

# **ESCUELA POLITÉCNICA NACIONAL**

**FACULTAD DE CIENCIAS**

**ESTUDIO DE VICTIMIZACIÓN EN EL DISTRITO  
METROPOLITANO DE QUITO MEDIANTE ANÁLISIS DE  
DATOS CATEGÓRICOS.**

**PROYECTO PREVIO A LA OBTENCION DEL TITULO DE MASTER EN  
ESTADÍSTICA APLICADA**

**MAURICIO ENRIQUE ABRIL DONOSO**

(mauricioabrildo@yahoo.com)

**DIRECTOR: MSc. ALFONSO JUAN ELÍAS CASTRO BALAREZO**

(acastro@epn.edu.ec)

**Quito, Julio de 2009**

## DECLARACION

Yo, Mauricio Enrique Abril Donoso, declaro que el trabajo aquí descrito es de mi autoría; que no ha sido previamente presentada para ningún grado o calificación profesional; y, que he consultado las referencias bibliográficas que se incluye en este documento.

La Escuela Politécnica Nacional, puede hacer uso de los derechos correspondientes a este trabajo, según lo establecido por la Ley de Propiedad Intelectual, por su Reglamento y por la normatividad institucional vigente.

-----  
Ing. Mat Mauricio Enrique Abril Donoso

## CERTIFICACION

Certifico que el presente trabajo fue desarrollado por Mauricio Enrique Abril Donoso, bajo mi supervisión.

-----

MSc. Alfonso Juan Elías Castro Balarezo

**DIRECTOR DE PROYECTO**

## **AGRADECIMIENTO**

Deseo expresar mi gratitud a la Facultad de Ciencias, así como también al MSc. Alfonso Castro, mi maestro en la Escuela Politécnica Nacional, tanto en pregrado como en el posgrado.

## **DEDICATORIA**

*A la memoria de mi padre Néstor Enrique †.*

## CONTENIDO

DECLARACIÓN.....	II
CERTIFICACIÓN.....	III
CONTENIDO.....	VI
RESUMEN.....	VIII
PRESENTACIÓN.....	IX
<b>CAPITULO 1 INTRODUCCION.....</b>	<b>1</b>
1.1 CONSIDERACIONES PREVIAS.....	1
1.2 PRESENTACIÓN DEL PROBLEMA.....	2
1.3 FORMULACIÓN Y SISTEMATIZACIÓN DEL PROBLEMA.....	4
1.4 PRESENTACIÓN E INTERPRETACIÓN LEGAL DE LOS CINCO TIPOS DE VIOLENCIA QUE SE ESTUDIARÁN.....	5
<b>CAPITULO 2 ANÁLISIS ESTADÍSTICO DESCRIPTIVO DE LOS DIFERENTES TIPOS DE VIOLENCIA.....</b>	<b>14</b>
2.1 METODOLOGÍA PARA EL ANÁLISIS ESTADÍSTICO.....	14
2.2 ANÁLISIS DE MUERTES VIOLENTAS.....	15
2.3 ANÁLISIS DE DELITOS CONTRA LOS VEHÍCULOS.....	31
2.4 ANÁLISIS DE DELITOS CONTRA LA PROPIEDAD.....	41
2.5 ANÁLISIS DE VIOLENCIA INTRAFAMILIAR.....	51
2.6 ANÁLISIS DE MALTRATO INFANTIL.....	59
<b>CAPITULO 3 CONSTRUCCIÓN DE ÍNDICES MEDIANTE MODELOS LOGÍSTICOS BINARIOS Y MULTINOMIALES.....</b>	<b>68</b>
3.1 MODELO LOGÍSTICO MULTINOMIAL.....	69
3.2 MODELO LOGÍSTICO PARA MUERTES VIOLENTAS.....	72
3.3 MODELO LOGÍSTICO PARA DELITOS CONTRA LOS VEHÍCULOS.....	76
3.4 MODELO LOGÍSTICO PARA DELITOS CONTRA LA PROPIEDAD.....	79
3.5 MODELO LOGÍSTICO PARA VIOLENCIA INTRAFAMILIAR.....	83
3.6 MODELO LOGÍSTICO PARA MALTRATO INFANTIL.....	88
<b>CAPITULO 4 CIONSTRUCCIÓN DE INDICES MEDIANTE MODELOS LOG LINEALES... </b>	<b>92</b>
4.1 MODELOS LOG LINEALES EN TEBLAS DE CONTINGENCIA.....	92
4.2 MODELO LOG LINEAL PARA MUERTES VIOLENTAS.....	101
4.3 MODELO LOG LINEAL PARA DELITOS CONTRA LOS VEHÍCULOS.....	105
4.4 MODELO LOG LINEAL PARA DELITOS CONTRA LA PROPIEDAD.....	109
4.5 MODELO LOG LINEAL PARA VIOLENCIA INTRAFAMILIAR.....	113
4.6 MODELO LOG LINEAL PARA MALTRATO INFANTIL.....	117

<b>CAPITULO 5</b>	<b>COMPARACIÓN DE LOS INDICES CONSTRUÍDOS CON LOS DOS METODOS.....</b>	<b>122</b>
5.1	COMPARACIÓN ESTADÍSTICA ENTRE LOS MODELOS.....	122
5.2	COMPARACIÓN DESCRIPTIVA DE LOS FENÓMENOS ESTUDIADOS...	125
5.3	ELECCIÓN DE LOS MEJORES MODELOS E ÍNDICES.....	125
<b>CAPITULO 6</b>	<b>CONCLUSIONES Y RECOMENDACIONES.....</b>	<b>127</b>
6.1	CONCLUSIONES.....	127
6.2	RECOMENDACIONES.....	129
	<b>REFERENCIAS BIBLIOGRAFICAS.....</b>	<b>131</b>
	<b>ANEXOS</b>	
<b>Anexo A</b>	<b>MÉTODOS EXPLORATORIOS.....</b>	<b>132</b>
A.1	MÉTODOS UNIVARIANTES.....	132
A.2	MÉTODOS MULTIVARIANTES.....	133
A.3	MODELOS DE ANÁLISIS DE VARIANZA.....	136
<b>Anexo B</b>	<b>ANÁLISIS ESTADÍSTICO CON DATOS PERDIDOS.....</b>	<b>140</b>
B.1	DISEÑO DE DATOS PÉRDIDOS.....	140

## RESUMEN

Una dificultad que se tiene en la modelación de eventos que se da en las ciencias sociales cuando se estudia la inseguridad que se produce y percibe en los ciudadanos de Quito, es que la gran mayoría de variables que se tiene en los fenómenos de violencia y delincuencia son de tipo cualitativo. Una metodología que nos permite modelar y estudiar asociaciones entre variables categóricas, es el análisis de datos categóricos, con la construcción de modelos log lineales y modelos logísticos multinomiales.



## PRESENTACION

El estudio de datos categóricos es un tipo de análisis estadístico orientado a modelar e inferir variables cualitativas y categóricas. Su aparición e importancia se da a partir de 1960, desarrollándose hasta la actualidad.

La importancia de trabajar con esta metodología radica en que se puede modelar estadísticamente variables dependientes cualitativas o categóricas, independientemente del tipo de variable que sean las explicativas. Las distribuciones que se utiliza en este tipo de modelación de las variables explicativas son Binomial, Multinomial y Poisson. Las principales aplicaciones se encuentran en las ciencias sociales, en ramas como la Sociología, Antropología, Derecho y Economía Social, entre otras.

En el proyecto que se plantea, se va a trabajar en modelaciones de diferentes fenómenos de violencia, que se suscitan en el Distrito Metropolitano de Quito (DMQ), ya que uno de los principales problemas que aquejan a las grandes ciudades y más aun si es la capital de un País es la inseguridad y la violencia de sus ciudadanos, para esto se contará con información oficial que es recabada por el Observatorio Metropolitano de Seguridad Ciudadana, entidad que depende del Municipio del DMQ, y que se encarga de producir información estadística descriptiva de los hechos de violencia que se suscitan en la ciudad.

El proyecto de Titulación se divide como sigue: En el Capitulo 1 se da una introducción, de los aspectos considerados para decidir la temática y la importancia en la ciudadanía y autoridades para tener una comprensión y entendimiento real del fenómeno de la violencia, además se detallan los tipos de violencia que serán considerados para la modelación y el tratamiento legal y las consideraciones prácticas de estos tipos de violencia. En el Capitulo 2 se muestra la descripción de todas las variables

que están asociadas a los tipos de violencia que se estudiarán; se muestra también el tratamiento desde la estadística, la cual contempla un análisis exploratorio descriptivo univariante, por medio de frecuencias relativas de cada una de las variables e histogramas, se muestra también un análisis multivariante, para determinar las relaciones entre variables y sus categorías, esto se realiza mediante el análisis de correspondencias múltiples, y el análisis de la varianza. En el Capítulo 3 se realiza una descripción teórica de los modelos logísticos multinomiales (para modelos binomiales, se considera la variable respuesta sólo dos categorías), las estimaciones de los parámetros, validación del modelo estimado e interpretación del modelo; la modelación se realiza para los cinco tipos de violencia que se estudia en este trabajo. El Capítulo 4 presenta la teoría de los modelos log lineales, que son aplicados a tablas de contingencia, es decir que en estos modelos la variable respuesta o dependiente son las frecuencias que se encuentran en las casillas de la tabla de contingencia que se forma del cruce de las variables categóricas. En el Capítulo 5 se realizan comparaciones entre los modelos estimados con las dos metodologías regresión: logística multinomial y modelo log lineal; las comparaciones son en función de los estimadores de validación de los modelos estimados. También se realiza una comparación descriptiva entre los modelos, básicamente para la propuesta de política, ya que por el mismo tratamiento legal de los diferentes tipos de violencia, es conveniente tener modelos con una variable dependiente y modelos que consideren las frecuencias en las tablas de contingencia. Finalizamos este trabajo con el Capítulo 6 en el cual se enuncian algunas conclusiones y recomendaciones sobre el comportamiento de la violencia, grupos vulnerables y algunas acciones que pueden ayudar en la reducción y prevención de los hechos delictivos.

# Capítulo 1.

## Introducción

### 1.1.- Consideraciones previas.-

Es importante para el desarrollo de una sociedad, con visión progresista, garantizar seguridad y generar en sus habitantes una percepción de tranquilidad y confianza en el sistema de justicia, sistema que se debe encargar del control, prevención y juzgamiento de todos los hechos delictivos que se pueden producir en una sociedad.

Para que el sistema de justicia entregue información confiable, debe contar con datos estadísticos válidos y confiables, lo que permitirá evaluar técnicamente su funcionamiento y poder optimizar los recursos e ir mejorando el servicio a la ciudadanía, información que hasta ahora es escasa e incluso inexistente.

Si bien la falta de información, o la no correcta interpretación de los datos que se tienen en instituciones encargadas de controlar la delincuencia en nuestro País y Ciudad, no permite tener una planificación apropiada y destinar recursos hacia acciones que ayuden a reducir óptimamente los niveles de violencia y con esto garantizar una adecuada convivencia ciudadana.

Otro inconveniente que se tiene por la falencia de información estadística es que no se puede formular y evaluar proyectos de seguridad y convivencia ciudadana de manera adecuada. Además, por la falta de información se tiene falencia y poca claridad para formular políticas de seguridad y convivencia, por parte del Estado y del Gobierno Local, es claro que con la creación de la Subsecretaría de Seguridad, se pretende formular políticas de seguridad en el Ecuador, pero hasta ahora no se observa una correcta interpretación y entendimiento de la seguridad ciudadana y no hay propuestas claras ni en acciones ni en políticas de dicha Subsecretaría.

La generación de violencia que se produce en un país, ciudad, barrio o en cualquier comunidad u organización humana por tan grande o pequeña que

sea, está asociada a diversos factores multicausales que deben ser estudiados, los que no podrán ser combatidos eficaz y eficientemente, si no se tiene un sistema de información oportuno y apropiado, que permita: formular políticas de seguridad, tomar decisiones óptimas como la focalización de recursos y la evaluación de su uso, en base a resultados medibles.

## **1.2.- Presentación del problema.-**

En la actualidad un tema que preocupa mucho a las grandes y pequeñas ciudades es el que se relaciona con la seguridad y convivencia ciudadana, ya que es un derecho de toda persona vivir con la garantía de su seguridad y la de sus familiares. En nuestro País, constitucionalmente, es el Estado el que debe garantizar la seguridad de sus habitantes, por medio de instituciones como: la Policía Nacional, Ministerio Público y los tribunales de justicia, pero lamentablemente poco y nada ha hecho el Estado y las instituciones mencionadas, para mejorar el sistema de justicia y tener niveles de violencia tolerables.

Como se pueden definir los niveles de violencia tolerables, si no se cuenta con estadísticas ni datos que permitan definir indicadores e índices de violencia y evaluar el comportamiento de la violencia en el DMQ.

Además, existe información muy general sobre la violencia que se genera en el País, y si se quiere tener información de violencia desagregada por ciudades es casi imposible conseguirla. Por esto, gobiernos locales como los de Quito y Guayaquil se han apropiado del tema de seguridad y han asignado recursos propios para brindar a sus habitantes niveles de seguridad tolerable y aceptable, además de implementar centros propios de recolección y procesamiento de información de violencia en sus jurisdicciones.

Se enfoca este estudio a la construcción de indicadores de violencia para el DMQ, esta ciudad tiene un sistema confiable de recolección de datos asociados a los diversos hechos de violencia que se suscitan en la ciudad. Esta información está disponible a partir del año 2003 hasta la actualidad.

Notemos que al estudiar la violencia dentro del DMQ, se tienen en primer lugar que establecer las fuentes de información con que se puede contar y la metodología que se va a utilizar para el desarrollo de dicho estudio. Además, se tiene que decidir qué temas relacionados con la seguridad ciudadana se van a estudiar.

La fuente de información que se ha definido para la elaboración de este proyecto es el Observatorio Metropolitano de Seguridad Ciudadana, el cual cuenta con información concerniente a hechos de violencia, que se producen en todo el DMQ.

La información que recopila el Observatorio se centra en los siguientes tipos de violencia:

**Tabla 1.2.1.- Tipos de violencia que se recaba datos en el Observatorio de Seguridad.**

<b>TIPO DE VIOLENCIA</b>	<b>DESCRIPCIÓN</b>	<b>FUENTE DE DATOS</b>	<b>TRATAMIENTO LEGAL</b>
<b>Muertes Violentas</b>	Se consideran muertes provocadas: homicidios y suicidios; y muertes no provocadas: muertes en accidentes de tránsito y muertes accidentales.	Departamento Médico Legal de la Policía Nacional del Ecuador.	Se trata dentro de lo legal en base al Código Penal y en base a un problema de salud Pública.
<b>Delitos a los vehículos</b>	Se considera cuando la totalidad del vehículo es robado o hurtado.	Ministerio Público de Pichincha.	Se trata en base a lo estipulado en el Código de Procedimiento Penal.
<b>Delitos Contra la Propiedad</b>	Son objeto de delito a la propiedad: las personas, los domicilios, los bancos, los locales comerciales, empresas, etc.	Ministerio Público de Pichincha.	Se trata en base a lo estipulado en el Código de Procedimiento Penal.
<b>Violencia Intra-familiar</b>	Es la violencia física, psicológica y sexual que puede sufrir una mujer o su familia.	Comisarías de la Mujer y La Familia adjuntas al Ministerio de Gobierno.	Ley 103 de la Mujer y la Familia.
<b>Maltrato Infantil</b>	Es la violencia física, psicológica y sexual que puede sufrir un niño, niña o adolescente menor a los 18 años.	DINAPEN, Policía especializada en niños, niñas y adolescentes.	Código de la Niñez y la Adolescencia.

Como se observa en la tabla 1.2.1, se va a trabajar con cinco tipos diferentes de violencia que se presentan en el DMQ; es claro también que las instituciones que generan los datos solo en los delitos a la propiedad y los vehículos es la misma, en los otros tipos de delitos la fuente es diferente. Todas las fuentes de datos envían su base al Observatorio de Seguridad, el cual almacena, procesa y analiza dichos datos y los transforma en información; es posible que algún hecho de violencia genere otro hecho de violencia pero no necesariamente existe una generación causal entre ellos.

Es por esto que al ser diferentes los tipos de violencia, las variables que los describen serán también diferentes. Esto crea la necesidad de trabajar la construcción de los indicadores de violencia diferenciado a cada tipo de violencia que se describió y luego de esto poder relacionarlos entre sí, y construir un indicador general que describa la violencia que se produce en nuestra ciudad.

Por este motivo es necesario construir índices de violencia que permitan mostrar la información de violencia a todas las personas e instituciones interesadas en los temas sobre seguridad y convivencia ciudadana. Los índices que se quiere construir deben ser confiables, entendibles y ágiles, lo que garantizará que la información estadística que se proporcione sea confiable, y útil para la toma de decisiones encaminadas a mejorar la seguridad y convivencia de todos los habitantes del DMQ.

### **1.3.- Formulación y sistematización del problema.-**

Como se mencionó anteriormente el estudio de la violencia en el DMQ, se lo tiene que realizar para cada tipo de violencia debido a la heterogeneidad que existe entre ellas, porque cada tipo de violencia se produce de manera independiente y porque las variables asociadas a cada una de ellas son también diferentes.

Las poblaciones donde se manifiestan los diferentes tipos de violencia también son diferentes ya sea por edades, educación, lugar de residencia, etc. Una vez

que se estudie el problema individual (por tipo de violencia) se procederá a realizar un estudio general de la violencia en el DMQ, y presentar un diagnóstico general de la violencia que se tiene en la ciudad. Luego que se realiza el diagnóstico para la ciudad, se deben plantear acciones en los campos políticos y de control, a mediano y largo plazo, para reducir los niveles de violencia que se tiene en el DMQ. Estas acciones deben influir positivamente en la percepción de la ciudadanía sobre su seguridad y su entorno. Con esto se logra influenciar de forma directa en la confianza y tranquilidad de vivir en nuestra Ciudad.

Notemos que se debe investigar y estudiar la percepción de la violencia que tienen los ciudadanos del DMQ, para identificar que la acción que se toma es la que brinda mayor confianza y por ende un mayor sentimiento de seguridad en la ciudadanía; esto también deben reflejar las estadísticas que se manejen para medir el comportamiento de la seguridad e inseguridad en nuestra Ciudad, por parte de las instituciones que proporcionan la información estadística. La información que proporcionen los índices tiene que ir encaminada a ayudar a la toma de decisiones de las instituciones que se encargan del control y prevención de la violencia en la ciudad.

Debido a que la mayoría de la información que proporciona el Observatorio es de tipo cualitativo, se ha decidido utilizar el análisis de datos categóricos y particularmente los modelos logísticos binomial y multinomial, para la construcción de los índices.

## **1.4.- Presentación e interpretación legal de los cinco tipos de violencia que se estudiarán.**

### **1.4.1.- Descripción legal de muertes violentas.-**

Las muertes violentas o por causas externas, se las puede enmarcar en dos ramas diferentes; una, la que trata la medicina (según la OPS), y la otra dentro del derecho y la legalidad (Según el Código de Procedimiento Penal). Dentro

de la medicina, se clasifican en muertes provocadas, esto es los homicidios y suicidios, y las muertes no provocadas, que son las muertes en accidentes de tránsito y accidentales; el tratamiento que da la medicina a las muertes violentas es un tratamiento epidemiológico, y con esto busca intervenir y reducir en un tiempo mucho menor al que se reduciría al tratarlo como una epidemia simple. El mayor énfasis, que pone la medicina, es en accidentes de tránsito, ya que es la causa principal de muerte a nivel mundial.

Según el ámbito legal, los homicidios son los únicos tratados como un procedimiento legal penal. Las muertes en accidentes de tránsito son tratados según lo estipulado en el código de tránsito y no se considera como un acto criminal. Los suicidios y muertes accidentales no tienen un tratamiento penal ni supone que fuere generado por un criminal, y se realiza la autopsia al cadáver para descartar que fuese provocada por un acto delictivo.

#### **1.4.1.1.- Homicidios.-**

El Código Penal hace referencia solo al homicidio, el cual está estipulado en el art.448.

Art. 448.- “[Presunción de dolo en homicidio, heridas, golpes o lesiones].- se califican de voluntarios el homicidio, las heridas, los golpes y lesiones, mientras no se pruebe lo contrario, o conste de falta de intención por las circunstancias del hecho, calidad y localización de las heridas, o de instrumentos con que se hicieron”.

Es importante señalar que dentro de los delitos contra la vida y particularmente además de los homicidios se define también: Asesinato, Parricidio, Infanticidio, Instigación al suicidio, homicidio preterintencional, homicidio preterintencional por suministro de sustancias, homicidio agravado en razón del parentesco, Homicidio intencional, Muerte producida en riña, Homicidio causado por deportista.



### 1.4.1.2.- Suicidios.-

Ahora tratemos de entender el suicidio, enfoquémonos en una interpretación del fenómeno. Los suicidios al igual que los homicidios son muertes violentas provocadas y que se generan en gran medida como consecuencia del desarrollo poblacional de los pueblos.

Delimitemos en primer lugar históricamente al suicidio, pues si bien como toda manifestación humana ha sufrido cambios en el imaginario colectivo, este fenómeno ha pasado de ser, si tomamos como ejemplo el occidente cristiano, un acto que podía ratificar la piedad de los mártires perseguidos del cristianismo primitivo a ser una conducta condenable por la Iglesia Católica.

Una segunda demarcación está relacionada a la cultura, hay sociedades en las que el suicidio es considerado un acto sublime y que glorifica a quien lo realiza. El término suicidio se dice que fue acuñado en el siglo XVIII y proviene de la fusión latina de *sui* –sí mismo- y *cidius* –matar-.

En la actualidad el término suicidio está referido sólo al acto consumado de matarse a sí mismo, y se le entiende como la culminación catastrófica de la conducta suicida que engloba a todas las manifestaciones de la intención de muerte. Según Pérez y Reytor (1995) las diferentes formas de conducta suicida son:

“Acto suicida.- Todo hecho por el que un individuo se causa a sí mismo una lesión, cualquiera sea el grado de intención letal y del conocimiento del verdadero móvil.”

“Parasuicidio.- Acto no mortal en el que una persona, de forma deliberada, se autolesiona o ingiere un medicamento en dosis superior a la prescrita o reconocida como terapéutica” (Sarason y Sarason, 1999).

“Conductas autodestructoras directas o consientes.- Corresponden a las tentativas de suicidio y al suicidio consumado.”

“Conductas autodestructivas indirectas o inconscientes.- Tiene un carácter insidioso y prolongado y no son reconocidos como actos suicidas.”

“Autolesión intencionada.- Término propuesto para conceptualizar el intento suicida como una conducta de afrontamiento.”

“Ideación suicida.- Comprende desde la idea fugaz de la dificultad de vivir a la idea suicida transitoria, prolongada, permanente, impulsiva o planeada.”

#### **1.4.1.3.- Muertes en accidentes de tránsito.-**

Históricamente se consideraba que los accidentes de los vehículos de motor eran sucesos azarosos y ocurrían a los demás como consecuencia inevitable del transporte. En particular el término “accidente” puede dar la impresión de inevitabilidad e impredecibilidad, es decir, de suceso imposible de controlar. Pero los choques causados por el tránsito son, por el contrario, sucesos que cabe someter al análisis racional y acciones correctoras. Ahora todas las investigaciones que se realizan sobre accidentalidad vial se enfocan a la prevención de lesiones y muertes, ya que la Organización Mundial de la Salud<sup>1</sup> (OMS), las califican como lesiones no provocadas.

En la actualidad las lesiones causadas por el tránsito se enfocan como un problema de salud pública de primera magnitud, y no como un mero derivado de la movilidad de los vehículos. El sector sanitario se beneficiaría mucho de una mejor prevención de dichas lesiones porque se reducirían las hospitalizaciones y la gravedad de los traumatismos. También se beneficiaría si, tras garantizarse condiciones más seguras para los peatones y ciclistas, un mayor número de personas adoptaran el hábito saludable de caminar o ir en bicicleta sin temor por su seguridad.

El sector sanitario tiene importantes funciones que desempeñar en el tratamiento de la seguridad vial:

- Descubrir características de las lesiones causadas por el tránsito; para esto se basa en la recopilación de datos que permitan describir el fenómeno.

---

<sup>1</sup> La Organización Mundial de la Salud fue creada en 1948 como organismo especializado de las Naciones Unidas para que actuara como autoridad directiva y coordinadora en los asuntos sanitarios y de salud pública a nivel internacional.

- Investigar causas de choques y lesiones causadas por el tránsito.
- Explorar maneras de prevenir y reducir la gravedad de las lesiones causadas por el tránsito.
- Ayudar a poner en práctica diversos entornos intervenciones aparentemente prometedoras, especialmente en el área del comportamiento humano.
- Recomendar a las instituciones involucradas en el control del tránsito y transporte, que las lesiones son un problema de primera magnitud y es importante que adopten mejores estrategias en materia de seguridad vial.

Ahora si se enfoca la seguridad vial como un problema de equidad social, muchos estudios demuestran que los choques de vehículos de motor tienen un impacto desproporcionado en los sectores pobres y vulnerables de la sociedad (Nantulya 2003 y Laflamme 2000). Las personas más pobres representan la mayoría de las víctimas y carecen de apoyo permanente en caso de las lesiones de larga duración.

#### **1.4.1.4.- Muertes accidentales.-**

Se define como muerte accidental aquella que se produce como resultado de la interacción entre un sujeto y un agente externo sin que exista la intención de causar daño o lesión, o no se conociere el resultado de dicha interacción.

Las muertes accidentales al igual que las muertes en accidentes de tránsito son consideradas como muertes violentas no intencionales. Este tipo de muerte fueron también asociadas a factores probabilísticos y se consideraban que dependían del azar, pero a medida que se aumentaron los mecanismos de recolección de información y con la tecnificación de los médicos forenses, se observa que la mayoría de estas muertes se da en lugares de trabajo y se relaciona directamente con la imprudencia y falta de seguridad y prevención en actividades peligrosas. El accidente de trabajo se define como todo suceso repentino que sobrevenga por causa o con ocasión del trabajo y que produzca en el trabajador una lesión orgánica, una perturbación funcional, una invalidez o la muerte. También es accidente de trabajo aquel que se produce durante la

ejecución de órdenes del empleador o durante la ejecución de una labor bajo su autoridad, aún fuera de lugar y de las horas de trabajo. Igualmente, se considera accidente de trabajo el que se produzca durante el traslado de los trabajadores desde su residencia a los lugares de trabajo o viceversa, cuando el transporte lo suministre el empleador.

No obstante, las muertes accidentales no están ligadas solo con las que se producen en el trabajo, sino que se investigarán otras posibles causas y poblaciones vulnerables.

#### **1.4.2.- Descripción legal de delitos contra la propiedad.**

El código penal califica los delitos contra la propiedad en los siguientes artículos:

Art. 547.- “[Hurto].- Son reos de hurto los que, sin violencia ni amenaza contra las personas, ni fuerza en las cosas, sustrajeren fraudulentamente una cosa ajena, con ánimo de apropiarse.”

“Reforma Art.547.- Además, son considerados como reos de hurto los individuos de reconocida conducta delictuosa, que habitualmente se dedicaren a la comisión de delitos contra la propiedad y que se hallaren registrados como tales en las Oficinas de Seguridad del Estado. La pena para esta clase de delincuentes será de uno a tres años de prisión.”

“Art. 550.- [Robo].- El que, mediante violencias o amenazas contra las personas o fuerza en las cosas, sustrajere fraudulentamente una cosa ajena, con ánimo de apropiarse, es culpado de robo, sea que la violencia tenga lugar antes del acto para facilitararlo, en el momento de cometerlo, o después de cometido para procurar su impunidad.”

#### **1.4.3.- Descripción legal de delitos contra los vehículos.-**

La descripción legal de los delitos a vehículos se enmarca legalmente también en los artículos que rigen el juzgamiento de los delitos a la propiedad. La diferencia se da en el tratamiento en el Ministerio Público, ya que las unidades

de investigación son diferentes. Cabe recalcar que los delitos a vehículos se dan, cuando todo el vehículo es robado, asaltado y robado y hurtado, no se considera en el estudio el robo o hurto de partes o piezas de un vehículo.

#### **1.4.4.- Descripción de violencia intrafamiliar.-**

La violencia intrafamiliar se enmarca dentro de las agresiones a la mujer y su núcleo familiar; este tipo de violencia preocupa ya que es muy poco denunciado, o se denuncia al producirse por más de una ocasión. La falta de denuncia puede deberse al esquema social en que vivimos, es decir que en muchas mujeres de clase media y baja aceptan el maltrato por parte de su cónyuge o conviviente de manera natural.

Para efectos de la modelación nos basaremos en la Ley 103 “Ley contra la violencia a la mujer y la familia”, en sus cuatro primeros artículos.

**Art. 1.- “Fines de esta ley.-** La presente Ley tiene por objeto proteger la integridad física, psicológica y la libertad sexual de la mujer y los miembros de su familia, mediante la prevención y sanción de la Violencia Intrafamiliar y los demás atentados contra sus derechos y los de su familia. Sus normas deben orientar las políticas del Estado y la comunidad sobre la materia.”

**Art. 2.- “Violencia intrafamiliar.-** Se considera Violencia Intrafamiliar toda acción u omisión que consista en maltrato físico, psicológico o sexual, ejecutado por un miembro de la familia en contra de la mujer o demás integrantes del núcleo familiar.”

**Art. 3.- “Ámbito de aplicación.-** Para los efectos de esta Ley se considera miembros del núcleo familiar a los cónyuges, descendientes, ascendientes, hermanos y sus parientes hasta el segundo grado de afinidad.”

La protección de esta Ley será extensiva a los ex cónyuges, convivientes, ex convivientes, a las personas con quienes se mantengan o se haya mantenido una relación consensual de pareja, así como a quienes comparten el hogar del agresor o el agredido.

**Art. 4.- “Formas de violencia intrafamiliar.-** Para los efectos de esta Ley se consideran:

- a) **Violencia física.-** Todo acto de fuerza que cause daño, dolor o sufrimiento físico en las personas agredidas, cualquiera que sea el medio empleado y sus consecuencias, sin considerarse el tiempo que se requiera para su recuperación;
- b) **Violencia psicológica.-** Constituye toda acción u omisión que cause daño, dolor, perturbación emocional, alteración psicológica, o disminución la autoestima de la mujer o del familiar agredido. Es también la intimidación, o amenaza mediante la utilización del apremio moral sobre otro miembro de familia infundiendo miedo o temor a sufrir un mal grave e inminente en su persona o en la de sus ascendientes, descendientes o afines hasta el segundo grado.
- c) **Violencia sexual.-** Sin perjuicio de violación y otros delitos contra la libertad sexual, se considera violencia sexual todo maltrato que constituya imposición en el ejercicio de la sexualidad de una persona, y que la obligue a tener relaciones u otras prácticas sexuales con el agresor o con terceros, mediante el uso de fuerza física, intimidación amenazas o cualquier otro medio coercitivo.”

#### **1.4.5.- Descripción legal maltrato infantil.-**

El maltrato infantil es el primer tipo de violencia que se presenta en una sociedad, es difícil de detectarlo y poco denunciado por las víctimas que en este caso son niños y adolescentes. En muchos casos se ha detectado que son los padres los que propician algún tipo de maltrato en los primeros años de edad de los menores.

Según el código de la niñez y adolescencia define el maltrato como:

**Art 67.- “Concepto de maltrato.-** Se entiende por maltrato toda conducta, de acción u omisión, que provoque o pueda provocar daño a la integridad o salud física, psicológica o sexual de un niño, niña o adolescente, por parte de cualquier persona, incluidos sus progenitores, otros parientes, educadores y personas a cargo de su cuidado; cualquiera sea el medio utilizado para el

efecto, sus consecuencias y el tiempo utilizado para la recuperación de la víctima. Se incluye en esta calificación el trato negligente o descuido grave o reiterado en el cumplimiento de las obligaciones para con los niños, niñas y adolescentes, relativas a la prestación de alimentos, alimentación, atención médica educación o cuidados diarios; y su utilización en la mendicidad.”

“Maltrato psicológico es el que ocasiona perturbación emocional, alteración psicológica o disminución de la autoestima en el niño, niña o adolescente agredido. Se incluye en esta modalidad las amenazas de causar un daño en su persona o bienes o en los de sus progenitores, otros parientes o personas encargadas de su cuidado.”

**Art. 68.- “Concepto de abuso sexual.-** Sin perjuicio de lo que dispone el Código Penal sobre la materia, para los efectos del presente Código constituye abuso sexual todo contacto físico, sugerencia de naturaleza sexual, a los que se somete a un niño, niña o adolescentes, aun con su aparente consentimiento, mediante seducción, chantaje, intimidación, engaños, amenazas, o cualquier otro medio.”

En los dos artículos anteriores se sustenta la modelación en lo relacionado al maltrato infantil.

## **Capítulo 2.**

### **Análisis estadístico descriptivo de los diferentes tipos de violencia.**

#### **2.1.- Metodología para el análisis estadístico.-**

En este capítulo se realizará un estudio estadístico descriptivo de los datos, que van a ser utilizados en la construcción de los índices de violencia. Para esto iniciamos el estudio con métodos estadísticos univariantes por cada variable asociada a cada tipo de violencia; estos métodos son tanto para variables cuantitativas como para variables cualitativas.

Los métodos exploratorios de las variables son muy útiles ya que nos permiten entender el comportamiento de éstas, observar su tendencia y poder comprender la importancia en la descripción del fenómeno de estudio. Debido a que la gran mayoría de las variables asociadas a los tipos de violencia que se han planteado para este estudio y construcción de índices son categóricas, los métodos que se utilizarán son: representación gráfica mediante diagramas de barras de las frecuencias relativas, histogramas de frecuencia y comparaciones mediante diagramas de cajas.

Siguiendo con el análisis se procederá a realizar una recodificación de las variables cuantitativas en cualitativas, y con esto se procederá a realizar un análisis factorial de correspondencias múltiples, por cada tipo de violencia para los que se construirán índices, con el objeto de identificar las variables que intervendrán en la construcción de los índices de victimización.

Este estudio descriptivo univariante y multivariante<sup>2</sup> de las variables asociadas con los tipos de violencia, finaliza con un análisis de varianza, relacionando variables y sus modalidades en diferentes subpoblaciones, lo que nos permite comparar y probar igualdad de medias en dichas subpoblaciones.

---

<sup>2</sup> Para más detalle de los métodos empleados en este estudio ver el Anexo A.



## 2.2.- Análisis de muertes violentas.-

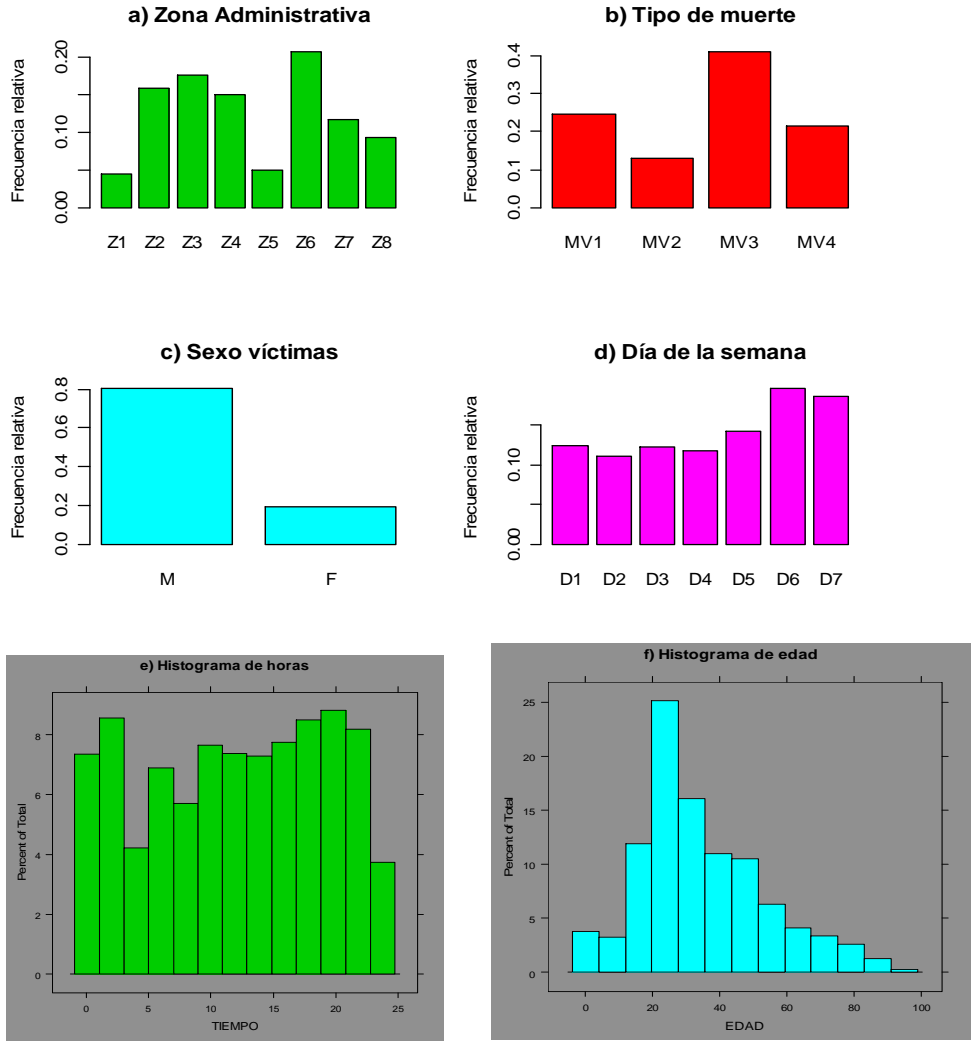
### 2.2.1.- Gráficos de frecuencias relativas.-

Iniciamos la exploración, mostrando las variables que están asociadas a muertes violentas y que se derivan de los protocolos de autopsias, que se realizan en el Departamento Médico Legal.

**Tabla 2.2.1.- Descripción y codificación de variables**

NOMBRE DE VARIABLE	TIPO DE VARIABLE	MODALIDADES	CODIFICACIÓN
ZONA ADMINISTRATIVA	Cualitativa	Calderón Centro Eloy Alfaro La Delicia Los Chillos Norte Quitumbe Tumbaco	Z1 Z2 Z3 Z4 Z5 Z6 Z7 Z8
TIPO DE MUERTE	Cualitativa	Homicidios Suicidios Accidentes de tránsito Accidentales	MV1 MV2 MV3 MV4
SEXO	Cualitativa	Masculino Femenino	M F
DÍA	Cualitativa	Lunes Martes Miércoles Jueves Viernes Sábado Domingo	D1 D2 D3 D4 D5 D6 D7.
HORAS	Cuantitativa continua	0-24	
EDAD (años)	Cuantitativa discreta	1-99	

Figura 2. 2.1.- Frecuencias relativas.



Las figuras a), b), c) y d), son representaciones gráficas de frecuencias relativas de las variables cualitativas, mientras que las figuras e) y f), son histogramas de frecuencias de las dos variables cuantitativas, todas estas variables describen a las muertes violentas. Los resultados que se deducen son los siguientes:

- La zona Centro, Eloy Alfaro, La Delicia y Norte, registran la mayor frecuencia de muertes, mientras que en las administraciones de Calderón y Los Chillos se tiene la menor frecuencia.
- La muerte en accidentes de tránsito, es la que se suscita mayoritariamente en nuestra ciudad, seguida por los homicidios, muertes accidentales y suicidios.

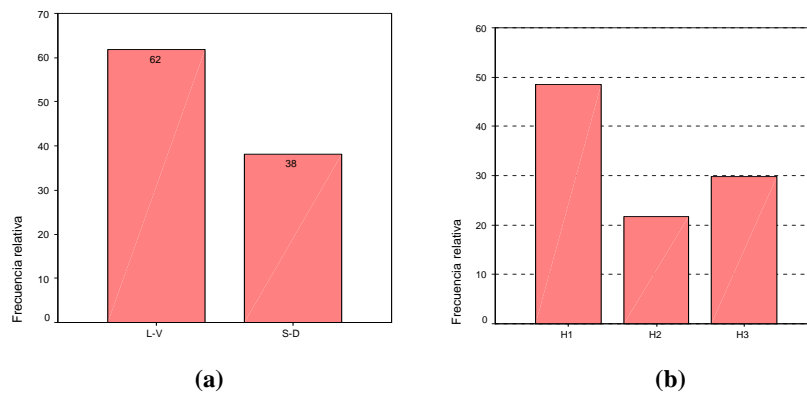
- La mayoría de muertes se da en personas de sexo masculino, y se tiene poca presencia en personas de sexo femenino; esto hace suponer que los hombres tienen mayor exposición a sufrir estos tipos de violencia.
- En los fines de semana: sábado y domingo, se registra la mayor cantidad de muertes violentas. Por esta razón procedemos a recodificar esta variable en dos modalidades como se observa en la Tabla 2.2.2.
- De acuerdo al histograma de horas, se observa que el comportamiento es similar en horas de la noche a partir de las 19:00 hasta las 23:00, y entre las 0:00 hasta la 1:59, otro comportamiento se clasificaría entre las 2:00 hasta las 6:59, y la última clasificación horaria se tiene entre las 7:00 hasta las 18:59. Con estas apreciaciones se procedió a transformar la variable hora de cuantitativa a cualitativa, considerando horas de la noche, horas de la madrugada y horas de la mañana como categorías (ver tabla 2.2.2 y figura 2.2.2(a) y 2.2.2(b)); con esto se tiene, que la mayor cantidad de muertes se tiene en horas de la noche y madrugada, y con una pequeña diferencia en horas de la mañana y tarde.
- Observando el histograma de frecuencias para edades, la mayoría de muertes se dan entre los 20 y 30 años, seguido de edades de 30 a 40, luego se registran menores eventos de violencia en edades de la niñez y en edades avanzadas.

Para los análisis, se transforma, las variables edad y hora que son cuantitativas en variables cualitativas como se indica en el siguiente cuadro.

**Tabla 2.2.2.- Recodificación de variables cuantitativas.-**

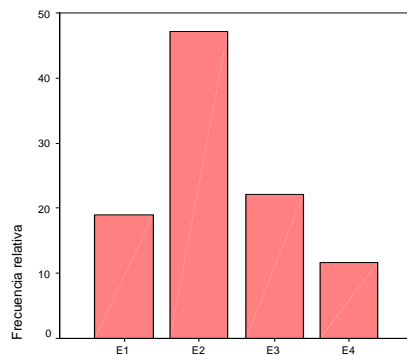
<b>NOMBRE DE VARIABLE</b>	<b>TIPO DE VARIABLE</b>	<b>MODALIDADES</b>	<b>CODIFICACIÓN</b>
RANGO HORAS	Cualitativa	19:00-23:59 y 0:00-5:59 6:00-12:59 13:00-18:59	H1= Noche/Madrugada. H2= Mañana H3= Tarde
RANGO EDAD (años)	Cualitativa	0-18. 19-39. 40-59. >60.	E1= 0-18. E2= 19-39. E3= 40-59. E4= >60.
DÍA	Cualitativa	Lunes-Viernes Sábado-Domingo	L-V S-D

**Figura 2.2.2.- Recodificación de variable día y rango horario.**



La figura 2.2.2 (b), muestra el comportamiento de muertes según las horas de ocurrencia, esto se evidencia del análisis del histograma de la variable hora.

**Figura 2.2.3.- Recodificación de rangos de edades.-**



Al igual que lo realizado para horas, la variable edad fue también recodificada y creadas categorías, como se muestra en la figura 2.2.3.

## **2.2.2.- Análisis Multivariante de muertes violentas**

### **2.2.2.1.- Análisis factorial de correspondencia múltiples.-**

Para aplicar el análisis factorial de correspondencias múltiples, utilizaremos todas las variables que fueron analizadas de manera univariante, esto es, las cuatro variables cualitativas, y las dos variables cuantitativas, que fueron recodificadas y transformadas en cualitativas.

Tabla 2.2.3.- Valores propios

Dimensión	Autovalores	Inercia	% de la inercia explicada	% acumulado
1	1,579	0,263	26,312	26,312
2	1,297	0,216	21,612	47,924
3	1,200	0,200	20,003	67,926
<b>Total</b>	<b>4,076</b>	<b>0,679</b>	<b>67,926</b>	
<b>Media</b>	<b>1,359</b>	<b>0,226</b>		

En la tabla 2.2.3, se muestran los valores propios asociados al análisis de correspondencias múltiples (ACM). Como se observa en tres dimensiones se recoge el 68% aproximadamente de la representación de las categorías. Es importante observar el comportamiento en gráficos bidimensionales, relacionando los tres factores.

Tabla 2.2.4.- Inercia y dimensión de las modalidades

Modalidades	Frecuencia marginal	Factor 1	Factor 2	Factor 3	Inercia modalidad
MV1	727	-1,1026	0,0993	0,1414	0,1258
MV2	382	0,1041	1,6940	0,0352	0,1452
MV3	1215	0,5023	-0,5705	0,5515	0,0983
MV4	640	0,2467	-0,0874	-1,2293	0,1307
M	2386	-0,2837	-0,1447	-0,1040	0,0325
F	578	1,1822	0,5458	0,4286	0,1342
L-V	1819	0,2469	-0,0160	-0,3427	0,0644
S-D	1124	-0,3851	0,0005	0,5654	0,1035
E1	561	0,8367	1,2829	0,2468	0,1351
E2	1390	-0,6264	0,0339	0,2304	0,0885
E3	655	0,0061	-0,5702	-1,0510	0,1298
E4	344	1,1681	-1,2080	0,6735	0,1473
Z1	100	0,1948	-0,4151	-0,1305	0,1610
Z2	357	-0,1543	-0,5331	-0,4288	0,1466
Z3	396	-0,2577	0,7802	-0,2663	0,1444
Z4	338	0,1076	0,1382	0,4382	0,1477
Z5	115	0,3667	-0,8904	0,1456	0,1602
Z6	468	0,0056	-0,2000	0,2823	0,1404
Z7	265	-0,3005	0,7203	-0,7679	0,1518
Z8	209	0,3115	-0,1242	0,6293	0,1549
H1	707	-0,8603	-0,0019	0,4174	0,1269
H2	591	0,2751	0,1584	-0,3146	0,1334
H3	816	0,5920	-0,0917	-0,0932	0,1208

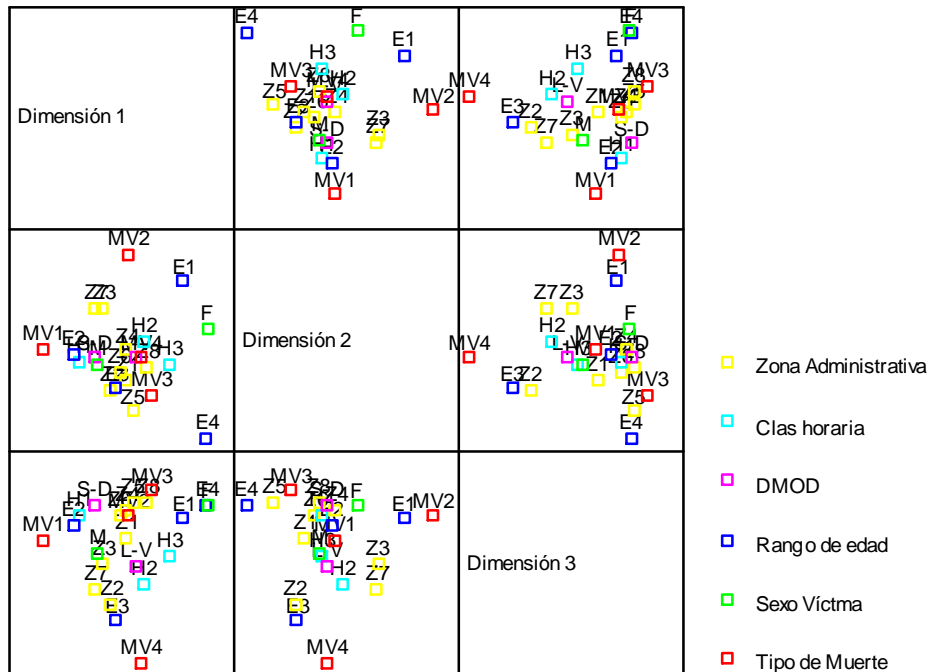
La tabla 2.2.4 muestra las coordenadas de las proyecciones de las modalidades sobre tres factores, además de esto se deduce también la inercia que contribuye cada modalidad. Observemos las modalidades de muertes violentas; se constata que muertes accidentales (MV3) tiene una inercia muy baja respecto de las otras modalidades. Para modalidades de sexo, se observa que para sexo masculino se tiene la menor inercia respecto del sexo femenino; según día de ocurrencia se observa que de lunes-viernes tienen la menor inercia respecto de los fines de semana; según el rango de edad, de 20-39 años (E2) tiene la menor inercia, respecto de las otras modalidades; para zona administrativa, se observa que no existe una diferencia marcada de la inercia entre las modalidades, al igual que en las modalidades de rango horario.

**Tabla 2.2.5.- Inercia y dimensión de las variables**

Variables	Dimensión			Inercia por variable
	Factor 1	Factor 2	Factor 3	
Sexo Víctima	0,34	0,05	0,04	0,15
Día	0,08	0,002	0,12	0,06
Clasificación horaria	0,28	0,02	0,10	0,13
Zona Administrativa	0,03	0,21	0,15	0,13
Rango edad	0,48	0,52	0,22	0,40
Tipo de Muerte	0,36	0,50	0,57	0,48

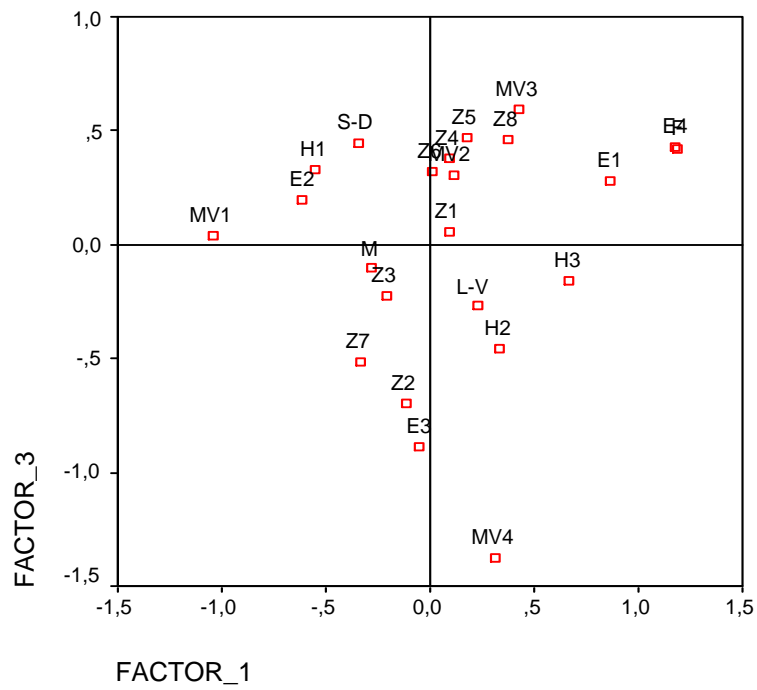
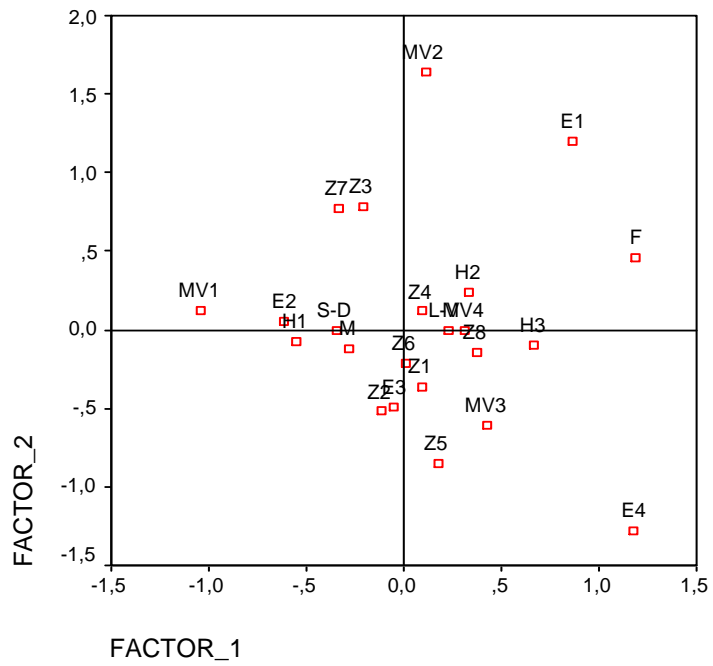
Al igual que en la tabla anterior, en la tabla 2.2.5 se muestra, las coordenadas de la proyección de las variables en tres factores y la inercia asociada a cada variable. Es claro que las variables que tienen la mayor inercia son tipo de muerte y rango de edad, lo que significa que tienen la mayor contribución absoluta, por lo que estas variables deben ser incluidas en análisis futuros. Entre clasificación horaria y zona administrativa se observa que tienen la misma inercia; se decidirá en función de la representación gráfica cuál de estas dos variables es conveniente utilizar en el estudio.

Figura2.2.4.- Relación de categorías.-

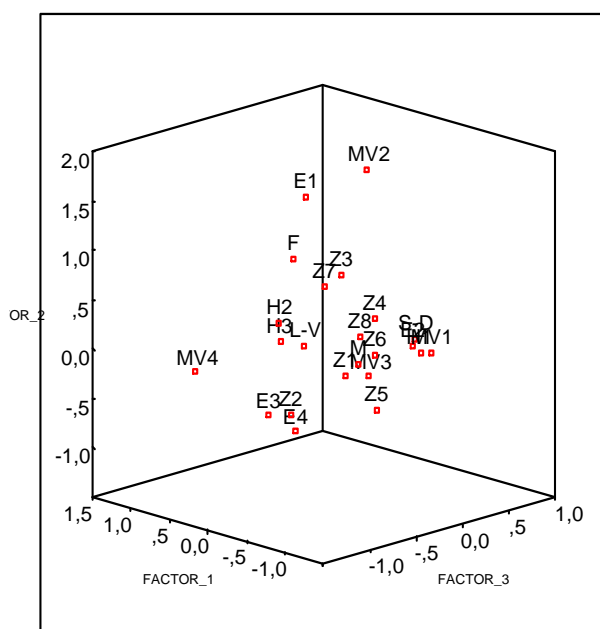
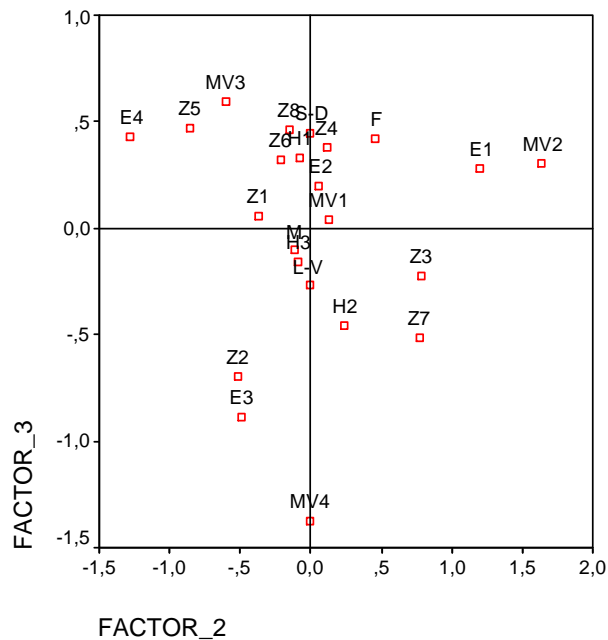


En la figura anterior se muestra como se relacionan las categorías de las variables cualitativas. Se tiene el gráfico en tres dimensiones, según esto no es muy clara la figura, por lo que se procede a realizar gráficos por pares de factores. Así tenemos.

Figura 2.2.6.- Comportamiento de las modalidades por factores.-







Con esto se tienen los resultados siguientes:

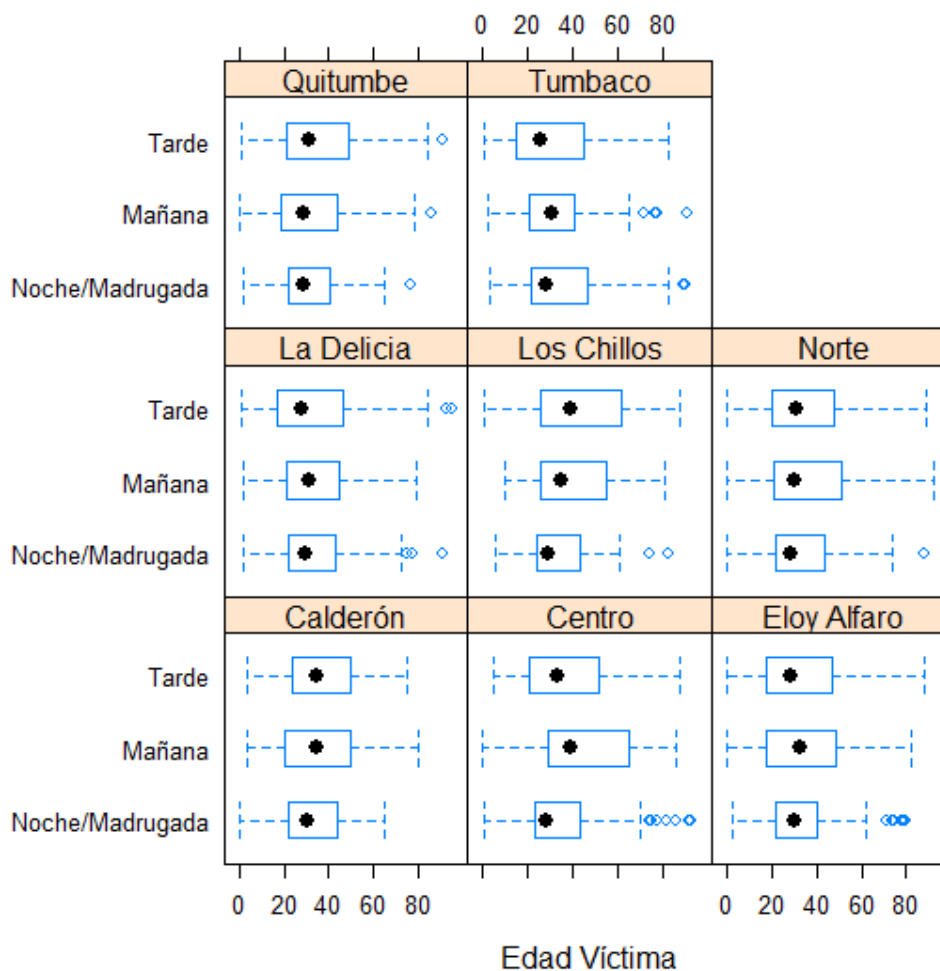
- Los homicidios se dan fundamentalmente en hombres de edades entre los 19 y 39 años, en horas de la noche y madrugada, los días sábados y domingos.
- Las muertes violentas en mujeres se dan fundamentalmente entre los 0 y 18 años o en mujeres de tercera edad mayores a 60 años. Entre los 0 y 18 años la causa fundamental es el suicidio, los que se han producido

fundamentalmente en las administraciones Eloy Alfaro Z3 y Quitumbe Z7.

- Para personas mayores de 60 años el mayor riesgo es de muertes por accidentes de tránsito y estas se suscitan en la administración del Valle de los Chillos en horas de la tarde.
- En la administración zonal Centro Z2, las edades de mayor riesgo de muerte son de 49 a 59 años, y se asocian también con muertes en accidentes de tránsito.

### 2.2.2.2.- Gráficos de cajas.-

Figura 2.2.7.- Gráfico de cajas para edades según clasificación horaria y zona administrativa.



La figura 2.2.7, está relacionada con la prueba de medias, esto es, se prueba igualdad de promedio de edades en los rangos de horas, y esto lo

discriminamos por zona administrativa; es claro que la mediana se ubica en edades de 20 a 40 años, esto es, el 50% de las muertes se dan en edades productivas. El rango horario, no tiene una diferencia marcada sea por edad de la víctima o por zona administrativa. Se muestra la tabla ANOVA para Z1 administración Calderón; para las otras zonas resumimos los valores del estadístico F y el nivel de significancia.

**Tabla 2.2.6.- Tabla ANOVA**

TABLA ANOVA CALDERÓN					
Horas	Suma de cuadrados	Grados de libertad	Media cuadrática	F	Sig.
Inter-grupos	468,7046992	2	234,3523496	0,768334465	0,466836201
Intra-grupos	27146,19747	89	305,0134548		
Total	27614,90217	91			

La tabla ANOVA muestra los estadísticos F y el nivel de significación, los cuales nos permiten rechazar o no la hipótesis nula de igualdad de medias. En la administración Z1 Calderón no se rechaza la hipótesis nula de igualdad de medias, esto es, la edad promedio de muertes en H1 es 32 años, en H2 es de 35 años y en H3 de 37 años, la prueba nos indica que no existe diferencia entre las dos edades en los tres rangos horarios.

**Tabla 2.2.7.- Resumen de tabla ANOVA**

ZONA	F	Sig.
CENTRO	5,738498865	0,003557855
ELOY ALFARO	0,430343004	0,650614617
LA DELICIA	0,153067004	0,858137885
LOS CHILLOS	2,249788515	0,110528759
NORTE	1,481981263	0,228329238
QUITUMBE	1,019026534	0,362515447
TUMBACO	0,825313101	0,439670356

El comportamiento de la hora de muerte y la edad en cada zona administrativa es la siguiente:

En la administración Centro, se rechaza la hipótesis nula de igualdad de medias. En H1 el promedio de edad es 35 años, en H2 el promedio de edad es de 45 años y en H3 39 años.

En Z3 administración Eloy Alfaro, no se rechaza la hipótesis nula de igualdad de medias. El promedio de edad en H1 es de 33 años, H2 es de 34 años y en H3 de 32 años.

En Z4 administración La Delicia no se rechaza la hipótesis nula de igualdad de las medias; el promedio de edad en H1 y H2 es de 34 años y en H3 es de 33 años.

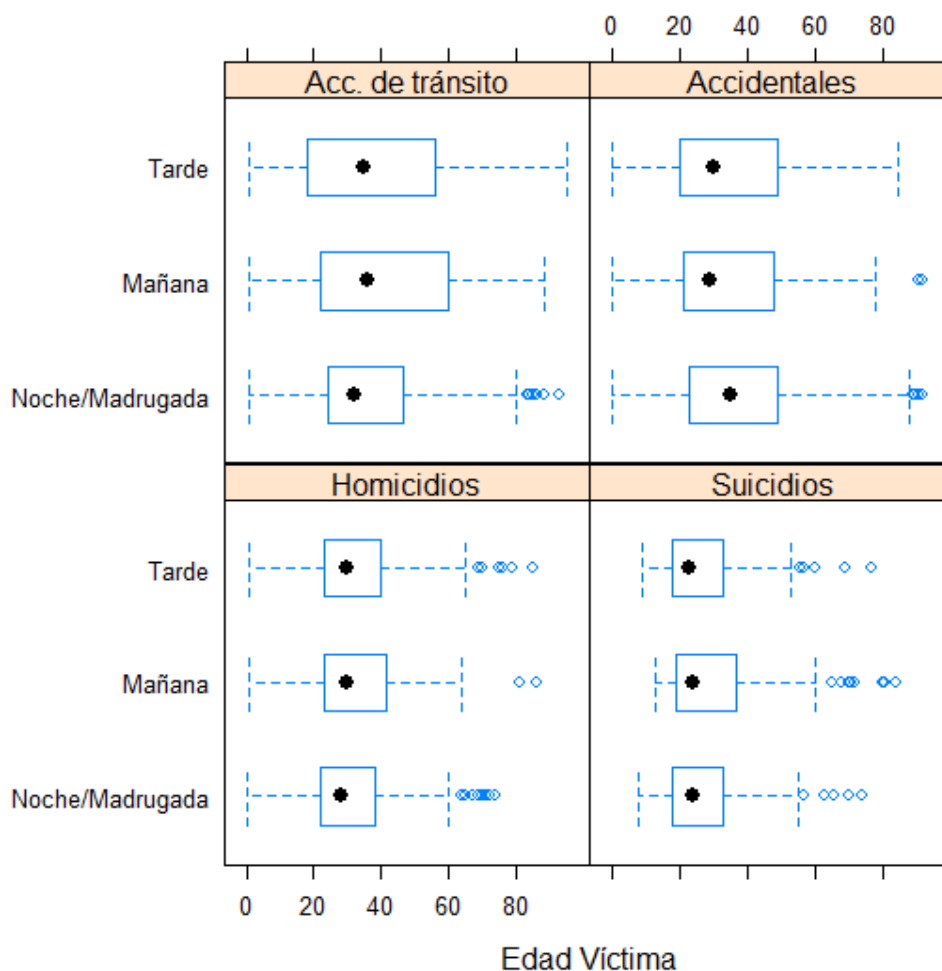
En Z5 administración Los Chillos, no se rechaza la hipótesis nula de igualdad de medias; así tenemos, que en H1 la edad promedio de muertes es de 34 años, en H2 la edad promedio de muerte es de 42 años y en H3 de 41 años.

En Z6 administración Norte, no se rechaza la hipótesis nula de igualdad de medias; el promedio de edad de muerte es, en H1 33 años, en H2 37 años y en H3 de 35 años.

En Z7 administración Quitumbe no se rechaza la hipótesis nula de igualdad de medias. El promedio de edad de muerte es, en H1 y H2 es de 32 años, en H3 la edad promedio es de 35 años.

En Z8 administración de Tumbaco, no rechazamos la hipótesis nula de igualdad de medias. El promedio de edad en H1 es de 36 años, en H2 es de 33 años y H3 de 31 años.

Figura 2.2.8.- Gráfico de cajas de edades según clasificación horaria y tipo de muerte.



Observando el comportamiento de las muertes por edad de la víctima y rango horario, se tiene que las muertes accidentales y en accidentes de tránsito las edades tiene mayor amplitud, produciéndose en el 50% de las mismas en edades de hasta 40 años, mientras que el restante 50% se extiende hasta edades de hasta 60%, las dispersiones mayores se tiene en horas de la mañana y tarde, mientras que en horas de noche y madrugada se observa que se tiene menor dispersión de edades de las víctimas.

Tabla 2.2.8.- Resumen de ANOVA

MUERTE VIOLENTA	F	Sig.
HOMICIDIOS	1,77043816	0,17106355
SUICIDIOS	2,12480955	0,12102544
ACCIDENTES DE TRÁNSITO	4,83608552	0,00810283
ACCIDENTALES	3,6029829	0,02786569

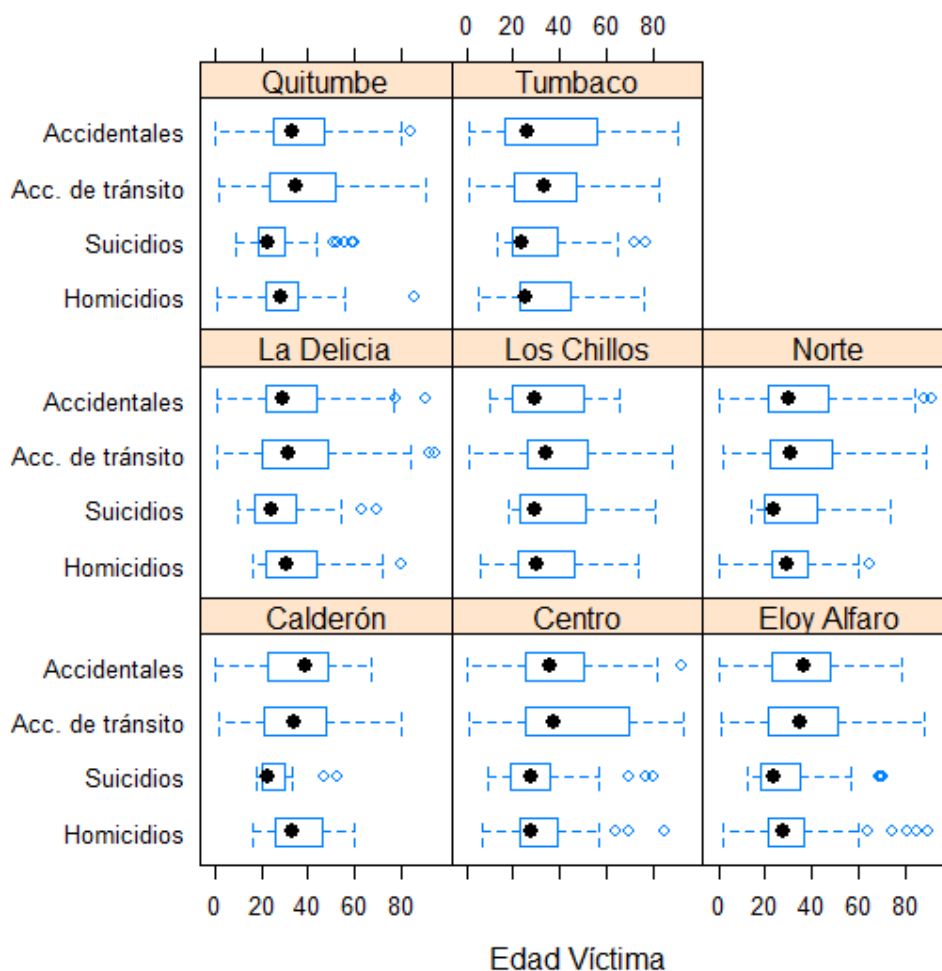
Para las muertes violentas, homicidios (MV1) se observa que no se rechaza la hipótesis nula de igualdad de medias; las edades promedio son las siguiente, para H1 31 años, H2 32 años y H3 33 años.

En los suicidios (MV2), no se rechaza la hipótesis nula de igualdad de medias, esto es las edades promedio de suicidios en H1 es de 28 años, en H2 de 32 años y en H3 de 26 años.

Las muertes en accidentes de tránsito (MV3), se rechaza la hipótesis nula de igualdad de medias, esto es, la edad promedio en H1 es de 36 años, mientras que en h2 la edad promedio de muertes es de 41 años y en H3 de 38 años.

Para las muertes accidentales, se rechaza la hipótesis nula de igualdad de medias, esto indica que existe diferencia entre H1 y H2, la edad promedio de muertes accidentales en H1 es de 38 años, en H2 es de 32 años y en H3 de 34 años.

Figura 2.2.9.- Gráfico de cajas para edad según tipo de muerte y zona administrativa



Por último, estudiaremos el comportamiento de las edades de la víctima, según el tipo de muerte y la administración zonal. Para esto se utilizará también la prueba de igualdad de medias, y los gráficos de cajas, así los suicidios son el tipo de muerte que tiene un comportamiento diferente por administración zonal, así en la administración de Calderón y Quitumbe, se observa que se dan en personas jóvenes en casi el 75% de los casos. Otro tipo de muerte que tiene un comportamiento diferente es el homicidio, así en las administraciones de Quitumbe, Tumbaco y Centro, cerca del 75% de casos se tiene en edades de 20 hasta 40 años de edad.

**Tabla 2.2.9.- Resumen de ANOVA**

ZONA	F	Sig.
CALDERÓN	0,769741558	0,513760083
CENTRO	9,556811425	4,42E-06
ELOY ALFARO	4,379513466	0,004769105
LA DELICIA	2,425592633	0,065520893
LOS CHILLOS	0,549214943	0,649709613
NORTE	2,850543196	0,03702298
QUITUMBE	6,54436014	0,0002779
TUMBACO	0,273283104	0,844627947

En la administración Z1 Calderón, la prueba no rechaza la hipótesis nula de igualdad de medias; esto es, las edades promedio de muertes violentas 36 años en homicidios, 27 en suicidios, 35 en accidentes de tránsito y 34 en accidentales, no son diferentes estadísticamente.

La administración Z2 Centro, se rechaza la hipótesis nula de igualdad de medias de edades por tipo de muerte violenta; así, tenemos: homicidios tiene una edad promedio de 32 años, suicidios 31 años, accidentes de tránsito 45 años y accidentales 39 años.

En Z3 Eloy Alfaro, se rechaza la hipótesis nula de igualdad de medias, las edades promedio según el tipo de muerte son, para homicidios 31 años, suicidios 29 años, accidentes de tránsito 37 años y accidentales 35 años. Para la administración Z4 La Delicia, no rechazamos la hipótesis nula de igualdad de medias; las edades promedio de muerte en la zona no son diferentes estadísticamente. Para homicidios la edad promedio es de 35 años, para suicidios de 27 años, accidentes de tránsito de 35 y accidentales de 33.

E Z5 Valle de los Chillos, no se rechaza la hipótesis nula de igualdad de medias, ahora, las edades promedio por tipo de muerte son las siguientes: homicidios 36 años, suicidios 36 años, accidentes de tránsito 40 años y accidentales 34 años.

En Z6 administración Norte, se rechaza la hipótesis nula de igualdad de medias; esto es, las edades promedio por tipo de muerte son como sigue, para homicidios la edad promedio es de 31 años, suicidios 32 años, accidentes de tránsito 37 años y accidentales 35 años.



Para la administración Z7 Quitumbe, se rechaza la hipótesis nula de igualdad de medias, las edades promedio para homicidios son 30 años, para suicidios es 27 años, para accidentes de tránsito es 38 y para accidentales es 37 años.

Por último para la administración Z8 Tumbaco no se rechaza la hipótesis nula de igualdad de medias, esto es, no existe diferencia estadística entre edad promedio de muertes violentas, sin embargo las edades promedio por tipo de muerte son, homicidios 34 años, suicidios 33 años, accidentes de tránsito 34 años y accidentales 37 años.

## 2.3.- Análisis de delitos contra los vehículos.-

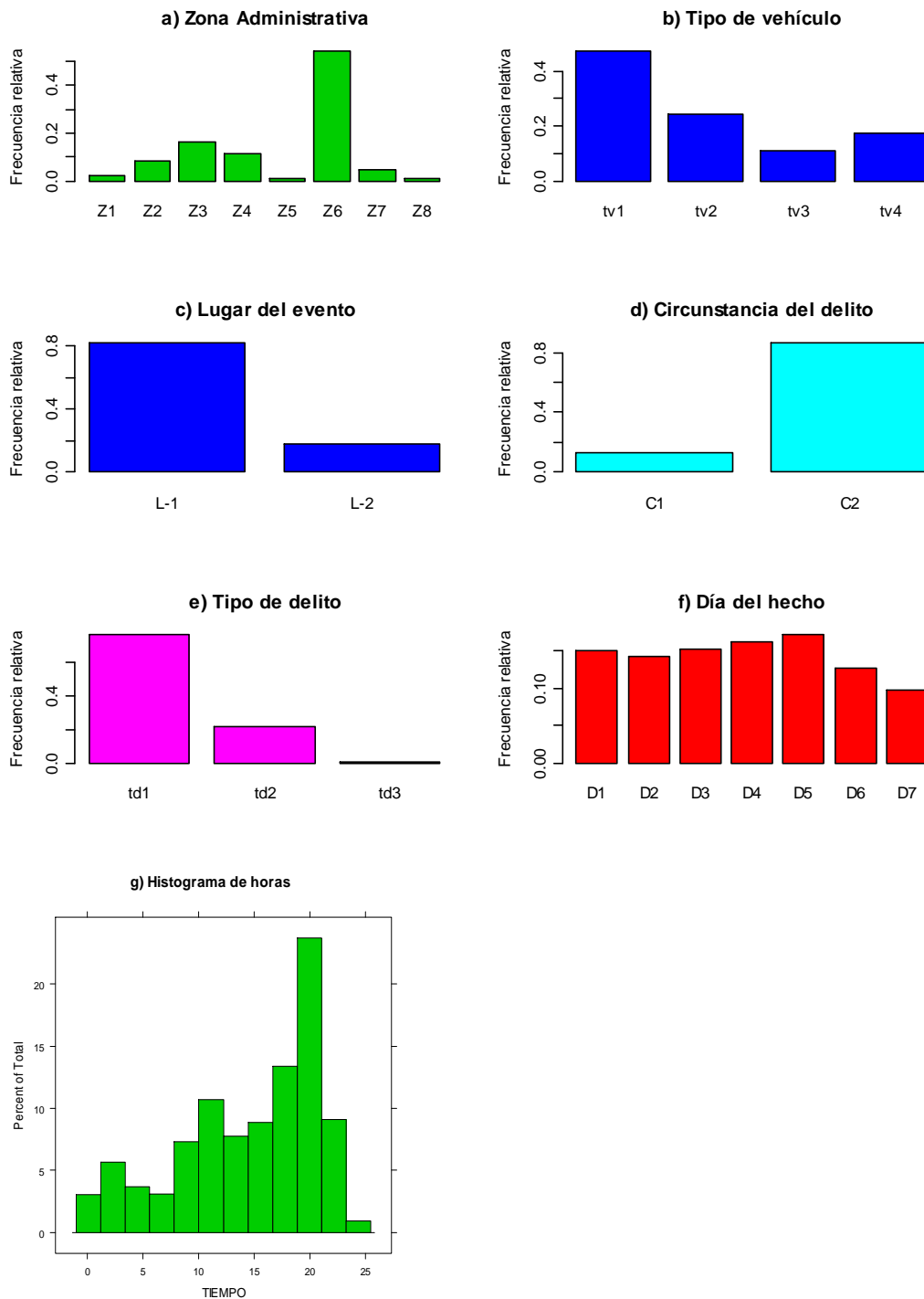
### 2.3.1.- Gráficos de frecuencias relativas.-

Enseguida procedemos a realizar el análisis exploratorio univariante de delitos a los vehículos. Las variables, que se consideran, son tomadas del código penal, y que constan en las denuncias que se registran en el Ministerio Público de Pichincha en la unidad de delitos contra los vehículos.

Tabla 2.3.1.- Descripción de variables de delitos contra vehículos

NOMBRE DE VARIABLE	TIPO DE VARIABLE	MODALIDAD	CODIFICACIÓN
ZONA ADMINISTRATIVA	Cualitativa	Calderón Centro Eloy Alfaro La Delicia Los Chillos Norte Quitumbe Tumbaco	Z1 Z2 Z3 Z4 Z5 Z6 Z7 Z8
TIPO DE VEHÍCULO	Cualitativa	Automóvil Camioneta Todo terreno Motocicleta Otro.	TV1 TV2 TV3 TV4 TV5
LUGAR	Cualitativa	Vía pública Estacionamiento	L1 L2
CIRCUNSTANCIA	Cualitativa	Estacionado Interceptado.	C1 C2
TIPO DELITO	Cualitativa	Robo. Asalto y robo. Hurto.	TD1 TD2 TD3
HORAS	Cuantitativa continua		0-24
DÍA DEL HECHO	Cualitativa	Lunes Martes Miércoles Jueves Viernes Sábado Domingo.	D1 D2 D3 D4 D5 D6 D7

**Figura 2.3.1.- Frecuencias relativas.**



De la figura anterior se puede observar los siguientes resultados:

- En la zona Norte se registra el mayor número de denuncias, en general en el resto de las zonas administrativas se tienen frecuencias relativas similares.

- Los automóviles ocupan el primer lugar de los vehículos denunciados, seguidos por camionetas y motocicletas.
- Es en la vía pública donde se cometen mayoritariamente delitos contra los vehículos, siendo los más afectados los que se encuentran estacionados en la vía pública.
- El robo es el delito que más se produce en nuestra ciudad.
- El mayor número de delitos contra los vehículos se tiene de lunes a viernes, teniéndose los fines de semana un menor registro de estos delitos.
- Según la distribución de las horas de los robos se observa que entre las 19:00 hasta las 21:00, son las horas donde se produce el mayor número de delitos a vehículos.

## **2.3.2.- Multivariante delitos contra los vehículos.-**

### **2.3.2.1.- Análisis factorial de correspondencia múltiples.-**

**Tabla 2.3.2.- Valores propios.-**

<b>Dimensión</b>	<b>Autovalores</b>	<b>Inercia</b>	<b>% de la inercia explicada</b>	<b>% acumulado</b>
1	1,920	0,274	27,432	27,432
2	1,309	0,187	18,699	46,131
3	1,251	0,179	17,875	64,006
Total	4,480	0,640	64,006	
Media	1,493	0,213		

La tabla 2.3.2 muestra los valores propios que se obtienen del ACM, y que se utilizan para la formación de los factores de proyección de las modalidades. Los dos primeros factores tienen las mayores representaciones de las modalidades, alcanzando el 49% aproximadamente.

Tabla 2.3.3.- Factores e inercia por modalidad

Modalidades	Frecuencia marginal	Factor 1	Factor 2	Factor 3	Inercia modalidad
Z1	46	-0,2138	1,6125	1,5123	0,1396
Z2	170	0,5028	0,4407	-1,1271	0,1309
Z3	328	0,1470	0,5394	0,8910	0,1199
Z4	229	-0,4994	-0,2486	0,5767	0,1268
Z5	25	-1,6414	1,7212	0,6228	0,1411
Z6	1068	0,0793	-0,4136	-0,2603	0,0680
Z7	87	-0,1577	1,4104	-0,3821	0,1368
Z8	22	-0,7494	0,1653	-0,4620	0,1413
tv1	959	-0,3042	-0,3148	-0,1140	0,0756
tv2	498	0,1148	0,0092	0,8020	0,1079
tv3	221	0,5878	-0,3496	0,4464	0,1274
tv4	359	0,2884	1,0584	-1,0924	0,1177
L-1	1671	-0,1010	-0,2326	0,0760	0,0257
L-2	366	0,4580	1,0769	-0,3562	0,1172
C1	274	-2,1758	0,2331	-0,4006	0,1236
C2	1759	0,3382	-0,0323	0,0600	0,0195
td1	1571	0,4588	-0,0347	0,0979	0,0327
td2	439	-1,6697	0,0492	-0,3150	0,1121
td3	27	0,4125	1,4184	-0,7000	0,1410
t1	256	-0,4149	1,7418	0,2901	0,1249
t2	376	0,4395	0,0108	-0,0295	0,1165
t3	510	0,4401	-0,0682	-0,5127	0,1071
t4	839	-0,3624	-0,5238	0,2630	0,0840
L-V	1580	0,0623	-0,1880	-0,2797	0,0320
S-D	456	-0,2112	0,6606	0,9595	0,1109

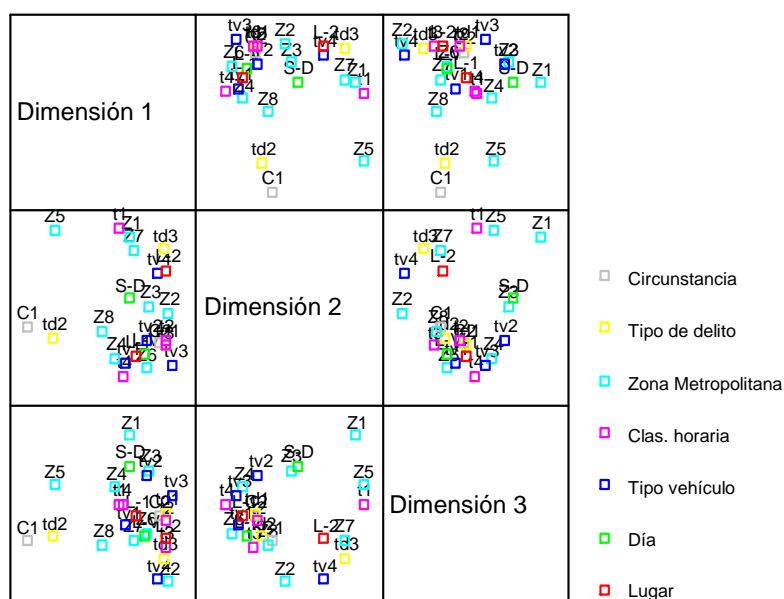
La tabla 2.3.3 resume las coordenadas de las proyecciones de las modalidades en tres factores, y muestra también la inercia asociada con cada modalidad. Las categorías que presentan las inercias menores son circunstancia estacionado, lugar vía pública, tipo de vehículo automóvil y zona administrativa norte.

Tabla 2.3.4.- Factores e inercia por variable

Variables	Dimensión			Inercia por variable
	Factor 1	Factor 2	Factor 3	
Día	0,02	0,18	0,19	0,13
Zona Metropolitana	0,10	0,52	0,30	0,31
Lugar	0,05	0,25	0,11	0,14
Circunstancia	0,73	0,00	0,03	0,26
Tipo de delito	0,76	0,13	0,05	0,31
Tipo vehículo	0,10	0,19	0,47	0,26
Clas. Horaria	0,16	0,03	0,09	0,09

Al igual que en la tabla anterior, en la tabla 2.3.4, se muestra las coordenadas de las variables en tres factores y su respectiva inercia. El resultado que se obtuvo por las categorías de cada variable, se comprueba también en la tabla anterior, siendo las variables zona metropolitana y tipo de delito las de mayor inercia, seguida de tipo de vehículo y circunstancia.

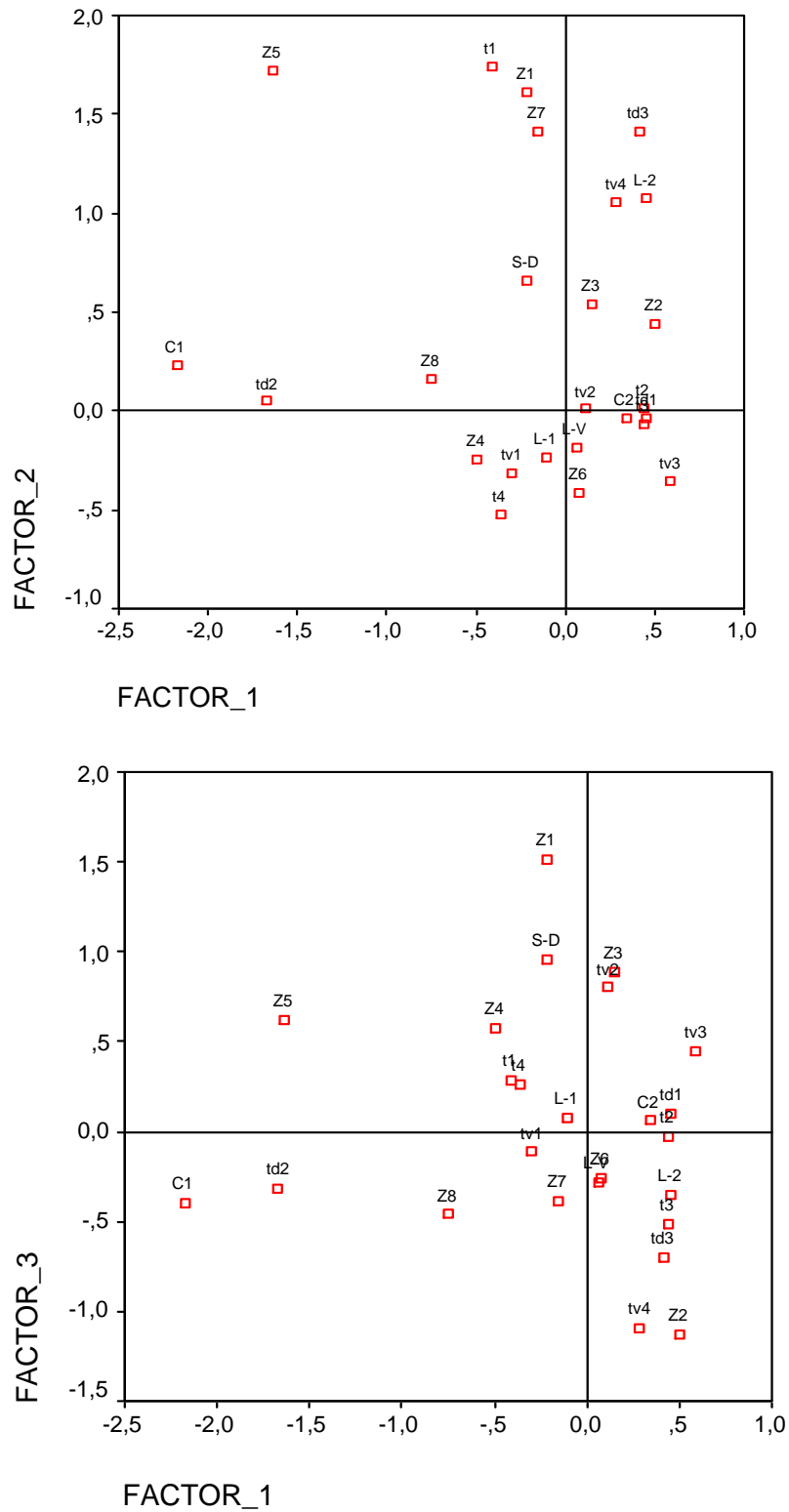
Figura 2.3.2.- Relación entre categorías.

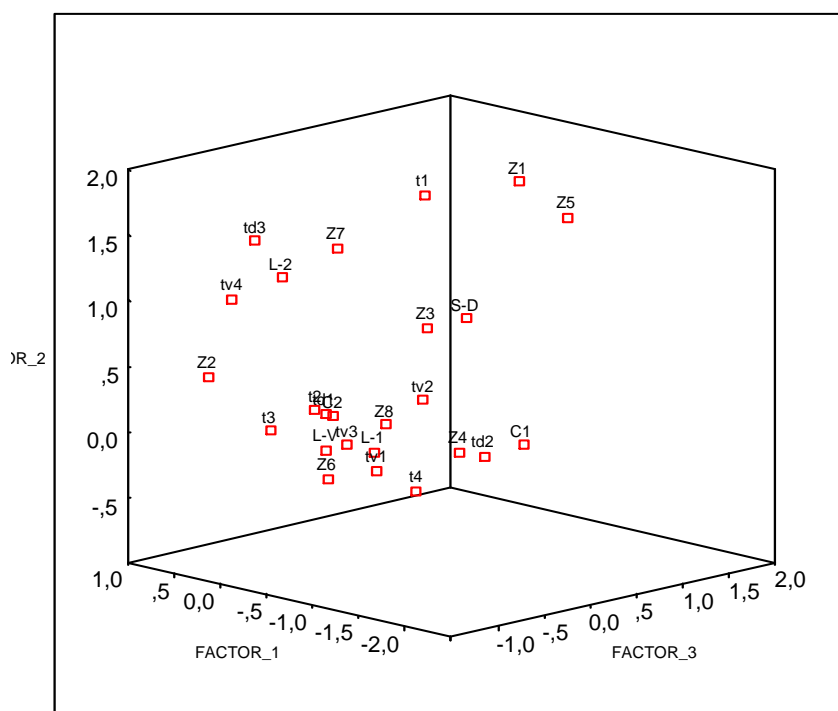
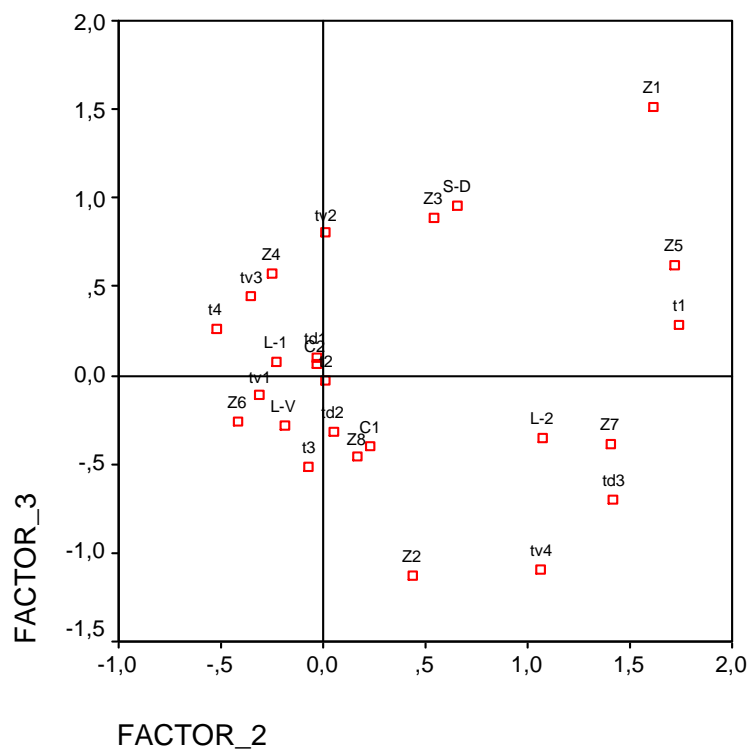


La figura muestra el comportamiento de las categorías de las variables en tres dimensiones, es claro que no se tiene una visualización adecuada del

comportamiento, por esto, se procede a realizar gráficos separados en dos dimensiones.

Figura 2.3.3.- Comportamiento de las modalidades por factores.-





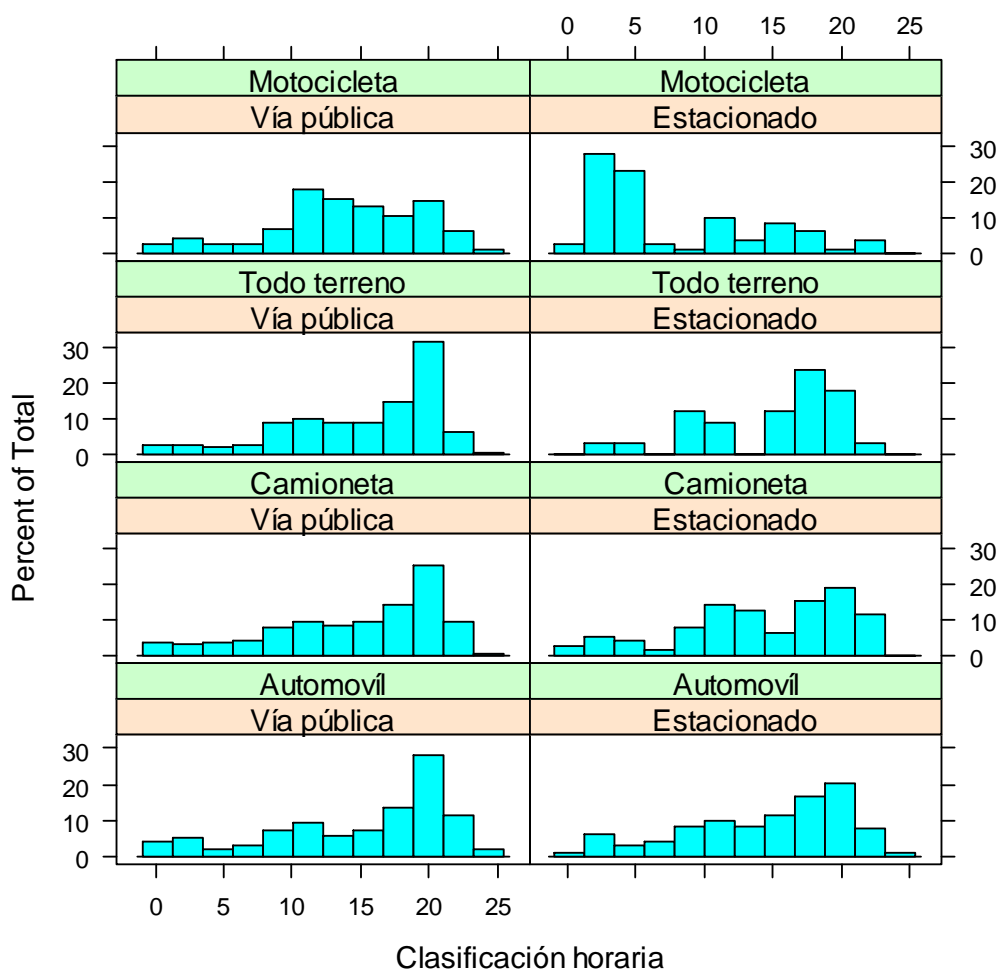
Al analizar el comportamiento de las categorías de las diferentes variables que describen el comportamiento de las categorías de las diferentes variables que describen el delito a los vehículos, se observa lo siguiente:

- En la administración Los Chillos Z5, se observa que el delito que más se comete es el asalto y robo de vehículos, cuando estos se encuentran estacionados.
- En las administraciones de Calderón Z1 y Quitumbe Z7, se tiene la mayor probabilidad de sufrir este delito es en horas de la madrugada, siendo los sábados y domingos, los días de mayor riesgo, a este comportamiento se puede asociar también a la administración Valle de los Chillos (por la proyección que tiene respecto al factor 2).
- Los automóviles son más probables de sufrir delitos en horas de la noche, al estar estacionados en la vía pública, los días lunes, martes, miércoles y jueves, siendo las administraciones de La Delicia Z4 y Norte Z6, donde más se tendría este fenómeno. Este comportamiento es similar en los vehículos todo terreno.
- En las administraciones Centro y Eloy Alfaro, las motocicletas sufren los mayores delitos al ser hurtadas de los estacionamientos.



### 2.3.2.2.- Gráficos multivariante.-

Figura 2.3.4.- Histograma de horas según tipo de vehículo y lugar del delito.



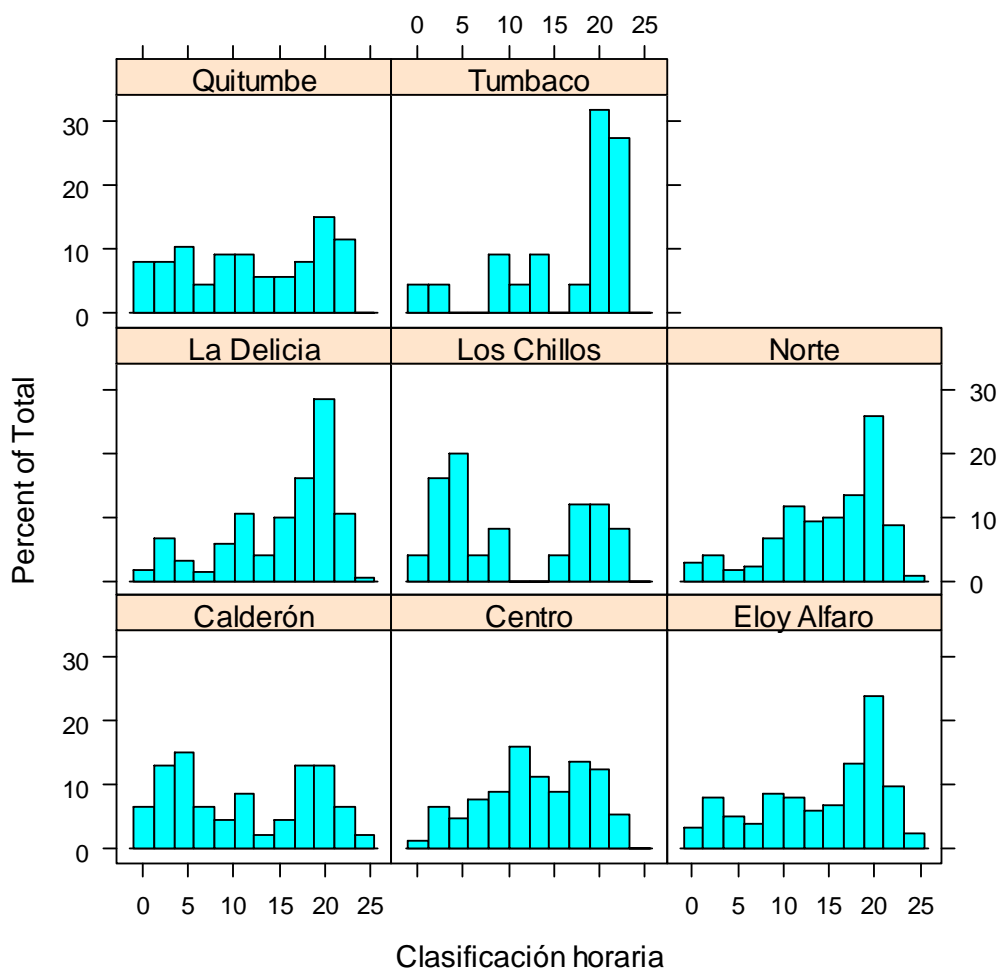
Se tiene un comportamiento según la clasificación horaria como sigue:

Los automóviles, camionetas y vehículo todo terreno, tienen mayor probabilidad de sufrir delitos en el siguiente rango horario, de 19:00 horas a 21:00 horas.

Las motocicletas tienen riesgo en la vía pública, ya que el rango horario en que sufren los delitos es de 10:00 horas hasta las 20:00 horas.

Cuando los delitos se producen en estacionamientos, estos se suscitan en horas de la madrugada y los afectados son las motocicletas. Los automóviles son afectados en horas de la tarde y noche entre las 18:00 horas y las 20:00 horas. Las camionetas tienen dos rangos de horas en donde sufren delitos en estacionamientos, estos son: de 10:00 a 15:00 horas y de 19:00 a 21:00 horas, este comportamiento es similar en vehículos todo terreno.

Figura 2.3.5.- Histograma de horas según zona administrativa.-



Los resultados son los siguientes:

En las administraciones Calderón, Valle de los Chillos y Quitumbe, los delitos se producen en horas de la madrugada de 3:00 a 5:00 horas y en horas de la noche de 18:00 a 21:00 horas.

Del mismo modo, en las administraciones Eloy Alfaro, La Delicia, Norte y Tumbaco, los delitos son más frecuentes en horas de la noche de 18:00 a 21:00 horas. Mientras que en la zona Centro se tiene un comportamiento similar en el rango horario que comprende las 10:00 hasta las 20:00 horas.

## 2.4.- Análisis de delitos contra la propiedad.-

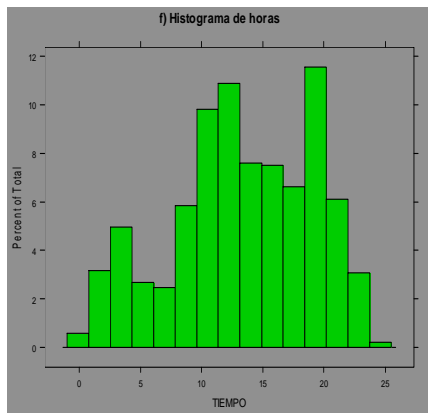
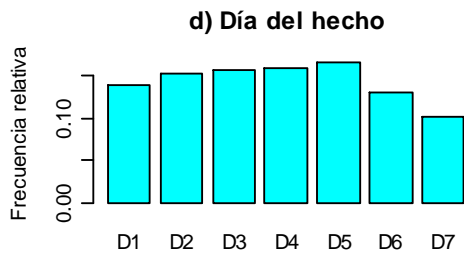
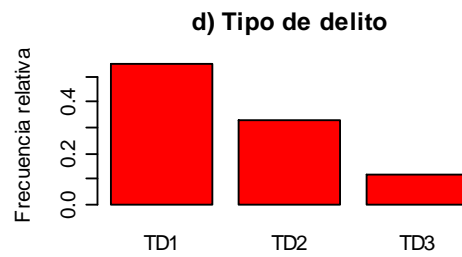
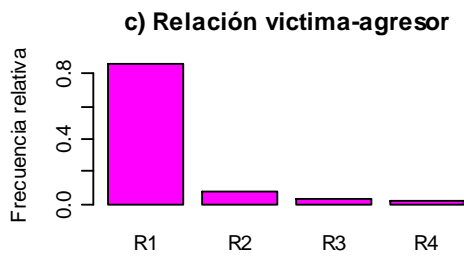
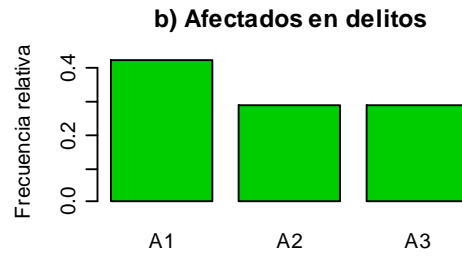
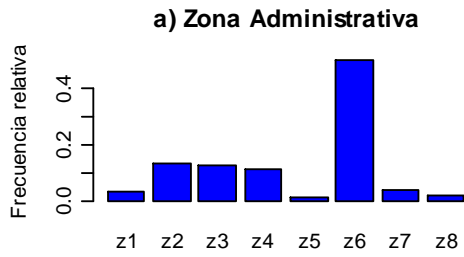
Otro tipo de delito que se va a describir es el delito contra la propiedad, estos delitos, al igual que los delitos contra los vehículos son tipificados en el código de procedimiento penal, y las variables son tomadas de las denuncias que se registran en el Ministerio Público de Pichincha en la unidad de delitos contra la propiedad.

**Tabla 2.4.1.- Descripción de variables de delitos contra la propiedad.-**

NOMBRE DE VARIABLE	TIPO DE VARIABLE	MODALIDADES	CODIFICACIÓN
ZONA ADMINISTRATIVA	Cualitativa	Calderón Centro Eloy Alfaro La Delicia Los Chillos Norte Quitumbe Tumbaco	Z1 Z2 Z3 Z4 Z5 Z6 Z7 Z8
AFFECTADO	Cualitativa	Personas Domicilios Empresas	A1 A2 A3
RELACIÓN VÍCTIMA-AGRESOR	Cualitativa	Ninguna Conocido Empleado Otra.	R1 R2 R3 R4
TIPO DELITO	Cualitativa	Robo. Asalto y robo. Hurto.	TD1 TD2 TD3
DÍA DEL HECHO	Cualitativa	Lunes Martes, Miércoles Jueves Viernes Sábado, Domingo.	D1 D2 D3 D4 D5 D6 D7
HORAS	Cuantitativa continua		0-24

### 2.4.1.- Gráficos de frecuencias.-

Figura 2.4.1.- Frecuencias relativas.-



En la figura se muestra un análisis gráfico de las variables asociadas a la denuncia de delitos contra la propiedad, los principales resultados son los siguientes:

- La zona administrativa que más denuncias registra es la Norte, esto puede obedecer a que en esta administración se encuentra el ministerio Público de Pichincha, luego le siguen las zonas Centro, Eloy Alfaro y la Delicia. En las otras zonas administrativas tienen una frecuencia relativa similar.
- Las personas son quienes sufren la mayor cantidad de delitos, y con frecuencia relativa similar se tiene a los domicilios y empresas.
- El principal delito que se denuncia es el robo, seguido del asalto y robo y en menor cantidad se tiene los hurtos.
- De lunes a viernes son cuando se producen la mayor frecuencia de delitos, los mismos que se producen en horas de la tarde y noche, los días sábado y domingo se tiene una menor proporción. Además se observa que no hay ninguna relación entre los delincuentes y las víctimas.

## **2.4.2.- Multivariante delitos contra la propiedad.-**

### **2.4.2.1.- Análisis factorial de correspondencia múltiples.-**

**Tabla 2.4.2.- Valores propios**

<b>Dimensión</b>	<b>Autovalores</b>	<b>Inercia</b>	<b>% de la inercia explicada</b>	<b>% acumulado</b>
1	1,876	0,313	31,269	31,269
2	1,598	0,266	26,636	57,905
3	1,334	0,222	22,229	80,135
Total	4,808	0,801	80,135	
Media	1,603	0,267		

La tabla 2.4.2, muestra los valores propios que se tiene del ACM, y que generan los factores y centros de gravedad. Se observa que el primer factor tiene la mayor representatividad de las modalidades con el 31% aproximadamente, mientras que entre los tres primeros factores alcanzan el 80% de la varianza.

Tabla 2.4.3.- Coordenadas e inercias

Modalidades	Frecuencia marginal	Factor 1	Factor 2	Factor 3	Inercia modalidad
TD1	7515	0,4173	0,5457	0,3269	0,0743
TD2	4497	-1,0392	-0,2362	-0,3338	0,1114
TD3	1553	1,2433	-2,0510	-0,5801	0,1476
A1	5733	-0,8219	-0,1225	-0,4384	0,0962
A2	3896	0,8897	0,8369	-0,2646	0,1188
A3	3936	0,4165	-0,6870	0,9144	0,1183
z1	407	0,6678	0,6361	0,0697	0,1617
z2	1703	0,0876	-0,2200	-0,5015	0,1457
z3	1588	0,4227	0,1992	-0,5084	0,1472
z4	1427	0,3369	0,5719	0,0712	0,1491
z5	236	0,9579	0,5515	-1,2318	0,1638
z6	6114	-0,3656	-0,2495	0,4049	0,0915
z7	532	0,7832	0,5679	-1,0897	0,1601
z8	288	0,7236	-0,0465	-0,2802	0,1631
D1	1852	0,0193	-0,2338	0,2436	0,1439
D2	2019	-0,1998	-0,0740	0,0933	0,1419
D3	2072	-0,0879	-0,2092	0,2810	0,1412
D4	2108	-0,0609	0,0154	0,1399	0,1408
D5	2200	-0,0813	-0,0644	-0,0519	0,1396
D6	1724	0,2081	0,2722	-0,2940	0,1455
D7	1350	0,4343	0,5750	-0,4773	0,1501
R1	11563	-0,1893	0,1590	0,2073	0,0246
R2	1053	1,0128	-0,4054	-2,3974	0,1537
R3	472	1,9491	-3,2078	0,5962	0,1609
R4	259	1,6596	-0,2050	-0,6755	0,1635
t1	1472	0,3557	0,5139	0,6886	0,1486
t2	2874	0,1426	-0,0381	-0,1712	0,1314
t3	3756	0,1380	0,0546	0,0425	0,1205
t4	3232	-0,8485	-0,1346	-0,3137	0,1270

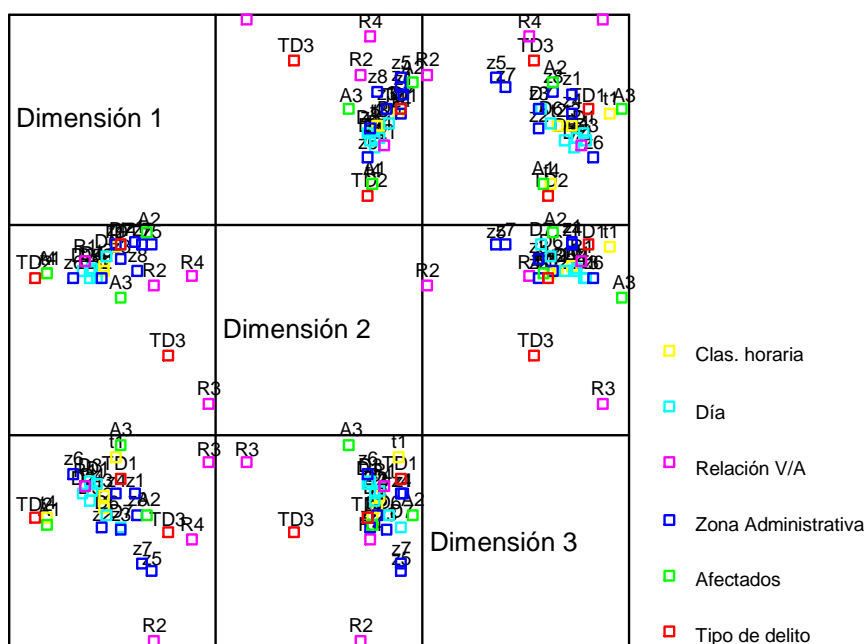
La tabla 2.4.3 muestra la inercia por modalidad y las coordenadas de proyección de estas sobre tres factores. Las categorías que tienen la menor inercia son relación V/A ninguna, tipo de delito robo, administración zonal norte y afectado personas.

Tabla 2.4.4.-

Variables	Dimensión			Inercia por variable
	Factor 1	Factor 2	Factor 3	
Tipo de delito	0,63	0,67	0,14	0,48
Afectados	0,56	0,34	0,34	0,42
Zona Administrativa	0,16	0,10	0,21	0,16
Día	0,03	0,06	0,06	0,05
Relación V/A	0,29	0,39	0,50	0,40
Clas. horaria	0,19	0,03	0,08	0,10

Al igual que la tabla anterior, en la tabla 2.4.4, se muestran las coordenadas de las proyecciones de las variables y las inercias respectivas, siendo tipo de delito, afectado y relación víctima agresor las que presentan la mayor inercia, mientras que las de menor inercia son día, zona administrativa y clasificación horaria.

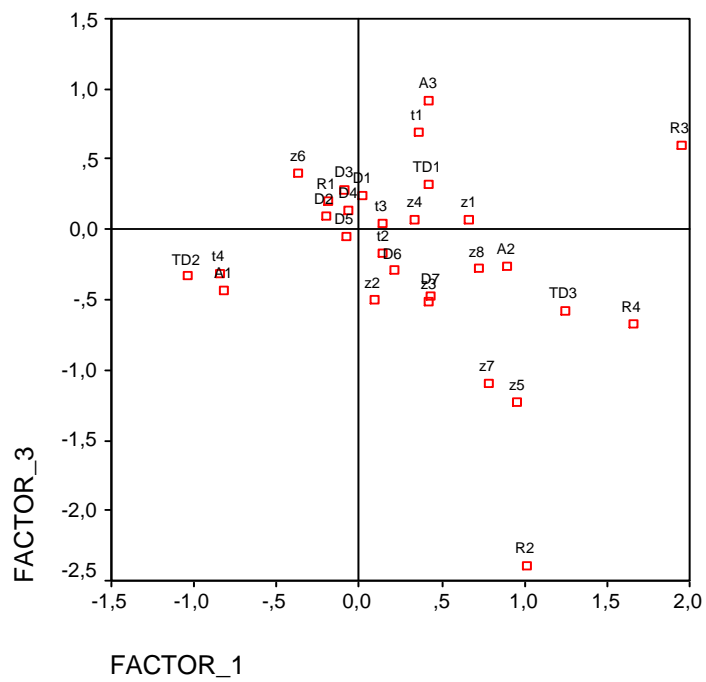
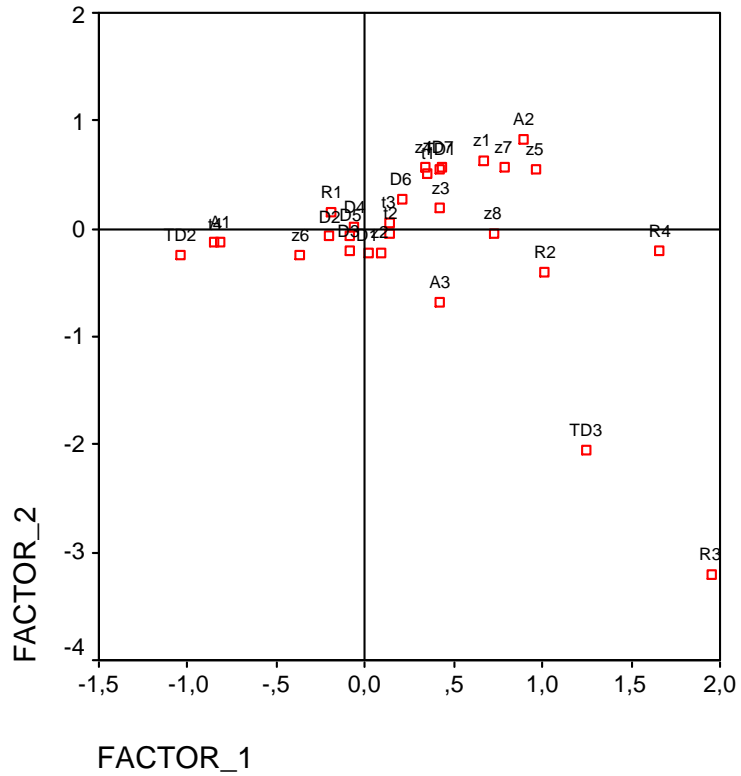
Figura 2.4.2.- Relaciones entre categorías.-



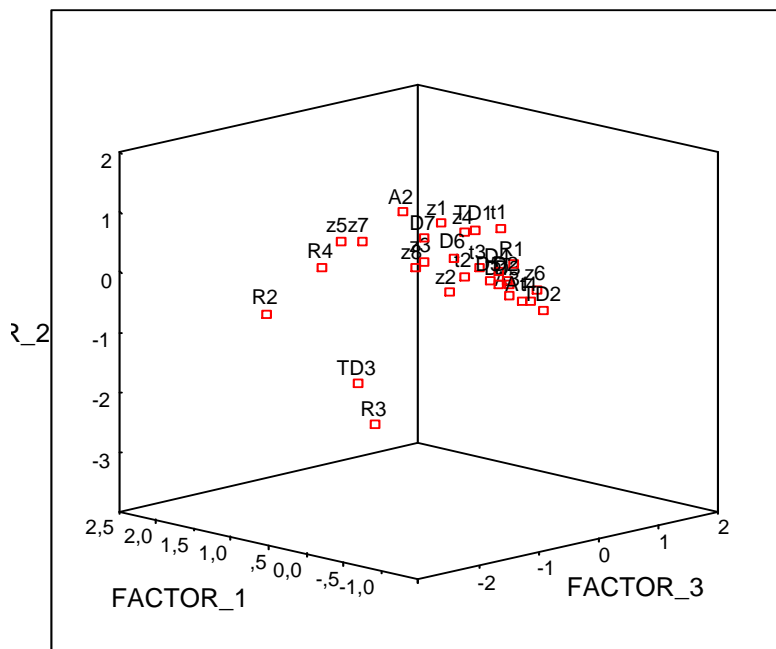
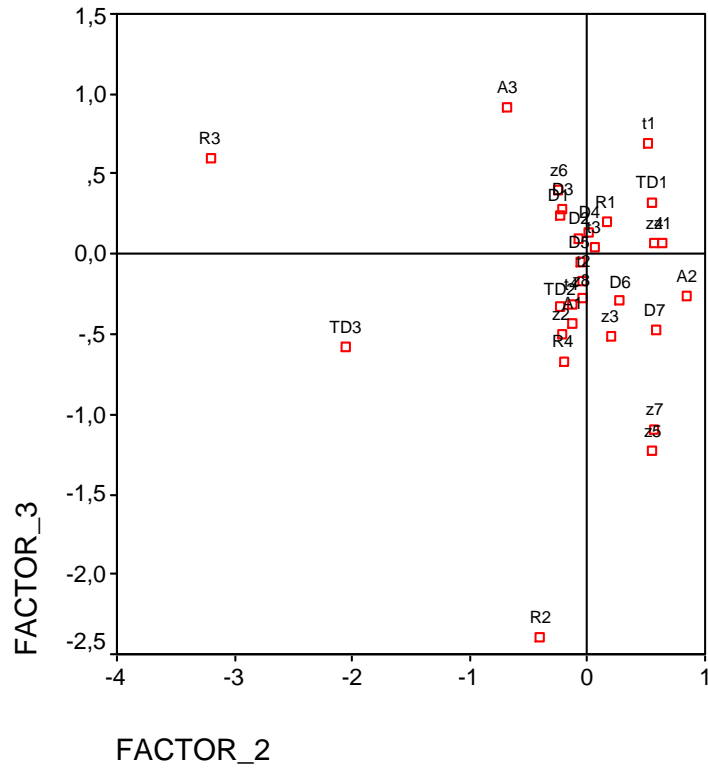
Del gráfico del análisis factorial de correspondencias en tres factores, no se puede interpretar de forma clara debido a la gran cantidad de categorías de las

variables, por esta razón, se procede a realizar gráficos separados y uno en tres dimensiones.

Figura 2.4.3.- Comportamiento de las modalidades por factores





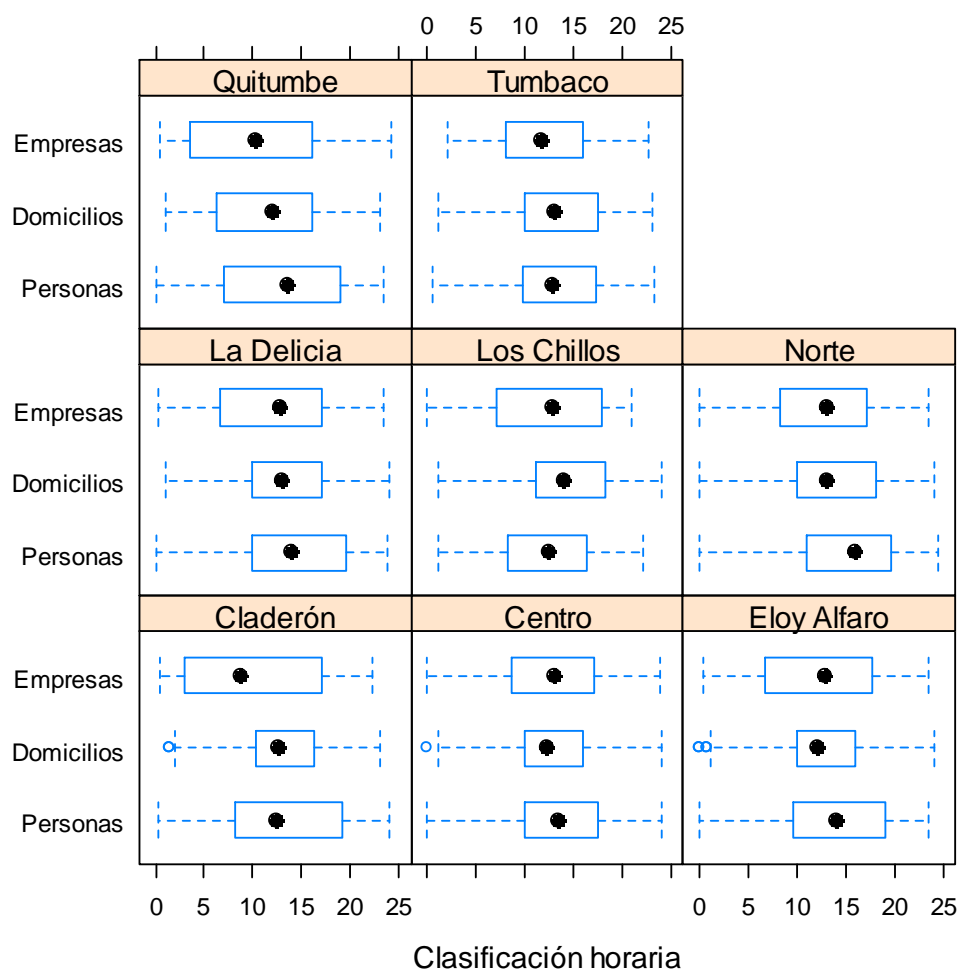


Con esto se puede indicar los siguientes resultados:

- Las personas sufren de asalto y robo, estos se dan en horas de la noche, esto por lo general, se producen en la administración Norte Z6 y Centro Z2 (por el factor 2).
- Las empresas sufren de hurtos, que generalmente, lo cometen sus empleados.
- Los delitos a domicilios se dan en cuatro zonas administrativas Calderón Z1, Quitumbe Z7, Los Chillos Z5 y Tumbaco Z8.
- La administración de La Delicia, el día domingo, en horas de la madrugada y el delito de robo, tienen como afectados las personas y hogares, ya que estos se encuentran a los extremos de estas categorías.

## 2.4.2.- Gráficos multivariante.-

Figura 2.4.4.- Gráfico de cajas para horas, según afectado y zona administrativa



En el comportamiento de afectado por administración zonal, se observa que no existen diferencias significativas, por horas, excepto en la zona norte, donde las personas sufren asaltos en horas de la tarde y los domicilios en horas de la mañana. En calderón se observa que el comportamiento en los delitos a domicilios se da en la gran mayoría más del 75% de estos en horas de la mañana.

Tabla 2.4.5.- tabla ANOVA

ANOVA CALDERÓN					
TIEMPO	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Inter-grupos	671,7667607	2	335,8833803	9,185975135	0,000130749
Intra-grupos	12212,64454	334	36,564804		
Total	12884,4113	336			

En la tabla se muestra los estadísticos de la prueba de igualdad de medias, en este caso, para la administración Z1 Calderón, se rechaza la hipótesis nula de igualdad de medias, esto es, que existe diferencia entre las horas promedio de ocurrencia de delitos según el afectado, así para A1 el promedio de hora es de 13 horas, para A2 el promedio de hora es de 13:00, para A3 el promedio es de 10 horas.

**Tabla 2.4.6.- Resumen de tabla ANOVA**

ZONA	F	Sig.
CENTRO	5,908057304	0,002785384
ELOY ALFARO	11,89315724	7,59177E-06
LA DELICIA	9,930842151	5,27921E-05
LOS CHILLOS	2,886681634	0,05818476
NORTE	131,8682524	1,34941E-56
QUITUMBE	3,894515782	0,021071733
TUMBACO	1,810398131	0,165922441

En la tabla anterior se muestra en resumen, los estadísticos que determinan si se acepta o rechaza la hipótesis nula de igualdad de medias. Los resultados son los siguientes:

En Z2 administración Centro, se rechaza la hipótesis nula de igualdad de medias, así, las horas promedio son como sigue, en A1 es las 14 horas, para A2 13 horas y para A3 la hora promedio es de 12 horas.

En Z3 administración Eloy Alfaro, se rechaza la hipótesis nula de igualdad de medias, para A1 se tiene en promedio 14 horas, para A2 12 horas y para A3 12 horas.

Para Z4 administración La Delicia se rechaza la hipótesis nula de igualdad de medias, así tenemos las horas promedio según el afectado, para A1 es de 14 horas, para A2 la hora promedio es de 13 horas, para A3 la hora promedio es de 12 horas.

Para Z5 administración Valle de los Chillos, no se rechaza la hipótesis nula de igualdad de medias, esto es, las horas promedio por tipo de afectado no son diferentes estadísticamente, el promedio para A1 es de 12 horas, para A2 el promedio es de 14 horas y para A3 el promedio es de 12 horas.

En Z6 administración Norte, se rechaza la hipótesis nula de igualdad de medias, esto es los promedios de horas por tipo de afectado es diferente, el promedio en A1 es de 15 horas, para A2 el promedio es de 13 horas y para A3 el promedio es de 12 horas.

En Z7 administración Quitumbe, se rechaza la hipótesis nula de igualdad de medias, los promedios de horas son los siguientes, A1 13 horas, A2 11 horas y A3 11 horas en promedio.

En Z8 administración Tumbaco, no se rechaza la hipótesis nula de igualdad de medias, en promedio para A1 se tiene 13 horas, para A2 13 horas y para A3 12 horas, no se tiene diferencia estadística entre los promedios.

## 2.5.- Análisis de violencia intrafamiliar.-

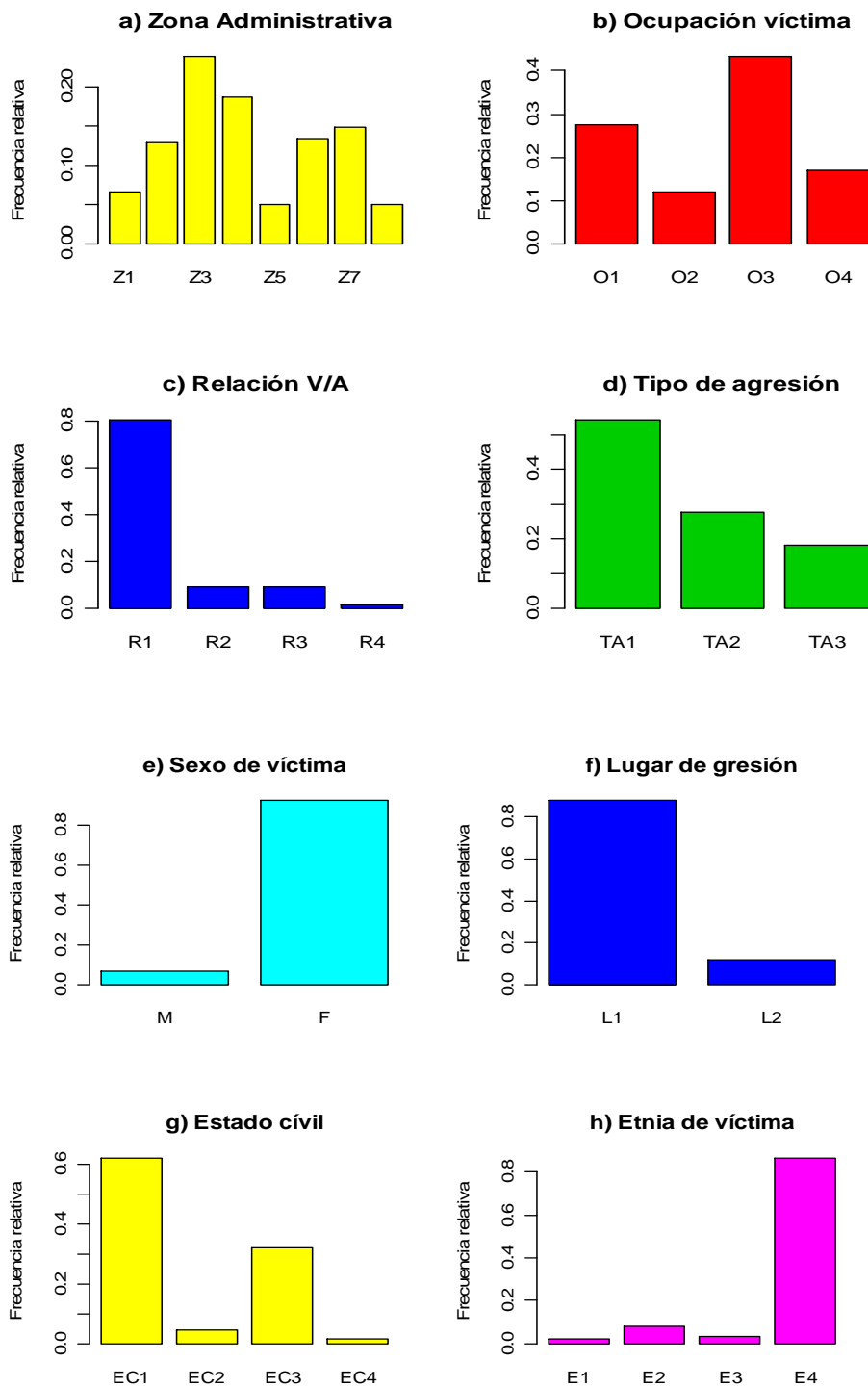
Las variables que describen este tipo de violencia, están asociadas con las denuncias que son receptadas en las comisarías de la Mujer y la Familia, estas denuncias se sustentan en la ley 103.

Tabla 2.5.1.- Descripción de variables de violencia intrafamiliar.-

NOMBRE DE VARIABLE	TIPO DE VARIABLE	MODALIDAD	CODIFICACIÓN
ZONA ADMINISTRATIVA	Cualitativa	Calderón Centro Eloy Alfaro La Delicia Los Chillos Norte Quitumbe Tumbaco	Z1 Z2 Z3 Z4 Z5 Z6 Z7 Z8
OCUPACIÓN	Cualitativa	Empleada Informal Ama de casa Otra.	O1 O2 O3 O4
RELACIÓN VÍCTIMA-AGRESOR	Cualitativa	Pareja. Expareja Familiar Otra.	R1 R2 R3 R4
TIPO DE AGRESIÓN	Cualitativa	Física. Psicológica. Sexual.	TA1 TA2. TA3
SEXO	Cualitativa	Masculino Femenino	M F
LUGAR DEL HECHO	Cualitativa	Domicilio Fuera del domicilio	L1 L2
ESTADO CIVIL	Cualitativa	Casado Divorciado Soltero Viudo	EC1 EC2 EC3 EC4
ETNIA	Cualitativa	Negra Blanca Indio Mestiza	E1 E2 E3 E4

## 2.5.1.- Gráficos de frecuencias.-

Figura 2.5.1.- Frecuencias relativas



En la figura se muestra el comportamiento de las variables que describen este tipo de violencia, los resultados son los siguientes:

- La zona Eloy Alfaro, La Delicia, Quitumbe, Norte y Centro son las que registran la mayor cantidad de denuncias.
- De acuerdo a la ocupación de la víctima se tiene que son las más de casa las que más denuncian algún tipo de maltrato, les siguen las empleadas domésticas, y víctimas que tiene otro tipo de ocupaciones.
- La pareja actual de la víctima o su conviviente, es la que comúnmente realiza alguno de los maltratos que se considera en la ley de la mujer y la familia, y es en la casa donde se efectúa el maltrato.
- EL tipo de maltrato que mayor registro presenta es el físico, seguido del psicológico y el sexual en menor frecuencia.
- La mayor frecuencia de agresiones se tiene en mujeres que en hombres, que en muchas ocasiones no denuncian oportunamente el maltrato, sino cuando es con mucha frecuencia, o es denunciado por terceras personas cercanas al hogar de la agredida.
- Es al interior del hogar donde se suscitan las agresiones, esto ase considerar, que siendo el hogar en muchas ocasiones el lugar que brinda seguridad a una familia, en el caso de maltrato intrafamiliar se convierte en el escenario donde se produce el maltrato.
- Según el estado civil de las víctimas, son las mujeres casadas las que registran la mayor cantidad de denuncias de maltrato.
- En general las mujeres que denuncian mayoritariamente los casos de maltrato, son de raza mestiza y en menor frecuencia las mujeres de raza blanca.

## **2.5.2.- Multivariante de violencia intrafamiliar.-**

### **2.5.2.1.- Análisis factorial de correspondencia múltiple.-**

**Tabla 2.5.2.- Valores propios**

<b>Dimensión</b>	<b>Autovalores</b>	<b>Inercia</b>	<b>% de la inercia explicada</b>	<b>% acumulado</b>
1	1,638	0,205	20,474	20,474
2	1,549	0,194	19,368	39,842
3	1,433	0,179	17,912	57,755
Total	4,620	0,578	57,755	
Media	1,540	0,193		

La tabla 2.5.2 muestran los valores propios del ACM que se asocian con este tipo de violencia. Los niveles de representatividad de los factores es baja, tal que entre los tres primeros factores se observa que tienen un 57% de representatividad de las modalidades.

**Tabla.-2.5.3.-**

Modalidades	Frecuencia marginal	Factor 1	Factor 2	Factor 3	Inercia modalidad
Z1	525	-0,4593	0,1626	-0,4731	0,1186
Z2	1021	0,0649	0,0317	0,4303	0,1125
Z3	1892	0,1623	-0,1658	0,4527	0,1018
Z4	1477	-0,2633	0,3199	-0,6139	0,1069
Z5	395	-0,0473	-0,5431	0,2072	0,1202
Z6	1057	0,4402	0,3867	-0,2856	0,1120
Z7	1179	-0,2283	-0,4243	0,5703	0,1105
Z8	400	-0,3483	-0,0126	-0,1044	0,1201
E1	178	-0,8002	1,6269	-0,8509	0,1228
E2	658	-0,0027	0,7493	-1,2313	0,1169
E3	267	-0,9649	-0,0506	-0,6468	0,1217
E4	6985	0,0336	-0,1227	0,1611	0,0394
O1	2283	0,6897	-0,2515	-0,5491	0,0970
O2	1010	0,3180	-0,3735	-0,4389	0,1126
O3	3603	-0,6476	0,1847	0,4803	0,0808
O4	1410	0,3256	0,4948	-0,2519	0,1077
TA1	5085	-0,1780	-0,0481	0,2234	0,0627
TA2	2591	0,7461	-0,3813	0,0810	0,0932
TA3	1696	-0,6571	0,7047	-0,9010	0,1042
R1	7474	-0,2992	-0,1863	-0,2194	0,0334
R2	840	1,5104	1,7556	0,2296	0,1147
R3	860	0,8067	-0,4255	1,6296	0,1145
R4	133	1,7721	0,8697	0,0674	0,1234
L1	8279	-0,1394	-0,1464	0,0621	0,0235
L2	1148	0,9811	1,0461	-0,6098	0,1109
M	724	2,2691	-1,9720	-0,9975	0,1161
F	9362	-0,1476	0,1522	0,0856	0,0102
EC1	6171	-0,1845	-0,5122	-0,3123	0,0493
EC2	449	1,7728	1,2932	0,2345	0,1195
EC3	3209	0,1472	0,8162	0,4000	0,0857
EC4	142	0,8804	-0,7630	4,5333	0,1233

La tabla 2.5.3 se presenta las coordenadas de las proyecciones de las modalidades y las inercias de estas. Existen muchas categorías que tienen inercias bajas, así, sexo de víctima femenino, lugar de agresión al interior del



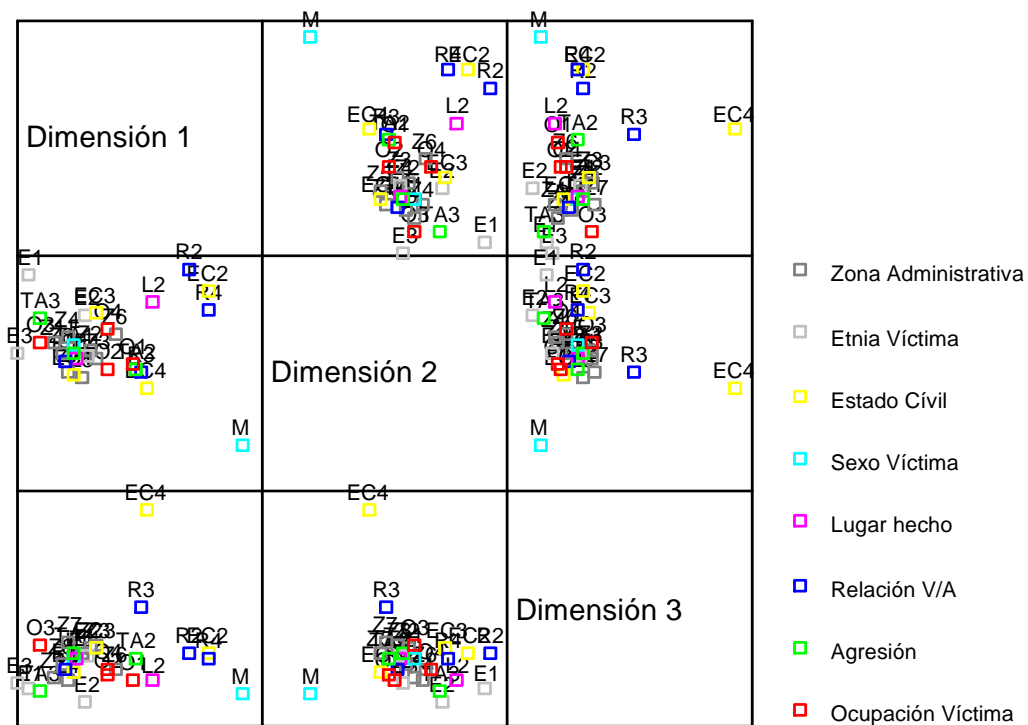
hogar, relación de la víctima con el agresor pareja o conviviente, etnia mestiza, estado civil casada y también en la modalidad soltera se observa una baja inercia. Es importante señalar que en las modalidades de zona administrativa, no se evidencia una inercia de alguna de estas bajas.

**Tabla 2.5.4.-**

Variables	Dimensión			Inercia por variable
	Factor 1	Factor 2	Factor 3	
Ocupación Víctima	0,28	0,07	0,17	0,18
Agresión	0,23	0,12	0,17	0,17
Relación V/A	0,35	0,30	0,25	0,30
Lugar hecho	0,12	0,14	0,05	0,10
Sexo Víctima	0,39	0,30	0,08	0,25
Estado Civil	0,18	0,45	0,39	0,34
Etnia Víctima	0,04	0,09	0,15	0,09
Zona Administrativa	0,06	0,07	0,18	0,10

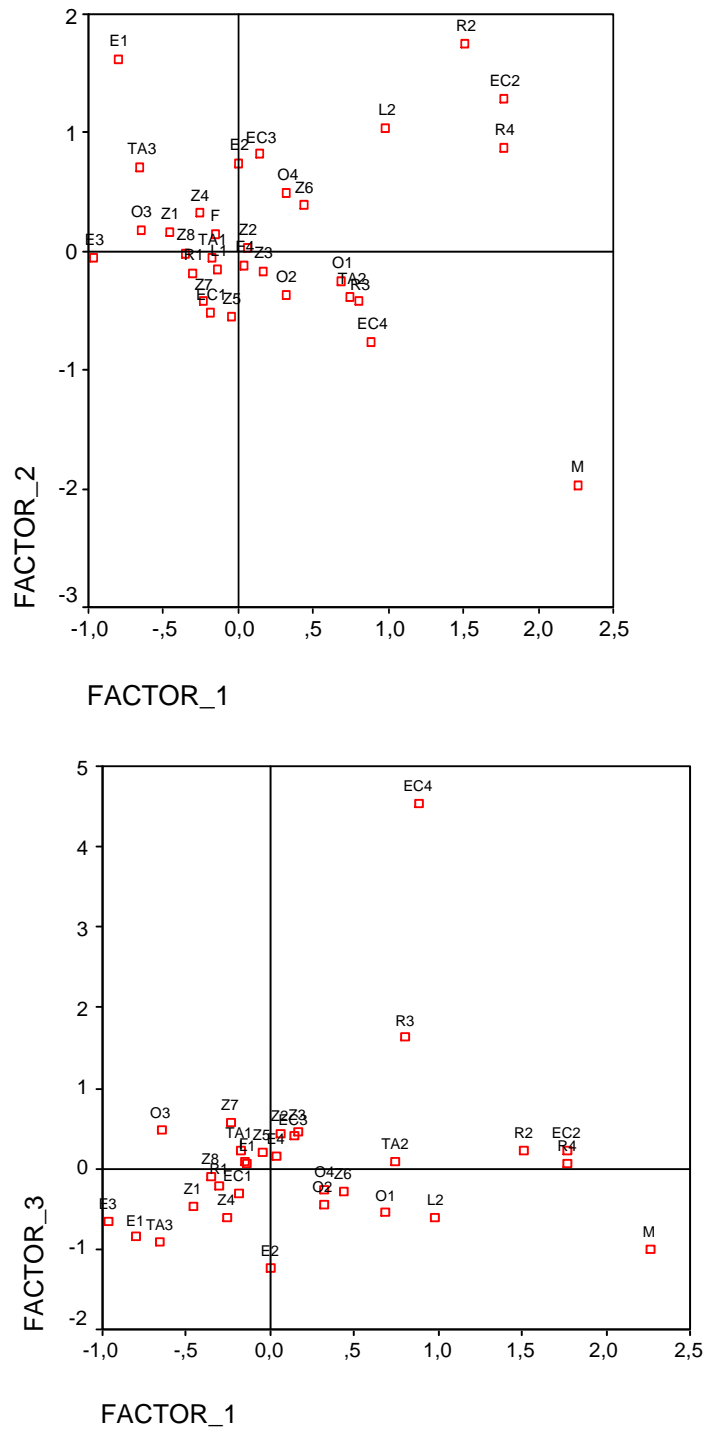
Al igual que la tabla anterior en la tabla 2.5.4, muestra también las coordenadas de las proyecciones de las variables y la inercia de cada una de estas. Las variables de mayor inercia son estado civil, relación de la víctima con el agresor y sexo de la víctima, mientras que las de menor inercia son etnia de la víctima y lugar del hecho.

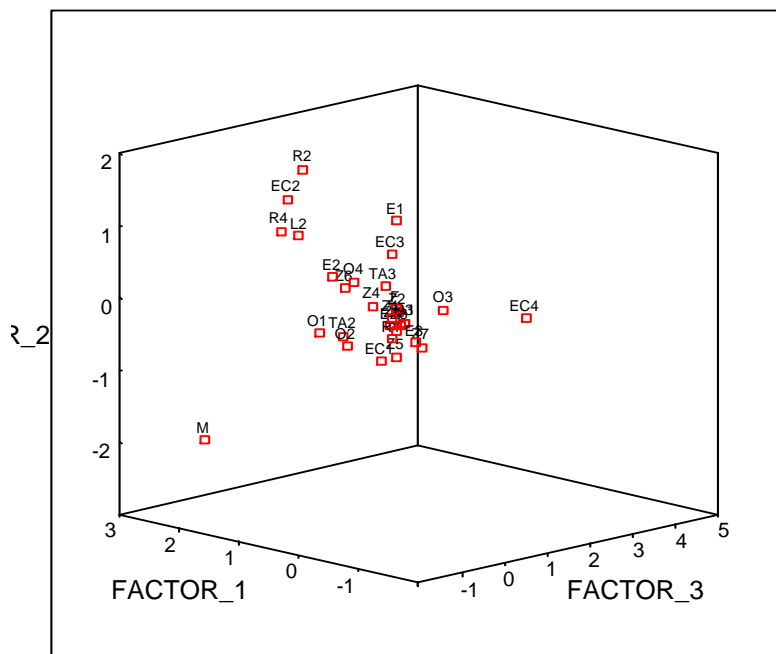
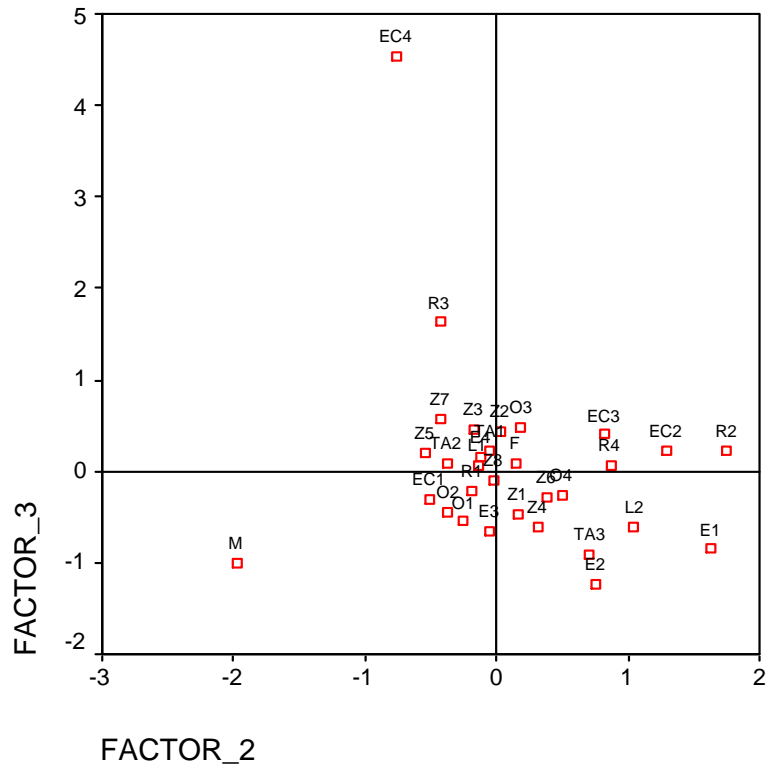
Figura 2.5.2.- Relación entre categorías.-



De las posiciones de las categorías de las variables, en los tres factores que se representa en la figura, no se puede observar las relaciones entre estas adecuadamente, por lo que se procede a graficar los factores en figuras separadas y en dos dimensiones.

Figura 2.5.3.- Comportamiento de las modalidades por factores.-





Con esta representación se tiene una mejor visualización del comportamiento de las categorías, y se puede indicar los siguientes resultados:

- Cuando la víctima es de estado civil divorciada, sufre agresiones por parte de su ex pareja y de agresores que tiene otro tipo de relación con las víctimas como conocidos o amigos, tanto las agresiones realizadas por la ex pareja y por conocidos o amigos, se realizan fuera del hogar.
- Las mujeres indígenas y de raza negra son propensas a sufrir maltrato sexual y por lo general la ocupación que tienen son las de amas de casa.
- Las víctimas de estado civil viudas, sufren de agresión psicológica, siendo los agresores, sus familiares, y las ocupaciones que tienen estas víctimas son empleadas<sup>3</sup> domésticas y trabajadoras informales.
- La agresión física se asocia cuando el agresor es la pareja, en el interior del hogar, y el estado civil es casada.

## **2.6.- Análisis de maltrato infantil.-**

### **2.6.1.- Gráficos de frecuencia relativa.-**

Las variables que describen este tipo de violencia, se toma de las denuncias de maltrato infantil que se registran en la DINAPEN, y que se tipifica en el código de la niñez y la adolescencia.

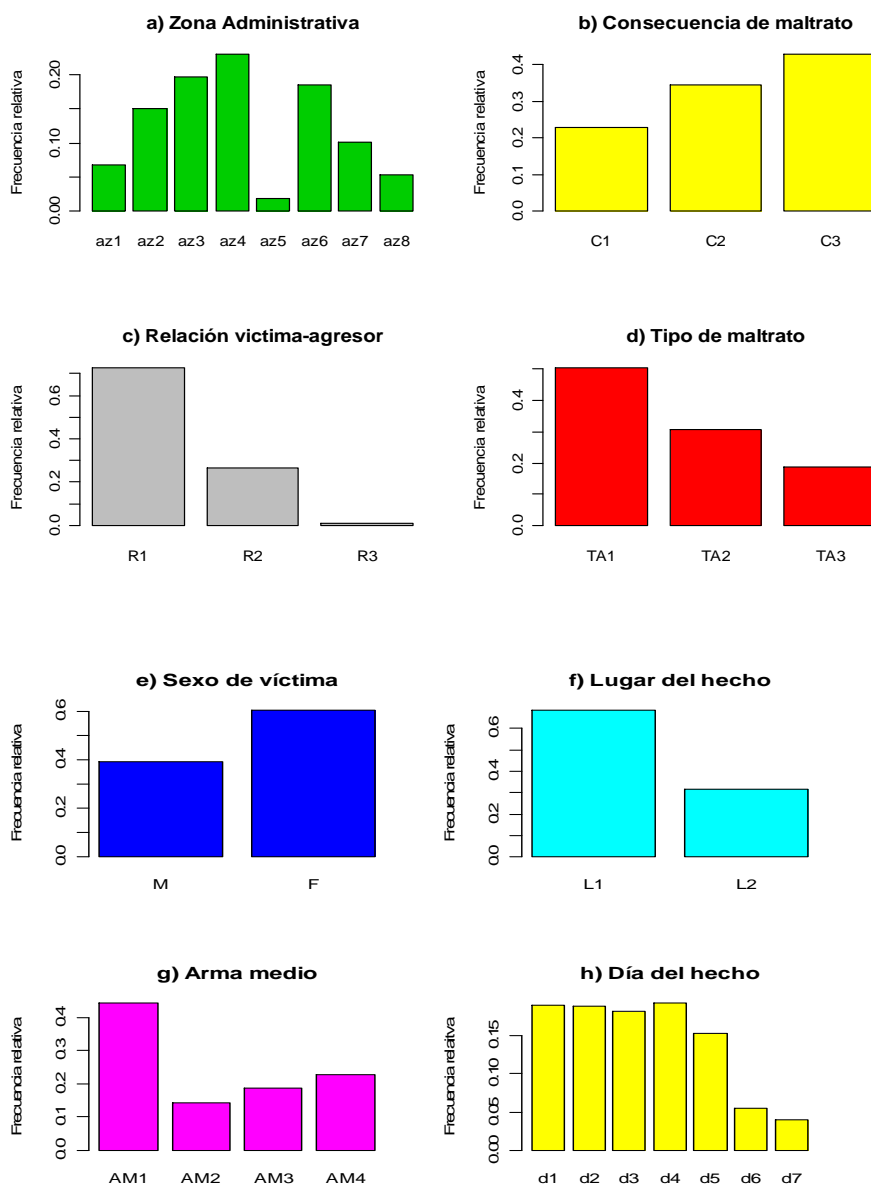
---

<sup>3</sup> Según la ley 103 de la mujer y la familia, una empleada doméstica que trabaja puertas adentro forma parte también de la familia, y la ley se extiende hacia ellas.

Tabla 2.6.1.- Descripción de variables de maltrato infantil.-

NOMBRE DE VARIABLE	TIPO DE VARIABLE	MODALIDADES	CODIFICACIÓN
ZONA ADMINISTRATIVA	Cualitativa	Calderón Centro Eloy Alfaro La Delicia Los Chillos Norte Quitumbe Tumbaco	Z1 Z2 Z3 Z4 Z5 Z6 Z7 Z8
CONSECUENCIA	Cualitativa	Lesión leve Lesión grave Perturbación emocional	C1 C2 C3
RELACIÓN VÍCTIMA-AGRESOR	Cualitativa	Padres Familiar Ninguna	R1 R2 R3
TIPO DE AGRESIÓN	Cualitativa	Física. Psicológica. Sexual.	TA1 TA2 TA3
EDAD	Cuantitativa discreta		0-18 años
SEXO	Cualitativa	Masculino Femenino.	M F
LUGAR DEL HECHO	Cualitativa	Domicilio, Fuera del domicilio	L1 L2
ARMA-MEDIO	Cualitativa	Golpes Contundente Abuso sexual Otro	AM1 AM2 AM3 AM4

**Figura 2.6.1.- Frecuencias relativas.-**



En la figura se tiene una representación gráfica de las variables categóricas que describen el maltrato infantil, teniéndose los siguientes resultados:

- Las zonas administrativas que más denuncias registran son Centro, Eloy Alfaro, La Delicia y Norte.
- La consecuencia que más frecuente ocurre es la perturbación emocional seguida de la lesión grave y lesión leve.
- Son los padres quienes realizan la mayoría de los maltratos, seguidos de los familiares de las víctimas.

- El maltrato que más es denunciado es el físico, luego el psicológico y por último el sexual.
- La mayor cantidad de denuncias de maltrato infantil se tiene en personas del sexo femenino, y es en la casa donde se efectúan en la mayoría de los casos el maltrato.
- Los golpes son el medio de maltrato más común que es denunciado, y el abuso sexual es también denunciado como medio de maltrato.
- En relación a los días de ocurrencia se observa que de lunes a viernes se tiene la mayor cantidad de maltrato.

## **2.6.2.- Multivariante maltrato infantil.-**

### **2.6.2.1.- Análisis de correspondencias múltiples.-**

**Tabla 2.6.2.- Valores propios**

<b>Dimensión</b>	<b>Autovalores</b>	<b>Inercia</b>	<b>% de la inercia explicada</b>	<b>% acumulado</b>
1	2,713	0,388	38,751	38,751
2	2,622	0,375	37,464	76,214
3	1,469	0,210	20,985	97,200
Total	6,804	0,972	97,200	
Media	2,268	0,324		

En la tabla 2.6.2 se muestran los valores propios asociados al ACM, que se utilizan en la determinación de los factores. Así los tres primeros factores recogen el 97% de la representación de las modalidades.



Tabla 2.6.3.-

Modalidades	Frecuencia marginal	Factor 1	Factor 2	Factor 3	Inercia modalidad
M	662	0,0925	-0,4913	-0,1254	0,0868
F	1025	-0,0846	0,3137	0,0787	0,0561
L1	1126	0,4070	-0,1247	-0,2492	0,0475
L2	516	-0,8811	0,2361	0,5502	0,0992
TA1	851	0,8310	-0,1918	0,1619	0,0708
TA2	517	-1,0795	-0,8161	-0,0568	0,0991
TA3	319	-0,5471	1,8229	-0,3472	0,1158
AM1	741	0,5110	-0,2929	0,7251	0,0801
AM2	238	1,0957	0,0063	-1,1570	0,1227
AM3	315	-0,5545	1,8377	-0,3452	0,1162
AM4	382	-1,2586	-0,9443	-0,4024	0,1105
C1	381	0,9614	-0,3480	1,0058	0,1106
C2	576	0,4484	0,7016	-0,6370	0,0941
C3	718	-0,8876	-0,3755	-0,0227	0,0821
R1	1201	0,2189	-0,3416	-0,2845	0,0412
R2	435	-0,5271	0,8885	0,7399	0,1060
R3	18	-0,8500	0,8581	1,0345	0,1413
E1	128	-1,3419	-1,1783	-0,2272	0,1320
E2	275	0,0849	-0,2116	-0,5050	0,1196
E3	501	0,3092	-0,0401	-0,4402	0,1004
E4	706	0,0725	0,3665	0,5547	0,0831

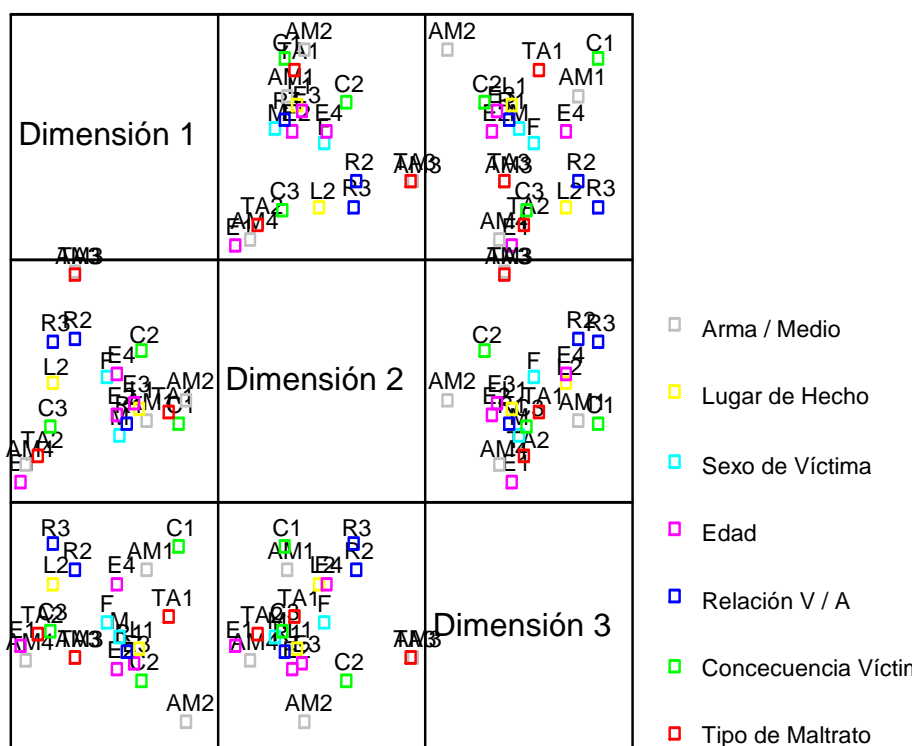
La tabla 2.6.3, muestra las coordenadas de las proyecciones de las modalidades en tres factores y las inercias de cada modalidad. Las categorías que tienen la menor inercia son, relación de la víctima con el agresor padres, el lugar de la agresión el hogar, tipo de maltrato el físico y el sexo de la víctima femenino, aunque no difiere mucho del sexo masculino.

Tabla 2.6.4.-

Variables	Dimensión			Inercia por variable
	Factor 1	Factor 2	Factor 3	
Tipo de Maltrato	0,762109678	0,851045259	0,038003234	0,55038606
Sexo de Víctima	0,007702513	0,15453516	0,00940338	0,05721368
Lugar de Hecho	0,347993733	0,027424002	0,137204054	0,17087393
Arma / Medio	0,700180806	0,870223024	0,4730621	0,68115531
Consecuencia Víctima	0,612680267	0,25543693	0,363548745	0,41055531
Relación Víctima / Agresor	0,113473036	0,294466775	0,212894868	0,20694489
Rango de edad	0,168397628	0,169325512	0,234866429	0,19086319

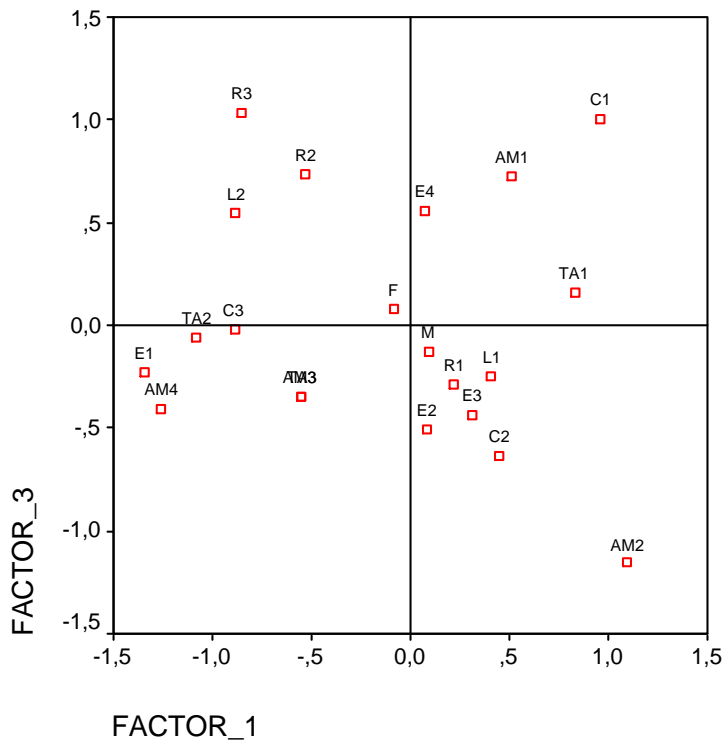
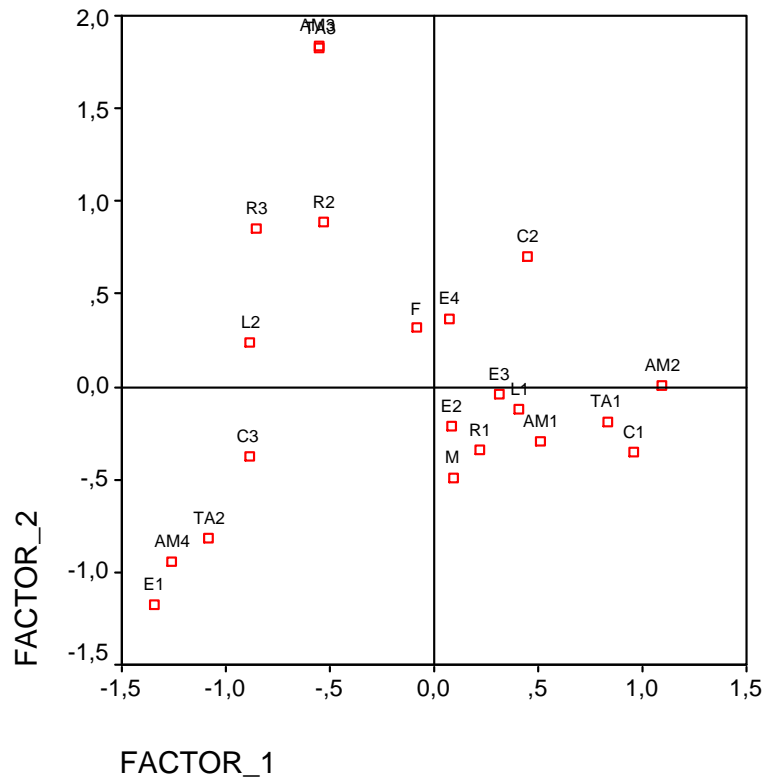
Al igual que en la tabla anterior, se muestra en la tabla 2.6.4, las coordenadas de las proyecciones de las variables en tres factores y las inercias de estas. Las variables que tienen mayor inercia son arma medio, tipo de maltrato, consecuencia de la víctima.

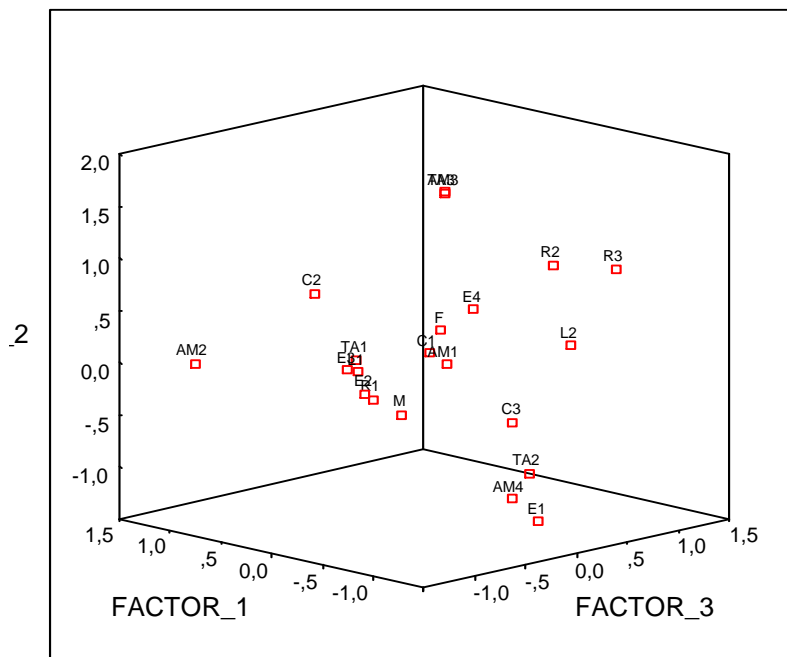
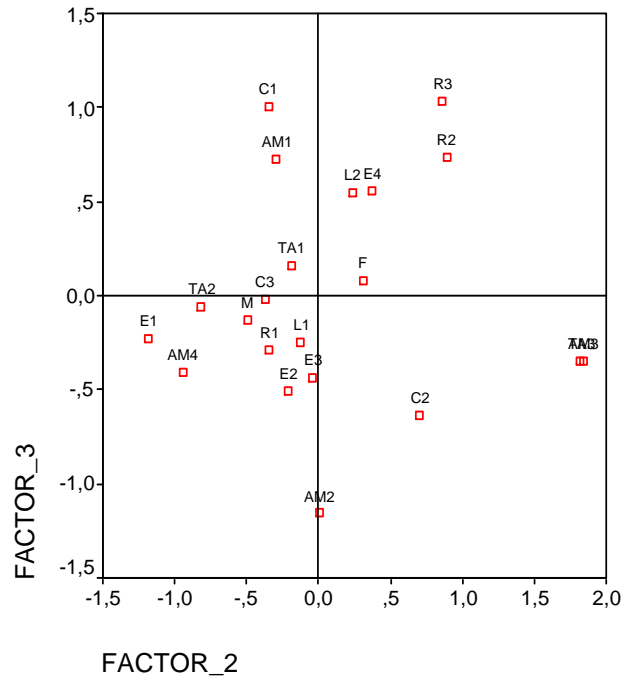
Figura 2.6.2.- Relación entre categorías.-



Los resultados de la figura anterior se presentan por los tres factores, como se observa, no se tiene una representación adecuada de las categorías, por esto, se procede a realizar gráficos separados por cada dos factores.

Figura 2.6.3.- Comportamiento de las modalidades por factores.-





Ahora se obtiene del comportamiento de las categorías de las variables asociadas con el maltrato infantil son las siguientes:

- Los niños menores a un año de edad, sufren de maltrato psicológico, teniendo como consecuencia la perturbación emocional, los medios son burlas, agredir su autoestima, etc.
- Las mujeres tiene mayor riesgo de maltrato en edades de 11 a 18 años, las agresiones por lo general son fuera del hogar y con lesiones graves en la víctima.
- El maltrato sexual lo realizan familiares y agresores que no tienen relación con las víctimas, estas agresiones por lo general lo cometen fuera del hogar de la víctima y obviamente causan lesiones graves.
- Los maltratos físicos se los realizan con armas contundentes y golpes, esta agresión causa lesiones leves en la víctima.
- En los niños varones, de edades de 2 a 5 años y de 6 a 10 años, las agresiones se las cometen en el hogar por medio de golpes, siendo los agresores los propios padres de las víctimas.

## **Capítulo 3.**

# **Construcción de modelos logísticos binarios y multinomiales.**

En este capítulo se procede a construir los modelos logísticos, binarios y multinomiales, los que serán base de la determinación de los índices de violencia. Todos estos modelos están sustentados en las variables que fueron determinadas al realizar el análisis exploratorio univariante y multivariante, de los diferentes tipos de violencia, los que se estudiaron en el capítulo anterior.

Además de esto se realizará, la imputación de valores perdidos (en los modelos que se requiera realizar este procedimiento); para esto nos basamos en la matriz de diseño de valores perdidos y el procedimiento de estimación para datos cualitativos y para tablas de contingencia (ver ANEXO B).

Iniciamos este capítulo, con la presentación de la estimación de los modelos logísticos multinomiales, siguiendo las estimaciones de los modelos: muertes violentas, delitos a vehículos, delitos a la propiedad, violencia intrafamiliar y maltrato infantil.

### 3.1.- Modelo logístico multinomial.-

El modelo logístico multinomial (o multicategorico) es una generalización del modelo logístico binomial: para esto tenemos las siguientes consideraciones:

Sea  $Y$  una variable dependiente (o respuesta) nominal u ordinal, con más de dos categorías y con distribución multinomial.

Se tienen variables explicativas, las cuales pueden ser nominales ordinales, categóricas o continuas.

#### 3.1.1.-Distribución multinomial.-

Sea  $Y$  la variable respuesta con  $J$  categorías.

Sean  $\{\pi_1, \dots, \pi_J\}$  las probabilidades de respuesta de las categorías, tal que se

cumple  $\sum_{j=1}^J \pi_j = 1 \quad \forall j, \quad \pi_j \geq 0$ .

Se tiene  $n$  observaciones independientes de  $Y$ , con:  $y_1$  el número de respuestas de la categoría 1,  $y_2$  el número de respuestas de la categoría 2 y así sucesivamente; entonces se tiene:

$$Y = \begin{bmatrix} y_1 \\ \vdots \\ y_J \end{bmatrix}, \text{ con } \sum_{j=1}^J y_j = n$$

La distribución multinomial se define como

$$f(y|n) = \frac{n!}{y_1! \dots y_J!} \pi_1^{y_1} \dots \pi_J^{y_J} \quad (1)$$

Es claro que si  $J=2$ , (1) es la distribución binomial.

Sea  $Y_1, \dots, Y_J$  variables aleatorias independientes con distribución de Poisson( $\theta_j$ ). La función de distribución conjunta

$$f(y) = \prod_{j=1}^J \frac{\theta_j^{y_j}}{y_j!} \text{ donde } Y = \begin{bmatrix} y_1 \\ \vdots \\ y_J \end{bmatrix}, \text{ con } \sum_{j=1}^J y_j = n$$

Sea  $n = Y_1 + \dots + Y_J$ ,  $n$  tiene una distribución  $Poisson(\theta_1 + \dots + \theta_J)$  (Kalbfleisch, 1985, página 142)

Con esto, la distribución condicional<sup>4</sup> de  $Y$  dado  $n$  es

---

<sup>4</sup> El resultado siguiente (2) se obtiene por la definición de distribución condicional

$$f(y|n) = \left[ \prod_{j=1}^J \frac{\theta_j^{y_j} e^{-\theta_j}}{y_j!} \right] / \frac{(\theta_1 + \dots + \theta_J)^n e^{-(\theta_1 + \dots + \theta_J)}}{n!} \quad (2)$$

La distribución anterior es equivalente (Dobson, 2001, Capítulo 8) a:

$$f(y|n) = \left( \frac{\theta_1}{\sum \theta_k} \right)^{y_1} \dots \left( \frac{\theta_J}{\sum \theta_k} \right)^{y_J} \frac{n!}{y_1 \dots y_J!} \quad (3)$$

Observación.-

- Si  $\pi_j = \frac{\theta_j}{\left( \sum_{k=1}^K \theta_k \right)}$ ,  $j = 1, \dots, J$ , entonces, (3) es la misma que (1)
- $\sum_{j=1}^J \pi_j = 1$

$$E(Y_j) = n\pi_j$$

- Además se tiene que:  $Var(Y_j) = n\pi_j(1 - \pi_j)$   
 $Cov(Y_j, Y_k) = -n\pi_j\pi_k$  (Agresti 2000)

### 3.1.2.- Regresión logística multinomial.-

Cualquier categoría de la variable respuesta ( $Y$ ), es elegida como la categoría de referencia. Supondremos que es la primera categoría. El modelo logit para las categorías restantes se define por:

$$\text{logit}(\pi_j) = \log\left(\frac{\pi_j}{\pi_1}\right) = x_j' \beta_j \text{ para } j = 2, \dots, J, \quad (4)$$

Donde  $x_j$  es una fila de la matriz de diseño y  $\beta_j$  es el vector de los parámetros

Con la expresión de (4), se tiene  $J-1$  ecuaciones *logit* simultáneas, con lo que se obtiene con las primeras derivadas de las  $J-1$  ecuaciones los estimadores

$\hat{\beta}_j$  de  $\beta_j$ , además de esto se prueba la hipótesis nula de

$$\begin{cases} H_0 : \hat{\beta}_j = 0 & j = 1, \dots, J-1 \\ H_1 : \hat{\beta}_j \neq 0 \end{cases}$$

Con las estimaciones anteriores, se tiene las probabilidades estimadas como:

$$\hat{\pi}_j = \hat{\pi}_1 \exp\left(x_j' \hat{\beta}_j\right), \quad j = 2, \dots, J \quad (5)$$

Pero se sabe que  $\hat{\pi}_1 + \dots + \hat{\pi}_j = 1$ , con esto se tiene



$$\hat{\pi}_1 = \frac{1}{1 + \sum_{j=2}^J \exp\left(x_j' \hat{\beta}_j\right)} \quad (6)$$

De donde,

$$\hat{\pi}_j = \frac{\exp\left(x_j' \hat{\beta}_j\right)}{1 + \sum_{j=2}^J \exp\left(x_j' \hat{\beta}_j\right)} \quad (7)$$

### 3.1.3.- Residuos de Pearson Chi-cuadrado.-

$$r_i = \frac{o_i - e_i}{\sqrt{e_i}} \quad (8)$$

$o_i$  = frecuencia observada

$e_i$  = frecuencia esperada

Los estadísticos de validación son similares a los de regresión logística binomial.

#### 3.1.3.1.- Estadístico Chi-cuadrado.-

$$X^2 = \sum_{i=1}^N r_i^2 \quad (9)$$

#### 3.1.3.2.- Desviación.-

Se define en términos de valores máximos de la función log-verosimilitud, para el modelo estimado y el modelo maximal<sup>5</sup>.

$l(\hat{\beta})$  para el modelo estimado

$l(\hat{\beta}_{max})$  para el modelo maximal

$$D = 2[l(\hat{\beta}_{max}) - l(\hat{\beta})] \quad (10)$$

#### 3.1.3.3.- Razón de verosimilitud del estadístico Chi-cuadrado.-

Se define en base al valor máximo de la función log-verosimilitud para el modelo minimal<sup>6</sup>,  $l(\hat{\beta}_{min})$  y  $l(\hat{\beta})$ , se define.

$$C = 2[l(\hat{\beta}) - l(\hat{\beta}_{min})] \quad (11)$$

#### 3.1.3.4.-

<sup>5</sup> El modelo maximal se refiere a que en el modelo a más de los parámetros de las estimaciones de las categorías de las variables se contempla también las estimaciones de las interacciones de las categorías de estas en base a (4).

<sup>6</sup> Modelo minimal se refiere a que en el modelo estimado solo se tiene el primer estimador, esto es la constante en cada estimación en base a (4).

$$R^2 = \frac{l(\hat{\beta}_{min}) - l(\hat{\beta})}{l(\hat{\beta}_{min})} \quad (12)$$

$X^2, D \sim X^2_{(N-p)}$ , con  $p$  el número de parámetros estimados.

$C \sim X^2_{(p-(J-1))}$ , el modelo minimal, tiene un parámetro para cada logit.

El modelo se interpreta en términos de la disparidad ( $OR$ ). Consideremos  $Y$  variable respuesta con  $J$  categorías.

$$X = \begin{cases} 1 & \text{presencia} \\ 0 & \text{ausencia} \end{cases}$$

$$OR_j = \frac{\pi_{jp} / \pi_{1p}}{\pi_{ja} / \pi_{1a}}$$

Donde,  $\pi_{jp}$  y  $\pi_{ja}$ , son probabilidades de respuesta de la categoría  $j$ , de acuerdo a la relación si es presencia o ausencia.

Para el modelo:

$$\log\left(\frac{\pi_j}{\pi_1}\right) = \beta_{0j} + \beta_{1j}x, \quad j = 2, \dots, J \quad (13)$$

El logaritmo de la disparidad es

$$\log\left(\frac{\pi_{ja}}{\pi_{1a}}\right) = \beta_{0j}, \quad \text{cuando } x=0 \text{ indica ausencia de exposición y}$$

$$\log\left(\frac{\pi_{jp}}{\pi_{1p}}\right) = \beta_{0j} + \beta_{1j}, \quad \text{cuando } x=1 \text{ indica presencia de exposición}$$

Ahora:

$$\log(OR_j) = \log\left(\frac{\pi_{jp}}{\pi_{1p}}\right) - \log\left(\frac{\pi_{ja}}{\pi_{1a}}\right) = \beta_{1j} \quad (14)$$

Luego,  $OR_j = \exp(\beta_{1j})$ , la cual tiene un intervalo de confianza dado por:

$$\exp[b_{1j} \pm 1,96 * s.e(\hat{\beta}_{1j})] \quad (15)$$

### 3.2.- Índice para muertes violentas.-

A continuación estimamos el modelo logístico multinomial, para describir las muertes violentas; para esto utilizaremos las siguientes variables, las cuales fueron definidas con el análisis univariante y multivariante que se realizó en el capítulo 2.

Tabla 3.1.- Definición de variables

NOMBRE DE VARIABLE	TIPO DE VARIABLE	MODALIDADES	CODIFICACIÓN
TIPO DE MUERTE (variable respuesta)	Cualitativa	Homicidios- Suicidios Accidentes de tránsito Accidentales	MV1 MV2 MV3
SEXO (variable explicativa)	Cualitativa	Masculino Femenino	M F
RANGO EDAD (años) (variable explicativa)	Cualitativa	0-19. 20-39. 40-59. >60.	E1= 0-19. E2= 20-39. E3= 40-59. E4= >60.

Además de la tabla de variables, se muestra la matriz de diseño de los valores perdidos, los cuales son estimados con los procedimientos definidos para estos (ver anexo B).

Para la imputación de valores perdidos se parte de la matriz de diseño<sup>7</sup> de los datos perdidos de las variables tipo de muerte, zona administrativa, clasificación horaria y rangos de edad.

### 3.2.1.- Estimación del modelo.-

Iniciamos el modelo estimándolo en función de las variables explicativas y la variable respuesta o dependiente (la variable dependiente es el tipo de muerte).

$$\begin{cases} H_0 : \hat{\beta}_j = 0 & j = 1, \dots, J - 1 \\ H_1 : \hat{\beta}_j \neq 0 \end{cases}$$

<sup>7</sup> Sea:

$Y = (y_{ij})$ , la matriz  $(n \times K)$  sin valores perdidos.

$$M = (m_{ij}) \text{ con } \begin{cases} m_{ij} = 1 & \text{si } y_{ij} \text{ es perdido} \\ m_{ij} = 0 & \text{si } y_{ij} \text{ observado} \end{cases}$$

La matriz  $M$  define el diseño de los datos perdidos.

Tabla 3.2.- Estimación de parámetros

Estimaciones de los parámetros								
Tipo de Muerte	Categorías	B	Error típ.	gl	Sig.	Exp(B)	Intervalo de confianza al 95% para Exp(B)	
							Límite inferior	Límite superior
Homicidio-Suicidio	Intersección	-0,33	0,20	1	0,1100805			
	0-19	0,74	0,21	1	0,0004236	2,092301	1,39	3,15
	20-39	1,19	0,19	1	5,858E-10	3,301837	2,26	4,82
	40-59	0,41	0,21	1	0,0453231	1,508215	1,01	2,26
	Masculino	0,06	0,14	1	0,6381958	1,065606	0,82	1,39
Acc. tránsito	Intersección	1,27	0,16	1	0,000			
	0-19	-0,57	0,17	1	0,001096	0,565507	0,40	0,80
	20-39	-0,35	0,16	1	0,0252629	0,704541	0,52	0,96
	40-59	-0,58	0,17	1	0,0004852	0,559699	0,40	0,78
	Masculino	-0,29	0,13	1	0,0219741	0,749076	0,59	0,96

En la tabla 3.2, se muestran las estimaciones de los parámetros para cada una de las categorías de las variables cualitativas, considerando las categorías de referencia para la variable dependiente es la muerte accidental, se considera también la redundancia de las variables en el modelo y los grados de libertad. Además de esto, se muestra también los niveles de significación de los parámetros. En esta fase nos focalizamos en determinar el modelo; la validación se realizará más adelante, el modelo estimado es el siguiente:

$$x_1 = \begin{cases} 1 & \text{si } 0 - 19 \text{ años} \\ 0 & \text{si no} \end{cases}, x_2 = \begin{cases} 1 & \text{si } 20 - 39 \text{ años} \\ 0 & \text{si no} \end{cases}, x_3 = \begin{cases} 1 & \text{si } 40 - 59 \text{ años} \\ 0 & \text{si no} \end{cases},$$

$$x_4 = \begin{cases} 1 & \text{hombre} \\ 0 & \text{mujer} \end{cases}$$

$$\ln\left(\frac{\pi_1}{\pi_3}\right) = -0.327 + 0.738x_1 + 1.194x_2 + 0.411x_3 + 0.064x_4 \text{ y}$$

$$\ln\left(\frac{\pi_2}{\pi_3}\right) = 1.271 - 0.57x_1 - 0.35x_2 - 0.58x_3 - 0.289x_4$$

### 3.2.2.- Validación del modelo.-

Lo que sigue es la validación del modelo que se estimó anteriormente; para esto, se prueba la hipótesis siguiente:

$\left\{ \begin{array}{l} H_0: \text{Los coeficientes son cero} \\ H_1: \text{Los coeficientes no son cero} \end{array} \right.$

Figura 3.2.- Estadísticos del modelo estimado

Información del ajuste del modelo				
Modelo	-2 log verosimilitud	Chi-cuadrado	gl	Sig.
Sólo la intersección	252,482			
Final	95,528	156,954	8	,000

Iniciamos la validación mostrando el estadístico de máxima verosimilitud sobre el modelo, como muestra la figura anterior, se rechaza la hipótesis nula de que los coeficientes finales del modelo son cero. Ahora, probamos la hipótesis siguiente, sobre el modelo estimado y sobre los efectos:

$\left\{ \begin{array}{l} H_0: \text{El modelo es adecuado} \\ H_1: \text{El modelo no es adecuado} \end{array} \right.$

Figura 3.3.- Pruebas estadísticas sobre las variables independientes

Bondad de ajuste			
	Chi-cuadrado	gl	Sig.
Pearson	5,594	6	,470
Desviación	5,588	6	,471

Como se muestra en la figura, las pruebas de bondad de ajuste, indican que el modelo estimado es el adecuado, ya que la desviación y el estadístico de Pearson, son similares y no rechazan la hipótesis nula.

Una vez que se muestran los estadísticos sobre la estimación del modelo, se procede a estudiar el comportamiento de los residuos.

Figura 3.4.- Prueba no paramétrica sobre los residuos

Prueba de Kolmogorov-Smirnov para una muestra			RESIDUOS
N			48
Parámetros normales <sup>a,b</sup>	Media		-,0039
	Desviación típica		,96567
Diferencias más extremas	Absoluta		,076
	Positiva		,076
	Negativa		-,069
Z de Kolmogorov-Smirnov			,526
Sig. asintót. (bilateral)			,945

a. La distribución de contraste es la Normal.

b. Se han calculado a partir de los datos.

En el cuadro anterior se presenta la prueba de Kolmogorov-Smirnov que nos permite probar si los residuos siguen una distribución normal, esto es:

$$\left\{ \begin{array}{l} H_0: \text{La distribución de contraste es normal} \\ H_1: \text{La distribución de contraste es diferente a la normal} \end{array} \right.$$

Es claro que la significación asintótica es mayor que 0,05, lo que nos indica que no se debe rechazar la hipótesis nula (asumiendo un nivel de significación de 0,05).

Figura 3.5.-



La figura muestra cómo se comportan los residuos en la recta de la normal, es claro que tienen una tendencia lineal, al igual que el comportamiento de los mismos.

### 3.2.3.- Interpretación del modelo.-

Una vez que se ha estimado y validado el modelo lo que nos queda es interpretar los resultados, esto es, interpretar la variable dependiente en función de las variables explicativas; para esto nos basamos en la tabla 3.2 y en la columna séptima, donde constan las disparidades.

Para los homicidios y suicidios respecto de las muertes accidentales, se tiene que las edades de menos de 18 años y de 19 a 39 años tienen el mayor riesgo de sufrir este tipo de violencia, al igual que las personas de sexo masculino.

Para las muertes en accidentes de tránsito respecto de las muertes accidentales, no se tiene un comportamiento diferenciado por rango de edad, ni por el sexo de la víctima.

### 3.3.- Modelo para delitos contra los vehículos.

Enseguida se procede a estimar el modelo para los delitos contra los vehículos, para esto utilizaremos las variables siguientes:

Tabla 3.2.- Definición de las variables.-

NOMBRE DE VARIABLE	TIPO DE VARIABLE	MODALIDAD	CODIFICACIÓN
TIPO DE VEHÍCULO (Variable respuesta)	Cualitativa	Automóvil Camioneta Todo terreno Motocicleta	TV1 TV2 TV3 TV4
LUGAR (variable explicativa)	Cualitativa	Vía pública Estacionamiento	L1 L2
CIRCUNSTANCIA (variable explicativa)	Cualitativa	Estacionado Interceptado.	C1 C2

Para la construcción del modelo utilizaremos la variable cualitativa tipo de vehículo como la variable respuesta (o dependiente), y el resto de variables de la tabla anterior como variables explicativas (o independientes).

#### 3.3.1.- Estimación del modelo.-

Iniciamos la construcción del modelo con la estimación, esto es, definir los valores de los parámetros que intervendrán en el mismo.

$$\begin{cases} H_0 : \hat{\beta}_j = 0 & j = 1, \dots, J - 1 \\ H_1 : \hat{\beta}_j \neq 0 \end{cases}$$

Figura 3.4.- Definición de parámetros del modelo

Estimaciones de los parámetros								
Tipo vehículo	Categorías	B	Error típ.	gl	Sig.	Exp(B)	Intervalo de confianza al 95% para Exp(B)	
							Límite inferior	Límite superior
Automóvil	Intersección	0,2770	0,1273	1	0,02951876			
	Vía pública	0,8983	0,1480	1	1,2723E-09	2,4555	1,8373	3,2816
	Estacionado	0,0279	0,1777	1	0,87502588	1,0283	0,7259	1,4567
Camioneta	Intersección	-0,3025	0,1481	1	0,04105235			
	Vía pública	0,8832	0,1699	1	2,0191E-07	2,4187	1,7336	3,3747
	Estacionado	-0,4767	0,2128	1	0,02507668	0,6208	0,4091	0,9421
Todo terreno	Intersección	-1,1273	0,1968	1	1,0178E-08			
	Vía pública	0,9798	0,2207	1	8,9843E-06	2,6640	1,7286	4,1055
	Estacionado	-1,3166	0,3466	1	0,00014555	0,2680	0,1359	0,5287

La figura anterior muestra las estimaciones de los parámetros estimados en el modelo; además de esto muestra el error típico de cada parámetro no redundante, los grados de libertad, el nivel de significación y las disparidades, incluyendo sus intervalos de confianza. En esta parte lo que nos interesa es la estimación del modelo, el cual toma la categoría de referencia tv4 (delitos a motocicletas), y en base a esta categoría estimara los modelos logísticos para las demás categorías, el modelo estimado es el siguiente:

$$x_1 = \begin{cases} 1 & \text{Calle} \\ 0 & \text{si no} \end{cases}, x_2 = \begin{cases} 1 & \text{estacionado} \\ 0 & \text{si no} \end{cases},$$

$$\ln\left(\frac{\pi_1}{\pi_4}\right) = 0.277 + 0.89x_1 + 0.028x_2$$

$$\ln\left(\frac{\pi_2}{\pi_4}\right) = -0.303 + 0.883x_1 - 0.477x_2$$

$$\ln\left(\frac{\pi_3}{\pi_4}\right) = -1.127 + 0.98x_1 - 1.317x_2$$

### 3.3.2.- Validación del modelo.-

Ahora lo que sigue es validar y aceptar el modelo estimado. Para esto se prueba la hipótesis nula siguiente:



$\left\{ \begin{array}{l} H_0: \text{El modelo es adecuado} \\ H_1: \text{El modelo no es adecuado} \end{array} \right.$

Figura 3.6.- Pruebas estadísticas sobre las variables que intervienen en el modelo.-

	Chi-cuadrado	gl	Sig.
Pearson	3,048	3	,384
Desviación	3,065	3	,382

Las pruebas de bondad de ajuste, indican que no se debe rechazar la hipótesis nula de la adecuación y validez del modelo estimado, como se observa, los estadísticos de Pearson y la Desviación son similares.

Ahora lo que sigue es estudiar el comportamiento de los residuos, con la siguiente prueba de hipótesis sobre los residuos.

$\left\{ \begin{array}{l} H_0: \text{La distribución de contraste es normal} \\ H_1: \text{La distribución de contraste es diferente a la normal} \end{array} \right.$

Figura 3.7.- prueba no paramétrica sobre los residuos.

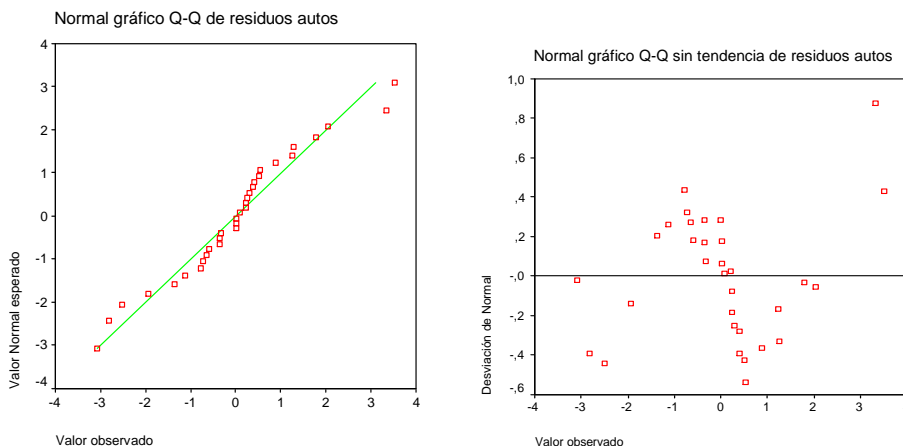
			RESIDUOS
N			16
Parámetros normales	a,b	Media	,0188
		Desviación típica	,55330
Diferencias más extremas		Absoluta	,183
		Positiva	,167
		Negativa	-,183
Z de Kolmogorov-Smirnov			,731
Sig. asintót. (bilateral)			,659

a. La distribución de contraste es la Normal.

b. Se han calculado a partir de los datos.

En efecto el cuadro anterior nos indica que no se rechaza la hipótesis nula de que los residuos siguen una distribución normal.

Figura 3.8.- Distribución de los residuos.



En la figura no se evidencia claramente un comportamiento normal ya que algunos residuos salen de la tendencia lineal, en los extremos inferior y superior de la recta. Además no evidencia claramente el comportamiento aleatorio de los residuos.

### 3.3.3.- Interpretación del modelo.-

Una vez estimado y validado el modelo, lo que sigue es interpretar los resultados del mismo, esto es:

Para los delitos a automóviles, se tiene que en las vías públicas, y estacionados, se tiene un mayor riesgo que cuando se encuentran en un estacionamiento, respecto de las motos. El comportamiento anterior es similar en camionetas y vehículos todo terreno, siendo estos los de mayor riesgo ( $\text{Exp}(B)=2,664$ ) pero la modalidad es de intercepción de los mismo, es decir cuando están en movimiento.

### 3.4.-Modelo para delitos contra la propiedad.

Para la construcción del modelo de delitos contra la propiedad, utilizamos las siguientes variables:

Tabla 3.3.- Definición de variables.-

NOMBRE DE VARIABLE	TIPO DE VARIABLE	MODALIDADES	CODIFICACIÓN
AFFECTADO (variable respuesta)	Cualitativa	Personas Domicilios Empresas	A1 A2 A3
TIPO DELITO (variable explicativa)	Cualitativa	Robo. Asalto y robo. Hurto.	TD1 TD2 TD3
DÍA DEL HECHO (variable explicativa)	Cualitativa	Lunes Martes, Miércoles Jueves Viernes Sábado, Domingo.	D1 D2 D3 D4 D5 D6 D7

Para este modelo, la variable respuesta o dependiente es el afectado con tres categorías, el resto de variables de la tabla anterior son las variables explicativas o independientes. Enseguida se procede a la definición del modelo logístico multinomial.

### 3.4.1.- Estimación del modelo.-

Para la estimación de este modelo, se parte también de la estimación de los valores perdidos, las variables en las que se estimaron valores perdidos fueron: afectado, zona administrativa y clasificación horaria.

Una vez estimados los valores perdidos, se procede con la estimación del modelo mediante el paquete estadístico, el cual nos entrega el siguiente resultado, sobre la prueba de hipótesis siguiente:

$$\begin{cases} H_0 : \hat{\beta}_j = 0 & j = 1, \dots, J - 1 \\ H_1 : \hat{\beta}_j \neq 0 \end{cases}$$

Figura 3.7.- Estimación de los parámetros del modelo

Estimaciones de los parámetros								
Afectados	Categorías	B	Error típ.	gl	Sig.	Exp(B)	Intervalo de confianza al 95% para Exp(B)	
							Límite inferior	Límite superior
Personas	Intersección	-0,5042	0,0982	1	2,824E-07			
	Robo	0,5620	0,0706	1	1,6472E-15	1,7541	1,5276	2,0143
	Asalto y robo	2,0169	0,0737	1	1,048E-164	7,5149	6,5036	8,6834
	Lunes	-0,3875	0,0974	1	6,9214E-05	0,6787	0,5608	0,8215
	Martes	-0,1845	0,0969	1	0,0569074	0,8315	0,6877	1,0054
	Miércoles	-0,2180	0,0959	1	0,02306952	0,8041	0,6663	0,9705
	Jueves	-0,2119	0,0961	1	0,02747291	0,8090	0,6701	0,9767
	Viernes	-0,1208	0,0956	1	0,20667551	0,8862	0,7348	1,0690
	Sábado	-0,0503	0,1013	1	0,6196738	0,9509	0,7796	1,1599
Domicilios	Intersección	-0,1450	0,0963	1	0,13230682			
	Robo	1,1991	0,0720	1	3,0808E-62	3,3172	2,8805	3,8201
	Asalto y robo	-0,4781	0,0984	1	1,1806E-06	0,6199	0,5112	0,7518
	Lunes	-0,9774	0,0974	1	1,0875E-23	0,3763	0,3109	0,4554
	Martes	-0,7208	0,0964	1	7,7647E-14	0,4864	0,4026	0,5876
	Miércoles	-0,8530	0,0958	1	5,2935E-19	0,4261	0,3532	0,5141
	Jueves	-0,7468	0,0951	1	4,1148E-15	0,4739	0,3933	0,5710
	Viernes	-0,6953	0,0948	1	2,1694E-13	0,4989	0,4144	0,6007
	Sábado	-0,4526	0,0992	1	5,0574E-06	0,6360	0,5236	0,7725

Como antes, la figura anterior muestra las estimaciones de los parámetros del modelo, el error típico, los grados de libertad, el nivel de significación, y la razón de disparidades. No interpretaremos los estadísticos que se tiene en la figura, sino más adelante en la validación e interpretación del modelo. El modelo estimado es el siguiente:

$$x_1 = \begin{cases} 1 & \text{Robo} \\ 0 & \text{si no} \end{cases}, x_2 = \begin{cases} 1 & \text{Asalto/robo} \\ 0 & \text{si no} \end{cases}, x_3 = \begin{cases} 1 & \text{lunes} \\ 0 & \text{si no} \end{cases}, x_4 = \begin{cases} 1 & \text{martes} \\ 0 & \text{si no} \end{cases},$$

$$x_5 = \begin{cases} 1 & \text{miércoles} \\ 0 & \text{si no} \end{cases}, x_6 = \begin{cases} 1 & \text{jueves} \\ 0 & \text{si no} \end{cases}, x_7 = \begin{cases} 1 & \text{viernes} \\ 0 & \text{si no} \end{cases}, x_8 = \begin{cases} 1 & \text{sábado} \\ 0 & \text{si no} \end{cases},$$

$$\ln\left(\frac{\pi_1}{\pi_2}\right) = -0.504 + 0.56x_1 + 2.017x_2 - 0.388x_3 - 0.184x_4 - 0.218x_5 - 0.212x_6 - 0.121x_7 - 0.05x_8$$

$$\ln\left(\frac{\pi_2}{\pi_3}\right) = -0.145 + 1.199x_1 - 0.478x_2 - 0.977x_3 - 0.721x_4 - 0.853x_5 - 0.747x_6 - 0.695x_7 - 0.453x_8$$

### 3.4.2.- Validación del modelo.-

Ahora lo que sigue es validar estadísticamente el modelo estimado, para esto se presenta estadísticos que nos permitirá aceptar o rechazar el modelo estimado, para esto se prueba la hipótesis:

$$\left\{ \begin{array}{l} H_0: \text{El modelo es adecuado} \\ H_1: \text{El modelo no es adecuado} \end{array} \right.$$

Figura 3.9.- Pruebas estadísticas sobre las variables.-

Bondad de ajuste			
	Chi-cuadrado	gl	Sig.
Pearson	28,420	24	,243
Desviación	28,328	24	,246

La figura, muestra las pruebas de bondad de ajuste, la que nos indica que no se debe rechazar la validez del modelo estimado, esto es, que el modelo que se estimo es el adecuado.

Ahora estudiemos el comportamiento de los residuos.

$$\left\{ \begin{array}{l} H_0: \text{La distribución de contraste es normal} \\ H_1: \text{La distribución de contrate es diferente a la normal} \end{array} \right.$$

Figura 3.10.- Prueba no paramétrica sobre los residuos.

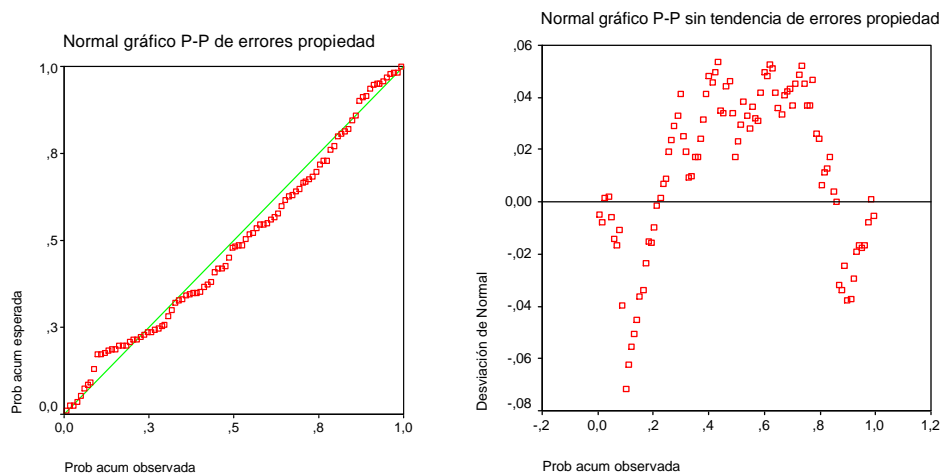
Prueba de Kolmogorov-Smirnov para una muestra			Residuos de Pearson
N			63
Parámetros normales <sup>a,b</sup>	Media		,0007
	Desviación típica		,82231
Diferencias más extremas	Absoluta		,078
	Positiva		,078
	Negativa		-,072
Z de Kolmogorov-Smirnov			,620
Sig. asintót. (bilateral)			,836

a. La distribución de contraste es la Normal.

b. Se han calculado a partir de los datos.

El cuadro muestra que no se rechaza la hipótesis nula de que los residuos siguen una distribución normal.

Figura 3.11.-Distribución de los residuos



En la figura anterior se observa que existe un comportamiento casi normal, saliéndose algunos residuos de la tendencia lineal, al igual que el comportamiento aleatorio de los mismos.

### 3.4.3.- Interpretación del modelo.-

Ahora nos queda interpretar el modelo estimado y validado, en base a las categorías de las variables.

Para los delitos contra las personas respecto de las empresas, se tiene que el asalto y robo es el mayor riesgo al que está expuesta, según los días se observa que los de mayor riesgo son de martes a sábado. Ahora los domicilios se observa que son los robos el delito que tiene mayor probabilidad de presentarse, según los días son de martes a sábado los de mayor riesgo.

### 3.5.- Modelo para violencia intrafamiliar.

Para la construcción del modelo logístico multinomial asociado a la violencia intrafamiliar consideramos las siguientes variables:

Tabla 3.4.- Definición de variables.-

NOMBRE DE VARIABLE	TIPO DE VARIABLE	MODALIDAD	CODIFICACIÓN
TIPO DE AGRESIÓN (variable respuesta)	Cualitativa	Física. Psicológica. Sexual.	TA1 TA2. TA3
RELACIÓN VÍCTIMA- AGRESOR (variable explicativa)	Cualitativa	Pareja. Ex pareja Familiar Otra.	R1 R2 R3 R4
LUGAR DEL HECHO (variable explicativa)	Cualitativa	Domicilio Fuera del domicilio	L1 L2
ESTADO CIVIL (variable explicativa)	Cualitativa	Casado Divorciado Soltero Viudo	EC1 EC2 EC3 EC4

La variable tipo de agresión es nuestra variable respuesta o dependiente, mientras que el resto de variables son consideradas como las variables explicativas o independientes.

### 3.5.1.- Estimación del modelo.-

Iniciamos este proceso estimando los datos perdidos de acuerdo a la matriz de diseño para estos y con los procedimientos para datos perdidos cualitativos.

Enseguida se procede a mostrar el modelo y sus parámetros relacionando la variable respuesta con las variables explicativas; para esto probamos la hipótesis siguiente:

$$\begin{cases} H_0 : \hat{\beta}_j = 0 & j = 1, \dots, J - 1 \\ H_1 : \hat{\beta}_j \neq 0 \end{cases}$$

Figura 3.12.- Definición de parámetros.-

Estimaciones de los parámetros								
Agresión	Categorías	B	Error típ.	gl	Sig.	Exp(B)	Intervalo de confianza al 95% para Exp(B)	
							Límite inferior	Límite superior
Física	Intersección	0,9178	0,4211	1	0,02929965			
	Pareja	-0,0397	0,2663	1	0,88152642	0,9611	0,5703	1,6196
	Ex-pareja	0,1120	0,2838	1	0,69308252	1,1185	0,6413	1,9507
	Familiar	1,9768	0,3454	1	1,0446E-08	7,2195	3,6687	14,2072
	Domicilio	0,4007	0,0892	1	7,0978E-06	1,4928	1,2533	1,7781
	Casado	-0,1457	0,3371	1	0,66567729	0,8644	0,4464	1,6738
	Divorciado	-0,0152	0,3696	1	0,96719352	0,9849	0,4773	2,0326
	Soltero	-0,2238	0,3385	1	0,50849078	0,7994	0,4118	1,5522
Psicológica	Intersección	0,8608	0,4366	1	0,04863628			
	Pareja	-0,8809	0,2702	1	0,00111371	0,4144	0,2440	0,7038
	Ex-pareja	0,0776	0,2876	1	0,78741037	1,0806	0,6150	1,8987
	Familiar	1,6781	0,3489	1	1,5167E-06	5,3552	2,7024	10,6120
	Domicilio	0,4516	0,1007	1	7,3448E-06	1,5709	1,2894	1,9138
	Casado	0,0133	0,3544	1	0,97014207	1,0134	0,5059	2,0297
	Divorciado	0,0507	0,3887	1	0,89618345	1,0520	0,4911	2,2539
	Soltero	-0,4816	0,3564	1	0,17663571	0,6178	0,3072	1,2423

En la figura anterior se tiene las estimaciones de todos los parámetros que intervienen en el modelo, además de esto se presentan los errores típicos de los parámetros, el nivel de significación las disparidades y los intervalos de las disparidades entre otros estadísticos, el modelo estimado es el siguiente:

$$x_1 = \begin{cases} 1 & \text{Pareja} \\ 0 & \text{si no} \end{cases}, x_2 = \begin{cases} 1 & \text{ex - pareja} \\ 0 & \text{si no} \end{cases}, x_3 = \begin{cases} 1 & \text{familiar} \\ 0 & \text{si no} \end{cases}, x_4 = \begin{cases} 1 & \text{hogar} \\ 0 & \text{si no} \end{cases},$$

$$x_5 = \begin{cases} 1 & \text{casado} \\ 0 & \text{si no} \end{cases}, x_6 = \begin{cases} 1 & \text{divorciado} \\ 0 & \text{si no} \end{cases}, x_7 = \begin{cases} 1 & \text{soltero} \\ 0 & \text{si no} \end{cases},$$

$$\ln\left(\frac{\pi_1}{\pi_3}\right) = 0.918 - 0.04x_1 + 0.112x_2 + 1.977x_3 + 0.401x_4 - 0.146x_5 - 0.015x_6$$

$$- 0.224x_7$$

$$\ln\left(\frac{\pi_2}{\pi_3}\right) = 0.861 - 0.881x_1 + 0.078x_2 + 1.678x_3 + 0.452x_4 + 0.013x_5 + 0.051x_6$$

$$- 0.482x_7$$



### 3.5.2.- Validación del modelo.-

Enseguida se procede a validar el modelo estadísticamente, para esto se muestran los estadísticos y pruebas siguientes.

$$\left\{ \begin{array}{l} H(0): \text{El modelo es adecuado} \\ H(1): \text{El modelo no es adecuado} \end{array} \right.$$

Figura 3.13.- Pruebas estadísticas sobre el modelo.-

Bondad de ajuste			
	Chi-cuadrado	gl	Sig.
Pearson	56,254	48	,193
Desviación	58,910	48	,134

La figura anterior muestra las pruebas de bondad de ajuste, sobre la validez del modelo estimado, estos estadísticos nos indican que no se debe rechazar la hipótesis nula, esto es, el modelo estimado es válido.

Ahora lo que sigue es estudiar el comportamiento de los residuos, probando la siguiente hipótesis:

$$\left\{ \begin{array}{l} H_0: \text{La distribución de contraste es normal} \\ H_1: \text{La distribución de contraste es diferente a la normal} \end{array} \right.$$

Figura 3.14.- Pruebas no paramétricas de los residuos.

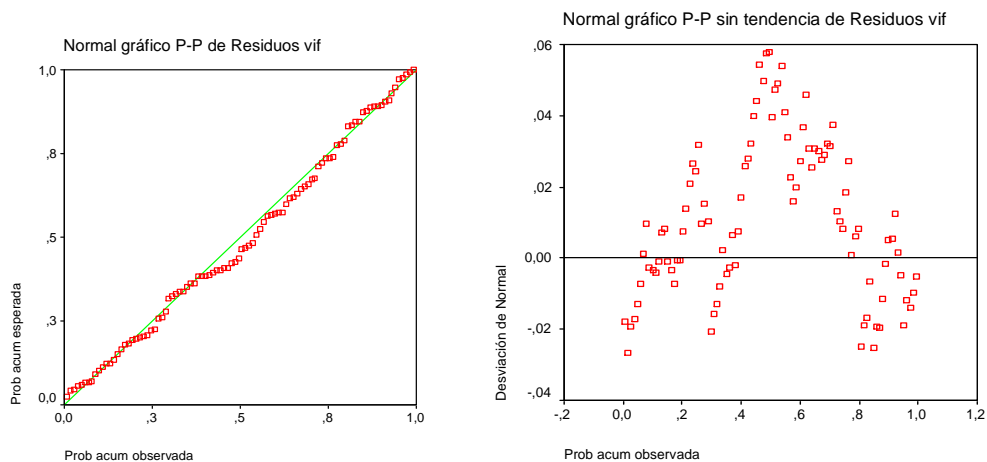
Prueba de Kolmogorov-Smirnov para una muestra			Residuos víf
N			96
Parámetros normales <sup>a,b</sup>	Media		-,0064
	Desviación típica		,94520
Diferencias más extremas	Absoluta		,063
	Positiva		,063
	Negativa		-,033
Z de Kolmogorov-Smirnov			,617
Sig. asintót. (bilateral)			,841

a. La distribución de contraste es la Normal.

b. Se han calculado a partir de los datos.

Como se observa en el cuadro anterior, la prueba de normalidad sobre los residuos, nos indica que no se rechaza la hipótesis; esto es, se acepta que los residuos siguen una distribución normal.

Figura 3.15.-Distribución de los residuos



La figura muestra que los residuos se ubican sobre la recta normal, lo que nos ratifica la prueba no paramétrica de normalidad. Se puede ver también que los residuos presentan un comportamiento aleatorio.

### 3.5.3.- Interpretación del modelo.-

Una vez que se ha estimado y validado el modelo lo que sigue es interpretarlo, los resultados que se tiene son los siguientes.

Para la agresión física en relación al maltrato sexual, se tiene que son los familiares de las víctimas los que producen este tipo de violencia, seguido de las exparejas; el lugar de ocurrencia del hecho es al interior del hogar, mientras que según el estado civil de la víctima, no se observa una diferencia entre alguna de las categorías.

Para el maltrato psicológico se observa que los mayores maltratantes son también los familiares; este tipo de maltrato se produce también dentro del hogar y el estado civil de mayor riesgo para que se produzca este tipo de maltrato es casado y divorciado.

### 3.6.- Modelo para maltrato infantil.

Por último, construimos el modelo logístico multinomial para el maltrato infantil, para esto utilizaremos las variables siguientes:

Tabla 3.5.- Definición de variables.-

NOMBRE DE VARIABLE	TIPO DE VARIABLE	MODALIDADES	CODIFICACIÓN
TIPO DE AGRESIÓN (variable dependiente)	Cualitativa	Física. Psicológica. Sexual.	TA1 TA2 TA3
EDAD (variable explicativa)	Cuantitativa discreta	0-4 5-10 11-14 15-18	E1 E2 E3 E4
ARMA MEDIO (variable explicativa)	Cualitativa	Golpes Contundente Abuso sexual Otro	AM1 AM2 AM3 AM4

La tabla muestra las variables cualitativas, tanto la variable respuesta o dependiente que es tipo de agresión, como las variables explicativas o independientes que son el resto de las variables, con esto se procede a determinar el modelo.

### 3.6.1.- Estimación del modelo.-

A continuación presentamos la definición del modelo, para esto, se presenta las estimaciones de los parámetros, probando la siguiente hipótesis:

$$\begin{cases} H_0 : \hat{\beta}_j = 0 & j = 1, \dots, J - 1 \\ H_1 : \hat{\beta}_j \neq 0 \end{cases}$$

Figura 3.16.- Definición de los parámetros del modelo.-

Estimaciones de los parámetros								
Tipo de Maltrato	Categorías	B	Error típ.	gl	Sig.	Exp(B)	Intervalo de confianza al 95% para Exp(B)	
							Límite inferior	Límite superior
Físico	Intersección	2,7197	0,8507	1	0,00138827			
	0-4	2,4671	2,8012	1	0,3784598	11,7882	0,0487	2856,3555
	5-10	0,8841	1,0742	1	0,41050674	2,4207	0,2948	19,8740
	11-14	-0,3607	0,7414	1	0,62655657	0,6972	0,1630	2,9812
	Golpes	2,5231	0,9382	1	0,00716157	12,4671	1,9822	78,4110
	Contundente	2,0963	1,0257	1	0,04098232	8,1360	1,0897	60,7462
	Abuso sexual	-8,5753	1,2603	1	1,0159E-11	0,0002	0,0000	0,0022
Psicológico	Intersección	4,6486	0,8374	1	2,8376E-08			
	0-4	2,6217	2,7964	1	0,34849356	13,7593	0,0573	3303,2017
	5-10	0,7351	1,0760	1	0,4944655	2,0858	0,2532	17,1855
	11-14	-0,3889	0,7477	1	0,60298855	0,6778	0,1566	2,9346
	Golpes	-0,6166	0,9273	1	0,50605659	0,5398	0,0877	3,3228
	Contundente	-3,8557	1,1278	1	0,00062935	0,0212	0,0023	0,1930
	Abuso sexual	-10,4589	1,2482	1	5,3199E-17	0,0000	0,0000	0,0003

En la figura, se muestra la estimación de los parámetros asociados a las variables explicativas, además de esto se presenta también los errores de estos parámetros, los grados de libertad, el nivel de significación, la razón de disparidades y sus intervalos de confianza. Aquí lo que nos interesa es la determinación de los parámetros del modelo. A continuación, procedemos a validar el modelo con pruebas estadísticas sobre el mismo; el modelo estimado es el siguiente:

$$x_1 = \begin{cases} 1 & \text{golpes} \\ 0 & \text{otro} \end{cases}, x_2 = \begin{cases} 1 & \text{contundente} \\ 0 & \text{otro} \end{cases}, x_3 = \begin{cases} 1 & \text{abuso sexual} \\ 0 & \text{otro} \end{cases}, x_4 = \begin{cases} 1 & 0 - 4 \text{ años} \\ 0 & \text{si no} \end{cases},$$

$$x_5 = \begin{cases} 1 & 5 - 10 \text{ años} \\ 0 & \text{si no} \end{cases}, x_6 = \begin{cases} 1 & 11 - 14 \text{ años} \\ 0 & \text{si no} \end{cases},$$

$$\ln\left(\frac{\pi_1}{\pi_3}\right) = 2.72 + 2.523x_1 + 2.096x_2 - 8.575x_3 + 2.467x_4 + 0.884x_5 - 0.361x_6$$

$$\ln\left(\frac{\pi_2}{\pi_3}\right) = 4.649 - 0.617x_1 - 3.856x_2 - 10.459x_3 + 2.622x_4 + 0.735x_5 - 0.389x_6$$

### 3.6.2.- Validación del modelo.-

Para esto, procedemos a presentar pruebas estadísticas sobre el modelo estimado, para lo que se prueba la siguiente hipótesis:

$$\left\{ \begin{array}{l} H(0): \text{El modelo es adecuado} \\ H(1): \text{El modelo no es adecuado} \end{array} \right.$$

Figura 3.17.- Pruebas estadísticas sobre el modelo.-

Bondad de ajuste			
	Chi-cuadrado	gl	Sig.
Pearson	20,017	18	,332
Desviación	24,670	18	,134

En la figura anterior se muestra la prueba de bondad de ajuste, la cual nos indica que no se debe rechazar la hipótesis nula de que el modelo estimado es el adecuado, se muestra también, la información de ajuste del modelo, con dos estadísticos, estos son, el estadístico de Pearson y la Desviación, además de los grados de libertad del modelo y el nivel de significación estadística

Ahora proseguimos con el estudio de los residuos, para lo que se prueba la hipótesis de normalidad de residuos.

$$\left\{ \begin{array}{l} H0: \text{La distribución de contraste es normal} \\ H1: \text{La distribución de contraste es diferente a la normal} \end{array} \right.$$

Figura 3.18.- Prueba no paramétrica sobre los residuos

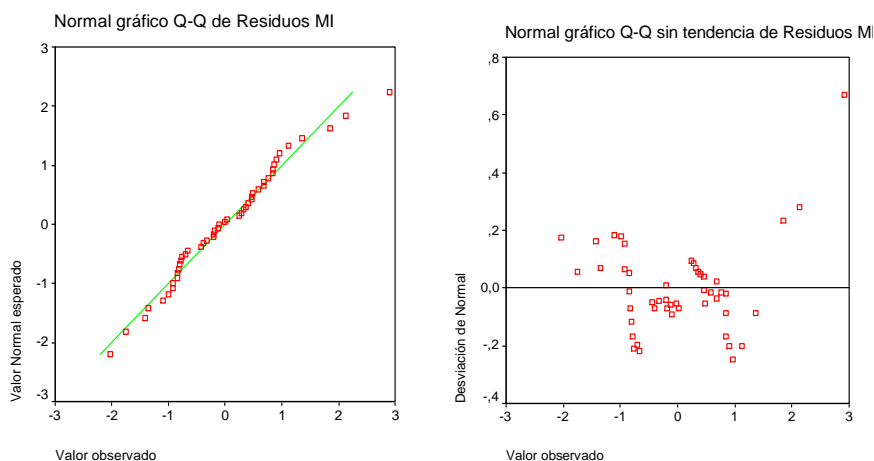
Prueba de Kolmogorov-Smirnov para una muestra			Residuos MI
N			48
Parámetros normales <sup>a,b</sup>	Media		,0170
	Desviación típica		,99851
Diferencias más extremas	Absoluta		,083
	Positiva		,083
	Negativa		-,051
Z de Kolmogorov-Smirnov			,577
Sig. asintót. (bilateral)			,893

a. La distribución de contraste es la Normal.

b. Se han calculado a partir de los datos.

Como se muestra en el cuadro anterior, no se rechaza la hipótesis nula de que los residuos siguen una distribución normal.

Figura 3.19.-Distribución de los residuos



La figura muestra también que se tiene una superposición de los residuos sobre la recta de la normal, además, se tiene una tendencia aleatoria de los residuos, con lo que se tiene la aceptación de que estos cumplen con las hipótesis del modelo.

### 3.6.3.- Interpretación del modelo.-

Por último, lo que sigue es interpretar el modelo estimado y validado para cada una de las categorías de la variable dependiente.

Para el maltrato físico se tiene que el medio utilizado son los golpes tanto con puños como con armas contundentes, la edad de mayor riesgo son de menos de cuatro años.

Mientras que para el maltrato psicológico, se observa que no existe una causa por el tipo de arma que sea considerada, mientras que la edad de mayor riesgo es de menos de cuatro años. Esto muestra que son las primeras edades de los infantes que ellos están en mayor riesgo, resulta un poco incomprensible ya que se supondría que en estas edades están al cuidado de sus padres.

## **Capítulo 4**

### **Construcción de Modelos**

#### **Log lineales.**

Al igual que en capítulo anterior, en éste nos concentramos en modelar los tipos de violencia descritos anteriormente, basados en la construcción y estimación de modelos log lineales, que están sustentados en la selección de variables que fueron seleccionadas en el capítulo 2 mediante métodos exploratorios univariantes y multivariantes, todas estas variables son categóricas y al cruzarlas se forman tablas de contingencia resumiendo la información de forma descriptiva.

Se construye un modelo log lineal para cada tipo de violencia; como se verá más adelante en este tipo de modelación, no interesa definir variables explicativas y exploratorias. Dentro de la estimación del modelo, se consideraran los pasos siguientes: selección de variables, estimación del modelo, validación estadística y gráfica del mismo y se finaliza con la interpretación.

#### **4.1.- Modelos log lineales en tablas de contingencia.-**

Las variables respuesta y las variables exploratorias son todas categóricas. Las observaciones consisten en las cantidades de frecuencia ubicadas en las celdas de una tabla de contingencia. Los modelos log lineales, describen el diseño de asociación e interacción entre un conjunto de variables categóricas.

### 4.1.1.- Modelos log lineales para tablas de contingencia de dos dimensiones.-

#### 4.1.1.1.- Distribuciones de probabilidad.

*X* tiene *I* modalidades.

*Y* tiene *J* modalidades.

Tabla 4.1.- Diseño tabla de contingencia en dos dimensiones.

	$Y_1$	$Y_2$	...	$Y_j$	...	$Y_J$	TOTAL
$X_1$	$n_{11}$	$n_{12}$	...	$n_{1j}$	...	$n_{1J}$	$n_{1.}$
$X_2$	$n_{21}$	$n_{22}$	...	$n_{2j}$	...	$n_{2J}$	$n_{2.}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$X_i$	$n_{i1}$	$n_{i2}$	...	$n_{ij}$	...	$n_{iJ}$	$n_{i.}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$X_I$	$n_{I1}$	$n_{I2}$	...	$n_{Ij}$	...	$n_{IJ}$	$n_{I.}$
TOTAL	$n_{.1}$	$n_{.2}$	...	$n_{.j}$	...	$n_{.J}$	$n = n_{..}$

Iniciamos mostrando las distribuciones de probabilidad asociadas con las tablas de contingencia.

#### a) Distribución de Poisson

Sea  $\{n_{ij}\}$  variable aleatoria independiente de Poisson con parámetros  $\lambda_{ij} > 0$ ; la distribución de probabilidad conjunta es:

$$f(n; \lambda) = \prod_{i=1}^n \prod_{j=1}^J \frac{\lambda_{ij}^{n_{ij}} e^{-\lambda_{ij}}}{n_{ij}!} \quad (1)$$

#### Observación

Si  $n = n_{..}$  (por el diseño del estudio), por la propiedad aditiva de la distribución de Poisson,  $\mathbf{n}$  también tiene una distribución de Poisson con parámetro

$$\lambda_{..} = \sum \sum \lambda_{ij}.$$

#### b) Distribución Multinomial

Si la restricción es el total de las observaciones,  $\{n_{ij}\}$  es condicional en  $\mathbf{n}$



$$f(y | n) = \frac{\prod_{i=1}^I \prod_{j=1}^J \frac{\lambda_{ij}^{n_{ij}} e^{-\lambda_{ij}}}{n_{ij}!}}{\frac{\lambda_{..}^n e^{-\lambda_{..}}}{n!}} \quad (2)$$

$$f(y | n) = n! \prod_{i=1}^I \prod_{j=1}^J \frac{\theta_{ij}^{n_{ij}}}{n_{ij}!} \quad (3), \text{ donde } \theta_{ij} = \frac{\lambda_{ij}}{\lambda_{..}}; \quad 0 \leq \theta_{ij} \leq 1 \wedge \sum_i \sum_j \theta_{ij} = 1$$

Por la restricción sabemos que  $\lambda_{..} = \prod \prod \lambda_{ij}^{n_{ij}}$

Además  $e^{-\lambda_{..}} = \prod \prod e^{-\lambda_{ij}}$

Con esto se tiene la distribución multinomial.

### c) Distribución producto de multinomiales

Otra restricción se tiene cuando se fijan los totales fila o totales columna, en lugar de fijar el total general.

En este caso la distribución de probabilidad es multinomial para cada fila (o columna); así para la j-ésima fila con total fila fija, se tienen la distribución

$$f(n_{i1}, \dots, n_{ij} | n_{i.}) = n_{i.}! \prod_{j=1}^J \frac{\theta_{ij}^{n_{ij}}}{n_{ij}!} \quad (4)$$

$$\sum_j \theta_{ij} = 1$$

Asumimos que las filas (o columnas) son independientes. Si los totales en las filas son fijos, la distribución conjunta de todos los  $n_{ij}$  es:

$$f(n | n_{i.}, i = 1, \dots, I) = \prod_{i=1}^I n_{i.}! \prod_{j=1}^J \frac{\theta_{ij}^{n_{ij}}}{n_{ij}!} \quad (5), \text{ que es la distribución del producto}$$

multinomiales.

Se quiere probar la hipótesis nula de que las distribuciones son iguales.

#### 4.1.1.2.- Modelo log lineal.-

Si  $n_1, \dots, n_N$  siguen una distribución de Poisson  $P(\lambda_1, \dots, \lambda_N)$ , las frecuencias esperadas están dadas por:

$$E(n_i) = \lambda_i.$$

Si  $n_1, \dots, n_N$  sigue una distribución multinomial con probabilidades  $\theta_1, \dots, \theta_N$ , tal

que  $\sum_{i=1}^N \theta_i = 1$  y frecuencia total  $n = \sum_{i=1}^N n_i$ , se tiene que:

$$E(n_i) = n\theta_i$$

Para el producto de multinomiales se tiene:

$$E(n_{ijk}) = n_{i.k} \theta_{ijk} \text{ (Para totales fila fijo)}$$

$$E(n_{ijk}) = n_{.jk} \theta_{ijk} \text{ (Para totales columna fijo)}$$

#### 4.1.1.2.1.- Modelo log lineal para una tabla bidimensional.-

Se tiene dos variables cualitativas

$X$  con  $I$  modalidades.

$Y$  con  $J$  modalidades.

Se tiene  $N=IJ$  celdas con  $\{Y_{ij}\}$  variables aleatorias independientes con distribución de Poisson, tal que  $\mu_{ij} = E(Y_{ij})$

Las frecuencias observadas en cada celda se nota por  $\{n_{ij}\}$ . En dos dimensiones, solo pueden ocurrir dos modelos: las dos variables son independientes o las dos variables están asociadas.

##### Independencia del modelo.

Se tiene la estructura de  $\{\mu_{ij}\}$

$$\mu_{ij} = \mu \alpha_i \beta_j \text{ (Agresti, 2002, pg 132)}$$

Todas las hipótesis pueden ser formuladas como modelos multiplicativos para las frecuencias esperadas.

Si la hipótesis es que la variable fila y columna son independientes, se tiene:

$\theta_{ij} = \theta_i \theta_j$ , talque  $\theta_i$  es la probabilidad marginal fila y  $\theta_j$  es la probabilidad marginal columna, para las que se cumple.

$$\sum \theta_i = 1 \text{ y } \sum \theta_j = 1$$

Para la distribución multinomial tenemos:

$$\mu_{ij} = E(n_{ij}) = n\theta_i\theta_j$$

$$\log \mu_{ij} = \lambda + \lambda_i^X + \lambda_j^Y \quad (7)$$

Este es el modelo loglineal de independencia, tal que:

$$\lambda_i^X \text{ es el efecto fila} \quad \wedge \quad \lambda_j^Y \text{ es el efecto columna}$$

Donde la función logaritmo es la función natural de enlace del modelo lineal generalizado  $\eta_i = \ln(E(n_i)) = x_i'\beta \quad i = 1, \dots, n$

En este modelo las restricciones son:

$$\lambda_i^X = \lambda_j^Y = 0$$

#### 4.1.1.2.2.- Estimación del modelo log lineal.-

En función a la descripción anterior tenemos:

$$\eta_i = \ln E(y_{ij}) = \ln n\theta_i\theta_j.$$

$$\eta_i = \mu + \alpha_i + \beta_j$$

$$\ln(\mu_{ij}) = \lambda + \lambda_i^X + \lambda_j^Y \quad (8)$$

Donde  $\ln(\mu_{ij})$  depende aditivamente de un término basado en el tamaño de la muestra, un término basado en la probabilidad de la *fila*  $i$ ,  $\lambda_i^X$ , y un término basado en la probabilidad de la *columna*  $j$ ,  $\lambda_j^Y = 0$ .

El valor estimado que satisface el modelo es:

$$\hat{\mu} = \frac{n_{i.} * n_{.j}}{n}, \text{ donde la hipótesis que probamos son:}$$

$$\left\{ \begin{array}{l} H0: \text{Independencia entre variables.} \\ H1: \text{No independencia.} \end{array} \right.$$

#### 4.1.1.2.3.- Interpretación de los parámetros.-

Para tablas bidimensionales  $I \times J$ , los modelos lineales generalizados tratan las  $N=IJ$  celdas como  $N$  observaciones independientes de Poisson. Consideremos el modelo de independencia

$$\ln(\mu_{ij}) = \lambda + \lambda_i^X + \lambda_j^Y \quad (9)$$

Este modelo, no distingue entre variables respuesta y variables explicativas, trata a las dos variables como respuestas, modelando  $\{\mu_{ij}\}$  tpor medio de las combinaciones de las categorías de cada variable.

$$\eta_{ij} = \ln(\mu_{ij}) = \lambda + \lambda_i^X + \lambda_j^Y$$

$$\frac{P(\text{exito})}{P(\text{fracso})} = \frac{\pi}{1-\pi} \Rightarrow \ln\left(\frac{\pi}{1-\pi}\right) = \ln\left(\frac{\mu_{i1}}{\mu_{i2}}\right) \quad (10)$$

$$= \lambda_1^Y - \lambda_2^Y$$

El logit para  $Y$  no depende de  $X$ , tal que  $\pi$  es la probabilidad de éxito

El modelo log lineal de independencia corresponde al modelo  $\text{logit}(\pi) = \alpha$  que es idéntico en cada categoría de  $X$

Así tenemos que, la probabilidad de respuesta en la columna 1 es:

$$\exp(\alpha) = \exp(\lambda_1^Y - \lambda_2^Y) \quad (11)$$

#### 4.1.1.2.5.- Modelo saturado.-

Estadísticamente, variables dependientes, satisfacen un modelo loglineal más completo.

$$\ln(\mu_{ij}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY} \quad (12)$$

Los parámetros  $\{\lambda_{ij}^{XY}\}$  son términos de asociación entre las modalidades de  $X$  y de  $Y$ . Sobre esta asociación se prueba la hipótesis siguiente:

$$H_0 = \lambda_{ij}^{XY} = 0 \quad \forall i, j \text{ (Modelo de independencia)}$$

$$\ln(\theta) = \ln\left(\frac{\mu_{11}\mu_{22}}{\mu_{12}\mu_{21}}\right) = \lambda_{11}^{XY} + \lambda_{22}^{XY} - \lambda_{12}^{XY} - \lambda_{21}^{XY} \quad (13)$$

Así, los  $\{\lambda_{ij}^{XY}\}$  determinan el logaritmo de las disparidades.

Si  $\ln(\theta) = 1 \Rightarrow X$  y  $Y$  son independientes.

#### 4.1.1.3.- Estimación de máxima verosimilitud.-

##### i) Para la distribución de Poisson

Para la distribución de Poisson la función de máxima verosimilitud es:

$$l = \sum_{i=1}^N (n_i \ln \lambda_i - \lambda_i - \ln(n_i!)) \quad (14)$$

##### ii) Para la distribución multinomial

$$l = \ln(n!) + \sum_{i=1}^N (n_i \ln \theta_i - \ln(n_i!)) \quad (15)$$

$$l = cte + \sum_{i=1}^N n_i \ln(E(n_i)) \quad (16) \quad E(n_i) = n\theta_i \quad \sum \theta_i = 1$$

iii) Para la distribución del producto de multinomiales

$$l = cte + \sum_{i=1}^N n_i \ln(E(n_i)) \quad (17)$$

#### 4.1.1.4.- Prueba de hipótesis y la mejor estimación.-

Para el modelo maximal existen N parámetros, que pueden ser tomados sin pérdida de generalidad  $E(Y_1), \dots, E(Y_N)$ . Pueden ser estimadas por las correspondientes frecuencias observadas  $y_1, \dots, y_N$ . Para alguna de las tres distribuciones de probabilidad la función de máxima verosimilitud tiene la forma

$$l(\hat{\beta}_{\max}; y) = cte + \sum_{i=1}^N y_i \ln(y_i) \quad (18)$$

Para otro modelo  $e_i$  denota la estimación esperada de las frecuencias con la función de máxima verosimilitud.

$$l(\hat{\beta}; y) = cte + \sum_{i=1}^N y_i \ln(e_i) \quad (19) \text{ (Las cotas son las mismas)}$$

Con esto se tiene el estadístico de razón de verosimilitud que se lo conoce como la desviación:

$$D = 2 \left[ \sum_i y_i \ln \left( \frac{y_i}{e_i} \right) \right] \quad (20)$$

El estadístico Ji-cuadrado queda como sigue:

$$X^2 = \sum_{i=1}^N \frac{(o_i - e_i)^2}{e_i} \rightarrow \chi_{N-p}^2 \quad (21) \text{ } o=\text{observado}; e=\text{estimado}$$

Los residuos estandarizados para cada celda pueden definirse como:

$$r_i = \frac{o_i - e_i}{\sqrt{e_i}} \quad (22) \quad (\text{Dobson, 1990, pg 133})$$

#### 4.1.2.- Modelos log lineales para tablas en tres vías.-

Asumimos como antes que tenemos una muestra de tamaño n, distribuida sobre  $IJK=N$  celdas. Sobre la muestra multinomial (márgenes no fijos), la

probabilidad de que la observación caiga en la celda  $ijk$  es  $\pi_{ijk}$ ,  $\forall i = 1, \dots, I$ ;  $j = 1, \dots, J$ ;  $k = 1, \dots, K$ , siendo el valor esperado de la celda  $\mu_{ijk} = n\pi_{ijk}$ .

#### 4.1.2.1.- Modelo log lineal y estructura de independencia.-

Sea  $\{\mu_{ijk}\}$  la frecuencia esperada en la tabla de contingencia. Un único factor en el modelo  $\{\mu_{ijk}\}$  representa la distribución marginal. Para tablas de tres vías se tienen algunos tipos de independencia; estos son:

##### 4.1.2.1.1.- Independencia mutua.-

$$\pi_{ijk} = \pi_{i..} \pi_{.j.} \pi_{..k}$$

$$\mu_{ijk} = n\pi_{i..} \pi_{.j.} \pi_{..k}$$

$$\ln(\mu_{ijk}) = \ln(n) + \ln(\pi_{i..}) + \ln(\pi_{.j.}) + \ln(\pi_{..k}) \quad (23)$$

$$\ln(\mu_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z$$

##### 4.1.2.1.2.- Independencia conjunta.-

La variable  $X$  es parcialmente independiente de  $Y$  y  $Z$ , si:

$$\pi_{ijk} = \pi_{i..} \pi_{.jk} \quad \forall i, j, k$$

La variable compuesta  $YZ$ , la cual tiene  $JK$  niveles de  $Y$  y  $Z$  es mutuamente independiente de  $X$ .

$$\ln(\mu_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{jk}^{YZ} \quad (24).$$

La interacción entre la segunda y última variable ( $Y$  y  $Z$ ) es reflejada por la diferencia particionada en los dos sub-bloques de la primera variable.

##### 4.1.2.1.3.- Independencia condicional.-

$X$ - $Y$  son condicionalmente independiente dado  $Z$ , cuando se alcanza la independencia en cada tabla parcial, en la que  $Z$  es fijo.

Si  $\pi_{ij|k} = P(X = i, Y = j | Z = k) \Rightarrow$

$\pi_{ij|k} = \pi_{i,j|k} \pi_{.jk} \quad \forall i, j, k$ , con las estimaciones siguientes:

$$\pi_{ij|k} = \frac{\pi_{ijk}}{\pi_{.j.}}, \quad \pi_{i,j|k} = \frac{\pi_{i.k}}{\pi_{i..}}, \quad \pi_{.jk} = \frac{\pi_{.jk}}{\pi_{.j.}}$$

$$\ln(\mu_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{jk}^{XZ} + \lambda_{jk}^{YZ} \quad (25)$$

Estructura de independencia de  $X$  y  $Y$  dados los dos niveles de  $Z$ .

#### 4.1.2.1.4.- No existe interacción en tres vías.-

Este modelo es cercano al modelo saturado. Aunque los niveles de las variables interactúan unos con otros, no existe interacción entre las tres variables.

$$\ln(\mu_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ} \quad (26)$$

#### 4.1.2.1.5.- Interpretación de los parámetros del modelo.-

El modelo log lineal es el siguiente:

$$\ln(\mu_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ} + \lambda_{ijk}^{XYZ} \quad (27)$$

Este se lo conoce como modelo saturado, el cual permite observar la razón de disparidades entre un par de variables cruzada con niveles de la tercera variable.

En el modelo de asociación homogénea se elimina el tercer factor  $\{\lambda_{ijk}^{XYZ}\}$ .

La razón de disparidades condicional entre  $X$ - $Y$   $\theta_{XY(K)}$  describe la asociación condicional entre  $X$  y  $Y$  en la  $k$ -ésima tabla parcial.

$$\ln(\theta_{XY(K)}) = \ln\left(\frac{\mu_{11k} * \mu_{22k}}{\mu_{12k} * \mu_{21k}}\right) = \lambda_{11}^{XY} + \lambda_{22}^{XY} - \lambda_{12}^{XY} - \lambda_{21}^{XY} \quad (28)$$

La razón de disparidad es la misma para todos los niveles de  $Z$ . Son similares las disparidades para las otras interacciones.

#### Estimación de los parámetros del modelo log-lineal.-

##### Estadístico mínimo suficiente.-

Se tiene una tabla de contingencia de tres vías, tal que las cantidades de las celdas tienen una distribución de Poisson  $\{Y_{ijk} = n_{ijk}\}$  con distribución de probabilidad conjunta dado por:

$$f(n; \mu) = \prod_{i=1}^n \prod_{j=1}^J \prod_{k=1}^K \frac{\mu_{ijk}^{n_{ijk}} e^{-\mu_{ijk}}}{n_{ijk}!} \quad (29)$$

La función logaritmo de la verosimilitud es:

$$l(\mu) = \sum_i \sum_j \sum_k n_{ijk} \log(\mu_{ijk}) - \sum_i \sum_j \sum_k \mu_{ijk}$$

Remplazando la ecuación del modelo log-lineal saturado tenemos:

$$l(\mu) = n\lambda + \sum_i n_{i..} \lambda_i^X + \sum_j n_{.j.} \lambda_j^Y + \sum_k n_{..k} \lambda_k^Z + \sum_i \sum_j n_{ij.} \lambda_{ij}^{XY} + \sum_i \sum_k n_{i.k} \lambda_{ik}^{XZ} + \sum_j \sum_k n_{.jk} \lambda_{jk}^{YZ} + \sum_i \sum_j \sum_k n_{ijk} \lambda_{ijk}^{XYZ} - \sum_i \sum_j \sum_k \exp(\lambda + \dots + \lambda_{ijk}^{XYZ}) \quad (30)$$

La distribución de Poisson pertenece a la familia exponencial, entonces los estimadores de los parámetros son estadísticos suficientes.

Dependiendo del tipo de independencia, y del modelo, sea este saturado o no, se estiman los parámetros de la ecuación anterior.

Así para el modelo de independencia mutua, se tienen las estimaciones siguientes:

$$\hat{\lambda}_i^X = \ln(\pi_{i..}) - \frac{1}{I} \sum_i \ln(\pi_{i..}) \quad (31)$$

$$\hat{\lambda}_j^Y = \ln(\pi_{.j.}) - \frac{1}{J} \sum_j \ln(\pi_{.j.}) \quad (32)$$

$$\hat{\lambda}_k^Z = \ln(\pi_{..k}) - \frac{1}{K} \sum_k \ln(\pi_{..k}) \quad (33)$$

$$\hat{\lambda} = \ln(n) + \frac{1}{I} \sum_i \ln(\pi_{i..}) + \frac{1}{J} \sum_j \ln(\pi_{.j.}) + \frac{1}{K} \sum_k \ln(\pi_{..k}) \quad (34)$$

Sujeto a las restricciones:

$$\sum \lambda_i^X = \sum \lambda_j^Y = \sum \lambda_k^Z = 0 \quad (35)$$

Así se tienen las estimaciones para los demás modelos.

#### Validación del modelo estimado.-

Para validar el modelo estimado, se utiliza la prueba de hipótesis siguiente.

$$\begin{cases} H(0) = \text{el modelo estimado es adecuado} \\ H(1) = \text{el modelo estimado no es adecuado} \end{cases}$$

Para aceptar o rechazar la hipótesis nula, se utiliza al igual que en los modelos logísticos multinomial los estadísticos de bondad de ajuste, estos son la desviación y el estadístico Chi-cuadrado.

$$D = 2 \left[ \sum_i y_i \ln \left( \frac{y_i}{e_i} \right) \right] \quad (36)$$

$$X^2 = \sum_{i=1}^N \frac{(o_i - e_i)^2}{e_i} \rightarrow \chi_{N-p}^2 \quad (37) \quad o = \text{observado}; e = \text{estimado}$$



## 4.2.- Modelo log lineal para muertes violentas

Definamos las variables que consideramos para la construcción del modelo log lineal, por medio de un análisis univariante y multivariante, con esto se tiene las variables siguientes:

Tabla 4. 2.1.- Definición de variables

NOMBRE DE VARIABLE	TIPO DE VARIABLE	MODALIDADES	CODIFICACIÓN
TIPO DE MUERTE	Cualitativa	Homicidios- Suicidios Accidentes de tránsito Accidentales	MV1 MV2 MV3
SEXO	Cualitativa	Masculino Femenino	M F
RANGO EDAD (años)	Cualitativa	0-18. 19-39. 40-59. >60.	E1= 0-19. E2= 20-39. E3= 40-59. E4= >60.

En la tabla 3.1, se muestran las variables categóricas que utilizaremos para la elaboración de los modelos log lineales.

### 4.2.1.- Estimación del modelo.-

Una vez que se definió la tabla de contingencia, se procederá a la definición del modelo log lineal; para esto, utilizaremos el módulo del paquete SPSS (loglinear-models) tanto, en la definición del modelo por medio de un procedimiento iterativo jerárquico, cuanto para la validación del modelo con un método descriptivo y representaciones gráficas de los errores del modelo definido. Iniciamos la definición del modelo probando las interacciones entre variables; para esto probamos dos hipótesis de interacción que son:

$$\begin{cases} H(0): \text{ todos los efectos debido a las interacciones de orden } k \text{ o superior } = 0 \\ H(0): \text{ todos los efectos debido a las interacciones de orden } k = 0 \end{cases}$$

Figura4. 2.1.- Pruebas sobre los efectos de interacción de variables

```

***** H I E R A R C H I C A L   L O G   L I N E A R   * * * * *
Tests that K-way and higher order effects are zero.

      K      DF      L.R. Chisq      Prob      Pearson Chisq      Prob      Iteration
      3         6          5,588      ,4709          5,594      ,4702          4
      2        17         301,863      ,0000          313,854      ,0000          2
      1        23         2458,971      ,0000          3246,263      ,0000          0

Tests that K-way effects are zero.

      K      DF      L.R. Chisq      Prob      Pearson Chisq      Prob      Iteration
      1         6         2157,108      ,0000          2932,410      ,0000          0
      2        11         296,275      ,0000          308,260      ,0000          0
      3         6          5,588      ,4709          5,594      ,4702          0

```

La figura anterior muestra los resultados para aceptar o rechazar las dos pruebas de hipótesis de los efectos de las variables.

En las dos pruebas se tiene que los efectos de orden dos son significativos, para el efecto de tercer orden no se tiene evidencia estadística de algún nivel de significación. A continuación se muestra la estimación del modelo mediante un procedimiento iterativo.

Figura4. 2.2.- Proceso iterativo de la estimación del modelo

```

Step 2

The best model has generating class

      TMUER_1*AGE1
      TMUER_1*SEXO
      AGE1*SEXO

Likelihood ratio chi square =      5,58811      DF = 6      P =      ,471
***** H I E R A R C H I C A L   L O G   L I N E A R   * * * * *

The final model has generating class

      TMUER_1*AGE1
      TMUER_1*SEXO
      AGE1*SEXO

The Iterative Proportional Fit algorithm converged at iteration 0.
The maximum difference between observed and fitted marginal totals is      ,179
and the convergence criterion is      ,567

```

El procedimiento iterativo estima en la segunda iteración el mejor modelo, reportando interacciones de segundo orden como se observa en la figura; las iteraciones de segundo orden no salen del modelo estimado.

#### 4.2.2.- Validación del modelo.-

Lo que sigue es aceptar o rechazar el modelo estimado, esta decisión se tomará probando la hipótesis siguiente:

$$\begin{cases} H(0) = \text{el modelo estimado es adecuado} \\ H(1) = \text{el modelo estimado no es adecuado} \end{cases}$$

**Figura 4.2.3.- Pruebas de bondad de ajuste**

##### Goodness-of-fit Statistics

	Chi-Square	DF	Sig.
Likelihood Ratio	5,5880	6	,4709
Pearson	5,5936	6	,4702

Para esta prueba se tiene dos estadísticos el de máxima verosimilitud y el de Pearson; como vemos en la figura no se rechaza la hipótesis nula, con lo que aceptamos el modelo estimado.

El mejor modelo estimado es:

**Figura 4.2.4.- Modelo estimado**

##### Model and Design Information

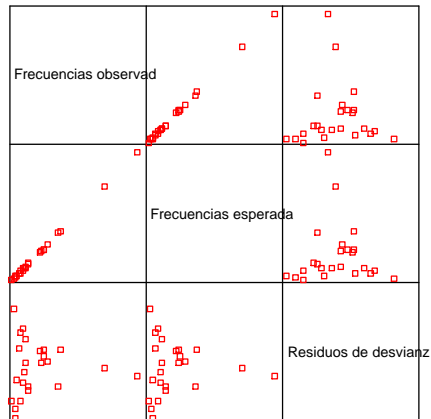
Model: Poisson

Design: Constant + AGE1 + SEXO + TMUER\_1 + SEXO\*AGE1 + TMUER\_1\*AGE1 + TMUER\_1\*SEXO

Además de las pruebas sobre la estimación del modelo y las iteraciones, se presentan pruebas gráficas, que representan la dispersión de las frecuencias observadas estimadas y la dispersión de los residuos, complementando con pruebas de normalidad sobre los residuos.

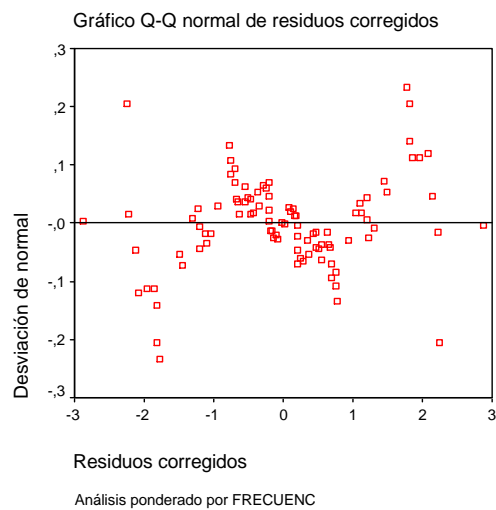
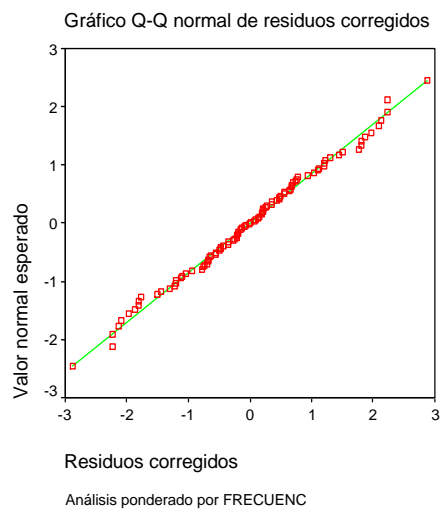
Figura 4.2.5.-

Modelo de Poisson



Tenemos un gráfico de dispersión de las frecuencias observadas, esperadas y residuos.

Figura4. 2.6.- Pruebas de normalidad de residuos



Es claro en la figura, la distribución de los residuos sobre la recta normal y la dispersión aleatoria de los mismos, nos muestra que estos siguen aproximadamente una distribución normal.

#### 4.2.3.- Interpretación del modelo.-

Una vez estimado y validado, el modelo log lineal para describir las frecuencias de muertes violentas, se procede a interpretar en términos de asociaciones de independencia o de dependencia entre las variables categóricas.

Se puede indicar que existen interacciones de segundo orden entre Sexo y Edad, Sexo y Tipo de muerte, y entre Tipo de muerte y Edad. Con esto se puede indicar que existe una dependencia condicional del tipo de muerte con la edad y el sexo de la víctima.

### 4.3.- Modelo log lineal para delitos contra vehículos

En la tabla se muestra las variables que intervendrán en la estimación del modelo, y con las cuales se construye la tabla de contingencia.

**Tabla4. 3.1.- Definición de las variables.-**

NOMBRE DE VARIABLE	TIPO DE VARIABLE	MODALIDAD	CODIFICACIÓN
TIPO DE VEHÍCULO	Cualitativa	Automóvil Camioneta Todo terreno Motocicleta	TV1 TV2 TV3 TV4
LUGAR	Cualitativa	Vía pública Estacionamiento	L1 L2
CIRCUNSTANCIA	Cualitativa	Estacionado Interceptado.	C1 C2

Como se observa en la tabla anterior, todas las variables son de tipo cualitativo y categórico.

#### 4.3.1.- Estimación del modelo.-

Ahora, sigue estimar los coeficientes del modelo planteado, para esto se presenta las estimaciones que da el paquete por medio de un algoritmo de iteración, procedemos a realizar dos pruebas de hipótesis sobre los efectos de las interacciones de orden  $k$  o superior y las de orden  $k$ .

$$\begin{cases} H(0): \text{ todos los efectos debido a las interacciones de orden } k \text{ o superior } = 0 \\ H(0): \text{ todos los efectos debido a las interacciones de orden } k = 0 \end{cases}$$

Figura 4.3.1.- Pruebas sobre los efectos de interacción de variables

```

***** H I E R A R C H I C A L   L O G   L I N E A R   * * * * *
Tests that K-way and higher order effects are zero.

      K      DF      L.R. Chisq      Prob      Pearson Chisq      Prob      Iteration
      3       3          3,065      ,3817          3,048      ,3843          3
      2      10         113,503      ,0000          112,098      ,0000          2
      1      15        2796,385      ,0000          3563,561      ,0000          0

Tests that K-way effects are zero.
.
      K      DF      L.R. Chisq      Prob      Pearson Chisq      Prob      Iteration
      1       5        2682,882      ,0000          3451,463      ,0000          0
      2       7         110,437      ,0000          109,050      ,0000          0
      3       3          3,065      ,3817          3,048      ,3843          0

```

Como se muestra en la figura anterior se tiene interacciones de segundo. Con esto no rechazamos la hipótesis nula de que las interacciones de tercer orden son cero.

Figura4. 3.2.- Proceso iterativo de la estimación del modelo

```

If Deleted Simple Effect is      DF      L.R. Chisq Change      Prob      Iter
VEHICULO*CIRCUNST                3          30,161      ,0000          2
VEHICULO*LUGAR1                  3          42,324      ,0000          2
CIRCUNST*LUGAR1                   1          41,069      ,0000          2

Step 2

The best model has generating class

    VEHICULO*CIRCUNST
    VEHICULO*LUGAR1
    CIRCUNST*LUGAR1

Likelihood ratio chi square =      3,06539      DF = 3      P =      ,382

The final model has generating class

    VEHICULO*CIRCUNST
    VEHICULO*LUGAR1
    CIRCUNST*LUGAR1

The Iterative Proportional Fit algorithm converged at iteration 0.
The maximum difference between observed and fitted marginal totals is      ,280
and the convergence criterion is      ,665

```

El modelo se estima en la segunda iteración, como se observa en la figura existen interacciones de segundo orden.

### 4.3.2.- Validación del modelo.-

Lo que sigue es validar el modelo tanto con pruebas estadísticas sobre las estimaciones, como con pruebas gráficas sobre los residuos. La prueba de hipótesis que contrastamos es la siguiente:

$$\begin{cases} H(0) = \text{el modelo es adecuado} \\ H(1) = \text{el modelo estimado no es adecuado} \end{cases}$$

Figura 4.3.3.- Pruebas de bondad de ajuste

Goodness-of-fit Statistics

	Chi-Square	DF	Sig.
Likelihood Ratio	3,0654	3	,3817
Pearson	3,0480	3	,3843

Como se observa en la figura anterior se tiene una buena estimación del modelo ya que los estadísticos de máxima verosimilitud y de Pearson no rechazan la hipótesis nula y tienen valores similares, ahora el modelo estimado es el que sigue:

Figura 4.3.4.- Modelo estimado

Model and Design Information

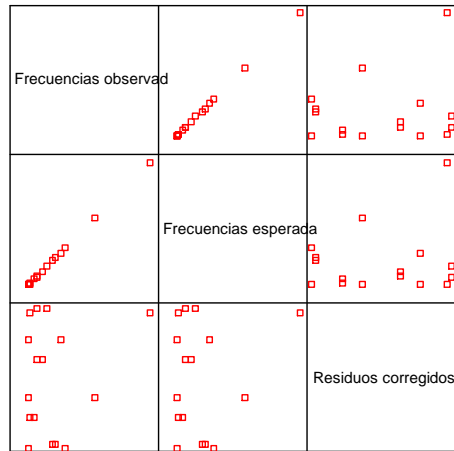
Model: Poisson  
 Design: Constant + CIRCUNST + LUGAR1 + VEHICULO + LUGAR1\*CIRCUNST + VEHICULO  
 \*CIRCUNST + LUGAR1\*VEHICULO

Este es el modelo que mejor estima las frecuencias esperadas para los delitos contra los vehículos.

Ahora mostremos las pruebas gráficas.

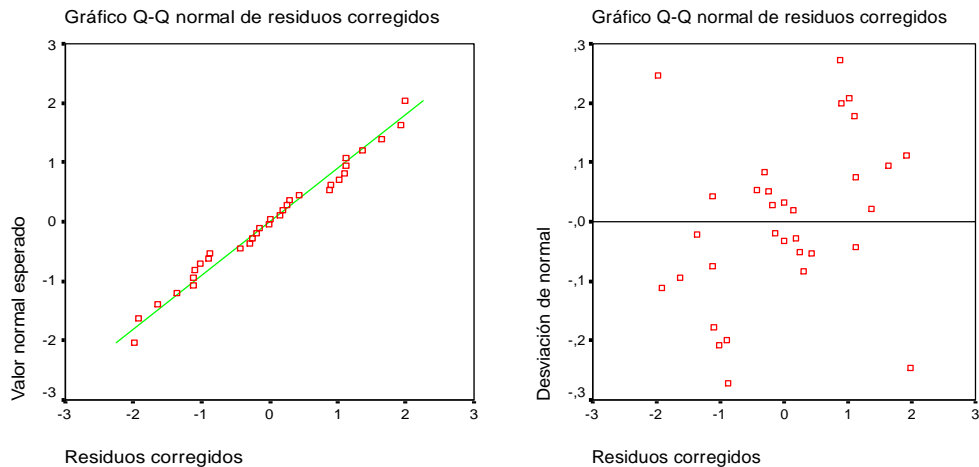
Figura 3.3.5.-

## Modelo de Poisson



La figura anterior muestra el comportamiento de las frecuencias observadas y estimadas con el modelo que se construyó, además de la dispersión de los residuos tanto con las frecuencias estimadas y las frecuencias observadas, se tiene un comportamiento aleatorio de los residuos.

Figura 3.3.6.- Pruebas de normalidad de residuos



Al representar los residuos en la recta normal se observa casi una total alineación de estos sobre la recta, lo cual indica que los residuos siguen aproximadamente una distribución normal, al igual que la dispersión aleatoria de los residuos estimados corregidos según su desviación normal.



### 4.3.3.- Interpretación del modelo.-

Una vez construido y validado el modelo, se procede a su interpretación. Como se observa en el modelo existen tres interacciones de segundo orden. Ahora las interacciones de segundo orden son: Lugar1-Circunstancia, vehiculo-lugar1 y Lugar-Circunstancia, con esto se tiene, que el tipo de vehículo, lugar y circunstancia son condicionalmente dependientes.

### 4.4.- Modelo log lineal para delitos contra la propiedad

Al igual que los delitos contra los vehículos, tenemos que estimar un modelo log lineal para los delitos a la propiedad. Las variables que consideraremos para la estimación del modelo son:

**Tabla 4.4.1.- Definición de variables.-**

NOMBRE DE VARIABLE	TIPO DE VARIABLE	MODALIDADES	CODIFICACIÓN
AFECTADO	Cualitativa	Personas	A1
		Domicilios	A2
		Empresas	A3
TIPO DELITO	Cualitativa	Robo.	TD1
		Asalto y robo.	TD2
		Hurto.	TD3
DÍA DEL HECHO	Cualitativa	Lunes	D1
		Martes,	D2
		Miércoles	D3
		Jueves	D4
		Viernes	D5
		Sábado,	D6
		Domingo.	D7

Con las tres variables cualitativas y sus categorías se procede a construir la tabla de contingencia y estimar el modelo log lineal.

#### 4.4.1.- Estimación del modelo.-

Ahora lo que se quiere es estimar las interacciones de orden  $k$  y superior y las interacciones de orden  $k$  entre las variables, para lo cual se procede a realizar las pruebas de hipótesis siguientes:

$$\left\{ \begin{array}{l} H(0): \text{ todos los efectos debido a las interacciones de orden } k \text{ o superior } = 0 \\ H(0): \text{ todos los efectos debido a las interacciones de orden } k = 0 \end{array} \right.$$

Figura 4.4.1.- Pruebas sobre los efectos de interacción de variables

```

***** H I E R A R C H I C A L   L O G   L I N E A R *****
Tests that K-way and higher order effects are zero.

      K      DF      L.R. Chisq      Prob      Pearson Chisq      Prob      Iteration
      3      24      28,329      ,2464      28,420      ,2427      6
      2      52      3910,142      ,0000      3798,873      ,0000      2
      1      62      9039,968      ,0000      8941,205      ,0000      0
Tests that K-way effects are zero.

      K      DF      L.R. Chisq      Prob      Pearson Chisq      Prob      Iteration
      1      10      5129,826      ,0000      5142,331      ,0000      0
      2      28      3881,813      ,0000      3770,453      ,0000      0
      3      24      28,329      ,2464      28,420      ,2427      0

```

La figura anterior muestra las estimaciones de orden K y superior sobre las interacciones de las variables en la construcción del modelo log lineal, la figura muestra que existen iteraciones de segundo orden. No se tiene interacción de tercero orden.

Figura 4.4.2.- Proceso iterativo de la estimación del modelo

```

***** H I E R A R C H I C A L   L O G   L I N E A R *****
If Deleted Simple Effect is          DF      L.R. Chisq Change      Prob      Iter
TDELITO*AFECTADO                      4          3696,698      ,0000      2
AFECTADO*DIAHEC                       12          174,854      ,0000      2
□
Step 3

The best model has generating class

TDELITO*AFECTADO
AFECTADO*DIAHEC

Likelihood ratio chi square = 38,59043      DF = 36      P = ,353
The final model has generating class

TDELITO*AFECTADO
AFECTADO*DIAHEC

The Iterative Proportional Fit algorithm converged at iteration 0.
The maximum difference between observed and fitted marginal totals is      ,000
and the convergence criterion is      ,577

```

El método iterativo, estima el modelo en la tercera iteración, el cual consta de dos interacciones de segundo orden, no se tiene interacciones de tercer orden.

#### 4.4.2.- Validación del modelo.-

Ahora se procede a validar el modelo estimado con el proceso iterativo, para esto, se procede a presentar pruebas estadísticas y pruebas gráficas, que permite aceptar o rechazar el modelo estimado, para esto se prueba la hipótesis siguiente:

$$\begin{cases} H(0) = \text{el modelo es adecuado} \\ H(1) = \text{el modelo estimado no es adecuado} \end{cases}$$

**Figura 4.4.3.- Pruebas de bondad de ajuste**

Goodness-of-fit Statistics			
	Chi-Square	DF	Sig.
Likelihood Ratio	38,5903	36	,3533
Pearson	37,6323	36	,3944

La figura anterior muestra la validación estadística del modelo estimado, esto es, se muestra los estadísticos Ji cuadrado de máxima verosimilitud y de Pearson. Con esto tenemos que no rechazamos la hipótesis nula, entonces aceptamos el modelo estimado como el que mejor estima las frecuencias.

**Figura 4.4.4.- Modelo estimado**

Model and Design Information

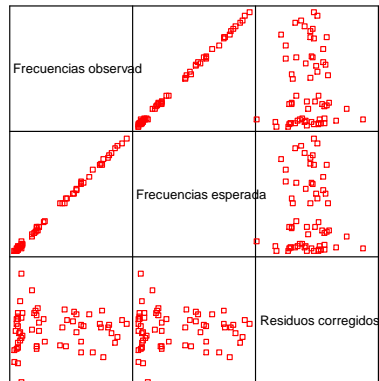
Model: Poisson

Design: Constant + AFECTADO + DIAHEC + TDELITO + AFECTADO\*DIAHEC + TDELITO\*AFECTADO

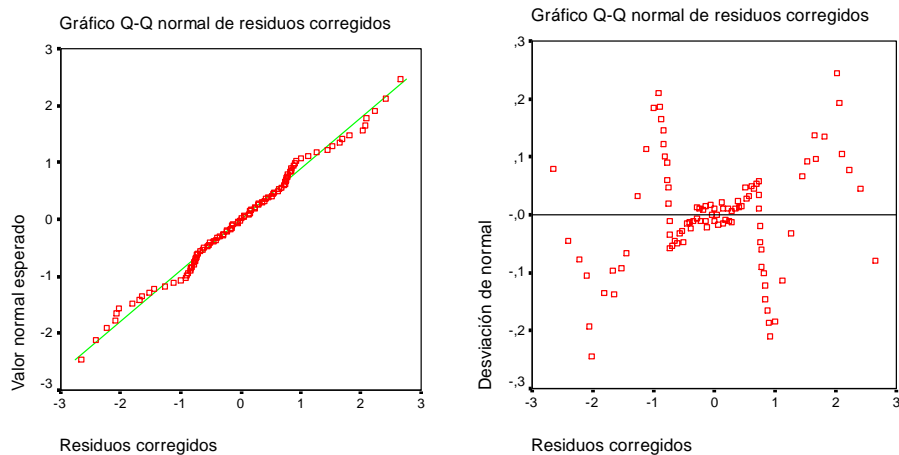
Ahora observemos las pruebas gráficas sobre las estimaciones y los residuos.

**Figura 4.4.5.-**

Modelo de Poisson



En la figura se tiene que las frecuencias observadas y estimadas son similares, esto se da por la tendencia lineal de los puntos. De igual manera se observa un comportamiento aleatorio de los residuos, entre la frecuencia observada y con la estimada por el modelo.

**Figura 4.4.6.- Pruebas de normalidad de residuos**

La representación de los residuos estandarizados muestra que estos siguen una distribución aproximadamente normal, y tienen un comportamiento aleatorio.

#### 4.4.3.- Interpretación del modelo.-

Una vez estimado y validado el modelo, lo que queda es interpretarlo. Las interacciones que se tiene son: Afectado-tipo de delito y afectado-día del hecho, no se tiene interacción entre Tipo de delito-día de hecho, con lo que se

tiene una independencia condicional por medio de afectado entre día del hecho y tipo de delito.

#### 4.5.- Modelo log lineal para violencia intrafamiliar.-

Las variables que intervendrán en la construcción de la tabla de contingencia y en la estimación del modelo log lineal son las siguientes:

**Tabla 4.5.1.- Definición de variables.-**

NOMBRE DE VARIABLE	TIPO DE VARIABLE	MODALIDAD	CODIFICACIÓN
RELACIÓN VÍCTIMA-AGRESOR	Cualitativa	Pareja. Expareja Familiar Otra.	R1 R2 R3 R4
TIPO DE AGRESIÓN	Cualitativa	Física. Psicológica. Sexual.	TA1 TA2. TA3
LUGAR DEL HECHO	Cualitativa	Domicilio Fuera del domicilio	L1 L2
ESTADO CIVIL	Cualitativa	Casado Divorciado Soltero Viudo	EC1 EC2 EC3 EC4

Se tiene cuatro variables en la construcción del modelo, todas son de tipo cualitativo.

##### 4.5.1.- Estimación del modelo.-

Ahora realizamos el proceso iterativo para determinar las interacciones entre variables y estimar el mejor modelo que se adecua a las frecuencias observadas. Primero probamos las hipótesis sobre las interacciones de los efectos de las variables, que son las siguientes:

$$\left\{ \begin{array}{l} H(0): \text{ todos los efectos debido a las interacciones de orden } k \text{ o superior } = 0 \\ H(0): \text{ todos los efectos debido a las interacciones de orden } k = 0 \end{array} \right.$$

**Figura 4.5.1.- Pruebas sobre los efectos de interacción de variables**

Tests that K-way and higher order effects are zero.

K	DF	L.R. Chisq	Prob	Pearson Chisq	Prob	Iteration
4	18	14,320	,7080	12,841	,8010	3
3	57	67,374	,1636	65,672	,2016	5
2	86	1632,718	,0000	2187,588	,0000	2
1	95	32133,530	,0000	99491,695	,0000	0

Tests that K-way effects are zero.

K	DF	L.R. Chisq	Prob	Pearson Chisq	Prob	Iteration
1	9	30500,812	,0000	97304,107	,0000	0
2	29	1565,343	,0000	2121,917	,0000	0
3	39	53,054	,0660	52,831	,0687	0
4	18	14,320	,7080	12,841	,8010	0

Según la figura anterior se observa, que las interacciones representativas estadísticamente son las de segundo orden. No se tendrían interacciones de tercero y cuarto orden.

**Figura 4.5.2.- Proceso iterativo de la estimación del modelo**

Step 6

The best model has generating class

```
AGRESI*VICAGRE*HLUGAR
AGRESI*ESTVICT
VICAGRE*ESTVICT
```

Likelihood ratio chi square = 61,16889 DF = 54 P = ,234

□

\* \* \* \* \* H I E R A R C H I C A L L O G L I N E A R \* \* \* \* \*

The final model has generating class

```
AGRESI*VICAGRE*HLUGAR
AGRESI*ESTVICT
VICAGRE*ESTVICT
```

The Iterative Proportional Fit algorithm converged at iteration 0.

The maximum difference between observed and fitted marginal totals is ,449  
and the convergence criterion is 2,492

Como indica la figura se tiene la convergencia del modelo en la sexta iteración. En la que se tiene dos relaciones de segundo orden y una de tercer orden. Es claro que no se evidencio en la prueba de interacciones que existan ni de tercero ni de cuarto orden, pese a esto en el proceso iterativo se tiene una interacción de tercer orden.

#### 4.5.2.- Validación del modelo.-

Ahora lo que sigue es aceptar o rechazar el modelo estimado, para esto se prueba la hipótesis nula siguiente:

$$\begin{cases} H(0) = \text{el modelo estimado es adecuado} \\ H(1) = \text{el modelo estimado no es adecuado} \end{cases}$$

Para validar el modelo estimado se presentan pruebas estadísticas de máxima verosimilitud y de Pearson, los cuales permiten aceptar o rechazar la hipótesis nula.

**Figura 4.5.3.- Pruebas de bondad de ajuste**

Goodness-of-fit test statistics

Likelihood ratio chi square =	22,50061	DF = 24	P = ,549
Pearson chi square =	21,26870	DF = 24	P = ,623

La figura muestra las estimaciones de los estadísticos Ji cuadrado de máxima verosimilitud y de Pearson sobre la validez del modelo, los dos estadísticos nos indica que no se rechaza la hipótesis nula, con lo cual aceptamos el modelo estimado, como el que mejor estima las frecuencias esperadas.

**Figura 4.5.4.- Modelo estimado**

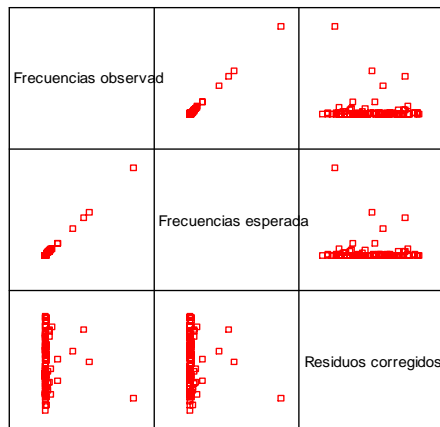
Model and Design Information

Model: Poisson  
 Design: Constant + AGRESI + ESTVICT + HLUGAR + VICAGRE + AGRESI\*ESTVICT +  
 VICAGRE\*ESTVICT + AGRESI\*VICAGRE\*HLUGAR

A continuación se muestra pruebas gráficas tanto para las frecuencias estimadas con el modelo, como para los residuos.

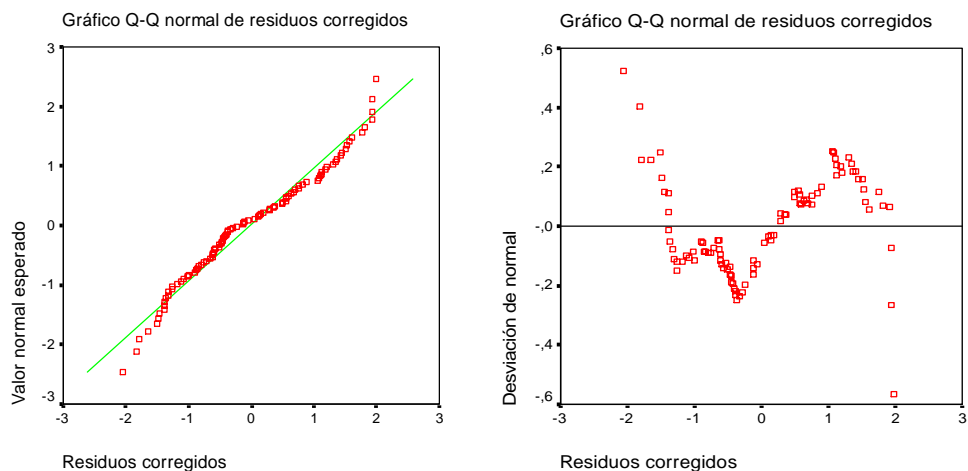
Figura 4.5.5.-

## Modelo de Poisson



Según la figura anterior, se tiene que el modelo estimado, predice adecuadamente las frecuencias esperadas, esto por la tendencia lineal entre la frecuencia observada y la frecuencia estimada con el modelo. Al relacionar los residuos con la frecuencia observada y estimada no se tiene un comportamiento aleatorio, lo que indicaría que este modelo no es adecuado.

Figura 4.5.6.- Pruebas de normalidad de residuos



Según la figura de normalidad, se tiene un comportamiento aproximadamente normal de los residuos, al igual que la aleatoriedad de los mismos.

### 4.5.3.- Interpretación del modelo.-

Una vez que se ha estimado y aceptado el modelo, lo que sigue es interpretar este modelo.



Por las interacciones de segundo orden podemos concluir que el estado civil de la víctima es condicionalmente dependiente del tipo de agresión y de la relación de la víctima con el agresor. Por último de la interacción de tercer orden se puede indicar que el tipo de agresión se relaciona con el lugar donde se produce la agresión por medio de la relación entre la víctima y el agresor.

#### 4.6.- Modelo log lineal para maltrato infantil

En la tabla siguiente se muestra las variables que son consideradas para construir el modelo log lineal para el maltrato infantil.

Tabla 4.6.1.- Definición de variables.-

NOMBRE DE VARIABLE	TIPO DE VARIABLE	MODALIDADES	CODIFICACIÓN
TIPO DE AGRESIÓN	Cualitativa	Física. Psicológica. Sexual.	TA1 TA2 TA3
EDAD	Cuantitativa discreta	0-4 5-10 11-14 15-18	E1 E2 E3 E4
SEXO	Cualitativa	Masculino Femenino.	M F
LUGAR DEL HECHO	Cualitativa	Domicilio, Fuera del domicilio	L1 L2

En la tabla anterior se tiene tres variables cualitativas categóricas, que son: relación víctima agresor, tipo de agresión y sexo de la víctima. Mientras que se tiene una variable cuantitativa que es la edad de la víctima, para efecto de la construcción del modelo log lineal, se procede a recodificar la variable en una categórica, para lo cual se procede a formar grupos de edades.

##### 4.3.5.1.- Estimación del modelo.

Enseguida se procede a construir el modelo log lineal para describir el maltrato infantil, el cual nos ayudara a estimar los comportamientos e interacciones de las diferentes categóricas de las variables cualitativas.

En primer lugar estimamos los efectos de las interacciones entre las variables que intervendrán en la construcción y definición del modelo, para esto se prueba las dos hipótesis siguientes:

$$\left\{ \begin{array}{l} H(0): \text{ todos los efectos debido a las interacciones de orden } k \text{ o superior } = 0 \\ H(0): \text{ todos los efectos debido a las interacciones de orden } k = 0 \end{array} \right.$$

**Figura 4.6.1.- Pruebas sobre los efectos de interacción de variables**

```

Tests that K-way and higher order effects are zero.

      K      DF      L.R. Chisq      Prob      Pearson Chisq      Prob      Iteration
      4      18          3,118      1,0000          3,153      1,0000          4
      3      57          32,756      ,9959          33,738      ,9940          5
      2      86          563,618      ,0000          813,467      ,0000          2
      1      95          3880,022      ,0000          5825,469      ,0000          0
-----

Tests that K-way effects are zero.

      K      DF      L.R. Chisq      Prob      Pearson Chisq      Prob      Iteration
      1       9      3316,403      ,0000          5012,002      ,0000          0
      2      29      530,862      ,0000          779,729      ,0000          0
      3      39      29,638      ,8604          30,585      ,8301          0
      4      18          3,118      1,0000          3,153      1,0000          0

```

Esta prueba muestra que efectos de tercer y cuarto orden no son significativos. Por el contrario la prueba de los k efectos muestra que las interacciones de cuarto orden son no significativas, existen interacciones sólo de segundo orden.

**Figura 4.6.2.- Proceso iterativo de la estimación del modelo**

```

Step 7

Likelihood ratio chi square = 33,07403      DF = 60      P = ,998
-----

□

* * * * * H I E R A R C H I C A L      L O G      L I N E A R * * * * *

The final model has generating class

AGE*LUGAR
TIMALTRA*LUGAR
TIMALTRA*SEXVICTI
TIMALTRA*AGE
SEXVICTI*AGE

The Iterative Proportional Fit algorithm converged at iteration 0.
The maximum difference between observed and fitted marginal totals is ,073
and the convergence criterion is ,250

```

En la figura anterior se muestra una parte del procedimiento iterativo para la estimación del modelo, este se produce en la séptima iteración, es claro que existen interacciones de segundo orden, estas son: age-lugar, timaltra-lugar, timaltra-sexvicti, timaltra-age, sexvicti-age, con esto el modelo estimado es el siguiente:

### 4.3.5.2.- Validación del modelo.-

Una vez estimado el modelo, se procede a su validación, para esto, se realizan pruebas estadísticas sobre el modelo estimado y pruebas sobre los residuos, estas son estadísticas y gráficas.

Iniciamos probando si el modelo que se estimó con el método iterativo es el adecuado, para esto se procede a probar la siguiente hipótesis:

$$\left\{ \begin{array}{l} H(0) = \text{el modelo estimado es adecuado} \\ H(1) = \text{el modelo estimado no es adecuado} \end{array} \right.$$

**Figura 4.6.3.- Pruebas de bondad de ajuste**

Goodness-of-fit Statistics			
	Chi-Square	DF	Sig.
Likelihood Ratio	33,0740	24	,1025
Pearson	34,4389	24	,0772

La figura anterior muestra dos estadísticos que nos permite aceptar o rechazar esta hipótesis, los estadísticos son de máxima verosimilitud y el estadístico de Pearson, los dos nos indican que no debemos rechazar la hipótesis nula, esto es, se acepta el modelo estimado como el que mejor estima los valores de las frecuencias esperadas.

**Figura 4.6.4.- Modelo estimado**

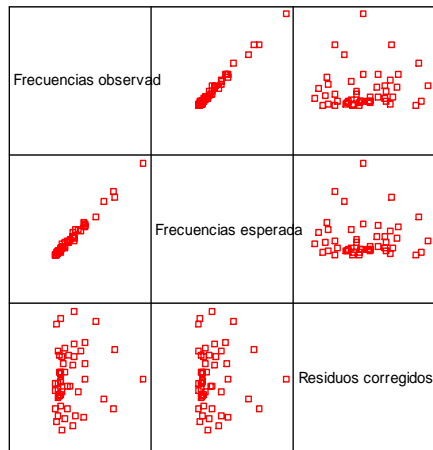
```
Model and Design Information

Model: Poisson
Design: Constant + AGE + SEXVICTI + TIMALTRA + AGE*LUGAR + SEXVICTI*AGE
        + TIMALTRA*AGE + TIMALTRA*LUGAR + TIMALTRA*SEXVICTI
```

Ahora se muestran algunas pruebas gráficas sobre el modelo y los residuos, con lo que se terminará de validar el modelo.

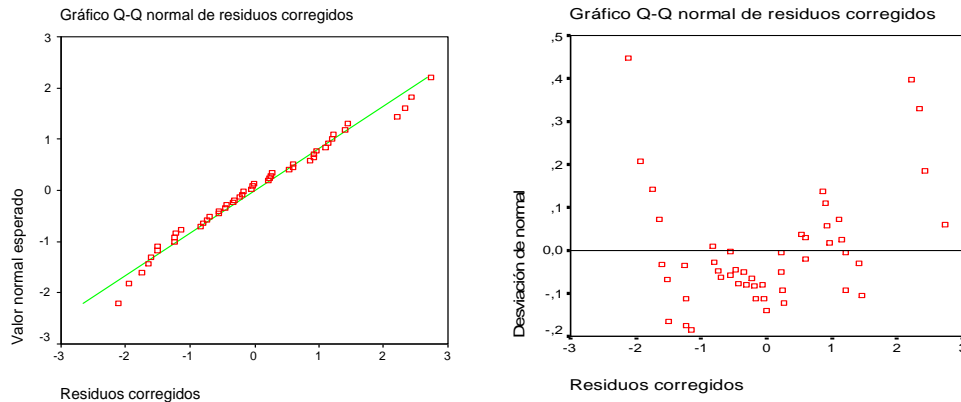
Figura 4.6.5.-

## Modelo de Poisson



La figura muestra el comportamiento de las frecuencias observadas y las estimadas, es claro que se tiene una tendencia lineal lo que indica que los valores estimados y observados son similares, además de esto se tiene también la distribución de los residuos tanto con los valores observados como con los valores estimados, el comportamiento de los residuos es casi aleatorio.

Figura 4.6.6.- Pruebas de normalidad de residuos



La figura anterior muestra la representación de los residuos corregidos sobre la recta normal, como vemos los residuos se superponen casi linealmente sobre esta, con esto podemos indicar que los residuos siguen aproximadamente una distribución normal, el comportamiento de los residuos corregidos según su desviación, la mayoría de estos se distribuyen aleatoriamente

#### **4.3.5.3.- Interpretación del modelo.-**

Ahora lo que nos queda es interpretar el modelo log lineal, para esto utilizaremos las interacciones de segundo orden que intervienen en el mismo.

Como se observa en el modelo, las interacciones que intervienen son: edad-lugar, sexo-lugar, tipo de maltrato-edad, tipo de maltrato-lugar, tipo de maltrato-sexo, la única interacción de segundo orden que no interviene es la de sexo y lugar, con esto se puede indicar que existe una dependencia condicional entre tipo de maltrato y las demás variables, y existe independencia condicional entre sexo y lugar de ocurrencia. Además de esto, el sexo de la víctima se relaciona con el lugar de ocurrencia de la agresión a través del tipo de maltrato y a través de la edad de la víctima.

## **Capítulo 5**

### **Comparación de los modelos construidos con los dos métodos.**

En este capítulo, nos concentramos en comparar estadísticamente los modelos estimados con las dos metodologías, tanto con modelos logísticos multinomiales, como con modelos log lineales, con el propósito de determinar las debilidades y fortalezas de cada uno.

Otro aspecto que debemos tomar en cuenta para la elección del modelo, es la predicción que este hace para valores futuros, y el interés que tenga para realizar propuestas de política pública, además de poder definir acciones, programas y proyectos, que se encaminen a reducir y erradicar los diferentes tipos de violencia y victimización que hemos estudiado.

Es claro también que se debe tener en cuenta que el modelo que se elija debe focalizarse en satisfacer las demandas de seguridad y convivencia de los ciudadanos, ya que todo plan encaminado a reducir la violencia no ha podido satisfacer adecuadamente las necesidades y demandas en seguridad que tienen los ciudadanos.

#### **5.1.- Comparación estadística entre los modelos.-**

Para realizar la comparación estadística entre los dos modelos que se construyeron por cada tipo de violencia, se utilizará la prueba de bondad de ajuste con los estadísticos que permiten validar a cada modelo, estos son, la desviación y el estadístico de Pearson; junto con esto se mostrarán los valores conforme la ley Chi-Cuadrado, los grados de libertad para cada uno de los modelos estimados y el nivel de significancia que tiene cada estadístico, con lo que se tendrá la robustez de la prueba.

### 5.1.1.- Comparación de modelos de muertes violentas.-

Se presenta un cuadro comparativo de los estadísticos de bondad de ajuste para los dos modelos.

**Tabla 5.1.- Estadísticos de bondad de ajuste para los modelos logístico multinomial y log lineal para muertes violentas**

ESTADÍSTICO	Chi-cuadrado		Grados de libertad		Nivel de significación	
	Regresión logística multinomial	Modelo log-lineal	Regresión logística multinomial	Modelo log-lineal	Regresión logística multinomial	Modelo log-lineal
Pearson	5,594	5,593	6	6	0,47	0,47
Desviación	5,588	5,588	6	6	0,471	0,47

Los estadísticos de validación de los dos modelos, son similares, esto es, cualquiera de los modelos, se puede utilizar en la construcción del índice, es conveniente el que sea el más sencillo de estimar, procesar y de interpretar.

### 5.1.2.- Comparación de modelos de delitos contra los vehículos.-

Seguimos con la comparación de los delitos contra los vehículos.

**Tabla 5.2.- Estadísticos de bondad de ajuste para los modelos logístico multinomial y log lineal para delitos contra los vehículos**

ESTADÍSTICO	Chi-cuadrado		Grados de libertad		Nivel de significación	
	Regresión logística multinomial	Modelo log-lineal	Regresión logística multinomial	Modelo log-lineal	Regresión logística multinomial	Modelo log-lineal
Pearson	3,048	3,048	3	3	0,384	0,38
Desviación	3,065	3,065	3	3	0,384	0,38

Este tipo de violencia se observa que los estadísticos de validación de los modelos no difieren, son prácticamente los mismos, al igual que el nivel de significancia de los estadísticos.

### 5.1.3.- Comparación de modelos de delitos contra la propiedad.-

La tabla para los delitos contra la propiedad es la siguiente.

**Tabla 5.3.- Estadísticos de bondad de ajuste para los modelos logístico multinomial y log lineal para delitos contra la propiedad**

ESTADÍSTICO	Chi-cuadrado		Grados de libertad		Nivel de significación	
	Regresión logística multinomial	Modelo log-lineal	Regresión logística multinomial	Modelo log-lineal	Regresión logística multinomial	Modelo log-lineal
Pearson	28,42	37,63	24	36	0,243	0,3944
Desviación	28,328	38,59	24	36	0,246	0,3533

La tabla muestra que los estimadores del modelo log lineal, son más fuertes que los del modelo logístico multinomial, tanto en los grados de libertad como en el nivel de significancia.

#### 5.1.4.- Comparación de modelos de violencia intrafamiliar.-

Tabla 5.4.- Estadísticos de bondad de ajuste para los modelos logístico multinomial y log lineal para violencia Intra familiar.

ESTADÍSTICO	Chi-cuadrado		Grados de libertad		Nivel de significación	
	Regresión logística multinomial	Modelo log-lineal	Regresión logística multinomial	Modelo log-lineal	Regresión logística multinomial	Modelo log-lineal
Pearson	56,254	21,268	48	24	0,193	0,623
Desviación	58,91	22,5	48	24	0,134	0,549

Los estadísticos del modelo logístico son mayores a los del modelo log lineal, esto es, existe mayor dispersión de los residuos en el modelo logístico, además de esto, el modelo logístico tiene el doble de grados de libertad que el modelo log lineal, sin embargo el nivel de significancia del modelo log lineal es superior al modelo logístico, esto se da por la estimación de los residuos entre lo estimado y lo observado.

#### 5.1.5.- Comparación de modelos de maltrato infantil.-

A continuación se muestra la tabla comparativa para los modelos de maltrato infantil.

Tabla 5.5.- Estadísticos de bondad de ajuste para los modelos logístico multinomial y log lineal para maltrato infantil

ESTADÍSTICO	Chi-cuadrado		Grados de libertad		Nivel de significación	
	Regresión logística multinomial	Modelo log-lineal	Regresión logística multinomial	Modelo log-lineal	Regresión logística multinomial	Modelo log-lineal
Pearson	20,017	34,44	18	24	0,332	0,0772
Desviación	24,67	33,44	18	24	0,134	0,1025

El modelo logístico multinomial es más robusto que el modelo log lineal; con esto es conveniente utilizar el modelo logístico, basta ver el nivel de significancia. Este resultado al igual que en el caso anterior se da por la diferencia que existe entre los valores observados y los estimados; ahora se tiene un mayor grado de libertad en el modelo log lineal.



## **5.2.- Comparación descriptiva de los fenómenos estudiados.-**

Para la aceptación de uno u otro modelo, enfoquémonos en la aplicabilidad de estos.

El modelo logístico multinomial, se encarga de estimar las categorías de una variable dependiente, en función de otras variables que se las conoce como independientes. Mientras que el modelo log lineal, lo que busca es estimar las relaciones entre las categorías de las variables cualitativas que definen la tabla de contingencia; en este modelo no se considera ninguna variable dependiente o independiente. Sin embargo la utilización de uno u otro modelo, dependerá de la validez descriptiva que tenga del fenómeno de estudio, esto es, lo que interesa es describir adecuadamente los hechos de violencia y en base a esto formular planes acciones y proyectos que reduzcan los niveles de violencia que se tiene.

## **5.3.- Elección de los mejores modelos e índices.-**

Como se mencionó antes, para la elección del modelo, este dependerá de la estimación estadística y de la necesidad y requerimiento de fomentar política pública.

### **5.3.1.- Muertes violentas**

Como muestra la tabla 5.1, los estadísticos de validación son similares, esto nos indica que estadísticamente cualquiera de los dos modelos estimados sea logístico o loglineal es adecuado en la descripción del fenómeno de estudio. Dada la definición desde la OMS de las muertes por causas externas, se tiene que considerar las acciones por el tipo de muerte, esto es, nos conviene describir los tipos de muerte, por lo que se utilizará el modelo logístico multinomial.

### **5.3.2.- Delitos contra los vehículos.-**

Estos modelos al igual que los de muertes violentas, los estadísticos de validación son iguales. Lo que nos interesa es como generar política para este tipo de violencia, por esto, nos preocuparemos de la modelación de las

frecuencias de la tabla de contingencia, y del nivel de asociación de las categorías de las variables que forman la tabla de contingencia, por lo que se utilizará el modelo log lineal.

### **5.3.3.- Delitos contra la propiedad.-**

Estadísticamente el mejor modelo es el log lineal, pero recordemos, que a diferencia de los delitos contra los vehículos, las categorías de la variable dependiente del modelo logístico, entre las cuales, están como afectados las personas, los domicilios, locales comerciales, es claro que no todos los afectados requieren el mismo tratamiento y políticas de prevención, control y atención de los tipos de violencia que pueden sufrir, por esta razón, el modelo que se empleará es el modelo logístico multinomial.

### **5.3.4.- Violencia Intra familiar.-**

Según los valores de la tabla 5.4, es claro que los estadísticos de validación del modelo son mejores estadísticamente los valores del modelo log lineal, este criterio de selección del modelo se complementa con efectos prácticos del tratamiento de este tipo de violencia. Si bien en la ley 103 se definen tres tipos de violencia, que son física, psicológica y sexual, es por demás obvio que el cometimiento de uno de estos tipos de maltrato, generan los demás tipos de violencia (en menor grado el maltrato psicológico), es por esto que es también aconsejable el utilizar el modelo log lineal y no el modelo logístico multinomial.

### **5.3.5.- Maltrato infantil.-**

Según la tabla 5.5, los estadísticos de validación, indican que estadísticamente es mejor utilizar el modelo logístico multinomial y no el modelo loglineal. Recordemos que para el maltrato infantil nos basamos en el código de la niñez y adolescencia, donde para el modelo logístico multinomial, la variable dependiente es el tipo de maltrato, que son físico, psicológico y sexual.

## Capítulo 6

### Conclusiones y recomendaciones

En este capítulo se muestran los resultados más importantes de la investigación, así como también algunas recomendaciones que pueden ayudar al mejoramiento del mismo.

#### 6.1.- Conclusiones.-

- Los modelos logísticos multinomial y log lineal, son los de mayor utilidad en la modelación de fenómenos asociados con las ciencias sociales, ya que permiten estudiar las relaciones entre variables cualitativas o categóricas. En el modelo logístico multinomial se relaciona una variable dependiente multicategórica, con variables explicativas sean estas cualitativas o cuantitativas, mientras que en el modelo loglineal, se busca estudiar las relaciones entre las categorías de las variables cualitativas representadas en una tabla de contingencia; además para estos dos modelos existen mecanismos de validación.
- Según la construcción de los modelos para los diferentes tipos de delitos, la vulnerabilidad de los ciudadanos, se tiene por diferentes causas, las que varían según el tipo de violencia a la que estamos expuestos.
- Dependiendo del tipo de muerte, se observa que el mayor riesgo se da en hombres y en edades de 15 hasta los 30 años; es claro que son edades productivas, y que es importante fomentar acciones preventivas, como el restringir totalmente el portar armas en la noche, sean que estas cuenten o no con permisos de portación de las mismas, y también la portación de armas blancas.
- Es importante analizar el comportamiento de las muertes en mujeres, ya que pese a que estas tienen un menor riesgo, este se centra en edades productivas y de gestación. Como medidas preventivas es importante fomentar la denuncia de delitos menores y que por lo general

desemboca en lo que se conoce como feminicidio y donde no funcionarían acciones de control ya que por lo general esto se da en los hogares de la víctima.

- Los robos de hogares tienen un comportamiento puntual, esto es, se dan en horas de la mañana y la tarde; los lapsos son de 10:00 hasta las 12:00 en la mañana y de 14:00 hasta las 17:00, de lunes a viernes, lo que se debe a que en estas horas la mayoría de hogares permanecen abandonados o con muy pocas personas (por lo general de servicio o ancianos), lo que facilita el cometimiento del delito en los rangos y horarios expuestos anteriormente.
- Con todas las limitantes que se tiene en la información y subregistro de los delitos contra las mujeres y los niños niñas y adolescentes, se logró estimar un modelo log lineal para violencia intra familiar y un modelo logístico multinomial para maltrato infantil.
- Si observamos el maltrato a niños niñas y adolescentes, el tipo de violencia y afectado depende también de la edad y el sexo de la víctima. Así, en los niños el mayor tipo de violencia es la física, y en edades de 0 hasta los cuatro años, mientras que las mujeres tienen mayor maltrato en edades superiores a los 10 años y el tipo de maltrato que se presenta es el sexual junto al psicológico.
- La calidad de información en maltrato infantil y violencia intrafamiliar, es limitada, por el tratamiento legal, así, si una persona sufre de agresión sexual esta conlleva agresión psicológica y física, lo mismo ocurre con la agresión física, pero por el tratamiento que se da en la ley, no se evidencia estas particularidades que se producen en estos tipos de violencia.
- Como se mencionó anteriormente, uno de los tipos de violencia, que presenta mayor complejidad en el análisis, es el de violencia intra familiar, no sólo por el subregistro que existe, sino también, por la complejidad de los tipos de violencia que se tiene estipulado en la ley 103; pese a esto el modelo creado puede ayudarnos a interpretar el comportamiento de estos delitos; así, el maltrato físico y sexual es el de

mayor riesgo para las mujeres, siendo la pareja actual, como esposo, novio o conviviente, quien ejecuta estos maltratos.

## **6.2.- Recomendaciones.-**

- Es conveniente, realizar una mejor recolección de la información de delitos contra la propiedad, ya que se observa que existen subregistros en los datos, así, existen delitos que no son denunciados, y por tanto no consta en las estadísticas institucionales, esto se da por la desconfianza que tiene la ciudadanía en las instituciones de justicia, lo que fomenta la no denuncia de estos hechos de violencia, para reducir esta inconsistencia de información, se puede almacenar los registros de los auxilios detallados en los partes policiales puede ayudar a estimar y modelar mejor el fenómeno de estudio, y que no se almacenan y procesan como estadísticas de violencia, al igual que la ejecución periódica de encuestas de victimización.
- Para reducir el robo y asalto a personas, se debe implementar la vigilancia de lugares de concentración masiva de personas, como las estaciones norte y sur del trolebús, paradas de buses de transporte público, locales comerciales y parques de la ciudad.
- Además de medidas de vigilancia, es necesario intervenir sobre las economías ilegales de comercialización de artículos robados, ya que reduciendo la oferta de estos, y los recursos económicos que generan, se impacta directamente en el cometimiento de robos de enseres, celulares, etc.
- Se hace imprescindible fomentar la implantación de estacionamientos públicos donde se pueda dejar con la garantía de que el vehículo es vigilado y está seguro, ya que en la ciudad el factor de riesgo es la vía pública y cuando este está estacionado y abandonado.
- La descripción y modelación estadística de la violencia intra familiar, tiene deficiencia desde el tratamiento que se da a este tipo de violencia en la ley 103 de la mujer y la familia. La ley trata tres tipos de maltrato: el físico, el psicológico y el sexual, pero es claro que dependiendo del

grado de violencia del maltrato, se van a presentar el maltrato físico, psicológico y sexual, y no se podría discernir por cual tipo de maltrato acusar al victimario.

- El maltrato infantil, se lo debe investigar a mayor detalle, sea con grupos focales o con entrevistas directas a niños por parte de personal capacitado en el tratamiento de violencia infantil, ya que por medio de la recolección de denuncias, se tiene una cifra que no es real, las variables que se tienen en denuncias, no ayudan a determinar las causas que producen y generan este tipo de violencia completamente.
- Es conveniente que se puedan contrastar los resultados que se obtuvieron con uno u otro modelo, con otras investigaciones, tales como por medio de encuestas, o estudios cualitativos o de las ciencias sociales.
- Otro estudio que podría mejorar el realizado, es investigar, el tratamiento que se da a todas las denuncias y los resultados finales que se han logrado, ya que para una total comprensión del fenómeno delictual, no se debería dejar de lado la parte administradora de justicia y la investigativa.
- De los tipos de violencia, es la violencia intra familiar y el maltrato a niños, niñas y adolescentes, los que registran la mayor cantidad de subregistro, ya que existe una subordinación aceptada por las víctimas, debido a factores que no se abordaron en este estudio.

## Referencias bibliográficas

1. Abdelmonem A, Virginia C y May S, (2004): "Computer-Aided Multivariate Analysis", fourth edition, *Chapman & Hall/CRC 2004*.
2. Agresti, A. (1996): "An introduction to categorical data analysis", John Wiley & Sons, INC 1996.
3. Ayala, J; Iturralde, T; Rodríguez, A: "Construcción de índices simplificados de riesgo país: Aproximación a los casos de Europa y América", Universidad del País Vasco-Euskal Herriko Unibertsitatea, Cuadernos de Gestión Vol. 2. Nº 2, 2002
4. Barreiro, J; Ruzo, E; Losada, F: "Modelo logit multinomial y regresión con variables ficticias: una aplicación regional al sector lácteo". Working Paper Series Economic Development Nº 81 (<http://www.usc.es/economet/eea.htm>).
5. Escobar, J y Romero, J: "Métodos de construcción de índices de precios de vivienda", webpondo.org
6. Giudici, P. (2003): "Applied data mining", *John Wiley & Sons*, 2003.
7. Organización Mundial de la Salud, (2004): "Informe mundial sobre prevención de los traumatismos causados por el tránsito".
8. Lévy, J. y Varela, M. (2003): "Análisis Multivariable para las Ciencias Sociales", Prentice Hall, 2003.
9. Rencher, A. (2002): "Methods of Multivariate Analysis". John Wiley & Sons, INC. Publication, 2003.
10. Alcaldía Mayor de Bogotá, "Investigación sobre el suicidio en Bogotá", Universidad Nacional de Colombia, 2001.
11. Saposnikow J, Salazar Jy Carrillo F. (2000): "Convivencia y seguridad: un reto a la gobernabilidad", Banco Interamericano de Desarrollo. Universidad de Alcalá, 2000.
12. Theus, M and Lauer, S: "Visualizing Loglinear Models", *Journal of Computational and Graphical Statistics*, volume 8, Number 3, pages 396-412.

# ANEXO A

## Métodos exploratorios univariantes y multivariantes.

El análisis exploratorio está compuesto por técnicas estadísticas analíticas y gráficas, que se encargan de evidenciar las características más importantes de los datos representados por variables cuantitativas y cualitativas.

Para el propósito de nuestro estudio utilizaremos métodos exploratorios univariantes y multivariantes, que serán detallados a continuación.

### **A.1 Métodos univariantes.-**

A continuación, se describe los métodos empleados en la descripción de los datos de violencia, los métodos están asociados con variables cuantitativas y cualitativas.

Nos enfocaremos básicamente en métodos gráficos, ya que la mayoría de variables en todos los tipos de violencia que analizamos son cualitativas, en los métodos gráficos representaremos las frecuencias relativas de las categorías de las variables, mientras que para las variables cuantitativas emplearemos histogramas de frecuencias.

#### **A.1.1 Distribución de frecuencia.-**

La distribución de frecuencia, viene a ser un ordenamiento de los datos que permite observar los valores de frecuencia de cada una de las categorías.

Para mayor claridad visual, las distribuciones de frecuencia se representan en forma gráfica que se los conoce como diagramas de frecuencia, estos se usan para representar variables cualitativas ordinales y cuantitativas discretas. Existen diversas formas de presentar las frecuencias de una distribución.

#### **A.1.2 Histogramas de frecuencia.-**

Está formado por un conjunto de rectángulos que representan la frecuencia de cada categoría. Representa gráficamente las frecuencias correspondientes a los valores observados.

En los histogramas la altura de cada rectángulo es igual a la probabilidad de que  $X$  asuma el valor que corresponde al punto medio de su base.



Las gráficas de barras es similar a los histogramas, en general estos dos tipos de representaciones se usan en las estadísticas descriptivas para comunicar visualmente la información proporcionada por una distribución de probabilidad o una distribución de datos reales.

### **A.1.3 Distribución de frecuencia relativa.-**

El término relativo indica la representación de proporciones o fracciones del total. La frecuencia relativa se calcula dividiendo la frecuencia de cada uno de los valores entre el total, que viene a ser la suma de las frecuencias. La frecuencia relativa tiene la ventaja de servir como referencia.

## **A.2 Métodos multivariantes.-**

Nos concentraremos en dos tipos de análisis multivariante, estos son: el análisis de correspondencias múltiples y los modelos de descomposición de la varianza.

### **A.2.1 Análisis de correspondencia múltiple (ACM).-**

El método de análisis de correspondencias múltiples estudia las relaciones entre cualquier número de características, cada una de ellas con varias modalidades, estas relaciones se representan en un gráfico bidimensional.

El ACM está diseñado para analizar tablas disyuntivas<sup>8</sup> completas, que son tablas de variables cualitativas.<sup>9</sup>

#### **NOMENCLATURA.-**

Una tabla disyuntiva completa **Z** queda descrita mediante:

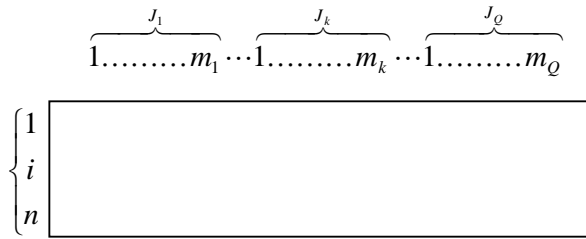
- Un conjunto de individuos  $I=1, \dots, i, \dots, n$
- Un conjunto de variables o preguntas  $J_1, \dots, J_k, \dots, J_Q$
- Un conjunto de categorías para cada pregunta  $1, \dots, m_k$

El número total de modalidades  $J = \sum_k m_k$

---

<sup>8</sup> Recoge la información de variables que tienen varias modalidades. La codificación utilizada es el código lógico, la pertenencia a una modalidad se indica con el valor 1 y la no pertenencia, con el 0. Cada individuo escoge una y sólo una modalidad.

<sup>9</sup> Siempre es posible transformar una variable métrica en cualitativa, dividiendo su intervalo de variación en clases de equivalencia sucesivas.



$\mathbf{Z}$  es la tabla  $I \times J$ . El elemento  $z_{ij}$  puede tomar el valor de 0 o 1, según que el individuo  $i$  haya elegido la modalidad  $j$  o no.

### A.2.3 El método del ACM.-

Veamos las particularidades de un análisis factorial de correspondencias aplicado a una tabla disyuntiva completa.

#### a) Significado de la terminología.-

Los elementos de  $\mathbf{Z}$ ,  $z_{ij} = k_{ij}$  son 0 o 1

$k_i = \sum_j k_{ij} = Q$ , el número de preguntas

$f_{ij} / f_i = k_{ij} / k_i = 1/Q$ , el inverso del número de preguntas o 0 según que el individuo haya elegido o no la modalidad  $j$ .

Donde  $f_{ij} / f_i = k_{ij} / k_i$  son los perfiles fila (ver Grande, Abascal página 85).

$k_j = \sum_i k_{ij}$ , es el número de individuos que poseen la modalidad  $j$ .

#### b) La matriz a diagonalizar.-

Para obtener los factores<sup>10</sup> es necesario diagonalizar la matriz  $V$  que en este caso particular se convierte en

$$V = \frac{1}{Q} D^{-1} B \quad (\text{ver Grande, Abascal página 124})$$

Donde  $B = Z'Z$ , es la tabla de Burt. Es una matriz simétrica formada por  $Q^2$  bloques:

- Los bloques de la diagonal son tablas diagonales que cruzan una pregunta con sí misma  $Z_k'Z_k$ .
- Los bloques fuera de la diagonal son tablas de contingencia obtenidas cruzando las preguntas de dos en dos  $Z_k'Z_{k'}$ .

<sup>10</sup> El factor es un agrupamiento de modalidades y se interpreta, a partir de su correlación entre modalidades de distintas variables. Estas correlaciones son las proyecciones de las modalidades sobre los factores.

0	$C_{12}$		$C_{1Q}$
$C_{21}$	0		$C_{2Q}$
		0	
$C_{Q1}$	$C_{Q2}$		0

La matriz D es diagonal cuyos elementos son los de la matriz de Burt, los efectivos de cada modalidad.

**c) Formulas de transición.-**

Sustituyendo los valores del apartado a) en las formulas de transición del análisis factorial de correspondencias

$$F_{\alpha}(i) = \frac{1}{\sqrt{\lambda_{\alpha}}} \sum_j \frac{f_{ij}}{f_i} G_{\alpha}(j)$$

$$G_{\alpha}(j) = \frac{1}{\sqrt{\lambda_{\alpha}}} \sum_i \frac{f_{ij}}{f_j} F_{\alpha}(i)$$

Obtenemos

$$F_{\alpha}(i) = \frac{1}{\sqrt{\lambda_{\alpha}}} \frac{1}{Q} \sum_j k_{ij} G(j)$$

$$G_{\alpha}(j) = \frac{1}{\sqrt{\lambda_{\alpha}}} \frac{1}{k_j} \sum_i k_{ij} F(i)$$

**d) Centros de gravedad en las nubes y subnubes.-**

1. El centro de gravedad de los puntos de variables N(J) en AFC es  $\sqrt{f_i}$ .

En este caso es la distribución uniforme  $1/\sqrt{n}$ . En efecto

$$k_i = Q, \quad \sum_i k_i = nQ \text{ luego } f_i = 1/n \text{ y } \sqrt{f_i} = 1/\sqrt{n}.$$

2. El centro de gravedad de las modalidades de cada pregunta, es el mismo que el de la nube de modalidades N(J),  $1/\sqrt{n}$ . En efecto el centro de gravedad de la sub tabla se obtiene a partir de su distribución marginal.

3. Como el AFC es centrado y el centro de gravedad de las modalidades de una pregunta coincide con el conjunto J, y con el origen, las modalidades de cada cuestión están centradas en torno al origen, no pudiendo tener todos los mismos signos.

**e) Ayudas a la interpretación.-**

Se define la contribución de una variable  $J_k$  al factor  $\alpha$  como la suma de las contribuciones de las modalidades de la variable

$$CTA_{\alpha}(J_k) = \sum_{j \in J_k} CTA_{\alpha}(j)$$

**f) Las inercias.-**

- 1.- Si G representa el centro de gravedad, la inercia debida a la modalidad j es

$$I(j) = f_j d^2(j, G) = f_j \sum_i \left( \frac{f_{ij}}{f_j \sqrt{f_i}} - \sqrt{f_i} \right)^2 = \frac{k_j}{nQ} \sum_i \left( \frac{k_{ij}/nQ}{k_j 1/n} - 1/\sqrt{n} \right)^2 = \frac{1}{Q} \left( 1 - \frac{k_j}{n} \right)$$

- 2.- La inercia de una pregunta es la suma de inercias de las modalidades.

$$I(J_k) = \sum_{j \in J_k} I(j) = \sum_{j \in J_k} \frac{1}{Q} \left( 1 - \frac{k_j}{n} \right) = \frac{1}{Q} (m_k - 1).$$

- 3.- La inercia total es la suma de las inercias de todas las preguntas.

$$I = \sum_k I(J_k) = \sum_k \frac{1}{Q} (m_k - 1) = \frac{J}{Q} - 1.$$

**A.3 Modelos de análisis de varianza.-**

Lo que se quiere conseguir con estos modelos, es descomponer la varianza total en sumas de varianzas parciales correspondientes a fuentes de variación distintas. Las diferentes varianzas se comparan entre sí mediante un contraste Fisher-Snedecor.

La tabla de datos es la siguiente:

$Y_{11}$	...	$Y_{1n_1}$	Submuestra 1
$Y_{21}$	...	$Y_{2n_2}$	Submuestra 2
$\vdots$		$\vdots$	
$Y_{I1}$	...	$Y_{In_I}$	Submuestra I

Las observaciones por tratamiento no son necesariamente iguales.

El modelo que se plantea es el siguiente:

$$y_{ij} = \mu_i + \varepsilon_{ij}$$

Donde:

$\mu_i$  = media del tratamiento  $i$ .

$\varepsilon_{ij}$  = error aleatorio.

Las hipótesis sobre los errores son las siguientes:

1.  $\forall i, \forall j, \varepsilon_{ij} \rightarrow N(0, \sigma^2)$
2.  $\varepsilon_{ij}$  son independientes

La estimación de los parámetros se lo realiza por el método de mínimos cuadrados.

$$\text{Minimizar } f(\mu_1, \dots, \mu_I) = \sum_{i=1}^I \sum_{j=1}^{n_i} (y_{ij} - \mu_i)^2$$

Los estimadores que se obtienen son:

$$\hat{\mu}_i = \bar{y}_i$$

Estos estimadores son iguales con el método de máxima verosimilitud.

### Propiedades.-

1.  $E(\bar{y}_i) = \mu_i, \text{Var}(\bar{y}_i) = \sigma^2 / n_i$
2.  $\frac{1}{\sigma^2} \sum_{i=1}^I \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 \rightarrow \chi_{n-I}^2$
3.  $y_{ij} \rightarrow N(\mu_i, \sigma^2) \quad \forall i, j$ , con  $y_{ij}$  independientes.
4.  $\bar{y}_i \rightarrow N(\mu_i, \sigma^2 / n_i)$

### Teorema de Cochran.-

Sea  $\mathbf{Y}$  un vector aleatorio de  $\mathfrak{R}^n$ , que sigue una ley  $N_n(\mu, \Sigma)$ . Sean  $E_1, \dots, E_k$  son subespacios vectoriales de  $\mathfrak{R}^n$  dos a dos  $\Sigma^{-1}$  ortogonales, tal que

$$\mathfrak{R}^n = E_1 \oplus \dots \oplus E_k.$$

$Y = Y^1 + \dots + Y^k$ ;  $Y^j \in E_j$  Se tiene

1.  $\forall j, Y^j \rightarrow N(\mu_j, \Sigma)$
2.  $Y^1, \dots, Y^k$  son independientes

$$3. (Y^j - \mu^j)' \Sigma^{-1} (Y^j - \mu^j) \rightarrow \chi^2_{\dim(E_j)}$$

Con esto podemos tener la siguiente notación:

$$1. \text{STC} = \sum_{i=1}^I \sum_{j=1}^{n_j} (y_{ij} - \bar{y}_{..})^2$$

$$2. \text{SEC} = \sum_{i=1}^I n_i (\bar{y}_{i.} - \bar{y}_{..})^2$$

$$3. \text{SRC} = \sum_{i=1}^I \sum_{j=1}^{n_j} (y_{ij} - \bar{y}_{i.})^2$$

Donde:

STC= Suma total de cuadrado

SEC= Suma esperada de cuadrado

SRC= Suma de residuos de cuadrado.

**Teorema.-**

$$\text{STC} = \text{SEC} + \text{SRC}$$

Con todo esto se tiene, la tabla ANOVA, que es en donde se resume la información del análisis de varianza.

FUENTE	GRADOS DE LIBERTAD	SUMA DE CUADRADOS	MEDIA DE SUMA DE CUADRADOS	F
Tratamientos	I-1	$\sum_{i=1}^I n_i (\bar{y}_{i.} - \bar{y}_{..})^2$	$\frac{1}{I-1} \text{SEC}$	MSEC/MSRC
Error	n-I	$\sum_{i=1}^I \sum_{j=1}^{n_j} (y_{ij} - \bar{y}_{i.})^2$	$\frac{1}{n-i} \text{SRC}$	
Total	n-1	$\sum_{i=1}^I \sum_{j=1}^{n_j} (y_{ij} - \bar{y}_{..})^2$		

**Teorema.-** Si  $\mu_1 = \dots = \mu_I$  entonces  $F \rightarrow F_{(I-1, n-I)}$

Corolario.-

$$H_0 : \mu_1 = \dots = \mu_I = \mu$$

$$H_1 : \mu_i \neq \mu$$

Se rechaza  $H_0$  al nivel  $\alpha$  si y sólo si  $F > F_{(I-1)(n-I)}(\alpha)$

Las pruebas de igualdad de medias las combinamos con gráficos de caja, para visualizar el comportamiento de las medias en distintas poblaciones.

Para su construcción se utilizan cinco estadísticos de la distribución de frecuencias: el mínimo, el primer cuartil, la mediana, el tercer cuartil y el máximo; también se indica la posición de la media.

Un diagrama de caja se construye como sigue:

1. Ordenar los datos de la muestra y obtener el valor mínimo, el máximo, y los tres cuartiles  $Q_1, Q_2, Q_3$ .
2. Dibujar un rectángulo cuyos extremos son  $Q_1$  y  $Q_3$ , e indicar la posición de la mediana ( $Q_2$ ) mediante una línea.
3. Calcular unos límites admisibles superior e inferior que van a servir para identificar los valores atípicos. Estos límites se calculan con:

$$LI = Q_1 - 1.5(Q_3 - Q_1)$$

$$LS = Q_3 + 1.5(Q_3 - Q_1)$$

4. Considerar como valores atípicos los situados fuera del intervalo  $(LI, LS)$ .
5. Dibujar una línea que baya desde cada extremo del rectángulo central hasta el valor más alejado no atípico, es decir, que está dentro del intervalo  $(LI, LS)$ .
6. Identificar todos los datos que están fuera del intervalo  $(LI, LS)$ , marcándolos como atípicos.

## ANEXO B

### Análisis estadístico con datos perdidos.

Se preocupa del estudio de la matriz de datos cuando algunas de las entradas no son observadas.

#### B.1 Diseño de datos perdidos.-

Describe cuáles valores son observados y cuales están perdidos en la matriz de datos; además, define cuál mecanismo de datos perdidos conviene a la relación entre valores perdidos y observados en las variables de la matriz de datos.

Sea:

$Y = (y_{ij})$ , la matriz  $(n \times K)$  sin valores perdidos.

$$M = (m_{ij}) = \begin{cases} m_{ij} = 1 & \text{si } y_{ij} \text{ es perdido} \\ m_{ij} = 0 & \text{si } y_{ij} \text{ observado} \end{cases}$$

La matriz  $M$  define el diseño de los datos perdidos.

Algunos métodos para estimar datos perdidos se aplican a todo diseño de datos perdidos, mientras que otros métodos están restringidos a un diseño especial.

Supondremos que valores perdidos ocultan valores reales que son significativos para el análisis.

#### B.1.1 Mecanismos que se da a los datos perdidos.-

Tenemos:

$Y = (y_{ij})$  y

$$M = (m_{ij}) = \begin{cases} m_{ij} = 1 & \text{si } y_{ij} \text{ es perdido} \\ m_{ij} = 0 & \text{si } y_{ij} \text{ observado} \end{cases}$$

El mecanismo de los datos perdidos es caracterizado por la distribución condicional de  $M$  dado  $Y$ , es decir,

$f(M | Y, \Phi)$ ,  $\Phi$  es un parámetro desconocido.

Si los valores perdidos no dependen de los valores de los datos  $Y$ , perdidos u observados, esto es si,



$f(M | Y, \Phi) = f(M | \Phi) \quad \forall Y, \Phi$ . Se dice que los datos son perdidos completamente al azar. (MCAR)

**Observación.-**

Esto no quiere decir que el diseño por sí mismo es aleatorio, sino, que los valores perdidos no dependen de los datos.

Sea  $Y_{Obs}$ , las componentes observadas o entradas de  $Y$ , y  $Y_{mis}$ , las componentes perdidas. Suponemos que los valores perdidos dependen solamente de  $Y_{Obs}$  de  $Y$ , y no de las componentes que son perdidas.

$$f(M | Y, \Phi) = f(M | Y_{Obs}, \Phi) \quad \forall Y_{mis}, \Phi, \text{ perdidos y aleatorios (MAR).}$$

El mecanismo se llama no perdido y aleatorio (NMAR) si la distribución de  $M$  depende de los valores perdidos de la matriz  $Y$ .

Sea,

$$Y = (y_1, \dots, y_n)^T$$

$$M = (M_1, \dots, M_n) = \begin{cases} M_i = 0 & \text{si } y_i \text{ observada.} \\ M_i = 1 & \text{si } y_i \text{ no observada} \end{cases}$$

Suponemos que la distribución conjunta de  $(y_i, M_i)$  es independiente de las unidades. En particular, la probabilidad de que una unidad es observada no depende de los valores de  $Y$  o  $M$  para otras unidades.

$$f(Y, M | \theta, \phi) = f(Y | \theta) f(M | Y, \phi) = \prod_{i=1}^n f(y_i | \theta) \prod_{i=1}^n f(M_i | y_i, \phi)$$

Donde:

$f(y_i | \theta)$ , es la densidad de  $y_i$  con parámetro  $\theta$  desconocido.

$f(M_i | y_i, \phi)$ , es la densidad de la distribución de Bernoulli para el indicador binario  $M_i$ .

$P(M_i = 1 | y_i, \phi)$ , si  $y_i$  es perdido.

$P(M_i = 1 | y_i, \phi) = \phi$ , si los perdidos son independientes de  $y_i$ , entonces el mecanismo es (MCAR).

Si el mecanismo depende de  $y_i$  es (NMAR).

**B.1.2 Modelos para tablas de contingencia clasificadas parcialmente, ignorando el mecanismo de datos perdidos.**

Los datos son variables categóricas, tenemos  $n$  individuos y  $V$  variables categóricas, que se representa en una tabla de contingencia  $V$ -dimensional, con  $C$  celdas definidas por niveles conjuntos de las variables.

Las entradas en la tabla son:

$$\{n_{jk\dots t}\} \quad Y_1 = j, Y_2 = k, \dots, Y_V = t$$

Si la matriz de datos tiene valores perdidos, algunos de los casos, son clasificados parcialmente en la tabla.

Los casos que se clasifican totalmente producen una tabla  $V$ -dimensional de cantidades  $\{r_{jk\dots t}\}$ , y los casos clasificados parcialmente, forman sub tablas suplementarias de baja dimensión, definidas por un subconjunto de variables  $(Y_1, \dots, Y_V)$ , que son observadas.

Se utilizan métodos de Máxima Verosimilitud y de Bayes, para estimar los valores perdidos.

Un tipo más general de valores perdidos, ocurre cuando, el nivel  $j$  de una variable categórica particular  $Y_1$ , no es conocida, pero si se conoce que el caso cae dentro de un subconjunto  $S$  de valores de  $Y_1$ .

Si  $Y_1$  es perdido, pero por lo menos un valor  $Y_1^*$  de  $Y_1$  es observado, entonces  $S$  es un subconjunto propio de los valores posibles de  $Y_1$ .

### **B.1.3 Verosimilitud factorada para datos monótonos multinomiales.**

Se asume que,  $\{n_{jk\dots t}\}$  tiene una distribución multinomial con parámetros  $n$  y  $\theta = \{\pi_{jk\dots t}\}$ , además de esto asumimos también que el mecanismo de los datos perdidos es ignorable y el diseño es monótono.

$$L(\theta | \{n_{jk\dots t}\}) = \prod_{j,k,\dots,t} \pi_{jk\dots t}^{n_{jk\dots t}}, \quad \sum_{j,k,\dots,t} \pi_{jk\dots t} = 1$$

Para datos completos, la estimación de ML es:

$$\hat{\pi}_{jk\dots t} = \frac{n_{jk\dots t}}{n}$$

Para muestras grandes, se tiene la siguiente estimación de la varianza:

$$\text{Var}(\pi_{jk\dots t} - \hat{\pi}_{jk\dots t}) = \hat{\pi}_{jk\dots t}(1 - \hat{\pi}_{jk\dots t})/n.$$

Para la inferencia de Bayes tenemos:

$$P(\{\pi_{jk\dots t}\}) \propto \prod_{j,k,\dots,t} \pi_{jk\dots t}^{\alpha_{jk\dots t}-1}, \quad \pi_{jk\dots t} > 0, \quad \sum_{j,k,\dots,t} \pi_{jk\dots t} = 1 \text{ Distribución a priori.}$$

Mientras que la distribución a posteriori es:

$$P(\{\pi_{jk\dots t}\} | \{n_{jk\dots t}\}) \propto \prod_{j,k,\dots,t} \pi_{jk\dots t}^{\alpha_{jk\dots t} + n_{jk\dots t} - 1}, \quad \sum_{j,k,\dots,t} \pi_{jk\dots t} = 1$$

El objetivo es obtener respuestas análogas para datos perdidos.

### B.1.4 Estimación de máxima verosimilitud para diseño monótono.-

Consideramos la estimación ML, para una tabla de contingencia en dos vías con un margen suplementario en una vía, luego se extenderá a tablas multi-vía.

Tenemos dos variables categóricas:

$Y_1$  con  $J$  niveles, tal que  $j=1,2,\dots,J$

$Y_2$  con  $K$  niveles, tal que  $k=1,2,\dots,K$ .

Se tiene  $r$  observaciones  $\{(y_{i1}, y_{i2}), i=1,\dots,r\}$  con  $y_{i1}$ ,  $y_{i2}$  observado.

Se tiene  $m=n-r$  observaciones  $\{y_{i1}, i=r+1,\dots,n\}$  con  $y_{i1}$  observado y  $y_{i2}$  perdido (o no observado).

$$\theta = (\pi_{11}, \pi_{12}, \dots, \pi_{JK}) \quad \text{y} \quad \sum_{j=1}^J \sum_{k=1}^K \pi_{jk} = \pi_{++} = 1$$

La verosimilitud de los datos puede ser escrito como:

$$L(\Phi | \{r_{jk}, m_j\}) = \left( \prod_{j=1}^J \pi_{j+}^{r_{j+} + m_j} \right) \times \left( \prod_{j=1}^J \prod_{k=1}^K \pi_{k,j}^{r_{k,j}} \right)$$

El primer término es la verosimilitud para la distribución multinomial del margen  $r_{j+} + m_j$ , con parámetros  $(n, \pi_{j+})$ .

El segundo término es la verosimilitud para el producto de  $J$  distribuciones condicionales multinomiales de  $\{r_{jk}\}$  dado  $r_{j+}$  con parámetros  $r_{j+}$  y probabilidades:

$$\pi_{k,j} = P(Y_2 = k | Y_1 = j) = \pi_{jk} / \pi_{j+}, \quad k = 1, \dots, K$$

$$\phi_1 = \{\pi_{j+}, j=1, \dots, J\} \quad y \quad \phi_2 = \{\pi_{k,j}, j=1, \dots, J; k=1, \dots, K\}$$

Maximizando cada componente por separado, obtenemos los estimadores ML.

$$\hat{\pi}_{j+} = \frac{r_{j+} + m_j}{n} \quad y \quad \hat{\pi}_{k,j} = \frac{r_{jk}}{r_{j+}}$$

$$\hat{\pi}_{jk} = \hat{\pi}_{j+} \hat{\pi}_{k,j} = \frac{r_{jk} + (r_{jk}/r_{j+})m_j}{n}$$

El estimador ML distribuye efectivamente  $r_{jk}/r_{j+}$  de las observaciones no clasificadas  $m_j$  en la celda (j, k).

Para el análisis bayesiano, suponemos que especificamos las distribuciones de *Dirichlet* para  $\{\pi_{j+}\}$  y  $\{\pi_{k,j}\}$ .

$$P(\phi) \propto \left( \prod_{j=1}^J \pi_{j+}^{n_{j0}-1} \right) \times \left( \prod_{j=1}^J \prod_{k=1}^K \pi_{k,j}^{r_{jk0}-1} \right) \text{ DISTRIBUCIÓN A PRIORI.}$$

La distribución a posteriori es un producto de distribuciones independientes para  $\{\pi_{j+}\}$  y  $\{\pi_{k,j}\}$

$$P(\phi | \text{datos}) \propto \left( \prod_{j=1}^J \pi_{j+}^{n_{j0} + r_{j+} + m_j - 1} \right) \times \left( \prod_{j=1}^J \prod_{k=1}^K \pi_{k,j}^{r_{jk} + r_{jk0} - 1} \right) \text{ DISTRIBUCIÓN A POSTERIORI}$$

Una derivada  $(\pi_{jk}^{(d)})$ . de  $(\pi_{jk})$ . de la distribución a posteriori es fácilmente obtenida por la primera derivada  $(\pi_{j+}^{(d)})$ . y  $(\pi_{k,j}^{(d)})$  para sus distribuciones a posteriori en la ecuación anterior, y ajustando  $(\pi_{jk}^{(d)}) = \pi_{j+}^{(d)} \pi_{k,j}^{(d)}$ , análogo a la ecuación (13.5). Derivando de una función de Dirichletes fácilmente explicado usando variables aleatorias Ji-cuadrado o gamma.

### B.1.5 Precisión de la estimación.-

La matriz de covarianza asintótica asociada con los estimadores de ML, puede ser obtenida calculando la matriz de información para los parámetros en la forma facturada de la verosimilitud, invirtiendo esta matriz.

Se puede calcular estas varianzas y covarianzas directamente; así, tenemos.

$$\text{Var} \hat{\pi}_{jk} = E(\text{Var} \hat{\pi}_{jk} | \{n_{j+}\}) + \text{Var}(E(\hat{\pi}_{jk} | \{n_{j+}\})),$$

Donde  $\{n_{j+}\}$  es el conjunto de cantidades marginales de  $Y_1$ . Luego

$$\begin{aligned} \text{Var} \hat{\pi}_{jk} &= E\{\hat{\pi}_{j+} \pi_{k,j} (1 - \pi_{k,j}) / r_{j+}\} + \text{Var}\{\hat{\pi}_{j+} \pi_{k,j}\} \\ &= \pi_{j+}^2 \pi_{k,j} (1 - \pi_{k,j}) / r_{j+} + \pi_{k,j}^2 \pi_{j+} (1 - \pi_{j+}) / n, \end{aligned}$$

Asintóticamente al orden  $1/r_{j+}^2$ .

$$\text{Var} \hat{\pi}_{jk} \approx \frac{\pi_{jk} (1 - \pi_{jk})}{r} \left\{ 1 - \frac{(\pi_{k,j} - \pi_{jk})(n-r)}{(1 - \pi_{jk})n} + c_j \frac{1 - \pi_{k,j}}{1 - \pi_{jk}} \right\},$$

Dónde  $c_j = r\pi_{j+} / r_{j+} - 1$ . Substituyendo la estimación de los parámetros tenemos:

$$\text{Var}(\pi_{jk} - \hat{\pi}_{jk}) \approx \frac{\hat{\pi}_{jk} (1 - \hat{\pi}_{jk})}{r} \left\{ 1 - \frac{\hat{\pi}_{k,j} \hat{\pi}_{jk} (n-r)}{1 - \hat{\pi}_{jk} n} + c_j \frac{1 - \hat{\pi}_{k,j}}{1 - \hat{\pi}_{jk}} \right\}.$$

El análisis de Bayes de la varianza posterior asintótica de la varianza de  $\pi_{jk}$ , es un resultado similar. La covarianza estimamos:

$$\text{Cov}(\pi_{jk} - \hat{\pi}_{jk}, \pi_{jl} - \hat{\pi}_{jl}) \approx \frac{-\hat{\pi}_{jk} \hat{\pi}_{jl}}{r} \left\{ 1 + \frac{(1 - \hat{\pi}_{j+}) (n-r)}{\hat{\pi}_{j+} n} + \frac{c_j}{\hat{\pi}_{j+}} \right\}, \quad k \neq l$$

$$\text{Cov}(\pi_{ik} - \hat{\pi}_{ik}, \pi_{jl} - \hat{\pi}_{jl}) \approx \frac{-\hat{\pi}_{ik} \hat{\pi}_{jl}}{n}, \quad i \neq j$$

### B.1.6 Estimación de Bayes y verosimilitud para muestras multinomiales con diseño general de datos perdidos.

Para los datos incompletos multinomiales, que no forman un diseño monótono, requieren un procedimiento iterativo para estimar por ML, el paquete utilizado es R, el cual tiene un modulo con los algoritmos para estimar datos perdidos.