

ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL

**Instituto de Ciencias Matemáticas**

**Ingeniería en Estadística Informática**

 **“Modelo de Detección de Fraudes en Aseguramiento de Vehículos utilizando Componentes Principales y Análisis RIDIT”**

**TESIS DE GRADO**

Previa a la obtención del Título de:

 INGENIERA EN ESTADÍSTICA INFORMÁTICA

Presentada por:

**Heydi Mariana Roa López**

**GUAYAQUIL - ECUADOR**

**AÑO**

**2004**

**AGRADECIMIENTO**

A todas las personas que de una u otra manera colaboraron en la realización de este trabajo y específicamente a la gran ayuda de mi Director el Msc. Fernando Sandoya Sánchez.

**DEDICATORIA**

A Dios

A mis padres

A mis hermanos

Y a mis familiares y amigos

**TRIBUNAL DE GRADUACIÓN**

|  |  |
| --- | --- |
| **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_****ING. PABLO ÁLVAREZ** **PRESIDENTE DEL TRIBUNAL** | **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_****MSC. FERNANDO SANDOYA****DIRECTOR DE TESIS** |
|  |  |
|  |  |
| **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_****ING. JORGE FERNÁNDEZ****VOCAL** | **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_****ING. SORAYA SOLÍS****VOCAL** |
|  |  |
|  |  |

**DECLARACIÓN EXPRESA**

### “La responsabilidad del contenido de esta tesis de grado, me corresponde exclusivamente; y el patrimonio intelectual de la misma a la ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL”

(Reglamento de graduación de la ESPOL)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Heydi Mariana Roa López

**RESUMEN**

El presente trabajo tiene como propósito proporcionar una metodología dirigida a cuantificar la probabilidad de fraude en las declaraciones de siniestros vehiculares, denominada Análisis PRIDIT, específicamente cuando se tienen variables cualitativas involucradas puesto que en el Ecuador no son muy numerosos los estudios dedicados a la detección de fraudes con este tipo de variables.

En el primer capítulo se presenta información acerca de lo que representa el fraude en el mercado asegurador de vehículos, en el segundo capítulo se da a conocer los antecedentes del cobro de siniestros en los seguros de automóviles en nuestro país, Ecuador, desde 1.993 hasta 2.003; así como también la situación actual de este tipo de seguro frente a los demás ramos de seguros del mercado ecuatoriano. El tercer capítulo presenta información del Análisis de Componentes Principales, técnica que es utilizada para el desarrollo de la nueva metodología aquí propuesta. En el cuarto capítulo se realiza la descripción, desarrollo y presentación de resultados de la aplicación del Análisis PRIDIT, para finalmente presentar las respectivas conclusiones y recomendaciones.

**INDICE GENERAL**

**Pág.**

RESUMEN II

INDICE GENERAL III

ABREVIATURAS IV

SIMBOLOGÍA V

INDICE DE TABLAS VI

INDICE DE CUADROS VII

INDICE DE GRÁFICOS VIII

INDICE DE FIGURAS IX

INTRODUCCIÓN 1

1. **EL FRAUDE EN EL MERCADO ASEGURADOR DE AUTOMÓVILES**
	1. 1.1.Introducción…………………………………………………………….... 3
	2. 1.2.Técnica estadística para validación de los indicadores de fraude…. 12
2. **COBRO DE SINIESTROS EN SEGUROS DE VEHÍCULOS DURANTE LA ÚLTIMA DÉCADA EN EL ECUADOR**
	1. 2.1. Introducción……………………………………………………………….15
	2. 2.2..Información Contable de Siniestros Pagados en Seguros de Vehículos de los Archivos de la Superintendencia de Bancos………………………………………………………………..…. 16
	3. 2.3. Cálculo de Índices de los Montos Totales de Siniestros Pagados… 22
	4. 2.4. Datos de Seguros Equinoccial del Ecuador sobre Pagos de Siniestros en Seguros Vehiculares y Desglose por Causas……………...……. 31
3. **ANÁLISIS DE COMPONENTES PRINCIPALES**
	1. 3.1. Introducción……………………………………………………………… 36
	2. 3.2. Objetivos del Análisis de Componentes Principales………………… 38
	3. 3.3. Obtención de las Componentes Principales…………………………. 39
	4. 3.4. Varianza Explicada……………………………………………………… 42
	5. 3.5. Calificaciones de los componentes principales……………………… 44
	6. 3.6. Estimación de Componentes Principales…………………………….. 45
	7. 3.7. Determinación del Número de Componentes Principales………….. 47

 3.7.1. Método 1……………..……………….………………………….. 48

 3.7.2. Método 2……………….…………..….…………………………. 49

* 1. 3.8. Resultados Útiles sobre ACP………………………………………….. 50

 3.8.1. Eigenvalores iguales a cero……………..……………………... 50

 3.8.2. Vectores de Carga de Componentes…………………….…… 52

* 1. 3.9. ACP sobre la Matriz de Correlaciones Ρ……………………………... 54

 3.9.1. Datos Estandarizados (valores Z)……………….…………..... 56

* + - 1. Calificaciones de Componentes Principales………. 57
			2. Vectores de Correlaciones de Componentes……... 58
			3. Matriz de Correlaciones de la Muestra…………….. 58
			4. Determinación del Número de Componentes Principales…………………………………………….. 59
	1. Posibles Interpretaciones de las Componentes Principales…….. 60
	2. El Uso de Componentes en Análisis Subsecuentes……………… 62
1. **USO DE COMPONENTES PRINCIPALES Y PUNTUACIONES RIDIT EN LA DETECCIÓN DE FRAUDE EN EL COBRO DE SEGUROS DE VEHÍCULOS**
	1. 4.1. ¿Qué son los RIDIT’S y el Análisis PRIDIT?………………..…… 64
	2. Tipo de Datos….……………………………………………………. 66
		1. Descripción y Codificación de la Variables…...………..... 68
	3. Validez de las Variables Instrumentales……………………….... 72
	4. Apreciación Global de PRIDIT en el Caso de Datos Ordinales Discretos……..……………………………………………………. 73
	5. Asignación de Puntuaciones Numéricas a las Categorías de Variables Ordinales Cualitativas………………..………………. 74
	6. Desarrollo de una Medida de Poder Discriminatorio de una Variable…...……………………………………………………….. 81
		1. Relación entre el Análisis PRIDIT y la Prueba No Paramétrica Wilcoxon…………………………………... 82
	7. Evaluación del Poder Discriminatorio de las Variables Indicadoras de Fraude y Obtención de una Puntuación Global de Sospecha de Fraude para una Demanda Entera………… 93
	8. Clasificación de las Demandas por medio de las Ponderaciones PRIDIT…………………………………………………………... 100
	9. Prueba de Consistencia para Variaciones de Tiempo…….. 102
2. **CONCLUSIONES Y RECOMENDACIONES**

**APÉNDICE**

Apéndice A: Demostraciones de Lemas y Teoremas

**ANEXOS**

Anexo 1: Datos Originales

Anexo 2: Datos Transformados con Puntuaciones RIDIT y Clases

**BIBLIOGRAFIA**

**ABREVIATURAS**

ICEA Investigación Cooperativa entre Entidades Aseguradoras

RIDIT Relative to an Identified Distribution

ACP Análisis de Componentes Principales

SCREE Score de los Eigenvalores

G1 Grupo 1 o Grupo Fraudulento

G2 Grupo 2 o Grupo No Fraudulento

**SIMBOLOGÍA**

 Número índice

Xt Monto del siniestro pagado en el período t

X0 Monto del siniestro pagado en el período base

 Matriz de datos

 Matriz de Varianzas y Covarianzas

Yj  j-ésima Componente Principal

aj j-ésimo eigenvector normalizado de 

 j-ésimo eigenvalor de 

A Matriz de eigenvectores de 

 Matriz de Varianzas y Covarianzas de Y

 Matriz de Componentes Principales

Var (Yj) Varianza de la j-ésima Componente Principal

tr() Traza de la Matriz de Varianza y Covarianza

tr() Traza de la Matriz de Varianzas y Covarianzas de Y

**SIMBOLOGÍA**

Var(Xi) Varianza de la i-ésima variable original

 Sumatoria desde j igual a 1 hasta p de los eigenvalores

n Número de unidades experimentales

S Matriz de Varianzas y Covarianzas Muestral

 Estimador del j-ésimo eigenvalor

 Estimador del j-ésimo eigenvector

d Dimensionalidad del espacio en el que están los datos

k Número de eigenvalores iguales a cero

 Vector de carga de las Componentes Principales

 Matriz de Correlación

 j-ésimo eigenvalor de la Matriz de Correlación

 j-ésimo eigenvector de la Matriz de Correlación

Xrj Valor de la r-ésima categoría de la j-ésima variable indicadora

**SIMBOLOGÍA**

Zrj Valor estandarizado de la r-ésima categoría de la j-ésima variable indicadora de fraude

kt Número de categorías disponibles y ordenadas para la variable indicadora t

 Vector de proporciones de respuesta observada de las demandas

Bti Puntuación RIDIT de la i-ésima categoría de la variable t

 Valor esperado de las puntuaciones RIDIT

H0 Hipótesis Nula

H1 Hipótesis Alterna

Rti Puntuación RIDIT original de Bross de la i-ésima categoría de la variable t

n1 Tamaño de la muestra 1

n2 Tamaño de la muestra 2

**SIMBOLOGÍA**

**** Estadístico de prueba U de Mann-Whitney

 Media del estadístico 

 Varianza del estadístico 

 Media de las puntuaciones RIDIT de la variable t

 Estadístico de prueba de Wilcoxon

N Número de demandas

 Vector multinomial de la probabilidad de respuesta del Grupo 1

 Vector multinomial de la probabilidad de respuesta del Grupo 2

 Proporción de demandas que pertenecen al grupo fraudulento

(-1) Proporción de demandas que pertenecen al grupo no fraudulento

 Puntuación esperada para una demanda fraudulenta

**SIMBOLOGÍA**

 Medida de Poder Discriminatorio de las t variables indicadoras de fraude

 Función indicadora de A

F Matriz de puntuaciones individuales PRIDIT

 Valor RIDIT de la i-ésima categoría de la variable t

 Ponderación PRIDIT estimada de las t variables indicadoras de fraude

E[F’F] Valor esperado de la matriz F transpuesta por la matriz F

 Matriz diagonal de las varianzas singulares de un Análisis de Factores

q Tipo de grupo

 Puntuación RIDIT de una demanda seleccionada al azar del grupo q

**INDICE DE TABLAS**

**Pág.**

**TABLA I** Variación Porcentual Corriente Año a Año para el Período 1.993-2.003……………………………………... 21

**TABLA II** Monto Total de Indemnización Pagada e Índices con Año Base 1.993=100………………………………………….. 24

**TABLA III** Cálculo de Puntuaciones RIDIT………………………… 78

**TABLA IV** Cálculo de Puntuaciones RIDIT para la variable VEHUSO…………………………………………………... 79

**TABLA V** Prueba de Kolmogorov-Smirnov de las Puntuaciones RIDIT………………………………………………………. 80

**TABLA VI** Ponderaciones PRIDIT para las Variables Indicadoras de Fraude……………………………………………………... 98

**TABLA VII** Medida de Poder Discriminatorio para una Demanda Completa………………………………………………….. 99

**TABLA VIII** Clasificación de las Demandas……………………..…. 101

**INDICE DE CUADROS**

**Pág.**

**CUADRO 2.1** Monto Total de Indemnización Liquidada por Siniestros para el Período 1.993 – 2.003…………………………... 18

CUADRO 2.2 Monto y Porcentaje de Siniestros Pagados por Segmento de Realización en el Ramo de Seguros Generales…... 27

**CUADRO 2.3** Porcentaje por Segmento de Realización de Siniestros Pagados…………………………………………………… 29

**CUADRO 2.4** Montos Pagados por Siniestros Vehiculares por Seguros Equinoccial (Período 2.000-2.003)……………………... 33

**CUADRO 2.5** Número de Siniestros Pagados por Seguros Equinoccial

 ……………………………………………………………… 34

**CUADRO 4.1** Variables Utilizadas en la Investigación……………….. 67

**INDICE DE GRÁFICOS**

**Pág.**

**GRÁFICO 2.1** Monto Total de Siniestros Pagados en el Ecuador en el Ramo de Seguros de Vehículos Período 1.993-2.003 ……………………………………………………………… 19

**GRÁFICO 2.2** Variación Porcentual Corriente del Monto Total de Siniestros Pagados Período 1.993-2.003……………… 22

**GRÁFICO 2.3** Indices del Monto Total de Siniestros Liquidados con Año Base 1.993=100…………………………..……………… 25

**GRÁFICO 2.4** Porcentaje de Pago e Siniestros en Seguros de Vehículos con Relación al Resto de Segmentos……... 30

**GRÁFICO 2.5** Porcentaje de Pago e Siniestros en Seguros de Vehículos con Relación al Resto de Segmentos Primer Trimestre 2.004…………………………………………... 31

**INDICE DE FIGURAS**

**Pág.**

**FIGURA 1.1** Fraude: Caso 1……………………………………………. 7

**FIGURA 1.2** Fraude: Caso 2……………………………………………. 8

**FIGURA 1.3** Fraude: Caso 3……………………………………………. 8

**INTRODUCCIÓN**

A pesar de que el fraude se da en casi todas las ramas, los fraudes en los seguros se han convertido en una práctica común. El mercado asegurador considera el fraude como un factor ineludible de riesgo y, hoy en día las entidades luchan por desarrollar un foco de acción frente al mismo.

Precisamente, el tema de investigación realizado es un modelo de detección de fraude en las declaraciones de siniestros de vehículos utilizando una nueva técnica denominada PRIDIT, el mismo que es aplicable a las compañías aseguradoras.

El **objetivo general** del presente trabajo de investigación es:

* La aplicación de un modelo de detección de fraude a una cartera real de seguro de automóviles, enfocado en las declaraciones de siniestros de automóviles para clasificar y cuantificar el nivel de fraude de cada una de estas declaraciones, realizadas por los asegurados.

Los **objetivos específicos** son:

* Reducir la incertidumbre e incrementar las oportunidades de clasificar las demandas correcta y eficientemente a cada grupo (fraudulentas / no fraudulentas) sin importar el tipo de variables que intervengan.
* Transformar respuestas categóricas en un conjunto de valores numéricos que estén dentro de un intervalo [-1,1], lo cual refleje la relativa anormalidad de una respuesta en particular.
* Determinar una ponderación de fraude para cada variable involucrada en el análisis.
* Determinar una medida de poder discriminatorio que permita clasificar las demandas en fraudulentas y no fraudulentas.