

ESCUELA POLITÉCNICA NACIONAL



FACULTAD DE CIENCIAS

CARRERA DE INGENIERÍA MATEMÁTICA

**ELABORACIÓN DE PERFILES DE INDIVIDUOS Y HOGARES DE
ACUERDO AL GRADO DE EXPOSICIÓN A LA VICTIMIZACIÓN.
UN ANÁLISIS PARA QUITO Y GUAYAQUIL.**

**PROYECTO PREVIO A LA OBTENCIÓN DEL
TÍTULO DE INGENIERO MATEMÁTICO**

LARA CAPA WASHINGTON RICARDO

Correo: wricardo_lc@hotmail.com

DIRECTOR: Dr. HOLGER CAPA SANTOS

Correo: holger.capa@epn.edu.ec

DECLARACIÓN:

Yo, Washington Ricardo Lara Capa, declaro bajo juramento que el trabajo aquí escrito es de mi autoría; que no ha sido previamente presentado para ningún grado o calificación profesional; y que he consultado las referencias bibliográficas que se incluyen en este documento.

A través de la presente declaración cedo mis derechos de propiedad intelectual correspondientes a este trabajo, a la Escuela politécnica Nacional, según lo establecido por la Ley de Propiedad Intelectual, por su reglamento y por la normatividad institucional vigente.

Washington Ricardo Lara Capa

CI: 171590514-5

CERTIFICACIÓN

Certifico que el presente trabajo fue desarrollado por el Sr. Washington Ricardo Lara Capa, bajo mi supervisión.

Dr. Holger Capa Santos

DIRECTOR DE PROYECTO

AGRADECIMIENTOS

A Dios, por bendecirme durante toda mi vida y, una vez más, al permitirme alcanzar esta meta.

A mis padres, por darme el amor y formación necesaria para alcanzar mis sueños.

Al Dr. Holger Capa Santos, pilar fundamental en el desarrollo de este proyecto, por su guía y apoyo permanente.

A la Escuela Politécnica Nacional, en especial a los docentes del Departamento de Matemática, por formarme de manera competitiva en conocimiento e integridad.

A todos quienes hicieron posible alcanzar este primer escalón en mi formación profesional.

CONTENIDO

<i>CAPÍTULO 1:</i> _____	<i>1</i>
<i>INTRODUCCIÓN</i> _____	<i>1</i>
<i>CAPÍTULO 2:</i> _____	<i>7</i>
<i>PRINCIPALES RESULTADOS SOBRE VICTIMIZACIÓN</i> ____	<i>7</i>
<i>A PERSONAS</i> _____	<i>7</i>
<i>2.1 ANTECEDENTES</i> _____	<i>7</i>
<i>2.2 DEFINICIONES</i> _____	<i>9</i>
<i>2.3 ANÁLISIS DESCRIPTIVO</i> _____	<i>10</i>
<i>2.4 ANÁLISIS PERCEPCIÓN</i> _____	<i>25</i>
<i>CAPITULO 3:</i> _____	<i>37</i>
<i>SCORE DE VICTIMIZACIÓN</i> _____	<i>37</i>
<i>3.1 SELECCIÓN DE LA MUESTRA</i> _____	<i>38</i>
<i>3.2 DEFINICIÓN DE BUENOS Y MALOS (VICTIMIZADOS / NO VICTIMIZADOS)</i> _____	<i>39</i>
<i>3.3 ANÁLISIS DE LA INFORMACIÓN</i> _____	<i>39</i>
<i>3.4 CONSTRUCCIÓN DEL MODELO</i> _____	<i>43</i>
<i>3.5 VALIDACIÓN</i> _____	<i>54</i>
<i>3.6 APLICACIÓN</i> _____	<i>56</i>

<i>CAPITULO 4</i> _____	82
<i>PERFILES DE VICTIMIZACIÓN</i> _____	82
ETAPAS DE DESARROLLO PARA UN ANÁLISIS DE	
AGRUPAMIENTO _____	83
AGRUPAMIENTO PARA QUITO _____	88
AGRUPAMIENTO PARA GUAYAQUIL _____	109
<i>CAPITULO 5:</i> _____	119
CONCLUSIONES Y RECOMENDACIONES _____	119
<i>BIBLIOGRAFÍA</i> _____	126
<i>ANEXO 1</i> _____	127
VARIABLES GENERADAS POR LOS ÁRBOLES _____	127
DE DECISIÓN _____	127
<i>ANEXO 2</i> _____	137
VARIABLES INICIALMENTE CONSIDERADAS EN EL	
DESARROLLO DE LA REGRESIÓN LOGÍSTICA _____	137
<i>ANEXO 3:</i> _____	155
DISTRIBUCIONES DE CONGLOMERADOS PARA	
GUAYAQUIL _____	155

CONTENIDO

(Cuadros)

CAPÍTULO 2

<i>Cuadro 2.1: Tamaños Muestrales</i>	<i>9</i>
<i>Cuadro 2.2: Tamaño Muestral por Provincia</i>	<i>9</i>
<i>Cuadro 2.3: Victimización por Región</i>	<i>10</i>
<i>Cuadro 2.4: Prueba Chi-Cuadrado</i>	<i>10</i>
<i>Cuadro: 2.5: Victimización por provincia</i>	<i>11</i>
<i>Cuadro 2.6: Prueba Chi-Cuadrado</i>	<i>13</i>
<i>Cuadro 2.7: Victimización por Rango de Edad</i>	<i>13</i>
<i>Cuadro 2.8: Prueba Chi-Cuadrado</i>	<i>14</i>
<i>Cuadro 2.9: Victimización por Género</i>	<i>15</i>
<i>Cuadro 2.10: Prueba Chi-Cuadrado</i>	<i>15</i>
<i>Cuadro 2.11: Victimización por Etnia</i>	<i>16</i>
<i>Cuadro 2.12: Victimización por Nivel de Educación</i>	<i>17</i>
<i>Cuadro 2.13: Prueba Chi-Cuadrado</i>	<i>18</i>
<i>Cuadro 2.14: Victimización por Estado Civil</i>	<i>18</i>
<i>Cuadro 2.15: Prueba Chi-Cuadrado</i>	<i>19</i>
<i>Cuadro 2.16: Victimización por Ocupación</i>	<i>20</i>
<i>Cuadro 2.17: Prueba Chi-Cuadrado</i>	<i>21</i>
<i>Cuadro 2.18: Porcentaje del Tipo de Victimización</i>	<i>21</i>
<i>Cuadro 2.19: Comparación entre Percepción y Victimización por Región</i>	<i>22</i>
<i>Cuadro 2.20: Comparación entre Percepción y Victimización por Provincia</i>	<i>24</i>
<i>Cuadro 2.21: Comparación entre Percepción y Victimización por Rango de Edad</i>	<i>25</i>
<i>Cuadro 2.22: Comparación entre Percepción y Victimización por Género</i>	<i>26</i>
<i>Cuadro 2.23: Comparación entre Percepción y Victimización por Estado Civil</i>	<i>27</i>
<i>Cuadro 2.24: Comparación entre Percepción y Victimización por Nivel de Educación</i>	<i>28</i>

Cuadro 2.25: Comparación entre Percepción y Victimización por Ocupación.....29

Cuadro 2.26: Victimización a Hogares y Personas por Provincia31

CAPÍTULO 3

Cuadro 3.1: DISTRIBUCIÓN SCORE Y DISTRIBUCIÓN EQUITATIVA49

Cuadro 3.2: ANÁLISIS EXPLORATORIO54

Cuadro 3.3: CHÍ-CUADRADO VARIABLES ANSWER TREE55

Cuadro 3.4: ÁRBOL GENERADO POR PROVINCIA58

Cuadro 3.5: ÁRBOL GENERADO POR NIVEL DE INSTRUCCIÓN59

Cuadro 3.6: ÁRBOL GENERADO POR INGRESO PERCÁPITA59

Cuadro 3.7: ÁRBOL GENERADO POR PARENTEZCO CON EL JEFE DE HOGAR.....60

*Cuadro 3.8: ÁRBOL GENERADO POR PORCENTAJE DE PERSONAS QUE TRABAJA POR CANTON
.....60*

Cuadro 3.9: ÁRBOL GENERADO POR EDAD60

Cuadro 3.10: ÁRBOL GENERADO POR INSTRUCCIÓN Y EDAD60

Cuadro 3.11: ÁRBOL GENERADO POR ÁREA60

Cuadro 3.12: CHI-CUADRADO VARIABLES ANSWER TREE60

Cuadro 3.13: VARIABLES SELECCIONADAS EN EL MODELO61

Cuadro 3.14: RESUMEN DEL MODELO62

Cuadro 3.15: PRUEBA OMNIBUS DE LOS COEFICIENTES64

Cuadro 3.16: RESUMEN VARIABLES SELECCIONADAS64

Cuadro 3.17: DISTRIBUCIÓN DE BUENOS Y MALOS65

Cuadro 3.18: PRUEBA DE HOSMER LEMESHOW67

Cuadro 3.19: TABLA DE CONTINGENCIA DE HOSMER LEMESHOW68

Cuadro 3.20: PRUEBA DE KOLMOGOROV SMIRNOV69

Cuadro 3.21: DISTRIBUCIÓN PARA CÁLCULO DE GINI69

Cuadro 3.22: ANÁLISIS ESTADÍSTICO VARIABLES EN EL MODELO70

Cuadro 3.23: ANÁLISIS DE DISCRIMINACIÓN K-S Y ODDS72

Cuadro 3.24: PRUEBA DE KOLMOGOROV - SMIRNOV73

<i>Cuadro 3.25: DISTRIBUCIÓN PARA CÁLCULO DE GINI</i>	74
<i>Cuadro 3.26: DISTRIBUCIÓN SCORE POR QUINTILES</i>	75
<i>Cuadro 3.27: PERFILES DEFINIDOS POR PUNTAJE DE SCORE</i>	75
<i>Cuadro 3.28: FRECUENCIAS VARIABLES POR GRUPO PERFIL</i>	76
<i>Cuadro 3.29: PORCENTAJE DE INCIDENCIA DE VARIABLES POR GRUPO PERFIL</i>	76
<i>Cuadro 3.30: DESCRIPCIÓN DE GRUPO PERFIL</i>	76

CAPÍTULO 4

<i>Cuadro 4.1: VARIABLES CONSIDERADAS</i>	81
<i>Cuadro 4.2: CRITERIO BAYESIANO DE SCHWARZ</i>	85
<i>Cuadro 4.3: CARACTERÍSTICAS DE GRUPOS DEFINIDOS</i>	91
<i>Cuadro 4.4: VICTIMIZACIÓN POR GRUPO</i>	92
<i>Cuadro 4.5: PRUEBA DE HOMOGEIDAD DE VARIANZAS PARA LA VICTIMIZACIÓN A PERSONAS</i>	93
<i>Cuadro 4.6: PRUEBA DE IGUALDAD DE MEDIAS PARA LA VICTIMIZACIÓN A PERSONAS</i>	94
<i>Cuadro 4.7: COMPARACIONES MÚLTIPLES PARA LA VICTIMIZACIÓN A PERSONAS</i>	94
<i>Cuadro 4.8: PERCEPCIÓN DE VICTIMIZACIÓN POR GRUPO</i>	96
<i>Cuadro 4.9: PRUEBA DE HOMOGENEIDAD DE LAS VARIANZAS PARA LA PERCEPCIÓN</i>	96
<i>Cuadro 4.10: PRUEBA DE IGUALDAD DE MEDIAS PARA LA PERCEPCIÓN</i>	96
<i>Cuadro 4.11: COMPARACIONES MÚLTIPLES PARA LA PERCEPCIÓN</i>	97
<i>Cuadro 4.12: ÍNDICE DE VICTIMIZACIÓN A HOGARES</i>	99
<i>Cuadro 4.13: GRUPOS BARRIOS DEFINIDOS POR ELÁRBOL DE DECISIÓN</i>	101
<i>Cuadro 4.14: VICTIMIZACIÓN A HOGARES POR GRUPO</i>	102
<i>Cuadro 4.15: VICTIMIZACIÓN A PERSONAS POR GRUPO</i>	102
<i>Cuadro 4.16: VICTIMIZACIÓN A HOGARES Y PERSONAS</i>	103
<i>Cuadro 4.17: VICTIMIZACIÓN GRUPOS BARRIOS Y PERFILES SOCIALES</i>	104
<i>Cuadro 4.18: REPRESENTATIVIDAD GRUPOS BARRIOS Y PERFILES SOCIALES</i>	105
<i>Cuadro 4.19: INCIDENCIA DE VICTIMIZACIÓN EN GRUPOS BARRIOS Y PERFILES SOCIALES</i>	105

<i>Cuadro 4.20: ÍNDICE DE VICTIMIZACIÓN A PERSONAS</i>	<i>106</i>
<i>Cuadro 4.21: ÍNDICE DE PERCEPCIÓN DE VICTIMIZACIÓN</i>	<i>106</i>
<i>Cuadro 4.22: ÍNDICE DE VICTIMIZACIÓN A HOGARES</i>	<i>107</i>
<i>Cuadro 4.23: CARACTERÍSTICAS DE GRUPOS DEFINIDOS</i>	<i>107</i>
<i>Cuadro 4.24: VICTIMIZACIÓN A PERSONAS POR GRUPO.....</i>	<i>108</i>
<i>Cuadro 4.25: PERCEPCIÓN DE VICTIMIZACIÓN POR GRUPO</i>	<i>108</i>
<i>Cuadro 4.26: PRUEBA DE IGUALDAD DE MEDIAS PARA LA VICTIMIZACIÓN</i>	<i>108</i>
<i>Cuadro 4.27: COMPARACIONES MÚLTIPLES PARA LA VICTIMIZACIÓN A PERSONAS.....</i>	<i>109</i>
<i>Cuadro 4.28: COMPARACIONES MÚLTIPLES PARA LA PERCEPCIÓN DE LAVICTIMIZACIÓN</i>	<i>110</i>
<i>Cuadro 4.29: GRUPOS BARRIOS DEFINIDOS POR EL ÁRBOL.....</i>	<i>112</i>
<i>Cuadro 4.30: GRUPOS DE BARRIOS EN GUAYAQUIL.....</i>	<i>113</i>
<i>Cuadro 4.31: VICTIMIZACIÓN A HOGARES POR GRUPO</i>	<i>113</i>
<i>Cuadro 4.32: PERCEPCIÓN DE VICTIMIZACIÓN POR GRUPO</i>	<i>113</i>
<i>Cuadro 4.33: VICTIMIZACIÓN POR GRUPOS BARRIOS Y PERFILES SOCIALES</i>	<i>114</i>
<i>Cuadro 4.34: SEGMENTOS CON MAYOR VICTIMIZACIÓN</i>	<i>114</i>
<i>Cuadro 4.35: REPRESENTATIVIDAD GRUPOS BARRIOS Y PERFILES SOCIALES</i>	<i>115</i>

CONTENIDO

(GRÁFICOS)

CAPÍTULO 2

<i>Gráfico 2.1: Nivel de Victimización por región</i>	<i>10</i>
<i>Gráfico 2.2: Distribución de la Población por Provincia</i>	<i>12</i>
<i>Gráfico 2.3: Victimización por Provincia</i>	<i>12</i>
<i>Gráfico 2.4: Nivel de Victimización por Rango de Edad</i>	<i>14</i>
<i>Gráfico 2.5: Nivel de Victimización por Género</i>	<i>15</i>
<i>Gráfico 2.6: Nivel de Victimización por Etnia</i>	<i>16</i>
<i>Gráfico 2.7: Nivel de Victimización por Nivel de Educación</i>	<i>17</i>
<i>Gráfico 2.8: Nivel de Victimización por Estado Civil</i>	<i>19</i>
<i>Gráfico 2.9: Nivel de Victimización por Ocupación.....</i>	<i>20</i>
<i>Gráfico 2.10: Porcentaje del Tipo de Victimización</i>	<i>22</i>
<i>Gráfico 2.11: Comparación entre Percepción y Victimización por Región</i>	<i>23</i>
<i>Gráfico 2.12: Nivel Percepción y Victimización por Provincia</i>	<i>25</i>
<i>Gráfico 2.13: Comparación entre Percepción y Victimización por Rango de Edad</i>	<i>26</i>
<i>Gráfico 2.14: Nivel de Percepción y Victimización por Género</i>	<i>27</i>
<i>Gráfico 2.15: Nivel de Percepción y Victimización por Estado Civil.....</i>	<i>28</i>
<i>Gráfico 2.16: Nivel de Percepción y Victimización por Nivel de Educación</i>	<i>29</i>
<i>Gráfico 2.17: Comparación entre Percepción y Victimización por Ocupación</i>	<i>30</i>
<i>Gráfico 2.18: Nivel de Victimización a Hogares y Personas por Provincia</i>	<i>32</i>

CAPÍTULO 3

<i>Gráfico 3.1: FUNCIÓN DE DISTRIBUCIÓN DE BUENOS Y MALOS</i>	<i>48</i>
<i>Gráfico 3.2: CURVA DE LORENZ</i>	<i>49</i>
<i>Gráfico 3.3: ÁREAS CURVA DE LORENZ</i>	<i>50</i>

<i>Gráfico 3.4: ÁRBOL DE DECISIÓN POR PROVINCIA</i>	57
<i>Gráfico 3.5: DISTRIBUCIÓN BUENOS Y MALOS</i>	66
<i>Gráfico 3.6: DISTRIBUCIÓN K-S</i>	68
<i>Gráfico 3.7: ODDS - DISTRIBUCIÓN DE BUENOS Y MALOS</i>	72

CAPÍTULO 4

<i>Gráfico 4.1: DISTRIBUCIÓN PORCENTUAL DE GRUPO OCUPACIÓN POR CONGLOMERADO</i>	86
<i>Gráfico 4.2: DISTRIBUCIÓN PORCENTUAL DEL RANGO DE EDAD POR CONGLOMERADO</i>	86
<i>Gráfico 4.3: DISTRIBUCIÓN PORCENTUAL DEL GENERO POR CONGLOMERADO</i>	87
<i>Gráfico 4.4: DISTRIBUCIÓN PORCENTUAL DEL ESTADO CIVIL POR CONGLOMERADO</i>	87
<i>Gráfico 4.5: DISTRIBUCIÓN PORCENTUAL DEL NIVEL DE INSTRUCCIÓN POR CONGLOMERADO</i>	88
<i>Gráfico 4.6: PRUEBA CHÍ - CUADRADO GRUPO OCUPACIÓN</i>	89
<i>Gráfico 4.7: PRUEBA CHÍ - CUADRADO RANGO DE EDAD</i>	89
<i>Gráfico 4.8: PRUEBA CHI - CUADRADO GÉNERO</i>	90
<i>Gráfico 4.9: PRUEBA CHÍ - CUADRADO ESTADO CIVIL</i>	90
<i>Gráfico 4.10: PRUEBA CHI-CUADRADO NIVEL DE INSTRUCCIÓN</i>	91
<i>Gráfico 4.11: VICTIMIZACIÓN A PERSONAS POR GRUPO</i>	95
<i>Gráfico 4.12: PERCEPCIÓN POR GRUPO</i>	98
<i>Gráfico 4.13: VICTIMIZACIÓN Y PERCEPCIÓN</i>	98
<i>Gráfico 4.14: ÁRBOL GENERADO POR BARRIOS</i>	100
<i>Gráfico 4.15: VICTIMIZACIÓN A PERSONAS POR GRUPO</i>	103
<i>Gráfico 4.16: VICTIMIZACIÓN A HOGARES Y PERSONAS</i>	104
<i>Gráfico 4.17: VICTIMIZACIÓN A PERSONAS POR GRUPO</i>	107
<i>Gráfico 4.18: PERCEPCIÓN POR GRUPO</i>	110
<i>Gráfico 4.19: PERCEPCIÓN Y VICTIMIZACIÓN</i>	111

RESUMEN

El objetivo de este trabajo consiste en desarrollar modelos que cuantifiquen el riesgo de victimización a personas, tanto a nivel nacional como para las dos ciudades más grandes del Ecuador, Quito y Guayaquil, con el propósito de encontrar segmentos de alto riesgo sobre el cual aplicar posibles estrategias que permitan disminuir el riesgo a ser victimizados.

A nivel nacional se ha creado un indicador que cuantifique el grado de exposición de los individuos a ser victimizados, analizando los factores de riesgo en forma multivariada. Este indicador se construye mediante un modelo de regresión logística, en el cual las variables exógenas relevantes se han elegido con un proceso de árboles de decisión, obteniendo de esta manera un modelo compuesto, desarrollado en la lógica de un modelo de *score* generalmente utilizado en el campo financiero. En el desarrollo del modelo se presenta la descripción conceptual y metodológica de construcción de un modelo de *score* así como las pruebas estadísticas necesarias para validar su funcionamiento en la etapa de desarrollo y de prueba.

Para las ciudades más grandes, Quito y Guayaquil, debido a su representatividad a nivel nacional, se han creado grupos de acuerdo a sus similitudes que permitan enfocar estrategias de acción diferenciadas que minimicen la exposición a la victimización y por ende la victimización misma. Para esto, se desarrolla un modelo de conglomerados que describa el nivel de victimización de cada grupo. Este segundo análisis, al ser más específico para cada una de las ciudades, considera el estudio de la victimización a hogares y su incidencia en la victimización a personas.

CAPÍTULO 1:

INTRODUCCIÓN

Uno de los problemas más graves que actualmente vivimos en nuestro país es la inseguridad. Los medios de comunicación muestran constantemente delitos que ocurren en todo el país; los ciudadanos manifiestan reclamos frecuentes por la falta de seguridad en que se desenvuelven.

Un primer paso para tratar este terrible mal es conocer el significado de victimización, la incidencia de ésta en el país y en cada uno de los lugares que lo conforman.

El concepto de delito es diferente para los individuos. Éste tiene dos aspectos: el natural que considera al delito como una acción sin moral que afecta la vida de otro miembro de la sociedad directa o indirectamente; por otro lado, existen los delitos jurídicos que son resultado de la violación de una ley jurídica que por tanto tiene un marco legal. Estas dos definiciones muchas veces no coinciden debido al desequilibrio que existe entre la evolución de la ética y el derecho. Esta es una valiosa razón por la cual la victimización y su percepción desde el punto de la sociedad son motivos de estudio.

Otro punto a tratar en el estudio de este problema es la diferencia que existe entre la percepción de la victimización y la realidad de la misma. Entre lo sucedido denunciado y lo no denunciado; nunca será posible conocer la cifra real pero es posible tener cifras que se aproximen a través de diferente información que no necesariamente es de denuncia. La denuncia es solo una parte de la criminalidad real ya que existen eventos que no son denunciados por temor, falta de confianza en la gestión del organismo a cargo o porque simplemente la denuncia toma un tiempo y trabajo que los ciudadanos no pueden o quieren disponer. Sea cual sea

la causa es de vital importancia conocer las cifras que más cerca estén de la realidad ya que estas indican el marco de gobernabilidad que la sociedad tiene sobre sí misma. La forma de obtenerlas son las encuestas a la población cuyas edades mínimas varían entre 14 y 16 años; estas deberían indicar el comportamiento delictivo producido en la vida social que es declarado por las personas víctimas y sobre el conocimiento de víctimas en su grupo familiar. Este proceso pretende dar una medida empírica del delito en la sociedad.

La diferencia que existe entre la cantidad de denuncias policiales o judiciales y la cantidad de denuncias obtenidas en una encuesta, es llamada la “cifra negra de la criminalidad” y debe servir para conocer la diferencia de lo denunciado y lo real para posteriores estimaciones de victimización a partir de la información denunciada. Además, la información que es obtenida en las encuestas debe servir para validar la información obtenida en las denuncias observando que las tendencias sean similares pero con diferente escala. Otro punto que también es importante mencionar es que la información denunciada y la obtenida mediante una encuesta generalmente son diferentes ya que la denunciada es referente a hechos delictivos graves como crímenes, violaciones, robos de automotores, etc.; por otro lado, las encuestas contienen también información de robos, hurtos, arranches, etc., mostrando así una criminalidad diferente en las dos fuentes de información.

Es importante notar que aunque la información denunciada tenga deficiencia comparada con la obtenida en una encuesta, es de vital importancia como una medida, siempre y cuando se mantengan presentes sus limitaciones.

Adicionalmente, la encuesta es una metodología para el conocimiento del delito relativamente nueva, usada desde la década de los 90; aún está en etapa de perfeccionamiento en sus técnicas de recolección y de análisis de resultados. La información utilizada para el presente trabajo es la primera de este tipo a nivel nacional.

Como toda encuesta a la población, las de victimización tienen períodos de referencia; es decir, es una investigación de corte transversal en el cual deben haber ocurrido los hechos de victimización sobre los cuales se pregunta. Los resultados pueden ser utilizados para medir la extensión de la criminalidad, de la misma manera que el volumen comparativo de la denuncia referentes a un determinado tiempo.

Aunque una encuesta sea la manera más adecuada de acercarnos a la cifra real de la criminalidad, tiene sus limitaciones. Los delitos sexuales, por ejemplo, al ser investigados mediante operativos de campo pueden ser motivo de revictimización de los declarantes; penetrar en su intimidad sin la debida preparación del investigador y del investigado, no resulta en cifras consistentes y más aún no es éticamente correcto sin un protocolo de aceptación y un abordaje que permita restablecer a la víctima del *shock* que causa el recuerdo de su victimización. Sesgos similares y otros diferentes pueden ocurrir al investigar el maltrato infantil.

Percepción y Victimización

La población tiene una percepción de victimización que generalmente es distinta de la real, ya que la percepción está fundamentada en sucesos imaginarios ó sobredimensionados que alteran la verdadera naturalidad del delito. Esta sobredimensionada percepción se debe en gran medida a los medios de comunicación, que inducen al ciudadano a un combate a la delincuencia. El concepto de inseguridad se ha alimentado por sucesos exagerados o inusuales suscitados especialmente en las ciudades de Quito y Guayaquil. Es muy peligroso construir el miedo social que a falta de presencia fuerte del Estado en unos casos y de una justicia penal ágil en otros, ponga en manos de la población la justicia, el castigo y la ejecución de penas: la llamada justicia con las propias manos.

OBJETIVOS

Si bien manejar una cifra de victimización cercana a la realidad permite medir el nivel de delincuencia en el Ecuador y de posibles estrategias para combatirla, existen herramientas matemáticas que pueden generar decisiones focalizadas que ayuden a mitigar este mal de manera más efectiva. En este proyecto de titulación se utilizarán modelos estadísticos que permitan estudiar más profundamente este tema de gran importancia social, con la finalidad de poder definir estrategias de acción que minimicen la victimización en la medida de lo posible, tomando información de la encuesta Victimización en el Ecuador, contratada por la Unidad de Ejecución Especializada en Seguridad Pública (UEE) del Ministerio de Gobierno.

En este sentido, el objetivo general de este trabajo consiste en crear perfiles para personas, mediante la creación de un indicador que cuantifique el grado de exposición que los individuos tienen a ser victimizados y la integración de de grupos de acuerdo a sus similitudes, que permitan enfocar estrategias de acción diferenciadas para minimizar la exposición a la victimización y por ende la victimización misma. Se realiza un análisis nacional, así como también para Quito y Guayaquil. En el análisis nacional se desarrolla un modelo de *score* que cuantifique el riesgo que un individuo tiene de ser victimizado. Para las ciudades de Quito y Guayaquil se desarrolla un modelo de conglomerados que describe el perfil social de cada grupo permitiendo estudiar la victimización de manera diferenciada. Por tanto los objetivos específicos se plantean de la siguiente forma:

- I. Desarrollar un modelo de *Score* que permita cuantificar el grado de exposición que cada individuo tiene para ser victimizado.
- II. Validar el modelo de *Score* obtenido mediante indicadores estadísticos de discriminación del modelo, a través de dos muestras: una de desarrollo y otra de validación.
- III. Desarrollar un modelo de agrupamiento que permita clasificar a los individuos de una población dada y en función de ella estudiar la

victimización a personas conjuntamente con la victimización de hogares. Este estudio se realiza separadamente para Quito y Guayaquil.

Se consideran las siguientes hipótesis:

- I. El nivel de victimización nacional (y de otros indicadores) es diferente al de las provincias.
- II. Las características sobre la victimización entre las dos principales ciudades del Ecuador, Quito y Guayaquil, son diferentes.
- III. Es posible identificar perfiles de individuos con diferentes niveles de victimización mediante el uso de regresión logística.
- IV. Es posible generar segmentos de individuos en función de sus semejanzas y que tengan diferentes niveles de victimización.

Para alcanzar los objetivos planteados, en el segundo capítulo se detallan los principales resultados descriptivos obtenidos en la encuesta a nivel nacional, con un buen nivel de desagregación, como por ejemplo, por género, por edad, etc., así como ciertos análisis comparativos.

En el tercer capítulo se presenta la aplicación del modelo estadístico de regresión logística que intenta explicar el uso del mismo para la creación de un *score* de victimización a nivel nacional, que cuantifique el riesgo de que un ecuatoriano sea victimizado. Este puede mostrar un resultado en cierto sentido muy general, ocasionado por el nivel de agregación, que en este caso es a nivel país, pero que sin embargo como se verá es una adecuada herramienta para la discriminación de la victimización, brindando la posibilidad de priorizar y especializar la gestión en los individuos con mayor riesgo. La metodología aplicada podría implementarse para el nivel de detalle deseado, como puede ser por ciudad, cuyos resultados serán más precisos y útiles para cada comunidad.

En el capítulo cuatro se presenta un análisis más específico. Se han seleccionado las ciudades de Quito y Guayaquil para el estudio por ser de las más victimizadas y las más grandes del Ecuador. Se realiza un modelo de agrupación para los individuos de cada ciudad con el fin de determinar los diferentes niveles de victimización que cada grupo tiene y las posibles estrategias que se podrían aplicar a cada uno de ellos, considerando la tipología de cada grupo.

Para terminar, en el capítulo cinco se presentan las conclusiones y recomendaciones obtenidas a lo largo del desarrollo de este estudio.

CAPÍTULO 2:

PRINCIPALES RESULTADOS SOBRE VICTIMIZACIÓN

A PERSONAS

2.1 ANTECEDENTES

La metodología sobre la cual se obtuvo la información de la encuesta es mediante entrevista directa. Este método aunque supone costos operativos elevados tiene las ventajas de obtener información con una alta confiabilidad y cobertura, superando a métodos como las encuestas telefónicas y los correos electrónicos, que demandan la conciencia de la importancia de estos estudios para una declaración verídica por parte de los entrevistados, siendo su nivel de respuesta bastante bajo.

El muestreo fue aplicado sobre las viviendas, una opción mejor que la de los lugares aglomerados que no poseen un diseño muestral. Las viviendas, al tener un lugar específico, permiten tener un diseño muestral sobre el cuál se puede usar factores de expansión.

Población Objetivo

La población objetivo de estudio concierne a las personas mayores a 16 años, de las capitales provinciales de todas las provincias del país, con excepción de Galápagos.

La encuesta se la realizó por muestreo y la entrevista a un informante que tenga una edad mínima de 18 años (en la medida de lo posible se intentó entrevistar al jefe de hogar).

El diseño de la muestra es multietápico. A continuación se describen las fases correspondientes¹:

Las unidades primarias de muestreo son los estratos, que corresponden a cada uno de los cantones al que pertenece cada capital provincial. Esta estratificación permitirá calcular indicadores nacionales o regionales, de ser necesario.

En cada una de las capitales provinciales se estratificó la población en urbana y rural (sectores urbanos de las parroquias rurales; sectores amanzanados, no los dispersos, por cuanto los costos de recolección se elevan considerablemente). Se realizó una asignación proporcional de la muestra para cada estrato. Para evitar una exagerada dispersión geográfica en el área rural, previamente se seleccionaron parroquias rurales, con el criterio de que en cada una de ellas se investigarán, en general, 2 sectores censales, pero no más de 3.

En cada estrato se seleccionaron sectores censales con igual probabilidad.

En la cuarta fase se seleccionaron con probabilidad igual 12 viviendas en cada sector censal (también se seleccionarán 4 adicionales como posibles remplazos). En los casos en que exista más de un hogar en la vivienda (situación excepcional) se sorteó un hogar para la entrevista.

Finalmente, en la quinta fase, se entrevistó en cada hogar seleccionado a todos los miembros con al menos 16 años de edad. No se entrevistó a todos pero se recobró información sobre todos.

¹Mayores detalles se encuentran en el Informe Final sobre Estudio de Percepción y Victimización en el Ecuador-Diseño Metodológico, 2008.

Cuestionario de personas

Las preguntas se organizaron en capítulos consecutivos, que corresponden a los temas de investigación de la encuesta. La información es referente a²:

- INFORMACIÓN SOBRE PERSONAS
- DATOS DEMOGRÁFICOS DE LOS MIEMBROS DEL HOGAR
- DATOS EDUCATIVOS
- DATOS ÉTNICOS-CULTURALES
- MIGRACIÓN
- DATOS ECONÓMICOS
- DATOS SOBRE ROBOS Y ASALTOS
- DATOS SOBRE INSEGURIDAD
- DATOS SOBRE MALTRATO
- VIOLENCIA INTRAFAMILIAR
- PERCEPCIÓN DE INSEGURIDAD
- ACOSO Y ABUSO SEXUAL

2.2 DEFINICIONES

El indicador de victimización es obtenido mediante el análisis de varias variables que intentan comprender el fenómeno de la inseguridad. El concepto de victimización a personas, victimización a hogares y percepción de victimización se especifica en las siguientes definiciones:

Índice de Victimización a Personas (IVP): Indica el porcentaje de personas (16 años de edad en adelante) que fueron víctimas en el transcurso de los últimos 12 meses de al menos uno de los siguientes delitos: robo con fuerza, robo sin fuerza, ataques y amenazas.

² La base dispone de más información que para este análisis no es objeto de estudio.

Índice de Percepción de Inseguridad (IPI): Indica el porcentaje de personas que han manifestado cualquiera de las siguientes opciones:

- En las últimas dos semanas ha conversado con sus familiares, amigos o compañeros sobre delincuencia y seguridad.
- Comparado con un año atrás, considera que en su barrio el problema de la inseguridad frente al delito es más grave.
- Comparado con un año atrás, diría que su barrio es hoy menos seguro.
- Considera que hoy en la Ciudad la inseguridad respecto al delito es un problema bastante importante o muy importante.
- Se sienten nada o poco seguro al caminar en su barrio en la noche.
- Se sienten nada o poco seguro cuando está solo/a en su casa.
- Se sienten nada o poco seguro cuando está solo/a en su auto.

Índice de Victimización a Hogares (IVH): Indica el porcentaje de hogares que fueron víctimas en el transcurso de los últimos 12 meses de al menos uno de los siguientes delitos: robo a viviendas y robo de vehículos, motocicletas, bicicletas u otros accesorios.

2.3 ANÁLISIS DESCRIPTIVO

El objetivo del presente capítulo es mostrar los índices de victimización a personas y su percepción a nivel nacional en sus diferentes segmentaciones.

Se presentará un análisis descriptivo de las variables más importantes de la población. Para esto se ha utilizado un ponderador que fue utilizado en el estudio y que corresponden según el caso a la representatividad que tiene cada capital provincial en los indicadores nacionales. Para este análisis se han excluido a los menores de 16 años.

Tamaño de la muestra y asignación por provincias

El tamaño de la muestra fue obtenido considerando una precisión del 6,20 % y un nivel de confianza del 95% para variables multinomiales (respuestas múltiples) en cada capital provincial, resultando en una precisión a nivel nacional menor a 1,5%. En Guayas y pichincha la precisión fue del 4,20%. A continuación se muestra la distribución de los tamaños de muestra:

Cuadro 2.1: Tamaños Muestrales

Tamaños Muestrales			
Cantón capital	Precisión	No. Hogares	No. Sect. Cens.
Guayaquil	4,20%	732	61
Quito	4,20%	732	61
Zamora	6,20%	312	26
Otro cantón	6,20%	336	28
Total nacional		8.496	708

Considerando que los tamaños de muestra, pese a ser representativos para cada capital provincial, no se obtuvieron con asignación proporcional, se aplicó un ponderador de acuerdo al tamaño poblacional³:

³ Mayores detalles en el cálculo de la muestra se encuentran en el Informe Final sobre Estudio de Percepción y Victimización en el Ecuador-Diseño Metodológico, 2008.

Cuadro 2.2: Tamaño Muestral por Provincia

PROVINCIA	Frecuencia Personas	Ponderador	Población Estimada
AZUAY	944	228	215.125
BOLIVAR	864	20	17.376
CAÑAR	936	26	24.557
CARCHI	945	42	39.600
COTOPAXI	802	52	41.581
CHIMBORAZO	928	110	101.834
EL ORO	1.016	138	140.708
ESMERALDAS	962	73	70.650
GUAYAS	2.264	667	1.509.778
IMBABURA	953	91	86.976
LOJA	999	91	91.152
LOS RIOS	932	77	72.128
MANABI	1.013	132	134.188
MORONA SANTIAGO	840	13	11.024
NAPO	904	14	12.561
PASTAZA	933	19	17.509
PICHINCHA	2.075	604	1.252.677
TUNGURAHUA	930	143	132.946
ZAMORA	907	9	8.115
SUCUMBIOS	816	30	24.808
ORELLANA	871	14	11.990
SANTO DOMINGO DE LOS TSACHILAS	778	220	171.312
SANTA ELENA	1.029	22	22.352
Total	23.641	2.836	4.210.947

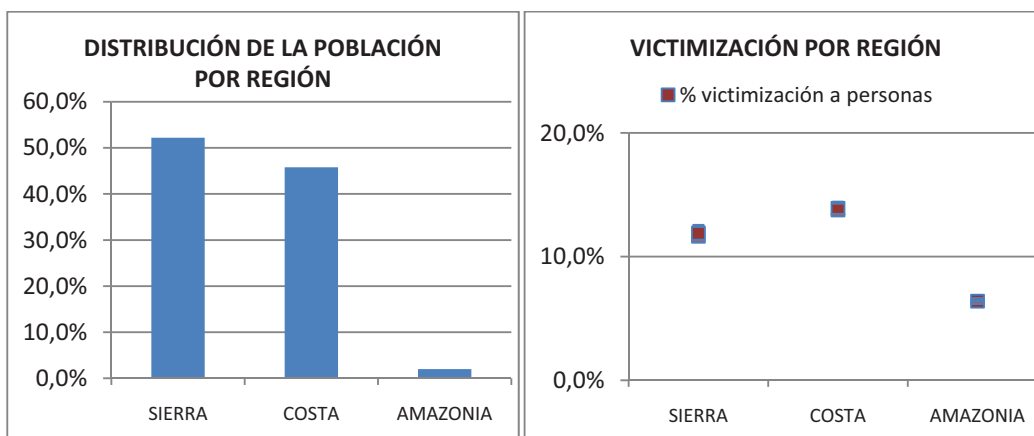
VICTIMIZACIÓN POR REGIÓN

Cuadro 2.3: Victimización por Región

Victimización por Región						
Región	Tamaño de muestra	Frecuencia población estimada	% población estimada	% victimización a personas	Victimización Lím.inf. (95%)	Victimización Lím.sup. (95%)
SIERRA	12.183	2.197.488	52,2%	11,9%	11,2%	12,5%
COSTA	6.187	1.927.452	45,8%	13,9%	13,4%	14,4%
AMAZONIA	5.271	86.007	2,0%	6,4%	6,3%	6,6%
Total	23.641	4.210.948	100,0%	12,7%	11,8%	13,5%

La mayor parte de la población⁴ se encuentra concentrada en la Costa y en la Sierra.

Gráfico 2.1: Nivel de Victimización por región



La Costa presenta la mayor victimización mientras que la Amazonía tiene el menor porcentaje de victimización. Se verifica que estadísticamente exista una diferencia en la distribución de la victimización de acuerdo a la región. Para esto se contrasta la hipótesis nula de independencia entre la victimización y la región⁵.

$$\begin{cases} H_0: \text{La victimización es independiente de la región} \\ H_1: \text{La victimización no es independiente de la región} \end{cases}$$

Cuadro 2.4: Prueba Chi-Cuadrado

Prueba Chi-Cuadrado			
	Valor	df	Asy mp. Sig. (2-sided)
Pearson Chi-Square	6798,618 ^a	2	,000
Likelihood Ratio	7339,933	2	,000
Linear-by-Linear Association	1114,330	1	,000
N of Valid Cases	4210947		

^a. No existen frecuencias esperadas menores a 5. La mínima frecuencia esperada es 10890,14.

⁴ Se considera como Población a la muestra expandida por el ponderador.

⁵ Para las posteriores variables se omitirá la formulación de las pruebas de hipótesis mostrando directamente el p-valor del estadístico chi-cuadrado.

La victimización muestra que existe relación entre la victimización y la Región.

VICTIMIZACIÓN POR PROVINCIA

Cuadro: 2.5: Victimización por provincia

VICTIMIZACIÓN POR PROVINCIA						
Provincia	Tamaño de muestra	Frecuencia población estimada	% población estimada	% victimización a personas	Victimización Lím.inf. (95%)	Victimización Lím.sup. (95%)
AZUAY	944	215.125	5,1%	5,6%	4,9%	6,3%
BOLIVAR	864	17.376	0,4%	2,2%	1,7%	2,7%
CAÑAR	936	24.557	0,6%	3,3%	2,7%	3,9%
CARCHI	945	39.600	0,9%	7,3%	6,5%	8,1%
COTOPAXI	802	41.580	1,0%	16,7%	15,4%	18,0%
CHIMBORAZO	928	101.835	2,4%	11,1%	10,1%	12,1%
EL ORO	1.016	140.708	3,3%	11,5%	10,6%	12,5%
ESMERALDAS	962	70.650	1,7%	8,5%	7,7%	9,4%
GUAYAS	2.264	1.509.778	35,9%	14,7%	14,0%	15,4%
IMBABURA	953	86.976	2,1%	5,9%	5,1%	6,6%
LOJA	999	91.151	2,2%	11,7%	10,7%	12,7%
LOS RIOS	932	72.127	1,7%	19,1%	17,9%	20,3%
MANABI	1.013	134.189	3,2%	6,7%	6,0%	7,5%
MORONA SANTIAGO	840	11.024	0,3%	7,7%	6,9%	8,6%
NAPO	904	12.561	0,3%	6,6%	5,8%	7,4%
PASTAZA	933	17.509	0,4%	2,1%	1,7%	2,6%
PICHINCHA	2.075	1.252.677	29,7%	14,7%	14,0%	15,5%
TUNGURAHUA	930	132.946	3,2%	9,0%	8,1%	9,9%
ZAMORA	907	8.115	0,2%	4,6%	4,0%	5,3%
SUCUMBIOS	816	24.808	0,6%	8,8%	7,9%	9,8%
ORELLANA	871	11.990	0,3%	7,5%	6,6%	8,3%
STO. DOM. TSACHILAS	778	171.313	4,1%	6,9%	6,1%	7,8%
SANTA ELENA	1.029	22.351	0,5%	7,9%	7,1%	8,7%
Total	23.641	4.210.946	100,0%	12,7%	11,8%	13,5%

Las provincias de Guayas y Pichincha son las más pobladas, representan el 65,6% de toda la población.

Gráfico 2.2: Distribución de la Población por Provincia

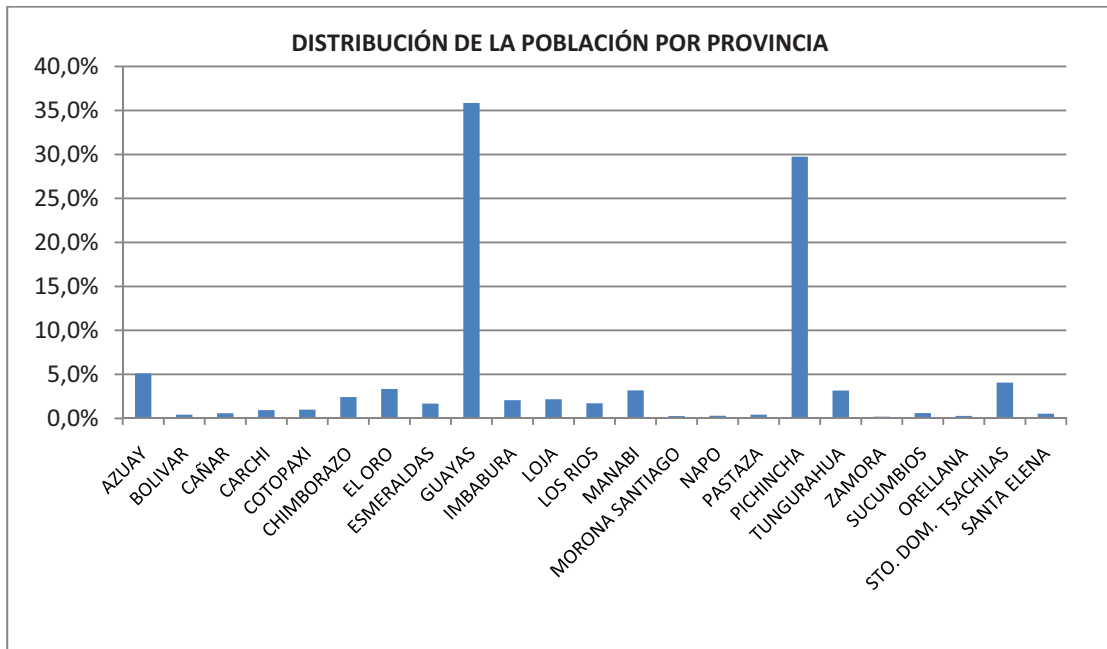
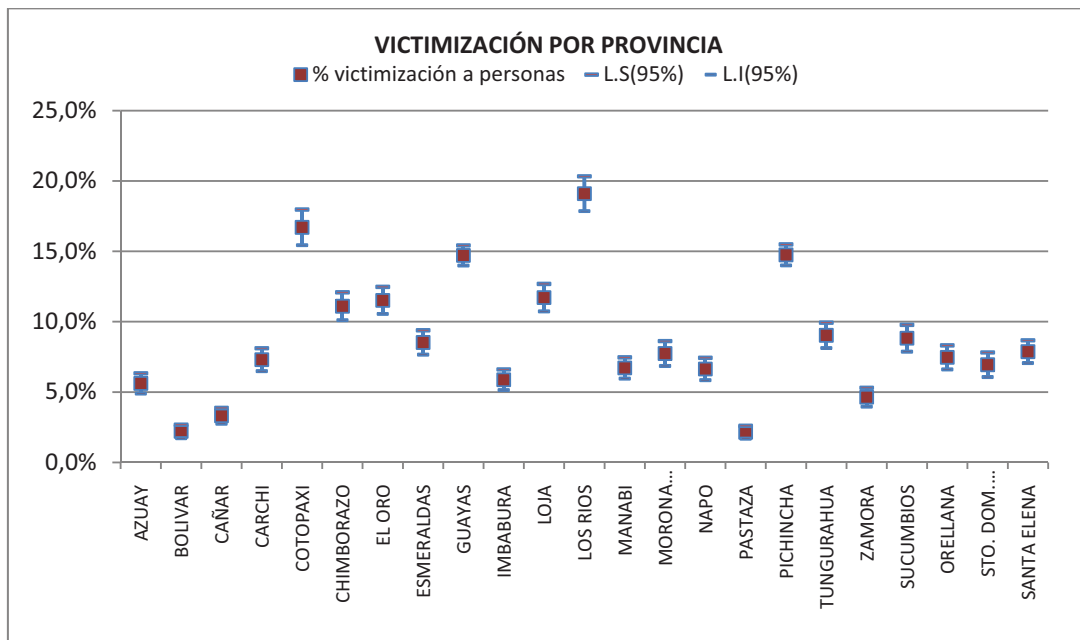


Gráfico 2.3: Victimización por Provincia



La victimización está relacionada con las Provincias. Las provincias de Cotopaxi y Los Ríos son las más elevadas con un porcentaje de victimización mayor al 15%

aunque no son significativas puesto que representan menos del 5% de la población. Guayas y Pichincha también se encuentran con una victimización por encima del promedio de la población; al realizar un análisis sobre estas dos provincias se cubriría gran parte de la población y se centraría el análisis en solo dos segmentos. A continuación se presenta la prueba chi-cuadrado para la distribución de victimizados en las provincias:

Cuadro 2.6: Prueba Chi-Cuadrado

Prueba Chi-Cuadrado			
	Valor	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	48409,667 ^a	22	,000
Likelihood Ratio	55355,320	22	,000
Linear-by-Linear Association	793,914	1	,000
N of Valid Cases	4210946		

^a. No existen frecuencias esperadas menores a 5. La mínima frecuencia esperada es 1027,51.

La victimización es diferente de acuerdo a las provincias.

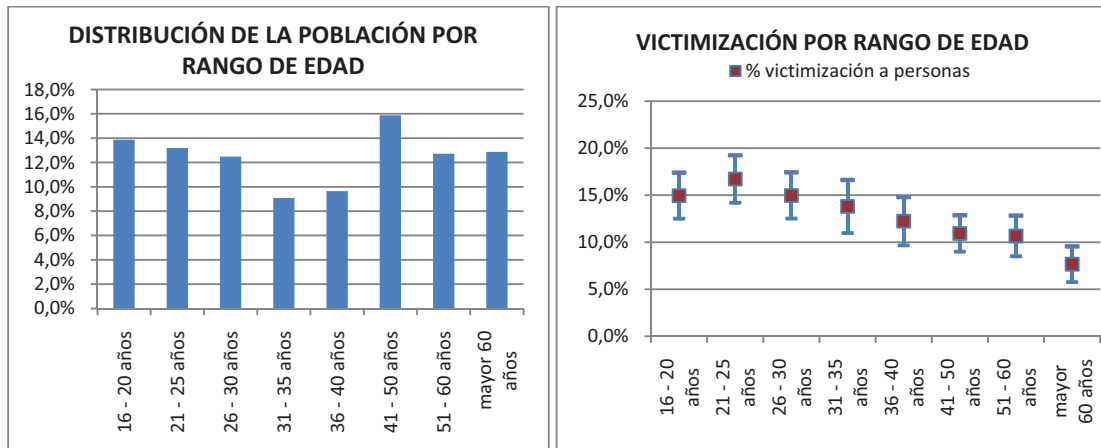
VICTIMIZACIÓN POR RANGO DE EDAD

Cuadro 2.7: Victimización por Rango de Edad

Victimización por Rango de edad						
Rango de edad	Tamaño de muestra	Frecuencia población estimada	% población estimada	% victimización a personas	Victimización Lím.inf. (95%)	Victimización Lím.sup. (95%)
No Responde	22	9.883	0,2%	0,0%	0,0%	0,0%
16 - 20 años	3.573	584.063	13,9%	14,9%	12,5%	17,4%
21 - 25 años	3.283	555.394	13,2%	16,7%	14,2%	19,2%
26 - 30 años	2.850	526.134	12,5%	15,0%	12,5%	17,4%
31 - 35 años	2.213	382.616	9,1%	13,8%	11,0%	16,6%
36 - 40 años	2.303	406.410	9,7%	12,2%	9,7%	14,8%
41 - 50 años	3.805	668.514	15,9%	10,9%	9,0%	12,9%
51 - 60 años	2.715	535.699	12,7%	10,7%	8,5%	12,8%
mayor 60 años	2.877	542.233	12,9%	7,7%	5,8%	9,6%
Total	23.641	4.210.948	100,0%	12,7%	11,8%	13,5%

Como se puede observar en el análisis realizado para la población mayor a 16 años, la distribución de la población estimada por rango de edad se muestra muy equilibrada, no hay rangos de edad que tengan una notable acumulación.

Gráfico 2.4: Nivel de Victimización por Rango de Edad



En los primeros rangos de edad el mayor porcentaje de victimización corresponde a la población de 21 a 25 años. Además, se observa una victimización decreciente de acuerdo al rango de edad por lo que se podría tener una incidencia de la victimización inversamente proporcional al rango de edad.

Cuadro 2.8: Prueba Chi-Cuadrado

Prueba Chi-Cuadrado			
	Valor	df	Asy mp. Sig. (2-sided)
Pearson Chi-Square	31490,911 ^a	8	,000
Likelihood Ratio	33742,474	8	,000
Linear-by-Linear Association	25321,385	1	,000
N of Valid Cases	4210946		

a. No existen frecuencias esperadas menores a 5. La mínima frecuencia esperada es 1251,38.

Con un p-valor de 0,000 se rechaza la hipótesis nula de independencia entre la victimización y el rango de edad.

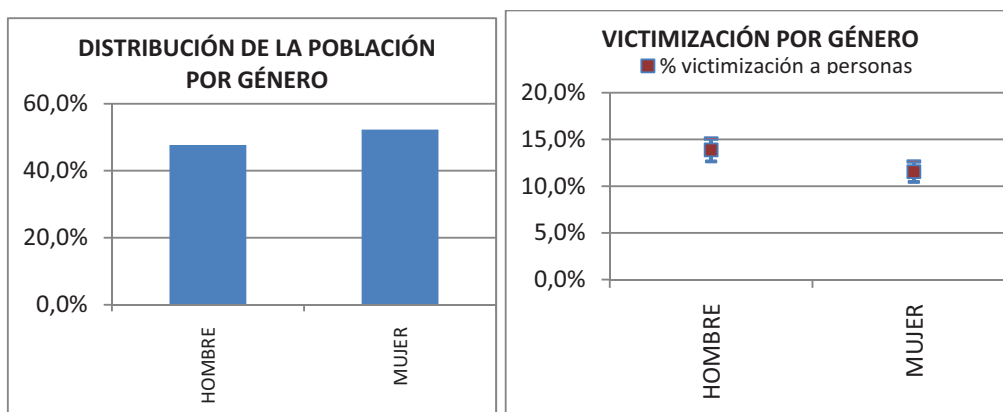
VICTIMIZACIÓN POR GÉNERO

Cuadro 2.9: Victimización por Género

Victimización por Género						
Género	Tamaño de muestra	Frecuencia población estimada	% población estimada	% victimización a personas	Victimización Lím.inf. (95%)	Victimización Lím.sup. (95%)
HOMBRE	11.184	2.010.205	47,7%	13,9%	0,0%	0,0%
MUJER	12.457	2.200.743	52,3%	11,6%	0,0%	0,0%
Total	23.641	4.210.948	100,0%	12,7%	11,8%	13,5%

La distribución por género muestra un equilibrio entre la cantidad de hombres y mujeres.

Gráfico 2.5: Nivel de Victimización por Género



Al igual que en el rango de edad se realiza una prueba chi-cuadrado para las proporciones de cada género.

Cuadro 2.10: Prueba Chi-Cuadrado

Prueba Chi-Cuadrado					
	Valor	df	Asy mp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	5145,840 ^a	1	,000		
Continuity Correction ^b	5145,629	1	,000		
Likelihood Ratio	5139,985	1	,000		
Fisher's Exact Test				,000	,000
Linear-by-Linear Association	5145,839	1	,000		
N of Valid Cases	4210948				

a. Cálculos para tabla 2x2

b. No existen frecuencias esperadas menores a 5. La mínima frecuencia esperada es 254531,1.

La prueba chi-cuadrado muestra una diferente distribución de victimizados de acuerdo al género. Los hombres son más victimizados que las mujeres, tienen 1,2 puntos porcentuales mayor al promedio de la población. Esto también se puede deducir de los intervalos de confianza correspondientes.

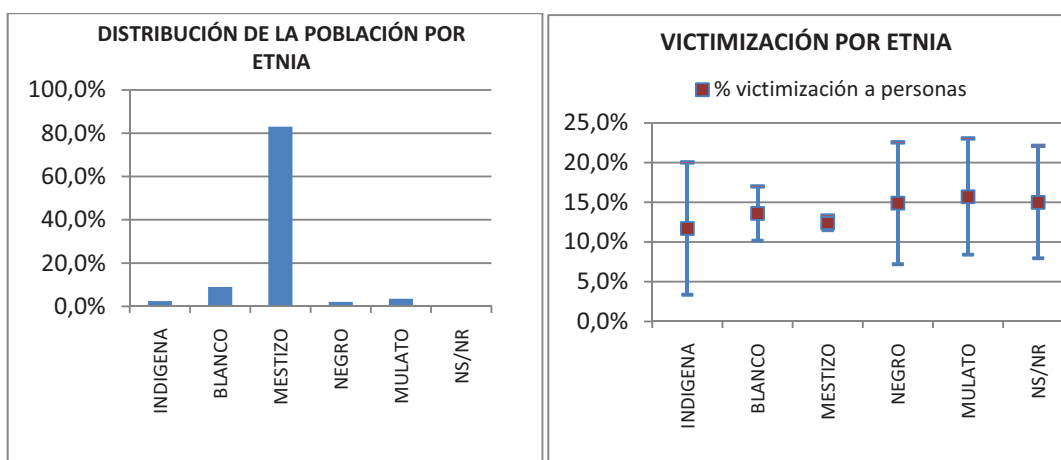
VICTIMIZACIÓN POR ETNIA

Cuadro 2.11: Victimización por Etnia

Victimización por Etnia						
Etnia	Tamaño de muestra	Frecuencia población estimada	% población estimada	% victimización a personas	Victimización Lím.inf. (95%)	Victimización Lím.sup. (95%)
INDIGENA	866	102.479	2,4%	11,7%	0,0%	0,0%
BLANCO	1.504	378.766	9,0%	13,6%	0,0%	0,0%
MESTIZO	20.155	3.493.180	83,0%	12,4%	0,0%	0,0%
NEGRO	553	86.417	2,1%	14,9%	0,0%	0,0%
MULATO	537	145.670	3,5%	15,7%	0,0%	0,0%
NS/NR	26	4.436	0,1%	15,0%	0,0%	0,0%
Total	23.641	4.210.948	100,0%	12,7%	11,8%	13,5%

La etnia se encuentra mayoritariamente conformada por la mestiza con el 83% seguida por la etnia blancos con el 9%.

Gráfico 2.6: Nivel de Victimización por Etnia



Las etnias indígena y mestiza tienen niveles de victimización similares, bajo el promedio, mientras que los blancos, mulatos y negros tienen índices superiores al promedio. Los intervalos de confianza bastante amplios aparecen por los tamaños de muestra reducidos.

VICTIMIZACIÓN POR NIVEL DE EDUCACIÓN

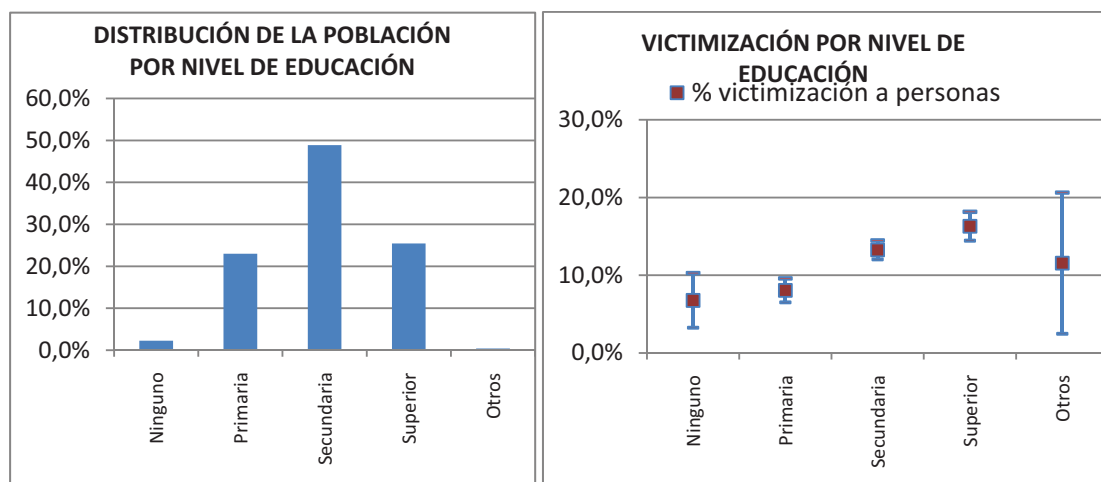
Cuadro 2.12: Victimización por Nivel de Educación

Cuadro 2.12

Victimización Nivel de Educación						
Nivel de Educación	Tamaño de muestra	Frecuencia población estimada	% población estimada	% victimización a personas	Victimización Lím.inf. (95%)	Victimización Lím.sup. (95%)
Ninguno	517	95.104	2,3%	6,8%	3,2%	10,3%
Primaria	6.419	965.891	23,0%	8,1%	6,5%	9,6%
Secundaria	10.746	2.052.455	48,9%	13,3%	12,0%	14,5%
Superior	5.861	1.068.148	25,4%	16,3%	14,4%	18,2%
Otros	74	17.135	0,4%	11,5%	2,5%	20,6%
Total	23.617	4.210.948	100,0%	12,7%	11,8%	13,5%

El cuadro del nivel de educación muestra que el 23% de la población estudió sólo primaria, 48,9% secundaria y el 25,4% superior.

Gráfico 2.7: Nivel de Victimización por Nivel de Educación



La victimización en los grupos de educación superior es mayor al promedio poblacional. Este resultado es consistente con la mayor victimización que se presenta entre los 21 y 25 años, edad en la que se desarrollan los estudios superiores, como se verá más adelante en la variable ocupación.

Cuadro 2.13: Prueba Chi-Cuadrado

Prueba Chi-Cuadrado			
	Valor	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	35020,309 ^a	4	,000
Likelihood Ratio	37062,848	4	,000
Linear-by-Linear Association	32358,028	1	,000
N of Valid Cases	4198733		

a. No existen frecuencias esperadas menores a 5. La mínima frecuencia esperada es 2173,21.

La victimización no es independiente al nivel de estudios; sin embargo, esta relación puede estar vinculada al rango de edad y a la ocupación.

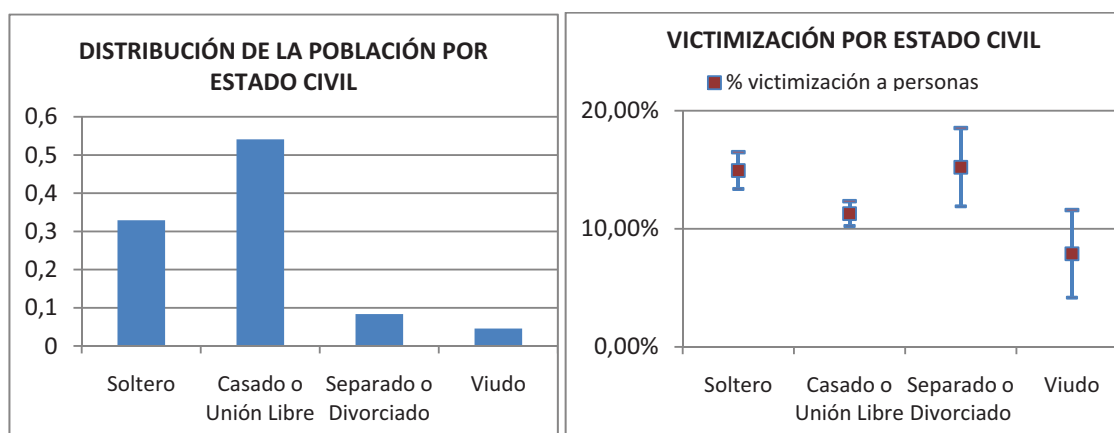
VICTIMIZACIÓN POR ESTADO CIVIL

Cuadro 2.14: Victimización por Estado Civil

Victimización Estado Civil						
Estado Civil	Tamaño de muestra	Frecuencia población estimada	% población estimada	% victimización a personas	Victimización Lím.inf. (95%)	Victimización Lím.sup. (95%)
Soltero	7.673	1.386.122	32,9%	14,93%	13,37%	16,49%
Casado o Unión Libre	13.307	2.279.329	54,1%	11,29%	10,24%	12,34%
Separado o Divorciado	1.657	352.924	8,4%	15,22%	11,90%	18,53%
Viudo	1.004	192.573	4,6%	7,88%	4,17%	11,59%
Total	23.641	4.210.948	100,0%	12,7%	11,8%	13,5%

Considerando que la población analizada es aquella de edad mayor o igual a 16 años, el 41,4% de ésta es casada, que al considerarla con la unión libre de 12,7% forman un porcentaje de parejas que conviven en 54,1%. Los solteros representan el 32,9%.

Gráfico 2.8: Nivel de Victimización por Estado Civil



Los viudos y casados son los menos victimizados mientras, los solteros, separados o divorciados son los más victimizados. Esto puede justificarse por el hecho que las personas sin obligaciones hogareñas pasan mayor tiempo fuera de su casa y, por ende, se encuentran más expuestos al peligro.

Cuadro 2.15: Prueba Chi-Cuadrado

Prueba Chi-Cuadrado			
	Valor	df	Asy mp. Sig. (2-sided)
Pearson Chi-Square	16428,143 ^a	3	,000
Likelihood Ratio	16679,280	3	,000
Linear-by-Linear Association	6275,196	1	,000
N of Valid Cases	4210946		

a. No existen frecuencias esperadas menores a 5. La mínima frecuencia esperada es 24383,46.

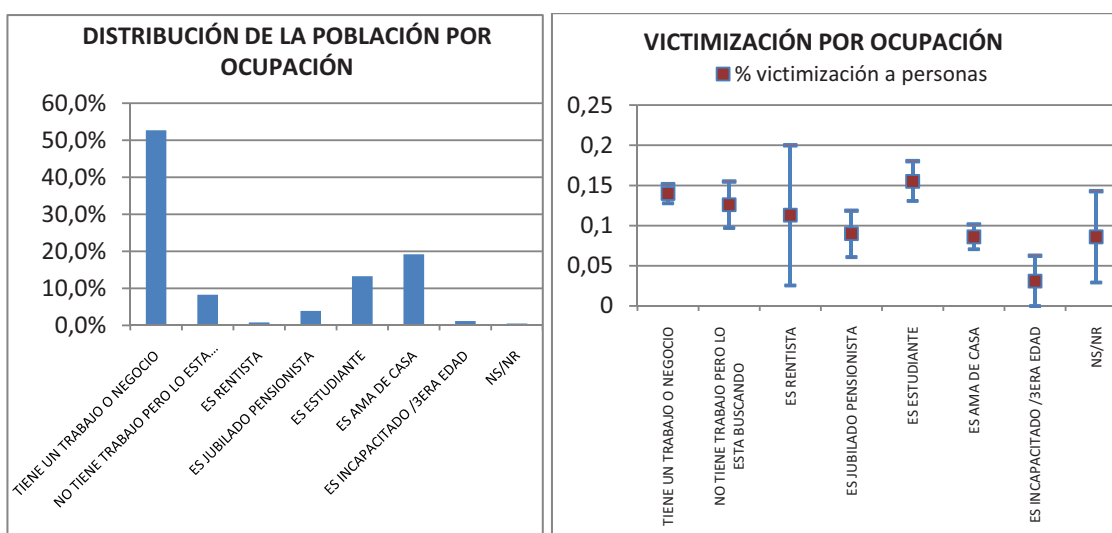
La victimización no es independiente del estado civil.

VICTIMIZACIÓN POR OCUPACIÓN

Cuadro 2.16: Victimización por Ocupación

Victimización por Ocupación						
Ocupación	Tamaño de muestra	Frecuencia población estimada	% población estimada	% victimización a personas	Victimización Lím.inf. (95%)	Victimización Lím.sup. (95%)
TIENE UN TRABAJO O NEGOCIO	12.764	2.221.172	52,7%	14,0%	12,8%	15,2%
NO TIENE TRABAJO PERO LO ESTA BUSCANDO	1.543	350.305	8,3%	12,6%	9,7%	15,5%
ES RENTISTA	158	33.702	0,8%	11,3%	2,5%	20,0%
ES JUBILADO PENSIONISTA	712	165.358	3,9%	9,0%	6,1%	11,9%
ES ESTUDIANTE	3.326	561.129	13,3%	15,5%	13,1%	18,0%
ES AMA DE CASA	4.697	808.451	19,2%	8,6%	7,1%	10,2%
ES INCAPACITADO /3ERA EDAD	357	51.005	1,2%	3,1%	0,0%	6,2%
NS/NR	84	19.823	0,5%	8,6%	2,9%	14,3%
Total	23.641	4.210.948	100,0%	12,7%	11,8%	13,5%

Gráfico 2.9: Nivel de Victimización por Ocupación



Los estudiantes con el 15,5% de victimización son los más victimizados en la población, seguidos por quienes tienen un trabajo o negocio. Son quienes se exponen más al peligro y poseen cosas que son atractivas para los delincuentes.

Cuadro 2.17: Prueba Chi-Cuadrado

Prueba Chi-Cuadrado			
	Valor	df	Asy mp. Sig. (2-sided)
Pearson Chi-Square	26279,942 ^a	7	,000
Likelihood Ratio	29092,555	7	,000
Linear-by-Linear Association	10610,295	1	,000
N of Valid Cases	4210945		

a. No existen frecuencias esperadas menores a 5. La mínima frecuencia esperada es 2509,97.

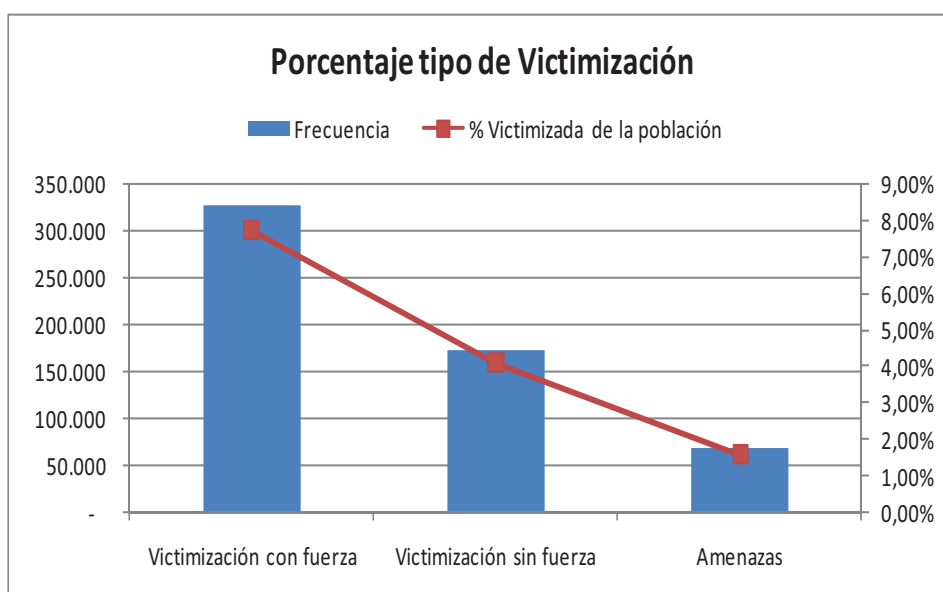
La victimización también se relaciona con la actividad de los individuos.

TIPO DE VICTIMIZACIÓN

Cuadro 2.18: Porcentaje del Tipo de Victimización

Porcentaje tipo de Victimización		
Victimización	Frecuencia	% Victimizada de la población
Victimización con fuerza	326.039	7,70%
Victimización sin fuerza	173.299	4,10%
Amenazas	68.510	1,60%
Victimización total	533.187	12,70%

Gráfico 2.10: Porcentaje del Tipo de Victimización



La victimización con fuerza es la más representativa de la victimización a personas, esta representa el 60,5% de todas las victimizaciones suscitadas a personas; esto implica un mayor peligro para las personas que sufren de esta situación.

2.4 ANÁLISIS PERCEPCIÓN

La percepción de la victimización como se indicó en el principio del presente capítulo indica la presencia indirecta de la victimización; es decir, los hechos sucedidos en su entorno o en medios de comunicación que producen un sentimiento de inseguridad. A continuación se presenta un análisis comparativo entre la percepción y la victimización⁶.

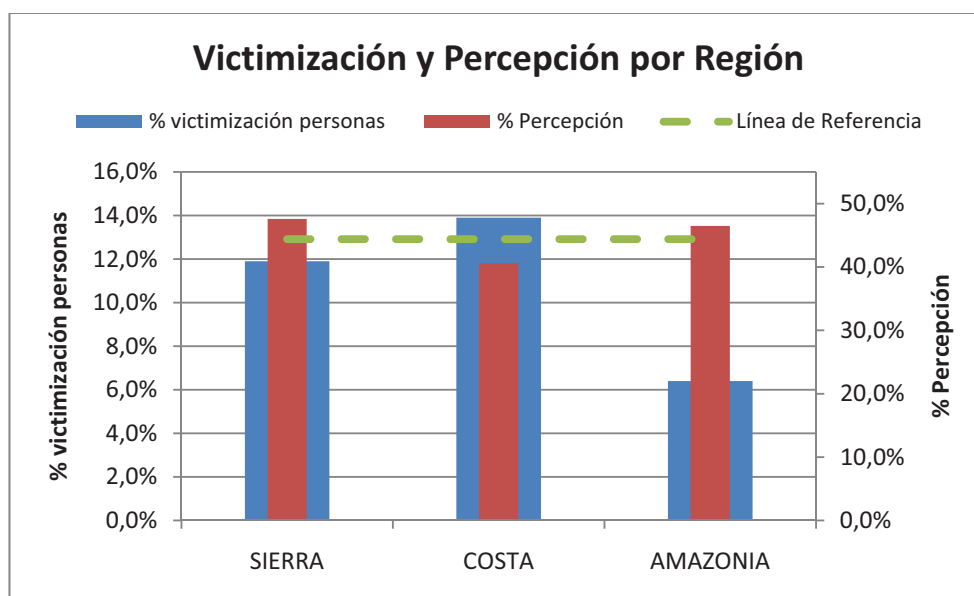
⁶ Debido al análisis comparativo entre la victimización y la percepción no se realizan pruebas de proporciones para la percepción.

REGIÓN

Cuadro 2.19: Comparación entre Percepción y Victimización por Región

Región		
Región	% victimización personas	% Percepción
SIERRA	11,9%	47,6%
COSTA	13,9%	40,6%
AMAZONIA	6,4%	46,5%
Total	12,7%	44,4%

Gráfico 2.11: Comparación entre Percepción y Victimización por Región



Nótese que la línea segmentada de color verde, línea de referencia, coincide con el promedio de victimización en la escala del eje izquierdo y la percepción de victimización en el eje derecho; esto permite una mejor interpretación comparativa de los valores, leyendo adecuadamente ambas escalas.

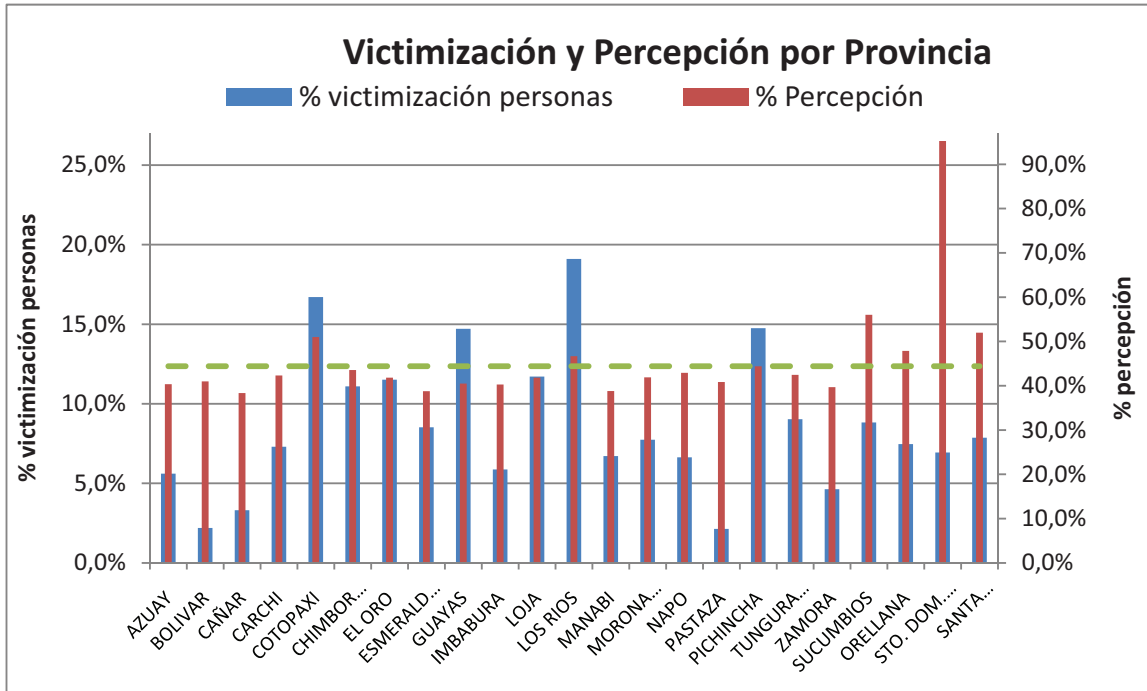
La región Costa tiene la percepción más baja pese a ser la más victimizada mientras que las otras regiones muestran una percepción mayor al promedio nacional y una victimización menor al promedio nacional; es decir, inversamente proporcional a la región Costa.

PROVINCIA

Cuadro 2.20: Comparación entre Percepción y Victimización por Provincia

Provincia		
Provincia	% victimización personas	% Percepción
AZUAY	5,6%	40,4%
BOLIVAR	2,2%	41,0%
CAÑAR	3,3%	38,4%
CARCHI	7,3%	42,3%
COTOPAXI	16,7%	51,0%
CHIMBORAZO	11,1%	43,5%
EL ORO	11,5%	41,8%
ESMERALDAS	8,5%	38,8%
GUAYAS	14,7%	40,5%
IMBABURA	5,9%	40,3%
LOJA	11,7%	41,8%
LOS RIOS	19,1%	46,7%
MANABI	6,7%	38,8%
MORONA SANTIAGO	7,7%	41,9%
NAPO	6,6%	42,9%
PASTAZA	2,1%	40,8%
PICHINCHA	14,7%	44,4%
TUNGURAHUA	9,0%	42,5%
ZAMORA	4,6%	39,7%
SUCUMBIOS	8,8%	56,0%
ORELLANA	7,5%	47,9%
STO. DOM. TSACHILAS	6,9%	95,2%
SANTA ELENA	7,9%	52,0%
Total	12,7%	44,4%

Gráfico 2.12: Nivel Percepción y Victimización por Provincia



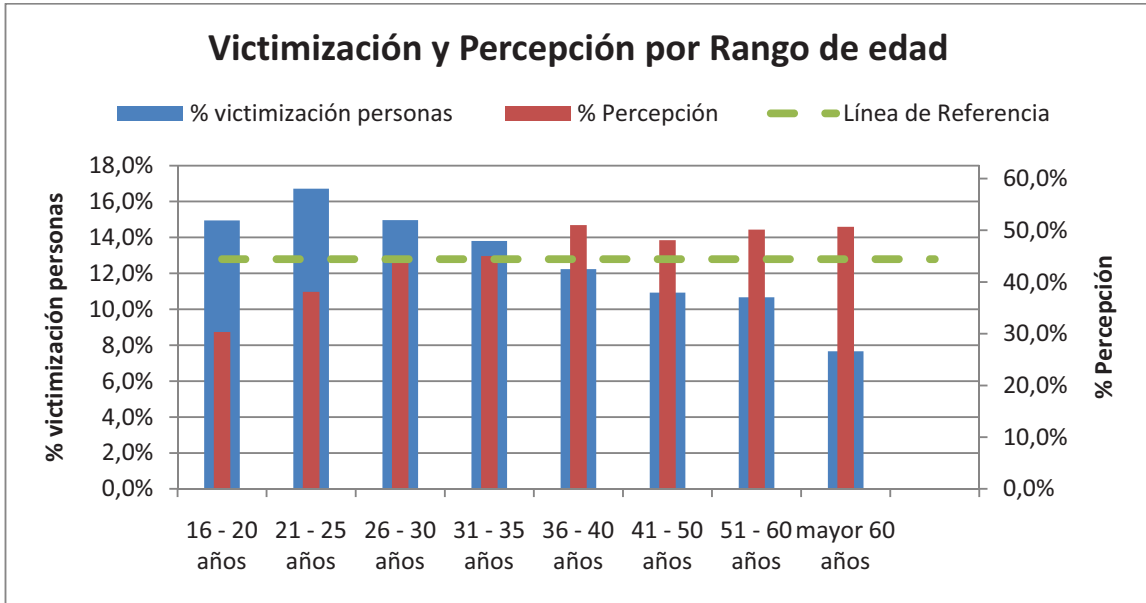
La percepción que se observa por provincia no explica la victimización que existe en cada una de ellas de forma general. Para las provincias: Los Ríos, Cotopaxi, Guayas y Pichincha, las más victimizadas, no se observa en la percepción un comportamiento diferente a las otras provincias, pues se encuentran cercanas al promedio. Esto podría significar que la victimización es un componente estructural.

RANGO DE EDAD

Cuadro 2.21: Comparación entre Percepción y Victimización por Rango de Edad

Rango de edad		
Rango de edad	% victimización personas	% Percepción
No Responde	0,0%	15,2%
16 - 20 años	14,9%	30,3%
21 - 25 años	16,7%	38,1%
26 - 30 años	15,0%	44,7%
31 - 35 años	13,8%	45,0%
36 - 40 años	12,2%	51,0%
41 - 50 años	10,9%	48,1%
51 - 60 años	10,7%	50,1%
mayor 60 años	7,7%	50,7%
Total	12,7%	44,4%

Gráfico 2.13: Comparación entre Percepción y Victimización por Rango de Edad



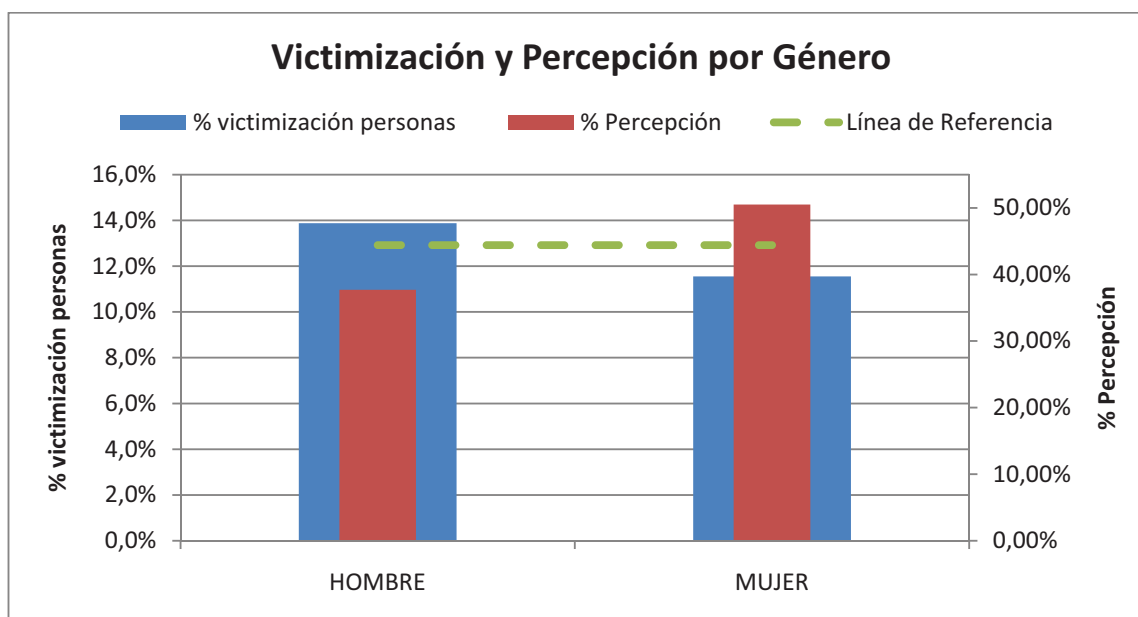
La victimización y la percepción son inversamente proporcionales. En los rangos de edad más bajos se tiene el mayor nivel de victimización y el menor porcentaje de percepción, lo que indica una menor precaución de este grupo con el peligro mientras que en los rangos de edad mayores se observa lo opuesto.

GÉNERO

Cuadro 2.22: Comparación entre Percepción y Victimización por Género

Género		
Género	% victimización personas	% Percepción
HOMBRE	13,9%	37,70%
MUJER	11,6%	50,50%
Total	12,7%	44,40%

Gráfico 2.14: Nivel de Percepción y Victimización por Género



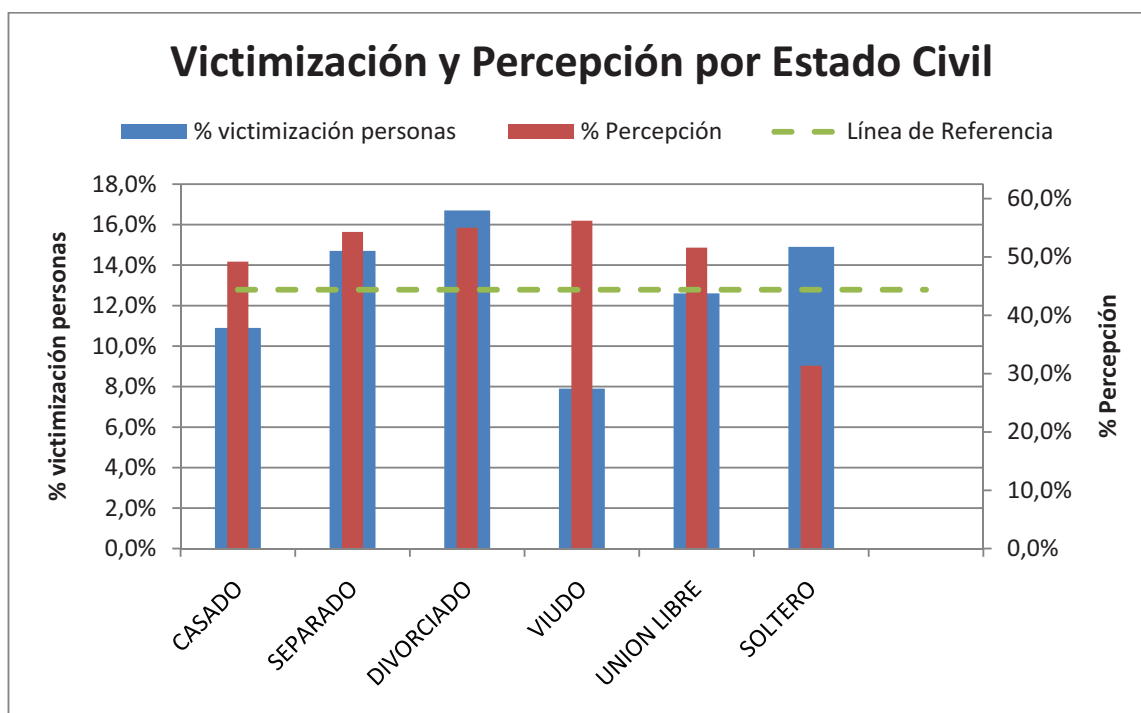
Los hombres muestran un mayor grado de victimización que el de las mujeres; este hecho podría deberse al mayor grado de exposición al peligro que podría explicarse por el menor grado de percepción al peligro que éstos tienen.

ESTADO CIVIL

Cuadro 2.23: Comparación entre Percepción y Victimización por Estado Civil

Estado Civil		
Estado civil	% victimización personas	% Percepción
CASADO	10,9%	49,2%
SEPARADO	14,7%	54,3%
DIVORCIADO	16,7%	55,0%
VIUDO	7,9%	56,2%
UNION LIBRE	12,6%	51,6%
SOLTERO	14,9%	31,4%
Total	12,7%	44,4%

Gráfico 2.15: Nivel de Percepción y Victimización por Estado Civil



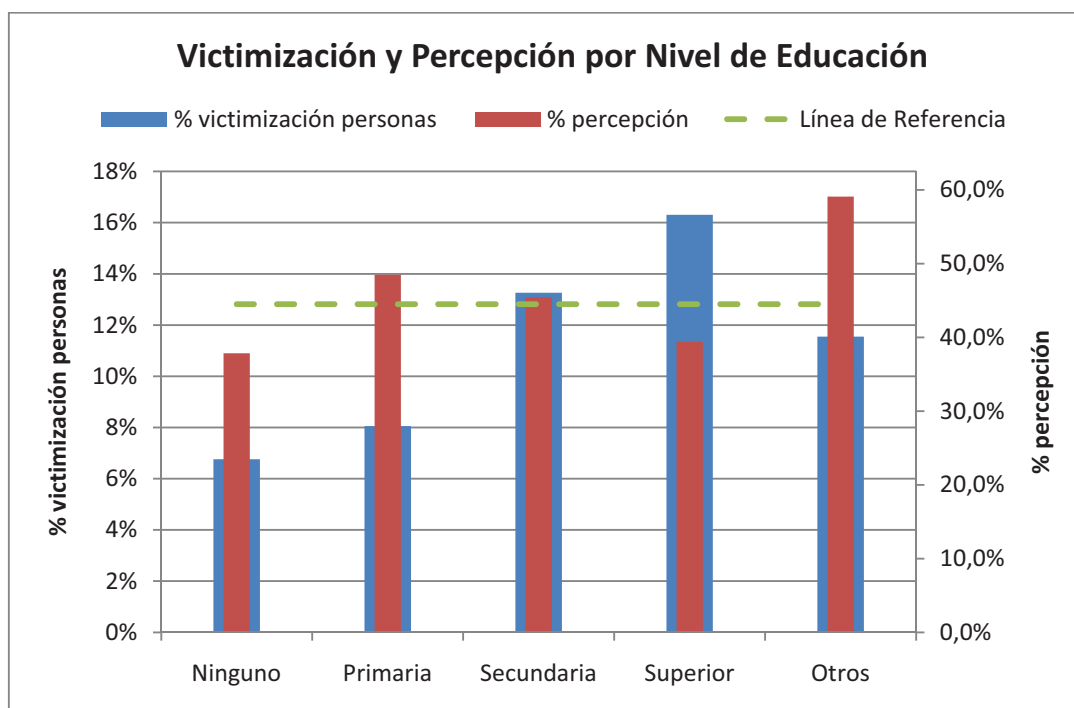
La percepción del soltero es la más baja, siendo esta una de las posibles razones de que la victimización de este grupo sea de las más altas. Los porcentajes de victimización y de percepción de los divorciados son altos.

NIVEL DE EDUCACIÓN

Cuadro 2.24: Comparación entre Percepción y Victimización por Nivel de Educación

Nivel de Educación		
Nivel de Educación	% victimización personas	% percepción
Ninguno	7%	37,8%
Primaria	8%	48,5%
Secundaria	13%	45,4%
Superior	16%	39,4%
Otros	12%	59,1%
Total	13%	44,5%

Gráfico 2.16: Nivel de Percepción y Victimización por Nivel de Educación



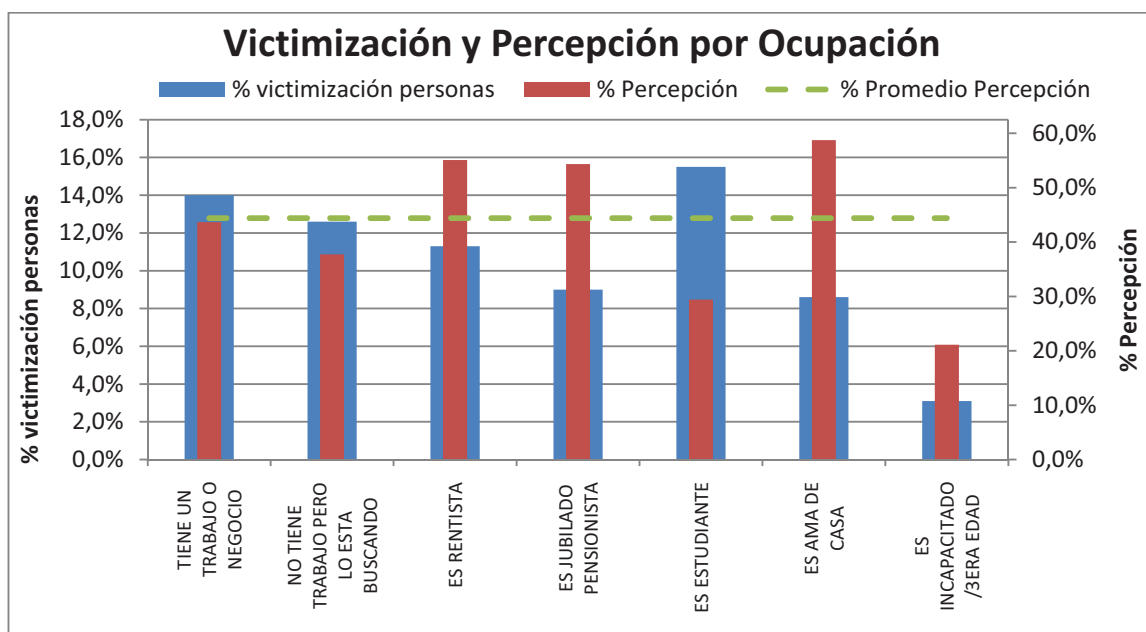
Las personas con educación universitaria tienen una percepción menor al promedio mientras el porcentaje de victimización es alto. Esto concuerda plenamente con el análisis hecho por rango de edad que muestra una mayor victimización y menor percepción para el rango de edad entre 21 y 25 años.

OCUPACIÓN

Cuadro 2.25: Comparación entre Percepción y Victimización por Ocupación

Victimización por Ocupación		
Ocupación	% victimización personas	% Percepción
TIENE UN TRABAJO O NEGOCIO	14,0%	43,7%
NO TIENE TRABAJO PERO LO ESTA BUSCANDO	12,6%	37,8%
ES RENTISTA	11,3%	55,1%
ES JUBILADO	9,0%	54,3%
PENSIONISTA	9,0%	54,3%
ES ESTUDIANTE	15,5%	29,4%
ES AMA DE CASA	8,6%	58,8%
ES INCAPACITADO /3ERA EDAD	3,1%	21,1%
NS/NR	8,6%	41,6%
Total	12,7%	44,4%

Gráfico 2.17: Comparación entre Percepción y Victimización por Ocupación



Las ocupaciones con percepción alta tienen porcentajes de victimización bajos. Los estudiantes tienen la percepción más baja con la victimización más alta. Esto confirma la relación inversa, en general, entre victimización y percepción; es decir, quien más se cuida también tiene menor probabilidad de ser victimizado.

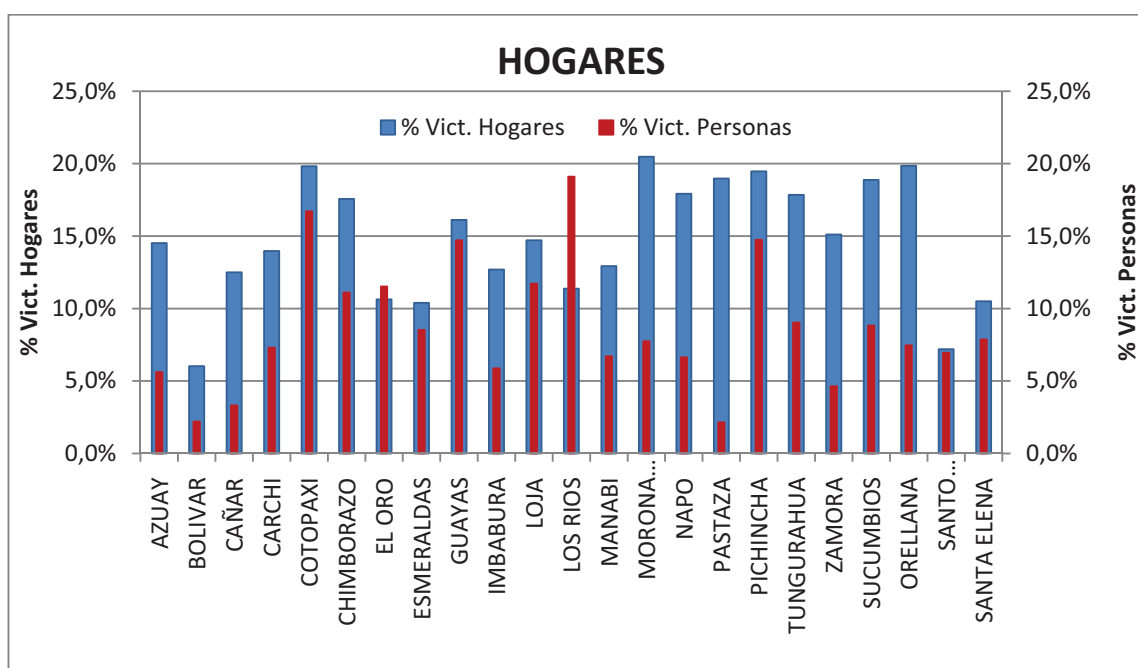
VICTIMIZACIÓN A HOGARES

En el estudio de la victimización a personas es de gran importancia estudiar la asociación o incidencia que la victimización a hogares pueda tener. Este tema es tratado con mayor detalle en el capítulo cuatro. A continuación se presenta los índices de victimización a hogares que se tiene en cada una de las provincias:

Cuadro 2.26: Victimización a Hogares y Personas por Provincia

PROVINCIA	Población	% Población	% Vict. Hogares	% Vict. Personas
AZUAY	215.124	5,1%	14,5%	5,6%
BOLIVAR	17.377	0,4%	6,0%	2,2%
CAÑAR	24.558	0,6%	12,5%	3,3%
CARCHI	39.600	0,9%	14,0%	7,3%
COTOPAXI	41.581	1,0%	19,8%	16,7%
CHIMBORAZO	101.835	2,4%	17,6%	11,1%
EL ORO	140.708	3,3%	10,6%	11,5%
ESMERALDAS	70.650	1,7%	10,4%	8,5%
GUAYAS	1.509.778	35,9%	16,1%	14,7%
IMBABURA	86.976	2,1%	12,7%	5,9%
LOJA	91.152	2,2%	14,7%	11,7%
LOS RIOS	72.127	1,7%	11,4%	19,1%
MANABI	134.188	3,2%	12,9%	6,7%
MORONA SANTIAGO	11.023	0,3%	20,5%	7,7%
NAPO	12.561	0,3%	17,9%	6,6%
PASTAZA	17.509	0,4%	19,0%	2,1%
PICHINCHA	1.252.677	29,7%	19,5%	14,7%
TUNGURAHUA	132.946	3,2%	17,8%	9,0%
ZAMORA	8.115	0,2%	15,1%	4,6%
SUCUMBIOS	24.808	0,6%	18,9%	8,8%
ORELLANA	11.989	0,3%	19,9%	7,5%
SANTO DOMINGO DE LOS TSACHILAS	171.312	4,1%	7,2%	6,9%
SANTA ELENA	22.352	0,5%	10,5%	7,9%
Total	4.210.946	100,0%	16,2%	12,7%

Gráfico 2.18: Nivel de Victimización a Hogares y Personas por Provincia



Existen provincias como Cotopaxi, Pichincha y Guayas que presentan niveles altos de victimización a personas y a viviendas. Otras provincias presentan niveles altos de victimización a hogares pero no a viviendas o viceversa; no se vislumbra una estructura diferenciada.

En el análisis descriptivo de cada una de las variables se observó que los jóvenes, estudiantes o universitarios y solteros son más victimizados que el resto de la población. También, el grupo con estas características tiene una percepción de inseguridad menor al resto de la población, siendo esta una de las posibles razones de su mayor victimización. Estos dos hechos pueden ser fruto del comportamiento natural y social que estos individuos tienen de acuerdo a su edad y estado civil.

Los divorciados también tienen un alto porcentaje de victimización que puede ser asociado a un comportamiento similar al de los jóvenes.

En cada una de las regiones se tiene niveles de inseguridad muy diferentes; por ejemplo la región Costa tiene un porcentaje de victimización de 13,9% casi dos puntos más que la Sierra con 11,95 y esta a su vez 5,5 puntos más que la región Amazónica que tiene un porcentaje de victimización de 6,4%.

Las provincias también mostraron niveles de inseguridad muy diferentes; las provincias de Cotopaxi, los Ríos, Pichincha y Guayas resultaron ser las más victimizadas a nivel nacional.

CAPITULO 3:

SCORE DE VICTIMIZACIÓN

Un modelo de *score*⁷ es un conjunto de decisiones que se obtienen mediante técnicas estadísticas que ayudan a estimar la probabilidad de que suceda un determinado evento de interés. Mediante esta probabilidad estimada es posible definir estrategias que ayuden a que el evento de estudio suceda o no. Son varias las técnicas estadísticas utilizadas para construir uno de estos modelos y se aplican en campos como la medicina, la banca, etc.

En el presente capítulo se utilizará un modelo de regresión logística para la creación del *score*, para usarlo en un tema de gran importancia social como es la victimización.

ESQUEMA METODOLÓGICO DE UN MODELO DE SCORE

En este capítulo se presentan los pasos para efectuar un modelo de *score*, así como una aplicación a los datos de victimización. Los puntos a tratar en este capítulo son:

- 1. Selección de la muestra**
- 2. Definición de buenos y malos**
- 3. Análisis de la información**
- 4. Construcción del modelo**
- 5. Validación del modelo**
- 6. Aplicación**

⁷ Se denomina *score* a la probabilidad multiplicada por un factor; expresada como un puntaje generalmente entre 0 y 1.000.

Los puntos del 1 al 5 representan el esquema metodológico de un modelo de *score*.

3.1 SELECCIÓN DE LA MUESTRA

Para desarrollar un modelo que calcule la probabilidad de un evento de interés es necesario tener una determinada experiencia, que permita definir a priori el tipo de información relevante que podría ser necesaria para el desarrollo del modelo; cuando el investigador no tiene la suficiente experiencia en el tema es necesario recurrir a un criterio experto, sobre todo cuando la información va ser extraída de una muestra.

La información de la muestra debe ser representativa de la población⁸.

Hay que recalcar que para la construcción del modelo de regresión logística se tomó el 70% de los elementos de la base total, dejándose el 30% como muestra de prueba para la validación del modelo.

Con el objetivo de desarrollar un modelo de *score* (o *scoring*) de victimización se considera la siguiente información:

- Información demográfica
- Información Económica
- Información Antropométrica
- Información de Victimización

En el presente caso, esta información ha sido recogida en la Encuesta de Victimización y Percepción en el Ecuador facilitada por la Unidad Ejecutora Especializada en Seguridad Pública (UEE)

⁸ Conocido como el *Throw the door population* en el libro *Credit Scoring and its applications*, Lyn C Thomas, 1997.

3.2 DEFINICIÓN DE BUENOS Y MALOS (VICTIMIZADOS / NO VICTIMIZADOS)

En el desarrollo de modelos de score es necesario definir el indicador de buenos y malos, desde el punto de vista estadístico, que corresponde al evento que deseamos analizar; el concepto de bueno y malo es crucial ya que se buscará la relación entre esta variable definida y la información disponible. Generalmente se define como Bueno al registro que presenta positivo a la característica de estudio. Para este trabajo el indicador ya está definido y se presenta a continuación:

VICTIMIZADO

Se considera como BUENO a los individuos que en la encuesta presentan afirmativo en el Índice de Victimización a Personas (IVP): personas (16 años de edad en adelante) que fueron víctimas en el transcurso de los últimos 12 meses de al menos uno de los siguientes delitos: robo con fuerza, robo sin fuerza, ataques y amenazas.

NO VICTIMIZADO

Se considera como MALO a personas de 16 años en adelante que no fueron víctimas en el transcurso de los últimos 12 meses de ninguno de los siguientes delitos: robo con fuerza, robo sin fuerza, ataques y amenazas.

3.3 ANÁLISIS DE LA INFORMACIÓN

En el proceso de desarrollo de cualquier modelo, la parte fundamental y que más tiempo necesita es la de extracción de información, pues es sobre ella que se

cimientos todo tipo de análisis y posteriormente acciones. Cuando la información no es adecuadamente obtenida los costos de corrección son elevados, pudiendo llegar a ser incluso más que el costo de obtención.

Para un análisis efectivo de la información se deben considerar los siguientes puntos:

- Consistencia de la información
- Tratamiento de datos de nulos
- Análisis Multivariado

CONSISTENCIA DE LA INFORMACIÓN

El análisis de la consistencia de la información, como se había mencionado anteriormente, es una de las partes más importantes en un estudio. Para realizar una prueba adecuada de consistencia es recomendable considerar una pequeña submuestra de individuos (parte de la muestra para el desarrollo del modelo), de tal manera que se pueda validar con la información física la veracidad de la misma (control de calidad). Esto ha sido realizado, como una etapa previa a la recepción de la información, por lo que no forma parte del presente trabajo.

TRATAMIENTO DE DATOS NULOS

Al disponer de una base entregada se trata de analizar la completitud de los datos obtenidos, de tal manera que se pueda determinar las variables que serán útiles en el estudio.

De la información disponible en la base se encontró que a excepción de la variable tipo de trabajo que tenía el 84% de la información en blanco, todas las variables consideradas no tenían valores perdidos. No se consideró en el análisis esta variable.

En total se tienen 23.641 registros de personas.

Las variables que originalmente se consideraron para el desarrollo del modelo son:

- **PROVINCIA**
- **DATOS DEMOGRAFICOS –SEXO**
- **DATOS DEMOGRAFICOS -AÑOS CUMPLIDOS**
- **DATOS EDUCATIVOS -NIVEL DE INSTRUCCIÓN**
- **DATOS EDUCATIVOS -ESTADO CIVIL**
- **DATOS ECONOMICOS –TRABAJO**
- **PROPORCIÓN DE PERSONAS QUE TRABAJAN FAMILIA**
- **PERSONAS POR DORMITORIO**
- **POBREZA (CRITERIOS COMUNIDAD ANDINA)**
- **POBREZA EXTREMA (CRITERIOS COMUNIDAD ANDINA)**
- **POBREZA (CRITERIOS COMUNIDAD ANDINA)**
- **PROPORCIÓN DE POBRES POR CANTÓN**

ANÁLISIS MULTIVARIADO

El modelo debería ser el más sencillo posible; que explique de mejor manera las relaciones entre las variables de interés, y que además sea congruente e interpretable. Hay que tener en cuenta que un mayor número de variables en el modelo implicará mayores errores de estimación. Un análisis univariado (como el realizado en el capítulo 2) sirve para tener una idea de la información contenida en los datos como un análisis exploratorio no obstante la selección de las variables; la agrupación de los atributos dentro de una misma variable o la agrupación de atributos de variables combinadas es posible realizarlo mediante análisis multivariado que permite realizar estos procesos en forma eficiente y sin criterios subjetivos. Para realizar este análisis se ha empleado la técnica de árboles de decisión que permite un agrupamiento de fácil interpretación.

A continuación se describe brevemente el funcionamiento de los árboles de decisión.

ÁRBOLES DE DECISIÓN

Se define como un árbol de decisión a una estructura que tiene forma de árbol en el cual las ramas presentan conjuntos de decisiones que a su vez generan sucesivas reglas de clasificación de un conjunto de datos en subgrupos disjuntos. Las ramificaciones se realizan en forma recursiva hasta el momento en que se cumplen determinados criterios de parada. Las ramificaciones empiezan considerando un nodo inicial que es a partir del cual el grupo de individuos se divide en grupos y subgrupos discriminando la variable objetivo en función de las variables explicativas.

El objetivo esencial de estos métodos es obtener grupo de individuos más homogéneos con respecto a la variable que se desea discriminar dentro de cada grupo y muy diferentes entre los grupos.

Dependiendo de la estructura del árbol, el número de ramas que se permiten generar a partir de cada nodo, son de dos tipos:

Árboles CART: que generan dos ramificaciones a partir de cada nodo.

Árboles CHAID: que generan dos o más ramificaciones a partir de cada nodo.

Su funcionamiento es como sigue:

A partir de una base de datos con una variable Y a discriminar, y un conjunto finito de variables explicativas, se trata de seleccionar a las variables que mejor discriminen a la variable Y, encontrando la mejor regla de clasificación de la variable respuesta.

CRITERIO DE BÚSQUEDA PARA LAS REGLAS DE CLASIFICACIÓN

Hay diferentes formas de efectuar divisiones en función de los valores que tomen las variables explicativas.

El método que se usará en el análisis multivariado para la selección de las variables será el CHAID. Esta es una técnica no paramétrica de árboles de clasificación que estudia distintas particiones seleccionando finalmente la mejor (el óptimo). El procedimiento utiliza una tabla de contingencia para cada partición, en la cual se estudia la dependencia de la variable objetivo en función de las variables explicativas; la partición seleccionada será aquella que muestre una mayor asociación con respecto a la variable objetivo; es decir, aquella que tenga el menor valor-p al realizar la prueba chi-cuadrado. Por esto el nombre del algoritmo CHAID de sus siglas en inglés: Chi-square Automatic Interaction Detection.

3.4 CONSTRUCCIÓN DEL MODELO

Para la fase de construcción se consideran las siguientes fases:

- La regresión logística
- La significación del modelo y las variables
- Multicolinealidad
- Ajuste global del modelo
- La interpretación del modelo

REGRESIÓN LOGÍSTICA

La ecuación de partida en los modelos de regresión logística es como sigue:

$$P(y=1|x) = \frac{e^{b_0 + \sum_{i=1}^n b_i x_i}}{1 + e^{b_0 + \sum_{i=1}^n b_i x_i}}$$

Donde:

$P(y=1|X)$: La probabilidad de que y tome el valor 1, dado X (las covariables del modelo).

b_0 : es la constante del modelo o término independiente

n : el número de covariables

b_i : los coeficientes de las covariables

x_i : las covariables que forman parte del modelo.

Esta se denomina regresión logística. En esta distribución, la relación entre la probabilidad de la variable dependiente (cualitativa dicotómica) y una covariable continua, no está definida por una recta (lo que correspondería un modelo lineal), sino que describe una forma sigmoidea (distribución logística).

El modelo de regresión logística es un modelo estadístico en el que se desea conocer la relación entre:

- Una variable dependiente cualitativa dicotómica (regresión logística binaria o Bernoulli) o con más de dos valores (regresión logística multinomial).
- Una o más variables explicativas independientes, o covariables, ya sean cualitativas o cuantitativas.

Siendo la ecuación inicial del modelo de tipo exponencial, su transformación logarítmica (logit) permite su uso como una función lineal.

LA SIGNIFICACIÓN DEL MODELO Y LAS VARIABLES

Como se observa, las covariables pueden ser cuantitativas o cualitativas. Las covariables cualitativas deben ser dicotómicas, tomando valores 0 para su ausencia y 1 para su presencia (esta codificación es importante, ya que cualquier otra codificación provocaría modificaciones en la interpretación del modelo). Pero si la covariable cualitativa tuviera más de dos categorías, para su inclusión en el modelo deberíamos realizar una transformación de la misma en varias covariables cualitativas dicotómicas ficticias o de diseño (variables *dummy*), que para este estudio fueron determinadas usando árboles de decisión, de forma que una de las categorías se tomaría como categoría de referencia. Con ello cada categoría entraría en el modelo de forma individual. También se puede crear la variable indicadora de la presencia de un grupo de categorías del total que tiene la variable original; por ejemplo, si una variable tiene 5 categorías puede crearse una variable indicadora que indique la presencia de dos de las cinco categorías. En general, si la covariable cualitativa posee n categorías, habrá que utilizar máximo $n-1$ covariables ficticias⁹.

Para la estimación de los coeficientes del modelo se recurre al cálculo de estimaciones de verosimilitud, es decir, estimaciones que maximicen la probabilidad de obtener los valores de la variable dependiente, dados los datos de las covariables. Estas estimaciones no son de cálculo directo, como ocurre en el caso de las estimaciones de los coeficientes de regresión lineal por el método de los mínimos cuadrados. Para el cálculo de estimaciones máximo-verosímiles se recurre a métodos iterativos, como el método de Newton-Raphson. Dado que el

⁹ Este procedimiento se encuentra en el documento: Modelos de Regresión Logística Incondicional; Servicio de Medicina Interna Hospital Punta de Europa Algeciras. <http://saei.org/hemero>

cálculo es complejo, normalmente hay que recurrir al uso de rutinas de programación o a paquetes estadísticos¹⁰.

El modelo final deberá ser probado mediante pruebas de significación del modelo total y de cada uno de los coeficientes de cada variable.

El siguiente paso será comprobar la significación estadística del modelo en general y cada uno de los coeficientes de regresión en el modelo.

El estadístico razón de verosimilitud: Se trata de ir contrastando cada modelo que surge de eliminar de forma aislada cada una de las covariables frente al modelo completo. La ausencia de significación implica que el modelo sin la covariable no empeora respecto al modelo completo (es decir, da igual su presencia o su ausencia), por lo que según la estrategia de obtención del modelo más reducido, dicha covariable debe ser eliminada del modelo ya que no aporta nada al mismo. Esta prueba no asume ninguna distribución concreta, por lo que es la más recomendada para estudiar la significación de los coeficientes.

El resultado de la prueba del estadístico razón de verosimilitud permite probar la hipótesis nula H_0 de que todos los coeficientes del modelo son iguales a cero. Esta se obtiene de la fórmula:

$$\text{Estadístico Razón de verosimilitud} = -2 [\ln(L_{mo}) - \ln(L_{ms})]$$

Donde:

$\ln(L_{mo})$ es el logaritmo natural de la verosimilitud del modelo inicial (sin variables).

$\ln(L_{ms})$ es el logaritmo atural de la verosimilitud del modelo seleccionado.

Este estadístico sigue una distribución chi-cuadrado con p (número de variables) grados de libertad y permite rechazar la hipótesis nula de que todos los

¹⁰ Para una presentación más detallada ver *Econometría* de Alfonso Novales, 1993, pag 541.

coeficientes son iguales a cero, si el p-valor es inferior al 0,05 (con un nivel de significación del 5%).

$$\begin{cases} H_0: \beta_i = 0 \quad \forall i \\ H_1: \exists i \text{ tal que } \beta_i \neq 0 \end{cases}$$

El estadístico de Wald: Contrasta la hipótesis de que un coeficiente aislado es distinto de 0. Su valor para un coeficiente concreto viene dado por el cociente entre el valor del coeficiente y su correspondiente error estándar, que siguen una distribución chi-cuadrado con un grado de libertad. La obtención de significación indica que dicho coeficiente es diferente de 0 y merece la pena su conservación en el modelo.

$$\begin{cases} H_0: \beta_i = 0 \\ H_1: \beta_i \neq 0 \end{cases}$$

LA MULTICOLINEALIDAD

Se dice que existe multicolinealidad cuando dos o más de las covariables del modelo mantienen una relación lineal.

Generalmente, siempre existe un cierto grado de multicolinealidad entre variables explicativas en un modelo; no obstante, se intenta que la existencia de esa multicolinealidad sea mínima de manera que no afecte mayormente en la precisión de los estimadores de los parámetros del modelo.

Cuando la colinealidad es perfecta, es decir, cuando una covariable puede determinarse según una ecuación lineal de una o más de las restantes covariables, debe eliminarse la covariable que actúa como dependiente. La presencia de variables que presentan multicolinealidad ocasionan aumentos exagerados en los errores estándar y del valor estimado para los coeficientes de la regresión, lo que resulta en estimaciones poco creíbles.

Un primer paso para analizar este aspecto puede ser examinar la matriz de coeficientes de correlación entre las covariables. Coeficientes de correlación muy elevados nos llevarán a investigar en mayor profundidad. Sin embargo este método, bueno para detectar colinealidad entre dos covariables, puede llevar a no poder detectar multicolinealidad entre más de dos de ellas.

Existen otros procedimientos analíticos para detectar multicolinealidad. Se puede olvidar, por el momento, de la variable dependiente y realizar modelos en los que una de las covariables actuará como variable dependiente y las restantes covariables como independientes de aquella. Para cada uno de estos modelos podemos calcularle su R^2 ; en caso de existir colinealidad en alguna variable o algunas variables, retirarlas de acuerdo al mayor valor de su R^2 .

Índice de Kaiser-Meyer-Olkin (KMO)

Otra forma de estudiar la multicolinealidad es mediante el índice Keiser-Meyer-Olkin. Este índice permite comparar las magnitudes de los coeficientes de correlación con las magnitudes de los coeficientes de correlación parcial. El índice KMO se calcula:

$$KMO = \frac{\sum_{i=1}^n \sum_{j=1}^n r_{ij}^2}{\sum_{i=1}^n \sum_{j=1}^n r_{ij}^2 + \sum_{i=1}^n \sum_{j=1}^n s_{ij}^2}$$

Donde:

r_{ij} es el coeficiente de correlación entre las variables i y j ($i \neq j$).

s_{ij} es el coeficiente de correlación parcial entre las variables i y j ($i \neq j$).

El índice KMO, que varía entre 0 y 1, se utiliza cuando se desea hacer un análisis factorial. Un KMO bajo indica que la intercorrelación entre las variables no es grande y por tanto no será de utilidad un análisis factorial debido a que se necesitaría un número de factores parecido al número de variables para tener un porcentaje aceptable de la información.

Para Keiser (1974):

- KMO entre 0,9 y 1 es considerado excelente para la aplicación de un análisis factorial, lo que implicaría multicolinealidad en el modelo de regresión logística.
- KMO entre 0,7 y 0,9 es considerado aceptable.
- KMO entre 0,6 y 0,7 es considerado mediocre o regular.
- KMO inferior a 0,6 es malo para aplicar un análisis factorial, lo que significaría ausencia de multicolinealidad en las variables explicativas de modelo de regresión logística.

AJUSTE GLOBAL DEL MODELO

Cuando se ha definido el modelo final se debe medir que tan bien se ajusta a los datos. Se entiende que dicho modelo tiene, por el método que se haya elegido, todas las covariables que son necesarias, una vez que se han resuelto posibles problemas de multicolinealidad. Medir el ajuste del modelo también permitirá comparar diferentes modelos permitiendo seleccionar el que mejor se ajuste.

Existen varias formas de medir la adecuación del modelo (pruebas de bondad de ajuste):

- **PRUEBA DE HOSMER-LEMESHOW**
- **ESTADISTICO DE KOLMOGOROV-SMIRNOV (COEFICIENTE K-S)**
- **COEFICIENTE GINI**

Prueba de Hosmer-Lemeshow

La prueba de Hosmer-Lemeshow agrupa a los individuos en deciles de la probabilidad estimada; a partir de esto se construye una tabla de contingencia de 10x2 entre los deciles de la probabilidad estimada y el indicador de buenos y malos. Por otro lado, se puede hacer la misma tabla con los valores estimados de buenos y malos, y sobre estas dos distribuciones, de valores observados y estimados, construirse el estadístico chi-cuadrado con el cual contrastar la hipótesis de que las dos distribuciones son independientes con un nivel de significación dado (0,05 en este trabajo).

Estadístico de Kolmogorov-Smirnov (Coeficiente K-S)

Otra forma de medir el ajuste del modelo es mediante las funciones de distribución empírica de buenos y malos, en las cuales se puede observar la máxima distancia, para determinar si son equivalentes estadísticamente. La prueba de hipótesis es la siguiente:

$$\begin{cases} H_0: \text{La distribución de los buenos es equivalente a la de los malos} \\ H_1: \text{La distribución de los buenos no es equivalente a la de los malos} \end{cases}$$

Se acepta o rechaza la hipótesis nula de acuerdo al valor alcanzado por la máxima distancia entre las dos distribuciones, esta distancia será comparada con el valor crítico de distancia máxima $D_{\max}(n, \alpha)$ con el que se aceptará o no la hipótesis nula con nivel de significación dado α para el número de datos n .

Formalmente el estadístico K-S se calcula como sigue:

$$D_{\max} = \max |F_b(x) - F_m(x)|$$

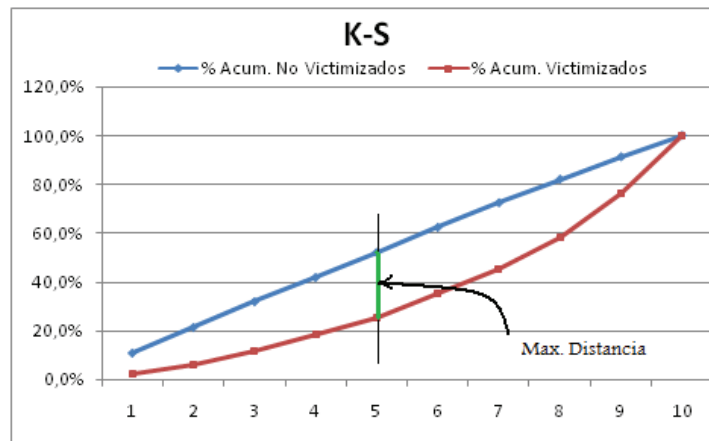
Donde:

$F_b(x)$ es la función de distribución empírica de los buenos en el valor x .

$F_m(x)$ es la función de distribución empírica de los malos en el valor x .

Dependiendo del área de investigación el valor de la máxima distancia entre la distribución empírica de los buenos y la distribución empírica de malos es un valor de aceptable discriminación.

Gráfico 3.1: FUNCIÓN DE DISTRIBUCIÓN DE BUENOS Y MALOS



Coefficiente Gini

El coeficiente de Gini se calcula sobre la curva de Lorenz, la que se explica a continuación:

La curva de Lorenz es una forma gráfica de presentar la distribución relativa de una variable en un dominio determinado. Por ejemplo, para nuestro estudio nos interesa estudiar en los deciles de la variable *score* de victimización la distribución acumulada del porcentaje de malos y buenos (como un punto del plano cartesiano). Interesa que en cada decil el porcentaje de buenos acumulado vaya aumentando mientras que en el porcentaje acumulado de malos suceda de forma inversa. Si el porcentaje de buenos y malos no cambiaría en cada uno de los deciles se puede decir que la variable *score* de victimización no discrimina a los

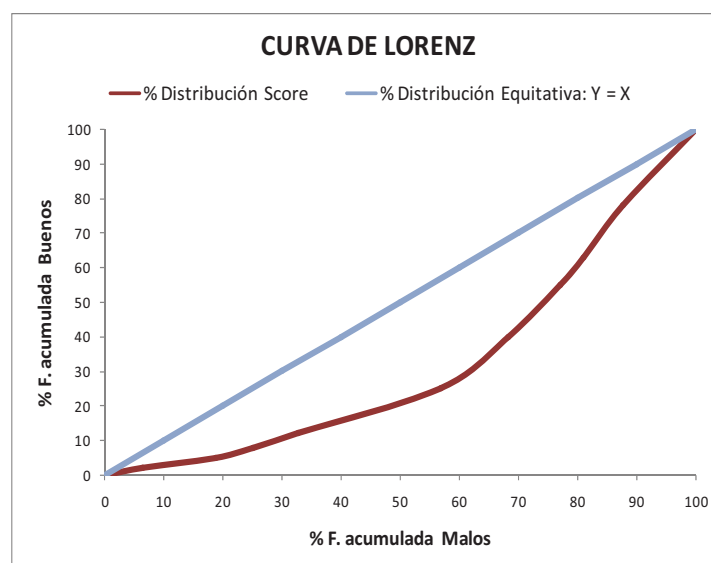
buenos y malos sino que todos estos se encuentran distribuidos en forma equitativa. A continuación se presenta un ejemplo:

Cuadro 3.1: DISTRIBUCIÓN SCORE Y DISTRIBUCIÓN EQUITATIVA

Deciles Score Victimización	% Distribución Score		% Distribución Equitativa: Y = X	
	% F. acumulada Malos	% F. acumulada Buenos	% F. acumulada Malos	% F. acumulada Buenos
D1	6	2	10	10
D2	19	5	20	20
D3	25	8	30	30
D4	33	12	40	40
D5	57	25	50	50
D6	68	40	60	60
D7	77	55	70	70
D8	81	63	80	80
D9	88	78	90	90
D10	100	100	100	100

En el gráfico se puede observar la diferencia que existe entre la curva de completa equidad y la curva de Lorenz del score de victimización.

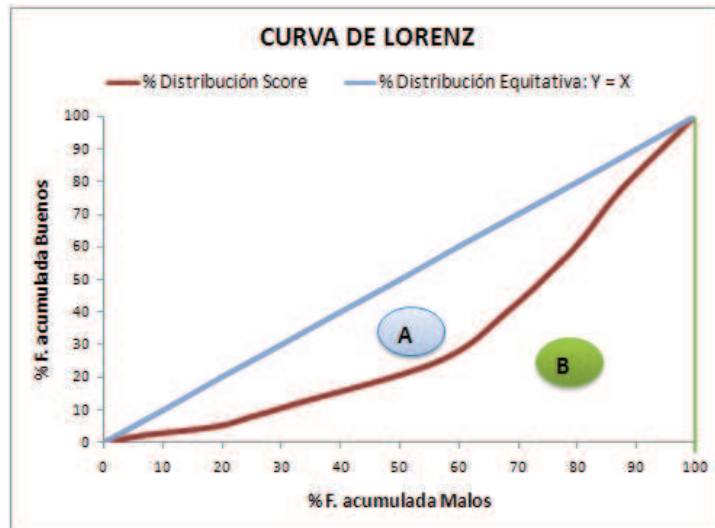
Gráfico 3.2: CURVA DE LORENZ



Mediante la construcción de la curva de Lorenz para el score de victimización en función de los buenos y malos es posible calcular el coeficiente de Gini. Éste es una medida de la desigualdad entre dos distribuciones. Este, a diferencia del K-S que mide la máxima distancia entre dos distribuciones, mide el área que existe

entre la curva de Lorenz (de la variable objeto de estudio) y la curva de completa equidad (no discriminativa).

Gráfico 3.3: ÁREAS CURVA DE LORENZ



El coeficiente de Gini se calcula como una razón de las áreas en el diagrama de la curva de Lorenz. Si el área entre la línea de perfecta igualdad y la curva de Lorenz es A, y el área por debajo de la curva de Lorenz es B, entonces el coeficiente de Gini es $A/(A+B)=2A$. Esta razón se expresa como porcentaje o como equivalente numérico de ese porcentaje, que es siempre un número entre 0 y 1. Este valor es una medida de discriminación entre el score de victimización y la función de equidad; mientras mayor sea su valor habrá una mayor diferencia y por ende mayor discriminación.

LA INTERPRETACIÓN DEL MODELO

El desarrollo de un modelo de *score* obedece a dos posibles objetivos:

- Un modelo de predicción
- Un modelo de explicación

Para desarrollar un modelo de predicción se debe estudiar además del poder de discriminación, su estabilidad en el transcurso del tiempo, así como el proceso, continuo y automático en caso de ser necesario, de recolección de la información que alimente el modelo y permita monitorearlo.

En el caso que se desee realizar un modelo de explicación se debe realizar el análisis de discriminación del modelo antes de referenciarlo como una herramienta de decisión para el tema objetivo en ese periodo de tiempo.

Para el presente trabajo es de interés validar la aplicabilidad de utilizar un modelo de regresión logística, generando reglas de decisión que permitan proponer estrategias de acción que minimicen, en la medida de lo posible, la victimización; es decir, se trata de un modelo explicativo.

3.5 VALIDACIÓN

La validación del modelo se debe realizar en una muestra que no haya formado parte de la muestra de desarrollo. El análisis debe contener las siguientes etapas:

- **Análisis de significación de las variables.**
- **Análisis de polaridad de las variables**
- **Ajuste del modelo**

Análisis de significación de las variables.

Utilizando tablas de contingencia se debe probar si las variables que resultaron significativas en la fase de desarrollo lo son en la muestra de prueba; es decir, con el estadístico chi-cuadrado se debe realizar un análisis de significación de cada una de las variables del modelo con la variable objetivo en la muestra de prueba. Si existen variables que son independientes de la variable objetivo se debe re-

hacer el modelo sin la inclusión de estas ya que no aseguran predicción ni interpretación.

Análisis de polaridad de las variables

Aun que las variables seleccionadas en el modelo del desarrollo sean significativas en la muestra de prueba es posible que estas no estén correctamente polarizadas en la muestra de prueba.

La polaridad se refiere al signo positivo o negativo que acompaña al coeficiente de cada variable. Si el signo del coeficiente es negativo quiere decir que la presencia de esa variable penaliza la probabilidad del evento (en la regresión logística) y se debe a que los individuos que tienen la presencia de esa variable (por ejemplo, los jubilados), en su mayoría no muestran la presencia de la variable objetivo (no son victimizados).

Se esperaría que la polaridad de los coeficientes de las variables en la base de desarrollo, se mantenga en la base de prueba ya que de otro modo la variable, que puede ser significativa en la base de desarrollo y de prueba, mostraría inestabilidad en la predicción, siendo positiva/negativa en la base de desarrollo y de signo opuesto en la base de prueba, por lo que también debería ser retirada del modelo. La aplicación es de la siguiente manera:

Si el porcentaje de buenos en una muestra es de $a\%$, una variable dicotómica premia al modelo (coeficiente positivo en la regresión logística) si de los individuos que presentan esta variable el $b\%$ son buenos con $b\% > a\%$. Una variable dicotómica está alineada cuando esta relación se mantiene en la base de desarrollo y de prueba.

El análisis expuesto en el párrafo anterior también puede realizarse con otra medida llamada en inglés *odds ratio*. Para un individuo dado, se define como el cociente entre la probabilidad de que y sea igual a 1 dado el patrón x y la probabilidad de que y sea igual a 0 dado x ; es decir:

$$OR(X) = \frac{P(Y=1|X)}{1-P(Y=1|X)}$$

Así obtener un *odds ratio* de por ejemplo 1/3 se interpreta diciendo que la probabilidad de ser bueno ($y=1$) es un tercio de la probabilidad de ser malo o que la probabilidad de ser malo es tres veces mayor que la probabilidad de ser bueno (dado el patrón x).

Para comprobar que la polaridad de las variables tenga sentido con el signo de los coeficientes en la base de prueba se compara el *odds ratio* de cada variable con el *odds ratio* de esa misma muestra, esperando que si el coeficiente de una variable premia la probabilidad de evento (es positivo) entonces el *odds* de esa variable sea mayor al *odds* de la muestra.

3.6 APLICACIÓN

ANÁLISIS DE LA INFORMACIÓN

Como se explicó en la sección tres de este capítulo, no se tienen valores en blanco a excepción de la variable tipo de trabajo que tenía un porcentaje significativo de valores en blanco (mayor al 5%) que no se consideró para el desarrollo del modelo.

El análisis exploratorio de cada una de las variables se muestra a continuación:

Cuadro 3.2: ANÁLISIS EXPLORATORIO

VARIABLE	DESCRIPCIÓN	Validos	Perdidos	Tipo Variable	Categorías Representativas	Categorías poco Representativas
Provincia	Indica la provincia del individuo	23.641	-	Nominal	Pichincha, Guayaquil	Todas excepto Pichincha, Guayas
Datos demográficos - Sexo	Indica el género del individuo	23.641	-	Nominal	Todas: masculino, femenino	
Datos demográficos - Años	Indica la edad en años del individuo	23.641	-	Continua	Media: 38,5 años	
Datos educativos - Nivel de Instrucción	Indica el nivel de instrucción del individuo	23.641	-	Nominal	Primaria; Secundaria; Sup. Universitaria; Educ. Media	Todas no Representativas
Datos educativos - Estado	Indica el estado civil del individuo	23.641	-	Nominal	Casado, Soltero	Divorciado
Datos económicos - Trabajo	Indica el tipo de actividad que realiza	23.641	-	Nominal	Trabajo, Estudiante, ama de casa	Rentista, Jubilado, Incapacitado
Proporción de personas que trabajan en la familia	Es el cociente entre las personas que trabajan y el total de integrantes de la familia	23.641	-	Continua	Media: 130,12 USD	
Personas por dormitorio	Es el cociente entre las personas y el número de cuartos en la vivienda	23.641	-	Continua	Media: 1,97	
Pobreza (CCA)	Indica si el individuo es pobre de acuerdo al Criterio de la Comunidad Andina	23.641	-	Nominal	No es representativa	No es representativa
Pobreza Extrema (CCA)	Indica si el individuo es pobre extremo de acuerdo al CCA	23.641	-	Nominal	No es representativa	No es representativa
Proporción de trabajadores por cantón	Es el cociente entre trabajadores y el total de individuos en el cantón	23.641	-	Continua	Media: 0,539	
Área Rural	Indica el área en que vive el individuo	23.641	-	Nominal	Todas Significativas: Rural / Urbana	

Existen variables indicadoras como la pobreza que no son representativas y que por tal razón no son consideradas en el modelo. La variable provincia tiene a la mayoría de categorías provincias no representativas; es decir contienen menos del 5% de la población. Estas categorías no representativas de esta variable, así como de otras serán agrupadas utilizando árboles de decisión.

ANÁLISIS MULTIVARIADO

Se generaron árboles de decisión para obtener la mejor segmentación que cada una de las variables debería considerar para la discriminación de la victimización. En este análisis se incorporó también el análisis de conjunto de dos variables combinadas.

La variable objetivo sobre la cual se realizará las particiones es el indicador de victimización de personas, para el cual se definió como bueno al individuo que es victimizado y malo al que no lo es. Considerando esta variable en el nodo inicial el algoritmo CHAID selecciona las variables más significativas para la discriminación de la variable objetivo.

Cuadro 3.3: CHÍ-CUADRADO VARIABLES ANSWER TREE

Predictor	S.	Chi-Square	D.	Adj.Prc
PROVINCIA	7 D..	551,5649	6	0,000000
NIVEL DE INSTRUCCIÓN	4 D..	145,3574	3	0,000000
3. DATOS DEMOGRAFICOS -ANOS CUM...	5 D..	119,2962	4	0,000000
12. DATOS ECONOMICOS -TRABAJO	3 D..	115,9732	2	0,000000
7. DATOS EDUCATIVOS -ESTADO CIVIL	3 D..	67,8295	2	0,000000
2. DATOS DEMOGRAFICOS -SEXO	*. A..	*	*	*
Proporción de personas que trabajan	*. A..	*	*	*
Pobreza (CRITERIOS COMUNIDAD ANDI...	*. A..	*	*	*

La clasificación que presenta el cuadro de arriba muestra las variables ordenadas de acuerdo al valor estadístico chi-cuadrado, estadístico que mide la dependencia de cada una de las variables con la variable objetivo. Se puede observar que la variable que más discrimina la victimización es la provincia, después la variable que más discrimina es la variable nivel de instrucción y así sucesivamente¹¹. Es importante saber que el nivel de discriminación de cada una de las variables sobre la variable objetivo considera una relación directa, es decir no considera variables combinadas por lo que este análisis muestra cuáles son las variables originales simples más significativas para el análisis. Una vez conocida la importancia de cada una de las variables será necesario realizar árboles de variables combinadas que pueden mostrar otras relaciones que el análisis con variables simples no lo hace, alcanzando mayores niveles de discriminación e interpretación.

¹¹ La variable con mayor chi-cuadrado tiene mayor relación con la variable dependiente.

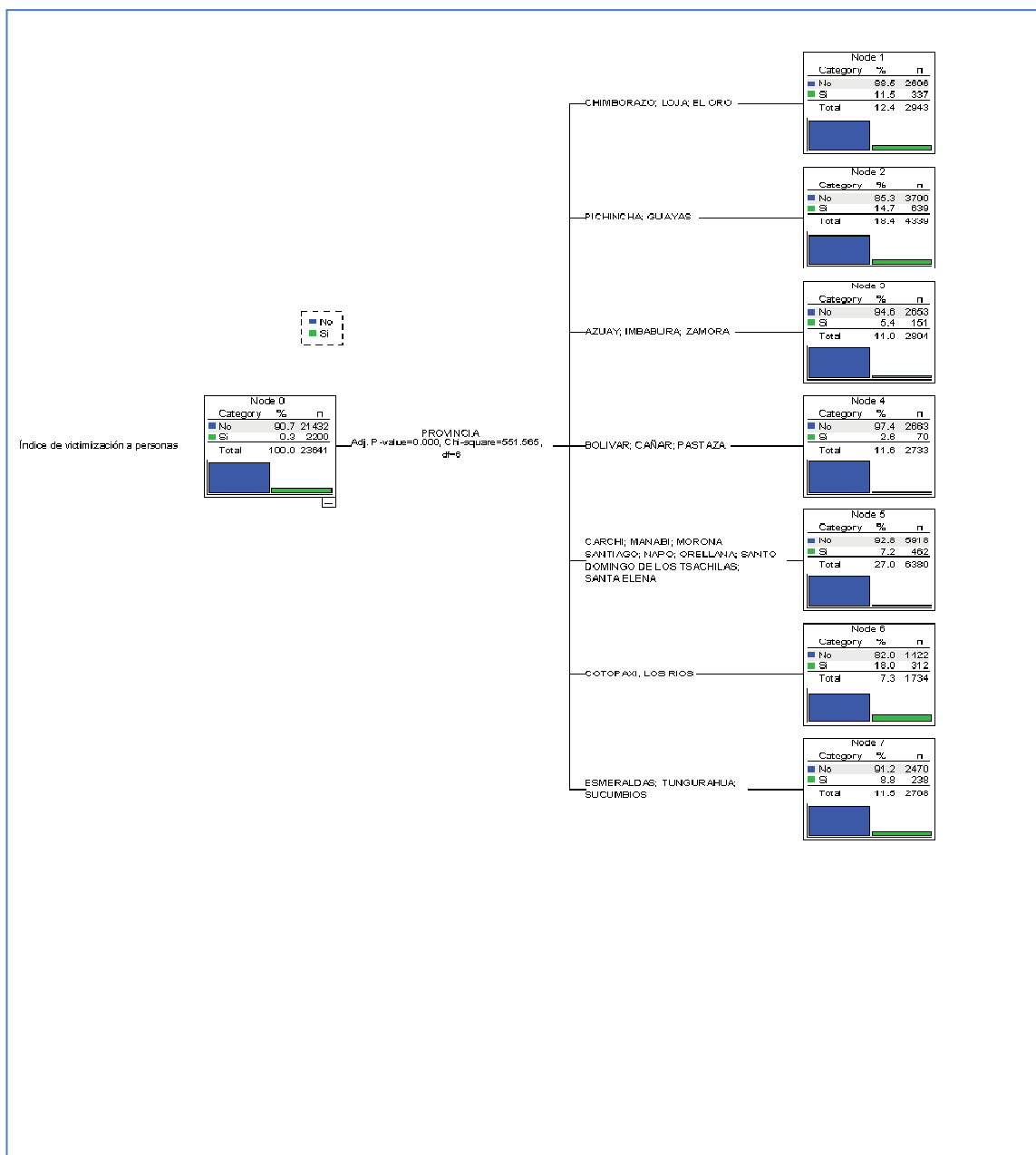
Este análisis es realizado en el paquete estadístico Answer tree 3.1. También puede efectuarse en el SPSS 15.0, con la dificultad que este programa no muestra la clasificación ranqueada de todas las variables por lo que se debe realizar el análisis generando el árbol para cada variable y posteriormente comparar su chi-cuadrado, mientras que el Answer tree 3.1 lo realiza automáticamente (es un programa especializado y específico para el uso de árboles de decisión).

Las variables que resultaron significativas después de este análisis son:

1. Categoría Victimización por Provincia
2. Categoría Victimización por Pobreza (Directa)
3. Categoría Victimización por Nivel de Instrucción
4. Categoría Victimización por Ingreso Per cápita
5. Categoría Victimización por Parentesco con el Padre
6. Categoría Victimización Porcentaje que trabajan por Cantón
7. Categoría Victimización por Rango de Edad
8. Categoría Victimización por Género
9. Categoría Victimización por Edad y Sexo
10. Categoría Victimización Instrucción y Edad
11. Categoría Victimización por Área

Por ejemplo para la variable Categoría Victimización por Provincia se desarrolló el siguiente análisis:

Gráfico 3.4: ÁRBOL DE DECISIÓN POR PROVINCIA



Se generó el árbol de decisión entre la variable dependiente Victimización y la variable Provincia.

Con base a este árbol se construyó la variable categórica para la variable Provincia.

Categoría Victimización por Provincia

Esta variable presenta las siguientes Categorías:

Cuadro 3.4: ÁRBOL GENERADO POR PROVINCIA

Categoría Victimización por Provincia
CHIMBORAZO; LOJA; EL ORO
PICHINCHA; GUAYAS
AZUAY; IMBABURA; ZAMORA
CARCHI; MANABÍ; MORONA S; NAPO; ORELLANA; STO DOMINGO; STA ELENA
COTOPAXI; LOS RÍOS
ESMERALDAS; TUNGURAHUA;SUCUMBÍOS
OTROS

Para Cada una de estas categorías se generó una variable dicotómica para incluirlas en el modelo a excepción de una, la que presente menor discriminación en el árbol:

V1: Si el individuo vive en una de las siguientes provincias: Chimborazo, Loja, El Oro.

V2: Si el individuo vive en una de las siguientes provincias: Pichincha, Guayas.

V3: Si el individuo vive en una de las siguientes provincias: Azuay, Imbabura, Zamora

V4: Si el individuo vive en una de las siguientes provincias: Carchi, Manabí, Morona Santiago, Napo, Orellana, Sto Domingo

V5: Si el individuo vive en una de las siguientes provincias: Cotopaxi, Los Ríos

V6: Si el individuo vive en una de las siguientes provincias: Esmeraldas, Tungurahua, Sucumbíos

Este proceso se desarrolló para todas las variables originales arriba mencionadas y se las ingresó para el desarrollo del modelo. La agrupación realizada para cada una de las variables es¹²:

Cuadro 3.5: ÁRBOL GENERADO POR NIVEL DE INSTRUCCIÓN

Categoría Victimización por Nivel de Instrucción
PRIMARIA; OTROS
SECUNDARIA
NINGUNO
SUPERIOR

Cuadro 3.6: ÁRBOL GENERADO POR INGRESO PERCÁPITA

Categoría Victimización por Percapita
ING PER CAPITA <=60
ING PERCAPITA > 60

Cuadro 3.7: ÁRBOL GENERADO POR PARENTEZCO CON JEFE DE HOGAR

Categoría Victimización por Parentezco
PADRES; SUEGROS; OTROS NO PARIENTES
JEFE; NIETO NIETO/A
CONYUGE; EMPLEADA; YERNA O NUERA
OTROS PARIENTES; HIJO/A

Cuadro 3.8: ÁRBOL GENERADO POR PORCENTAJE DE PERSONAS QUE TRABAJA POR CANTON

Categoría Victimización Porcentaje que trabajan por Cantón
MENOR O IGUAL A 48%
MAYOR A 48% Y MENOR A 57%
IGUAL AL 57%
MAYOR A 57% Y MENOR O IGUAL A 59%
MAYOR A 59%

¹² Al igual que en la Provincia se generaron variables dicotómicas para las categorías de cada variable definidas por lo árboles de decisión. Estos se muestran en el Anexo 1.

Cuadro 3.9: ÁRBOL GENERADO POR EDAD

Categoría Victimización por Rango de Edad
<= 18 AÑOS
MAYOR A 18 AÑOS Y MENOR O IGUAL A 35
MAYOR A 35 Y MENOR O IGUAL A 53 AÑOS
MAYOR A 53 Y MENOR O IGUAL A 63 AÑOS
MAYOR A 63 AÑOS

Cuadro 3.10: ÁRBOL GENERADO POR INSTRUCCIÓN Y EDAD

Categoría Victimización Instrucción Edad
SUPERIOR MENOR A 35 AÑOS
SUPERIOR MAYOR A 35 AÑOS
OTROS

Cuadro 3.11: ÁRBOL GENERADO POR ÁREA

Categoría Victimización por Area
AREA URBANA
AREA RURAL

En el cuadro de abajo se puede observar como la combinación de variables aumenta los niveles de discriminación.

Cuadro 3.12: CHI-CUADRADO VARIABLES ANSWER TREE

Predictor	Nodes	Split Type	Chi-Square
Categoría Victimización por Provincia	7	Default	551,5649
Categoría Victimización Instrucción Edad	4	Default	196,1437
Categoría Victimización por Nivel de Instruc...	4	Default	145,3574
Categoría Victimización por Edad y Sexo	5	Default	136,9872
Categoría Victimización por Rango de Edad	4	Default	115,6961
Categoría Victimización por Pobreza (Directa)	*	Arbitrary	*
Categoría Victimización por Percapita	*	Arbitrary	*
Categoría Victimización por Parentezco	*	Arbitrary	*

Observe que la combinación de la variable nivel de instrucción y edad alcanza un valor chi-cuadrado mayor al de las variables solas; en este caso, el nivel de educación alcanza el siguiente lugar (tercero) mientras que la edad sola se encuentra en quinto lugar. La asociación que tienen las variables combinadas con la variable objetivo muestra una perspectiva diferente al de las variables simples.

CONSTRUCCIÓN DEL MODELO

El modelo fue desarrollado con regresión logística, en el cual se desea estimar la probabilidad de ser victimizado. Las variables dicotómicas que se seleccionaron son:

Cuadro 3.13: VARIABLES SELECCIONADAS EN EL MODELO

Variable	Descripción	Tipo Variable
V_2	Provincia: Pichincha Guayas	Indicadora (1/0)
V_3	Provincia: Azuay Imbabura, Zamora	Indicadora (1/0)
V_4	Provincia: Bolívar Cañar Pastza	Indicadora (1/0)
V_5	Provincia: Carchi Manabí; Moro. Sant; Napo; Orellano; Sto. Dom. Tsáchilas; Sta. Elena	Indicadora (1/0)
V_6	Provincia: Coptopaxi, Los Ríos	Indicadora (1/0)
V_20	Edad mayor a 63 años	Indicadora (1/0)
V_28	Mujer entre 18 y 35 años	Indicadora (1/0)
V_30	Educación Superior menor a 35 años	Indicadora (1/0)
V_31	Area Rural	Indicadora (1/0)
P3_3	Edad	Continua

SIGNIFICACIÓN DEL MODELO Y DE LAS VARIABLES

Una vez consideradas todas las variables originales y transformadas en variables dicotómicas, se aplicó el método de máxima verosimilitud para la estimación del modelo con pasos hacia delante, que va seleccionando las variables más significativas hasta no tener ninguna con un p-valor menor a 0,05. Por cada inclusión de una variable se calcula la verosimilitud del modelo con la nueva variable.

Cuadro 3.14: RESUMEN DEL MODELO

RESUMEN DEL MODELO

Step	-2 Log likelihood	Cox & Snell R Cuadrado	Nagelkerke R Cuadrado
1	10068,822 ^a	,007	,015
2	9913,565 ^a	,016	,035
3	9825,450 ^a	,021	,047
4	9751,042 ^b	,026	,056
5	9707,472 ^b	,028	,062
6	9683,339 ^b	,030	,065
7	9659,800 ^b	,031	,068
8	9630,283 ^b	,033	,072
9	9623,534 ^b	,033	,072
10	9619,372 ^b	,034	,073

- a. La estimacion terminó en la iteración numero 5 porque los parámetros cambiaron en menos de 0,001
- b. La estimacion terminó en la iteración numero 5 porque los parámetros cambiaron en menos de 0,001

El estadístico de razón de verosimilitud se calcula para el último paso con el que se contrastará la hipótesis de que el modelo no es significativo mediante la diferencia de la verosimilitud inicial y final.

Verosimilitud del modelo inicial: $L_0=10.183,977$

Verosimilitud del modelo final: $L_f=9.619,372$

$$\text{Razón de Verosimilitud} = -2 \ln \left(\frac{L_0}{L_f} \right) = 564,605 \text{ con } 10 \text{ grados de libertad}$$

Esto se muestra en el siguiente cuadro:

Cuadro 3.15: PRUEBA OMNIBUS DE LOS COEFICIENTES**Prueba Omnibus para los Coeficientes**

		Chi-square	df	Sig.
Paso 1	Step	115,155	1	,000
	Block	115,155	1	,000
	Model	115,155	1	,000
Paso 2	Step	155,257	1	,000
	Block	270,412	2	,000
	Model	270,412	2	,000
Paso 3	Step	88,115	1	,000
	Block	358,527	3	,000
	Model	358,527	3	,000
Paso 4	Step	74,408	1	,000
	Block	432,934	4	,000
	Model	432,934	4	,000
Paso 5	Step	43,570	1	,000
	Block	476,505	5	,000
	Model	476,505	5	,000
Paso 6	Step	24,133	1	,000
	Block	500,638	6	,000
	Model	500,638	6	,000
Paso 7	Step	23,539	1	,000
	Block	524,177	7	,000
	Model	524,177	7	,000
Paso 8	Step	29,516	1	,000
	Block	553,693	8	,000
	Model	553,693	8	,000
Paso 9	Step	6,749	1	,009
	Block	560,443	9	,000
	Model	560,443	9	,000
Paso 10	Step	4,162	1	,041
	Block	564,605	10	,000
	Model	564,605	10	,000

Con este valor de chi-cuadrado (564,605) se rechaza la hipótesis nula de que todos los coeficientes del modelo son iguales a cero con un nivel de significación $\alpha=0,05$.

Significancia de las variables

Con el uso del estadístico de Wald se analizó la significación de cada una de las variables.

Cuadro 3.16: RESUMEN VARIABLES SELECCIONADAS

Variable	Descripción	B	S.E.	Wald	df	Sig.	Exp(B)
V_2	Provincia: Pichincha Guayas	0,42	0,07	31,48	1,00	0,00	1,52
V_3	Provincia: Azuay Imbabura, Zamora	-0,69	0,11	37,61	1,00	0,00	0,50
V_4	Provincia: Bolívar Cañar Pastza	-1,36	0,15	84,15	1,00	0,00	0,26
V_5	Provincia: Carchi Manabí; Moro. Sant; Napo; Orellano; Sto. Dom. Tsáchilas; Sta. Elena	-0,43	0,08	28,15	1,00	0,00	0,65
V_6	Provincia: Coptopaxi, Los Ríos	0,72	0,09	62,85	1,00	0,00	2,06
V_20	Edad mayor a 63 años	-0,49	0,16	9,61	1,00	0,00	0,61
V_28	Mujer entre 18 y 35 años	-0,19	0,09	4,07	1,00	0,04	0,83
V_30	Superior menor a 35 años	0,34	0,07	22,25	1,00	0,00	1,40
V_31	Area Rural	-0,57	0,11	27,32	1,00	0,00	0,57
P3_3	Edad	-0,01	0,00	9,05	1,00	0,00	0,99
Constant		-1,85	0,10	318,44	1,00	0,00	0,16

Como se observa en el cuadro precedente todos los coeficientes de las variables seleccionadas son estadísticamente diferentes de cero con nivel de significación menor del 0,05.

MULTICOLINEALIDAD

Kayser - Meyer - Olkin	
Kaiser-Meyer-Olkin Measure of Sampling	0,308

Para este análisis se obtuvo un KMO menor a 0,4 por lo que se concluye que la multicolinealidad es baja y no es un problema para la construcción del modelo.

AJUSTE GLOBAL DEL MODELO

Distribución de Buenos y Malos

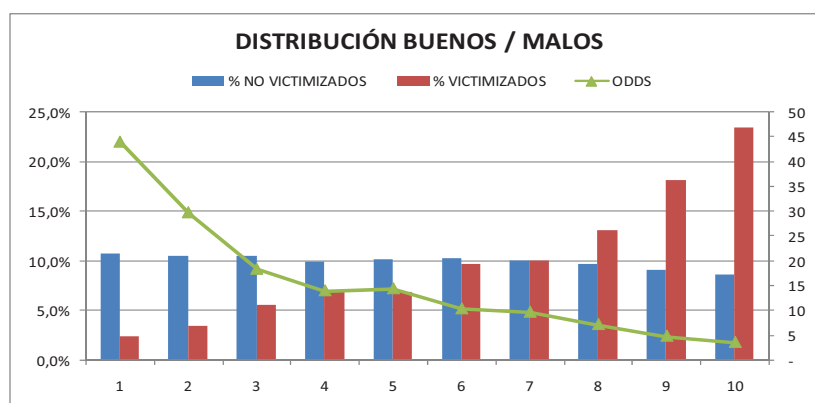
Cuadro 3.17: DISTRIBUCIÓN DE BUENOS Y MALOS

DECILES	DISTRIBUCIÓN BUENOS VS MALOS						KS	ODDS
	NO VICTIMIZADOS			VICTIMIZADOS				
	Frecuencia	% NO VICTIMIZADOS	% Acum. No Victimizados	Frecuencia	% VICTIMIZADOS	% Acum. Victimizados		
1	1627	10,8%	10,8%	37	2,42%	2,4%	0,08	44
2	1578	10,5%	21,4%	53	3,47%	5,9%	0,15	30
3	1582	10,5%	31,9%	86	5,63%	11,5%	0,20	18
4	1491	9,9%	41,8%	106	6,94%	18,5%	0,23	14
5	1535	10,2%	52,1%	106	6,94%	25,4%	0,27	14
6	1544	10,3%	62,3%	148	9,69%	35,1%	0,27	10
7	1513	10,1%	72,4%	155	10,15%	45,3%	0,27	10
8	1454	9,7%	82,1%	200	13,10%	58,3%	0,24	7
9	1374	9,2%	91,3%	278	18,21%	76,6%	0,15	5
10	1312	8,7%	100,0%	358	23,44%	100,0%	-	4
Total	15010	100,0%		1527	100,00%			10

Para medir el nivel de discriminación del modelo se divide la muestra de desarrollo en deciles y esta a su vez en dos grupos; en buenos, que para el estudio son los victimizados, y en malos o sea los no victimizados.

Mediante la división de los dos grupos se observa la distribución en cada uno de ellos en los deciles; se espera, como en efecto sucede, que para mayores deciles el porcentaje relativo de victimizados aumente mientras que disminuya para los no victimizados.

Gráfico 3.5: DISTRIBUCIÓN BUENOS Y MALOS



El porcentaje de victimizados comienza en el primer decil con 2,4% y aumenta a medida que crece el grupo decil, terminando en el último, con 23,44%, mostrando así un correcto ordenamiento. En los *odds* se puede observar que para el primer decil se tiene 44 no victimizados por cada victimizado, mientras que en el último la relación es de cuatro a uno.

Prueba de Hosmer y Lemeshow

La medida de ajuste global del modelo se calcula por el valor chi-cuadrado obtenido entre los valores observados y esperados en la muestra de desarrollo. La hipótesis a contrastar es que las distribuciones observada y esperada no son diferentes estadísticamente.

Cuadro 3.18: PRUEBA DE HOSMER LEMESHOW

Prueba de Hosmer y Lemeshow			
Step	Chi-square	df	Sig.
1	,000	0	.
2	,000	1	1,000
3	4,679	8	,791
4	8,771	8	,362
5	4,329	8	,826
6	3,927	8	,864
7	4,400	8	,819
8	4,402	8	,819
9	12,822	8	,118
10	10,416	8	,237

El valor observado chi-cuadrado es de 10,416, que proporciona un p-valor de 0,237, lo que lleva a aceptar la hipótesis nula de que las distribuciones no son diferentes estadísticamente. Abajo se presenta la tabla de contingencia:

Cuadro 3.19: TABLA DE CONTINGENCIA DE HOSMER LEMESHOW

Tabla de contingencia para la prueba de Hosmer y Lemeshow

		Índice de victimización a personas = No		Índice de victimización a personas = Si		Total
		Observed	Expected	Observed	Expected	
Step 10	1	1627	1629,599	41	38,401	1668
	2	1586	1582,676	59	62,324	1645
	3	1574	1570,561	82	85,439	1656
	4	1491	1510,777	122	102,223	1613
	5	1521	1523,818	127	124,182	1648
	6	1551	1529,444	120	141,556	1671
	7	1507	1492,847	158	172,153	1665
	8	1437	1437,924	204	203,076	1641
	9	1396	1409,434	270	256,566	1666
	10	1320	1322,919	344	341,081	1664

Estadístico Kolmogorov-Smirnov (estadístico K-S)

En el estadístico de Kolmogorov-Smirnov se contrastó la hipótesis de que las distribuciones de buenos y malos siguen distribuciones iguales sobre la variable de *Score*.

Cuadro 3.20: PRUEBA DE KOLMOGOROV SMIRNOV

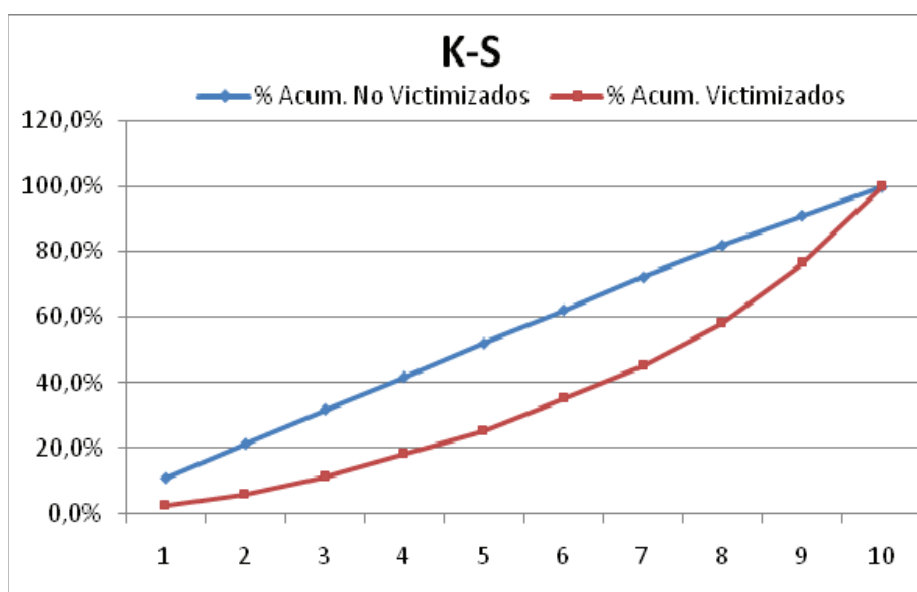
Prueba Kolmogorov - Smirnov

		Score Mod2
Most Extreme	Absolute	,266
Differences	Positive	,266
	Negative	,000
Kolmogorov-Smirnov Z		9,902
Asymp. Sig. (2-tailed)		,000

a. Grouping Variable: Índice de victimización a personas

Con un nivel de significación de 0,05 rechazamos la hipótesis de que las distribuciones de buenos y malos son equivalentes.

Gráfico 3.6: DISTRIBUCIÓN K-S



El KS del modelo alcanza una máxima distancia de 0,27, estadístico con el cual se rechazó la hipótesis de que la distribución de victimizados y no victimizados son equivalentes.

Coeficiente de Gini

Se calcula el coeficiente de Gini que es el área expresada en porcentaje que existe entre la función del score de victimización y la función de igualdad, que no discrimina a buenos y malos.

Cuadro 3.21: DISTRIBUCIÓN PARA CÁLCULO DE GINI

DECILES	DISTRIBUCIÓN BUENOS VS MALOS						GINI
	NO VICTIMIZADOS			VICTIMIZADOS			
	% NO	% Acum. No	% Acum. GINI	%	% Acum.	% Acum. GINI	
1	10,8%	10,8%	10,8%	2,42%	2,4%	2,4%	0,26%
2	10,5%	21,4%	10,5%	3,47%	5,9%	8,3%	0,87%
3	10,5%	31,9%	10,5%	5,63%	11,5%	17,4%	1,84%
4	9,9%	41,8%	9,9%	6,94%	18,5%	30,0%	2,98%
5	10,2%	52,1%	10,2%	6,94%	25,4%	43,9%	4,49%
6	10,3%	62,3%	10,3%	9,69%	35,1%	60,5%	6,22%
7	10,1%	72,4%	10,1%	10,15%	45,3%	80,4%	8,10%
8	9,7%	82,1%	9,7%	13,10%	58,3%	103,6%	10,04%
9	9,2%	91,3%	9,2%	18,21%	76,6%	134,9%	12,35%
10	8,7%	100,0%	8,7%	23,44%	100,0%	176,6%	15,43%
Total general	100,0%			100,00%			37,42%

El coeficiente de GINI alcanza un área de discriminación de 37,42%.

Este coeficiente al igual que el estadístico de Kolmogorov-Smirnov nos dará medidas de discriminación del modelo en área (Gini) y en Distancia (K-S) con la cual compararemos la muestra de prueba.

VALIDACIÓN DEL MODELO

Una vez definido el modelo final se realiza la validación del mismo, con base en la muestra de prueba. Esta se seleccionó como un 30% de la base original, de manera aleatoria, y no fue utilizada para establecer el modelo.

ANÁLISIS ESTADÍSTICO

Es indispensable evaluar la consistencia de cada una de las variables para lo cual se realiza el análisis estadístico de cada una de ellas:

Cuadro 3.22: ANÁLISIS ESTADÍSTICO VARIABLES EN EL MODELO

		NO VICTIMIZADOS	VICTIMIZADOS	ODDS				
		6422	682	0,11				
Variable	Descripción	B	Exp(B)	SIG TEST	VICTIMIZADOS	NO VICTIMIZADOS	ODDS	RELEVANCIA
V_2	Provincia: Pichincha Guayas	0,418	1,519	0,000	205	1150	0,178	19%
V_3	Provincia: Azuay Imbabura, Zamora	-0,691	0,501	0,000	43	745	0,058	11%
V_4	Provincia: Bolívar Cañar Pastza	-1,359	0,257	0,000	16	785	0,020	11%
V_5	Provincia: Carchi Manabí; Moro. Sant; Napo; Orellano; Sto. Dom. Tsachilas; Sta. Elena	-0,428	0,652	0,039	163	1772	0,092	27%
V_6	Provincia: Coptopaxi, Los Ríos	0,721	2,057	0,000	82	422	0,194	7%
V_20	Edad mayor a 63 años	-0,493	0,611	0,003	50	712	0,070	11%
V_28	Mujer entre 18 y 35 años	-0,189	0,828	0,000	60	1045	0,057	16%
V_30	Educación Superior menor a 35 años	0,338	1,402	0,000	156	936	0,167	15%
V_31	Indicador Area Rural	-0,569	0,566	0,000	46	773	0,060	12%
P3_3	Edad	-0,008	0,992					

Significación de las variables

Para cada una de las variables artificiales (*dummy*) seleccionadas se realiza un análisis de correspondencia entre la variable independiente y la variable dependiente para la muestra de prueba. En esta etapa de la modelación se encontraron variables *dummy* que aunque fueron seleccionadas en el desarrollo del modelo no resultaron ser significativas, a partir de las pruebas de hipótesis correspondientes, por lo que considerarlas ocasionaría un error de estimación y de interpretación del *score* de victimización; por esta razón se volvió a la fase de desarrollo en la cual se retiró dicha variable y así hasta llegar al modelo final. Se seleccionaron solamente las variables *dummy* que resultaron significativas en la muestra de prueba. Adicionalmente, se validó que las variables seleccionadas sean representativas; es decir, que contengan al menos el 5% de la muestra de prueba.

Polaridad de las variables

Una vez que las variables significativas sean seleccionadas puede ser que no estén correctamente polarizadas en la muestra de prueba; es decir, se desea ver las variables que en el modelo tengan coeficientes positivos tengan *odds* mayores al *odds* de la muestra de prueba y si tienen coeficiente negativo tengan *odds* menores al *odds* de la muestra de prueba. Que el *odds* de la variable explicativa en la muestra de prueba no tenga relación con el signo del coeficiente indicaría que aunque la variable sea significativa para explicar la variable objetivo en las dos muestras, de desarrollo y de prueba, el sentido en que la variable explica es opuesto en cada una de ellas, por lo que la variable podría considerarse como inestable y debería ser retirada del modelo.

En la base de prueba se puede observar que todas las variables están alineadas; es decir, la relación del *odds* de cada variable con el *odds* de la muestra tiene relación con el signo del coeficiente de cada variable.

Para llegar a este resultado se realizaron varios modelos en los cuales se fueron excluyendo variables que resultaron significativas pero no estaban alineadas en la base de prueba.

Luego de realizar las pruebas respectivas de significación y polaridad de las variables¹³, entre modelos obtenidos con la muestra de desarrollo y con la muestra de prueba, se retuvieron las variables que se muestran en el cuadro precedente y que conforman el modelo final.

AJUSTE DEL MODELO

Una vez que el modelo seleccionado considera variables significativas y alineadas en la muestra de prueba, se debe medir como se ajusta el modelo a la muestra de prueba estableciendo este como el escenario más probable de discriminación.

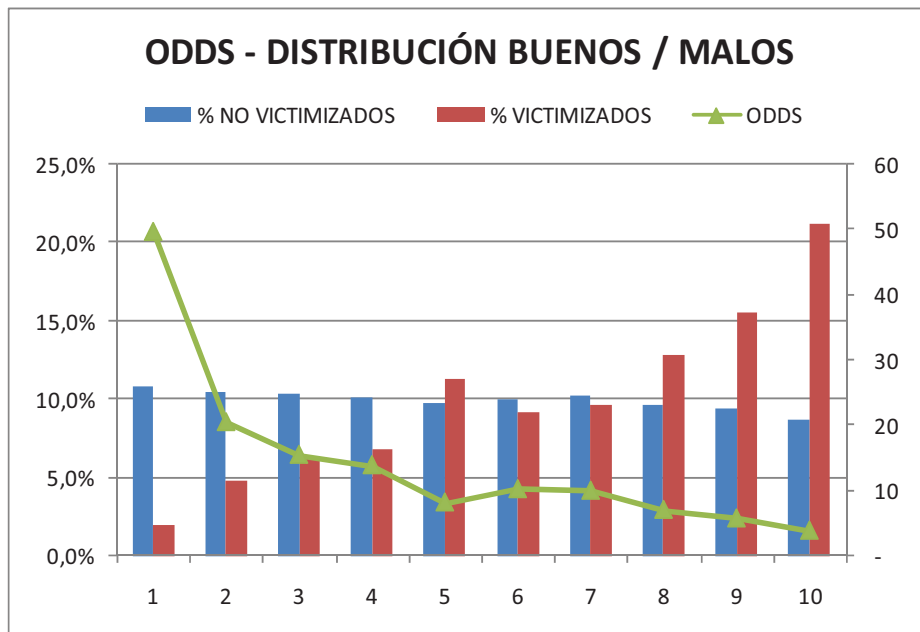
Distribución de Buenos y Malos

Cuadro 3.23: ANÁLISIS DE DISCRIMINACIÓN K-S Y ODDS

DECILES	DISTRIBUCIÓN BUENOS VS MALOS						KS	ODDS
	NO VICTIMIZADOS			VICTIMIZADOS				
	Frecuencia	% NO VICTIMIZADOS	% Acum. No Victimizados	Frecuencia	% VICTIMIZADOS	% Acum. Victimizados		
1	696	10,8%	10,8%	14	2,05%	2,1%	0,09	50
2	677	10,5%	21,4%	33	4,84%	6,9%	0,14	21
3	667	10,4%	31,8%	43	6,30%	13,2%	0,19	16
4	654	10,2%	41,9%	47	6,89%	20,1%	0,22	14
5	631	9,8%	51,8%	77	11,29%	31,4%	0,20	8
6	645	10,0%	61,8%	63	9,24%	40,6%	0,21	10
7	660	10,3%	72,1%	66	9,68%	50,3%	0,22	10
8	622	9,7%	81,8%	88	12,90%	63,2%	0,19	7
9	609	9,5%	91,3%	106	15,54%	78,7%	0,13	6
10	561	8,7%	100,0%	145	21,26%	100,0%	-	4
Total	6422	100,0%		682	100,0%			9

¹³ Detalles de este análisis se muestran en el Anexo 2.

Gráfico 3.7: ODDS - DISTRIBUCIÓN DE BUENOS Y MALOS



En la distribución de buenos y malos para cada uno de los deciles se observa que el modelo ordena bien a los buenos y malos en la muestra de prueba, pues a mayores deciles corresponden un mayor porcentaje de victimizados.

Estadísticos de Kolmogorov - Smirnov

Al igual que en la muestra de desarrollo se prueba si las distribuciones de buenos y malos no son equivalentes en la muestra de prueba.

Cuadro 3.24: PRUEBA DE KOLMOGOROV - SMIRNOV

Prueba Kolmogorov-Smirnov ^a		Score Mod2
Most Extreme Differences	Absolute	,223
	Positive	,223
	Negative	,000
Kolmogorov-Smirnov Z		5,538
Asymp. Sig. (2-tailed)		,000

a. Grouping Variable: Índice de victimización a personas

Con una significancia menor a 0.05 rechazamos la hipótesis nula de que las distribuciones de buenos y malos sean equivalentes por lo que el modelo discrimina las distribuciones de buenos y malos en la muestra de prueba. La distancia máxima entre la distribución de buenos y malos fue de 0,223, muy cercana a la obtenida en la muestra de desarrollo: 0,266. Es normal que la discriminación disminuya en las muestras de prueba ya que son más pequeñas que las bases de desarrollo.

Coeficiente de Gini

El coeficiente de Gini, como se ha mencionado anteriormente, mide el área entre dos distribuciones, la de equidad o no discriminativa y la distribución objeto de estudio.

Cuadro 3.25: DISTRIBUCIÓN PARA CÁLCULO DE GINI

DECILES	DISTRIBUCIÓN BUENOS VS MALOS						GINI
	NO VICTIMIZADOS			VICTIMIZADOS			
	% NO VICTIMIZADOS	% Acum. No Victimizados	% Acum. GINI No Victimizados	% VICTIMIZADOS	% Acum. Victimizados	% Acum. GINI Victimizados	
1	10,8%	10,8%	10,8%	2,42%	2,1%	2,1%	0,22%
2	10,5%	21,4%	10,5%	3,47%	6,9%	8,9%	0,94%
3	10,4%	31,8%	10,4%	5,63%	13,2%	20,1%	2,09%
4	10,2%	41,9%	10,2%	6,94%	20,1%	33,3%	3,39%
5	9,8%	51,8%	9,8%	6,94%	31,4%	51,5%	5,06%
6	10,0%	61,8%	10,0%	9,69%	40,6%	72,0%	7,23%
7	10,3%	72,1%	10,3%	10,15%	50,3%	90,9%	9,34%
8	9,7%	81,8%	9,7%	13,10%	63,2%	113,5%	10,99%
9	9,5%	91,3%	9,5%	18,21%	78,7%	141,9%	13,46%
10	8,7%	100,0%	8,7%	23,44%	100,0%	178,7%	15,61%
Total general	100,0%			100,00%			31,66%

El coeficiente de GINI es de 31,66% menor al de desarrollo 37,42%, pero no muy diferente de ésta.

Una vez aprobadas todas las pruebas estadísticas aplicadas al modelo, se procede a utilizarlo para la creación de perfiles de victimización a nivel nacional.

INTERPRETACIÓN DEL MODELO

Como se mencionó anteriormente, un modelo de regresión logística obedece a dos posibles objetivos: la interpretación y la predicción. En este estudio, considerando que la información fue obtenida mediante una encuesta, fundamentalmente con datos cualitativos, se ha puesto énfasis en la interpretación, buscando así encontrar patrones que permitan focalizar y diferenciar la gestión de acuerdo a la exposición a la victimización.

Para encontrar los perfiles de acuerdo a la exposición a la victimización cuantificada en el modelo de *score*¹⁴, presentado en una escala de cero a mil, es importante estudiar la incidencia de cada una de las variables del modelo en diferentes grupos de victimización.

Se ha dividido la muestra de prueba en quintiles sobre la cual se definirán perfiles¹⁵:

Cuadro 3.26: DISTRIBUCIÓN SCORE POR QUINTILES

	Rango Puntaje Score	% No Victimizado	% Vicitimizado
QUINTIL 1	- 46,10	96,69%	3,31%
QUINTIL 2	46,11 - 70,62	93,62%	6,38%
QUINTIL 3	70,63 - 97,71	90,11%	9,89%
QUINTIL 4	97,72 - 143,80	89,28%	10,72%
QUINTIL 5	143,81 -	82,34%	17,66%

A partir del porcentaje de victimizados para cada quintil, se observa que los quintiles uno y dos, entre 0 y 70,62 puntos en el puntaje *score*, tienen un nivel de victimización menor a 7%; los quintiles 2 y 3, entre 70,63 y 97,71 puntos en el puntaje *score*, tienen un porcentaje de victimización similar, alrededor del 10%, y el último quintil, mayor o igual a 143,81 puntos en el puntaje *score*, tiene el porcentaje de victimización más alto con el 17,66%. En función del porcentaje de victimización de cada quintil se han agrupado a los quintiles uno con el dos y el

¹⁴ Un modelo de *score* presenta la probabilidad generalmente en una escala de 0 a 1000 y no de 0 a 1.

¹⁵ Se generan estrategias sobre la muestra de prueba ya que se considera a esta como el escenario más probable y pesimista al no formar parte de la muestra de construcción del modelo.

tres con cuatro, obteniendo los siguientes perfiles referenciados por el puntaje de score.

Cuadro 3.27: PERFILES DEFINIDOS POR PUNTAJE DE SCORE

	Rango Puntaje Score	% No Victimizado	% Vicitimizado
GRUPO 1	- 70,6	95,16%	4,84%
GRUPO 2	70,9 - 143,	89,69%	10,31%
GRUPO 3	144, -	82,34%	17,66%

Para definir el perfil de cada grupo se estudia la incidencia de cada una de las variables en el mismo; para esto, se considera la matriz de frecuencias entre los grupos y las variables seleccionadas en el modelo:

Cuadro 3.28: FRECUENCIAS VARIABLES POR GRUPO PERFIL

	V2	V3	V4	V5	V6	V20	V28	V30	V31	TOTAL
GRUPO 1	37	670	801	1010	10	584	673	166	643	2831
GRUPO 2	567	118		925	111	178	385	362	172	2852
GRUPO 3	751				383		47	564	4	1421
TOTAL	1355	788	801	1935	504	762	1105	1092	819	7104

Cuadro 3.29: PORCENTAJE DE INCIDENCIA DE VARIABLES POR GRUPO PERFIL

	V2	V3	V4	V5	V6	V20	V28	V30	V31
GRUPO 1	1%	24%	28%	36%	0%	21%	24%	6%	23%
GRUPO 2	20%	4%	0%	32%	4%	6%	13%	13%	6%
GRUPO 3	53%	0%	0%	0%	27%	0%	3%	40%	0%
TOTAL	19%	11%	11%	27%	7%	11%	16%	15%	12%

Observando los perfiles de cada grupo y comparando la incidencia que tiene cada variable entre el grupo y la población, se establecen los perfiles de cada grupo.

Cuadro 3.30: DESCRIPCIÓN DE GRUPO PERFIL

PERFIL	VARIABLE	DETALLE
GRUPO 3	Provincia:	Pichincha, Guayaquil Cotopaxi, Los Ríos
	Educación y edad:	Educación Superior y edad menor a 35 años
GRUPO 2	Provincia:	Provincias: Carchi Manabí; Moro. Sant; Napo; Orellano; Sto. Dom. Tsáchilas; Sta. Elena
	Provincia-Edad-Género	Si son de las provincias del grupo 3 son de la tercera edad o mujeres entre 18 y 35 años
GRUPO 1	Provincia:	Azuay, Imbabura, Zamora Bolívar, Cañar, Pastza
	Edad	Edad mayor a 63 años
	Edad y Género	Mujer entre 18 y 35 años
	Area	Area Rural

Como se puede observar, las variables seleccionadas por el modelo son la provincia, el área rural o urbana, edad y género. La provincia resultó ser la más importante, contribuyendo casi con todas sus variables indicadoras, mientras que las demás contribuyen con el modelo con una de sus variables indicadoras representando solo una parte de la variable original. Este hecho se presenta debido a que al realizarse un score de victimización a nivel nacional, las variables seleccionadas son aquellas estadísticamente significativas a nivel nacional; por tanto, la variable provincia es la más importante pues la realidad socio-económica de cada provincia es diferente. Se debería realizar un análisis como el presentado en este capítulo para las provincias grandes, de manera que se puedan evidenciar factores socioeconómicos más específicos que discriminen el tipo de victimización por lugar, edad, género o situación económica, que permitan definir acciones específicas propias de cada provincia.

El modelo de este capítulo, como se ha visto en las páginas anteriores, ordena y discrimina la victimización considerando la información más relevante a nivel nacional; de este modo el grupo tres tiene el mayor riesgo de victimización, en el cual se evidencia que las personas de educación superior y de edad menor a 35 años tienen un grado de exposición o preferencia por parte de la delincuencia

superior al resto de la población. Las provincias de Pichincha y Guayas, las dos más grandes del Ecuador, se encuentran en el grupo de mayor riesgo de victimización. A continuación se describe un detalle de cada grupo:

GRUPO 1:

El grupo uno está conformado por tres subgrupos: el primero, los individuos de las provincias de Azuay, Imbabura, Zamora, Bolívar, Cañar y Pastaza. Estas provincias tienen porcentajes de victimización bajas, menores al 6%, en relación con el índice de victimización nacional que es del 12,7%. En el segundo, individuos de la tercera edad o mujeres de 18 a 35 años que pertenecen a las provincias: Carchi, Manabí, Morona Santiago, Napo, Orellana, Santo Domingo de los Tsáchilas y Santa Elena. Finalmente en el tercero se encuentran los individuos que viven en áreas rurales que no son de la provincia de Pichincha y Guayas. Este grupo tiene un índice de victimización del 4,84%, mostrando que no tiene niveles altos de victimización. En el caso de las provincias de Carchi, Manabí, Morona Santiago, Napo, Orellana, Santo Domingo de los Tsáchilas y Santa Elena puede verse que el comportamiento de las mujeres e individuos de la tercera edad los mantiene en riesgo de victimización bajo. Los que pertenecen a áreas rurales mantienen un nivel bajo de victimización debido a la realidad comercial y económica que no es tentativa para la delincuencia.

GRUPO 2:

Considera dos grupos: los que pertenecen a las provincias de: Carchi, Manabí, Morona Santiago, Napo, Orellana, Santo Domingo de los Tsáchilas y Santa Elena, que tienen educación superior y son menores a treinta años; el segundo subgrupo que considera a los individuos de las provincias de Pichincha y Guayas, que son de la tercera edad, mujeres de treinta y cinco años o viven en áreas rurales. El porcentaje de victimización de este grupo es de 10,31%, por lo que

puede verse que existe una notable diferencia en la exposición al peligro en los individuos hombres menores a treinta y cinco años, de educación superior al resto de individuos, ya que éstos tienen aproximadamente cuatro puntos porcentuales más de exposición a la victimización que aquellos que viven en las mismas provincias con otras características.

GRUPO 3

El grupo 3 está conformado por tres subgrupos: el primero, que está en las provincias de Pichincha o Guayas y no son de la tercera edad, ni viven en áreas rurales, ni son mujeres de edad menor a 35 años; el segundo: Cotopaxi y Los Ríos; el tercero es el conformado por los que tienen estudios superiores con edad menor a 35 años y que no viven en las provincias agrupadas en ninguno de los grupos antes mencionados. Pichincha y Guayas tienen un porcentaje de victimización de 14,7%, mientras que Cotopaxi y los Ríos presentan niveles de 16,7% y 19,1%, respectivamente, siendo los más altos en todo el país.

CAPITULO 4

PERFILES DE VICTIMIZACIÓN

La heterogeneidad de la población constituye la materia prima de un análisis cuantitativo y cualitativo; sin embargo, el estudio del individuo u objeto particular aislado resulta un elemento de heterogeneidad demasiado pequeño con relación al objetivo de análisis, por lo cual se trata entonces de agrupar a los individuos con el objetivo de hacer el análisis sobre los grupos y no sobre cada uno de los individuos, utilizando técnicas estadísticas adecuadas.

Las técnicas de agrupación recurren a cálculos algorítmicos que realizan operaciones recursivas y repetitivas que pueden considerar una jerarquía en el número de agrupaciones ascendente/descendente o considerar un número determinado de grupos sobre el cual se hace la clasificación de cada uno de los individuos.

En general se puede decir que para un conjunto de información para varios individuos se puede medir la asociación que existe entre ellos a través de la similitud o distancia que exista entre sus variables asociadas.

Para realizar un análisis de agrupamiento se debe considerar que el análisis y resultado del modelo sirven solo para ese diseño y que éste se caracteriza por la elección de individuos, las variables utilizadas, criterio de asociación utilizado y nivel de agrupación elegido; todas las especificidades definen diferentes soluciones y, por tanto, la solución final depende del interés particular del investigador.

Los análisis discriminante y de agrupación son diferentes, pues el primero intenta explicar una estructura definida a priori mientras el segundo intenta determinarla.

En el presente capítulo se intenta encontrar conglomerados para las ciudades de Quito y Guayaquil, que permitan identificar localmente perfiles de exposición a la victimización sin considerar alguna estructura *a-priori*.

Adicionalmente, al análisis de conglomerados, se estudiará la victimización en hogares y la relación que ésta pueda tener con la victimización a personas.

Como una aproximación a las ciudades de Quito y Guayaquil se han considerado las cabeceras cantonales de estas ciudades, de acuerdo a la división política-administrativa que entrega el INEC.

ETAPAS DE DESARROLLO PARA UN ANÁLISIS DE AGRUPAMIENTO

1. SELECCIÓN DE LA MUESTRA DE DATOS
2. SELECCIÓN Y TRANSFORMACIÓN DE LAS VARIABLES A UTILIZAR
3. SELECCIÓN DE DISTANCIA O SIMILITUD Y CÁLCULO DE LA MISMA
4. SELECCIÓN DEL CRITERIO DE AGRUPACIÓN
5. DETERMINACIÓN DE LA ESTRUCTURA CORRECTA

1. SELECCIÓN DE LA MUESTRA

Se debe adecuar al máximo la muestra que será objeto de análisis, considerando los valores en blanco y atípicos. Este aspecto ya se trató en el capítulo precedente.

2. SELECCIÓN DE VARIABLES

No se deben elegir variables indiscriminadamente. El grupo de variables seleccionadas representan una determinada estructura; la inclusión de una variable no relevante para el análisis podría definir una estructura de agrupación diferente que puede no estar alineada con el objetivo. En este estudio el objetivo es determinar la victimización de los individuos de Quito y Guayaquil de

acuerdo su perfil social. Luego de los resultados analizados en los dos capítulos precedentes, a continuación se presentan las variables utilizadas:

Cuadro 4.1: VARIABLES CONSIDERADAS

Nombre	Tipo Variable
RANGO DE EDAD	Categórica
GÉNERO	Categórica
ESTADO CIVIL	Categórica
NIVEL DE ESTUDIOS	Categórica
OCUPACIÓN	Categórica

3. SELECCIÓN DEL CRITERIO DE AGRUPACIÓN

La técnica utilizada para determinar los conglomerados es la de agrupación en dos fases o bietápica. Este método puede trabajar con variables continuas o categóricas, con grandes cantidades de información, y se realiza en dos etapas:

La primera, construye un árbol de características en la cual se agrupan los individuos en pequeños conglomerados disminuyendo de esta manera significativamente el tamaño de la base.

La segunda, agrupa los pequeños conglomerados (de la primera etapa) con la técnica aglomerativa jerárquica tratando a los pequeños conglomerados como individuos.

4. SELECCIÓN DE DISTANCIA O SIMILITUD Y CÁLCULO DE LA MISMA

La distancia utilizada para este análisis es la distancia log-verosimilitud debido a que las variables son categóricas. A continuación se especifica el cálculo de la distancia utilizada:

Notaciones:

k^A Número total de variables continuas

k^B Número total de variables categóricas

L_k Número de categorías de la k-ésima variable

R_k Rango de la k-ésima variable continua

N Número de datos

N_k Número de datos en el conglomerado k.

$\hat{\mu}_k$ Media estimada de la k-ésima variable continua.

$\hat{\sigma}_k^2$ Es la varianza estimada de la k-ésima variable continua

$\hat{\mu}_{jk}$ Es la media estimada de la k-ésima variable continua en el conglomerado j.

$\hat{\sigma}_{kj}^2$ Es la varianza estimada de la k-ésima variable continua en el conglomerado j.

N_{jkl} Número de casos en el conglomerado j cuya variable categórica k toma la categoría l.

N_{kl} Número de casos en la categoría l de la variable categórica k.

$d(j,s)$ Distancia entre los conglomerados j y s.

$\langle j,s \rangle$ Índice que muestra el nuevo conglomerado formado por la combinación de los conglomerados j y s.

La distancia log-verosimilitud que se utilizará, se define de la siguiente manera:

$$d(i,j) = \xi_i + \xi_j - \xi_{\langle i,j \rangle}$$

donde:

$$\xi_v = -N_v \left(\sum_{k=1}^{K^A} \frac{1}{2} \log \left(\hat{\sigma}_k^2 + \hat{\sigma}_{vk}^2 \right) + \sum_{k=1}^{K^B} \hat{E}_{vk} \right)$$

ε_v puede interpretarse como una medida de dispersión dentro del *cluster* v .

ε_v consiste de dos partes;

El primer sumando de la ecuación anterior, mide la dispersión de las variables continuas dentro del conglomerado v . El segundo sumando es una medida de entropía utilizada para medir la dispersión de las variables categóricas, donde;

$$\hat{E}_{vk} = - \sum_{l=1}^{L_k} \frac{N_{vkl}}{N_v} \log \frac{N_{vkl}}{N_v}$$

5. DETERMINACIÓN DE LA ESTRUCTURA CORRECTA

El número adecuado de conglomerados para la segunda etapa se obtiene mediante el Criterio Bayesiano de Schwarz (BIC), que consiste en el siguiente análisis:

Primero se realiza una serie de particiones sobre los conglomerados pequeños definidos en la primera fase del conglomerado bietápico, utilizando la técnica jerárquica aglomerativa. Para cada número de conglomerados J se calcula el BIC de la siguiente manera:

$$\text{BIC}(J) = -2 \sum_{j=1}^J \xi_j + m_j \log(N)$$

Donde:

$$m_j = J \left\{ 2K^A + \sum_{k=1}^{K^B} (L_{K^B} - 1) \right\}$$

El máximo número de conglomerados K se obtiene mediante la razón entre el $BIC(J)$ y el $BIC(1)$, para el primer valor de J , con el cual se obtenga un cociente menor a $0,04$ ¹⁶, valor obtenido en simulaciones realizadas por los técnicos de SPSS. Una vez que se tiene el máximo número posible de conglomerados K , el objetivo es encontrar el número óptimo de conglomerados (de acuerdo a este modelo de agrupación). Para esto se utiliza la razón de cambio $R_1(J)$ para el BIC en todos los J conglomerados ($J=1, \dots, K$) definida como:

$$R_1(J) = \frac{dBIC(J)}{dBIC(1)}$$

donde:

$$dBIC(J) = BIC(J) - BIC(J+1)$$

Representa la distancia cuando los $J+1$ conglomerados pasan a formar J conglomerados.

El número de conglomerados final se obtiene con el cociente entre las dos razones de cambio más grandes $R_1(\cdot)$, definidas como $R_1(J_1)$ y $R_1(J_2)$. Cuando el cociente $R_1(J_1)/R_1(J_2)$ es mayor que $1,15$ ¹⁷ el número de conglomerados es igual a J_1 de otro modo el número de conglomerados es igual al máximo entre J_1 y J_2 ¹⁸.

¹⁶ Este valor está basado en estudios de simulación de los Autores de SPSS *Two Step Conglomeradoing*. Mayores detalles en el documento SPSS TWO STEP CONGLOMERADO – A FIRST EVALUATION by Johann Bacher, Knut Wensing, Melanie, 2004-02.

¹⁷ Idem referencia 16.

¹⁸ Mayores detalles documento en la referencia 16.

AGRUPAMIENTO PARA QUITO

ANÁLISIS DESCRIPTIVO DE QUITO

La provincia de Pichincha es una de la más grandes del País cuyo índice de victimización a personas es de 14,7%, victimización a hogares del 21% y con una percepción de inseguridad del 44%; este alto porcentaje de percepción de inseguridad se debe también a la gran difusión de delitos que son noticia de titulares en los diferentes medios de comunicación.

CONGLOMERADOS EN QUITO

A continuación se muestra el cuadro resumen del BIC con el cual se define el número seleccionado de grupos.

Cuadro 4.2: CRITERIO BAYESIANO DE SCHWARZ

Auto-Clustering				
Number of Clusters	Schwarz's Bayesian Criterion (BIC)	BIC Change ^a	Ratio of BIC Changes ^b	Ratio of Distance Measures ^c
1	21614,216			
2	18484,643	-3129,573	1,000	1,861
3	16868,499	-1616,144	,516	1,420
4	15772,549	-1095,950	,350	1,042
5	14726,030	-1046,519	,334	1,868
6	14231,628	-494,403	,158	1,228
7	13855,500	-376,128	,120	1,021
8	13490,005	-365,495	,117	1,034
9	13141,403	-348,602	,111	1,014
10	12799,735	-341,668	,109	1,289
11	12566,494	-233,241	,075	1,110
12	12370,517	-195,977	,063	1,045
13	12189,103	-181,414	,058	1,019
14	12013,815	-175,288	,056	1,037
15	11849,823	-163,992	,052	1,029

a. The changes are from the previous number of clusters in the table.

b. The ratios of changes are relative to the change for the two cluster solution.

c. The ratios of distance measures are based on the current number of clusters against the previous number of clusters.

La técnica del conglomerado a dos fases seleccionó 5 grupos mediante el criterio Bayesiano de Schwarz. A continuación se describe la distribución de cada una de las variables para cada conglomerado.

Gráfico 4.1: DISTRIBUCIÓN PORCENTUAL DE GRUPO OCUPACIÓN POR CONGLOMERADO

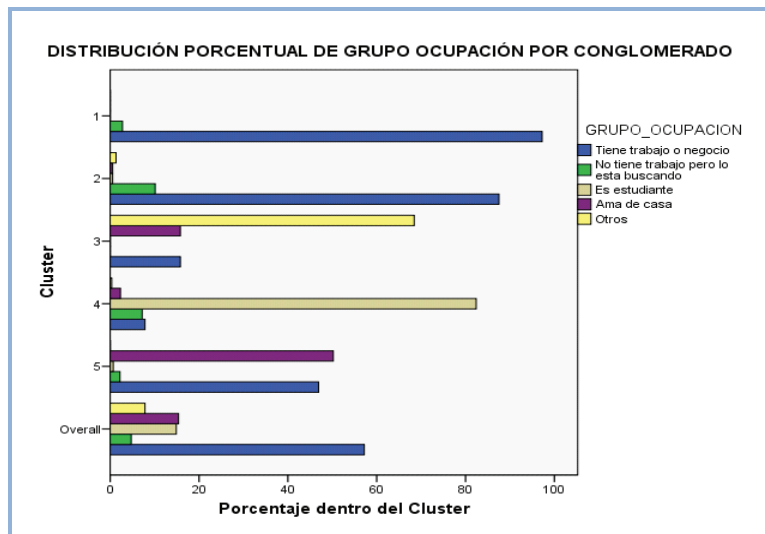


Gráfico 4.2: DISTRIBUCIÓN PORCENTUAL DEL RANGO DE EDAD POR CONGLOMERADO

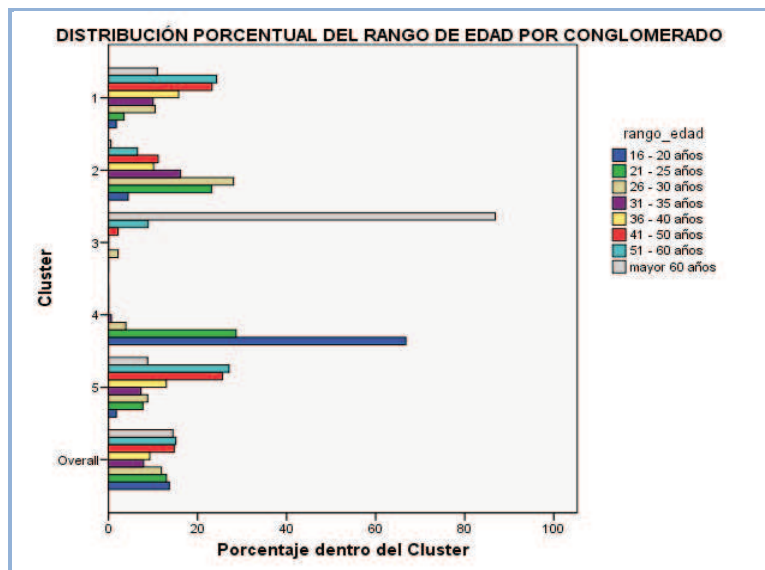


Gráfico 4.3: DISTRIBUCIÓN PORCENTUAL DEL GENERO POR CONGLOMERADO

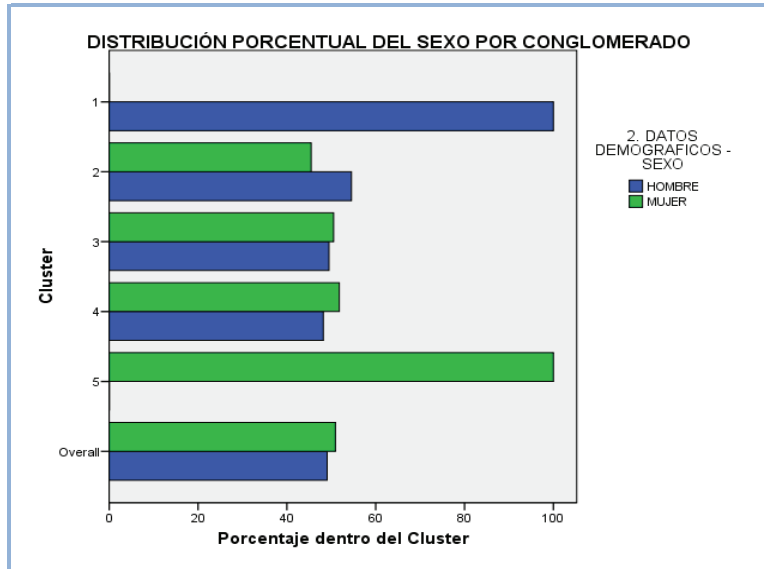


Gráfico 4.4: DISTRIBUCIÓN PORCENTUAL DEL ESTADO CIVIL POR CONGLOMERADO

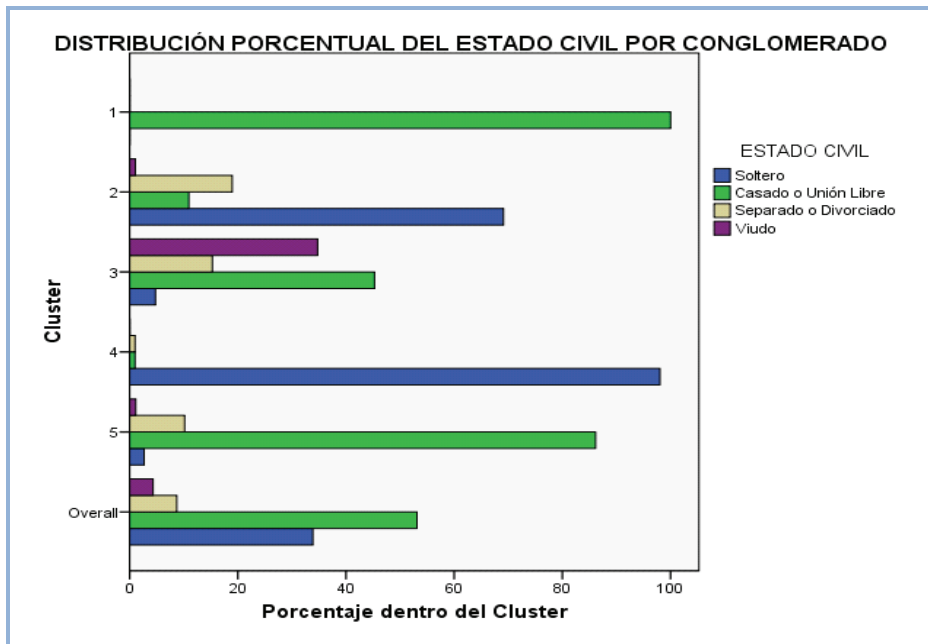
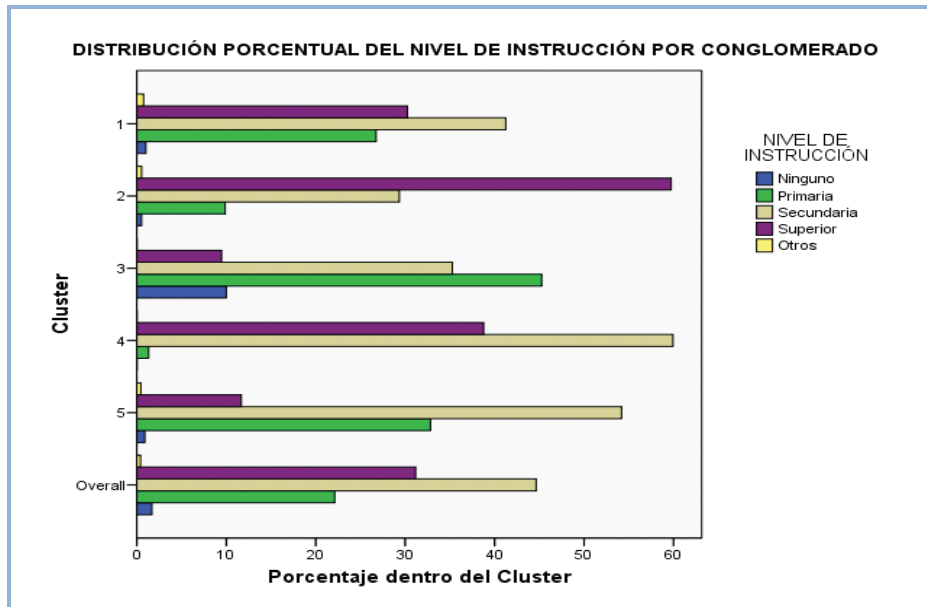


Gráfico 4.5: DISTRIBUCIÓN PORCENTUAL DEL NIVEL DE INSTRUCCIÓN POR CONGLOMERADO



La interpretabilidad de cada conglomerado viene en función de la importancia de cada variable para la formación del conglomerado o grupo. A continuación se muestra la incidencia de cada variable en la formación de cada grupo; las barras superiores a la línea segmentada en azul muestran que las variables son representativas para el grupo (se contrasta la hipótesis nula H_0 : igualdad de la distribución en el conglomerado y la de la población):

Gráfico 4.6: PRUEBA CHÍ - CUADRADO GRUPO OCUPACIÓN

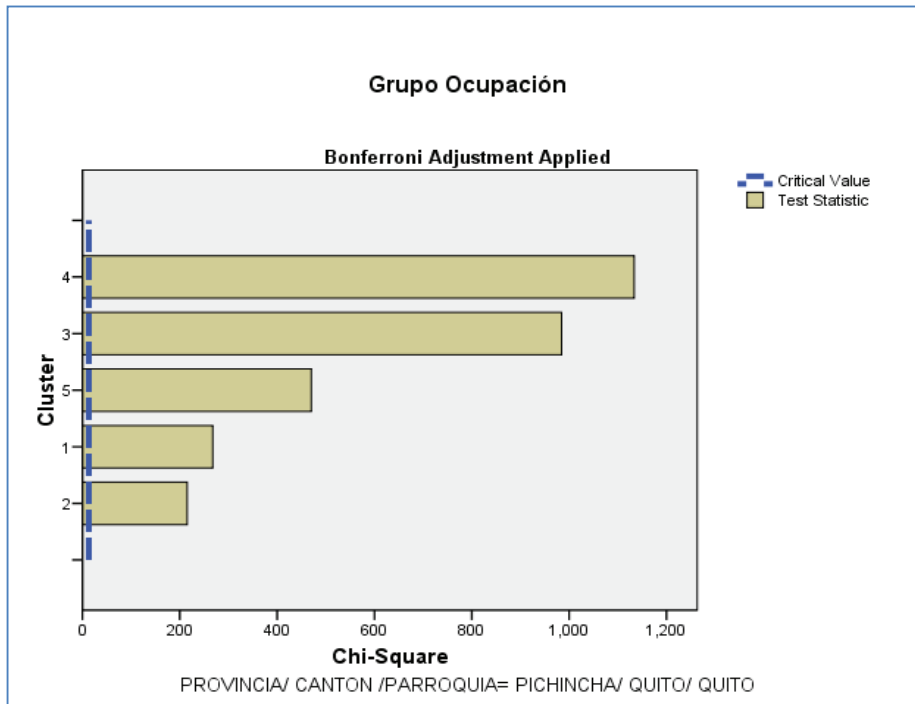


Gráfico 4.7: PRUEBA CHÍ - CUADRADO RANGO DE EDAD

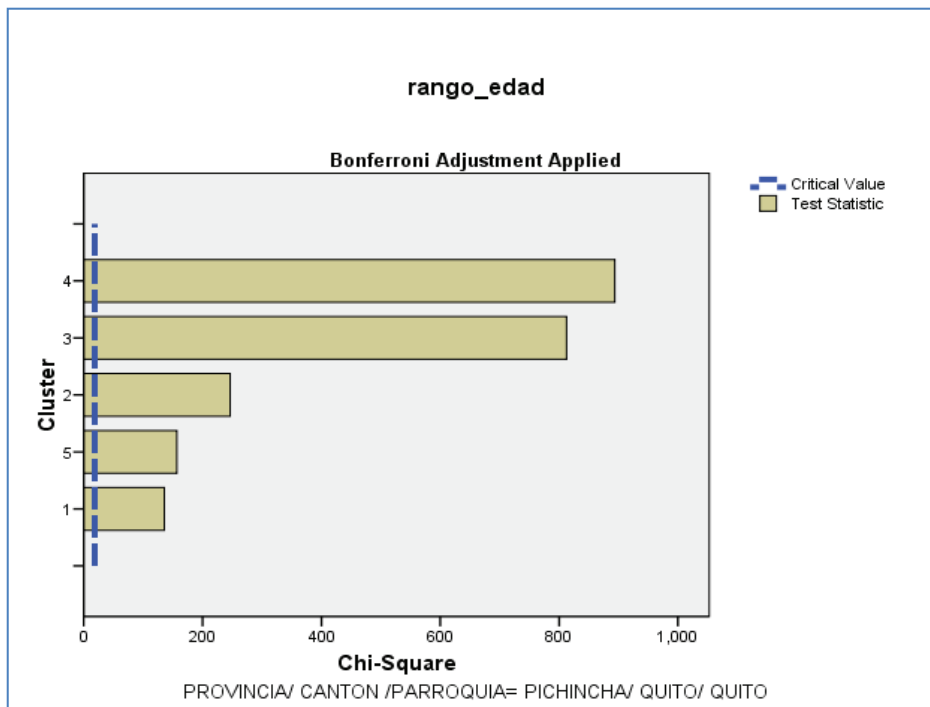


Gráfico 4.8: PRUEBA CHI - CUADRADO GÉNERO

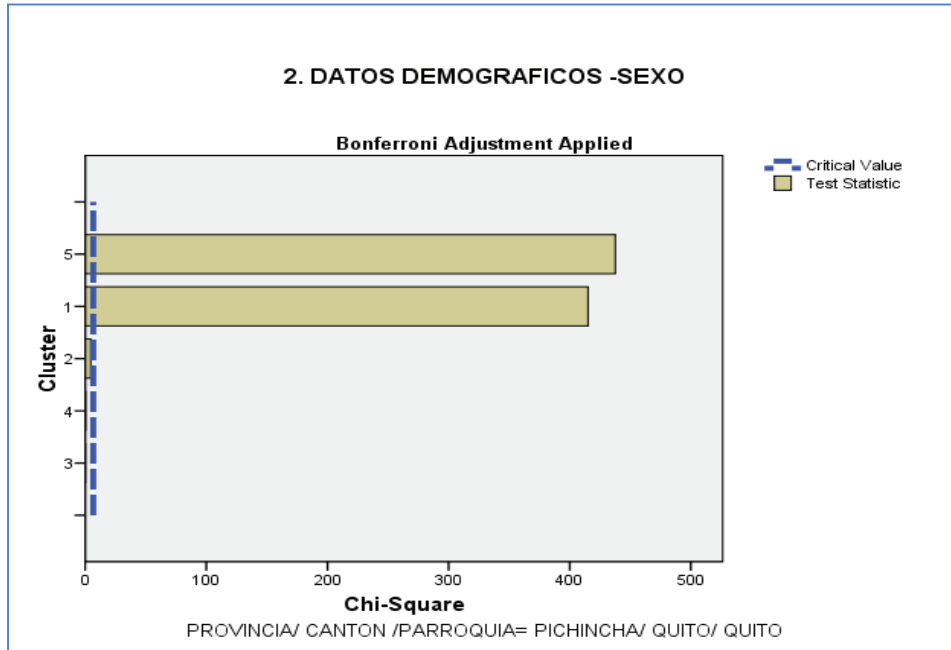


Gráfico 4.9: PRUEBA CHÍ - CUADRADO ESTADO CIVIL

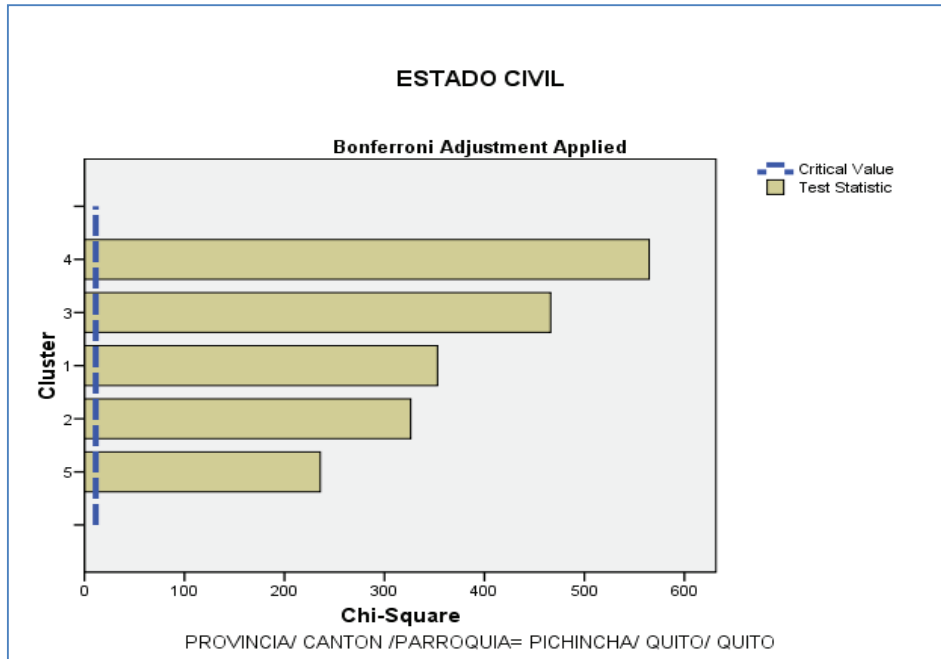
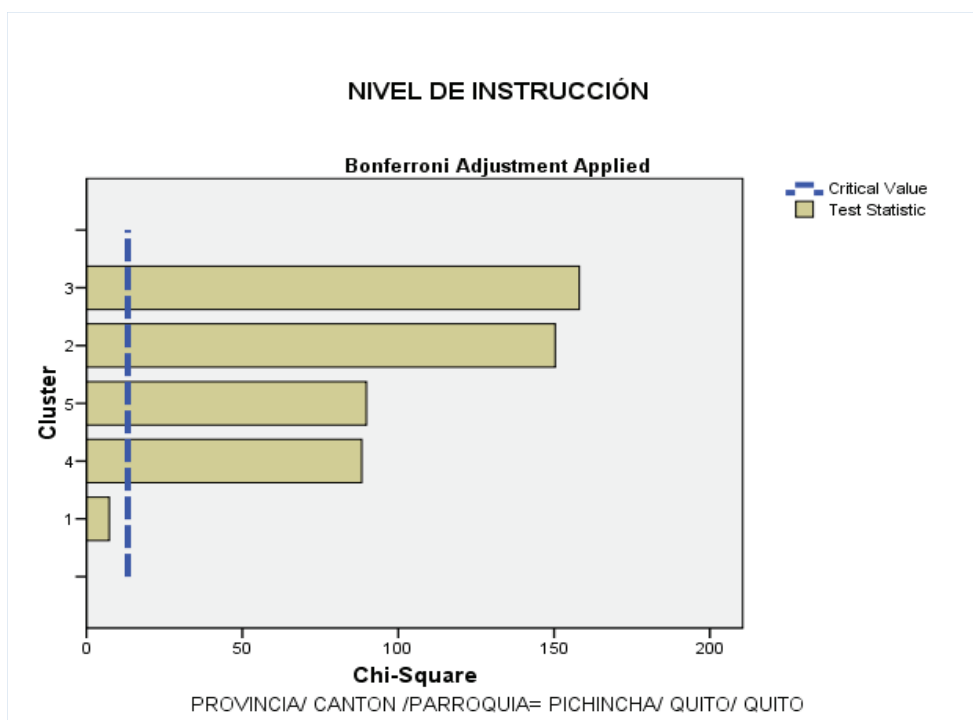


Gráfico 4.10: PRUEBA CHI-CUADRADO NIVEL DE INSTRUCCIÓN



Con base en las distribuciones e incidencia de cada una de las variables se muestran las características de cada grupo:

Cuadro 4.3: CARATERÍSTICAS DE GRUPOS DEFINIDOS

	Grupo Ocupación	Rango de Edad	Sexo	Estado Civil	Nivel Instrucción
Grupo 1	Trabajo o Negocio	Mayor a 30 años	Hombre	Casado / Unión Libre	
Grupo 2	Trabajo o Negocio	Entre 20 y 30 años	-	Soltero / Separado	Secundaria / Superior
Grupo 3	Otros	Mayor a 60 años	-	-	-
Grupo 4	Es Estudiante	de 16 a 25 años	-	Soltero	Secundaria / Superior
Grupo 5	Trabajo / Ama de Casa	Mayor a 30 años	Mujer	Casado / Unión Libre	-

GRUPO 1

Este grupo se caracteriza por los hombres casados o que viven en unión libre y son mayores a 30 años de edad. A este grupo se lo llamará hombres-familia.

GRUPO 2

Son individuos que tienen empleo o su propio negocio, que mayoritariamente tienen una edad entre 20 y 30 años y que son solteros o separados. A este grupo se lo llamará solteros-trabajo.

GRUPO 3

Se caracterizan por aquellas personas de la tercera edad. A este grupo se le denominará adultos-mayores.

GRUPO 4

Son individuos solteros entre 16 y 25 años que son estudiantes de secundaria o universidad. A este grupo se lo llamará jóvenes-estudiantes.

GRUPO 5

Son mujeres casadas mayores a 30 años que tienen trabajo o son amas de casa. A este grupo se lo denominaremos mujeres-familia.

De acuerdo a los grupos definidos se observa el nivel de victimización y percepción en cada uno de ellos.

VICTIMIZACIÓN

Cuadro 4.4: VICTIMIZACIÓN POR GRUPO

		Indice de victimización a personas		Total
		No	Si	
CONGLOMERADOS QUITO	hombres - familia	84,50%	15,50%	100,00%
	solteros - trabajo	78,70%	21,30%	100,00%
	adultos - mayores	88,90%	11,10%	100,00%
	jóvenes - estudiantes	84,00%	16,00%	100,00%
	mujeres - familia	86,10%	13,90%	100,00%
Total		84,00%	16,00%	100,00%

Se observan diferentes niveles de victimización para cada uno de los conglomerados por lo que se realiza un análisis ANOVA de la victimización en

función de los grupos. Para esto se realiza una prueba de homogeneidad de las varianzas para sustentar el uso del estadístico F. La prueba es de la siguiente manera:

H_0 : Todas las varianzas de los grupos son iguales

H_1 : Existe al menos un grupo con varianza diferente a los demás grupos

**Cuadro 4.5: PRUEBA DE HOMOGEIDAD DE VARIANZAS
PARA LA VICTIMIZACIÓN A PERSONAS**

Prueba de Homogeneidad de Varianzas			
Índice de victimización a personas			
Levene Statistic	df1	df2	Sig.
12,780	4	1731	,000

Con un p-valor de 0,000 se rechaza la hipótesis nula de igualdad de las varianzas (con los niveles de significación usuales), por lo que se procede a utilizar el estadístico de Welch o el estadístico de Brown-Forsythe para análisis ANOVA para grupos de varianzas desiguales. La prueba de hipótesis se plantea de la siguiente manera.

H_0 : Las medias de los grupos son iguales

H_1 : Existe al menos un grupo con una media desigual al resto de grupos

Cuadro 4.6: PRUEBA DE IGUALDAD DE MEDIAS PARA LA VICTIMIZACIÓN A PERSONAS

Prueba Robusta de Igualdad de Medias				
Índice de victimización a personas				
	Statistic ^a	df1	df2	Sig.
Welch	3,130	4	759,826	,014
Brown-Forsythe	3,391	4	1599,098	,009

a. Asymptotically F distributed.

Para un nivel de significación del 5%, en ambos casos, se rechaza hipótesis nula de igualdad de medias para los conglomerados; por esta razón, se realiza una prueba múltiple comparativa de medias para encontrar los grupos que presentan diferentes niveles de victimización.

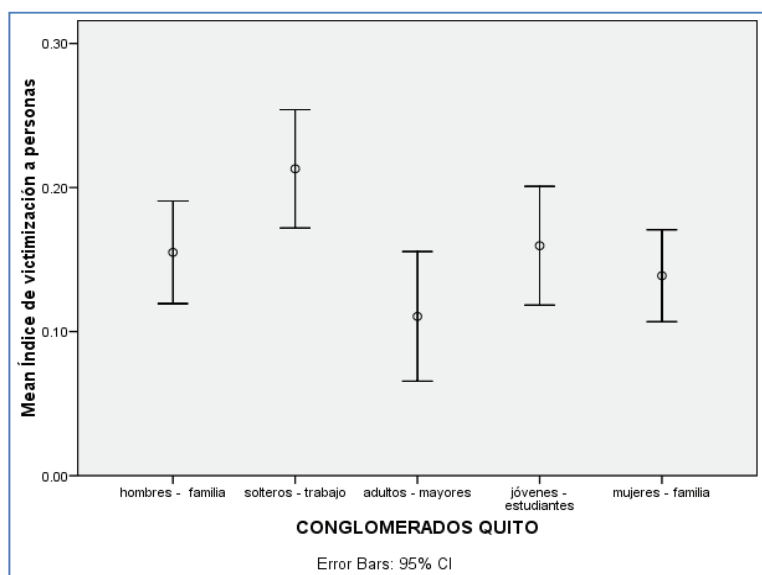
Cuadro 4.7: COMPARACIONES MÚLTIPLES PARA LA VICTIMIZACIÓN A PERSONAS

Comparaciones Múltiples						
Dependent Variable: Índice de victimización a personas						
Tamhane						
(I) CONGLOMERADOS QUITO	(J) CONGLOMERADOS QUITO	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
hombres - familia	solteros - trabajo	-,05799	,02765	,309	-,1356	,0197
	adultos - mayores	,04447	,02913	,744	-,0375	,1264
	jóvenes - estudiantes	-,00461	,02769	1,000	-,0824	,0732
	mujeres - familia	,01623	,02433	,999	-,0521	,0845
solteros - trabajo	hombres - familia	,05799	,02765	,309	-,0197	,1356
	adultos - mayores	,10246*	,03093	,010	,0155	,1895
	jóvenes - estudiantes	,05338	,02958	,524	-,0297	,1365
	mujeres - familia	,07422	,02646	,050	-,0001	,1485
adultos - mayores	hombres - familia	-,04447	,02913	,744	-,1264	,0375
	solteros - trabajo	-,10246*	,03093	,010	-,1895	-,0155
	jóvenes - estudiantes	-,04908	,03096	,701	-,1362	,0380
	mujeres - familia	-,02824	,02800	,977	-,1071	,0506
jóvenes - estudiantes	hombres - familia	,00461	,02769	1,000	-,0732	,0824
	solteros - trabajo	-,05338	,02958	,524	-,1365	,0297
	adultos - mayores	,04908	,03096	,701	-,0380	,1362
	mujeres - familia	,02084	,02650	,996	-,0536	,0953
mujeres - familia	hombres - familia	-,01623	,02433	,999	-,0845	,0521
	solteros - trabajo	-,07422	,02646	,050	-,1485	,0001
	adultos - mayores	,02824	,02800	,977	-,0506	,1071
	jóvenes - estudiantes	-,02084	,02650	,996	-,0953	,0536

*. The mean difference is significant at the .05 level.

Se observa en el intervalo de confianza para las medias de las diferencias de cada pareja de grupos que existen dos diferentes entre sí, mostrados con (*), y que corresponden a los grupos adultos-mayores y solteros-trabajo. Esto se puede ver mejor en el siguiente gráfico:

Gráfico 4.11: VICTIMIZACIÓN A PERSONAS POR GRUPO



Los intervalos de confianza para la media son diferentes para los grupos solteros-trabajo y adultos-mayores. Podrían existir diferencias estadísticas para la media en los demás grupos, sin embargo con este tamaño de muestra, entre 200 y 400 para cada grupo, no se pueden realizar inferencias más desagregadas con mucha confiabilidad.

PERCEPCIÓN

Cuadro 4.8: PERCEPCIÓN DE VICTIMIZACIÓN POR GRUPO

		Percepción		Total
		No	Si	
CONGLOMERADOS QUITO	hombres - familia	61,00%	39,00%	100,00%
	solteros - trabajo	64,90%	35,10%	100,00%
	adultos - mayores	44,20%	55,80%	100,00%
	jóvenes - estudiantes	72,60%	27,40%	100,00%
	mujeres - familia	39,00%	61,00%	100,00%
Total		56,30%	43,70%	100,00%

Al igual que en la victimización se hace un ANOVA de la percepción de la victimización en función de los grupos establecidos.

Cuadro 4.9: PRUEBA DE HOMOGENEIDAD DE LAS VARIANZAS PARA LA PERCEPCIÓN

Prueba de Homogeneidad de Varianzas			
Percepción de victimización			
Levene Statistic	df1	df2	Sig.
22,894	4	1731	,000

Con un p-valor de 0,000 se rechaza la hipótesis de igualdad de varianzas para cada uno de los grupos por lo que se aplican los estadísticos de Welch y Brown-Forsythe, que no asumen este supuesto.

Cuadro 4.10: PRUEBA DE IGUALDAD DE MEDIAS PARA LA PERCEPCIÓN

Prueba Robusta de Igualdad de Medias				
Percepción de victimización				
	Statistic ^a	df 1	df2	Sig.
Welch	31,237	4	749,159	,000
Brown-Forsythe	30,725	4	1463,605	,000

a. Asy mptotically F distributed.

Con un p-valor de 0,00 se rechaza la hipótesis nula de igualdad de varianzas de percepción de inseguridad para los grupos, por lo que se hace un análisis múltiple de medias para encontrar aquellas que sean diferentes.

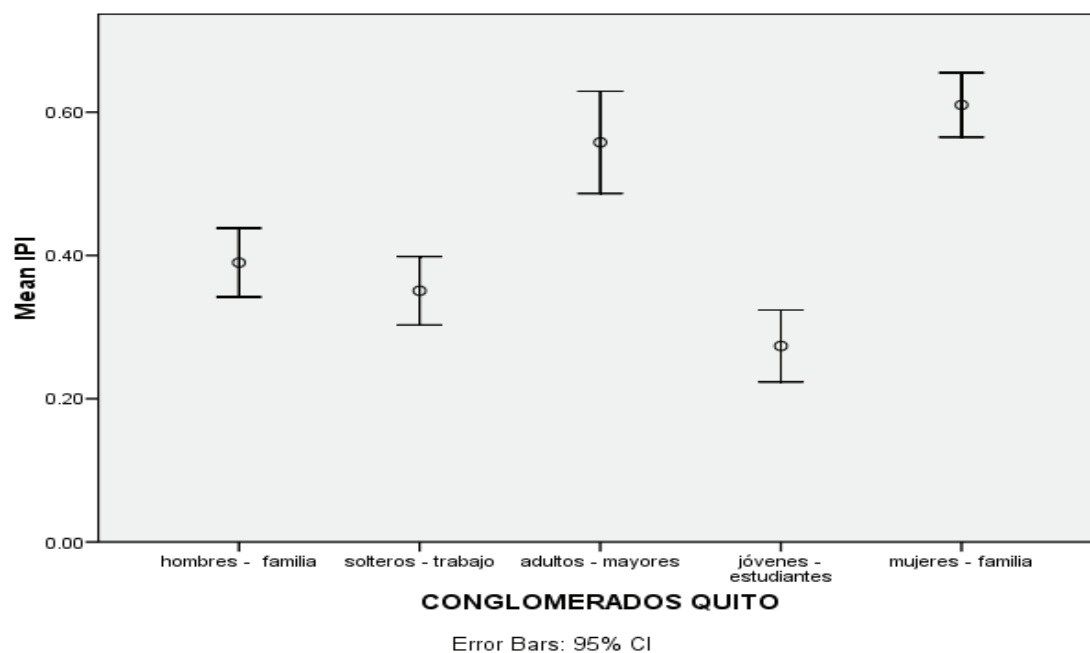
Cuadro 4.11: COMPARACIONES MÚLTIPLES PARA LA PERCEPCIÓN

Comparaciones Múltiples						
Dependent Variable: Percepción de victimización						
Tamhane						
(I) CONGLOMERADOS QUITO	(J) CONGLOMERADOS QUITO	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
hombres - familia	solteros - trabajo	,03935	,03448	,947	-,0575	,1362
	adultos - may ores	-,16789*	,04360	,001	-,2907	-,0451
	jóvenes - estudiantes	,11638*	,03530	,010	,0173	,2155
	mujeres - familia	-,22013*	,03349	,000	-,3141	-,1261
solteros - trabajo	hombres - familia	-,03935	,03448	,947	-,1362	,0575
	adultos - may ores	-,20725*	,04357	,000	-,3300	-,0845
	jóvenes - estudiantes	,07703	,03525	,256	-,0220	,1760
	mujeres - familia	-,25948*	,03344	,000	-,3533	-,1656
adultos - may ores	hombres - familia	,16789*	,04360	,001	,0451	,2907
	solteros - trabajo	,20725*	,04357	,000	,0845	,3300
	jóvenes - estudiantes	,28428*	,04421	,000	,1598	,4088
	mujeres - familia	-,05224	,04278	,920	-,1728	,0683
jóvenes - estudiantes	hombres - familia	-,11638*	,03530	,010	-,2155	-,0173
	solteros - trabajo	-,07703	,03525	,256	-,1760	,0220
	adultos - may ores	-,28428*	,04421	,000	-,4088	-,1598
	mujeres - familia	-,33652*	,03427	,000	-,4328	-,2403
mujeres - familia	hombres - familia	,22013*	,03349	,000	,1261	,3141
	solteros - trabajo	,25948*	,03344	,000	,1656	,3533
	adultos - may ores	,05224	,04278	,920	-,0683	,1728
	jóvenes - estudiantes	,33652*	,03427	,000	,2403	,4328

*. The mean difference is significant at the .05 level.

Como se observa en la comparación múltiple, existen varios grupos que presentan promedios de percepción diferentes.

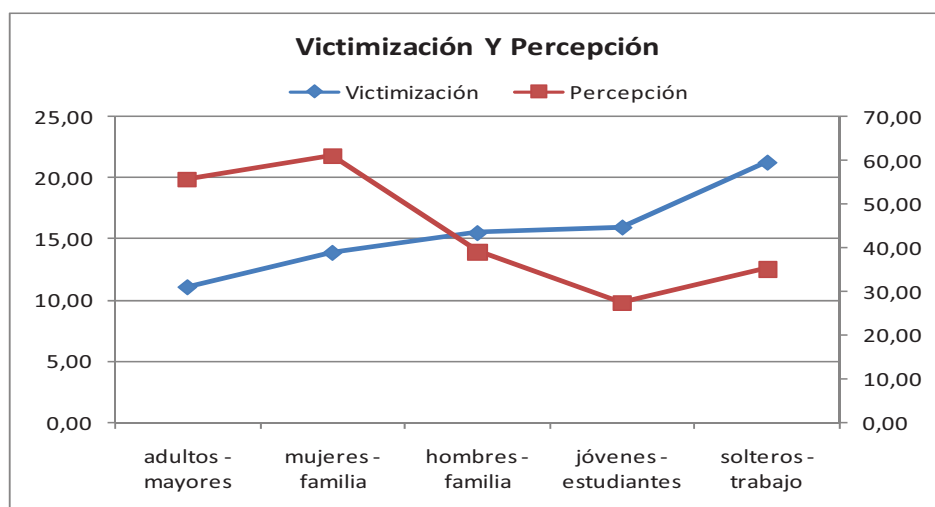
Gráfico 4.12: PERCEPCIÓN POR GRUPO



Los adultos-mayores junto con las mujeres-familia muestran una mayor percepción mientras que los jóvenes estudiantes muestran la menor percepción.

PERCEPCIÓN Y VICTIMIZACIÓN

Gráfico 4.13: VICTIMIZACIÓN Y PERCEPCIÓN



La victimización y la percepción de cada grupo muestra como los grupos más victimizados tienen menos percepción y los grupos menos victimizados tienen mayor percepción. Este hecho muestra la aversión al riesgo que tiene cada uno de los grupos y como consecuencia de esa aversión la victimización misma.

VICTIMIZACIÓN HOGARES

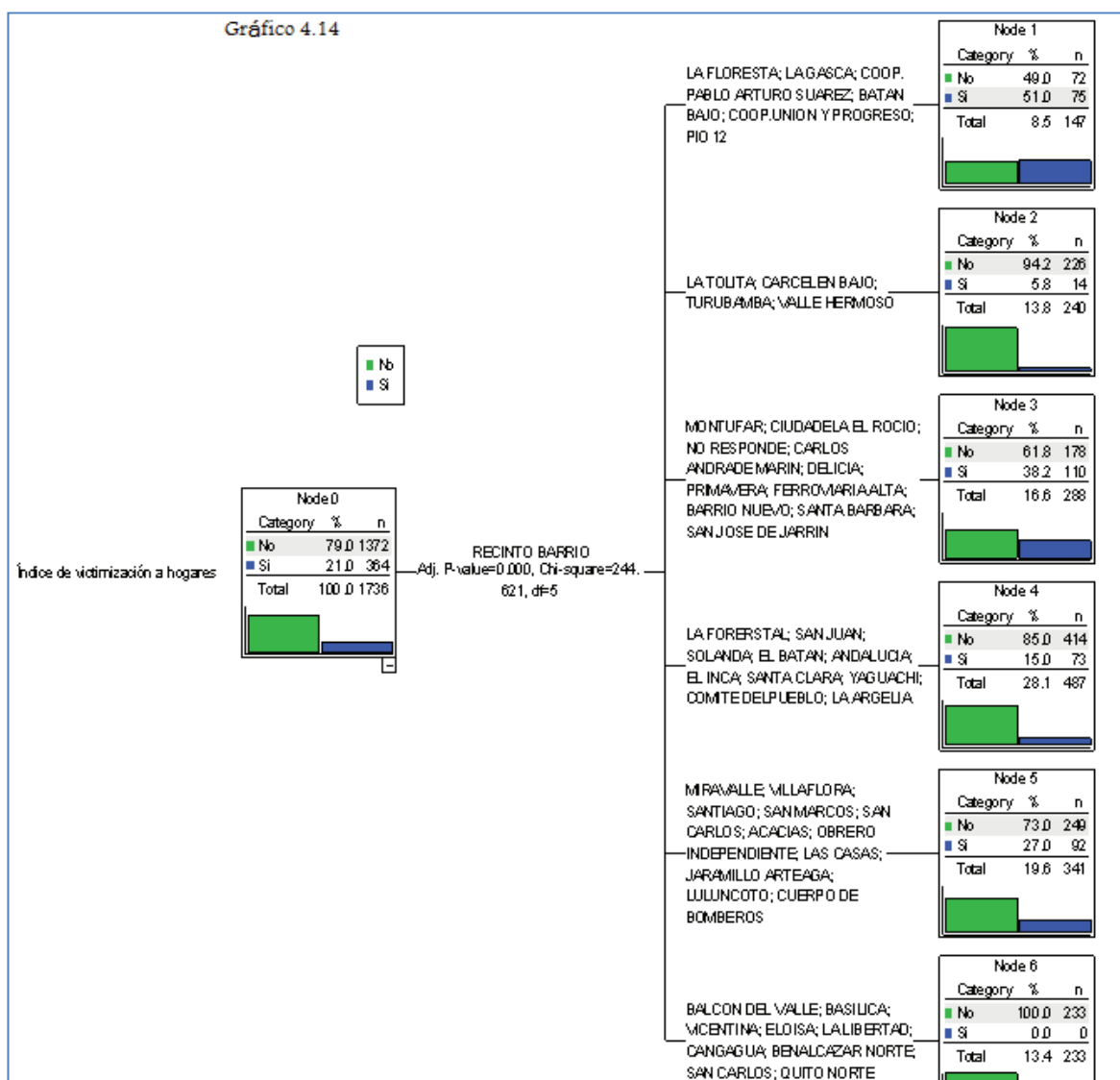
Otro factor importante es el estudio de la victimización a hogares que es el porcentaje de hogares que fueron víctimas en alguno de los siguientes delitos: robo de viviendas, vehículos o motorizados. El objetivo de estudiar esta victimización es conocer su incidencia y la relación que ésta pueda tener con la victimización a personas.

Cuadro 4.12: ÍNDICE DE VICTIMIZACIÓN A HOGARES

Índice de victimización a hogares					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	561	79,0	79,0	79,0
	Si	149	21,0	21,0	100,0
Total		710	100,0	100,0	

Para la cabecera cantonal de Quito tenemos un índice de victimización a hogares del 21%. Se han empleado árboles de decisión para agrupar los barrios en función de la exposición de los hogares a ser victimizados (en función de los individuos en cada hogar).

Gráfico 4.14: ÁRBOL GENERADO POR BARRIOS



El árbol de decisión define seis nodos que muestran diferentes niveles de victimización a hogares en los diferentes grupos de barrios. Los grupos definidos por el árbol se muestran a continuación:

Cuadro 4.13: GRUPOS BARRIOS DEFINIDOS POR ELÁRBOL DE DECISIÓN

Nodo	Barrio
Nodo 1	PIO 12 LA FLORESTA COOP. UNION Y PROGRESO COOP. PABLO ARTURO SUAREZ LA GASCA BATAN BAJO
Nodo 2	CARCELEN BAJO TURUBAMBA VALLE HERMOSO LA TOLITA
Nodo 3	BARRIO NUEVO MONTUFAR SAN JOSE DE JARRIN DELICIA CARLOS ANDRADE MARIN CIUDADELA EL ROCIO SANTA BARBARA FERROVIARIA ALTA PRIMAVERA NO RESPONDE
Nodo 4	SOLANDA EL BATAN YAGUACHI LA ARGELIA SAN JUAN EL INCA LA FORERSTAL COMITE DELPUEBLO ANDALUCIA SANTA CLARA
Nodo 5	OBRERO INDEPENDIENTE JARAMILLO ARTEAGA ACACIAS SANTIAGO LULUNCOTO MIRAVALLE SAN MARCOS VILLAFLORA LAS CASAS SAN CARLOS CUERPO DE BOMBEROS
Nodo 6	BENALCAZAR NORTE SAN CARLOS QUITO NORTE CANGAGUA BALCON DEL VALLE LA LIBERTAD ELOISA BASILICA VICENTINA

Cuadro 4.14: VICTIMIZACIÓN A HOGARES POR GRUPO

% within Grupos Victimización Hogares				
		Índice de victimización a hogares		Total
		No	Si	
Grupos Victimización Hogares	Grupo 1	49,0%	51,0%	100,0%
	Grupo 2	94,2%	5,8%	100,0%
	Grupo 3	61,8%	38,2%	100,0%
	Grupo 4	85,0%	15,0%	100,0%
	Grupo 5	73,0%	27,0%	100,0%
	Grupo 6	100,0%		100,0%
Total		79,0%	21,0%	100,0%

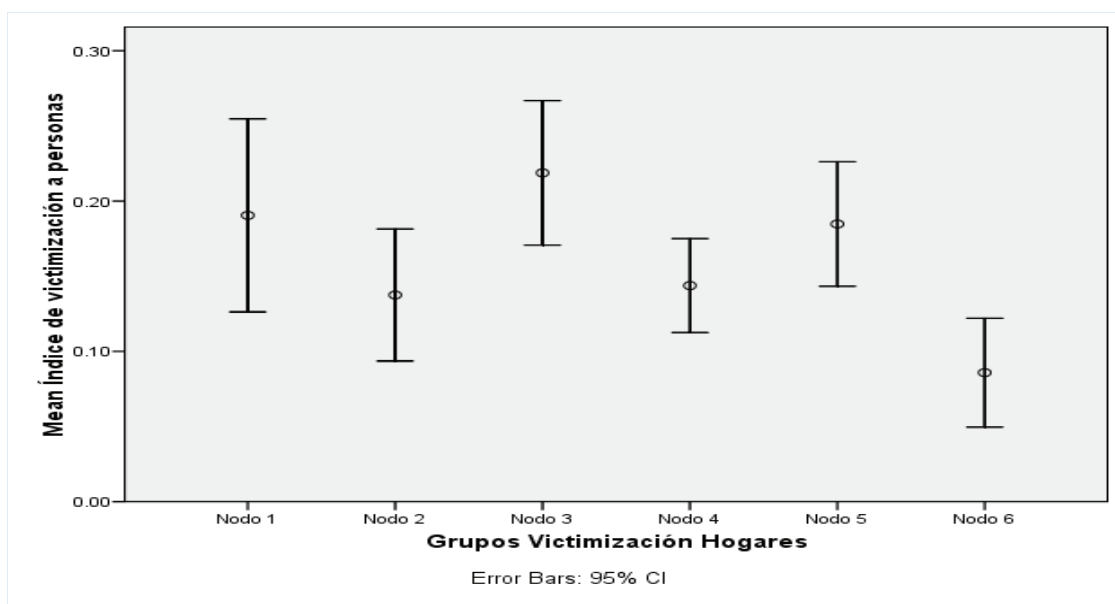
Como se observa en el cuadro precedente, existen grupos de barrios con un alto nivel de victimización a hogares como el nodo uno y tres, con niveles mayores al 38% y que son diferentes estadísticamente. El nodo seis no muestra victimización (cabe mencionar que la encuesta fue diseñada para cabeceras cantonales y no para subgrupos contenidos en ésta, pero que sin embargo son utilizados en este trabajo con el objetivo mostrar como se pueden aplicar técnicas matemáticas en el tratamiento de este tema social).

Se observa también los índices de victimización a personas en los grupos de barrios agrupados en función de la victimización a hogares.

Cuadro 4.15: VICTIMIZACIÓN A PERSONAS POR GRUPO

% within Grupos Victimización Hogares				
		Índice de victimización a personas		Total
		No	Si	
Grupos Victimización Hogares	Grupo 1	81,0%	19,0%	100,0%
	Grupo 2	86,3%	13,8%	100,0%
	Grupo 3	78,1%	21,9%	100,0%
	Grupo 4	85,6%	14,4%	100,0%
	Grupo 5	81,5%	18,5%	100,0%
	Grupo 6	91,4%	8,6%	100,0%
Total		84,0%	16,0%	100,0%

Gráfico 4.15: VICTIMIZACIÓN A PERSONAS POR GRUPO



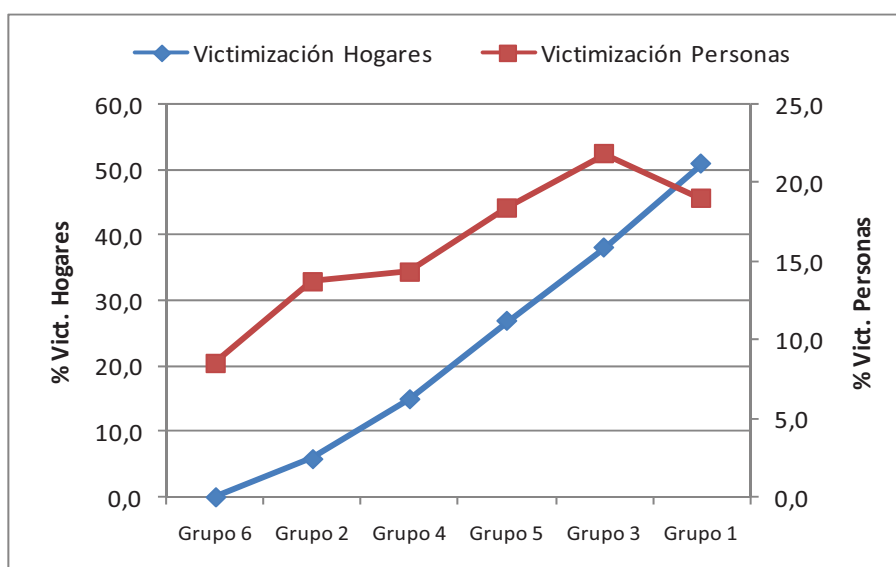
La victimización a personas de los barrios en los grupos tres y seis son diferentes estadísticamente mostrando así que la victimización a personas depende no solo del perfil social sino también de la ubicación geográfica.

Se observa la victimización a personas y hogares en los grupos de barrios establecidos para observar gráficamente la relación que existe entre ambas:

Cuadro 4.16: VICTIMIZACIÓN A HOGARES Y PERSONAS

Vict. Hogares Y Vict. Personas		
Grupo	Victimización Hogares	Victimización Personas
Grupo 6	0,0	8,6
Grupo 2	5,8	13,8
Grupo 4	15,0	14,4
Grupo 5	27,0	18,5
Grupo 3	38,2	21,9
Grupo 1	51,0	19,0
Total	21,0	16,0

Gráfico 4.16: VICTIMIZACIÓN A HOGARES Y PERSONAS



La victimización a hogares y personas muestra un coeficiente de correlación lineal (R -cuadrado de 0,76), mostrando de esta manera una relación lineal fuerte entre la victimización de hogares y la victimización de personas. Este valor podría indicar que la victimización de los diferentes perfiles de individuos se suscita de acuerdo a los barrios en que habitan. En este punto es importante conocer el origen de la victimización, si es producto de un comportamiento social, situación geográfica/económica o ambas a la vez. Para esto se estudia la victimización a personas en la matriz que conforman los perfiles sociales y los grupos de barrios.

Cuadro 4.17: VICTIMIZACIÓN GRUPOS BARRIOS Y PERFILES SOCIALES

% Victimización	Nodos (Grupos de Barrio)						Total Perfiles Sociales
	Nodo 1	Nodo 2	Nodo 3	Nodo 4	Nodo 5	Nodo 6	
hombres - familia	28,1%	6,0%	20,3%	18,5%	16,7%	2,4%	15,5%
solteros - trabajo	18,4%	27,7%	32,7%	21,7%	19,8%	9,8%	21,3%
adultos - mayores	6,3%	0,0%	18,5%	5,7%	19,3%	5,1%	11,1%
jóvenes - estudiantes	24,1%	11,4%	16,9%	11,1%	21,0%	14,6%	16,0%
mujeres - familia	12,5%	16,0%	21,0%	9,2%	16,0%	10,0%	13,9%
Total Grupos Barrio	19,0%	13,8%	21,9%	14,4%	18,5%	8,6%	16,0%

La matriz de los perfiles sociales y los grupos de barrios nos permite visualizar la victimización de cada una de ellas o ambas al mismo tiempo, permitiendo definir estrategias diferenciadas. La victimización del perfil solteros-trabajo se mantiene alta en la mayoría de grupos barrios por lo que se recomendaría diseñar una estrategia específica para este perfil social. Por otro lado, se observa que los barrios que pertenecen al grupo barrio tres tienen una victimización alta en cada uno de sus perfiles sociales, por lo que se debería crear una estrategia específica para este conjunto de barrios.

Si se aplicaran estrategias sobre los barrios y grupos sociales más victimizados, como en este caso, son los grupos barrios 1 y 3, los solteros-trabajo y hombres-familia para los perfiles sociales, y se estimara una reducción del 20% de la victimización en estos grupos; se obtuviera una variación negativa del 13,6% de la victimización de toda la ciudad, enfocándonos en solo el 60% de los ciudadanos, como se muestra en el cuadro a continuación:

Cuadro 4.18: REPRESENTATIVIDAD GRUPOS BARRIOS Y PERFILES SOCIALES

% Población	Nodos (Grupos de Barrio)						
Perfiles Sociales	Nodo 1	Nodo 2	Nodo 3	Nodo 4	Nodo 5	Nodo 6	Total Nodos
hombres - familia	1,8%	3,9%	4,0%	7,1%	3,8%	2,4%	23,0%
solteros - trabajo	2,2%	2,7%	3,0%	6,1%	4,7%	3,5%	22,2%
adultos - mayores	0,9%	0,9%	1,6%	2,0%	3,3%	2,2%	10,9%
jóvenes - estudiantes	1,7%	2,0%	3,4%	4,7%	3,6%	2,4%	17,7%
mujeres - familia	1,8%	4,3%	4,7%	8,1%	4,3%	2,9%	26,2%
Total Grupos	8,5%	13,8%	16,6%	28,1%	19,6%	13,4%	100,0%

Cuadro 4.19: INCIDENCIA DE VICTIMIZACIÓN EN GRUPOS BARRIOS

Y PERFILES SOCIALES

% Victimización	Nodos (Grupos de Barrio)						
Perfiles Sociales	Nodo 1	Nodo 2	Nodo 3	Nodo 4	Nodo 5	Nodo 6	Total Nodos
hombres - familia	22,5%	4,8%	16,2%	14,8%	13,3%	1,9%	12,4%
solteros - trabajo	14,7%	22,1%	26,2%	17,4%	15,8%	7,9%	17,0%
adultos - mayores	5,0%	0,0%	14,8%	5,7%	19,3%	5,1%	10,4%
jóvenes - estudiantes	19,3%	11,4%	13,6%	11,1%	21,0%	14,6%	14,9%
mujeres - familia	10,0%	16,0%	16,8%	9,2%	16,0%	10,0%	13,0%
Total Grupos	15,2%	12,3%	17,5%	12,5%	16,9%	8,0%	13,8%

Los segmentos seleccionados (en amarillo en el cuadro 4.18) representan el 60% de la población. Alcanzar una disminución del 20% de la victimización en estos segmentos, de características similares, representaría una disminución del 13,6% en toda la ciudad.

AGRUPAMIENTO PARA GUAYAQUIL

Para la ciudad de Guayaquil se realiza un análisis similar al de la ciudad de Quito, razón por la cual se mencionan los resultados en forma resumida¹⁹.

Cuadro 4.20: ÍNDICE DE VICTIMIZACIÓN A PERSONAS

Índice de victimización a personas						
PROVINCIA			Frequency	Percent	Valid Percent	Cumulative Percent
GUAYAS	Valid	No	1902	85,3	85,3	85,3
		Si	328	14,7	14,7	100,0
	Total		2230	100,0	100,0	

Cuadro 4.21: ÍNDICE DE PERCEPCIÓN DE VICTIMIZACIÓN

Percepción de victimización						
PROVINCIA			Frequency	Percent	Valid Percent	Cumulative Percent
GUAYAS	Valid	No	1342	60,2	60,2	60,2
		Si	888	39,8	39,8	100,0
	Total		2230	100,0	100,0	

¹⁹ Detalles del agrupamiento se muestran en el Anexo 3.

Cuadro 4.22: ÍNDICE DE VICTIMIZACIÓN A HOGARES

Índice de victimización a hogares						
PROVINCIA			Frequency	Percent	Valid Percent	Cumulative Percent
GUAYAS	Valid	No	744	87,0	87,0	87,0
		Si	111	13,0	13,0	100,0
	Total		855	100,0	100,0	

En el análisis de conglomerados se obtuvieron los siguientes grupos:

Cuadro 4.23: CARACTERÍSTICAS DE GRUPOS DEFINIDOS

	Grupo Ocupación	Rango de Edad	Sexo	Estado Civil	Nivel Instrucción
Grupo 1	Estudiante, trabajo o Negocio	Menores a 30 años		Soltero	Secundaria / Superior
Grupo 2	Amas de casa	Mayores a 30 años	Mujer	Casado/Unión Libre	-
Grupo 3	Trabajo o Negocio	Mayores a 30 años	Hombre	Casado/Unión Libre	-

De acuerdo a las características de cada grupo se han puesto nombres a cada uno de los grupos para una mejor comprensión de los mismos. Al grupo 1 se lo ha llamado como jóvenes, al grupo 2 como mujeres-familia y al grupo 3 como hombres-familia

Los porcentajes de Victimización y Percepción para cada conglomerado se muestran a continuación:

Cuadro 4.24: VICTIMIZACIÓN A PERSONAS POR GRUPO

% within CONGLOMERADOS_GUAYAQUIL				
		Índice de victimización a personas		Total
		No	Si	
CONGLOMERADOS_	jóvenes	81,1%	18,9%	100,0%
GUAYAQUIL	mujeres - familia	90,1%	9,9%	100,0%
	hombres - familia	84,6%	15,4%	100,0%
Total		85,3%	14,7%	100,0%

Cuadro 4.25: PERCEPCIÓN DE VICTIMIZACIÓN POR GRUPO

% within CONGLOMERADOS_GUAYAQUIL				
		Percepción de victimización		Total
		No	Si	
CONGLOMERADOS_	jóvenes	73,0%	27,0%	100,0%
GUAYAQUIL	mujeres - familia	50,0%	50,0%	100,0%
	hombres - familia	58,7%	41,3%	100,0%
Total		60,1%	39,9%	100,0%

Se observa que la victimización para cada grupo es muy diferente no obstante se hace un análisis Anova para cada grupo:

Cuadro 4.26: PRUEBA DE IGUALDAD DE MEDIAS PARA LA VICTIMIZACIÓN

Prueba Robusta de Igualdad de Medias				
Índice de victimización a personas				
	Statistic ^a	df1	df2	Sig.
Welch	11,985	2	1378,115	,000
Brown-Forsythe	10,919	2	1952,723	,000

a. Asymptotically F distributed.

Con un p-valor de 0,00 se rechaza la hipótesis de igualdad de medias de victimización para los grupos.

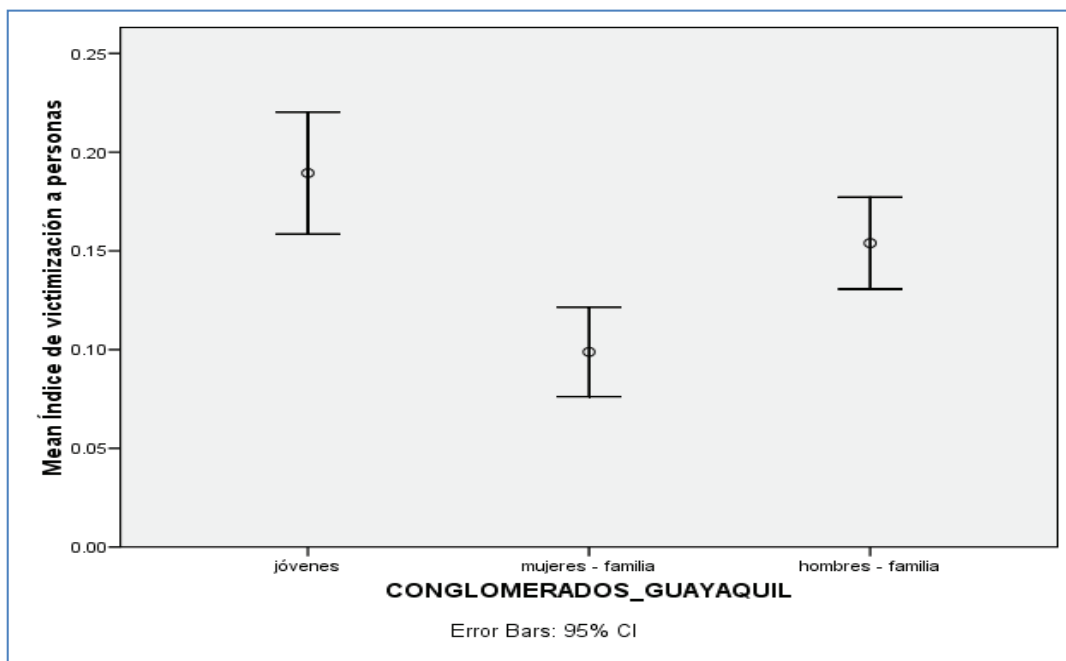
Cuadro 4.27: COMPARACIONES MÚLTIPLES PARA LA VICTIMIZACIÓN A PERSONAS

Comparaciones Múltiples						
Dependent Variable: Índice de victimización a personas						
Tamhane						
(I) CONGLOMERADOS_ GUAYAQUIL	(J) CONGLOMERADOS_ GUAYAQUIL	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
jóvenes	mujeres - familia	,09060*	,01950	,000	,0440	,1372
	hombres - familia	,03548	,01968	,200	-,0116	,0825
mujeres - familia	jóvenes	-,09060*	,01950	,000	-,1372	-,0440
	hombres - familia	-,05513*	,01655	,003	-,0947	-,0156
hombres - familia	jóvenes	-,03548	,01968	,200	-,0825	,0116
	mujeres - familia	,05513*	,01655	,003	,0156	,0947

*. The mean difference is significant at the .05 level.

Para la victimización se observa que existe diferencia estadística entre los tres grupos:

Gráfico 4.17: VICTIMIZACIÓN A PERSONAS POR GRUPO



Existe diferencia estadística de los dos grupos más victimizados con el grupo mujeres-familia. Para la percepción el análisis es similar. A continuación se muestra el cuadro de las comparaciones múltiples para la percepción.

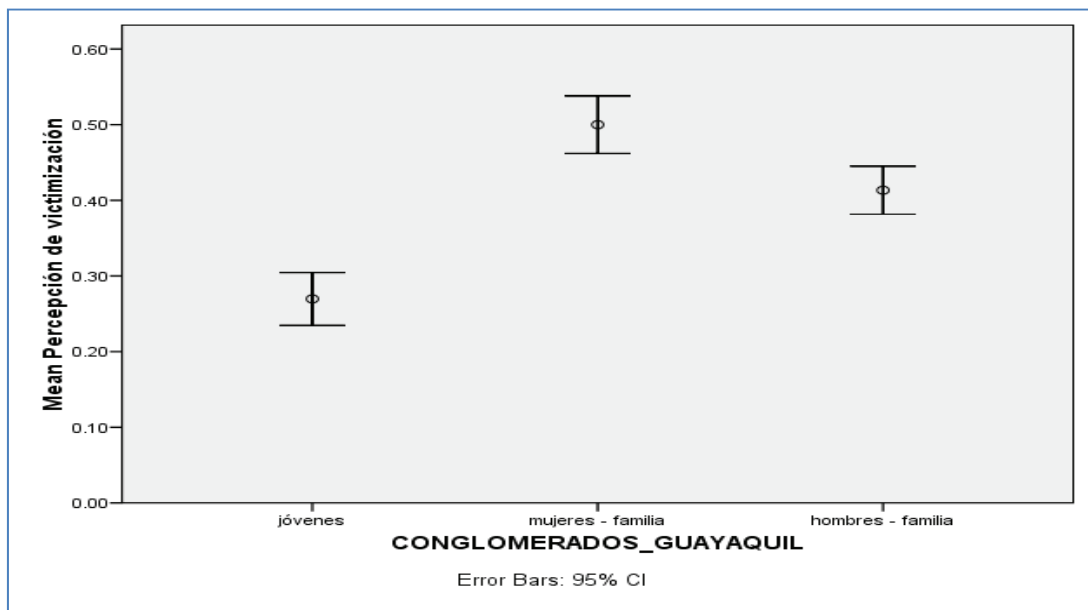
Cuadro 4.28: COMPARACIONES MÚLTIPLES PARA LA PERCEPCIÓN DE LA VICTIMIZACIÓN

Comparaciones Múltiples						
Dependent Variable: Percepción de victimización						
Tamhane						
(I) CONGLOMERADOS_ GUAYAQUIL	(J) CONGLOMERADOS_ GUAYAQUIL	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
jóvenes	mujeres - familia	-,23034*	,02630	,000	-,2932	-,1675
	hombres - familia	-,14368*	,02404	,000	-,2012	-,0862
mujeres - familia	jóvenes	,23034*	,02630	,000	,1675	,2932
	hombres - familia	,08665*	,02522	,002	,0264	,1469
hombres - familia	jóvenes	,14368*	,02404	,000	,0862	,2012
	mujeres - familia	-,08665*	,02522	,002	-,1469	-,0264

*. The mean difference is significant at the .05 level.

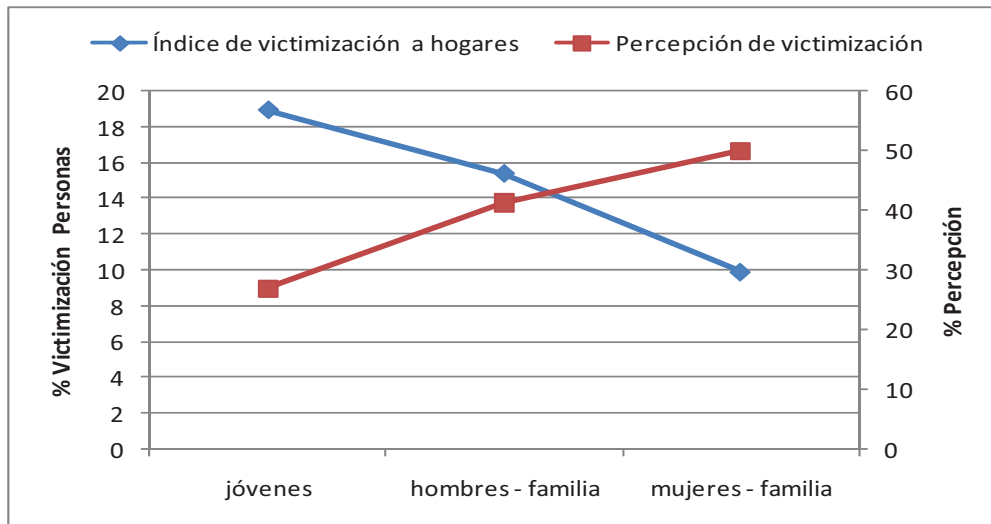
Para la percepción se observa que todos los grupos tienen una media estadísticamente diferente.

Gráfico 4.18: PERCEPCIÓN POR GRUPO



La victimización y la percepción de victimización muestran un comportamiento inversamente proporcional:

Gráfico 4.19: PERCEPCIÓN Y VICTIMIZACIÓN



VICTIMIZACIÓN A HOGARES

Mediante árboles de decisión se obtienen 5 grupos de victimización a hogares, que muestran los siguientes resultados:

Cuadro 4.29: GRUPOS BARRIOS DEFINIDOS POR EL ÁRBOL

Grupo	Barrio
Grupo 1	FLORESTA GUASMO MARTHA ROLDOS SAN EDUARDO SANTIAGO ROLDOS GUERROS DEL FORTIN URDESA SANTA MONICA
Grupo 2	PANCHO JACOME ATARAZANA NIGERIA
Grupo 3	GARAY GUAYACANES SAMANES LA CHALA TECHO PARA TODOS CLAVELES 1 DIGNIDAD POPULAR URDENOR COOPERATIVA OLIVOS
Grupo 4	BARRIO LINDO SUBURBIO ALBORADA ROSALES SAUCES ORQUIDEAS PRIMERO DE MAYO FLOR DE FORTIN PASCUALES PISO Y TECHO NO RESPONDE
Grupo 5	25 DE JULIO ACACIAS COOPERATIVA 5 DE DICIEMBRE

Cuadro 4.30: GRUPOS DE BARRIOS EN GUAYAQUIL

Grupos Barrios Guayaquil					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Grupo 1	998	45,0	45,0	45,0
	Grupo 2	631	28,4	28,4	73,4
	Grupo 3	310	14,0	14,0	87,3
	Grupo 4	98	4,4	4,4	91,8
	Grupo 5	183	8,2	8,2	100,0
	Total	2220	100,0	100,0	

Cuadro 4.31: VICTIMIZACIÓN A HOGARES POR GRUPO

% within Grupos Hogares Guayaquil				
		Índice de victimización a hogares		Total
		No	Si	
Grupos	1	83,8%	16,2%	100,0%
Hogares	2	91,6%	8,4%	100,0%
Guayaquil	3	66,8%	33,2%	100,0%
	4	100,0%		100,0%
	5	76,0%	24,0%	100,0%
Total		83,7%	16,3%	100,0%

Cuadro 4.32: PERCEPCIÓN DE VICTIMIZACIÓN POR GRUPO

% within Grupos Hogares Guayaquil				
		Percepción de victimización		Total
		No	Si	
Grupos	1	55,2%	44,8%	100,0%
Hogares	2	64,0%	36,0%	100,0%
Guayaquil	3	68,1%	31,9%	100,0%
	4	63,3%	36,7%	100,0%
	5	57,9%	42,1%	100,0%
Total		60,1%	39,9%	100,0%

Los grupos barrios de hogares muestran que existen niveles de victimización a personas diferenciados para cada grupo; así, por ejemplo, los grupos de barrios 1 y 3 representan el 50% de la ciudad con los niveles de victimización más altos.

PERFILES Y GRUPOS DE HOGARES

Cuadro 4.33: VICTIMIZACIÓN POR GRUPOS BARRIOS Y PERFILES SOCIALES

% Victimización	Grupos Barrios Guayaquil					
Perfiles Sociales	Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5	Total Perfiles Sociales
jóvenes	26%	9%	21%	14%	17%	19%
mujeres - familia	11%	5%	17%	15%	9%	10%
hombres - familia	17%	13%	16%	12%	14%	15%
Total Barrios	18%	9%	18%	13%	14%	15%

Una posible estrategia sería trabajar sobre el perfil jóvenes y los grupos barrios uno y tres que muestran niveles de victimización más altos:

Cuadro 4.34: SEGMENTOS CON MAYOR VICTIMIZACIÓN

% Victimización	Grupos Barrios Guayaquil					
Perfiles Sociales	Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5	Total Perfiles Sociales
jóvenes	26%	9%	21%	14%	17%	19%
mujeres - familia	11%	5%	17%	15%	9%	10%
hombres - familia	17%	13%	16%	12%	14%	15%
Total Barrios	18%	9%	18%	13%	14%	15%

Cuadro 4.35: REPRESENTATIVIDAD GRUPOS BARRIOS Y PERFILES SOCIALES

% Población	Grupos Barrios Guayaquil					
Perfiles Sociales	Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5	Total Perfiles Sociales
jóvenes	12%	8%	4%	1%	3%	28%
mujeres - familia	13%	9%	4%	1%	2%	30%
hombres - familia	20%	11%	6%	2%	3%	42%
Total Barrios	45%	28%	14%	4%	8%	100%

Considerando la representatividad que cada grupo seleccionado (marcado en amarillo en el cuadro 4.32) tiene sobre la población, se podrían generar estrategias que minimicen la victimización de manera focalizada y optimizando recursos.

Las estadísticas del lugar, hora del día, tipo de agresión y objeto robado podrán definir si la victimización se debe al perfil social o a la ubicación geográfica de residencia del individuo, permitiendo así definir estrategias para el perfil social o el sector, según corresponda.

El estudio de las estadísticas del lugar, hora del día, tipo de agresión y objeto robado no son parte del presente estudio debido a que en la encuesta no se encuentra declarado el lugar del robo y además no se cuenta con tamaños de muestra suficientes de casos como para inferir resultados confiables.

Finalmente, los índices de victimización en cada barrio y perfil social pueden ser una medida de la gestión de las organizaciones encargadas de generar estrategias para que este índice disminuya.

CAPITULO 5:

CONCLUSIONES Y RECOMENDACIONES

PARA EL ANÁLISIS DESCRIPTIVO

1. En el análisis descriptivo de cada una de las variables se observó que los jóvenes, estudiantes o universitarios y solteros son más victimizados que el resto de la población. También se observó que el grupo con estas características tiene una percepción de inseguridad menor al resto de la población, siendo esta una de las posibles razones de su mayor victimización. Estos dos hechos pueden ser fruto del comportamiento natural y social que estos individuos tienen de acuerdo a su edad y estado civil. En general, se observa que existe una relación inversa entre el nivel de victimización y el nivel de percepción de inseguridad.
2. Los divorciados también tienen un alto porcentaje de victimización que puede ser asociado a un comportamiento similar al de los jóvenes.
3. En cada una de las regiones se tienen niveles de inseguridad muy diferentes; por ejemplo, la región Costa tiene un porcentaje de victimización de 13,9%, casi dos puntos más que la Sierra con 11,95% y 5,5 puntos porcentuales más que la región Amazónica, que tiene un porcentaje de victimización de 6,4%.
4. Las provincias también mostraron niveles de inseguridad muy diferentes; las provincias de Cotopaxi, los Ríos, Pichincha y Guayas resultaron ser las más victimizadas a nivel nacional con un porcentaje de victimización mayor al promedio nacional de 12,7%.

PARA EL SCORE DE VICTIMIZACIÓN

5. Para desarrollar un modelo de *score* al igual que para cualquier otro modelo es indispensable contar con información consistente; usualmente, la obtención de la información demanda el 80% del desarrollo de todo el modelo. Se requiere siempre contrastar el modelo de desarrollo con una muestra de prueba.
6. La definición de elemento bueno o malo debe ser cuidadosamente establecida, pues a partir de ella serán seleccionadas y agrupadas las diferentes variables explicativas. Una definición exageradamente flexible puede terminar en modelos con baja capacidad discriminativa o baja capacidad predictiva aunque el proceso de modelamiento sea el correcto.
7. En el análisis multivariado es indispensable encontrar las variables y la combinación de las mismas que permitan la construcción de un modelo que eficiente, que equilibre los requerimientos de menor información y mayor explicación. Para esto, el uso de árboles de decisión permite encontrar las variables y la combinación de las mismas, que discriminen mejor los buenos y malos elementos, permitiendo una interpretación lógica y evitando criterios subjetivos por parte del investigador; además, permite encontrar relaciones que podrían indicar características hasta ese momento no conocidas.
8. La regresión logística, para objetivos explicativos, es un modelo que debería considerar solamente variables indicadoras (*dummy*), ya que estas indican la incidencia de un determinado evento con la probabilidad del evento de estudio, permitiendo así la interpretación del modelo y la construcción de perfiles.
9. Antes de estudiar la discriminación de un modelo de *score* se debe hacer el análisis estadístico del modelo en general y luego de cada una de sus

variables, verificando que cada una de estas sea representativa (el número de elementos afirmativos debe ser al menos el 5% de la población), que tenga correlación con la variable objetivo y que se encuentre alineada de acuerdo al *odds* de cada variable, esperando la misma relación en la muestra de desarrollo y la de prueba.

10. Mediante la discriminación en cada rango percentil se pueden obtener perfiles y mediante estos obtener estrategias diferenciadas de decisión que permitan realizar una gestión más oportuna y eficiente para tratar de disminuir la victimización (los ejemplos presentados constituyen solamente una primera aproximación, pues se requeriría mayor información para análisis más profundos). El análisis de *odds*, por ejemplo, mostró en la distribución por deciles en el *score* una relación de 20 no victimizados a un victimizado en el primer decil mientras que el último decil, una relación de 4 a 1.
11. La clasificación de perfiles a partir de quintiles en el *Score* de Victimización mostró que geográficamente las provincias de Pichincha, Guayas, Cotopaxi y Los Ríos son las más victimizadas y los individuos de educación superior con edad menor a 35 años tienen también un alto índice de victimización. Las variables seleccionadas para el modelo de regresión logística son la provincia, el área geográfica (rural o urbana), edad y género.
12. La provincia resultó ser la más importante, contribuyendo casi con todas las variables indicadoras, mientras que las demás variables contribuyen con el modelo con una parte de las variables indicadoras de sus categorías. Este hecho se presenta debido a que al realizarse un *score* de victimización a nivel nacional, las variables seleccionadas son aquellas estadísticamente significativas a nivel nacional; por tanto, la variable provincia es la más importante pues la realidad socio-económica de cada provincia es diferente. Este hecho conduce a que variables como género, edad, tipo de ocupación,

nivel de instrucción, etc no sean tan relevantes en el modelo; sin embargo, para estudios más detallados (por ejemplo, por ciudad) estas variables serían de gran importancia.

PARA LOS PERFILES DE VICTIMIZACIÓN

13. El objetivo de crear perfiles de victimización es identificar grupos homogéneos de individuos que por sus características puedan ser victimizados en diferente forma e intensidad. En este trabajo se ha utilizado el análisis de conglomerados, con una metodología bastante reciente (a dos etapas), que permite trabajar con grandes cantidades de información y determinar automáticamente el número óptimo de conglomerados.
14. Para el análisis de conglomerados la selección de las variables es de suma importancia ya que estas son las que definirán los grupos a formarse; por tanto, no se deben considerar variables que no tengan interpretabilidad para el estudio y si contribuyan para la agrupación. En este sentido, las variables que se seleccionaron fueron: el rango de edad, el género, estado civil, nivel de estudios y la ocupación.
15. Para la ciudad de Quito el análisis determinó cinco grupos definidos como; jóvenes estudiantes; solteros con trabajo; padres de familia; madres de familia y adultos mayores. Los solteros con trabajo mostraron ser los más victimizados con un índice de victimización del 21,3%, los jóvenes estudiantes con un 16%, los padres de familia con un 15.5%, las madres de familia 13,9% y los adultos mayores en un 11,1%. La percepción de los jóvenes estudiantes fue 27,4%, solteros con trabajo 35%, padres de familia 39%, adultos mayores 56% y madres de familia 61%.
16. Para estos grupos se realizaron estudios ANOVA de la victimización y percepción, encontrando que los conglomerados de jóvenes estudiantes y

solteros con trabajo tienen una mayor victimización que los adultos mayores y que la percepción es inversamente proporcional a la victimización en estos conglomerados; estos resultados son similares a los encontrados en el análisis descriptivo a nivel nacional.

17. En el estudio de la victimización también se incluyó en el análisis el estudio de la victimización a hogares. Se generó un árbol de decisión que agrupara los barrios de acuerdo a su nivel de victimización con el objetivo de estudiar la asociación que pueda existir entre la victimización a personas y la victimización a hogares. Se generaron seis grupos de barrios; el grupo con mayor victimización a hogares mostró un índice del 51% mientras el más bajo un índice del 0%. En este punto, es importante mencionar que el marco muestral de la información fue considerado a nivel de cabeceras cantonales y por tanto su uso a un nivel más desagregado intenta mostrar solamente una propuesta como referente para el estudio de este tema, pues sería poco confiable realizar inferencias con tan pocos datos para cada barrio; por eso se realizaron análisis sobre grupos de barrios.
18. La victimización a personas y hogares mostró una correlación lineal de 0,87, lo que sugiere que la victimización no solo se debe por el perfil social del individuo sino también por la posición geográfica en la que habita. El estudio profundo de la posible causa de victimización considerando estos dos factores puede generar estrategias que minimicen la exposición a la victimización. Para esto, es relevante disponer de otras características de la victimización, como son la hora, el lugar, el objeto robado, el objeto usado para la victimización y el tipo de agresión.
19. En la encuesta no se dispone del lugar del delito, información importante para corroborar si la victimización es por sector o perfil social. Se recomienda considerar esta información en estudios posteriores.

20. El análisis realizado para Guayaquil es similar al realizado para Quito. Se obtuvieron tres perfiles sociales y cinco grupos barrios. Los grupos barrios uno y tres son los más victimizados con un índice del 18%; el perfil social jóvenes es el más victimizado con el 19%.

ESTRATEGIAS PLANTEADAS

21. En Quito, la victimización del perfil solteros-trabajo se mantiene alta en la mayoría de grupos barrios (o independiente de estos), por lo que se recomendaría investigar profundamente sobre el tipo de victimización de este grupo acerca de la hora, lugar, razón y tipo de victimización con el objetivo de generar controles de vigilancia y dar información a este grupo sobre el tipo de victimización al que son expuestos en determinadas condiciones. En la encuesta se encuentran detalles del objeto robado, día de la semana, hora del día y tipo de victimización (pero no se encuentra el lugar del robo), sin embargo no es posible inferir resultados estadísticos debido a que el número de casos en cada provincia no son suficientes ya que la unidad de estudio son las cabeceras cantonales y no subconjuntos más pequeños a estos, debido a que la encuesta es a nivel nacional. Por tal razón, no es posible generar estrategias totalmente definidas para las ciudades, sin embargo pueden ser consideradas para estudios posteriores en los que los tamaños de muestra por ciudades sean más grandes.

22. En Quito, se observa que los barrios que pertenecen al grupo barrio uno y tres tienen una victimización alta en cada uno de sus perfiles sociales, por lo que podría ser una estrategia generar comités barriales de vigilancia que muestren el conocimiento de la delincuencia en el barrio, generar marchas pacíficas de rechazo que intimiden a la delincuencia al saberse conocida. Otra alternativa podría ser instalar alarmas en cada uno de los hogares con la intención de brindar apoyo en caso de ser victimizado solicitando de igual manera mayor ayuda a la policía en los diferentes sectores del barrio.

Incluso la policía del sector podría formar parte de estos comites barriales con el objetivo de ayudar a organizarlos frente a la victimización y perduren el mayor tiempo posible.

23. Si se aplicaran estrategias sobre los barrios y grupos sociales más victimizados (en este caso fueron los grupos barrios 1 y 3) y los perfiles sociales (solteros con trabajo y padres de familia) y se alcanzara una reducción del 20% de la victimización en estos grupos, se obtendría una reducción del 13,6% de la victimización de toda la ciudad, enfocándose en el 60% de los ciudadanos.

24. Con el uso de la matriz de victimización por perfil social y situación geográfica se puede generar un sistema de medida que evalúe la gestión social y de seguridad policial encargadas de controlar la inseguridad existente en la ciudad. Adicionalmente, esta matriz puede orientar sobre el tipo de gestión a realizar: por perfil social o perfil geográfico. Generar esta información en los sucesivos años permitira ver la tendencia que tiene la victimización en función de la gestión direccionada.

25. El análisis descriptivo mostró que la percepción resultó ser inversamente proporcional a la victimización, por lo que podría considerarse que los individuos que tienen alta percepción son menos expuestos a la victimización; en general, se podría considerar crear campañas que aumenten la percepción de victimización en los individuos que tengan una mayor probabilidad de ser victimizados.

BIBLIOGRAFÍA

BACHER JOHANN, WENSING KNUT, SPSS Two Step Conglomerado – A First Evaluation, Melanie, 2004-02

CABALLERO GRANADO FRANCISCO JAVIER; Modelos de Regresión Logística Incondicional; Servicio de Medicina Interna Hospital Punta de Europa Algeciras. <http://saei.org/hemero>

DIJK JAN VAN, KESTEREN JOHN VAN, SMIT PAUL; Victimización en la Perspectiva Internacional; ICESI, Instituto Ciudadano de Estudios Sobre la Inseguridad; 2008

DOBSON J. ANNETE; An introduction to Generalized Linear Models; Chapman and Hall; 1990.

LYN C THOMAS, EDELMAN DAVID B, CROOK JONATHAN N., Credit Scoring and its Applications; SIAM, Society for industrial Applied Mathematics; 2002.

MARTORI CAÑAS JUAN CARLOS; Análisis Estadístico con SPSS para Windows Volumen II Estadística Multivariante. 2003

NOVALES CINCA ALFONSO; Econometría; 1993

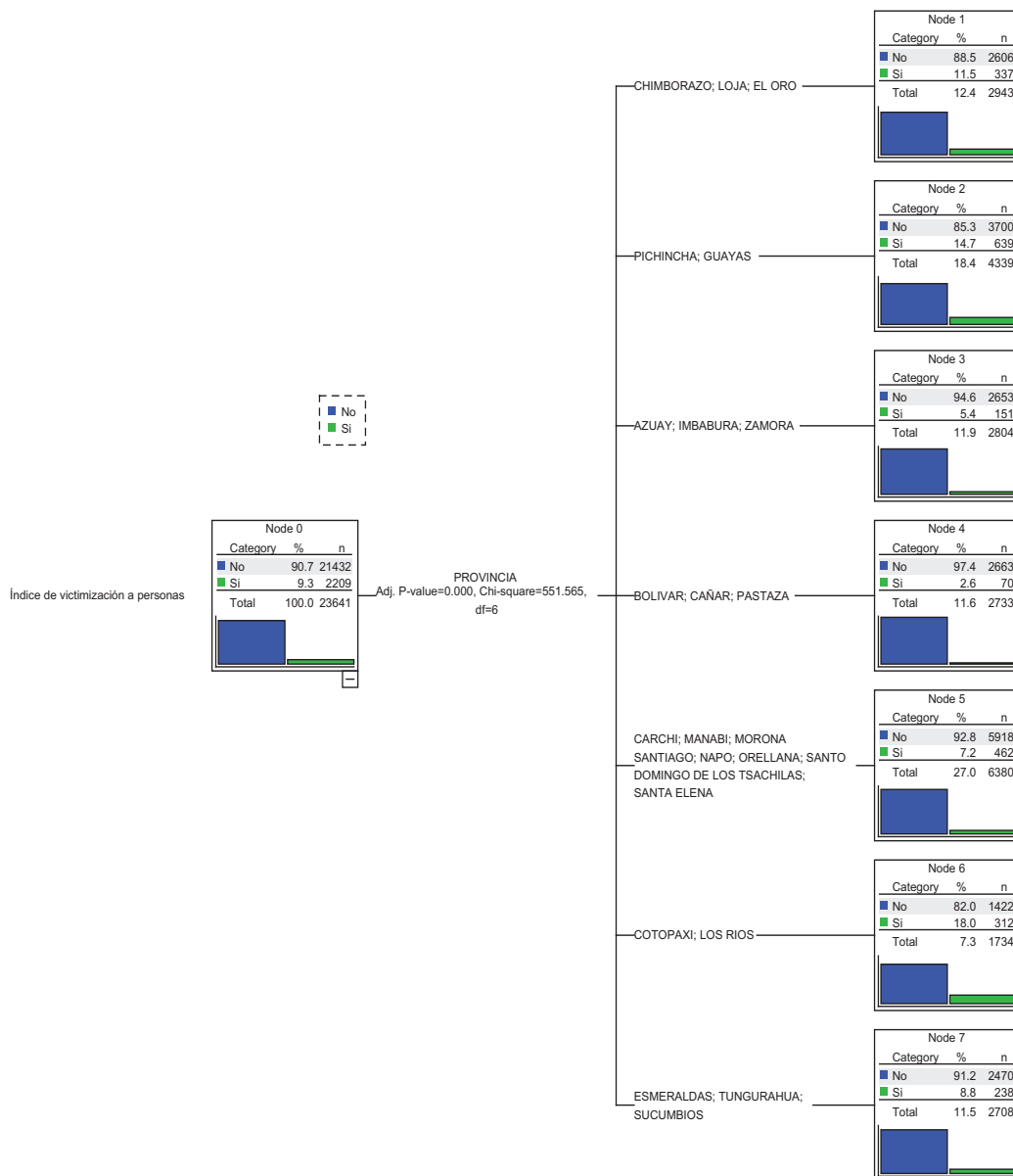
SOZZO MÁXIMO; Inseguridad, Prevención y Policía; Crearimagen; 2008.

URIEL ALDAS EZEQUIEL JOAQUÍN; Análisis Multivariante Aplicado; Thomson; 2005.

ANEXO 1

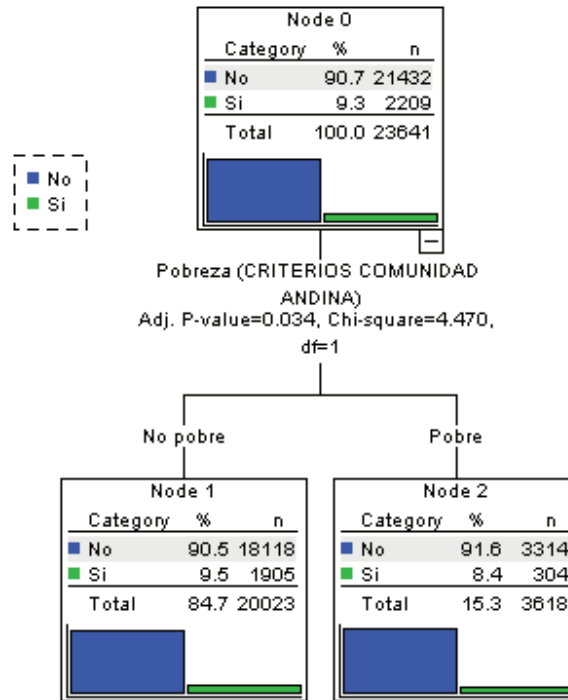
VARIABLES GENERADAS POR LOS ÁRBOLES DE DECISIÓN

PROVINCIA



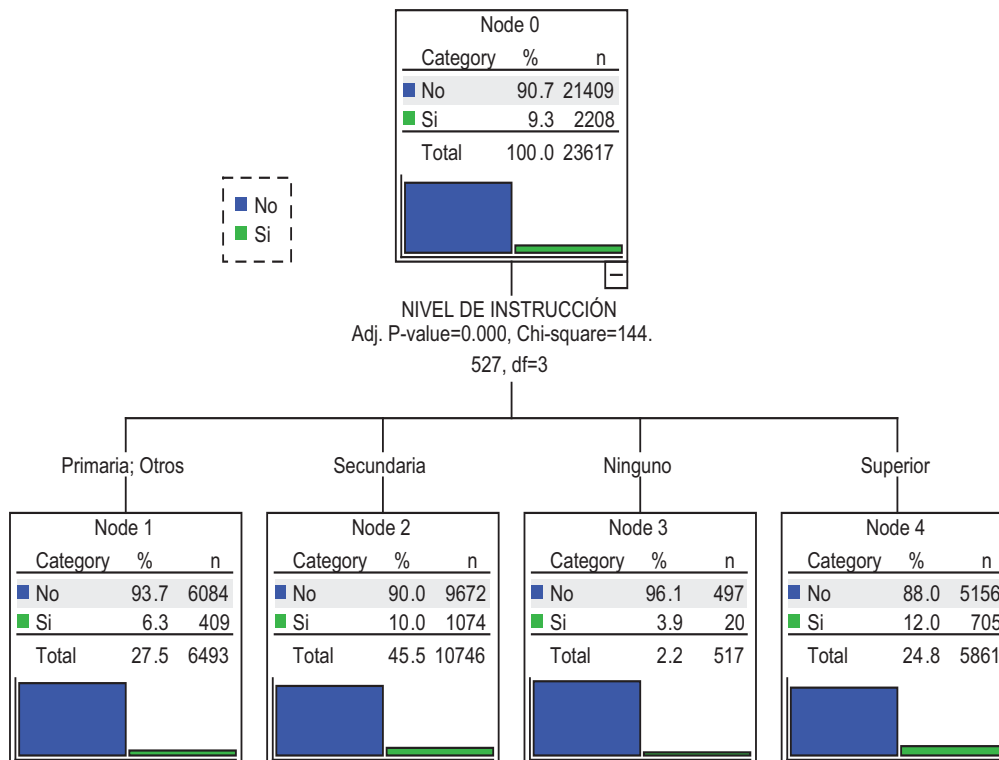
POBREZA

Índice de victimización a personas



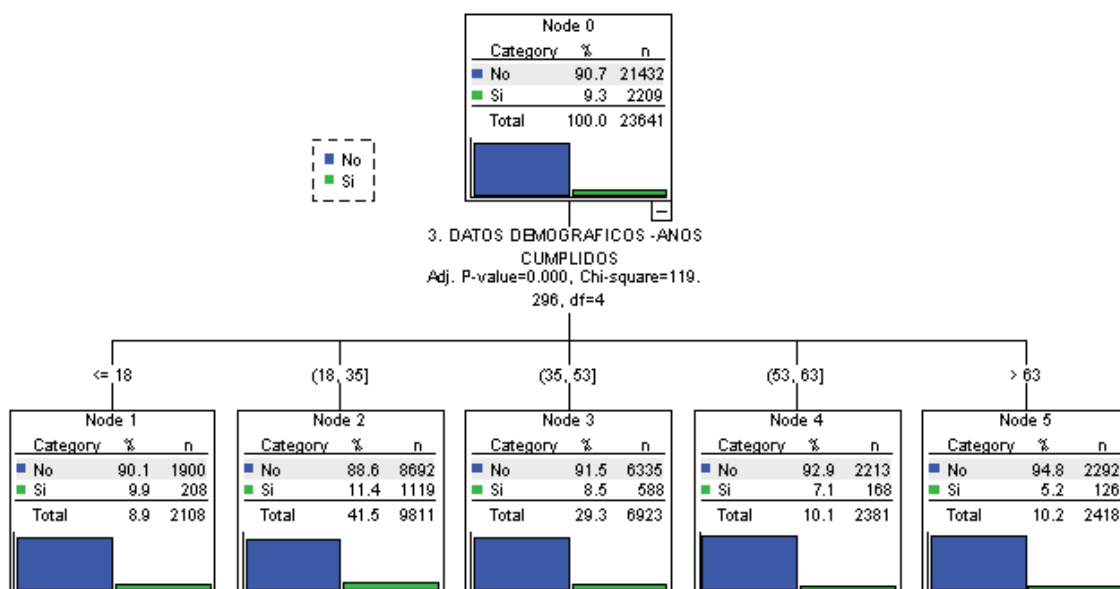
NIVEL DE INSTRUCCIÓN

Índice de victimización a personas



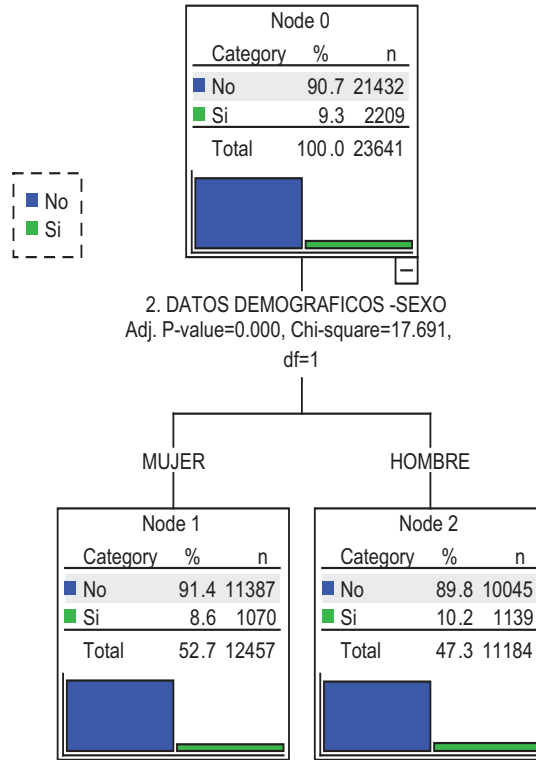
EDAD

Índice de victimización a personas



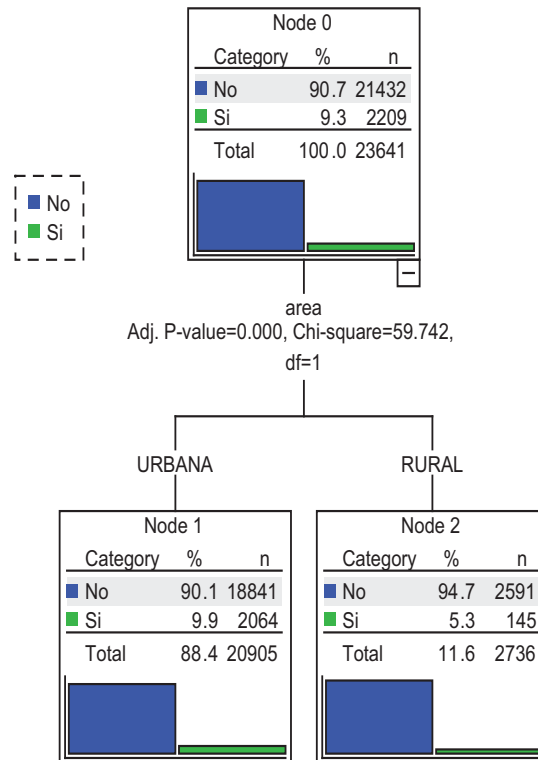
GÉNERO

Índice de victimización a personas



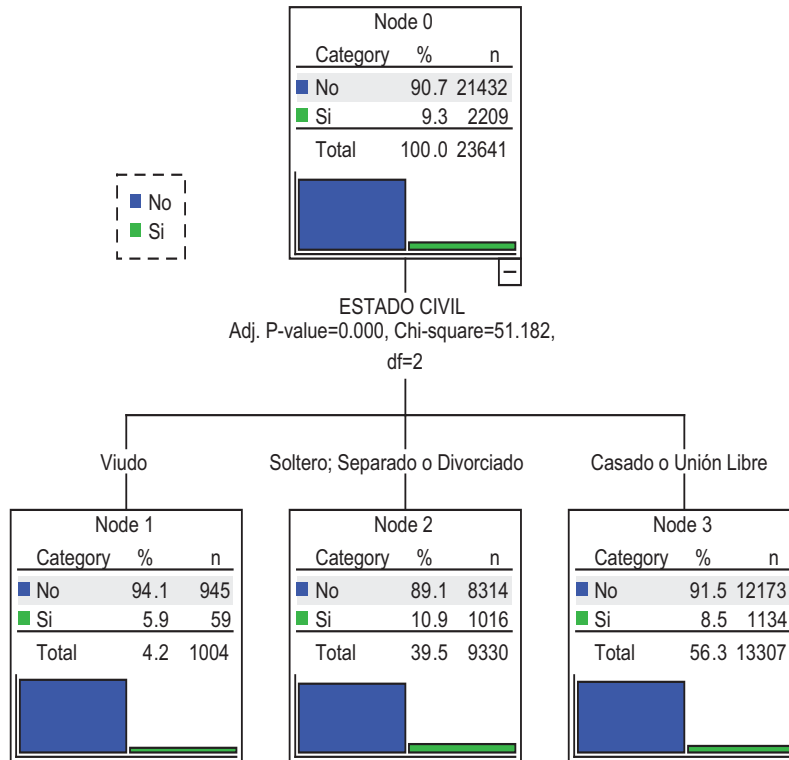
AREA GEOGRÁFICA

Índice de victimización a personas

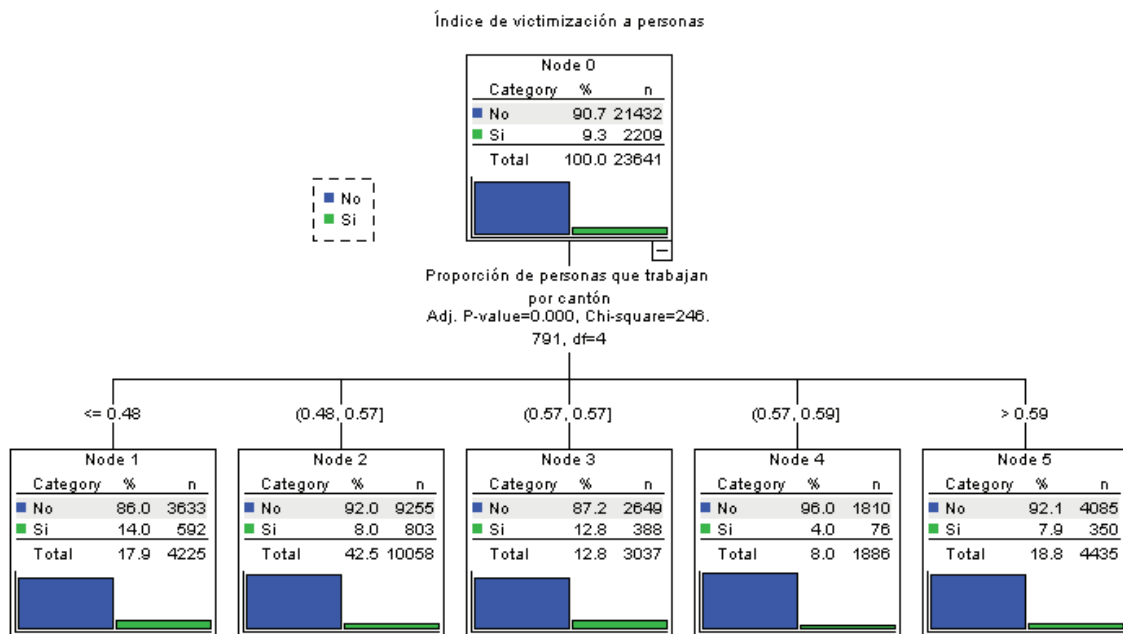


ESTADO CIVIL

Índice de victimización a personas

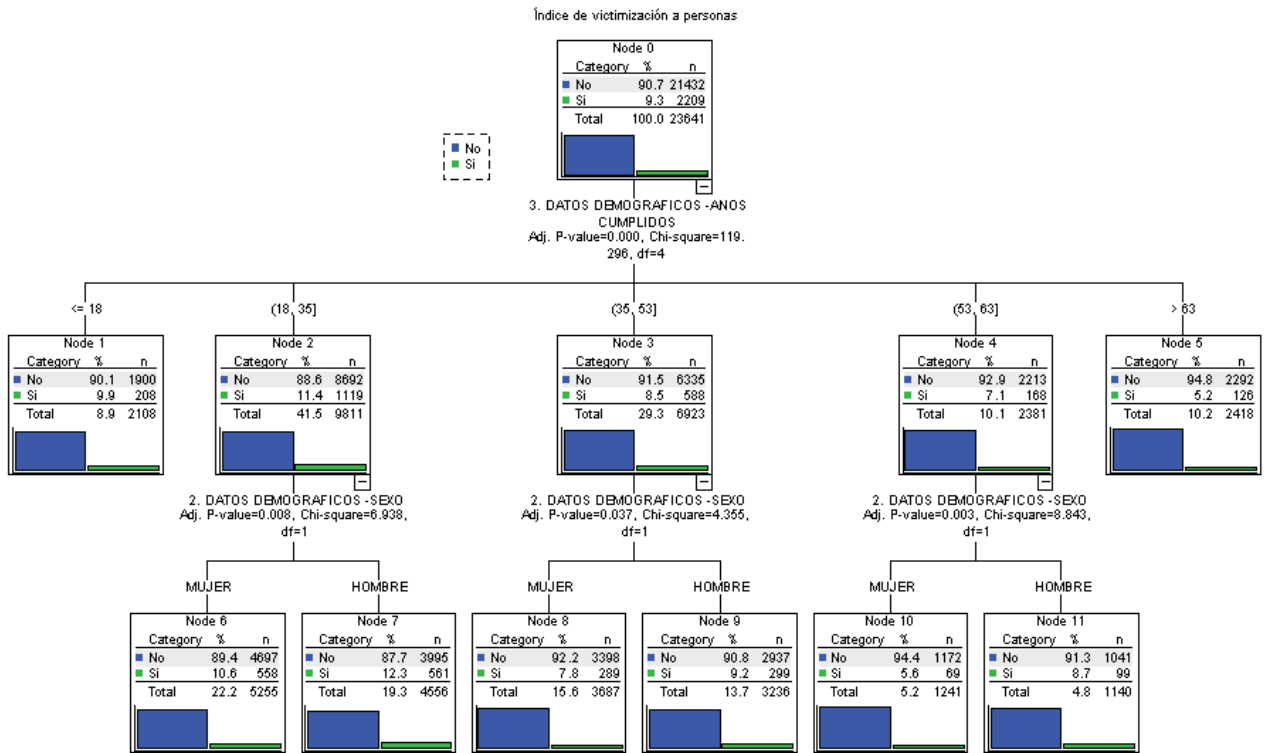


PROPORCIÓN DE PERSONAS QUE TRABAJAN POR CANTÓN

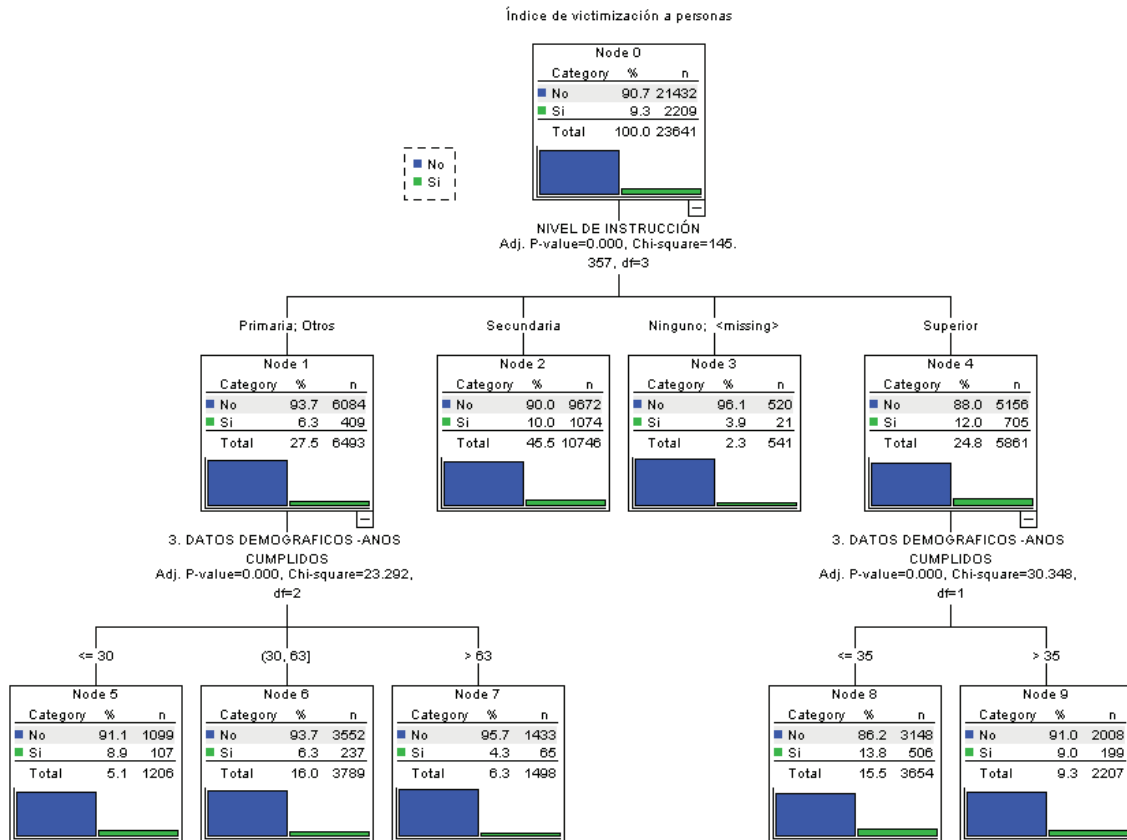


VARIABLES COMBINADAS

GÉNERO-EDAD



NIVEL DE INSTRUCCIÓN-EDAD



ANEXO 2

VARIABLES INICIALMENTE CONSIDERADAS EN EL DESARROLLO DE LA REGRESIÓN LOGÍSTICA

Se consideran todas las variables indicadoras para cada uno de los grupos definidos por los árboles de decisión sobre las variables originales.

Variables not in the Equation^a

Step	Variables	Score	df	Sig.
0	V_1	15,795	1	,000
	V_2	122,509	1	,000
	V_3	41,169	1	,000
	V_4	108,210	1	,000
	V_5	45,593	1	,000
	V_6	142,049	1	,000
	V_7	1,739	1	,187
	V_8	54,795	1	,000
	V_9	4,423	1	,035
	V_10	45,085	1	,000
	V_11	34,828	1	,000
	V_12	32,178	1	,000
	V_13	90,189	1	,000
	V_14	22,274	1	,000
	V_15	39,630	1	,000
	V_16	17,232	1	,000
	V_17	57,968	1	,000
	V_18	5,964	1	,015
	V_19	6,360	1	,012
	V_20	47,364	1	,000
	V_21	8,282	1	,004
	V_22	12,172	1	,000
	V_23	34,034	1	,000
	V_24	10,027	1	,002
	V_25	,010	1	,920
	V_26	7,746	1	,005
	V_27	,402	1	,526
	V_28	26,628	1	,000
	V_29	34,787	1	,000
	V_30	70,904	1	,000
	V_31	42,846	1	,000
	V_32	36,398	1	,000

a. Residual Chi-Squares are not computed because of redundancies.

En el paso 14 se muestra el primer conjunto de variables seleccionadas para el modelo original.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1(a)	V_6	,910	,079	133,788	1	,000	2,484
	Constant	-2,380	,029	6722,624	1	,000	,093
Step 2(b)	V_2	,815	,063	168,651	1	,000	2,260
	V_6	1,117	,081	189,060	1	,000	3,054
	Constant	-2,586	,035	5367,898	1	,000	,075
Step 3(c)	V_1	,707	,081	76,993	1	,000	2,028
	V_2	,965	,066	211,732	1	,000	2,624
	V_6	1,266	,084	227,374	1	,000	3,546
	Constant	-2,735	,041	4405,766	1	,000	,065
Step 4(d)	V_1	,712	,081	77,751	1	,000	2,038
	V_2	,978	,066	216,233	1	,000	2,658
	V_6	1,267	,084	226,413	1	,000	3,550
	V_17	,424	,054	60,713	1	,000	1,529
	Constant	-2,934	,050	3463,981	1	,000	,053
Step 5(e)	V_1	,726	,081	80,472	1	,000	2,067
	V_2	1,001	,067	224,767	1	,000	2,720
	V_6	1,265	,085	223,988	1	,000	3,543
	V_11	,481	,057	70,352	1	,000	1,617
	V_17	,546	,057	92,641	1	,000	1,727
	Constant	-3,190	,060	2797,347	1	,000	,041
Step 6(f)	V_1	,601	,082	53,608	1	,000	1,824
	V_2	,875	,068	165,320	1	,000	2,399
	V_4	-,937	,145	41,853	1	,000	,392
	V_6	1,139	,086	177,322	1	,000	3,125
	V_11	,482	,057	70,494	1	,000	1,619
	V_17	,545	,057	91,883	1	,000	1,724
	Constant	-3,064	,062	2460,257	1	,000	,047
Step 7(g)	V_1	,580	,082	49,815	1	,000	1,786
	V_2	,861	,068	159,545	1	,000	2,366
	V_4	-,948	,145	42,721	1	,000	,388
	V_6	1,132	,086	174,532	1	,000	3,103
	V_11	,575	,060	92,581	1	,000	1,777
	V_17	,491	,058	72,526	1	,000	1,634
	V_32	,358	,058	37,412	1	,000	1,430
	Constant	-3,222	,068	2240,972	1	,000	,040
Step 8(h)	V_1	,586	,082	50,656	1	,000	1,796
	V_2	,873	,068	163,375	1	,000	2,394
	V_4	-,935	,145	41,532	1	,000	,393
	V_6	1,140	,086	176,405	1	,000	3,128
	V_11	,599	,060	100,261	1	,000	1,820

	V_17	,408	,059	48,429	1	,000	1,503
	V_20	-,700	,126	30,774	1	,000	,496
	V_32	,326	,059	31,004	1	,000	1,385
	Constant	-3,133	,069	2056,071	1	,000	,044
Step 9(i)	V_1	,544	,083	43,388	1	,000	1,723
	V_2	,850	,068	154,234	1	,000	2,339
	V_4	-,933	,145	41,366	1	,000	,393
	V_6	1,142	,086	176,047	1	,000	3,132
	V_11	,595	,060	98,682	1	,000	1,812
	V_17	,405	,059	47,668	1	,000	1,500
	V_20	-,698	,126	30,506	1	,000	,498
	V_31	-,589	,109	29,345	1	,000	,555
	V_32	,325	,059	30,688	1	,000	1,384
	Constant	-3,066	,070	1929,310	1	,000	,047
Step 10(j)	V_1	,511	,083	37,911	1	,000	1,667
	V_2	,844	,068	151,790	1	,000	2,325
	V_4	-,951	,145	42,931	1	,000	,386
	V_6	1,132	,086	172,644	1	,000	3,101
	V_11	,613	,060	103,496	1	,000	1,846
	V_17	,295	,065	20,338	1	,000	1,342
	V_20	-,704	,126	31,023	1	,000	,495
	V_30	,327	,079	17,323	1	,000	1,387
	V_31	-,563	,109	26,757	1	,000	,569
	V_32	,284	,060	22,572	1	,000	1,328
	Constant	-3,059	,070	1916,925	1	,000	,047
Step 11(k)	V_1	,507	,083	37,191	1	,000	1,660
	V_2	,832	,069	147,178	1	,000	2,298
	V_4	-,955	,145	43,275	1	,000	,385
	V_6	1,116	,086	167,283	1	,000	3,052
	V_11	,633	,061	109,176	1	,000	1,883
	V_17	,243	,067	13,223	1	,000	1,274
	V_20	-,789	,128	37,802	1	,000	,454
	V_28	-,310	,090	11,857	1	,001	,734
	V_30	,312	,079	15,787	1	,000	1,366
	V_31	-,532	,109	23,702	1	,000	,587
	V_32	,265	,060	19,498	1	,000	1,303
	Constant	-2,980	,073	1664,536	1	,000	,051
Step 12(l)	V_1	,506	,083	37,081	1	,000	1,659
	V_2	,835	,069	147,874	1	,000	2,304
	V_4	-,955	,145	43,223	1	,000	,385
	V_6	1,111	,086	165,534	1	,000	3,037
	V_11	,805	,081	98,251	1	,000	2,236
	V_12	,319	,100	10,160	1	,001	1,375
	V_17	,252	,067	14,223	1	,000	1,287
	V_20	-,771	,129	36,009	1	,000	,462
	V_28	-,327	,090	13,180	1	,000	,721
	V_30	,336	,079	18,098	1	,000	1,400

Step 13(m)	V_31	-,526	,109	23,141	1	,000	,591
	V_32	,417	,076	29,785	1	,000	1,518
	Constant	-3,200	,101	1002,436	1	,000	,041
	V_1	,431	,086	25,050	1	,000	1,538
	V_2	,760	,072	111,560	1	,000	2,139
	V_3	-,341	,111	9,477	1	,002	,711
	V_4	-1,030	,147	49,162	1	,000	,357
	V_6	1,037	,089	135,680	1	,000	2,820
	V_11	,799	,081	97,011	1	,000	2,224
	V_12	,313	,100	9,806	1	,002	1,367
	V_17	,247	,067	13,623	1	,000	1,280
	V_20	-,766	,129	35,463	1	,000	,465
	V_28	-,326	,090	13,092	1	,000	,722
	V_30	,351	,079	19,592	1	,000	1,420
Step 14(n)	V_31	-,524	,109	22,985	1	,000	,592
	V_32	,416	,076	29,643	1	,000	1,516
	Constant	-3,123	,103	910,801	1	,000	,044
	V_1	,233	,108	4,649	1	,031	1,263
	V_2	,563	,098	33,244	1	,000	1,755
	V_3	-,539	,129	17,461	1	,000	,583
	V_4	-1,228	,161	58,131	1	,000	,293
	V_5	-,294	,102	8,296	1	,004	,746
	V_6	,839	,111	57,363	1	,000	2,314
	V_11	,797	,081	96,362	1	,000	2,218
	V_12	,309	,100	9,541	1	,002	1,361
	V_17	,255	,067	14,498	1	,000	1,290
	V_20	-,772	,129	36,020	1	,000	,462
	V_28	-,319	,090	12,555	1	,000	,727
V_30	,343	,079	18,641	1	,000	1,408	
V_31	-,524	,109	22,928	1	,000	,592	
V_32	,413	,076	29,261	1	,000	1,512	
Constant	-2,924	,123	567,500	1	,000	,054	

a Variable(s) entered on step 1: V_6.

b Variable(s) entered on step 2: V_2.

c Variable(s) entered on step 3: V_1.

d Variable(s) entered on step 4: V_17.

e Variable(s) entered on step 5: V_11.

f Variable(s) entered on step 6: V_4.

g Variable(s) entered on step 7: V_32.

h Variable(s) entered on step 8: V_20.

i Variable(s) entered on step 9: V_31.

j Variable(s) entered on step 10: V_30.

k Variable(s) entered on step 11: V_28.

l Variable(s) entered on step 12: V_12.

m Variable(s) entered on step 13: V_3.

n Variable(s) entered on step 14: V_5.

Sobre este primer modelo se estudia la significancia y la polaridad en la base de prueba para posteriormente seleccionar las variables del modelo final. A continuación se presenta el análisis de significancia de cada una de las variables en la muestra de prueba:

V_1 * Índice de victimización a personas

Crosstab

Count

Indicador_Muestra			Índice de victimización a personas		Total
			No	Si	No
,00	V_1	,00	5610	581	6191
		1,00	812	101	913
	Total		6422	682	7104
1,00	V_1	,00	13216	1291	14507
		1,00	1794	236	2030
	Total		15010	1527	16537

Chi-Square Tests

Indicador_Muestra		Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
,00	Pearson Chi-Square	2,581(b)	1	,108		
	Continuity Correction(a)	2,391	1	,122		
	Likelihood Ratio	2,488	1	,115		
	Fisher's Exact Test				,117	,063
	Linear-by-Linear Association	2,581	1	,108		
	N of Valid Cases	7104				
	1,00	Pearson Chi-Square	15,795(c)	1	,000	
Continuity Correction(a)		15,471	1	,000		
Likelihood Ratio		14,864	1	,000		
Fisher's Exact Test					,000	,000
Linear-by-Linear Association		15,794	1	,000		
N of Valid Cases		16537				

a Computed only for a 2x2 table

b 0 cells (.0%) have expected count less than 5. The minimum expected count is 87,65.

c 0 cells (.0%) have expected count less than 5. The minimum expected count is 187,45.

V_2 * Índice de victimización a personas

Crosstab

Count

Indicador_Muestra			Índice de victimización a personas		Total
			No	Si	No
,00	V_2	,00	5272	477	5749
		1,00	1150	205	1355
	Total		6422	682	7104
1,00	V_2	,00	12460	1093	13553
		1,00	2550	434	2984
	Total		15010	1527	16537

Chi-Square Tests

Indicador_Muestra		Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
,00	Pearson Chi-Square	58,977(b)	1	,000		
	Continuity Correction(a)	58,192	1	,000		
	Likelihood Ratio	53,066	1	,000		
	Fisher's Exact Test				,000	,000
	Linear-by-Linear Association	58,969	1	,000		
	N of Valid Cases	7104				
	1,00	Pearson Chi-Square	122,509(c)	1	,000	
Continuity Correction(a)		121,737	1	,000		
Likelihood Ratio		109,885	1	,000		
Fisher's Exact Test					,000	,000
Linear-by-Linear Association		122,502	1	,000		
N of Valid Cases		16537				

a Computed only for a 2x2 table

b 0 cells (.0%) have expected count less than 5. The minimum expected count is 130,08.

c 0 cells (.0%) have expected count less than 5. The minimum expected count is 275,54.

V_3 * Índice de victimización a personas

Crosstab

Count

Indicador_Muestra			Índice de victimización a personas		Total
			No	Si	No
,00	V_3	,00	5677	639	6316
		1,00	745	43	788
	Total		6422	682	7104
1,00	V_3	,00	13102	1419	14521
		1,00	1908	108	2016
	Total		15010	1527	16537

Chi-Square Tests

Indicador_Muestra		Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
,00	Pearson Chi-Square	17,533(b)	1	,000		
	Continuity Correction(a)	17,000	1	,000		
	Likelihood Ratio	20,105	1	,000		
	Fisher's Exact Test				,000	,000
	Linear-by-Linear Association	17,530	1	,000		
	N of Valid Cases	7104				
1,00	Pearson Chi-Square	41,169(c)	1	,000		
	Continuity Correction(a)	40,644	1	,000		
	Likelihood Ratio	46,934	1	,000		
	Fisher's Exact Test				,000	,000
	Linear-by-Linear Association	41,166	1	,000		
	N of Valid Cases	16537				

a Computed only for a 2x2 table

b 0 cells (.0%) have expected count less than 5. The minimum expected count is 75,65.

c 0 cells (.0%) have expected count less than 5. The minimum expected count is 186,15.

Chi-Square Tests

Indicador_Muestra		Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
,00	Pearson Chi-Square	60,128(b)	1	,000		
	Continuity Correction(a)	59,145	1	,000		
	Likelihood Ratio	83,129	1	,000		
	Fisher's Exact Test				,000	,000
	Linear-by-Linear Association	60,120	1	,000		
	N of Valid Cases	7104				
1,00	Pearson Chi-Square	108,210(c)	1	,000		
	Continuity Correction(a)	107,342	1	,000		
	Likelihood Ratio	140,663	1	,000		
	Fisher's Exact Test				,000	,000
	Linear-by-Linear Association	108,203	1	,000		
	N of Valid Cases	16537				

a Computed only for a 2x2 table

b 0 cells (.0%) have expected count less than 5. The minimum expected count is 76,90.

c 0 cells (.0%) have expected count less than 5. The minimum expected count is 178,40.

V_5 * Índice de victimización a personas

Crosstab

Count

Indicador_Muestra			Índice de victimización a personas		Total
			No	Si	No
,00	V_5	,00	4650	519	5169
		1,00	1772	163	1935
	Total		6422	682	7104
1,00	V_5	,00	10864	1228	12092
		1,00	4146	299	4445
	Total		15010	1527	16537

Chi-Square Tests

Indicador_Muestra		Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
,00	Pearson Chi-Square	4,241(b)	1	,039		
	Continuity Correction(a)	4,057	1	,044		
	Likelihood Ratio	4,346	1	,037		
	Fisher's Exact Test				,042	,021
	Linear-by-Linear Association	4,240	1	,039		
	N of Valid Cases	7104				
1,00	Pearson Chi-Square	45,593(c)	1	,000		
	Continuity Correction(a)	45,184	1	,000		
	Likelihood Ratio	48,401	1	,000		
	Fisher's Exact Test				,000	,000
	Linear-by-Linear Association	45,590	1	,000		
	N of Valid Cases	16537				

a Computed only for a 2x2 table

b 0 cells (.0%) have expected count less than 5. The minimum expected count is 185,76.

c 0 cells (.0%) have expected count less than 5. The minimum expected count is 410,44.

V_6 * Índice de victimización a personas

Crosstab

Count

Indicador_Muestra			Índice de victimización a personas		Total
			No	Si	No
,00	V_6	,00	6000	600	6600
		1,00	422	82	504
	Total		6422	682	7104
1,00	V_6	,00	14010	1297	15307
		1,00	1000	230	1230
	Total		15010	1527	16537

Chi-Square Tests

Indicador	Muestra	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)		
,00	Pearson Chi-Square	27,806(b)	1	,000				
	Continuity Correction(a)	26,985	1	,000				
	Likelihood Ratio	23,833	1	,000				
	Fisher's Exact Test						,000	,000
	Linear-by-Linear Association	27,802	1	,000				
	N of Valid Cases	7104						
1,00	Pearson Chi-Square	142,049(c)	1	,000				
	Continuity Correction(a)	140,832	1	,000				
	Likelihood Ratio	115,155	1	,000				
	Fisher's Exact Test						,000	,000
	Linear-by-Linear Association	142,041	1	,000				
	N of Valid Cases	16537						

a Computed only for a 2x2 table

b 0 cells (,0%) have expected count less than 5. The minimum expected count is 48,39.

c 0 cells (,0%) have expected count less than 5. The minimum expected count is 113,58.

V_11 * Índice de victimización a personas

Crosstab

Count

Indicador	Muestra		Índice de victimización a personas		Total
			No	Si	No
,00	V_11	,00	4016	393	4409
		1,00	2406	289	2695
	Total		6422	682	7104
1,00	V_11	,00	9563	856	10419
		1,00	5447	671	6118
	Total		15010	1527	16537

Chi-Square Tests

Indicador	Muestra	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
,00	Pearson Chi-Square	6,314(b)	1	,012	,013	,007
	Continuity Correction(a)	6,107	1	,013		
	Likelihood Ratio	6,238	1	,013		
	Fisher's Exact Test					
	Linear-by-Linear Association	6,313	1	,012		
	N of Valid Cases	7104				
1,00	Pearson Chi-Square	34,828(c)	1	,000	,000	,000
	Continuity Correction(a)	34,501	1	,000		
	Likelihood Ratio	34,151	1	,000		
	Fisher's Exact Test					
	Linear-by-Linear Association	34,826	1	,000		
	N of Valid Cases	16537				

a Computed only for a 2x2 table

b 0 cells (,0%) have expected count less than 5. The minimum expected count is 258,73.

c 0 cells (,0%) have expected count less than 5. The minimum expected count is 564,93.

V_12 * Índice de victimización a personas

Crosstab

Count

Indicador	Muestra		Índice de victimización a personas		Total
			No	Si	No
,00	V_12	,00	4624	538	5162
		1,00	1798	144	1942
	Total		6422	682	7104
1,00	V_12	,00	10784	1201	11985
		1,00	4226	326	4552
	Total		15010	1527	16537

Chi-Square Tests

Indicador	Muestra	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)		
,00	Pearson Chi-Square	14,705(b)	1	,000				
	Continuity Correction(a)	14,360	1	,000				
	Likelihood Ratio	15,428	1	,000				
	Fisher's Exact Test						,000	,000
	Linear-by-Linear Association	14,703	1	,000				
	N of Valid Cases	7104						
1,00	Pearson Chi-Square	32,178(c)	1	,000				
	Continuity Correction(a)	31,838	1	,000				
	Likelihood Ratio	33,734	1	,000				
	Fisher's Exact Test						,000	,000
	Linear-by-Linear Association	32,176	1	,000				
	N of Valid Cases	16537						

a Computed only for a 2x2 table

b 0 cells (.0%) have expected count less than 5. The minimum expected count is 186,44.

c 0 cells (.0%) have expected count less than 5. The minimum expected count is 420,32.

V_17 * Índice de victimización a personas

Crosstab

Count

Indicador	Muestra		Índice de victimización a personas		Total
			No	Si	No
,00	V_17	,00	3854	340	4194
		1,00	2568	342	2910
	Total		6422	682	7104
1,00	V_17	,00	8886	750	9636
		1,00	6124	777	6901
	Total		15010	1527	16537

Chi-Square Tests

Indicador	Muestra	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)		
,00	Pearson Chi-Square	26,311(b)	1	,000				
	Continuity Correction(a)	25,893	1	,000				
	Likelihood Ratio	25,921	1	,000				
	Fisher's Exact Test						,000	,000
	Linear-by-Linear Association	26,308	1	,000				
	N of Valid Cases	7104						
1,00	Pearson Chi-Square	57,968(c)	1	,000				
	Continuity Correction(a)	57,554	1	,000				
	Likelihood Ratio	57,205	1	,000				
	Fisher's Exact Test						,000	,000
	Linear-by-Linear Association	57,964	1	,000				
	N of Valid Cases	16537						

a Computed only for a 2x2 table

b 0 cells (.0%) have expected count less than 5. The minimum expected count is 279,37.

c 0 cells (.0%) have expected count less than 5. The minimum expected count is 637,23.

V_20 * Índice de victimización a personas

Crosstab

Count

Indicador	Muestra		Índice de victimización a personas		Total
			No	Si	No
,00	V_20	,00	5710	632	6342
		1,00	712	50	762
	Total		6422	682	7104
1,00	V_20	,00	13430	1451	14881
		1,00	1580	76	1656
	Total		15010	1527	16537

Chi-Square Tests

Indicador_Muestra		Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)		
,00	Pearson Chi-Square	9,081(b)	1	,003				
	Continuity Correction(a)	8,693	1	,003				
	Likelihood Ratio	9,986	1	,002				
	Fisher's Exact Test						,002	,001
	Linear-by-Linear Association	9,079	1	,003				
	N of Valid Cases	7104						
1,00	Pearson Chi-Square	47,364(c)	1	,000				
	Continuity Correction(a)	46,750	1	,000				
	Likelihood Ratio	56,093	1	,000				
	Fisher's Exact Test						,000	,000
	Linear-by-Linear Association	47,361	1	,000				
	N of Valid Cases	16537						

a Computed only for a 2x2 table

b 0 cells (.0%) have expected count less than 5. The minimum expected count is 73,15.

c 0 cells (.0%) have expected count less than 5. The minimum expected count is 152,91.

V_28 * Índice de victimización a personas

Crosstab

Count

Indicador_Muestra			Índice de victimización a personas		Total
			No	Si	No
,00	V_28	,00	5377	622	5999
		1,00	1045	60	1105
	Total		6422	682	7104
1,00	V_28	,00	12503	1350	13853
		1,00	2507	177	2684
	Total		15010	1527	16537

Chi-Square Tests

Indicador_Muestra		Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
,00	Pearson Chi-Square	26,223(b)	1	,000		
	Continuity Correction(a)	25,657	1	,000		
	Likelihood Ratio	29,860	1	,000		
	Fisher's Exact Test				,000	,000
	Linear-by-Linear Association	26,219	1	,000		
	N of Valid Cases	7104				
1,00	Pearson Chi-Square	26,628(c)	1	,000		
	Continuity Correction(a)	26,253	1	,000		
	Likelihood Ratio	28,802	1	,000		
	Fisher's Exact Test				,000	,000
	Linear-by-Linear Association	26,626	1	,000		
	N of Valid Cases	16537				

a Computed only for a 2x2 table

b 0 cells (.0%) have expected count less than 5. The minimum expected count is 106,08.

c 0 cells (.0%) have expected count less than 5. The minimum expected count is 247,84.

V_30 * Índice de victimización a personas

Crosstab

Count

Indicador_Muestra			Índice de victimización a personas		Total
			No	Si	No
,00	V_30	,00	5486	526	6012
		1,00	936	156	1092
	Total		6422	682	7104
1,00	V_30	,00	12798	1177	13975
		1,00	2212	350	2562
	Total		15010	1527	16537

Crosstab

Count

Indicador_Muestra			Índice de victimización a personas		Total
			No	Si	No
,00	V_30	,00	5486	526	6012
		1,00	936	156	1092
	Total		6422	682	7104
1,00	V_30	,00	12798	1177	13975
		1,00	2212	350	2562
	Total		15010	1527	16537

V_31 * Índice de victimización a personas**Crosstab**

Count

Indicador_Muestra			Índice de victimización a personas		Total
			No	Si	No
,00	V_31	,00	5649	636	6285
		1,00	773	46	819
	Total		6422	682	7104
1,00	V_31	,00	13192	1428	14620
		1,00	1818	99	1917
	Total		15010	1527	16537

Chi-Square Tests

Indicador_Muestra		Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)		
,00	Pearson Chi-Square	16,927(b)	1	,000				
	Continuity Correction(a)	16,412	1	,000				
	Likelihood Ratio	19,272	1	,000				
	Fisher's Exact Test						,000	,000
	Linear-by-Linear Association	16,925	1	,000				
	N of Valid Cases	7104						
	1,00	Pearson Chi-Square	42,846(c)	1			,000	
Continuity Correction(a)	42,299	1	,000					
Likelihood Ratio	49,297	1	,000					
Fisher's Exact Test				,000	,000			
Linear-by-Linear Association	42,844	1	,000					
N of Valid Cases	16537							

a Computed only for a 2x2 table

b 0 cells (.0%) have expected count less than 5. The minimum expected count is 78,63.

c 0 cells (.0%) have expected count less than 5. The minimum expected count is 177,01.

V_32 * Índice de victimización a personas

Crosstab

Count

Indicador_Muestra			Índice de victimización a personas		Total
			No	Si	No
,00	V_32	,00	3918	378	4296
		1,00	2504	304	2808
	Total		6422	682	7104
1,00	V_32	,00	9200	815	10015
		1,00	5810	712	6522
	Total		15010	1527	16537

Chi-Square Tests

Indicador	Muestra	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
,00	Pearson Chi-Square	8,042(b)	1	,005		
	Continuity Correction(a)	7,810	1	,005		
	Likelihood Ratio	7,951	1	,005		
	Fisher's Exact Test				,005	,003
	Linear-by-Linear Association	8,041	1	,005		
	N of Valid Cases	7104				
1,00	Pearson Chi-Square	36,398(c)	1	,000		
	Continuity Correction(a)	36,067	1	,000		
	Likelihood Ratio	35,838	1	,000		
	Fisher's Exact Test				,000	,000
	Linear-by-Linear Association	36,396	1	,000		
	N of Valid Cases	16537				

a Computed only for a 2x2 table

b 0 cells (,0%) have expected count less than 5. The minimum expected count is 269,57.

c 0 cells (,0%) have expected count less than 5. The minimum expected count is 602,23.

Las variables que presentan un p-valor menor a 0,05 y muestran estar correctamente polarizadas son seleccionadas para el modelo final que se presenta en el capítulo 3.

ANEXO 3:

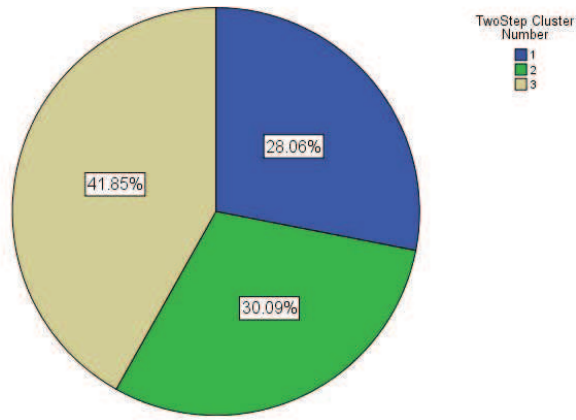
**DISTRIBUCIONES DE CONGLOMERADOS PARA
GUAYAQUIL**

A continuación se presentan los cuadros del análisis de conglomerados para la ciudad de Guayaquil.

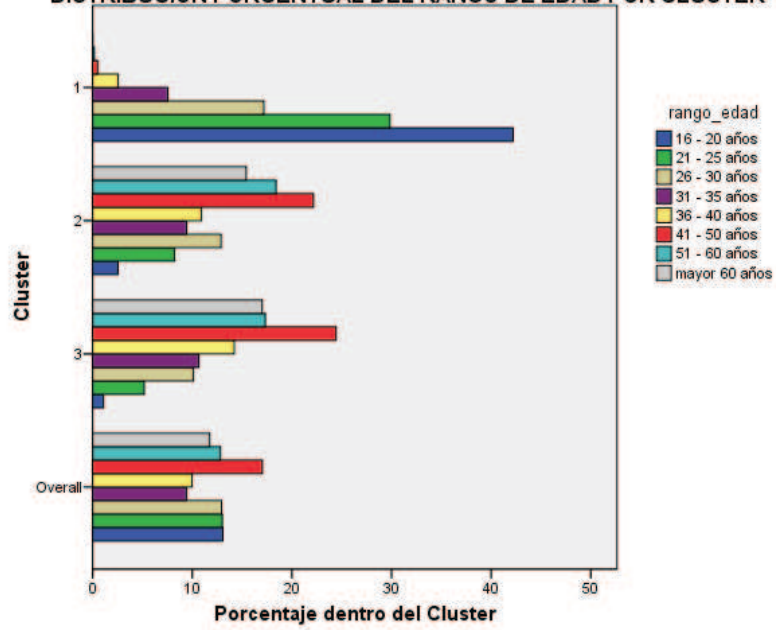
Auto-Conglomeradoing

Number of Conglomerados	Schwarz's Bayesian Criterion (BIC)	BIC Change(a)	Ratio of BIC Changes(b)	Ratio of Distance Measures(c)
1	28106,541			
2	24469,229	-3637,313	1,000	1,699
3	22388,119	-2081,110	,572	1,669
4	21199,515	-1188,604	,327	1,169
5	20203,669	-995,846	,274	1,422
6	19546,550	-657,118	,181	1,027
7	18910,215	-636,336	,175	1,049
8	18310,256	-599,959	,165	1,298
9	17881,575	-428,681	,118	1,125
10	17516,611	-364,964	,100	1,063
11	17181,784	-334,826	,092	1,048
12	16869,109	-312,675	,086	1,029
13	16569,342	-299,767	,082	1,077
14	16301,553	-267,789	,074	1,186
15	16098,696	-202,858	,056	1,074

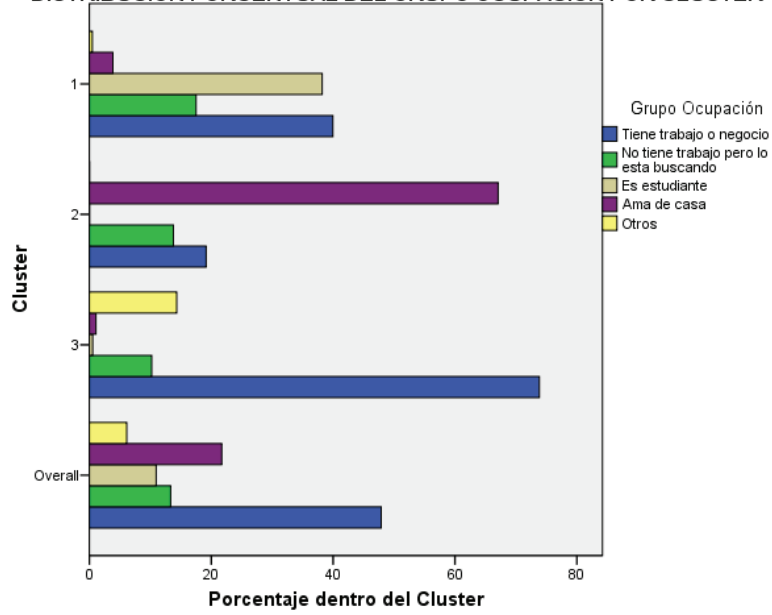
DISTRIBUCIÓN CLUSTER



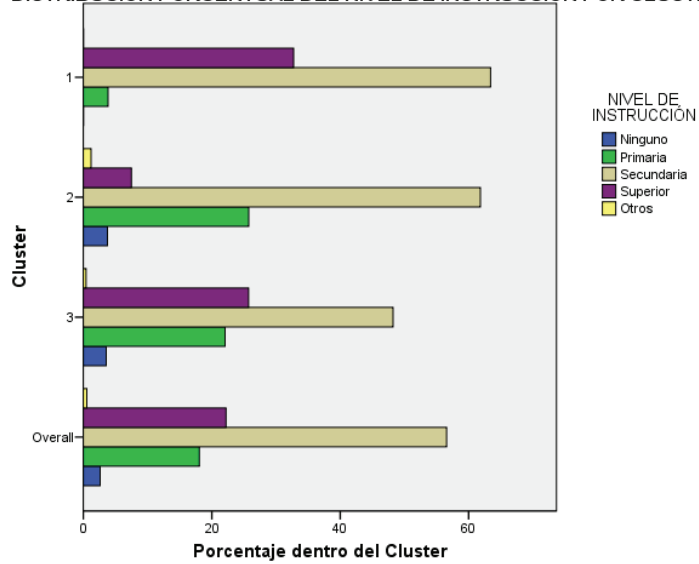
DISTRIBUCIÓN PORCENTUAL DEL RANGO DE EDAD POR CLUSTER



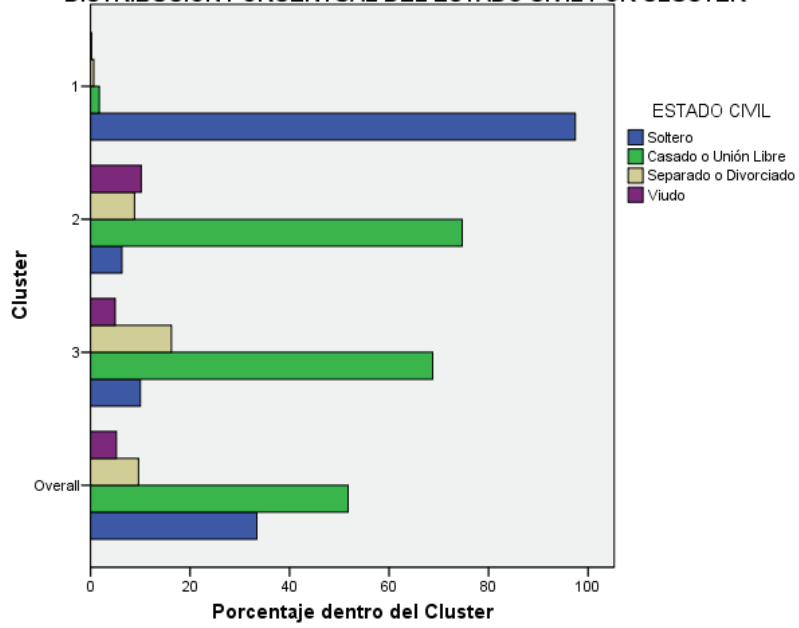
DISTRIBUCIÓN PORCENTUAL DEL GRUPO OCUPACIÓN POR CLUSTER



DISTRIBUCIÓN PORCENTUAL DEL NIVEL DE INSTRUCCIÓN POR CLUSTER



DISTRIBUCIÓN PORCENTUAL DEL ESTADO CIVIL POR CLUSTER



DISTRIBUCIÓN PORCENTUAL DEL GÉNERO POR CLUSTER

