

ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL

Instituto de Ciencias Matemáticas

Ingeniería en Estadística Informática

Análisis Estadístico de la Sustracción de Vehículos en la ciudad
de Guayaquil. Año 1999.

TESIS DE GRADO

Previa la obtención del Título de:

INGENIERO EN ESTADÍSTICA INFORMATICA

Presentado por:

Beissy A. Pacheco Ochoa

GUAYAQUIL – ECUADOR

2001

TRIBUNAL DE GRADUACIÓN

Mat. Jorge Medina.
DIRECTOR DEL ICM

Mat. Eduardo Rivadeneira
DIRECTOR DE TESIS

Ing. Félix Ramírez.
VOCAL

Ing. Jorge Duque.
VOCAL

DECLARACIÓN EXPRESA

“La responsabilidad del contenido de esta Tesis de Grado, me corresponden exclusivamente; y el patrimonio intelectual de la misma a la ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL”

Beissy Pacheco

DEDICATORIA

A DIOS

A MIS PADRES, Paky y Richard

A MI ABUELITO, Manuel

A MIS HERMANOS

A ROBERTO

AGRADECIMIENTO

A Dios, por ser siempre fiel.

Al Mat. Eduardo Rivadeneira, mi amigo y apoyo fundamental en este gran proyecto.

A Mis Padres, por dejarme la mejor de las herencias, la educación.

RESUMEN

El siguiente estudio representa un análisis estadístico de los robos de vehículos ocurridos en la ciudad de Guayaquil durante el año 1999; los datos

.

En su primera parte se hace una mirada retrospectiva de la mujer y su evolución social política y económica en el Mundo y en nuestro país.

En el capítulo dos se presenta el marco teórico, que consiste en Introducción al método factorial seguido de una breve descripción del Análisis de Componentes principales , Análisis Factorial de Correspondencia y el Análisis Factorial de Correspondencias Múltiples, en el cual se fundamentará nuestro estudio.

En el tercer capítulo se presenta el análisis univariado y multivariado de las variables. Finalmente se muestran las conclusiones a las que nos condujo nuestro análisis.

INDICE GENERAL

	Pág.
RESUMEN.....	I
INDICE GENERAL.....	II
INDICE DE GRAFICOS.....	III
INDICE DE TABLAS.....	IV
INTRODUCCIÓN.....	V
I. PLANTEAMIENTO DEL PROBLEMA.....	1
1.1 Definición del Problema.....	1
1.2 Alcance.....	2
1.3 Organismo Encargado de la Captura y Aprensión de los delincuentes.....	2
1.3.1 Policía Judicial.....	2
1.3.2 Misión.....	2
1.4 Areas de Competencia de la PJ.....	3
1.4.1 Delitos Contra la Propiedad	3
1.4.2 Delitos Contra la Administración y Fé Pública.....	3
1.4.3 Vehículos.....	4
1.4.4 Delitos Contra las Personas	4
1.4.5 Otros Casos de Menores.....	4
1.5 Variables que Utilizaremos para Nuestro Estudio.....	5

1.6	Marco Legal.....	11
1.6.1	Recepción de Denuncias.....	11
1.6.2	Procedimientos de aprensión.....	11
1.6.3	Generalidades de la investigación Sobreseimiento y Juicio.....	13
1.7	Vehículos Recuperados.....	16
1.8	Procedimiento para Estimar los Próximos Lugares de Robo	17
1.8.1	Tipos de Análisis Estadísticos Realizados en la Actualidad.....	18
1.9	El Aporte de la Investigación Científica a la Solución del Problema Planteado.....	18
1.9.1	Propuesta.....	18
II.	MARCO TEORICO.....	20
2.1	Importancia de los Datos.....	20
2.2	Histogramas de Frecuencias.....	23
2.2.1	Ordenando los Datos.....	23
2.2.1.1	Ventajas del Ordenamiento de Datos.....	23
2.2.2	Distribución de Frecuencias.....	24
2.2.3	Histogramas.....	25
2.3	Probabilidad y Variables Aleatorias.....	26
2.3.1	Probabilidad.....	26

2.3.2	Variables Aleatorias.....	28
2.3.3	Valor Esperado.....	31
2.3.4	distribuciones de Probabilidad.....	31
2.4	Medidas de Tendencia Central y Dispersión....	32
2.4.1	La Media	32
2.4.2	La Moda	35
2.4.3	La Mediana.....	36
2.4.4	La Varianza.....	37
2.5	Pruebas de Hipótesis.....	39
2.5.1	Definición	39
2.5.2	La confianza.....	41
2.5.3	Estadístico de Prueba Región de Rechazo y Valor P.....	42
2.5.4	Pruebas con Parámetros Poblacionales.....	45
	2.5.4.1 Pruebas de Medias con Varianza Conocida.....	45
	2.5.4.2 Pruebas de Medias con Varianza Desconocida.....	46
	2.5.4.3 Pruebas de Diferencias de Medias con Varianza Conocida.....	46
	2.5.4.4 Pruebas de Diferencias de Medias con Varianza Desconocida.....	47

2.6	Tablas de Contingencia.....	50
2.6.1	Definición.....	50
2.6.2	Hipótesis.....	52
2.6.3	Contribución de las Tablas de Contingencia al Estudio de la Sustracción de Vehículos en la Ciudad de Guayaquil.....	53
2.7	Componentes Principales.....	53
2.7.1	introducción al Análisis Factorial.....	53
2.7.2	Nociones Básicas.....	55
2.7.3	La Matriz de Datos.....	56
2.7.4	El Análisis Preliminar.....	57
2.7.5	Las Componentes principales....	58
2.7.6	Cálculo de los valores propios....	61
2.7.7	La Matriz de Factores.....	63
III.	ANÁLISIS ESTADÍSTICO.....	65
3.1	Análisis Univariado.....	65
3.1.1	Día del Robo.....	66
3.1.2	Mes del Robo.....	75
3.1.3	Hora del Robo.....	82
3.1.4	Marca del Vehículo.....	92
3.1.5	Tipo de Vehículo.....	99
3.1.6	Color del Vehículo.....	102

3.1.7	Entorno del Robo.....	106
3.1.8	Número de Personas que Realizan el Robo.....	111
3.1.9	Acción Delincuencial Para Realizar El Robo.....	114
3.1.10	Armas.....	116
3.1.11	Locomoción.....	118
3.1.12	Zona.....	120
3.2	Análisis Bivariado.....	132
3.2.1	Día Vs. Hora.....	133
3.2.2	Marca Vs. Tipo.....	137
3.2.3	Marca Vs. Zona.....	141
3.2.4	Num. Vs. Zona.....	151
3.2.5	Locomoción Vs. Zona.....	154
3.2.6	Entorno Vs. Acción.....	157
3.2.7	Hora Vs. Zona.....	160
3.2.8	Marca Vs. Color.....	164
3.3	Análisis Multivariado.....	172
3.3.1	Día Vs. Hora Vs. Zona.....	173
IV.	CONCLUSIONES.....	184
V.	RECOMENDACIONES.....	191
VI.	BIBLIOGRAFÍA.....	193

INTRODUCCION.

Este trabajo investigativo se produce en base a la necesidad de encontrar una solución factible para frenar la creciente ola de delincuencia en la que está sumergido el país. Si bien es cierto, el objetivo de este análisis es determinar, o en su defecto, estimar las relaciones existentes entre las variables que rodean las múltiples sustracciones de vehículos que suceden en la actualidad, consideramos a este proyecto el primer paso hacia la solución del problema.

Se han elegido diferentes métodos estadísticos de análisis, entre ellos podemos contar con las pruebas de hipótesis, los intervalos de confianza, las tablas de contingencia y finalmente el análisis de componentes principales.

Esperamos que este estudio sea de utilidad para las próximas generaciones en la búsqueda de la solución al problema del creciente índice de delincuencia.

Capítulo I.

1. PLANTEAMIENTO DEL PROBLEMA.

1.1 Definición del Problema.

Guayaquil es una ciudad de aproximadamente 2 millones de habitantes que corresponden aproximadamente al 16% de la población total del Ecuador. “En la ciudad de Guayaquil, semanalmente se suscitan un promedio de 40 robos de vehículos” (según revela un informe que fue analizado por el Grupo de Tarea que participa de las actividades antidelictivas en la provincia del Guayas) que lamentablemente no son evitados con eficiencia, por la fuerza pública. Una de las causas de este fenómeno se debe desde mi punto de vista, a la no utilización de métodos estadísticos multivariados para el análisis de los procedimientos delincuenciales. Mediante este estudio me dispongo a proponer 2 métodos de análisis multivariado de datos para encontrar las causas potenciales de este problema.

1.2 Alcance.

El objetivo de mi análisis es determinar las circunstancias bajo las que se cometen los robos de vehículos en Guayaquil, además quiero proponer soluciones mediante métodos de prevención basados en los resultados de mi análisis.

1.3 Organismo Encargado de la Captura y Aprensión de los Delincuentes.

1.3.1 Policía Judicial.

Es un cuerpo auxiliar de la administración de justicia, integrado por personal especializado de la Policía Nacional. Se rige, en lo que fuere aplicable, por la Ley Orgánica de la Policía Nacional, Código de Procedimiento Penal, Ley Orgánica de la Función Jurisdiccional y demás Leyes y Reglamentos Pertinentes

1.3.2 Misión.

“Prevención e investigación de las infracciones Penales y la aprehensión de los presuntos infractores”

1.4 Áreas de Competencia de la PJ

El departamento de inteligencia de la Policía Judicial se encarga de velar por la seguridad de la ciudadanía previniendo y combatiendo cinco diferentes tipos de delitos:

1.4.1 Delitos contra la propiedad

- a) Estruches: robos en casas y departamentos
- b) Robo a locales comerciales y oficinas
- c) Robo a bancos
- d) Hurto: Robo sin utilizar la violencia
- e) Abigeato
- f) Robo de transportes urbanos: Colectivos
- g) Robo en Carreteras: Transporte interprovincial
- h) Robos de cargas en carreteras: Contenido de camiones
- i) Extorsión
- j) Apropiación indebida
- k) Estafas

1.4.2 Delitos contra la administración y fé pública

- a) Cheques sin fondos
- b) Seudoagente: Suplantar identidad
- c) Falsificaciones: Firmas y documentos
- d) Falsificaciones de productos
- e) Falsificación de monedas
- f) Violación de domicilio
- g) Destrucción de bienes
- h) Tenencia ilegal de armas de fuego
- i) Tenencia ilegal de arma blanca
- j) Evasiones
- k) Invasiones
- l) Peculado
- m) Asociación ilícita: Pandillas mayores

1.4.3 Vehículos

- a) Asalto y robo de carros: Robos de carros por la fuerza
- b) Robo de carros: Robos de carros por hurto
- c) Robo/hurto de motos
- d) Robo de accesorios
- e) Robo o hurto de bicicletas

1.4.4 Delitos contra las personas

- a) Homicidios
- b) Suicidios
- c) Linchamientos
- d) Otras causas de muerte
- e) Tentativa de homicidio: Intento de homicidio
- f) Intimidaciones: Amenazas de muerte
- g) Agresiones físicas
- h) Plagio: Secuestro
- i) Tentativa de plagio: Intento de secuestro
- j) Rapto
- k) Abandono de menor
- l) Estrupo: Violación de padres a hijos
- m) Violación
- n) Tentativa de violación: Intento de violación
- o) Atentado al pudor
- p) Acoso Sexual
- q) Desaparecidos
- r) Trata de blancas
- s) Delitos varios

1.4.5 Otros casos de menores

- a) Entrega o venta de drogas a menores: poseer drogas o vender drogas a menores
- b) Pandillas juveniles
- c) Escándalo de menores
- d) Batidas de menores: Operativos policiales
- e) Faltamiento de palabra de menores

Para nuestro estudio nos basaremos únicamente la sección 1.4.3 a) y b)

1.5 Variables que Utilizaremos para Nuestro Estudio

Dentro de lo que son robos de vehículos la PJ ha recolectado datos que hemos clasificado en los siguientes grupos:

a) Fecha: día/mes/año

b) Día

- Lunes
- Martes
- Miércoles
- Jueves
- Viernes
- Sábado
- Domingo

c) Mes

- Enero
- Febrero
- Marzo
- Abril
- Mayo
- Junio
- Julio
- Agosto
- Septiembre
- Octubre
- Noviembre
- Diciembre

d) Hora:

- H00: desde las 00H00 hasta las 00:59
- H01: desde las 01H00 hasta las 01:59
- H02: desde las 02H00 hasta las 02:59
- H03: desde las 03H00 hasta las 03:59
- H04: desde las 04H00 hasta las 04:59
- H05: desde las 05H00 hasta las 05:59
- H06: desde las 06H00 hasta las 06:59
- H07: desde las 07H00 hasta las 07:59
- H08: desde las 08H00 hasta las 08:59
- H09: desde las 09H00 hasta las 09:59
- H10: desde las 10H00 hasta las 10:59
- H11: desde las 11H00 hasta las 11:59
- H12: desde las 12H00 hasta las 12:59
- H13: desde las 13H00 hasta las 13:59
- H14: desde las 14H00 hasta las 14:59
- H15: desde las 15H00 hasta las 15:59
- H16: desde las 16H00 hasta las 16:59
- H17: desde las 17H00 hasta las 17:59
- H18: desde las 18H00 hasta las 18:59
- H19: desde las 19H00 hasta las 19:59
- H20: desde las 20H00 hasta las 20:59
- H21: desde las 21H00 hasta las 21:59
- H22: desde las 22H00 hasta las 22:59
- H23: desde las 23H00 hasta las 23:59

e) Zona¹:

- S1
- S2
- S3
- S4
- S5
- S6
- C1
- C2
- C3
- C4

¹ Las delimitaciones de cada zona se especifican en la sección 3.1.11

- C5
- O1
- O2
- O3
- O4
- N1
- N2
- N3
- N4
- N5
- N6
- N7
- N8
- N9
- N10
- N11
- N12
- N13
- N14
- N15
- N16
- N17

f) Marca: Marca del vehículo robado.

- ALFAROM: Alpha Romeo
- AUSTIN
- BMW
- CHEVRLT: Chevrolet
- CHRYSLER
- CITROEN
- DACIA
- DAEWOO
- DAIHATSU
- DATSUN
- DELTA
- DODGE
- ENCAVA
- ESCODA
- FIAT
- FORD
- GMC
- HINO

- HONDA
- HYUNDAI
- INTERNTL: International
- IZUSU
- JEEP
- JOHNDDEE: Jhonn Deere
- KIA
- LADA
- LEXUS
- MACK
- MAZDA
- MERCBENZ: Mercedes Benz
- MERCURY
- MITSUBSH: Mitsubishi
- MORRIS
- MUSTANG
- NISSAN
- OPELREC: Opel Record
- PEUGEOT
- PLYMOUTH
- SAEHAN
- SANGYONG
- SCANIA
- SEAT
- SKODA
- SUBARU
- SUZUKI
- TOYOTA
- TRIUMPH
- VW: Volkswagen
- VOLVO
- WILLYS
- ZASTAVA

g) Tipo

- Auto
- Campero
- Camioneta
- Otros:
 - Bus
 - Camión
 - Furgoneta
 - Furgón
 - Micro Bus
 - Montacarga
 - Montero
 - Mula
 - Retroexcavadora
 - Tractor
 - Trailer

h) Color

- Amarillo
 - Amarillo
 - Mostaza
- Naranja
 - Anaranjado
 - Dorado
- Azul
- Bicolor
- Blanco
- Café
- Celeste
- Crema
 - Crema
 - Beige
 - Brandy
 - Champagne
 - Habano
- Gris
 - Gris
 - Plomo
 - Plateado

- Morado
 - Morado
 - Lila
 - Violeta
- Negro
- Rojo
 - Rojo
 - Ladrillo
 - Tomate
 - Vino
- Rosado
- Turquesa
- Verde

i) Entorno: Circunstancias del vehículo previo al robo

- Estacionado: Vehículo Parqueado en la acera
- Domicilio: Llegando, saliendo o parqueado en el garaje del domicilio
- Andando: Conduciendo libremente
- Semáforo: Al esperar la luz verde
- Flete: Llevando pasajeros en una carrera de taxi

j) Num: Número de personas que realizan el robo

- Uno
- Dos
- Tres
- Cuatro
- Cinco
- Seis
- Siete
- Ocho
- Nueve: nueve o mas

k) Acción: Robo con fuerza o sin fuerza

- AR: Robo con Asalto (robo con violencia)
- Robo: Robo sin Asalto(robo sin violencia)

l) Armas:

- Fuego: Pistola, revolver, etc.
- Blanca: Cuchillo, navaja, etc
- Ninguna

m) locomoción: medio de transporte de los delincuentes

- A pie
- Vehículo

1.6 Marco Legal

1.6.1 Recepción de Denuncias.

La Policía Judicial es el instrumento legalmente encargado de receptor denuncias de sustracción de vehículos. El interesado debe acudir a la oficina de recepción de denuncias, y explicar la situación bajo la cual se suscitó el robo, proporcionar información acerca de vehículo (marca, modelo, etc) y sus datos personales (nombre, dirección, etc), acompañados de alguna descripción de los delincuentes en el caso de tenerla.

1.6.2 Procedimientos de Aprensión.

El artículo 64 del código penal explica que “con el objeto de investigar un delito de acción pública, a pedido del fiscal, el juez competente podrá ordenar la detención de una persona contra la cual haya presunciones de responsabilidad”, es

decir que una persona puede ser arrestada para ser investigada si existen sospechas sobre ella. Además, los agentes de la Policía Judicial pueden aprehender a una persona que halla sido sorprendida cometiendo un delito (inmediatamente después de haberlo realizado) si se le encontraran armas, instrumentos, huellas o documentos relativos al delito recién cometido.

Una vez descubierto el sospechoso se procede a su captura y aprehensión, el aprehendido es ingresado a una patrulla y luego conducido a los calabozos de la PJ para ser reclutado en las cárceles de Guayaquil. A partir de ese momento se inicia una investigación del caso. Si dentro de las 24 horas siguientes al arresto se determina que el sospechoso es inocente, este quedará en libertad. Si por el contrario al cabo de 24 horas la investigación no arroja resultados, se dictará auto de prisión preventiva si fuera procedente². El sospechoso será transferido al Centro de Detención Provisional ubicado en la ciudad de Pascuales. La prisión preventiva no podrá exceder de 6 meses contados a partir de

² La prisión preventiva es procedente cuando existen indicios suficientes sobre la existencia del delito e indicios claros y precisos de que el implicado es autor o cómplice del delito.

la fecha en que se haga efectivo el auto de prisión preventiva.

1.6.3 Generalidades de la Investigación, Sobreseimiento y Juicio

La Policía Judicial realizará la investigación del caso bajo la dirección y control del ministerio público, a fin de reunir todas las pruebas y evidencias posibles y de evitar que los sospechosos se oculten o fuguen.

Dentro de un lapso máximo de 90 días, el fiscal del caso debe dar inicio a la instrucción fiscal que consiste en que el imputado en presencia de su abogado, presente al fiscal su versión sobre las circunstancias y móviles del hecho y sobre su participación o la de otras personas.

Si el fiscal considera que existe evidencia suficiente de que existió el delito y que el sospechoso estuvo implicado en él, debe requerir por escrito al juez que dicte el auto de llamamiento a juicio mediante un dictamen acusatorio, si por el contrario el Fiscal no considera que existen evidencias

suficientes para enjuiciar al sospechoso, emitirá su dictamen sin acusación.

A partir de este momento el juez cuenta con diez días para convocar a las partes a una audiencia preliminar donde se escucharán (y transcribirán en un acta) las declaraciones del imputado, el abogado defensor, el Fiscal, el sospechoso particular o sus abogados. Inmediatamente después de escuchar las declaraciones, el Juez leerá a los presentes su resolución.

Si el Fiscal hubiera dictado un dictamen acusatorio, y como producto de la instrucción fiscal el Juez coincidiera con la decisión del Fiscal, el Juez deberá dictar un auto de llamamiento a juicio.

Si se diera el caso que el Fiscal hubiese declarado su dictamen sin acusación y que el Juez considere que existen evidencias suficientes para acusar el Juez deberá enviar las actuaciones a un Fiscal Superior para que ratifique o contradiga el pronunciamiento del inferior.

Una vez dictado el auto de llamamiento a juicio, el Juez deberá transportar el proceso al tribunal penal o a la oficina de sorteos.

Si el Fiscal Superior hubiere ratificado el pronunciamiento del inferior, el juez debe admitir el dictamen y dictar un auto de sobreseimiento. El sobreseimiento se dicta cuando no existen evidencias suficientes del delito, de que el acusado halla participado en el delito, o que el hecho no constituya un delito en si. Si el sobreseimiento es *provisional*, se suspende la etapa del juicio, si es *definitivo* dará por finalizado el juicio, y si es *provisional del proceso y definitivo a favor del imputado* impedirá que el sospechoso vuelva a ser enjuiciado por el mismo proceso. El sobreseimiento no puede exceder de 5 años si es provisional del proceso, ni de 3 años si es provisional del imputado. Si se vencieran estos plazos el Juez deberá dictar *sobreseimiento en firme* que consiste en dictar un auto de sobreseimiento definitivo del proceso y del imputado.

Luego de haberse dictado el auto de llamamiento a juicio y con el objeto de probar la existencia del delito y la implicación

del imputado se iniciará la etapa de juicio. Según el artículo 285 del código penal “El día y hora fijados, el presidente del tribunal después de verificar la presencia del acusado, del fiscal, del acusador particular si lo hubiere, del ofendido, de los testigos, peritos o traductores, debe declarar abierto el juicio”. Posterior al juicio se realizará un debate en el que participarán el Fiscal y el defensor (o en su defecto el acusado). Luego del debate el tribunal procederá a deliberar acerca proceso y de las pruebas, al fin de la deliberación el tribunal dictará la sentencia Si el acusado fuere absuelto, el presidente ordenará su inmediata libertad sino deberá cumplir la condena impuesta.

1.7 Vehículos recuperados.

Si se rescata algún vehículo producto de la investigación, este es transportado a los patios la PJ (ver figura 1.1) hasta ser reconocido por sus dueños.

En los patios de la Policía Judicial se encuentran los automóviles recuperados



Figura 1.1 Patios de la Policía Judicial

1.8 Procedimiento para estimar los próximos lugares de robo

En un mapa de la ciudad de Guayaquil que posee el departamento de inteligencia de la PJ se colocan semanalmente señaladores de colores que indican la ubicación de los lugares donde se han suscitado incidentes delictivos durante el transcurso de la semana, los colores son asignados por tipo de delito y los señaladores son colocados en base a los resultados del análisis univariado semanal realizado por ellos. Luego se organizan operativos para vigilar aquellos lugares donde se han cometido más delitos; esto resulta muy poco productivo, ya que por lo general los delincuentes no regresan a cometer mas delitos en el mismo lugar durante el transcurso de la semana.

1.8.1 Tipos de análisis estadísticos realizados en la actualidad.

Lamentablemente en la PJ se han venido realizando hasta ahora únicamente el análisis univariado de los datos de algunas de las variables que poseen para el estudio de la delincuencia. Dejando sin analizar las posibles relaciones entre esas variables, quedando como consecuencia muchas preguntas sin respuesta. Los miembros del departamento de inteligencia de la P.J. grafican histogramas de frecuencias, gráficas de corridas y diagramas de pastel anuales, mensuales y semanales que imprimen a manera de informes. Ellos se basan en estos resultados para estimar en qué lugares se suscitan la mayor cantidad de actos delictivos.

1.9 El aporte de la investigación científica a la solución del problema planteado.

1.9.1 Propuesta.

Como ya había mencionado, me voy a basar en dos herramientas estadísticas multivariadas de análisis:

- a) Análisis Univariado de las Variables de Interés
 - Histogramas de Frecuencias
 - Gráficas de Corridas
 - Diagramas de Pastel

- Pruebas de hipótesis
 - Intervalos de Confianza
- b) Análisis Bivariado
- Tablas de Contingencia
- c) Análisis Multivariado
- Análisis de Componentes Principales

A partir del análisis univariado obtendremos tendencias de robo respecto al día, marca de carro, modelo, hora, etc. Además haremos inferencias acerca de algunas de estas variables, y obtendremos intervalos de confianza de otras.

Las Tablas de Contingencia, nos ayudarán a determinar si existe alguna relación de dependencia entre parejas de variables y el Análisis de Componentes Principales nos guiará en la decisión de elegir cuales son las variables que poseen mayor relevancia para nuestro análisis, y las agrupará mediante combinaciones lineales en conjuntos de variables que tengan características similares entre si.

Todo esto con el objeto de poder obtener conclusiones relevantes acerca de los procedimientos y planes de robo de los delincuentes.

Capítulo II.

2. MARCO TEÓRICO.

2.1 Importancia de los datos

Para dar sentido a la recolección de datos, es decir, para poder obtener información relevante a partir de los datos, debemos agrupar los datos unos con otros de manera que estos tengan diferencias significativas que los caractericen tanto dentro como fuera de cada uno de sus grupos. Sólo así podremos llegar a obtener conclusiones lógicas que nos lleven a la solución de un problema o simplemente para conocer lo que estos datos significan o representan.

Según Levin y Rubin¹ “es necesario tomar en cuenta que todos los grupos relevantes deben estar representados en los datos”, es decir que se debe tratar en lo posible de no omitir grupos de datos que sean representativos para nuestro estudio, por ejemplo, si estamos estudiando la preferencia de los hombres ecuatorianos de acuerdo a su raza a una cierta marca de perfume y los clasificamos simplemente en hombres blancos y hombres negros, estaremos omitiendo al grupo de los mulatos y de los indios de tal manera que al realizar el análisis correspondiente estaremos obteniendo resultados incompletos.

También “debemos asegurarnos de trabajar con datos confiables” ya que los resultados de los análisis estadísticos que se realicen van a depender única y directamente de los datos colectados. Además “los datos deben estar basados en suposiciones e interpretaciones correctas”. Debemos de poder reconocer el significado de cada una de las variables de manera que no existan contradicciones, ni dudas, ya que de eso dependen las conclusiones y el análisis.

¹Richard Levin y David Rubin, autores del libro “Estadística Aplicada para Administradores” sexta edición, año 1996, página12.

2.2 Histogramas de frecuencias

2.2.1 Ordenando los Datos

Es muy común encontrarse con muestras de datos en desorden, el primer paso a seguir para graficar un histograma de frecuencias es ordenar los datos. Por ejemplo, a continuación se presenta una tabla de datos que contiene datos de promedios de calificaciones.

Tabla 2.1
Promedios de calificaciones de Matemáticas de estudiantes de tercer año de secundaria.

16	15	17	08	12
20	16	17	14	18
13	16	16	18	15
17	11	17	16	15

Procedemos ahora a ordenar los datos en forma ascendente

Tabla 2.2
Promedios de calificaciones de estudiantes de tercer año de secundaria ordenados en forma ascendente.

08	11	12	13	14
15	15	15	16	16
16	16	16	17	17
17	17	18	18	20

2.2.1.1 Ventajas del ordenamiento de datos

Al arreglar los datos podemos notar características que antes eran difíciles notar a simple vista:

Podemos notar inmediatamente la cota superior (máximo) y la cota inferior (mínimo).- En la tabla 2.2 vemos que la mínima nota es 08 y la máxima es 20.

Podemos notar si existen datos que se repiten.- Del ejemplo anterior (tabla 2.2) podemos observar que los números 15, 16, 17 y 18 se repiten 3, 5, 4 y 2 veces respectivamente

Podemos dividir los datos en intervalos.- De la tabla 2.2 notamos que en la primera mitad se encuentran las notas desde 08 hasta 16, mientras que en la segunda mitad las notas van desde 16 hasta 20, esto nos hace notar que más de la mitad de los estudiantes aprobaron Matemáticas ese año.

2.2.2 Distribución de Frecuencias

Una tabla de frecuencias contiene información acerca de la distribución de frecuencias de una variable. Para construir la tabla de frecuencias de un conjunto de datos primero dividimos los datos en intervalos (o clases, no necesariamente del mismo tamaño) que contienen el conjunto de datos de la muestra que cae en ese intervalo, de tal manera que cada uno de los datos estén contenidos

en uno y solamente uno de los intervalos. Utilizaremos la tabla 2.2 para generar una tabla de frecuencias en base a los promedios de calificaciones del ejemplo anterior.

Tabla 2.3
Distribución de frecuencias de los promedios de calificaciones de Matemáticas de estudiantes de tercer año de secundaria.

Clase	Frecuencia
05-08	1
09-12	2
13-16	10
17-20	7

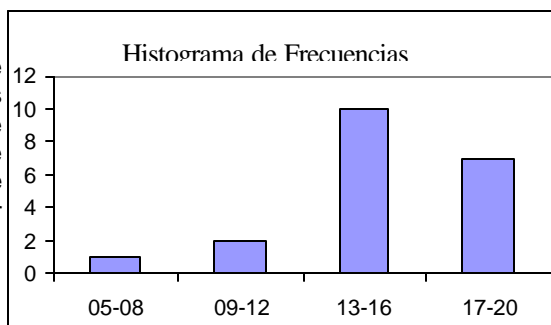
2.2.3 Histogramas

Se puede graficar histogramas en el plano cartesiano, en el espacio euclidiano, y en cualquier espacio de dimensión finita, pero es preferible y para cuestiones prácticas graficarlo en \mathbb{R}^2 y \mathbb{R}^3 , en este texto nos referiremos únicamente a histogramas graficados en el plano cartesiano.

Un histograma de frecuencias consiste en un conjunto de rectángulos colocados verticalmente sobre el eje de las abscisas (eje de las clases: eje **X**) cuyo alto corresponde a la frecuencia observada dentro de cada clase (como se vio anteriormente uno para cada clase: eje **Y**). Si los intervalos que describen a las clases son todos del mismo tamaño el ancho de cada uno de los rectángulos es el mismo, en caso contrario si los intervalos son de diferente tamaño el ancho

será proporcional al tamaño de la clase al que corresponda el rectángulo. La figura 2.1 ilustra un histograma de frecuencias para los datos de la tabla 2.2 dividiéndolo en las clases descritas en la tabla 2.3

Figura 2.1
Histograma de frecuencias de los promedios de calificaciones de Matemáticas de estudiantes de tercer año de secundaria

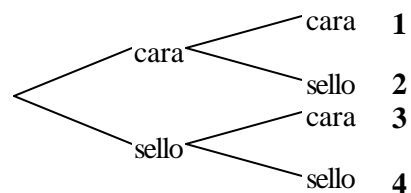


2.3 Probabilidad y Variables Aleatorias.

2.3.1 Probabilidad

Analizamos un ejemplo para entender el concepto. Si hacemos un experimento que consiste en lanzar una moneda al aire una vez, todos los posibles resultados son dos, cara y sello, pero si la lanzamos dos veces el número de resultados posibles son 4 como lo muestran la figura 2.2 y la tabla 2.4:

Figura 2.2.
Diagrama de tallos y hojas que resume los 4 posibles resultados de



lanzar una moneda dos veces

1	Cara	cara
2	Cara	sello
3	sello	cara
4	sello	sello

Tabla 2.4.

resume los 4 posibles resultados de lanzar una moneda dos veces.

A partir de estas figuras podemos notar que el número de posibilidades de obtener dos caras de estos lanzamientos es 1 de 4, lo mismo sucede para dos sellos, pero que el número de posibilidades de obtener una cara y un sello es 2 de 4. Podemos resumir esto en la tabla 2.5

cara y cara	1 de 4
cara y sello ó sello y cara	2 de 4
sello y sello	1 de 4

Tabla 2.5.

resume las posibilidades de obtener cada uno de los resultados después lanzar una moneda dos veces.

En términos matemáticos podemos decir que 1 de 4 es $\frac{1}{4}$ y que 2 de 4 son $\frac{2}{4}$, estas dos fracciones definen dos probabilidades, la primera es la probabilidad de obtener dos caras y la segunda es la probabilidad de obtener una cara y un sello en dos lanzamientos de una moneda.

Para entender la definición de probabilidad, definamos primero “espacio muestral”.

“Un espacio muestral es el conjunto de todos los resultados posibles de un experimento”

Probabilidad es una función que asigna a cada evento del espacio muestral un número entre cero y uno.

$$p: \begin{array}{l} S \xrightarrow{\quad} [0,1] \\ x \longrightarrow p(x), \end{array}$$

donde S es un espacio muestral y x es un evento de S y $p(x)$ que se define como la probabilidad de que ocurra el evento x .

Una función de probabilidad debe cumplir con los siguientes axiomas.

i. $p(S)=1$

ii. $0 \leq p(x) \leq 1$

iii. $\sum_{i=1}^n p(x_i) = 1, \forall x_i \in S$

2.3.2 Variables Aleatorias

Una variable aleatoria x es una función definida sobre un espacio muestral S , donde, a cada elemento a de S (llamado evento), se le asigna un número real $x(a)$.

$$x: S \rightarrow \mathfrak{R}$$

$$a \longrightarrow x(a), \text{ donde } a \in S$$

Así, retomando el ejemplo anterior del lanzamiento de monedas, se puede definir una variable aleatoria x como el número de caras que se observan en dos lanzamientos de una moneda, de donde se sabe que x puede tomar los valores de 2, 1 ó 0. La tabla de probabilidades para cada uno de estos sucesos se muestra a continuación.

Sea

x : número de caras que se observan en dos lanzamientos de una moneda

Evento	x	$p(x=x)$
cara y cara	2	$\frac{1}{4}$
cara y sello ó sello y cara	1	$\frac{2}{4}$
sello y sello	0	$\frac{1}{4}$

Tabla 2.6

resume las probabilidades de obtener cada uno de los resultados posibles después lanzar una moneda dos veces.

Como se observa la columna $p(x=x)$ muestra las probabilidades de que la variable aleatoria x tome los valores 2, 1 y 0 que son $\frac{1}{4}$, $\frac{2}{4}$ y $\frac{1}{4}$ respectivamente.

Existen dos tipos de variables aleatorias discretas y continuas

- **Variables aleatorias discretas**

Este tipo de variables aleatorias se caracterizan porque su dominio es un conjunto finito o infinito contable de valores, en otras palabras, estas variables aleatorias pueden tomar un número limitado de valores.

- **Variables aleatorias continuas**

Las variables aleatorias continuas se definen en un dominio continuo o sea pueden tomar un valor cualquiera dentro de un intervalo dado.

Ejemplos de variables aleatorias discretas y continuas:

- **Discretas**

x: grado escolar de un niño de primaria

$$\mathbf{dom(x) = \{1, 2, 3, 4, 5, 6\}}$$

y: número de páginas de un libro

$$\mathbf{dom(y) = \{1, 2, 3, \dots, n, \dots\}}$$

- **continuas**

w: peso en kilogramos de una persona

$$\mathbf{dom(w) = (0.5, 1000)}$$

z: número real

$$\mathbf{dom(z) = (-\infty, +\infty)}$$

2.3.3 Valor Esperado.

Asociado a cada variable aleatoria **x** existe un número llamado valor esperado $E[\mathbf{x}]$ que como su nombre lo define es el valor que se esperaría obtener al seleccionar aleatoriamente un valor de los

que puede tomar una variable aleatoria. Matemáticamente este valor es calculado dependiendo del tipo de variable que se estudie, si es discreta o continua. Las fórmulas matemáticas para obtener este valor se presentarán en la sección 2.4.

2.3.4 Densidades de probabilidad.

Acerca de este tema nos limitaremos a decir que así como hicimos con el ejemplo de las monedas, se puede definir una función de probabilidad para cada variable aleatoria. Esta función se llama función de densidad de probabilidad de la variable aleatoria o simplemente función de distribución de x ó $f(x)=p(x=x)$.

2.4 Medidas de Tendencia Central y Dispersión.

Las medidas de tendencia muestran mediante cálculos matemáticos algunas características de la población.

Las medidas de tendencia central nos dan una noción del valor hacia donde se agrupan la mayoría de los valores que toma la variable de estudio. Estos valores se calculan por medio de parámetros poblacionales (en el caso de estudiar poblaciones), o estimadores de estos parámetros (en el caso de trabajar con muestras).

2.4.1 La Media.

Supongamos que estudiamos una población cualquiera, se puede calcular la tendencia central de los datos por medio de la media poblacional μ , este valor es un parámetro, es decir una constante, y se calcula mediante las fórmulas.

Para variables aleatorias discretas.

$$m_x = E[x] = \sum_{i=1}^N x_i p(x = x_i),$$

donde $E[x]$ es el valor esperado de x (2) y $p(x = x_i)$ es la probabilidad de que la variable aleatoria \mathbf{X} tome el valor x_i y N es el tamaño de la población.

Para variables aleatorias continuas.

$$m_x = E[x] = \int_{-\infty}^{\infty} xf(x) dx,$$

donde $f(x)$ es la función de distribución de x .

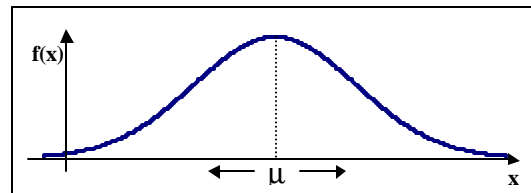
Para el caso de querer estimar la tendencia central de una población por medio de una muestra, podemos hacerlo calculando un estimador de la media poblacional, examinemos el caso de \bar{x} (equis barra), la media aritmética, cuya fórmula se deriva de la anterior.

$$\bar{x} = E[X] = \sum xP(X = x) = \sum_{i=1}^n x\left(\frac{1}{n}\right) = \frac{x}{n},$$

donde x_i representa cada uno de los valores que toma la muestra y n es el tamaño de la muestra.

Para ilustrar mejor el significado de μ y \bar{x} observemos la figura 2.3

Figura 2.3
Ilustración de la
media poblacional



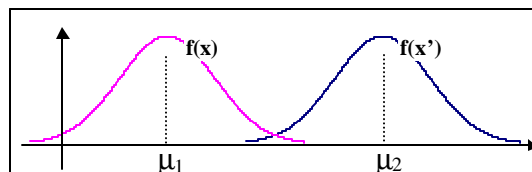
Tenemos aquí una población \mathbf{X} y hemos graficado su distribución de puntos $f(\mathbf{X})$. Podemos notar que los valores

² Ver sección 2.3.3

que toma X se agrupan alrededor de μ , teniendo la máxima frecuencia en μ y desplazándose hacia ambos lados de este valor. Esto es debido a que la distribución de la población es simétrica. Para distribuciones asimétricas la media se mueve del valor con frecuencia máxima hacia la derecha o izquierda, dependiendo del caso.

Si dibujamos la distribución de otra población X' en el mismo eje como en la figura 2.4 podemos comparar que las dos poblaciones tienen distinta media y que los valores de X' se agrupan hacia la derecha de los valores de X , es decir que la media de X' es mayor que la de X , esto es obvio dado que $\mu_2 > \mu_1$.

Figura 2.4
Comparación entre medias



2.4.2 La Moda.

Otra manera de medir la tendencia central es mediante el cálculo de la moda. La moda se define como el valor que más se repite entre el conjunto de datos. Es decir para

calcular su valor se debe contar cuantos números corresponden a cada dato y el dato con mayor frecuencia de repeticiones es la moda. También podemos dividir los datos en clases y contar los números que corresponden a dicha clase, así sabremos cual es la “clase modal”. Examinemos un ejemplo ilustrativo. La tabla 2.7 contiene el tiempo que tardó una cajera de un banco en atender a cada uno de los clientes que llegaron a su ventanilla durante una jornada de trabajo de 2 horas.

Tabla 2.7
Tiempo de atención (en minutos) a clientes en una ventanilla de un banco durante dos horas.

3	6	3	3	4
4	5	4	5	4
5	4	4	5	3
3	4	6	4	4
5	3	4	4	

Al ordenar y contar los datos obtenemos la tabla 2.8

Tabla 2.8
Distribución de frecuencias del tiempo de atención a clientes en una ventanilla de un banco durante dos horas.

Minutos	Frecuencia
3	6
4	11
5	5
6	2

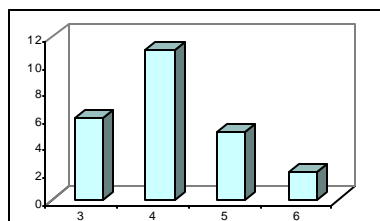


Figura 2.5
Histograma de frecuencias del tiempo de atención a clientes en una ventanilla de un banco durante dos horas.

De aquí obtenemos que la moda es 4 ya que es el valor que más se repite.

Cuando en una distribución de frecuencias se encuentra más de una moda, esta distribución se llama bimodal, trimodal, etc, dependiendo del número de modas que contenga.

2.4.3 La Mediana.

El valor de la mediana \tilde{x} depende directamente del número de datos que contenga la población. Para calcular la mediana primero debemos de ordenar los datos, luego calculamos la posición en la que se encuentra la mediana dependiendo del tamaño de la población (o de la muestra), si es par o impar, y al final calculamos su valor. Dicho en otras palabras, si el tamaño de la población es impar se encontrará en la posición $(n+1)/2$ y si es par la mediana se define como es promedio entre los valores que se encuentran en las posiciones $n/2$ y $(n+2)/2$. Es decir,

$$\tilde{x} = \begin{cases} x_{\left(\frac{n+1}{2}\right)}, & \text{si } n \text{ es impar, y} \\ \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n+2}{2}\right)}}{2}, & \text{si } n \text{ es par} \end{cases}$$

2.4.4 La Varianza

La varianza (σ^2 o s^2) es una medida de dispersión, es decir, indica que tan separados (o dispersos) se encuentran los datos. Este valor se calcula obteniendo la suma del cuadrado de las diferencias entre cada uno de los datos y la media y dividiendo este valor para N (el tamaño de la población) o para n-1 (n es el tamaño de la muestra). Es decir,

Varianza poblacional.

$$s^2 = \frac{\sum_{i=1}^N (x - m)^2}{N} = \frac{\sum x^2}{N} - m^2$$

Varianza muestral.

$$s^2 = \frac{\sum_{i=1}^n (x - \bar{x})^2}{n - 1} = \left(\frac{1}{n - 1} \right) (\sum x^2 - n\bar{x}^2)$$

A partir de la varianza se puede calcular el valor de la desviación estándar (σ o s) que se obtiene tomando la raíz cuadrada de la primera.

$$s = \sqrt{s^2} \quad \text{y} \quad \sigma = \sqrt{\sigma^2}$$

La desviación estándar nos da una noción de que tan dispersos de la media están los datos. De estudios realizados se sabe que por lo menos el 75% de los valores caen entre $\mu+2\sigma$ y $\mu-2\sigma$ y que por lo menos 89% de los valores caen entre $\mu+3\sigma$ y $\mu-3\sigma$. Veremos más adelante que lo aquí expresado es muy útil, ya que a partir de una muestra podemos calcular la media y la desviación estándar para obtener un intervalo (con un cierto grado de confianza) que contengan un gran grupo de datos dentro de ellos y así estimar poblaciones, sin necesidad de analizar todos los datos de la población.

2.5 Pruebas de Hipótesis.

2.5.1 Definición

Una hipótesis estadística es un supuesto acerca del valor de un parámetro poblacional en base a una muestra. Cuando se quiere probar o contradecir una hipótesis estadística se realizan lo que se llaman “Pruebas de Hipótesis”, mediante

estas pruebas se compara una característica de una variable aleatoria con una constante ó con la misma característica de otra variable aleatoria, tomando en cuenta un cierto grado de error llamado confianza (que es elegido por el investigador),. Se pueden formular infinidad de pruebas de hipótesis, entre las más usadas se encuentran las siguientes:

- Pruebas comparando un parámetro poblacional con un valor en referencia.
- Pruebas acerca de la distribución de probabilidad de una variable aleatoria.
- Pruebas comparando el valor de un parámetro poblacional de una variable aleatoria con los de otra.
- Pruebas acerca de la independencia entre dos variables aleatorias.

Las pruebas de hipótesis estadísticas se formulan postulando un supuesto, versus la negación del supuesto. El supuesto se llama hipótesis nula y se denota por H_0 , y la negación se denomina hipótesis alterna y se denota por H_1 , Observemos los siguientes contrastes acerca la media μ de una población.

1. **H₀: $\mu = 4$**
Vs.
H₁: $\mu \neq 4$
2. **H₀: $\mu \geq 4$**
Vs.
H₁: $\mu < 4$

El primer contraste de hipótesis postula que la media es igual a 4 versus la media no es igual a 4, el segundo postula que la media es mayor o igual a 4 versus la media es menor que 4. En ambos casos, para poder elegir entre la hipótesis nula y la alterna se deben calcular dos valores, compararlos y de acuerdo al resultado aceptar o rechazar la hipótesis nula.

2.5.2 La confianza

Se llama nivel de confianza al grado de error que se puede permitir que suceda en la estimación, como ya se mencionó, este error es manejable (elegido por el investigador). El nivel de confianza es un número real contenido en el intervalo $[0,100]$ y se mide en porcentaje. Podremos hablar entonces de 95% de confianza. El nivel de confianza se relaciona con

el nivel de significancia denotado por la letra griega α , que representa la probabilidad de rechazar la hipótesis nula cuando es falsa o probabilidad de cometer error tipo 1.

Una vez elegido el nivel de confianza $c\%$, se obtiene α mediante el siguiente cálculo:

$$\alpha = 1 - \frac{c}{100}$$

Entonces, podemos dejar fija la probabilidad de cometer error tipo 1, o sea la probabilidad de concluir que H_0 es verdadero cuando en realidad es falso, y así tener control sobre el error de la estimación.

2.5.3 Estadístico de prueba, Región de Rechazo y Valor p.

Para decidir entre aceptar o rechazar la hipótesis nula, se debe calcular un valor \hat{q} en base a una función matemática. Una vez obtenido este valor se puede optar por dos métodos:

1. \hat{q} se compara con el **estadístico de prueba** θ_0 que es el valor que toma una variable aleatoria θ , definida con

características de acuerdo a la confianza, a los datos y a suposiciones acerca de la población.

Dependiendo del resultado de esta comparación ($\hat{q} = \theta_0$ ó $\hat{q} < \theta_0$ ó $\hat{q} > \theta_0$) se acepta o rechaza la hipótesis nula.

2. Se obtiene el **valor p** de la prueba mediante el cálculo de la fórmula:

$$\text{valor } p = p(\mathbf{q} \geq \hat{\mathbf{q}}) = \int_{\hat{\mathbf{q}}}^{+\infty} f(\mathbf{q}) d\mathbf{q}, \text{ y se toma la decisión de aceptar o rechazar } H_0.$$

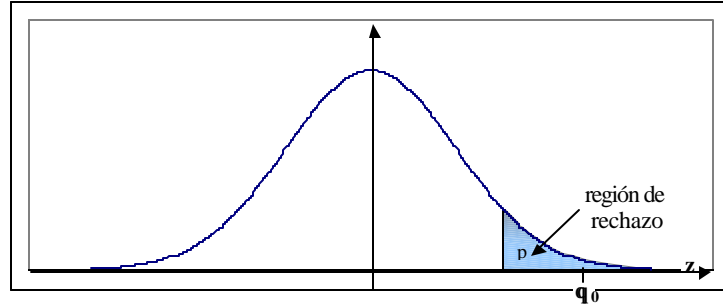
El valor p tiene un uso significativo en la estadística, ya que a partir de este valor y un nivel de significancia dado se puede decidir si aceptar o rechazar la hipótesis nula. Por ejemplo, a 95% de confianza ($\alpha = 0.05$) se acepta la hipótesis nula si $p \geq 0.1$; se la rechaza si $p \leq 0.05$; y cuando $0.1 < p < 0.05$ se dice que hay un región de indecisión (no se puede concluir algo), y no se puede tomar una decisión únicamente en base a la información obtenida, para mejorar la región, se debe optar por aumentar el número de unidades muestreadas, o cambiar el nivel de significancia.

El valor p es la probabilidad de que la variable aleatoria θ tome un valor mayor a \hat{q} ; esto no es más que el área bajo la curva de la función de distribución de probabilidad de la variable aleatoria θ . Gráficamente el valor p de la prueba define una región llamada **región de rechazo**, esto es debido a que dependiendo de la magnitud de esta región se acepta o rechaza la hipótesis nula.

Por ejemplo, asumamos que θ es Normal con media 0 y varianza 1 (normal estándar).

Figura 2.6

Grafica la función de distribución de una variable aleatoria Normal Estándar



Supongamos que el valor del **estadístico de prueba** es θ_0 (valor obtenido a partir de la tabla de la distribución Z), el área bajo la curva de z cuando $z \geq \theta_0$ es la llamada **región de rechazo** que define el **valor p** de la prueba, y para este caso se obtiene mediante el cálculo de la fórmula.

$$\text{valor } p = p(z \geq \hat{q}_0) = \int_{q_0}^{+\infty} f(z) dz,$$

donde z es una variable aleatoria normal estándar y $f(z)$ es la función de distribución de probabilidad de z .

Supongamos que $p = 0.0001$; debido a que $p < 0.05$ se dice que p es pequeño y se puede rechazar la hipótesis nula, según como se le halla definido. A continuación definiremos como se establecen los contrastes de hipótesis.

2.5.4 Pruebas con parámetros poblacionales

Para estas pruebas, primero se deben hacer supuestos acerca de la distribución de probabilidad de la variable aleatoria, para luego calcular el valor del estadístico de prueba en base a información obtenida de la muestra.

2.5.4.1 Pruebas de Medias con Varianza Conocida

Si la población es distribuida Normalmente o el tamaño de la muestra es lo suficientemente grande para asumir normalidad ($n \geq 30$) y se conoce la varianza, la región de rechazo viene dada de la siguiente forma:

Tabla 2.9
Regiones de rechazo para pruebas de hipótesis de medias con varianza conocida, asumiendo normalidad.

Prueba	Rechace H_0 en favor de H_1 si
$H_0: \mu = \mu_0$ Vs. $H_1: \mu \neq \mu_0$	$ z_0 > z_{\alpha/2}$
$H_0: \mu \leq \mu_0$ Vs. $H_1: \mu > \mu_0$	$z_0 > z_\alpha$
$H_0: \mu \geq \mu_0$ Vs. $H_1: \mu < \mu_0$	$z_0 < -z_\alpha$

Donde el estadístico de prueba es $z_0 = \frac{\bar{X} - m_0}{s/\sqrt{n}}$.

2.5.3.2 Pruebas de Medias con Varianza Desconocida

Si la población es normal o el tamaño de la muestra es lo suficientemente grande para asumir normalidad y no se conoce la varianza, la región de rechazo viene dada de la siguiente forma:

Tabla 2.10
Regiones de rechazo para pruebas de hipótesis de medias con varianza desconocida, asumiendo normalidad.

Prueba	Rechace H_0 en favor de H_1 si
$H_0: \mu = \mu_0$ Vs. $H_1: \mu \neq \mu_0$	$ t_0 > t_{\alpha/2, n-1}$
$H_0: \mu \leq \mu_0$ Vs. $H_1: \mu > \mu_0$	$t_0 > t_{\alpha, n-1}$
$H_0: \mu \geq \mu_0$ Vs. $H_1: \mu < \mu_0$	$t_0 < -t_{\alpha, n-1}$

Donde el estadístico de prueba es $t_0 = \frac{\bar{X} - m_0}{s/\sqrt{n}}$.

2.5.3.3 Pruebas de Diferencia de Medias con Varianza Conocida

Cuando se conoce la varianza y la población es Normal o el tamaño de la muestra es lo suficientemente grande para asumir normalidad ($n \geq 30$) la región de rechazo viene dada de la siguiente forma:

Tabla 2.11
Regiones de rechazo para pruebas de hipótesis de diferencia de medias con varianza conocida, asumiendo normalidad.

Prueba	Rechace H_0 en favor de H_1 si
$H_0: \mu_1 - \mu_2 = \mu_0$ Vs. $H_1: \mu_1 - \mu_2 \neq \mu_0$	$ z_0 > z_{\alpha/2}$
$H_0: \mu_1 - \mu_2 \leq \mu_0$ Vs. $H_1: \mu_1 - \mu_2 > \mu_0$	$z_0 > z_{\alpha}$
$H_0: \mu_1 - \mu_2 \geq \mu_0$ Vs. $H_1: \mu_1 - \mu_2 < \mu_0$	$z_0 < -z_{\alpha}$

Donde el estadístico de prueba es $z_0 = \frac{(\bar{X}_1 - \bar{X}_2) - m_0}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$.

2.5.3.4 Pruebas de Diferencia de Medias con Varianza Desconocida

Si se asume que ambas poblaciones son distribuidas normalmente o el tamaño de las muestras (n_1 y n_2) son lo suficientemente grandes para asumir normalidad y no se conoce la varianza, pero se asume que las varianzas de

ambas poblaciones son iguales ($\sigma_1 = \sigma_2$), la región de rechazo viene dada de la siguiente forma:

Tabla 2.12
Regiones de rechazo para pruebas de hipótesis de diferencia de medias con varianzas desconocida pero iguales, asumiendo normalidad.

Prueba	Rechace H_0 en favor de H_1 si
$H_0: \mu_1 - \mu_2 = \mu_0$ Vs. $H_1: \mu_1 - \mu_2 \neq \mu_0$	$ t_0 > t_{\alpha/2, n_1+n_2-2}$
$H_0: \mu_1 - \mu_2 \leq \mu_0$ Vs. $H_1: \mu_1 - \mu_2 > \mu_0$	$t_0 > t_{\alpha, n_1+n_2-2}$
$H_0: \mu_1 - \mu_2 \geq \mu_0$ Vs. $H_1: \mu_1 - \mu_2 < \mu_0$	$t_0 < -t_{\alpha, n_1+n_2-2}$

Donde el estadístico de prueba es $t_0 = \frac{(\bar{X}_1 - \bar{X}_2) - m_0}{s_p \sqrt{1/n_1 + 1/n_2}}$ y

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Si por el contrario se asume que las varianzas de ambas poblaciones no son iguales ($\sigma_1 \neq \sigma_2$), la región de rechazo viene dada de la siguiente forma:

Tabla 2.13
Regiones de rechazo para pruebas de hipótesis de diferencia de medias con varianzas desconocida y diferentes, asumiendo normalidad.

Prueba	Rechace H_0 en favor de H_1 si
$H_0: \mu_1 - \mu_2 = \mu_0$ Vs. $H_1: \mu_1 - \mu_2 \neq \mu_0$	$ t_0 > t_{\alpha/2, n}$
$H_0: \mu_1 - \mu_2 \leq \mu_0$ Vs. $H_1: \mu_1 - \mu_2 > \mu_0$	$t_0 > t_{\alpha, n}$
$H_0: \mu_1 - \mu_2 \geq \mu_0$ Vs. $H_1: \mu_1 - \mu_2 < \mu_0$	$t_0 < -t_{\alpha, n}$

Donde el estadístico de prueba es $t_0 = \frac{(\bar{X}_1 - \bar{X}_2) - m_0}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$ y

$$n = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \right)}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}$$

2.6 Tablas de contingencia.

2.6.1 Definición

Son arreglos de filas y columnas sobre las que se contrastan 2 características medibles de un conjunto de entes o individuos. A cada característica se la puede dividir en un conjunto de criterios mutuamente excluyentes (no necesariamente exhaustivos) que corresponden a los rótulos de las filas y columnas de la tabla (uno para cada criterio).

Tendremos entonces n criterios para la primera característica y p criterios para la segunda, que en la tabla representan n filas y p columnas respectivamente. El valor de cada celda X_{ij} de la tabla se determina mediante el conteo del número de individuos que cumplen con el criterio i de la primera característica y a la vez con el criterio j de la segunda. A continuación se muestra la tabla completa.

Tabla 2.14. Muestra del ingreso de datos en una Tabla de Contingencia.

		Característica 2						
		Criterio 1	Criterio 2	...	Criterio j	...	Criterio p	Suma de Filas
Característica 1	Criterio 1	X_{11}	X_{12}	...	X_{1j}	...	X_{1p}	$X_{1.}$
	Criterio 2	X_{21}	X_{22}	...	X_{2j}	...	X_{2p}	$X_{2.}$
	...	⋮	⋮		⋮		⋮	⋮
	Criterio i	X_{i1}	X_{i2}	...	X_{ij}	...	X_{ip}	$X_{i.}$
	...	⋮	⋮		⋮		⋮	⋮
	Criterio n	X_{n1}	X_{n2}	...	X_{nj}	...	X_{np}	$X_{n.}$
	Suma de Columnas	$X_{.1}$	$X_{.2}$...	$X_{.j}$..	$X_{.p}$	$X_{..}$

donde X_{ij} es la cantidad de individuos que cumplen con el criterio i y el criterio j. Nótese que la sumatoria de los valores de cada fila $X_{i.}$ corresponden al número de individuos que cumplen con el criterio i de la característica 1 y la sumatoria de los valores de cada columna $X_{.j}$ corresponde al número

de individuos que cumplen con el criterio j de la característica 2.

Además, la sumatoria de los valores de todas las filas debe ser igual a la sumatoria de los valores de todas las columnas y a la vez al número total de individuos en estudio; en otras palabras.

$$\sum_{i=1}^n X_{i.} = \sum_{j=1}^p X_{.j} = X_{..}$$

Una vez completada la tabla se procede a analizar el contraste de hipótesis siguiente.

2.6.2 Hipótesis

H₀: La característica 1 es independiente de la característica 2

Vs.

H₁: No es verdad H₀⁽³⁾

A (1- α)100% de confianza se rechaza la hipótesis nula H₀ a favor de la hipótesis alterna H₁ si:

$$X^2 > C_{\alpha, (n-1)(p-1)}^2$$

³ H₁ también puede ser interpretado como: la característica 1 y la característica 2 son dependientes.

donde
$$X^{2^*} = \sum_{i=1}^n \sum_{j=1}^p \frac{(f_{o,i,j} - f_{e,i,j})^2}{f_{e,i,j}}, \quad f_{e,i,j} = (X_{i,j} - e_{i,j})^2 \frac{1}{e_{i,j}},$$

$$e_{i,j} = (X_{i.} X_{.j}) \frac{1}{X_{..}}, \quad \text{y} \quad \chi^2_{2\alpha, (n-1)(p-1)}$$
 es el valor

correspondiente a una variable aleatoria Ji-cuadrado con $(1-n)(1-p)$ grados de libertad.

2.6.3 Contribución de las tablas de contingencia al estudio de la sustracción de vehículos en la ciudad de Guayaquil.

Trataremos de probar que existe relación entre cada par de escenarios bajo los que se han suscitado robos de vehículos, por ejemplo intentaremos probar que existe relación entre la hora y el lugar del robo.

2.7 Componentes Principales.

2.7.1 Introducción al Análisis Factorial

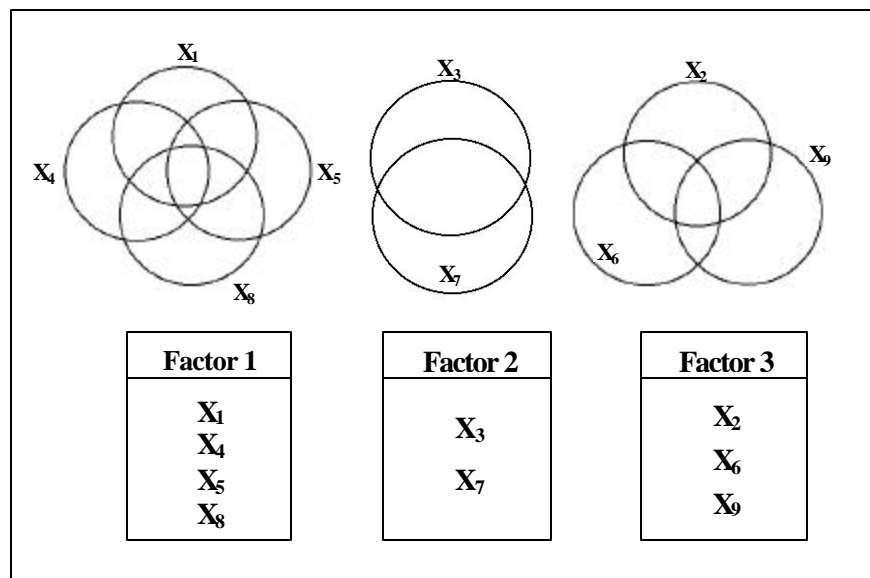
En ocasiones nos encontramos con casos en los que tenemos que tratar con muchas variables, lo que se puede derivar en complicados análisis de los resultados. Existen técnicas estadísticas para reducir variables sin perder

información, una de estas técnicas es el Análisis de Factores o “Análisis Factorial”.

El Análisis Factorial es una técnica multivariada que identifica variables relacionadas y las resume en un conjunto de nuevas variables o factores con el objeto de remover la redundancia y la información duplicada de un conjunto de variables correlacionadas, agrupándolas en algunos factores que contienen variables con características similares dentro del factor y diferentes fuera del factor.

Supongamos por ejemplo que estudiamos 9 variables de interés X_1 , X_2 , X_3 , X_4 , X_5 , X_6 , X_7 , X_8 y X_9 , el Análisis Factorial dividirá las variables en grupos llamados factores, por ejemplo en 3 factores como lo muestra la figura 2.5

Figura 2.5
Ejemplo de variables correlacionadas agrupadas en factores.



Cada factor contiene un grupo de variables altamente correlacionadas entre ellas y no correlacionadas entre las demás que no pertenecen a ese factor, permitiéndonos trabajar con menos variables y simplificando el análisis.

Es obvio que en la práctica no se darán casos perfectos como el aquí mostrado en los que las variables de un factor no están correlacionadas con las variables de otro factor, pero se puede esperar que exista poca correlación entre estas variables y mucha entre las variables de un mismo factor.

2.7.2 Nociones Básicas.

Componentes Principales es una técnica multivariada de interdependencia que se deriva del Análisis Factorial, en la que cada uno de los factores se expresa como una combinación lineal de las variables (de estudio) que pertenecen a él.

Al aplicar esta técnica al principio se obtienen igual número de factores que de variables, para luego ser reducidos en un número menor de acuerdo a un criterio (de importancia) que nosotros elegimos. Cada factor explica una fracción de la

varianza total inherente en el conjunto de datos, así, cada factor tratará de explicar la mayor cantidad de varianza posible. El primer factor siempre contendrá más información que el segundo y más información que la que pueda proporcionar una sola variable, el segundo factor contendrá mas información que el tercero y así en adelante. De tal manera que el o los últimos factores deberán contener individualmente menos información que la que pueda contener una sola variable. Es por esto que podemos eliminar factores, ya que asumimos como poco representativos los que proporcionen menos información que una sola variable.

2.7.3 La Matriz de Datos

Primero ingresamos la matriz original de $n \times p$ que contiene n objetos o entes como rótulos de las filas y p variables o características de los objetos como rótulos de las columnas así como lo muestra la tabla 2.15.

Tabla 2.15
Clasificación e ingreso de datos

		Variables			
		X_1	X_2	...	X_p
Objetos	O_1				
	O_2				
	\vdots				
	O_n				

2.7.4 El Análisis Preliminar

El Análisis de Componentes Principales se basa en la matriz de correlaciones o en la de covarianzas, que son calculadas en base a una matriz de datos como la definida en la tabla 2.15, para nuestro estudio nos basaremos en la matriz de Covarianzas. Para obtener esta matriz se calculan las covarianzas $\sigma_{i,j}$ entre cada par de variables X_i y X_j ; $i,j=1,2,\dots,p$ de la siguiente manera:

$$s_{i,j} = \text{cov}(X_i, X_j) = E[X_i \cdot X_j] - E[X_i]E[X_j]$$

Con estos valores completamos la matriz de correlaciones que obviamente será cuadrada de $p \times p$. Esta matriz se verá como la mostrada en la tabla 2.16

Tabla 2.16
Matriz de covarianzas

		Variables			
		$X_i \backslash X_j$	X_1	X_2	...
Variables	X_1	σ_{11}	σ_{12}	...	σ_{1p}
	X_2	σ_{21}	σ_{22}	...	σ_{2p}
	\vdots	\vdots	\vdots		\vdots
	X_p	σ_{p1}	σ_{p2}	...	σ_{pp}

2.7.5 Las Componentes Principales.

Son algebraicamente combinaciones lineales particulares de las p variables aleatorias $X_1, X_2, X_3 \dots X_p$. Geométricamente, estas combinaciones lineales representan la selección de un nuevo sistema coordenado obtenido por la rotación del sistema original con $X_1, X_2, X_3 \dots X_p$ como los ejes coordenados iniciales. Los nuevos ejes representan las direcciones con máxima variabilidad y proveen una descripción simple de la estructura de la covarianza.

Como ya se mencionó, los componentes principales dependen de la matriz de covarianza Σ (o de la matriz de correlación ρ) de $X_1, X_2, X_3 \dots X_p$. y su desarrollo no requiere un supuesto normal multivariado.

Sea el vector aleatorio $\underline{\mathbf{X}}^T = [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_p]$ con matriz de covarianzas Σ y con valores propios $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$.

Considere las combinaciones lineales.

$$Y_1 = \underline{\mathbf{a}}_1^T \underline{\mathbf{X}} = a_{11}X_1 + a_{12}X_2 + \dots + a_{1p}X_p \quad ^4$$

$$Y_2 = \underline{\mathbf{a}}_2^T \underline{\mathbf{X}} = a_{21}X_1 + a_{22}X_2 + \dots + a_{2p}X_p$$

...

$$Y_p = \underline{\mathbf{a}}_p^T \underline{\mathbf{X}} = a_{p1}X_1 + a_{p2}X_2 + \dots + a_{pp}X_p$$

Y dado que,

$$\text{Si } \mathbf{Z} = \underline{\mathbf{c}}\underline{\mathbf{X}}, \text{ entonces } \mathbf{m}_z = E[\mathbf{z}] = E[\underline{\mathbf{c}}\underline{\mathbf{X}}] = \underline{\mathbf{c}}\mathbf{m}_x$$

$$\text{y } \mathbf{S}_z = \text{cov}[\mathbf{z}] = \text{cov}[\underline{\mathbf{c}}\underline{\mathbf{X}}] = \underline{\mathbf{c}}\mathbf{S}_x\underline{\mathbf{c}}^T$$

Entonces,

$$\text{var}[Y_i] = \underline{\mathbf{a}}_i^T \mathbf{S} \underline{\mathbf{a}}_i, \quad i = 1, 2, \dots, p$$

$$\text{y, } \text{cov}[Y_i, Y_k] = \underline{\mathbf{a}}_i^T \mathbf{S} \underline{\mathbf{a}}_k \quad k = 1, 2, \dots, p$$

Los componentes principales son aquellas combinaciones lineales no correlacionadas Y_1, Y_2, \dots, Y_p cuyas varianzas $\text{var}[Y_i]$ sean lo más grande posible.

⁴ Denotaremos $\underline{\mathbf{X}}$ como el vector \mathbf{X}

La primera componente principal es la combinación lineal $\underline{a}_1^T \underline{X}$ que maximice el valor de $\text{var}[\underline{a}_1^T \underline{X}] = \underline{a}_1^T \mathbf{S} \underline{a}_1$, sujeto a la condición $\underline{a}_1^T \underline{a}_1 = 1$.

La segunda componente principal es la combinación lineal $\underline{a}_2^T \underline{X}$ que maximice el valor de $\text{var}[\underline{a}_2^T \underline{X}]$, sujeto a las condiciones $\underline{a}_2^T \underline{a}_2 = 1$ y $\text{cov}[\underline{a}_1^T \underline{X}, \underline{a}_2^T \underline{X}] = 0$.

Y la i -ésima componente principal es la combinación lineal $\underline{a}_i^T \underline{X}$ que maximice el valor de $\text{var}[\underline{a}_i^T \underline{X}]$, sujeto a las condiciones $\underline{a}_i^T \underline{a}_i = 1$ y $\text{cov}[\underline{a}_i^T \underline{X}, \underline{a}_k^T \underline{X}] = 0$ para $k < i$.

Una manera sencilla de encontrar las componentes principales de un conjunto de datos es calculando los valores y vectores propios de su matriz de covarianzas. Sea \mathbf{S} la matriz de covarianzas asociada con el vector aleatorio $\underline{X}^T = [X_1, X_2, \dots, X_p]$, y sean $(\lambda_1, \underline{e}_1), (\lambda_2, \underline{e}_2), \dots, (\lambda_p, \underline{e}_p)$, los pares ordenados valor propio, vector propio de \mathbf{S} donde $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$.

La i -ésima componente principal estará dada por:

$$Y_i = \underline{e}_i^T \underline{X} = e_{1,i} X_1 + e_{2,i} X_2 + \dots + e_{p,i} X_p, \quad i=1,2,\dots,p$$

Además, bajo estas condiciones es claro que;

$$\text{var}[Y_i] = \underline{e}_i^T \underline{S} \underline{e}_i = 1, \quad i = 1, 2, \dots, p$$

$$\text{y, cov}[Y_i, Y_j] = \underline{e}_i^T \underline{S} \underline{e}_j = 0, \quad i \neq j; j = 1, 2, \dots, p$$

2.7.6 Cálculo de los valores propios.

A partir de la matriz de covarianzas podremos calcular los valores propios que corresponderán a cada componente obteniendo las soluciones de la ecuación:

$$\Sigma_{p \times p} - \lambda I_{p \times p} = 0;$$

donde $\Sigma_{p \times p}$ es la matriz de covarianzas de las p variables originales, λ son cada uno de los $k \leq p$ valores propios que vamos a calcular y la matriz $I_{p \times p}$ es la matriz identidad.

Evaluando los valores correspondientes en esta ecuación obtendremos un polinomio de grado menor o igual a p que nos llevarán al cálculo de p valores propios. Al cabo de este paso proseguiremos a ordenar los valores propios de mayor a menor. Como nuestro objetivo es encontrar $k < p$ componentes debemos elegir un criterio de selección.

El valor propio λ_i representa que tanta información contiene la i -ésima componente; es decir, si es mayor que

$$\frac{\sum_{i=1}^p \lambda_i}{p}$$

la i -ésima componente contendrá más información que la que debería proporcionar una variable (en caso contrario menos). Es decir que para nuestro estudio elegiremos el número de componentes a estudiarse de acuerdo al número de valores propios mayores al promedio de los λ s. Los vectores propios correspondientes a cada valor propio serán los vectores de cada componente principal a estudiarse. Así; tendremos tantas componentes como valores propios mayores al promedio de los λ s.

Un valor importante es la proporción de información que contiene la i -ésima componente principal. Este valor se calcula dividiendo el i -ésimo valor propio para el total acumulado entre todos los valores propios:

$$\% \text{ de la varianza total dado por el } i\text{-ésimo componente} = \frac{\lambda_i}{\lambda_1 + \lambda_2 + \dots + \lambda_p}$$

2.7.7 La Matriz de Factores.

En esta matriz la información nos empieza a dar la cara; la matriz de factores se verá de la siguiente manera:

Tabla 2.16
Matriz de factores

	Y_1	Y_2	...	Y_k
X_1	e_{11}	e_{12}	...	e_{1k}
X_2	e_{21}	e_{22}	...	e_{2k}
\vdots	\vdots	\vdots		\vdots
X_k	e_{p1}	e_{p2}	...	e_{pk}

Esta es una matriz cuyas columnas son los vectores propios correspondientes a los valores propios mayores a 1. cada uno de sus valores $e_{j,i}$ son los coeficientes de la i -ésima componente para la j -ésima variable y son directamente proporcionales a las correlaciones entre este componente y esta variable.

Por medio de esta matriz elegiremos en que componente debemos ubicar a cada variable. Mientras mayor sea $e_{j,i}$ más se relacionarán Y_i con X_j y viceversa. Por cuestiones prácticas; es usual añadir una columna r_{YX_k} a un lado de cada vector propio (como referencia) para evitar el problema de interpretación de la matriz de covarianzas causado por las diferentes escalas de medición de las variables. En otras

palabras ubicaremos a la variable X_j en la componente cuyo valor $[\text{abs}(r_{Y_i X_k})]$ se aproxime más a 1. La ecuación para el cálculo de las correlaciones entre Y_i y X_k es:

$$r_{Y_i X_k} = \frac{\text{cov}(Y_i, X_k)}{\sqrt{\text{var}(Y_i)}\sqrt{\text{var}(X_k)}} = \frac{\mathbf{I}_i e_{ki}}{\sqrt{\mathbf{I}_i} \sqrt{\mathbf{s}_{kk}}} = \frac{e_{ki} \sqrt{\mathbf{I}_i}}{\sqrt{\mathbf{s}_{kk}}}$$

donde: $i, k = 1, 2, \dots, p$

Capítulo III.

3. ANÁLISIS ESTADÍSTICO.

3.1 Análisis Univariado

El procedimiento que utilizaremos para realizar el análisis univariado de las variables de estudio será exhaustivo para cada una de ellas.

Analizaremos la distribución de cada variable de acuerdo a cada una de las situaciones dentro de las que se presentaron. Así por ejemplo, analizaremos la frecuencia diaria de las sustracciones de vehículos comparándolos de acuerdo al mes, de acuerdo a la hora, al lugar, etc.; en el caso de ser necesario imputaremos datos, de acuerdo a la distribución de probabilidades de la variable;

probaremos contrastes de hipótesis acerca de la independencia de las modalidades de cada variable; y finalmente dejaremos planteados interrogantes que serán resueltos en las secciones posteriores a esta.

3.1.1 Día del Robo

Observemos el histograma de frecuencias de la figura 3.1 que ilustra la frecuencia relativa de la media de ocurrencias obtenida de la tabla 3.1. Aquí podemos observar que aparentemente existe una tendencia en el número de vehículos sustraídos cada día de la semana. Ya que según parece las sustracciones van aumentando desde el domingo hasta el viernes.

Tabla 3.1: Frecuencias absoluta y relativa del total de sustracciones de vehículos en Guayaquil en el año 1999 contadas por día de la semana.

Día	Frec. Abs.	Frec. Rel.
LUNES	392	0.14
MARTES	394	0.14
MIERCOLES	398	0.14
JUEVES	438	0.16
VIERNES	433	0.16
SABADO	363	0.13
DOMINGO	346	0.13

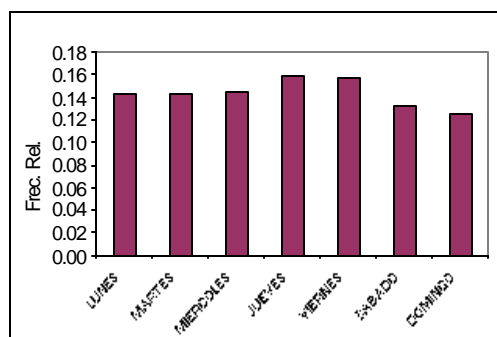


Figura 3.1: Histograma de frecuencias relativas del total de sustracciones de vehículos anuales contadas por día de la semana.

Para probarlo o contradecirlo debemos realizar un contraste de hipótesis para probar la uniformidad de la distribución de la variable “día de la semana”:

Ho: la distribución de sustracciones diarias es uniforme

Vs.

H₁: el número de sustracciones depende del día

Para probar este contraste de hipótesis como se explica en la sección 2.5 debemos formar una tabla de frecuencias observadas y esperadas (ver tabla 3.2),

Tabla 3.2:

Frecuencias observada y esperada y valor del estadístico χ^2 del número de sustracciones de vehículos en Guayaquil en el año 1999 contadas por día de la semana.

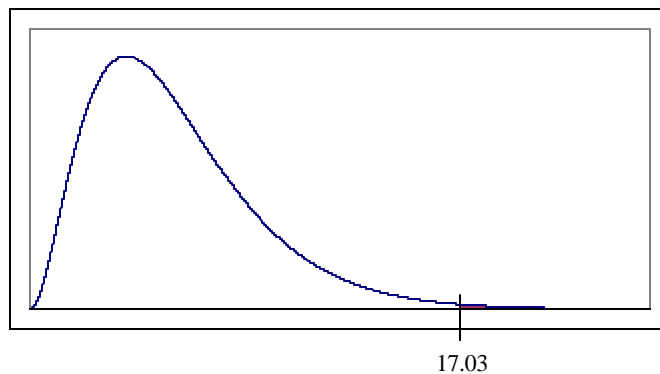
Día	Frec. Observada f_o	Frec. Esperada f_e	$f_o - f_e$	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
Domingo	392	1/7*2990=394.86	-2.86	8.16	0.02
Lunes	394	394.86	-0.86	0.73	0.00
Martes	398	394.86	3.14	9.88	0.03
Miércoles	438	394.86	43.14	1861.31	4.71
Jueves	433	394.86	38.14	1454.88	3.68
Viernes	363	394.86	-31.86	1014.88	2.57
Sábado	346	394.86	-48.86	2387.02	6.05
Totales	2764	1*2764			17.06

$$\sum_{i:1}^7 \frac{(f_{oi} - f_{ei})^2}{f_{ei}} = 17.06 \sim c^2_{a,6}$$

El valor del estimador del estadístico $\chi^2_{a,6}$ para esta prueba es 17.06, y comparando con el valor real del $\chi^2_{a,6}$ a 95% de confianza que es $\chi^2_{0.05,6} = 12.59 < 17.06$, entonces se rechaza la hipótesis nula, quedando evidenciado que existe diferencia significativa entre el número de sustracciones que se dan cada día de la semana.

Figura 3.2.

Distribución χ^2 con 6 grados de libertad para la prueba de bondad de ajuste



Viéndolo desde otro punto de vista, el área bajo la curva cuando $x \geq 17.03$ es $p = 0.009$ que es un valor pequeño para aceptar la hipótesis nula. En otras palabras, podemos decir que existe suficiente evidencia estadística para afirmar que el número de sustracciones que se den en una fecha determinada depende directamente del día de la semana al

que esta fecha corresponde. Hagamos estadística descriptiva para obtener mayor información.

Tabla 3.3:

Estadística descriptiva del número de sustracciones de vehículos en Guayaquil en el año 1999 contadas por día de la semana.

	lunes	martes	miércoles	jueves	viernes	sábado	domingo
Media	7.54	7.58	7.65	8.42	8.17	6.98	6.65
Mediana	7.50	7.00	8.00	8.00	8.00	6.00	6.00
Moda	5	11	9	8	6	6	5
Varianza	13.31	11.90	9.21	14.52	15.11	11.82	10.74
Desv. est.	3.65	3.45	3.03	3.81	3.89	3.44	3.28
Error	0.51	0.48	0.42	0.53	0.53	0.48	0.45
Mínimo	1	2	1	1	2	1	1
Máximo	16	14	16	17	18	18	15
Totales	392	394	398	438	433	363	346

Vemos aquí que aparentemente se están suscitando mayor cantidad de sustracciones de vehículos los días jueves y viernes, y si observamos con detenimiento nos daremos cuenta que las evidencias son substanciales, ya que la se podría decir que la media y la moda para ambos días coinciden. Incluso los valores máximos y los totales además del histograma de frecuencias que ilustra la figura 3.1 para ambos días sugieren esta teoría. Por otra parte el día domingo es el aparente día de descanso (por decirlo así), ya que se puede observar en la tabla que existe consistencia en los resultados por lo que podemos suponer que en este día ocurren menos sustracciones. Para probar estas teorías vamos a analizar todos los contrastes de hipótesis acerca de

la diferencia entre parejas de medias, entonces para cada par de medias vamos a probar los siguientes contrastes:

Primer contraste:

$$H_0: \mu_i = \mu_j$$

Vs.

$$H_1: \mu_i \neq \mu_j$$

Que es lo mismo que:

$$H_0: \mu_i - \mu_j = 0$$

Vs.

$$H_1: \mu_i - \mu_j \neq 0$$

Rechace H_0 a favor de H_1 si $|t_0| > t_{\frac{\alpha}{2}, n}$,

Segundo contraste:

$$H_0: \mu_i - \mu_j \leq 0$$

Vs.

$$H_1: \mu_i - \mu_j > 0$$

Rechace H_0 a favor de H_1 si $t_0 > t_{\alpha, n}$,

$$\text{donde } t_0 = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{s_1^2/n_1 + s_2^2/n_2}} \text{ y } n = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}{\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}}$$

Para $i = 1, 2, 3, 4, 5, 6$; $j = 2, 3, 4, 5, 6, 7$; $i < j$

Donde 1: lunes; 2: martes; 3: miércoles; 4: jueves; 5: viernes;
6: sábado; 7: domingo.

Para decidir de se acepta o rechaza la hipótesis nula en cada caso observemos los resultados de la tabla 3.4.

Tabla 3.4: Prueba t acerca de las diferencias entre las medias de sustracciones diarias de vehículos.

Contraste (Ho) ¹	t _o	n	t _[α/2,n]	t _[α,n]	t _o < t _[α/2,n]	t _o > t _[α,n]
lunes-martes	-0.06	101.68	1.98	----	No	----
lunes-miércoles	-0.18	98.73	1.98	----	No	----
lunes-jueves	-1.21	101.81	1.98	----	No	----
lunes-viernes	-0.86	102.80	1.98	----	No	----
lunes-sábado	0.80	101.64	1.98	----	No	----
lunes-domingo	1.30	100.85	1.98	----	No	----
martes-miércoles	-0.1	100.4	1.984	----	No	----
martes-jueves	-1.2	101	1.984	----	No	----
martes-viernes	-0.8	102	1.984	----	No	----
martes-sábado	0.88	102	1.984	----	No	----
martes-domingo	1.4	101.7	1.984	----	No	----
miércoles-jueves	-1.1	97.13	1.984	----	No	----
miércoles-viernes	-0.8	98.09	1.984	----	No	----
miércoles-sábado	1.06	100.5	1.984	----	No	----
miércoles-domingo	1.61	101.4	1.984	----	No	----
jueves-viernes	0.34	103	1.984	----	No	----
jueves-sábado	2.03	100.9	1.984	1.66	Si	Si
jueves- domingo	2.54	99.76	1.9842	1.66	Si	Si
viernes- sábado	1.66	101.9	1.984	----	No	----
viernes- domingo	2.16	100.7	1.984	1.66	Si	Si
sábado- domingo	0.5	101.8	1.984	----	No	----

A 95% de confianza encontramos que únicamente existen diferencias significativas entre el número de robos que ocurren el jueves y sábado, jueves y domingo, viernes y domingo, y además que el día jueves ocurren mas robos que

¹ Se asume que el contraste de hipótesis para estas pruebas es tal como está definido en la parte superior de esta página. Así por ejemplo para la primera fila de la tabla 3.4 el contraste correspondiente sería Ho: $\mu_1 - \mu_2 = 0$ Vs. H₁: $\mu_1 - \mu_2 \neq 0$.

durante el fin de semana y que ocurren mas robos el viernes que el domingo.

Ahora, construyamos intervalos de confianza para cada uno de los días de la semana, de tal manera que podamos establecer los días de mayor y menor ocurrencia de sustracciones.

Tabla 3.5: intervalos de confianza al 95% para el promedio de sustracciones de vehículos anuales en Guayaquil por día de la semana.

Día	mínimo	Máximo
LUNES	6.52	8.55
MARTES	6.62	8.54
MIÉRCOLES	6.81	8.50
JUEVES	7.36	9.48
VIERNES	7.10	9.24
SABADO	6.02	7.94
DOMINGO	5.74	7.57

Los resultados obtenidos en las dos tablas anteriores se pueden resumir en las siguientes conclusiones,

- no existen diferencias significativas entre los días **lunes, martes, miércoles, jueves y viernes** que son los días en los que suceden el mayor número de robos en la semana que están entre 6 y 9,
- tampoco existen diferencias significativas entre los días **sábado y domingo** y durante estos días ocurren menos sustracciones que en los demás encontrándose entre 6 y 8 sustracciones por día y,

Vemos entonces los días de menor riesgo son los dos últimos días de la semana donde suceden de 6 a 8 sustracciones diarias, mientras que durante el resto de la semana este valor aumenta y puede llegar a 9 o 10 robos diarios.

Otro punto interesante es calcular el promedio de sustracciones diarios de vehículos, sin tomar en cuenta a que día pertenece, este valor es $2764(\text{robos}) \div 365(\text{días}) = 7.57(\text{robos/día})$, es decir que **la media de robos por día es 8.**

Sería interesante conocer los días (las fechas) en los que han sucedido la mayor cantidad de sustracciones de vehículos, consideremos grandes los valores mayores a $\bar{x} + 2s$, sabemos que $\bar{x} = 7.57$ y al calcular la desviación estándar obtuvimos $s = 3.53$, entonces $\bar{x} + 2s = 7.57 + 2(3.53) = 14.63$, entonces encontremos las fechas en las que han sucedido esta cantidad de robos. La tabla 3.6 muestra las fechas en las que han sucedido mas de 14 robos.

Fecha	Cantidad
04-Feb-99	17
07-Feb-99	16
11-Feb-99	18
17-Jul-99	18
14-Sep-99	16
28-Sep-99	16
01-Oct-99	16
11-Oct-99	16
28-Oct-99	17
04-Nov-99	18
13-Nov-99	17
16-Nov-99	16
18-Nov-99	18
25-Nov-99	16
26-Nov-99	19
01-Dec-99	18

Tabla 3.6: lista de las fechas en las que han sucedido mas de 15 sustracciones de vehículos.

Vemos aquí que los días en los que suceden la mayor cantidad de sustracciones no tienen alguna característica especial que los distinga de los demás. Analicemos ahora la variable mes.

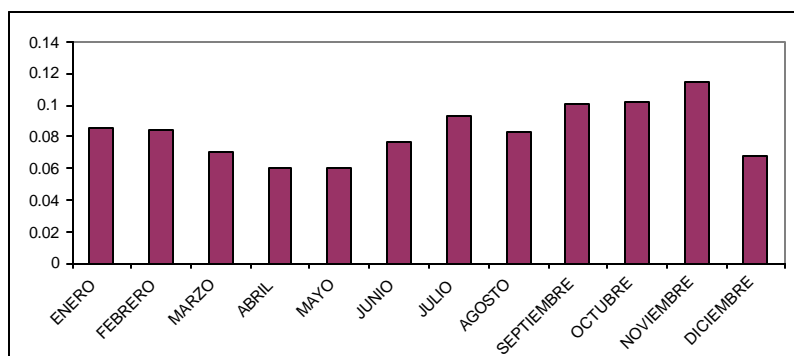
3.1.2 Mes del Robo

De nuevo empezamos estudiando el histograma de frecuencias de esta variable (figura 3.3), que se deriva de la tabla 3.7.

Mes	Frec. Abs.	Frec. Rel.
ENERO	238	0.09
FEBRERO	234	0.08
MARZO	196	0.07
ABRIL	166	0.06
MAYO	167	0.06
JUNIO	211	0.08
JULIO	258	0.09
AGOSTO	230	0.08
SEPTIEMBRE	278	0.10
OCTUBRE	280	0.10
NOVIEMBRE	317	0.11
DICIEMBRE	189	0.07

Tabla 3.7: Frecuencia absoluta y relativa de la suma del número de sustracciones de vehículos por mes ocurridas durante el año 1999.

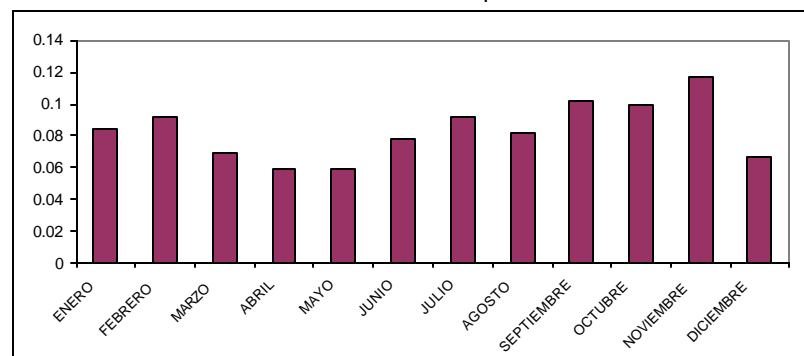
Figura 3.4: Histograma de frecuencias relativas del total de sustracciones de vehículos anuales contadas por mes del año.



Aparentemente existe una diferencia pronunciada entre la frecuencia de robos que se dan de mes a mes, puesto que en el histograma se ve que hay un amplio decrecimiento

entre Enero y Abril, volviendo a ascender a partir de Mayo hasta Noviembre, y con un decaimiento para Diciembre. Antes de continuar con el análisis nos encontramos con un problema, ya que el número de días que posee cada mes cambia de mes a mes, esto afecta directamente en la varianza debido al error por el sesgo, para evitar esto aumentaremos 7 valores, de tal manera que todos los meses del año tendrán 31 días. Ahora, observemos como queda el nuevo histograma de frecuencias relativas con los datos aumentados.

Figura 3.5: Histograma de frecuencias relativas (mejorado) del total de sustracciones de vehículos anuales contadas por mes del año.



Hagamos la prueba de bondad de ajuste para verificar veracidad de esta teoría, y probemos el siguiente contraste:

Ho: la distribución de sustracciones mensuales es uniforme

Vs.

H₁: no es verdad Ho

Tabla 3.8:

Frecuencias observada y esperada y valor del estadístico χ^2 del número de sustracciones de vehículos en Guayaquil en el año 1999 contadas por mes del año.

Mes	Frec. Observada f_o	Frec. Esperada f_e	$f_o - f_e$	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
ENERO	238	$(1/12) * 2815 = 234.58$	3.42	11.67	0.05
FEBRERO	258	234.58	23.42	548.34	2.34
MARZO	196	234.58	-38.58	1488.67	6.35
ABRIL	166	234.58	-68.58	4703.67	20.05
MAYO	167	234.58	-67.58	4567.51	19.47
JUNIO	218	234.58	-16.58	275.01	1.17
JULIO	258	234.58	23.42	548.34	2.34
AGOSTO	230	234.58	-4.58	21.01	0.09
SEPTIEMBRE	287	234.58	52.42	2747.51	11.71
OCTUBRE	280	234.58	45.42	2062.67	8.79
NOVIEMBRE	328	234.58	93.42	8726.67	37.20
DICIEMBRE	189	234.58	-45.58	2077.84	8.86
Totales	2815				118.42

$$\sum_{i:1}^{12} \frac{(f_{oi} - f_{ei})^2}{f_{ei}} = 118.42 \sim \chi^2_{a,11}$$

El valor p de una ji cuadrado con 11 grados de libertad es $p = 3.77 \times 10^{-20}$ que obviamente es muy pequeño comparado con 0.05, es decir que a 95% de confianza rechazamos la hipótesis nula respecto a la igualdad entre todas las medias de los meses del año. Proseguimos a realizar los cálculos de la estadística descriptiva de estas variables (ver tabla 3.9).

Tabla 3.9:

Estadística descriptiva del número de sustracciones de vehículos en Guayaquil en el año 1999 contadas por mes del año.⁽²⁾

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
Media	7.68	8.36	6.32	5.53	5.39	7.03	8.32	7.42	9.27	9.03	10.57	6.10
Mediana	8	9	6	6	5	7	9	7	9	9	11	7
Moda	11.00	5.00	5.00	3.00	5.00	7.00	9.00	6.00	9.00	6.00	8.00	8.00
Varianza	11.89	17.42	8.16	6.26	5.71	5.14	10.69	13.05	8.34	13.50	12.81	13.09
Desv. est.	3.45	4.17	2.86	2.50	2.39	2.27	3.27	3.61	2.89	3.67	3.58	3.62
Error	1	1	1	0	0	0	1	1	1	1	1	1
Mínimo	1	2	1	1	2	3	1	3	2	3	5	1
Máximo	14	17	12	11	11	13	18	18	14	16	18	16
Totales	238	234	196	166	167	211	258	230	278	280	317	189

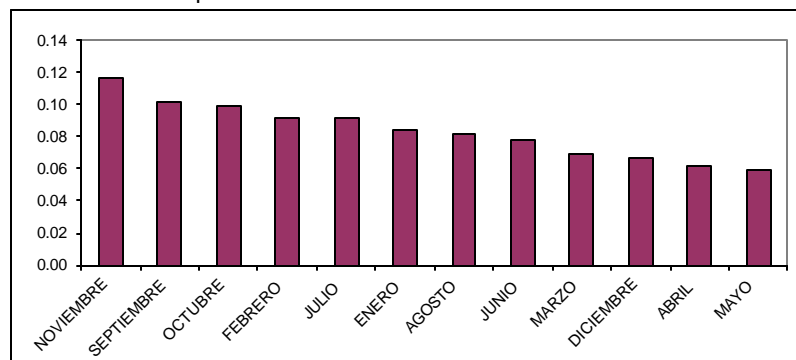
Vemos aquí existe una clara diferencia entre la media mediana y moda del mes de Noviembre. También podemos notar que en los meses de Abril y Mayo ocurren un menor número de sustracciones de vehículos que en los demás meses. Vemos que los totales

de las columnas también coinciden con esta teoría. Hagamos la prueba de diferencia de medias y formemos intervalos de confianza para conocer los meses de mayor y menor incidencia de robos.

² Debido a que el número de días de cada mes se diferencian uno de otro, se han imputado 8 datos (con la media mensual) de tal manera que se lograron 31 días en cada mes todo esto con el objeto de minimizar el error por sesgo.

De acuerdo al diagrama de Pareto presentado a continuación, se puede esperar que ocurran mayor cantidad de robos diarios en los meses de Noviembre, Septiembre y Octubre, y que ocurran mucho menos robos en Diciembre, Abril y Mayo.

Figura 3.6: Diagrama de Pareto del total de sustracciones de vehículos anuales contadas por mes del año.



Dividamos el dominio en grupos de tal manera que la probabilidad de que cada uno suceda sea la mismo o casi la misma.

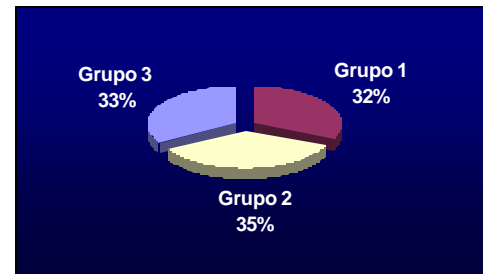
Tabla 3.10: Segmentación de sustracciones de acuerdo al mes del robo tomando como base la proporción de robos acumulada.

Grupo	1	2	3
Meses	Noviembre, Septiembre, Octubre	Febrero, Julio, Enero, Agosto	Junio, Marzo, Diciembre, Abril, Mayo
Porcentaje acumulado	32	35	33

Tabla 3.11: intervalos de confianza al 95% para el promedio diario de sustracciones de vehículos de acuerdo al mes en que ocurrieron.

Día	Mínimo	Máximo
Enero	6	9
Febrero	7	10
Marzo	5	7
Abril	5	6
Mayo	5	6
Junio	6	8
Julio	7	10
Agosto	6	9
Septiembre	8	10
Octubre	8	10
Noviembre	9	12
Diciembre	5	7

Figura 3.6: Segmentación de sustracciones de acuerdo al mes del robo tomando como base la proporción de robos acumulada.



Hemos obtenido tres grupos donde en cada grupo se encuentra aproximadamente el 33% del total de los robos de vehículos dados durante el año.

Grupo 1 (4-8 robos):

- Marzo (5-7)
- Abril (5-6)
- Mayo (5-6)
- Junio (6-8)
- Diciembre (5-7)

Grupo 2 (6-10 robos):

- Enero (6-9)
- Febrero (7-10)
- Julio (7-10)
- Agosto (6-9)

Grupo 3 (7-12 robos):

- Septiembre (8-10)
- Octubre (8-10)
- Noviembre (9-12)

podemos concluir lo siguiente:

- En los meses de **Septiembre, Octubre y Noviembre** suceden la mayor cantidad de robos al día: entre 7 y 12 robos.
- En los meses de **Enero, Febrero, Julio y Agosto** disminuye la actividad delictiva para los roba_carros (entre 6 y 10 robos).
- Y, en los meses de **Marzo, Abril, Mayo, Junio y Diciembre** sucede la menor cantidad de robos de vehículos en promedio, de 4 a 8 robos.

3.1.3 Hora del Robo

Por lo general, se dice que el número de robos aumenta durante las horas en las que las personas viajan desde el hogar hacia el lugar de trabajo y viceversa, mas no cuando los autos están estacionados, sería interesante verificar que se cumpla esta teoría. Por lo pronto verifiquemos si esta teoría es cierta en lo que respecta a la hora del robo y más adelante examinaremos si la mayor cantidad de robos se dan cuando el vehículo está en movimiento.

Antes de empezar a realizar el análisis encontramos que el número total de registros era de 2698 pero que habían 182 datos faltantes, por esta razón debimos imputarlos, para esto simulamos horas de robo mediante la generación de números aleatorios de la siguiente manera:

1. se obtuvo la función de distribución de la variable hora,
2. se calculó la función de densidad (distribución acumulada) de la variable hora,
3. se generaron números aleatorios y
4. finalmente (de acuerdo al número aleatorio) se asignaron valores a la variable.

La tabla 3.12 resume los cálculos realizados.

Tabla 3.12: funciones de distribución y densidad de la variable hora para usarse al imputar datos.

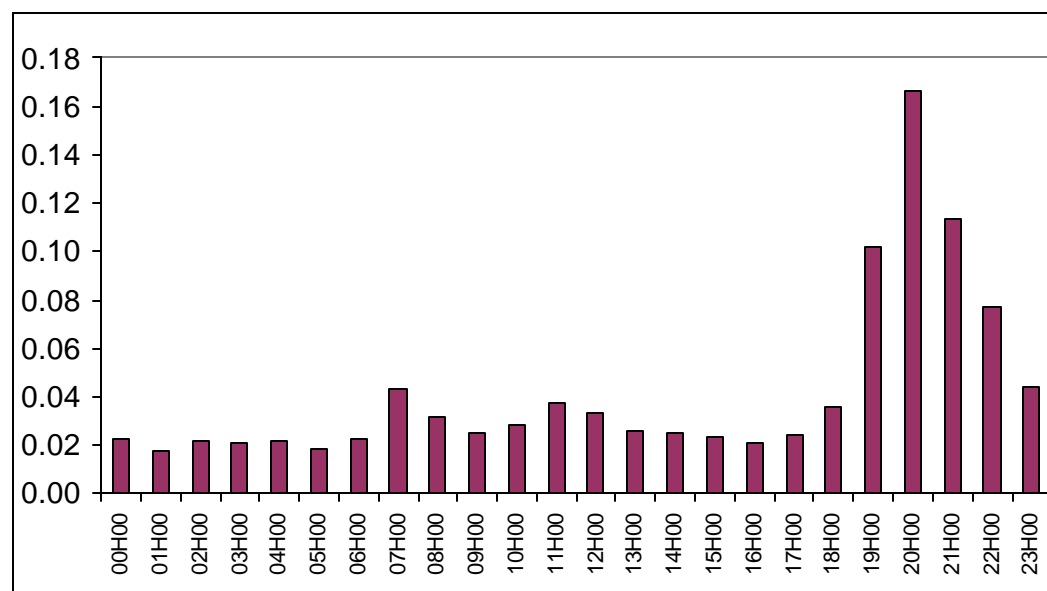
hora	Cuenta	p(x)	F(x)	
			desde	hasta
00H00	66	0.023	0.000	0.023
01H00	54	0.019	0.024	0.043
02H00	59	0.021	0.044	0.063
03H00	66	0.023	0.064	0.087
04H00	68	0.024	0.088	0.111
05H00	56	0.020	0.112	0.131
06H00	64	0.023	0.132	0.154
07H00	118	0.042	0.155	0.195
08H00	82	0.029	0.196	0.225
09H00	72	0.026	0.226	0.250
10H00	85	0.030	0.251	0.280
11H00	104	0.037	0.281	0.317
12H00	92	0.033	0.318	0.350
13H00	71	0.025	0.351	0.375
14H00	71	0.025	0.376	0.400
15H00	68	0.024	0.401	0.424
16H00	57	0.020	0.425	0.444
17H00	68	0.024	0.445	0.469
18H00	105	0.037	0.470	0.506
19H00	279	0.099	0.507	0.605
20H00	460	0.163	0.606	0.768
21H00	313	0.111	0.769	0.879
22H00	211	0.075	0.880	0.954
23H00	130	0.046	0.955	1.000

Examinemos el histograma de frecuencias de la figura 3.7 obtenida a partir de la tabla 3.13.

Tabla 3.13: Frecuencias absoluta y relativa del total de sustracciones de vehículos en Guayaquil en el año 1999 contadas por hora del día.

Hora ⁽³⁾	Frec. Abs.	Frec. Rel.
00H00	63	0.02
01H00	48	0.02
02H00	59	0.02
03H00	57	0.02
04H00	61	0.02
05H00	51	0.02
06H00	62	0.02
07H00	119	0.04
08H00	87	0.03
09H00	70	0.03
10H00	78	0.03
11H00	103	0.04
12H00	92	0.03
13H00	71	0.03
14H00	70	0.03
15H00	65	0.02
16H00	57	0.02
17H00	66	0.02
18H00	98	0.04
19H00	281	0.10
20H00	459	0.17
21H00	313	0.11
22H00	212	0.08
23H00	122	0.04

Figura 3.7: Histograma de frecuencias relativas del total de sustracciones de vehículos anuales contadas por hora del día.



³ Se debe asumir 0H00 como 0H00 £ X < 1H00; 1H00 como 1H00 £ X < 2H00 y así sucesivamente.

Se puede observar claramente que la mayoría de robos se están dando desde las 19H00 hasta después de las 23H00 con una cota superior entre las 20H00 y las 21H00, y es notorio también que durante el intervalo sobrante suceden una cantidad casi constante de sustracciones que constituyen (cada una) entre el 2 y 4 por ciento del total de sustracciones diarias. Probemos si el número de sustracciones depende de la hora. Establezcamos el contraste de hipótesis que vamos a probar.

H_0 : la distribución de sustracciones por hora es uniforme

Vs.

H_1 : el número de sustracciones depende de la hora

Tabla 3.14:

Frecuencias observada y esperada y valor del estadístico χ^2 del número de sustracciones de vehículos en Guayaquil en el año 1999 contadas por hora del día.

Hora	Frec. Observada f_o	Frec. Esperada f_e	$f_o - f_e$	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
00H00	63	115.17	-52.17	2721.36	23.63
01H00	48	115.17	-67.17	4511.36	39.17
02H00	59	115.17	-56.17	3154.69	27.39
03H00	57	115.17	-58.17	3383.36	29.38
04H00	61	115.17	-54.17	2934.03	25.48
05H00	51	115.17	-64.17	4117.36	35.75
06H00	62	115.17	-53.17	2826.69	24.54
07H00	119	115.17	3.83	14.69	0.13
08H00	87	115.17	-28.17	793.36	6.89
09H00	70	115.17	-45.17	2040.03	17.71
10H00	78	115.17	-37.17	1381.36	11.99
11H00	103	115.17	-12.17	148.03	1.29
12H00	92	115.17	-23.17	536.69	4.66

Hora	Frec. Observada f_o	Frec. Esperada f_e	$f_o - f_e$	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
13H00	71	115.17	-44.17	1950.69	16.94
14H00	70	115.17	-45.17	2040.03	17.71
15H00	65	115.17	-50.17	2516.69	21.85
16H00	57	115.17	-58.17	3383.36	29.38
17H00	66	115.17	-49.17	2417.36	20.99
18H00	98	115.17	-17.17	294.69	2.56
19H00	281	115.17	165.83	27500.69	238.79
20H00	459	115.17	343.83	118221.36	1026.52
21H00	313	115.17	197.83	39138.03	339.84
22H00	212	115.17	96.83	9376.69	81.42
23H00	122	115.17	6.83	46.69	0.41
Totales	2764				2044.42

$$\sum_{i:1}^{24} \frac{(f_{oi} - f_{ei})^2}{f_{ei}} = 2044.42 \sim \chi^2_{a,23}$$

Debido a que el estadístico de prueba ji cuadrado es muy grande $\chi^2_{a,23} = 2044.42$ su valor p correspondiente es muy pequeño $p = 5.75 \times 10^{-430}$ por lo que existe demasiada evidencia estadística para rechazar la hipótesis nula de igualdad de medias, que es algo que ya suponíamos iba a suceder. Ahora, examinemos la tabla 3.15 que detalla la estadística descriptiva de la variable hora.

Tabla 3.15:

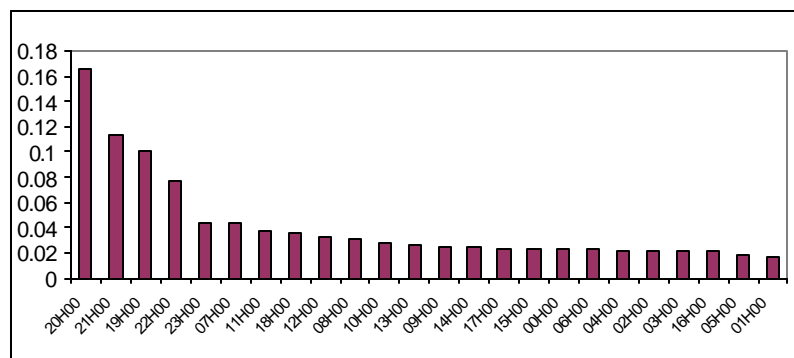
Estadística descriptiva del número de sustracciones de vehículos en Guayaquil en el año 1999 contadas por hora del día.

	00H00	01H00	02H00	03H00	04H00	05H00	06H00	07H00	08H00	09H00	10H00	11H00
Media	1.19	1.09	1.16	1.10	1.05	1.06	1.19	1.24	1.19	1.11	1.15	1.13
Mediana	1	1	1	1	1	1	1	1	1	1	1	1
Moda	1	1	1	1	1	1	1	1	1	1	1	1
Varianza	0.06	0.04	0.06	0.04	0.03	0.04	0.06	0.05	0.05	0.05	0.04	0.04
Desv. est.	0.19	0.08	0.17	0.09	0.05	0.06	0.20	0.27	0.18	0.13	0.13	0.14
Error	0.44	0.29	0.42	0.30	0.22	0.24	0.44	0.52	0.43	0.36	0.36	0.37
Mínimo	1	1	1	1	1	1	1	1	1	1	1	1
Máximo	3	2	3	2	2	2	3	3	3	3	2	3
Totales	63	48	59	57	61	51	62	119	87	70	78	103

	12H00	13H00	14H00	15H00	16H00	17H00	18H00	19H00	20H00	21H00	22H00	23H00
Media	1.19	1.09	1.16	1.10	1.05	1.06	1.19	1.24	1.19	1.11	1.15	1.13
Mediana	1	1	1	1	1	1	1	1	1	1	1	1
Moda	1	1	1	1	1	1	1	1	1	1	1	1
Varianza	0.06	0.04	0.06	0.04	0.03	0.04	0.06	0.05	0.05	0.05	0.04	0.04
Desv. est.	0.19	0.08	0.17	0.09	0.05	0.06	0.20	0.27	0.18	0.13	0.13	0.14
Error	0.44	0.29	0.42	0.30	0.22	0.24	0.44	0.52	0.43	0.36	0.36	0.37
Mínimo	1	1	1	1	1	1	1	1	1	1	1	1
Máximo	3	2	3	2	2	2	3	3	3	3	2	3
Totales	63	48	59	57	61	51	62	119	87	70	78	103

Podemos observar que durante la mayoría de las horas del día están sucediendo un robo en promedio por hora, a excepción de las horas entre las 19H00 y las 22H00. Observemos el diagrama de Pareto que se muestra a continuación:

Figura 3.8: Diagrama de Pareto del total de sustracciones de vehículos anuales contadas por hora del día.

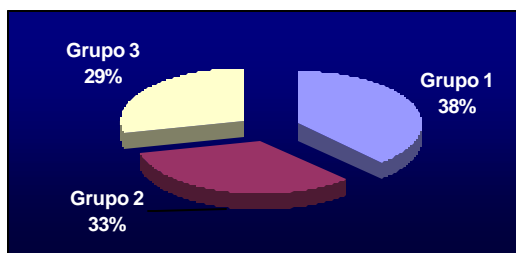


Aquí podemos notar claramente que las horas de mayor incidencia delictiva se dan entre las 19 y las 23 horas, que sobresalen claramente de las demás en el gráfico. Dividamos estas modalidades en tres grupos, como se hizo con los meses del año, y observemos que sucede.

Tablas 3.16: Segmentación de sustracciones de acuerdo a la hora del robo tomando como base la proporción de robos acumulada.

Grupo	1	2	3
Hora	19H00-20H00 20H00-21H00 21H00-22H00	7H00-8H00	0H00-1H00
		8H00-9H00	1H00-2H00
		10H00-11H00	2H00-3H00
		11H00-12H00	3H00-4H00
		12H00-13H00	4H00-5H00
		18H00-19H00	5H00-6H00
		22H00-23H00	6H00-7H00
		23H00-24H00	9H00-10H00
			13H00-14H00
			14H00-15H00
			15H00-16H00
			16H00-17H00
			17H00-18H00
		Porcentaje acumulado	38

Figura 3.9: Segmentación de sustracciones de acuerdo a la hora del robo tomando como base la proporción de robos acumulada.



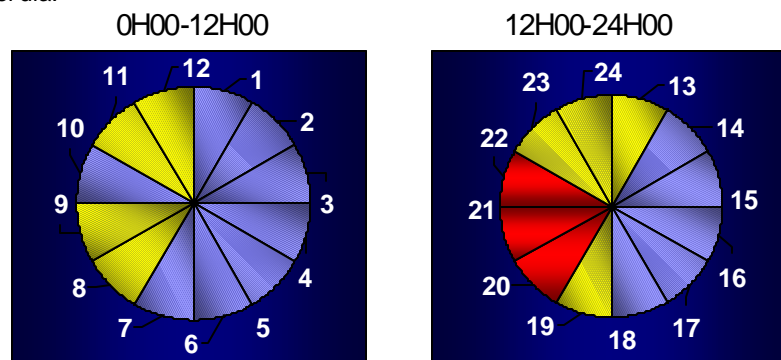
De acuerdo a la tabla 3.16 y la figura 3.9 hemos obtenido 3 grupos que se pueden resumir en la siguiente tabla

Tablas 3.17: Resumen de la segmentación de sustracciones de acuerdo a la hora del robo tomando como base la proporción de robos acumulada.

Grupo	Riesgo Alto	Riesgo Medio	Riesgo Bajo
Hora	19H00-22H00	7H00-9H00	0H00-7H00
		10H00-13H00	9H00-10H00
		18H00-19H00	13H00-18H00
		22H00-24H00	
Porcentaje acumulado	38	33	29

Si recordamos la hipótesis postulada acerca del supuesto de que el número de robos aumenta durante las horas en que los vehículos viajan desde el hogar hacia el trabajo y viceversa, vemos que efectivamente el número de sustracciones es mayor durante las horas de la noche en las que se viaja de regreso a casa después de un día normal de trabajo. Por el contrario, en las horas de la madrugada, cuando la gente descansa en sus hogares y durante las horas en las que generalmente se trabaja número de robos se hace mínimo, así como lo muestran los dos relojes de la figura 3.10

Figura 3.10: Cuadro indicativo del riesgo de robo de acuerdo a la hora del día.



Color rojo: riesgo alto; amarillo: riesgo medio; azul: riesgo bajo.

Vemos entonces que efectivamente entre las 7 y 10 de la noche (horas en las que viajamos del lugar de trabajo hacia

el hogar) ocurren una mayor cantidad de sustracciones de vehículos.

Las horas de riesgo medio son de 7 a 9 AM (horas en las que nos dirigimos desde el hogar hacia lugar de trabajo); de 10AM a 13PM (horas del almuerzo); de 6 a 7 PM y de 10 PM a 12 AM.

Las horas de la madrugada desde las cero horas hasta las 7 de la mañana, de 9 a 10 AM y por la tarde de 1 a 6 PM son las horas de menor riesgo, horas en las que el tráfico vehicular es mínimo.

3.1.4 Marca del Vehículo.

La marca es una de las variables mas importantes a investigar, ya que para el dueño de un vehículo debería resultar por lo menos curioso saber si su carro se encuentra dentro de la lista de los mas solicitados por la delincuencia. Se puede esperar que la Marca Chevrolet caiga dentro de este grupo, ya que es notorio que la mayoría de los vehículos que circulan en la ciudad tienen esta marca.

De aquí en adelante graficaremos el diagrama de Pareto de la variable en estudio, pero antes observemos la distribución de la proporción de vehículos robados por marca durante el año 1999.

Tabla 3.18: Frecuencias absoluta y relativa del total de sustracciones de vehículos en Guayaquil en el año 1999 contadas por marca de carro.

Marca	Frec. Abs.	Frec. Rel.
ALFA ROMEO	1	0.00
AUSTIN	4	0.00
BMW	5	0.00
CHEVROLET	916	0.33
CHRYSLER	2	0.00
CITROEN	1	0.00
DACIA	1	0.00
DAEWOO	48	0.02
DAIHATSU	19	0.01
DATSUN	134	0.05
DELTA	2	0.00
DODGE	3	0.00
ENCAVA	1	0.00
FIAT	123	0.04
FORD	121	0.04
GMC	2	0.00
HINO	5	0.00
HONDA	12	0.00
HYUNDAI	82	0.03
INTERNATIONAL	3	0.00
ISUZU	29	0.01
JEEP	19	0.01
JHONN DEERE	1	0.00
KIA	13	0.00

Marca	Frec. Abs.	Frec. Rel.
LADA	74	0.03
LEXUS	1	0.00
MACK	2	0.00
MAZDA	184	0.07
MERCEDES BENZ	9	0.00
MERCURY	1	0.00
MITSUBISHI	153	0.06
MORRIS	2	0.00
MUSTANG	1	0.00
NISSAN	208	0.08
OPEL RECORD	1	0.00
PEUGEOT	5	0.00
PLYMOUTH	1	0.00
SAEHAN	2	0.00
SANGYONG	1	0.00
SCANIA	2	0.00
SEAT	2	0.00
SKODA	65	0.02
SUBARU	5	0.00
SUZUKI	230	0.08
TOYOTA	186	0.07
TRIUMPH	1	0.00
VOLKSWAGEN	73	0.03
VOLVO	3	0.00

Marca	Frec. Abs.	Frec. Rel.
WILLYS	2	0.00
ZASTAVA	3	0.00

Figura 3.10: Diagrama de Pareto de la Proporción de sustracciones de vehículos anuales contadas por marca de carro.

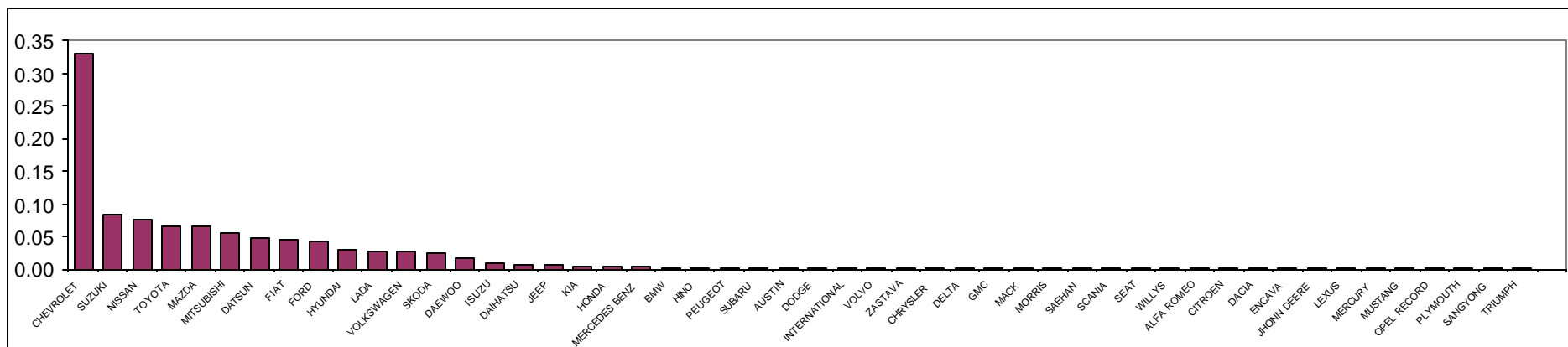
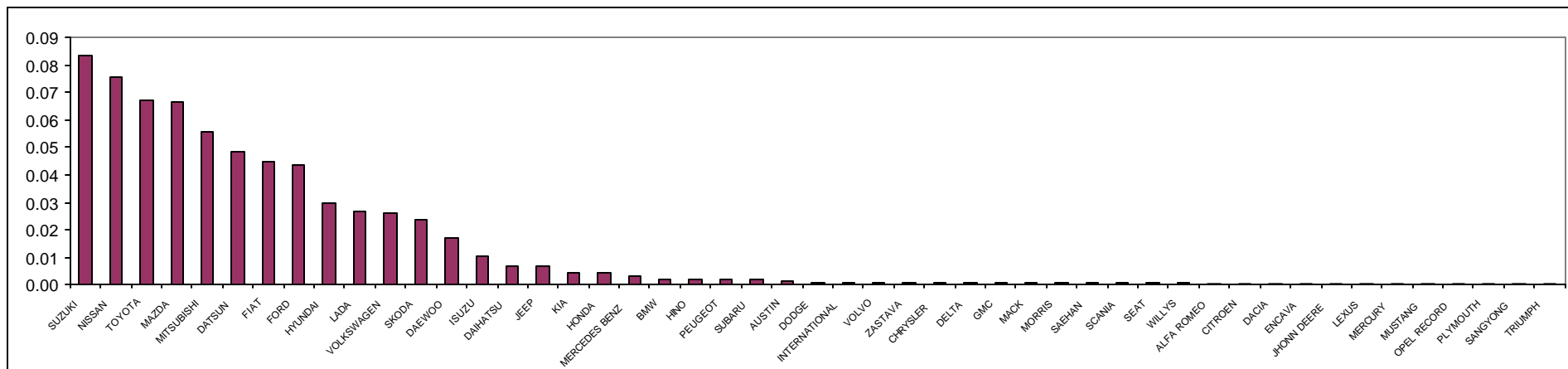


Figura 3.11: Diagrama de Pareto de la proporción de sustracciones de vehículos anuales contadas por marca de carro, eliminando los valores aberrantes.



Podemos notar muy claramente que el número de robos de carros con la marca Chevrolet sobrepasa grandemente a la cantidad de sustracciones de vehículos de las demás marcas ya que esta marca acumula aproximadamente el 33% de los robos anuales; es claro que esta variable en si misma representa un grupo que se distingue de los demás, es por eso que la hemos considerado un valor aberrante y hemos graficado otro diagrama de Pareto omitiendo esta variable, con el objeto de observar mejor la distribución de las demás variables.

Hagamos la prueba de linealidad usando el estadístico Ji-Cuadrado para probar que existen diferencias significativas entre la cantidad de robos de cada una de las marcas.

Tabla 3.19:

Frecuencias observada y esperada y valor del estadístico χ^2 del número de sustracciones de vehículos en Guayaquil en el año 1999 contadas por marca de carro.

Marca	Frec. Observada f_o	Frec. Esperada f_e	$f_o - f_e$	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
ALFA ROMEO	1	55.28	-54.28	2946.32	53.30
AUSTIN	4	55.28	-51.28	2629.64	47.57
BMW	5	55.28	-50.28	2528.08	45.73
CHEVROLET	916	55.28	860.72	740838.92	13401.57
CHRYSLER	2	55.28	-53.28	2838.76	51.35
CITROEN	1	55.28	-54.28	2946.32	53.30

Marca	Frec. Observada f_o	Frec. Esperada f_e	$f_o - f_e$	$(f_o - f_e)^2$	$(f_o - f_e)^2$ f_e
DACIA	1	55.28	-54.28	2946.32	53.30
DAEWOO	48	55.28	-7.28	53.00	0.96
DAIHATSU	19	55.28	-36.28	1316.24	23.81
DATSUN	134	55.28	78.72	6196.84	112.10
DELTA	2	55.28	-53.28	2838.76	51.35
DODGE	3	55.28	-52.28	2733.20	49.44
ENCAVA	1	55.28	-54.28	2946.32	53.30
FIAT	123	55.28	67.72	4586.00	82.96
FORD	121	55.28	65.72	4319.12	78.13
GMC	2	55.28	-53.28	2838.76	51.35
HINO	5	55.28	-50.28	2528.08	45.73
HONDA	12	55.28	-43.28	1873.16	33.88
HYUNDAI	82	55.28	26.72	713.96	12.92
INTERNATIONAL	3	55.28	-52.28	2733.20	49.44
ISUZU	29	55.28	-26.28	690.64	12.49
JEEP	19	55.28	-36.28	1316.24	23.81
JHONN DEERE	1	55.28	-54.28	2946.32	53.30
KIA	13	55.28	-42.28	1787.60	32.34
LADA	74	55.28	18.72	350.44	6.34
LEXUS	1	55.28	-54.28	2946.32	53.30
MACK	2	55.28	-53.28	2838.76	51.35
MAZDA	184	55.28	128.72	16568.84	299.73
MERCEDES BENZ	9	55.28	-46.28	2141.84	38.75
MERCURY	1	55.28	-54.28	2946.32	53.30
MITSUBISHI	153	55.28	97.72	9549.20	172.74
MORRIS	2	55.28	-53.28	2838.76	51.35
MUSTANG	1	55.28	-54.28	2946.32	53.30
NISSAN	208	55.28	152.72	23323.40	421.91
OPEL RECORD	1	55.28	-54.28	2946.32	53.30
PEUGEOT	5	55.28	-50.28	2528.08	45.73
PLYMOUTH	1	55.28	-54.28	2946.32	53.30
SAEHAN	2	55.28	-53.28	2838.76	51.35
SANGYONG	1	55.28	-54.28	2946.32	53.30
SCANIA	2	55.28	-53.28	2838.76	51.35
SEAT	2	55.28	-53.28	2838.76	51.35
SKODA	65	55.28	9.72	94.48	1.71
SUBARU	5	55.28	-50.28	2528.08	45.73
SUZUKI	230	55.28	174.72	30527.08	552.23
TOYOTA	186	55.28	130.72	17087.72	309.11
TRIUMPH	1	55.28	-54.28	2946.32	53.30
VOLKSWAGEN	73	55.28	17.72	314.00	5.68
VOLVO	3	55.28	-52.28	2733.20	49.44
WILLYS	2	55.28	-53.28	2838.76	51.35
ZASTAVA	3	55.28	-52.28	2733.20	49.44
Totales	2764				17153.19

$$\sum_{i:1}^{51} \frac{(f_{oi} - f_{ei})^2}{f_{ei}} = 17153.19 \sim c^2_{a,50}$$

Debido a que el estadístico de prueba ji cuadrado es muy grande $c^2_{a,53} = 17153.19$ su valor p correspondiente es muy pequeño (cercano a cero) por lo que existe demasiada evidencia estadística para rechazar la hipótesis de igualdad de medias. Esto implica que existen diferencias significativas entre el número de robos que suceden por cada marca de carro.

En la tabla que se muestra a continuación resumimos 4 subgrupos de marcas de vehículos de acuerdo a la proporción de robos que se han dado.

Tabla 3.20: Segmentación de sustracciones de acuerdo a la marca del vehículo tomando como base la proporción de robos acumulada.

Tipo de Riesgo	Marcas				%
Muy alto	Chevrolet				33
Alto	Nissan, Suzuki, Toyota				23
Medio	Mazda, Mitsubishi, Datsun, Fiat				21
Bajo	Alfa Romeo Austin BMW Chrysler Citroen Dacia Daewoo Daihatsu Delta Dodge	Encava Ford GMC Hino Honda Hyundai International Isuzu Jeep John Deere Kia	Lada Lexus Mack Merc. Benz Mercury Morris Mustang Opel Rec. Peugeot Plymouth	Saehan Sang Yong Scania Seat Skoda Subaru Triumph Volkswagen Volvo Willys Zastava	23

Podemos entonces decir que la marca potencialmente susceptible a ser robada es Chevrolet, ya que el 33% de los vehículos robados corresponden a esta marca. De los vehículos restantes hemos obtenido tres grupos que contienen aproximadamente $(100 - 33)\% / 3 = 66\% / 3 = 22\%$ del total de robos anuales. De esta segmentación vemos que seguido de la marca Chevrolet, se encuentran las marcas Nissan, Suzuki y Toyota que constituyen un grupo al que hemos nombrado de riesgo alto, seguido por las marcas Mazda, Mitsubishi, Datsun y Fiat que componen el grupo de riesgo medio y finalmente el grupo que incluye todas las marcas restantes, al que hemos nombrado de riesgo bajo.

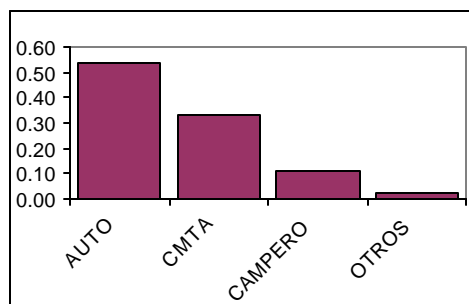
3.1.5 Tipo de Vehículo.

Ya que contamos con información disponible hagamos estadística alrededor del tipo de vehículo robado. La variable tipo puede tomar uno de cuatro valores: auto, campero, camioneta, otros. Dibujemos el diagrama de Pareto de la variable en estudio.

Tabla 3.21: Frecuencias absoluta y relativa del número de sustracciones de vehículos de acuerdo a su tipo.

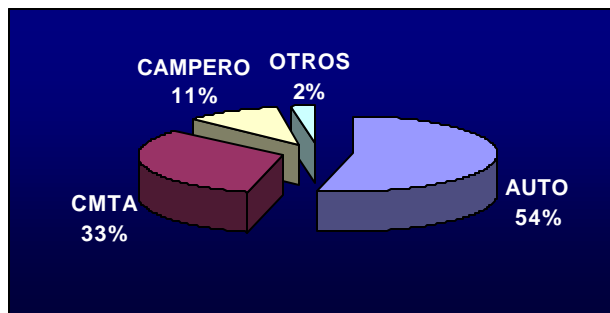
Tipo	Frec. Abs.	Frec. Rel.
AUTO	1486	0.54
CMTA	908	0.33
CAMPERO	301	0.11
OTROS	69	0.02

Figura 3.12: Diagrama de Pareto de la proporción de sustracciones de vehículos de acuerdo a su tipo.



Claramente se observa que el tipo de vehículo del que más se han reportado sustracciones ha sido el automóvil (54%), seguido de la camioneta (33%) y el campero (11%), dejando muy atrás a los demás tipos, por ejemplo los buses, camiones, furgonetas, tráileres, etc. (2%).

Figura 3.13: Diagrama circular de la proporción de sustracciones ocurridas en 1999 de acuerdo al tipo de vehículo.



Para confirmar lo dicho observemos los resultados de la prueba de Bondad de ajuste.

Tabla 3.22:

Frecuencias observada y esperada y valor del estadístico χ^2 del número de sustracciones de vehículos en Guayaquil en el año 1999 contadas por tipo de carro.

Tipo	Frec. Observada f_o	Frec. Esperada f_e	$f_o - f_e$	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
AUTO	1486	691.00	795.00	632025.00	914.65
CAMPERO	301	691.00	-390.00	152100.00	220.12
CMTA	908	691.00	217.00	47089.00	68.15
OTROS	69	691.00	-622.00	386884.00	559.89
Totales	2764				1762.80

$$\sum_{i:1}^4 \frac{(f_{oi} - f_{ei})^2}{f_{ei}} = 1762.80 \sim c^2_{a,3}$$

Una vez mas el estadístico de prueba ji cuadrado es muy grande $c^2_{a,3} = 1762.80$ para rechazar la hipótesis nula de igualdad de medias por lo que podemos concluir que existen

diferencias significativas entre el número de robos que suceden de acuerdo al tipo de carro.

Concluimos entonces que los casos de robos de vehículos pesados no son muy comunes; y mientras más liviano es el carro, mayor es la probabilidad de ser sustraído por los delincuentes.

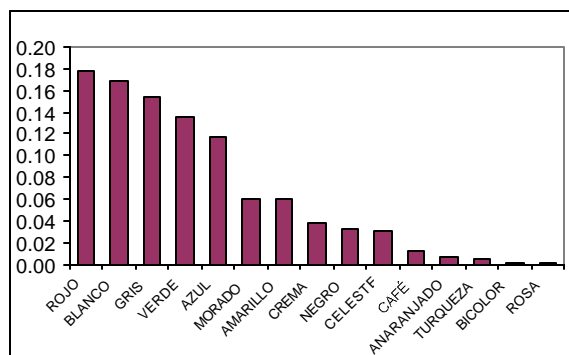
3.1.6 Color del Vehículo.

Esta variable es muy importante ya que es posible que el color de un vehículo influya en la decisión de ser robado o no. Debido a que la lista de colores es demasiado extensa se decidió agruparlos de acuerdo a sus tonos, es así que se agruparon el color *rojo* con el tomate, ladrillo y vino; el *crema* con el beige, brandy, champagne y habano, entre otros. De este resumen obtuvimos 15 variables, que hemos graficado en el diagrama de Pareto de la figura que se muestra a continuación.

Tabla 3.23: Frecuencias absoluta y relativa del número de sustracciones anuales de la variable color

Tipo	Frec. Abs.	Frec. Rel.
ROJO	490	0.18
BLANCO	465	0.17
GRIS	426	0.15
VERDE	374	0.14
AZUL	325	0.12
MORADO	168	0.06
AMARILLO	167	0.06
CREMA	103	0.04
NEGRO	91	0.03
CELESTE	82	0.03
CAFÉ	33	0.01
ANARANJADO	16	0.01
TURQUEZA	15	0.01
BICOLOR	5	0.00
ROSA	4	0.00

Figura 3.14: Diagrama de Pareto de la proporción de vehículos robados de acuerdo a su color.



Vemos aquí que los colores de mayor riesgo son el rojo y blanco, seguido del gris, verde y el azul, dejando a los demás colores con baja ponderación. Probemos la hipótesis de diferencia de medias y comparemos el valor del estadístico Ji-cuadrado con el de la variable aleatoria para aceptar o rechazar la hipótesis nula de igualdad de medias.

Tabla 3.24:

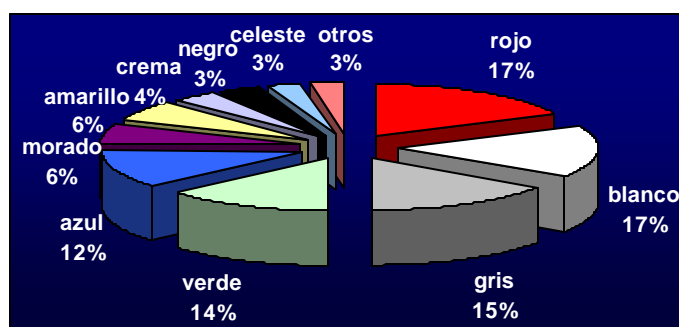
Frecuencias observada y esperada y valor del estadístico χ^2 del número de sustracciones de vehículos en Guayaquil en el año 1999 contadas por color del carro.

Color	Frec. Observada f_o	Frec. Esperada f_e	$f_o - f_e$	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
AMARILLO	167	184.27	-17.27	298.14	1.62
ANARANJADO	16	184.27	-168.27	28313.67	153.66
AZUL	325	184.27	140.73	19805.87	107.48
BICOLOR	5	184.27	-179.27	32136.54	174.40
BLANCO	465	184.27	280.73	78811.20	427.70
CAFÉ	33	184.27	-151.27	22881.60	124.18
CELESTE	82	184.27	-102.27	10458.47	56.76
CREMA	103	184.27	-81.27	6604.27	35.84
GRIS	426	184.27	241.73	58435.00	317.12
MORADO	168	184.27	-16.27	264.60	1.44
NEGRO	91	184.27	-93.27	8698.67	47.21
ROJO	490	184.27	305.73	93472.87	507.27
ROSA	4	184.27	-180.27	32496.07	176.35
TURQUEZA	15	184.27	-169.27	28651.20	155.49
VERDE	374	184.27	189.73	35998.74	195.36
Totales	2764				2481.88

$$\sum_{i:1}^{15} \frac{(f_{oi} - f_{ei})^2}{f_{ei}} = 2481.88 \sim \mathbf{c}_{a,14}^2$$

Al valor del estadístico de prueba ji cuadrado es $\chi^2_{a,14}=2481.88$ le corresponde un valor $p \approx 0$, que es muy pequeño para poder concluir que no existen diferencias significativas entre el número de vehículos robados de acuerdo a su color. Es decir que el número de robos es dependiente del color del carro.

Figura 3.15: Diagrama circular de la proporción de robos durante el año 1999 de acuerdo al color del vehículo.



Dividamos los colores en tres grupos que acumulen aproximadamente 1/3 del total de robos registrados.

Tablas 3.25: Resumen de la segmentación de sustracciones de acuerdo al color del vehículo tomando como base la proporción de robos acumulada.

Grupo	Riesgo Alto	Riesgo Medio	Riesgo Bajo
Color	Blanco Rojo	Azul Gris Verde	Amarillo Anaranjado Bicolor Café Celeste Crema Morado Negro Rosa Turquesa
Porcentaje acumulado	35	41	25

Efectivamente, de acuerdo al análisis aquí mostrado se podría decir que los colores de mayor riesgo son el blanco y el rojo, pero habría que comparar la proporción de vehículos robados con la de vehículos en circulación de estos dos colores, para poder concluir si en realidad el color influye en el robo, o simplemente tienen una relación directamente proporcional. Lastimosamente no contamos con esa información, por lo que sólo diremos que el 35% de los vehículos robados fueron de color rojo y blanco. Además, contrario a lo que se podría pensar los vehículos de color amarillo, café, crema y negro, no representan un peso alto, sino por el contrario representan a la minoría.

3.1.7 Entorno del Robo.

Se ha definido esta variable como: circunstancia en la que se encontraba el vehículo previo al robo, es decir:

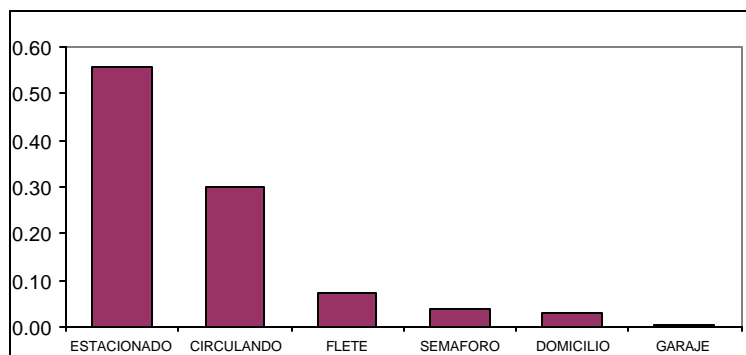
- Si el vehículo estaba estacionado en la acera;
- Si fue sustraído llegando a la casa, saliendo de la casa o en el garaje de la casa;
- Si fue interceptado mientras era conducido por la calle;
- Si el conductor fue atacado esperando mientras la luz roja del semáforo cambiaba a verde; ó
- Si el conductor fue sorprendido por los pasajeros en una carrera de taxi mientras los dirigía a su destino.

El objetivo es identificar bajo que circunstancias suceden la mayor cantidad de robos, claro está que estas modalidades (de la variable) se han obtenido de testimonios, es decir que podemos afirmar que las modalidades aquí descritas constituyen en si una característica del “Modus Operandis” de los delincuentes.

Tabla 3.26: Frecuencias absoluta y relativa del número de sustracciones anuales de acuerdo al entorno

Entorno	Frec. Abs.	Frec. Rel.
ESTACIONADO	1542	0.56
CIRCULANDO	830	0.30
FLETE	199	0.07
SEMAFORO	100	0.04
DOMICILIO	82	0.03
GARAJE	11	0.00

Figura 3.16: Diagrama de Pareto de la proporción de vehículos robados de acuerdo al entorno.



De acuerdo a la figura aquí expuesta la mayor proporción de vehículos robados se agrupa en la modalidad “Estacionado”, de acuerdo a las declaraciones de los denunciantes, las sustracciones ocurrieron mientras el conductor se detuvo a comprar algo en una farmacia o en una despensa o farmacia, o mientras comía en un restaurante, mas adelante analizaremos si en este tipo de sustracciones se utilizó la violencia, ya que es posible que al delincuente no le atemorice el ser visto o no.

Una vez más hagamos la prueba de diferencia de medias para confirmar lo dicho

Tabla 3.27:

Frecuencias observada y esperada y valor del estadístico χ^2 del número de sustracciones de vehículos en Guayaquil en el año 1999 contadas por cada una de las modalidades de la variable entorno.

Entorno	Frec. Observada f_o	Frec. Esperada f_e	$f_o - f_e$	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
CIRCULANDO	830	460.67	369.33	136407.11	296.11
DOMICILIO	82	460.67	-378.67	143388.44	311.26
ESTACIONADO	1542	460.67	1081.33	1169281.78	2538.24
FLETE	199	460.67	-261.67	68469.44	148.63
GARAJE	11	460.67	-449.67	202200.11	438.93
SEMAFORO	100	460.67	-360.67	130080.44	282.37
Totales	2764				4015.54

$$\sum_{i:1}^6 \frac{(f_{oi} - f_{ei})^2}{f_{ei}} = 4015.54 \sim \mathbf{c^2}_{a,5}$$

Efectivamente, el valor p correspondiente al estadístico Ji cuadrado aquí obtenido es demasiado pequeño para concluir igualdad de medias.

Ahora formemos grupos que indiquen el tipo de riesgo de cada una de las modalidades de la variable entorno.

Tablas 3.28: Resumen de la segmentación de sustracciones de acuerdo al entorno del vehículo robado tomando como base la proporción de robos acumulada.

Grupo	Riesgo Alto	Riesgo Medio	Riesgo Bajo
Entorno	Estacionado	Circulando	Domicilio Flete Garaje Semáforo
Porcentaje acumulado	56	30	14

La modalidad circulando también representa una alta ponderación, ya que agrupa 30% de los casos por si sola. Estos casos ocurren cuando el conductor se encuentra dentro del vehículo con el motor andando. De acuerdo a los testimonios el conductor se puede encontrar con una persona apuntando un arma de frente en el camino; o que otro carro se atravesase por el frente y obligue al conductor a bajar de su vehículo, todo esto sucede lejos de la casa del conductor (por lo menos a dos cuadras).

En los demás casos que representan una minoría el conductor fue obligado a bajar del vehículo cerca de su casa (máximo a dos cuadras), saliendo o llegando a ella; o el vehículo fue sustraído cuando se encontraba estacionado dentro del garaje de la casa o en el garaje. Esto hace suponer que existe cierto tipo de seguridad (o menor riesgo

de robo) mientras el vehículo se encuentra dentro del entorno del hogar del propietario. También dentro de estos casos (de menor ponderación) se encuentran aquellos en los que el conductor decidió hacer un flete o una carrera de taxi y fue atacado por los pasajeros; y aquellos casos en los que el conductor fue sorprendido esperando que el semáforo cambie de rojo a verde.

Para calcular el promedio de robos diarios cuando el vehículo está estacionado obtengamos la razón entre 1542 y 366 o sea $1542 \div 366 = 4.21$, es decir que de los 8 robos que suceden (en promedio) en el día, aproximadamente cuatro suceden cuando el vehículo está estacionado (lejos de casa) y dos ($830 \div 366 = 2.26$) cuando el vehículo está en movimiento y los dos restantes se dividen en el “grupo de riesgo bajo” que se compone por las modalidades domicilio, flete, garaje y semáforo.

3.1.8 Número de Personas que Realizan el Robo (num).

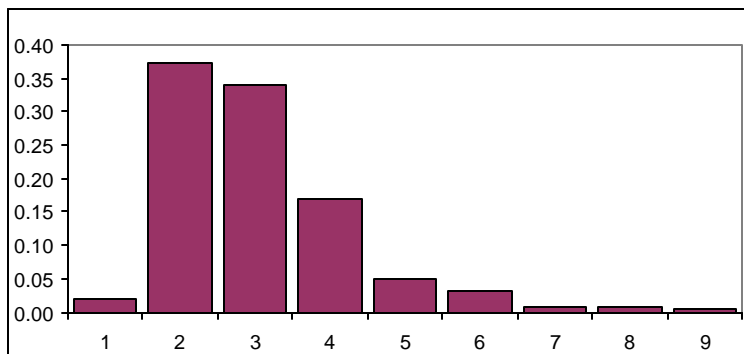
Esta variable puede tomar valores desde uno hasta nueve, donde nueve equivale a 9 o mas personas (Ver sección 1.5 literal k). Se esperaría que este número sea mayor que uno en la mayoría de los casos, ya que estamos asumiendo que el delito se sustraer vehículos es parte del crimen organizado; y que tampoco sea muy grande ya que (por sentido común) no sería eficiente. Pero a pesar de esto hemos encontrado algunos valores menores que 2 y mayores a 6, incluso se presentaron casos en los que actuaron 15 delincuentes.

A continuación se encuentra la tabla que muestra las frecuencias absoluta y relativa de la variable Num.

Tabla 3.29: Frecuencias absoluta y relativa del numero de sustracciones anuales de acuerdo al número de delincuentes.

Num	Frec. Abs.	Frec. Rel.
1	52	0.02
2	1028	0.37
3	941	0.34
4	469	0.17
5	135	0.05
6	89	0.03
7	18	0.01
8	20	0.01
9	12	0.00

Figura 3.17: Histograma de frecuencias de la proporción de vehículos robados de acuerdo al número de delincuentes.



Vemos que nuestra hipótesis concuerda con los resultados, ya que los datos se agrupan entre 2 y 6, pero para ser más precisos los datos se estarían agrupando entre 2 y 3. Analicemos los resultados de la prueba Ji-Cuadrado de la tabla 3.30.

Tabla 3.30:

Frecuencias observada y esperada y valor del estadístico χ^2 del número de sustracciones de vehículos en Guayaquil en el año 1999 contadas por cada una de las modalidades de la variable número de delincuentes.

Num.	Frec. Observada f_o	Frec. Esperada f_e	$f_o - f_e$	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
1	52	345.50	-293.50	86142.25	249.33
2	1028	345.50	682.50	465806.25	1348.21
3	941	345.50	595.50	354620.25	1026.40
4	469	345.50	123.50	15252.25	44.15
5	135	345.50	-210.50	44310.25	128.25
6	89	345.50	-256.50	65792.25	190.43
7	18	345.50	-327.50	107256.25	310.44
8	20	345.50	-325.50	105950.25	306.66
9	12	345.50	-333.50	111222.25	321.92
Totales	2764				3925.77

$$\sum_{i:1}^9 \frac{(f_{oi} - f_{ei})^2}{f_{ei}} = 3925.77 \sim c^2_{a,8}$$

El estadístico $c^2_{a,8} = 3925.77$ implica un valor p cercano a cero. Podemos afirmar entonces que, como ya sospechábamos, existen diferencias significativas entre el número de personas que roban un vehículo. Observemos como queda la segmentación de esta variable de acuerdo a su ponderación.

Tablas 3.31: Resumen de la segmentación de sustracciones de acuerdo al número de delincuentes tomando como base la proporción de robos acumulada.

Grupo	Mayoría 1	Mayoría 2	Minoría
Número	2	3	1, 4 ó mas
Porcentaje acumulado	37	34	29

De acuerdo a la tabla 3.31 en la mayoría de las sustracciones actúan 2 o 3 personas, ya que las dos cantidades agrupan el 71% de los casos, esto hace pensar que para el crimen organizado es muy fácil cometer el delito de sustraer un vehículo.

3.1.9 Acción Delincuencial para Realizar el Robo.

Siguiendo con el análisis del “Modus Operandis” de la delincuencia. Nos encontramos con esta variable “Acción” que se divide en dos modalidades:

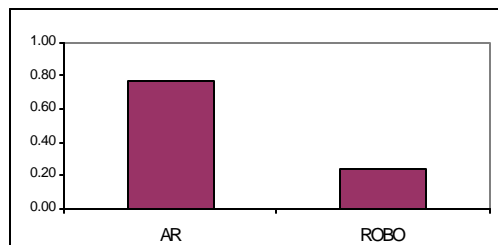
- Robo con Asalto (AR), y
- Robo sin Asalto (Robo)

Si la mayoría de sustracciones se dan cuando el vehículo está estacionado, se puede esperar que la modalidad robo sin asalto suceda en mayor proporción que la modalidad robo con asalto; para probarlo o contradecirlo primero dibujemos el diagrama de Pareto de esta variable.

Tabla 3.32: Frecuencias absoluta y relativa del número de sustracciones anuales de acuerdo al tipo de acción del robo.

Acción	Frec. Abs.	Frec. Rel.
AR	2121	0.77
ROBO	643	0.23

Figura 3.18: Diagrama de Pareto de la proporción de vehículos robados de acuerdo al tipo de acción del robo.



Contrario a lo que se pensaba el número de robos con asalto es mucho mayor al de robos sin asalto, esto si tiene explicación lógica, ya que es posible que la mayoría de las sustracciones ocurran en presencia de la persona que conduce el vehículo, ya que en el caso de que el vehículo esté estacionado, el conductor pudo haber estado cerca del lugar, de tal manera que se percató de lo que estaba sucediendo, e intentó detener a los delincuentes. También de acuerdo a declaraciones, cuando el conductor ingresa al lugar de destino, los delincuentes ingresan detrás de el, lo apuntan con un revolver y lo obligan a entregarle las llaves del carro.

En resumen, la mayoría de las sustracciones se dan en presencia del conductor del vehículo, o de alguna persona que tenga interés en que el vehículo no sea sustraído.

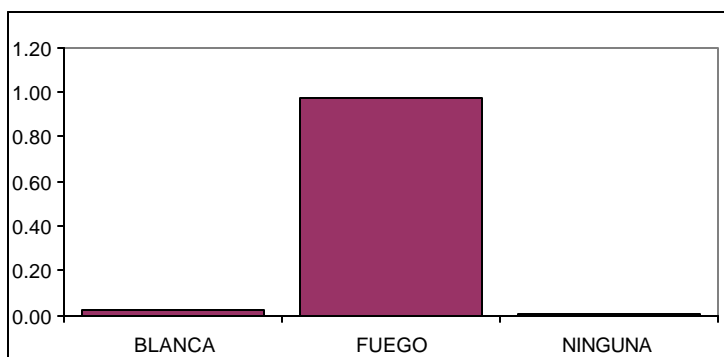
3.1.9 Armas.

Ya que sabemos que la mayoría de las sustracciones de vehículos suceden por medio de la fuerza, nos interesa saber que tipo de arma utilizaron, si usaron arma blanca, de fuego, o ningún arma en el robo ya que así tendremos una idea del riesgo que corre una persona de ser agredida de gravedad en caso de que intenten robar su vehículo.

Tabla 3.33: Frecuencias absoluta y relativa del número de sustracciones anuales de acuerdo al tipo de arma usada.

Num	Frec. Abs.	Frec. Rel.
BLANCA	42	0.02
FUEGO	2075	0.98
NINGUNA	4	0.00

Figura 3.19: Histograma de frecuencias de la proporción de vehículos robados de acuerdo al tipo de arma usada.



Definitivamente, si nos encontráramos en el caso de ser robados, el riesgo de ser agredido seriamente es alto, ya que casi en el 100% de los asaltos los delincuentes están provistos de un arma (o armas) de fuego.

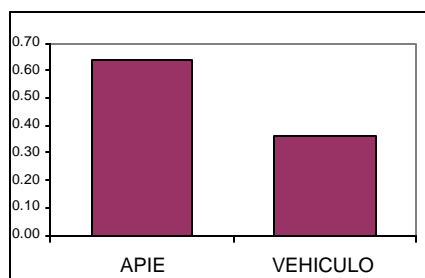
3.1.11 Locomoción.

Mediante esta variable queremos determinar si se pudo identificar algún vehículo como medio de transporte de los delincuentes que realizaron el robo; esto es con el objeto de seguir descubriendo características acerca del Modus Operandis de este tipo de delincuentes. Observemos el diagrama de Pareto de la distribución de las sustracciones de vehículos de acuerdo al tipo de locomoción de los sustractores.

Tabla 3.34: Frecuencias absoluta y relativa del número de sustracciones anuales de acuerdo al tipo locomoción de los sustractores.

Locomoción	Frec. Abs.	Frec. Rel.
APIE	1771	0.64
VEHICULO	993	0.36

Figura 3.20: Diagrama de Pareto de la proporción de vehículos robados de acuerdo al tipo de locomoción de los sustractores.



Vemos que en la mayoría de los casos ningún testigo ha observado un vehículo (a más del sustraído) que se pueda

relacionar con el robo. Asumiendo que los datos son confiables podríamos pensar que:

- ✓ Mientras el vehículo estaba estacionado, los sustractores se acercaron a pie y sustrajeron el vehículo.
- ✓ Los sustractores conocían el recorrido del vehículo a sustraer, lo esperaron en un lugar determinado y cuando el vehículo se encontró en ese lugar, lo sustrajeron.
- ✓ Debido a la tensión del momento, la víctima no se percató si en realidad había un vehículo relacionado con el robo.

3.1.11 Zona.

Para analizar los sectores potenciales de robo hemos dividido la ciudad en 34 zonas y para eliminar el sesgo hemos procurado dividir la ciudad en sectores que tengan el mismo tamaño y que se puedan identificar como zonas claramente distinguibles de las demás. Además se ha procurado nombrar los sectores de acuerdo a su ubicación; es decir a las ciudadelas del centro se las ha nombrado con la letra C, a las del norte con la letra N, a las del sur con la letra S, a las del oeste con la letra O, y a las del este con la letra E, además debido a que la cantidad de robos ocurridos en la Av. Perimetral es representativa, se ha optado por nombrar una zona con la letra P. A continuación se describen gráficamente las delimitaciones de cada una de las zonas en las que se dividió la ciudad.

Figura 3.21: Zona S1



Figura 3.22: Zona S2

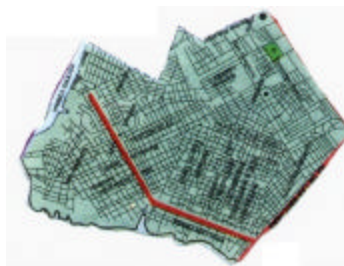


Figura 3.23: Zona S3



Figura 3.24: Zona S4



Figura 3.25: Zona S5



Figura 3.26: Zona S6



Figura 3.27: Zona C1



Figura 3.28: Zona C2



Figura 3.29: Zona C3

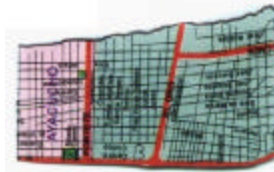


Figura 3.30: Zona C4



Figura 3.31: Zona C5



Figura 3.32: Zona O1



Figura 3.33: Zona O2



Figura 3.34: Zona O3



Figura 3.35: Zona O4

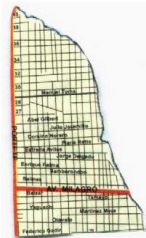


Figura 3.36: Zona O5



Figura 3.37: Zona N1



Figura 3.38: Zona N2



Figura 3.40: Zona N4



Figura 3.42: Zona N6



Figura 3.44: Zona N8



Figura 3.39: Zona N3



Figura 3.41: Zona N5



Figura 3.43: Zona N7



Figura 3.45: Zona N9



Figura 3.46: Zona N10**Figura 3.47:** Zona N11

N12: Comprende el sector donde se ubica el Quinto Guayas

N13: Comprende la urbanización Monte Bello

N14: Comprende la urbanización Bastión Popular

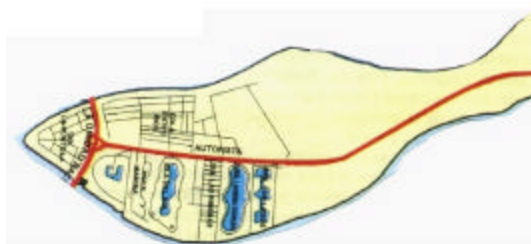
N15: Comprende la ciudadela Los Vergeles

N16: Comprende el recinto Pascuales

N17: Comprende la ciudadela Las Orquídeas

N0: Comprende la Urb. Puerto Azul

P: Comprende la vía Perimetral

Figura 3.48: Zona E1

A continuación se presenta el mapa completo que une cada uno de los sectores anteriormente descritos.

Figura 3.49: Plano Urbano de Guayaquil⁴. Parte I.

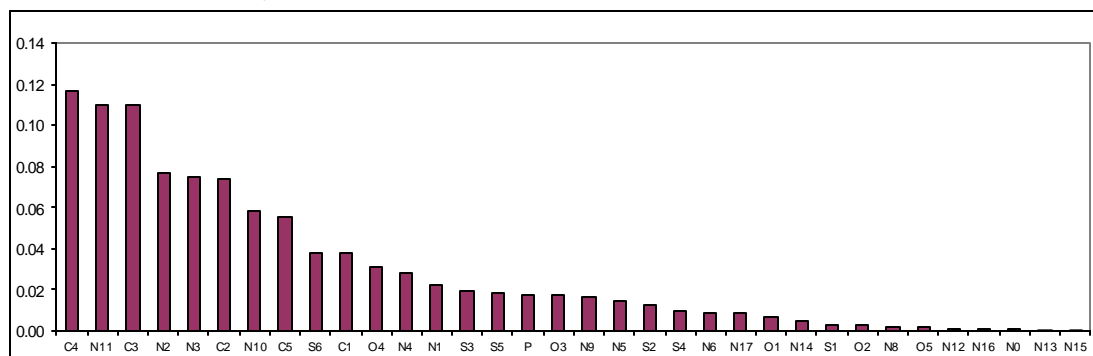
⁴ Este mapa es una actualización parcial correspondiente al año de 1999 y es propiedad de “3E” Ediciones Educativas Ecuatorianas-Guayaquil-Ecuador; lámina #519

Figura 3.49: Plano Urbano de Guayaquil: Parte II.

Tabla 3.35: Frecuencias absoluta y relativa del total de sustracciones de vehículos en Guayaquil en el año 1999 contadas por zona del robo.

Zona	Frec. Abs.	Frec. Rel.	Zona	Frec. Abs.	Frec. Rel.
C1	104	0.04	N17	23	0.01
C2	205	0.07	N2	212	0.08
C3	303	0.11	N3	208	0.08
C4	322	0.12	N4	78	0.03
C5	154	0.06	N5	40	0.01
N0	2	0.00	N6	25	0.01
N1	62	0.02	N8	6	0.00
N10	160	0.06	N9	45	0.02
N11	305	0.11	O1	19	0.01
N12	3	0.00	O2	7	0.00
N13	1	0.00	O3	47	0.02
N14	12	0.00	O4	85	0.03
N15	1	0.00	O5	6	0.00
N16	3	0.00	P	48	0.02

Figura 3.50: Diagrama de Pareto del total de sustracciones de vehículos en Guayaquil en el año 1999 contadas por zona del robo.



En la gráfica observamos que las zonas de más peligro son C4, N11 y O3 ya que se distinguen claramente de las demás y cada una acumula del 10 al 12% de los robos anuales, mientras que existe una gran proporción de zonas que acumulan menos del 4% de los robos cada una. Antes de

obtener conclusiones relevantes examinemos el valor del estadístico de prueba Ji-Cuadrado.

Tabla 3.36:

Frecuencias observada y esperada y valor del estadístico χ^2 del número de sustracciones de vehículos en Guayaquil en el año 1999 contadas por zona del robo.

Zona	Frec. Observada f_o	Frec. Esperada f_e	$f_o - f_e$	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
C1	104	81.29	22.71	515.56	6.34
C2	205	81.29	123.71	15303.15	188.24
C3	303	81.29	221.71	49153.50	604.64
C4	322	81.29	240.71	57939.32	712.71
C5	154	81.29	72.71	5286.15	65.02
N0	2	81.29	-79.29	6287.56	77.34
N1	62	81.29	-19.29	372.26	4.58
N10	160	81.29	78.71	6194.62	76.20
N11	305	81.29	223.71	50044.32	615.60
N12	3	81.29	-78.29	6129.97	75.40
N13	1	81.29	-80.29	6447.15	79.31
N14	12	81.29	-69.29	4801.67	59.07
N15	1	81.29	-80.29	6447.15	79.31
N16	3	81.29	-78.29	6129.97	75.40
N17	23	81.29	-58.29	3398.20	41.80
N2	212	81.29	130.71	17084.03	210.15
N3	208	81.29	126.71	16054.38	197.49
N4	78	81.29	-3.29	10.85	0.13
N5	40	81.29	-41.29	1705.20	20.98
N6	25	81.29	-56.29	3169.03	38.98
N8	6	81.29	-75.29	5669.20	69.74
N9	45	81.29	-36.29	1317.26	16.20
O1	19	81.29	-62.29	3880.56	47.73
O2	7	81.29	-74.29	5519.62	67.90
O3	47	81.29	-34.29	1176.09	14.47
O4	85	81.29	3.71	13.73	0.17
O5	6	81.29	-75.29	5669.20	69.74
P	48	81.29	-33.29	1108.50	13.64
S1	9	81.29	-72.29	5226.44	64.29
S2	34	81.29	-47.29	2236.73	27.51
S3	53	81.29	-28.29	800.56	9.85
S4	27	81.29	-54.29	2947.85	36.26
S5	50	81.29	-31.29	979.32	12.05
S6	105	81.29	23.71	561.97	6.91
Totales	2764				3685.15

$$\sum_{i:1}^{34} \frac{(f_{oi} - f_{ei})^2}{f_{ei}} = 3685.15 \sim \chi^2_{a,33}$$

El valor p correspondiente al valor de la variable aleatoria Ji-Cuadrado evaluado en 3685.12 es muy pequeño; decimos entonces que existe suficiente evidencia estadística para rechazar la hipótesis nula de igualdad de medias.

En la tabla que se muestra a continuación resumimos 3 subgrupos de marcas de vehículos de acuerdo a la proporción de robos que se han dado.

Tabla 3.37: Segmentación de sustracciones de acuerdo a la zona del robo del vehículo tomando como base la proporción de robos acumulada.

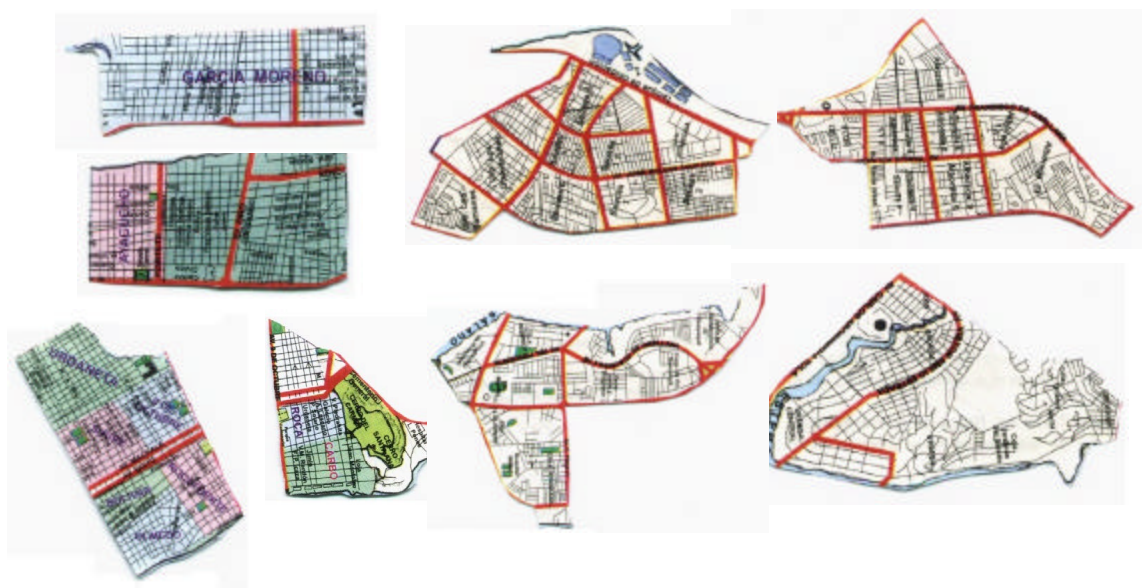
Tipo de Riesgo	Zonas	%
Muy alto	C3, C4, N11	34
Medio	C2, C5, N2, N3, N10	34
Bajo	C1, N0, N1, N4, N5, N6, N7, N8, N9, N12, N13, N14, N15, N16, N17, O1, O2, O3, O4, O5, P, S1, S2, S3, S4, S5, S6	32

Vemos aquí que dentro de las zonas de mayor riesgo se encuentran sectores ubicados en el centro (C3 y C4) y en el norte de la ciudad (N11); y lo mismo sucede con las zonas de riesgo medio (C2, C5, N2, N3 y N10). Si observamos con detenimiento el mapa de la figura 3.50 y de acuerdo a los resultados aquí obtenidos se puede notar claramente que la

zona de alto riesgo de robo es la que rodea al núcleo (centro) de la ciudad que se puede resumir en 3 sectores (Ver figura 3.52):

- a. Desde la Av. José Vicente Trujillo, hasta la calle Gómez Rendón y desde el río Guayas hasta la Av. Quito
- b. Las parroquias Urdaneta, García Moreno, Sucre, Nueve de Octubre, Ayacucho, Bolívar, Olmedo, Rocafuerte, Roca y Carbo
- c. Cdlas Urdesa; Kennedy; Miraflores; Urdenor; Atarazana; Saucos (del 1 al 9) y Alborada (del 1 al 12, excepto "10"), Urb. Los Samanes y Urb. Los Guayacanes

Figura 3.52: Zona de alto riesgo de robo



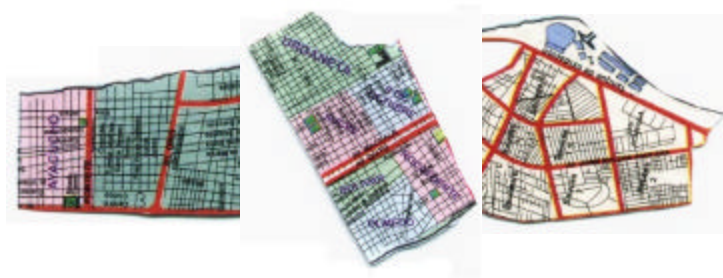
Siendo los de riesgo potencial los siguientes:

Sector a (Ver página 123).

Parroquias Ayacucho, Urdaneta, Sucre, 9 de Octubre,
Bolívar, Olmedo y Rocafuerte

Ciudadelas Sauces (del 2 al 7), Urb. Los Samanes y Urb.
Los Guayacanes

Figura 3.51: Zona de Riesgo Potencial de Robo



Para aclarar las interrogantes generadas debemos realizar algunos análisis estadísticos más elaborados como el análisis bivariado, y de ser necesario el multivariado.

Recojamos las interrogantes planteadas y presentémoslas a continuación:

¿Existe alguna relación entre el día y la hora?

¿Existe alguna relación entre la marca y el tipo?

¿Existe alguna relación entre la marca y la zona?

¿Existe alguna relación entre el número de delincuentes y la zona del robo?

¿Existe alguna relación entre el medio de locomoción y la zona?

¿Existe alguna relación entre el entorno y la acción?

¿Existe alguna relación entre la hora y la zona?

¿Existe alguna relación entre la marca y el color del vehículo?

¿Se podría afirmar que el día, la hora y la zona tienen algo que ver?

3.2 Análisis Bivariado

Luego de realizar el análisis univariado hemos encontrado necesario ampliar nuestro estudio y continuar con el análisis bivariado con el objeto de aclarar las dudas generadas en el estudio antes realizado.

Aquí nos disponemos a contrastar parejas de variables de tal manera que podamos concluir si los resultados provenientes de una variable tienen relación con los de otra. En otras palabras, queremos saber si los cambios en las modalidades de una variable responden a cambios en las modalidades de otra; por ejemplo, si concluyéramos que el día del robo se relaciona con la hora del robo podremos identificar con mayor claridad los sectores de riesgo y no riesgo, y podremos obtener más conclusiones relevantes acerca del Modus Operandis de los roba-carros.

En este apartado haremos uso de las tablas de contingencia descritas en la sección 2.6.

3.2.1 Día Vs. Hora

Para esta prueba vamos a contrastar las siguientes hipótesis:

H_0 : El día del robo no se relaciona con la hora

Vs.

H_1 : No es verdad H_0 .

Para llenar la tabla $r \times c$ pondremos en la celda $X_{i,j}$ el valor correspondiente a la suma del número de vehículos sustraídos el día i y a la hora j . Observemos como queda la tabla para este contraste.

Tabla 3.38: Tabla de contingencia Día Vs. Hora

		Día							Total
		Lun	Mar	Mier	Jue	Vier	Sab	Dom	
Hora	H00	6	9	5	6	10	12	14	62
	H01	6	7	4	7	10	9	3	46
	H02	7	5	8	7	15	13	4	59
	H03	9	4	4	9	7	14	12	59
	H04	7	6	9	14	7	7	10	60
	H05	7	7	7	9	6	10	5	51
	H06	14	7	10	9	11	6	5	62
	H07	23	21	24	19	16	10	7	120
	H08	12	13	8	20	19	7	7	86
	H09	9	9	11	12	11	11	7	70
	H10	12	17	5	9	9	18	9	79
	H11	12	12	16	16	16	18	13	103
	H12	10	8	11	20	19	13	11	92
	H13	12	13	12	8	8	7	11	71
	H14	9	13	9	11	11	7	11	71
	H15	5	10	15	12	10	6	7	65
	H16	10	8	2	8	9	11	9	57
	H17	9	7	12	9	10	8	12	67
	H18	11	15	15	17	21	8	12	99
	H19	50	45	45	42	44	26	29	281
	H20	72	83	74	66	53	44	65	457
	H21	37	40	49	56	45	37	49	313
	H22	26	22	23	35	48	32	25	211
	H23	17	13	20	17	18	29	9	123
Total	392	394	398	438	433	363	346	2764	

Tabla 3.39: Valores $e_{i,j}$ obtenida a partir de la tabla 3.38

$e_{i,j}$	Lun	Mar	Mier	Jue	Vier	Sab	Dom
H00	8.793	8.84	8.93	9.82	9.71	8.14	7.76
H01	6.524	6.56	6.62	7.29	7.21	6.04	5.76
H02	8.368	8.41	8.5	9.35	9.24	7.75	7.39
H03	8.368	8.41	8.5	9.35	9.24	7.75	7.39
H04	8.509	8.55	8.64	9.51	9.4	7.88	7.51
H05	7.233	7.27	7.34	8.08	7.99	6.7	6.38
H06	8.793	8.84	8.93	9.82	9.71	8.14	7.76
H07	17.02	17.1	17.3	19	18.8	15.8	15
H08	12.2	12.3	12.4	13.6	13.5	11.3	10.8
H09	9.928	9.98	10.1	11.1	11	9.19	8.76
H10	11.2	11.3	11.4	12.5	12.4	10.4	9.89
H11	14.61	14.7	14.8	16.3	16.1	13.5	12.9
H12	13.05	13.1	13.2	14.6	14.4	12.1	11.5
H13	10.07	10.1	10.2	11.3	11.1	9.32	8.89
H14	10.07	10.1	10.2	11.3	11.1	9.32	8.89
H15	9.219	9.27	9.36	10.3	10.2	8.54	8.14
H16	8.084	8.13	8.21	9.03	8.93	7.49	7.14
H17	9.502	9.55	9.65	10.6	10.5	8.8	8.39
H18	14.04	14.1	14.3	15.7	15.5	13	12.4
H19	39.85	40.1	40.5	44.5	44	36.9	35.2
H20	64.81	65.1	65.8	72.4	71.6	60	57.2
H21	44.39	44.6	45.1	49.6	49	41.1	39.2
H22	29.92	30.1	30.4	33.4	33.1	27.7	26.4
H23	17.44	17.5	17.7	19.5	19.3	16.2	15.4

Tabla 3.40: Valores $f_{i,j}$ obtenida a partir de la tabla 3.38

$f_{i,j}$	Lun	Mar	Mier	Jue	Vier	Sab	Dom
H00	0.89	0	1.7	1.49	0.01	1.83	5.01
H01	0.04	0.03	1	0.01	1.08	1.45	1.32
H02	0.22	1.38	0	0.59	3.59	3.56	1.55
H03	0.05	2.31	2.4	0.01	0.54	5.04	2.88
H04	0.27	0.76	0	2.12	0.61	0.1	0.82
H05	0.01	0.01	0	0.1	0.5	1.63	0.3
H06	3.08	0.38	0.1	0.07	0.17	0.56	0.98
H07	2.1	0.89	2.6	0	0.42	2.11	4.28
H08	0	0.04	1.6	2.98	2.27	1.63	1.32
H09	0.09	0.1	0.1	0.07	0	0.36	0.35
H10	0.06	2.92	3.6	0.99	0.92	5.6	0.08
H11	0.47	0.49	0.1	0.01	0	1.48	0
H12	0.71	1.99	0.4	2.02	1.46	0.07	0.02
H13	0.37	0.82	0.3	0.94	0.88	0.58	0.5
H14	0.11	0.82	0.1	0.01	0	0.58	0.5
H15	1.93	0.06	3.4	0.28	0	0.75	0.16
H16	0.45	0	4.7	0.12	0	1.65	0.49
H17	0.03	0.68	0.6	0.25	0.02	0.07	1.56
H18	0.66	0.06	0	0.11	1.94	1.92	0.01

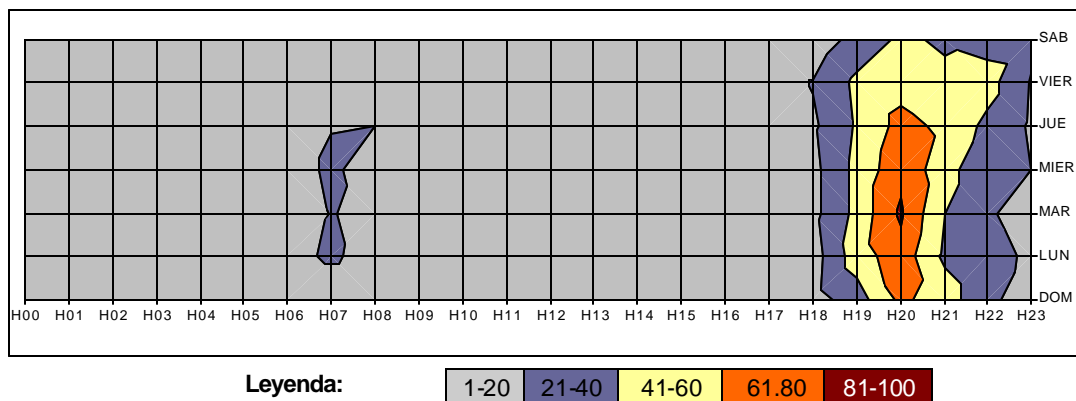
H19	2.58	0.61	0.5	0.14	0	3.22	1.08
H20	0.8	4.89	1	0.57	4.83	4.28	1.06
H21	1.23	0.48	0.3	0.83	0.33	0.41	2.46
H22	0.51	2.17	1.8	0.07	6.76	0.66	0.08
H23	0.01	1.17	0.3	0.32	0.08	10.2	2.66

A partir de las tablas 3.39 y 3.40 obtenemos los valores necesarios para evaluar en la ecuación siguiente, para obtener el valor del estadístico de prueba Ji-cuadrado χ^2 .

$$\chi^2 = \sum_{i=1}^{24} \sum_{j=1}^7 \frac{(f_{o_{i,j}} - f_{e_{i,j}})^2}{f_{e_{i,j}}} = 186.28 \quad \chi^2_{138}$$

El valor p correspondiente a esta variable es 0.004, decimos entonces que existe suficiente evidencia estadística para rechazar la hipótesis de independencia, es decir que existe una relación entre la hora y el día del robo. Analicemos gráficamente esta teoría dibujando las superficies de nivel del cruce de las modalidades de las variables Día y Hora.

Figura 3.54: Superficies de nivel del cruce de las variables día Vs. Hora.



De acuerdo a la figura 3.54 la zona de mayor riesgo ocurre el martes a las 20H00 (zona roja); seguido muy de cerca por la zona naranja que ocurre aprox. de jueves a domingo desde las 19H30 hasta las 20H30, luego observamos a la zona amarilla (riesgo medio), que encierra la unión de las zonas viernes a sábado de 18H00 a 22H00; jueves a domingo de 18H30 a 19H30 y de 20H30 a 21H30. Las zonas restantes representan las zonas de riesgo bajo y muy bajo. Si generalizamos podemos decir que debemos tener mayor precaución de lunes a domingo desde las 18H30 hasta las 21H30 debiendo estar muy alerta los martes a las 20H00.

3.2.2 Marca Vs. Tipo

Con esta prueba queremos descubrir como se distribuye el número de sustracciones respecto a la marca y el tipo de carro, es decir, si existe alguna relación, y de existir dicha relación definir cual es. El contraste es el siguiente:

H_0 : La marca del vehículo no se relaciona con el tipo de

vehículo

Vs.

H_1 : No es verdad H_0 .

De aquí en adelante presentaremos únicamente la tabla de contingencia del contraste y automáticamente calcularemos el valor del estadístico de prueba⁵⁵, y de existir alguna relación entre las variables pasaremos a graficar las superficies de nivel.

Tabla 3.38: Tabla de contingencia Marca Vs. Tipo

		Tipo				Total
		Auto	Campero	Cmta	Otros	
Marca	ALFAROM	1	0	0	0	1
	AUSTIN	4	0	0	0	4
	BMW	5	0	0	0	5
	CHEVRLT	416	169	327	4	916
	CHRYSLER	1	1	0	0	2
	CITROEN	1	0	0	0	1
	DACIA	1	0	0	0	1

⁵⁵ las tablas $e_{i,j}$ y $f_{i,j}$ se pueden encontrar en el disquete adjunto a la tesis

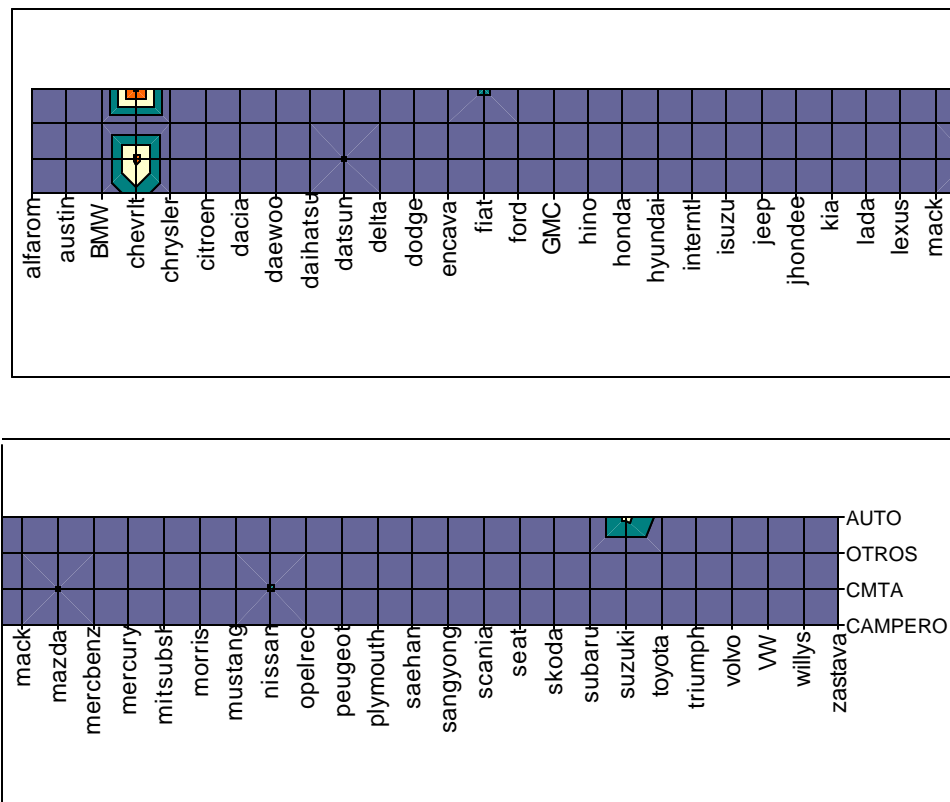
DAEWOO	48	0	0	0	48
DAIHATSU	4	4	3	8	19
DATSUN	28	0	105	1	134
DELTA	0	0	0	2	2
DODGE	0	1	2	0	3
ENCAVA	0	0	0	1	1
FIAT	118	0	5	0	123
FORD	29	13	70	9	121
GMC	0	0	1	1	2
HINO	0	0	0	5	5
HONDA	10	2	0	0	12
HYUNDAI	82	0	0	0	82
INTERNTL	0	0	0	3	3
ISUZU	3	1	22	3	29
JEEP	0	19	0	0	19
JHONDEE	0	0	0	1	1
KIA	3	3	0	7	13
LADA	74	0	0	0	74
LEXUS	1	0	0	0	1
MACK	0	0	0	2	2
MAZDA	77	0	107	0	184
MERCBENZ	4	0	0	5	9
MERCURY	1	0	0	0	1
MITSUBSH	54	50	41	8	153
MORRIS	2	0	0	0	2
MUSTANG	1	0	0	0	1
NISSAN	88	11	108	1	208
OPELREC	1	0	0	0	1
PEUGEOT	5	0	0	0	5
PLYMOUTH	1	0	0	0	1
SAEHAN	0	0	2	0	2
SANGYONG	0	1	0	0	1
SCANIA	0	0	0	2	2
SEAT	2	0	0	0	2
SKODA	48	0	17	0	65
SUBARU	5	0	0	0	5
SUZUKI	229	0	1	0	230
TOYOTA	68	25	91	2	186
TRIUMPH	1	0	0	0	1
VOLVO	3	0	0	0	3
VW	64	0	5	4	73
WILLYS	0	1	1	0	2
ZASTAVA	3	0	0	0	3
Total	1486	301	908	69	2764

$$X^{2*} = \sum_{i=1}^{50} \sum_{j=1}^4 \frac{(f_{o_{i,j}} - f_{e_{i,j}})^2}{f_{e_{i,j}}} = 2262.875 \quad \chi^2_{147}$$

De acuerdo al valor del estadístico $\chi^2=2262.875$ el valor p correspondiente es 0.000 que demuestra suficiente evidencia estadística para rechazar la hipótesis de independencia, es decir existe una relación entre el tipo y la marca del carro.

A continuación las curvas de nivel que describen la relación entre las variables en estudio.

Figura 3.55: Superficies de nivel del cruce de las variables Marca Vs. Tipo.



Vemos aquí que dentro de los casos más comunes se encuentran los autos y camionetas Chevrolet, los autos suzuki,

las camionetas Nissan, las camionetas Mazda y los autos Fiat; mientras que las demás marcas y tipos de vehículos tienen muy poca representatividad en los robos.

3.2.3 Marca Vs. Zona

Vamos a probar si existe relación entre la marca del vehículo y la zona del robo, para determinar si se puede zonas de riesgo para cada marca. El contraste de hipótesis se define como sigue:

H_0 : La marca del vehículo no se relaciona con la zona del robo

Vs.

H_1 : No es verdad H_0 .

Observemos los resultados obtenidos a partir de la tabla de contingencia que se muestra a continuación.

Tabla 3.38: Tabla de contingencia Día Vs. Hora

		Zona					
		C1	C2	C3	C4	C5	N0
Marca	ALFAROM	0	0	0	0	0	0
	AUSTIN	0	1	0	0	1	0
	BMW	0	0	1	0	0	0
	CHEVRLT	32	59	107	122	46	0
	CHRYSLER	0	1	0	0	0	0
	CITROEN	0	0	0	0	0	0
	DACIA	0	0	0	0	0	0
	DAEWOO	0	4	10	2	3	0
	DAIHATSU	3	1	2	1	2	0
	DATSUN	9	9	11	15	11	0
	DELTA	0	0	0	0	0	0
	DODGE	0	0	0	0	0	0
	ENCAVA	0	0	0	0	0	0
	FIAT	4	14	16	22	5	0
	FORD	9	9	10	14	7	0
	GMC	0	0	0	2	0	0
	HINO	0	0	1	1	0	0

HONDA	1	1	3	1	2	0
HYUNDAI	2	9	6	5	4	0
INTERNTL	0	0	0	0	0	0
ISUZU	3	1	3	3	1	0
JEEP	1	1	3	1	2	0
JHONDEE	0	0	1	0	0	0
KIA	1	1	2	2	1	0
LADA	5	5	4	4	3	0
LEXUS	0	0	1	0	0	0
MACK	1	0	0	0	0	0
MAZDA	8	11	26	15	8	0
MERCBENZ	1	0	0	0	1	0
MERCURY	0	0	0	0	0	0
MITSUBSH	4	11	16	13	7	0
MORRIS	0	0	0	1	0	0
MUSTANG	0	0	0	0	0	0
NISSAN	7	21	21	29	11	0
OPELREC	0	0	0	1	0	0
PEUGEOT	0	0	0	1	0	0
PLYMOUTH	0	0	0	0	0	0
SAEHAN	0	0	0	0	0	0
SANGYONG	0	0	0	0	0	0
SCANIA	0	0	1	0	0	0
SEAT	1	0	0	0	0	0
SKODA	2	8	13	10	4	0
SUBARU	0	0	2	0	0	0
SUZUKI	7	17	18	26	13	0
TOYOTA	3	15	15	16	14	1
TRIUMPH	0	1	0	0	0	0
VOLVO	0	0	0	1	1	0
VW	0	6	9	12	6	1
WILLYS	0	0	0	0	1	0
ZASTAVA	0	0	0	2	0	0
Total	104	206	302	322	154	2

	Zona						
	N1	N10	N11	N12	N13	N14	
ALFAROM	0	0	0	0	0	0	0
AUSTIN	1	0	0	0	0	0	0
BMW	0	1	0	0	0	0	0
CHEVRLT	21	60	116	0	1	4	4
CHRYSLER	0	0	0	0	0	0	0
CITROEN	0	0	0	0	0	0	0
DACIA	0	0	0	0	0	0	0
DAEWOO	0	4	5	0	0	0	0
DAIHATSU	1	0	0	0	0	0	0
DATSUN	2	3	11	1	0	1	1
DELTA	0	0	0	0	0	0	0
DODGE	0	0	1	0	0	0	0
ENCAVA	0	0	0	0	0	0	0

FIAT	4	2	13	0	0	0
FORD	1	8	13	0	0	1
GMC	0	0	0	0	0	0
HINO	0	0	0	0	0	0
HONDA	0	0	0	0	0	0
HYUNDAI	3	5	10	1	0	0
INTERNTL	0	0	0	0	0	0
ISUZU	0	0	2	0	0	0
JEEP	0	0	0	0	0	0
JHONDEE	0	0	0	0	0	0
KIA	0	0	0	0	0	0
LADA	3	8	6	0	0	1
LEXUS	0	0	0	0	0	0
MACK	0	0	0	0	0	0
MAZDA	3	6	19	0	0	1
MERCBENZ	1	2	0	0	0	0
MERCURY	0	0	0	0	0	0
MITSUBSH	2	12	18	0	0	2
MORRIS	0	0	0	0	0	0
MUSTANG	0	1	0	0	0	0
NISSAN	2	14	25	0	0	0
OPELREC	0	0	0	0	0	0
PEUGEOT	0	0	1	0	0	0
PLYMOUTH	0	0	0	0	0	0
SAEHAN	0	0	2	0	0	0
SANGYONG	0	0	0	0	0	0
SCANIA	0	0	0	0	0	0
SEAT	0	0	0	0	0	0
SKODA	1	4	4	0	0	0
SUBARU	0	0	0	1	0	0
SUZUKI	5	12	32	0	0	0
TOYOTA	6	15	22	0	0	2
TRIUMPH	0	0	0	0	0	0
VOLVO	0	0	0	0	0	0
VW	2	4	5	0	0	0
WILLYS	0	0	1	0	0	0
ZASTAVA	1	0	0	0	0	0
Total	59	161	306	3	1	12

	Zona					
	N15	N16	N17	N2	N3	N4
ALFAROM	0	0	0	0	0	0
AUSTIN	0	0	0	0	0	0
BMW	0	0	0	1	1	0
CHEVRLT	0	2	12	66	64	25
CHRYSLER	0	0	0	1	0	0
CITROEN	0	0	0	0	0	0
DACIA	0	0	0	0	0	0
DAEWOO	0	1	0	2	1	3
DAIHATSU	0	0	0	0	2	0
DATSUN	0	0	0	2	5	5
DELTA	0	0	0	0	1	0

DODGE	0	0	0	0	0	0
ENCAVA	0	0	0	0	0	0
FIAT	0	0	0	7	10	4
FORD	1	0	0	16	12	1
GMC	0	0	0	0	0	0
HINO	0	0	0	0	0	0
HONDA	0	0	0	2	2	0
HYUNDAI	0	0	0	5	7	3
INTERNTL	0	0	1	0	0	0
ISUZU	0	0	0	1	2	0
JEEP	0	0	0	7	2	1
JHONDEE	0	0	0	0	0	0
KIA	0	0	0	1	0	0
LADA	0	0	1	5	1	1
LEXUS	0	0	0	0	0	0
MACK	0	0	0	0	0	0
MAZDA	0	0	4	14	11	11
MERCBENZ	0	0	0	1	0	1
MERCURY	0	0	0	1	0	0
MITSUBSH	0	0	0	27	14	3
MORRIS	0	0	0	1	0	0
MUSTANG	0	0	0	0	0	0
NISSAN	0	0	0	10	15	5
OPELREC	0	0	0	0	0	0
PEUGEOT	0	0	0	1	0	0
PLYMOUTH	0	0	0	0	0	0
SAEHAN	0	0	0	0	0	0
SANGYONG	0	0	0	0	0	0
SCANIA	0	0	0	0	0	0
SEAT	0	0	0	0	1	0
SKODA	0	0	0	2	6	2
SUBARU	0	0	0	0	0	1
SUZUKI	0	0	2	19	33	8
TOYOTA	0	0	3	14	8	5
TRIUMPH	0	0	0	0	0	0
VOLVO	0	0	0	0	1	0
VW	0	0	0	7	8	1
WILLYS	0	0	0	0	0	0
ZASTAVA	0	0	0	0	0	0
Total	1	3	23	213	207	80

Marca	Zona					
	N5	N6	N8	N9	O1	O2
ALFAROM	0	0	0	0	0	1
AUSTIN	0	0	0	1	0	0
BMW	1	0	0	0	0	0
CHEVRLT	9	6	1	17	8	1
CHRYSLER	0	0	0	0	0	0
CITROEN	1	0	0	0	0	0
DACIA	0	0	0	0	0	0
DAEWOO	0	0	0	2	0	0
DAIHATSU	0	0	0	0	0	0

DATSUN	0	3	0	8	2	3
DELTA	0	0	0	0	0	0
DODGE	2	0	0	0	0	0
ENCAVA	0	0	0	0	0	0
FIAT	2	1	0	1	0	0
FORD	2	4	1	1	0	0
GMC	0	0	0	0	0	0
HINO	0	1	0	0	0	0
HONDA	0	0	0	0	0	0
HYUNDAI	1	3	0	1	0	0
INTERNTL	0	0	0	1	0	0
ISUZU	0	0	0	1	1	0
JEEP	0	0	0	0	0	0
JHONDEE	0	0	0	0	0	0
KIA	1	0	0	0	0	0
LADA	0	1	1	1	2	0
LEXUS	0	0	0	0	0	0
MACK	0	0	0	0	0	0
MAZDA	3	1	0	2	4	0
MERCBENZ	0	0	0	0	0	0
MERCURY	0	0	0	0	0	0
MITSUBSH	4	0	0	1	0	0
MORRIS	0	0	0	0	0	0
MUSTANG	0	0	0	0	0	0
NISSAN	4	2	2	2	0	0
OPELREC	0	0	0	0	0	0
PEUGEOT	0	0	0	0	0	0
PLYMOUTH	0	0	0	0	0	0
SAEHAN	0	0	0	0	0	0
SANGYONG	1	0	0	0	0	0
SCANIA	1	0	0	0	0	0
SEAT	0	0	0	0	0	0
SKODA	1	0	0	0	0	0
SUBARU	0	0	0	0	0	0
SUZUKI	2	1	0	4	1	1
TOYOTA	3	2	1	1	1	1
TRIUMPH	0	0	0	0	0	0
VOLVO	0	0	0	0	0	0
VW	1	0	0	1	0	0
WILLYS	0	0	0	0	0	0
ZASTAVA	0	0	0	0	0	0
Total	39	25	6	45	19	7

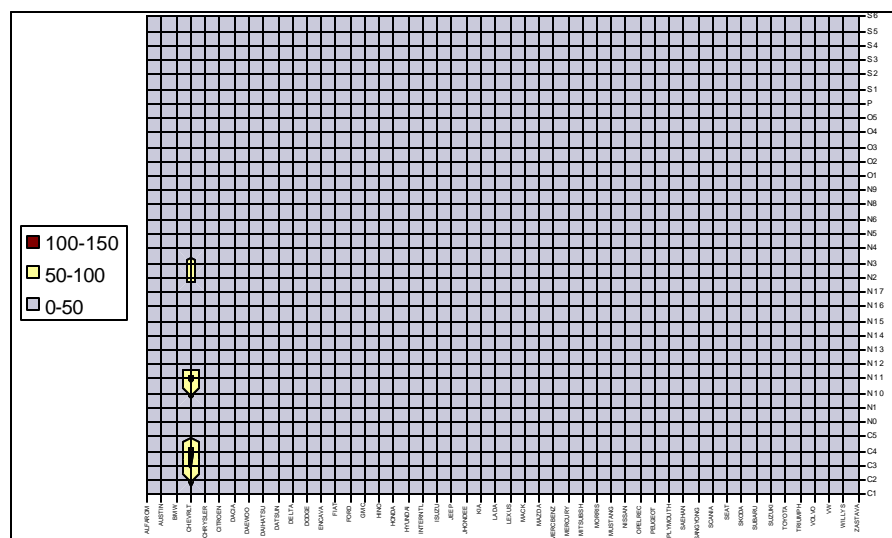
	Zona				Total
	S3	S4	S5	S6	
ALFAROM	0	0	0	0	1
AUSTIN	0	0	0	0	4
BMW	0	0	0	0	5
CHEVRLT	17	5	16	37	916
CHRYSLER	0	0	0	0	2
CITROEN	0	0	0	0	1
DACIA	0	0	0	0	1

DAEWOO	0	1	2	5	48
DAIHATSU	0	0	0	2	19
DATSUN	3	6	3	1	134
DELTA	0	0	1	0	2
DODGE	0	0	0	0	3
ENCAVA	0	0	0	0	1
FIAT	2	1	3	5	123
FORD	2	0	0	1	121
GMC	0	0	0	0	2
HINO	0	0	0	0	5
HONDA	0	0	0	0	12
HYUNDAI	1	0	2	6	82
INTERNTL	0	0	0	0	3
ISUZU	1	0	1	0	29
JEEP	0	0	0	0	19
JHONDEE	0	0	0	0	1
KIA	0	0	0	1	13
LADA	4	2	3	4	74
LEXUS	0	0	0	0	1
MACK	0	0	0	0	2
MAZDA	3	2	6	6	184
MERCBENZ	0	0	0	0	9
MERCURY	0	0	0	0	1
MITSUBSH	1	1	3	4	153
MORRIS	0	0	0	0	2
MUSTANG	0	0	0	0	1
NISSAN	6	5	2	6	208
OPELREC	0	0	0	0	1
PEUGEOT	0	0	0	2	5
PLYMOUTH	0	0	1	0	1
SAEHAN	0	0	0	0	2
SANGYONG	0	0	0	0	1
SCANIA	0	0	0	0	2
SEAT	0	0	0	0	2
SKODA	1	0	1	3	65
SUBARU	0	0	0	0	5
SUZUKI	5	0	5	12	230
TOYOTA	4	4	1	7	186
TRIUMPH	0	0	0	0	1
VOLVO	0	0	0	0	3
VW	2	0	0	4	73
WILLYS	0	0	0	0	2
ZASTAVA	0	0	0	0	3
Total	52	27	50	106	2764

$$X^{2*} = \sum_{i=1}^{50} \sum_{j=1}^{34} \frac{(f_{o_i,j} - f_{e_i,j})^2}{f_{e_i,j}} = 2768.596 \quad \chi^2_{1617}$$

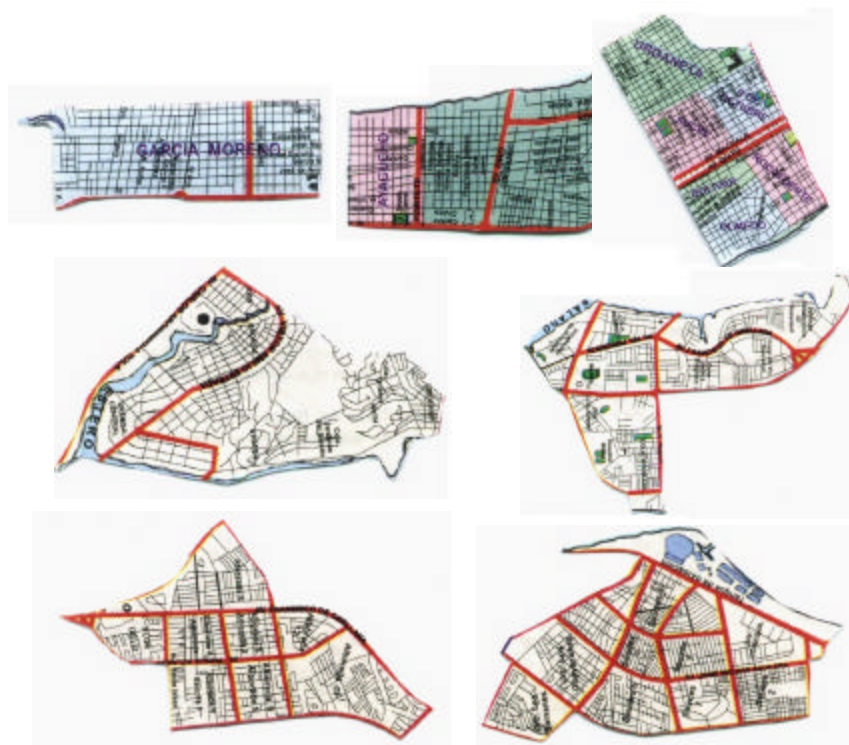
Una vez más existe evidencia suficiente para rechazar la hipótesis de independencia; hemos encontrado aquí que la marca del carro sustraído tiene relación con el sector donde se sustrajo. Estudiemos la figura 3.56 que ilustra las relaciones de estas variables por medio de curvas de nivel.

Figura 3.56: Superficies de nivel del cruce de las variables Marca Vs. Zona.



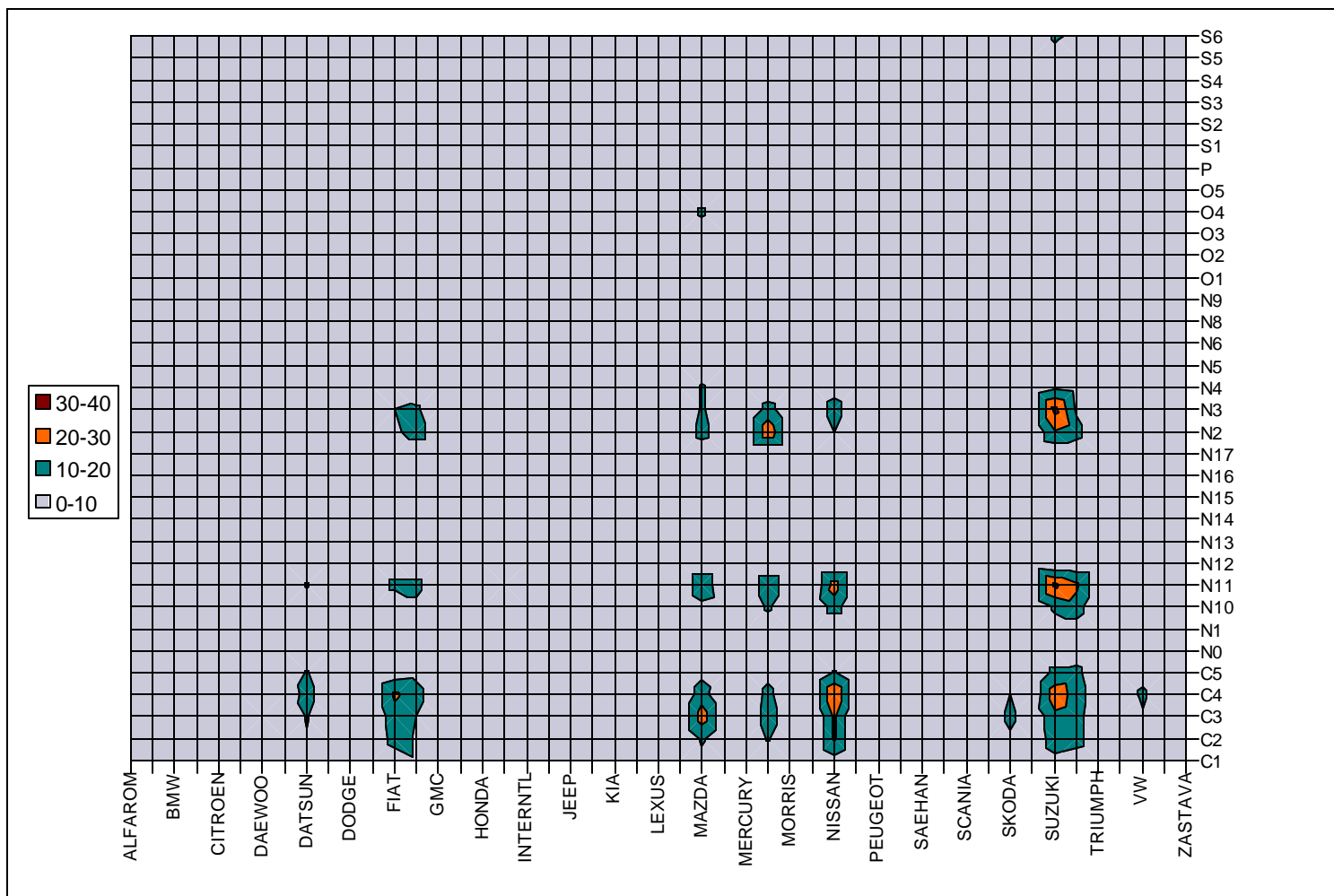
Si observamos con atención, todas las áreas marcadas como riesgosas pertenecen a la marca Chevrolet, esto es debido a que (de acuerdo a lo anteriormente analizado) esta marca tiene una alta representatividad en los resultados ocasionada por la magnitud de los datos. De acuerdo a la gráfica, los robos estarían sucediendo con mayor frecuencia en las zonas C2, C3, C4, N2, N3, N10 y N11 (Ver figura 3.57).

Figura 3.57: Zonas de mayor riesgo de robo de vehículos marca Chevrolet.



Vemos aquí que en las zonas de mayor incidencia delictiva están sucediendo más sustracciones de la marca que más roban. Es posible entonces que debido a que la marca Chevrolet tiene una alta representatividad (ya que ocupa un 33% del total de los robos) haya producido que las zonas anteriormente descritas sean las de mayor riesgo. Para estudiar lo que sucede con las demás zonas debemos dibujar otro diagrama de curvas de nivel, que no incluya a la variable Chevrolet.

Figura 3.58: Superficies de nivel del cruce de las variables Marca Vs. Zona excluyendo la marca modalidad Chevrolet.



De acuerdo a la figura 3.57 las zonas de mayor incidencia delictiva conservan el comportamiento descrito en la sección 3.1.11, que concuerdan con la marca Chevrolet; es decir que dicha marca no estaría sesgando los resultados. Por medio de este segundo gráfico nos damos cuenta que las áreas de mayor incidencia se describen por medio de las intersecciones entre las zonas y las marcas de mayor riesgo; se confirma entonces el peligro de robo que existe para los vehículos con marcas de alto riesgo en zonas de alto riesgo.

3.2.4 Num Vs. Zona

Esta prueba es una de las tantas que usaremos para tratar de delimitar las zonas (si existieran) en las que operan diferentes bandas a través del número de integrantes. Para esta prueba contrastaremos el siguiente par de hipótesis:

H_0 : El número de delincuentes no se relaciona con la zona del

robo

Vs.

H_1 : No es verdad H_0 .

Observemos la tabla 3.39 describe la tabla de contingencia para esta prueba.

Tabla 3.39: Tabla de contingencia Num Vs. Zona

	Num					
	UNO	DOS	TRES	CUATRO	CINCO	SEIS
C1	0	40	33	18	3	9
C2	1	83	71	35	5	8
C4	4	107	111	63	19	12
C5	2	68	48	21	4	5
N0	0	1	1	0	0	0
N1	2	20	15	15	4	3
N2	5	93	72	28	9	3
N3	3	77	73	35	13	5
N4	5	25	23	20	2	2
N5	1	7	20	8	1	2
N6	0	9	8	3	2	2
N8	0	2	3	0	1	0
N9	1	14	20	8	1	0
N10	1	73	57	22	5	3
N11	8	129	109	36	13	4
N12	0	2	0	1	0	0
N13	0	0	0	1	0	0
N14	1	4	3	2	2	0
N15	0	0	0	0	0	1
N16	0	1	2	0	0	0
N17	0	3	9	5	1	3

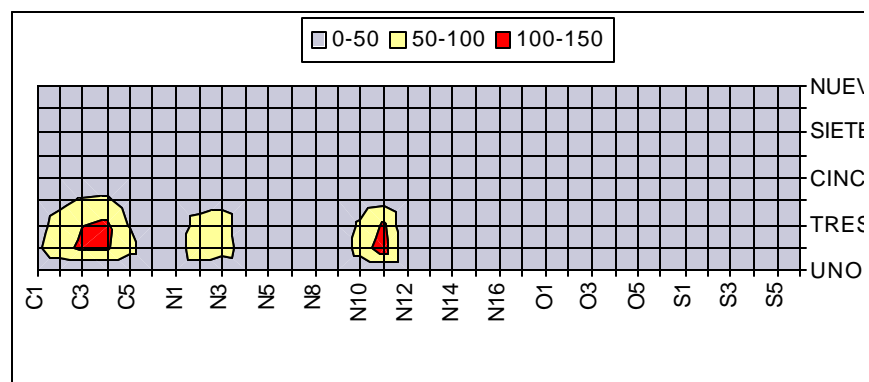
O1	1	10	6	1	0	1
O2	0	2	4	0	0	1
O3	1	14	14	12	3	2
O4	2	25	32	16	5	2
O5	0	2	1	0	2	0
P	1	6	20	12	4	1
S1	1	4	2	1	0	0
S2	2	12	11	5	2	2
S3	0	18	23	8	2	0
S4	2	10	7	5	2	1
S5	2	23	11	8	3	3
S6	2	37	44	15	5	2
Total	53	1033	952	462	129	83

	Num			
	SIETE	OCHO	NUEVE	Total
C1	0	1	0	104
C2	2	1	0	206
C3	2	3	1	302
C4	0	3	3	322
C5	2	0	4	154
N0	0	0	0	2
N1	0	0	0	59
N2	1	2	0	213
N3	1	0	0	207
N4	2	0	1	80
N5	0	0	0	39
N6	0	0	1	25
N8	0	0	0	6
N9	1	0	0	45
N10	0	0	0	161
N11	4	3	0	306
N12	0	0	0	3
N13	0	0	0	1
N14	0	0	0	12
N15	0	0	0	1
N16	0	0	0	3
N17	1	1	0	23
O1	0	0	0	19
O2	0	0	0	7
O3	0	1	0	47
O4	1	2	0	85
O5	1	0	0	6
P	0	3	2	49
S1	0	0	0	8
S2	0	0	0	34
S3	0	1	0	52
S4	0	0	0	27
S5	0	0	0	50
S6	1	0	0	106
Total	19	21	12	2764

$$X^{2*} = \sum_{i=1}^{34} \sum_{j=1}^9 \frac{(f_{o_{i,j}} - f_{e_{i,j}})^2}{f_{e_{i,j}}} = 356.76 \sim \chi^2_{264}$$

El valor p correspondiente a esta prueba es 0.000; por lo que podemos afirmar que existe suficiente evidencia estadística para rechazar la hipótesis de independencia entre las variables num y zona

Figura 3.59: Superficies de nivel del cruce de las variables Num Vs. Zona.



Claramente se observa que en las zonas de mayor riesgo (C2, C3, C4, C5, N2, N3, N10 y N11) están sucediendo robos donde intervienen de 1 a 4 sustractores

3.2.5 Locomoción Vs. Zona

En concordancia con el objetivo de descubrir las características del Modus Operandis de los delincuentes roba-carros descubriremos si existe una relación entre el medio en el que se movilizan y la zona en la que realizan la sustracción del vehículo. Mediante este análisis trataremos descubrir si este tipo de bandas se pueden ubicar por sectores.

El contraste de hipótesis a probar es el siguiente:

H_0 : El medio de locomoción no se relaciona con la zona del robo

Vs.

H_1 : No es verdad H_0 .

A continuación se encuentra la tabla de contingencia correspondiente a esta prueba.

Tabla 3.40: Tabla de contingencia Locomoción Vs. Zona

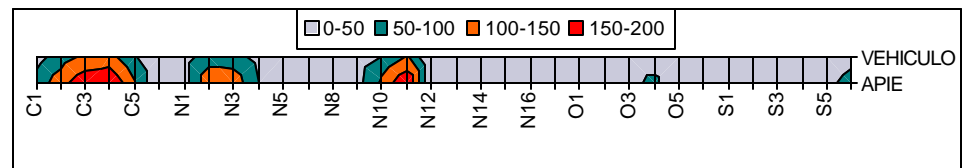
		Locomoción		
		APIE	VEHICULO	TOTAL
Zona	C1	76	28	104
	C2	127	79	206
	C3	191	111	302
	C4	198	124	322

C5	102	52	154
N0	0	2	2
N1	32	27	59
N2	134	79	213
N3	134	73	207
N4	50	30	80
N5	15	24	39
N6	18	7	25
N8	4	2	6
N9	28	17	45
N10	111	50	161
N11	199	107	306
N12	3	0	3
N13	1	0	1
N14	11	1	12
N15	0	1	1
N16	0	3	3
N17	10	13	23
O1	17	2	19
O2	5	2	7
O3	31	16	47
O4	64	21	85
O5	2	4	6
P	24	25	49
S1	5	3	8
S2	27	7	34
S3	28	24	52
S4	18	9	27
S5	38	12	50
S6	68	38	106
Total	1771	993	2764

$$X^{2*} = \sum_{i=1}^{34} \sum_{j=1}^2 \frac{(f_{o_{i,j}} - f_{e_{i,j}})^2}{f_{e_{i,j}}} = 69.605 \sim \chi^2_{33}$$

Al calcular $\int_{69.605}^{\infty} f(x)dx = 0.000$ tenemos suficiente evidencia estadística para rechazar la hipótesis de independencia entre las variables Locomoción y Zona. Analicemos las curvas de nivel para esta prueba.

Figura 3.59: Superficies de nivel de las variables Locomoción Vs. Zona



En la figura 3.59 vemos que las proporciones se mantienen invariantes, ya que ambas variables son directamente proporcionales, es decir que los robos a pie que son los más representativos se ubican en las zonas de mayor riesgo, y los robos en vehículo (menos representativos) se ubican en las zonas de menor riesgo. Podemos concluir que existe consistencia en los procedimientos de robo de estas dos variables en todo Guayaquil.

3.2.6 Entorno Vs. Acción.

Aquí investigamos el tipo de riesgo de acuerdo al entorno del robo; estudiaremos las relaciones (de existir alguna) entre el entorno del robo y la acción de los sustractores.

Contrastaremos las dos hipótesis siguientes:

H_0 : El entorno del robo no se relaciona con la acción de los delincuentes

Vs.

H_1 : No es verdad H_0 .

Primero hagamos la prueba de independencia.

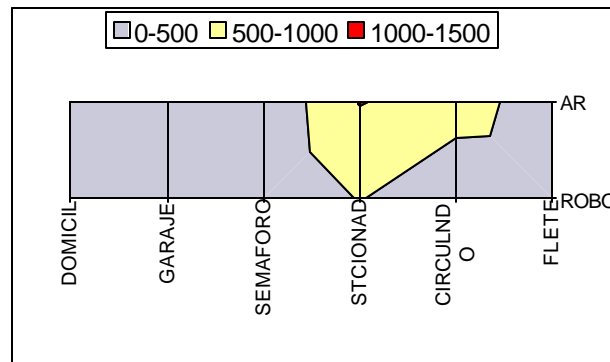
Tabla 3.40: Tabla de contingencia Entorno Vs. Acción.

		Acción		Total
		ROBO	AR	
Entorno	DOMICIL	7	70	77
	GARAJE	1	10	11
	SEMAFORO	16	84	100
	STCIONAD	534	1022	1556
	CIRCULNDO	66	755	821
	FLETE	19	180	199
Total		643	2121	2764

$$X^{2^*} = \sum_{i=1}^6 \sum_{j=1}^2 \frac{(f_{o_i,j} - f_{e_i,j})^2}{f_{e_i,j}} = 256.95 \sim \mathbf{C}^2$$

De esta prueba obtenemos valor $p = 0.000$, rechazamos entonces H_0 y decimos que existe relación entre las variables entorno y acción.

Figura 3.60: Superficies de nivel de las variables Entorno Vs. Acción



De acuerdo al gráfico se puede decir que suceden muchísimos asaltos cuando el vehículo está estacionado (1000-1500) y muchos cuando el carro está circulando, por otro lado suceden muchos robos sin asalto cuando el vehículo está estacionado, y pocos cuando está circulando. Es decir que existe mucho mayor riesgo de ser robado y asaltado cuando el vehículo está en movimiento que de ser solo asaltado; esto era de esperarse, ya que si el vehículo es sustraído cuando el conductor se encuentra dentro va a tener que suceder un enfrentamiento entre el conductor y el sustractor. Es importante que el conductor sepa que si mientras maneja su vehículo (o mantenga su vehículo estacionado) llegara a encontrarse en la

situación de víctima de robo, no ofrezca resistencia alguna, ya que estaría arriesgando su integridad física y su vida.

3.2.7 Hora Vs. Zona.

Queremos determinar que relación existe entre las modalidades de las variables hora y zona. Las hipótesis a contrastar son las siguientes:

H_0 : La hora del robo no se relaciona con la zona del robo

Vs.

H_1 : No es verdad H_0 .

Observemos los resultados de la tabla de contingencia 3.41

Tabla 3.42: Tabla de contingencia Hora Vs. Zona

		Hora											
		H00	H01	H02	H03	H04	H05	H06	H07	H08	H09	H10	H11
Zona	C1	3	1	5	2	5	1	3	4	3	2	3	0
	C2	2	3	6	5	5	2	7	12	8	6	3	6
	C3	6	6	6	5	1	3	8	21	11	6	7	15
	C4	3	3	3	7	11	12	12	11	8	5	17	6
	C5	1	5	4	1	3	5	7	10	6	9	6	9
	N0	0	0	0	1	0	0	0	0	1	0	0	0
	N1	0	0	0	1	0	2	1	2	2	3	5	3
	N10	8	3	10	4	2	2	0	1	3	4	0	5
	N11	8	5	3	5	5	2	3	13	6	2	9	10
	N12	0	0	0	0	0	0	0	0	0	1	0	1
	N13	0	0	0	0	0	0	0	0	0	0	0	0
	N14	1	0	0	1	0	0	1	0	1	1	0	2
	N15	0	0	0	0	0	0	0	0	0	0	0	0
	N16	0	0	0	0	0	0	0	1	0	0	0	0
	N17	0	0	0	0	0	0	0	2	0	0	0	2
	N2	10	2	1	1	4	4	5	12	6	6	3	11
	N3	3	3	2	4	3	1	3	8	6	7	8	7
	N4	0	2	3	0	1	2	2	3	3	2	5	6
N5	0	2	1	0	0	0	0	2	5	0	0	2	
N6	0	1	2	2	2	3	0	0	2	0	0	0	
N8	0	0	0	0	0	0	0	0	0	1	0	0	

N9	0	1	0	2	1	3	0	1	0	1	0	1
O1	1	0	1	1	1	1	0	0	2	0	0	0
O2	1	1	0	1	0	0	0	0	0	0	0	0
O3	1	1	0	3	1	1	1	2	3	3	2	5
O4	5	2	3	1	3	2	1	3	4	2	2	1
O5	0	0	0	0	1	0	1	0	0	0	0	0
P	1	1	1	0	3	0	2	1	1	2	3	1
S1	1	0	0	1	0	1	0	0	1	0	0	0
S2	2	2	0	1	3	1	2	2	0	0	0	0
S3	1	0	2	6	3	1	2	2	0	1	0	3
S4	2	2	2	2	1	0	0	1	0	0	1	2
S5	0	0	2	2	1	1	0	4	3	2	3	2
S6	2	0	2	0	0	1	1	2	1	4	2	3
Total	62	46	59	59	60	51	62	120	86	70	79	103

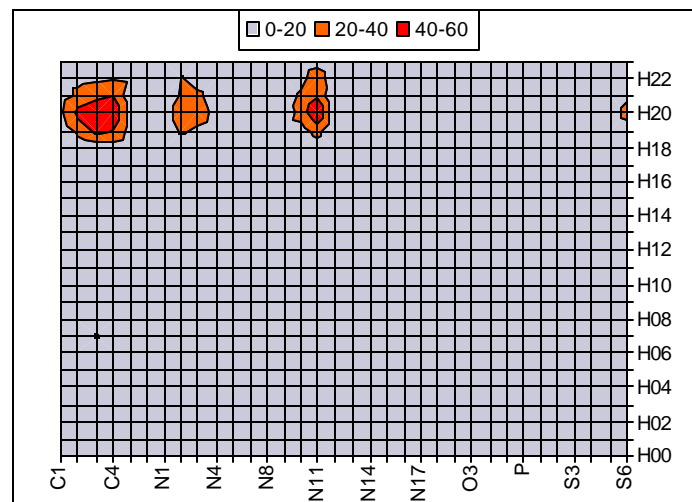
	Hora											
	H12	H13	H14	H15	H16	H17	H18	H19	H20	H21	H22	H23
C1	4	1	2	3	2	2	1	9	19	12	12	5
C2	1	6	0	7	2	3	7	24	46	26	12	7
C3	12	7	4	3	3	3	11	44	57	37	16	10
C4	10	5	7	6	6	9	10	39	56	40	19	17
C5	9	2	6	5	8	4	6	9	10	8	13	8
N0	0	0	0	0	0	0	0	0	0	0	0	0
N1	1	3	3	2	0	2	5	3	8	8	2	3
N10	2	3	5	5	4	6	7	12	29	24	14	8
N11	9	12	12	5	5	6	14	26	60	37	36	13
N12	0	0	0	0	0	0	0	0	0	0	1	0
N13	0	0	0	0	0	0	0	0	0	1	0	0
N14	0	0	1	1	0	1	0	0	1	0	0	1
N15	0	0	0	0	1	0	0	0	0	0	0	0
N16	0	0	0	0	0	0	0	1	0	0	0	1
N17	0	1	2	1	1	0	0	7	6	1	0	0
N2	10	8	6	6	3	4	6	22	33	22	21	7
N3	9	10	6	5	6	12	11	16	34	26	8	9
N4	7	4	3	1	4	3	2	8	9	4	2	4
N5	3	1	0	0	2	1	1	4	5	4	5	1
N6	2	1	1	0	0	0	1	1	4	1	2	0
N8	0	0	0	0	0	0	0	1	0	2	0	2
N9	1	0	0	1	2	2	3	8	5	9	3	1
O1	1	0	0	0	1	0	0	1	4	1	2	2
O2	0	0	0	0	0	1	2	0	0	0	1	0
O3	0	0	2	1	0	0	1	2	5	6	4	3
O4	3	2	1	3	0	2	1	13	12	8	9	2
O5	0	0	1	0	0	1	0	0	0	0	1	1
P	1	0	2	1	1	2	5	6	8	4	3	0
S1	0	0	1	0	1	0	0	1	0	0	1	0
S2	2	2	1	2	1	2	0	1	2	2	3	3
S3	0	0	1	0	0	0	0	5	8	7	4	6
S4	0	0	0	1	0	0	1	1	3	3	3	2

S5	2	1	3	0	1	1	1	6	6	3	3	3
S6	3	2	1	6	3	0	3	11	27	17	11	4
Total	92	71	71	65	57	67	99	281	457	313	211	123

$$X^{2*} = \sum_{i=1}^{34} \sum_{j=1}^{24} \frac{(f_{o_{i,j}} - f_{e_{i,j}})^2}{f_{e_{i,j}}} = 980.605 \sim \chi^2_{759}$$

Debido a que el valor del estadístico ji-cuadrado es muy alto tenemos suficiente evidencia estadística para rechazar H_0 ; es decir la hora y la zona tienen relación entre si.

Figura 3.60: Superficies de nivel de las variables Hora Vs. Zona



Vemos que el riesgo de robo en las zonas C2,C3, C4 y N2 es mayor desde las 19H00 hasta antes de las 22H00; en las zonas N3 y N10 desde las 20H00 hasta antes de las 22H00; en la zona N11 desde las 19H00 hasta antes de las 23H00 y en la

zona S6 alrededor de las 20H00. Podemos notar que a pesar que la zona S6 no se encuentra entre las de mayor riesgo, aparece en la gráfica representada de color naranja alrededor de las 20H00, esto implica que sigue el mismo comportamiento que las demás variables; esto implica que el riesgo es alto a las mismas horas. Estas conclusiones son importantes, ya que ahora existe suficiente evidencia estadística para afirmar que debemos tener mucha precaución en los sectores de riesgo únicamente en las horas de riesgo (aquí descritas), sin tener que preocuparse todo el tiempo, ya que a las demás horas el número de sustracciones no es representativo. De todos modos no debemos descuidarnos en las horas de menor riesgo en los sectores azules, ya que esta porción ocupa un 32% de los robos anuales que equivalen aproximadamente a 884 que representan alrededor de 2 robos diarios.

3.2.8 Marca Vs Color.

Es posible que exista alguna relación entre la marca del vehículo y su color; de ser así, sería muy interesante descubrir como está establecida dicha relación. A priori podemos esperar que la relación concuerde con las características definidas en la sección 3.1, pero sabemos que es posible que no suceda así.

Una vez mas contrastemos las hipótesis correspondientes a esta prueba:

H_0 : La Marca del vehículo no se relaciona con el Color del
vehículo

Vs.

H_1 : No es verdad H_0 .

analicemos la tabla de contingencia de las variables a estudiar.

Tabla 3.43: Tabla de contingencia Marca Vs. Color

	Color					
	Amarillo	Azul	Bicolor	Blanco	Café	Celeste
ALFAROM	0	0	0	0	1	0
AUSTIN	0	3	0	0	0	0
BMW	0	0	0	1	0	0
CHEVRLT	39	112	3	123	8	34
CHRYSLER	0	0	0	1	0	0
CITROEN	0	0	0	0	0	0
DACIA	0	0	0	1	0	0
DAEWOO	3	6	0	7	0	0
DAIHATSU	2	2	0	4	0	1
DATSUN	18	17	0	17	1	6
DELTA	0	0	0	2	0	0
DODGE	0	0	0	0	0	0
ENCAVA	0	0	0	1	0	0
FIAT	1	15	0	19	0	0
FORD	5	14	0	12	1	5
GMC	0	0	0	1	0	0
HINO	1	0	0	4	0	0
HONDA	0	3	0	3	0	0
HYUNDAI	25	7	0	9	0	0
INTERNTL	0	0	1	1	0	0
ISUZU	2	4	0	8	1	0
JEEP	0	0	0	3	1	1
JHONDEE	1	0	0	0	0	0
KIA	0	1	0	7	0	0
LADA	38	1	0	11	0	1
LEXUS	0	0	0	0	0	0
MACK	0	0	0	1	0	0
MAZDA	7	30	1	38	4	0
MERCBENZ	1	1	0	2	1	1
MERCURY	0	0	0	0	0	0
MITSUBSH	1	18	0	29	2	1
MORRIS	0	0	0	0	1	0
MUSTANG	0	0	0	1	0	0
NISSAN	3	21	0	41	1	8
OPELREC	0	0	0	0	0	1
PEUGEOT	0	1	0	0	0	0
PLYMOUTH	0	0	0	0	0	0
SAEHAN	0	0	0	1	0	0
SANGYONG	0	0	0	0	0	0
SCANIA	0	1	0	0	0	0
SEAT	0	0	0	0	0	0
SKODA	2	17	0	16	1	1
SUBARU	1	0	0	2	1	0
SUZUKI	3	26	0	39	7	13
TOYOTA	10	24	0	41	2	7
TRIUMPH	0	0	0	0	0	0
VOLVO	0	0	0	0	0	0
VW	4	1	0	17	0	2

WILLYS	0	0	0	0	0	0
ZASTAVA	0	0	0	2	0	0
Total	167	325	5	465	33	82

	Color					
	Crema	Gris	Morado	Naranja	Negro	Rojo
ALFAROM	0	0	0	0	0	0
AUSTIN	0	0	0	0	0	1
BMW	0	1	0	0	0	1
CHEVRLT	29	164	68	2	26	171
CHRYSLER	0	0	0	0	0	1
CITROEN	0	1	0	0	0	0
DACIA	0	0	0	0	0	0
DAEWOO	4	10	5	0	1	4
DAIHATSU	1	4	0	0	1	2
DATSUN	3	12	8	1	3	40
DELTA	0	0	0	0	0	0
DODGE	0	1	1	0	1	0
ENCAVA	0	0	0	0	0	0
FIAT	7	24	5	1	5	29
FORD	3	20	6	1	6	30
GMC	0	0	0	0	1	0
HINO	0	0	0	0	0	0
HONDA	0	4	0	0	1	1
HYUNDAI	0	11	3	0	3	10
INTERNTL	0	0	0	0	0	0
ISUZU	0	2	0	0	2	8
JEEP	2	5	1	1	3	0
JHONDEE	0	0	0	0	0	0
KIA	0	1	1	0	1	0
LADA	10	2	3	0	1	6
LEXUS	1	0	0	0	0	0
MACK	0	0	1	0	0	0
MAZDA	6	30	8	3	2	27
MERCBENZ	0	1	0	0	1	1
MERCURY	0	0	0	0	0	1
MITSUBSH	8	31	20	2	6	14
MORRIS	0	0	0	0	0	1
MUSTANG	0	0	0	0	0	0
NISSAN	5	43	17	1	2	35
OPELREC	0	0	0	0	0	0
PEUGEOT	2	0	0	0	0	1
PLYMOUTH	0	0	1	0	0	0
SAEHAN	0	0	1	0	0	0
SANGYONG	0	0	0	0	0	1
SCANIA	0	0	0	0	0	1
SEAT	0	0	0	0	0	1
SKODA	0	4	1	0	1	13
SUBARU	1	0	0	0	0	0
SUZUKI	9	31	10	3	14	44
TOYOTA	10	16	6	0	3	30

TRIUMPH	1	0	0	0	0	0
VOLVO	0	1	0	0	1	0
WW	1	7	3	1	4	17
WILLYS	0	0	0	0	1	1
ZASTAVA	0	0	0	0	0	0
Total	103	426	169	16	90	492

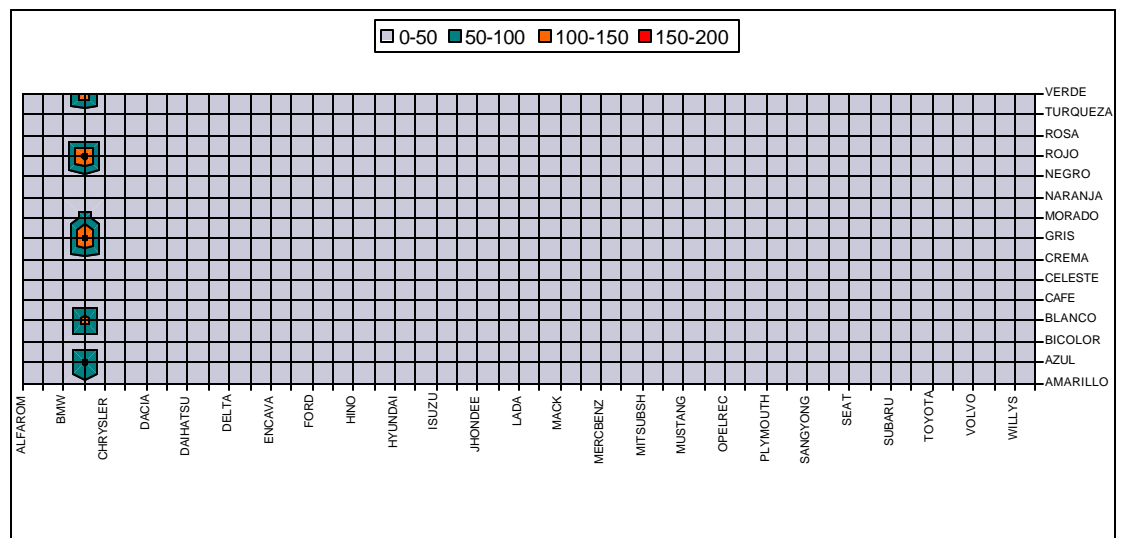
	Color			Total
	Rosa	Turqueza	Verde	
ALFAROM	0	0	0	1
AUSTIN	0	0	0	4
BMW	0	0	2	5
CHEVRLT	0	3	134	916
CHRYSLER	0	0	0	2
CITROEN	0	0	0	1
DACIA	0	0	0	1
DAEWOO	0	0	8	48
DAIHATSU	0	0	2	19
DATSUN	0	1	7	134
DELTA	0	0	0	2
DODGE	0	0	0	3
ENCAVA	0	0	0	1
FIAT	0	1	16	123
FORD	0	0	18	121
GMC	0	0	0	2
HINO	0	0	0	5
HONDA	0	0	0	12
HYUNDAI	2	1	11	82
INTERNTL	0	0	1	3
ISUZU	0	0	2	29
JEEP	0	0	2	19
JHONDEE	0	0	0	1
KIA	0	0	2	13
LADA	0	0	1	74
LEXUS	0	0	0	1
MACK	0	0	0	2
MAZDA	0	0	28	184
MERCBENZ	0	0	0	9
MERCURY	0	0	0	1
MITSUBSH	0	1	20	153
MORRIS	0	0	0	2
MUSTANG	0	0	0	1
NISSAN	0	0	31	208
OPELREC	0	0	0	1
PEUGEOT	0	0	1	5
PLYMOUTH	0	0	0	1
SAEHAN	0	0	0	2
SANGYONG	0	0	0	1
SCANIA	0	0	0	2
SEAT	0	1	0	2
SKODA	0	0	9	65
SUBARU	0	0	0	5
SUZUKI	0	6	25	230
TOYOTA	0	1	36	186
TRIUMPH	0	0	0	1
VOLVO	0	0	1	3
VW	0	0	16	73

WILLYS	0	0	0	2
ZASTAVA	0	0	1	3
Total	2	15	374	2764

$$X^{2*} = \sum_{i=1}^{50} \sum_{j=1}^{15} \frac{(f_{o_{i,j}} - f_{e_{i,j}})^2}{f_{e_{i,j}}} = 1586.052 \sim \chi^2_{686}$$

De esta prueba hemos obtenido un valor $p = 0.000$; de aquí que podemos rechazar la hipótesis nula de independencia entre las variables marca y color.

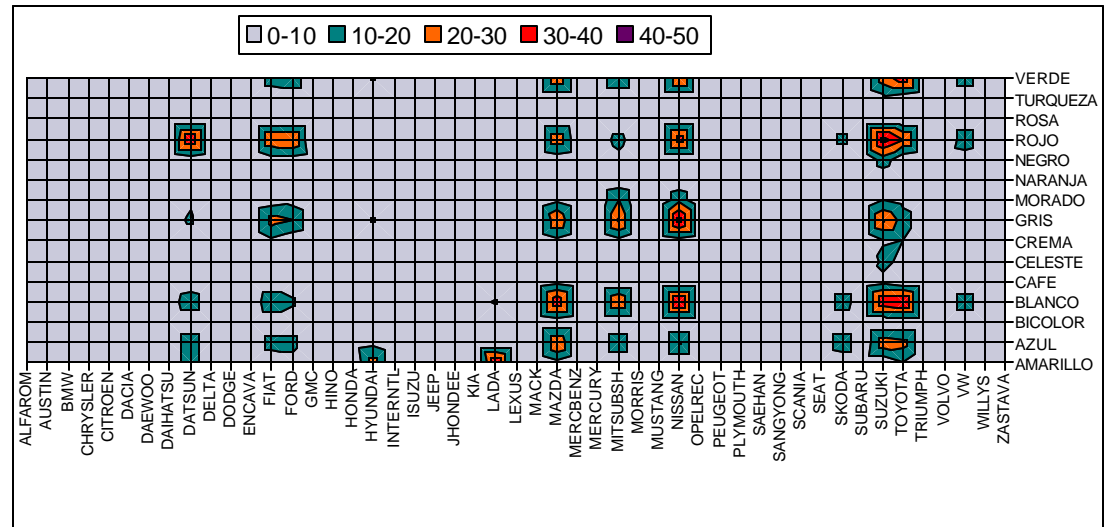
Figura 3.60: Superficies de nivel de las variables Marca Vs. Color.



En esta figura aparecen como únicas zonas de riesgo las correspondientes a la marca Chevrolet; esto era de esperarse debido a que esta marca ocupa un 33% del total por si sola.

Sería interesante examinar que sucede con las demás marcas, para determinar la relación marca-color.

Figura 3.60: Superficies de nivel de las variables Marca Vs. Color.



De acuerdo a ambas figuras hemos formado la siguiente tabla:

Tabla 3.59: Tipo de riesgo de robo por marca de acuerdo al color.

Marca	Color									
	Amarillo	Azul	Blanco	Celeste	Gris	Morado	Negro	Rojo	Verde	
Chevrolet		○	○		○	○		○	○	
Datsun								○		
Fiat					○			○		
Ford								○		
Hyundai	○									
Lada	○									
Mazda		○		○				○		
Mitsubishi				○	○	○				
Nissan			○		○	○		○	○	
Skoda								○		
Suzuki		○	○		○			○	○	
Toyota		○	○		○			○	○	
Volkswagen										

Referencia.- ○: Riesgo Alto; ◐: Riesgo Medio; Vacías: Riesgo Bajo

A diferencia de lo que sucedía con los análisis previos aquí aparecen las modalidades que no tenían mucha relevancia en el análisis univariado, como por ejemplo los colores amarillo, celeste, morado y negro y las marcas Ford, Hyundai, Lada, Skoda y Volkswagen.

3.3 Análisis Multivariado

Para poder culminar nuestro estudio es necesario realizar un análisis más elaborado; de tal manera que podamos establecer relaciones entre más de 2 variables.

En esta parte nos disponemos a investigar como se agrupa el siguiente grupo de variables:

1. Situación que propicia el robo: ¿Se puede establecer alguna relación entre el día, la hora y la zona del robo?

Para decidir como agrupar las variables vamos a hacer uso del Análisis de Componentes Principales estudiado en el capítulo 2. Mediante este método estadístico obtendremos grupos de variables que tienen características de variabilidad similares dentro del grupo y diferentes fuera de él.

A continuación estudiaremos las relaciones entre el día, la hora y la zona del robo:

3.3.1 Día Vs. Hora Vs. Zona

El número total de modalidades de las variables que nos disponemos estudiar es 65. El objetivo es agruparlas de acuerdo a sus características de variabilidad en no mas de 10 variables con el objeto de encontrar conjuntos mas pequeños claramente distinguibles unos de otros.

El primer paso a seguir es calcular la matriz de covarianzas de las 65 variables originales, luego calcularemos sus valores propios para determinar que porcentaje de la varianza total está contenido en cada componente y con que tanta información contribuye cada uno. A continuación de este procedimiento determinaremos que componentes consideraremos importantes y elegiremos un número razonable de ellos tomando en cuenta la cantidad de información proporcionada por cada una.

La tabla 3.60 muestra la matriz de covarianzas de las variables originales; esta es una matriz triangular superior debido a que toda matriz de covarianzas es simétrica ($A=A^T$).

Inicialmente obtendremos 65 componentes (igual al total de variables); de estos elegiremos los que contribuyan con mayor información que una sola variable; para este caso el porcentaje de la varianza contenido en cada componente debería ser mayor que $100 / 65$; en otras palabras

$$\% / \text{variable} \geq 100\% / 65 \text{ variables} = 1.54\% \text{ por variable}$$

Tabla 3.62: Valores propios, % de varianza y % de varianza acumulada de la matriz de covarianzas

Comp.	Eigen Valores	% de la Varianza	Porcentaje Acumulado
1	12.12	15.52	15.52
2	11.41	14.61	30.13
3	9.99	12.79	42.92
4	9.80	12.54	55.46
5	9.11	11.67	67.13
6	8.16	10.45	77.58
7	3.54	4.54	82.11
8	1.26	1.61	83.72
9	1.05	1.35	85.07
10	0.87	1.12	86.19
11	0.79	1.01	87.19
12	0.71	0.91	88.10
13	0.63	0.81	88.92
14	0.62	0.79	89.71
15	0.55	0.70	90.42
16	0.51	0.66	91.07
17	0.47	0.60	91.68
18	0.42	0.54	92.22
19	0.40	0.52	92.73
20	0.36	0.46	93.19
21	0.32	0.40	93.60
22	0.30	0.38	93.98
23	0.29	0.38	94.36
24	0.28	0.36	94.71
25	0.26	0.33	95.04
26	0.24	0.30	95.35
27	0.23	0.30	95.64
28	0.21	0.27	95.92
29	0.21	0.26	96.18
30	0.20	0.25	96.44
31	0.19	0.24	96.68
32	0.19	0.24	96.91
33	0.18	0.23	97.14
34	0.16	0.21	97.35
35	0.16	0.20	97.55
36	0.15	0.20	97.74
37	0.15	0.19	97.93
38	0.15	0.19	98.12
39	0.14	0.18	98.30
40	0.13	0.16	98.47
41	0.12	0.16	98.62
42	0.12	0.15	98.77
43	0.11	0.15	98.92
44	0.10	0.13	99.06
45	0.10	0.13	99.18
46	0.09	0.12	99.30
47	0.09	0.11	99.42
48	0.08	0.10	99.51
49	0.07	0.09	99.60
50	0.06	0.07	99.67
51	0.05	0.07	99.74
52	0.05	0.06	99.80
53	0.04	0.05	99.85
54	0.03	0.04	99.89
55	0.02	0.03	99.91
56	0.02	0.02	99.94
57	0.01	0.02	99.95
58	0.01	0.02	99.97
59	0.01	0.01	99.98
60	0.01	0.01	99.99
61	0.01	0.01	99.99
62	0.00	0.00	100.00
63	0.00	0.00	100
64	0.00	0.00	100
65	0.00	0.00	100

Vemos claramente que las 8 primeras componentes contribuyen con mas de 1.54% por variable, acumulando entre ellas un 83.72% de la varianza total. Esto es bueno, ya que podremos reemplazar las 65 variables originales por un número reducido de componentes que explicarán una gran parte de la información total.

Proseguimos a calcular la matriz de vectores propios correspondientes a cada valor propio que acumule más del 1.54% de la varianza total.

Tabla 3.61: Valores propios de los componentes principales de las variables día, hora y zona.

	1	2	3	4	5	6	7	8
LUN	0.40	-0.02	0.20	0.01	0.16	0.23	0.42	0.03
MAR	0.47	-0.13	-0.25	-0.24	-0.23	0.41	0.41	0.06
MIER	0.46	-0.29	-0.16	-0.33	-0.22	-0.04	0.48	0.03
JUE	0.50	0.06	0.05	0.10	0.27	0.19	0.30	-0.01
VIER	-0.09	0.48	0.02	0.13	0.33	0.17	0.37	0.00
SAB	-0.11	-0.03	-0.07	-0.09	-0.37	0.28	0.05	0.01
DOM	-0.06	0.14	-0.08	-0.12	-0.17	0.34	0.25	0.05
H00	-0.01	0.01	-0.02	0.02	-0.17	-0.04	0.10	0.02
H01	-0.02	0.03	0.00	0.02	0.00	0.14	0.05	0.00
H02	-0.03	0.04	0.01	-0.02	-0.05	0.16	0.09	0.00
H03	0.02	0.01	0.01	-0.01	-0.18	-0.02	0.06	0.00
H04	0.06	0.04	0.11	-0.02	-0.01	-0.02	0.04	0.00
H05	0.02	0.00	-0.01	0.00	-0.01	0.13	0.02	0.03
H06	-0.01	0.01	0.06	0.02	0.03	0.03	0.07	0.40
H07	0.01	0.01	0.03	0.00	0.10	0.08	0.10	0.41
H08	0.00	0.19	0.01	0.04	0.06	0.02	0.09	0.01
H09	0.01	-0.12	0.01	0.01	0.02	0.02	0.14	-0.07
H10	-0.01	-0.02	0.00	0.17	-0.05	0.04	0.04	0.03
H11	-0.01	0.02	0.00	-0.02	-0.03	0.02	0.25	-0.02
H12	0.02	0.20	0.01	-0.01	0.00	0.02	0.08	0.01
H13	0.01	-0.03	-0.02	0.01	0.03	0.01	0.21	0.01
H14	0.00	0.02	-0.02	0.02	0.03	-0.01	0.20	0.04
H15	0.03	0.03	-0.04	-0.04	0.15	0.02	0.06	0.04
H16	0.01	0.01	0.03	0.14	-0.01	-0.01	0.03	0.01
H17	0.00	-0.01	0.01	-0.02	0.03	-0.03	0.07	-0.32
H18	0.00	0.14	-0.02	0.02	0.04	-0.01	0.07	0.03
H19	0.20	0.05	0.06	0.06	0.14	0.09	0.29	0.39

H20	0.24	-0.08	-0.06	0.10	0.14	-0.05	0.20	-0.84
H21	0.08	0.09	0.00	-0.03	0.06	-0.07	0.42	0.44
H22	-0.04	0.21	0.03	0.04	-0.03	0.07	0.22	0.10
H23	0.04	-0.01	0.03	-0.04	-0.06	0.21	0.13	-0.04
N0	0.00	0.00	0.00	-0.06	0.00	0.00	0.00	0.00
N1	0.03	-0.01	-0.01	-0.01	-0.02	-0.15	0.05	-0.03
N2	0.20	0.07	0.03	0.08	0.04	-0.04	0.13	-0.04
N3	0.21	0.02	-0.08	0.03	0.12	0.08	0.21	0.04
N4	-0.12	0.00	0.00	0.02	0.01	0.11	0.08	0.04
N5	0.00	0.03	0.01	-0.02	0.03	0.03	0.17	0.00
N6	-0.01	0.03	-0.02	-0.02	0.00	-0.01	0.16	0.02
N8	0.00	0.07	0.00	0.00	-0.01	0.01	0.00	0.01
N9	0.00	-0.01	-0.01	0.14	-0.01	0.00	0.04	0.01
N10	0.22	-0.02	-0.06	0.00	0.01	-0.01	0.12	0.00
N11	0.26	0.04	-0.01	0.03	0.06	0.03	0.20	0.37
N12	-0.01	0.01	0.07	0.00	-0.01	-0.01	0.00	0.00
N13	0.01	0.00	0.07	0.00	0.00	0.00	0.00	0.00
N14	0.00	0.00	-0.01	0.00	0.00	0.12	0.00	-0.01
N15	0.01	0.00	0.05	0.00	0.00	0.00	0.00	0.00
N16	-0.01	0.01	0.07	0.00	-0.01	0.01	0.01	0.00
N17	0.01	0.13	-0.02	0.01	0.01	0.00	0.02	0.00
O1	-0.02	0.02	0.12	0.02	0.00	0.01	0.02	0.00
O2	0.00	0.08	0.00	0.00	0.00	0.00	-0.01	0.00
O3	-0.01	0.01	-0.01	0.00	-0.13	0.03	0.04	0.02
O4	-0.04	0.03	-0.03	0.12	0.01	0.02	0.09	0.02
O5	0.00	0.00	0.00	-0.11	0.00	0.00	-0.08	0.00
P	-0.12	0.03	0.00	0.00	0.05	0.03	0.08	0.02
S1	0.00	-0.01	-0.08	0.01	0.01	0.00	0.00	-0.01
S2	-0.03	0.01	0.12	0.01	0.02	-0.01	0.05	0.02
S3	-0.01	0.01	0.03	-0.14	-0.01	0.02	0.06	-0.01
S4	0.00	-0.01	-0.01	0.11	0.00	0.01	0.05	-0.02
S5	-0.01	0.14	0.04	0.00	0.00	0.01	0.03	-0.01
S6	-0.05	0.02	0.14	0.05	0.03	-0.01	0.09	-0.06
C1	0.03	0.02	0.01	-0.04	0.16	0.03	0.08	0.00
C2	0.14	0.08	0.02	-0.04	-0.02	0.07	0.16	-0.16
C3	0.18	0.03	0.14	0.04	0.10	0.11	0.22	-0.04
C4	0.15	0.09	0.01	0.07	0.08	0.01	0.17	0.06
C5	0.03	0.09	0.03	-0.02	0.02	-0.01	0.18	-0.46

Cuando se trabaja con la matriz de covarianzas resulta complicado establecer que valores con altos y cuales no. Es por eso que es útil calcular los coeficientes de correlación bivariados entre las componentes principales y las modalidades de las variables en estudio. En la tabla 3.62 se encuentran dichos coeficientes para el ejercicio en cuestión.

3.3 Análisis Multivariado

Para poder culminar nuestro estudio es necesario realizar un análisis más elaborado; de tal manera que podamos establecer relaciones entre más de 2 variables.

En esta parte nos disponemos a investigar como se agrupa el siguiente grupo de variables:

1. Situación que propicia el robo: ¿Se puede establecer alguna relación entre el día, la hora y la zona del robo?

Para decidir como agrupar las variables vamos a hacer uso del Análisis de Componentes Principales estudiado en el capítulo 2. Mediante este método estadístico obtendremos grupos de variables que tienen características de variabilidad similares dentro del grupo y diferentes fuera de él.

A continuación estudiaremos las relaciones entre el día, la hora y la zona del robo:

3.3.1 Día Vs. Hora Vs. Zona

El número total de modalidades de las variables que nos disponemos estudiar es 65. El objetivo es agruparlas de acuerdo a sus características de variabilidad en no mas de 10 variables con el objeto de encontrar conjuntos mas pequeños claramente distinguibles unos de otros.

El primer paso a seguir es calcular la matriz de covarianzas de las 65 variables originales, luego calcularemos sus valores propios para determinar que porcentaje de la varianza total está contenido en cada componente y con que tanta información contribuye cada uno. A continuación de este procedimiento determinaremos que componentes consideraremos importantes y elegiremos un número razonable de ellos tomando en cuenta la cantidad de información proporcionada por cada una.

La tabla 3.60 muestra la matriz de covarianzas de las variables originales; esta es una matriz triangular superior debido a que toda matriz de covarianzas es simétrica ($A=A^T$).

Inicialmente obtendremos 65 componentes (igual al total de variables); de estos elegiremos los que contribuyan con mayor información que una sola variable; para este caso el porcentaje de la varianza contenido en cada componente debería ser mayor que $100 / 65$; en otras palabras

$$\% / \text{variable} \geq 100\% / 65 \text{ variables} = 1.54\% \text{ por variable}$$

Tabla 3.62: Valores propios, % de varianza y % de varianza acumulada de la matriz de covarianzas

Comp.	Eigen Valores	% de la Varianza	Porcentaje Acumulado	Comp.	Eigen %Valores	% de la Varianza	Porcentaje Acumulado
1	12.12	15.52	15.52	34	0.16	0.21	97.35
2	11.41	14.61	30.13	35	0.16	0.20	97.55
3	9.99	12.79	42.92	36	0.15	0.20	97.74
4	9.80	12.54	55.46	37	0.15	0.19	97.93
5	9.11	11.67	67.13	38	0.15	0.19	98.12
6	8.16	10.45	77.58	39	0.14	0.18	98.30
7	3.54	4.54	82.11	40	0.13	0.16	98.47
8	1.26	1.61	83.72	41	0.12	0.16	98.62
9	1.05	1.35	85.07	42	0.12	0.15	98.77
10	0.87	1.12	86.19	43	0.11	0.15	98.92
11	0.79	1.01	87.19	44	0.10	0.13	99.06
12	0.71	0.91	88.10	45	0.10	0.13	99.18
13	0.63	0.81	88.92	46	0.09	0.12	99.30
14	0.62	0.79	89.71	47	0.09	0.11	99.42
15	0.55	0.70	90.42	48	0.08	0.10	99.51
16	0.51	0.66	91.07	49	0.07	0.09	99.60
17	0.47	0.60	91.68	50	0.06	0.07	99.67
18	0.42	0.54	92.22	51	0.05	0.07	99.74
19	0.40	0.52	92.73	52	0.05	0.06	99.80
20	0.36	0.46	93.19	53	0.04	0.05	99.85
21	0.32	0.40	93.60	54	0.03	0.04	99.89
22	0.30	0.38	93.98	55	0.02	0.03	99.91
23	0.29	0.38	94.36	56	0.02	0.02	99.94
24	0.28	0.36	94.71	57	0.01	0.02	99.95
25	0.26	0.33	95.04	58	0.01	0.02	99.97
26	0.24	0.30	95.35	59	0.01	0.01	99.98
27	0.23	0.30	95.64	60	0.01	0.01	99.99
28	0.21	0.27	95.92	61	0.01	0.01	99.99
29	0.21	0.26	96.18	62	0.00	0.00	100.00
30	0.20	0.25	96.44	63	0.00	0.00	100
31	0.19	0.24	96.68	64	0.00	0.00	100
32	0.19	0.24	96.91	65	0.00	0.00	100
33	0.18	0.23	97.14				

Vemos claramente que las 8 primeras componentes contribuyen con mas de 1.54% por variable, acumulando entre ellas un 83.72% de la varianza total. Esto es bueno, ya que podremos reemplazar las 65 variables originales por un número reducido de componentes que explicarán una gran parte de la información total.

Proseguimos a calcular la matriz de vectores propios correspondientes a cada valor propio que acumule más del 1.54% de la varianza total.

Tabla 3.61: Valores propios de los componentes principales de las variables día, hora y zona.

	1	2	3	4	5	6	7	8
LUN	0.40	-0.02	0.20	0.01	0.16	0.23	0.42	0.03
MAR	0.47	-0.13	-0.25	-0.24	-0.23	0.41	0.41	0.06
MIER	0.46	-0.29	-0.16	-0.33	-0.22	-0.04	0.48	0.03
JUE	0.50	0.06	0.05	0.10	0.27	0.19	0.30	-0.01
VIER	-0.09	0.48	0.02	0.13	0.33	0.17	0.37	0.00
SAB	-0.11	-0.03	-0.07	-0.09	-0.37	0.28	0.05	0.01
DOM	-0.06	0.14	-0.08	-0.12	-0.17	0.34	0.25	0.05
H00	-0.01	0.01	-0.02	0.02	-0.17	-0.04	0.10	0.02
H01	-0.02	0.03	0.00	0.02	0.00	0.14	0.05	0.00
H02	-0.03	0.04	0.01	-0.02	-0.05	0.16	0.09	0.00
H03	0.02	0.01	0.01	-0.01	-0.18	-0.02	0.06	0.00
H04	0.06	0.04	0.11	-0.02	-0.01	-0.02	0.04	0.00
H05	0.02	0.00	-0.01	0.00	-0.01	0.13	0.02	0.03
H06	-0.01	0.01	0.06	0.02	0.03	0.03	0.07	0.40
H07	0.01	0.01	0.03	0.00	0.10	0.08	0.10	0.41
H08	0.00	0.19	0.01	0.04	0.06	0.02	0.09	0.01
H09	0.01	-0.12	0.01	0.01	0.02	0.02	0.14	-0.07
H10	-0.01	-0.02	0.00	0.17	-0.05	0.04	0.04	0.03
H11	-0.01	0.02	0.00	-0.02	-0.03	0.02	0.25	-0.02
H12	0.02	0.20	0.01	-0.01	0.00	0.02	0.08	0.01
H13	0.01	-0.03	-0.02	0.01	0.03	0.01	0.21	0.01
H14	0.00	0.02	-0.02	0.02	0.03	-0.01	0.20	0.04
H15	0.03	0.03	-0.04	-0.04	0.15	0.02	0.06	0.04
H16	0.01	0.01	0.03	0.14	-0.01	-0.01	0.03	0.01
H17	0.00	-0.01	0.01	-0.02	0.03	-0.03	0.07	-0.32
H18	0.00	0.14	-0.02	0.02	0.04	-0.01	0.07	0.03
H19	0.20	0.05	0.06	0.06	0.14	0.09	0.29	0.39

H20	0.24	-0.08	-0.06	0.10	0.14	-0.05	0.20	-0.84
H21	0.08	0.09	0.00	-0.03	0.06	-0.07	0.42	0.44
H22	-0.04	0.21	0.03	0.04	-0.03	0.07	0.22	0.10
H23	0.04	-0.01	0.03	-0.04	-0.06	0.21	0.13	-0.04
N0	0.00	0.00	0.00	-0.06	0.00	0.00	0.00	0.00
N1	0.03	-0.01	-0.01	-0.01	-0.02	-0.15	0.05	-0.03
N2	0.20	0.07	0.03	0.08	0.04	-0.04	0.13	-0.04
N3	0.21	0.02	-0.08	0.03	0.12	0.08	0.21	0.04
N4	-0.12	0.00	0.00	0.02	0.01	0.11	0.08	0.04
N5	0.00	0.03	0.01	-0.02	0.03	0.03	0.17	0.00
N6	-0.01	0.03	-0.02	-0.02	0.00	-0.01	0.16	0.02
N8	0.00	0.07	0.00	0.00	-0.01	0.01	0.00	0.01
N9	0.00	-0.01	-0.01	0.14	-0.01	0.00	0.04	0.01
N10	0.22	-0.02	-0.06	0.00	0.01	-0.01	0.12	0.00
N11	0.26	0.04	-0.01	0.03	0.06	0.03	0.20	0.37
N12	-0.01	0.01	0.07	0.00	-0.01	-0.01	0.00	0.00
N13	0.01	0.00	0.07	0.00	0.00	0.00	0.00	0.00
N14	0.00	0.00	-0.01	0.00	0.00	0.12	0.00	-0.01
N15	0.01	0.00	0.05	0.00	0.00	0.00	0.00	0.00
N16	-0.01	0.01	0.07	0.00	-0.01	0.01	0.01	0.00
N17	0.01	0.13	-0.02	0.01	0.01	0.00	0.02	0.00
O1	-0.02	0.02	0.12	0.02	0.00	0.01	0.02	0.00
O2	0.00	0.08	0.00	0.00	0.00	0.00	-0.01	0.00
O3	-0.01	0.01	-0.01	0.00	-0.13	0.03	0.04	0.02
O4	-0.04	0.03	-0.03	0.12	0.01	0.02	0.09	0.02
O5	0.00	0.00	0.00	-0.11	0.00	0.00	-0.08	0.00
P	-0.12	0.03	0.00	0.00	0.05	0.03	0.08	0.02
S1	0.00	-0.01	-0.08	0.01	0.01	0.00	0.00	-0.01
S2	-0.03	0.01	0.12	0.01	0.02	-0.01	0.05	0.02
S3	-0.01	0.01	0.03	-0.14	-0.01	0.02	0.06	-0.01
S4	0.00	-0.01	-0.01	0.11	0.00	0.01	0.05	-0.02
S5	-0.01	0.14	0.04	0.00	0.00	0.01	0.03	-0.01
S6	-0.05	0.02	0.14	0.05	0.03	-0.01	0.09	-0.06
C1	0.03	0.02	0.01	-0.04	0.16	0.03	0.08	0.00
C2	0.14	0.08	0.02	-0.04	-0.02	0.07	0.16	-0.16
C3	0.18	0.03	0.14	0.04	0.10	0.11	0.22	-0.04
C4	0.15	0.09	0.01	0.07	0.08	0.01	0.17	0.06
C5	0.03	0.09	0.03	-0.02	0.02	-0.01	0.18	-0.46

Quando se trabaja con la matriz de covarianzas resulta complicado establecer que valores con altos y cuales no. Es por eso que es útil calcular los coeficientes de correlación bivariados entre las componentes principales y las modalidades de las variables en estudio. En la tabla 3.62 se encuentran dichos coeficientes para el ejercicio en cuestión.

Conclusiones

El siguiente trabajo científico consiste en un Análisis Estadístico de las Sustracciones de Vehículos en la Ciudad de Guayaquil ocurridos en el año de 1999. El objetivo de esta investigación es encontrar las relaciones que existen entre las variables que intervienen en el robo de un vehículos entre ellas podemos mencionar al marca, el color, el sitio del robo, la hora, el día, etc.

Hemos encontrado que

Los siguientes puntos constituyen un resumen de lo que hemos podido extraer a partir de nuestro análisis:

1. Durante los días **jueves y viernes** ocurren la mayor cantidad de robos de la semana, pero estos días no tienen diferencias significativas respecto al lunes, martes y miércoles. Sin embargo durante el **sábado y el domingo** ocurren la menor cantidad de sustracciones distinguiéndose estadísticamente de las demás. Además descubrimos que diariamente suceden alrededor de 8 robos lo cual resulta una cifra alarmante. Otro punto a destacar es el haber concluido que los días feriados no influyen de alguna manera en la cantidad de robos diarios.

2. Los meses de **Septiembre, Octubre y Noviembre** son los de mayor riesgo acumulando entre ellos un 32% del total de robos mensuales (7-12 robos diarios) seguidos de los meses de Enero, Febrero, Julio y Agosto (6-10), mientras que durante los meses comprendidos **desde Marzo hasta Junio y el mes de Diciembre** se sustraen vehículos en menor cantidad (4-8).

3. **Entre las 7 y las 10 de la noche** el índice de robos de vehículos encuentra su cota superior mientras que las horas de riesgo medio pertenecen a los intervalos de [7,9)horas, [10,13)horas, [18,21)horas y [22,24)horas; y las horas de menor riesgo se distribuyen en los intervalos **[0,7)horas, [9,10)horas, [13,18)horas**; de aquí se concluyó que el riesgo se torna mayor durante las horas de llegar al trabajo o salir del trabajo (al final del día o a la hora del almuerzo), y menor durante las horas en las que regularmente se permanece dentro de la oficina o dentro del hogar.

4. Encontramos que la marca con mayor riesgo de ser sustraída es **Chevrolet**, a la que calificamos como marca de “riesgo muy alto”, a esta marca le siguen las marcas Nissan Suzuki y Toyota

a las cuales definimos como de riesgo medio dejando todas las demás marcas dentro del grupo de riesgo bajo.

5. Vimos también que el tipo de vehículo preferido por la delincuencia es el **auto** con un 54%, seguido de la camioneta con 33%, del campero con 11% y dejando a los **vehículos pesados** rezagados con un 2%. Concluimos que mientras más liviano el vehículo, mayor es la probabilidad de ser sustraído.
6. Los colores favoritos fueron el **rojo, blanco**, gris, verde y azul; agrupando los dos primeros un 35% y los tres últimos un 41%. Todos los demás colores agrupan el 25% restante.
7. Descubrimos que contrario a lo que se pensaba, la proporción de vehículos sustraídos mientras se encontraban **estacionados** supera a las demás ya que cuenta con un 56% del total, seguido por la modalidad circulando (mientras el vehículo se encontraba en movimiento) que agrupa el 30% del total, dejando a las demás circunstancias con un 14% de representatividad.
8. Encontramos también que en la mayoría de los casos se mencionaron a **2** ó 3 delincuentes como autores del robo (con

37 y 34% respectivamente), mientras que los casos en los que se mencionaron **1 o más de 4** representan la minoría.

9. También descubrimos que la mayoría de las sustracciones suceden por medio de un **asalto**, es decir que en la mayoría de los casos el delito sucede en presencia de alguna persona que tenga interés en que el vehículo no sea sustraído.
10. Existe un riesgo potencial de ser herido de gravedad cuando sucede la sustracción de un vehículo, ya que en un 98% de los casos se pudo identificar por lo menos un **arma de fuego** como medio del robo.
11. Se concluyó que por lo general los sustractores se transportan **vía pie**, ya que, en un 64% de los casos no se pudo identificar algún vehículo relacionado con el robo.
12. Por último, descubrimos que los sectores de riesgo potencial pertenecen al centro y al norte de la ciudad (**C3, C4 y N11**), los de riesgo medio son C2, C5, N2, N3 y N10, todos los demás sectores de la ciudad se pueden considerar de riesgo bajo ya que al pertenecer 27 sectores a este grupo y representar el 32% del total anualmente estarían sucediendo

aproximadamente $2764(32/100)(1/27) = 33$ robos anuales por zona.

Mediante el análisis bivariado hemos resuelto algunas interrogantes, debemos destacar que hubiese sido imposible tratar de obtener conclusiones confiables sin realizar dicho análisis. Examinemos los resultados derivados de esta análisis y resumamos lo más importante:

13. Descubrimos que el mayor riesgo de robo ocurre los días martes de 19h00 hasta antes de las 20h00 (de 81 a 100), seguido por los días de jueves a domingo a la misma hora (de 61 a 80) estas dos combinaciones entre lugar y hora son las de mayor riesgo; también encontramos que todos los días de la semana a partir de las 20h00 hasta antes de las 19h00 ocurren mucho menos sustracciones que en los demás casos (de 1 a 20, excepto de lunes a jueves de 7h00 a 8h59).

14. También corroboramos que dentro de los robos más comunes se encuentran los autos y las camionetas Chevrolet; y los autos Suzuki.

15. Determinamos que la marca Chevrolet tiene una alta incidencia de robos en las zonas de mayor riesgo (C2, C3, C4, N2, N3,

N10 y N11).; la tabla 3.60 ilustra las zonas (de acuerdo a la marca) en las que el número de vehículos robados es mayor.

Tabla 3.60

	C2	C3	C4	C5	N10	N11	N2	N3	O4	S6
Chevrolet	○	○	○		○	○	○	○		
Datsun										
Fiat			○							
Ford										
Mazda	○									
Mitsubishi							○			
Nissan			○			○				
Skoda										
Suzuki			○			○		○		
Toyota						○				
VW										

Referencia.- ○: Riesgo Alto; ◐: Riesgo Medio; **Vacías**: Riesgo Bajo

16. Pudimos concluir que en las zonas de mayor índice de robos (C2, C3, C4, C5, N2, N3, N10 y N11) intervienen de 1 a 4 sustractores.

17. Entre otras cosas pudimos determinar que en las zonas rojas ocurren mas sustracciones donde no se identifica un vehículo como medio del robo.

18. En concordancia con el análisis univariado, los vehículos robados al estar estacionados o circulando han sido sustraídos mediante asalto y robo.

19. También existe concordancia entre las horas y las zonas pico, ya que en los momentos en que han ocurrido una mayor cantidad de sustracciones (19H00-21H59) los vehículos se han encontrado en las zonas de mayor riesgo (C2, C3, C4, N2, N3, N10, N11).

La tabla 3.59 ilustra claramente la distribución de la pareja Marca Vs. Color.

Recomendaciones

1. Indiscutiblemente debemos tener mucha precaución y cautela al conducir nuestro vehículo por las calles de Guayaquil, ya que en el momento menos pensado podemos ser víctimas de un asalto o un robo, no solamente vehicular sino de cualquier tipo.
2. Debemos procurar no transitar las calles de los sectores de alto riesgo de robo (Centro, Sauces, Alborada, etc) en horas pico (de 7 a 10 de la noche) para no aumentar la probabilidad de ser asaltado o robado.
3. Nunca deje su vehículo encendido o con alguna puerta abierta mientras realiza una diligencia, ni por un segundo (es el detalle que más aprovechan los delincuentes).
4. Salga de paseo tranquilamente (claro está, dejando asegurada su casa) durante los feriados y fines de semana ya que los roba-carros también descansan estos días.
5. Si la marca de su carro es Chevrolet, Nissan, Suzuki o Toyota, ponga atención en los lugares y las horas que transita por la ciudad, estas marcas son de riesgo potencial. Lo mismo sucede con los vehículos livianos (tipo automóvil) y con los vehículos color rojo, blanco, gris, verde y azul.

6. Evite responder un asalto con agresión, puede ser peligroso para su vida, ya que en la mayoría de los casos los delincuentes se encuentran armados y listos para cualquier cosa y por lo general los asaltos son realizados por más de una persona.

BIBLIOGRAFÍA

1. KACHIGAN SAM K., Multivariate Statistical Analysis. Segunda Edición. Editorial Radius Press. 1989.
2. ILDEFONSO GRANDE ESTEBAN, ABASCAL FERNÁNDEZ ELENA, Métodos multivariantes para la Investigación Comercial. Primera Edición. Editorial Ariel. 1989.
3. ZIKMUND WILLIAM G., Investigación de Mercados. Sexta edición. Editorial Prentice Hall, 1998.
4. MALHOTRA NARRES. Investigación de Mercados. Segunda edición. Editorial Prentice Hall, 1997.
5. MIGUEL S., BIGNE E., LEVY J. P., CUENCA A. C., MIGUEL Ma. J.. Investigación de Mercados. Primera edición. Editorial Mc Graw Hill, 1997.
6. DILLON W, MADDEN T., FIRTLE N.. La Investigación de Mercados Entorno de Marketing. Tercera edición. Editorial Mc Graw Hill, 1997.
7. NRESH K. MAA. Investigación de Mercados un Enfoque Práctico. Segunda edición. Editorial Prentice Hall, 1997.
8. MITRA AMITAVA, Quality Control and Improvement. Segunda edición. Editorial Prentice Hall, 1998.

9. SPIEGEL MURRIA R., Probabilidad y Estadística. Segunda edición. Editorial Mc Graw Hill, 1991.
10. TOMALA CHRISTIAN, Análisis Estadístico de las Causas de Accidentes de Tránsito en la Ciudad de Guayaquil (Tesis, Instituto de Ciencias Matemáticas Escuela Superior Politécnica del Litoral, 2000)
11. CONTRERAS HELIANA, Estadísticas electorales del Ecuador Periodo de 1978-1998 (Tesis, Instituto de Ciencias Matemáticas Escuela Superior Politécnica del Litoral, 1999)
12. RIVADENEIRA EDUARDO, PATIÑO MARIO Guía para Elaboración de Tesis de Grado. Primera Edición. Escuela Superior Politécnica del Litoral. 1999.
13. <http://www.monografias.com/trabajos/tesisgrado/tesisgrado.shtml>
14. CODIGO PENAL
15. CODIGO CIVIL