



UNIVERSIDAD CENTRAL DE VENEZUELA
FACULTAD DE CIENCIAS
ESCUELA DE MATEMÁTICAS
POSTGRADO EN MODELOS ALEATORIOS



ANÁLISIS COMPARATIVO Y APLICACIÓN DE
MODELOS DE REPRESENTACIÓN DE
FACTORES DE RIESGO EN EL RAMO DE
AUTOMÓVIL CASCO DEL MERCADO
ASEGURADOR VENEZOLANO LAPSO 2015

Magister Scientiarum en Modelos Aleatorios

AUTOR:

Especialista. Carlos Javier Yépez Ballesteros

TUTOR:

Doctor. Ricardo Rios

Caracas, 2018



UNIVERSIDAD CENTRAL DE VENEZUELA
FACULTAD DE CIENCIAS
COMISIÓN DE ESTUDIOS DE POSTGRADO



Comisión de Estudios
de Postgrado

VEREDICTO

Quienes suscriben, miembros del jurado designado por el Consejo de la Facultad de Ciencias de la Universidad Central de Venezuela, para examinar el Trabajo de Grado presentado por: Carlos J. Yépez Ballesteros, Cédula de Identidad 10.811.059, bajo el título "ANÁLISIS COMPARATIVO Y APLICACIÓN DE MODELOS DE REPRESENTACIÓN DE FACTORES DE RIESGO EN EL RAMO DE AUTOMÓVIL CASCO DEL MERCADO ASEGURADOR VENEZOLANO LAPSO 2015", a fin de cumplir con el requisito legal para optar al grado académico de **MAGÍSTER SCIENTIARUM, MENCIÓN MODELOS ALEATORIOS**, dejan constancia de lo siguiente:

1.- Leído como fue dicho trabajo por cada uno de los miembros del jurado, se fijó el día 30 de abril de 2018 a las 11:00 AM., para que el autor lo defendiera en forma pública, lo que éste hizo en LAPROESA, mediante un resumen oral de su contenido, luego de lo cual respondió satisfactoriamente a las preguntas que le fueron formuladas por el jurado, todo ello conforme con lo dispuesto en el Reglamento de Estudios de Postgrado.

2.- Finalizada la defensa del trabajo, el jurado decidió **Aprobar** por considerar, sin hacerse solidario con la ideas expuestas por el autor, que se ajusta a lo dispuesto y exigido en el Reglamento de Estudios de Postgrado.

Para dar este veredicto, el jurado estimó que el trabajo proporciona información de calidad al mercado asegurador Venezolano que sirva de referencia para la planificación incorporando valores a riesgo y generación de controles de gestión.

En fe de lo cual se levanta la presente ACTA, a los 30 días del mes de abril del año 2018, conforme a lo dispuesto en el Reglamento de Estudios de Postgrado, actuó como Coordinador del jurado Ricardo Ríos.



Figura 1: Veredicto1 - Primera parte


Angie Pineda

Nombre Apellido / C.I. 15.368.508
Institución (UCV)


Elvia Flores

Nombre Apellido / C.I. 4.128.598
Institución (UCV)


Ricardo Rios

Nombre Apellido / C.I. 3.949.476
Institución (UCV)
Tutor



Figura 2: Veredicto2 - Segunda parte

Resumen

Se estudió el 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 compararon dos modelos matemáticos de representación, a través de K medias y el análisis de componentes principales, incorporando los modelos de agrupamiento, se describieron las particularidades de los modelos, las similitudes y sus diferencias. En el estudio se detectaron las características relevantes entre los modelos del aprendizaje estadístico en los datos con múltiples variables; Las variables monto facturado del siniestro, valor del vehículo y la marca modelo son las que arrojaron mayor información. Se utilizaron los reclamos para incorporarlos en el modelo aleatorio de pérdida agregada bajo incertidumbre, y se estimó la máxima pérdida probable de los reclamos con la información de las aseguradoras dadas del mercado asegurador venezolano. Se estimó 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.

Dedico este trabajo de grado:

A mis padres Gladys y Roger.

A mis hijas Carla y Gabriela.

A mi nieto Santiago Moises.

*A mi esposa Carmen Teresa García de Yépez, que con gran constancia comprensión,
ha sido pilar fundamental del núcleo familiar.*

A mis hermanos: Bernardo, Jesús, Roger, Roger David y Eduardo.

*A mis seres queridos que ya no están con nosotros: mi hermana María Antonia
Yépez, mi hermano José Ángel Yépez, mi tía Gisela, mis abuelas y abuelos.*

Reconocimientos

Agradezco al tutor Ricardo Rios por su constancia y dedicación, a los profesores Mairene Colina y Benito Hernandez por el apoyo y consideración. De igualmanera agradezco a la Secretaria Carlota, a los Profesores Ángel Colmenares, Mayarín López, Jorge Días, Daniel Barraez, Raul Barrera, a los estudiantes Johanna, José Muñoz, Jorge Vinseiro, que de forma particular colaboraron en la dar oportunas observaciones al presente trabajo.

Índice general

Índice de figuras	xvii
Índice de tablas	xxvii
1. Introducción	1
2. Planteamiento, objetivos, importancia antecedentes y materiales	5
2.1. Planteamiento del problema	5
2.1.1. Problema	5
2.1.2. Interrogantes de la investigación	7
2.2. Objetivo general	7
2.3. Objetivos específicos	8
2.4. Importancia de la investigación y la justificación	8
2.5. Antecedentes	9
2.6. Materiales	11
2.6.1. Fuentes de información	11
2.6.2. Universo	12

ÍNDICE GENERAL

2.6.3. Unidades estadísticas	12
2.6.4. Variables	12
2.6.5. Coberturas	15
3. Marco teórico	17
3.1. Aspectos metodológicos	17
3.1.1. Tipo y diseño de la investigación	18
3.1.2. Métodos utilizados	18
Método estadístico que fundamenta la estrategia utilizada	18
Método matemático, modelo aleatorio de pérdidas agregadas	21
3.2. Metodología matemática en la ciencia de los datos	28
3.2.1. Estrategia de análisis	28
4. Análisis e interpretación de resultados	31
4.1. Conglomerado a través del análisis de componentes principales	58
4.2. Conglomerado a través de k medias	60
4.3. Conglomerado a través del análisis discriminante de variables normalizadas	68
4.4. Estudio longitudinal, conglomerado a través del análisis de componentes principales	69
4.5. Estudio longitudinal, conglomerado a través de una variación de K me- dias con distancia propuesta por Chiodi	70
4.6. Determinación de la estimación de las pérdidas agregadas	71
4.6.1. Sin aplicar factor de deflactación estimado a los montos de sinies- tros individuales	71
4.6.2. Aplicando factor de deflactación, estimado, a los montos de si- niestros individuales	73

5. Conclusiones y recomendaciones	79
5.1. Conclusiones	79
5.2. Recomendaciones	80
A. Paquetes utilizados	83
A.1. Paquetes estadísticos	83
A.2. Paquetes para la transcripción	83
B. Selección de datos obtenidos	85
B.1. Filtrado de los datos	85
C. Estudio por cada variable	87
C.1. Estudio descriptivo transversal	87
C.1.1. Presentación de histogramas	87
C.1.2. Presentación de correlación de las variables	103
C.1.3. Presentación de gráfico de dispersión	104
D. Conglomerado a través de k medias primera opción	115
D.1. Conglomerado a través de K medias, con variables originales	115
E. Conglomerado a través de k medias con variables numéricas normalizadas y variables categóricas transformadas en ficticias	123
E.1. K medias, con variables numéricas normalizadas	123
F. Conglomerado a través de k medias con variables normalizadas y variables categóricas transformadas en ficticias normalizadas	131

F.1. K medias, variables numéricas normalizadas y las categóricas transformadas en ficticias normalizadas	131
G. Conglomerado a través de bagging de weka 3.7	139
G.1. Proceso	139
H. Conglomerado a través de bosque aleatorio	141
H.1. Proceso	141
I. Conglomerado a través del análisis de componentes principales	143
I.1. Análisis de componentes principales	143
J. Conglomerado a través k medias con la variable marca modelo	147
K. Conglomerado a través del análisis discriminante, apoyado en k medias	155
L. Estudio longitudinal, conglomerado a través del análisis de componentes principales	163
L.1. Primer proceso: asignación revisión de datos	163
L.2. Segundo proceso: resultados	165
M. Estudio longitudinal, conglomerado a través de k medias, con k=2 medias y la distancia propuesta por Chiodi	167
M.1. Primer proceso: asignación para k=2, distancia euclídeana	167
M.2. Segundo proceso de asignación para k=2, distancia Chiodi	170
M.3. Tercer proceso: asignación para k=2, confirmación distancia de Chiodi .	172
N. Determinación de la estimación de las pérdidas agregadas	175

N.1. Asociar la función de ocurrencia	176
N.2. Asociar la función de monto siniestro individual	181
Ñ. Vocabulario	197
Ñ.1. Tema asegurador	197
Ñ.2. Tema Siglas	198
Bibliografía	201

Índice de figuras

1.	Veredicto1	II
2.	Veredicto2	III
4.1.	Histograma número de vehículos asegurados por año de construcción . .	32
4.2.	Histograma número de vehículos por valor del vehículo expresado en millones de Bolívares	33
4.3.	Histograma número de vehículos por el monto total de siniestro del vehículo	34
4.4.	Histograma número de vehículos por el número total de siniestro del vehículo	35
4.5.	Histograma número de vehículos por cada marca y modelo asegurado . .	36
4.6.	Histograma número de vehículos por cada número de ocupantes de los vehículos	37
4.7.	Histograma número de vehículos por cada estado civil observado	39
4.8.	Histograma número de vehículos por cada sexo observado	40
4.9.	Histograma número de vehículos por cada modalidad observada	41
4.10.	Histograma número de vehículos por cada tipo de contratante observado	42
4.11.	Histograma número de vehículos por cada tipo de vehículo observado . .	43
4.12.	Histograma número de vehículos por densidad observada	44

ÍNDICE DE FIGURAS

4.13. Histograma número de vehículos por el número de renovaciones observadas	45
4.14. Histograma número de vehículos por cada edad observada	46
4.15. Histograma número de vehículos por la exposición al riesgo observada .	47
4.16. Dispersión tasa de riesgo, monto total facturado, número de siniestro y tasa de uso por cada marca modelo	49
4.17. Dispersión tasa de riesgo, monto total facturado, número de siniestro y tasa de uso por cada valor del vehículo	50
4.18. Dispersión tasa de riesgo, monto total facturado, número de siniestro y tasa de uso por cada estado civil	51
4.19. Dispersión tasa de riesgo, monto total facturado, número de siniestro y tasa de uso por cada sexo	52
4.20. Dispersión tasa de riesgo, monto total facturado, número de siniestro y tasa de uso por cada modalidad	53
4.21. Dispersión tasa de riesgo, monto total facturado, número de siniestro y tasa de uso por cada lapso de tiempo asegurado del vehículo	54
4.22. Dispersión tasa de riesgo, monto total facturado, número de siniestro y tasa de uso por cada edad del conductor del vehículo asegurado	55
4.23. Dispersión tasa de riesgo, monto total facturado, número de siniestro y tasa de uso por cada tiempo de exposición del vehículo	56
4.24. Dispersión tasa de riesgo, monto total facturado, número de siniestro y tasa de uso por cada año de construcción del vehículo	57
4.25. Covarianza de las principales variables	58
4.26. Autovalores del análisis de componentes principales	59
4.27. Conglomerado a través de k medias, con k=7, todas las variables origi- nales sin modificación	60
4.28. Conglomerado a través de k medias, con K=7 medias, variables núme- ricas normalizadas, variables categóricas transformadas en ficticias sin excluir variables	63

4.29. Conglomerado a través de k medias, con K=7 medias, variables numéricas normalizadas, variables categóricas transformadas en ficticias normalizadas sin excluir variables	64
4.30. Conglomerado a través de bagging de weka 3.7	65
4.31. Conglomerado a través del aprendizaje del bosque aleatorio	66
4.32. Conglomerado a través de k medias, con k=7 con solo la marca modelo observando el número de siniestro y el monto facturado total	67
4.33. Conglomerado a través del análisis discriminante, con K=2 solo a la variables marca modelo	68
4.34. Estudio longitudinal, conglomerado a través del análisis de componentes principales, gráfico de las curvas de las marcas modelos por monto total facturado acumulado y agrupado	69
4.35. Estudio longitudinal, conglomerado a través de k medias con distancia de Chiodi, muestra de dos grupos, tercera asignación	70
4.36. Estimación de las pérdidas agregadas, máxima pérdida probable, función binomial negativa compuesta lognormal mixta wakeby	72
4.37. Estimación de las pérdidas agregadas, máxima pérdida probable, función de probabilidad acumulada binomial negativa compuesta lognormal mixta valor extremo general, valor en riesgo	76
4.38. Estimación de las pérdidas agregadas, máxima pérdida probable, función de probabilidad de densidad binomial negativa compuesta lognormal mixta valor extremo general, valor en riesgo	77
C.1. Histograma número de vehículos por año de construcción	88
C.2. Histograma número de vehículos por valor del vehículo	89
C.3. Histograma número de vehículos por el monto total de siniestro del vehículo	90
C.4. Histograma número de vehículos por el número total de siniestro del vehículo	91
C.5. Histograma número de vehículos por cada marca y modelo asegurado	92

ÍNDICE DE FIGURAS

C.6. Histograma número de vehículos por cada número de ocupantes de los vehículos	93
C.7. Histograma número de vehículos por cada estado civil observado	95
C.8. Histograma número de vehículos por cada sexo observado	96
C.9. Histograma número de vehículos por cada modalidad observada	97
C.10. Histograma número de vehículos por cada tipo de contratante observado	98
C.11. Histograma número de vehículos por cada tipo de vehículo observado	99
C.12. Histograma número de vehículos por densidad poblacional observada	100
C.13. Histograma número de vehículos por el número de renovaciones observadas	101
C.14. Histograma número de vehículos por cada edad observada	102
C.15. Histograma número de vehículos por la exposición al riesgo observada	103
C.16. Dispersión tasa de riesgo, monto total facturado, número de siniestro y tasa de uso por cada marca modelo	105
C.17. Dispersión tasa de riesgo, monto total facturado, número de siniestro y tasa de uso por cada valor de vehículo	106
C.18. Dispersión tasa de riesgo, monto total facturado, número de siniestro y tasa de uso por cada estado civil	107
C.19. Dispersión tasa de riesgo, monto total facturado, número de siniestro y tasa de uso por cada sexo	108
C.20. Dispersión tasa de riesgo, monto total facturado, número de siniestro y tasa de uso por cada modalidad	109
C.21. Dispersión tasa de riesgo, Monto Total Facturado, Número de siniestro y tasa de uso por cada tiempo asegurado del vehículo	110
C.22. Dispersión tasa de riesgo, monto total facturado, número de siniestro y tasa de uso por cada edad del conductor del vehículo	111
C.23. Dispersión tasa de riesgo, monto total facturado, número de siniestro y tasa de uso por cada tiempo de exposición del vehículo	112

C.24. Dispersión tasa de riesgo, monto total facturado, número de siniestro y tasa de uso por cada tiempo de exposición del vehículo	113
D.1. Conglomerado a través de k medias, con K=2, con todas las variables originales	116
D.2. Conglomerado a través de k medias, con K=3, con todas las variables originales	117
D.3. Conglomerado a través de k medias, con K=4, con todas las variables originales	118
D.4. Conglomerado a través de k medias, con K=5, con todas las variables originales	119
D.5. Conglomerado a través de k medias, con K=6, con todas las variables originales	120
D.6. Conglomerado a través de k medias, con K=7, con todas las variables originales	121
E.1. Conglomerado a través de k medias, con K=2, con variables numéricas normalizadas y variables categóricas transformadas en ficticias de 0 y 1	124
E.2. Conglomerado a través de k medias, con K=3, con variables numéricas normalizadas y variables categóricas transformadas en ficticias de 0 y 1	125
E.3. Conglomerado a través de k medias, con K=4, con variables numéricas normalizadas y variables categóricas transformadas en ficticias de 0 y 1	126
E.4. Conglomerado a través de k medias, con K=5, con variables numéricas normalizadas y variables categóricas transformadas en ficticias de 0 y 1	127
E.5. Conglomerado a través de k medias, con K=6, con variables numéricas normalizadas y variables categóricas transformadas en ficticias de 0 y 1	128
E.6. Conglomerado a través de k medias, con K=7, con variables numéricas normalizadas y variables categóricas transformadas en ficticias de 0 y 1	129

ÍNDICE DE FIGURAS

F.1. Conglomerado a través de k medias, con $K=2$, con variables numéricas normalizadas y variables categóricas transformadas en ficticias de 0 y 1 normalizadas	132
F.2. Conglomerado a través de k medias, con $K=3$, con variables numéricas normalizadas y variables categóricas transformadas en ficticias de 0 y 1 normalizadas	133
F.3. Conglomerado a través de k medias, con $K=4$, con variables numéricas normalizadas y variables categóricas transformadas en ficticias de 0 y 1 normalizadas	134
F.4. Conglomerado a través de k medias, con $K=5$, con variables numéricas normalizadas y variables categóricas transformadas en ficticias de 0 y 1 normalizadas	135
F.5. Conglomerado a través de k medias, con $K=6$, con variables numéricas normalizadas y variables categóricas transformadas en ficticias de 0 y 1 normalizadas	136
F.6. Conglomerado a través de k medias, con $K=7$, con variables numéricas normalizadas y variables categóricas transformadas en ficticias de 0 y 1 normalizadas	137
G.1. Conglomerado a través de bagging con Weka 3.7	140
H.1. Conglomerado a través del aprendizaje del bosque aleatorio	142
I.1. Estudio transversal de conglomerado a través del análisis de componentes principales, se presenta la matriz de covarianza	144
I.2. Estudio transversal de conglomerado a través del análisis de componentes principales, se presenta los autovalores	145
J.1. Conglomerado a través de k medias, con $K=2$, con las variables numéricas normalizadas, con solo la variable categórica "marca modelo" transformada en variable ficticia de valores 0 y 1 normalizada	148

J.2. Conglomerado a través de k medias, con $K=3$, con las variables numéricas normalizadas, con solo la variable categórica "marca modelo" transformada en variable ficticia de valores 0 y 1 normalizada 149

J.3. Conglomerado a través de k medias, con $K=4$, con las variables numéricas normalizadas, con solo la variable categórica "marca modelo" transformada en variable ficticia de valores 0 y 1 normalizada 150

J.4. Conglomerado a través de k medias, con $K=5$, con las variables numéricas normalizadas, con solo la variable categórica "marca modelo" transformada en variable ficticia de valores 0 y 1 normalizada 151

J.5. Conglomerado a través de k medias, con $K=6$, con las variables numéricas normalizadas, con solo la variable categórica "marca modelo" transformada en variable ficticia de valores 0 y 1 normalizada 152

J.6. Conglomerado a través de k medias, con $K=7$, con las variables numéricas normalizadas, con solo la variable categórica "marca modelo" transformada en variable ficticia de valores 0 y 1 normalizada 153

K.1. Parámetros del análisis discriminante a un conglomerado con $K=2$. . . 156

K.2. Parámetros del análisis discriminante a un conglomerado con $K=3$. . . 157

K.3. Parámetros del análisis discriminante a un conglomerado con $K=4$. . . 158

K.4. Parámetros del análisis discriminante a un conglomerado con $K=5$. . . 159

K.5. Parámetros del análisis discriminante a un conglomerado con $K=6$. . . 160

K.6. Parámetros del análisis discriminante a un conglomerado con $K=7$. . . 161

L.1. Estudio longitudinal, conglomerado a través del análisis de componentes principales, gráfico de las curvas de las marcas modelos por monto total facturado acumulado 164

L.2. Estudio longitudinal, conglomerado a través del análisis de componentes principales, tabla de autovalores 165

ÍNDICE DE FIGURAS

L.3. Estudio longitudinal, conglomerado a través del análisis de componentes principales, gráfico de las curvas de las marcas modelos por monto total facturado acumulado y agrupado	166
M.1. Estudio longitudinal, conglomerado a través de k medias, con k=2 grupo uno, primera asignación, distancia euclideana	168
M.2. Estudio longitudinal, conglomerado a través de k medias, con k=2 grupo dos, primera asignación, distancia euclideana	169
M.3. Estudio longitudinal, conglomerado a través de k medias con distancia de Chiodi, k=2 grupo uno, segunda asignación	170
M.4. Estudio longitudinal, conglomerado a través de k medias con distancia de Chiodi, k=2 grupo dos, segunda asignación	171
M.5. Estudio longitudinal, conglomerado a través de k medias con distancia de Chiodi, k=2 grupo uno, tercera asignación, confirmación	172
M.6. Estudio longitudinal, conglomerado a través de k medias con distancia de Chiodi, k=2 grupo dos, tercera asignación, confirmación	173
M.7. Estudio longitudinal, conglomerado a través de k medias con distancia de Chiodi, k=2, se muestra los dos grupos, tercera asignación, confirmación	174
N.1. Función de distribución acumulada binomial negativa	176
N.2. Parámetros de la binomial negativa	177
N.3. Bondad de ajuste de la variable discreta 1 y 2	178
N.4. Bondad de ajuste de la variable discreta 3 y 4	179
N.5. Bondad de ajuste de la variable discreta 5	180
N.6. Gráfico P P binomial negativa	180
N.7. Función de densidad de probabilidad de la binomial negativa	181
N.8. Bondad de ajuste 1 al 2	182
N.9. Bondad de ajuste 3 al 4	183

N.10.Bondad de ajuste 5 al 6	184
N.11.Bondad de ajuste 7	185
N.12.Bondad de ajuste 8 al 9	186
N.13.Bondad de ajuste 10 al 12	187
N.14.Bondad de ajuste 13 al 14	188
N.15.Bondad de ajuste 15 al 16	189
N.16.Bondad de ajuste 17 al 18	190
N.17.Resultado de ajuste de la lognormal	191
N.18.Resultado del cálculo del ajuste de la lognormal	191
N.19.Resultado del cálculo del ajuste de la wakeby del 1 al 33	192
N.20.Resultado del cálculo del ajuste de la wakeby del 34 al 63	193
N.21.Resultado del cálculo del ajuste de la wakeby del 61 y 25	194
N.22.Función de distribución acumulada de la wakeby	195
N.23.Función de densidad de la wakeby	195
N.24.Función calculada para hallar el valor en riesgo	196

Índice de tablas

2.1. Variables sobre el contrato	13
2.2. Variables sobre el vehículo	13
2.3. Variables sobre el resumen de reclamos por vehículo	14
2.4. Variables sobre el detalle de reclamos	14
2.5. Variables utilizadas, modelo longitudinal de series cortas	15
2.6. Variables utilizadas, modelo pérdidas agregadas	15
4.1. Índice nacional de precios año 2015	74
4.2. Estimación del índice nacional de precios año 2016	75

Introducción

En el aprendizaje estadístico, la agrupación es una técnica matemática fundamental para determinar las variables que dan mayor información en los datos correspondientes al Seguro (definido en el apéndice [Ñ.1](#)) en el ramo de casco de automóvil.

Los objetivos del presente trabajo especial de grado son: Analizar las comparaciones de los modelos matemáticos de agrupación y presentar un modelo aleatorio llamado la máxima pérdida probable, dentro de la visualización del riesgo de casco de automóvil con la metodología del modelo probabilístico de pérdidas agregadas, identificada como: Cálculo del riesgo operacional.

Se realiza un análisis comparativo de modelos de conglomerados en datos con múltiples variables; además, se desarrolla el análisis de componentes principales para encontrar agrupaciones de individuos homogéneos en lo referente al patrón que explican las variables estudiadas. Se aporta una aplicación en la estimación de la máxima pérdida probable (MPP [Ñ.2](#)).

La situación incluye grupo ad hoc no bien clasificado, siendo esto una evidente fuente de incongruencias entre el monto total estimado, en bolívares, de reclamos para desarrollar las tarifas y el más cercano a la realidad, monto total de reclamos ocurridos a posteriori.

Por limitaciones en cuanto a “datos dispersos” y “poca información de los datos reales con incertidumbre de su validez”, de acuerdo con Panjer([1](#)), sólo se estima la MPP, en bolívares del seguros de casco de automóvil en la cobertura de pérdida parcial, basándose en datos históricos diarios desde 01 de enero de 2015 al 31 de diciembre de

1. INTRODUCCIÓN

2015 con el fin de estimar el monto de siniestro anual, del año 2016.

Se debe hacer la MPP, ya que las empresas aseguradoras en el seguro de casco del automóvil están inmersos en una alta volatilidad, variaciones constantes de costos de los repuestos y honorarios de reparación con respecto a estimaciones realizadas en anteriores períodos.

Se logra identificar y se coloca en las conclusiones, las características relevantes entre las técnicas matemáticas de “k medias” y el “análisis de componentes principales”, la “discriminación” de los vehículos y las tendencias obtenidas por procedimientos dentro del campo en estudio. Aportando una opción objetiva para abordar datos con múltiples variables y datos longitudinales mediante el “análisis de componentes principales” descrito por Vásquez y Ramírez(2), que paralelamente se reduce las dimensiones estudiadas. Rodríguez(3), Jorion(4) y Venegas(5) muestran modelos matemáticos de múltiples variables en el marco del riesgo.

Los conglomerados se utilizan para focalizar determinadas estructuras correspondientes a grupos homogéneos que son disimiles a los otros grupos del universo estudiado, lo cual permite obtener indicadores; con fines referenciales, estos conglomerados se utilizan para la aplicación de un indicador del modelo aleatorio de riesgo operacional, llamado por más de diez matemáticos, entre ellos: Hogg(6), Daykin(7) y Panjer(1), modelo de pérdidas agregadas bajo incertidumbre para evaluar el riesgo de pérdida parcial en bolívares por causas determinadas previamente, en un contrato de seguros a los vehículos expuestos a dichos acontecimientos y amparados por empresas registradas como aseguradoras en el mercado asegurador venezolano, para ello se procede a estimar la MPP. A los efectos anteriores se evalúa un modelo aleatorio probabilístico de pérdidas agregadas basado en dos modelos matemáticos: modelo probabilístico de las ocurrencias que afectan la pérdida parcial a vehículos en exposición y el modelo probabilístico de la severidad utilizada para dejar el vehículo en las mismas condiciones que tenía antes de la reclamación del siniestro, el reclamo está asociado a cada uno de los eventos previamente convenido en el contrato de seguros, como causantes de la pérdida parcial del vehículo expuesto al riesgo. Todo lo anterior soportado y programado en Charpentier(8) y en Kass(9).

La estructura general de las secciones que contiene el cuerpo del presente trabajo especial de grado es:

En el capítulo II compuesta por el planteamiento del problema de investigación, interrogantes, el objetivo general, los objetivos específicos, justificación de la investigación, importancia de la investigación, antecedentes de la investigación, los materiales fuentes de información, el universo, las unidades estadísticas, las variables y la cobertura.

En el capítulo III marco teórico, se aclara los aspectos metodológicos por medio de la puntualización del tipo y el diseño de la investigación, los métodos a manejar y la estrategia de análisis a utilizar para desarrollar la investigación.

En el capítulo IV se describe el análisis e interpretación de resultados.

En el capítulo V se expresan las conclusiones y recomendaciones.

Se indican las referencias bibliográficas utilizadas.

Se presentan los anexos correspondientes.

Planteamiento, objetivos, importancia antecedentes y materiales

2.1. Planteamiento del problema

2.1.1. 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

2. PLANTEAMIENTO, OBJETIVOS, IMPORTANCIA ANTECEDENTES Y MATERIALES

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 perteneciente 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(7) “...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...”. 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 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

Jorion(4) 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,
1. Permitir y facilitar la asignación de recursos en forma objetiva, y
1. 3. Evaluar el desempeño operacional.

2.1.2. Interrogantes de la investigación

1. ¿Cuáles características se obtienen en la comparación de los modelos elaborados para la obtención de conglomerados de análisis de múltiples variables tomando en cuenta los factores de riesgo en el seguros de casco del automóvil, a través de técnicas de aprendizaje estadístico no supervisado?
2. ¿Cómo establecer una unificación de criterio para agrupar determinados vehículos en una compañía de seguros con información del mercado asegurador venezolano solo cuando esta compañía no tenga suficiente o ninguna información?
3. ¿Cómo aplicar la comparación de las técnicas para la obtención de conglomerados en los factores de riesgo en el seguro de casco del automóvil en el mercado asegurador de Venezuela en el lapso anual 2015?
4. ¿Cómo se estima, para un conglomerado obtenido, los indicadores de los factores de riesgo bajo incertidumbre, en el seguro de casco del automóvil en el mercado asegurador de Venezuela en el lapso anual 2015?

2.2. Objetivo general

Analizar comparativamente e integrando los resultados en el uso del modelo de K medias con el uso del análisis de componentes principales elaborados para obtener agrupaciones, tomando en cuenta los factores de riesgo en el seguros de casco del automóvil, a través de técnicas de aprendizaje estadístico no supervisado.

2.3. Objetivos específicos

1. Realizar una unificación de criterio para agrupar determinados vehículos en una compañía de seguros con información de seis (6) compañías aseguradoras 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 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 en el lapso anual 2015.
3. Describir 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 del mercado asegurador de Venezuela en el lapso anual 2015.

2.4. Importancia de la investigación y la justificación

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

1. Permite mejorar la selección del modelo de agrupamiento, para datos transversales con múltiples variables.
2. Proporciona información 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 del 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(3), se evidencia la importancia de obtener los datos en forma efectiva y eficaz, a continuación, Rodríguez(3), 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 pérdidas 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(3), refleja la importancia de los datos procesados y de agrupar los datos por vehículo.

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

2.5. Antecedentes

El ramo de automóvil en el seguro y las ciencias actuariales ha sido abordado por gran cantidad de autores, entre ellos, Bowers(10), Daykin(7); mientras que Kass(9) realiza una generación pseudo-aleatoria de la cartera de seguros de automóvil y donde las pólizas tienen los 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 2.1 y 2.2) solo se podrá realizar una comparación aproximada. Kass(9) 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

2. PLANTEAMIENTO, OBJETIVOS, IMPORTANCIA ANTECEDENTES Y MATERIALES

de riesgo son diferentes al presente proyecto, se realizará una comparación aproximada. También, Kass(9) 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 trabajo del conductor (empleado de oficina, conductor habitual, o conductor dinámico), donde determinó el ranking de cada compañía de seguros; mientras en el presente proyecto se plantea realizar grupos de vehículos y no de empresas aseguradoras. Otra forma de abordar el tema lo describe Charpentier(8) 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; se puede realizar una leve comparación a pesar que los factores de riesgos son diferentes a los planteados en el presente anteproyecto. Finalmente, Charpentier(8) 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 este proyecto 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 casco.

En Venezuela, el tema de agrupamiento se ha abordado para otras ramas de la ciencia como lo es la salud; mientras se han realizado más de veinte (20) trabajos para calcular el costo del seguro, entre ellos Chirinos(11) desarrolla una tarifa de automóvil casco, López(12) encuentra el resultado técnico en automóvil individual, Thonon(13) aplica el análisis y la utilización de la estadística para la realización de la tarifa de automóvil.

El tema del aprendizaje estadístico lo ha trabajado Salazar(14) en su trabajo de grado sobre “Caracterización de deudores de tarjetas de crédito de una institución financiera a través de técnicas de aprendizaje estadístico”.

Vásquez y Ramírez(2) abordan el Análisis de Segmentación, Análisis de Componentes Principales y Ramírez(15) la teoría de probabilidad, que será utilizado plenamente en el presente trabajo especial de grado.

Días(16) en su trabajo de pasantía muestra la descripción de la superintendencia de la actividad aseguradora, institución donde se fundamentó el presente trabajo especial de grado.

Además, en el tema de conglomerado de datos longitudinales, Chiodi(17) realiza una técnica de agrupamiento, como una variante de la “k medias” multivariante de tres vías, de matrices cubicas con un enfoque de serie creciente de acuerdo a la función de mortalidad descrita por Gompertz.

Asimismo, Cuesta(18) realiza una investigación sobre la agrupación de la técnica k medias recortadas, desarrollando la codificación del paquete en el lenguaje R, llamado tkmeans, el cual tiene por objetivo obtener k grupos donde el promedio de las variables múltiples, llamados centroides, sean mínimos dentro de cada grupo y por el contrario, el promedio de las variables múltiples, centroides, sean máximas de cada individuos con respecto a los otros individuos ubicados en los otros k-1 grupos; la técnica no encuentra, por si sola, un número k óptimo de agrupaciones, lo cual lo hace menos potentes que otras técnicas que operan múltiples variables.

Para la agrupación a través de la construcción de curvas principales, Yépez(19) identifica las principales curvas que describen el patrón de evolución del comportamiento de la producción de petróleo obteniendo doce grupos de campos petroleros, y además, como resultado fundamental, se estimó la “máxima disminución probable no programada” en el valor volumétrico de la producción petrolera. A los efectos anteriores se simula, mediante Monte Carlo, un modelo probabilístico de pérdidas agregadas basado, a su vez, en el modelo probabilístico de las ocurrencias que afectan la “disminución” y el modelo probabilístico de la severidad asociada a cada uno de los eventos descritos, como causantes de la “disminución” de la producción de petróleo.

2.6. Materiales

2.6.1. 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.

2.6.2. Universo

Universo de estudio definido por seis (6) compañías de seguros que son parte parcial del mercado asegurador venezolano.

2.6.3. Unidades estadísticas

En el presente trabajo 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.

2.6.4. Variables

Las variables utilizadas en el modelo de agrupación transversal, son:

Tabla 2.1: 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 2015.

Tabla 2.2: 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 2015.

2. PLANTEAMIENTO, OBJETIVOS, IMPORTANCIA ANTECEDENTES Y MATERIALES

Tabla 2.3: 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 2015.

Tabla 2.4: 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 2015.

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

Tabla 2.5: 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 facturado total	Por cada Marca-Modelo, por mes acumulado.

Fuente: Sudeaseg 2015.

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

Tabla 2.6: Variables utilizadas, modelo pérdidas agregadas

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

Fuente: Sudeaseg 2015.

2.6.5. 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.

A continuación se presenta las bases que constituyen el soporte teórico del presente trabajo.

Marco teórico

3.1. Aspectos metodológicos

A continuación, se presenta la estructura del trabajo:

- a) Se Realiza agrupaciones por el aprendizaje estadístico no supervisado a través de k medias para visualizar la forma de los grupos; se contruye 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 arbol 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 Utiliza 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. Simultaneamente, se incorpora la teoría de valores extremos para optimizar el cálculo de la estimación futura de las pérdidas.

3.1.1. Tipo y diseño de la investigación

El presente estudio es 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(20). La investigación es efectivamente comparativa, debido a que busca las similitudes y diferencias de varias técnicas de agrupación. La investigación es de tipo cuantitativa y no experimental ya que hace uso de datos procedentes de históricos reales, sin manipulación por parte del investigador, de acuerdo a Sampieri(21). 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.

3.1.2. Métodos utilizados

A continuación se nombran los métodos utilizados.

Método estadístico que fundamenta la estrategia utilizada

Análisis de componentes principales: El análisis de componentes principales se utiliza para los datos transversales, según Yépez(19) la técnica de múltiples variables conocida como análisis de componentes principales (ACP), que puede ser presentada según el enfoque geométrico de K. Pearson(22) citado por Vásquez y Ramírez(23), 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, en este caso, a la caracterización de los vehículos amparados en el ramo casco del automóvil asegurado de acuerdo a los reclamos de gastos desde 01 de enero de 2015 al 31 de diciembre de 2015. El ACP al ser aplicado sobre datos transversales definidos por los vehículos asegurados, se utilizará para describir los patrones de grupos de vehículos asegurados en el momento bajo estudio. Esta información se organiza sobre una matriz de datos centrada 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 Ramírez (2007) (23), al aplicar el ACP sobre la matriz

$X_c X_c^t$ y considerar las coordenadas de las variables proyectadas sobre el α -ésimo eje, se obtiene la expresión:

$$\varphi_{px1}^\alpha = X_{pxn}^t U_{nx1}^\alpha = \sum_{i=1}^n u_{i\alpha} x_i \quad (3.1)$$

Siendo: U_{nx1}^α autovector normalizado asociado con λ_α , el α -ésimo mayor autovalor de la matriz $X_c X_c^t$:

$$X_c X_c^t = \begin{pmatrix} u_{1\alpha} \\ u_{2\alpha} \\ \vdots \\ u_{i\alpha} \\ \vdots \\ u_{n\alpha} \end{pmatrix}_{nx1}$$

x_i es el vector que describe (ver tablas 2.1, 2.2 y 2.3) el valor de las p variable del i-ésimo vehículo asegurado, en el momento del estudio:

$$x_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ij} \\ \vdots \\ x_{ip} \end{pmatrix}_{px1}$$

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:

$$\frac{\lambda_\alpha}{\sum_{\alpha=1}^n \lambda_\alpha} * 100 \quad (3.2)$$

El cociente 3.2 mide la proporción de la variabilidad total de los vehículos expuestos al riesgo que es captada por φ^α , la α -ésima componente principal.

2. φ^α es una combinación lineal de los vectores X_i , que describen los valores de cada uno de los vehículos expuestos al riesgo bajo estudio, en consecuencia puede señalarse que: ***φ^α es la α -ésima componente principal de los valores de los vehículos asegurados.***

3. MARCO TEÓRICO

3. La importancia del i -ésimo vehículo en la formación de la α -ésima componente principal φ^α , puede obtenerse al evaluar el coeficiente $u_{i\alpha}$ de manera que: $u_{i\alpha} - 2$
Mide la contribución del i -ésimo vehículo asegurado a la variabilidad por la α -ésima componente principal de los valores de los mismos.
4. La contribución de la α -ésima componente principal para explicar el patrón entre las variables del i -ésimo vehículo asegurado, se define en la forma:

$$\frac{\varphi_{i\alpha}^2}{d^2(x_i, \bar{x})} \quad (3.3)$$

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

Agrupación k medias: Según Cuesta(18) 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 cada grupo y por el contrario, el promedio de las variables múltiples, centroides, sean máximas de cada individuos con respecto 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 estudio y se aplica la técnica de k medias.

Análisis discriminante con variables normalizadas: Según Díaz, Morales (2012) (24) 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, Morales (2012) (24).

Estudio longitudinal, conglomerado a través del análisis de componentes principales: También Yépez (2010, pág 49) (19) 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 cuya filas se corresponden con las marcas modelos de los vehículos amparados, y cuyas columnas describen la pérdida monetaria acumulada mensual en bolívares a lo largo del período bajo estudio.

Estudio longitudinal, conglomerado a través de una modificación de k medias con una distancia propuesta por Chiodi: Según Chiodi (1987, p.445) (17) 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 Chiodi (1987, p.445) (17) indicando que el problema puede ser crucial si los datos están constituidos por series temporales cortas.

Método matemático, modelo aleatorio de pérdidas agregadas

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 Jorion, P.(2007) (4) 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 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 Venegas(5), el riesgo se fundamenta en lo siguiente: “Si S es una variable aleatoria definida en (Ω, f, P_θ) , donde θ es un vector de parámetros asociado con la distribución de S ... el valor en riesgo (VaR_{1-q}^S) con un nivel de confianza del $(1 - q)100\%$, se define como el valor en la distribución de esa variable aleatoria tal que, en un período de tiempo dado, $[t, T]$ la probabilidad:

$$P_\theta \{VaR_{1-q}^S \leq S\} = 1 - q \text{ ó } P_\theta \{S \geq VaR_{1-q}^S\} = 1 - q \quad (3.4)$$

A partir de esta definición, se desprende, que:

$$VaR_{1-q}^S = \text{infimo} \{S \in R | P_\theta \{S < s\} \leq 1 - q\} \quad (3.5)$$

Como se puede observar, la 3.5 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

3. MARCO TEÓRICO

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 importante resaltar que esa metodología está basada en el supuesto de normalidad de la variable “S”. En la presente investigación el objetivo consiste en encontrar la distribución probabilística de la pérdida global monetaria causada a una cartera de vehículos asegurados en seis (6) compañías de seguros, una parte parcial del mercado asegurador venezolano, en el seguro de “casco de automóvil”, cobertura de “pérdida parcial” y solo a un grupo de vehículos previamente seleccionado, utilizando el enfoque de la “Función Probabilística Compuesta” desarrollado por Daykin(7), que requiere del uso de distribuciones de valores extremos. Según Jorion(4), 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 esten excluidos en el contrato de seguros.

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 Daykin(7) es el enfoque más apropiado. El modelo probabilístico de pérdidas agregadas, siendo un modelo econométrico de orden uno AR(1), se define a través de la combinación de los siguientes dos modelos: En el contexto del proceso de siniestros anuales, se asumen n 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 su ocurrencia futura, las causas de siniestro son los eventos propiamente materializados, los cuales pueden ser el volcamiento, la colisión con otros vehículos u objetos fijos y cualquier causa que no este especificada como exclusiones en el contrato de seguros:

$$E_1, E_2, \dots, E_n, \text{ con } n = 0, 1, 2, \dots \quad (3.6)$$

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”:

$$X_1, X_2, \dots, X_n, \text{ con } n = 0, 1, 2, \dots \quad (3.7)$$

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

$$S = X_1 + X_2 + \dots + X_n = \sum_{i=1}^n X_i, \text{ con } n = 0, 1, 2, \dots \quad (3.8)$$

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 y el modelo de pérdidas agregadas.

1. Modelo aleatorio de ocurrencia de eventos (siniestros/reclamos)
Siguiendo la formulación de Daykin(7), puede considerarse que el número N 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:

$$p_N = \text{Prob}(N = n), \text{ con } N = 0, 1, 2, \dots \quad (3.9)$$

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 Ramírez(15), la variable aleatoria: K : Número de eventos (siniestros/reclamos) que ocurren

3. MARCO TEÓRICO

anualmente. Sigue una distribución probabilística de “Binomial negativa”, con función de masa definida por:

$$p_N = \text{Prob}(N = r) = f(r) = \binom{N+r-1}{r} p^N (1-p)^r, \text{ con } N = 0, 1, 2, \dots \quad (3.10)$$

Donde:

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

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

Valor esperado:

$$E(r) = \frac{N(1-p)}{p} \quad (3.11)$$

Varianza:

$$\text{Var}(r) = \frac{N(1-p)}{p^2} \quad (3.12)$$

Relación del valor esperado y la varianza.

$$E(r) \geq \text{Var}(r) \quad (3.13)$$

Estimación de los parámetros N y p .

Los parámetros N y p , 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 Hogg(6), llamado el método de los momentos, también desarrollado de manera automática por el programa gratuito “EasyFit(r)”. Con las funciones 3.11 y 3.12 se expresa los parámetros “N” y “p”:

$$N = \frac{(E[r])^2}{\text{Var}[r] - E[r]} \text{ y } p = \frac{E[r]}{\text{Var}[r]} \quad (3.14)$$

Se deducen los estimadores consistentes de N y p reemplazando 3.11 y 3.12 por sus estimadores \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 contraste de Bondad de Ajuste, desarrollado de manera automática por el programa gratuito “EasyFit(r)”, para corroborar que la distribución empírica se comporta conforme a la función teórica “Binomial negativa”. Según se describe en Panjer(1), 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:

$$Q = \sum_{j=1}^J \frac{(a_j - A_j)^2}{A_j} \quad (3.15)$$

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 “a” 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 $A_j = a \text{Prob}(r = j), j = 1, 2, \dots, J$.

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

Sea “N” 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 siniestros en el grupo de vehículos asegurados en las seis (6) empresas de seguros estudiadas, y sea “X” 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 “X”. Siguiendo a Daykin (op. cit), 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) la Generalizada de Valores Extremos tres parámetros.

La variable aleatoria “X”, en el intervalo $(0, \text{Percentil}_{99,97})$, sigue la distribución LogNormal de tres parámetros:

$$X_i \sim \text{LogNorm}(\mu, \sigma, \gamma); \text{ con } i=1,2,\dots,N \quad (3.16)$$

Parámetros:

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

μ : parámetro continuo;

γ : parámetro continuo de ubicación ($\gamma \equiv 0$ equivale a la distribución Lognormal de dos parámetros)

Dominio:

$\gamma < x < +\infty$

Función de densidad de probabilidad:

$$f(x) = \frac{1}{\sigma(x-\gamma)\sqrt{2\Pi}} e^{\left(-\frac{1}{2}\left(\frac{\ln(x-\gamma)-\mu}{\sigma}\right)^2\right)} \quad (3.17)$$

Función de distribución acumulada:

$$F(x) = \frac{1}{\sigma\sqrt{2\Pi}} \int_{\gamma}^x \frac{1}{(t-\gamma)} e^{\left(-\frac{1}{2}\left(\frac{\ln(t-\gamma)-\mu}{\sigma}\right)^2\right)} dt \quad (3.18)$$

La variable aleatoria “X”, en el intervalo (Percentil_{99,971},Percentil₁₀₀), sigue la distribución Generalizada de Valores Extremos tres parámetros que corresponde a la generalización conjunta de las distribuciones de Fréchet, Weibull y Gumbel:

$$X_i \sim \text{GenExtreV}(\mu, \sigma, \kappa); \text{ con } i=1,2,\dots,N \quad (3.19)$$

Parámetros:

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

μ : parámetro continuo de ubicación;

κ : parámetro continuo de forma

Dominio:

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

$-\infty < x < +\infty$ para $\kappa = 0$

Función de densidad de probabilidad:

$$f(x) = \begin{cases} \frac{1}{\sigma} e^{\left(-\left(1+\kappa z\right)^{-\frac{1}{\kappa}}\right)} \left(1 + \kappa z\right)^{\left(-1-\frac{1}{\kappa}\right)} & \text{si } \kappa \neq 0 \\ \frac{1}{\sigma} e^{\left(-z - e^{-z}\right)} & \text{si } \kappa = 0 \end{cases} \quad (3.20)$$

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

Función de distribución acumulada:

$$F(x) = \begin{cases} \frac{1}{\sigma} \int_{\mu-\frac{\sigma}{\kappa}}^x e^{\left(-\left(1+\kappa\left(\frac{t-\mu}{\sigma}\right)^{-\frac{1}{\kappa}}\right)\right)} \left(1 + \kappa\left(\frac{t-\mu}{\sigma}\right)\right)^{\left(-1-\frac{1}{\kappa}\right)} dt & \text{si } \kappa \neq 0 \\ \frac{1}{\sigma} \int_{-\infty}^x e^{\left(-\left(\frac{t-\mu}{\sigma}\right) - e^{-\left(\frac{t-\mu}{\sigma}\right)}\right)} dt & \text{si } \kappa = 0 \end{cases} \quad (3.21)$$

Siguiendo a Hogg(6), 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” 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.

2. 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:

$$S = X_1 + X_2 + \dots + X_N = \sum_{i=1}^N X_i, \text{ con } N = 0, 1, 2, \dots \quad (3.22)$$

Se supone que montos de siniestros, X_i , causado por la ocurrencia del i -ésimo evento es independiente de la variable aleatoria N , y que además se distribuye de acuerdo al modelo mixto antes propuesto. Entonces, el modelo probabilístico de pérdidas agregadas debido a la ocurrencia de N eventos, queda descrito por la distribución de variable aleatoria compuesta S que describe el monto acumulado indemnizado en siniestros de la pérdida anual de seis (6) empresas del mercado asegurador correspondiente al seguro de casco de automóviles, en la cobertura de pérdida parcial.

Distribución de la variable compuesta:
variable

$$S = \sum_{i=1}^N X_i, \text{ con } N = 0, 1, 2, \dots \quad (3.23)$$

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

$$n = 0 \quad n = 1 \text{ y } x_1 \leq s \quad n = 2 \text{ y } x_1 + x_2 \leq s \quad n = 3 \text{ y } x_1 + x_2 + x_3 \leq s \quad (3.24)$$

Función de Distribución Acumulada

$$F_S(s) = \sum_{n=0}^{\infty} p_n Pr(S \leq s | N = n) = \sum_{n=0}^{\infty} p_n F_S^{*n}(s) \quad (3.25)$$

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

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

$F_S^{*n}(s) = F_S^* F_S^{*(n-1)}(s)$, con $n > 0$. Se detalla algunas convoluciones:

$$\begin{aligned} F_S^{*n}(s) &= \int_0^s F_S^{*(n-1)}(s-y) f_S(y) dy \\ F_S^{*(n-1)}(s) &= \int_0^s F_S^{*(n-2)}(s-y) f_S(y) dy \\ &\vdots \\ &\vdots \\ &\vdots \\ F_S^{*(n-(n-2))}(s) &= \int_0^s F_S^{*(n-(n-1))}(s-y) f_S(y) dy \\ F_S^{*(n-(n-1))}(s) &= \int_0^s F_S^{*(n-n)}(s-y) f_S(y) dy \\ F_S^{*(n-n)}(s) &= F_S^{*(0)}(s) = 1, \text{ siempre y cuando } s \geq 0. \end{aligned} \quad (3.26)$$

Para el presente estudio la variable en ningún momento será menor a cero. Daykin, autor cuyas referencias han signado estos desarrollos, junto con las de Yépez, 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. Con la información descrita anteriormente Jorion(4) propone que la empresa de seguros pueda realizar las tres actividades siguientes:

- a. 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.
- b. 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.
- c. 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.

3.2. Metodología matemática en la ciencia de los datos

3.2.1. 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 de-

ficientes; 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 estudio.
- * Se determinan los datos incorrectos, de ser significativo, se procede de la misma forma que el item anterior.

En otras ocasiones los datos estan completos y no se aprecia incongruencias aparentes; pero no aportan información, ya que la forma de registrarlos es incorrecta o no estan 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 comparandolos con indicadores construidos especialmente para dicha revisión, de no aportar información se procede igual que en el item 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.

Análisis e interpretación de resultados

A continuación, se presenta el análisis y la interpretación de resultados de los métodos de cálculo desarrollados en el marco teórico. Los datos utilizados de la información de los vehículos asegurados son extraídos del Proyecto de Unificación de Solicitudes 2015 (Fase I), se trabajó con los archivos DS (Detalle de Siniestros) la información correspondiente a los siniestros pagados y pendientes de cada una de las empresas de seguros, los archivos identificados por RC (Recibos) donde se identifica la exposición de riesgo del vehículo y los archivos señalados por VEH (Vehículos) que describen los factores de riesgos. Se muestra los histogramas, correlación y gráficos de dispersión correspondientes a los potenciales factores de riesgo, estas variables están descritas en la “tabla sobre el vehículo” [2.2](#):

4. ANÁLISIS E INTERPRETACIÓN DE RESULTADOS

1. Año de construcción del vehículo asegurado.

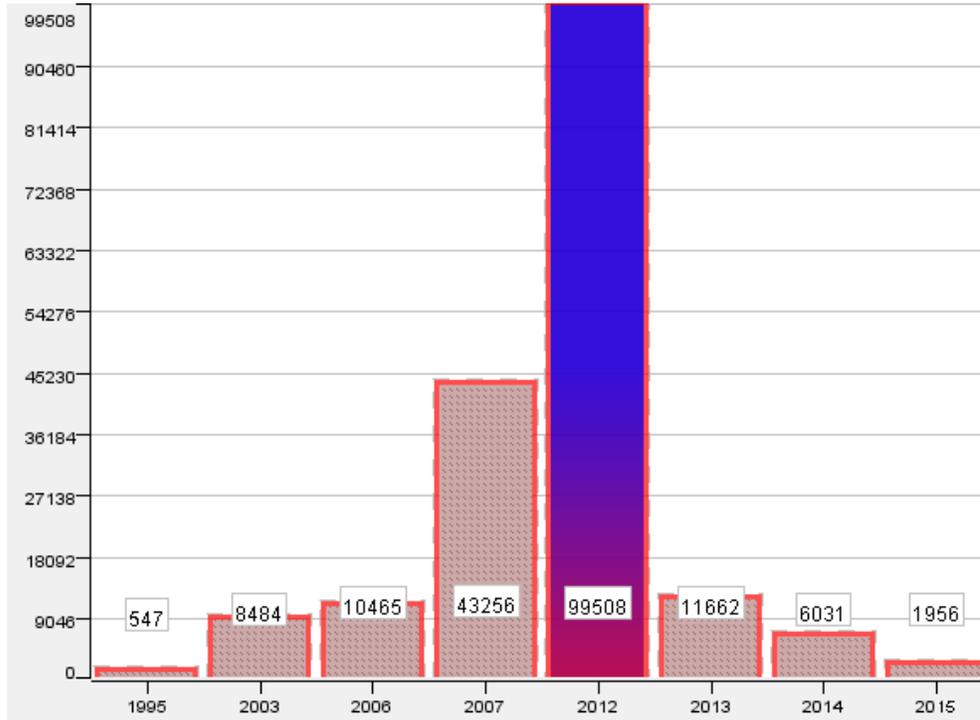


Figura 4.1: Histograma número de vehículos asegurados por año de construcción - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura 4.1, tipo de variable: numérica discreta, valores: desde 1980 hasta 2015, en primera instancia se puede seleccionar la variable para realizar los otros estudios antes de la agrupación transversal, además, se evidencia el bajo aseguramiento de carros construidos después del 2012.

2. Valor del vehículo al 2015 expresado en millones de Bolívares.

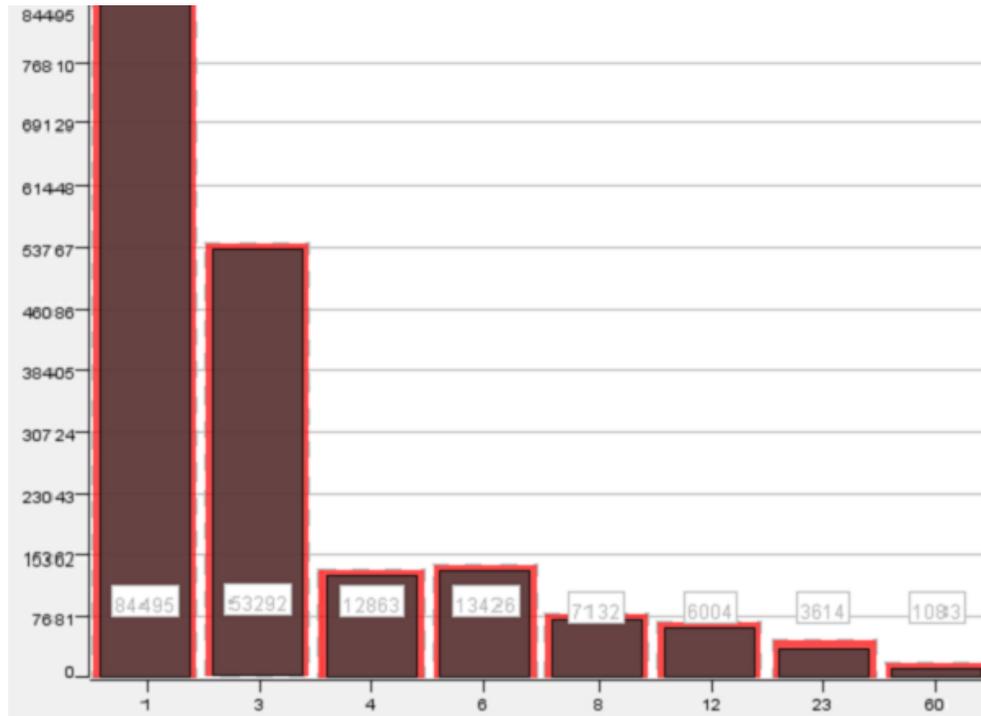


Figura 4.2: Histograma número de vehículos por valor del vehículo expresado en millones de Bolívares - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura 4.2, tipo de variable: numérica continua establecida por Daykin(7), valores: desde 1 Millón hasta 51 Millones de Bolívares, en primera instancia se puede seleccionar la variable para realizar los otros estudios antes de la agrupación transversal, además, se evidencia el alto aseguramiento de carros con valores por debajo de 3 millones de Bolívares.

4. ANÁLISIS E INTERPRETACIÓN DE RESULTADOS

3. Monto total de siniestro del vehículo expresado en millones de Bolívares.

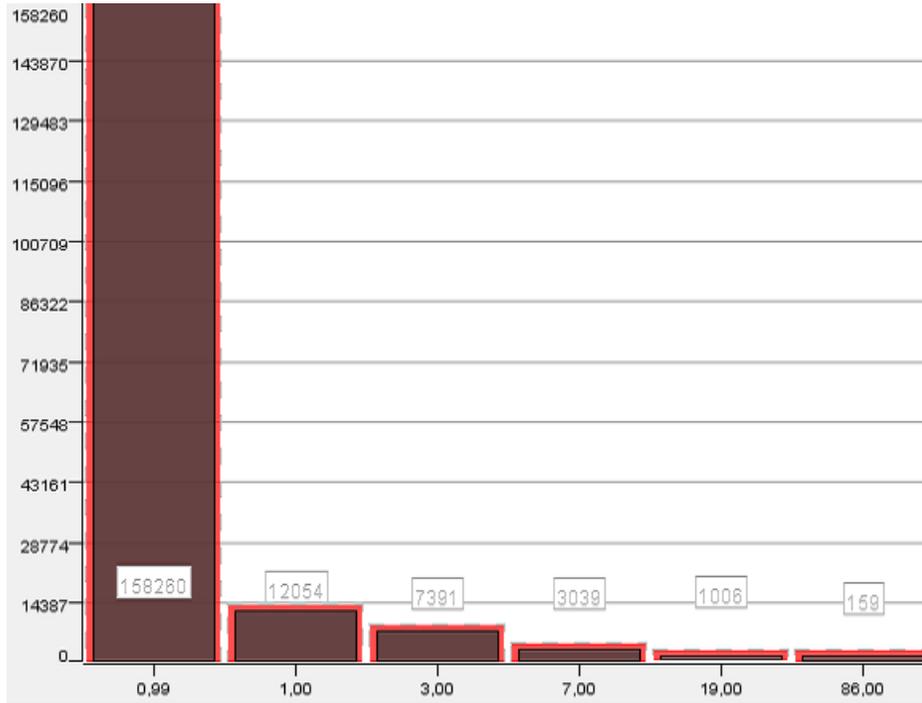


Figura 4.3: Histograma número de vehículos por el monto total de siniestro del vehículo - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura 4.3, tipo de variable: numérica continua, valores: desde 0 hasta 8,6, en primera instancia se puede seleccionar la variable para realizar los otros estudios antes de la agrupación transversal, además, esta variable es la razón de los estudios por lo que se considera la variable y_1 .

4. Número total de siniestros del vehículo.

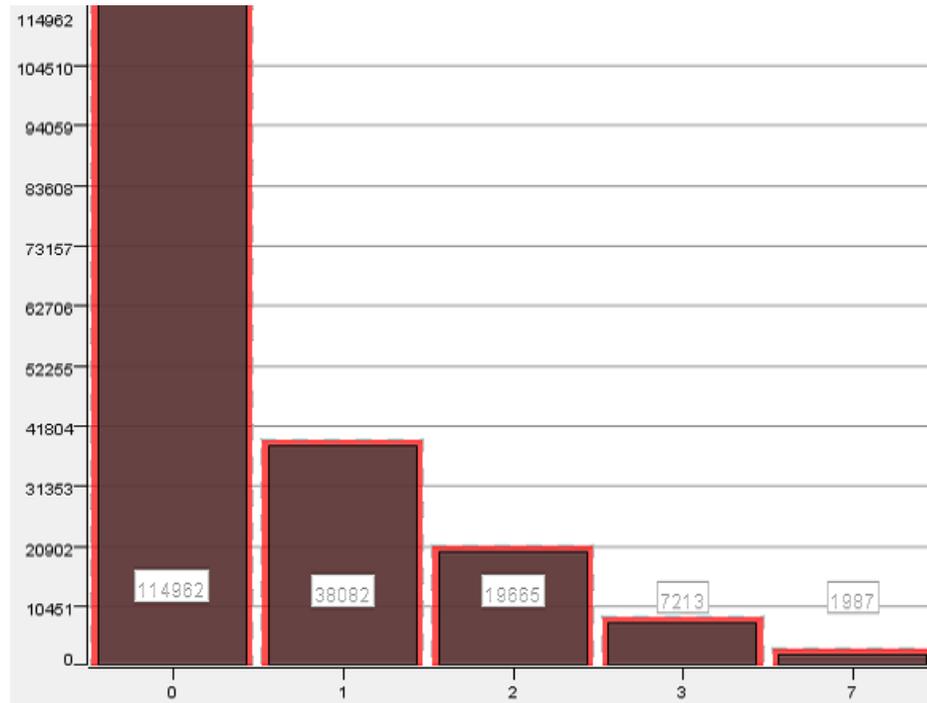


Figura 4.4: Histograma número de vehículos por el número total de siniestro del vehículo - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura 4.4, tipo de variable: numérica discreta, valores: desde 0 hasta 7, en primera instancia se puede seleccionar la variable para realizar los otros estudios antes de la agrupación transversal, además, esta variable es la razón de los estudios por lo que se considera la variable y_2 .

4. ANÁLISIS E INTERPRETACIÓN DE RESULTADOS

5. Marca modelo del vehículo.

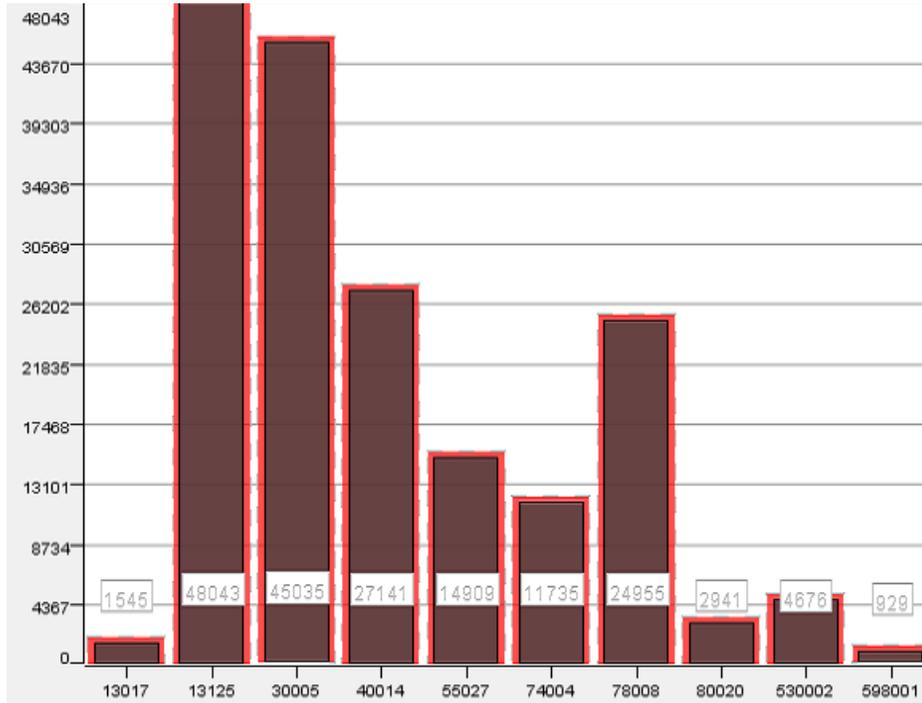


Figura 4.5: Histograma número de vehículos por cada marca y modelo asegurado - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura 4.5, tipo de variable: categórica sin orden, valores: código desde 001000 al 500000, en primera instancia se puede seleccionar la variable para realizar los otros estudios antes de la agrupación transversal, además, esta variable es una de las principales para discriminar grupos y realizar tarifas.

6. Número de ocupantes del vehículo.

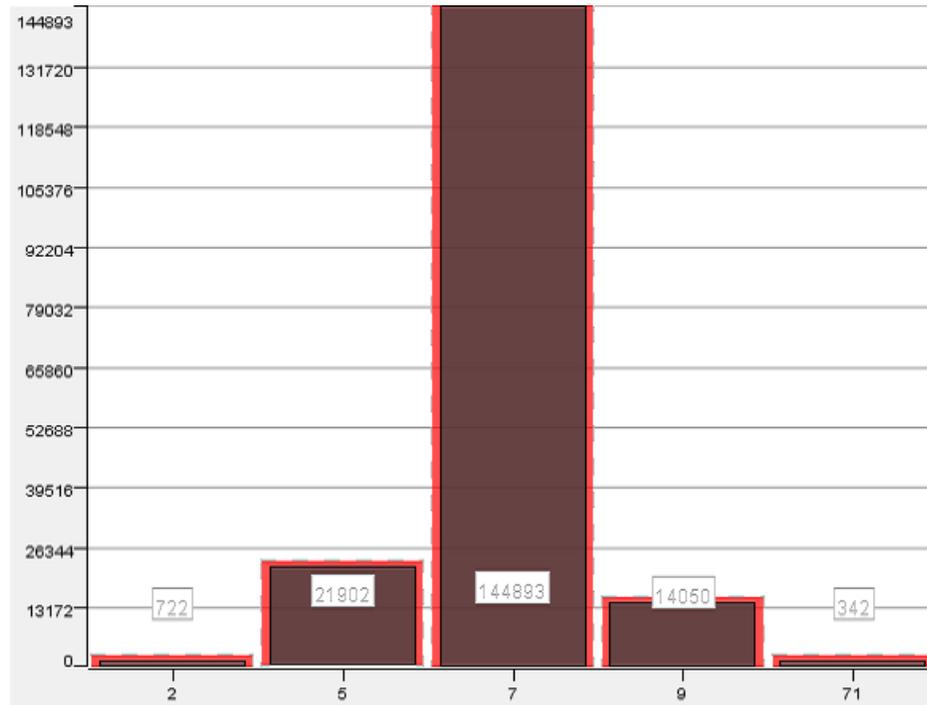


Figura 4.6: Histograma número de vehículos por cada número de ocupantes de los vehículos - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura 4.6, tipo de variable: categórica de orden, valores: desde 1 hasta 71, en primera instancia se puede seleccionar la variable para realizar los otros estudios antes de la agrupación transversal, aunque, como se dice en las limitaciones, se trabajara con vehículos con ocupantes desde 2 hasta 8.

4. ANÁLISIS E INTERPRETACIÓN DE RESULTADOS

7. Entidad geográfica de circulación del vehículo.

Con respecto a la entidad geográfica de circulación del vehículo asegurado, tipo de variable: categórica sin orden, valores: código 010000 al 600000, no se puede seleccionar la variable para realizar los otros estudios antes de la agrupación transversal, porque se observa que la información es de carácter administrativo sin considerar la verdadera zona de circulación.

8. Estado civil del conductor del vehículo.

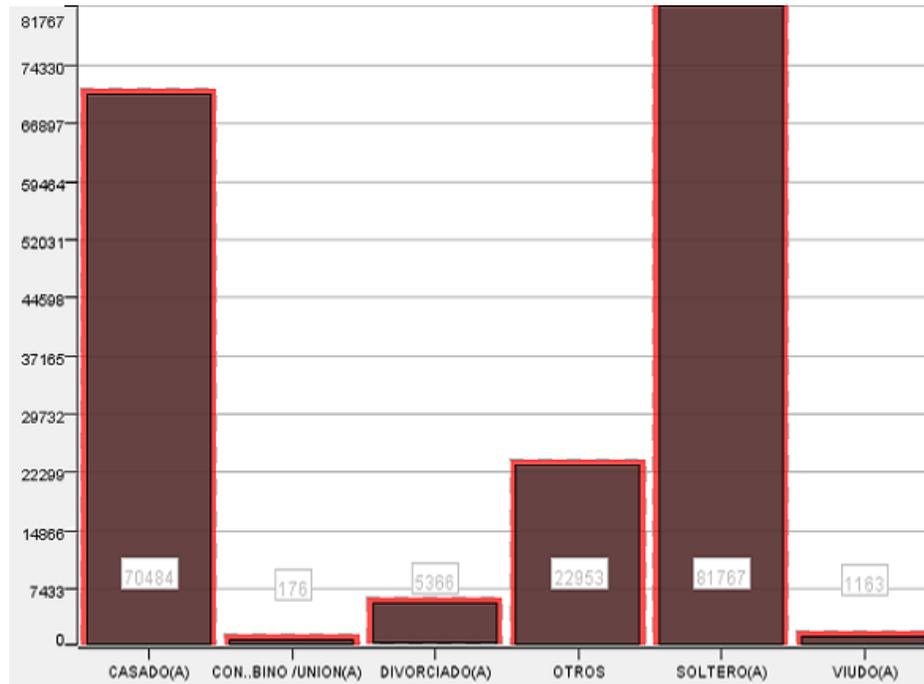


Figura 4.7: Histograma número de vehículos por cada estado civil observado -

Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura 4.7, tipo de variable: categórica sin orden, valores: 1 Casado(a); 2 Soltero(a); 3 Divorciado(a); 4 Viudo(a); 5 Concubino(a), en primera instancia se puede seleccionar la variable para realizar los otros estudios antes de la agrupación transversal, además, se observa que los casados y solteros tienen la mayoría de los vehículos.

4. ANÁLISIS E INTERPRETACIÓN DE RESULTADOS

9. Sexo del conductor del vehículo.

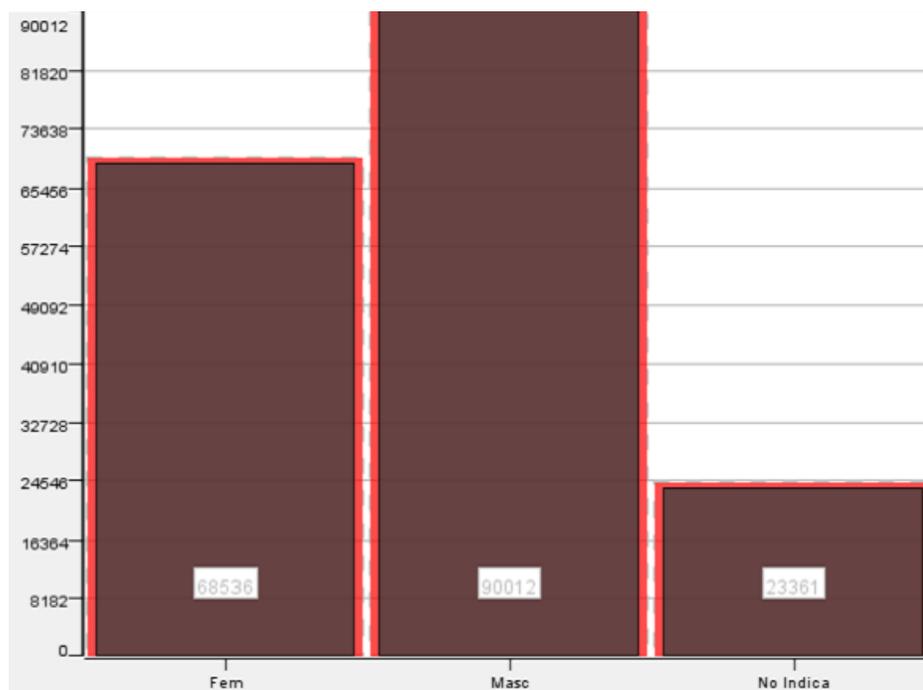


Figura 4.8: Histograma número de vehículos por cada sexo observado - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura 4.8, tipo de variable: categórica sin orden, valores: 1 Mujer, 2 Hombre, 3 No informaron, en primera instancia se puede seleccionar la variable para realizar los otros estudios antes de la agrupación transversal, además, se observa que el sexo masculino tienen la mayoría de los vehículos.

10. Modalidad del vehículo.

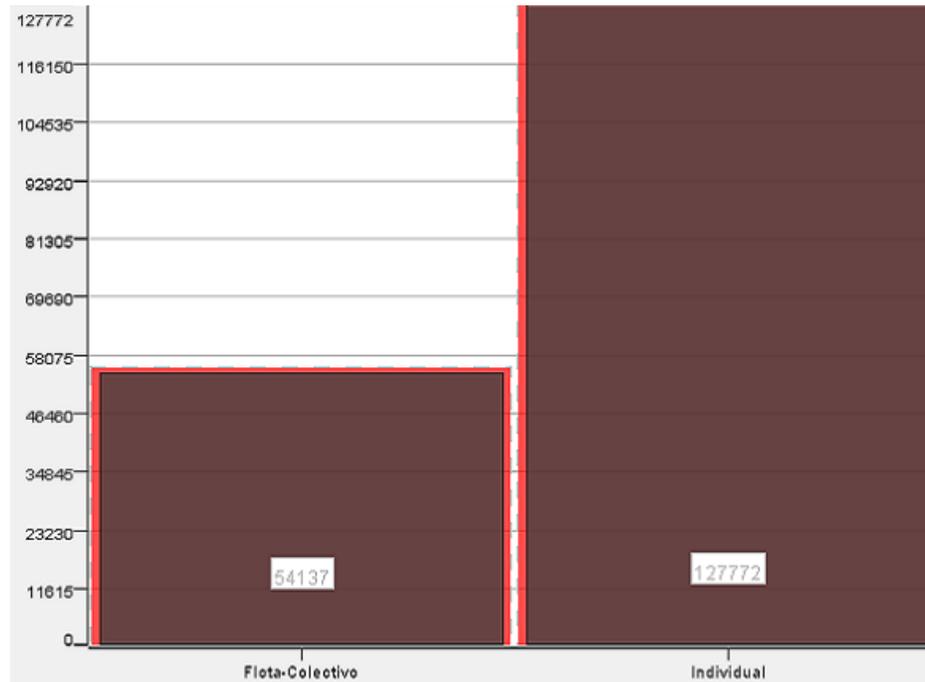


Figura 4.9: Histograma número de vehículos por cada modalidad observada -

Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura 4.9, tipo de variable: categórica sin orden, valores: 1 Individual, 2 Colectivo o Flota, en primera instancia se puede seleccionar la variable para realizar los otros estudios antes de la agrupación transversal, además, se observa que los contratos realizados individualmente tienen la mayoría de los vehículos.

4. ANÁLISIS E INTERPRETACIÓN DE RESULTADOS

11. Contratante gubernamental o no del seguro.

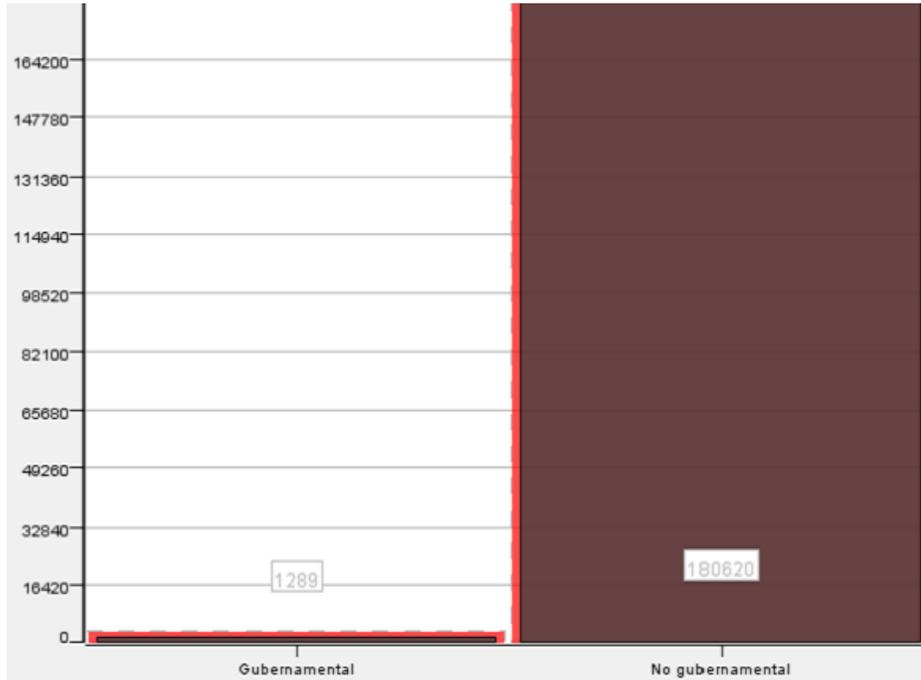


Figura 4.10: Histograma número de vehículos por cada tipo de contratante observado - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura 4.10, tipo de variable: categórica sin orden, valores: 1 Gubernamental, 2 No gubernamental, no se puede seleccionar la variable para realizar los otros estudios antes de la agrupación transversal, porque se observa que la información está más del 99% contenida en no gubernamental, la discriminación puede ser descartada.

12. Tipo de vehículo.

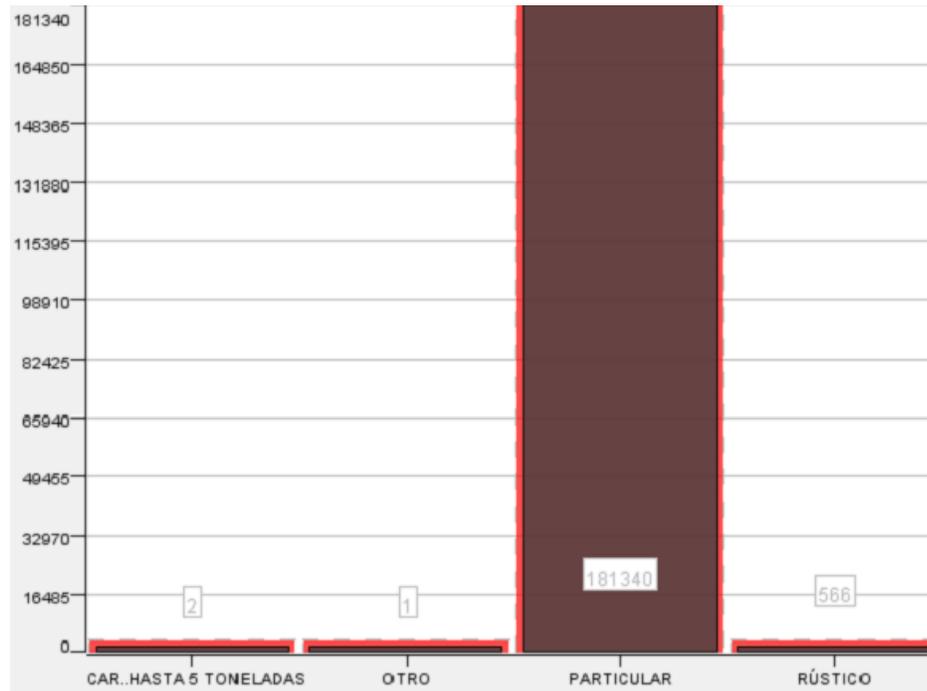


Figura 4.11: Histograma número de vehículos por cada tipo de vehículo observado - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura 4.11, tipo de variable: categórica sin orden, valores: 1 Particular, 2 Rústico, 3 Pick Up, 4 Carga hasta 5 TN, 99 Otros, no se puede seleccionar la variable para realizar los otros estudios antes de la agrupación transversal, ya que se observa que el código particular tienen la mayoría de los vehículos cuando existen vehículos de todos los tipos identificado como particular.

4. ANÁLISIS E INTERPRETACIÓN DE RESULTADOS

13. Densidad poblacional donde circula personas por kilómetro cuadrado.

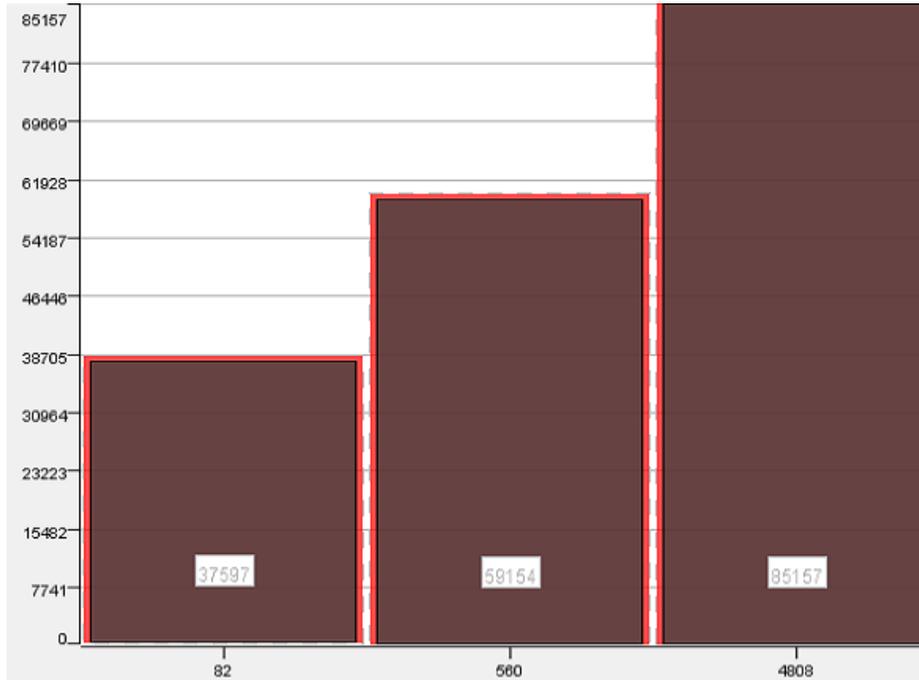


Figura 4.12: Histograma número de vehículos por densidad observada - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura 4.12, tipo de variable: numérica discreta, no se puede seleccionar la variable para realizar los otros estudios antes de la agrupación transversal, ya que esta variable es dependiente a la Zona geográfica que también fue excluida del estudio.

14. Tiempo de asegurado en años.

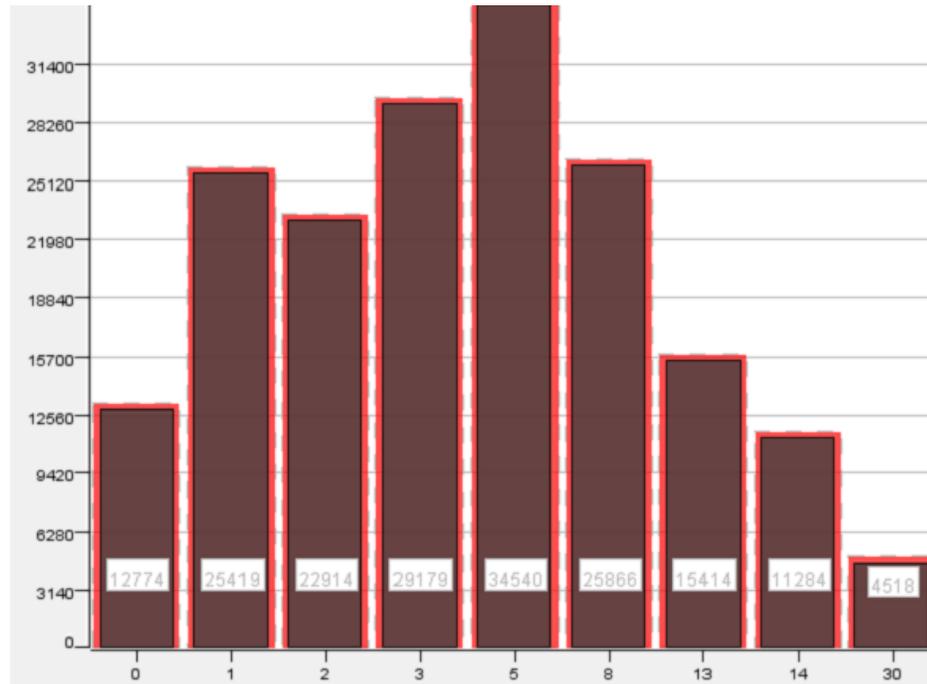


Figura 4.13: Histograma número de vehículos por el número de renovaciones observadas - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura 4.13, tipo de variable: categórica de orden, en primera instancia se puede seleccionar la variable para realizar los otros estudios antes de la agrupación transversal, además, se observa que los contratos que tienen entre 4 y 5 años contienen la mayoría de los vehículos.

4. ANÁLISIS E INTERPRETACIÓN DE RESULTADOS

15. Edad del conductor del vehículo.

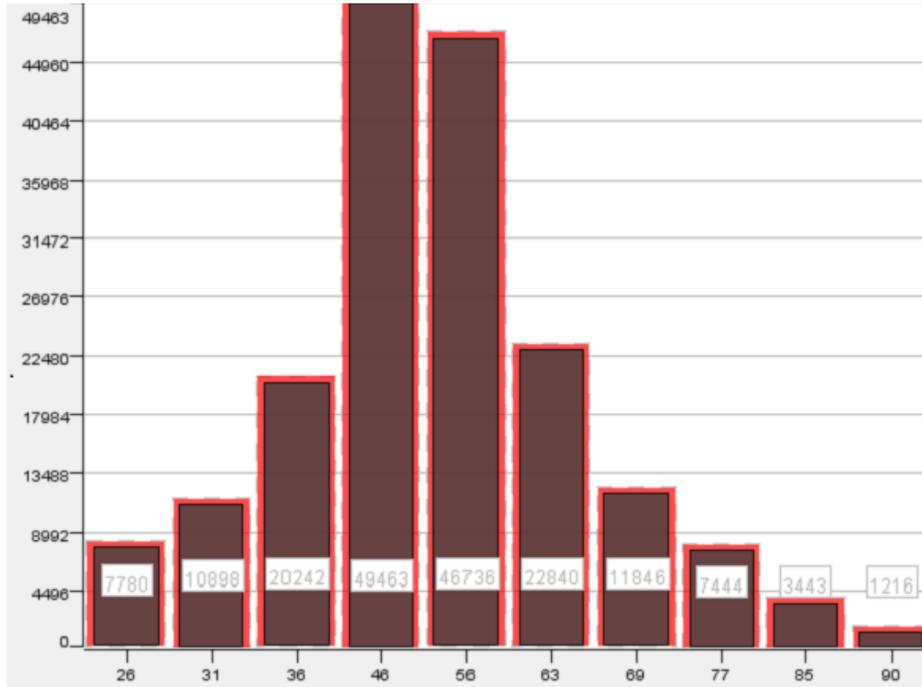


Figura 4.14: Histograma número de vehículos por cada edad observada - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura 4.14, tipo de variable: numérica discreta, en primera instancia se puede seleccionar la variable para realizar los otros estudios antes de la agrupación transversal, además, se observa que la clase con los conductores de edad entre 37 y 46 años, contienen la mayoría de los vehículos.

16. Factor de exposición al riesgo del vehículo.

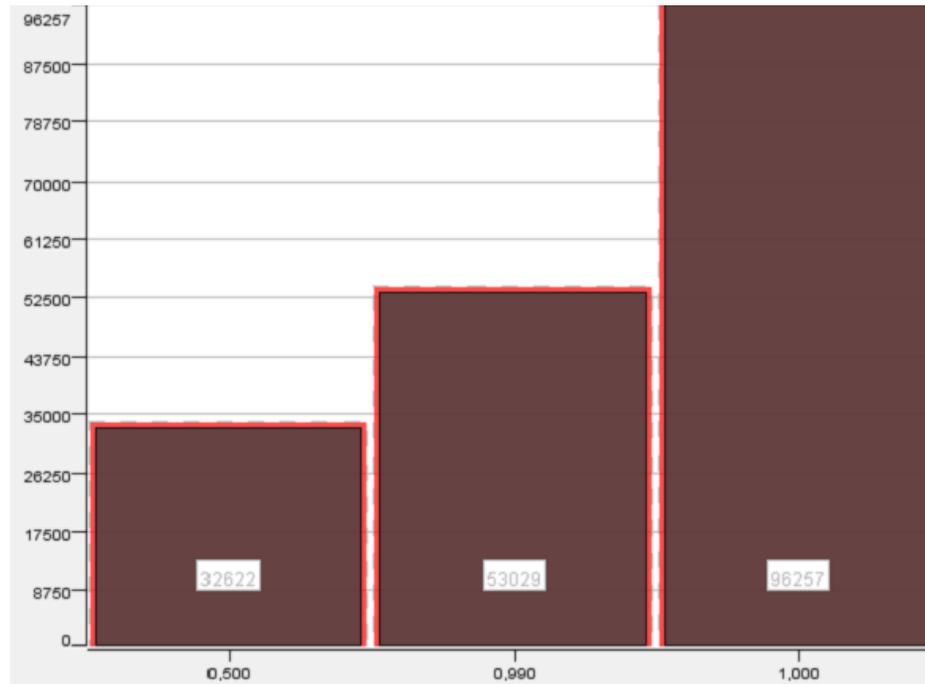


Figura 4.15: Histograma número de vehículos por la exposición al riesgo observada - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura 4.15, tipo de variable: numérica continua, en primera instancia se puede seleccionar la variable para realizar los otros estudios antes de la agrupación transversal, además, se observa que la clase con los factores de exposición al riesgo entre 0,991 y 1, contienen la mayoría de los vehículos.

4. ANÁLISIS E INTERPRETACIÓN DE RESULTADOS

Análisis de correlación de las siguientes variables:

1. Modalidad,
2. Valor del vehículo,
3. Año,
4. Número de ocupante,
5. Tipo de vehículo,
6. Estado civil,
7. Sexo,
8. Habitante por kilometro cuadrado,
9. Marca modelo,
10. Edad del conductor,
11. Tiempo asegurado,
12. Número de siniestros,
13. Monto Total Facturado Siniestro.

Revisando dichas variables, no existe una marcada correlación entre las anteriores variables, por lo que se debe realizar un estudio más exhaustivo.

Presentación de gráfico de dispersión:

A continuación, se presentan gráficos de dispersión con respecto a variables compuestas que son: Tasa natural de riesgo posterior, tasa de uso posterior. Estas variables son calculadas a través de las siguientes variables, para Tasa natural de riesgo posterior se utiliza el Monto Total Facturado de Siniestros entre la Suma Total de los valores de los vehículos y para la tasa de uso posterior se utiliza el Número de siniestros entre la suma total del factor de exposición de riesgo.

1. Marca modelo.

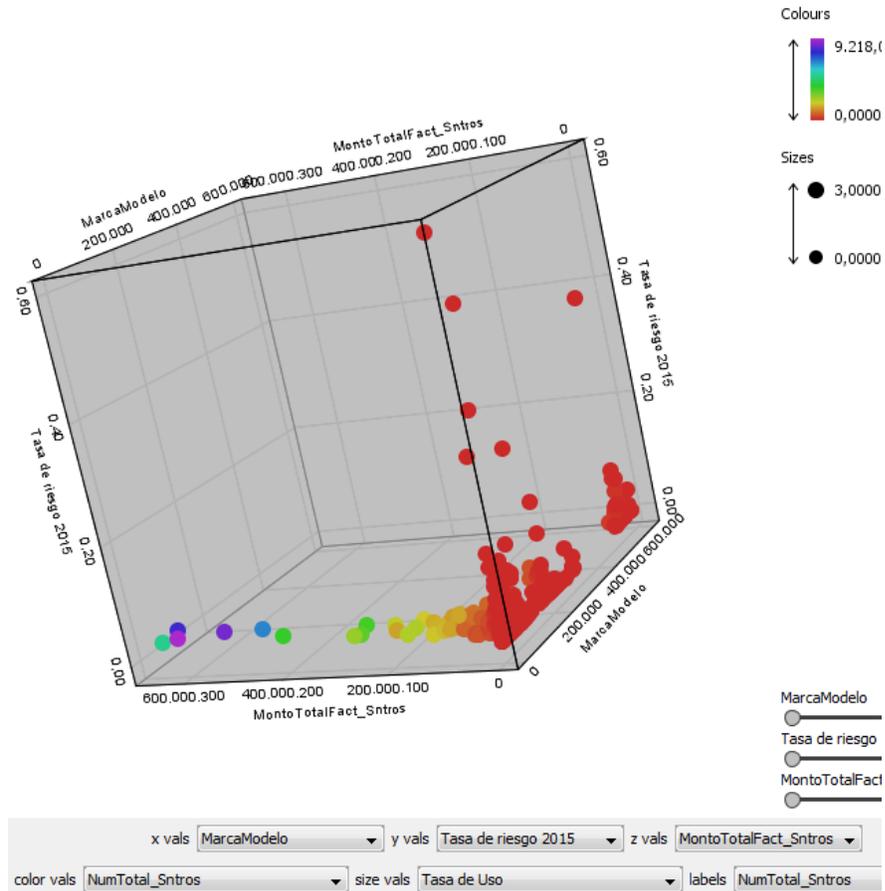


Figura 4.16: Dispersión tasa de riesgo, monto total facturado, número de siniestros y tasa de uso por cada marca modelo - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura 4.16, en primera instancia se observa que a mayor cantidad de siniestros y de exposición de cada Marca Modelo la tasa natural de riesgo es robusta estando al rededor del 0,02.

4. ANÁLISIS E INTERPRETACIÓN DE RESULTADOS

2. Valor del vehículo.

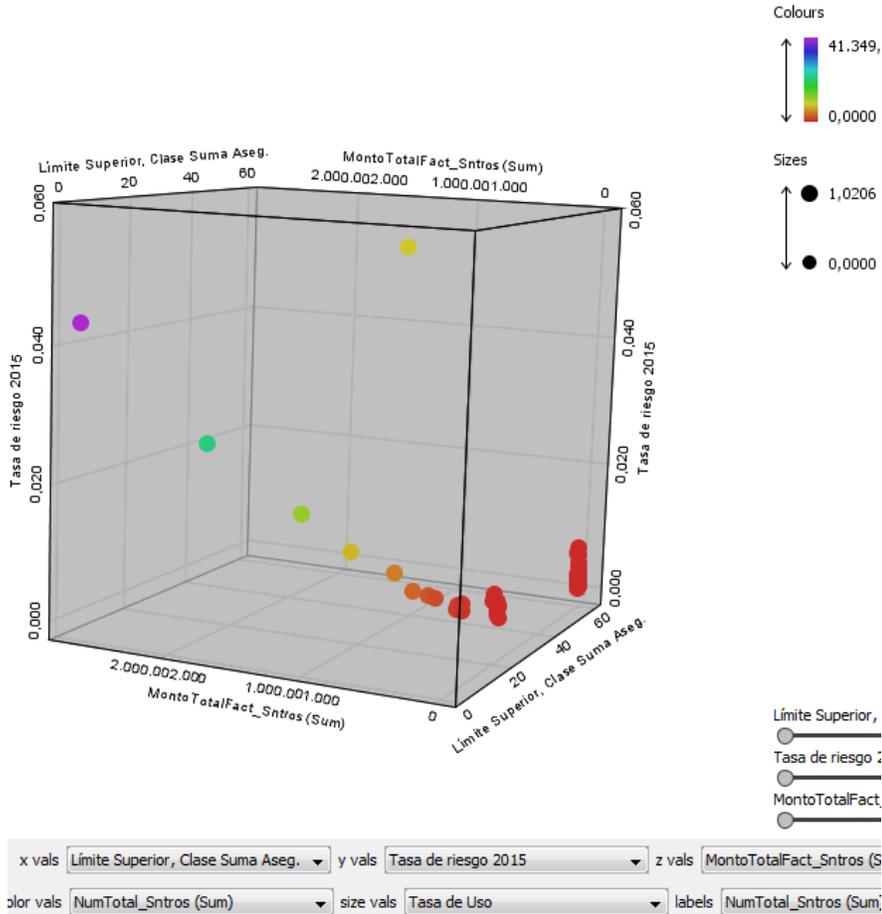


Figura 4.17: Dispersión tasa de riesgo, monto total facturado, número de siniestro y tasa de uso por cada valor del vehículo - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura 4.17, en primera instancia se observa que a mayor cantidad de siniestros y de exposición de cada valor del vehículo la tasa natural de riesgo es robusta estando al rededor del 0,02.

3. Estado civil.

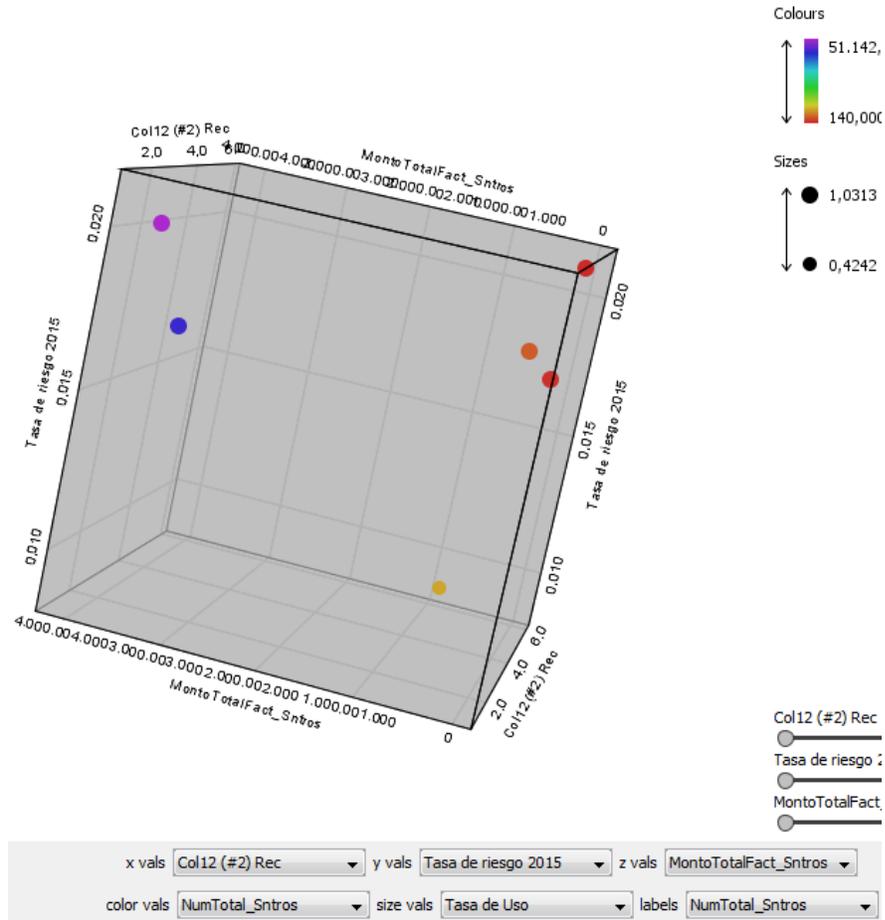


Figura 4.18: Dispersión tasa de riesgo, monto total facturado, número de siniestro y tasa de uso por cada estado civil - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura 4.18, en primera instancia se observa que a mayor cantidad de siniestros y de exposición de cada estado civil la tasa natural de riesgo es robusta estando al rededor del 0,02.

4. ANÁLISIS E INTERPRETACIÓN DE RESULTADOS

4. Tipo de sexo del conductor habitual.

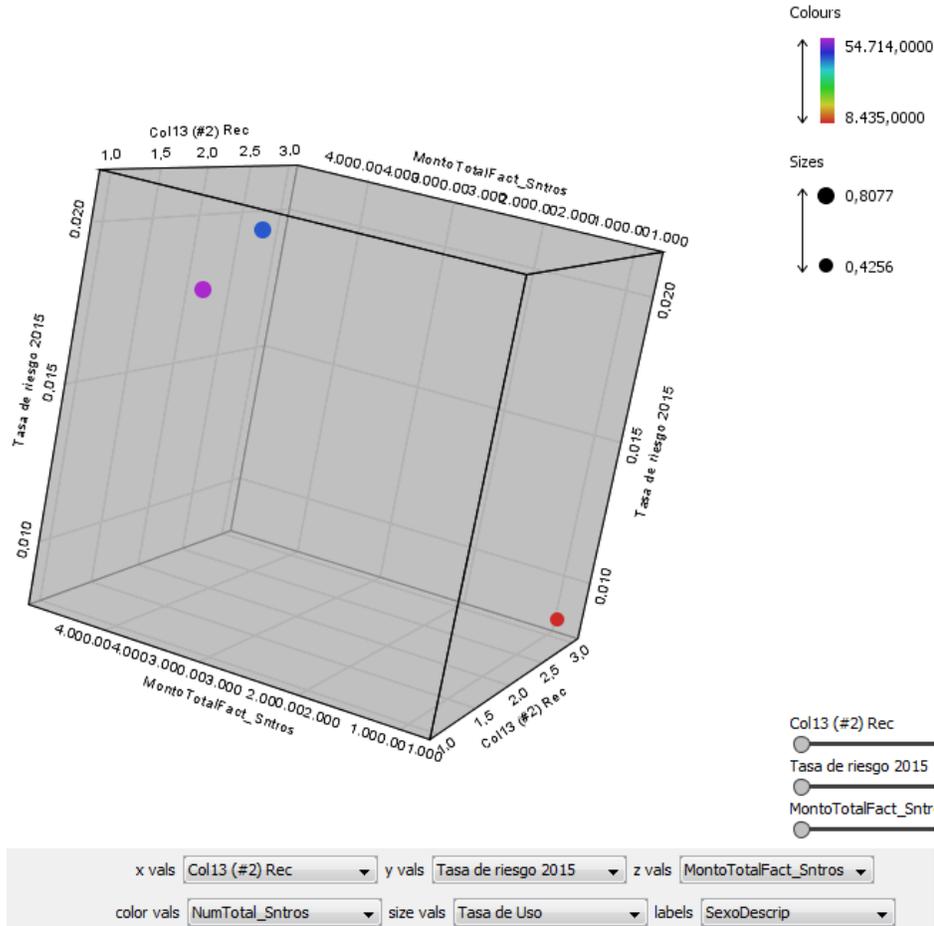


Figura 4.19: Dispersión tasa de riesgo, monto total facturado, número de siniestro y tasa de uso por cada sexo - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura 4.19, en primera instancia se observa que a mayor cantidad de siniestros y de exposición de cada sexo la tasa natural de riesgo es contundente estando al rededor del 0,02.

5. Modalidad.

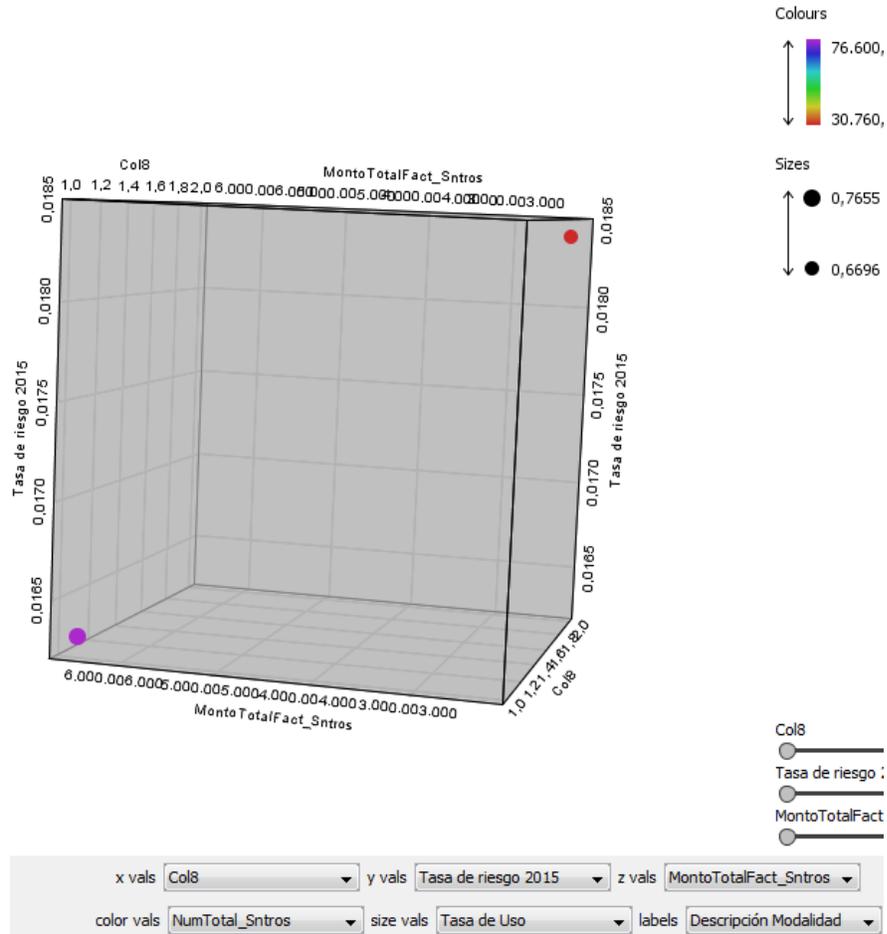


Figura 4.20: Dispersión tasa de riesgo, monto total facturado, número de siniestro y tasa de uso por cada modalidad - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura 4.20, en primera instancia se observa que a mayor cantidad de siniestros y de exposición de cada modalidad la tasa natural de riesgo es robusta estando al rededor del 0,02.

4. ANÁLISIS E INTERPRETACIÓN DE RESULTADOS

6. Lapso de tiempo asegurado.

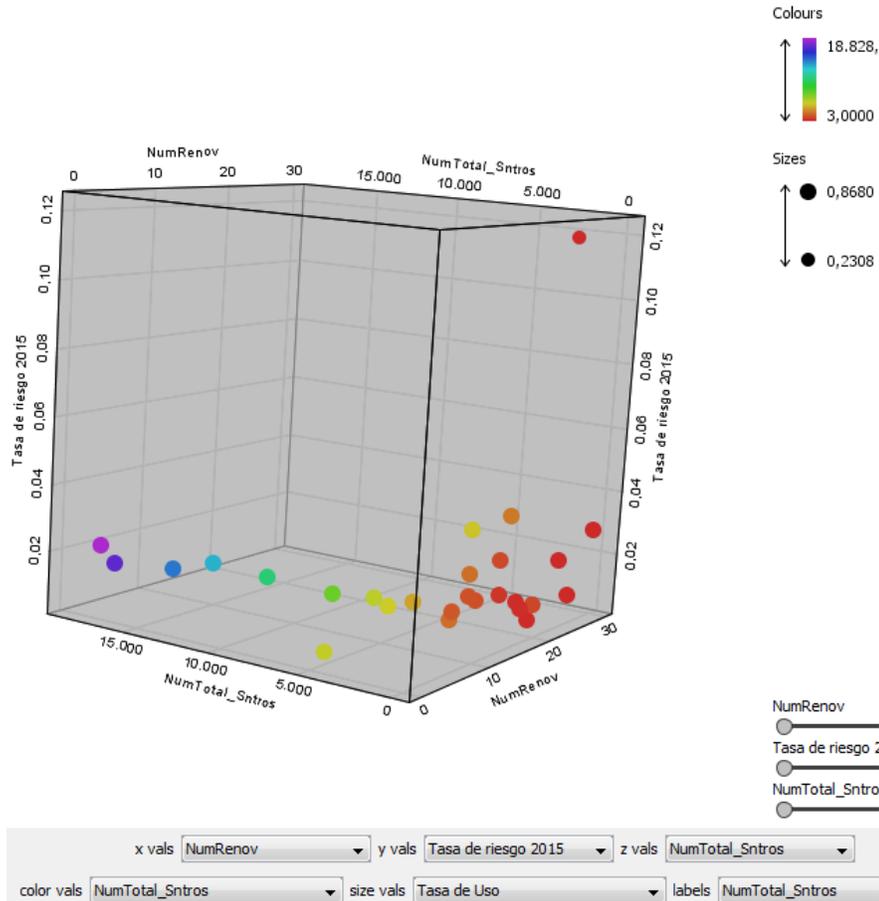


Figura 4.21: Dispersión tasa de riesgo, monto total facturado, número de siniestro y tasa de uso por cada lapso de tiempo asegurado del vehículo - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura 4.21, en primera instancia se observa que a mayor cantidad de siniestros y de exposición de cada “lapso de tiempo asegurado del vehículo” la tasa natural de riesgo es contundente estando al rededor del 0,02.

7. Edad del conductor.

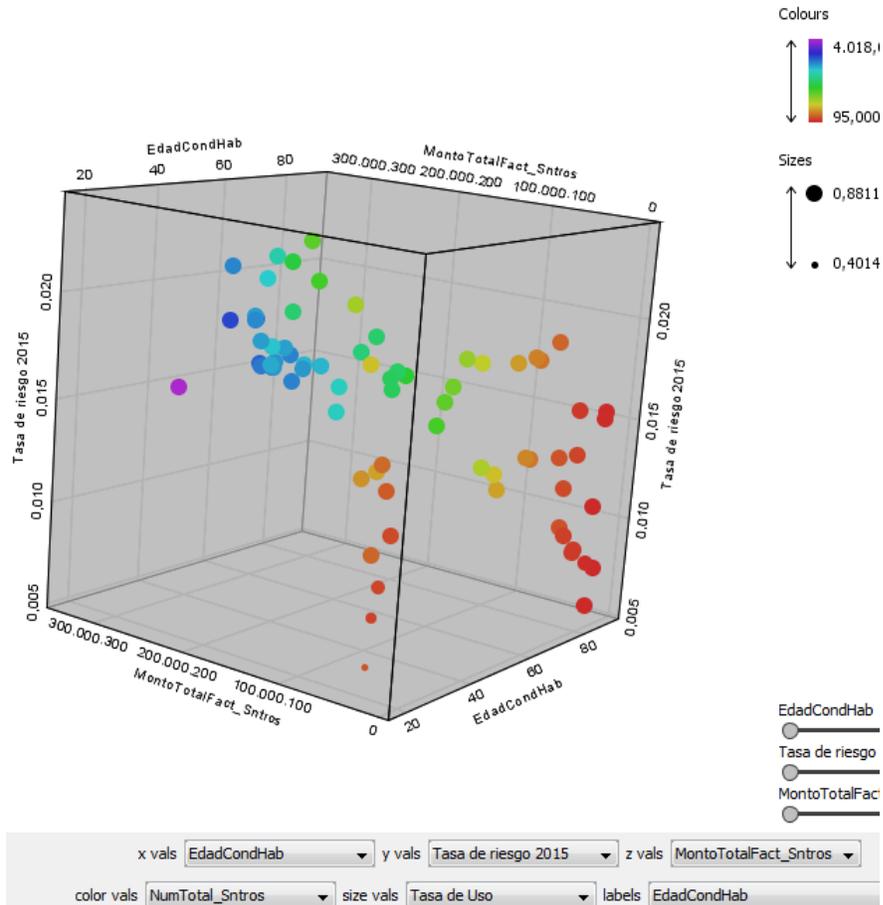


Figura 4.22: Dispersión tasa de riesgo, monto total facturado, número de siniestro y tasa de uso por cada edad del conductor del vehículo asegurado - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura 4.22, en primera instancia se observa que a mayor cantidad de siniestros y de exposición de cada edad del conductor la tasa natural de riesgo es contundente estando al rededor del 0,02.

4. ANÁLISIS E INTERPRETACIÓN DE RESULTADOS

8. Exposición al riesgo.

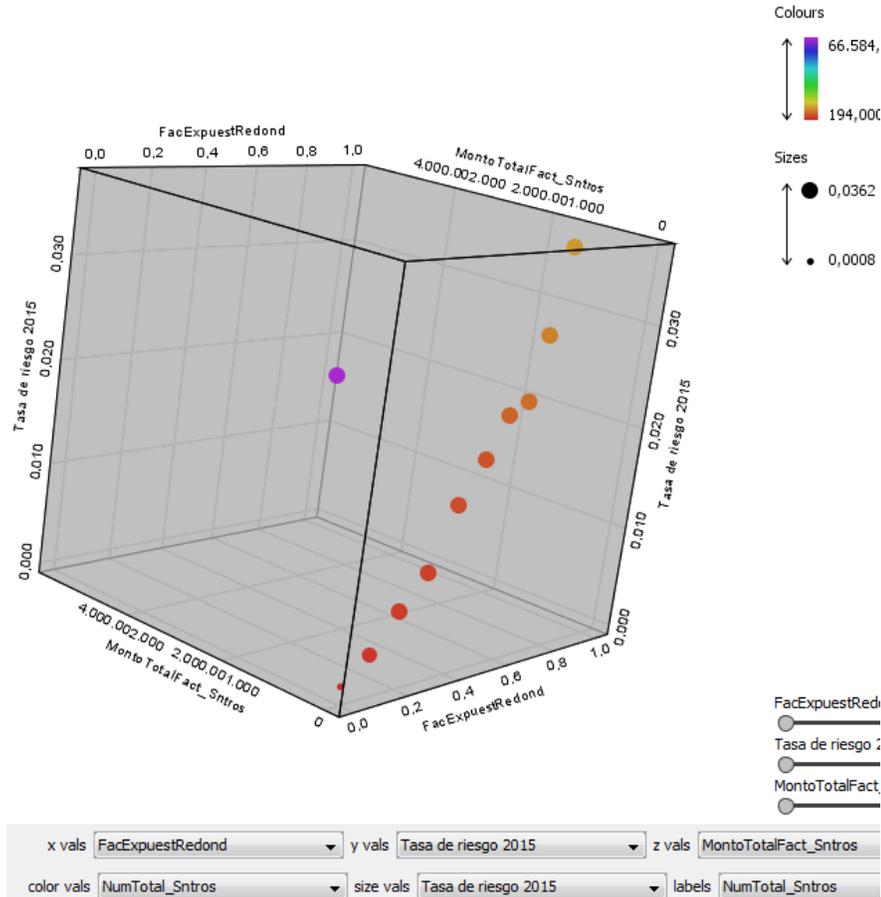


Figura 4.23: Dispersión tasa de riesgo, monto total facturado, número de siniestro y tasa de uso por cada tiempo de exposición del vehículo - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura 4.23, en primera instancia se observa que a mayor cantidad de siniestros de cada exposición la tasa natural de riesgo es contundente estando al rededor del 0,02.

9. Año de construcción de vehículo asegurado.

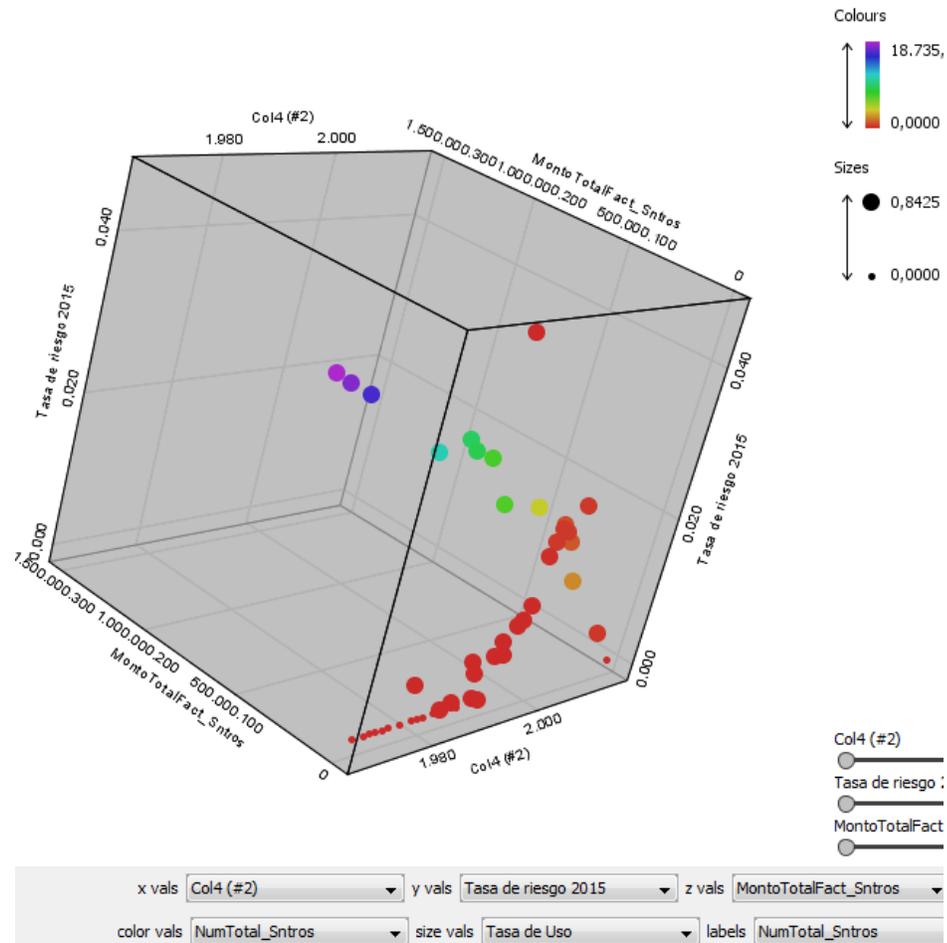


Figura 4.24: Dispersión tasa de riesgo, monto total facturado, número de siniestro y tasa de uso por cada año de construcción del vehículo - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura 4.24, en primera instancia se observa que a mayor cantidad de siniestros y de exposición de cada año la tasa natural de riesgo es contundente estando al rededor del 0,02.

4.1. Conglomerado a través del análisis de componentes principales

Al analizar los resultados dados al aplicar la técnica de componentes principales se observó poca covarianza entre las variables “año”, “edad del conductor”, “número total de siniestros de las marcas modelos” y “monto total facturado por marca modelo”, lo cual se muestra en el siguiente cuadro:

Row ID	D Col4 (#2)	D EdadCo...	D NumTot...	D MontoT...
Col4 (#2)	1.046	-0.122	0.049	0.04
EdadCondHab	-0.122	1.036	-0.015	-0.013
NumTotal_Snt...	0.049	-0.015	0.998	0.397
MontoTotalFa...	0.04	-0.013	0.397	0.977

Figura 4.25: Covarianza de las principales variables - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

4.1 Conglomerado a través del análisis de componentes principales

De igual forma se realizaron los cálculos de los autovalores, dando como resultado que es significativo mantener cuatro (4) grupos por los primeros autovalores que contienen más del 60 por ciento de la información registrada, los mismos autovalores se grafican a continuación:

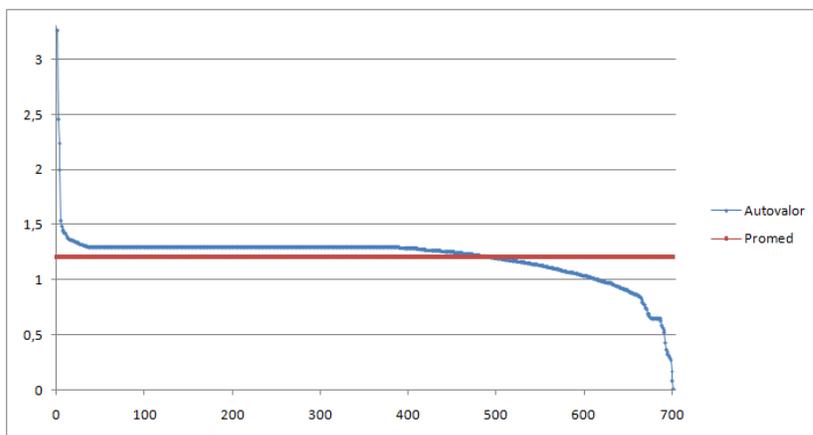


Figura 4.26: Autovalores del análisis de componentes principales - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

4.2. Conglomerado a través de k medias

1. Se realizó grupos a través de k medias con todas las variables conjuntamente, estas variables son originales sin efectuar modificaciones sobre ellas. Conglomerado con k=7 medias de todas las variables originales.

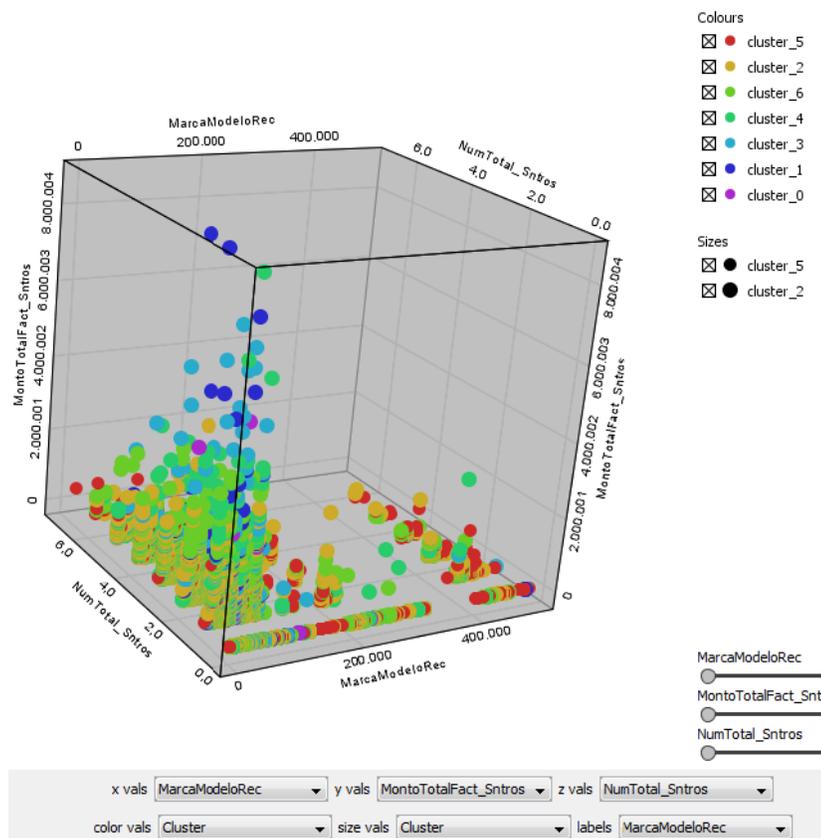


Figura 4.27: Conglomerado a través de k medias, con k=7, todas las variables originales sin modificación - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura 4.27, los grupos no son claros al ser graficados con la variable de monto facturado de siniestros. Todos los demás agrupaciones están en el apendice D.

2. Se realizó grupos a través de k medias con las variables numéricas normalizadas, con las variables categóricas transformadas en variables ficticias de valores 0 y 1, sin excluir ninguna de las variables estudiadas.

4. ANÁLISIS E INTERPRETACIÓN DE RESULTADOS

Conglomerado con $K=7$ medias, variables numéricas normalizadas, variables categóricas transformadas en ficticias, sin exclusión.

Con respecto a la figura 4.28, los grupos se dividen entre los que tienen siniestros y monto facturado de siniestros y los demás grupos con los que no tienen siniestros. Todos los demás agrupaciones están en el apéndice 5.

3. Se realizó grupos a través de k medias con las variables numéricas normalizadas, con las variables categóricas transformadas en variables ficticias de valores 0 y 1 normalizadas sin excluir variables.

Conglomerado con $K=7$ medias, variables numéricas normalizadas, variables categóricas transformadas en ficticias normalizadas sin excluir variables.

Con respecto a la figura 4.29, se agrupan por varios factores donde se aprecia que la mejor agrupación está asociado con las "marcas modelos".

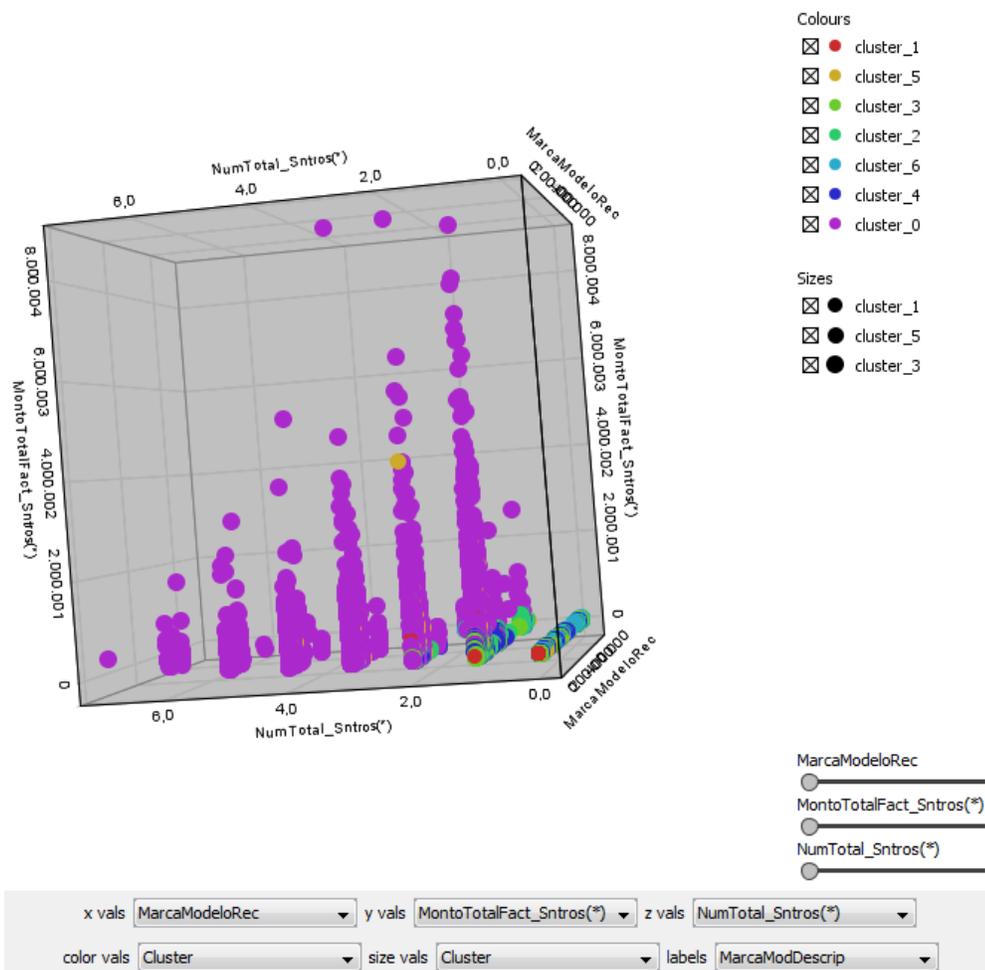


Figura 4.28: Conglomerado a través de k medias, con K=7 medias, variables numéricas normalizadas, variables categóricas transformadas en ficticias sin excluir variables - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

4. ANÁLISIS E INTERPRETACIÓN DE RESULTADOS

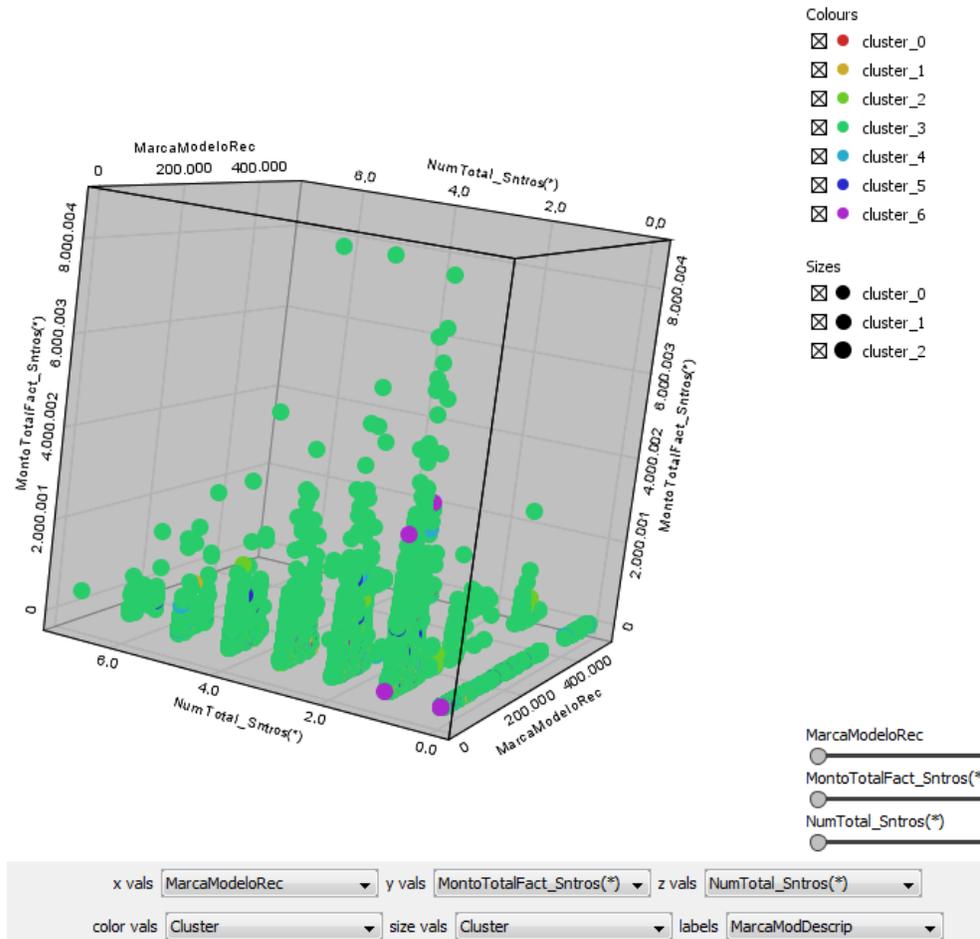


Figura 4.29: Conglomerado a través de k medias, con K=7 medias, variables numéricas normalizadas, variables categóricas transformadas en ficticias normalizadas sin excluir variables - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

4. Conglomerado a través de bagging de weka 3.7.
Se realizó los siguientes grupos.

```

Bagging
All the Base classifiers:

REPtree
-----
Col191 < 250062.5
| Col142 < 2010.5
| | Col142 < 2007.5
| | | Col191 < 151187.5
| | | | MontoTotalFact_Sntros < 7421.65 : 75766.56 (96/5411087603.23) [39/5946576928.67]
| | | | MontoTotalFact_Sntros >= 7421.65 : 427002.5 (5/35909490405.76) [3/53232644008.64]
| | | | Col191 >= 151187.5
| | | | | Col162 < 2.5 : 75359.08 (17/2475440163.2) [9/3153100381.8]
| | | | | Col162 >= 2.5 : 33503.88 (446/1284702965.86) [227/833526129.22]
| | | | Col142 >= 2007.5
| | | | | Col191 < 164225
| | | | | Col191 < 161500
| | | | | | Col191 < 31920
| | | | | | Col142 < 2008.5 : 77240.8 (4/9336018.19) [1/28007910.06]
| | | | | | Col142 >= 2008.5 : 559006.86 (6/2) [1/1]
| | | | | | Col191 >= 31920
| | | | | | | Col191 < 105850
| | | | | | | | NumTotal_Sntros < 0.5
| | | | | | | | | Col132 < 1.5 : 546001 (2/0) [0/0]
| | | | | | | | | Col132 >= 1.5
| | | | | | | | | | Col142 < 2009.5 : 74809.95 (44/336940103.51) [17/89884885.76]
| | | | | | | | | | Col142 >= 2009.5 : 388184.33 (2/58798005289) [4/52886325280.5]
| | | | | | | | | | NumTotal_Sntros >= 0.5 : 539001.17 (6/244988333.47) [0/0]
| | | | | | | | Col191 >= 105850
| | | | | | | | | EdadCondHab < 45.5 : 46993.86 (78/4600519382.09) [33/9021156465.98]
| | | | | | | | | EdadCondHab >= 45.5
| | | | | | | | | | EdadCondHab < 53.5 : 403519.5 (4/54143760490.69) [2/33988717863.56]
| | | | | | | | | | EdadCondHab >= 53.5
| | | | | | | | | | | Col191 < 135650 : 198002 (2/0) [1/0]
| | | | | | | | | | | Col191 >= 135650 : 22662.71 (6/621686835.56) [1/124337367.11]
| | | | | | | | Col191 >= 161500
| | | | | | | | | Col191 < 163437.5 : 531978.39 (27/0) [11/25816292055.36]
| | | | | | | | | Col191 >= 163437.5
| | | | | | | | | | Col142 < 2009 : 13105 (6/0) [1/0]
| | | | | | | | | | Col142 >= 2009 : 412777.75 (5/45436674017.44) [3/48155011026.16]
| | | | | | | | Col191 >= 164225
| | | | | | | | | Col191 < 208093
| | | | | | | | | | EdadCondHab < 26.5
| | | | | | | | | | Col191 < 179600 : 545003 (2/0) [0/0]
| | | | | | | | | | Col191 >= 179600 : 13105 (3/0) [1/0]
| | | | | | | | | | EdadCondHab >= 26.5 : 43971.95 (153/12909633607.66) [55/1624180286.48]
| | | | | | | | Col191 >= 208093
| | | | | | | | | Col191 < 221143
| | | | | | | | | | NumRenov < 3.5 : 21755.08 (8/294821239) [4/547434787.75]

```

Figura 4.30: Conglomerado a través de bagging de weka 3.7 - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura 4.30, en primera instancia se realizaron 20.700 árboles, lo que indican que la marca modelo es el factor de riesgo que mejor explica a los grupos.

4. ANÁLISIS E INTERPRETACIÓN DE RESULTADOS

5. Conglomerado a través de bosque aleatorio.
Se realizó los siguientes grupos.

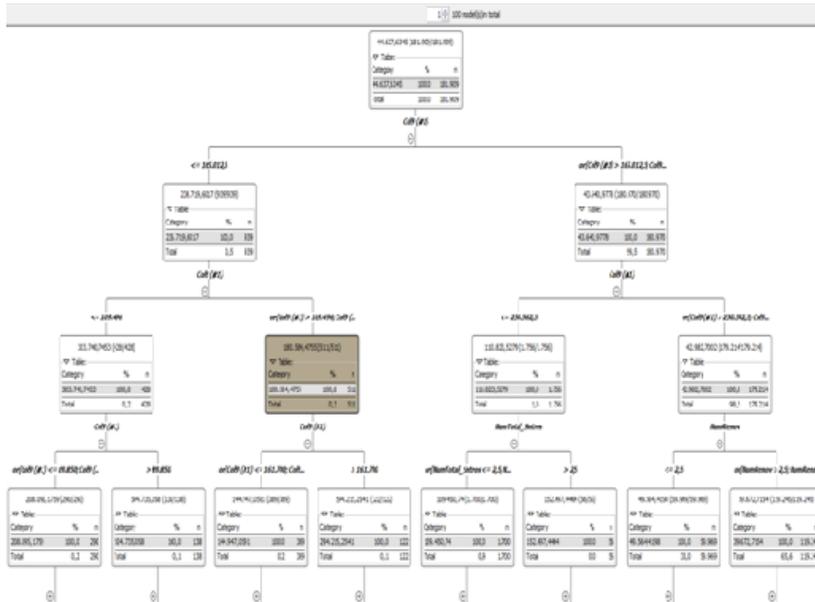


Figura 4.31: Conglomerado a través del aprendizaje del bosque aleatorio - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura 4.31, se realizaron separaciones aleatorias donde identifican la marca modelo como el factor de riesgo a separar.

6. Conglomerado con K medias por marca modelo solamente.

Al analizar los resultados anteriores se determina realizar una agrupación con solo la “marca modelo” observando los montos facturados total y el número de siniestros ocurridos a cada “marca modelo”.

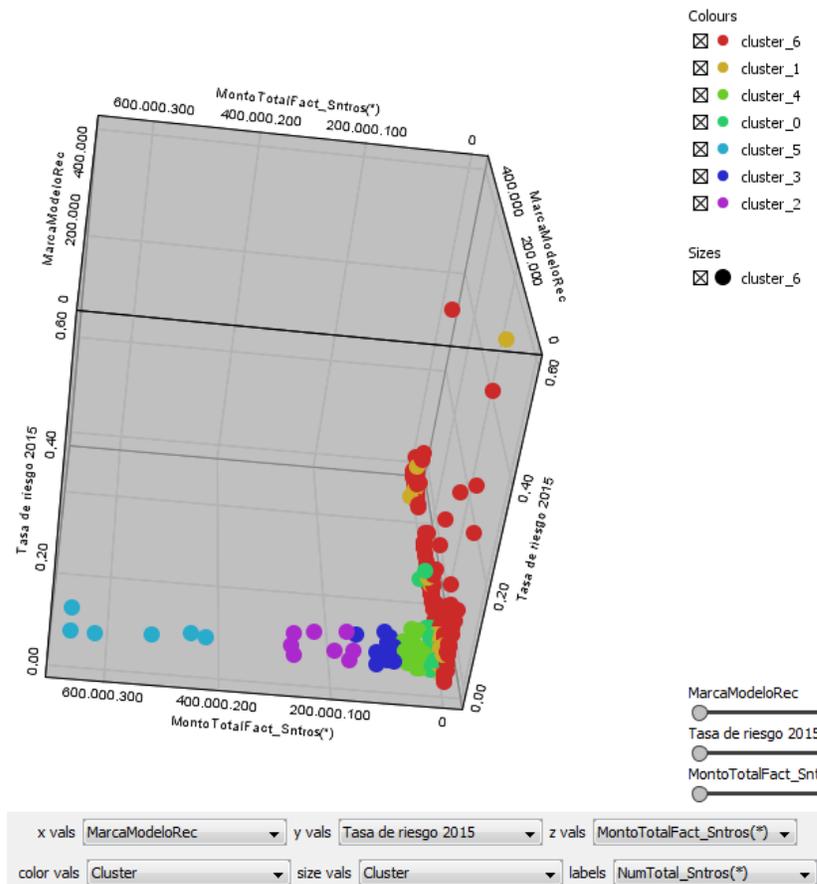


Figura 4.32: Conglomerado a través de k medias, con k=7 con solo la marca modelo observando el número de siniestro y el monto facturado total - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura 4.32, se realizaron separaciones bien identificadas para las “marcas modelos” que contienen en total más de Bs. 400.000.000 que estan en el grupo cinco (5) y así sucesivamente para cada grupo, lo anterior se detalla en el apendice j.

4.3. Conglomerado a través del análisis discriminante de variables normalizadas

Al culminarse el cálculo con la técnica del análisis discriminante con una agrupación previamente obtenida por la k medias solo “marca modelo” con número de siniestros y monto total facturado de siniestros, se obtuvo varios resultados que se muestran en el apéndice k, del cual el primer grupo tienen el siguiente gráfico con el valor del parámetro en el eje de las equis(x’s).

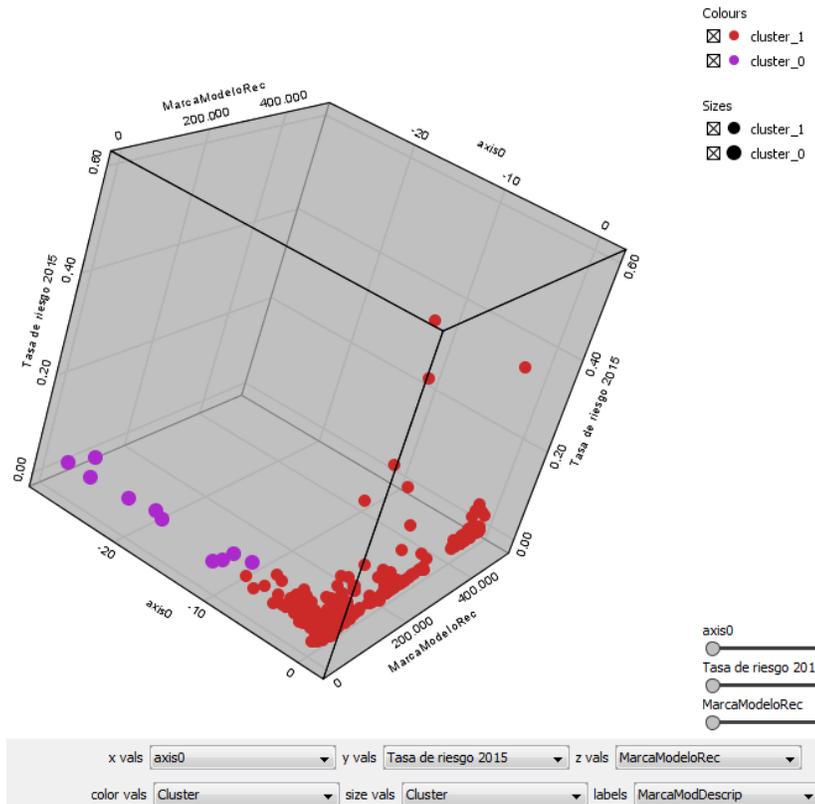


Figura 4.33: Conglomerado a través del análisis discriminante, con K=2 solo a la variables marca modelo - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

En el mismo se muestra la importancia de la separación por montos total facturado de siniestros de las marcas modelos de los vehículos asegurados en el 2015.

4.4. Estudio longitudinal, conglomerado a través del análisis de componentes principales

Después de utilizar el paquete llamado “FactorMiner” de “R” en “R studio” conjuntamente con “latex” con “knitr”, se obtuvo que con tres (3) grupos se visualiza la totalidad de la información de todos los vehículos asegurados en el año 2015, se muestra gráfico con una muestra del segundo y tercer grupo, para el primer grupo las curvas estan por debajo de las anteriores.

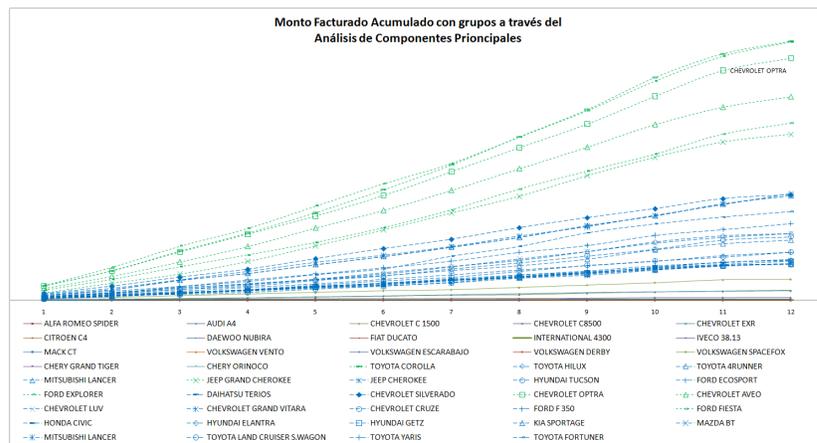


Figura 4.34: Estudio longitudinal, conglomerado a través del análisis de componentes principales, gráfico de las curvas de las marcas modelos por monto total facturado acumulado y agrupado - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Los demás resultados se aprecian en el apendice “L”.

4.6. Determinación de la estimación de las pérdidas agregadas

4.6.1. Sin aplicar factor de deflactación estimado a los montos de siniestros individuales

Con el grupo de mayor número de siniestros y monto total facturado de siniestro, este grupo se determinó en la última aglomeración k medias, se procedió a calcular, suponiendo sin inflación monetaria por efectos prácticos, la máxima pérdida probable, obteniendo el valor en riesgo del 98,3%.

4. ANÁLISIS E INTERPRETACIÓN DE RESULTADOS

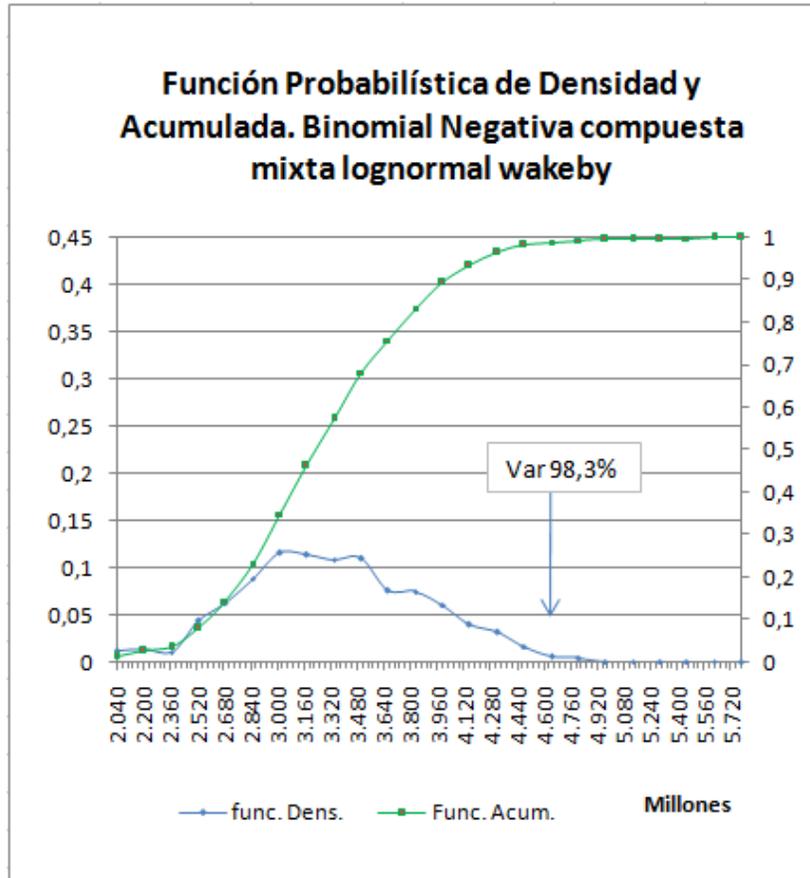


Figura 4.36: Estimación de las pérdidas agregadas, máxima pérdida probable, función binomial negativa compuesta lognormal mixta wakeby - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Las demás realizaciones de la obtención de la asociación de las probabilidades se muestran en el apéndice “N”.

4.6.2. Aplicando factor de deflactación, estimado, a los montos de siniestros individuales

Con el grupo de mayor número de siniestros y monto total facturado de siniestros se procedió a calcular, acumulando los montos individuales de siniestro al factor inflacionario estimado, la máxima pérdida probable, obteniendo el valor en riesgo al 99,8 % de confianza de 21.718 millones de Bolívares, lo cual se incrementa considerablemente a lo calculado sin incremento. Adicionalmente, se observó un cambio de la función de probabilidad asociada, el cual resultó “binomial negativa compuesta mixta lognormal con valor extremo general”.

A continuación, se muestra el cuadro con los factores de incremento anual, donde se indica el índice nacional de precios hasta el 31 de diciembre del 2015 y la estimación de dicho índice a través de una regresión multivariante polinómica realizada en la Institución.

4. ANÁLISIS E INTERPRETACIÓN DE RESULTADOS

Tabla 4.1: Índice nacional de precios año 2015

Mes	Año	INPC	Factor incremento anual
01	2015	905	
02	2015	949	
03	2015	1.000	
04	2015	1.064	
05	2015	1.149	
06	2015	1.262	
07	2015	1.398	
08	2015	1.571	
09	2015	1.752	
10	2015	1.951	
11	2015	2.169	
12	2015	2.358	

Fuente: BCV 2015

Tabla 4.2: Estimación del índice nacional de precios año 2016

Mes	Año	INPC	Factor incremento anual
01	2016	2.627	2,902762431
02	2016	2.878	3,032665964
03	2016	3.125	3,125000000
04	2016	3.419	3,213345865
05	2016	3.944	3,432550044
06	2016	4.491	3,558637084
07	2016	5.177	3,703147353
08	2016	6.098	3,881604074
09	2016	7.667	4,376141553
10	2016	10.189	5,222450026
11	2016	11.430	5,269709544
12	2016	12.823	5,438083121

Fuente: Est. A. Colmenares 2016

4. ANÁLISIS E INTERPRETACIÓN DE RESULTADOS

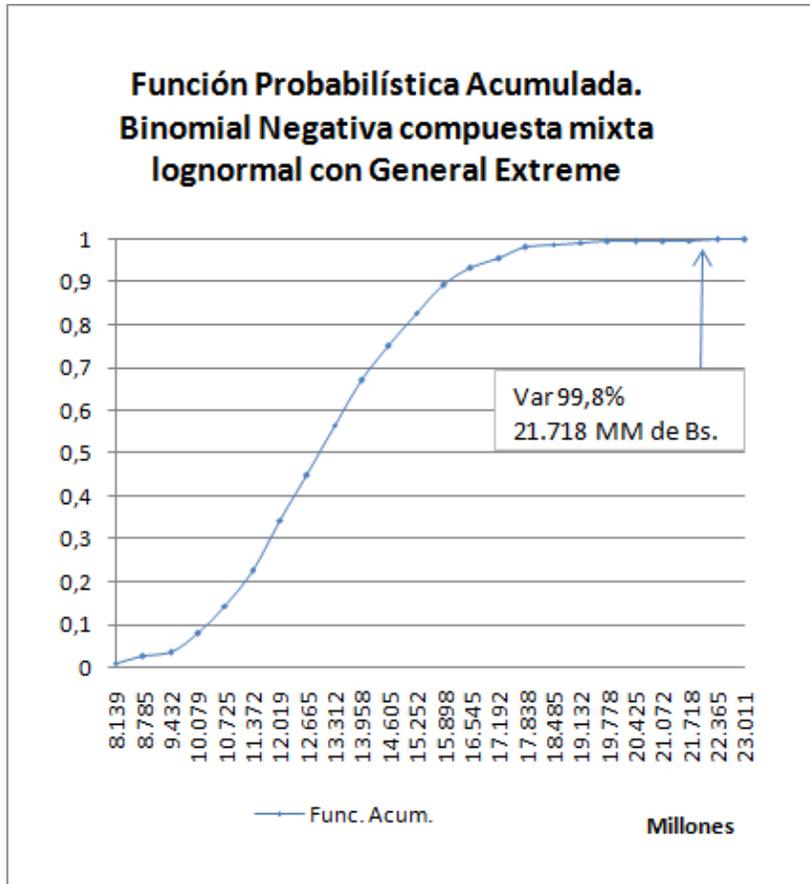


Figura 4.37: Estimación de las pérdidas agregadas, máxima pérdida probable, función de probabilidad acumulada binomial negativa compuesta lognormal mixta valor extremo general, valor en riesgo - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

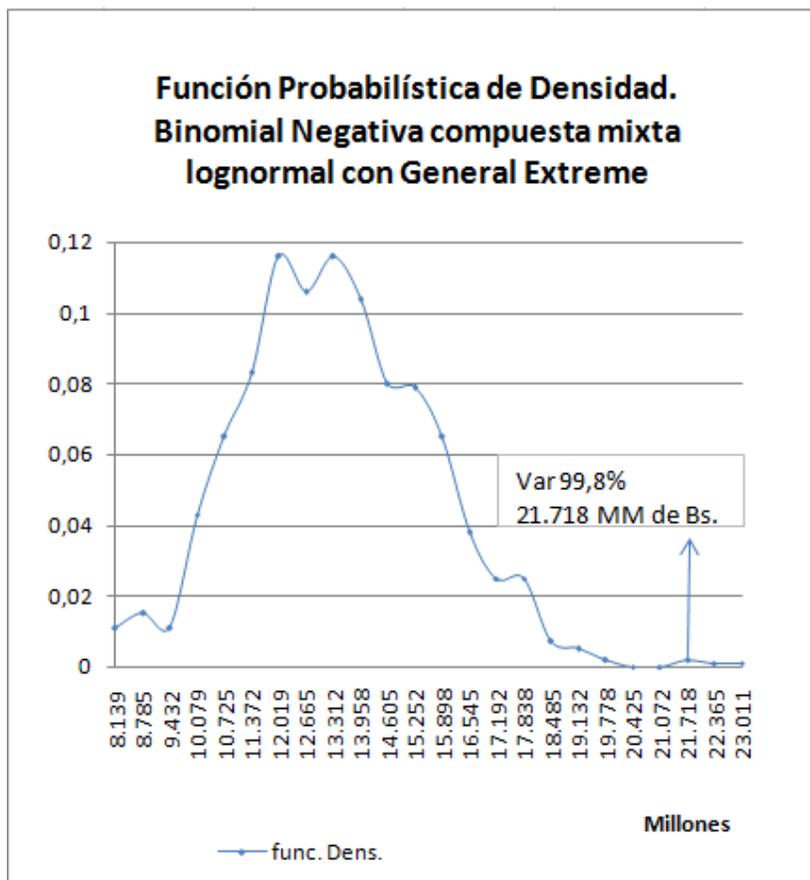


Figura 4.38: Estimación de las pérdidas agregadas, máxima pérdida probable, función de probabilidad de densidad binomial negativa compuesta lognormal mixta valor extremo general, valor en riesgo - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Conclusiones y recomendaciones

Después de realizar los análisis respectivos se realiza las siguientes conclusiones de la presente pasantía.

5.1. Conclusiones

1. 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.
2. El análisis de k medias permitió una agrupación de los vehículos asegurados de acuerdo al patrón que ellos forman en los factores de riesgo: valor del vehículo, marca modelo, monto de la factura total del siniestro de los registros obtenidos de seis (6) aseguradores del mercado; en todos sus contratos de seguros en el ramo de casco de automóvil, es decir, de la cobertura pérdida parcial. Al fijar los conglomerados, $k = 2$, se aprecia que en el primer grupo, el de mayor cantidad de vehículos, entorno al cual se agrupan por ejemplo, los auomóviles de marca modelo: Chevrolet Aveo, Chevrolet Optra, Toyota Yaris; de edad de los conduc-

5. CONCLUSIONES Y RECOMENDACIONES

tores entre 30 a 35 años; de sexo: masculino; describen una tendencia marcada al incremento de las pérdidas. Con los otros factores se observa estabilidad. Más interesante aún resulta el hecho de que los vehículos, que en líneas generales siguen el patrón de la primera agrupación con más número de vehículos, tienden a presentar fluctuaciones en los montos de siniestros, cuyo comportamiento debe explicarse por el alto riesgo de ocurrencia en el factor marca modelo. En efecto, a manera de ilustración el comportamiento de los siniestros de los vehículos Chevrolet Optra, Toyota Yaris, Chevrolet Aveo, queda en líneas generales descrito por el primer grupo. En el tipo de factor: femenino en sexo, edad de 20 a 30 años, presentó niveles bajos de montos facturados y en los factores o tipo de factor: masculino en sexo, edad de 30 a 35 años, obtuvo un incremento considerable. Lo que se observa que las variables monto facturado del siniestro, valor del vehículo y la marca modelo son las que arrojan mayor información.

3. El modelo de ocurrencia, distribuido como binomial negativa, arroja como tasa media de ocurrencia en el período bajo estudio de 40.678 eventos esperados anual y un máximo de ocurrencia probable de 70.510 eventos con un nivel de confianza de 98,9%.
4. El modelo de severidad de los eventos arroja como función probabilística a la log-normal tres (3) parámetros: $\sigma = 1,35683888$, $\mu = 11,7628997$, $v = -677,98995$ en el rango percentilar (0 a 99,97) y la función probailística valor extremo general (modelo Fréchet y Weibull) tres (3) parámetros: $K = -0,282470569$, $\sigma = 5.157.965,38$, $\mu = 18.066.046,90$ en el rango percentilar (99,98 a 99,990) en el restante tiende 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 de 322.478,70 Bolívares y una máxima pérdida probable por evento de 35.268.820,48 Bolívares con un nivel de confianza del 99,995%.
5. El modelo de pérdidas agregadas arroja el "valor a riesgo" para la máxima pérdida probable de 21.718.000.000 Bolívares con un nivel de confianza del 99,8%.

5.2. Recomendaciones

Esta investigación constituye sólo un ensayo para que sirva de referencia en el desarrollo de la gerencia de riesgo de las compañías de seguros y de observatorio para los organismos de control, en el marco del cálculo integral de riesgo en las compañías de seguros. Específicamente, el cálculo de las pérdidas monetaria del riesgo operacional

de acuerdo a los estatutos internacionales sobre la materia, en el comité de Basilea II (1995), según Basle committee on banking supervision (1995) (20).

A continuación se mencionan algunos elementos que podrían ser útiles y seguirían esencialmente los mismos procedimientos utilizados en esta investigación:

1. los análisis de k medias, componentes principales, discriminantes, particiones de arbol, particiones de bosque aleatorio, k medias para datos longitudinales con distancia sugerida por Chiodi y curvas principales, debería seguir utilizandose para obtener mapeos de riesgo de los cambios en las pérdidas por eventos que estan asegurados en el contrato de casco del automóvil. Además, procesar los métodos de conglomerados para varios años y realizar análisis evolutivo.
2. Con respecto al modelo de pérdidas agregadas, debería desarrollarse para todos los grupos que tengan el número de información mínima suficiente, identificando los grupos con poca información y determinando si son óptimos para realizar cálculos a través de la teoría de la credibilidad.
3. El modelo probabilístico de ocurrencia de eventos, deberá contemplarse su actualización para capturar la estacionariedad de la variable ocurrencia.
4. Para el modelo probabilístico de severidad de montos de facturas de reclamos, deberá contemplarse su actualización para obtener el cambio de estructura inflacionaria y de costo de la variable monto de factura por siniestro individual.
5. Estos procedimientos deberían replicarse en todas las empresas aseguradoras y en los organismos supervisores del país.

Paquetes utilizados

A.1. Paquetes estadísticos

En la investigación se utilizaron los siguientes programas:

Excel de Microsoft Office(r)

R project(r)

Knime(r)

EasyFit(r).

A.2. Paquetes para la transcripción

En el presente trabajo especial de grado se utilizó Latex junto con Rstudio a través de Knitr para conjugar transcripción del documento y procesamiento de algunos resultados.

Selección de datos obtenidos

B.1. Filtrado de los datos

1. Se incorpora vehículos menores de 5 Toneladas; se excluye las motos, los cuales se deben estudiar aparte.
2. La información de las “entidades federales de circulación” es incongruente con la realidad, por lo tanto se excluye de la investigación.
3. Se excluye la variable que identifica sí el tomador del seguro es una institución gubernamental o no, por tener menos del 1 por cada 100 vehículos.
4. El factor de exposición al riesgo será excluido en el estudio de la agrupación por sesgar los vehículos que no tienen siniestros.

Estudio por cada variable

C.1. Estudio descriptivo transversal

Se realizó un estudio descriptivo de las variables utilizadas para la realización de conglomerados de datos transversales.

C.1.1. Presentación de histogramas

C. ESTUDIO POR CADA VARIABLE

1. Año del vehículo, tipo de variable: numérica discreta, valores: desde 1980 hasta 2015

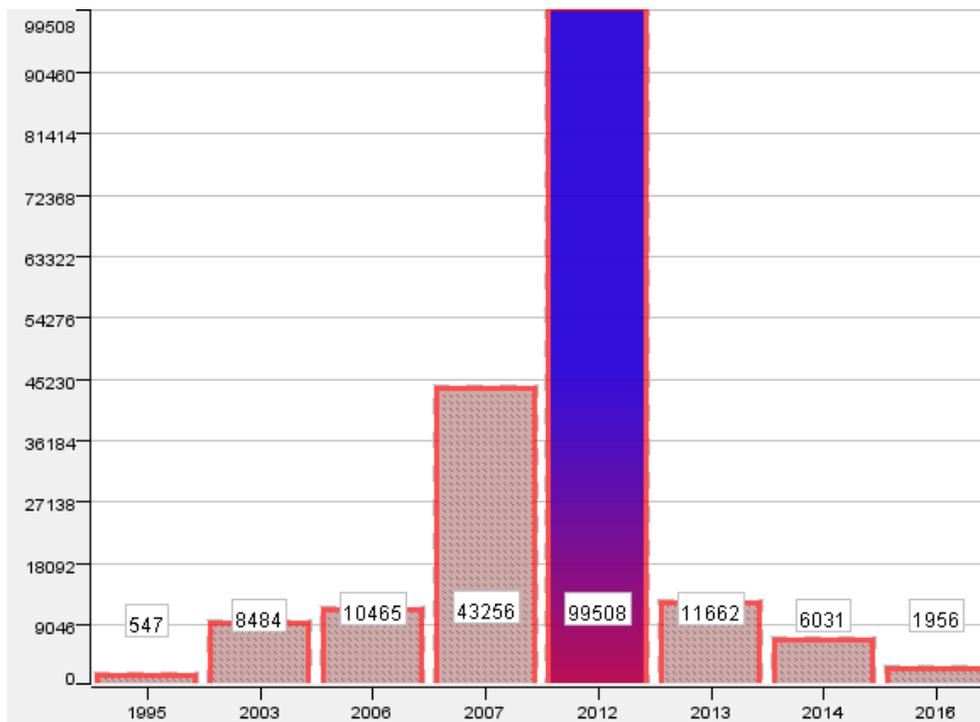


Figura C.1: Histograma número de vehículos por año de construcción - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura C.1, en primera instancia se puede seleccionar la variable para realizar los otros estudios antes de la agrupación transversal, además, se evidencia el bajo aseguramiento de carros construidos después del 2012.

2. Valor del Vehículo al 2015 expresado en millones de Bolívares, tipo de variable: numérica continua, valores: desde 1 hasta 51.

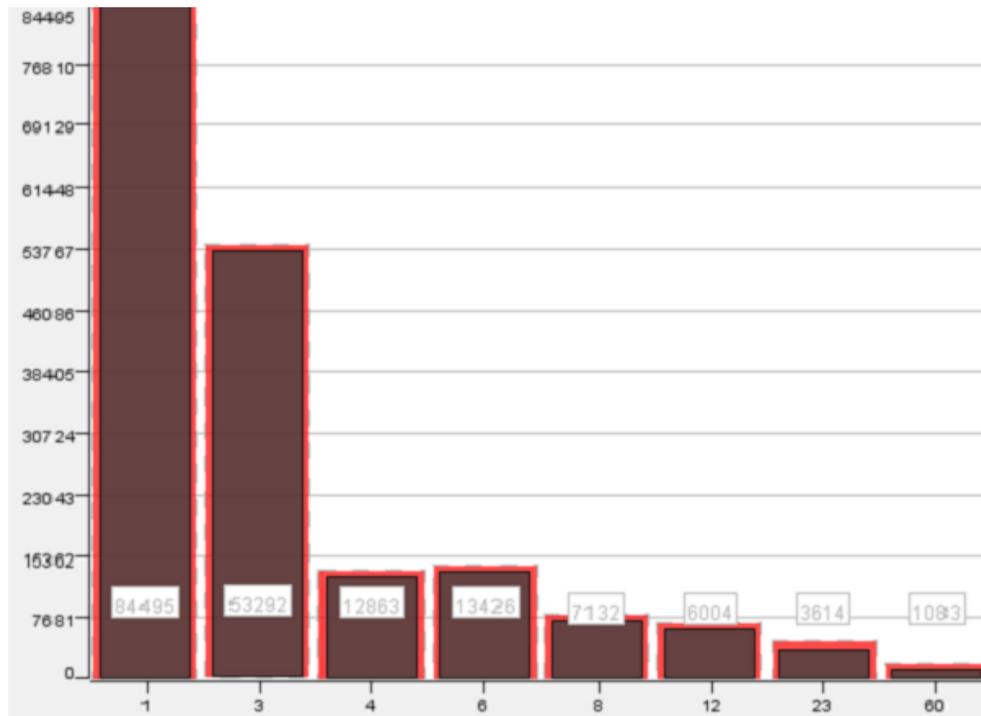


Figura C.2: Histograma número de vehículos por valor del vehículo - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura C.2, en primera instancia se puede seleccionar la variable para realizar los otros estudios antes de la agrupación transversal, además, se evidencia el alto aseguramiento de carros con valores por debajo de 3 millones de Bs.

C. ESTUDIO POR CADA VARIABLE

3. Monto total de siniestro del vehículo expresado en millones de Bolívares, tipo de variable: numérica continua, valores: desde 0 hasta 8,6.

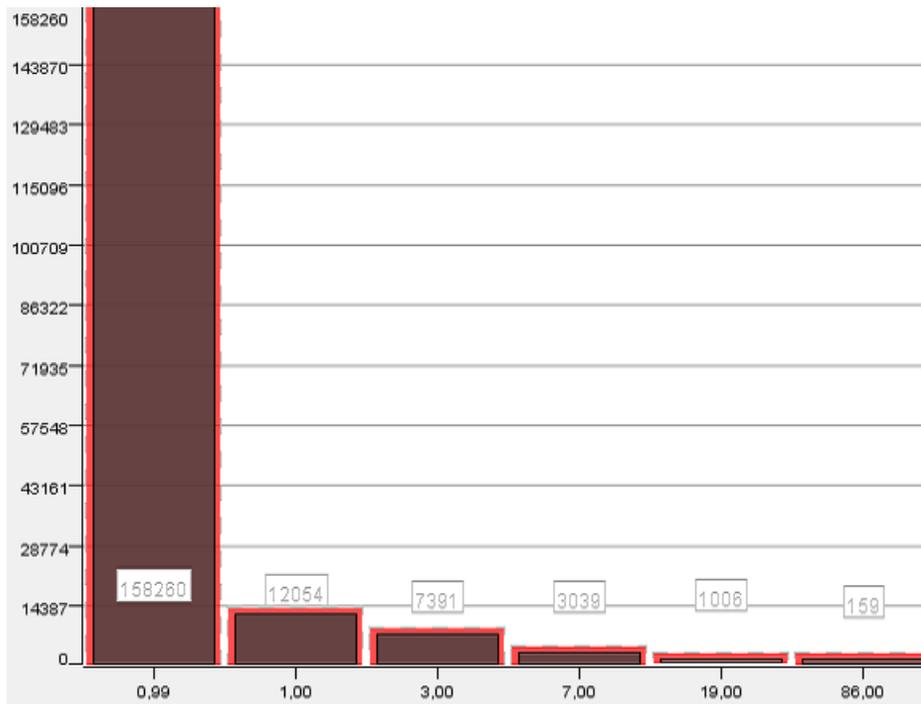


Figura C.3: Histograma número de vehículos por el monto total de siniestro del vehículo - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura [C.3](#), en primera instancia se puede seleccionar la variable para realizar los otros estudios antes de la agrupación transversal, además, esta variable es la razón de los estudios por lo que se considera la variable y1.

4. Número total de siniestros del vehículo, tipo de variable: numérica discreta, valores: desde 0 hasta 7.

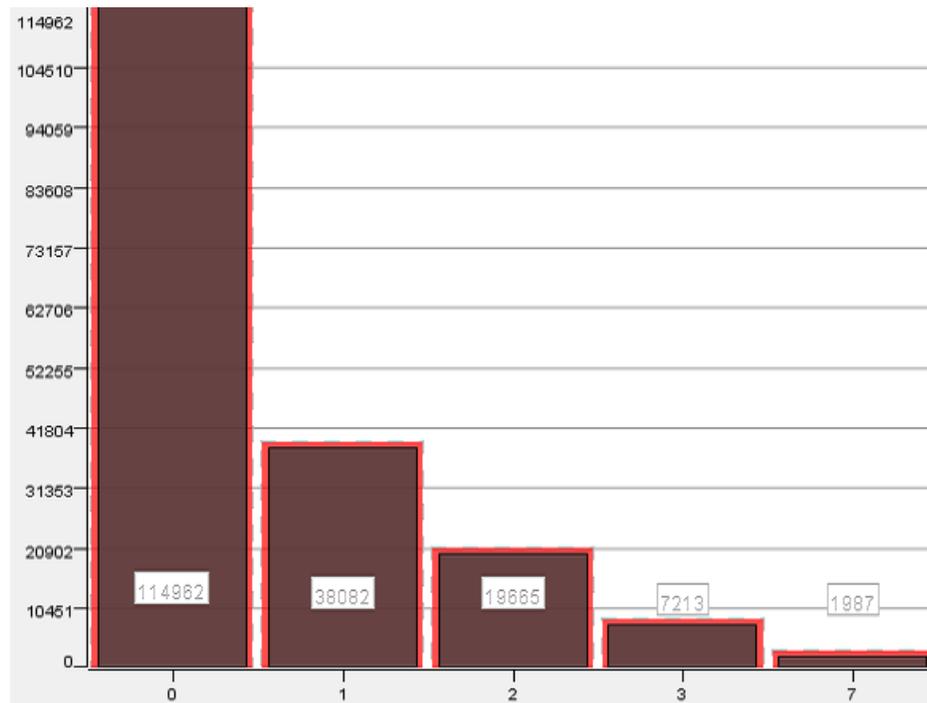


Figura C.4: Histograma número de vehículos por el número total de siniestro del vehículo - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura C.4, en primera instancia se puede seleccionar la variable para realizar los otros estudios antes de la agrupación transversal, además, esta variable es la razón de los estudios por lo que se considera la variable y2.

C. ESTUDIO POR CADA VARIABLE

5. Marca modelo del vehículo, tipo de variable: categórica sin orden, valores: código desde 001000 al 500000.

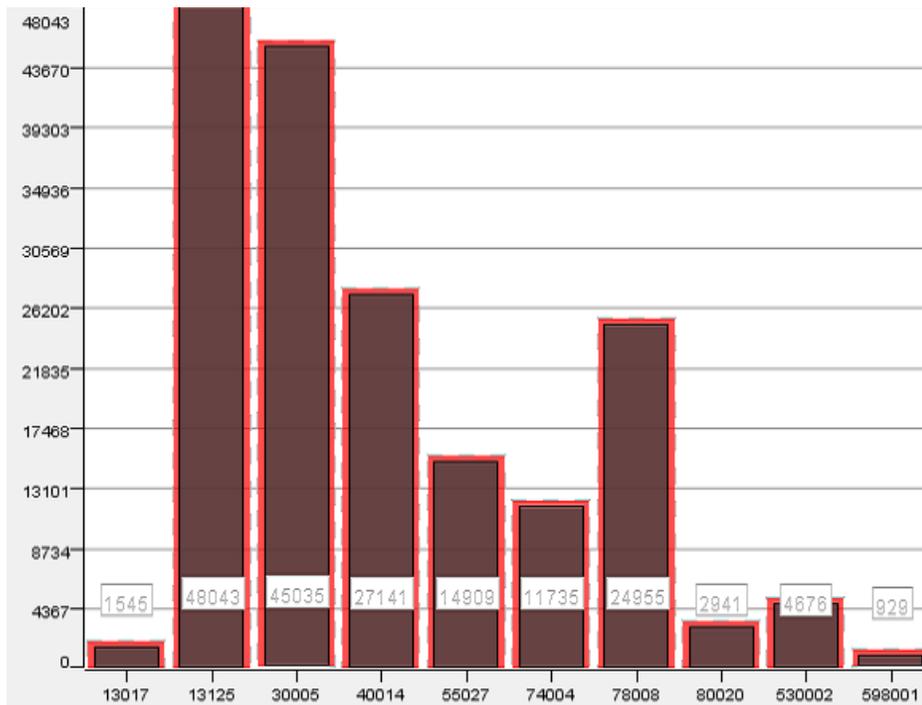


Figura C.5: Histograma número de vehículos por cada marca y modelo asegurado - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura C.5, en primera instancia se puede seleccionar la variable para realizar los otros estudios antes de la agrupación transversal, además, esta variable es una de las principales para discriminar grupos y calcular la máxima pérdida probable.

6. Número de ocupantes del vehículo, tipo de variable: categórica de orden, valores: desde 1 hasta 71.

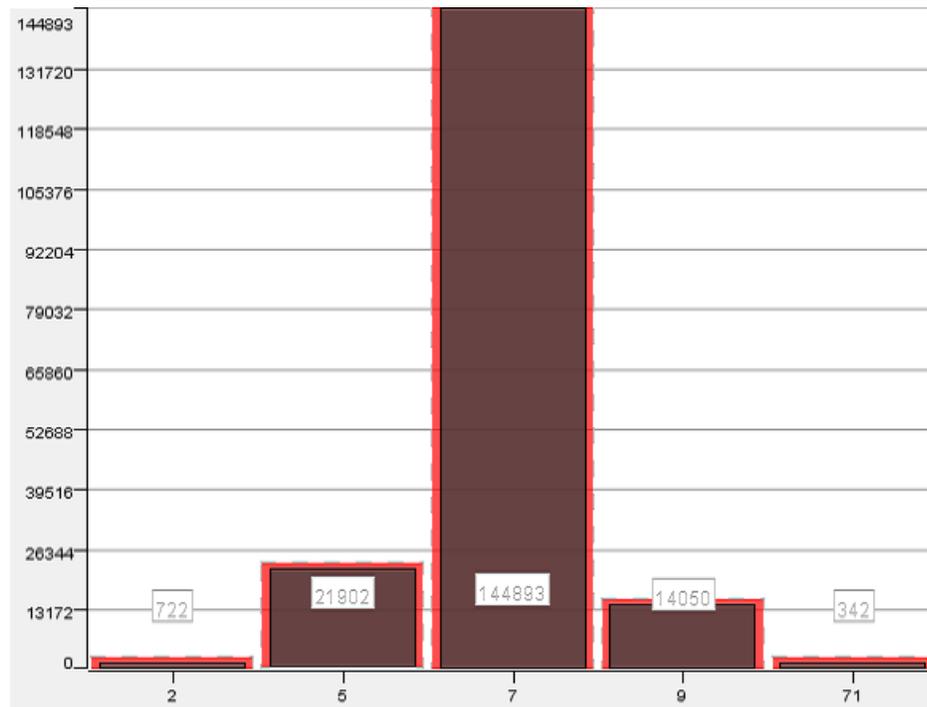


Figura C.6: Histograma número de vehículos por cada número de ocupantes de los vehículos - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura C.6, en primera instancia se puede seleccionar la variable para realizar los otros estudios antes de la agrupación transversal, aunque, como se dice en las limitaciones, se trabajara con vehículos con ocupantes desde 2 hasta 8.

C. ESTUDIO POR CADA VARIABLE

7. Entidad geográfica de circulación del vehículo, tipo de variable: categórica sin orden, valores: código 010000 al 600000.

Con respecto a la entidad geográfica, no se puede seleccionar la variable para realizar los otros estudios antes de la agrupación transversal, porque se observa que la información es de carácter administrativo sin considerar la verdadera zona de circulación.

8. Estado civil del conductor del vehículo, tipo de variable: categórica sin orden, valores: 1 Casado(a); 2 Soltero(a); 3 Divorciado(a); 4 Viudo(a); 5 Concubino(a).

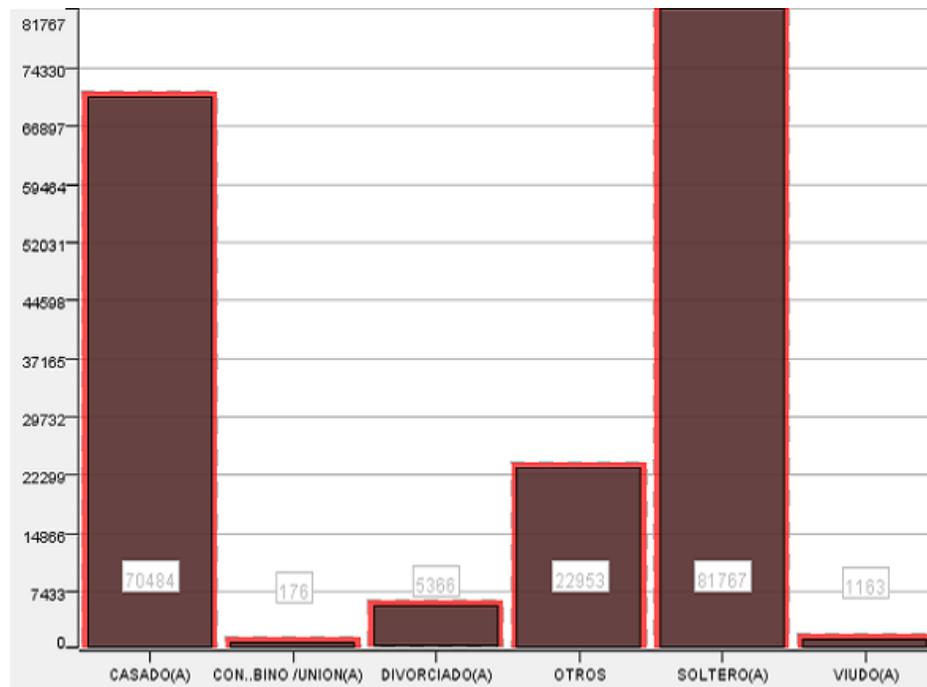


Figura C.7: Histograma número de vehículos por cada estado civil observado -

Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura C.7, en primera instancia se puede seleccionar la variable para realizar los otros estudios antes de la agrupación transversal, además, se observa que los casados y solteros tienen la mayoría de los vehículos.

C. ESTUDIO POR CADA VARIABLE

9. Sexo del conductor del vehículo, tipo de variable: categórica sin orden, valores: 1 Mujer, 2 Hombre, 3 No informaron.

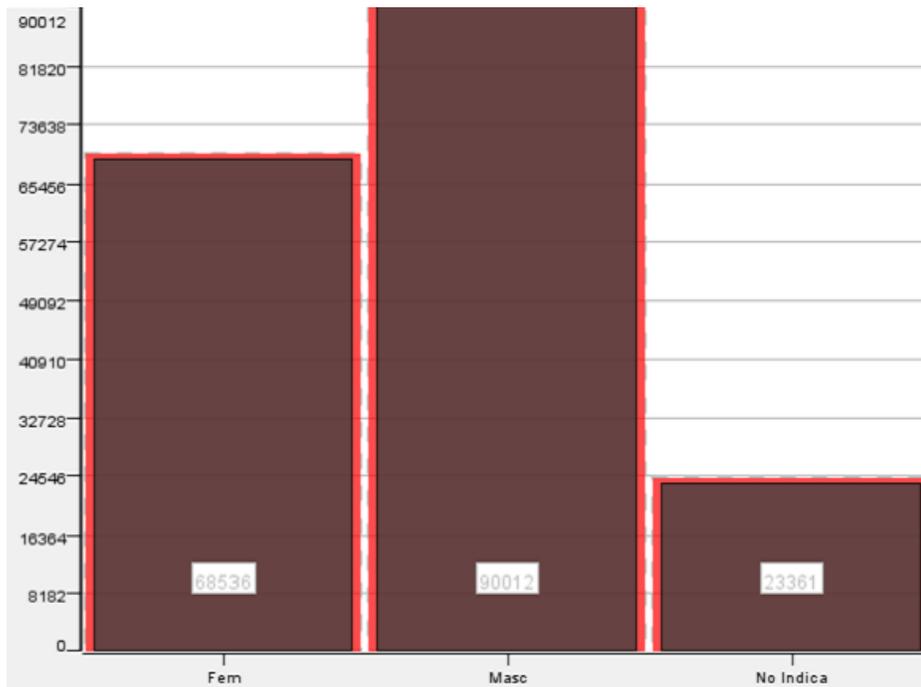


Figura C.8: Histograma número de vehículos por cada sexo observado - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura [C.8](#), en primera instancia se puede seleccionar la variable para realizar los otros estudios antes de la agrupación transversal, además, se observa que el sexo masculino tienen la mayoría de los vehículos.

10. Modalidad del vehículo, tipo de variable: categórica sin orden, valores: 1 Individual, 2 Colectivo o Flota.

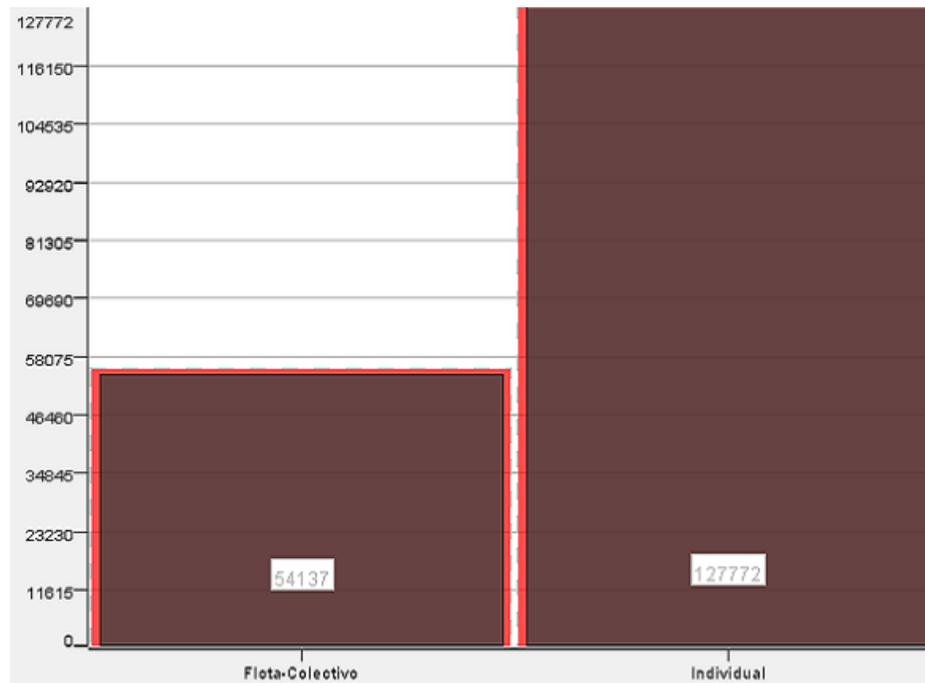


Figura C.9: Histograma número de vehículos por cada modalidad observada -

Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura C.9, en primera instancia se puede seleccionar la variable para realizar los otros estudios antes de la agrupación transversal, además, se observa que los contratos realizados individualmente tienen la mayoría de los vehículos.

C. ESTUDIO POR CADA VARIABLE

11. Contratante gubernamental o no del seguro, tipo de variable: categórica sin orden, valores: 1 Gubernamental, 2 No gubernamental.

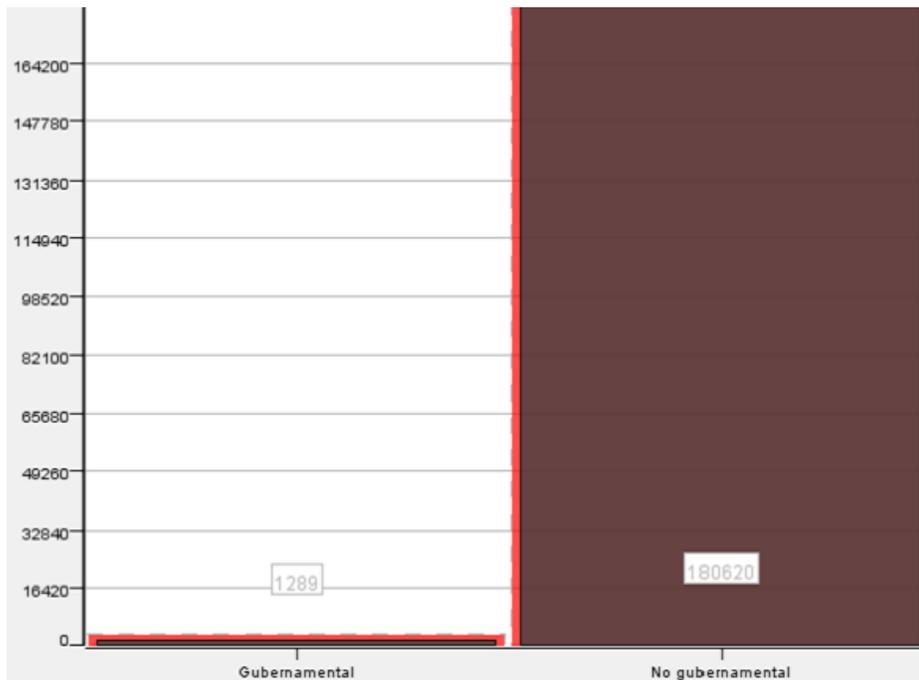


Figura C.10: Histograma número de vehículos por cada tipo de contratante observado - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura C.10, no se puede seleccionar la variable para realizar los otros estudios antes de la agrupación transversal, porque se observa que la información está más del 99% contenida en no gubernamental, la variable puede ser descartada.

12. Tipo de vehículo, tipo de variable: categórica sin orden, valores: 1 Particular, 2 Rústico, 3 Pick Up, 4 Carga hasta 5 TN, 99 Otros.

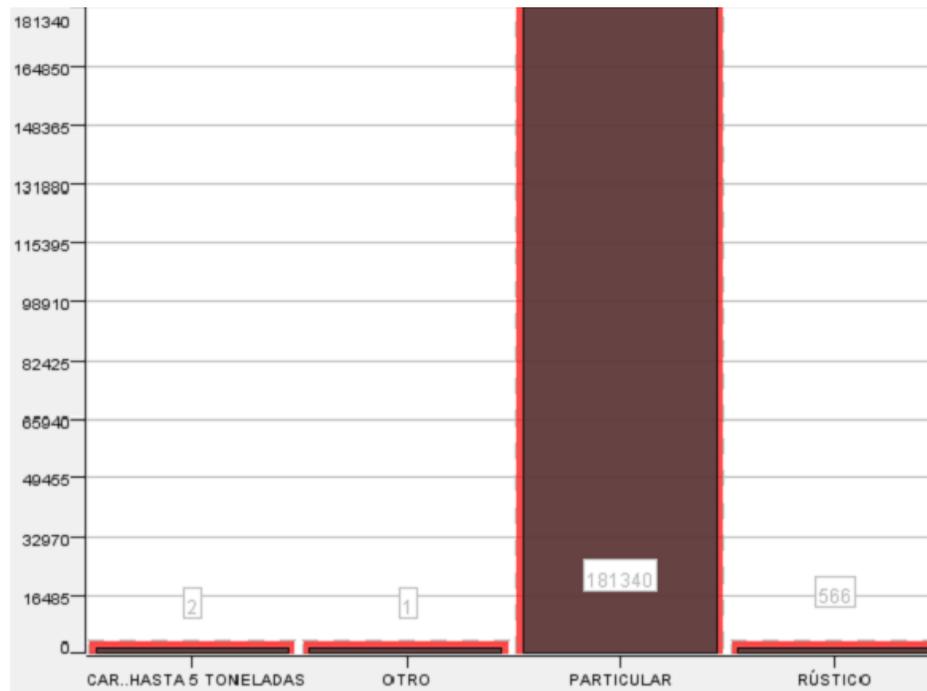


Figura C.11: Histograma número de vehículos por cada tipo de vehículo observado - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura C.11, no se puede seleccionar la variable para realizar los otros estudios antes de la agrupación transversal, ya que se observa que el código particular tienen la mayoría de los vehículos cuando existen vehículos de todos los tipos identificado como particular.

13. Densidad poblacional donde circula el vehículo, personas por kilómetro cuadrado, tipo de variable: numérica discreta, valores: desde 1 hasta 4.809.

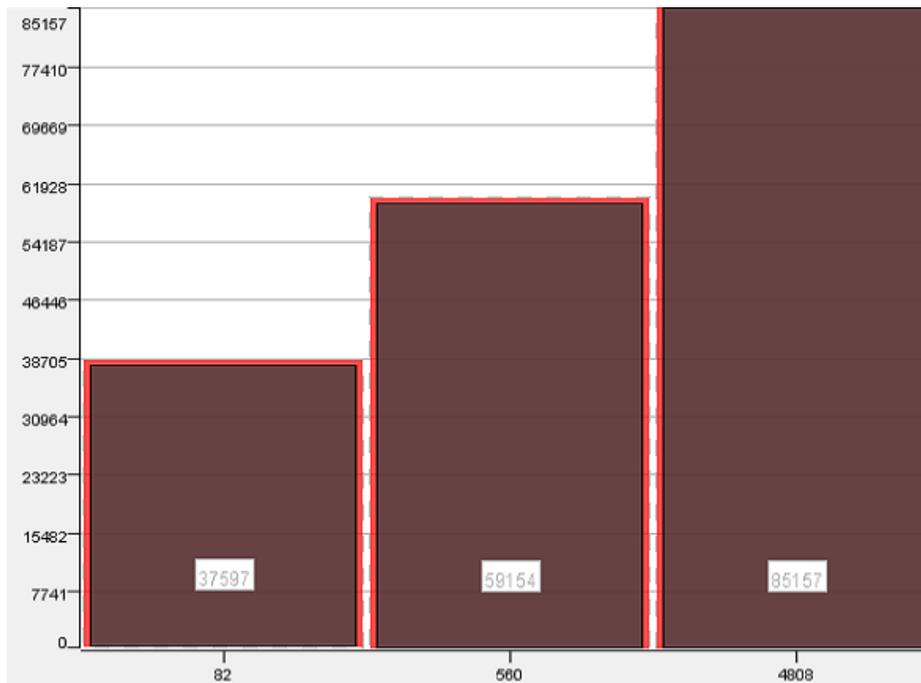


Figura C.12: Histograma número de vehículos por densidad poblacional observada - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura C.12, no se puede seleccionar la variable para realizar los otros estudios antes de la agrupación transversal, ya que esta variable es dependiente de la zona geográfica que fue excluida del estudio.

14. Tiempo de asegurado en años, tipo de variable: categórica de orden, valores: desde 0 hasta 30.

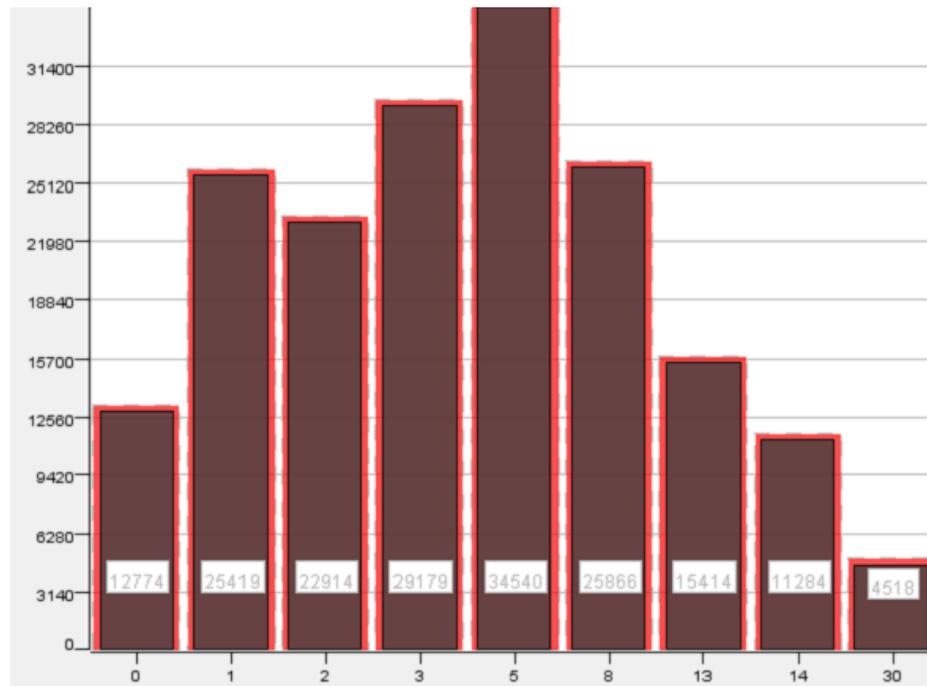


Figura C.13: Histograma número de vehículos por el número de renovaciones observadas - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura C.13, en primera instancia se puede seleccionar la variable para realizar los otros estudios antes de la agrupación transversal, además, se observa que los contratos que tienen entre 4 y 5 años contienen la mayoría de los vehículos.

C. ESTUDIO POR CADA VARIABLE

15. Edad del conductor del vehículo, tipo de variable: numérica discreta, valores: desde 18 hasta 90.

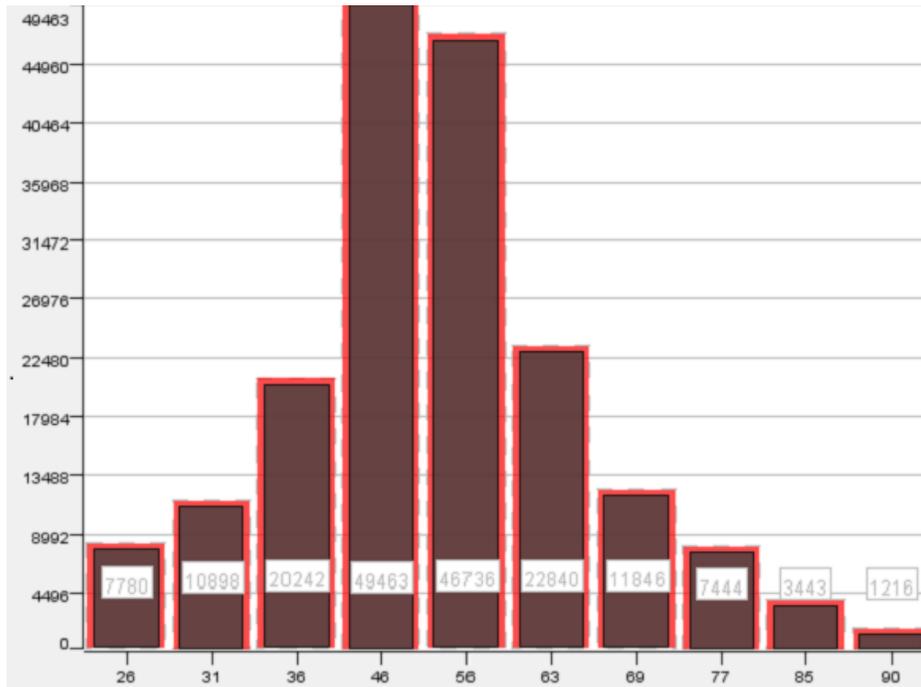


Figura C.14: Histograma número de vehículos por cada edad observada - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura C.14, en primera instancia se puede seleccionar la variable para realizar los otros estudios antes de la agrupación transversal, además, se observa que la clase con los conductores de edad entre 37 y 46 años, contienen la mayoría de los vehículos.

16. Factor de exposición al riesgo del vehículo, tipo de variable: numérica continua, valores: desde 0,0027 hasta 1.

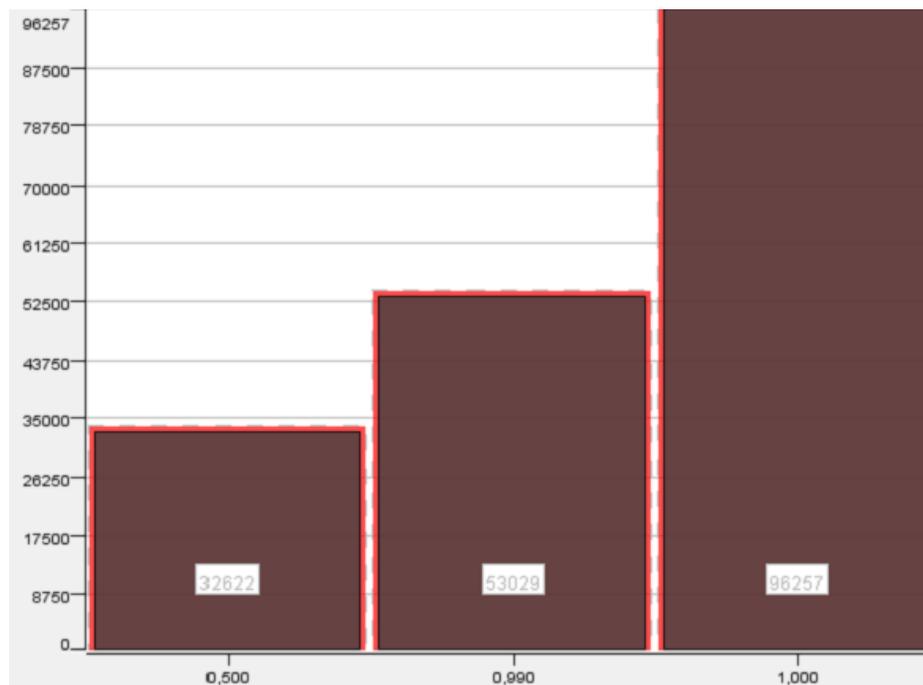


Figura C.15: Histograma número de vehículos por la exposición al riesgo observada - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura C.15, en primera instancia se puede seleccionar la variable para realizar los otros estudios antes de la agrupación transversal, además, se observa que la clase con los factores de exposición al riesgo entre 0,991 y 1, contienen la mayoría de los vehículos.

C.1.2. Presentación de correlación de las variables

1. Modalidad,
2. Valor del vehículo,

C. ESTUDIO POR CADA VARIABLE

3. Año,
4. Número de ocupante,
5. Tipo de vehículo,
6. Estado civil,
7. Sexo,
8. Habitante por kilometro cuadrado,
9. Marca modelo,
10. Edad del conductor,
11. Tiempo asegurado,
12. Número de siniestros,
13. Monto total facturado siniestro.

Revisando la matriz de correlaciones, no existe una marcada correlación entre las variables mostradas, por lo que se debe realizar un estudio exhaustivo.

C.1.3. Presentación de gráfico de dispersión

A continuación, se presentan gráficos de dispersión con respecto a variables compuestas que son: Tasa natural de riesgo posterior, tasa de uso posterior. Estas variables son cálculas a través de las siguientes variables, para Tasa natural de riesgo posterior se utiliza el "monto total facturado de siniestros.^{ent}re la "suma total de los valores de los vehículos para la tasa de uso posterior se utiliza el "número de siniestros.^{ent}re la "suma total del factor de exposición de riesgo".

1. Marca modelo,

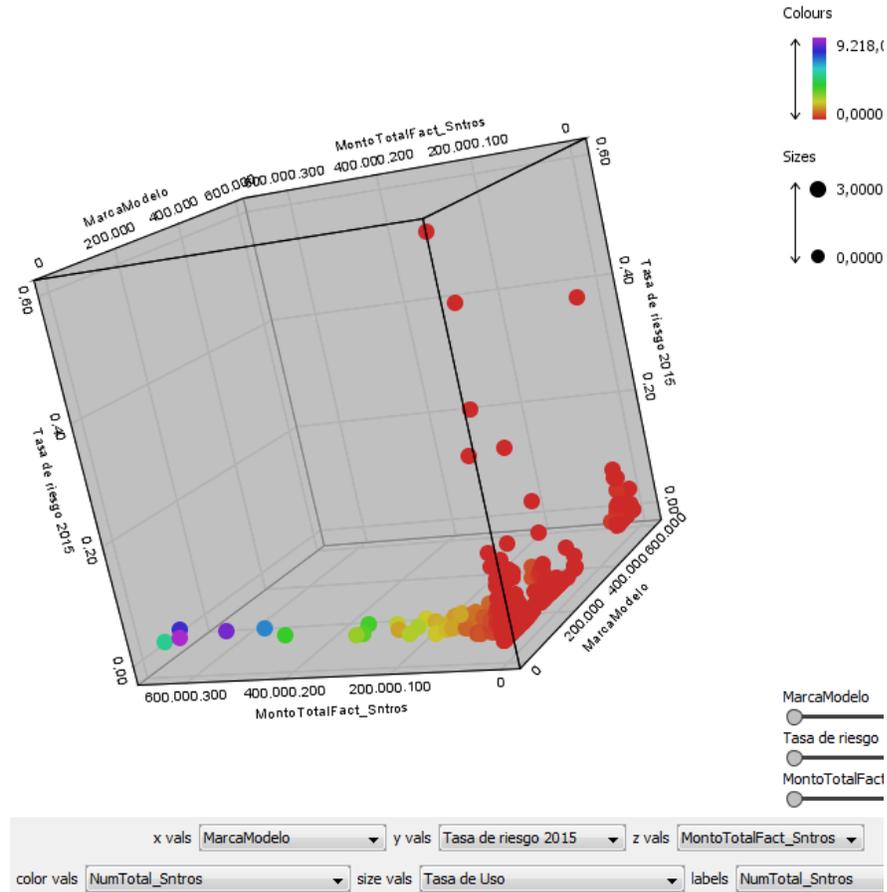


Figura C.16: Dispersión tasa de riesgo, monto total facturado, número de siniestro y tasa de uso por cada marca modelo - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura C.16, en primera instancia se observa que a mayor cantidad de siniestros y de exposición de cada marca modelo la tasa natural de riesgo es contundente estando al rededor del 0,02.

C. ESTUDIO POR CADA VARIABLE

2. Valor del vehículo,

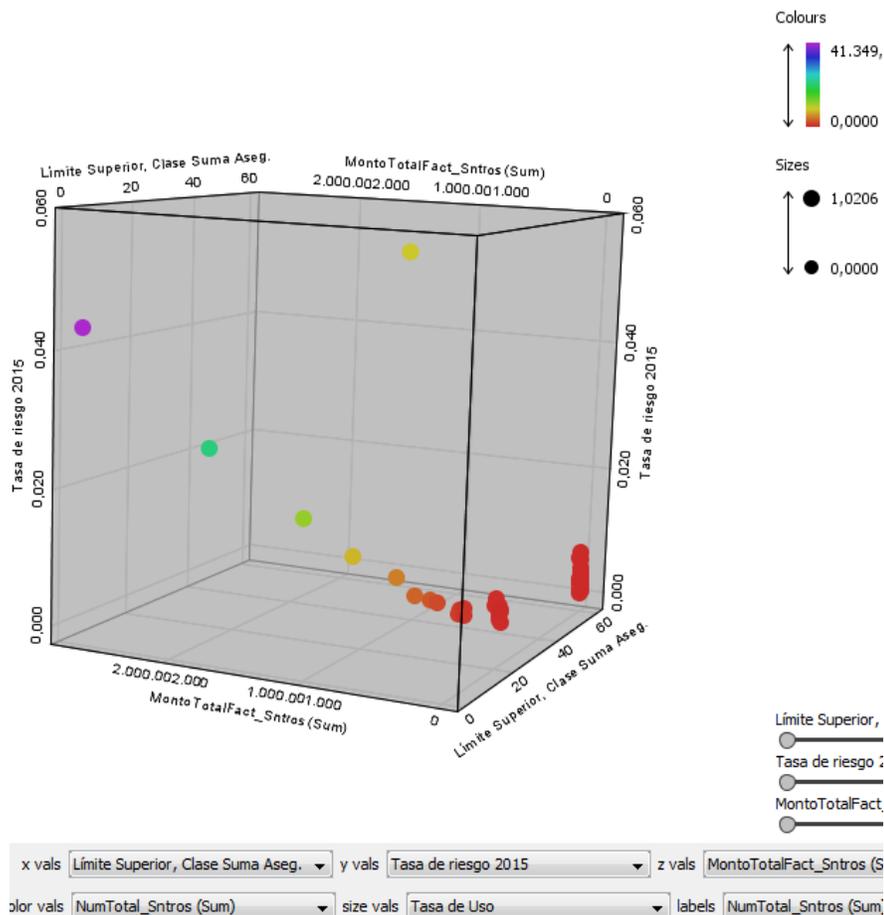


Figura C.17: Dispersión tasa de riesgo, monto total facturado, número de siniestro y tasa de uso por cada valor de vehículo - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura C.17, en primera instancia se observa que a mayor cantidad de siniestros y de exposición de cada valor del vehículo, la tasa natural de riesgo es precisa, estando al rededor del 0,02.

3. Estado civil,

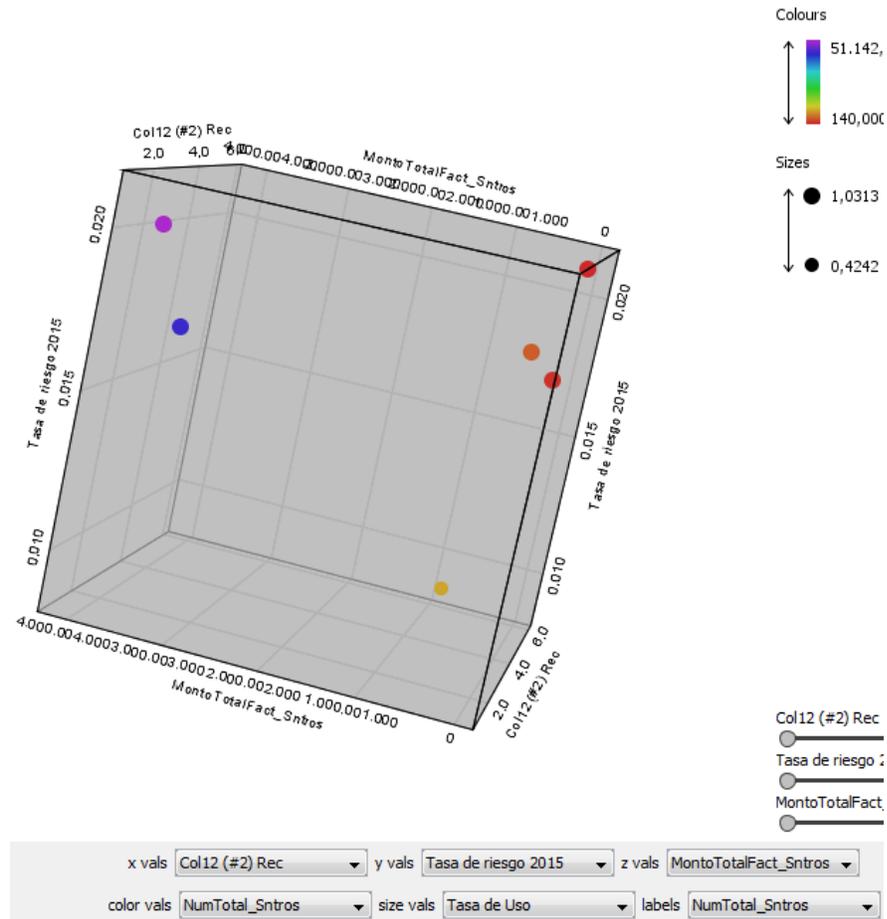


Figura C.18: Dispersión tasa de riesgo, monto total facturado, número de siniestro y tasa de uso por cada estado civil - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura C.18, en primera instancia se observa que a mayor cantidad de siniestros y de exposición de cada estado civil la tasa natural de riesgo es adecuada, estando al rededor del 0,02.

4. Sexo,

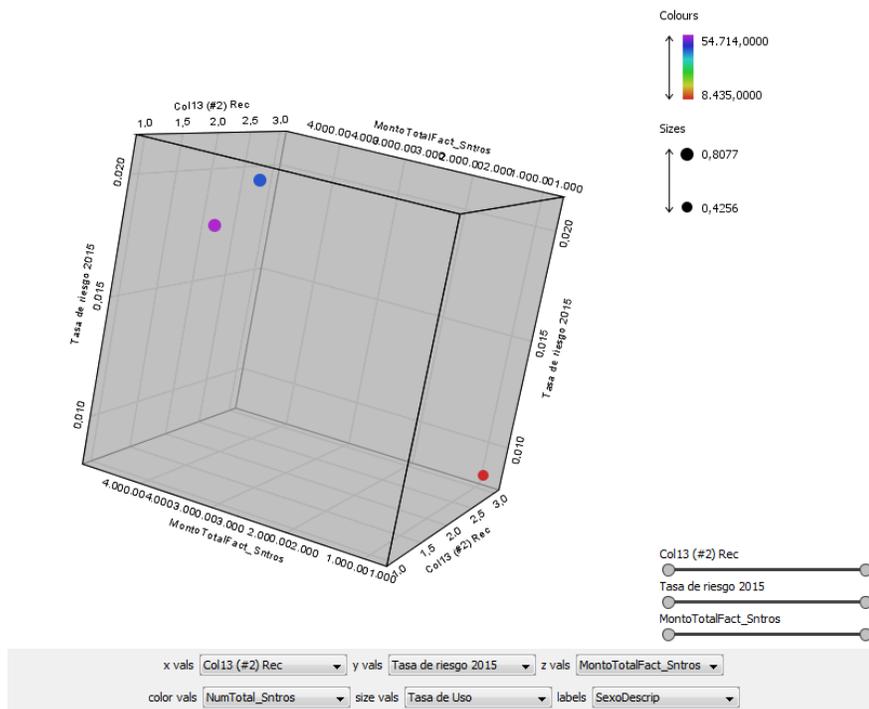


Figura C.19: Dispersión tasa de riesgo, monto total facturado, número de siniestro y tasa de uso por cada sexo - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura C.19, en primera instancia se observa que a mayor cantidad de siniestros y de exposición de cada sexo la tasa natural de riesgo es adecuada, estando al rededor del 0,02.

6. Tiempo asegurado,

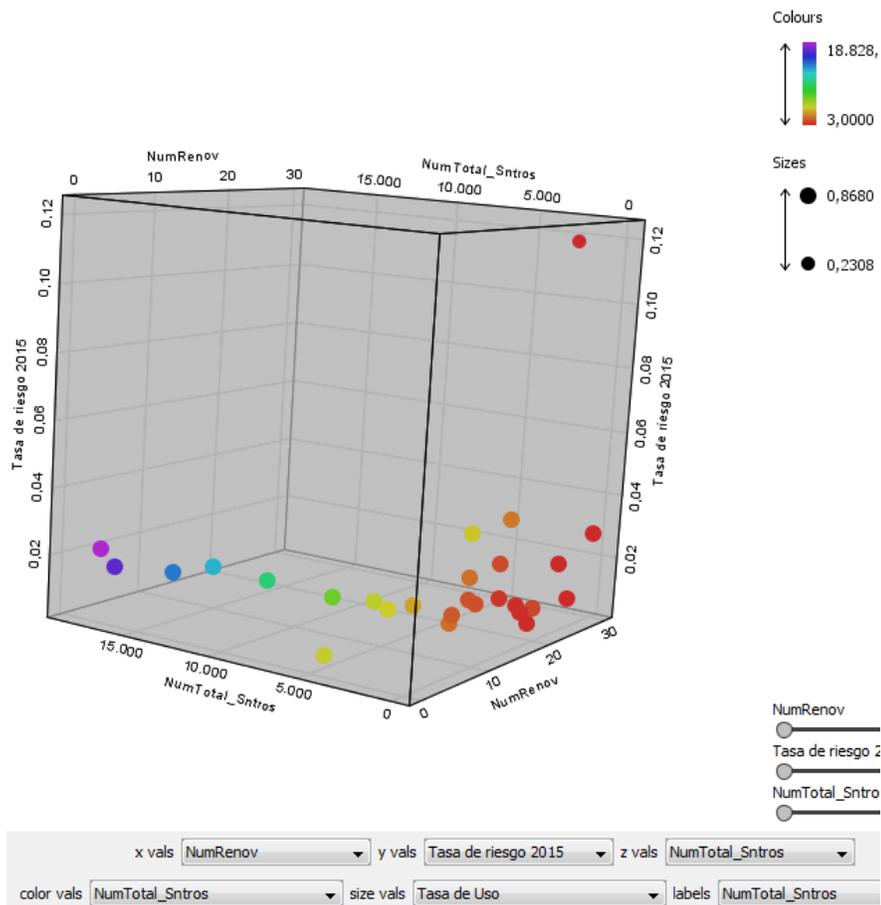


Figura C.21: Dispersión tasa de riesgo, Monto Total Facturado, Número de siniestro y tasa de uso por cada tiempo asegurado del vehículo - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura C.21, en primera instancia se observa que a mayor cantidad de siniestros y de exposición de cada "tiempo asegurado del vehículo" la tasa natural de riesgo es adecuada, estando al rededor del 0,02.

7. Edad del conductor,

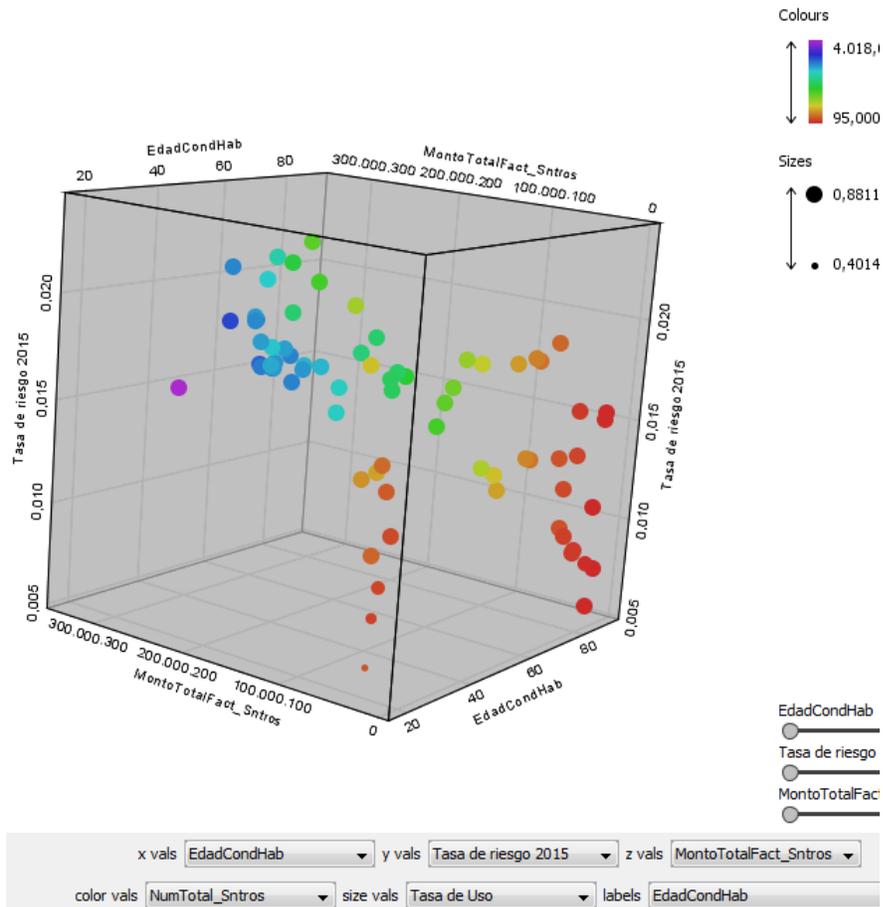


Figura C.22: Dispersión tasa de riesgo, monto total facturado, número de siniestro y tasa de uso por cada edad del conductor del vehículo - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura C.22, en primera instancia se observa que a mayor cantidad de siniestros y de exposición de cada edad del conductor la tasa natural de riesgo es adecuada, estando al rededor del 0,02.

8. Exposición al riesgo,

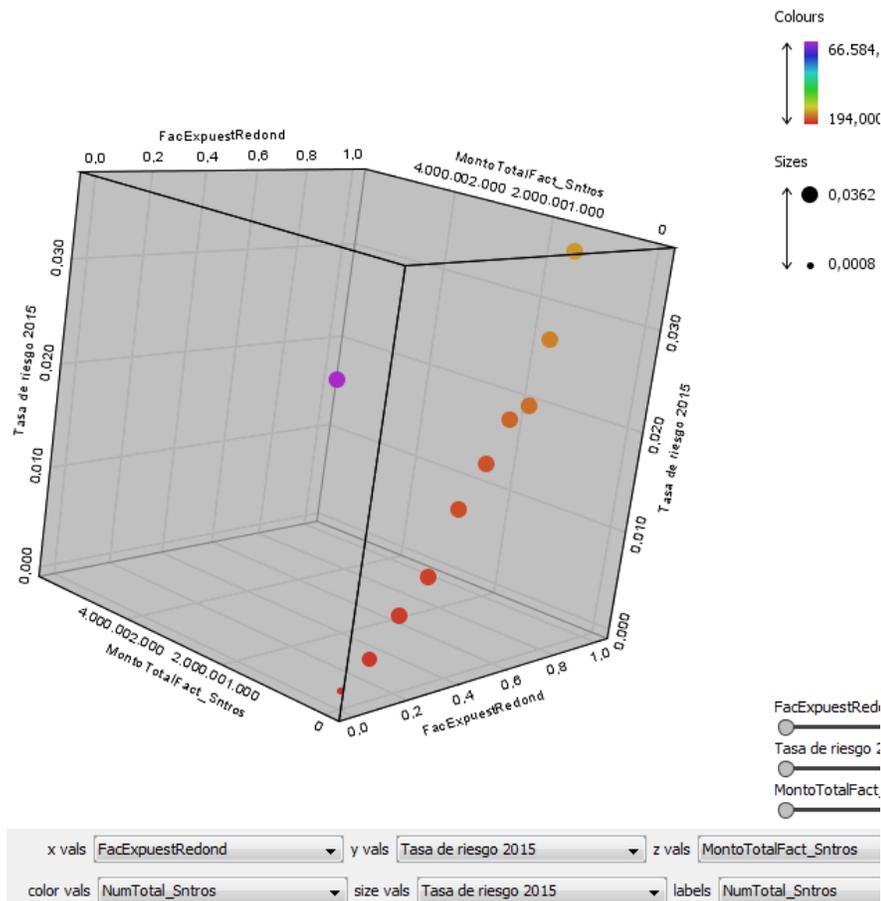


Figura C.23: Dispersión tasa de riesgo, monto total facturado, número de siniestro y tasa de uso por cada tiempo de exposición del vehículo - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura C.23, en primera instancia se observa que a mayor cantidad de siniestros de cada exposición la tasa natural de riesgo es adecuada, estando al rededor del 0,02.

9. Año,

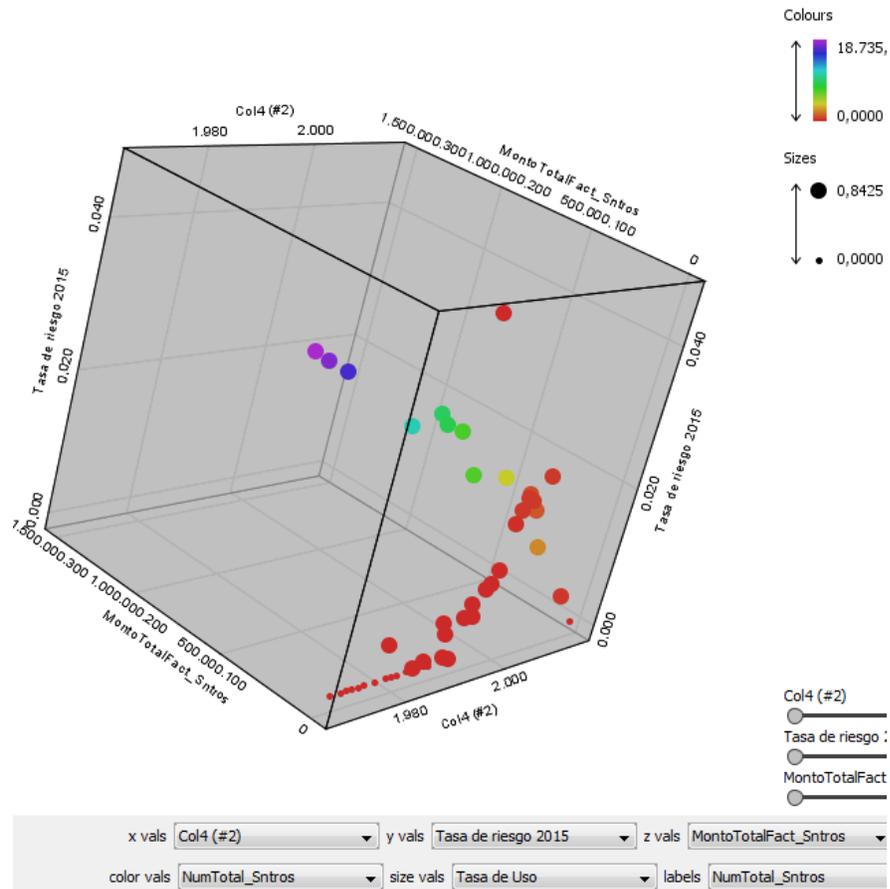


Figura C.24: Dispersión tasa de riesgo, monto total facturado, número de siniestro y tasa de uso por cada tiempo de exposición del vehículo - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura C.24, en primera instancia se observa que a mayor cantidad de siniestros y de exposición de cada Año la tasa natural de riesgo es adecuada, estando al rededor del 0,02.

Conglomerado a través de k medias

primera opción

D.1. Conglomerado a través de K medias, con variables originales

Se realizó grupos a través de k medias con todas las variables conjuntamente, estas variables son originales sin efectuar modificaciones sobre ellas.

1. Conglomerado con $K=2$ medias.

Con respecto a la figura [D.1](#), los grupos no son claros al ser graficados con la variable de monto facturado de siniestros.

D. CONGLOMERADO A TRAVÉS DE K MEDIAS PRIMERA OPCIÓN

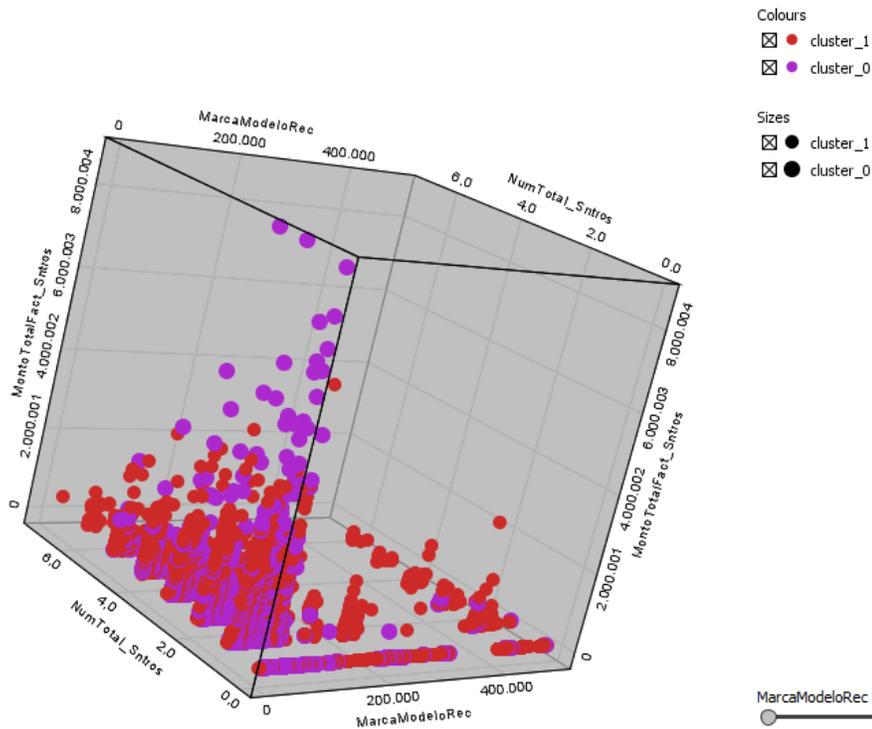


Figura D.1: Conglomerado a través de k medias, con $K=2$, con todas las variables originales - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

2. Conglomerado con $K=3$ medias.

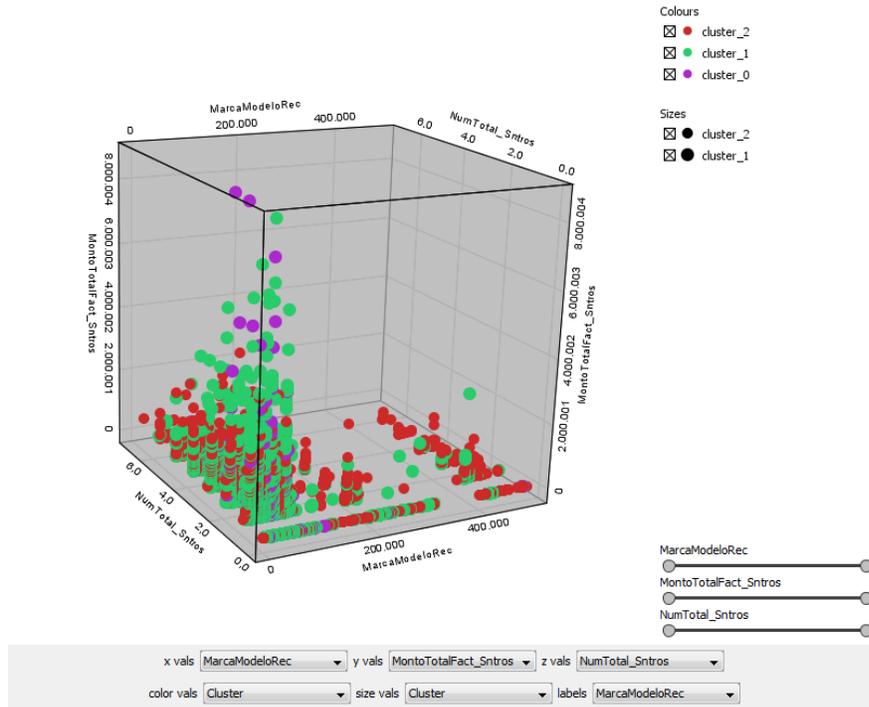


Figura D.2: Conglomerado a través de k medias, con $K=3$, con todas las variables originales - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura D.2, los grupos no son claros al ser graficados con la variable de monto facturado de siniestros.

3. Conglomerado con K=4 medias.

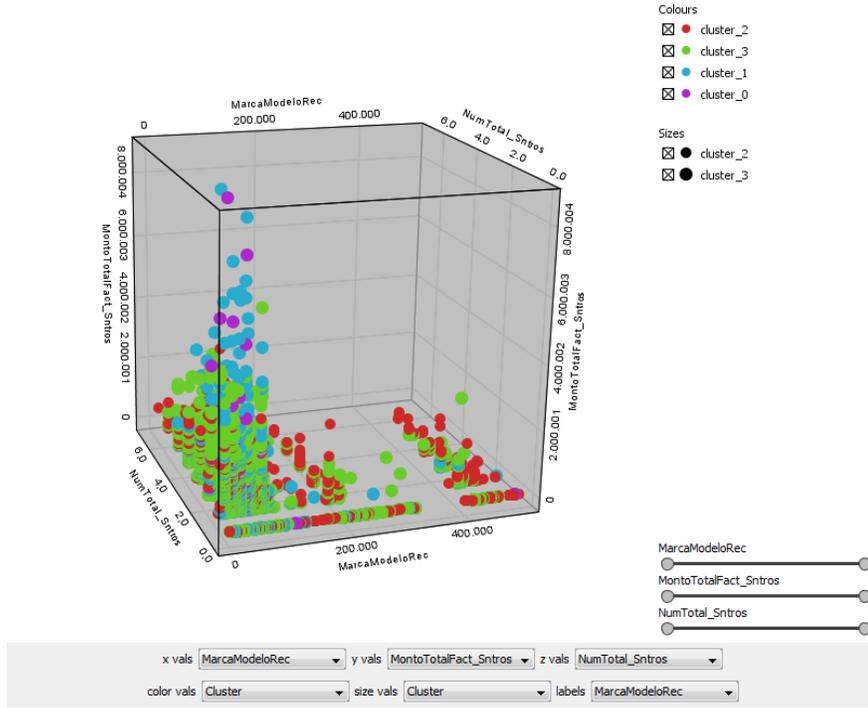


Figura D.3: Conglomerado a través de k medias, con K=4, con todas las variables originales - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura D.3, los grupos no son claros al ser graficados con la variable de monto facturado de siniestros.

4. Conglomerado con $K=5$ medias.

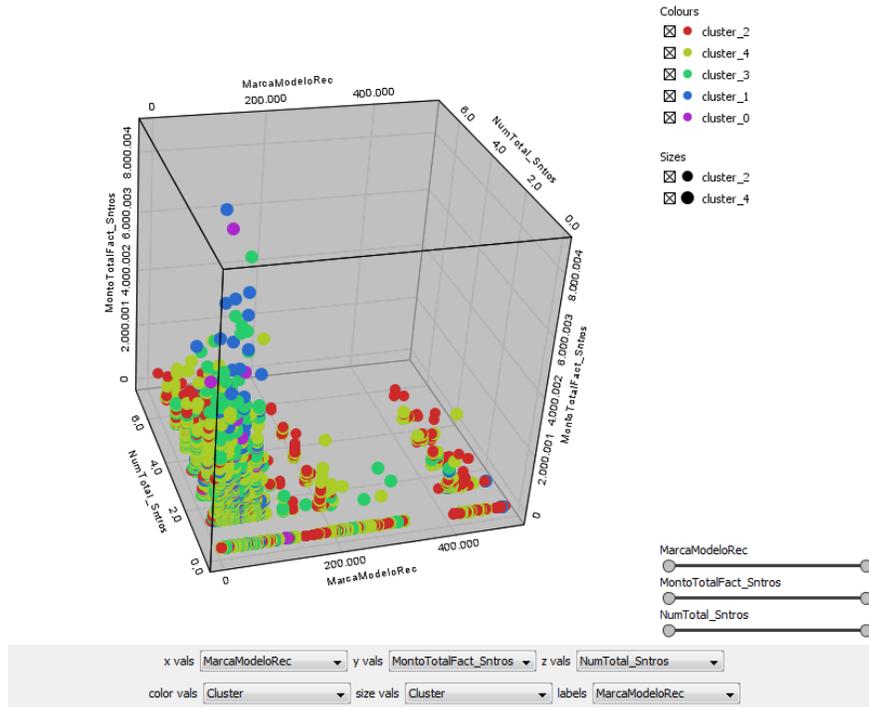


Figura D.4: Conglomerado a través de k medias, con $K=5$, con todas las variables originales - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura D.4, los grupos no son claros al ser graficados con la variable de monto facturado de siniestros.

5. Conglomerado con K=6 medias.

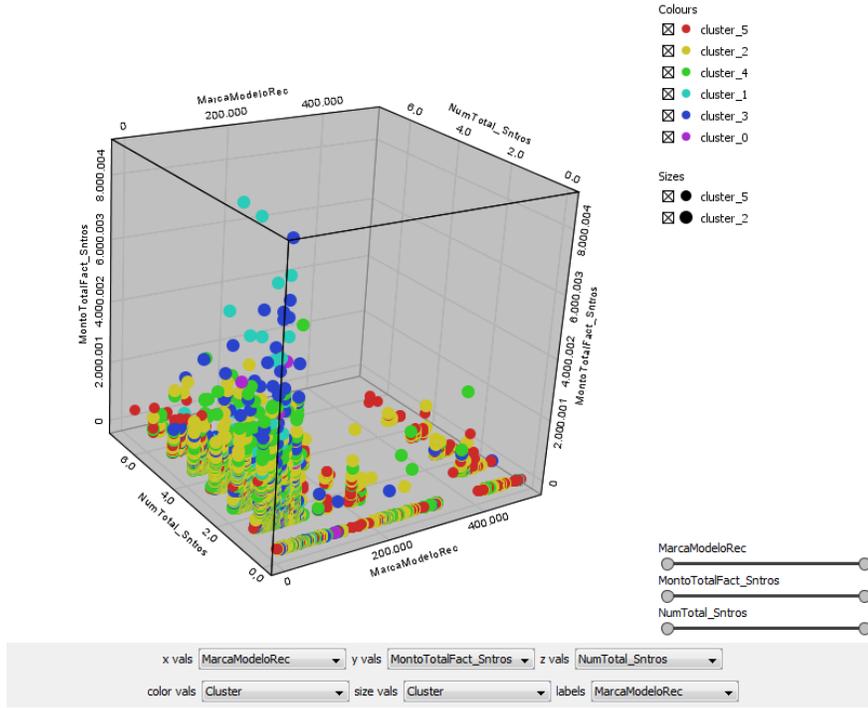


Figura D.5: Conglomerado a través de k medias, con K=6, con todas las variables originales - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura D.5, los grupos no son claros al ser graficados con la variable de monto facturado de siniestros.

6. Conglomerado con $K=7$ medias.

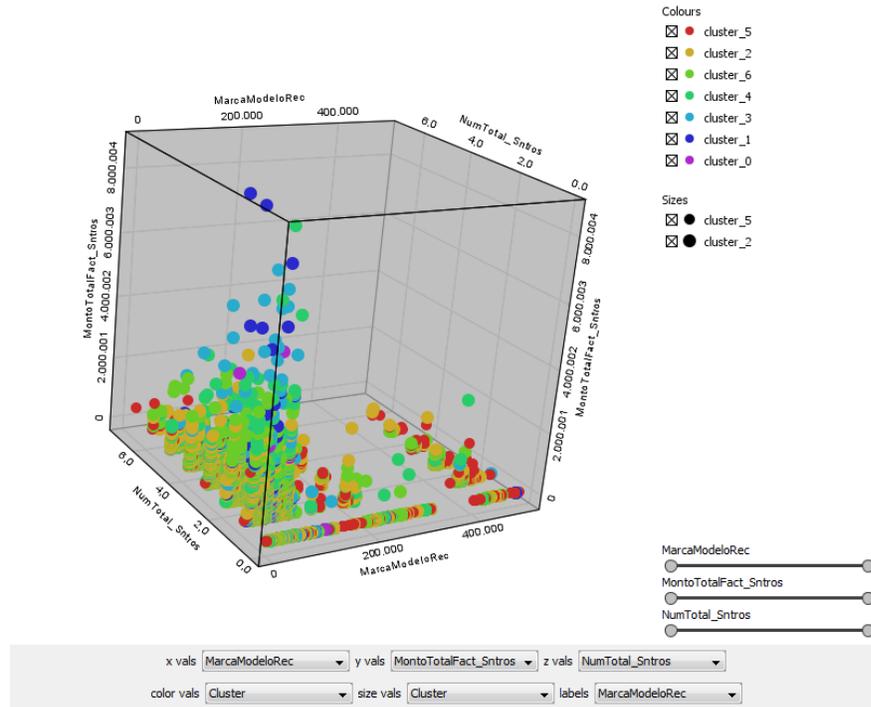


Figura D.6: Conglomerado a través de k medias, con $K=7$, con todas las variables originales - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura D.6, los grupos no son claros al ser graficados con la variable de monto facturado de siniestros.

Conglomerado a través de k medias con
variables numéricas normalizadas y
variables categóricas transformadas en
ficticias

E.1. K medias, con variables numéricas normalizadas

Se realizó grupos a través de k medias con las variables numéricas normalizadas, con las variables categóricas transformadas en variables ficticias de valores 0 y 1, sin excluir ninguna de las variables estudiadas.

E. CONGLOMERADO A TRAVÉS DE K MEDIAS CON VARIABLES NUMÉRICAS NORMALIZADAS Y VARIABLES CATEGÓRICAS TRANSFORMADAS EN FICTICIAS

1. Conglomerado con $K=2$ medias.

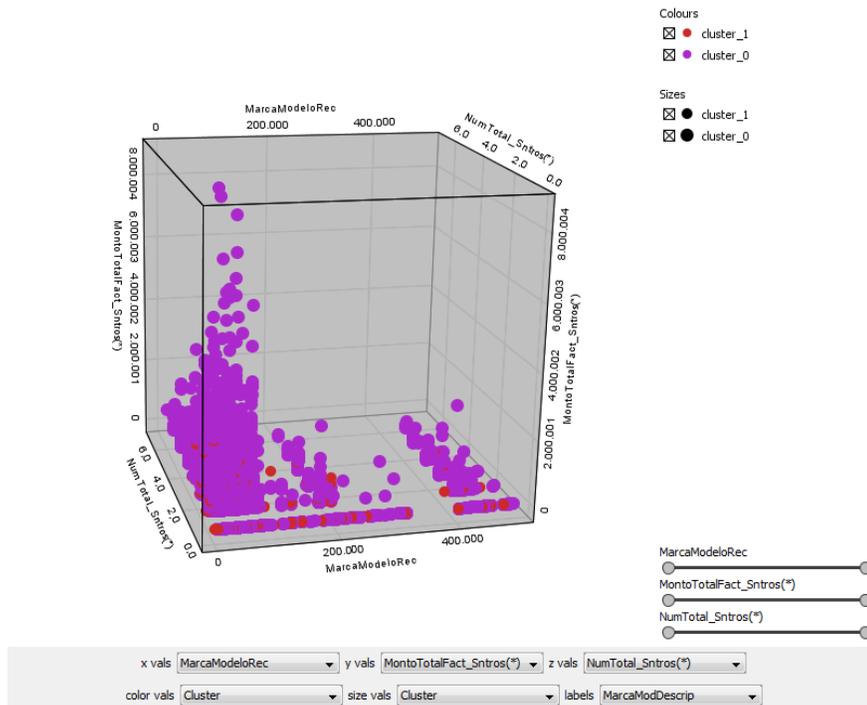


Figura E.1: Conglomerado a través de k medias, con $K=2$, con variables numéricas normalizadas y variables categóricas transformadas en ficticias de 0 y 1 -

Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura E.1, los grupos se dividen entre los que tienen siniestros, monto facturado de siniestros y los demás grupos con los que no tienen siniestros.

2. Conglomerado con K=3 medias.

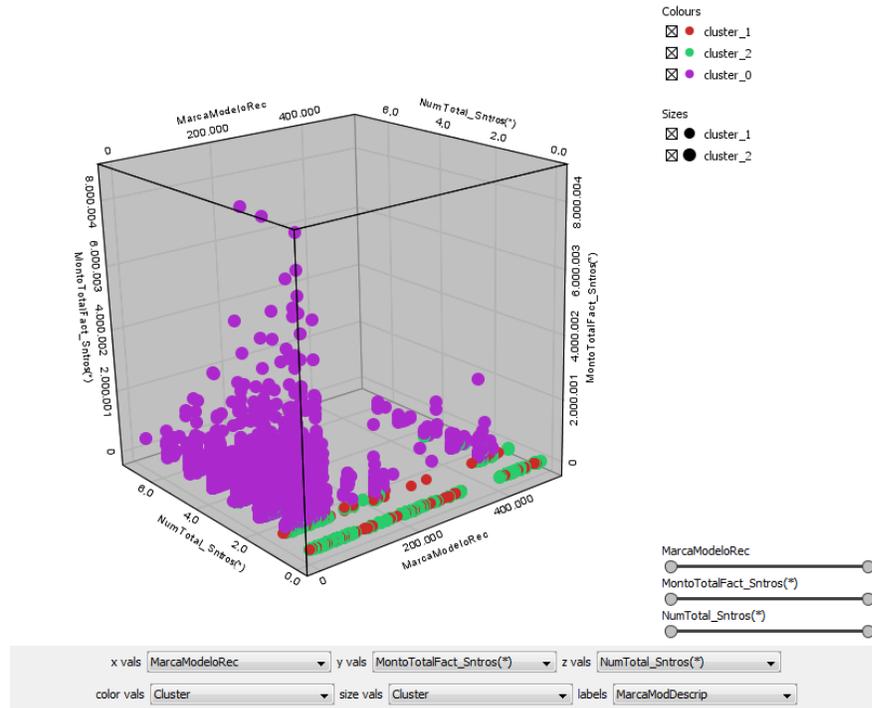


Figura E.2: Conglomerado a través de k medias, con K=3, con variables numéricas normalizadas y variables categóricas transformadas en ficticias de 0 y 1 - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura E.2, los grupos se dividen entre los que tienen siniestros, monto facturado de siniestros y los demás grupos con los que no tienen siniestros.

E. CONGLOMERADO A TRAVÉS DE K MEDIAS CON VARIABLES NUMÉRICAS NORMALIZADAS Y VARIABLES CATEGÓRICAS TRANSFORMADAS EN FICTICIAS

3. Conglomerado con K=4 medias.

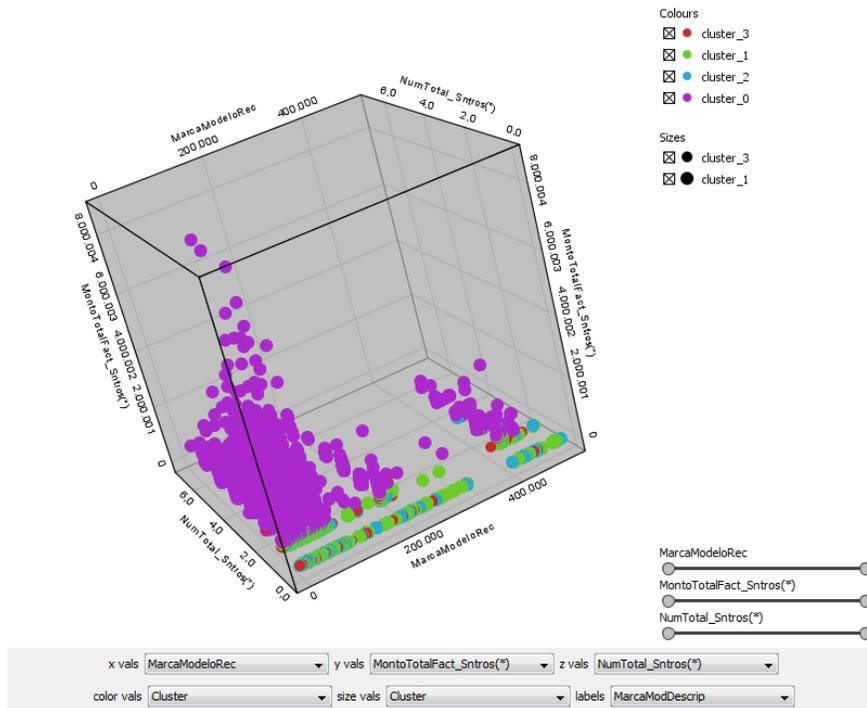


Figura E.3: Conglomerado a través de k medias, con K=4, con variables numéricas normalizadas y variables categóricas transformadas en ficticias de 0 y 1 -

Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura E.3, los grupos se dividen entre los que tienen siniestros, monto facturado de siniestros y los demás grupos con los que no tienen siniestros.

4. Conglomerado con K=5 medias.

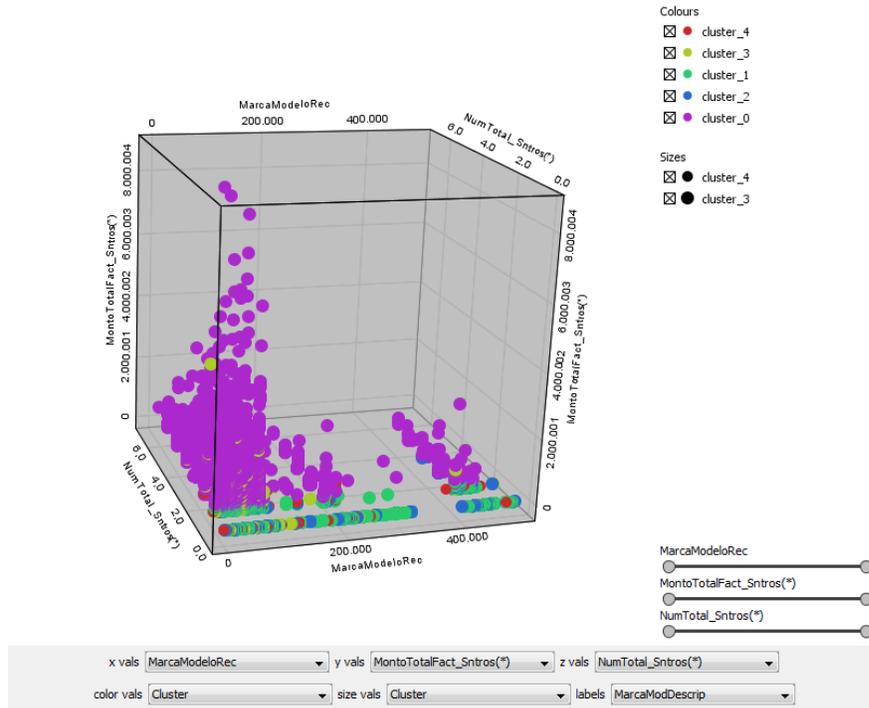


Figura E.4: Conglomerado a través de k medias, con K=5, con variables numéricas normalizadas y variables categóricas transformadas en ficticias de 0 y 1 - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura E.4, los grupos se dividen entre los que tienen siniestros, monto facturado de siniestros y los demás grupos con los que no tienen siniestros.

E. CONGLOMERADO A TRAVÉS DE K MEDIAS CON VARIABLES NUMÉRICAS NORMALIZADAS Y VARIABLES CATEGÓRICAS TRANSFORMADAS EN FICTICIAS

5. Conglomerado con K=6 medias.

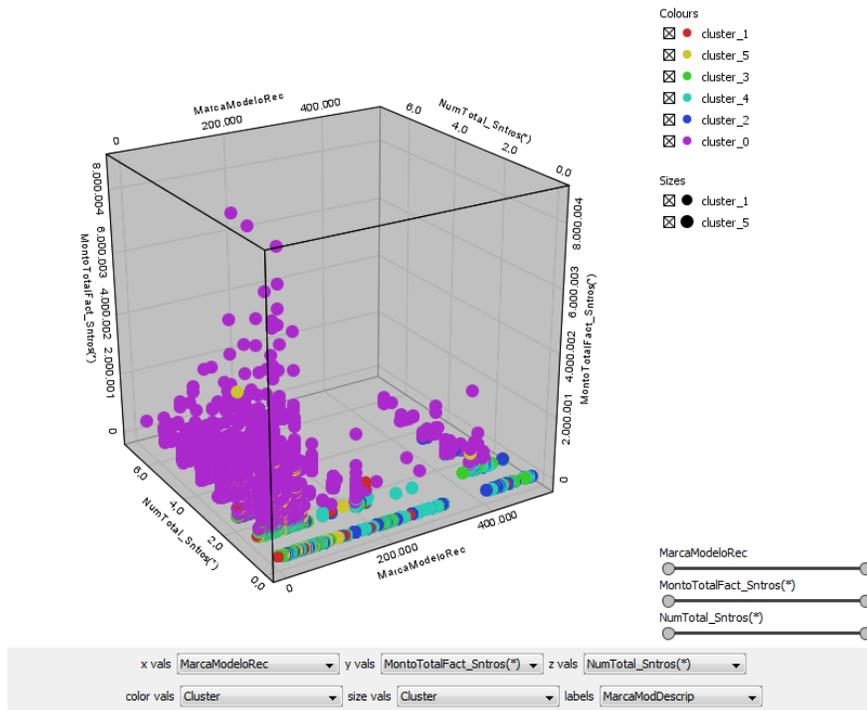


Figura E.5: Conglomerado a través de k medias, con K=6, con variables numéricas normalizadas y variables categóricas transformadas en ficticias de 0 y 1 -

Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura E.5, los grupos se dividen entre los que tienen siniestros, monto facturado de siniestros y los demás grupos con los que no tienen siniestros.

6. Conglomerado con K=7 medias.

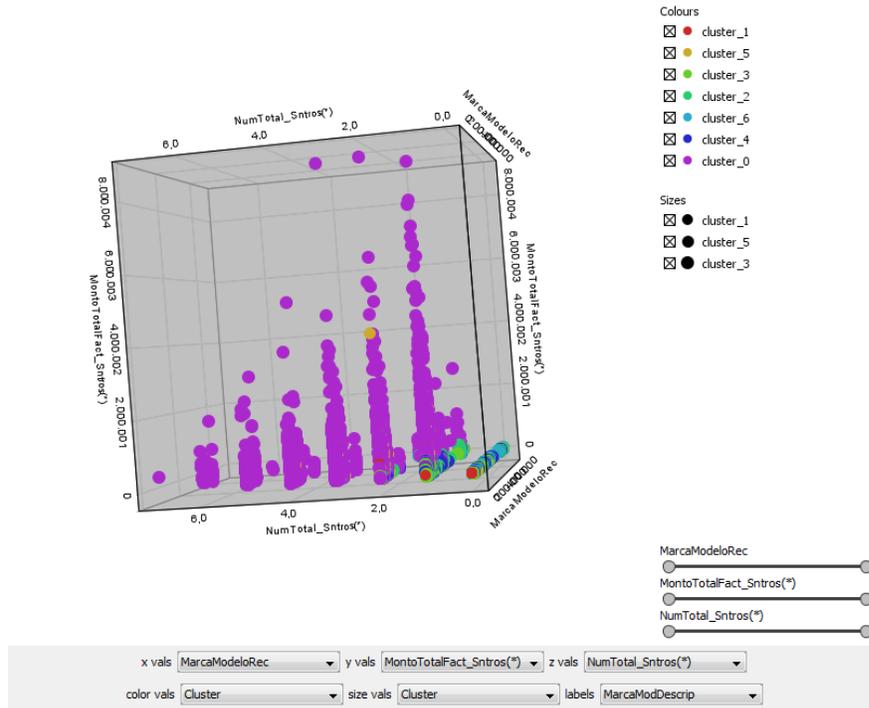


Figura E.6: Conglomerado a través de k medias, con K=7, con variables numéricas normalizadas y variables categóricas transformadas en ficticias de 0 y 1 - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura E.6, los grupos se dividen entre los que tienen siniestros, monto facturado de siniestros y los demás grupos con los que no tienen siniestros.

Conglomerado a través de k medias con
variables normalizadas y variables
categóricas transformadas en ficticias
normalizadas

**F.1. K medias, variables numéricas normalizadas y las ca-
tegóricas transformadas en ficticias normalizadas**

Se realizó grupos a través de k medias con las variables numéricas normalizadas, con las variables categóricas transformadas en variables ficticias de valores 0 y 1, solo con las variables seleccionadas.

F. CONGLOMERADO A TRAVÉS DE K MEDIAS CON VARIABLES NORMALIZADAS Y VARIABLES CATEGÓRICAS TRANSFORMADAS EN FICTICIAS NORMALIZADAS

1. Conglomerado con K=2 medias.

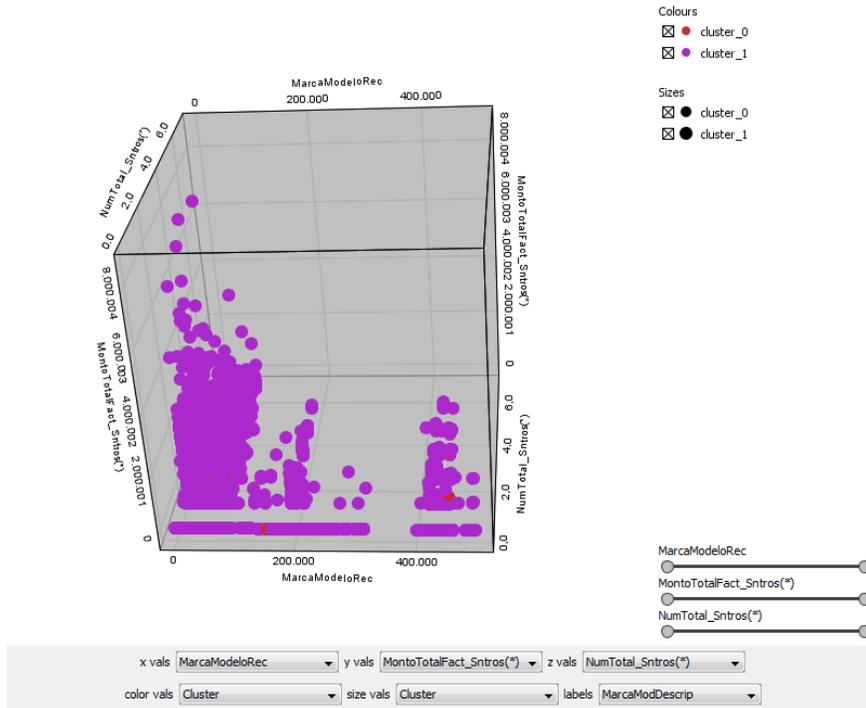


Figura F.1: Conglomerado a través de k medias, con K=2, con variables numéricas normalizadas y variables categóricas transformadas en ficticias de 0 y 1 normalizadas - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura F.1, los grupos se dividen entre los que tienen siniestros, monto facturado de siniestros y los demás grupos con los que no tienen siniestros.

2. Conglomerado con K=3 medias.

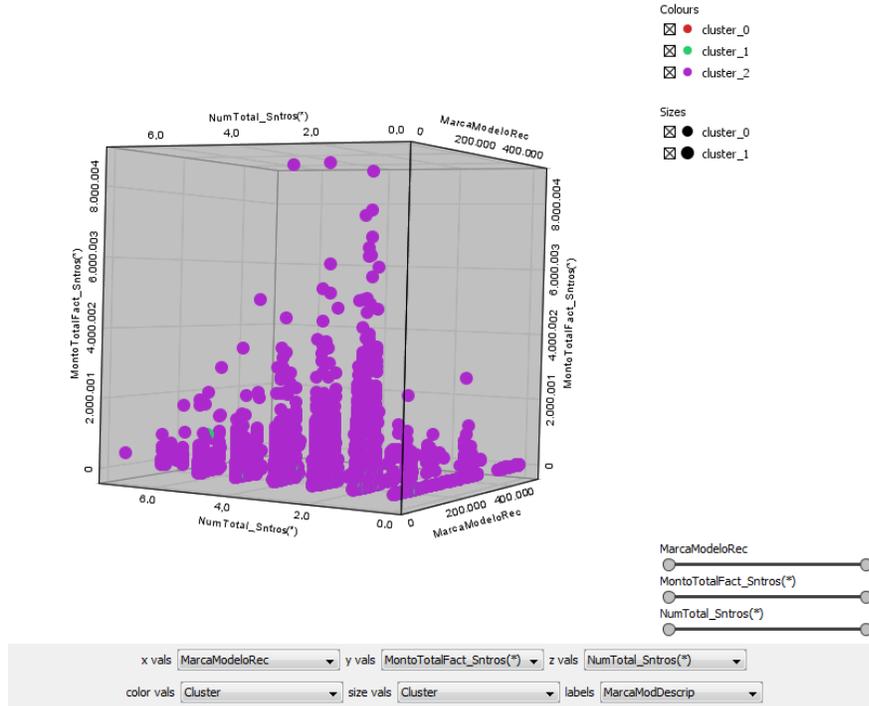


Figura F.2: Conglomerado a través de k medias, con K=3, con variables numéricas normalizadas y variables categóricas transformadas en ficticias de 0 y 1 normalizadas - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura F.2, los grupos se dividen entre los que tienen siniestros, monto facturado de siniestros y los demás grupos con los que no tienen siniestros.

F. CONGLOMERADO A TRAVÉS DE K MEDIAS CON VARIABLES NORMALIZADAS Y VARIABLES CATEGÓRICAS TRANSFORMADAS EN FICTICIAS NORMALIZADAS

3. Conglomerado con K=4 medias.

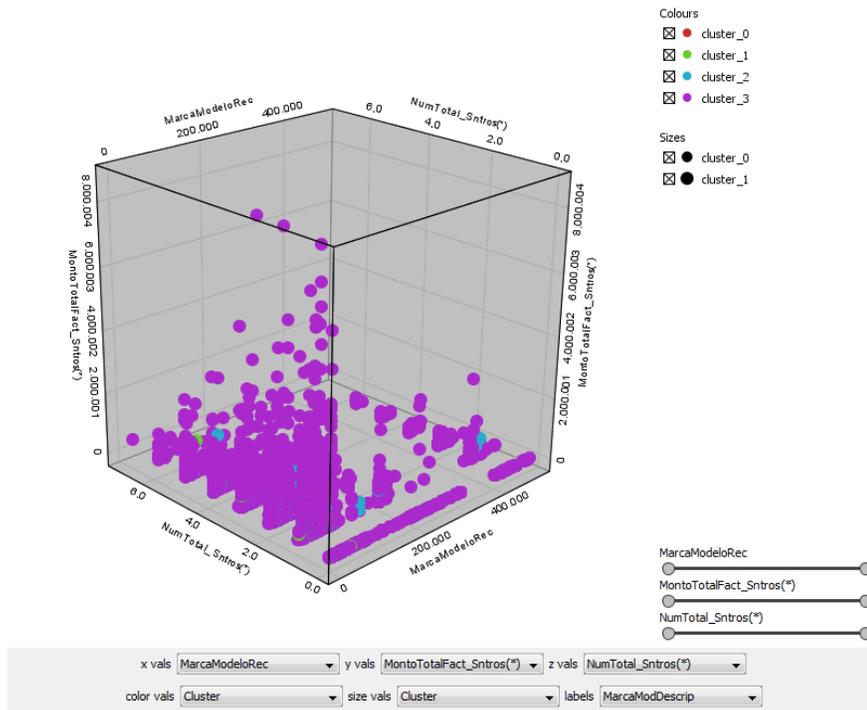


Figura F.3: Conglomerado a través de k medias, con K=4, con variables numéricas normalizadas y variables categóricas transformadas en ficticias de 0 y 1 normalizadas - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura F.3, los grupos se dividen entre los que tienen siniestros, monto facturado de siniestros y los demás grupos con los que no tienen siniestros.

4. Conglomerado con K=5 medias.

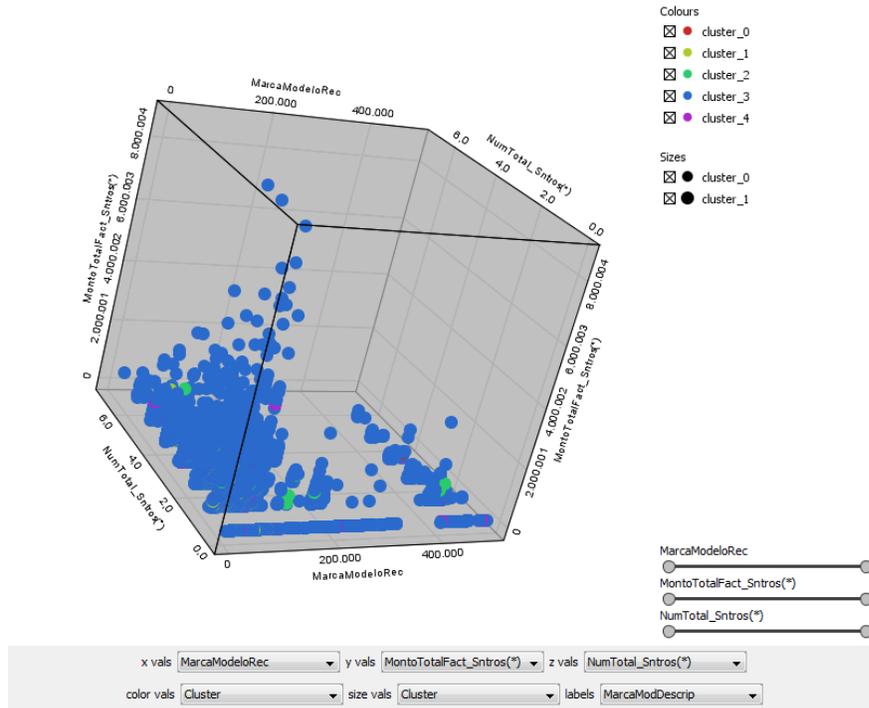


Figura F.4: Conglomerado a través de k medias, con K=5, con variables numéricas normalizadas y variables categóricas transformadas en ficticias de 0 y 1 normalizadas - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura F.4, los grupos se dividen entre los que tienen siniestros, monto facturado de siniestros y los demás grupos con los que no tienen siniestros.

F. CONGLOMERADO A TRAVÉS DE K MEDIAS CON VARIABLES NORMALIZADAS Y VARIABLES CATEGÓRICAS TRANSFORMADAS EN FICTICIAS NORMALIZADAS

5. Conglomerado con K=6 medias.

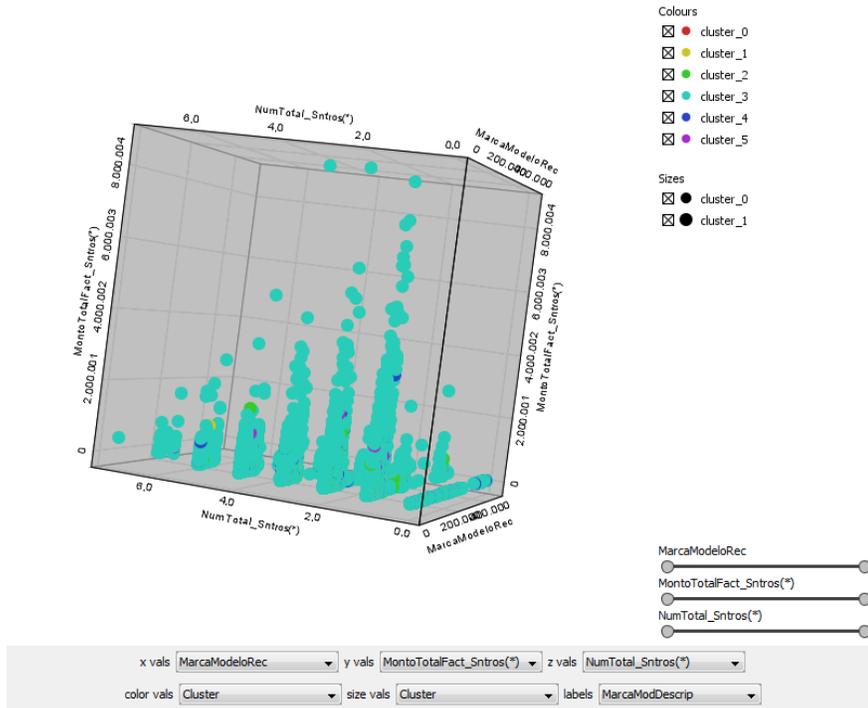


Figura F.5: Conglomerado a través de k medias, con K=6, con variables numéricas normalizadas y variables categóricas transformadas en ficticias de 0 y 1 normalizadas - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura F.5, los grupos se dividen entre los que tienen siniestros, monto facturado de siniestros y los demás grupos con los que no tienen siniestros.

6. Conglomerado con K=7 medias.

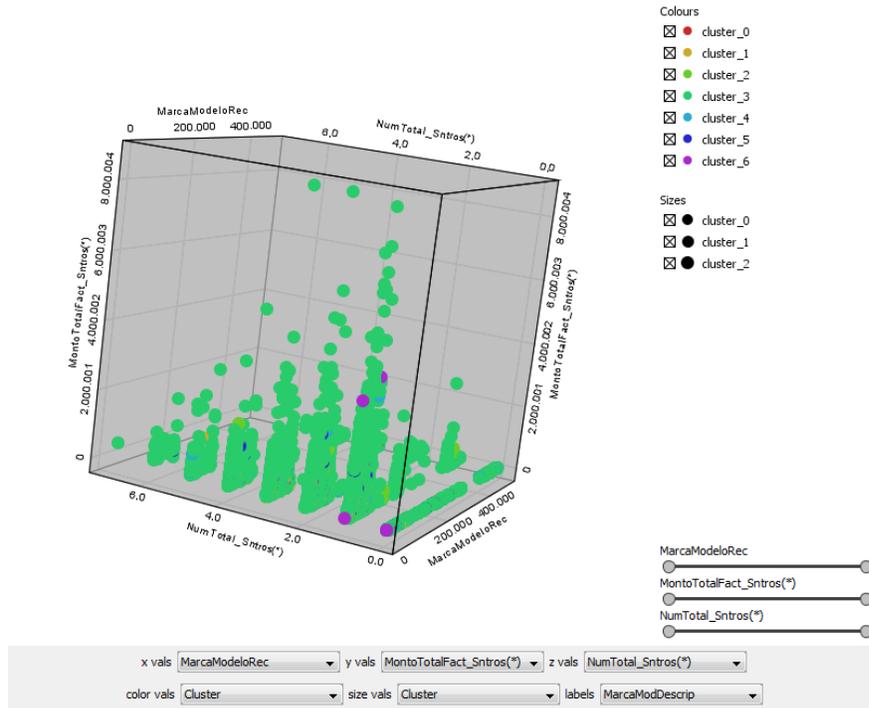


Figura F.6: Conglomerado a través de k medias, con K=7, con variables numéricas normalizadas y variables categóricas transformadas en ficticias de 0 y 1 normalizadas - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura F.6, los grupos se dividen entre los que tienen siniestros, monto facturado de siniestros y los demás grupos con los que no tienen siniestros.

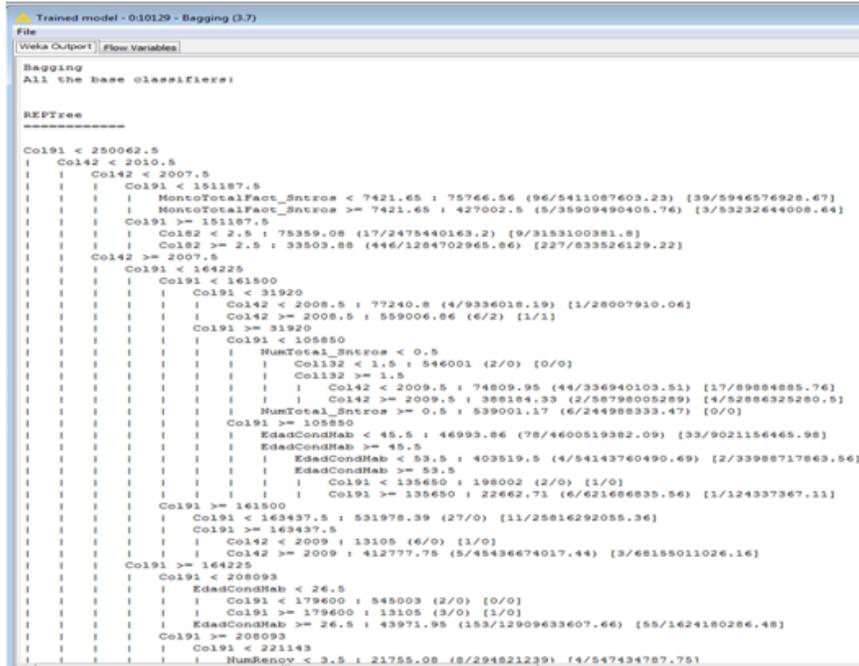
Apéndice G

Conglomerado a través de bagging de weka 3.7

G.1. Proceso

G. CONGLOMERADO A TRAVÉS DE BAGGING DE WEKA 3.7

Se realizó los siguientes grupos.



```
Trained model - 0:10129 - Bagging (3.7)
File
[Weka Output] | Flow Variables
Bagging
All the Base classifiers:
REPTree
-----
Col191 < 250062.5
|
| Col142 < 2010.5
| |
| | Col191 < 151187.5
| | |
| | | MontoTotalFact_Sntros < 7421.65 : 75766.56 (96/5411087603.23) [39/5946576928.47]
| | | MontoTotalFact_Sntros >= 7421.65 : 427002.5 (5/35909490405.76) [3/53232644008.64]
| | | Col191 >= 151187.5
| | | Col182 < 2.5 : 75359.08 (17/2475440163.2) [9/3153100381.8]
| | | Col182 >= 2.5 : 33503.88 (446/1284702965.86) [227/833526129.22]
| | | Col142 >= 2007.5
| | | Col191 < 164225
| | | | Col191 < 161500
| | | | | Col191 < 31920
| | | | | | Col142 < 2008.5 : 77240.8 (4/9336018.19) [1/28007910.06]
| | | | | | Col142 >= 2008.5 : 559006.86 (6/2) [1/1]
| | | | | Col191 >= 31920
| | | | | | Col191 < 105850
| | | | | | | NumTotal_Sntros < 0.5
| | | | | | | | Col132 < 1.5 : 546001 (2/0) [0/0]
| | | | | | | | Col132 >= 1.5
| | | | | | | | | Col142 < 2009.5 : 74809.95 (44/336940103.51) [17/89884885.76]
| | | | | | | | | Col142 >= 2009.5 : 388184.33 (2/58798005289) [4/52886325280.5]
| | | | | | | | | NumTotal_Sntros >= 0.5 : 539001.17 (6/244988333.47) [0/0]
| | | | | | | | Col191 >= 105850
| | | | | | | | | EdadCondHab < 45.5 : 46993.86 (78/4600519382.09) [33/9021156465.98]
| | | | | | | | | EdadCondHab >= 45.5
| | | | | | | | | | EdadCondHab < 53.5 : 403519.5 (4/54143760490.69) [2/33988717863.56]
| | | | | | | | | | EdadCondHab >= 53.5
| | | | | | | | | | | Col191 < 135650 : 198002 (2/0) [1/0]
| | | | | | | | | | | Col191 >= 135650 : 22662.71 (6/621686835.56) [1/124337367.11]
| | | | | | | | | | | Col191 < 161500
| | | | | | | | | | | | Col191 < 163437.5 : 531978.39 (27/0) [11/25816292055.36]
| | | | | | | | | | | | Col191 >= 163437.5
| | | | | | | | | | | | | Col142 < 2009 : 13105 (6/0) [1/0]
| | | | | | | | | | | | | Col142 >= 2009 : 412777.75 (5/45436674017.44) [3/68155011026.16]
| | | | | | | | | | | | Col191 >= 164225
| | | | | | | | | | | | | Col191 < 208093
| | | | | | | | | | | | | | EdadCondHab < 26.5
| | | | | | | | | | | | | | | Col191 < 179600 : 848003 (2/0) [0/0]
| | | | | | | | | | | | | | | Col191 >= 179600 : 13105 (3/0) [1/0]
| | | | | | | | | | | | | | | EdadCondHab >= 26.5 : 43971.95 (153/12909633607.66) [55/1624180286.48]
| | | | | | | | | | | | | | Col191 >= 208093
| | | | | | | | | | | | | | | Col191 < 221143
| | | | | | | | | | | | | | | | NumRenov < 3.5 : 21755.08 (8/294821239) [4/547434787.75]
```

Figura G.1: Conglomerado a través de bagging con Weka 3.7 - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura G.1, en primera instancia se realizaron 20.700 árboles.

Conglomerado a través de bosque aleatorio

H.1. Proceso

Se realizó los siguientes grupos.

Con respecto a la figura [H.1](#), en primera instancia se realizaron separaciones aleatorias.

H. CONGLOMERADO A TRAVÉS DE BOSQUE ALEATORIO

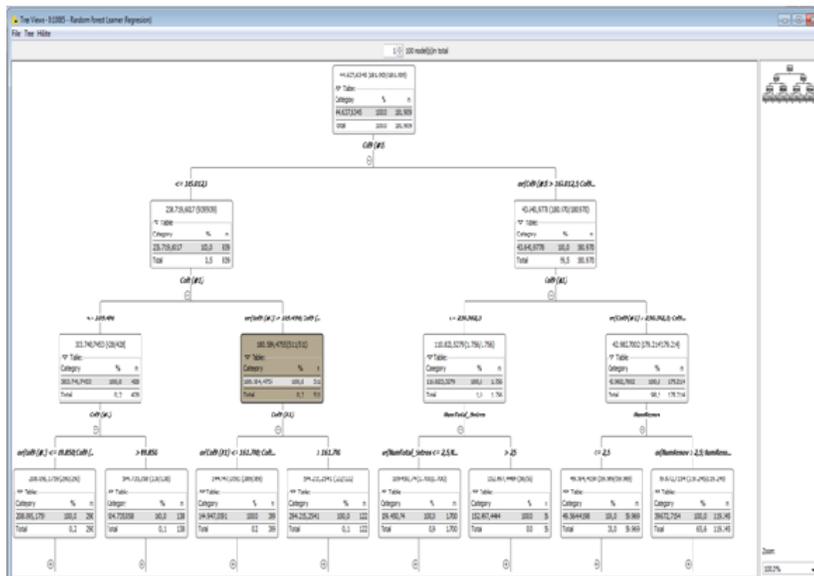


Figura H.1: Conglomerado a través del aprendizaje del bosque aleatorio - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Conglomerado a través del análisis de componentes principales

I.1. Análisis de componentes principales

Se realizó análisis de componentes principales para obtener los grupos óptimos.

I. CONGLOMERADO A TRAVÉS DEL ANÁLISIS DE COMPONENTES PRINCIPALES

1. Matriz de covarianza.

The screenshot shows a software window titled "Covariance matrix - 0:9459 - PCA Compute". Below the title bar is a menu bar with "File". Below the menu bar is a table with the following data:

Row ID	D Col4 (#2)	D EdadCo...	D NumTot...	D MontoT...
Col4 (#2)	1.046	-0.122	0.049	0.04
EdadCondHab	-0.122	1.036	-0.015	-0.013
NumTotal_Snt...	0.049	-0.015	0.998	0.397
MontoTotalFa...	0.04	-0.013	0.397	0.977

Figura I.1: Estudio transversal de conglomerado a través del análisis de componentes principales, se presenta la matriz de covarianza - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura I.1 Estudio transversal de conglomerado a través del análisis de componentes principales, se presenta la matriz de covarianza, se observa la covariabilidad presentada por las variables valor del vehículo, edad del conductor, número de siniestros y monto facturado de siniestros subtotalizado por vehículo, lo que muestra que existe poca covariabilidad entre las variables.

2. Autovalores.

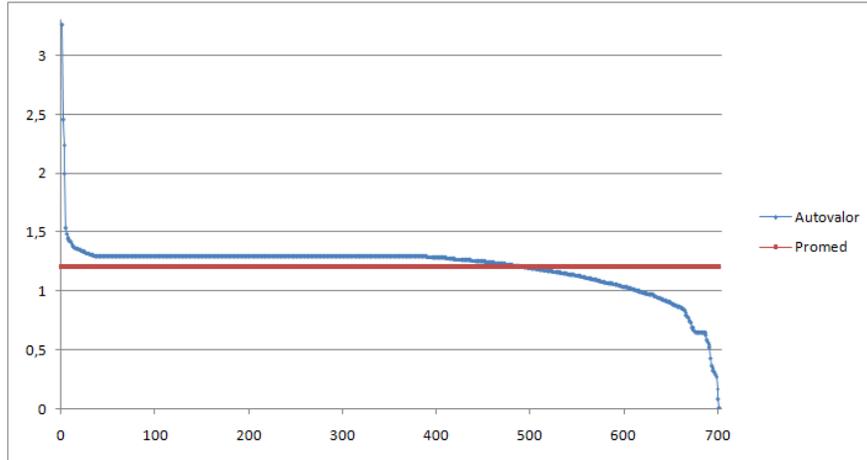


Figura I.2: Estudio transversal de conglomerado a través del análisis de componentes principales, se presenta los autovalores - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura I.2 Estudio transversal de conglomerado a través del análisis de componentes principales, se presenta los autovalores, los grupos óptimos a través del análisis de componentes principales se obtiene con los autovalores, contabilizando los que, la suma de los mayores, tengan la mayor cantidad de información, en este punto se puede demostrar con el test, que es necesario realizar cinco (5) grupos.

Conglomerado a través k medias con la variable marca modelo

Se realizó grupos a través de k medias con las variables numéricas normalizadas, con solo la variable categórica "marca modelo" transformada en variables ficticia de valores 0 y 1 normalizada.

J. CONGLOMERADO A TRAVÉS K MEDIAS CON LA VARIABLE MARCA MODELO

1. Conglomerado con K=2 medias, marca modelo.

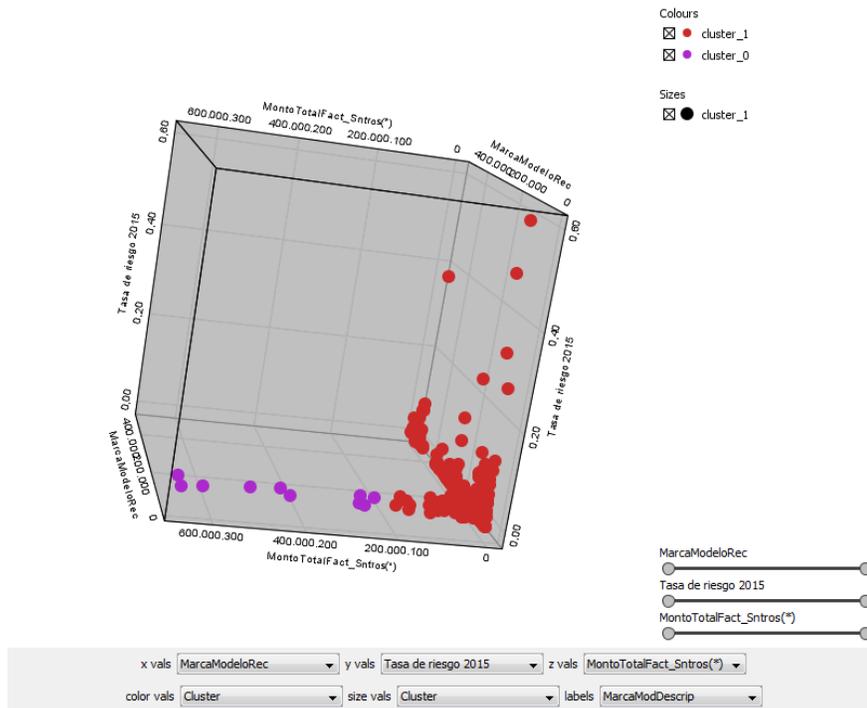


Figura J.1: Conglomerado a través de k medias, con K=2, con las variables numéricas normalizadas, con solo la variable categórica "marca modelo" transformada en variable ficticia de valores 0 y 1 normalizada - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura J.1, los grupos se dividen entre las marcas modelos que tienen el total del monto facturado de siniestros superior a Bs. 200 Millones y las demás marcas modelos.

2. Conglomerado con $K=3$ medias, marca modelo.

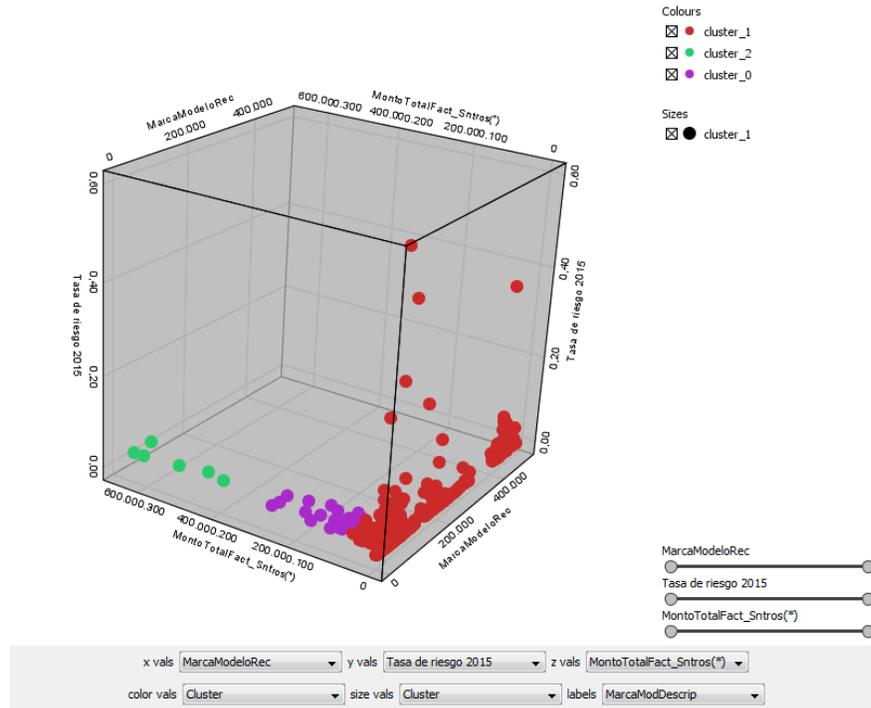


Figura J.2: Conglomerado a través de k medias, con $K=3$, con las variables numéricas normalizadas, con solo la variable categórica "marca modelo" transformada en variable ficticia de valores 0 y 1 normalizada - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura J.2, los grupos se dividen entre las marcas modelos que tienen total monto facturado de siniestros: a) Menores de Bs. 100 Millones, b) Aproximadamente entre Bs. 100 Millones y Bs. 300 Millones, c) Mayores de Bs. 300 Millones.

J. CONGLOMERADO A TRAVÉS K MEDIAS CON LA VARIABLE MARCA MODELO

3. Conglomerado con K=4 medias, marca modelo.

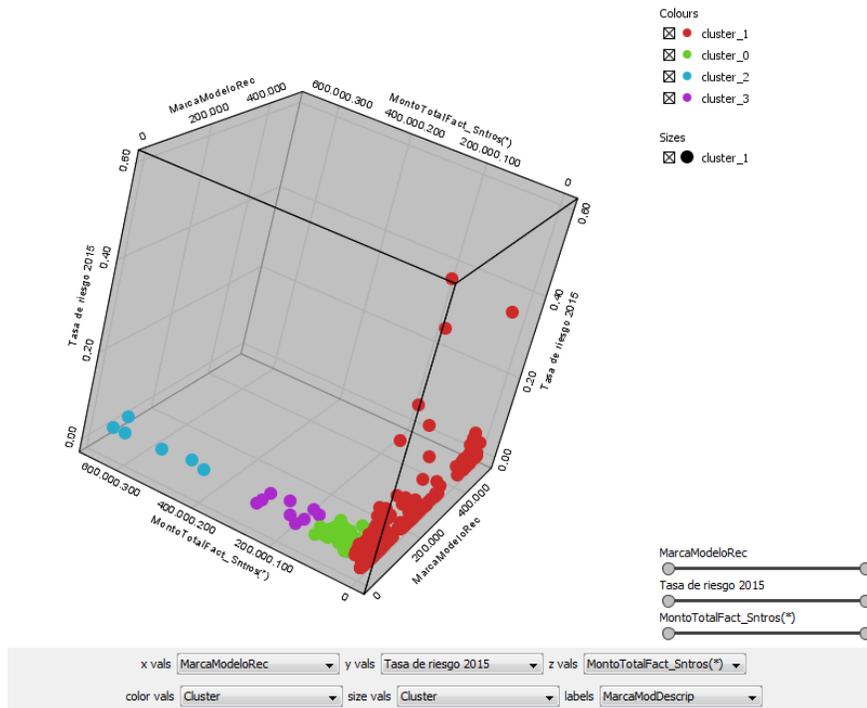


Figura J.3: Conglomerado a través de k medias, con K=4, con las variables numéricas normalizadas, con solo la variable categórica "marca modelo" transformada en variable ficticia de valores 0 y 1 normalizada - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura J.3, los grupos se dividen entre las marcas modelos que tienen total monto facturado de siniestros: a) Menores de Bs. 50 Millones, b) Aproximadamente entre Bs. 50 Millones y Bs. 180 Millones, c) Aproximadamente entre Bs. 180 Millones y Bs. 300 Millones, d) Mayores de Bs. 300 Millones.

4. Conglomerado con K=5 medias, marca modelo.

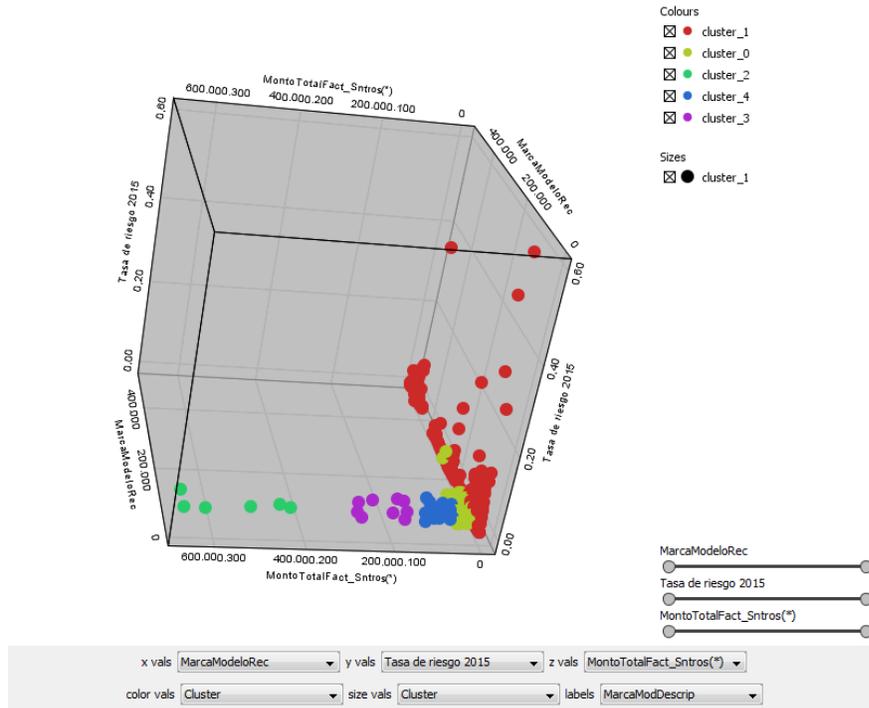


Figura J.4: Conglomerado a través de k medias, con K=5, con las variables numéricas normalizadas, con solo la variable categórica "marca modelo" transformada en variable ficticia de valores 0 y 1 normalizada - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura J.4, los grupos se dividen entre las marcas modelos que tienen total monto facturado de siniestros: a) Menores de Bs. 20 Millones, b) Aproximadamente entre Bs. 20 Millones y Bs. 50 Millones, c) Aproximadamente entre Bs. 50 Millones y Bs. 120 Millones, d) Aproximadamente entre Bs. 120 Millones y Bs. 300 Millones, e) Mayores de Bs. 300 Millones.

5. Conglomerado con K=6 medias, marca modelo.

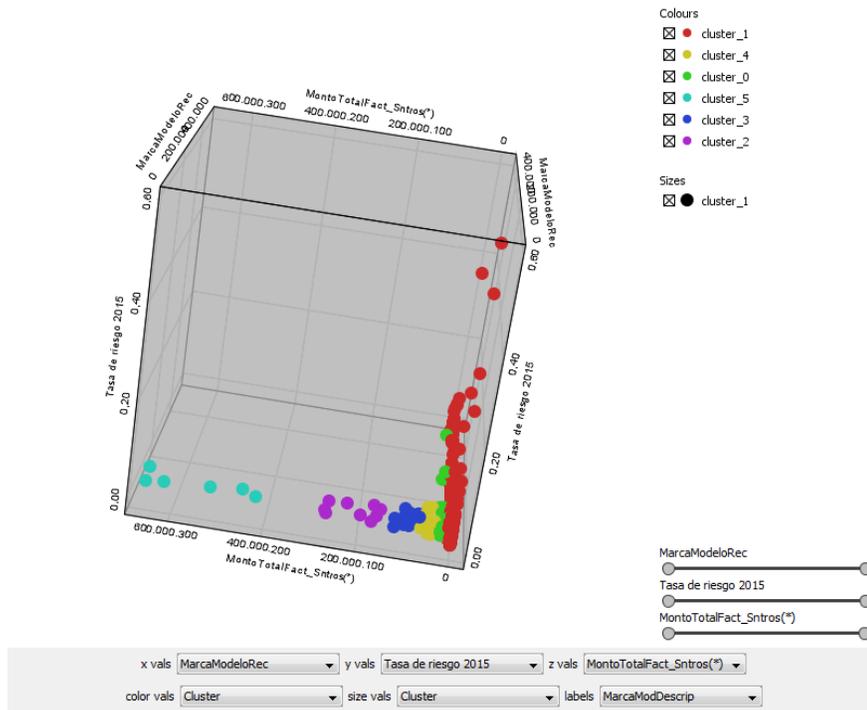


Figura J.5: Conglomerado a través de k medias, con K=6, con las variables numéricas normalizadas, con solo la variable categórica "marca modelo" transformada en variable ficticia de valores 0 y 1 normalizada - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura J.5, los grupos se dividen entre las marcas modelos que tienen total monto facturado de siniestros: a) Menores de Bs. 5 Millones, b) Aproximadamente entre Bs. 5 Millones y Bs. 10 Millones, c) Aproximadamente entre Bs. 10 Millones y Bs. 50 Millones, d) Aproximadamente entre Bs. 50 Millones y Bs. 120 Millones, e) Aproximadamente entre Bs. 120 Millones y Bs. 300 Millones, f) Mayores de Bs. 300 Millones.

6. Conglomerado con K=7 medias, marca modelo.

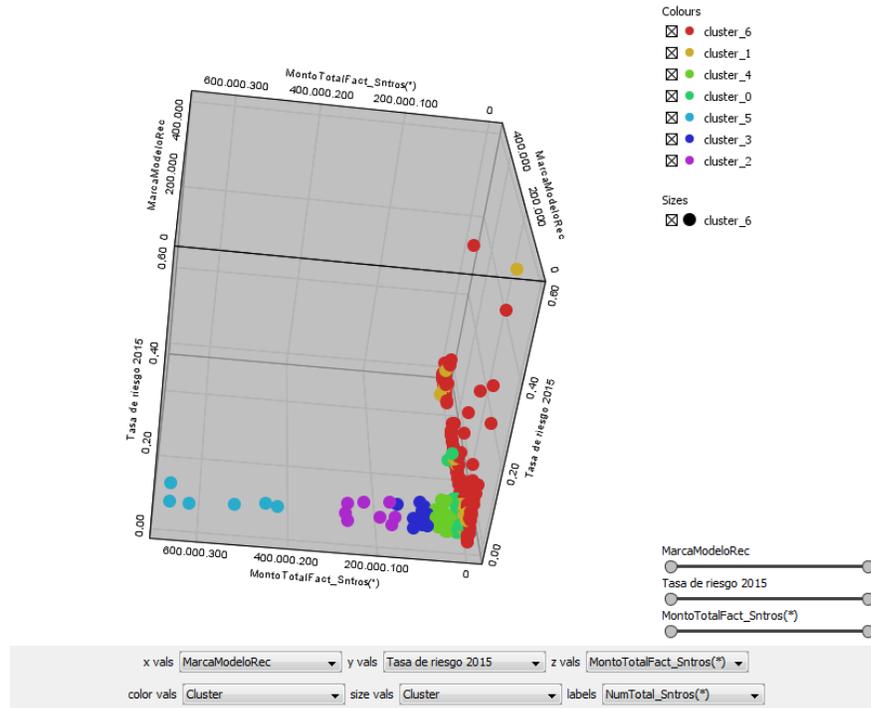


Figura J.6: Conglomerado a través de k medias, con K=7, con las variables numéricas normalizadas, con solo la variable categórica "marca modelo" transformada en variable ficticia de valores 0 y 1 normalizada - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura J.6, los grupos se dividen entre las marcas modelos que tienen total monto facturado de siniestros: a) Menores de Bs. 1 Millón, b) Aproximadamente entre Bs. 1 Millón y Bs. 5 Millones, c) Aproximadamente entre Bs. 5 Millones y Bs. 10 Millones, d) Aproximadamente entre Bs. 10 Millones y Bs. 50 Millones, e) Aproximadamente entre Bs. 50 Millones y Bs. 120 Millones, f) Aproximadamente entre Bs. 120 Millones y Bs. 300 Millones, g) Mayores de Bs. 300 Millones.

Conglomerado a través del análisis discriminante, apoyado en k medias

Se realizó análisis discriminante sobre los grupos encontrados a través de k medias con las variables numéricas normalizadas, con las variables categóricas marca modelo transformadas en variables ficticias de valores 0 y 1.

K. CONGLOMERADO A TRAVÉS DEL ANÁLISIS DISCRIMINANTE, APOYADO EN K MEDIAS

1. Conglomerado a través del análisis discriminante para K=2 medias.

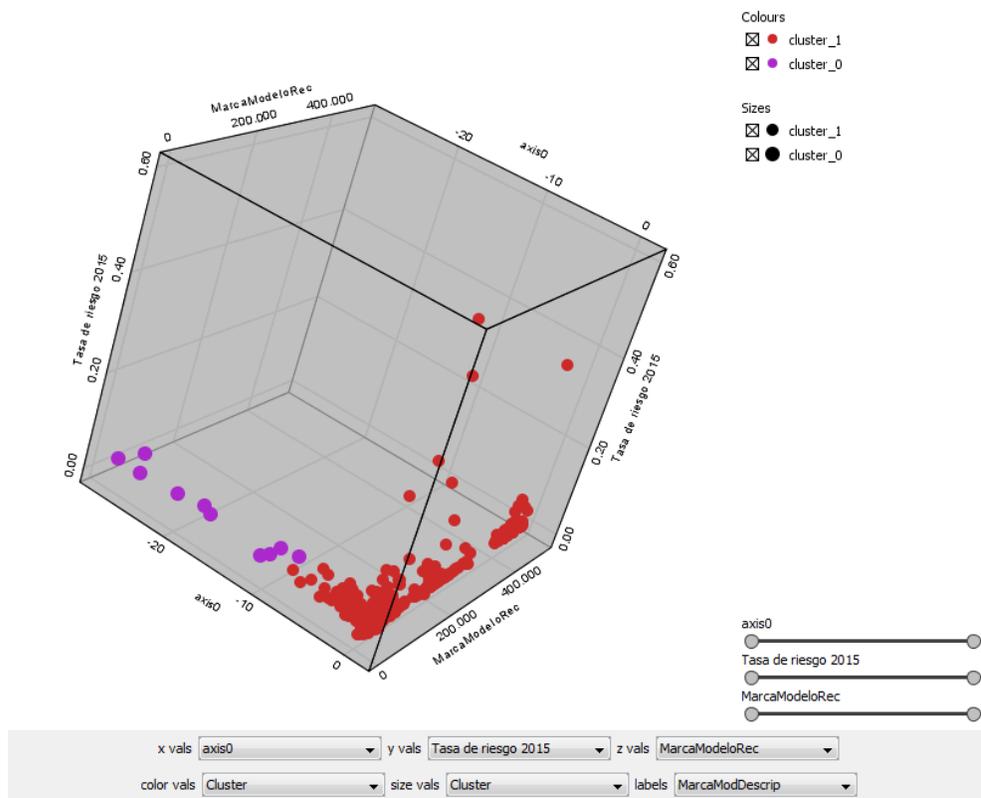


Figura K.1: Parámetros del análisis discriminante a un conglomerado con K=2

- Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura K.1, los grupos se dividen entre las marcas modelos que tienen total del monto facturado de siniestros, el resultado del análisis muestra que el parámetro 1 es igual a 10 en la ecuación discriminante.

2. Conglomerado a través del análisis discriminante para $k=3$ medias.

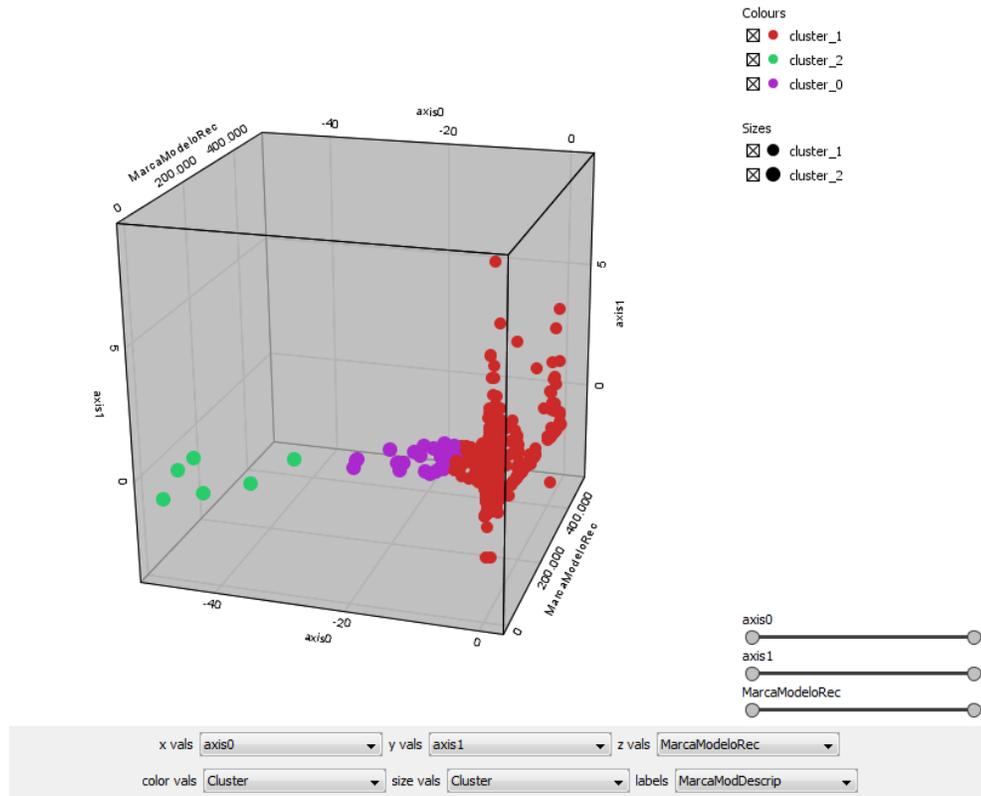


Figura K.2: Parámetros del análisis discriminante a un conglomerado con $K=3$

- Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura K.2, los grupos se dividen entre las marcas modelos que tienen total monto facturado de siniestro, el resultado del análisis muestra que el parámetro 1 es igual a 10, el parámetro 2 es igual a 30 en la ecuación discriminante.

K. CONGLOMERADO A TRAVÉS DEL ANÁLISIS DISCRIMINANTE, APOYADO EN K MEDIAS

3. Conglomerado a través del análisis discriminante para K=4 medias.

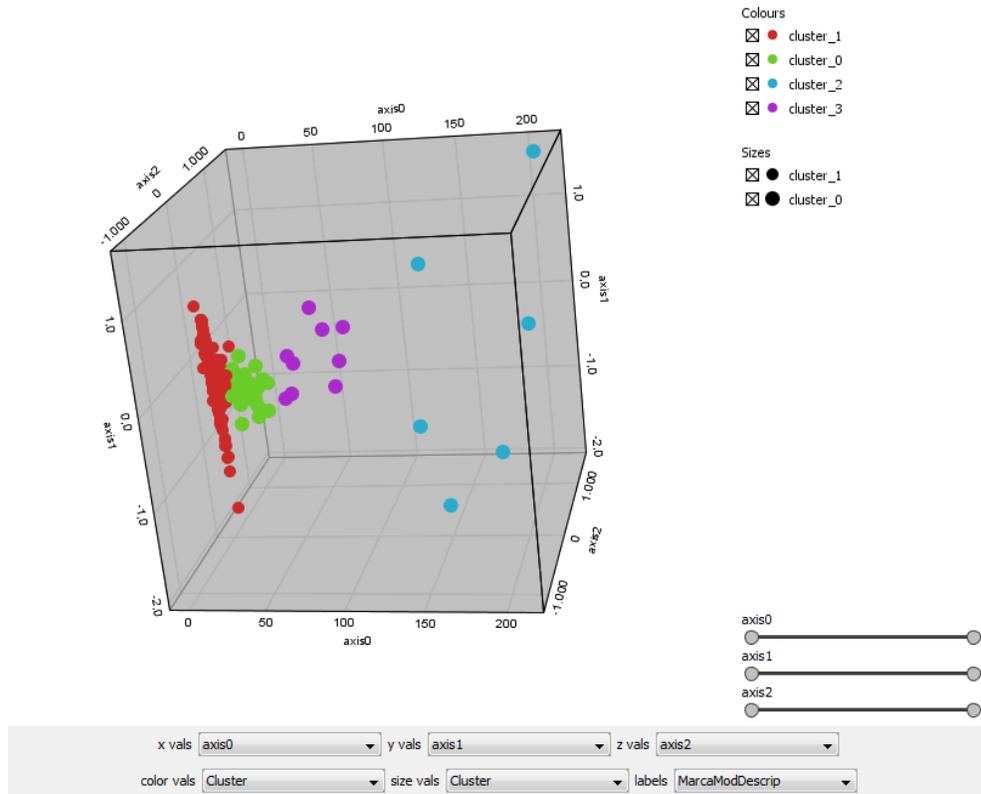


Figura K.3: Parámetros del análisis discriminante a un conglomerado con K=4

- Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura K.3, los grupos se dividen entre las marcas modelos que tienen total monto facturado de siniestros, el resultado del análisis muestra que el parámetro 1 es igual a 10, el parámetro 2 es igual a 30, el parámetro 3 es igual a 70 aproximadamente en la ecuación discriminante.

4. Conglomerado a través del análisis discriminante para $K=5$ medias.

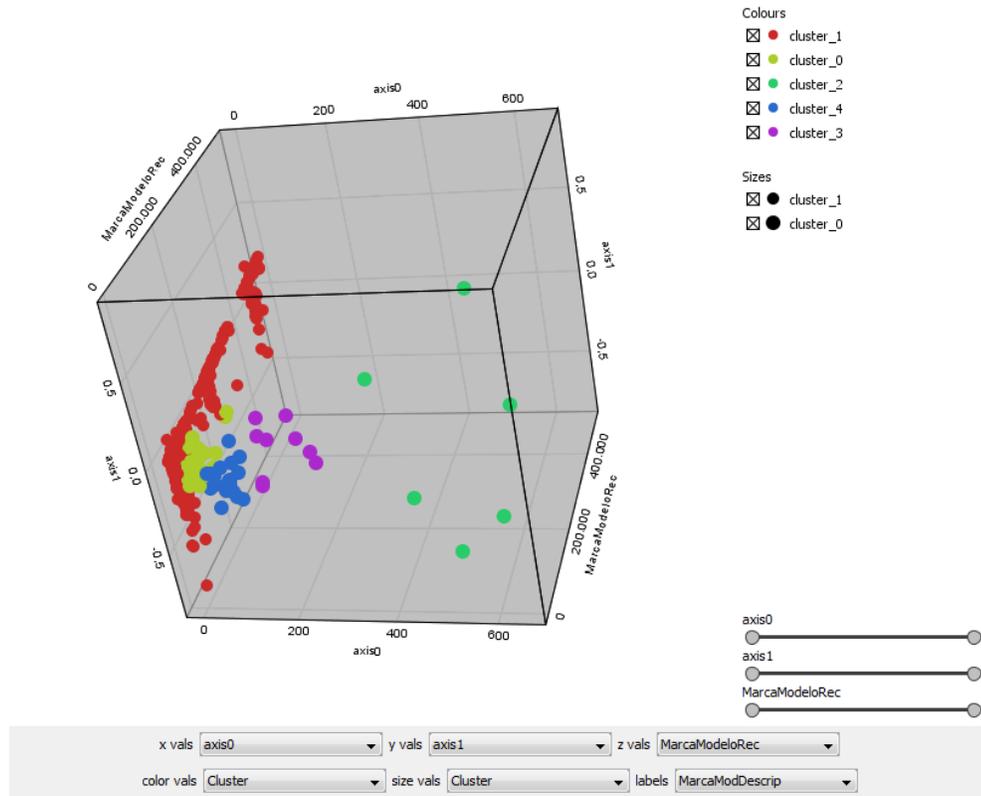


Figura K.4: Parámetros del análisis discriminante a un conglomerado con $K=5$

- Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura K.4, los grupos se dividen entre las marcas modelos que tienen total monto facturado de siniestros, el resultado del análisis muestra que el parámetro 1 es igual a 10, el parámetro 2 es igual a 30, el parámetro 3 es aproximadamente a 70 y el parámetro 4 está cerca de 200 en la ecuación discriminante.

K. CONGLOMERADO A TRAVÉS DEL ANÁLISIS DISCRIMINANTE, APOYADO EN K MEDIAS

5. Conglomerado a través del análisis discriminante para K=6 medias.

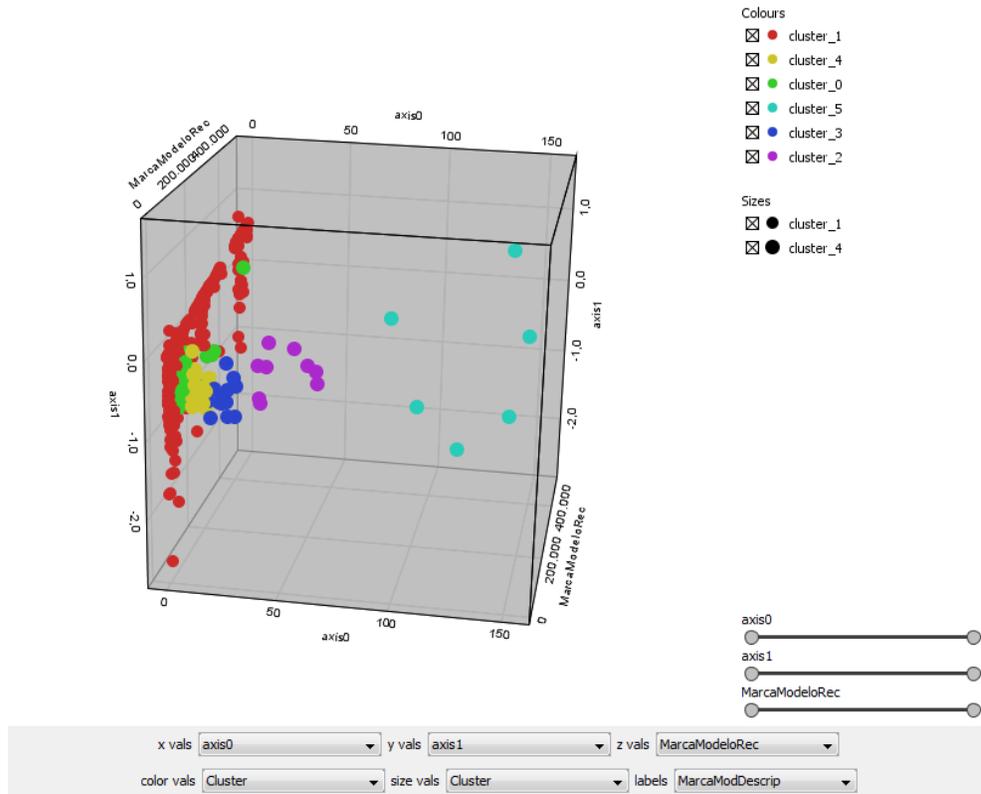


Figura K.5: Parámetros del análisis discriminante a un conglomerado con K=6

- Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura K.5, los grupos se dividen entre las marcas modelos que tienen total monto facturado de siniestros, el resultado del análisis muestra que el parámetro 1 es igual a 4, el parámetro 2 es igual a 6, el parámetro 3 es aproximadamente a 10, el parámetro 4 está cerca de 25, el parámetro 5 está cerca de 50 en la ecuación discriminante.

6. Conglomerado a través del análisis discriminante para la agrupación con $K=7$ medias.

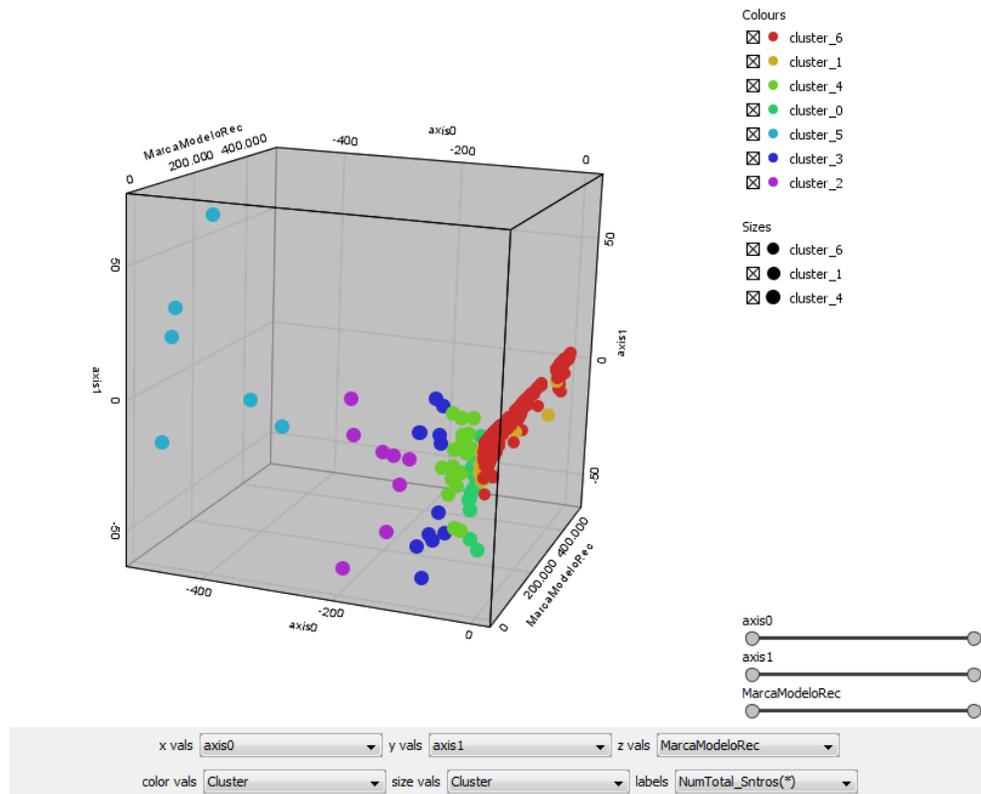


Figura K.6: Parámetros del análisis discriminante a un conglomerado con $K=7$

- Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura K.6, los grupos se dividen entre las marcas modelos que tienen total monto facturado de siniestros, el resultado del análisis muestra que el parámetro 1 es igual a 1, el parámetro 2 es igual a 6, el parámetro 3 es aproximadamente a 10, el parámetro 4 está cerca de 25, el parámetro 5 está cerca de 120, el parámetro 6 está cerca de 400 en la ecuación discriminante.

Estudio longitudinal, conglomerado a
través del análisis de componentes
principales

L.1. Primer proceso: asignación revisión de datos

Se realizó los cálculos correspondientes del análisis de componentes principales con el paquete de R factorMiner, dando los siguientes resultados:

L. ESTUDIO LONGITUDINAL, CONGLOMERADO A TRAVÉS DEL ANÁLISIS DE COMPONENTES PRINCIPALES

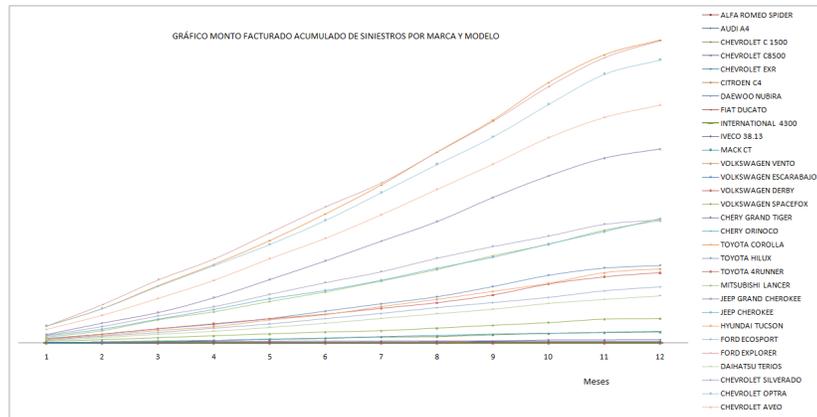


Figura L.1: Estudio longitudinal, conglomerado a través del análisis de componentes principales, gráfico de las curvas de las marcas modelos por monto total facturado acumulado - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura L.1, en primera instancia se muestra las curvas descritas para posteriormente realizar la correspondiente agrupación.

L.2. Segundo proceso: resultados

La cantidad de grupos óptimos resultantes es de Tres (3).

Autovalores:

	eigenvalue	percentage of variance	cumulative percentage of variance
comp 1	11,97246	99,77054	99,77054
comp 2	0,01781	0,14841	99,91895
comp 3	0,00462	0,03848	99,95743
comp 4	0,00193	0,01609	99,97352
comp 5	0,00156	0,01296	99,98648
comp 6	0,00057	0,00476	99,99124
comp 7	0,00034	0,00284	99,99409
comp 8	0,00023	0,00195	99,99604
comp 9	0,00019	0,00155	99,99759
comp 10	0,00013	0,00104	99,99863
comp 11	0,00010	0,00087	99,99950
comp 12	0,00006	0,00050	100,00000

Figura L.2: Estudio longitudinal, conglomerado a través del análisis de componentes principales, tabla de autovalores - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura L.2, se muestra la tabla con autovalores y donde se aprecia que con tres (3) grupos se obtiene más del noventa y nueve por ciento de la información.

L. ESTUDIO LONGITUDINAL, CONGLOMERADO A TRAVÉS DEL ANÁLISIS DE COMPONENTES PRINCIPALES

Resultado con gráfico:

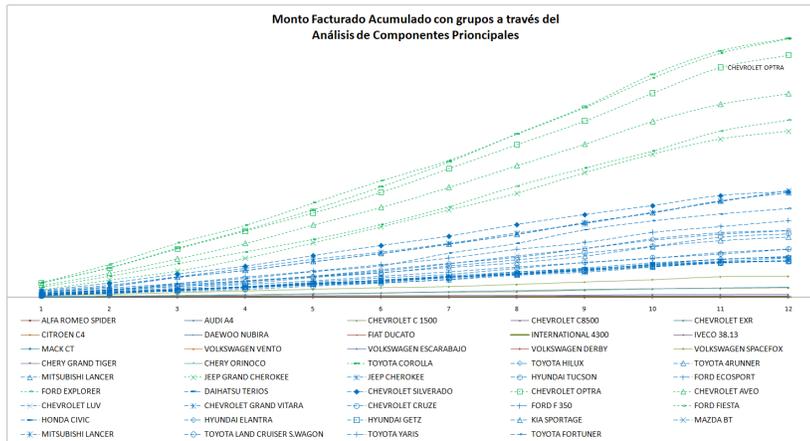


Figura L.3: Estudio longitudinal, conglomerado a través del análisis de componentes principales, gráfico de las curvas de las marcas modelos por monto total facturado acumulado y agrupado - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura L.3, se muestra las curvas descritas después de realizado la correspondiente agrupación a través del análisis de componentes principales longitudinal.

Estudio longitudinal, conglomerado a través de k medias, con $k=2$ medias y la distancia propuesta por Chiodi

M.1. Primer proceso: asignación para $k=2$, distancia euclideana

Se realizó los siguientes grupos, de acuerdo a la distancia euclideana y dos puntos semillas previamente seleccionados que son la curva con mayor montos y la curva de menor montos para encontrar la máxima separación.

M. ESTUDIO LONGITUDINAL, CONGLOMERADO A TRAVÉS DE K MEDIAS,
 CON K=2 MEDIAS Y LA DISTANCIA PROPUESTA POR CHIODI

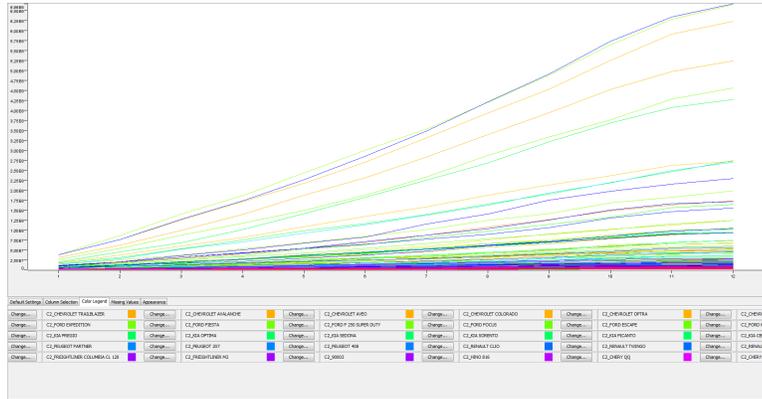


Figura M.1: Estudio longitudinal, conglomerado a través de k medias, con k=2 grupo uno, primera asignación, distancia euclideana - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura M.1, se muestra el primer grupo de la primera asignación, distancia euclideana.

M.1 Primer proceso: asignación para $k=2$, distancia euclídeana

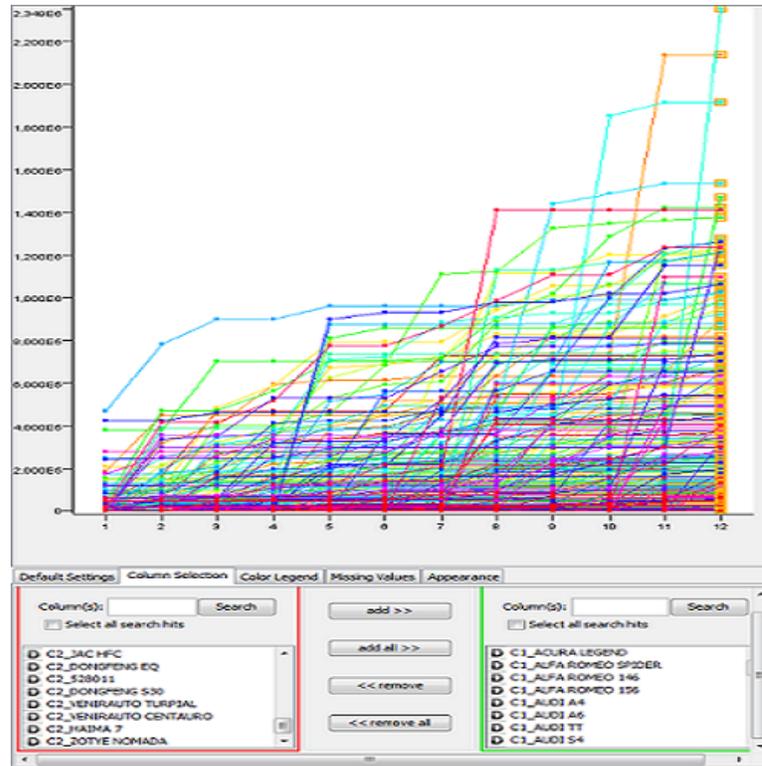


Figura M.2: Estudio longitudinal, conglomerado a través de k medias, con $k=2$ grupo dos, primera asignación, distancia euclídeana - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura M.2, se muestra el segundo grupo de la primera asignación, distancia euclídeana.

M.2. Segundo proceso de asignación para k=2, distancia Chiodi

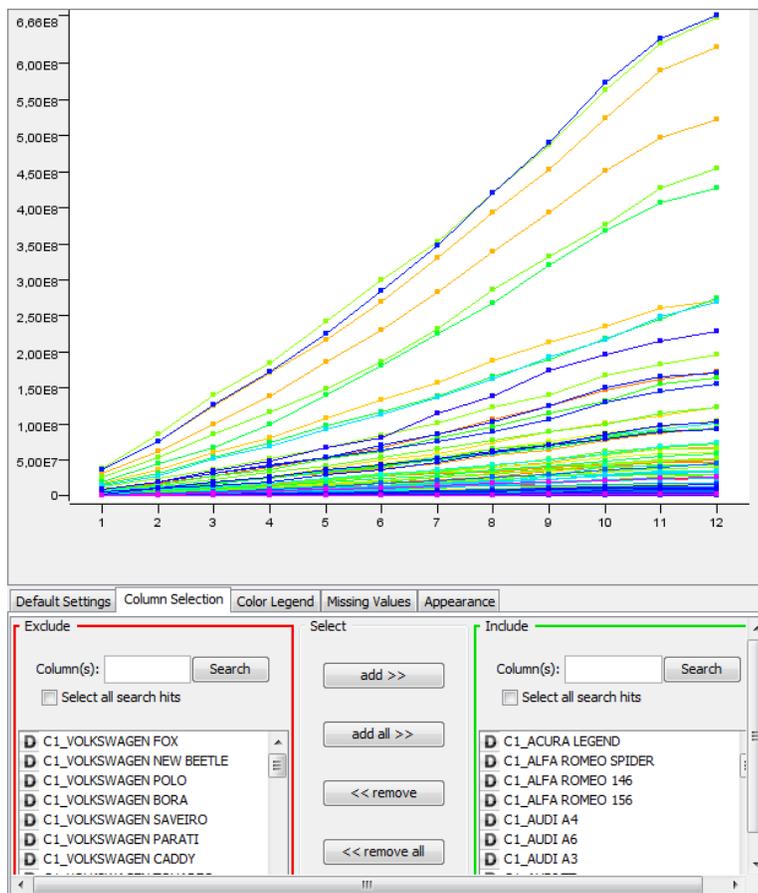


Figura M.3: Estudio longitudinal, conglomerado a través de k medias con distancia de Chiodi, k=2 grupo uno, segunda asignación - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

M.2 Segundo proceso de asignación para $k=2$, distancia Chiodi

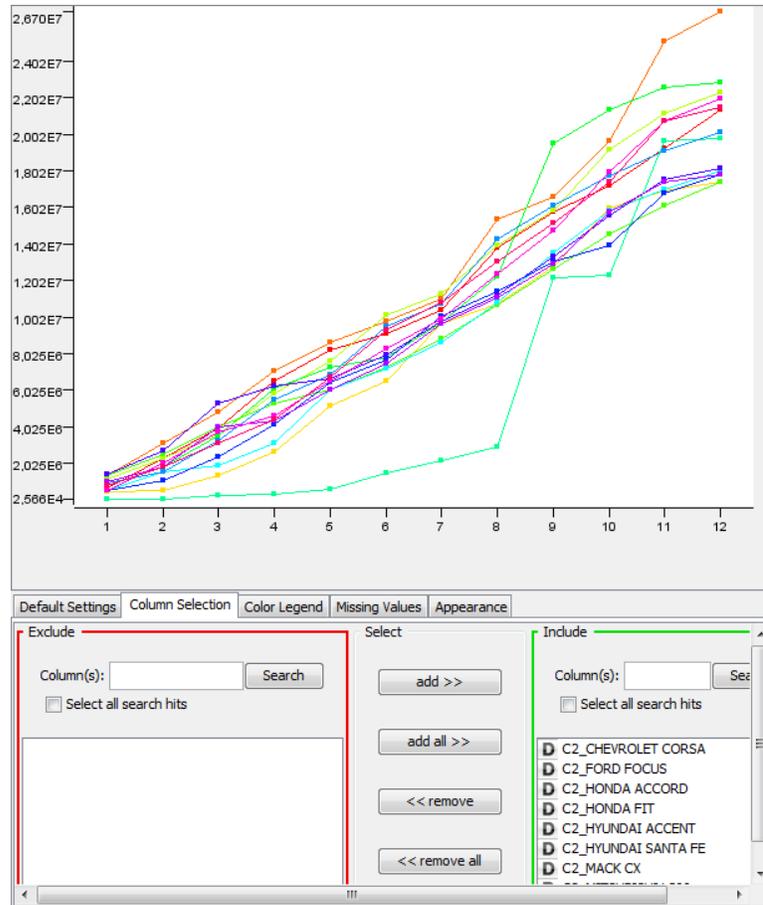


Figura M.4: Estudio longitudinal, conglomerado a través de k medias con distancia de Chiodi, $k=2$ grupo dos, segunda asignación - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura M.4, se muestra el segundo grupo de la segunda asignación, distancia de Chiodi.

M.3. Tercer proceso: asignación para k=2, confirmación distancia de Chiodi

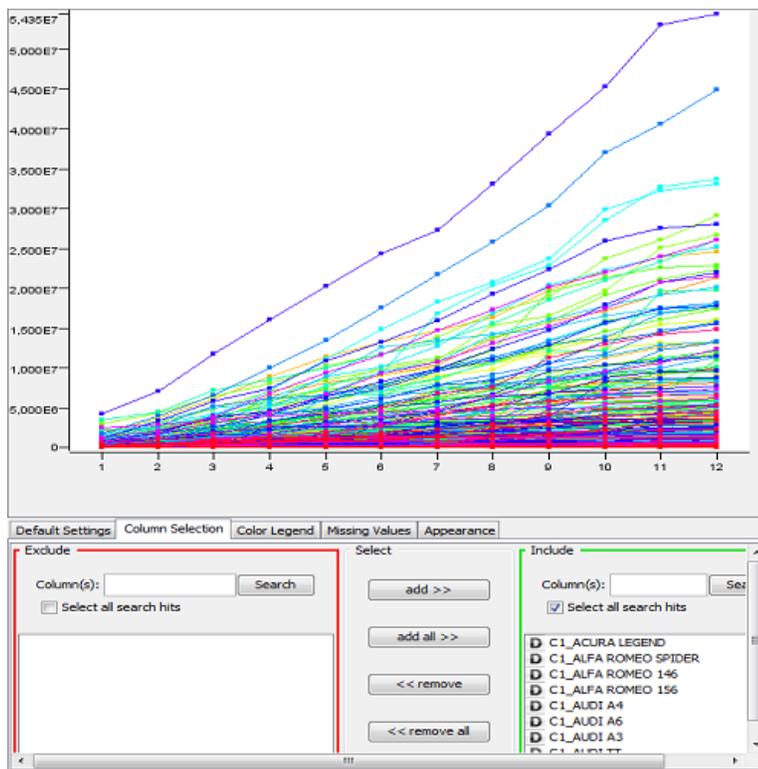


Figura M.5: Estudio longitudinal, conglomerado a través de k medias con distancia de Chiodi, k=2 grupo uno, tercera asignación, confirmación - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

M.3 Tercer proceso: asignación para $k=2$, confirmación distancia de Chiodi

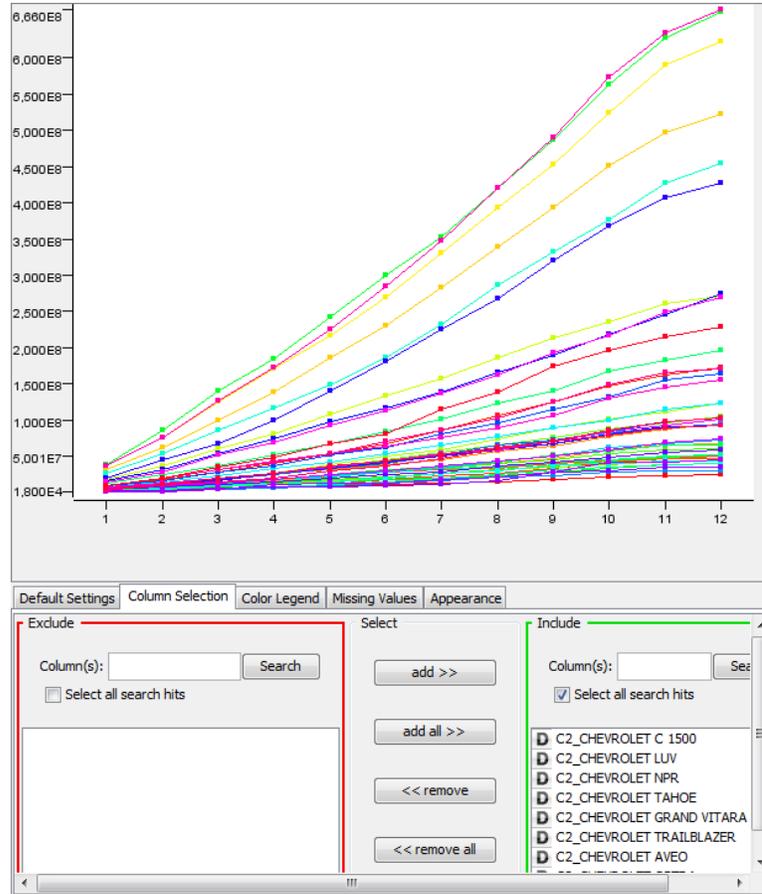


Figura M.6: Estudio longitudinal, conglomerado a través de k medias con distancia de Chiodi, $k=2$ grupo dos, tercera asignación, confirmación - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura M.6, se muestra el segundo grupo de la tercera asignación, confirmación distancia de Chiodi.

M. ESTUDIO LONGITUDINAL, CONGLOMERADO A TRAVÉS DE K MEDIAS, CON K=2 MEDIAS Y LA DISTANCIA PROPUESTA POR CHIODI

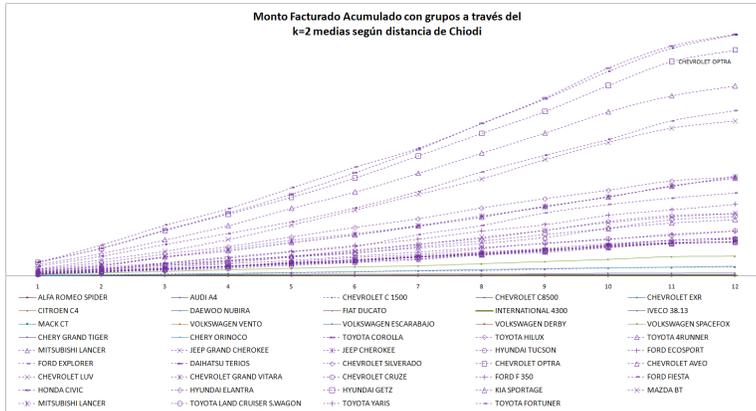


Figura M.7: Estudio longitudinal, conglomerado a través de k medias con distancia de Chiodi, k=2, se muestra los dos grupos, tercera asignación, confirmación - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Con respecto a la figura M.7, se da una muestra de los dos grupo de la tercera asignación, confirmación distancia de Chiodi.

Apéndice N

Determinación de la estimación de las
pérdidas agregadas

N.1. Asociar la función de ocurrencia

Se realizó la siguientes asociaciones:

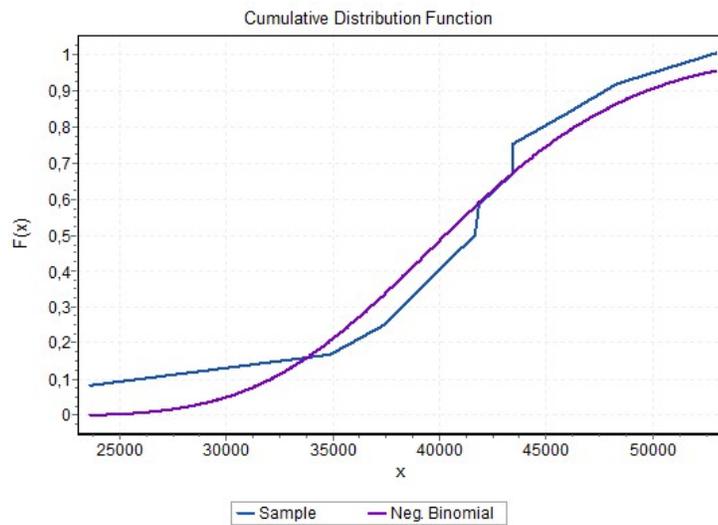


Figura N.1: Función de distribución acumulada binomial negativa - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Fitting Results		
#	Distribution	Parameters
1	D. Uniform	a=28891 b=53171
2	Geometric	p=2,43712225E-5
3	Logarithmic	$\theta=0,999998154$
4	Neg. Binomial	n=34 p=8,35123574E-4
5	Poisson	$\lambda=41031$
6	Bernoulli	No fit (data max > 1)
7	Binomial	No fit
8	Hypergeometric	No fit

Figura N.2: Parámetros de la binomial negativa - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

N. DETERMINACIÓN DE LA ESTIMACIÓN DE LAS PÉRDIDAS AGREGADAS

Goodness of Fit - Details					
D. Uniform [#1]					
Kolmogorov-Smirnov					
Sample Size	12				
Statistic	0,182495229				
P-Value	0,755633822				
Rank	2				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	0,29577	0,33815	0,37543	0,41918	0,44905
Reject?	No	No	No	No	No
Anderson-Darling					
Sample Size	12				
Statistic	4,17179389				
Rank	3				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	1,37490912	1,9286236	2,50175596	3,28919661	3,90741594
Reject?	Yes	Yes	Yes	Yes	Yes

[\[top\]](#)

Geometric [#2]					
Kolmogorov-Smirnov					
Sample Size	12				
Statistic	0,489337404				
P-Value	0,003582625				
Rank	3				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	0,29577	0,33815	0,37543	0,41918	0,44905
Reject?	Yes	Yes	Yes	Yes	Yes
Anderson-Darling					
Sample Size	12				
Statistic	3,89707655				
Rank	2				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	1,37490912	1,9286236	2,50175596	3,28919661	3,90741594
Reject?	Yes	Yes	Yes	Yes	No

Figura N.3: Bondad de ajuste de la variable discreta 1 y 2 - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Logarithmic [#3]					
Kolmogorov-Smirnov					
Sample Size	12				
Statistic	0,803276488				
P-Value	7,22959585E-8				
Rank	5				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	0,29577	0,33815	0,37543	0,41918	0,44905
Reject?	Yes	Yes	Yes	Yes	Yes
Anderson-Darling					
Sample Size	12				
Statistic	11,6038247				
Rank	4				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	1,37490912	1,9286236	2,50175596	3,28919661	3,90741594
Reject?	Yes	Yes	Yes	Yes	Yes

[\[top\]](#)

Neg. Binomial [#4]					
Kolmogorov-Smirnov					
Sample Size	12				
Statistic	0,166868248				
P-Value	0,838872302				
Rank	1				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	0,29577	0,33815	0,37543	0,41918	0,44905
Reject?	No	No	No	No	No
Anderson-Darling					
Sample Size	12				
Statistic	0,564373882				
Rank	1				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	1,37490912	1,9286236	2,50175596	3,28919661	3,90741594
Reject?	No	No	No	No	No

Figura N.4: Bondad de ajuste de la variable discreta 3 y 4 - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

N. DETERMINACIÓN DE LA ESTIMACIÓN DE LAS PÉRDIDAS AGREGADAS

Poisson [#5]					
Kolmogorov-Smirnov					
Sample Size	12				
Statistic	0,58199397				
P-Value	2,17250225E-4				
Rank	4				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	0,29577	0,33815	0,37543	0,41918	0,44905
Reject?	Yes	Yes	Yes	Yes	Yes
Anderson-Darling					
Sample Size	12				
Statistic	21,9868992				
Rank	5				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	1,37490912	1,9286236	2,50175596	3,28919661	3,90741594
Reject?	Yes	Yes	Yes	Yes	Yes

Figura N.5: Bondad de ajuste de la variable discreta 5 - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

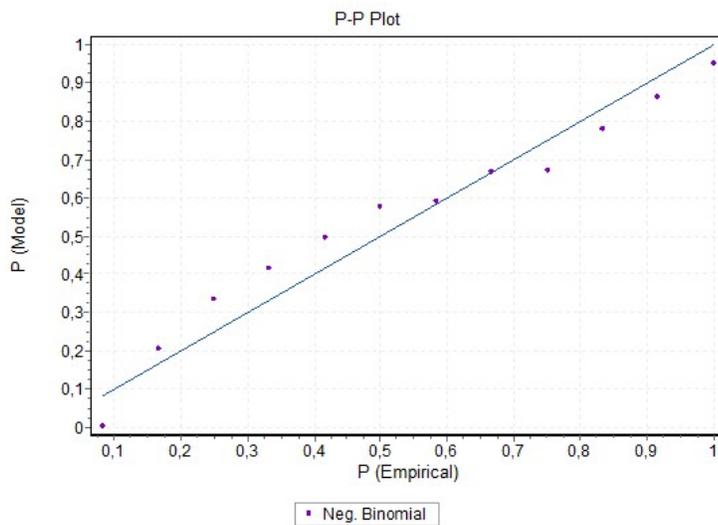


Figura N.6: Gráfico P P binomial negativa - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

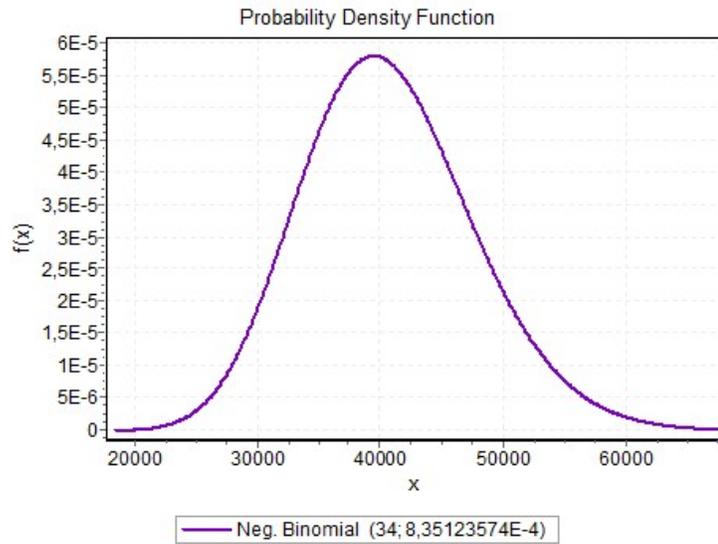


Figura N.7: Función de densidad de probabilidad de la binomial negativa -

Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

N.2. Asociar la función de monto siniestro individual

Se realizó una revisión exhaustiva y se determinó que el mejor ajuste es con dos funciones asociadas, establecidas por la teoría de los valores extremos.

N. DETERMINACIÓN DE LA ESTIMACIÓN DE LAS PÉRDIDAS AGREGADAS

Goodness of Fit - Details					
Beta [#1]					
Kolmogorov-Smirnov					
Sample Size	41031				
Statistic	0,0944				
P-Value	0				
Rank	24				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	0,0053	0,00604	0,0067	0,00749	0,00804
Reject?	Yes	Yes	Yes	Yes	Yes
Anderson-Darling					
Sample Size	41031				
Statistic	773,15				
Rank	20				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	1,3749	1,9286	2,5018	3,2892	3,9074
Reject?	Yes	Yes	Yes	Yes	Yes
Chi-Squared					
Deg. of freedom	9				
Statistic	INF				
P-Value	0				
Rank	41				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	12,242	14,684	16,919	19,679	21,666
Reject?	Yes	Yes	Yes	Yes	Yes
[top]					
Burr [#2]					
Kolmogorov-Smirnov					
Sample Size	41031				
Statistic	0,02596				
P-Value	0				
Rank	11				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	0,0053	0,00604	0,0067	0,00749	0,00804
Reject?	Yes	Yes	Yes	Yes	Yes
Anderson-Darling					
Sample Size	41031				
Statistic	50,505				
Rank	8				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	1,3749	1,9286	2,5018	3,2892	3,9074
Reject?	Yes	Yes	Yes	Yes	Yes
Chi-Squared					
Deg. of freedom	9				
Statistic	83,387				
P-Value	3,4195E-14				
Rank	5				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	12,242	14,684	16,919	19,679	21,666
Reject?	Yes	Yes	Yes	Yes	Yes

Figura N.8: Bondad de ajuste 1 al 2 - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

N.2 Asociar la función de monto siniestro individual

Burr (4P) [#3]					
Kolmogorov-Smirnov					
Sample Size	41031				
Statistic	0,48276				
P-Value	0				
Rank	48				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	0,0053	0,00604	0,0067	0,00749	0,00804
Reject?	Yes	Yes	Yes	Yes	Yes
Anderson-Darling					
Sample Size	41031				
Statistic	13565,0				
Rank	49				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	1,3749	1,9286	2,5018	3,2892	3,9074
Reject?	Yes	Yes	Yes	Yes	Yes
Chi-Squared					
Deg. of freedom	9				
Statistic	14651,0				
P-Value	0				
Rank	26				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	12,242	14,684	16,919	19,679	21,666
Reject?	Yes	Yes	Yes	Yes	Yes

[top]

Cauchy [#4]					
Kolmogorov-Smirnov					
Sample Size	41031				
Statistic	0,22233				
P-Value	0				
Rank	35				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	0,0053	0,00604	0,0067	0,00749	0,00804
Reject?	Yes	Yes	Yes	Yes	Yes
Anderson-Darling					
Sample Size	41031				
Statistic	3794,2				
Rank	33				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	1,3749	1,9286	2,5018	3,2892	3,9074
Reject?	Yes	Yes	Yes	Yes	Yes
Chi-Squared					
Deg. of freedom	9				
Statistic	1049,6				
P-Value	0				
Rank	18				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	12,242	14,684	16,919	19,679	21,666
Reject?	Yes	Yes	Yes	Yes	Yes

Figura N.9: Bondad de ajuste 3 al 4 - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

N. DETERMINACIÓN DE LA ESTIMACIÓN DE LAS PÉRDIDAS AGREGADAS

Chi-Squared [#5]					
Kolmogorov-Smirnov					
Sample Size	41031				
Statistic	0,75068				
P-Value	0				
Rank	53				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	0,0053	0,00604	0,0067	0,00749	0,00804
Reject?	Yes	Yes	Yes	Yes	Yes
Anderson-Darling					
Sample Size	41031				
Statistic	2,3901E+6				
Rank	58				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	1,3749	1,9286	2,5018	3,2892	3,9074
Reject?	Yes	Yes	Yes	Yes	Yes

[top]

Chi-Squared (2P) [#6]					
Kolmogorov-Smirnov					
Sample Size	41031				
Statistic	0,99954				
P-Value	0				
Rank	59				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	0,0053	0,00604	0,0067	0,00749	0,00804
Reject?	Yes	Yes	Yes	Yes	Yes

[top]

Dagum [#7]					
Kolmogorov-Smirnov					
Sample Size	41031				
Statistic	0,99882				
P-Value	0				
Rank	57				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	0,0053	0,00604	0,0067	0,00749	0,00804
Reject?	Yes	Yes	Yes	Yes	Yes
Anderson-Darling					
Sample Size	41031				
Statistic	9,9115E+5				
Rank	57				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	1,3749	1,9286	2,5018	3,2892	3,9074
Reject?	Yes	Yes	Yes	Yes	Yes

Figura N.10: Bondad de ajuste 5 al 6 - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

N.2 Asociar la función de monto siniestro individual

Dagum [#7]					
Kolmogorov-Smirnov					
Sample Size	41020				
Statistic	0,998823704				
P-Value	0				
Rank	57				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	0,005297877	0,006038494	0,006705049	0,00749504	0,008043096
Reject?	Yes	Yes	Yes	Yes	Yes
Anderson-Darling					
Sample Size	41020				
Statistic	990604,781				
Rank	57				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	1,37490912	1,9286236	2,50175596	3,28919661	3,90741594
Reject?	Yes	Yes	Yes	Yes	Yes

Figura N.11: Bondad de ajuste 7 - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

N. DETERMINACIÓN DE LA ESTIMACIÓN DE LAS PÉRDIDAS AGREGADAS

Dagum (4P) [#8]					
Kolmogorov-Smirnov					
Sample Size	41031				
Statistic	0,98227				
P-Value	0				
Rank	56				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	0,0053	0,00604	0,0067	0,00749	0,00804
Reject?	Yes	Yes	Yes	Yes	Yes
Anderson-Darling					
Sample Size	41031				
Statistic	2,7659E+5				
Rank	55				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	1,3749	1,9286	2,5018	3,2892	3,9074
Reject?	Yes	Yes	Yes	Yes	Yes
Chi-Squared					
Deg. of freedom	9				
Statistic	7,2329E+6				
P-Value	0				
Rank	30				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	12,242	14,684	16,919	19,679	21,666
Reject?	Yes	Yes	Yes	Yes	Yes

[top]

Error [#9]					
Kolmogorov-Smirnov					
Sample Size	41031				
Statistic	0,26217				
P-Value	0				
Rank	36				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	0,0053	0,00604	0,0067	0,00749	0,00804
Reject?	Yes	Yes	Yes	Yes	Yes
Anderson-Darling					
Sample Size	41031				
Statistic	5651,9				
Rank	37				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	1,3749	1,9286	2,5018	3,2892	3,9074
Reject?	Yes	Yes	Yes	Yes	Yes
Chi-Squared					
Deg. of freedom	9				
Statistic	9,3093E+11				
P-Value	0				
Rank	38				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	12,242	14,684	16,919	19,679	21,666
Reject?	Yes	Yes	Yes	Yes	Yes

Figura N.12: Bondad de ajuste 8 al 9 - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

N.2 Asociar la función de monto siniestro individual

Error Function [#10]					
Kolmogorov-Smirnov					
Sample Size	41031				
Statistic	0,5				
P-Value	0				
Rank	50				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	0,0053	0,00604	0,0067	0,00749	0,00804
Reject?	Yes	Yes	Yes	Yes	Yes
Anderson-Darling					
Sample Size	41031				
Statistic	9731,6				
Rank	45				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	1,3749	1,9286	2,5018	3,2892	3,9074
Reject?	Yes	Yes	Yes	Yes	Yes

[top]

Exponential [#11]					
Kolmogorov-Smirnov					
Sample Size	41031				
Statistic	0,16107				
P-Value	0				
Rank	32				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	0,0053	0,00604	0,0067	0,00749	0,00804
Reject?	Yes	Yes	Yes	Yes	Yes
Anderson-Darling					
Sample Size	41031				
Statistic	2347,7				
Rank	30				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	1,3749	1,9286	2,5018	3,2892	3,9074
Reject?	Yes	Yes	Yes	Yes	Yes

[top]

Exponential (2P) [#12]					
Kolmogorov-Smirnov					
Sample Size	41031				
Statistic	0,16107				
P-Value	0				
Rank	33				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	0,0053	0,00604	0,0067	0,00749	0,00804
Reject?	Yes	Yes	Yes	Yes	Yes
Anderson-Darling					
Sample Size	41031				
Statistic	2347,9				
Rank	31				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	1,3749	1,9286	2,5018	3,2892	3,9074
Reject?	Yes	Yes	Yes	Yes	Yes

Figura N.13: Bondad de ajuste 10 al 12 - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

N. DETERMINACIÓN DE LA ESTIMACIÓN DE LAS PÉRDIDAS AGREGADAS

Fatigue Life [#13]					
Kolmogorov-Smirnov					
Sample Size	41031				
Statistic	0,11943				
P-Value	0				
Rank	27				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	0,0053	0,00604	0,0067	0,00749	0,00804
Reject?	Yes	Yes	Yes	Yes	Yes
Anderson-Darling					
Sample Size	41031				
Statistic	984,31				
Rank	25				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	1,3749	1,9286	2,5018	3,2892	3,9074
Reject?	Yes	Yes	Yes	Yes	Yes
Chi-Squared					
Deg. of freedom	9				
Statistic	1,7095E+7				
P-Value	0				
Rank	31				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	12,242	14,684	16,919	19,679	21,666
Reject?	Yes	Yes	Yes	Yes	Yes
[top]					
Fatigue Life (3P) [#14]					
Kolmogorov-Smirnov					
Sample Size	41031				
Statistic	0,10739				
P-Value	0				
Rank	26				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	0,0053	0,00604	0,0067	0,00749	0,00804
Reject?	Yes	Yes	Yes	Yes	Yes
Anderson-Darling					
Sample Size	41031				
Statistic	801,51				
Rank	21				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	1,3749	1,9286	2,5018	3,2892	3,9074
Reject?	Yes	Yes	Yes	Yes	Yes
Chi-Squared					
Deg. of freedom	9				
Statistic	3,0311E+7				
P-Value	0				
Rank	32				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	12,242	14,684	16,919	19,679	21,666
Reject?	Yes	Yes	Yes	Yes	Yes

Figura N.14: Bondad de ajuste 13 al 14 - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

N.2 Asociar la función de monto siniestro individual

Frechet [#15]					
Kolmogorov-Smirnov					
Sample Size	41031				
Statistic	0,07712				
P-Value	0				
Rank	19				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	0,0053	0,00604	0,0067	0,00749	0,00804
Reject?	Yes	Yes	Yes	Yes	Yes
Anderson-Darling					
Sample Size	41031				
Statistic	738,04				
Rank	19				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	1,3749	1,9286	2,5018	3,2892	3,9074
Reject?	Yes	Yes	Yes	Yes	Yes
Chi-Squared					
Deg. of freedom	9				
Statistic	748,93				
P-Value	0				
Rank	16				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	12,242	14,684	16,919	19,679	21,666
Reject?	Yes	Yes	Yes	Yes	Yes

[top]

Frechet (3P) [#16]					
Kolmogorov-Smirnov					
Sample Size	41031				
Statistic	0,08642				
P-Value	0				
Rank	23				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	0,0053	0,00604	0,0067	0,00749	0,00804
Reject?	Yes	Yes	Yes	Yes	Yes
Anderson-Darling					
Sample Size	41031				
Statistic	943,6				
Rank	24				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	1,3749	1,9286	2,5018	3,2892	3,9074
Reject?	Yes	Yes	Yes	Yes	Yes
Chi-Squared					
Deg. of freedom	9				
Statistic	2989,5				
P-Value	0				
Rank	19				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	12,242	14,684	16,919	19,679	21,666
Reject?	Yes	Yes	Yes	Yes	Yes

Figura N.15: Bondad de ajuste 15 al 16 - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

N. DETERMINACIÓN DE LA ESTIMACIÓN DE LAS PÉRDIDAS AGREGADAS

Gamma [#17]					
Kolmogorov-Smirnov					
Sample Size	41031				
Statistic	0,35712				
P-Value	0				
Rank	43				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	0,0053	0,00604	0,0067	0,00749	0,00804
Reject?	Yes	Yes	Yes	Yes	Yes
Anderson-Darling					
Sample Size	41031				
Statistic	7425,9				
Rank	43				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	1,3749	1,9286	2,5018	3,2892	3,9074
Reject?	Yes	Yes	Yes	Yes	Yes
Chi-Squared					
Deg. of freedom	9				
Statistic	3337,0				
P-Value	0				
Rank	21				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	12,242	14,684	16,919	19,679	21,666
Reject?	Yes	Yes	Yes	Yes	Yes

[\[top\]](#)

Gamma (3P) [#18]					
Kolmogorov-Smirnov					
Sample Size	41031				
Statistic	0,10578				
P-Value	0				
Rank	25				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	0,0053	0,00604	0,0067	0,00749	0,00804
Reject?	Yes	Yes	Yes	Yes	Yes
Anderson-Darling					
Sample Size	41031				
Statistic	909,13				
Rank	22				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	1,3749	1,9286	2,5018	3,2892	3,9074
Reject?	Yes	Yes	Yes	Yes	Yes

Figura N.16: Bondad de ajuste 17 al 18 - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

N.2 Asociar la función de monto siniestro individual

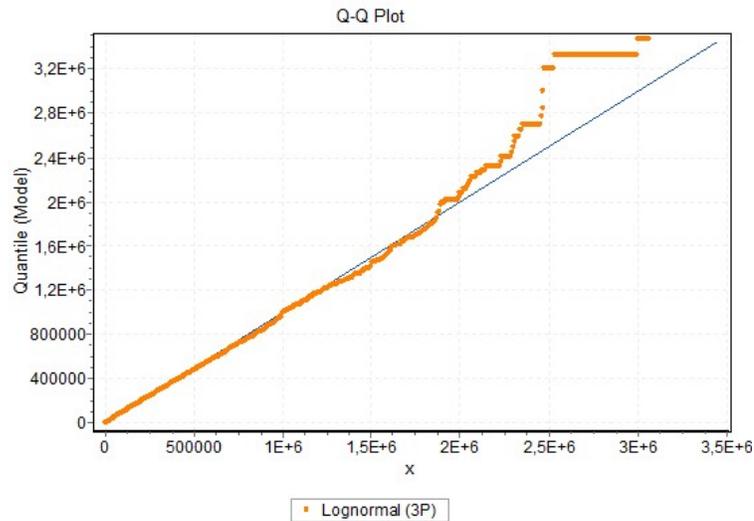


Figura N.17: Resultado de ajuste de la lognormal - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

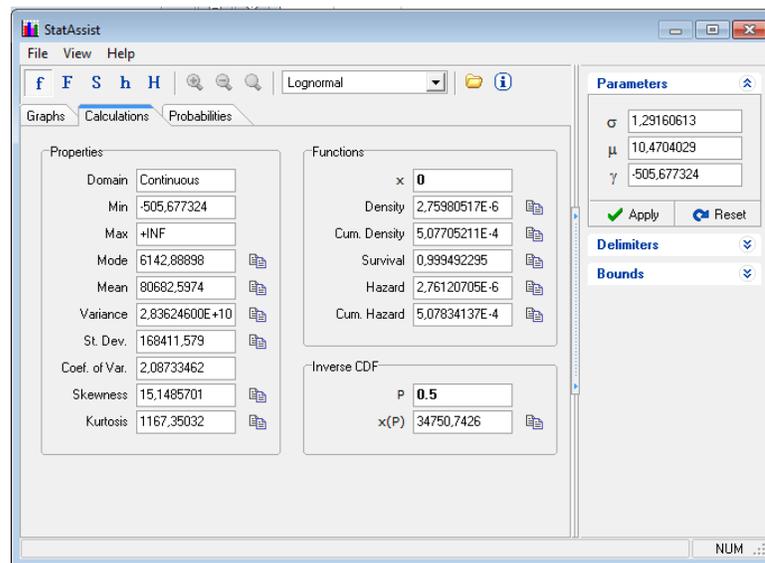


Figura N.18: Resultado del cálculo del ajuste de la lognormal - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

N. DETERMINACIÓN DE LA ESTIMACIÓN DE LAS PÉRDIDAS AGREGADAS

Fitting Results		
#	Distribution	Parameters
1	Beta	$\alpha_1=0,553435903$ $\alpha_2=0,982519617$ $a=3500000$ $b=6143028,32$
2	Burr	$k=4,14350718E-4$ $\sigma=270,294727$ $\beta=203802,839$
3	Burr (4P)	$k=0,407827023$ $\alpha=0,244343436$ $\beta=1,99554568$ $\gamma=3500000,0$
4	Cauchy	$\sigma=490843,154$ $\mu=4147391,3$
5	Chi-Squared	$v=4448942$
6	Chi-Squared (2P)	$v=1428836$ $\gamma=3132500$
7	Dagum	$k=166,765466$ $\alpha=0,472037324$ $\beta=22,71708$
8	Dagum (4P)	$k=1,76581867$ $\alpha=0,111746217$ $\beta=0,724758257$ $\gamma=3500000,0$
9	Erlang	$m=29$ $\beta=150721,142$
10	Erlang (3P)	$m=1$ $\beta=947907,997$ $\gamma=3499996,96$
11	Error	$k=1,92500603$ $\sigma=818871,028$ $\mu=4448942,95$
12	Error Function	$h=8,63514225E-7$
13	Exponential	$\lambda=2,24772493E-7$
14	Exponential (2P)	$\lambda=1,05380413E-6$ $\gamma=3500000$
15	Fatigue Life	$\alpha=0,16943422$ $\beta=4386003,91$
16	Fatigue Life (3P)	$\alpha=0,804001334$ $\beta=874732,453$ $\gamma=3287518,49$
17	Frechet	$\alpha=6,72758658$ $\beta=3938129,05$
18	Frechet (3P)	$\alpha=4,20421773$ $\beta=2205624,92$ $\gamma=1811782,13$
19	Gamma	$\alpha=29,51771$ $\beta=150721,142$
20	Gamma (3P)	$\alpha=0,457211822$ $\beta=1230630,09$ $\gamma=3500000$
21	Gen. Extreme Value	$k=0,068833936$ $\sigma=639578,341$ $\mu=4033255,47$
22	Gen. Gamma	$k=1,02709229$ $\alpha=32,3690552$ $\beta=150721,142$
23	Gen. Gamma (4P)	$k=0,779067373$ $\alpha=1,21221926$ $\beta=580202,288$ $\gamma=3500000$
24	Gen. Logistic	$k=0,214931948$ $\sigma=439147,215$ $\mu=4285007,84$
25	Gen. Pareto	$k=-0,292365475$ $\sigma=1405392,1$ $\mu=3361485,79$
26	Gumbel Max	$\sigma=638471,121$ $\mu=4080407,42$
27	Gumbel Min	$\sigma=638471,121$ $\mu=4817478,49$
28	Hypersecant	$\sigma=818871,028$ $\mu=4448942,95$
29	Inv. Gaussian	$\lambda=131322608,0$ $\mu=4448942,95$
30	Inv. Gaussian (3P)	$\lambda=2214985,51$ $\mu=1246809,17$ $\gamma=3202133,79$
31	Johnson SB	$\gamma=1,06979218$ $\delta=0,945089203$ $\lambda=4431871,74$ $\xi=3192192,53$
32	Kumaraswamy	$\alpha_1=1,05$ $\alpha_2=1,15$ $a=3500000,0$ $b=6143028,32$
33	Laplace	$\lambda=1,72702845E-6$ $\mu=4448942,95$

Figura N.19: Resultado del cálculo del ajuste de la wakeby del 1 al 33 - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

N.2 Asociar la función de monto siniestro individual

34	Levy	$\sigma=4323920,8$
35	Levy (2P)	$\sigma=359303,579 \quad \gamma=3431564,69$
36	Log-Gamma	$\alpha=7444,34179 \quad \beta=0,002054386$
37	Log-Logistic	$\alpha=9,42097415 \quad \beta=4238943,66$
38	Log-Logistic (3P)	$\alpha=1,85081134 \quad \beta=798097,489 \quad \gamma=3387360,37$
39	Log-Pearson 3	$\alpha=13,5274439 \quad \beta=0,048193376 \quad \gamma=14,6416191$
40	Logistic	$\sigma=451467,259 \quad \mu=4448942,95$
41	Lognormal	$\sigma=0,169004737 \quad \mu=15,2935523$
42	Lognormal (3P)	$\sigma=0,749143437 \quad \mu=13,7251252 \quad \gamma=3273373,11$
43	Nakagami	$m=6,89828778 \quad \kappa=2,04026841E+13$
44	Normal	$\sigma=818871,028 \quad \mu=4448942,95$
45	Pareto	$\alpha=4,4389452 \quad \beta=3500000$
46	Pareto 2	$\alpha=42,2470828 \quad \beta=189022749,0$
47	Pearson 5	$\alpha=36,1914176 \quad \beta=156488835,0$
48	Pearson 5 (3P)	$\alpha=4,36172963 \quad \beta=5432604,73 \quad \gamma=2865211,01$
49	Pearson 6	$\alpha_1=21186,799 \quad \alpha_2=36,241502 \quad \beta=7396,51857$
50	Pearson 6 (4P)	$\alpha_1=2,59985076 \quad \alpha_2=0,079029044$ $\beta=1,86573068E-5 \quad \gamma=3500000,0$
51	Pert	$m=4262657,35 \quad a=3500000 \quad b=6143028,32$
52	Phased Bi-Exponential	$\lambda_1=1,12949068E-6 \quad \gamma_1=3500000$ $\lambda_2=3,05414269E-7 \quad \gamma_2=1500000$
53	Power Function	$\alpha=0,794978942 \quad a=3500000 \quad b=6143028,32$
54	Rayleigh	$\sigma=3549742,89$
55	Rayleigh (2P)	$\sigma=1077557,51 \quad \gamma=3140252,87$
56	Reciprocal	$a=3500000,0 \quad b=6143028,32$
57	Rice	$v=4377532,46 \quad \sigma=787367,061$
58	Student's t	$v=2$
59	Triangular	$m=3703800,54 \quad a=3500000 \quad b=6143028,32$
60	Uniform	$a=3030616,73 \quad b=5867269,18$
61	Wakeby	$\alpha=1405392,1 \quad \beta=0,292365475 \quad \gamma=0$ $\delta=0 \quad \xi=3361485,79$
62	Weibull	$\alpha=6,59227933 \quad \beta=4569634,72$
63	Weibull (3P)	$\alpha=1,39517562 \quad \beta=1042034,54 \quad \gamma=3500000$
64	Johnson SU	No fit
65	Phased Bi-Weibull	No fit

Figura N.20: Resultado del cálculo del ajuste de la wakeby del 34 al 63 - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

N. DETERMINACIÓN DE LA ESTIMACIÓN DE LAS PÉRDIDAS AGREGADAS

Goodness of Fit - Details					
Wakeby [#61]					
Kolmogorov-Smirnov					
Sample Size	11				
Statistic	0,12329385				
P-Value	0,988608157				
Rank	1				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	0,30829	0,35242	0,39122	0,4367	0,4677
Reject?	No	No	No	No	No
Anderson-Darling					
Sample Size	11				
Statistic	0,183838556				
Rank	2				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	1,37490912	1,9286236	2,50175596	3,28919661	3,90741594
Reject?	No	No	No	No	No
Chi-Squared					
Deg. of freedom	1				
Statistic	0,003258382				
P-Value	0,954479687				
Rank	5				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	1,64237442	2,70554345	3,84145882	5,41189443	6,6348966
Reject?	No	No	No	No	No
[top]					
Gen. Pareto [#25]					
Kolmogorov-Smirnov					
Sample Size	11				
Statistic	0,12329385				
P-Value	0,988608157				
Rank	2				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	0,30829	0,35242	0,39122	0,4367	0,4677
Reject?	No	No	No	No	No
Anderson-Darling					
Sample Size	11				
Statistic	0,183838556				
Rank	3				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	1,37490912	1,9286236	2,50175596	3,28919661	3,90741594
Reject?	No	No	No	No	No
Chi-Squared					
Deg. of freedom	1				
Statistic	0,003258382				
P-Value	0,954479687				
Rank	6				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	1,64237442	2,70554345	3,84145882	5,41189443	6,6348966
Reject?	No	No	No	No	No

Figura N.21: Resultado del cálculo del ajuste de la wakeby del 61 y 25 - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

N.2 Asociar la función de monto siniestro individual

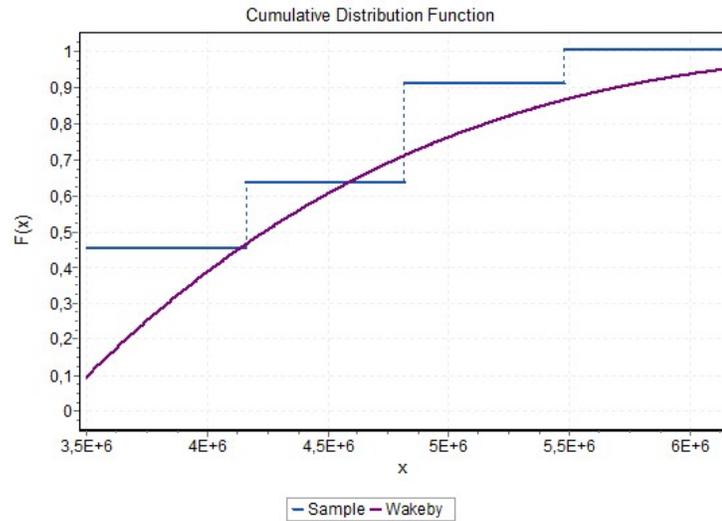


Figura N.22: Función de distribución acumulada de la wakeby - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

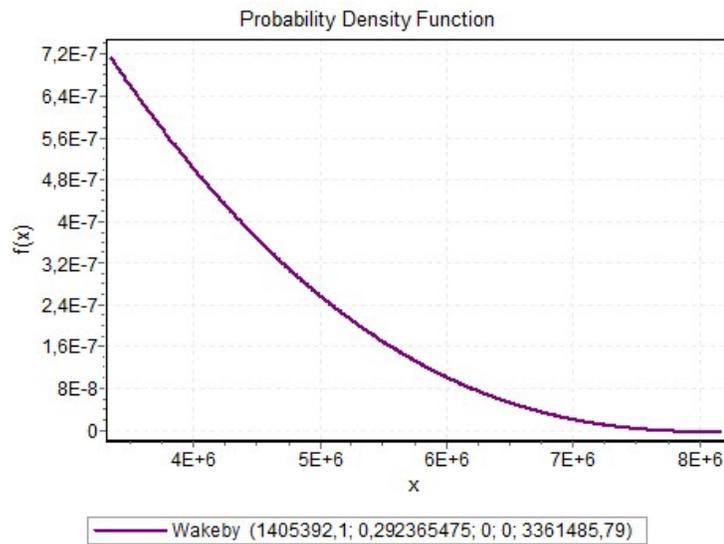


Figura N.23: Función de densidad de la wakeby - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

N. DETERMINACIÓN DE LA ESTIMACIÓN DE LAS PÉRDIDAS AGREGADAS

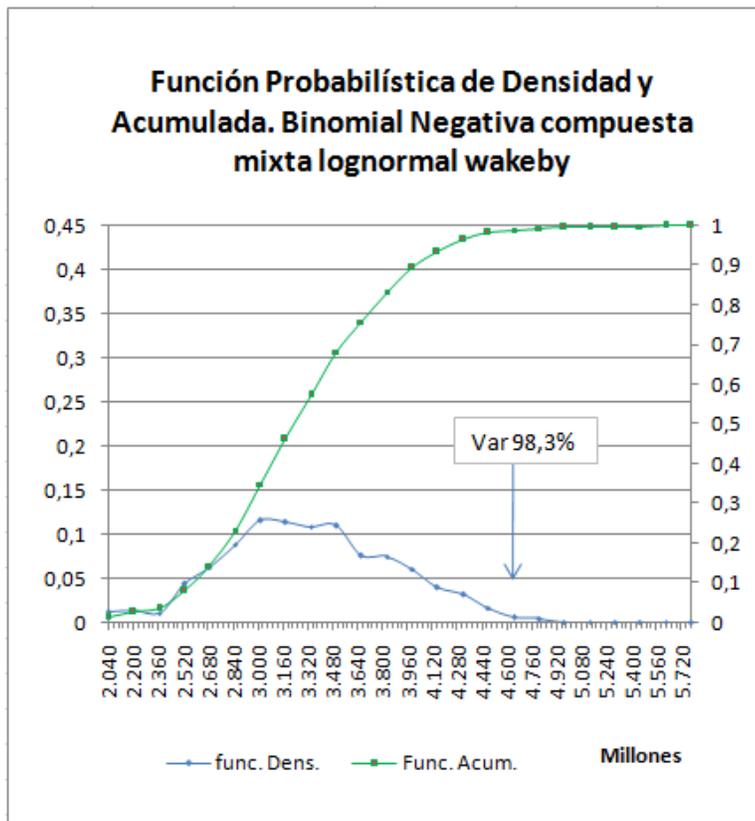


Figura N.24: Función calculada para hallar el valor en riesgo - Fuente: Superintendencia de la Actividad Aseguradora, Proyecto unificación, 2015

Vocabulario

Ñ.1. Tema asegurador

1. Seguro: Es el contrato comercial en virtud del cual una empresa de seguros, a cambio de una prima, asume las consecuencias de riesgos ajenos, que no se produzcan por acontecimientos que dependan enteramente de la voluntad del beneficiario, comprometiéndose a indemnizar, dentro de los límites pactados, el daño producido al tomador, al asegurado o al beneficiario, o a pagar un capital, una renta u otras prestaciones convenidas, todo subordinado a la ocurrencia de un evento denominado siniestro, cubierto por una póliza. Las disposiciones del contrato de seguro se aplicarán a los convenios mediante los cuales una persona se obliga a prestar un servicio o a pagar una cantidad de dinero en caso de que ocurra un acontecimiento futuro e incierto y que no dependa exclusivamente de la voluntad del beneficiario a cambio de una contraprestación, siempre que no exista una ley especial que lo regule.
2. Asegurado y beneficiario: Persona que en si misma, en sus bienes o en sus intereses económicos está expuesta al riesgo; y el beneficiario, aquél en cuyo favor se ha establecido la indemnización que pagará la empresa de seguros.
3. Tomador: la persona que obrando por cuenta propia o ajena, traslada los riesgos.
4. Objeto del contrato de seguros: Puede cubrir toda clase de riesgos si existe interés asegurable; salvo prohibición expresa de la ley.

5. Causa del contrato de seguros: Todo interés legítimo en la no materialización de un riesgo, que sea susceptible de valoración económica, puede ser causa de un contrato de seguros. Pueden asegurarse las personas y los bienes de lícito comercio en cuya conservación tenga el beneficiario un interés pecuniario legítimo.
6. Prima: Contraprestación que, en función del riesgo, debe pagar el tomador a la empresa de seguros en virtud de la celebración del contrato de seguros.
7. Periodo del seguro: Lapso de tiempo para el cual ha sido calculada la unidad de prima. En caso de que no se haya especificado y no pueda determinarse de acuerdo con el reglamento actuarial, se presume que la prima cubre el periodo de un (1) año.
8. Riesgo: Es el suceso futuro e incierto que no depende exclusivamente de la voluntad del tomador, del asegurado o del beneficiario, y cuya materialización da origen a la obligación de la empresa de seguros. Los hechos ciertos, salvo la muerte y los físicamente imposibles, no constituyen riesgo y son inasegurables. Tampoco constituye riesgo la incertidumbre subjetiva respecto a determinado hecho que se haya cumplido o no.
9. Siniestros: Es el acontecimiento futuro e incierto del cual depende la obligación de indemnizar por parte de la empresa de seguros. Si el siniestro ha continuado después de vencido el contrato, la empresa de seguros responde del valor de la indemnización en los términos del contrato. Pero si se inicia antes de la vigencia del contrato, y continúa después de que los riesgos hayan comenzado a correr por cuenta de la empresa de seguros, ésta queda relevada de su obligación de indemnizar.

Ñ.2. Tema Siglas

1. MPP: Máxima Pérdida Probable.
2. ACP: Análisis de Componentes Principales.
3. AD: Análisis Discriminante.
4. Ak: Agrupación a través de la técnica k-medias.
5. Bagging: Máquina de aprendizaje y minería de datos, bootstrap aggregating.

6. RF: Random Forest, bosque aleatorio; combinación de árboles predictores tal que cada árbol depende de los valores de un vector aleatorio probado independientemente y con la misma distribución para cada uno de estos.

Bibliografía

- [1] H. Panjer. (2006). *Operational Risk Modeling Analytics*. John Wiley & Sons series in probability and statistics, Inc., Hoboken, New Jersey. [1](#), [2](#), [24](#)
- [2] M. L. Vásquez de R. and G. J. Ramírez. (1999). *Contribuciones al análisis de segmentación*. Facultad de ciencias económicas y sociales- U.C.V., Oficina central de estadística e informática, Caracas. [2](#), [10](#)
- [3] E. Rodríguez T. (2004). *Administración del riesgo: Utilización de análisis de datos multivariantes para el estudio integral del riesgo*. s.n., Bogota. [2](#), [9](#)
- [4] P. Jorion. (2007). *Valor en Riesgo: El nuevo paradigma para el control de riesgo con derivados*. Limusa Noriega editores, México. [2](#), [7](#), [21](#), [22](#), [28](#)
- [5] F. Venegas M. (2008). *Riesgo financiero y económicos productos derivados y decisiones económicas bajo incertidumbre*. 2°. ed. Cengage Learning Editores, S.A. pp. 861-912, México. [2](#), [21](#)
- [6] R. Hogg and S. Klugman. (1984). *Loss distributions*. Wiley, pp. 217-233, New York. [2](#), [24](#), [26](#)
- [7] C. Daykin, T. Pentikäinen, and M. Pesonen. (1994). *Practical risk theory for actuaries*. Chapman & Hall, London. [2](#), [6](#), [9](#), [22](#), [23](#), [33](#)
- [8] A. Charpentier. (2015). *Computational actuarial science with R*. CRC Press Taylor & Francis Group, LLC, Boca raton, London, New York. [2](#), [10](#)
- [9] R. Kass, M. Goovaerts, J. Dhaene, and M. Denuit. (2008). *Modern actuarial risk theory using R*. Springer-Verlag, Berlin, Heidelberg. [2](#), [9](#), [10](#)

BIBLIOGRAFÍA

- [10] Bower, Gerber, Hickman, Jones, and Nesbitt. (1994). *Actuarial Mathematics*. The Society of actuaries, Schaumburg, Illinois. 9
- [11] C. Chirinos. (2001). *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. 10
- [12] A. López. (2001). *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. 10
- [13] T. Thonon. (2002). *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. 10
- [14] C. Salazar. (2012). *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. 10
- [15] G. Ramírez. (1999). *Teoría de la Probabilidad*. Trabajo de ascenso, no publicado. Universidad central de Venezuela., Caracas. 10, 23
- [16] J. Días. (2014). *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. 10
- [17] M. Chiodi. (1987). *The clustering of longitudinal. Multivariate data when time series are short*. Elsevier science publishers B.V., North-Holland. 11, 21
- [18] J. Cuesta-Albertos, A. Gordaliza, and C. Matrán. (1997). *Trimmed K-means: an attempt to robustify y quntizers. s.l.* Anuario de estadísticas. Vol. 25 (2), pp. 553-576, México. 11, 20
- [19] 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*. 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. 11, 18, 20
- [20] J. Hurtado de B. (2005). *¿Cómo formular objetivos de investigación?* Instituto universitario de tecnología “José Antonio Anzoátegui” Quirón Ediciones-Fundación Sypal. p. 25, Caracas. 18, 81
- [21] R. Sampieri, C. Fernández, and P. Baptista. (2006). *Metodología de la investigación*. 4°. ed. Mc Graw Hill, México. 18
- [22] K. Pearson. (1901). *On lines and planes of closest fit to systems of points in space*. Philosophical magazine, 2:559-572. <http://pbil.univ-lyon1.fr/R/pearson1901.pdf>, London. 18
- [23] M.L. Vásquez de R. y G.J. Ramírez. (2007). *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. 18
- [24] L. G. Díaz and M. A. Morales R. (2012). *Estadística multivariada: inferencia y métodos*. Universidad Nacional de Colombia, Bogotá, Colombia. 20