

Rea[®]

Revista Electrónica de la Actividad Aseguradora



REVISTA 1

Reseña documental del análisis comparativo y la aplicación de modelos de representación de factores de riesgo en el ramo del automóvil casco del mercado asegurador venezolano

Carlos Javier Yépez Ballesteros*¹

Sumario: I. Planteamiento del problema. II. Objetivos. III importancia de la investigación y la justificación. IV. Antecedentes. V. Materiales. 1. Fuentes de información. 2. Universo. 3. Unidades estadísticas. 4. Variables. 5. Coberturas. VI. Metodología. 1. Estructura del estudio. 2. Tipo y diseño del presente artículo. 3. Métodos utilizados. VII. Desarrollo de la reseña documental. 1. Método estadístico que fundamenta la estrategia a utilizar en una investigación cuantitativa. 2. Método matemático, modelo aleatorio de pérdidas agregadas. 3. Modelo aleatorio de pérdidas agregadas. 4. Modelo aleatorio de ocurrencia de eventos (siniestros/reclamos). 5. Modelo aleatorio de severidad asociado a cada evento (siniestro/reclamo). 6. Modelo aleatorio conjunto de pérdidas agregadas que une los dos modelos anteriores. VIII. Metodología matemática en la ciencia de los datos. IX. Desarrollo resumen genérico. X. Conclusiones y recomendaciones. XI. Referencias bibliográficas.

Resumen: El objetivo de la investigación es reseñar documentalmente el estudio del uso de las técnicas del aprendizaje estadístico con el fin de obtener agrupaciones para estimar un modelo aleatorio llamado la máxima pérdida probable de los reclamos en el seguro de casco de los automóviles, según los factores de riesgo. Se describieron la comparación de dos modelos matemáticos de representación, a través de K medias y el análisis de componentes principales, incorporando los modelos de agrupamiento, se mostró en teoría las particularidades de los modelos, las similitudes y sus diferencias. En el estudio se representó la metodología para detectar las características relevantes entre los modelos del aprendizaje estadístico en lo datos con múltiples variables; las variables monto facturado del siniestro,

1 Currículo: Magister Scientiarum mención Modelos Aleatorios (UCV), Especialista en Estadísticas (UCV), Licenciado en Ciencias Actuariales (UCV), Tesista en la Maestría en Estadísticas (UCV). Profesor Contratado Escuela de Altos Estudios de la Actividad Aseguradora y Reaseguros, Lic. FSH. Profesor con cargo convencional 8 horas en la Universidad Central de Venezuela. Profesor Contratado Instituto Universitario Tecnológico Alberto Adrianni. Funcionario Actuario VI en la Superintendencia de la Actividad Aseguradora. Correo: cjypezb@gmail.com.

*Agradecimientos al Doctor Ricardo Ríos por la co-redacción y tutoría del presente trabajo.

valor del vehículo y la marca modelo son las que, en teoría, pueden arrojar mayor información como factores de riesgo. Se relató la utilización de los reclamos para incorporarlos en el modelo aleatorio de pérdida agregada bajo incertidumbre, y se narró la manera de estimar la máxima pérdida probable de los reclamos con información que deben ser de aseguradoras seleccionadas previamente, del mercado asegurador venezolano. Se informó la técnica de estimar el anterior modelo probabilístico de pérdidas agregadas basado, a su vez, en el modelo probabilístico de las ocurrencias que afecta la pérdida y el modelo probabilístico de la severidad asociada a los eventos causantes de la pérdida en volumen monetario, de uno de los componentes descritos como factores de riesgo.

Palabra Clave: Aprendizaje estadístico, modelos aleatorios de múltiples variables.

Abstract: The objective of the research is to document the study of the use of statistical learning techniques in order to obtain groupings to estimate a random model called the maximum probable loss of claims in automobile helmet insurance, according to the risk factors. The comparison of two mathematical models of representation was described, through K means and principal components analysis, incorporating the grouping models, the particularities of the models, the similarities and their differences were shown in theory. The study represented the methodology to detect relevant characteristics between statistical learning models in multivariate data; The variables invoiced amount of the claim, value of the vehicle and the model make are those that, in theory, can provide the most information as risk factors. The use of the claims to incorporate them into the random model of aggregate loss under uncertainty was reported, and the way to estimate the maximum probable loss of the claims was narrated with information from insurers that must previously be selected from the Venezuelan insurance market. The technique of estimating the previous probabilistic model of aggregate losses based, in turn, on the probabilistic model of the occurrences that affects the loss and the probabilistic model of the severity associated with the events causing the loss in monetary volume, of one of the components described as risk factors, was reported.

Keyword: Statistical learning, multivariate random models.

I. PLANTEAMIENTO DEL PROBLEMA

Las empresas del mercado asegurador venezolano manejan su propia información de vehículos para utilizarlo en la estimación de los siniestros total anual, esto como insumo en la realización de las tarifas; esta información difieren entre sí, algunas aseguradoras, por su tamaño de cartera, se ven afectadas al no poder construir un conglomerado real, muchas veces la poca información obliga a estas aseguradoras a recurrir a otras compañías que pueden no tener la misma estructura o servicio y pueden tener un histórico de siniestros atípicos, produciéndose una distorsión, afectando la coherencia de la empresa que asume los conglomerados en desventaja comercial al agrupar un vehículo en un grupo que no describe su propia cartera de clientes y además, que al vehículo se le podría asignar tasas que no correspondan a los siniestros estimados reales.

Se puede apreciar a través de los conglomerados como la manera previa de realizar las tarifas. Las técnicas son variadas y extensas pero cada fenómeno abarca una situación particular donde unos modelos de conglomerados pueden ser pocos eficientes al momento de querer tener una agrupación lógica de los datos. Pasa con los datos con múltiples variables de un período previamente determinado y los datos longitudinales de múltiples variables, ya que los individuos, además de ser parecidos, también deben estar relacionados en la evolución del tiempo. Es necesario realizar una determinación de una distancia entre sujetos para datos temporales que permita aplicar modelos de clasificación, estableciendo criterios y procedimiento de comparación.

El modelo “análisis de componentes principales” se aplica para identificar de forma primaria los grupos con datos de manera

transversal, después se confirma los grupos con el “análisis discriminante” y determinar las dimensiones en estudio que más información describa, estas dimensiones son los factores de riesgo para los vehículos asegurados que están en estudio.

El contrato de seguros de casco de automóvil, está inmerso en el esquema de alta volatilidad (variación), causado por factores cuyo consecuente control requieren necesariamente una aproximación que utilice técnicas estadísticas en las que se combine una perspectiva de múltiples variables con elementos que evalúen el fenómeno basándose en la teoría estadística del riesgo. Los factores que alteran los reclamos pertenecientes a la cobertura de pérdida parcial en las pólizas de seguros de casco de automóvil se manifiestan en el vehículo, asegurado o conductor habitual y el tipo de contrato, que se aprecian antes de la ocurrencia de uno o más eventos, aislados o simultáneamente ocurridos, con mayor o menor intensidad, respecto de lo cual cabe resaltar que en numerosas oportunidades el volumen monetario que se factura es la respuesta a eventos aleatorios, que están previamente especificado en las condiciones generales y particulares del contrato de Seguros.

Según Daykin: “(...) alrededor de un siglo atrás se reconoció la necesidad de complementar el uso de técnicas clásicas determinísticas, con el análisis de la variabilidad (...), en el estudio de fenómenos complejos, dando origen de esta manera a la teoría del riesgo (...)”². El enfoque planteado por este autor es particularmente adecuado para analizar los procesos previos a la realización de tarifas del ramo de seguros de casco de automóvil sobre la cobertura de pérdida parcial.

En una empresa dedicada a la suscripción de contratos de seguros de automóvil resulta importante la descripción de la totalidad de los

² Daykin, T. Pentikäinen y M. Pesonen. *Practical risk theory for actuaries*. Chapman & Hall, London, 1994, p. 2.

montos de siniestros en el próximo año, así como la tendencia, de la reclamación, con un enfoque dirigido a evaluar el riesgo de la pérdida parcial, lo que permitiría incorporar esta información como referencia en la realización de la máxima pérdida probable del ramo de automóvil casco; Así, la ejecución adecuada de la tarifa tiene un impacto directo sobre la economía de la empresa.

De esta manera se trata de cumplir con tres actividades primordiales que según Jorion³, son: 1. Informar adecuadamente acerca de los riesgos con el objeto de que sean tomados en cuenta en la toma de decisiones, permitiendo establecer alertas tempranas y mitigación, 2. Permitir y facilitar la asignación de recursos en forma objetiva, y 3. Evaluar el desempeño operacional.

Con la presente investigación se pretende dar cuenta de los principales métodos estadísticos que contribuyen al estudio de Factores de Riesgo en el Seguros de Automóvil Casco, los cuales son: a) Análisis de componentes principales, b) Agrupación k medias, c) Análisis discriminante con variables normalizadas, d) Estudio longitudinal, conglomerado a través del análisis de componentes principales, e) Estudio longitudinal, conglomerado a través de una modificación de “K medias” con una distancia propuesta por Chiodi y los principales métodos matemáticos actuarial que contribuyen al estudio de Pérdidas Agregadas según el factor de riesgo previamente seleccionado, que es: a) Modelo aleatorio de pérdidas agregadas, b) Modelo aleatorio de ocurrencia de eventos (siniestros/reclamos), c) Modelo aleatorio de severidad asociado a cada evento (siniestro/reclamo), y d) Modelo aleatorio conjunto de pérdidas agregadas que une los dos modelos anteriores.

A. Interrogantes de la investigación

1. ¿Cuáles reseñas documentales se debe utilizar para

³ P. Jorion. *Valor en Riesgo: El nuevo paradigma para el control de riesgo con derivados*. Limusa Noriega editores, México, 2007, p. 2.

expresar las características que se captan en la comparación de los modelos elaborados en la obtención de conglomerados de análisis de múltiples variables con los factores de riesgo en los seguros de casco del automóvil, a través de técnicas de aprendizaje estadístico no supervisado?

2. ¿Cómo se reseña documentalmente el establecimiento de una unificación de criterio para agrupar determinados tipos de vehículos en una compañía de seguros con información del mercado asegurador venezolano, cuando esta compañía no tenga suficiente o ninguna información?
3. ¿Cómo se reseña documentalmente la aplicación de comparar las técnicas para la obtención de conglomerados con factores de riesgo en el seguro de casco del automóvil en el mercado asegurador de Venezuela?
4. ¿cómo se reseña documentalmente la estimación, para un conglomerado seleccionado, los indicadores de los factores de riesgo bajo incertidumbre, en el contrato, en el seguro de casco del automóvil en el mercado asegurador de Venezuela?

II. OBJETIVOS

A. Objetivo general

Reseñar documentalmente los análisis comparativos integrando los resultados en el uso del modelo de K medias con la aplicación del análisis de componentes principales elaborados para obtener agrupaciones, tomando en cuenta los factores de riesgo en el seguro de casco del automóvil, a través de técnicas de aprendizaje estadístico no supervisado.

B. Objetivos específicos

1. Explicar cualitativamente la realización de una unificación de criterio para agrupar determinados

vehículos en una compañía de seguros con información de parte parcial del mercado asegurador venezolano cuando la primera compañía no tenga suficiente o ninguna información referente a los reclamos de la marca y modelo de determinados vehículos.

2. Describir cualitativamente la aplicación de técnicas de aprendizaje estadísticos para la obtención de factores de riesgo a través de conglomerados en el seguro de casco del automóvil, de parte parcial del mercado asegurador venezolano.
3. Describir cualitativamente la estimación, para un conglomerado obtenido, de indicadores de los factores de riesgo bajo incertidumbre, en el seguro de casco de automóvil, en forma parcial el mercado asegurador venezolano.

III. IMPORTANCIA DE LA INVESTIGACIÓN Y LA JUSTIFICACIÓN

La importancia de la presente investigación cualitativa estriba en que la información que de ella se describa, eventualmente:

- A. Permite informar la forma de mejorar la selección del modelo de agrupamiento, para datos transversales con múltiples variables.
- B. Permite informar la forma de proporcionar conocimiento al mercado asegurador venezolano que sirva de referencia para la planificación, incorporando valores a riesgo y generación de controles de gestión.

“El desarrollo de metodología de responsabilidad de los datos en la empresa. . . es similar a la responsabilidad de cuidar la sangre que fluye por el sistema circulatorio el cuerpo humano. Sin una adecuada y perfeccionada respuesta, la

estructuración y mantenimiento de los datos estadísticos de la empresa... tienen un porcentaje muy alto que puede llevar al fracaso. Continuando con el símil, se habría afectado el sistema nervioso y de control, generando una pérdida total de la orientación y posibilidad de identificar peligros y dificultades”.

En la nota anterior de Rodríguez, se evidencia la importancia de obtener los datos en forma efectiva y eficaz, a continuación, Rodríguez, expresa la conveniencia de mantener los datos resguardados, adecuado a la presente investigación.

En el riesgo de la pérdida parcial en el ramo de automóvil casco, se tiene la necesidad de estudiar, cuando menos, tres elementos que deberán medirse en posteriores estudios: 1) Perfil longitudinal de la pérdida por marca modelo, 2) Variables propias del vehículo, 3) Condiciones externas del entorno: Zona geográfica de utilización del vehículo; pero antes también se tiene la necesidad de agrupar los vehículos desde el punto de vista de múltiples variables para el adecuado análisis del riesgo.

En el anterior párrafo se explica la prioridad de mantener los datos en resguardo de manera que sean confiables e inalterados, así como también tener, en el sistema de riesgo, perfiles de vehículos.

Rodríguez, refleja la importancia de los datos procesados y de agrupar los datos por vehículo.

El problema planteado se podría resolver mediante la posterior agrupación de los vehículos a través del apoyo de las nuevas curvas de marca y modelos descritos.

IV. ANTECEDENTES

El ramo de automóvil en el seguro y las ciencias actuariales ha

sido abordado por gran cantidad de autores, entre ellos, Bowers⁴, Daykin⁵; mientras que Kass⁶ realiza una generación pseudo-aleatoria de la cartera de seguros de automóvil y donde las pólizas tienen las siguientes variables consideradas por el autor como “factores de riesgo”: Sexo del conductor habitual, clase de actividad del vehículo, región geográfica del manejo del mismo y tipo de vehículo; como estos factores son diferentes a los planteados en el presente proyecto (ver tablas I y II) solo se podrá realizar una comparación aproximada.

Kass⁷ genera un modelo lineal generalizado para clasificar los vehículos asegurados y determinar un recargo o disminución de su tarifa, de acuerdo a su cantidad de reclamos en el período de cobertura, dentro del contexto del sistema de *bonus-malus*, con los siguientes factores de riesgos: Región de circulación del automóvil, intervalo etario del conductor habitual, el kilometraje del vehículo al concluir el período de estudio, el tipo de uso del vehículo si es privado-particular o empresarial, tipo de agrupación en el *bonus-malus* y el intervalo de clase que se encuentra el vehículo asegurado; donde se estableció grupos adecuados al *bonus-malus*; a pesar que los factores de riesgo son diferentes a la presente investigación, se realiza una comparación aproximada.

También, Kass⁸ determina los siguientes factores de riesgo para un análisis sencillo de la cartera: sexo del conductor habitual, tipo de región de circulación del vehículo (urbano, extra-urbano o en otra parte), tipo de vehículo (pequeño, mediano o grande), clase de

4 Bower, Gerber, Hickman, Jones and Nesbitt. *Actuarial Mathematics*. The Society of actuaries, Schaumburg, Illinois, 1994.

5 Daykin, T. Pentikäinen y M. Pesonen. *Practical risk theory for actuaries, op. cit.*

6 R. Kass, M. Goovaerts, J. Dhaene and M. Denuit. *Modern actuarial risk theory using R*. Springer-Verlag, Berlin, Heidelberg, 2008.

7 *Ídem.*

8 *Ídem.*

trabajo del conductor (empleado de oficina, conductor habitual o conductor dinámico), donde determinó el ranking de cada compañía de seguros; mientras en la presente investigación se plantea realizar grupos de vehículos y no de empresas aseguradoras. Otra forma de abordar el tema lo describe Charpentier^o donde realiza ranking con escala de “*bonus-malus*” a través de modelo lineal generalizado para datos longitudinales con los siguientes factores de riesgos: Año de observación, promedio de reclamos por unidad de expuestos, el código de la ciudad donde circula el vehículo, el ingreso per capital de la ciudad y la población por milla cuadrada de la ciudad.

Finalmente, Charpentier^o indica diez factores de riesgo: Número de contrato ID, número de reclamo durante el período de exposición NB, expuesto en años, poder del vehículo (categoría ordenada), edad del auto en años, edad del conductor habitual en años, marca del vehículo, tipo de combustible utilizado por el vehículo, región de conducción del vehículo en Francia, densidad de la población en la ciudad del conductor habitual donde vive y maneja; mientras que en el presente estudio se plantea realizar a través de aprendizaje estadístico, conglomerados para la posterior utilización en el desarrollo de la función de pérdida agregada en el ramo de seguros de automóvil del casco¹¹.

V. MATERIALES

1. Fuentes de información

La fuente de información que se utiliza en el modelo de pérdidas agregadas es obtenida a partir de una base de datos suministrada a través del “proyecto de unificación de solicitudes” generado por el área de investigación y trabajos especiales de la superintendencia de la actividad aseguradora.

2. Universo

Universo de estudio definido por seis (6) compañías de seguros que

son parte parcial del mercado asegurador venezolano.

3. Unidades estadísticas

En el presente estudio coinciden la unidad estadística de observación y la unidad estadística de análisis, y se describen a continuación.

Para la realización de los modelos de agrupación de datos transversal: la unidad estadística de observación y de análisis es el “vehículo” asegurado por las compañías de seguros.

Para la realización de los modelos de agrupación longitudinal: la unidad estadística de observación y de análisis es la “marca modelo” de los vehículos asegurados por las compañías de seguros.

Cada “marca modelo” está conformado por varios vehículos en exposición que constituyen la fuente primaria de observación; a su vez un vehículo puede tener cero, uno o varios reclamos. Asimismo, pueden existir varios vehículos que componen un contrato de flota, y en una compañía se agrupan varios contratos. Finalmente, el mercado asegurador venezolano unifica las compañías aseguradoras. Para la realización del modelo actuarial de pérdida agregadas: la unidad estadística de observación y de análisis es el “vehículo” asegurado por las compañías de seguros.

4. Variables

Las variables utilizadas en el modelo de agrupación transversal, son: Según el manual de la UPEL¹² (Universidad Pedagógica Experimental Libertador), se debe restringir el uso de líneas en la diagramación de los cuadros a las estrictamente necesarias para aumentar su claridad. Se utiliza sólo líneas horizontales para la separación entre el título del cuadro, los subtítulos de las columnas, el cuerpo de datos y las

12 Universidad Pedagógica Experimental Libertador. *Manual de Trabajos de Grado de Especialización y Maestría y Tesis Doctorales*. Fondo Editorial de la Universidad Pedagógica Experimental Libertador, Caracas, 2016, p. 50.

notas al pie.

Tabla I: Variables sobre el contrato

Variable	Descripción
Número ficticio de la empresa	Para identificar el registro.
Ubicación geográfica, siniestro	De los reclamos ocurridos, período estudiado.
Ubicación geográfica, suscripción	De la emisión del contrato del vehículo.
Fecha de emisión del contrato	Registro para identificar la renovación.
Fecha de inicio de vigencia	Comienzo del período de exposición.
Fecha fin de vigencia	Finalización del período de exposición.
Tipo de Modalidad	Individual o Flota.

Fuente: Sudeaseg.

Tabla II: Variables sobre el vehículo

Variable	Descripción
Año	Lapso de construcción del vehículo.
Marca	Según registro adquirido.
Modelo	Según registro adquirido.
Número de ocupantes	Cantidad de asientos del vehículo.
Tipo de vehículo	Peso del vehículo en kilogramos.
Precio, Suma Asegurada	Valor en Bolívares del vehículo, monto asegurado.

Edad del conductor habitual	No obligatorio, puede faltar.
Estado civil del conductor habitual	No obligatorio: Casado, soltero, viudo.
Sexo del conductor habitual	No obligatorio: Femenino, masculino.

Fuente: Sudeaseg.

Tabla III: Variables sobre el resumen de reclamos por vehículo

Variable	Descripción
Número de reclamos	Cuenta el número de siniestro.
Monto total facturado	Suma los montos facturados.

Fuente: Sudeaseg.

Tabla IV: Variables sobre el detalle de reclamos

Variable	Descripción
Fecha de ocurrencia	Momento que sucedió el siniestro.
Fecha de notificación	Momento que se registró en la empresa.
Causa	Evento que produjo la reclamación.
Monto facturado del reclamo	Volumen monetario en Bolívares.
Monto cubierto	Lo indemnizado por la empresa.
Monto pendiente	La cantidad a pagar por la empresa.

Monto pagado La cantidad que ha pagado la empresa.

Fuente: Sudeaseg.

Las variables utilizadas en el modelo de agrupación longitudinal de series cortas, son las siguientes:

Tabla V: Variables utilizadas, modelo longitudinal de series cortas

Variable	Descripción
Marca-Modelo	De los vehículos.
Número total de expuestos	Por cada Marca-Modelo, por mes acumulado.
Número total de eventos, reclamos	Por cada Marca-Modelo, por mes acumulado.
Monto total facturado	Por cada Marca-Modelo, por mes acumulado.

Fuente: Sudeaseg.

Las variables utilizadas en el modelo de la estimación de las futuras pérdidas agregadas:

Tabla VI: Variables utilizadas, modelo pérdidas agregadas

Variable	Descripción
Número de eventos en el año	Reclamos, siniestro total.
Monto facturado de cada reclamo	Volumen monetario en Bolívares.
Monto acumulado facturado en el año	Volumen acumulado de todos los eventos.

Fuente: Sudeaseg.

5. Coberturas

Horizontal: La investigación se sitúa en todo el territorio nacional de Venezuela.

Vertical: La investigación tiene como cobertura vertical la cantidad de ocurrencia de eventos (siniestros) cada mes, el monto facturado de cada siniestro estudiado y el monto total facturado mensual medido en Bolívares.

VI. METODOLOGÍA

La investigación es de tipo cualitativa documental, a continuación, se presenta la estructura del estudio comentado.

1. Estructura del estudio

- a) Se describe la forma de realizar agrupaciones por el aprendizaje estadístico no supervisado a través de k medias para visualizar la forma de los grupos; se construye agrupaciones con el análisis de componentes principales para definir la cantidad optimizada de grupos y con carácter confirmatorio: se aplica el análisis discriminante que adicionalmente permite determinar la importancia de los distintos factores de riesgo; se utiliza las técnicas de árbol y bosque aleatorio; se desarrolla grupos, con los datos de forma longitudinal, por medio del análisis de componentes principales y una variación del k medias con una distancia propuesta por Chiodi para series cortas.
- b) Se describe la forma de utilizar las matemáticas aplicadas, a través de la teoría de riesgo en el área del cálculo de la máxima pérdida probable, para encontrar una función de distribución probabilística particular, que describa el total de la pérdida obtenida por medio del método “*monte carlo*” y dos modelos probabilísticos: de ocurrencia de los eventos causantes del siniestro, y de severidad que es el valor monetario del siniestro individual asociado a cada evento anteriormente

indicado. Simultáneamente, se incorpora la teoría de valores extremos para optimizar el cálculo de la estimación futura de las pérdidas.

2. Tipo y diseño del presente artículo.

El presente estudio es de carácter cualitativo de una investigación de carácter comparativo ya que “se refiere a cuáles son las diferencias o semejanzas que existen entre dos o más contextos, seres, o unidades de estudio”, según Hurtado. La investigación es efectivamente documental de una investigación comparativa, debido a que describe la búsqueda de las similitudes y diferencias de varias técnicas de agrupación. La investigación es de tipo cualitativa y no experimental de una investigación cuantitativa no experimental, ya que describe un estudio cuantitativo que hace uso de datos procedentes de históricos reales, sin manipulación por parte del investigador, de acuerdo a Sampieri. El diseño es de tipo mixto, ya que combina el uso de información de corte longitudinal al investigar los cambios en el comportamiento de las Marcas Modelos de Vehículos en estudio a lo largo de períodos sucesivos en el tiempo, con información obtenida a partir de un corte transversal al considerar la ocurrencia de eventos de carácter ambiental, accidentes por colisión, volcamiento o accidente de tránsito.

3. Métodos utilizados

Recopilación de la información requerida para la investigación, colocación estructurada de forma de razonamiento lógico proposicional y ordenación por métodos de estudio univariante hasta los métodos multivariantes descrito a continuación.

VII. DESARROLLO DE LA RESEÑA DOCUMENTAL

1. Método estadístico que fundamenta la estrategia a utilizar

en una investigación cuantitativa

a) Análisis de componentes principales

El análisis de componentes principales se utiliza para los datos transversales, según Yépez¹³ la técnica de múltiples variables conocida como análisis de componentes principales (ACP), que puede ser presentada según el enfoque geométrico de K. Pearson¹⁴ citado por Vásquez y Ramírez¹⁵, tiene por objeto la búsqueda de planos de representación óptima de una nube de puntos, según el criterio de los mínimos cuadrados. La técnica genera nuevas variables denominadas componentes principales, las cuales retienen lo esencial de la información correspondiente. El ACP al ser aplicado sobre datos transversales definidos por los vehículos asegurados, se utilizará para describir los patrones de grupos de vehículos asegurados. Esta información se organiza sobre una matriz de datos centrada X_c, X_c' , cuyas filas se corresponden con los vehículos expuestos al riesgo, y cuyas columnas describen cada una de las variables en estudio como, por ejemplo: marca modelo, monto de siniestros facturado, valor del vehículo. En ese sentido, según Vásquez y Ramírez¹⁶, al aplicar el ACP sobre la matriz X_c, X_c', X_c, X_c' y considerar las coordenadas de las variables proyectadas sobre el α -ésimo eje, se obtiene la expresión:

13 C. Yépez. *Construcción de curvas principales de producción de petróleo, bajo incertidumbre y riesgo de disminución, en campos productores, PDVSA, división oriente, lapso 2003-2008, op.cit.*

14 K. Pearson. *On lines and planes of closest fit to systems of points in space.* Philosophical magazine, <http://pbil.univ-lyon1.fr/R/pearson1901.pdf>, London, 1901, p. 559-572.

15 M. L. Vásquez de R. and G. J. Ramírez. *Contribuciones al análisis de segmentación, op. cit.*

16 *Ídem.*

$$\varphi_{p \times 1}^{\alpha} = X_{p \times n}^T U_{n \times 1}^{\alpha} = \sum_{i=1}^n u_{i\alpha} x_i \quad (III.1)$$

Siendo:

$U_{n \times 1}^{\alpha}$ auto vector normalizado asociado con λ_{α} , el α -ésimo mayor auto valor de la matriz $X_c X_c^T X_c X_c^T$:

$$U_{n \times 1}^{\alpha} = \begin{pmatrix} u_{1\alpha} \\ u_{2\alpha} \\ \vdots \\ u_{i\alpha} \\ \vdots \\ u_{n\alpha} \end{pmatrix}_{n \times 1}$$

x_i es el vector que describe (ver tablas I, II y III) el valor de las variables del i -ésimo vehículo asegurado, en el momento del estudio:

$$x_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ij} \\ \vdots \\ x_{ip} \end{pmatrix}_{p \times 1}$$

Describiendo x_{ij} el valor asignado al vehículo asegurado i en la variable estudiada j . A partir de los resultados anteriores está claro que:

1. La importancia de la α -ésima componente puede medirse en términos de la variabilidad captada, mediante el siguiente cociente:

$$\frac{\lambda_{\alpha}}{\sum_{\alpha=1}^n \lambda_{\alpha}} * 100 \quad (III.2)$$

El cociente (III.2) mide la proporción de la variabilidad total de los vehículos expuestos al riesgo que es captada por φ^{α}

φ^α , la $\alpha\alpha$ -ésima componente principal.

2. $\varphi^\alpha \varphi^\alpha$ es una combinación lineal de los vectores X_1, X_2 , que describen los valores de cada uno de los vehículos expuestos al riesgo bajo estudio, en consecuencia, puede señalarse que:

$\varphi^\alpha \varphi^\alpha$ es la $\alpha\alpha$ -ésima componente principal de los valores de los vehículos asegurados.

3. La importancia del ii -ésimo vehículo en la formación de la $\alpha\alpha$ -ésima componente principal $\varphi^\alpha \varphi^\alpha$, puede obtenerse al evaluar el coeficiente $u_{i\alpha} u_{i\alpha}$ de manera que:

$u_{i\alpha}^2 u_{i\alpha}^2$ mide la contribución del ii -ésimo vehículo asegurado a la variabilidad por la $\alpha\alpha$ -ésima componente principal de los valores de los mismos.

4. La contribución de la $\alpha\alpha$ -ésima componente principal para explicar el patrón de entre las variables del ii -ésimo vehículo asegurado, se define en la forma:

$$\frac{\psi_{i\alpha}^2}{d^2(x_i, \bar{x})} \quad (III.3)$$

Siendo $\psi_{i\alpha} \psi_{i\alpha}$ la coordenada de proyección del ii -ésimo vehículo expuesto al riesgo sobre la $\alpha\alpha$ -ésima dirección de ajuste definida por la componente principal.

b) Agrupación k medias:

Según Cuesta¹⁷ la técnica de agrupación recortada “K medias” (tkmeans) tiene por objetivo obtener k grupos donde el promedio de las variables múltiples, llamados centroides, sean mínimos dentro de

17 J. Cuesta-Albertos, A. Gordaliza and C. Matrán. *Trimmed K-means: an attempt to robustify and quntizers*, op. cit.

cada grupo y por el contrario, el promedio de las variables múltiples, centroides, sean máximas de cada individuos con respectos a los otros individuos ubicados en los otros K-1 grupos. Aquí los individuos son los vehículos amparados en la cobertura de pérdida parcial del ramo de casco de automóvil de las empresas aseguradas en el estudio y se aplica la técnica de “K medias”.

c) Análisis discriminante con variables normalizadas

Según Díaz, Morales el análisis discriminante tienen dos objetivos, de una parte, está la separación o discriminación de grupos, y de otra, la predicción o asignación de un objeto, para este estudio sería los vehículos por asegurar, en uno de entre varios grupos previamente definidos, con base en los valores de las variables que lo identifican, por ejemplo: La Marca y Modelo del vehículo. Entre los autores que han desarrollado el análisis discriminante esta Welch (1937), citado por Rencher (1998), a partir del cual se obtiene algunas reglas de clasificación o discriminación según Díaz, Morales¹⁸.

d) Estudio longitudinal, conglomerado a través del análisis de componentes principales:

También Yépez¹⁹ aclara que el ACP al ser aplicado sobre datos longitudinales definidos por la pérdida monetaria de gastos por la cobertura de Pérdida Parcial en el ramo asegurador de Automóvil Casco, se utilizará para describir los patrones de la evolución de la pérdida acumulada en el período bajo estudio. Esta información se organiza sobre una matriz de datos centrada cuyas filas se corresponden con las marcas modelos de los vehículos amparados, y

18 L. G. Díaz y M. A. Morales. (2012). *Estadística multivariada: inferencia y métodos*. Universidad Nacional de Colombia, Bogotá, Colombia, 2012.

19 C. Yépez. *Construcción de curvas principales de producción de petróleo, bajo incertidumbre y riesgo de disminución, en campos productores, PDVSA, división oriente, lapso 2003-2008, op. cit., p. 49.*

cuyas columnas describen la pérdida monetaria acumulada mensual en bolívares a lo largo del período bajo estudio.

e) Estudio longitudinal, conglomerado a través de una modificación de “K medias” con una distancia propuesta por Chiodi:

Según Chiodi²⁰ cuando se examinan datos longitudinales de múltiples variables en serie de tiempo cortas, el conjunto de datos requiere una separación previa entre individuos según sólo la similitud de las formas de sus curvas de crecimiento con respecto a las varias variables, en esta investigación es la marca y modelo los individuos a través del monto acumulado de las pérdidas parciales de los meses estudiados, sigue Chiodi²¹ indicando que el problema puede ser crucial si los datos están constituidos por series temporales cortas.

2. Método matemático, modelo aleatorio de pérdidas agregadas

a) El riesgo de pérdida monetaria por siniestro a vehículos expuestos:

Existen diferentes enfoques para abordar la noción de riesgo, en este estudio se identifican solamente aspectos conceptuales que permiten situar con claridad el denominado riesgo a vehículos asegurados. El concepto de riesgo según Jorion²² es “el valor de la máxima pérdida esperada en un intervalo de tiempo determinado... fijado un nivel de confianza”. Señala así mismo el autor que “...el riesgo puede

20 M. Chiodi. *The clustering of longitudinal, Multivariate data when time series are short, op. cit.*, p. 445.

21 *Ídem.*

22 P. Jorion. *Valor en Riesgo: El nuevo paradigma para el control de riesgo con derivados*. Limusa Noriega editores, México, 2007.

ser definido en términos generales como la incertidumbre sobre los flujos futuros o resultados futuros”, lo que tiene interés básico en los enfoques del riesgo en las finanzas. Cabe señalar que en la presente investigación el riesgo se refiere al suceso futuro e incierto que no depende exclusivamente de la voluntad del tomador del contrato de seguros y cuya materialización da origen a la obligación de la empresa de seguros. Por consiguiente, lo que para Jorion significa incertidumbre en “flujos futuros”, equivale en esta investigación a incertidumbre en los “siniestros, reclamos o sucesos futuros”. Para Venegas²³, el riesgo se fundamenta en lo siguiente: “...Si “S” es una variable aleatoria definida en $(\Omega, \mathcal{F}, \mathbb{P}_\theta)(\Omega, \mathcal{F}, \mathbb{P}_\theta)$, donde θ es un vector de parámetros asociado con la distribución de “S”... el valor en riesgo $(VaR_{1-q}^S)(VaR_{1-q}^S)$ con un nivel de confianza del $(1 - q)(1 - q)$, se define como el valor en la distribución de esa variable aleatoria tal que, en un período de tiempo dado, $[t, T][t, T]$, para un intervalo de confianza del $(1 - q)100\%(1 - q)100\%$, la probabilidad:

$$\mathbb{P}_\theta\{VaR_{1-q}^S \leq S\} = 1 - q \text{ ó } \mathbb{P}_\theta\{S \leq VaR_{1-q}^S\} = q \quad (III. 4)$$

A partir de esta definición, se desprende, que:

$$VaR_{1-q}^S = -\text{infimo}\{S \in R | \mathbb{P}_\theta\{S > s\} \leq 1 - q\} \quad (III. 5)$$

Como se puede observar, la $VaR_{1-q}^S VaR_{1-q}^S$ es una estimación estadística del valor de “S” con cierto grado de confianza en un intervalo de tiempo dado”. Este autor enfoca el riesgo a nivel de Valor en Riesgo, VaR, concepto introducido en 1995 por J.P. Morgan, que ha sido publicado en un documento técnico que propone cuantificar el riesgo de mercado asociado a todas las posiciones de una entidad bancaria a través del cálculo de un único número. Es

23 F. Venegas M. *Riesgo financiero y económicos productos derivados y decisiones económicas bajo incertidumbre*. 2º. ed., Cengage Learning Editores, S.A., México, pp. 861-912,

importante resaltar que esa metodología está basada en el supuesto de normalidad de la variable “S”.

Según Jorion²⁴, el riesgo operacional se refiere a las pérdidas potenciales resultantes de utilización de sistemas de manejo, accidentes que afectan a los automóviles directamente, errores de terceras personas al colisionar sus vehículos sobre los automóviles asegurados, siempre y cuando se cumplan con las condiciones generales y particulares del contrato de seguros. Esto incluye riesgo de ejecución, que abarcan situaciones donde se falla en la ejecución del manejo del automóvil. También, el riesgo operacional, incluye los eventuales accidentes ocasionados por desastres naturales que no estén excluidos en el contrato de seguros.

3. Modelo aleatorio de pérdidas agregadas

Inicialmente, el abordaje del cálculo del riesgo se realizaba separadamente, es decir, privaba la teoría individual de estimación de riesgo para cada vehículo asegurado. Sin embargo, en la práctica para un efectivo cálculo del riesgo se hace necesario un abordaje que considere conjuntamente los distintos vehículos asegurados como un “colectivo”, que de acuerdo a Daykin²⁵ es el enfoque más apropiado. El modelo probabilístico de pérdidas agregadas, siendo un modelo econométrico de orden uno AR(1) (Auto Regresivo de orden 1), se define a través de la combinación de los siguientes dos modelos: En el contexto del proceso de siniestros anuales, se asumen “ m ” ocurrencias de eventos asociados (siniestros/reclamos) con uno o varias causas del siniestro, las mismas no son factores de riesgo. **Los factores de riesgos son las características del vehículo y del asegurado que aumenta o disminuye la probabilidad de**

24 P. Jorion. *Valor en Riesgo: El nuevo paradigma para el control de riesgo con derivados*, op. cit.

25 C. Daykin, T. Pentikäinen and M. Pesonen. *Practical risk theory for actuaries*, op. cit.

la ocurrencia futura de eventos cubiertos por el contrato de Seguros, las causas de siniestro son los eventos propiamente que materializa el riesgo, los cuales pueden ser el “volcamiento”, la “colisión” con otros vehículos u objetos fijos y cualquier causa que no esté especificada como exclusiones en el contrato de seguros:

$$E_1, E_2, \dots, E_n, \text{ con } n = 0, 1, 2, \dots \quad (\text{III.6})$$

Cada uno de estos eventos (siniestros/reclamos) induce una pérdida monetaria que se visualiza como pago del siniestro amparado, llamada severidad, descrita mediante otra variable “X”:

$$X_1, X_2, \dots, X_n, \text{ con } n = 0, 1, 2, \dots \quad (\text{III.7})$$

El valor monetario de la pérdida total en el año es obtenido con la siguiente forma:

$$S = X_1 + X_2 + \dots + X_n = \sum_{i=1}^n X_i, \text{ con } n = 0, 1, 2, \dots \quad (\text{III.8})$$

Se debe observar que en el comportamiento probabilístico del fenómeno subyacen las variables n, X y S :

1. n : Número de eventos (siniestros/reclamos) que ocurren en el intervalo de un año.
2. X : Valor monetario de la pérdida de cada una de las ocurrencias.
3. S : Valor monetario de la pérdida total en el año.

A continuación, se explican detalladamente los aspectos que permiten la estimación de los parámetros que caracteriza la distribución de las variables aleatorias N, X, S y el modelo de pérdidas agregadas.

4. Modelo aleatorio de ocurrencia de eventos (siniestros/reclamos)

Siguiendo la formulación de Daykin²⁶, puede considerarse que el

26
for actuaries, op. cit.

C. Daykin, T. Pentikäinen and M. Pesonen. *Practical risk theory*

número NN de eventos (siniestros/reclamos) que surge de un colectivo de riesgo, es una variable aleatoria discreta cuyo comportamiento se describe en términos de su distribución de probabilidades:

$$P_N = Prob(N = n), \text{ con } N = 0,1,2, \dots \quad (III.9)$$

En el cálculo colectivo del riesgo está compuesto por unidades estadísticas que en este estudio quedan descritas por los vehículos asegurados y están expuestos al riesgo de daño físico del automóvil llamado “casco del automóvil”, y los eventos (siniestros/reclamos) se refieren a los accidentes que puedan dar lugar a una reclamación y pago indemnizatorio al tomador o proveedor que repare los daños causados. La ocurrencia de los eventos (siniestros/reclamos) sigue una secuencia aleatoria y por lo tanto no es posible pronosticar el momento exacto de su ocurrencia, ni tampoco el número de eventos que ocurren anualmente. Sin embargo, si se puede establecer que cumplen con los preceptos del comportamiento estocástico de los procesos de Poisson: a) La ocurrencia de los eventos son independientes entre sí, b) El parámetro de ocurrencia depende del intervalo de tiempo dado, c) La probabilidad que dos eventos ocurra, al mismo tiempo, es cero. Sin embargo, ya que los eventos son independientes entre sí, entonces según Ramírez²⁷, la variable aleatoria: KK : Número de eventos (siniestros/reclamos) que ocurren anualmente. Sigue una distribución probabilística de “Binomial negativa”, con función de masa definida por:

$$P_N = Prob(N = r) = f(r) = \binom{N+r-1}{r} p^N (1-p)^r, \\ \text{con } N = 0,1,2, \dots \quad (III.10)$$

Donde:

p es la probabilidad continua de un solo suceso ($0 < p < 1$)
 $(0 < p < 1)$.

27 Vásquez de R. and G. J. Ramírez. *Contribuciones al análisis de segmentación*, op. cit.

El dominio es discreto entre $0 \leq r \leq +\infty$

Valor esperado:

$$E(r) = \frac{N(1-p)}{p} \quad (III.11)$$

Varianza:

$$Var(r) = \frac{N(1-p)}{p^2} \quad (III.12)$$

Relación del valor esperado y la varianza:

$$E(r) \geq Var(r) \quad (III.13)$$

Estimación de los parámetros NN y pp .

Los parámetros NN y pp , se estiman a partir de la distribución empírica observada del número de ocurrencias a lo largo de una muestra de “a” años, utilizando el método propuesto por Hogg²⁸, llamado el método de los momentos, también desarrollado de manera automática por el programa gratuito “EasyFit(r)”. Con las funciones III.11 y III.12 se expresa los parámetros “ NN ” y “ pp ”:

$$N = \frac{(E[r])^2}{Var[r] - E[r]}, \quad y \quad p = \frac{E[r]}{Var[r]} \quad (III.14)$$

Se deducen los estimadores consistentes de NN y pp reemplazando III.11 y III.12 por sus estimadores \bar{r} y $S^2\bar{r}$ y S^2 en estas expresiones. El inconveniente principal del método de los momentos es que los estimadores que resulta son en general bastante poco precisos, por lo cual se utiliza, después, un ajuste mayor con la función empírica, a través del comando “*Solver*” un algoritmo de programación lineal implementado en “Excel”. Se realiza, adicionalmente, tres (3) contrastes, entre ellas la Bondad de Ajuste, desarrollado de manera automática por el programa gratuito “*EasyFit(r)*”, para corroborar

28
233.

R. Hogg and S. Klugman. *Loss distributions*. Wiley, New York, 1984, pp. 217-

que la distribución empírica se comporta conforme a la función teórica “Binomial negativa”. Según se describe en Panjer²⁹, se utiliza para el contraste Chi-cuadrado (χ^2) el estadístico Q que asintóticamente sigue una distribución Chi-cuadrado con “ $J - 1$ ” grados de libertad:

$$Q = \sum_{j=1}^J \frac{(a_j - A_j)^2}{A_j} \tag{III.15}$$

Donde:

a : Es el número de años muestreados.

J : Es el máximo número de ocurrencias de la variable aleatoria r , observado en la muestra de “ aa ” años.

a_j : Es la frecuencia de las ocurrencias, medida por el número de años en la muestra en que se producen $j = 1, 2, \dots, J$ ocurrencias.

A_j : Es el número esperado de años en la muestra, en los que se producen j ocurrencias, que se obtiene en la forma $A_j = a * Prob(r = j), con j = 1, 2, \dots, J$.

5. Modelo aleatorio de severidad asociado a cada evento (siniestro/reclamo)

Sea “ NN ” la variable aleatoria cuyo comportamiento probabilístico describe el número de accidentes (eventos/siniestros/reclamos) ocurridos durante un año, que dan lugar al pago del monto total de

²⁹ H. Panjer. *Operational Risk Modeling Analytics*. John Wiley & Sons series in probability and statistics, Inc., Hoboken, New Jersey, 2006.

siniestros en el grupo de vehículos, y sea “ XX ” la variable aleatoria que mide la intensidad de la pérdida (monto pagado) debida a la ocurrencia del accidente (severidad asociada con la ocurrencia del evento). El problema que debe abordarse a estos efectos es el estudio de la distribución de probabilidad adecuada para la variable “ XX ”. Siguiendo a Daykin³⁰, en el presente estudio, y después de realizar numerosas pruebas y efectuar contrastes de “Bondad de ajuste” se conviene en proponer una distribución mixta, basada en el método de valores extremos:

Las distribuciones propuestas para el modelo mixto de severidad de un evento son: a) *LogNormal* tres parámetros y b) Generalizada de Valores Extremos tres parámetros.

La variable aleatoria “ XX ”, en el intervalo $(0, Percentil_{99,97})$ $(0, Percentil_{99,97})$, sigue la distribución *LogNormal* de tres parámetros:

$$X_i \sim \text{LogNorm}(\mu, \sigma, \gamma); \text{cont} = 1, 2, \dots, N \quad (\text{III. 16})$$

Parámetros:

σ : parámetro continuo ($\sigma > 0$) ($\sigma > 0$);

μ : parámetro continuo;

γ : parámetro continuo de ubicación ($\gamma = 0$ $\gamma = 0$ equivale a la distribución *LogNormal* de dos parámetros)

Dominio:

$$\gamma < x < +\infty \quad \gamma < x < +\infty.$$

Función de densidad de probabilidad:

$$f(x) = \frac{1}{\sigma(x-\gamma)\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln(x-\gamma)-\mu}{\sigma}\right)^2} \quad (\text{III. 17})$$

30 C. Daykin, T. Pentikäinen and M. Pesonen. *Practical risk theory for actuaries*, op. cit.

$$f(x) = \frac{1}{\sigma(x-\gamma)\sqrt{2\pi}} e^{(-\frac{1}{2}(\frac{\ln(x-\gamma)-\mu}{\sigma})^2)} \quad (III. 17)$$

Función de distribución acumulada:

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{\gamma}^x \frac{1}{(t-\gamma)} e^{(-\frac{1}{2}(\frac{\ln(t-\gamma)-\mu}{\sigma})^2)} dt \quad (III. 18)$$

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{\gamma}^x \frac{1}{(t-\gamma)} e^{(-\frac{1}{2}(\frac{\ln(t-\gamma)-\mu}{\sigma})^2)} dt \quad (III. 18)$$

La variable aleatoria “ X ”, en el intervalo (*Percentil*_{99,971}, *Percentil*₁₀₀) *Percentil*_{99,971}, *Percentil*₁₀₀), sigue la distribución Generalizada de Valores Extremos tres parámetros que corresponde a la generalización conjunta de las distribuciones de *Fréchet*, *Weibull* y *Gumbel*:

$$X_l \sim GenExtreV(\mu, \sigma, \kappa); \text{ con } l = 1, 2, \dots, N \quad (III. 19)$$

Parámetros:

σ : parámetro continuo de escala ($\sigma > 0$);

μ : parámetro continuo de ubicación;

κ : parámetro continuo de forma;

Dominio:

$$1 + \kappa \frac{(x-\mu)}{\sigma} > 0 \text{ para } \kappa \neq 0 \quad 1 + \kappa \frac{(x-\mu)}{\sigma} > 0 \text{ para } \kappa \neq 0;$$

$$-\infty < x < +\infty \text{ para } \kappa = 0 \quad -\infty < x < +\infty \text{ para } \kappa = 0;$$

Función de densidad de probabilidad:

$$f(x) = \begin{cases} \frac{1}{\sigma} e^{-(1+\kappa z)^{-\frac{1}{\kappa}}} (1 + \kappa z)^{(-1-\frac{1}{\kappa})} & \text{si } \kappa \neq 0 \\ \frac{1}{\sigma} e^{(-z - e^{-z})} & \text{si } \kappa = 0 \end{cases} \quad (III. 20)$$

Donde: $z = \frac{x-\mu}{\sigma}$; $z = \frac{x-\mu}{\sigma}$

Función de distribución acumulada:

$$F(x) = \begin{cases} \frac{1}{\sigma} \int_{\mu - \frac{x}{\kappa\sigma}}^x e^{-(1+\kappa(\frac{t-\mu}{\sigma}))^{-\frac{1}{\kappa}}} (1 + \kappa(\frac{t-\mu}{\sigma}))^{(-1-\frac{1}{\kappa})} dt & \text{si } \kappa \neq 0 \\ \frac{1}{\sigma} \int_{-\infty}^x e^{-(\frac{t-\mu}{\sigma})} - e^{-(\frac{t-\mu}{\sigma})} dt & \text{si } \kappa = 0 \end{cases}$$

(III. 21)

Siguiendo a Hogg³¹, los parámetros se estiman por el método de los momentos o de máxima verosimilitud, de acuerdo al caso, adicionalmente se utiliza el método de Newton, algoritmo numérico implementado en el programa computacional “Excel”, en la opción denominada “Solver” un comando de programación lineal de optimización. Adicionalmente, se utiliza el programa “EasyFit(r)” para confirmar los parámetros y efectuar el contraste de “Bondad de ajuste” para corroborar que la distribución empírica se distribuye conforme a la función teórica propuesta.

6. Modelo aleatorio conjunto de pérdidas agregadas que une los dos modelos anteriores

La situación que se plantea seguidamente es hallar el modelo probabilístico adecuado para describir la pérdida anual agregada:

$$S = X_1 + X_2 + \dots + X_N = \sum_{i=1}^N X_i, \text{ con } N = 0,1,2, \dots \text{ (III. 22)}$$

Se supone que montos de siniestros, X_i, X_j , causado por la ocurrencia del i -ésimo evento es independiente de la variable aleatoria NN , y que además se distribuye de acuerdo al modelo mixto antes propuesto. Entonces, el modelo probabilístico de pérdidas agregadas debido a

31 R. Hogg and S. Klugman. *Loss distributions, op. cit.*, pp. 217-233.

la ocurrencia de NN eventos, queda descrito por la distribución de variable aleatoria compuesta SS que describe el monto acumulado indemnizado en siniestros de la pérdida anual, correspondiente al seguro de casco de automóviles, en la cobertura de pérdida parcial. Distribución de la variable compuesta: Variable

$$S = \sum_{i=1}^N X_i, \text{ con } N = 0, 1, 2, \dots \quad (III. 23)$$

Con más detalle se explica la ocurrencia $\{S \leq s\} \{N = n\}$

Cuando $n = 0$ y $0 \leq s$;

Cuando $n = 1$ y $x_1 \leq s$;

Cuando $n = 2$ y $x_1 + x_2 \leq s$;

Cuando $n = 3$ y $x_1 + x_2 + x_3 \leq s$; (III. 24)

Función de Distribución Acumulada

$$F_S(s) = \sum_{n=0}^{\infty} p_n \text{Prob}(S \leq s | N = n) = \sum_{n=0}^{\infty} p_n F_S^{*n}(s) \quad (III. 25)$$

Siendo $F_S^{*n}(s) = \text{Prob}(\sum_{i=1}^n x_i \leq s | N = n)$

También se expresa en forma recurrente, la función de convolución:

$$F_S^{*n}(s) = F_S^* F_S^{*(n-1)}(s), \text{ con } n > 0.$$

Se detalla algunas convoluciones:

$$\begin{aligned}
 F_S^{*(n)}(s) &= \int_0^s F_S^{*(n-1)}(s-y)f_S(y)dy \\
 F_S^{*(n-1)}(s) &= \int_0^s F_S^{*(n-2)}(s-y)f_S(y)dy \\
 &\vdots \\
 F_S^{*(n-(n-2))}(s) &= \int_0^s F_S^{*(n-(n-1))}(s-y)f_S(y)dy \\
 F_S^{*(n-(n-1))}(s) &= \int_0^s F_S^{*(n-n)}(s-y)f_S(y)dy \\
 F_S^{*(n-n)}(s) &= F_S^{*(0)}(s) = 1, \quad \text{siempre y cuando } s \geq 0.
 \end{aligned}
 \tag{III. 26}$$

Para el presente estudio la variable en ningún momento será menor a cero. Daykin³², autor cuyas referencias han signado estos desarrollos, junto con las de Yépez³³, propone finalmente proceder a las estimaciones del modelo de pérdida agregada utilizando el método de simulación “*Monte Carlo*”. Así, al simular la distribución completa de “S”, se selecciona el valor del percentil 99,99; el cual determina el nivel de confianza del 99,99%, para la pérdida máxima probable esperada. Se resalta que este monto máximo es menor a la máxima pérdida posible que corresponde a la situación que todos los vehículos asegurados tengan uno o más siniestros, que tiene una probabilidad de cero, y, además, que el monto de la pérdida sea la máxima asegurada, lo cual también tiene una probabilidad que tiende a cero, de lo contrario el negocio del seguro no es viable.

32 C. Daykin, T. Pentikäinen and M. Pesonen. *Practical risk theory for actuaries*. op.cit.

33 C. Yépez. (2009). *Construcción de curvas principales de producción de petróleo, bajo incertidumbre y riesgo de disminución, en campos productores, PDVSA, división oriente, lapso 2003-2008*, op. cit.

Con la información descrita anteriormente Jorion³⁴ propone que la empresa de seguros pueda realizar las tres actividades siguientes:

- 1) Informando la máxima pérdida probable (percentil 99,99 de la función agregada de pérdida, el cual determina el nivel de confianza del 99,99%) a niveles directivos para la adecuada toma de decisión, ayudando a la alerta temprana para la mejor mitigación del riesgo con medidas a tiempo.
- 2) Permitiendo y facilitando la asignación de recursos en forma objetiva modificando el reglamento de cálculos de primas para suplir la desviación prevista en el modelo probabilísticos.
- 3) Evaluación del desempeño operacional, a través del índice de eficiencia que se construye con la desviación típica de las pérdidas (siniestros) observadas entre la desviación típica del Modelo de Pérdida Agregada.

VIII. METODOLOGÍA MATEMÁTICA EN LA CIENCIA DE LOS DATOS

1. Estrategia de análisis

Se realizará las siguientes fases:

Antes de aplicar los modelos matemáticos, la ciencia de los datos permite optimizar dicha información de forma integral, regularmente los investigadores tardan más del 80% del tiempo de la investigación en esta fase. Al tratarse de información real, los registros presentan una gran cantidad de incongruencias, tales como: datos omitidos, incorrectos, parcialmente modificados o no utilizable por depender de información deficientes; para solventar lo anterior se realiza los siguientes procesos:

*Se identifica los datos omitidos, de ser significativo, la variable (factor de riesgo) que representa dicha información es excluida del

34 C. Daykin, T. Pentikäinen and M. Pesonen. (1994). *Practical risk theory for actuaries*, op. cit.

estudio.

*Se determinan los datos incorrectos, de ser significativo, se procede de la misma forma que el ítem anterior.

En otras ocasiones los datos están completos y no se aprecia incongruencias aparentes; pero no aportan información, ya que la forma de registrarlos es incorrecta o no están alineados a las normas y políticas de la empresa aseguradora, se procede a la:

*Revisión de los datos de forma univariante a través de histogramas de frecuencia, de no aportar información el factor de riesgo (la variable) es excluida de la investigación.

*Revisión de los datos de forma univariante a través de gráficos de dispersión comparándolos con indicadores construidos especialmente para dicha revisión, de no aportar información se procede igual que en el ítem anterior.

A grandes rasgos se persigue lo siguiente:

*Se construye las agrupaciones para el estudio transversal.

*Igualmente se aplican métodos de agrupamiento para el estudio longitudinal.

*Se estima la máxima pérdida probable para un determinado grupo.

IX. DESARROLLO RESUMEN GENÉRICO

Se realiza una descripción de los resultados que se obtienen, detallados en otra entrega posterior, en la investigación, de forma resumida para que el lector tenga una visión clara de los resultados. El objetivo de la investigación es reseñar documentalmente el uso de las técnicas del aprendizaje estadístico. Se omite, por el momento, todo resultado numérico y gráfico.

X. CONCLUSIONES Y RECOMENDACIONES

Al haber comparado diferentes formas de conglomerar, permitió obtener parecidas agrupaciones; identificando los patrones que las

componen, apreciando la técnica k medias, el análisis de componentes principales, el análisis discriminante, las técnicas de particiones, las curvas principales y la k media con datos longitudinales a través de la distancia propuesta por Chiodi; lo cual se aprecia la importancia de aplicar todas las técnicas, que llevan a un resultado similar, para comprender mejor el agrupamiento de los registros pertenecientes a la cobertura de pérdida parcial en el ramo de contratos de seguros de casco del automóvil. Se puede demostrar las variables concretas que arrojan mayor información.

Se demostró que, realizando el cálculo de los Modelos de Pérdidas Agregadas, Modelo aleatorio de ocurrencia de eventos (siniestros/reclamos), Modelo aleatorio de severidad asociado a cada evento (siniestro/reclamo), y con el Modelo aleatorio conjunto de pérdidas agregadas que une los dos modelos anteriores, en un trabajo de campo desde el punto de vista matemático se demuestra en el modelo de ocurrencia, los parámetros de la tasa media de ocurrencia en el período previamente establecido de estudio la cantidad de eventos esperados anual y la máxima ocurrencia probable de eventos con un nivel de confianza de 98,9%. En el modelo de severidad de los eventos, se puede certificar como función probabilística a la lognormal de tres (3) parámetros (μ, σ, ν) en el rango de los percentiles (0 a 99,97) y la función probabilística valor extremo general (modelo “Fréchet” y “Weibull”) tres (3) parámetros (α, β, γ) en el rango de los percentiles (99,98 a 99,990) en el restante se puede expresar la tendencia a cero (0) la función de densidad y a uno (1) la función acumulada, tal como en la primera función. Lo que significa una tasa media de pérdida por cada evento expresada en Bolívares y una máxima pérdida probable por evento igualmente expresada en Bolívares con un nivel de confianza del 99,995%. Finalmente, se puede describir el modelo de pérdidas agregadas donde se certifica el “valor a riesgo” para la máxima pérdida probable expresada en Bolívares con un nivel de confianza del 99,8%.

XI. REFERENCIAS BIBLIOGRÁFICAS

- H. Panjer. *Operational Risk Modeling Analytics*. John Wiley & Sons series in probability and statistics, Inc., Hoboken, New Jersey. (2006).
- M. L. Vásquez de R. and G. J. Ramírez. *Contribuciones al análisis de segmentación*. Facultad de ciencias económicas y sociales- U.C.V., Oficina central de estadística e informática, Caracas. (1999).
- E. Rodríguez T. *Administración del riesgo: Utilización de análisis de datos multivariantes para el estudio integral del riesgo*. s.n., Bogotá. (2004).
- P. Jorion. *Valor en Riesgo: El nuevo paradigma para el control de riesgo con derivados*. Limusa Noriega editores, México. (2007).
- F. Venegas M. *Riesgo financiero y económicos productos derivados y decisiones económicas bajo incertidumbre*. 2°. ed., Cengage Learning Editores, S.A. , México. pp. 861-912. (2008).
- R. Hogg and S. Klugman. *Loss distributions*. Wiley, New York, pp. 217-233. (1984).
- C. Daykin, T. Pentikäinen and M. Pesonen. *Practical risk theory for actuaries*. Chapman & Hall, London. (1994).
- A. Charpentier. *Computational actuarial science with R*. CRC Press Taylor & Francis Group, LLC, Boca Raton, London, New York. (2015).
- R. Kass, M. Goovaerts, J. Dhaene and M. Denuit. *Modern actuarial risk theory using R*. Springer-Verlag, Berlin, Heidelberg. (2008).
- Bower, Gerber, Hickman, Jones and Nesbitt. *Actuarial Mathematics*. The Society of actuaries, Schaumburg, Illinois. (1994).
- C. Chirinos. *Desarrollo de tarifa de automóvil casco*. Trabajo de pasantía para optar a la licenciatura de ciencias actuariales no publicado, universidad central de Venezuela en la escuela de estadísticas y ciencias actuariales de la facultad de ciencias económicas y sociales, Caracas. (2001).
- A. López. *Resultado técnico de automóvil individual*. Trabajo

de pasantía para optar a la licenciatura de ciencias actuariales no publicado, universidad central de Venezuela en la escuela de estadísticas y ciencias actuariales de la facultad de ciencias económicas y sociales, Caracas. (2001).

T. Thonon. *Análisis y utilización de la estadística para la realización de la tarifa de automóvil*. Trabajo de pasantía para optar a la licenciatura de ciencias actuariales no publicado, universidad central de Venezuela en la escuela de estadísticas y ciencias actuariales de la facultad de ciencias económicas y sociales, Caracas. (2002).

C. Salazar. *Caracterización de deudores de tarjetas de crédito de una institución financiera a través de técnicas de aprendizaje estadístico*. Trabajo de especial de grado, no publicado. Universidad central de Venezuela en el postgrado de modelos aleatorios de la facultad de ciencias, Caracas. (2012).

G. Ramírez. *Teoría de la Probabilidad*. Trabajo de ascenso, no publicado. Universidad central de Venezuela, Caracas. (1999).

J. Días. *Sistema de información en la superintendencia de la actividad aseguradora*. Trabajo de pasantía para optar a la licenciatura de ciencias actuariales no publicado, universidad central de Venezuela en la escuela de estadísticas y ciencias actuariales de la facultad de ciencias económicas y sociales, Caracas. (2014).

M. Chiodi. *The clustering of longitudinal, Multivariate data when time series are short*. Elsevier science publishers B.V., North-Holland. (1987).

J. Cuesta-Albertos, A. Gordaliza and C. Matrán. *Trimmed K-means: an attempt to robustify and quantizers*. s.l. Anuario de estadísticas. Vol. 25 (2), México, pp. 553-576. (1997).

C. Yépez. *Construcción de curvas principales de producción de petróleo, bajo incertidumbre y riesgo de disminución, en campos productores, PDVSA, división oriente, lapso 2003-2008*. Trabajo de grado para optar al título de especialización en estadística, no publicado. Universidad Central de Venezuela en la comisión de

estudios para graduados área de postgrado programa integrado de postgrado en estadística FACES, Caracas. (2009).

C. Yépez. *Análisis comparativo y aplicación de modelos de representación de factores de riesgo en el ramo de automóvil casco del mercado asegurador venezolano lapso 2015*. Trabajo de grado para optar al título de especialización en estadística, no publicado. Universidad Central de Venezuela en la comisión de estudios para graduados área de postgrado programa integrado de postgrado en estadística FACES, Caracas. (2018).

J. Hurtado de B. *Metodología de la Investigación Holística*, Instituto universitario de tecnología “José Antonio Anzoátegui” Tercera Edición-Fundación Sypal, Caracas. (2000). p. 223.

R. Sampieri, C. Fernández y P. Baptista. *Metodología de la investigación*. 4º. ed, Mc Graw Hill, México. (2006).

K. Pearson *On lines and planes of closest fit to systems of points in space*. Philosophical magazine, . London. (1901). 2: 559-572. <http://pbil.univ-lyon1.fr/R/pearson1901.pdf>.

M.L. Vásquez de R. y G.J. Ramírez. *Análisis de Componentes Principales*. Material de Apoyo. Post Grado Especialización de Estadística en la Comisión de Estudios para Graduados, Área de Post Grado en Estadística y Actuariado. Programa Integrado de Post Grado en Estadística. Facultad de Ciencias Económicas y Sociales. Universidad Central de Venezuela, Caracas. (2007).

L. G. Díaz y M. A. Morales. *Estadística multivariada: inferencia y métodos*. Universidad Nacional de Colombia, Bogotá, Colombia. (2012).