineros de Chile*, Septiembre 2009.
- [9] “Informe Final de Evaluación Programa Plan Cuadrante de Seguridad Preventiva,” *Ministerio de Defensa, Subsecretaría de Carabineros*, Julio 2007.
- [10] A. Quérel, “L’Homme et le Développement de ses Facultés, ou Essai de Physique Sociale,” *Adamant Media Corporation*, 1835.
- [11] A.M. Guerry, “Essai sur la Statistique Morale de la France,” *Kessinger Publishing, Crochard, Paris*, 1833.
- [12] W. Spelman, “Criminal Careers of Public Places,” *Final Report to the U.S. National Institute of Justice, Austin, TX: LBJ School of Public Affairs, University of Texas at Austin*, 1995.
- [13] T. Adams-Fuller, “Historical Homicide Hot Spots: The Case of Three Cities,” *Doctoral Dissertation, Howard University, Washington, DC, USA*, 2001.
- [14] D. Paulsen, “Applying Neural Network analysis to Predict Commercial robbery victimization across different cities,” *Paper presented at the annual meeting of the American Society of Criminology, Los Angeles Convention Center, Los Angeles, CA, USA*, 2007.
- [15] H. Liua, D.E. Brown, “Criminal Incident Prediction Using a Point-Pattern-Based Density Model,” *International Journal of Forecasting*, vol. 19, p. 603–622, 2003.

- [16] C. Fraley, A. Raftery, “How Many Clusters? Which Clustering Method? Answers Via Model-Based Cluster Analysis,” *The Computer Journal, Department of Statistics, University of Washington, USA*, vol. 41, no. 8, pp. 578–588.
- [17] P. Wang, R. Mathieu, J. Ke, H.J. Cai, “Predicting Criminal Recidivism with Support Vector Machine,” *2010 International Conference Management and Service Science (MASS), Department Computer Information Systems, James Madison Univ., Harrisonburg, VA, USA*, pp. 1–9, 2010.
- [18] K. Baumgartner et al, “Constructing Bayesian Networks for Criminal Profiling from Limited Data,” *Knowledge-Based Systems Journal*, vol. 21, no. 7, 2008.
- [19] M. Jain, J. Tsai, J. Pita, C. Kiekintveld, S. Rathi, F. Ordóñez, M. Tambe, “Software Assistants for Randomized Patrol Planning for The LAX Airport Police and The Federal Air Marshals Service,” *Interfaces Journal*, vol. 40, no. 4, 2010.
- [20] N. Malleson, A. Heppenstall, L. See, “Crime Reduction Through Simulation: An Agent-Based Model of Burglary,” *Computers, Environment and Urban Systems Journal*, vol. 34, no. 3, pp. 236–250, 2010.
- [21] T. Bosse and C. Gerritsen, “Comparing Crime Prevention Strategies by Agent-Based Simulation,” *Proceeding WI-IAT '09 Proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology*, vol. 2, pp. 491–496.
- [22] D. McFadden, “The Measurement of Urban Travel Demand,” *Journal of Public Economics, Department of Economics, University of California, Berkeley, USA*, vol. 3, pp. 303–328, 1974.
- [23] D.A. Hensher, P.O. Barnard, T.P. Truong, “The Role of Stated Preference Methods in Studies of Travel Choice,” *Journal of Transport Economics and Policy*, vol. 22, no. 1, pp. 45–58, 1988.
- [24] P. Schmidt, R.P. Strauss, “The Prediction of Occupation Using Multiple Logit Models,” *International Economic Review*, vol. 16, no. 2, pp. 471–486, 1975.
- [25] M.J. Boskin, “A Conditional Logit Model of Occupational Choice,” *Journal of Political Economy*, vol. 82, no. 2, pp. 389–398, 1974.
- [26] J.V. Terza, “Ordinal Probit: A Generalization,” *Communications in Statistics, Theory and Methods*, vol. 14, pp. 1–12, 1985.
- [27] G.S. Becker, “The Economic Approach to Human Behavior,” *The University of Chicago Press, Chicago, USA*, 1976.
- [28] S.R. Chandukala, J. Kim, T.Otter, P.E. Rossi, G.M. Allenby, “Choice Models in Marketing: Economic Assumptions, Challenges and Trends,” *Foundations and Trends in Marketing Journal, Fisher College of Business Working Paper Series*, vol. 2, no. 2, 2007.
- [29] J.D. Ortuzar, L.G. Willumsen, “Modelling Transport,” *John Wiley & Sons, West Sussex, Inglaterra*, 2002.
- [30] “FSUTMS Mode Choice Modeling: Factors Affecting Transit Use and Access,” *Report No NCTR 397-02, prepared for National Center for Transit Research in cooperation with Florida Department of Transportation, Florida, USA*, 2002.
- [31] R. Claassen, A. Tegene, “Agricultural Land Use Choice: A Discrete Choice Approach,” *Agricultural and Resource Economics Review, Northeastern Agricultural and Resource Economics Association*, vol. 28, no. 1, 1999.

- [32] Y. Xue and D.E. Brown, "Spatial Analysis With Preference Specification of Latent Decision Makers for Criminal Event Prediction," *Decision Support Systems*, vol. 41, no. 3, p. 560–573, 2006.
- [33] W. Bernasco, "Modeling Micro-Level Crime Location Choice: Application of the Discrete Choice Framework to Crime at Places," *Journal of Quantitative Criminology, Netherlands Institute for the Study of Crime and Law Enforcement (NSCR) P.O. Box 71304 1008 BH Amsterdam The Netherlands, Springer Netherlands*, vol. 26, no. 1, pp. 113–138, 2010.
- [34] N.J. Horton, N.M. Laird, "Maximum Likelihood Analysis of Logistic Regression Models With Incomplete Covariate Data and Auxiliary Information," *Biometrics Journal*, vol. 57, no. 1, pp. 34–42, 2001.
- [35] D.W. Hosmer, S. Lemeshov, "Applied Logistic Regression," *Wiley Series in Probability and Statistics, John Wiley and Sons*, 2000.
- [36] S.J. Mason, N.E. Graham, "Areas Beneath the Relative Operating Characteristics (ROC) and Relative Operating Levels (ROL) Curves: Statistical Significance and Interpretation," *Quarterly Journal of the Royal Meteorological Society*, vol. 128, no. 584, p. 2145–2166, 2002.
- [37] A.A. Pérez, R. Kizys, L.M. Manzanedo, "Regresión Logística Binaria," *Proyecto e-Math, Universitat Oberta de Catalunya*.
- [38] W. Bernasco, P. Nieuwbeerta, "How do Residential Burglars Select Target Areas? A New Approach to the Analysis of Criminal Location Choice," *The British Journal of Criminology*, vol. 45, no. 3, pp. 296–315, 2005.
- [39] D.M. Lomba, "A Theoretical Model of a Criminal's Location Choice," *International Journal of Business Research, International Academy of Business and Economics, Chicago, USA*, vol. 8, no. 3, 2008.
- [40] J. Deutsch, S. Hakim, J. Weinblatt, "A Micro Model of the Criminal's Location Choice," *Journal of Urban Economics, Elsevier BV*, vol. 22, no. 2, pp. 198–208, 1987.
- [41] J. Deutsch, G.S. Epstein, "Changing a Decision Taken under Uncertainty: The Case of the Criminal's Location Choice," *Urban Studies, Department of Economics, Bar-Ilan University, Israel*, vol. 35, no. 8, pp. 1335–1343, 1998.
- [42] B.F.J. Manly, "The Design and Analysis of Research Studies," 1992.
- [43] A. Verma, R. Ramyaa, S. Maru, Y. Fan, R. Singh, "Rationalizing police patrol beats using Voronoi Tessellations," *IEEE International Conference of Intelligence and Security Informatics (ISI), Department of Criminal Justice, Indiana University, Bloomington, USA*, pp. 165–167, 2010.
- [44] M.H. Chen and D.K. Dey, "Variable Selection for Multivariate Logistic Regression Models," *Journal of Statistical Planning and Inference, Department of Statistics, University of Connecticut, USA*, vol. 111, no. 1-2, pp. 37–55, 2003.

Apéndices

Apéndice A

Método de Newton-Raphson

El método de Newton-Raphson es un algoritmo eficiente utilizado para encontrar aproximaciones de los ceros o raíces de una función real. También se usa para hallar el máximo o mínimo de una función, encontrando los ceros de su primera derivada.

Este es un método abierto, en el sentido de que su convergencia global no está garantizada. La única manera de alcanzar la convergencia es seleccionando un valor inicial lo suficientemente cercano a la raíz buscada. Así, si la función presenta múltiples puntos de inflexión o pendientes grandes en el entorno de esta raíz, entonces las probabilidades de que el algoritmo diverja aumentan. Luego, el método linealiza la función por la recta tangente en ese valor supuesto. La abscisa en el origen de dicha recta será, según el método, una mejor aproximación de la raíz que el valor anterior. Se realizarán sucesivas iteraciones hasta que el método haya convergido lo suficiente.

Sea $f : [a, b] \rightarrow \mathbb{R}$ una función derivable definida en el intervalo real $[a, b]$ y con f' su derivada. Si se empieza con un valor inicial x_0 , entonces se tiene:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

En cuanto a la obtención del algoritmo, tres son las formas principales por las que tradicionalmente se alcanza.

La primera de ellas es una simple interpretación geométrica. En efecto, atendiendo al desarrollo geométrico del método de la secante, podría pensarse en que si los puntos de iteración están lo suficientemente cerca (a una distancia infinitesimal), entonces la secante se sustituye por la tangente a la curva en el punto. Así pues, si por un punto de iteración trazamos la tangente a la curva, por extensión con el método de la secante, el nuevo punto de iteración se tomará como la abscisa en el origen de la tangente. Esto es equivalente a linealizar la función, es decir, f se reemplaza por una recta tal que contiene al punto $(x_0, f(x_0))$ y cuya pendiente coincide con la derivada de la función en ese punto, $f'(x_0)$. La nueva aproximación a la raíz, que sería x_1 , se logra al intersectar la función lineal con el eje X de abscisas. Matemáticamente:

$$f'(x_n) = \frac{f(x_n)}{x_n - x_{n+1}}$$

Una forma alternativa de obtener el algoritmo de Newton-Raphson es desarrollando la función $f(x)$

en serie de Taylor, para un entorno del punto x_n .

$$f(x) = f(x_n) + f'(x_n)(x - x_n) + (x - x_n)^2 \frac{f''(x_n)}{2!} + \dots$$

Si se trunca el desarrollo a partir del término de grado 2 y evaluamos en x_{n+1} se obtiene:

$$f(x_{n+1}) = f(x_n) + f'(x_n)(x_{n+1} - x_n)$$

Si además se acepta que x_{n+1} tiende a la raíz, se ha de cumplir que $f(x_{n+1}) = 0$. Luego, sustituyendo en la expresión anterior, se obtiene el algoritmo.

Hay que indicar que el método de Newton-Raphson puede interpretarse como un método de iteración de punto fijo. Así, si se asume $f(x) = 0$ entonces puede considerarse como un método de iteración de punto fijo.

$$g(x) = x + h(x)f(x)$$

Luego, se escoge $h(x)$ de manera que $g'(r) = 0$ donde r es la raíz buscada. Y dado que $g'(r)$ es:

$$g'(r) = 1 + h'(r)f(r) + h(r)f'(r) = 1 + h(r)f'(r)$$

Entonces:

$$h(r) = \frac{-1}{f'(r)}$$

Así, como $h(x)$ no tiene que ser única, se escoge de la forma más sencilla:

$$h(x) = \frac{-1}{f'(x)}$$

Por tanto, imponiendo subíndices, se obtiene la expresión que coincide con la del algoritmo de Newton-Raphson:

$$g(x_n) = x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

Apéndice B

Función de Verosimilitud

Como bien se muestra en la ecuación 3.13, se tiene que la función de verosimilitud es una generalización del modelo binomial logit.

$$\ln[L] = \sum_m \sum_j d_{mj} \ln[P_j(\beta, X_j)]$$

Donde,

$P_j = P(y_j = 1|X_j)$: probabilidad de que la alternativa j sea escogida dadas sus características X_j ,

$$d_{mj} = \begin{cases} 1, & \text{si el delito } k\text{-ésimo ocurrió en la alternativa } j \\ 0, & \text{en caso contrario} \end{cases}$$

Luego, la derivada tiene la siguiente forma:

$$\frac{\partial \ln[L]}{\partial \beta_i} = \sum_m (d_{mj} - P_j) x_{mj} \quad \forall j$$

Al igualar a cero la ecuación anterior, se obtiene lo siguiente:

$$\sum_m d_{mj} x_{mj} = \sum_m P_j x_{mj} \quad \forall j$$

Al dividir la expresión anterior por el total de observaciones, entonces se concluye que el valor estimado es igual al promedio de las observaciones, es decir:

$$\bar{x}_j = \hat{x}_j(\beta_j) \quad \forall j$$

Apéndice C

Diagramas de Voronoi

Concepto

Los polígonos de Voronoi es un método de interpolación bastante simple y se basa en la distancia euclidiana. Se originan al unir los puntos entre sí, trazando las simetrales¹ de los segmento de unión. Las intersecciones de estas simetrales determinan una serie de polígonos en un espacio bidimensional alrededor de un conjunto de puntos de control, de manera que el perímetro de los polígonos generados sea equidistante a los puntos vecinos.

Definición

Para cada conjunto discreto S de puntos en un espacio euclídeo y para casi² todo punto x , existe un punto de S que es el más cercano a x .

Si S contiene sólo dos puntos, a y b , entonces el conjunto de todos los puntos que equidistan de ambos es un hiperplano de codimensión 1.

Ese hiperplano es la frontera entre los puntos más cercanos a a que a b , y los puntos más cercanos a b que a a . De hecho ese hiperplano es el plano bisector del segmento que une a y b .

Más en general, el conjunto de puntos más cercanos a un punto c de S que a ningún otro punto de S (cuenca de atracción de c) es el interior de un politopo convexo (posiblemente no acotado) llamado Celda de Voronoi de c . El conjunto de todos esos politopos constituye una teselación completa del espacio euclídeo, denominado Teselación de Voronoi asociada a S .

Propiedades

1. El gráfico dual de un diagrama de Voronoi se corresponde con la triangulación de Delaunay para el mismo conjunto de puntos de S .
2. El par de puntos más cercano corresponde a dos células adyacentes en el diagrama de Voronoi.
3. Dos puntos son adyacentes en el *convex hull* si y sólo si sus células de Voronoi comparten un lado de longitud infinita.

¹Simetral: perpendicular que pasa por el punto medio de un segmento.

²Aquí el término *casi* se usa para indicar que existen excepciones en las cuales x puede equidistar de dos o más puntos de S .

Apéndice D

P-Valor

Definición

La elección del nivel de significación es en cierta forma arbitraria. Sin embargo, una vez obtenida la muestra, se puede calcular una cantidad que permite resumir el resultado del experimento de manera objetiva. Esta cantidad es el *p-valor* que corresponde al nivel de significación más pequeño posible que puede escogerse, para el cual todavía se aceptaría la hipótesis alternativa con las observaciones actuales. Cualquier nivel de significación escogido inferior al *p-valor* (simbólicamente pv) comporta aceptar H_0 . Claramente, al ser una probabilidad, se cumple que:

$$0 \leq pv \leq 1$$

El *p-valor* es una medida directa de lo verosímil que resulta obtener una muestra como la actual si es cierta H_0 . Los valores pequeños indican que es muy infrecuente obtener una muestra como la actual, en cambio, los valores altos que es frecuente. El *p-valor* se emplea para indicar cuánto (o cuán poco) contradice la muestra actual la hipótesis alternativa.

Informar sobre cuál es el *p-valor* tiene la ventaja de permitir que cualquiera decida que hipótesis acepta basándose en su propio nivel de riesgo α . Esto no es posible cuando se informa indicando sólo el resultado de la decisión, es decir, si se acepta o se rechaza H_0 con un *alpha* fijo.

Al proporcionar el *p-valor* obtenido con la muestra actual, la decisión se hará de acuerdo a la regla siguiente:

Si $pv \leq \alpha$, *aceptar* H_1

Si $pv \geq \alpha$, *aceptar* H_0

Apéndice E

Coeficientes y Significancia Método Backward Elimination

	Atributos	B	S,E,	Wald	df	Sig	Exp(B)
Step 1	X1	0,717	0,874	0,672	1	0,412	2,048
	X2	0,194	1,578	0,015	1	0,902	1,214
	X3	1,204	1,150	1,096	1	0,295	3,332
	X4	0,230	1,303	0,031	1	0,860	1,259
	X5	1,030	0,725	2,016	1	0,156	2,801
	X6	1,305	0,956	1,862	1	0,172	3,686
	X7	0,128	0,664	0,037	1	0,847	1,137
	X8	1,567	1,006	2,424	1	0,120	4,790
	X9	-0,889	1,478	0,361	1	0,548	0,411
	X10	-1,003	0,856	1,373	1	0,241	0,367
	X11	0,566	1,075	0,278	1	0,598	1,762
	X12	1,945	0,717	7,348	1	0,007	6,992
	X13	1,487	0,688	4,671	1	0,031	4,425
	X14	0,719	0,911	0,623	1	0,430	2,053
	X15	-0,395	1,106	0,127	1	0,721	0,674
	X16	0,430	1,323	0,106	1	0,745	1,537
	X17	1,096	0,725	2,284	1	0,131	2,991
	X18	1,796	0,856	4,402	1	0,036	6,024
	Constant	-3,676	0,896	16,824	1	0,000	0,025

Step 2	X1	0,710	0,872	0,662	1	0,416	2,033	
	X3	1,217	1,136	1,148	1	0,284	3,378	
	X4	0,222	1,301	0,029	1	0,864	1,249	
	X5	1,026	0,725	2,000	1	0,157	2,789	
	X6	1,301	0,955	1,857	1	0,173	3,673	
	X7	0,127	0,666	0,036	1	0,849	1,135	
	X8	1,565	1,006	2,422	1	0,120	4,783	
	X9	-0,871	1,464	0,354	1	0,552	0,419	
	X10	-1,004	0,860	1,363	1	0,243	0,366	
	X11	0,570	1,081	0,278	1	0,598	1,769	
	X12	1,951	0,716	7,424	1	0,006	7,032	
	X13	1,481	0,686	4,657	1	0,031	4,396	
	X14	0,715	0,911	0,617	1	0,432	2,045	
	X15	-0,388	1,104	0,123	1	0,725	0,678	
	X16	0,416	1,318	0,100	1	0,752	1,517	
	X17	1,102	0,723	2,324	1	0,127	3,011	
	X18	1,797	0,855	4,419	1	0,036	6,034	
	Constant	-3,667	0,893	16,846	1	0,000	0,026	
	Step 3	X1	0,735	0,859	0,732	1	0,392	2,085
X3		1,195	1,127	1,123	1	0,289	3,302	
X5		1,030	0,723	2,030	1	0,154	2,802	
X6		1,273	0,938	1,841	1	0,175	3,571	
X7		0,140	0,660	0,045	1	0,832	1,150	
X8		1,574	1,003	2,463	1	0,117	4,824	
X9		-0,823	1,442	0,326	1	0,568	0,439	
X10		-1,030	0,849	1,473	1	0,225	0,357	
X11		0,543	1,069	0,258	1	0,611	1,722	
X12		1,962	0,715	7,525	1	0,006	7,111	
X13		1,476	0,686	4,634	1	0,031	4,375	
X14		0,699	0,906	0,596	1	0,440	2,012	
X15		-0,403	1,100	0,134	1	0,715	0,669	
X16		0,397	1,316	0,091	1	0,763	1,488	
X17		1,095	0,722	2,302	1	0,129	2,990	
X18		1,777	0,845	4,420	1	0,036	5,913	
Constant		-3,651	0,886	16,972	1	0,000	0,026	
Step 4		X1	0,743	0,857	0,753	1	0,386	2,103
		X3	1,140	1,096	1,083	1	0,298	3,128
	X5	1,051	0,718	2,138	1	0,144	2,859	
	X6	1,281	0,939	1,864	1	0,172	3,602	
	X8	1,576	0,999	2,486	1	0,115	4,834	
	X9	-0,895	1,414	0,401	1	0,527	0,409	
	X10	-0,981	0,812	1,457	1	0,227	0,375	
	X11	0,583	1,049	0,309	1	0,579	1,791	
	X12	1,983	0,711	7,778	1	0,005	7,263	
	X13	1,465	0,683	4,606	1	0,032	4,326	
	X14	0,722	0,901	0,642	1	0,423	2,058	
	X15	-0,433	1,093	0,157	1	0,692	0,649	
	X16	0,331	1,276	0,067	1	0,795	1,392	
	X17	1,091	0,722	2,286	1	0,131	2,978	
	X18	1,833	0,807	5,151	1	0,023	6,251	
	Constant	-3,611	0,864	17,461	1	0,000	0,027	

Step 5	X1	0,746	0,855	0,760	1	0,383	2,108	
	X3	1,082	1,072	1,019	1	0,313	2,951	
	X5	1,074	0,713	2,267	1	0,132	2,926	
	X6	1,249	0,928	1,809	1	0,179	3,486	
	X8	1,546	0,983	2,473	1	0,116	4,692	
	X9	-0,813	1,379	0,347	1	0,556	0,444	
	X10	-0,976	0,815	1,433	1	0,231	0,377	
	X11	0,597	1,044	0,327	1	0,568	1,816	
	X12	1,992	0,712	7,830	1	0,005	7,333	
	X13	1,431	0,669	4,576	1	0,032	4,184	
	X14	0,754	0,890	0,717	1	0,397	2,126	
	X15	-0,475	1,082	0,192	1	0,661	0,622	
	X17	1,081	0,720	2,257	1	0,133	2,949	
	X18	1,858	0,804	5,345	1	0,021	6,409	
	Constant	-3,595	0,861	17,454	1	0,000	0,027	
	Step 6	X1	0,668	0,836	0,638	1	0,424	1,949
		X3	1,103	1,059	1,085	1	0,298	3,013
		X5	0,999	0,688	2,113	1	0,146	2,717
X6		1,281	0,914	1,965	1	0,161	3,599	
X8		1,518	0,973	2,434	1	0,119	4,562	
X9		-0,765	1,378	0,308	1	0,579	0,465	
X10		-0,973	0,819	1,412	1	0,235	0,378	
X11		0,526	1,032	0,259	1	0,611	1,691	
X12		1,938	0,699	7,679	1	0,006	6,944	
X13		1,415	0,668	4,481	1	0,034	4,117	
X14		0,583	0,799	0,534	1	0,465	1,792	
X17		1,064	0,719	2,190	1	0,139	2,897	
X18		1,884	0,800	5,546	1	0,019	6,581	
Constant		-3,530	0,838	17,737	1	0,000	0,029	
Step 7		X1	0,761	0,809	0,884	1	0,347	2,141
		X3	1,075	1,053	1,041	1	0,308	2,929
		X5	0,983	0,687	2,048	1	0,152	2,671
		X6	1,248	0,908	1,890	1	0,169	3,483
	X8	1,573	0,962	2,672	1	0,102	4,820	
	X9	-0,809	1,384	0,342	1	0,559	0,445	
	X10	-1,030	0,822	1,568	1	0,210	0,357	
	X12	2,022	0,683	8,760	1	0,003	7,556	
	X13	1,439	0,664	4,696	1	0,030	4,216	
	X14	0,507	0,791	0,411	1	0,522	1,660	
	X17	1,063	0,719	2,188	1	0,139	2,896	
	X18	1,941	0,795	5,955	1	0,015	6,965	
	Constant	-3,502	0,837	17,515	1	0,000	0,030	
	Step 8	X1	0,770	0,803	0,919	1	0,338	2,159
		X3	1,043	1,046	0,993	1	0,319	2,836
		X5	0,959	0,677	2,005	1	0,157	2,608
		X6	1,247	0,899	1,925	1	0,165	3,479
		X8	1,430	0,913	2,454	1	0,117	4,180
X10		-1,078	0,831	1,681	1	0,195	0,340	
X12		2,027	0,682	8,830	1	0,003	7,590	
X13		1,501	0,654	5,276	1	0,022	4,486	
X14		0,487	0,786	0,385	1	0,535	1,628	
X17		1,089	0,719	2,293	1	0,130	2,970	
X18		1,887	0,777	5,902	1	0,015	6,597	
Constant		-3,519	0,832	17,895	1	0,000	0,030	

Step 9	X1	0,818	0,799	1,048	1	0,306	2,266
	X3	1,156	1,045	1,222	1	0,269	3,176
	X5	1,000	0,679	2,168	1	0,141	2,717
	X6	1,225	0,892	1,886	1	0,170	3,403
	X8	1,513	0,902	2,814	1	0,093	4,539
	X10	-1,089	0,836	1,694	1	0,193	0,337
	X12	2,054	0,682	9,079	1	0,003	7,798
	X13	1,488	0,652	5,215	1	0,022	4,428
	X17	1,060	0,720	2,171	1	0,141	2,887
	X18	1,912	0,782	5,982	1	0,014	6,766
Constant	-3,491	0,838	17,374	1	0,000	0,030	
Step 10	X3	1,332	1,076	1,530	1	0,216	3,787
	X5	0,778	0,623	1,561	1	0,211	2,177
	X6	1,396	0,843	2,742	1	0,098	4,038
	X8	1,384	0,893	2,401	1	0,121	3,993
	X10	-1,089	0,824	1,749	1	0,186	0,336
	X12	1,935	0,681	8,080	1	0,004	6,923
	X13	1,673	0,630	7,044	1	0,008	5,328
	X17	0,985	0,710	1,927	1	0,165	2,678
	X18	1,863	0,775	5,775	1	0,016	6,446
	Constant	-3,199	0,752	18,098	1	0,000	0,041
Step 11	X3	1,536	1,075	2,040	1	0,153	4,645
	X6	1,300	0,810	2,572	1	0,109	3,669
	X8	1,464	0,913	2,570	1	0,109	4,323
	X10	-1,033	0,817	1,600	1	0,206	0,356
	X12	1,962	0,678	8,368	1	0,004	7,114
	X13	1,651	0,625	6,970	1	0,008	5,210
	X17	0,784	0,683	1,318	1	0,251	2,189
	X18	1,688	0,751	5,045	1	0,025	5,407
	Constant	-2,633	0,566	21,606	1	0,000	0,072
	Step 12	X3	1,468	1,033	2,017	1	0,156
X6		1,081	0,765	1,997	1	0,158	2,947
X8		1,233	0,855	2,080	1	0,149	3,432
X10		-1,003	0,814	1,517	1	0,218	0,367
X12		1,894	0,669	8,022	1	0,005	6,648
X13		1,525	0,608	6,295	1	0,012	4,596
X18		1,499	0,722	4,313	1	0,038	4,476
Constant		-2,279	0,440	26,886	1	0,000	0,102
Step 13	X3	1,528	1,043	2,146	1	0,143	4,611
	X6	1,104	0,754	2,145	1	0,143	3,017
	X8	1,107	0,855	1,679	1	0,195	3,026
	X12	1,856	0,662	7,861	1	0,005	6,395
	X13	1,487	0,603	6,079	1	0,014	4,422
	X18	1,572	0,715	4,828	1	0,028	4,815
	Constant	-2,420	0,428	31,988	1	0,000	0,089
	Step 14	X3	1,769	1,046	2,862	1	0,091
X6		1,047	0,741	1,998	1	0,157	2,849
X12		1,960	0,653	9,021	1	0,003	7,102
X13		1,477	0,597	6,117	1	0,013	4,378
X18		1,762	0,696	6,416	1	0,011	5,822
Constant		-2,350	0,417	31,753	1	0,000	0,095
Step 15	X3	1,612	1,029	2,456	1	0,117	5,014
	X12	1,753	0,627	7,810	1	0,005	5,770
	X13	1,426	0,584	5,963	1	0,015	4,161
	X18	1,907	0,688	7,684	1	0,006	6,733
	Constant	-2,143	0,368	33,865	1	0,000	0,117
Step 16	X12	1,922	0,613	9,823	1	0,002	6,835
	X13	1,273	0,576	4,887	1	0,027	3,573
	X18	2,001	0,684	8,567	1	0,003	7,400
	Constant	-2,030	0,349	33,876	1	0,000	0,131

Tabla E.1: Coeficientes y Significancia de los Atributos en cada Iteración Backward Elimination

Apéndice F

Códigos Fuentes

Generación Modelo

```
Leer los archivos [hurtos atributos esquinas] = ReadFiles('../Datos/BBDD_Hurtos_Numerica.xls',  
'../Datos/Atributos_Final_10.xlsx','../Datos/esquinas.xlsx');
```

```
Calcular bounding box xmin = min([atributos(:,1);  
hurtos(:,1)]);  
xmax = max([atributos(:,1);  
hurtos(:,1)]);  
ymin = min([atributos(:,2);  
hurtos(:,2)]);  
ymax = max([atributos(:,2);  
hurtos(:,2)]);
```

```
boundingBox = [xmin ymin;  
xmin ymax ;  
xmax ymin ;  
xmax ymax];  
k = convhull (boundingBox(:,1),boundingBox(:,2));  
boundingBox = [boundingBox(k,1),boundingBox(k,2)];
```

```
% Generar Diagramas de Voronoi [vx, vy] = voronoi(esquinas(:,1),esquinas(:,2));  
poligonos = Grilla(esquinas,'voronoi',boundingBox);
```

```
% Contar Hurtos y Atributos poligonos = CalculoEstadisticas(poligonos,hurtos,atributos);
```

```
% Generación de Matrices X e Y. % Yi es la cantidad de hurtos en el poligono i. % Xij es la cantidad  
de atributos j en el poligono i. threshold = 95; [Y,X] = GeneracionMatrices(poligonos,threshold);
```

Leer Archivos

```
function [hurtosOut atributos esquinas] = ReadFiles(hurtosFileName,atributosFileName, esquinas-
FileName) if nargin < 3 disp('Error: Se debe proveer al menos 3 nombres de archivos. Para: Hurtos,
atributos y esquinas') end [hurtos,hurtosTXT] = xlsread(hurtosFileName);
```

```
% Filtrado de Hurtos % Se define la variable que define los filtros filterRules. % La primera fila
contiene los campos a filtrar y la segunda contiene %la opcion de filtrado. % Cada fila siguiente
posee los valores aceptados o filtrados.
```

```
% Opciones de filtrado: % 'Solamente' Permanecen en la base de datos solo los hurtos que % cumplan
dicho campo a la perfección. % 'Eliminar' Elimina de la base de datos los hurtos que posean dicho
campo.
```

```
filterRules = 'DIA_DE_LA_SEMANA','Solamente',5 ;
numberOfFilterRules = size(filterRules,1);
rulesColumnIndexes = zeros(numberOfFilterRules,1);
```

```
for i = 1:numberOfFilterRules for j = 1:size(hurtosTXT,2) if strcmp(filterRules{i,1},hurtosTXT{1,j})
rulesColumnIndexes(i,1)=j;
end end end
```

```
i = 1;
while i <= size(hurtos,1) for j = 1:numberOfFilterRules if strcmp(filterRules{j,2},'Solamente') delete
= 1;
elseif strcmp(filterRules{j,2},'Eliminar') delete = 0;
end for k = 1:(size(filterRules{j,2})-2) if strcmp(filterRules{j,2},'Solamente') &&
(hurtos(i,rulesColumnIndexes(j,1))==filterRules{j,1,k+2}) delete = 0;
elseif strcmp(filterRules{j,2},'Eliminar') &&
(hurtos(i,rulesColumnIndexes(j,1))==filterRules{j,1,k+2}) delete = 1;
break;
end end if delete if(i == 1) hurtos = hurtos(2:end,:);
else hurtos = [hurtos(1:i-1,:);
hurtos(i+1:end,:)];
end i=i-1;
break;
end end i=i+1;
end hurtosOut = hurtos(:,11:12);
```

```
esquinas = xlsread(esquinasFileName);
atributos = xlsread(atributosFileName);
```

```
end
```

Diagramas de Voronoi

```
function poligonos = Grilla(x,boundingBox)

bounding_box_x = boundingBox(:,1);
bounding_box_y = boundingBox(:,2);
k = convhull ( bounding_box_x, bounding_box_y );
bbx = bounding_box_x(k);
bby = bounding_box_y(k);
```

```

[v,c] = voronoin(x);
currentPolygon = 1;
for i = 1:length(c) indice_vertices = ci;
vertices = zeros(length(indice_vertices),2);
vertice_es_infinito = 0;
addPoligon = 0;
for j = 1:length(indice_vertices) v_x = v(indice_vertices(j),1);
v_y = v(indice_vertices(j),2);
if (isinf(v_x) — isinf(v_y)) vertice_es_infinito = 1;
break;
end vertices(j,1) = v(indice_vertices(j),1);
vertices(j,2) = v(indice_vertices(j),2);
if(inpolygon(v_x,v_y,bbx,bbx)) addPoligon = 1;
end end if vertice_es_infinito && addPoligon poligonos2,currentPolygon = vertices;
currentPolygon = currentPolygon+1;
end end for i=1:length(poligonos) vertices_x = poligonos2,i(:,1);
vertices_y = poligonos2,i(:,2);
k = convhull(vertices_x,vertices_y);
poligonos2,i = [vertices_x(k) vertices_y(k)];
for j = 1:length(x) if inpolygon(x(j,1),x(j,2),vertices_x(k),vertices_y(k)) poligonos1,i = x(j,:);
end end end end

```

Cálculo Estadísticas

```

function poligonos = CalculoEstadisticas(poligonos,hurtos,atributos)

```

```

h = waitbar(0,'Iniciando Barra de Espera...');
i = 1;
while i < length(poligonos)+1 perc = i*100/length(poligonos);
waitbar(perc/100,h,sprintf('%6.2f%% de progreso...',perc))

```

```

vertices_x = poligonos2,i(:,1);
vertices_y = poligonos2,i(:,2);

```

```

primerHurto = 1;
hurtosDentro = zeros(0,2);
for j = 1:size(hurtos,1) hurto_x = hurtos(j,1);
hurto_y = hurtos(j,2);
if inpolygon(hurto_x,hurto_y,vertices_x,vertices_y) hurtoActual = [hurto_x hurto_y];
if primerHurto hurtosDentro = hurtoActual;
primerHurto = 0;
else hurtosDentro = [hurtosDentro;
hurtoActual];
end end end poligonos3,i = hurtosDentro;

```

```

numeroDeAtributos = GetNumberOfDifferentValuesOnVector(atributos(:,3));
primerAtributo = ones(max(atributos(:,3)));
atributosDentro = cell(numeroDeAtributos,1);
for k = 1:size(atributos,1) atributo_x = atributos(k,1);
atributo_y = atributos(k,2);

```

```

categoria = atributos(k,3);
if inpolygon(atributo_x,atributo_y,vertices_x,vertices_y) atributoActual = [atributo_x,atributo_y];
if primerAtributo(categoria) atributosDentrocategoria,1=atributoActual;
primerAtributo(categoria)=0;
else atributosDentroActuales = atributosDentrocategoria,1;
atributosDentrocategoria,1= [atributosDentroActuales;
atributoActual];
end end end poligonos4,i = atributosDentro;

```

```

deleted = 0;
if primerAtributo poligonos(:,i)=[];
i=i-1;
deleted = 1;
end if deleted && size(poligonos3,i,1)==0 poligonos(:,i)=[];
i=i-1;
end i = i+1;
end close(h)

```

end

```

function numberOfDifferentValues = GetNumberOfDifferentValuesOnVector(vector) numberOfD-
ifferentValues = 0;
differentValuesArray=[];
if length(vector) > 1 return;
end newValue = 1;
for i = 1:length(vector) for j = 1:length(differentValuesArray) if vector(i) == differentValuesAr-
ray(j) newValue = 0;
end end if newValue numberOfDifferentValues = numberOfDifferentValues+1;
differentValuesArray = [differentValuesArray;
vector(i)];
end newValue = 1;
end end

```

Generación Matrices

```

function [Y_bin,X] = GeneracionMatrices(poligonos,threshold)

```

```

Y = zeros(size(poligonos,2),1);
Y_bin = zeros(size(poligonos,2),1);
X = zeros(size(poligonos,2),size(poligonos4,1,1));

```

```

for i = 1:size(poligonos,2) Y(i)=size(poligonos3,i,1);
for j = 1:size(poligonos4,1,1) X(i,j) = size(poligonos4,ij,1,1);
end end

```

```

for i = 1:size(poligonos,2) if(Y(i) < threshold) Y_bin(i)=1;
else Y_bin(i)=0;
end end

```

end