**UNIVERSIDAD NACIONAL AUTÓNOMA DE NICARAGUA**
**FACULTAD DE CIENCIAS**
**DEPARTAMENTO DE MATEMÁTICA Y ESTADÍSTICA**

# ESTIMACIÓN DE UNA PRIMA PURA DIFERENCIADA POR FACTORES DE RIESGO PARA EL SEGURO DE AUTOMÓVIL UTILIZANDO EL MÉTODO DE TARIFICACIÓN A POSTERIORI

**MONOGRAFÍA**

**PARA OPTAR AL TÍTULO DE LICENCIADAS EN CIENCIAS ACTUARIALES Y FINANCIERAS**

PRESENTADA POR:

BRA. ILEANA MERCEDES ROMERO SALAZAR
BRA. FÀTIMA DE LOS ÀNGELES ROMERO MONTOYA

TUTORA: MSc. ÁNGELA ALTAMIRANO SLINGER

ASESOR: MSc. WILLIAM MILTON CARVAJAL HERRADORA

LEÒN, NOVIEMBRE 2007

# DEDICATORIA

Este trabajo monográfico lo dedico de forma especial a:

Dios, padre celestial, por iluminar mi camino hasta alcanzar el éxito, por haberme dado la oportunidad de realizar uno de los mayores sueños de mi vida, por estar siempre a mi lado, darme sabiduría y paciencia para resolver todos los conflictos que enfrenté día a día.

A mi mamá, que amo con todas las fuerzas de mi corazón María de Jesús Salazar Amador por su cariño y apoyo en el transcurso de toda mi vida, sin su ayuda no hubiese podido culminar mis estudios universitarios.

A mi papá, Cristóbal Ervin Romero Rocha, por su amor y confianza, por ser un padre responsable que lucha por la educación de sus hijos, sin su dedicación no hubiese podido cumplir este gran sueño.

A la universidad, mi segundo hogar, fruto de grandes luchas por haber formado en mí una persona profesional.

*Ileana Mercedes Romero Salazar*

# DEDICATORIA

Dedico este trabajo de manera especial a:

Dios, padre eterno y glorioso, por guiar mi camino hasta alcanzar el éxito, por haberme dado la fuerza y paciencia para solucionar los problemas, por permitirme alcanzar la gran ilusión de mi vida, concluir una carrera universitaria.

A mi mamá, Elizabeth Montoya Arrieta, que amo con toda mi alma y corazón por su cariño, comprensión y apoyo durante toda mi vida, sin su amor y constancia no hubiese podido alcanzar mis metas.

A mi papá, Pablo Absalón Romero Vanega, por su amor, amistad y confianza, por ser un padre ejemplar y maravilloso, sin su entrega no hubiese podido culminar mi carrera.

A mi abuelita, María Margarita Vanega Pineda, por su ternura, que desde los cielos me cuida y vela mis sueños, por haberme enseñado durante el tiempo que estuvo con nosotros, los frutos que surgen de la humildad y la paciencia para alcanzar las metas.

A la universidad, mi casa de estudios y mi segundo hogar, a la cual le debo cinco años de experiencias, madures y aprendizaje y que ha hecho de mí lo que soy ahora, toda una profesional con valores y principios éticos.

*Fátima de los Ángeles Romero Montoya*

# AGRADECIMIENTO

**A Dios**, dador de vida y sabiduría, que con su amor nos guió de su mano celestial hasta el final de nuestra primer meta de vida, concluir los estudios universitarios, sin su ayuda hubiese sido imposible alcanzar nuestra profesionalización.

**A nuestros padres** por su apoyo incondicional, moral, espiritual y económico a lo largo de toda nuestra vida, por habernos enseñado el valor de nuestro esfuerzo, dedicación y empeño.

**A nuestros profesores,** por habernos transmitido sus conocimientos en el transcurso de toda nuestra carrera, en especial a nuestra tutora MSc. Ángela Altamirano Slinger que a pesar del poco tiempo disponible para la realización de este trabajo nos apoyó de forma incondicional, a MSc. Alberto Cerda Campos que nos brindó su ayuda en todo momento y a MSc. Milton Carvajal por su valioso tiempo y excelente colaboración.

*Ileana Mercedes Romero Salazar*
*Fátima de los Ángeles Romero Montoya*

# ÍNDICE

# INTRODUCCIÓN

A medida que los mercados aseguradores se vuelven más sofisticados y la tecnología informática evoluciona, las compañías de seguros usan técnicas más complejas para analizar sus operaciones y diseñar sus estructuras de tarificación.

Las aseguradoras deben usar técnicas para ajustar modelos de tarificación a clases de riesgos, tales modelos pueden ser utilizados para determinar los factores de riesgo que de forma más eficiente, diferencian los buenos de los malos riesgos.

Uno de los objetivos de la Matemática Actuarial es la elaboración de Sistemas de Tarificación que garanticen la solvencia de la entidad aseguradora, pero que a la vez resulten lo más equitativo posible. Un problema que se presenta en el seguro de automóvil es la caracterización del riesgo individual a través de las variables de tarifa. Una solución a este problema puede darse mediante la agrupación de riesgos y la estimación de una prima pura del seguro en colectivos, más o menos homogéneos, combinando la información global disponible con la individual, esto es, usando el sistema de tarificación a posteriori en el que la prima de riesgo individual depende de la evolución de la siniestralidad del grupo homogéneo al que pertenece el individuo.

En nuestro país el cálculo de las primas en los seguros de automóvil, se realiza tomando como base, entre otros aspectos, tarifas de mercado, costos de repuestos del vehículo y la experiencia de costos de siniestralidad. La prima se obtiene multiplicando una tarifa estimada por millar de suma asegurada. No se consideran en el cálculo de las primas características relacionadas al conductor que son relevantes. Lo anterior nos motivó a realizar un estudio de estimación de primas incorporando la mayoría de variables disponibles que influyen sobre el número de siniestros y por ende en el cálculo de la prima.

En nuestro trabajo se muestra la influencia de factores de riesgo relacionados al vehículo y su conductor para el cálculo adecuado de primas puras, partiendo de la división de la cartera en grupos homogéneos, así como la estimación de primas utilizando la variable con mayor relación sobre el número de siniestros declarados por póliza. Se pretende que la información generada por la metodología empleada e incorporada en el sistema de tarificación proporcione a las Compañías de Seguros la garantía de ofrecer un producto a un precio más justo.

# OBJETIVOS

## OBJETIVO GENERAL:

Estimar una Prima Pura del Seguro de Automóvil, diferenciada por factores de riesgo aplicando el Método de Tarificación a Posteriori o experience-rating.

## OBJETIVOS ESPECÍFICOS:

- Identificar las etapas del proceso de Tarificación a Posteriori en el Seguro de Automóvil.
- Aplicar métodos estadísticos para determinar la influencia de los factores de riesgo sobre la prima pura del Seguro de Automóvil.
- Agrupar la cartera de Pólizas del Seguro de Automóvil según el factor de riesgo de mayor influencia sobre el número de siniestros declarados por póliza.
- Calcular el coste medio y la frecuencia de siniestralidad por grupos de riesgos.
- Comparar la heterogeneidad de los grupos de riesgos con relación a la cartera.
- Estimar primas de riesgos incorporando nueva información.

# CAPÍTULO 1
# MARCO TEÓRICO

## 1.1 EL SEGURO

El seguro es un servicio de seguridad ofrecido por una unidad económica o ente asegurador (sociedad anónima, mutua o cooperativa de seguros) a un cliente. Jurídicamente, se trata de un convenio oneroso por el que se establece la transferencia total o parcial a otra entidad distinta de la que puede sufrir las consecuencias económicas de determinados siniestros. El contrato en que se materializa este convenio se denomina póliza de seguros y el precio del servicio de seguridad es la prima de seguro, que es función del riesgo asegurable y de los restantes factores que integran el coste de la empresa, denominándose prima pura a la parte de la prima que atiende exclusivamente a la cobertura del riesgo.

El riesgo de forma genérica lo definimos como el acontecimiento incierto, independiente de la voluntad exclusiva de las partes y cuya realización implica, normalmente consecuencias desfavorables para el asegurado. Así, un contrato de seguro tiene dos componentes básicos e imprescindibles para su viabilidad: la existencia de un riesgo y el pago de un precio por su cobertura, es decir, la prima, que es el precio del servicio más el margen explícito de beneficio y deben cumplir los principios de equidad y suficiencia de acuerdo con la naturaleza de los riesgos asumidos por el asegurador.

El principio de equidad se refiere a que la prima o precio del seguro se ajuste al riesgo de siniestralidad de cada póliza. El principio de suficiencia consiste a que en términos esperados las primas sean suficientes para cubrir todos los riesgos de la cartera considerada, es decir, que permitan hacer rentable, en condiciones de estabilidad a largo plazo, a la empresa aseguradora.

Si un riesgo de una cartera es demasiado grande para una compañía pasará parte del mismo a otra u otras compañías, dando origen a los reaseguros o coaseguros. Desde el punto de vista actuarial los seguros se dividen en vida y no vida (o seguros generales). El estudio de estos últimos se aleja bastante del enfoque de los primeros, ya que incorpora técnicas más rigurosas del cálculo de probabilidades y la estadística matemática.

## 1.2 FACTORES DE RIESGO QUE INFLUYEN EN EL CÁLCULO DE LA PRIMA DE LOS SEGUROS DE AUTOMÓVIL

Existen dos factores generales que determinan el costo del seguro. El primer factor que se toma en cuenta es la suscripción y el segundo, la tarifa. Los aseguradores utilizan la suscripción para medir el riesgo asociado con el solicitante, o el grupo de solicitantes al que pertenece. A base de los resultados que arroje ese proceso de suscripción, se establece la tarifa, imponiendo un precio a base de lo que el asegurador entiende que le costará asumir la responsabilidad económica por cualquier reclamación potencial del solicitante de la póliza.

La imposición de tarifas de riesgos esta afectada por factores específicos, como experiencia previa como conductor, el área donde vive, género y edad, estado civil (si es casado o soltero), cobertura anterior del seguro, uso de su vehículo, marca y modelo del automóvil, son factores que comúnmente afectan el precio que pagará por su seguro, es decir todas aquellas variables que indique quien es usted y su historial de conducción.

Los factores siguientes tienen una gran influencia en las primas:

1. **Sexo**
   Los hombres tienen más accidentes en la carretera que las mujeres. Tal vez los hombres sean conductores más agresivos. O quizás conducen más que las mujeres. A las compañías de seguros no les importa la razón. Simplemente analizan los datos de los reclamos. Si su sexo tiene menos accidentes, paga menos.

2. **Edad**

   Los conductores de menos de 25 años de edad (y, para algunos aseguradores, de menos de 30 años de edad) se consideran a mayor riesgo de tener un accidente.

3. **Estado Civil**

   Los conductores casados tienden a tener menos accidentes que los conductores solteros. Quizás porque tienen que pensar en otra persona. Quizás porque hay un cónyuge que se queja con respecto a los malos hábitos de conducción. Una vez más, en realidad no importa. Si usted es soltero, tendrá que pagar más.

4. **Registro Personal de Conducción**

   Años de experiencia de conducción, accidentes, boletas por exceso de velocidad y ofensas por conducir en estado de ebriedad son todos factores que determinan cuánto riesgo usted presenta como conductor. Algunos piensan que si ha tenido un accidente, deberá pagar más para pagar por su reclamo. Sin embargo, en realidad estadísticamente aquellos que han tenido un accidente tienen mayor probabilidad de tener otro accidente.

5. **Uso del Vehículo**

   Las personas que viajan a diario a su lugar de trabajo tienen mayor riesgo que aquellos que sólo conducen para hacer los mandados o por recreación. Aquellos que usan su automóvil para sus negocios pagan primas más altas.

6. **Tipo de Vehículo**

   Los vehículos más grandes y más pesados se consideran a menor riesgo que aquellos más pequeños y ligeros. Los automóviles más caros son más costosos para reparar que los modelos económicos. Los automóviles deportivos están a mayor riesgo de los demás automóviles.

7. **Tamaño de su Deducible**

   Cuanto mayor sea su deducible, tanto menor será la posibilidad de que la compañía de seguros deberá pagar un reclamo y tanto menor su prima.

De la misma forma se pueden mencionar algunos factores que afectan en menor magnitud el costo del seguro:

- ¿Dónde estaciona el auto?.
- Aditamentos especiales de seguridad que tenga el auto.
- Rendimiento académico (si sigue en la escuela).
- Si el auto está pagado o si está financiado.

## 1.3 TARIFICACIÓN

La tarificación en los seguros, desde el punto de vista técnico, tiene como objetivo primordial el correcto cálculo de primas equitativas y suficientes, de manera que el tomador pague la prima más ajustada al riesgo que la póliza incorpora. Los principios técnicos en que se basa la elaboración de una tarifa constituirán el sistema de tarificación. Desde el punto de vista actuarial, se distinguen dos sistemas, la tarificación a Priori y la tarificación a Posteriori. En ambos casos, para poder realizar la valoración es necesario disponer de una correcta información, tanto a nivel individual como colectivo.

### 1.3.1 SISTEMAS DE TARIFICACIÓN

#### 1.3.1.1 Tarificación a priori o class-rating

El sistema de tarificación a priori nos permite asignar una prima a un riesgo que se incorpora a nuestra cartera sin tener necesariamente experiencia sobre la siniestralidad que conlleva. Únicamente es necesario conocer determinadas características para asignar una siniestralidad esperada y con ella una prima. Dado el objetivo de equidad y suficiencia en las primas, buscaremos la formación de grupos de riesgo homogéneos determinados por combinaciones de clases de tarifa, que tendrán internamente una siniestralidad esperada similar y por lo tanto poca dispersión entorno a su valor esperado.

Para la realización de una tarificación a priori partimos de la experiencia de una cartera para una determinada cobertura en un período fijado. Para cada póliza de la cartera, se obtienen los datos de la siniestralidad (número de siniestros, $X_i$ y sus correspondientes cuantías, $C_i$) y una serie de factores iniciales de riesgo, que pueden hacer referencia a características del objeto asegurado o a otros condicionamientos de este: características del asegurado, del tomador del seguro, condiciones socio-económicas que lo rodean, etc.

Los principios técnicos en que se basa la tarificación a priori consisten en un proceso que debe solucionar las siguientes fases:

1) La determinación de la estructura de tarifa: con la selección de las variables de tarifa la obtención de los grupos de tarifa, la inclusión de los gastos en la tarifa, y el tratamiento adecuado de los grandes riesgos.

2) El cálculo de un nivel adecuado de prima para cada grupo de tarifa.

3) Por último, la implementación de la tarifa al mercado competitivo.

Si asumimos como hipótesis del proceso de riesgo la equidistribución de las cuantías de los siniestros, la de independencia entre dichas cuantías y la de independencia entre el coste por siniestro y el número de siniestros, la prima pura P, la calculamos como el producto de la esperanza del número de siniestros por la esperanza de la cuantía de un siniestro:

$$P = E[X] * E[C] = \frac{\sum_{i=1}^{n} x_i}{N} * \frac{\sum_{i=1}^{n} c_i}{n}$$

Donde N: número total de pólizas que integran la cartera

$n$: número total de siniestros

$x_i$: i-ésimo siniestro por póliza

$c_i$ : Costo del $i$-ésimo siniestro

Por ello, las variables de tarifa pueden ser seleccionadas por separado respecto de ambas variables. La selección de dichas variables se realiza haciendo uso de métodos estadísticos de análisis multivariante.

Para la realización de un proceso de tarificación a priori, es conveniente que la experiencia en que se base la tarifa pertenezca a un intervalo temporal lo más cercano posible, siendo necesarias revisiones periódicas con datos actualizados que repitan el proceso con todas sus fases, comenzando por la selección de variables de tarifa.

### 1.3.1.2 Tarificación a posteriori o experience-rating

La tarificación a posteriori, en oposición a la tarificación a priori, parte de una prima inicial para cada unidad de riesgo, individuo o grupo, que se va modificando en períodos sucesivos de acuerdo con la experiencia individual o colectiva para dar lugar a las primas de los períodos sucesivos. En un sentido amplio, la expresión experience-rating se aplica a todo problema de actualización de tarifas mediante la incorporación de nueva información.

En la mayoría de los ramos de seguros generales, existen múltiples posibles fuentes de heterogeneidad de los riesgos. Una prima de riesgo basada en la siniestralidad observada de una cartera manifiestamente heterogénea puede resultar inadecuada si, probablemente, la estructura de la cartera se modifica. Para reducir este peligro, las pólizas normalmente se agrupan en diferentes estratos, en función de los factores de riesgo utilizados. A tal efecto se calculan las frecuencias de siniestralidad y las primas de los diferentes grupos de riesgo. Esta aproximación al problema es denominada Tarificación por Grupos.

La justificación de estos sistemas está en que dentro de cada clase de riesgo existe heterogeneidad, debido a la influencia de ciertos factores de riesgo no considerados (conocidos o desconocidos), o a la incorrecta agrupación de las clases de los sí considerados, provocando en cualquiera de los casos una heterogeneidad residual considerable . Esta heterogeneidad quedará recogida por la siniestralidad que irá teniendo cada póliza en los años sucesivos, muchas pólizas presentaran una frecuencia de siniestralidad próxima a la media del grupo, algunas sin embargo, tendrán una consideradamente superior tasa de siniestralidad, y otras notablemente inferior. Al considerar la experiencia propia de cada póliza obtendremos un mayor grado de equidad en las primas de los ejercicios posteriores, incorporando la información evolutiva de los riesgos mediante un sistema de bonificaciones y penalizaciones (sistema bonus-malus) de acuerdo con los resultados obtenidos.

En el seguro de automóvil su aplicación suele tomar como referencia el número de siniestros declarados en un período de un año; si no se han declarado siniestros se aplicaran descuentos o bonificaciones. Por cada siniestro declarado pueden reducirse los descuentos y/o aplicar recargos. Históricamente, lo usual ha sido basar las bonificaciones y recargos en el número de siniestros, aunque existen estudios basados en las cuantías de los siniestros declarados.

Los métodos de tarificación a posteriori pueden ser retrospectivos o prospectivos. Se da de forma retrospectiva cuando la prima es calculada tiempo después de ocurrido el siniestro, como por ejemplo, cuando se produce el reembolso de una parte de la prima satisfecha por un expuesto al riesgo (o grupo de riesgos) a la finalización del período de cobertura en base a una experiencia de siniestralidad positiva y de manera prospectiva cuando el cálculo de la prima se realiza con antelación a la ocurrencia del siniestro, un ejemplo es el Sistema de Bonificación por no Siniestralidad (BNS).

### 1.3.2 TARIFICACIÓN A POSTERIORI DEL SEGURO DE AUTOMÓVIL

Nos centramos ahora en el proceso de tarificación a posteriori del seguro de automóvil. Dicho seguro cubre los daños personales y los daños materiales ocasionados al vehículo como consecuencia de un hecho de la circulación. Ambas coberturas, la de daños personales y la de daños materiales, tienen frecuencia de siniestralidad y coste medio de un siniestro muy diferentes, por ello podemos diferenciar la información con el objetivo de calcular la esperanza del coste total por póliza por separado para cada cobertura.

Por el mismo motivo, cada cobertura adicional hasta llegar a un todo riesgo (responsabilidad civil, robo total o parcial a consecuencia de robo total, extensión territorial etc.) debe ser tratada por separado, tomando en cuenta que se debe contar con la información estadística necesaria para el estudio especifico por cobertura.

En este sentido, las primas puras totales por póliza se calcularán como la suma aritmética de las primas puras correspondientes a cada cobertura. Si denotamos por $X^c$ al número esperado de siniestros respecto a la cobertura $c$, y por $m^c$ la cuantía esperada de un siniestro respecto a la cobertura $c$ (siendo $c$: daños materiales, daños personales, daños propios, etc), la prima pura total se calculará como:

$$PT = \sum X^c * m^c$$

En el seguro de automóvil y dependiendo de la cobertura, los factores potenciales de riesgo a tener en cuenta pueden ser, por ejemplo:

a) **Factores relativos al vehículo asegurado:** el valor, la antigüedad, la categoría, la clase, el tipo, la marca, el modelo, el color, el tipo de combustible, la cilindrada, la potencia, el peso, o la relación potencia / peso, etc

b) **Factores relativos al (primer y/o segundo) conductor:** especialmente la antigüedad del carnet, la edad, el sexo y el resultado de la experiencia en el pasado. Aunque también pueden hacer referencia a condiciones socio-económicas que lo rodean como es el número de hijos, el estado civil, la profesión, etc.

c) **Factores relativos a la circulación:** la zona de circulación, el Departamento de la póliza (u otras clasificaciones realizadas a partir del lugar de conducción del vehículo), el uso del vehículo, los kilómetros anuales, etc.

d) **Factores relativos a la póliza:** antigüedad de la póliza, número de pagos anuales, etc.

### 1.3.3 PROCESO ESTADÍSTICO DE LA TARIFICACIÓN A POSTERIORI

El conocimiento de las principales variables que intervienen en los accidentes que sufren los conductores es una cuestión de gran trascendencia para las Compañías Aseguradoras que operan en el ramo de automóvil.

Conocer las características del conductor y/o vehículo que pueden influir en la siniestralidad nos permitirá obtener un mayor grado de equidad en la prima de los años posteriores, que la inicialmente cobrada, a través de la aplicación de un sistema de bonificación y penalización que dependerá de los resultados obtenidos.

Un análisis descriptivo de la variable endógena (número de siniestros) con relación a las variables explicativas (que pueden incluir características del vehiculo, como su uso y categoría, o, del conductor, sexo, edad, etc.) nos permitirá observar el comportamiento conjunto de cada par de variables. En el caso de variables cuantitativas el Método Estadístico de Regresión Lineal (simple o múltiple) es una alternativa aplicable para determinar la influencia de los factores de riesgo sobre las variables endógenas. La asociación entre variables cualitativas se puede medir a través de pruebas no paramétricas como la de Chi-cuadrado de Pearson, el coeficiente de contingencia, Phi y V de Cramer, etc.

Tomando en cuenta el grado de influencia que ejerce cada variable explicativa sobre el número de siniestros se procederá a estratificar, que no es más que la agrupación de dos o más variables cuyos datos reunirán características semejantes del grupo establecido.

La razón por la cual la estratificación es un procedimiento de gran importancia es que permitirá la obtención de una prima más equitativa, suficiente y acorde al riesgo expuesto.

Subdividiendo la cartera de pólizas en varios grupos se pretende que los datos de siniestralidad correspondientes a tales subgrupos resulten ser substancialmente más homogéneos que los de la cartera en su conjunto, es decir se obtendrá una frecuencia de siniestralidad y una prima de riesgo más especifica en cuanto a las características analizadas de cada grupo, el cálculo de la prima por grupo se realiza a través del método clásico usando la siguiente formula:

$$\Pr = \frac{n}{N} * \bar{c}$$

$$\Pr = \frac{Número\ siniestros}{Número\ expuestos} * \frac{\sum_{i=1}^{n} c_i}{n}$$

$$\Pr = fs * \bar{c}$$

Donde N: número total de pólizas que integran la cartera o grupo
n : número de siniestros
$c_i$ : Costo del $i$-ésimo siniestro
$\bar{c}$: Costo promedio de un siniestro
$fs$ : Frecuencia de siniestralidad

Empezamos calculando la media y la varianza de la distribución del número de siniestros correspondientes a los asegurados que integran cada grupo de riesgo, utilizando las distribuciones de frecuencia de siniestralidad de la muestra, mediante:

$$E[X] = \frac{Número\ siniestros}{Número\ expuestos} = \frac{n}{N}$$

$$Var(X) = E[X^2] - (E[X])^2$$

$X$ : Número de siniestros declarados por póliza

Los valores obtenidos serán utilizados para calcular los estimadores de los parámetros $P$ y $K$ de la distribución Binomial Negativa.

Para realizar el ajuste de la distribución Binomial Negativa se igualan las medias y varianzas teóricas y observadas:

$$Media = \frac{K(1-P)}{P}$$

$$Media * P = K(1-P)$$

$$K = \frac{Media * P}{(1-P)}$$

$$Varianza = \frac{K(1-P)}{P^2}$$

$$Varianza = \frac{\frac{Media * P}{(1-P)} * (1-P)}{P^2}$$

$$Varianza = \frac{Media * P}{P^2} = \frac{Media}{P}$$

$$P = \frac{Media}{Varianza}$$

Las probabilidades de ocurrencia de siniestros de la distribución Binomial Negativa y Poisson ajustadas se calculan para cada valor de la variable explicada por cada grupo de riesgo en estudio, las cuales se multiplican por el total de pólizas de cada grupo, obteniéndose las frecuencias ajustadas. Para comprobar que la distribución Binomial Negativa proporciona un modelo adecuado para la distribución de frecuencias de siniestralidad comparamos las frecuencias observadas y ajustadas, es decir le aplicamos una prueba de bondad de ajuste a cada grupo de riesgo, dependiente del resultado se podrá determinar si los datos se ajustan o no a una distribución Binomial Negativa.

Se calculan los parámetros de la distribución Gamma y se representan las distribuciones por grupo, esta distribución se utiliza como modelo descriptivo de la heterogeneidad de los riesgos para un determinado grupo, en caso de grupos homogéneos su gráfica resultará apuntada y estrecha con una desviación típica pequeña, para grupos heterogéneos será ancha, baja y con una gran desviación típica.

El proceso final del estudio de la heterogeneidad de los grupos de riesgo consiste en comparar la media y desviación estándar observada con la ajustada.

Se debe tener en cuenta que el análisis realizado de la heterogeneidad subyacente en las clases de riesgo sólo será válido si la cantidad de información estadística disponible es grande y en consecuencia resulta factible estimar con fiabilidad los parámetros de la distribución Binomial Negativa.

## 1.4 DISTRIBUCIONES ESTADÍSTICAS ÚTILES EN EL TRABAJO DE LOS SEGUROS GENERALES

### 1.4.1 DISTRIBUCIÓN NORMAL

La distribución normal ocupa una posición central en la ciencia de la estadística, principalmente por su rango de valores, es una distribución continua y tiene dos parámetros, su media $\mu$ y su desviación típica $\sigma$.

Estos parámetros nos definen la tendencia y la dispersión, respectivamente. La función de densidad de probabilidad es una campana simétrica en relación a la media, su expresión es básicamente complicada:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left[(x-\mu)/\sigma\right]^2\right\}, (-\infty < x < \infty)$$

En muy pocas ocasiones se hace uso de esta expresión, aunque el rango de una variable aleatoria normal $x$ va desde $-\infty$ a $\infty$, la probabilidad de que $X$ tome valores muy pequeños o muy grandes es pequeña, es decir hay una probabilidad del 95% de que $X$ tome un valor en el intervalo de magnitud dos desviaciones típicas alrededor de la media.

La probabilidad de que una variable aleatoria normal $X$ con media $\mu$ y desviación típica $\sigma$ tome un valor entre los valores a y b es:

$$\int_a^b \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left[(x-\mu)/\sigma\right]^2\right\} dx$$

No es posible expresar esta integral (o área) en términos de funciones matemáticas concretas. Es necesario efectuar una integración numérica, esta dificultad se supera efectuando un cambio de variable, de manera que la variable $X$, se transforme en una variable $Z$ que tiene media cero y desviación típica uno, definida por:

$$Z = \frac{X - \mu}{\sigma}$$

$$Z = \frac{X - mediaX}{desviación tipicaX}$$

Se dice que $X$ ha sido tipificada o estandarizada, porque si $X$ tiene una distribución normal, $Z$ también esta normalmente distribuida, a la variable $Z$ también se le conoce como normal unitaria por tener desviación típica uno, se denota como $N(0,1)$.

Debido a la importancia de la distribución normal $N(0,1)$, utilizamos un símbolo especial para representara su función de densidad de probabilidad: $\phi(x)$ en lugar de $f(x)$, denotaremos su función de distribución por $\Phi(x)$ en lugar de la usual $F(x)$.

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

$$\Phi(x) = \int_{-\infty}^{u} \phi(x) dx$$

La integral $\Phi(u)$ es el área existente bajo la curva $\phi(x)$ entre $x = -\infty$ y $x = u$, las tablas de $\phi(x)$ y $\Phi(x)$ son fácilmente disponibles y raramente tendremos que recurrir a las fórmulas con integración.

La probabilidad de que una variable aleatoria normal $X$ con media $\mu$ y desviación típica $\sigma$ tome un valor entre $\mu + A\sigma$ y $\mu + B\sigma$ es igual a la probabilidad de que la variable normal unitaria $Z$ este entre $A$ y $B$, que es

$\Phi(B)-\Phi(A)$ (el área bajo la curva de densidad normal unitaria entre $x=A$ y $x=B$). En otras palabras, escribiendo $a$ en lugar de $\mu+A\sigma$ y $b$ en lugar de $\mu+B\sigma$, la probabilidad de que una variable aleatoria normal $X$ con media $\mu$ y desviación típica $\sigma$ tome un valor entre $a$ y $b$ es:

$$\Phi[(b-\mu)/\sigma]-\Phi[(a-\mu)/\sigma]$$

Debemos resaltar una importante propiedad de la distribución normal: las combinaciones lineales de variables aleatorias normales independientes son también normales, y la media y varianza de una determinada combinación lineal pueden ser obtenidas utilizando las propiedades aditivas de las medias y varianzas.

### 1.4.2 DISTRIBUCIÓN GAMMA

La distribución Gamma es una distribución continua que tiene aplicaciones en un elevado número de procesos que se desarrollan en los seguros generales (análisis de la distribución del coste de un siniestro y de la heterogeneidad de los riesgos). La función de densidad de probabilidad es:

$$f(x)=\frac{\beta}{\Gamma(\alpha)}e^{-\beta x}(\beta x)^{\alpha-1} \qquad (0\leq x<\infty)$$

Con parámetros $\alpha$ y $\beta$ positivos y $\Gamma(\alpha)$ es un número que depende de $\alpha$, la media y varianza son respectivamente:

$$media=\frac{\alpha}{\beta}$$

$$\text{var}ianza=\frac{\alpha}{\beta^2}$$

La función de densidad de probabilidad de la distribución Gamma, es muy versátil con adecuadas combinaciones de los parámetros $\alpha$ y $\beta$. Cuando $\alpha=1$, la función de densidad toma su mayor valor en $x=0$, decreciendo a continuación. Para los demás valores de $\alpha$, $f(x)$ es nula para $x=0$, crece hasta alcanzar un máximo y decrece a continuación. La distribución es claramente no simétrica, es asimétrica positiva pero, cuando $\alpha$ crece, la asimetría decrece y la distribución se va haciendo más simétrica.

Se puede demostrar que la suma de $n$ variables aleatorias Gamma estocásticamente independientes entre sí, todas con función de densidad de probabilidad gamma es una variable aleatoria Gamma con parámetros $n\alpha$ y $\beta$. Para grandes valores de $n$, el Teorema Central de Limite nos garantiza que esta distribución será atribuiblemente Normal con media $n\alpha / \beta$ y varianza $n\alpha / \beta^2$ .

Si $X$ es una variable aleatoria Gamma con parámetros $\alpha$ y $\beta$, y $\alpha$ es grande, entonces $X$ será aproximadamente normal con media $\alpha / \beta$ y varianza $\alpha / \beta^2$, y

$$\frac{X-(\alpha / \beta)}{\sqrt{\alpha / \beta^2}}$$

Será aproximadamente normal unitaria, $N(0,1)$

### 1.4.3 DISTRIBUCIÓN DE POISSON

Es una distribución discreta utilizada generalmente para describir el proceso de ocurrencia de los siniestros con gran eficacia. Esta distribución corresponde a una variable aleatoria no negativa y entera que ocupa un lugar relevante en el ámbito de la estadística teórica.

Una aplicación de esta distribución en los Seguros Generales es el análisis del número de siniestros en una cartera de seguros de un determinado ramo.

Una variable aleatoria $X$ definida en el campo entero no negativo sigue una distribución de Poisson con parámetro $q$ si:

$$P(X=x)=e^{-q} \frac{q^x}{x!} (x=0,1,2,\ldots)$$

El parámetro $q$ debe ser positivo, y la media y varianza de la distribución vienen dadas por:

$$media = q$$
$$\text{var}\,ianza = q$$

El parámetro $q$ es a la vez media y varianza de la distribución. Una importante propiedad de la distribución de Poisson es que la suma de $n$ variables de Poisson independientes es a su vez una variable aleatoria de Poisson con media igual a la suma de las $n$ medias de las variables sumandos.

### 1.4.4 DISTRIBUCIÓN BINOMIAL NEGATIVA

Se dice que una variable aleatoria discreta $X$ sigue una distribución Binomial Negativa con parámetros $k$ y $p$ si:

$$P(X = x) = \binom{k + x - 1}{x} p^k (1 - p)^x, (x = 0,1,2,....)$$

La media y la varianza vienen dadas por:

$$Media = \frac{K(1 - P)}{P}$$

$$Varianza = \frac{K(1 - P)}{P^2}$$

La más importante aplicación de la distribución Binomial Negativa, desde el punto de vista de los seguros generales, es la que se refiere a la distribución de la frecuencia de siniestralidad cuando los riesgos no son homogéneos.

Consideremos una cartera de seguros en la cual la distribución del número de siniestros por asegurado y año es de Poisson con media $q$. Algunos asegurados son malos riesgos (es decir, su parámetro $q$ es elevado), mientras que otros son buenos riesgos (para ellos, $q$ es muy pequeño).

Sabiendo que la distribución Gamma es una distribución continua y muy versátil, supongamos que el parámetro $q$ correspondiente a diversos asegurados elegidos al azar siguen una distribución gamma, entonces, tenemos que:

a) $P(u < q < u + du) = \frac{\beta}{\Gamma(\alpha)} e^{\beta u} (\beta u)^{\alpha - 1} du$

Para un asegurado cuya frecuencia se encuentra en ese intervalo, la probabilidad de que declare $u$ siniestros viene dada por la ley de Poisson:

b) $P(X = x / u < q < u + du) = \dfrac{e^{-u} u^{x}}{x!}$

La probabilidad de que un asegurado, elegido al azar, declare $x$ siniestros en un año se calcula sumando (integrando) el producto de las ecuaciones b) y a) para todos los valores de $u$. Es decir,

c) $P(X = x) = \int_0^{\alpha} \left( \dfrac{e^{-u} u^{x}}{x!} \right) \dfrac{\beta}{\Gamma(\alpha)} e^{-\beta u} (\beta u)^{\alpha - 1} du$

Y, operando, resulta

$$P(X = x) = \binom{a + x - 1}{x} \left( \frac{\beta}{1 + \beta} \right)^{\alpha} \left( \frac{1}{1 + \beta} \right)^{x}$$

En otras palabras, la variable aleatoria correspondiente al número de siniestros declarados por un asegurado elegido al azar sigue una distribución binomial negativa con parámetros $k = \alpha$ y $p = \beta / (1 + \beta)$, donde $\alpha$ y $\beta$ son los parámetros de la distribución gamma representativa de la heterogeneidad de los riesgos, y

- $\alpha = k$
- $\beta = p / (1 - p)$

## 1.5 PRUEBAS NO PARAMÉTRICAS PARA DATOS NOMINALES

El grado de relación existente entre dos variables categóricas no puede ser establecido simplemente observando las frecuencias de una tabla de contingencia. Incluso auque la tabla recoja las frecuencias porcentuales en lugar de las absolutas, su simple observación no puede conducir a una conclusión definitiva por lo que se debe utilizar algunas medidas de asociación, preferiblemente acompañadas de su correspondiente prueba de significación.

### 1.5.1 PRUEBA CHI-CUADRADO

La opción **Chi-cuadrado** proporciona un estadístico propuesto por Pearson que permite contrastar la hipótesis de que los dos criterios de clasificación utilizados (las dos variables categóricas) son independientes. Para ellos compara las frecuencias observadas (obtenidas) con las esperadas (frecuencias que teóricamente debería haber en cada casilla si los dos criterios de clasificación fueran independientes), en caso de que esto pasase las frecuencias esperadas se estiman de la siguiente manera:

$$(\text{Frecuencia esperada})_{ij} = \frac{(\text{Total de la fila i})*(\text{Total de la columna j})}{\text{n}^{\text{o}} \text{ total de casos}}$$

Donde $i$, se refiere a una fila cualquiera y $j$ a una columna cualquiera. Es decir, bajo la condición de independencia, la frecuencia esperada de una casilla concreta se obtiene al dividir el producto de las frecuencias marginales por el número total de casos.

Obtenida las frecuencias esperadas para cada casilla, el estadístico $X^2$ o chi-cuadrado de Pearson se obtiene de la siguiente manera:

$$X^2 = \sum_i \sum_j \frac{(n_{ij} - m_{ij})^2}{m_{ij}};$$ Donde $n_{ij}$ se refiere a las

frecuencias observadas y $m_{ij}$ a las esperadas.

De la ecuación se desprende que el estadístico $X^2$ valdrá cero cuando las variables sean completamente independientes (pues las frecuencias observadas y las esperadas serán iguales) y que el valor del estadístico $X^2$ será tanto mayor cuanto mayor sea la discrepancia entre las frecuencias observadas y las esperadas es decir cuanto mayor sea la relación entre las variables.

Si los datos son compatibles con la hipótesis de independencia, la probabilidad asociada al estadístico $X^2$ será alta (mayor de 0.05). Si esa probabilidad es muy pequeña (menor que 0.05), se considera que los datos son incompatibles con la hipótesis de independencia y se concluirá que las variables estudiadas están relacionadas.

## 1.5.2 COEFICIENTE DE CONTINGENCIA C.

El coeficiente de contingencia trata de determinar el grado de asociación, comparando varios grupos o categorías de datos nominales clasificados en cuadros de contingencias con un diseño mayor de 2 x 2, su fórmula esta dada por:

$$C = \sqrt{\frac{\chi^2}{N + \chi^2}}, \text{ donde}$$

$C$ = Coeficiente de contingencia
$\chi^2$ = valor de chi-cuadrado calculado
$N$ = Número total de casos

El coeficiente de contingencia C tiene un campo de variación entre cero y uno, de manera que su valor es cero cuando existe carencia de asociación entre las variables, es decir cuando las variables son independientes. Cuando las variables muestran total asociación entre sí el coeficiente se aproxima a uno, pero sólo se alcanzará el valor uno en el caso ideal de infinitas modalidades.

# CAPÍTULO 2
# DISEÑO METODOLÓGICO

## 2.1 TIPO DE ESTUDIO:

El trabajo realizado es de tipo descriptivo y de corte transversal.

## 2.2 UNIVERSO Y MUESTRA:

Para la realización de este estudio se contó con una base de datos que fue proporcionada por una Compañía Aseguradora Privada Nicaragüense, la información corresponde a todas las pólizas del seguro de Automóvil de personas naturales, que estuvieron vigentes del 1$^{ro}$ de Octubre al 31 de Diciembre del año 2006, es decir, las pólizas emitidas del 1$^{ro}$ de Octubre del 2005 al 31 de Diciembre del 2006, la población total fue de 3,746 asegurados. En la base de datos proporcionada no se indicó si el asegurado era la misma persona que conducía el vehículo al momento del siniestro, asumimos que era la misma persona, a sabiendas de que este inconveniente podría afectar los resultados obtenidos. Nuestro interés primordial es mostrar la metodología de trabajo para el cálculo de una prima pura por grupo de riesgo.

Para la selección de la muestra se tomaron todas las pólizas que tenían completa la información sobre las variables que intervienen en el estudio que corresponden a 1,391 asegurados, lo que equivale al 37% del total de la población.

## 2.3 VARIABLES DE ESTUDIO:

Las variables que se utilizaron fueron: el número de siniestros que el asegurado ha declarado a su compañía durante el período considerado, como nuestra variable endógena, y las variables explicativas que incluyen características del vehículo como su categoría y uso; características del asegurado entre ellas edad, sexo y procedencia, así como el costo de los siniestros declarados.

## 2.4 OPERACIONALIZACIÓN DE LAS VARIABLES

| Variables | Tipo | Unidad de Medida | Escalas |
|---|---|---|---|
| Edad | Cuantitativa-Discreta | Años | 1 = Menores de 44<br>2 = Mayores de 43 |
| Sexo | Cualitativa | | F = Femenino<br>M = Masculino |
| Procedencia | Cualitativa | | Managua<br>Resto |
| Uso del vehículo | Cualitativa | | Particular<br>Comercial |
| Categoría de vehículo | Cualitativa | | 1 = Automóvil, Sedan, Coupe y Hatch Back.<br>2 = Jeep y Station wagon.<br>3 = Furgoneta Cerrada, Microbús y Panel Cerrado.<br>4 = Camioneta y Pick-Up.<br>5 = Camión. |
| Número de siniestros | Cuantitativa-Discreta | | X >= 0 |
| Costo de siniestros | Cuantitativa-Discreta | Dólares | X > 0 |

## 2.5 OBTENCIÓN DE LA INFORMACIÓN:

Se obtuvo información de diversas fuentes externas e internas. A través de varias entrevistas y cartas de solicitud se nos proporcionó una base de datos de asegurados del ramo de automóvil de una Compañía Aseguradora Privada Nicaragüense.

## 2.6 PROCESAMIENTO Y ANÁLISIS:

En primer lugar se hizo un análisis descriptivo de la relación de las variables de estudio, posteriormente se valoró la relación de las variables consideradas explicativas con la variable endógena “Número de Siniestros” a través de una prueba de Chi-Cuadrado de Pearson ($X^2$), con un nivel de confianza del 95% y un error del 5%. Después se procedió a identificar aquellas variables con mayor relación con la variable dependiente. Para el proceso de aproximación de primas por grupos se usó únicamente la variable independiente edad que es la de mayor relación con la variable explicada.

Para el procesamiento y manejo de los datos se utilizó el programa SPSS versión 12.00 para Windows y Microsoft Excel. Microsoft Word 2003 nos facilitó la realización del trabajo escrito. El Software R (Sistema para cálculos estadísticos y gráficos, versión 2006) nos permitió obtener las funciones de densidades y gráficas estadísticas. Con el Software Power Point se elaboró la presentación del trabajo.

Los resultados se presentan a través de tablas de contingencia y gráficos.

# CAPÍTULO 3
# RESULTADOS

Para los 1,391 asegurados de nuestra muestra se contabilizó el número de siniestros por póliza y se encontró que se distribuyen como lo indica la tabla siguiente:

**Tabla 3.1 Distribución del número de Siniestros declarados por póliza**

| | | Frecuencia | Porcentaje | Porcentaje Válido | Porcentaje Acumulado |
|---|---|---|---|---|---|
| Válido | 0 | 1228 | 88.3 | 88.3 | 88.3 |
| | 1 | 121 | 8.7 | 8.7 | 97.0 |
| | 2 o más | 42 | 3.0 | 3.0 | 100.0 |
| | Total | 1391 | 100.0 | 100.0 | |

Desde el punto de vista descriptivo se destaca un porcentaje elevado de valores nulos que toma la variable dependiente número de siniestros, exactamente el 88.30%, además sólo el 3% de los asegurados declaran más de 1 siniestros.

Los estadísticos descriptivos correspondientes a nuestra muestra en relación al número de siniestros se presentan en la tabla siguiente:

**Tabla 3.2 Estadísticos**

Número de Siniestros

| | |
|---|---|
| Válidos | 1391 |
| Perdidos | 0 |
| Media | 0.15 |
| Mediana | .00 |
| Moda | 0 |
| Desviación Estándar | 0.460 |
| Varianza | 0.211 |
| Rango | 4 |
| Mínimo | 0 |
| Máximo | 4 |

Tomando la categorización de vehículos:

1 = Automóvil, Sedan, Coupe y Hatch Back.

2 = Jeep y Station wagon.

3 = Furgoneta Cerrada, Microbús y Panel Cerrado.

4 = Camioneta y Pick-Up.

5 = Camión.

El porcentaje de siniestros declarados utilizando la variable categoría del vehículo se resume en la tabla 3.3.

**Tabla 3.3 Porcentaje de siniestros declarados, según categoría del vehículo**

| | | | Categoría de Vehículo | | | | | Total |
|---|---|---|---|---|---|---|---|---|
| | | | 1 | 2 | 3 | 4 | 5 | |
| núm_sin | 0 | Recuento | 624 | 264 | 36 | 270 | 34 | 1228 |
| | | Frecuencia esperada | 634.7 | 264.0 | 41.5 | 255.1 | 32.7 | 1228.0 |
| | | % de categoría | 86.8% | 88.3% | 76.6% | 93.4% | 91.9% | 88.3% |
| | 1 | Recuento | 65 | 28 | 9 | 17 | 2 | 121 |
| | | Frecuencia esperada | 62.5 | 26.0 | 4.1 | 25.1 | 3.2 | 121.0 |
| | | % de categoría | 9.0% | 9.4% | 19.1% | 5.9% | 5.4% | 8.7% |
| | 2 o más | Recuento | 30 | 7 | 2 | 2 | 1 | 42 |
| | | Frecuencia esperada | 21.7 | 9.0 | 1.4 | 8.7 | 1.1 | 42.0 |
| | | % de categoría | 4.2% | 2.3% | 4.3% | .7% | 2.7% | 3.0% |
| Total | | Recuento | 719 | 299 | 47 | 289 | 37 | 1391 |
| | | Frecuencia esperada | 719.0 | 299.0 | 47.0 | 289.0 | 37.0 | 1391.0 |
| | | % de categoría | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% |

En la tabla anterior podemos observar que la categoría de vehículos 4 muestra menos siniestros, ocurriendo lo contrario con la categoría 3 que presenta un alto porcentaje de siniestralidad, los vehículos de la categoría 2 muestran una distribución similar a la de la población global.

Aplicando la prueba chi-cuadrado para el estudio de la relación entre las variables número de siniestros y categoría de vehículo se obtuvo el siguiente resultado generado por el software SPSS:

**Tabla 3.4 Prueba Chi-cuadrado**

| | Valor | gl | Sig. Asintótica (bilateral) |
|---|---|---|---|
| Chi-cuadrado de Pearson | 20.132(a) | 8 | 0.01 |
| Razón de Verosimilitud | 21.086 | 8 | 0.007 |
| Número de casos válidos | 1391 | | |

a 4 cells (26.7%) have expected count less than 5. The minimum expected count is 1.12.

A partir del resultado anterior, con un 95% de confianza, podemos afirmar que el número de siniestros declarados por un asegurado depende de la categoría del vehículo asegurado, puesto que la prueba de chi-cuadrado anterior nos revela un nivel de significancia asintótica de 0.01, encontrándose en los niveles aceptables de asociación.

Haciendo un análisis similar para las variables número de siniestros y uso de vehículo se resumen los resultados en las tablas 3.5 y 3.6.

### Tabla 3.5 Porcentaje de siniestros declarados, según uso del vehículo

| | | | Uso Vehículo | | |
|---|---|---|---|---|---|
| | | | COMERCIAL | PARTICULAR | Total |
| núm_sin | 0 | Recuento | 4 | 1224 | 1228 |
| | | Frecuencia esperada | 3.5 | 1224.5 | 1228.0 |
| | | % de Uso_Veh | 100.0% | 88.2% | 88.3% |
| | 1 | Recuento | 0 | 121 | 121 |
| | | Frecuencia esperada | .3 | 120.7 | 121.0 |
| | | % de Uso_Veh | .0% | 8.7% | 8.7% |
| | 2 a más | Recuento | 0 | 42 | 42 |
| | | Frecuencia esperada | .1 | 41.9 | 42.0 |
| | | % de Uso_Veh | .0% | 3.0% | 3.0% |
| Total | | Recuento | 4 | 1387 | 1391 |
| | | Frecuencia esperada | 4.0 | 1387.0 | 1391.0 |
| | | % de Uso_Veh | 100.0% | 100.0% | 100.0% |

Con respecto al uso, se nota que los vehículos Comerciales no presentan siniestros, esto se debe posiblemente a la baja representatividad de este sobre el total de la población, contrario a los de uso particular que siguen la tendencia global.

### Tabla 3.6 Prueba Chi-cuadrado

| | Valor | gl | Sig. Asintótica (bilateral) |
|---|---|---|---|
| Chi-cuadrado de Pearson | 0.532(a) | 2 | 0.766 |
| Razón de Verosimilitud | 0.999 | 2 | 0.607 |
| Número de casos válidos | 1391 | | |

a. 3 cells (50.0%) have expected count less than 5. The minimum expected count is .12.

Revalidando el análisis descriptivo, la prueba no paramétrica del chi-cuadrado, con una significancia asintótica de 0.766 nos confirma que el uso del vehículo no influye en el número de siniestros declarados por póliza, este resultado quizá se deba a que nuestro estudio solo incluyó a personas naturales.

El análisis correspondiente a las variables número de siniestros y procedencia del asegurado se presenta en las tablas 3.7 y 3.8.

**Tabla 3.7 Porcentaje de siniestros declarados, según procedencia del asegurado**

| | | | Procedencia | | Total |
|---|---|---|---|---|---|
| | | | MANAGUA | RESTO | |
| Núm_Sin | 0 | Recuento | 1138 | 90 | 1228 |
| | | Frecuencia esperada | 1138.8 | 89.2 | 1228.0 |
| | | % de Procedencia | 88.2% | 89.1% | 88.3% |
| | 1 | Recuento | 112 | 9 | 121 |
| | | Frecuencia esperada | 112.2 | 8.8 | 121.0 |
| | | % de Procedencia | 8.7% | 8.9% | 8.7% |
| | 2 a más | Recuento | 40 | 2 | 42 |
| | | Frecuencia esperada | 39.0 | 3.0 | 42.0 |
| | | % de Procedencia | 3.1% | 2.0% | 3.0% |
| Total | | Recuento | 1290 | 101 | 1391 |
| | | Frecuencia esperada | 1290.0 | 101.0 | 1391.0 |
| | | % de Procedencia | 100.0% | 100.0% | 100.0% |

En la tabla 3.7 no se presentan diferencias significativas en cuanto al número de siniestros declarados tomando en consideración la procedencia del asegurado, de los asegurados de Managua el 3.10% declaran más de un siniestro, en relación al 2.00% declarados por asegurados del Resto del país.

**Tabla 3.8 Prueba Chi-cuadrado**

| | Valor | gl | Sig. Asintótica (bilateral) |
|---|---|---|---|
| Chi-cuadrado de Pearson | 0.404(a) | 2 | 0.817 |
| Razón de Verosimilitud | 0.454 | 2 | 0.797 |
| Número de casos válidos | 1391 | | |

a.1 cells (16.7%) have expected count less than 5. The minimum expected count is 3.05.

La prueba de bondad de ajuste, con significancia de 0.817 nos revela que el número de siniestros declarados por póliza es independiente de la procedencia del asegurado, este resultado podría estar influenciado por la poca representatividad de la categoría resto del país sobre la muestra con sólo 90 asegurados.

Del mismo modo, el análisis descriptivo y de relación para las variables número de siniestros y sexo, así como número de siniestros y edad se indica en las tablas 3.9, 3.10, 3.11 y 3.12.

**Tabla 3.9 Porcentaje de siniestros declarados, según sexo del asegurado**

| | | | Sexo | | Total |
|---|---|---|---|---|---|
| | | | F | M | |
| núm_sin | 0 | Recuento | 497 | 731 | 1228 |
| | | Frecuencia esperada | 493.5 | 734.5 | 1228.0 |
| | | % de Sexo | 88.9% | 87.9% | 88.3% |
| | 1 | Recuento | 41 | 80 | 121 |
| | | Frecuencia esperada | 48.6 | 72.4 | 121.0 |
| | | % de Sexo | 7.3% | 9.6% | 8.7% |
| | 2 a más | Recuento | 21 | 21 | 42 |
| | | Frecuencia esperada | 16.9 | 25.1 | 42.0 |
| | | % de Sexo | 3.8% | 2.5% | 3.0% |
| Total | | Recuento | 559 | 832 | 1391 |
| | | Frecuencia esperada | 559.0 | 832.0 | 1391.0 |
| | | % de Sexo | 100.0% | 100.0% | 100.0% |

La distribución del número de siniestros declarados por póliza no varia de forma significativa si consideramos la variable sexo, se puede observar que los asegurados de sexo masculino que declaran uno o más siniestros representan el 12.10% respecto a su sexo, en comparación a los de sexo femenino con 11.10%, lo que equivale a una diferencia de 1.00% entre ambos.

**Tabla 3.10 Prueba Chi-cuadrado**

| | Valor | gl | Sig. Asintótica (bilateral) |
|---|---|---|---|
| Chi-cuadrado de Pearson | 3.724(a) | 2 | 0.155 |
| Razón de Verosimilitud | 3.732 | 2 | 0.155 |
| Número de casos válidos | 1391 | | |

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 16.88.

El nivel de significancia obtenido entre las variables sexo y número de siniestros fue de 0.155, siendo mayor que 0.05, se encuentra fuera de los niveles de asociación, y por lo tanto se concluye que el número de siniestros declarados no esta influenciado por el sexo del asegurado.

**Tabla 3.11 Porcentaje de siniestros declarados, según tramos de edad**

| | | | Edad | | Total |
|---|---|---|---|---|---|
| | | | < = 43 años | > 43 años | |
| núm_sin | 0 | Recuento | 608 | 620 | 1228 |
| | | Frecuencia esperada | 613.6 | 614.4 | 1228 |
| | | % de edad | 87.50% | 89.10% | 88.30% |
| | 1 | Recuento | 53 | 68 | 121 |
| | | Frecuencia esperada | 60.5 | 60.5 | 121 |
| | | % de edad | 7.60% | 9.80% | 8.70% |
| | 2 a más | Recuento | 34 | 8 | 42 |
| | | Frecuencia esperada | 21 | 21 | 42 |
| | | % de edad | 4.90% | 1.10% | 3.00% |
| Total | | Recuento | 695 | 696 | 1391 |
| | | Frecuencia esperada | 695 | 696 | 1391 |
| | | % de edad | 100.00% | 100.00% | 100.00% |

El comportamiento de la variable dependiente en función de la edad muestra una tendencia diferente a las otras agrupaciones, en la tabla 3.11 se observa que a medida que aumenta la edad disminuye el número de siniestros declarados, del total de asegurados mayores de 43 años el 89.10% no declara siniestros y sólo el 1.10% declara más de dos siniestros.

La diferencia entre los dos grupos se ve marcada por los asegurados que declaran más de dos siniestros, donde los menores de 43 años superan en 3.80% al otro grupo de riesgo.

**Tabla 3.12 Prueba Chi-cuadrado**

| | Valor | gl | Sig. Asintótica (bilateral) |
|---|---|---|---|
| Chi-cuadrado de Pearson | 18.071(a) | 2 | 0.000 |
| Razón de Verosimilitud | 19.305 | 2 | 0.000 |
| Número de casos válidos | 1391 | | |

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 20.98.

Como el nivel de significancia obtenido fue de 0.000 afirmamos que el número de siniestros declarados por un asegurado depende de su edad, ya que este es menor que nuestro nivel de confianza.

Basándonos en los resultados anteriores seleccionamos la variable edad como factor de riesgo para el cálculo de la prima pura, ya que presentó la mayor relación con nuestra variable dependiente número de siniestros.

A continuación se presenta la distribución del número de siniestros declarados por póliza, por grupo de riesgo. Además el costo medio de un siniestro y el cálculo de la prima pura respectiva, por grupo.

## Tabla 3.13 Siniestralidad de una cartera de seguros de automóvil

| Grupo de edad | Menores de 44 años | Mayores de 43 años | Total Cartera |
|---|---|---|---|
| Nº Expuestos al riesgo | 695 | 696 | 1,391 |
| Nº Pólizas con n siniestros | | | |
| n = 0 | 608 | 620 | 1,228 |
| n = 1 | 53 | 68 | 121 |
| n = 2 | 30 | 6 | 36 |
| n = 3 | 4 | 1 | 5 |
| n = 4 | 0 | 1 | 1 |
| Número de Siniestros | 125 | 87 | 212 |
| Costo medio de Siniestro (US$) | **1,667.67** | **1,390.60** | **1,554.00** |

**Primas de Riesgo**

$$\Pr = fs * \bar{c}$$

$$\Pr(cartera) = 0.152408 * 1{,}554.00 = 236.84$$

$$\Pr(grupo1) = 0.179856 * 1{,}667.67 = 299.94$$

$$\Pr(grupo\,2) = 0.125 * 1{,}390.60 = 173.84$$

Considerando los datos de la tabla 3.13 se obtienen los parámetros característicos de la distribución Binomial Negativa y Gamma para cada grupo de riesgo, lo que se resume en la tabla 3.14.

**Tabla 3.14 Parámetros característicos de las distribuciones de frecuencia de siniestralidad**

| Grupo de edad | Menores de 44 años | Mayores de 43 años | Total Cartera |
|---|---|---|---|
| Número medio de Siniestros | 0.179856 | 0.125000 | 0.152408 |
| Varianza del número de siniestros | 0.268371 | 0.152478 | 0.211135 |
| Parámetros de la distribución Binomial Negativa | | | |
| $p$ | 0.670177 | 0.819788 | 0.721851 |
| $k$ | 0.365454 | 0.568627 | 0.395529 |
| Parámetros de la Distribución Gamma | | | |
| $\alpha$ | 0.365454 | 0.568627 | 0.395529 |
| $\beta$ | 2.031926 | 4.549020 | 2.595195 |

Se puede notar que existen diferencias significativas entre la media de la cartera y la que corresponde a los grupos de riesgo, si el asegurador utilizara la media de la cartera como referencia para el cálculo de la prima, es decir si éste aplicara una prima general, los menores de 44 años resultarían infratarificados con primas más bajas y los conductores mayores de 43 años serian fuertemente penalizados. Como es de esperarse la menor varianza la presentan los mayores de 43.

La distribución del número de siniestros observados y ajustados, así como su correspondiente prueba de bondad de ajuste a la distribución Binomial Negativa y Poisson se resumen en la tabla 3.15 y 3.16.

## Tabla 3.15 Distribución del número de siniestros en una cartera de 1,391 pólizas de automóvil

| Número de Siniestros por póliza | No. Pólizas (Observados) | Frecuencias Ajustadas | | | |
|---|---|---|---|---|---|
| | | P(X = x) | Poisson | P(X = x) | Binomial Negativa |
| n = 0 | 1,228 | 0.858638 | 1194.364893 | 0.879046 | 1222.7536789 |
| n = 1 | 121 | 0.130864 | 182.031170 | 0.096709 | 134.5225994 |
| n = 2 | 36 | 0.009972 | 13.871534 | 0.018770 | 26.1084623 |
| n = 3 | 5 | 0.000507 | 0.704712 | 0.004169 | 5.7988120 |
| n = 4 | 1 | 0.000019 | 0.026851 | 0.000984 | 1.3691910 |
| Total | 1391 | 0.999999 | 1390.999160 | 0.999678 | 1390.552744 |

## Tabla 3.16 Prueba de Bondad de Ajuste

| | Poisson | Binomial Negativa |
|---|---|---|
| Chi-cuadrado | 118.160 | 5.339 |
| Coef. Contingencia | 0.280 | 0.062 |

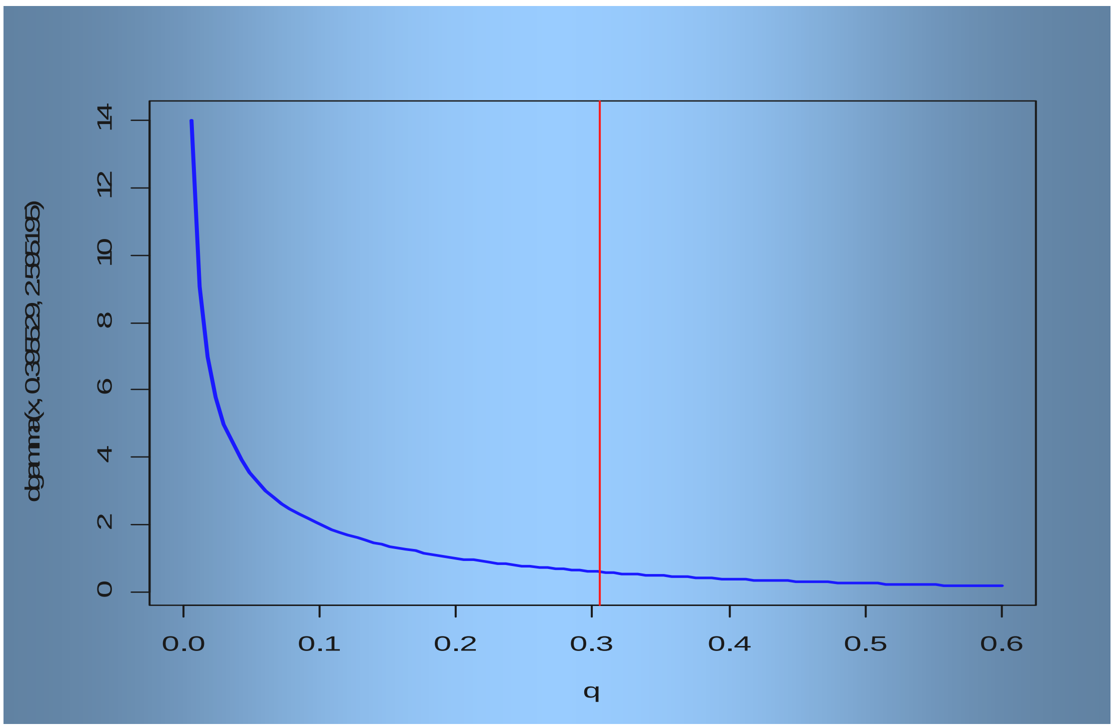
$$X^2 = \sum_i \sum_j \frac{\left(n_{ij} - m_{ij}\right)^2}{m_{ij}}$$

$$C = \sqrt{\frac{X^2}{N + X^2}}$$

La distribución Binomial negativa resulta obviamente ser un mejor modelo que la distribución de Poisson para la distribución del número de siniestros de la cartera total, esto se corrobora mediante los resultados de las pruebas de bondad de ajuste, tanto los valores del Chi-cuadrado como el Coeficiente de contingencia (cercano a cero), en ambos casos se observa que, con un ajuste a la Distribución Binomial negativa las distancia entre valores observados y ajustados son mínimas. En consecuencia los riesgos analizados son heterogéneos.

Para una mejor visualización de la heterogeneidad de la cartera se construyó la gráfica de la función de densidad de la frecuencia de siniestralidad, además se hizo el cálculo de la probabilidad que un asegurado elegido al azar presente una frecuencia de accidente mayor al doble de la media (0.304817).

**Figura 1. Función de densidad de probabilidad de la frecuencia de siniestralidad q de un asegurado, elegido al azar, correspondiente a la distribución de la tabla 3.15**




$$f(q) = \frac{2.60}{\Gamma(0.395529)} * e^{-2.60\,q} * (2.60\,q)^{-0.60}$$

$$f(q) = \frac{2.60}{2.2439} * e^{-2.60\,q} * (2.60\,q)^{-0.60}$$

$$F(q) = p(Q \leq q)$$

$$p(Q > q) = 1 - F(q)$$

$$p(Q > 0.304817) = 1 - p(q \leq 0.304817)$$

$$p(Q > 0.304817) = 1 - 0.841243 = 0.158757$$

Observando la forma de la gráfica de la función de densidad de la frecuencia de siniestralidad correspondiente a la cartera se deduce que es un grupo heterogéneo, que muestra una larga cola, es decir, que la frecuencia de siniestralidad no se concentra alrededor de su media. Considerando el área bajo la función de densidad de probabilidad a la derecha de $q = 0.304817$, podemos concluir que aproximadamente el 15.88% de los asegurados tienen una frecuencia de accidente superior al doble de su media.

A continuación se presenta un análisis similar al realizado para la cartera, utilizando la información correspondiente a cada grupo de riesgo.

**Tabla 3.17 Distribución del número de siniestros en un grupo de 695 pólizas de automóvil correspondientes a los asegurados con edades menores de 44 años**

| Número de Siniestros por póliza | No. Pólizas (Observados) | Frecuencias Ajustadas | | | |
|---|---|---|---|---|---|
| | | **P(X = x)** | **Poisson** | **P(X = x)** | **Binomial Negativa** |
| n = 0 | 608 | 0.835390 | 580.596330 | 0.863933 | 600.4335208 |
| n = 1 | 53 | 0.150250 | 104.423800 | 0.104134 | 72.3734765 |
| n = 2 | 30 | 0.013512 | 9.390630 | 0.023449 | 16.2969941 |
| n = 3 | 4 | 0.000810 | 0.562987 | 0.006098 | 4.2382069 |
| n = 4 | 0 | 0.000036 | 0.025314 | 0.001692 | 1.1761080 |
| Total | 695 | 0.999999 | 694.999061 | 0.999307 | 694.518306 |

## Tabla 3.18 Prueba de Bondad de Ajuste

| | Poisson | Binomial Negativa |
|---|---|---|
| Chi-cuadrado | 92.856 | 17.993 |
| Coef. Contingencia | 0.343 | 0.159 |

Según los resultados de las pruebas de bondad, el grupo de riesgo formado por los asegurados menores de 44 años no tiene un buen ajuste a ninguna de las dos distribuciones consideradas, aunque se nota que se acerca más a una distribución Binomial negativa. La ausencia de ajuste se debe probablemente al tamaño de muestra, ya que al aumentar la información estadística mejora la fiabilidad con que se estiman los parámetros $p$ y $k$.

## Tabla 3.19 Distribución del número de siniestros en un grupo de 696 pólizas de automóvil correspondientes a los asegurados con edades mayores de 43 años

| Número de Siniestros por pólizas | No. Pólizas (Observados) | Frecuencias Ajustadas | | | |
|---|---|---|---|---|---|
| | | P(X = x) | Poisson | P(X = x) | Binomial Negativa |
| n = 0 | 620 | 0.882497 | 614.217844 | 0.893158 | 621.6380245 |
| n = 1 | 68 | 0.110312 | 76.777231 | 0.091525 | 63.7014230 |
| n = 2 | 6 | 0.006895 | 4.798577 | 0.012936 | 9.0037250 |
| n = 3 | 1 | 0.000287 | 0.199941 | 0.001996 | 1.3892671 |
| n = 4 | 1 | 0.000009 | 0.006248 | 0.000321 | 0.2233627 |
| Total | 696 | 1.000000 | 695.999840 | 0.999936 | 695.955802 |

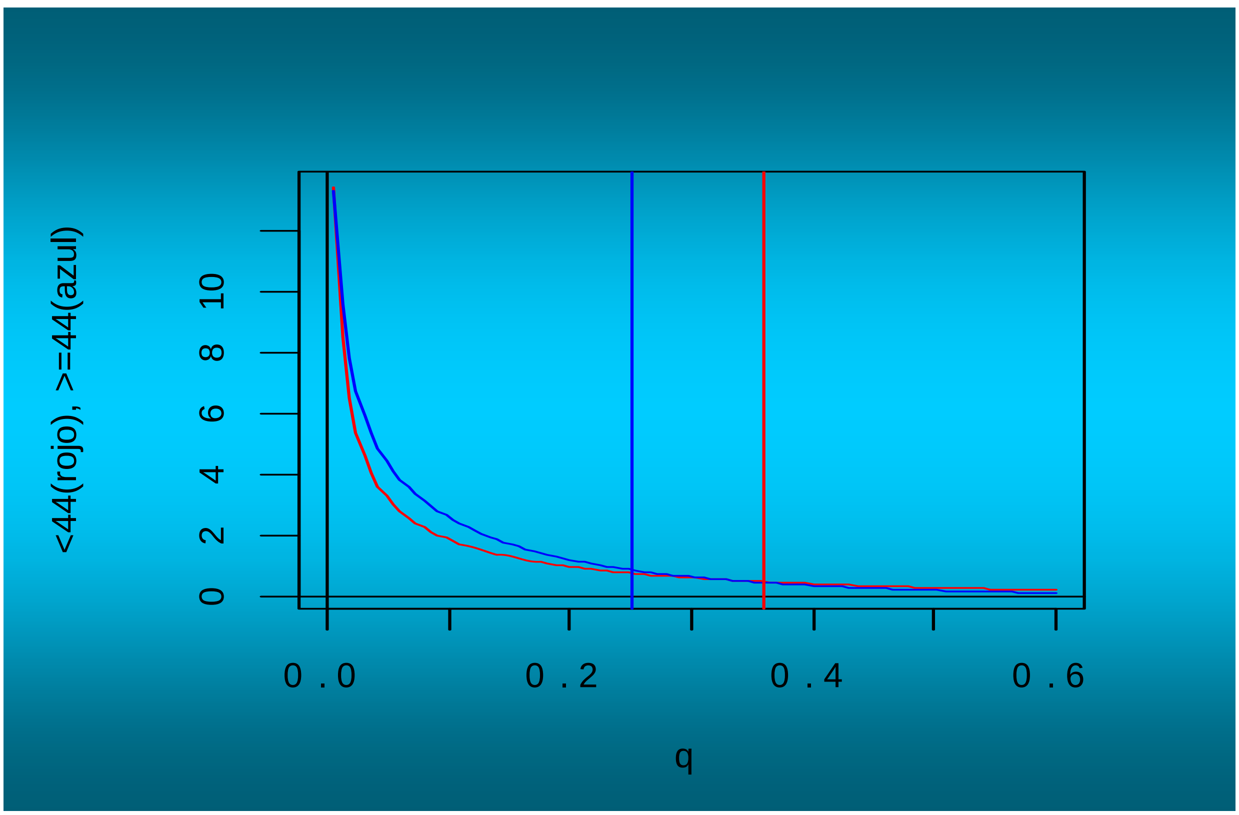
## Tabla 3.20 Prueba de Bondad de Ajuste

| | Poisson | Binomial Negativa |
|---|---|---|
| Chi-cuadrado | 162.614 | 4.106 |
| Coef. Contingencia | 0.435 | 0.077 |

Al igual que la cartera, los datos correspondientes a los asegurados mayores de 43 años se ajustan a una distribución Binomial negativa, podemos ver como el valor del Chi-cuadrado es menor con relación al obtenido con la distribución de Poisson, además un coeficiente de contingencia muy cercano a cero nos confirma este buen ajuste. Podemos entonces afirmar que los datos son heterogéneos.

La heterogeneidad por grupo se muestra en la figura 2, después se presenta el cálculo de la probabilidad, que un asegurado de cada grupo tenga una frecuencia de siniestralidad mayor al doble de la media del grupo al que corresponde.

**Figura 2. Función de densidad de probabilidad de la frecuencia de siniestralidad q de un asegurado, elegido al azar, correspondiente a la distribución de la tabla 10 (línea roja, menores de 44 años) y 11 (línea azul, mayores de 43 años).**



Según la forma de la figura 2. de la función de densidad de “q”, correspondiente a los grupos 1 y 2, nos muestra que ambos son heterogéneos, con la diferencia que los menores de 44 años presentan una heterogeneidad superior que los mayores de 43 años.

Para el primer grupo de riesgo la probabilidad que “q” sea mayor que 0.359712 es de 0.1586, en el caso del grupo 2 la probabilidad que sea mayor a 0.25 resulta ser 0.1552, esto nos indica que aproximadamente el 15% de los asegurados de cada grupo presentan una probabilidad mayor al doble de la media del grupo, hay que remarcar que la media del segundo grupo es menor que la del primero, indicando así, que los asegurados del grupo 2 presentaron una frecuencia de accidente menor.

$$f(q_1) = \frac{2.03}{\Gamma(0.365454)} * e^{-2.03q} * (2.03q)^{-0.63}$$

$$f(q_1) = \frac{2.03}{2.4345} * e^{-2.031926q} * (2.03q)^{-0.63}$$

$$p(Q_1 > 0.359712) = 1 - F(q \leq 0.359712)$$

$$p(Q_1 > 0.359712) = 1 - 0.841413 = 0.158587$$

$$f(q_2) = \frac{4.55}{\Gamma(0.568627)} * e^{-4.55q} * (4.55q)^{-0.43}$$

$$f(q_2) = \frac{4.55}{1.5658} * e^{-4.55q} * (4.55q)^{-0.43}$$

$$p(Q_2 > 0.25) = 1 - F(q \leq 0.25)$$

$$p(Q_2 > 0.25) = 1 - 0.844787 = 0.155213$$

Otra alternativa para el análisis de la heterogeneidad se realiza a través de la desviación típica, las que se observan en la tabla 3.21.

**Tabla 3.21 Media y desviación típica de la distribución de frecuencia de siniestralidad: Dos grupos de riesgo y la cartera total**

| Grupo de Riesgo | Distribución de la frecuencia de siniestralidad | |
|---|---|---|
| | Media | Desviación típica |
| Menores de 44 años | 0.1799 | 0.2975 |
| Mayores de 43 años | 0.1250 | 0.1658 |
| Cartera Total | 0.1524 | 0.2423 |

Observando las desviaciones típicas de cada grupo de riesgo podemos deducir que, en general, ninguno de los dos grupos formados por la muestra de asegurados resulta sustancialmente homogéneo, esto también se observa a través de sus funciones de densidad (figura 2), ambas resultan con grandes colas lo que representa que la frecuencia de siniestralidad de cada individuo del grupo es muy variable, es decir la media del grupo no es muy confiable.

Si bien, el grupo de los mayores de 43 años se ajusta adecuadamente a una binomial se observa que su desviación típica, siendo menor que la de la cartera, aun continúa elevada. La introducción de otro factor de riesgo, el aumento de la muestra y el incremento de los subintervalos podría posiblemente reducir de forma significativa la heterogeneidad residual que es muy alta.

Es de esperarse que el grupo de menores de 44 años presente una desviación típica mayor a la de la cartera debido al mal ajuste que ésta presenta a la distribución Binomial Negativa.

# CAPÍTULO 4
# CONCLUSIONES

Para aproximar primas del seguro de automóvil ajustadas al riesgo que asume una compañía de seguros utilizando variables de tarifa, se debe determinar la influencia de los factores de riesgo en el número de siniestros declarados por póliza, empleando en el cálculo de la prima las variables que presenten mayor relación respecto a la variable explicada. Se debe contar con una base de datos completa y suficiente como garantía de la relación entre las variables y la estimación del cálculo de la prima.

En nuestro estudio los factores de riesgo con influencia sobre el número de siniestros fueron la edad y la categoría del vehículo, de ambos tomamos el de mayor influencia que es la edad. Se estratificó la cartera en dos grupos de riesgos, estos grupos presentaron heterogeneidad, siendo más considerable la del primer grupo (menores de 44 años), es decir, en este intervalo de edades se encuentran los peores riesgos, dando como resultado una media mayor a la de la cartera y costos de siniestralidad más altos, en comparación al segundo grupo (mayores de 43 años) que, aunque presentan similares características en cuanto a su distribución poseen una media menor a la de la cartera.

Las primas puras obtenidas por grupo de riesgos fueron de US$ 299.94 para los menores de 44 años y de US$ 173.84 para los mayores de 43 años, observando la prima de la cartera que fue de US$ 236.84 podemos notar que existen diferencias significativas, considerando que el número de coberturas asignadas a cada póliza podría estar entre cinco y ocho. Si los datos del estudio hubiesen presentado una distribución homogénea por grupo las primas a cobrar serian las anteriores.

Los resultados obtenidos en nuestro estudio, no nos dan pautas para establecer confianza en la prima calculada basada en esta estratificación debido a la poca confiabilidad de los datos utilizados. Tomando en cuenta que en nuestro estudio se asumió que los variables relacionadas al asegurado corresponden al mismo individuo siniestrado.

Para disminuir la heterogeneidad residual de cada grupo de riesgo se podría incrementar la muestra, incluir un nuevo factor de riesgo o aumentar el número de subintervalos dentro del factor utilizado.

# CAPÍTULO 5
# RECOMENDACIONES

- Las Aseguradoras Nicaragüenses deben perfeccionar el sistema de recolección de datos estadísticos, tanto para el asegurado, su vehículo, la siniestralidad reportada por éste y datos relacionados con el conductor del vehículo al momento del siniestro, para que de esta manera se lleve un mejor control y los resultados de cualquier estudio específico a realizar se acerquen más a la realidad.

- La relación entre la Universidad y las Compañías de Seguros debe darse sobre una base de beneficio mutua. Las empresas de Seguros facilitando la información sobre problemas de interés a resolver, a través de estudios de investigación por parte de los estudiantes egresados.

- Que este trabajo sirva como base para el desarrollo de otros estudios de investigación relacionados con el cálculo de prima de seguros de automóvil o la ampliación del mismo utilizando estadísticas más confiables.

- Que los estudiantes de la carrera Licenciatura en Ciencias Actuariales y Financieras, profundicen las investigaciones relacionadas con el cálculo de primas de Seguros de Daños.

# BIBLIOGRAFÍA


- Bonilla Gildaberto, Como hacer una tesis de graduación con técnicas estadísticas, Primera edición, San Salvador, El Salvador, 1993, UCA editores, Universidad Centroamericana José Simeón Cañas.

- Boj del Val Eva y Claramunt Bielsa. Tarificación aposteriori del seguro de automóvil. Obtenido el 17 de Enero del 2007 en:
  http://www.ine.es/revistas/estaespa/160_5.pdf,
  http://dialnet.unirioja.es/servlet/oaiart.
  http://biblioteca.universia.net/html_bura/ficha/paramos/id/2126895.

- Hossack I.B, Pollard J.H y Zehnwirth (2001). Introducción a la estadística con aplicaciones a los seguros generales. Editorial MAPFRE, compuesto e impreso en los talleres de Fernández Ciudad, S.L, Catalina Suárez 19-28007 Madrid. Traducción de: Vegas Montaner Ángel, Universidad de Alcalá.

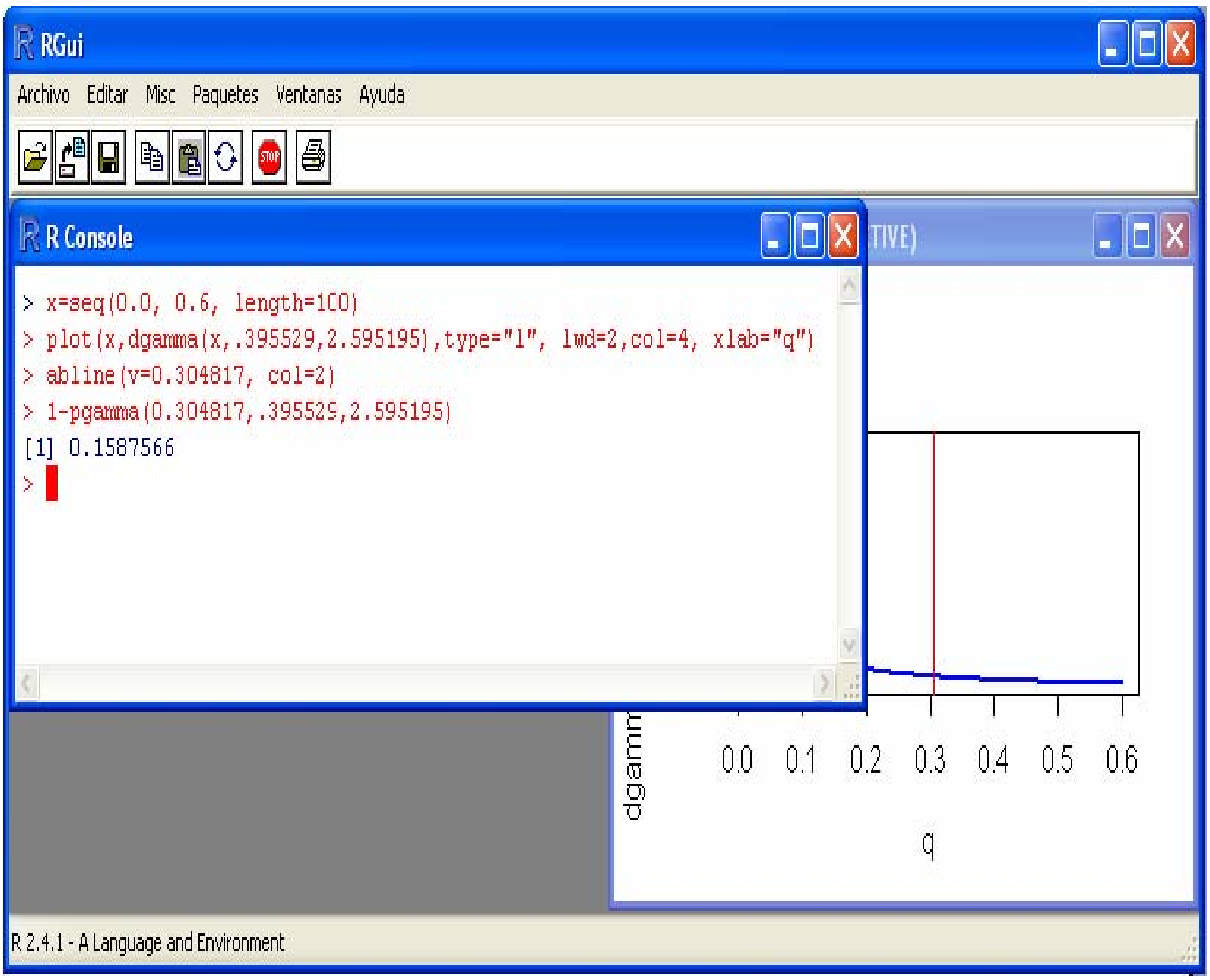
- Melgar Hiraldo María del Carmen, Ordaz Sanz José Antonio y Guerrero Casas Flor María. "Análisis de culpabilidad en los siniestros del seguro de automóvil". Departamento de de Economía, Métodos cuantitativos e historia económica, Universidad Pablo de Olavide, XIII Jornada de ASEPUMA. http://www.uv.es/asepuma/jornadas/santiago¨.

- Piura L., Julio. Metodología de la Investigación científica: Un enfoque integrador. Editorial PAVSA. Managua. 2006.

- Visauta Vinacua Bienvenido, Análisis estadístico con SPSS para Windows, Volumen I, Estadística básica, Segunda edición, Editorial McGraw-Hill/Interamericana de España, S.A.U 2002.

# ANEXOS

## Anexo 1. Función de densidad de q y cálculo de probabilidades de la cartera en el programa estadístico R.



**Anexo 2. Función de densidad de q y cálculo de probabilidades de los dos grupos de riesgo en el programa estadístico R.**