



CIMAT

Centro de Investigación en Matemáticas, A.C.

Análisis de percances viales para la implementación de una estrategia comercial.

Tesina

para obtener el grado de

Especialista en Métodos Estadísticos

presenta

LUIS IVÁN MARTÍNEZ SOTELO

Director de tesina

DR. JORGE RAÚL PÉREZ GALLARDO

Aguascalientes, Ags. 9 de septiembre de 2019

Índice general

1. Introducción	9
2. Situación Problemática	13
3. Análisis Descriptivo	17
3.1. Industria Automotriz	17
3.2. Accidentes a nivel nacional	25
3.3. Aseguradora AXXA	31
4. Identificación de Clusters	39
4.1. Planteamiento	39
4.2. Estimación del modelo	41
4.2.1. Existencia de cluster	41
4.2.2. Técnica de clustering	44
5. Índice de Costo Estatal	55
5.1. Estimación del índice	55
5.2. Índice de Costos	60
6. Pronósticos de Percances Viales	65
6.1. Descomposición de la serie	65
6.2. Pronóstico de accidentes	68
6.3. Pronósticos por Cluster	71
6.4. Accidentes Esperados vs Costos	72
7. Conclusiones Generales	75
8. Bibliografía	77

Índice de figuras

3.1. Panorama Sector Automotriz en México	18
3.2. Producción Anual de Vehículos (2015-2018)	19
3.3. Producción Mensual de Vehículos (2017-2018)	19
3.4. Ventas de Vehículos por mes (2015-2018)	20
3.5. Ventas por tipo de Vehículo (2015-2018)	21
3.6. Circulación de Vehículos (2015-2018)	21
3.7. Variación en Circulación de Vehículos	22
3.8. Autos en Circulación 2017 (Miles de Unidades)	23
3.9. Camiones Pasajeros en Circulación 2017	24
3.10. Camiones de Carga en Circulación 2017	24
3.11. Motocicletas en Circulación 2017	25
3.12. Número de Accidentes Nacionales 2017	26
3.13. Cantidad de Accidentes por día en 2017	27
3.14. Clasificación de Accidentes por Género 2017	28
3.15. Accidentes por tipo de Transporte 2017	28
3.16. Accidentes por causa 2017	29
3.17. Seguridad al Volante 2017	30
3.18. Cantidad de muertos en accidente 2017	30
3.19. Accidentes con personas heridas 2017	31
3.20. Cantidad de heridos en accidentes 2017	31
3.21. Cantidad de Accidentes AXXA	32
3.22. Cantidad de Accidentes AXXA (S/ Chihuahua)	32
3.23. Calendario accidentes AXXA	34
3.24. Gravedad de los Accidentes	35
3.25. Hospitalización	36
3.26. No. Fallecidos	37
3.27. Uso de ambulancias	37
3.28. Uso de grúas	38
3.29. Pérdida Total	38
4.1. Matriz de correlación	41
4.2. Matriz de correlación	41

4.3.	Datos Observados	42
4.4.	Datos Aleatorios	43
4.5.	Matriz de Distancias	44
4.6.	Gap statistic	45
4.7.	Índices	46
4.8.	Dendograma 9 Clusters	47
4.9.	Dendograma 5 Clusters	47
4.10.	Proyección Bidimensional de agrupaciones resultantes (9 grupos)	48
4.11.	Proyección Bidimensional de agrupaciones resultantes (5 grupos)	48
4.12.	Gráfico de silueta para agrupaciones resultantes (9 grupos) . .	49
4.13.	Gráfico de silueta para agrupaciones resultantes (5 grupos) . .	50
4.14.	Clasificación final de las 32 entidades federativas	51
4.15.	Comportamiento de Clusters	52
5.1.	Varianza Explicada	57
5.2.	Gráfica de cargas	57
5.3.	Gráfica de contribuciones	58
5.4.	Calidad de Representación	58
5.5.	Gráfica de doble proyección	59
5.6.	Semaforización del Índice de Costos	60
5.7.	Distribución del Índice	60
5.8.	Costo por Clúster	61
6.1.	Accidentes Mensuales	66
6.2.	Descomposición de la serie	67
6.3.	Tendencia y Estacionalidad	68
6.4.	Ajuste Holt-Winters	69
6.5.	Pronóstico	69
6.6.	Intervalos de Confianza	69
6.7.	Factores de Estacionalidad	70
6.8.	Residuales	71
6.9.	Pronósticos por Cluster	72

Resumen

En este trabajo se analiza la locación, frecuencia y gravedad de los accidentes viales en la república mexicana. Se presenta un análisis de interdependencia entre los tres factores mencionados para determinar comportamientos similares en distintas regiones geográficas del país y diseñar una estrategia financiera y operativa que minimice los costes y asegure un servicio de calidad por parte de la aseguradora mediante el uso eficiente de su capital humano. Se trabajó una muestra de 3 años de percances registrados por la INEGI y por la aseguradora AXXA en los 32 estados que conforman el país. El resultado del estudio muestra la formación de 5 regiones, las cuales difieren principalmente en los costos que representan para la aseguradora, la cantidad de autos en circulación y la cantidad de fallecidos; los estados de San Luis Potosí, Querétaro, Puebla, entre otros pertenecen a una misma región y representan los estados con mayor proyección de rentabilidad.

Capítulo 1

Introducción

La presente investigación se refiere al tema de seguros de automóviles, las compañías aseguradoras deben hacer frente a los siniestros que sufran sus asegurados, dichos percances generan un costo, el cual se puede desglosar en varios componentes, el presente estudio se enfoca en dos: costo de atención al siniestro y los gastos por capital humano.

Por regulación las aseguradoras deben tener recursos financieros líquidos que permitan solventar los accidentes a los que harán frente en el corto plazo, dado que se habla de accidentes futuros, la cantidad monetaria necesaria está en función de la probabilidad de ocurrencia de los accidentes.

La estimación eficiente de los accidentes esperados es fundamental para la elaboración de la estrategia financiera, ya que tener una reserva por debajo de la necesaria podría impedir que la aseguradora tenga los suficientes recursos líquidos para solventar los siniestros; en caso contrario, una reserva por encima de la necesaria causa una pérdida del valor del dinero ya que, al estar en la reserva, no genera los intereses que ofrecen otros mercados financieros.

Adicional a la cantidad de accidentes, una aseguradora debe considerar la gravedad de los mismos, esto para estimar el costo de reparación en que incurrirá en la reparación de los autos involucrados.

Otro aspecto importante a considerar por parte de las compañías de seguros es la calidad del servicio que otorgan a sus asegurados. La valoración del servicio por parte del asegurado se realiza con base a los tiempos de respuesta de la compañía, ya sea en el tiempo de arribo por parte de los ajustadores, el tiempo de diagnóstico de los daños, y el tiempo de reparación del vehículo, etc.

Para que el tiempo de atención por parte de la aseguradora sea el menor posible deben tener una administración del capital humano eficaz, en otras palabras, deben contar con la cantidad adecuada de personal para dar respuesta a los percances que se presenten. En la administración del personal, al igual que en el manejo de recursos financieros, el conocer la cantidad y gravedad de los siniestros futuros, además de su locación es fundamental, ya que una menor cantidad de personal afectara el servicio proporcionado por la compañía y una mayor cantidad a la necesaria crearía un gasto ineficiente de recursos.

Por lo comentado en los párrafos anteriores, el tener una metodología de pronóstico que permita estimar la locación, cantidad y gravedad de los percances viales es fundamental para la operación eficiente de una compañía de seguros.

La investigación de este tema se realizó por el interés de descubrir patrones de comportamiento y características de percances viales que permitan el diseño de una estrategia financiera y administrativa eficiente, la cual asegure un servicio de alta calidad minimizando el costo que implica la operación de una compañía de seguros.

Para la consecución del objetivo planteado, se consideró el registro de accidentes automovilísticos en el país entre los años 2015 y 2017.

Los datos utilizados son obtenidos de dos fuentes, la primera es el INEGI, específicamente las bases de datos relacionadas con la producción, ventas y circulación de vehículos de motor, así como la información relacionada con los accidentes ocurridos a nivel nacional. La segunda fuente es obtenida a través del portal del Instituto Internacional de Ciencia de Datos, y los cuales detallan la información sobre los percances viales de la aseguradora AXXA.

Los objetivos específicos del análisis son los siguientes:

Determinar regiones compuestas por los distintos estados de la república que tengan similitudes respecto a cantidad de automóviles y características de accidentes; esto permite identificar zonas donde se implementen estrategias comerciales similares.

Estimar el costo en que incurre la aseguradora al atender los percances viales presentados en cada una de las regiones identificadas. El costo está en función de los gastos operativos de la aseguradora tales como uso de ambulancia, grúas, pérdida total del auto, etc.

Pronosticar la cantidad de accidentes que se presentarán en el futuro en cada una de las regiones.

La consecución de los tres objetivos permite determinar el panorama que enfrentará la aseguradora y de esa manera administrar sus recursos financieros y operativos para lograr la rentabilidad deseada.

El análisis se estructura de la siguiente manera, el primer capítulo muestra un planteamiento formal de la situación a analizar, en el cual se especifican los objetivos, justificación, alcance, limitaciones y la metodología estadística.

El segundo capítulo presenta el análisis descriptivo de la situación del sector automovilístico y de los accidentes a nivel nacional. El siguiente capítulo consiste en la descripción, estimación del modelo de agrupación para conformar las regiones.

En el tercer capítulo se muestra el cálculo de costos por región. Por último, el cuarto capítulo describe la metodología de pronósticos, seguido de las conclusiones generales del estudio.

Capítulo 2

Situación Problemática

Planteamiento del problema

La compañía aseguradora se enfrenta a un escenario de incertidumbre respecto a los costos futuros, los cuales están en función la cantidad de accidentes y la gravedad de los mismos. Por lo tanto, la investigación se centra en diseñar una metodología de pronóstico que permita anticipar los gastos financieros y operativos de la aseguradora mediante la estimación de accidentes esperados así como su locación y costos.

Objetivo General

Definir una estrategia comercial para la administración de la aseguradora a lo largo de la república mexicana, por medio de la identificación de la cantidad de percances viales, los costos que representarán y la locación en la que ocurrirán.

Objetivos Específicos

- Definir regiones que compartan características de ocurrencia y causa de accidentes.
- Estimar el costo operativo de cada una de las regiones utilizando la información de gastos en que incurre la aseguradora para solventar los accidentes.
- Pronosticar la cantidad de accidentes futuros.

Justificación

La importancia del diseño de una estrategia comercial que contemple la incertidumbre inherente del sector de seguros es garantizar la solvencia financiera y capacidad operativa de las compañías dentro del sector. Estimando un escenario futuro de las condiciones del mercado en conjunto con los aspectos internos de las empresas se maximiza la rentabilidad esperada y se asegura la administración óptima del capital humano, logrando así un crecimiento sostenido y un incremento en la participación en el sector.

En este caso particular, se diseña una metodología para la aseguradora AXXA la cual permite entender comportamiento actual de las coberturas a los autos de sus asegurados, además de anticipar circunstancias venideras. El objetivo consiste en determinar las condiciones esperadas que faciliten el diseño e implementación de estrategias comerciales.

Alcances y Limitaciones

La interacción de los tres principales componentes de la metodología, locación, costos y accidentes futuros, permiten visualizar un panorama completo de las condiciones del mercado que enfrenta la aseguradora, por lo que aporta en gran medida la toma de decisiones administrativas y financieras. Asimismo, considerar distintas fuentes de información aumenta la visión del sector.

La información utilizada se limita a la relacionada con los accidentes ocurridos, complementar el presente estudio con datos relacionados con los ingresos de la aseguradora, las características de los clientes, los gastos fijos, entre otros; permitiría aumentar la precisión de la investigación y desarrollar una estrategia de mayor eficiencia.

Estrategia de resolución

El primer paso es la descripción visual de las principales características y relaciones correspondientes a los siniestros de automóviles, lo anterior con la finalidad de conocer a detalle el comportamiento y circunstancias en las que suceden. Se identifican factores clave que permitan establecer relaciones para lograr los objetivos planteados.

Para las siguientes etapas de la investigación e utilizan modelos de análisis multivariante; en específico dos modelos de interdependencia y uno de series temporales.

Las regiones son definidas utilizando un análisis de clustering jerárquico, el cual permite la asociación de distintos estados de la república en función de su similitud de distintas variables numéricas, específicamente se utiliza la cantidad de autos en circulación, cantidad de accidentes, edad promedio de los involucrados en el siniestro, número de muertos en accidentes, presencia de alcohol en los accidentes.

Una vez definidas las regiones, se estima un índice de costos haciendo uso del análisis de componentes principales, por medio del cual se crea un componente con las variables relacionadas a los costos que incurre la aseguradora al cubrir los siniestros reportados, estas variables son: nivel de lesión de los involucrados, gastos por hospital, fallecimiento, ambulancias, grúas, pérdida total del vehículo y nivel del daño.

La información utilizada fue obtenida de la base de datos de la aseguradora AXXA.

Es importante señalar que no se cuenta con información monetaria de los costos, por lo que para el cálculo del índice de costos se siguió la siguiente metodología: Las variables mencionadas en el párrafo anterior son variables dicotómicas por lo que se convirtieron a variables numéricas utilizando la cantidad mensual total por estado (ej. cantidad total de ambulancias utilizadas en el estado de Guanajuato). Una vez obtenida la nueva variable numérica, se supone que los costos unitarios de cada variable son iguales en todos los estados, de esta manera, un estado con mayor cantidad de grúas, ambulancias, casos con necesidad de hospital, etc; tendrá un mayor índice de costos.

Considerando la correlación de los Estados de la república mexicana con el componente, es posible calcular el índice de costos de manera individual, para después conjuntar los estados en su región correspondiente según se definió con el análisis de clustering.

Por último, se utiliza la metodología Holt-Winters para realizar el pronóstico de la cantidad de accidentes durante el año 2018, utilizando la información obtenida de la INEGI la cual contiene el registro de todos los accidentes ocurridos en el país durante los tres años anteriores.

Se realizan pronósticos mensuales para el periodo de tiempo mencionado, mediante el uso de la metodología Holt-Winters se estima el factor de estacionalidad y la tendencia de la cantidad de accidentes.

Una vez obtenido el pronóstico, el índice de costos y las regiones en las que

se dividirá el territorio, es posible detectar los patrones de comportamiento y diseñar una estrategia financiera y administrativa para maximizar la rentabilidad esperada de la aseguradora.

Capítulo 3

Análisis Descriptivo

El presente capítulo tiene como objetivo el describir visualmente las principales características y relaciones correspondientes a los siniestros de vehículos de motor, lo anterior con la finalidad de conocer a detalle el comportamiento de dicho fenómeno y las circunstancias en las cuales se presenta.

El análisis descriptivo se enfoca en tres principales componentes: La evolución del sector automotriz en México, Accidentes ocurridos a nivel nacional y Siniestros registrados por la aseguradora AXXA.

Se busca encontrar posibles relaciones que permitan identificar los factores clave para la estimación de la ocurrencia futura de siniestros y, por ende, la optimización de la estrategia financiera y la administración de recursos por parte de la aseguradora.

3.1. Industria Automotriz

Para conseguir el objetivo de describir el comportamiento de la ocurrencia de siniestros y sus características, se debe comenzar con la descripción del mercado global del cual se generan; en otras palabras, se debe presentar las condiciones del sector automotriz en México para dar una idea general del panorama en el cual los percances de vehículos de motor ocurren.

Son tres principales variables las que permiten describir el sector automotriz, estas son: Producción, Ventas de vehículos de motor y Vehículos registrados en circulación. Las primeras dos reflejan la concentración de vehículos en el futuro (la producción en el mediano-largo plazo y las ventas en el corto)

y la tercera variable indica las condiciones actuales.

La gráfica 3.1 muestra la relación entre las variables en los últimos 13 años (2005-2018). Cabe mencionar que la gráfica solo se refiere a la producción, venta y registro de automóviles particulares, esta última variable es representada por el tamaño de la burbuja.

Los datos fueron obtenidos del portal de la INEGI.

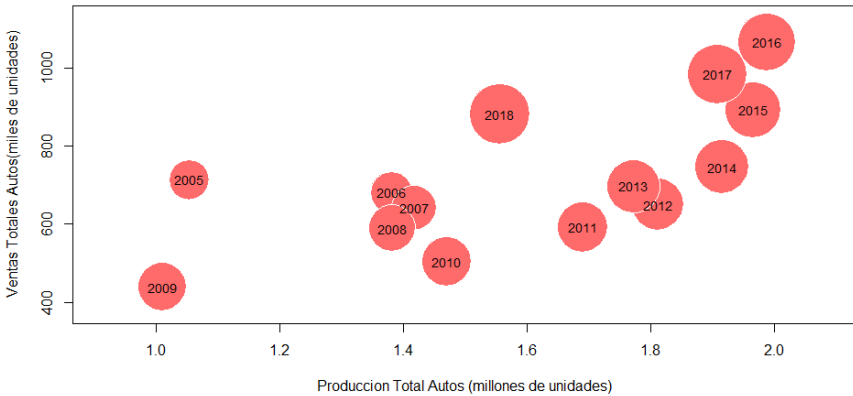


Figura 3.1: Panorama Sector Automotriz en México

Dos conclusiones generales se pueden obtener del análisis del gráfico, la primera es la existencia de la relación positiva entre la producción y venta de autos, de manera más notoria después del año 2011. La segunda es el incremento recurrente en la cantidad de autos en circulación.

De manera particular, se puede observar el efecto de la crisis mundial en 2008, la cual generó una baja considerable en la producción y ventas de autos en 2009. Así mismo, del año 2015 al 2018, la producción de autos ha ido a la baja, al igual que las ventas (aunque en menor medida), este fenómeno indica que el incremento en la cantidad de autos en circulación sea cada vez menor.

A continuación, se presenta el comportamiento individual de cada una de las variables.

Los gráficos 3.2 y 3.3 contemplan la producción total de vehículos de motor (autos, camiones de carga, camiones de pasajeros, motocicletas, etc.). A diferencia del gráfico de burbujas, el nivel de producción del 2016 no es el mayor, esto puede ser explicado a la incorporación de otros tipos de vehículos, es decir, en el año 2017 y 2018 la producción de autos disminuyó, pero se incrementó la producción de camiones, motos, etc. El histograma de la producción de camiones muestra un sesgo a la derecha, en el cual el nivel de producción es aún mayor que el de coches, esas observaciones pertenecen a los dos últimos años.

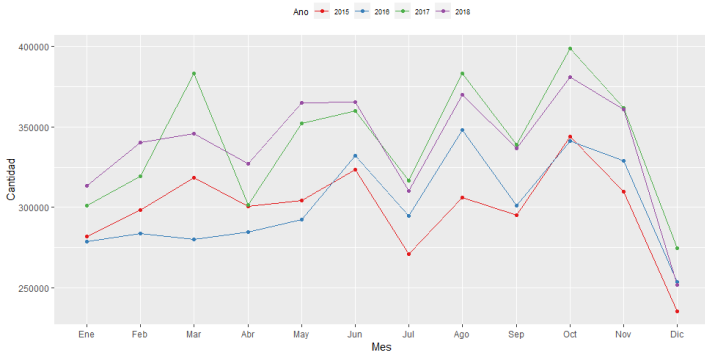


Figura 3.2: Producción Anual de Vehículos (2015-2018)

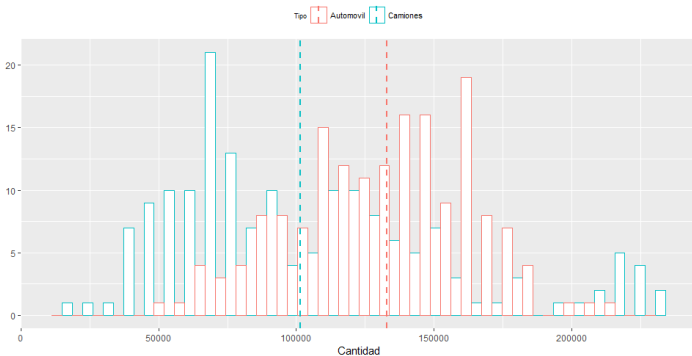


Figura 3.3: Producción Mensual de Vehículos (2017-2018)

Adicionalmente, se observa la posible existencia de estacionalidades en la producción de vehículos de motor; en específico, los meses de julio y diciembre muestran una caída abrupta en los niveles de producción, mientras que marzo, junio, agosto y octubre presentan alzas.

Respecto las ventas, los gráficos 3.4 y 3.5 muestran su comportamiento por mes y por tipo de auto.

En términos generales, las ventas de autos presentan un comportamiento estable con excepción del mes de diciembre, donde las ventas promedio son mayor al resto del año.

Si consideramos los trimestres del año, el segundo trimestre presenta el menor valor en ventas promedio (al menos de manera visual), mientras que el último muestra una tendencia a la alza.

Los autos importados son el tipo de vehículo con mayor número de ventas promedio, seguido de los autos subcompactos, aunque las ventas promedio de los autos compactos son muy similares a los anteriores.

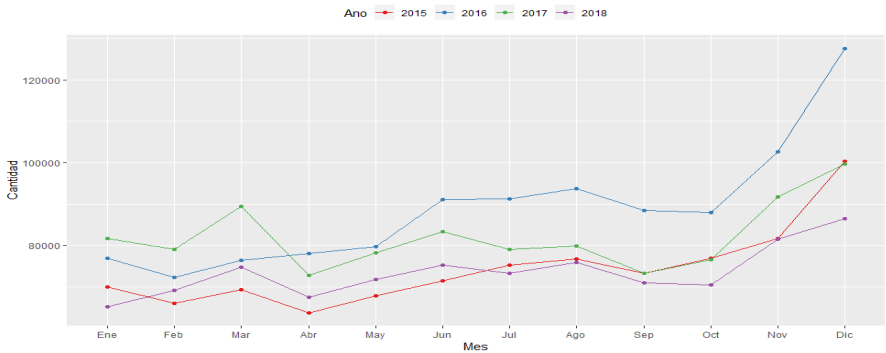


Figura 3.4: Ventas de Vehículos por mes (2015-2018)

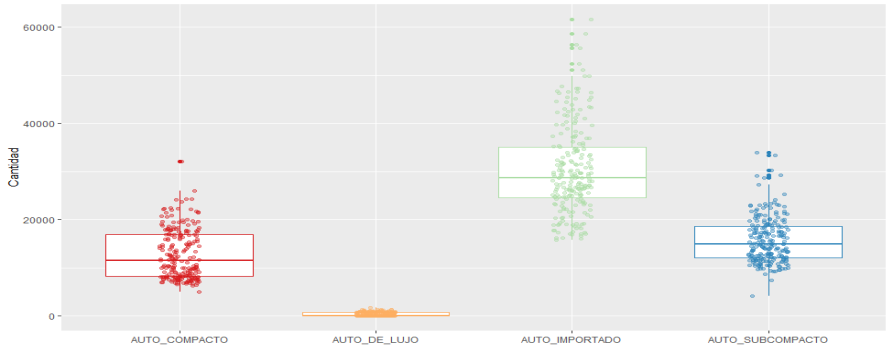


Figura 3.5: Ventas por tipo de Vehículo (2015-2018)

La variable con mayor relevancia es la cantidad de vehículos de motor en circulación, ya que define la población de interés. La gráfica 3.6 describe el comportamiento de dicha variable.

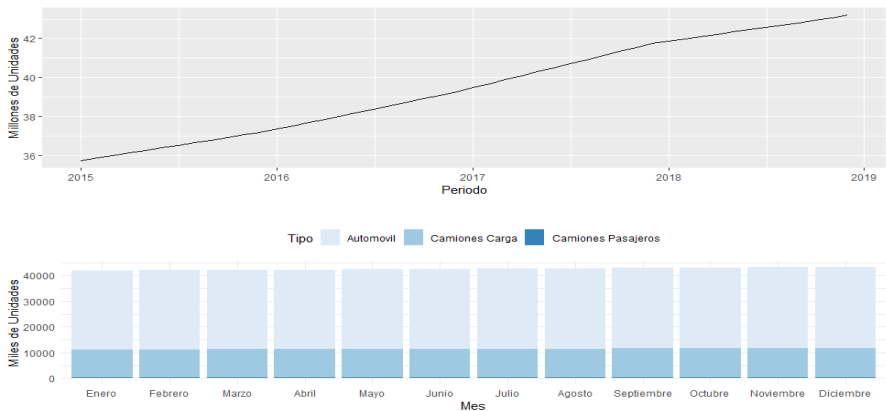


Figura 3.6: Circulación de Vehículos (2015-2018)

Se aprecia de forma clara la tendencia al alza en la cantidad de vehículos de motor en circulación, de los cuales el automóvil particular concentra más del 60% del total. Derivado de esto, el presente estudio se enfoca principalmente en los siniestros presentados en este tipo de vehículo.

El mapa de calor mostrado en la figura 3.7, presenta el incremento en la cantidad de autos en circulación por mes, dado que el gráfico de ventas mensuales mostraba que diciembre presentaba la mayor cantidad de ventas, es de esperarse que los mayores incrementos se observen en el mes de enero. Con excepción de ese mes, el resto de los meses sigue la tendencia predominante en el año en curso.

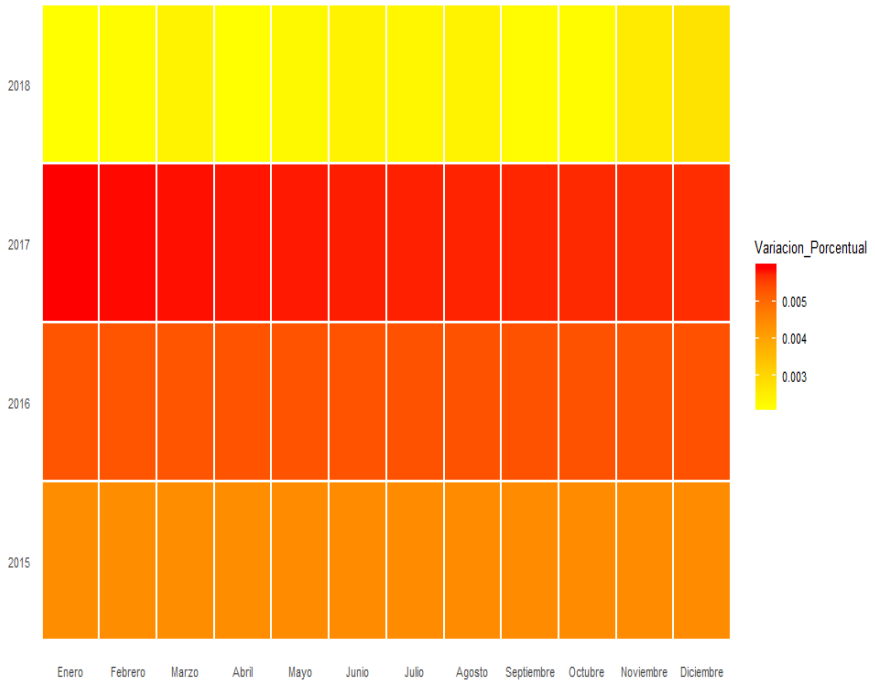


Figura 3.7: Variación en Circulación de Vehículos

Los años 2015, 2016 y 2017 presentan las mayores variaciones en el stock de autos registrados, mientras que 2018 muestra un decremento en los niveles de variación (acorde a lo observado en el primer gráfico donde se mostraban niveles muy bajos de producción y ventas para este año).

La locación es una variable importante a considerar, por lo que a continuación se muestran 4 mapas (Figuras 3.8-3.11) con la concentración de automóviles, camiones y motocicletas en los distintos estados de la república en el año 2017.

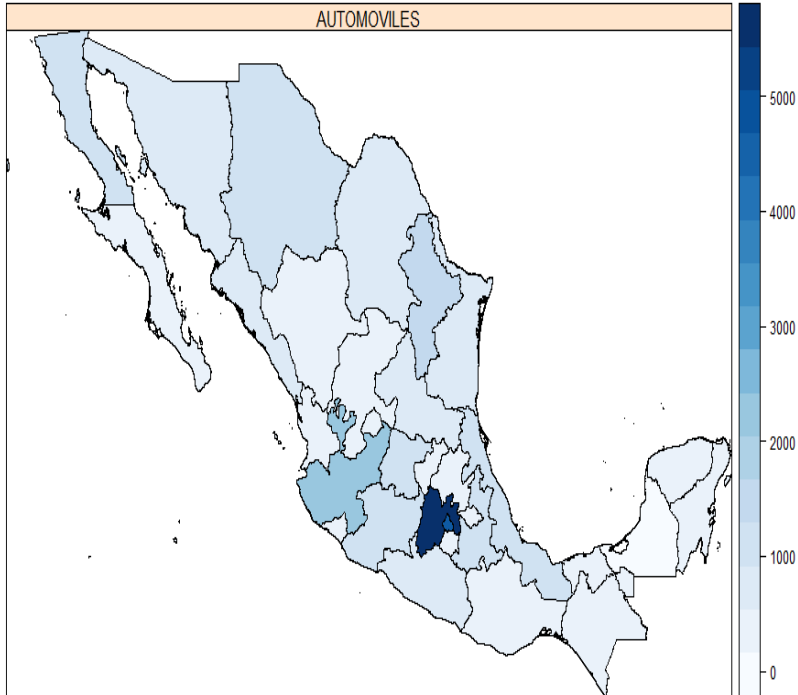


Figura 3.8: Autos en Circulación 2017 (Miles de Unidades)

Respecto a los automóviles, destacan el Estado de México, CDMX y Jalisco, en un segundo nivel aparecen Nuevo León, Chihuahua, Veracruz, Puebla, Guanajuato y Michoacán.

Considerando los 4 tipos de vehículos, los principales estados con mayores niveles de circulación son: Estado de México, CDMX, Nuevo León, Jalisco, Veracruz, Chihuahua, Guerrero, Puebla, Michoacán y Guanajuato.

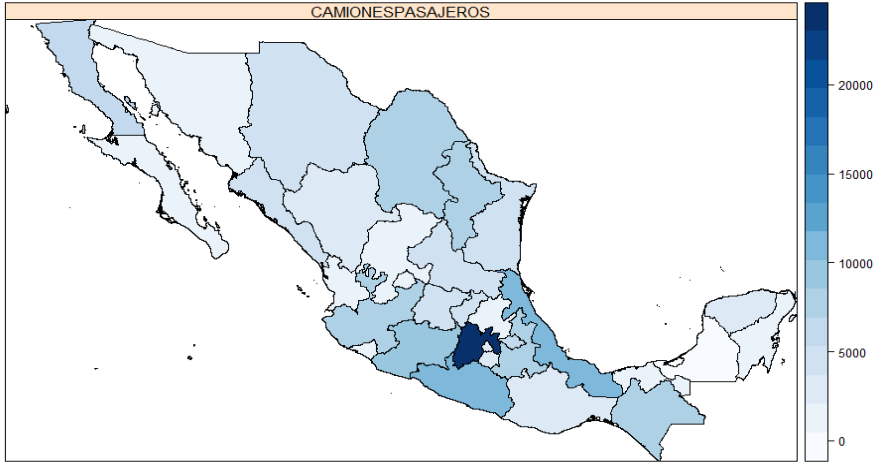


Figura 3.9: Camiones Pasajeros en Circulación 2017

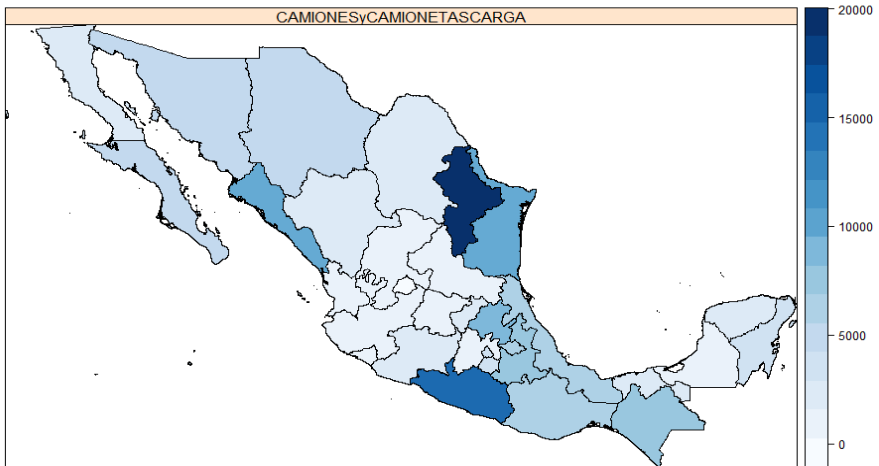


Figura 3.10: Camiones de Carga en Circulación 2017

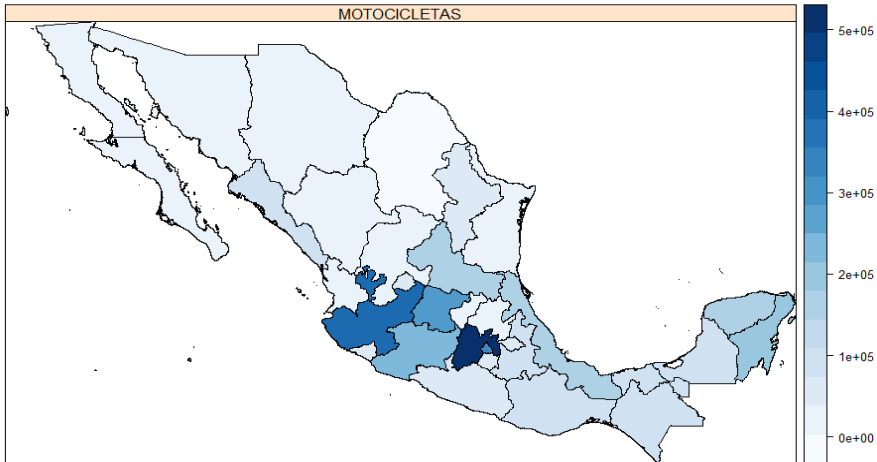


Figura 3.11: Motocicletas en Circulación 2017

3.2. Accidentes a nivel nacional

En la presente sección, se muestran los accidentes registrados a nivel nacional, esta información fue obtenida a través de la INEGI y se refiere a los siniestros ocurridos en el año 2017.

El objetivo de presentar la información de los percances a nivel nacional, es establecer una posible relación entre la cantidad de vehículos de motor en los estados y los accidentes ocurridos, así mismo, realizar una comparación entre dichos accidentes y a los que la aseguradora AXXA reporta. Esta comparación puede ser útil en varios sentidos: al establecer una relación entre el número de siniestros en el país contra los que AXXA registra, se puede identificar si el comportamiento es similar y si no lo es identificar un posible nicho de mercado para expandir la cobertura de seguros de vehículos de motor.

Adicionalmente, se cuenta con información sobre las circunstancias y gravedad de los siniestros, por lo que se identificará los posibles costos a los cuales se enfrenta la aseguradora y si estos coinciden con las características de los accidentes registrados a nivel nacional.

El primer punto, es identificar la relación entre la cantidad de vehículos con

los siniestros ocurridos. El mapa 3.12 muestra la distribución por estado de los accidentes registrados por la INEGI a nivel nacional de vehículos de motor.

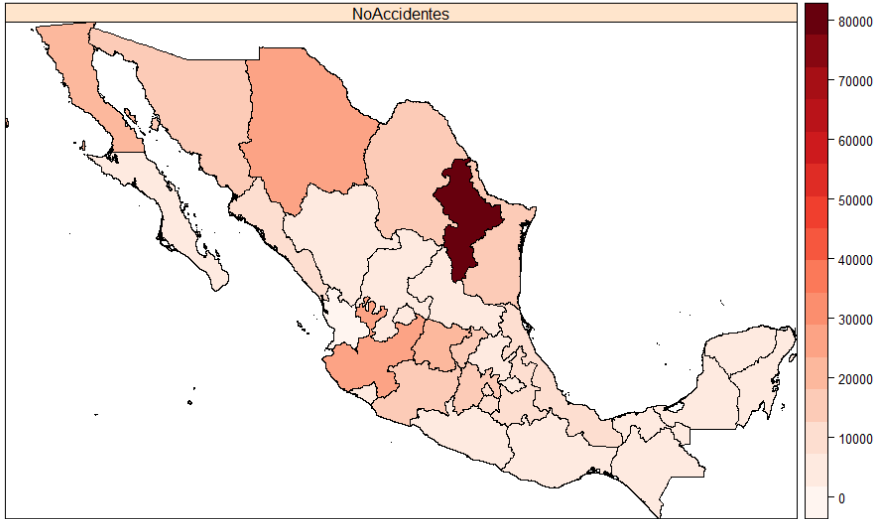


Figura 3.12: Número de Accidentes Nacionales 2017

El estado con mayor registro de accidentes es Nuevo León. Realizando la comparación entre los niveles de circulación y los accidentes, Nuevo León destaca por la cantidad de camiones en circulación, ya sea de carga o de pasajeros, referente a autos se encuentra dentro de los primeros 5.

Seguido de Nuevo León, se encuentra Jalisco, Chihuahua y Baja California, los cuales también presentan altos niveles de circulación de vehículos. El Estado de México ocupa el primer lugar en vehículos registrados, sin embargo, sus niveles de accidentes no son los mayores, aunque cabe resaltar que si se encuentra entre los 10 estados con mayor cantidad de siniestros.

Otra variable importante para describir el comportamiento de los siniestros es la fecha de ocurrencia, la figura 3.13 muestra un calendario del año 2017 en el cual se representa la cantidad de accidentes ocurridos en cada día del año.

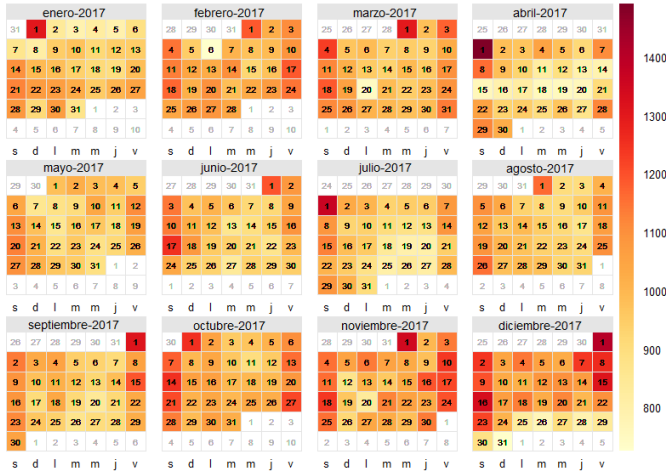


Figura 3.13: Cantidad de Accidentes por día en 2017

De manera general, resaltan tres puntos importantes: los fines de semana presentan mayor cantidad de accidentes, en el último trimestre del año los siniestros aumentan y en el primer día de cada mes la cantidad de accidentes aumenta.

A continuación, se muestra la gráfica 3.14 para incorporar información sobre el genero del responsable y la gravedad del siniestro.

Las observaciones relevantes del gráfico 3.14 son las siguientes:

- La proporción de accidentes fatales es pequeña, y es menos recurrente en los días martes y viernes.
- Las mujeres tienen una menor tasa de accidentes fatales.
- La cantidad de siniestros en los que un hombre es responsable es mayor.
- Los fines de semana presentan una mayor proporción de accidentes en los cuales los siniestros resultan fatales.

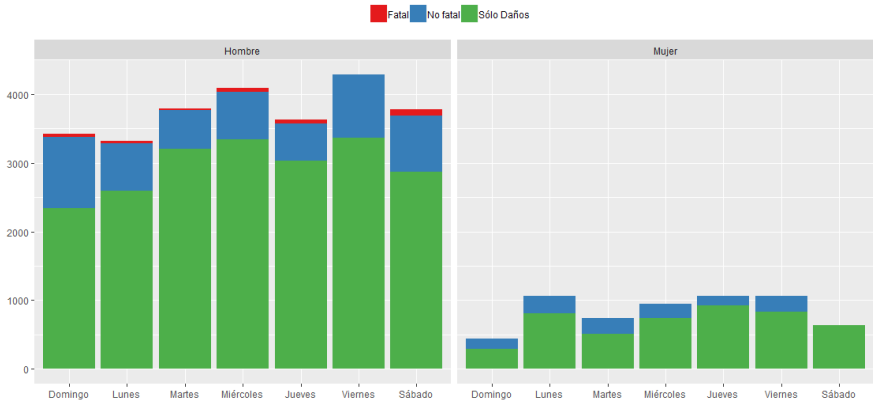


Figura 3.14: Clasificación de Accidentes por Género 2017

La figuras 3.15 y 3.16 muestran la distribución de accidentes en función del tipo de transporte y la causa del siniestro.

	Tipo Transporte	Cantidad Accidentes
1	AUTOMOVIL	431171
2	CAMIONETA PASAJEROS	87085
3	MICROBUS	5299
4	CAMION PASAJEROS	21442
5	OMNIBUS	2873
6	TRANVIA	42
7	CAMIONETA	50449
8	CAMION	16925
9	TRACTOR	9557
10	FERROCARRI	370
11	MOTOCICLETA	47175
12	BICICLETA	5703
13	OTRO VEHICULO	21019

Figura 3.15: Accidentes por tipo de Transporte 2017

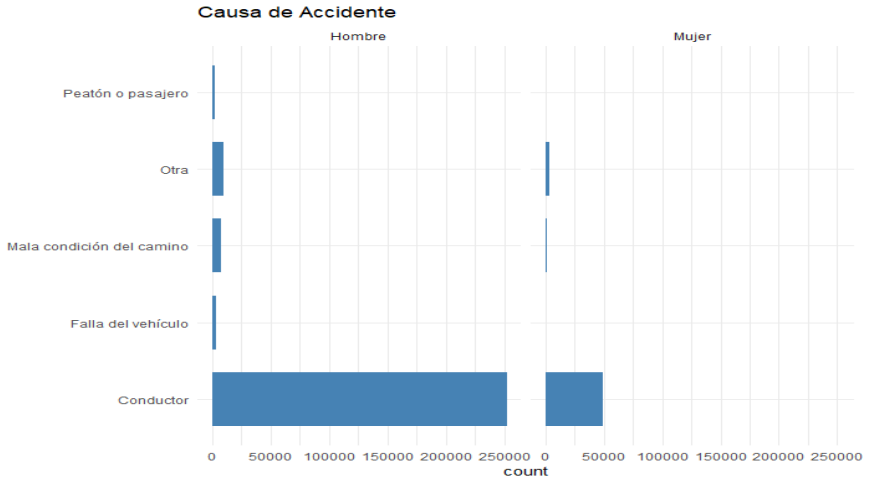


Figura 3.16: Accidentes por causa 2017

Un aspecto determinante en relación a los siniestros es la seguridad al volante; en específico, el uso de cinturón y el consumo de alcohol. Las figuras siguientes (3.17-3.20) muestran la relación de las dos variables en función de la cantidad de muertos y heridos en los siniestros a nivel nacional.

Respecto los hombres, son más propensos a no utilizar cinturón o al menos las circunstancias del accidente no permiten obtener la información sobre su uso.

En comparación con los hombres, la propensión de las mujeres de conducir bajo la influencia del alcohol es menor.

La distribución de personas muertas en accidentes es similar entre hombres y mujeres, en la cual en caso de que haya muertos, la mayoría de las ocasiones es 1 persona. La distribución del número de muertos en función del uso de cinturón muestra que cuando se usa cinturón, independientemente del género, la cantidad de muertos es menor.

Las figuras 3.19 y 3.20 muestra la información de accidentes cuando existen personas heridas.

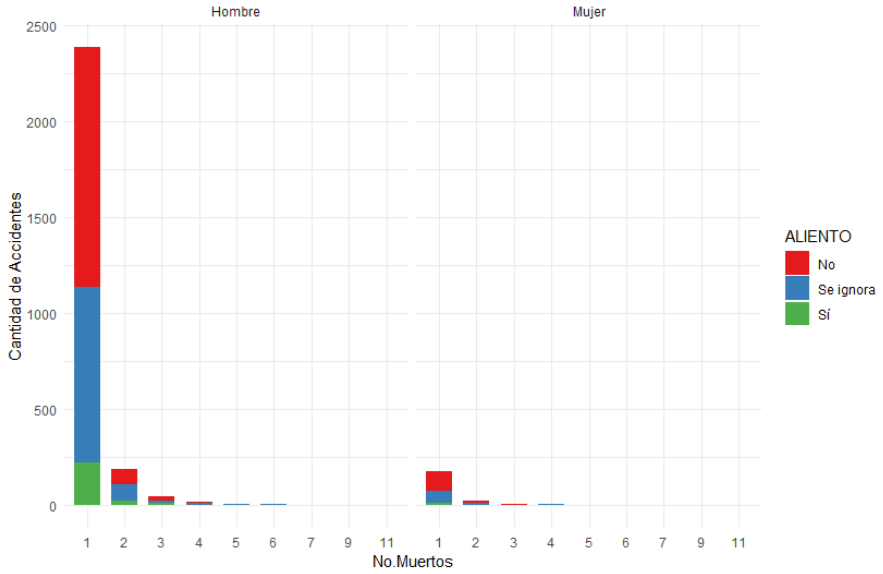


Figura 3.17: Seguridad al Volante 2017

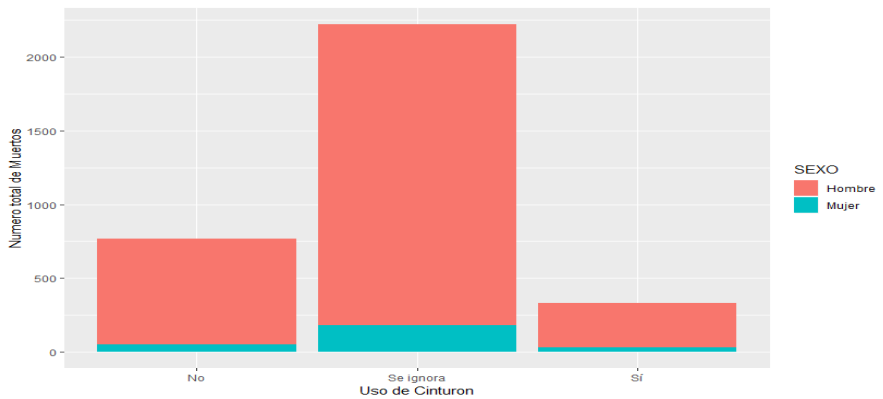


Figura 3.18: Cantidad de muertos en accidente 2017

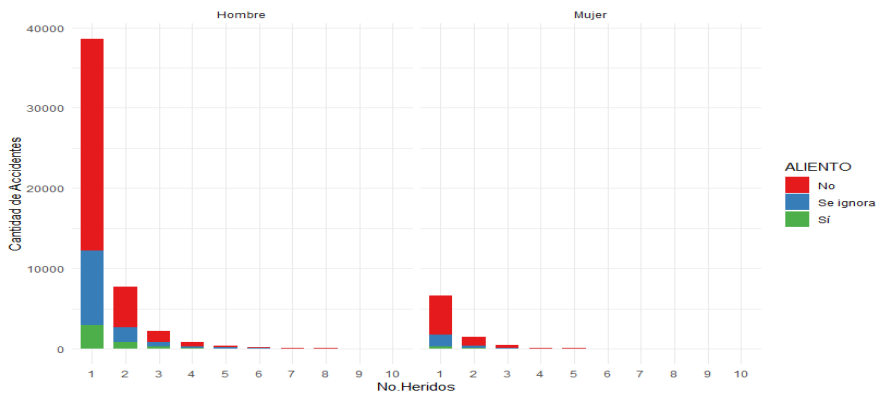


Figura 3.19: Accidentes con personas heridas 2017

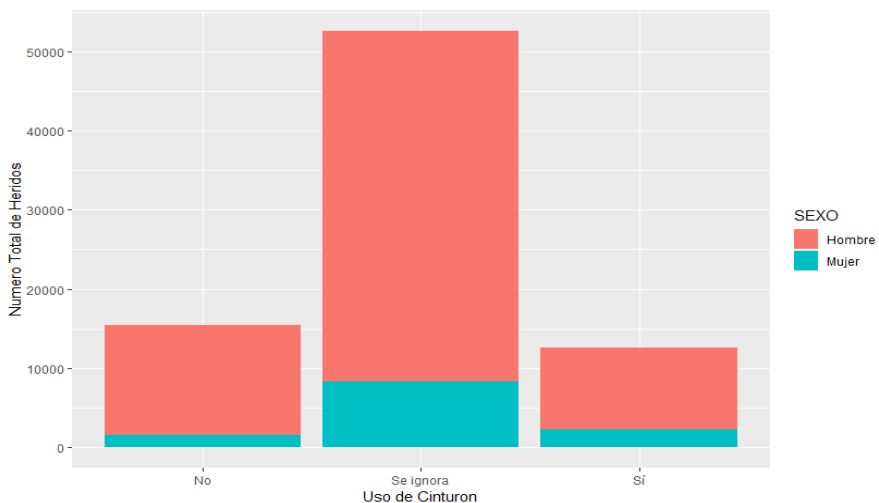


Figura 3.20: Cantidad de heridos en accidentes 2017

3.3. Aseguradora AXXA

La última sección del presente capítulo presenta la información respecto los siniestros registrados por la aseguradora AXXA. Los datos presentados

corresponden al año 2017, esto con objetivo de realizar la comparación entre la cantidad de autos y los accidentes a nivel nacional presentados en las secciones anteriores.

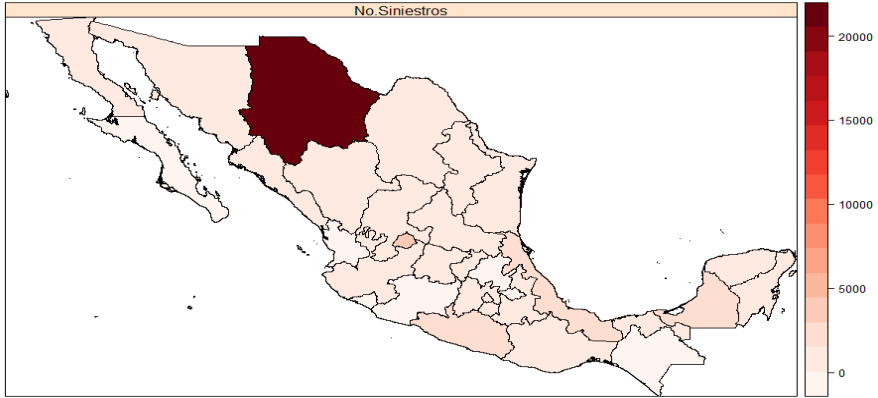


Figura 3.21: Cantidad de Accidentes AXXA

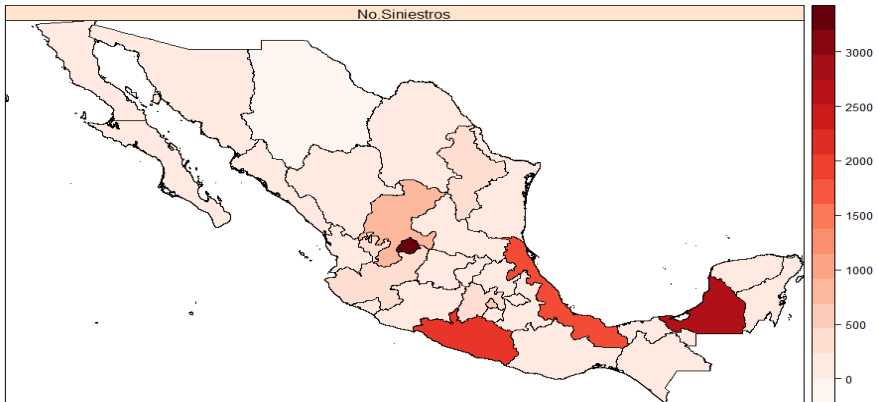


Figura 3.22: Cantidad de Accidentes AXXA (S/ Chihuahua)

El estado de Chihuahua presenta la mayor cantidad de siniestros, tal es la diferencia con otros estados que el primer mapa no permite identificar de manera clara cuales estados son los siguientes en numero de siniestros. Por lo tanto, se muestra un segundo mapa eliminando la información de Chihuahua, de esta manera se identifica de mejor manera que los estados críticos en número de accidentes son: Aguascalientes, Campeche, Veracruz, Guerrero y Zacatecas.

Referente al Estado de Chihuahua, en el número de vehículos en circulación se encuentra dentro de los primeros 7, resaltando en mayor medida en camiones de carga. Considerando el número de accidentes a nivel nacional, Chihuahua se encuentra en el top 3.

Aguascalientes y Campeche no resaltan los rubros de vehículos en circulación ni en accidentes a nivel nacional, el hecho de que AXXA presente un alto índice de siniestros en este estado pudiera ser por la concentración de clientes que tiene en esas regiones. Por último, Veracruz y Guerrero sobresalen en vehículos en circulación.

La figura 3.23 muestra el calendario del año 2017 y la cantidad de siniestros registrados por la aseguradora.

Retomando lo observado en la sección anterior, los siniestros de AXXA no coinciden con el patrón de accidentes nacionales, los fines de semana no se aprecia una tendencia clara de un aumento en la cantidad de siniestros, así mismo, del último trimestre del año solo diciembre presenta mayor cantidad de percances.

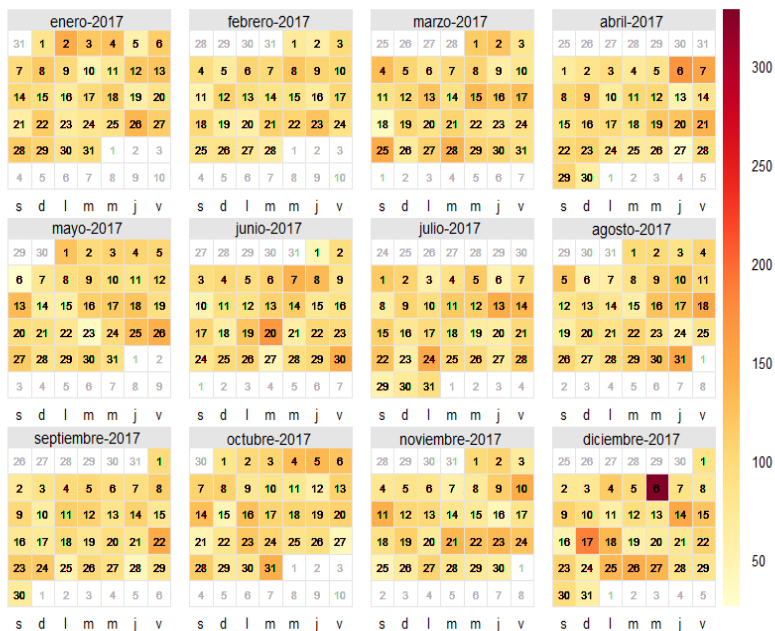


Figura 3.23: Calendario accidentes AXXA

El siguiente mosaico presenta la relación entre el nivel de daño y el punto de impacto en función del tipo de vehículo.

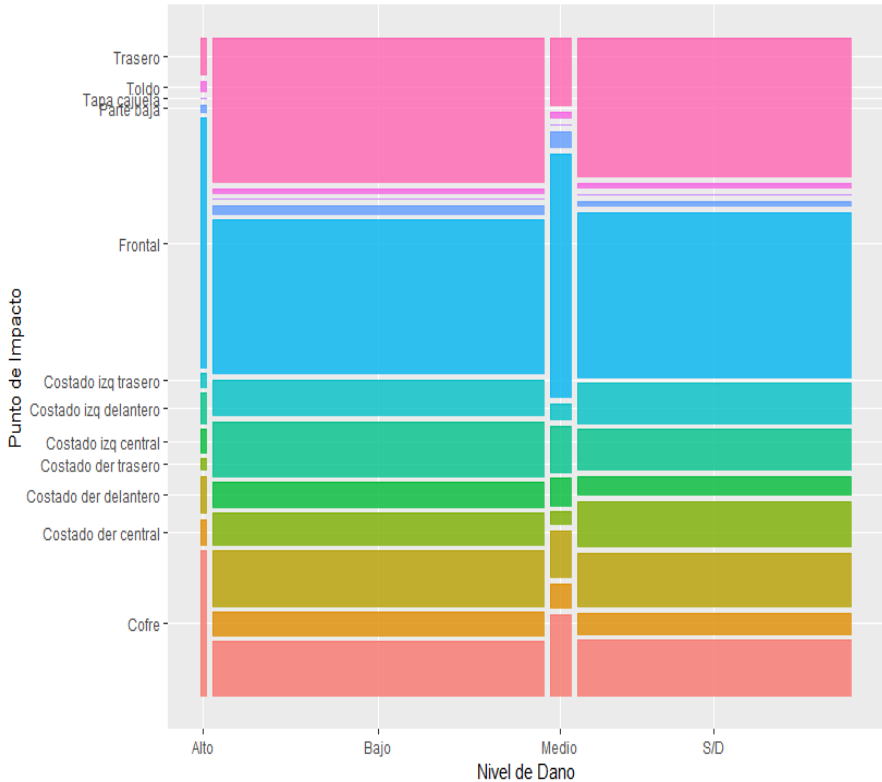


Figura 3.24: Gravedad de los Accidentes

- La mayor cantidad de siniestros presentan un nivel de daño bajo o sin daño alguno.
- Los dos principales puntos de impacto son: Trasero y Frontal.
- Cuando el daño es alto, los puntos de impacto más recurrentes son Frontal y Cofre.

Para finalizar la sección del análisis descriptivo, se muestra la información relacionada a los costos que incurre la aseguradora al atender a los vehículos asegurados (ver gráficas 3.25-3.29). Dado que no se tiene información sobre los costos monetarios que implica la atención del percance para la aseguradora, se muestran las proporciones en las cuales se presentan gastos adicionales a

la reparación de los automóviles.

Los principales costos identificados son: hospitalización de involucrados, fallecidos en siniestros, uso de ambulancia y grúas y pérdida total del automóvil.

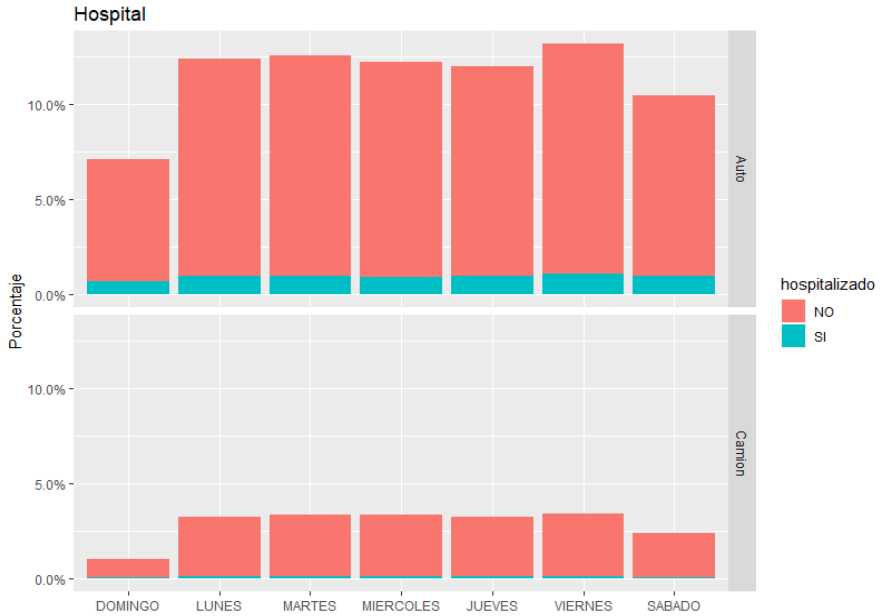


Figura 3.25: Hospitalización

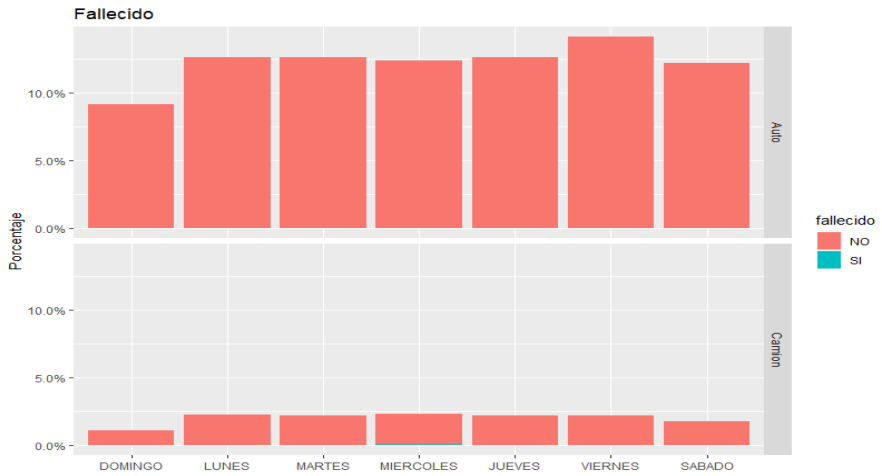


Figura 3.26: No. Fallecidos

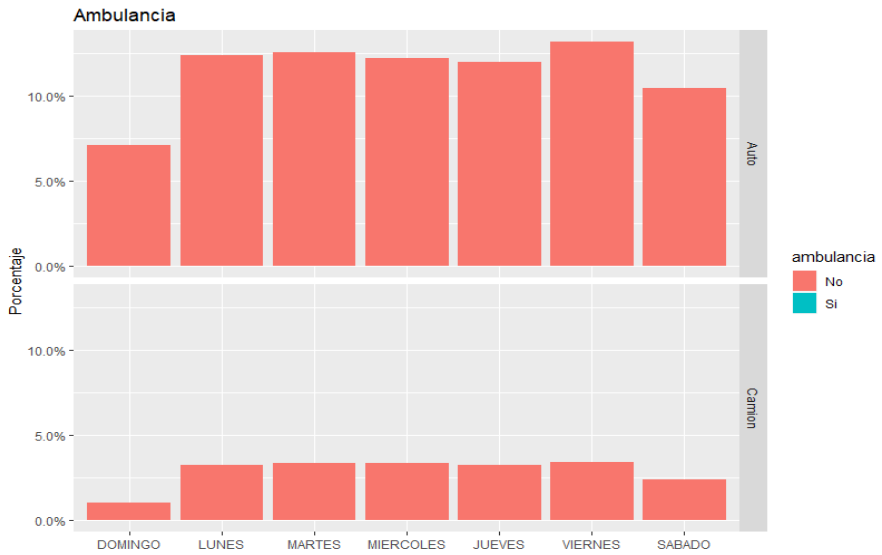


Figura 3.27: Uso de ambulancias

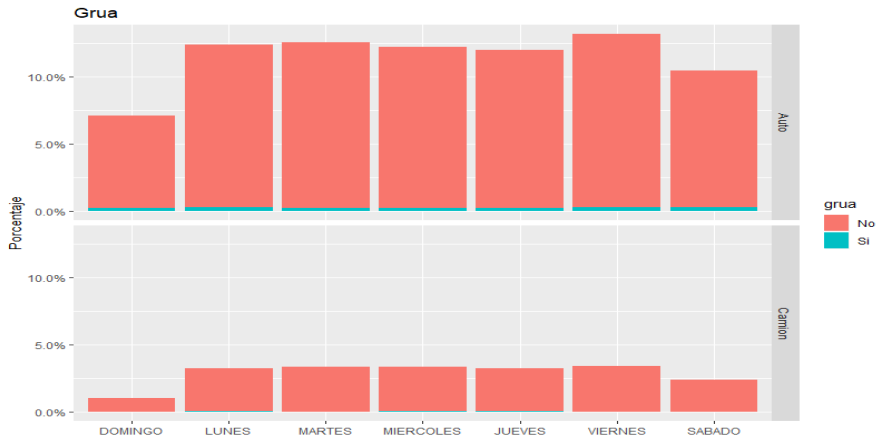


Figura 3.28: Uso de grúas

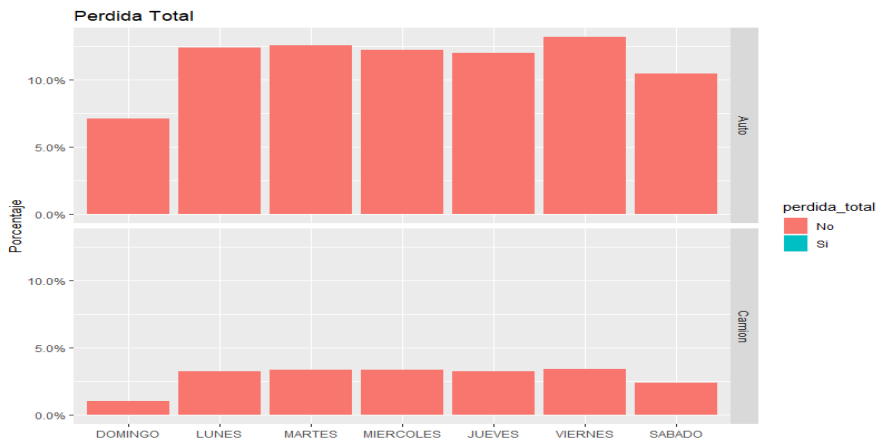


Figura 3.29: Pérdida Total

Derivado del análisis de las gráficas de costos, los que se presentan en mayor medida es la hospitalización de involucrados y el uso de grúas. Los demás, aunque sucedan en menor proporción pueden tener un coste esperado mayor, por lo que en capítulos posteriores se analizara en mayor medida la probabilidad y el impacto que podrían tener este tipo de costos para la aseguradora.

Capítulo 4

Identificación de Clusters

En el capítulo anterior, se mostraron las distintas características de los accidentes a nivel nacional, este segundo capítulo tiene como fin establecer relaciones entre los distintos estados de la república para encontrar similitudes que permitan a la aseguradora determinar regiones conformadas por estados homogéneos y así, implementar estrategias financieras y administrativas propias de cada región.

El capítulo muestra la metodología, estimación y presentación de resultados de una técnica de *clustering*, la cual permite identificar las regiones mencionadas y obtener las principales diferencias entre ellas. Específicamente, se utiliza el análisis de clustering jerárquico, definiendo como medida de asociación la distancia euclidiana y como técnica de agrupación el método de Ward.

De esta manera, la elaboración de una estrategia óptima de cobertura por parte de la aseguradora considerará dichas relaciones y diferencias haciendo más eficiente la administración de los recursos humanos y financieros.

4.1. Planteamiento

Para realizar el análisis se utiliza información obtenida de INEGI, en específico las bases de datos llamadas: *Accidentes de tránsito terrestre en zonas urbanas y suburbanas*; y *Vehículos de motor registrados en circulación*; para el año 2017.

Con base a la información indicada se construye una base de datos con las variables analizadas en el capítulo anterior: Número de automóviles par-

ticulares en circulación, cantidad de accidentes en el año, edad promedio del responsable del accidente, número total de muertos por siniestros, número total de heridos por siniestros, número de accidentes en días de fin de semana, cantidad de accidentes cuyos involucrados no usaban el cinturón de seguridad, cantidad de accidentes cuyos involucrados presentaban aliento alcohólico.

Cada variable es medida para cada uno de los 32 estados de la república.

El modelo propuesto a utilizar es el análisis de clustering, el cual permite clasificar una muestra de individuos en un número reducido de grupos mutuamente excluyentes basado en las similitudes existentes entre los mismos. Dado que se utiliza el análisis de clustering jerárquico, las variables deben de ser numéricas lo cual cumple la base de datos generada.

Previo a la estimación del modelo, se debe decidir si es conveniente el uso de todas las variables en la base de datos, pues variables con un alta correlación entre sí complicarían la interpretación de los resultados y no aportarían información adicional a la formación de los cluster.

La figura 4.1 muestra la matriz de correlación entre las 8 variables.

Observando la matriz, se perciben variables con un alto nivel de correlación entre sí, por lo que lo más conveniente es eliminarlas del análisis. Se utiliza un método de *paso a paso* como criterio para la eliminación de las variables, la primera en ser excluida es la cantidad de accidentes sin uso de cinturón, dado que presenta un coeficiente de correlación por encima de 0.80 con las variables que miden la cantidad de accidentes.

Una vez excluida la primera variable, se calcula nuevamente la matriz de correlación, y se eliminan las variable número de muertos en siniestros y número de accidentes en días de fin de semana, dado que presentan coeficientes de correlación mayores a 0.7.

La matriz de correlación de las variables seleccionadas se muestra en la figura 4.2

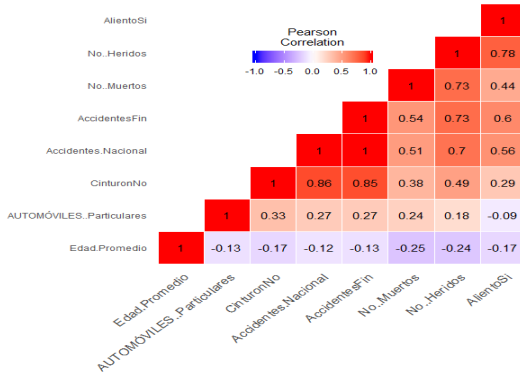


Figura 4.1: Matriz de correlación

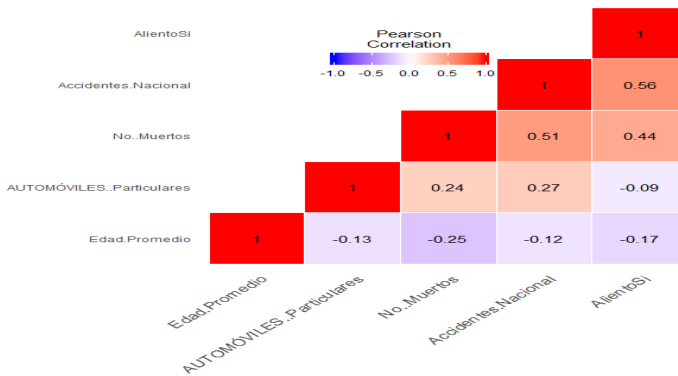


Figura 4.2: Matriz de correlación

4.2. Estimación del modelo

4.2.1. Existencia de cluster

El método de clustering genera agrupaciones aunque no sea evidente la existencia de similitud entre los datos, por lo tanto, para comenzar el análisis, el primer objetivo es comprobar la existencia de similitudes.

Para descubrir la existencia de agrupaciones evidentes, se realiza una comparación entre los datos observados y datos aleatorios formados con base a las observaciones. Los datos aleatorios se obtienen mediante una generación de datos basados en la distribución uniforme cuyos parámetros son obtenidos de los datos reales.

Después se calcula el estadístico de Hopkins para ambos conjuntos de datos, comparando los estadísticos se descubre la existencia de cluster en los datos reales.

Para visualizar ambos conjuntos de datos, se reduce su dimensionalidad a través del análisis de componentes principales, las gráficas 4.3 y 4.4 muestran ambos conjuntos de datos.

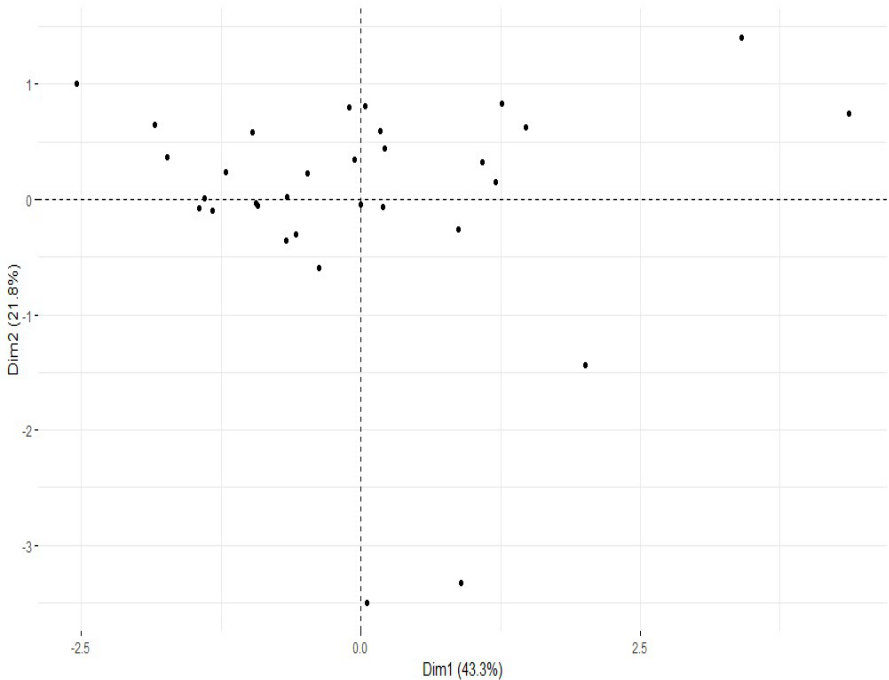


Figura 4.3: Datos Observados

El estadístico de Hopkins es usado para determinar la tendencia de un conjunto de datos a formar cluster; midiendo la probabilidad de que un conjunto de datos dado sea generado por una distribución uniforme; es decir prueba la

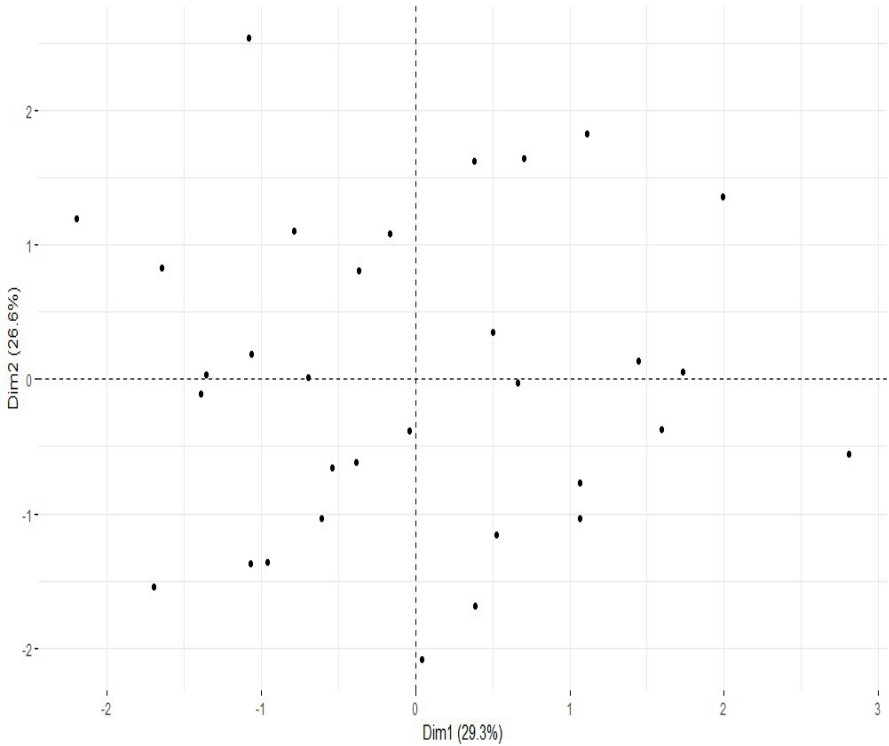


Figura 4.4: Datos Aleatorios

aleatoriedad espacial de los datos.

La hipótesis que se somete a evaluación es la siguiente:

H0: El conjunto de datos proviene de una distribución uniforme, es decir, no hay agrupaciones significativas.

H1: El conjunto de datos no proviene de una distribución uniforme, es decir, hay agrupaciones significativas.

Para valores del estadístico por encima de 0.5; se considera que el conjunto de datos tiene una tendencia a formar cluster, por encima de 0.75 se tiene un 90 % de nivel de confianza de la existencia de tal tendencia.

El estadístico de los datos observados es de 0.7153, mientras que el de los datos aleatorios es de 0.4566. Por lo tanto, se puede concluir que existen cluster en los datos observados.

4.2.2. Técnica de clustering

Se utilizan los métodos jerárquicos de cluster para la elaboración del análisis. Como medida de asociación se utiliza la distancia euclidiana.

En primer lugar los registros de la base de datos son estandarizados para evitar problemas debido a la escala de medición de cada variable. A partir de la matriz estandarizada de datos se procede al calculo de la matriz de distancias.

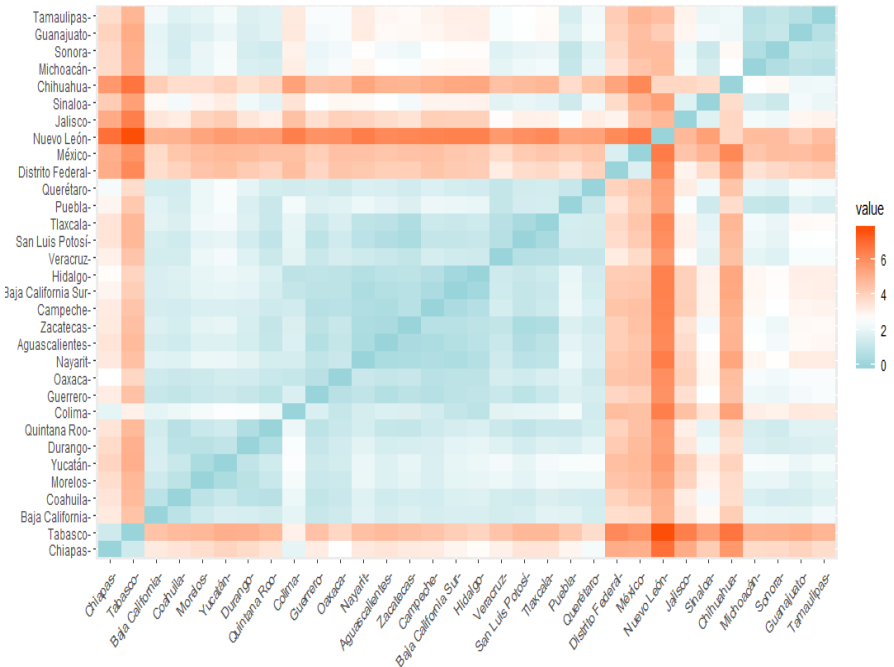


Figura 4.5: Matriz de Distancias

Del gráfico que muestra los valores de la matriz de distancias (ver figura 4.5) se aprecia a simple vista que Nuevo León y Chihuahua resaltan el distanciamiento que tienen con el resto de los estados, la misma situación se presenta

en el Estado de México, CDMX, Tabasco y Chiapas. Desde Baja California hasta Querétaro las distancias que presentan son visualmente cortas (en orden de la matriz, no en distancia geográfica).

La técnica de agrupación utilizada es el método de Ward, el cual es un procedimiento en el que, en cada etapa, se unen dos cluster para los cuales se tenga el menor incremento en el total de la suma de los cuadrados de las diferencias, dentro de cada cluster, de cada individuo al centroide del cluster. Este procedimiento fue presentado por primera vez en el año 1963 por Joe H. Ward.

El gap statistic recomienda que el número indicado de cluster sea de 9. Sin embargo, haciendo uso del paquete *NbClust* de R, el cual permite comparar 30 índices distintos, recomienda que sean 5. (Ver figuras 4.6 y 4.7)

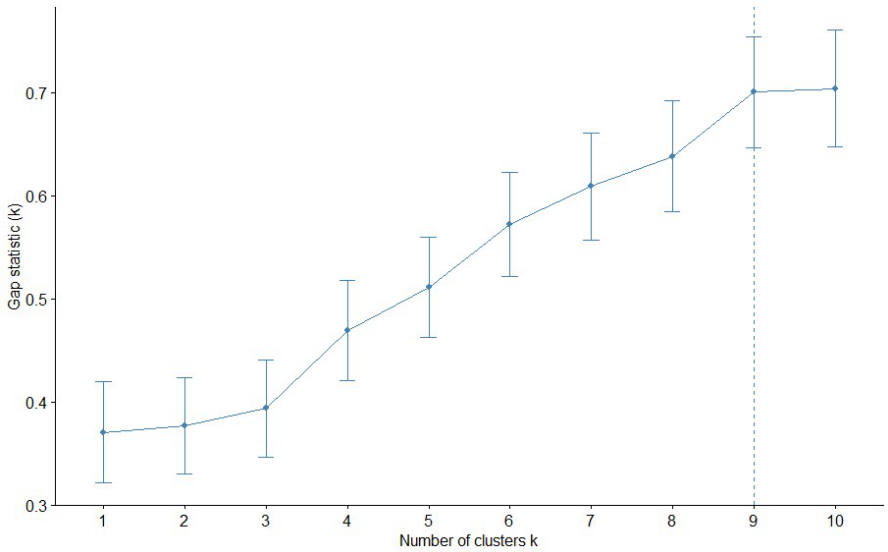


Figura 4.6: Gap statistic

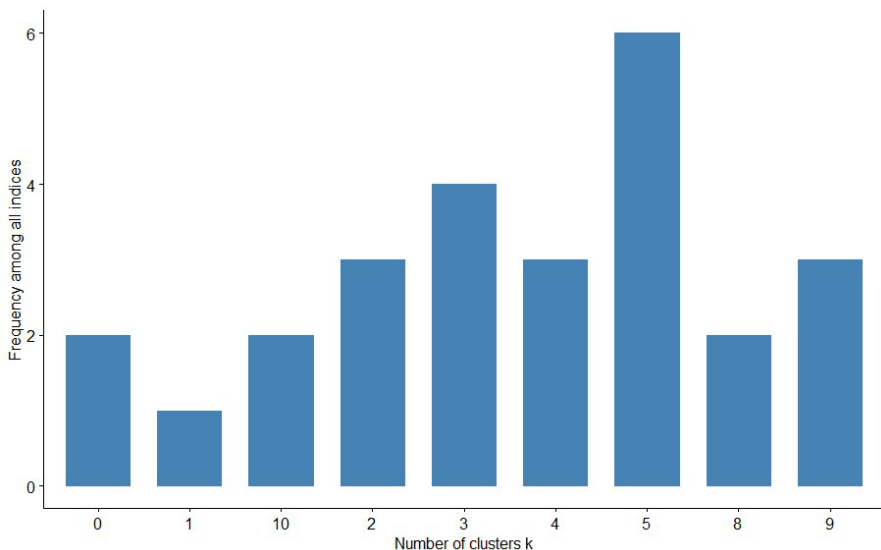


Figura 4.7: Índices

Se estima la técnica de clustering con ambos números de cluster y se comparan sus diferencias. Como se muestra en las figuras 4.8 y 4.9, tres agrupaciones resultantes son idénticas, mientras que tres grupos del dendograma de la figura 4.8 se unen para conformar el cluster con más elementos en el dendograma de la figura 4.9. También hay una fusión entre los tres últimos grupos que muestra el dendograma de la fig 4.8

Nuevo León por un lado, y CDMX y el Estado de México por otro están separados del resto. Chihuahua que parecía estar más cercano de Nuevo León, el método de clustering para cinco grupos lo coloca en el cluster de Guanajuato. Tabasco y Chaipas conforman un cluster que no sufre cambios al reducir el número de grupos.

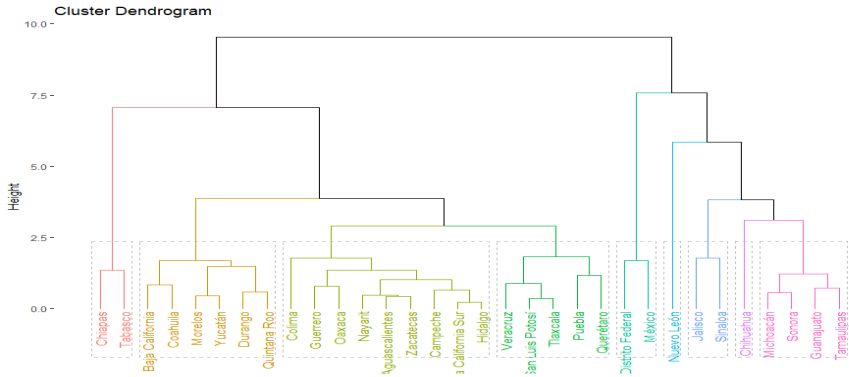


Figura 4.8: Dendrograma 9 Clusters

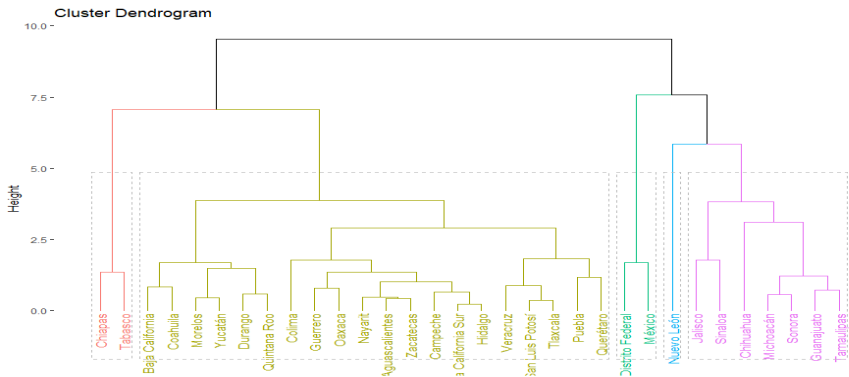


Figura 4.9: Dendrograma 5 Clusters

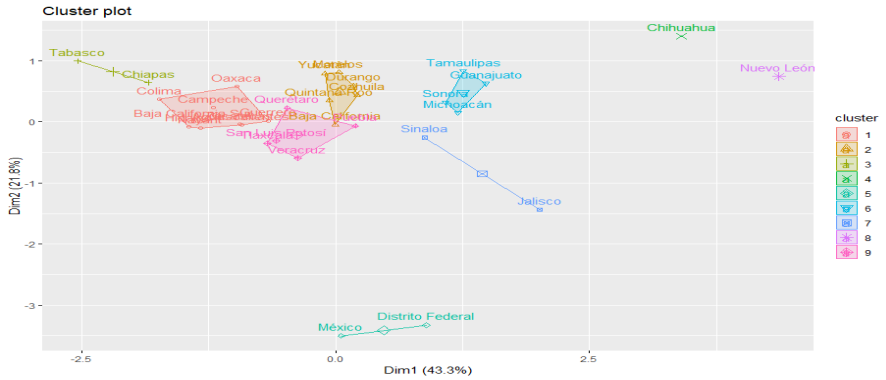


Figura 4.10: Proyección Bidimensional de agrupaciones resultantes (9 grupos)

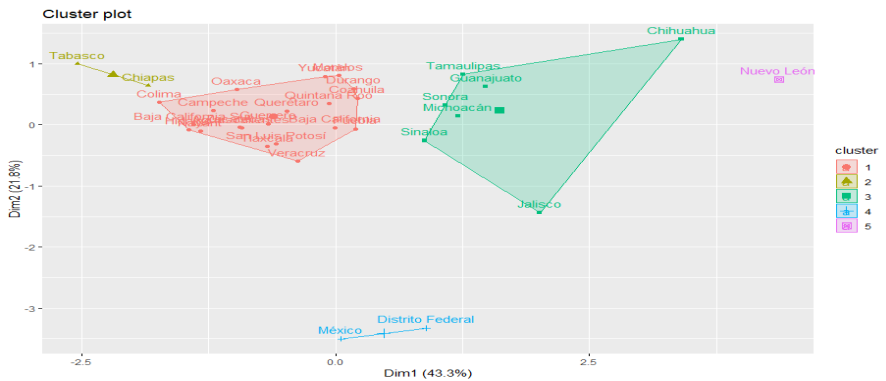


Figura 4.11: Proyección Bidimensional de agrupaciones resultantes (5 grupos)

Para decidir el resultado que mejor se adapta a los datos se utiliza la gráfica silueta, la cual presenta el grado de asociación de cada uno de los estados respecto al resto de miembros de su cluster en comparación con los estados en otros cluster.

El valor de la silueta va de -1 a 1, y un valor cercano a 1 representa que el estado ha sido bien clasificado.

Como se muestra en las figuras 4.12 y 4.13, el valor promedio de la silueta con 5 cluster es mayor al de 9, por lo que el modelo resultante es el de 5 agrupaciones.

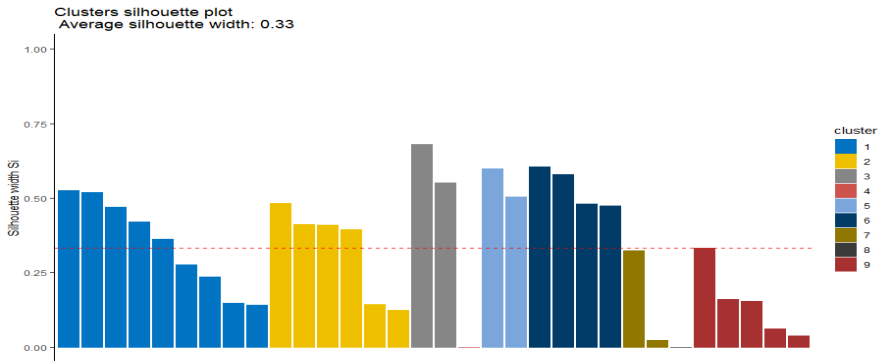


Figura 4.12: Gráfico de silueta para agrupaciones resultantes (9 grupos)

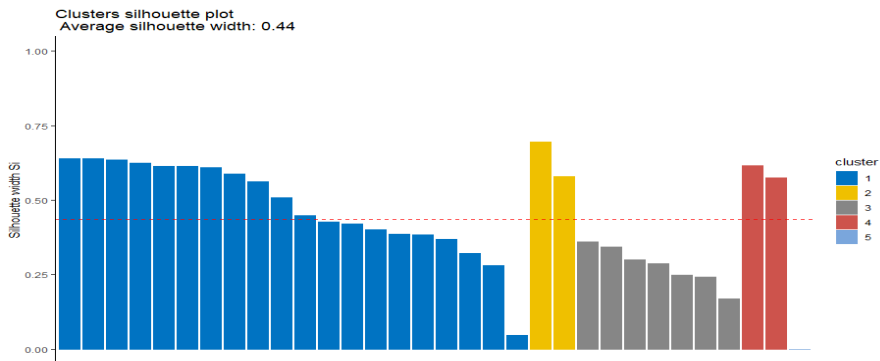


Figura 4.13: Gráfico de silueta para agrupaciones resultantes (5 grupos)

A continuación, se muestra la tabla con las clasificaciones de los estados y al cluster al que pertenecen.

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Aguascalientes	Chiapas	Chihuahua	CDMX	Nuevo León
Baja California	Tabasco	Guanajuato	México	
Baja California Sur		Jalisco		
Campeche		Michoacán		
Coahuila		Sinaloa		
Colima		Sonora		
Durango		Tamaulipas		
Guerrero				
Hidalgo				
Morelos				
Nayarit				
Oaxaca				
Puebla				
Querétaro				
Quintana Roo				
San Luis Potosí				
Tlaxcala				
Veracruz				
Yucatán				
Zacatecas				

Figura 4.14: Clasificación final de las 32 entidades federativas

Para comparar el comportamiento de los cluster en función de las variables utilizadas se presenta el gráfico de medias estandarizadas, el cual permite ver de manera clara las características de cada uno de los clusters.

MEDIA DE LOS CLUSTER

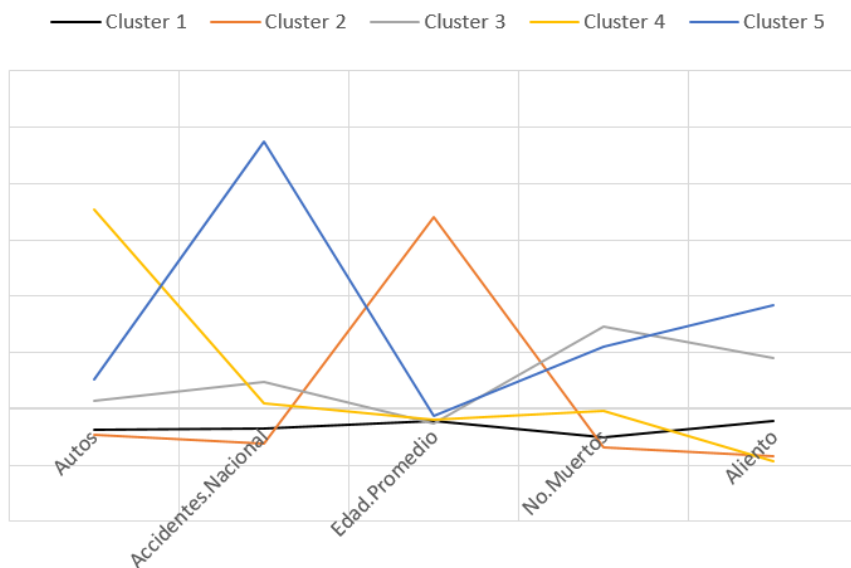


Figura 4.15: Comportamiento de Clusters

El cluster 4, destaca en la cantidad de autos en circulación y en los demás factores tiene un comportamiento similar al resto de los cluster, por lo tanto, se puede considerar el cluster 4 como el mayor nicho de mercado para una aseguradora, ya que presenta la mayor cantidad de clientes potenciales y su riesgo de accidentes es similar la media del país. Este cluster está compuesto por la CDMX y el Estado de México, por lo que los resultados del análisis son congruentes.

El cluster 5 resalta en dos ámbitos, el primero es el número de accidentes y el segundo es en el número de accidentes en el algún involucrado presento aliento alcohólico. Se debe mencionar, que este cluster tiene la segunda media de autos en circulación más alta, por lo que encontrar una cantidad tan elevada de accidentes no es tan inusual. Ocupa la segunda posición en número de muertos por accidente, por lo que considerando los resultados se puede decir que el cluster 5 es el que presenta mayores problemas en cuestión de accidentes viales. En este cluster se encuentra el estado de Nuevo León.

El cluster 3 presenta características similares al cluster anterior, aunque en magnitud, su media de accidentes es mucho menor que en Nuevo León, si

es mayor que la del resto del país y ocupa el primer lugar en número de muertos y segundo en accidentes con personas con aliento alcohólico involucrados; además tiene edad promedio más baja, este cluster también presenta muchas problemáticas respecto a los accidentes viales. En este cluster se encuentran los estados de Chihuahua, Jalisco, Guanajuato, Sinaloa y Sonora.

Un caso contrario a los dos grupos anteriores es el caso del cluster 1, donde sus niveles de accidentes, autos en circulación, etc; no resaltan respecto los otros grupos, se podría decir que son estados donde no hay mayores problemas de accidentes viales.

Se debe resaltar un punto importante, no es posible conocer la demanda a la que se enfrenta la aseguradora en los distintos cluster, ni la información financiera de cada uno de los mismos, por lo que no es posible categorizar un cluster como rentable o no rentable. Por ejemplo, el cluster 3 presenta mucha cantidad de accidentes fatales y no fatales, lo que a la aseguradora le representa costos de cobertura, a diferencia del cluster 1 que no ocurren siniestros, pero debido a esto quizá a la constante presencia de accidentes las personas en el cluster 1 estén más dispuestas a contratar un seguro y a un precio más alto, por lo que las primas compensarían los costos en los que se incurre.

El siguiente capítulo presenta una estimación de los costos de cada cluster en base a factores cualitativos como lo son el uso de grúas, lesiones, ambulancias, etc.

Capítulo 5

Índice de Costo Estatal

En el capítulo anterior, haciendo uso de la información nacional sobre accidentes viales, se formaron grupos de estado con características similares. En el presente capítulo se hará uso de esos clúster para determinar el costo al que se enfrenta la aseguradora en cada uno de ellos.

Una vez definidas las regiones (grupos de estados) y sus costos relacionados, la aseguradora es capaz de diseñar e implementar una estrategia comercial que maximice sus ganancias.

Cabe señalar que no se tiene información financiera sobre los accidentes de la aseguradora, los costos se estimaran con base a variables categóricas como uso de grúas, hospitalización del asegurado, ambulancias requeridas, etc; las cuales indirectamente representan costos de cobertura para la aseguradora.

5.1. Estimación del índice

Los datos requeridos para el análisis son obtenidos del Instituto Internacional de Ciencia de Datos, donde la aseguradora AXXA publica de manera gratuita la información sobre sus siniestros, la base de datos es llamada “Datos AXXA de percances viales”.

De la misma manera que en el capítulo anterior, se utiliza la información del año 2017.

La metodología a seguir consiste en crear un índice de costos estatal, el cual funcione como un indicador que contemple los gastos de cobertura que la aseguradora hace frente en cada uno de los estados; específicamente, se cuen-

Variable	Descripción
Nivel de Lesión	Contiene tres categorías, Alto, Medio y Bajo.
Hospitalización	Cantidad de personas que fueron hospitalizadas
Cantidad de fallecidos	Numero de muertos en los percances
Grúa	Numero de grúas requeridas
Ambulancia	Cantidad de ambulancias requeridas
Perdida total del vehículo	Cantidad de autos que se reportaron como perdida total
Fuga	Cantidad de percances donde el responsable huyo
Nivel de daño	Contiene tres categorías: Alto, medio, bajo.

tan el número de percances viales en los cuales se presentaron las siguientes características:

Las variables mencionadas conforman el índice por cada uno de los estados, seguido se conforman los clúster formados en el capítulo anterior y se comparan los costos de cada uno. La meta principal es definir las regiones más rentables para la aseguradora considerando las características de percances viales y los costos que generan para la aseguradora.

Para la estimación del índice de costos se utiliza el análisis de factores mediante el método de componentes principales, la cual es una técnica de interdependencia, que permite la reducción de la dimensión de la estructura de datos generando nuevas variables, para el presente caso el primer componente es usado para la creación del índice. Se utilizan las observaciones de las 32 entidades federativas con un total de 12 variables por estado.

La figura 5.1 muestra la varianza explicada por los componentes.

El primer componente explica el 73.1% de la varianza total, por lo que el índice a estimar explicara gran parte de la variabilidad de los costos.

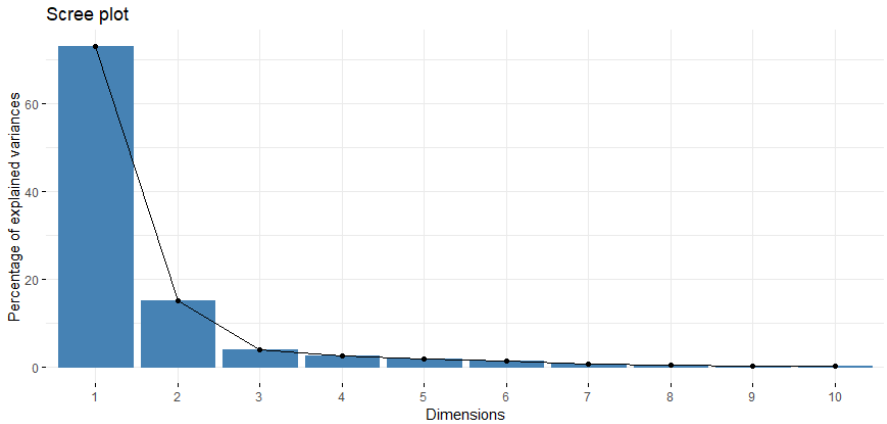


Figura 5.1: Varianza Explicada

La gráfica de cargas para los 2 primeros componentes se muestra en la figura 5.2, adicionalmente se muestra la información sobre su contribución a la formación de los componentes.

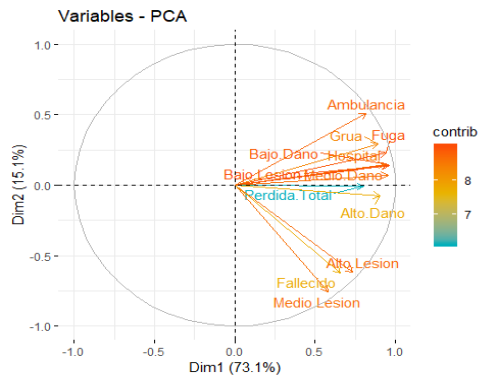


Figura 5.2: Gráfica de cargas

Como se mencionó anteriormente, el primer componente es el que será usado para la creación del índice, por lo que es importante conocer la contribución de las variables específicamente a este componente, la figura siguiente muestra el resultado.

El índice será construido considerando dichas contribuciones de las variables.

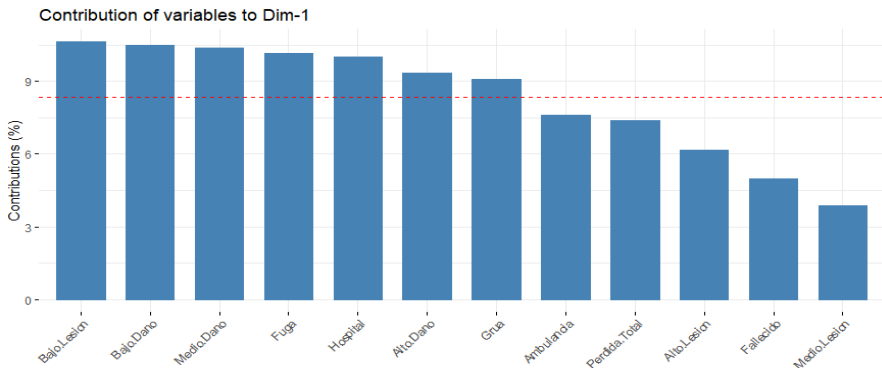


Figura 5.3: Gráfica de contribuciones

La gráfica 5.4 muestra la calidad de representación de los estados y su correlación con los componentes.

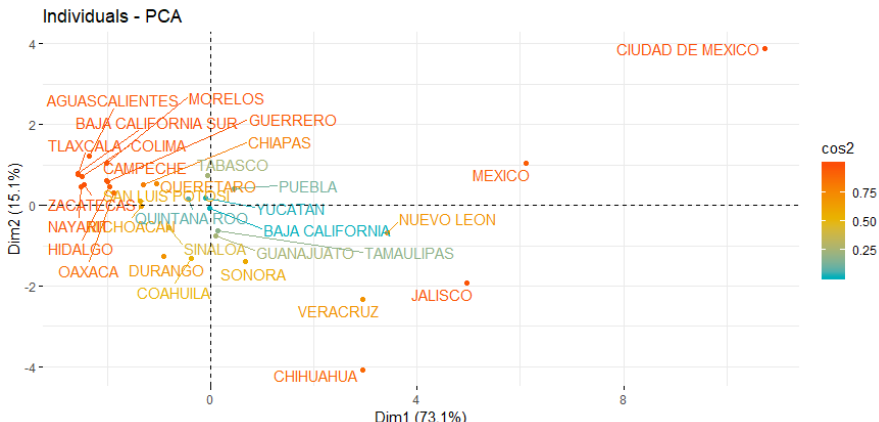


Figura 5.4: Calidad de Representación

CDMX, el Estado de México y Jalisco presentan los mayores costos, asimismo, la calidad de representación es de las más altas entre todos los estados.

Tlaxcala, Baja California Sur, Aguascalientes, entre otros; son los estados que representan el menor costo para la aseguradora. En términos generales, con excepción de Yucatán, Baja California y Quintana Roo, la calidad de representación de los estados es alta.

La gráfica .5 incorpora la información presentada en las anteriores figuras, asimismo, separa los estados por el clúster al cual pertenecen.

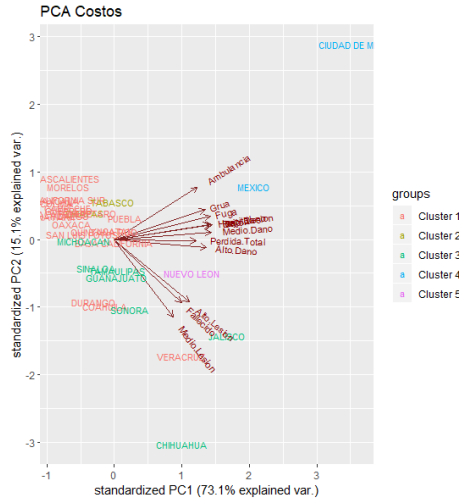


Figura 5.5: Gráfica de doble proyección

Visualmente, los clúster 4 y 5, los mas pequeños en cuanto a número de estados que los conforman, presentan los mayores costos operacionales para la aseguradora. El cluster 1, el que contiene mayor cantidad de estados, es el de menor costo, aunque el estado de Veracruz representa los mismos costos que el estado de Nuevo León.

5.2. Índice de Costos

El índice de costos es calculado a través de la correlación entre los distintos estados y el primer componente, una vez calculada dicha correlación se estandariza para obtener el índice definitivo.

AGUASCALIENTES	-0.792	MORELOS	-0.681
BAJA CALIFORNIA	-0.007	NAVARRIT	-0.85
BAJA CALIFORNIA SUR	-0.865	NUEVO LEON	1.154
CAMPECHE	-0.679	OAXACA	-0.637
CHIAPAS	-0.439	PUEBLA	0.153
CHIHUAHUA	0.994	QUERETARO	-0.354
CDMX	3.626	QUINTANA ROO	-0.147
COAHUILA	-0.126	SAN LUIS POTOSI	-0.458
COLIMA	-0.841	SINALOA	-0.271
DURANGO	-0.304	SONORA	0.231
GUANAJUATO	0.036	TABASCO	-0.017
GUERRERO	-0.671	TAMAULIPAS	0.048
HIDALGO	-0.66	TLAXCALA	-0.87
JALISCO	1.676	VERACRUZ	0.996
EDO. MEXICO	2.062	YUCATAN	-0.03
MICHOACAN	-0.456	ZACATECAS	-0.826

Figura 5.6: SemafORIZACIÓN del Índice de Costos

El mapa de calor de la figura 5.6 presenta los costos individuales por estados, la distribución nacional del índice se muestra en el siguiente histograma 5.7

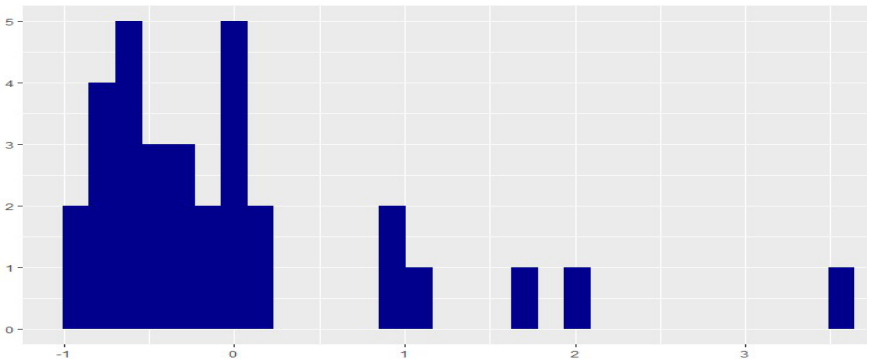


Figura 5.7: Distribución del Índice

La distribución del índice presenta un sesgo a la derecha derivado al alto costo en la CDMX y el Estado de México, la mayor concentración se da en un intervalo entre -1 y 0. Es importante señalar que el valor del índice intrínsecamente carece de utilidad, su beneficio se da de manera relativa es decir al momento de comparar con el resto de los estados.

Agrupando los estados de acuerdo a los resultados obtenidos en el capítulo anterior, podemos describir el costo por región.(Ver Figura 5.8)

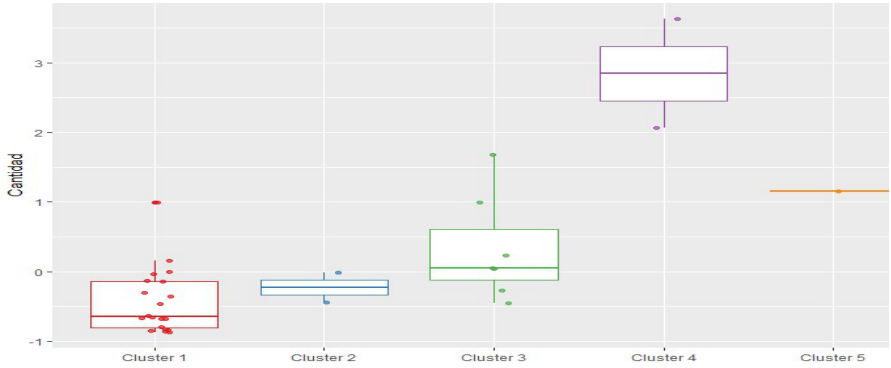


Figura 5.8: Costo por Clúster

El costo medio de la región 4 y 5 supera al resto de los estados, cabe señalar que debido a la varianza del clúster 3 existen estados dentro de la región que superan los costos de la región 4. La región 2 es el clúster con la menor varianza por lo que la predicción del gasto en esta región será más precisa. La región 1 presenta el menor costo medio; sin embargo, contiene estados con costos similares a la región 4.

Como parte final del presente capítulo se presenta la comparación entre las características de percances viales de las regiones y los costos que implican para la aseguradora.

REGIÓN	ESTADOS MIEMBRO	CARACTERÍSTICAS	COSTOS	CONCLUSIÓN
CLÚSTER 1	Aguascalientes, Baja California, Baja California Sur, Campeche, Coahuila, Colima, Durango, Guerrero, Hidalgo, Morelos, Nayarit, Oaxaca, Puebla, Querétaro, Quintana Roo, San Luis Potosí, Tlaxcala, Veracruz, Yucatán, Zacatecas	Sus niveles de accidentes, autos en circulación y número de muertos no resaltan respecto los otros grupos, se podría decir que son estados donde no hay mayores problemas de accidentes viales.	Es la región de menor costo, el estado de Veracruz es su miembro más costoso presentado un gasto similar a Chihuahua (perteneciente al clúster 3). La mayoría de sus estados miembro son los de menos gasto a nivel nacional e incluye estados de alta densidad poblacional.	Es una región con niveles bajos de accidentes y los percances que existen son de bajo costo. La baja circulación de automóviles en la región puede ser una limitante para el desarrollo de la región; sin embargo, cuenta con estados de alta densidad poblacional. Con base a su índice de costo, los estados de mayor proyección de rentabilidad son: San Luis Potosí, Querétaro, Durango, Quintana Roo, Coahuila, Yucatán, Baja California y Puebla.

REGIÓN	ESTADOS MIEMBRO	CARACTERÍSTICAS	COSTOS	CONCLUSIÓN
CLÚSTER 2	Chiapas, Tabasco	La edad promedio de los involucrados en accidentes es la mayor de todo el país.	Es la región con el segundo costo más bajo y la menor varianza, solo consta de dos estados, ambos con un coste similar.	Es la región con menos accidentes a nivel nacional, los percances son de bajo costo, la prima para personas mayores debe ser mayor a la del resto de país.
CLÚSTER 3	Chihuahua, Guanajuato, Jalisco, Michoacán, Sinaloa, Sonora, Tamaulipas	Su media de accidentes es menor que en Nuevo León, pero mayor que la del resto del país y ocupa el primer lugar en número de muertos y segundo en accidentes con personas con aliento alcohólico involucrados; además tiene edad promedio más baja.	Región con la mayor variabilidad de todas, Michoacán tiene un costo similar o más bajo que los estados de la región 2, Jalisco es el estado más costoso superando a Nuevo León.	La cantidad de accidentes que presenta esta región, podría aumentar la percepción de inseguridad vial y por ende la demanda de seguros, los estados de menor costo y por lo tanto de una mayor rentabilidad esperada son: Michoacán, Sinaloa, Guanajuato y Tamaulipas.

REGIÓN	ESTADOS MIEMBRO	CARACTERÍSTICAS	COSTOS	CONCLUSIÓN
CLÚSTER 4	CDMX, Estado de México	Destaca en la cantidad de autos en circulación y en los demás factores tiene un comportamiento similar al resto de los clúster.	Región más costosa	Mayor cantidad de clientes potenciales pero a un costo superior.
CLÚSTER 5	Nuevo León	Resalta en dos ámbitos, el primero es el número de accidentes y el segundo es en el número de accidentes en el que algún involucrado presento aliento alcohólico.	Es el 4to estado con mayor costo del país.	Alta cantidad de clientes potenciales con costo alto, coberturas que no apliquen en presencia de alcohol podrían disminuir los costos.

Capítulo 6

Pronósticos de Percances Viales

En los capítulos anteriores, se caracterizaron las distintas regiones del país con condiciones similares de accidentes de autos y los costos que estos representan para la aseguradora, para completar el análisis y la estrategia a seguir para la aseguradora sólo resta estimar la cantidad de accidentes que se presentaran.

Contemplando la cantidad esperada de accidentes, los costos que implican y las regiones donde ocurrirán, la aseguradora será capaz de administrar de manera eficiente sus reservas monetarias y su capital humano con el fin de proveer un servicio de calidad y aumentar su rentabilidad.

En el presente capítulo se pronostica la cantidad anual de accidentes que ocurrirán a nivel nacional haciendo uso de métodos de análisis de series de tiempo.

6.1. Descomposición de la serie

Para realizar la estimación se utiliza la información de los percances viales de la aseguradora AXXA, específicamente del año 2015 al 2017, pronosticando los accidentes del año 2018. La base de datos proporcionada por la aseguradora detalla los accidentes de manera individual, por lo que antes de comenzar con la estimación se transforman los datos para obtener el número de accidentes mensuales de los 3 años.

La gráfica 6.1 muestra los resultados.

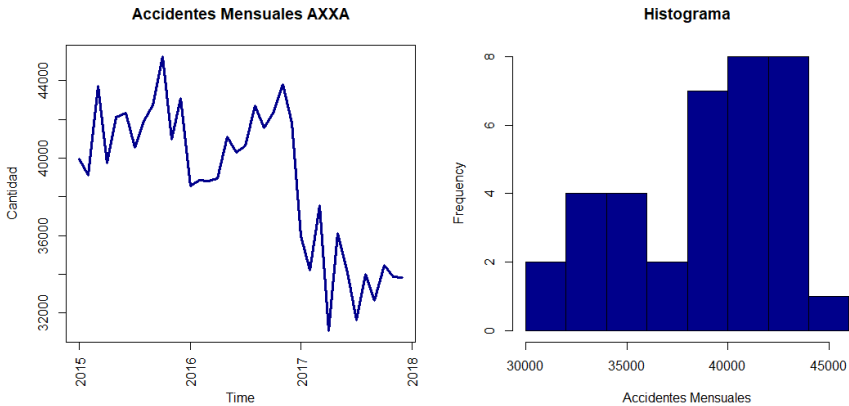


Figura 6.1: Accidentes Mensuales

En los primeros dos años, el comportamiento mensual de los accidentes presentaba una dinámica relativamente estable, el mes de enero se presentan pocos accidentes y una tendencia al alza en el resto del año con bajas en ciertos meses. El año 2017 difiere completamente, ya que, como primer punto el nivel de accidentes es menor a comparación del 2015 y 2016, y como segundo punto, su dinámica parece ser de un tipo de oscilación alrededor de una media.

Dado lo mencionado en el párrafo anterior, considerando los 3 años se visualiza una tendencia a la baja en la cantidad de percances que AXXA debe atender. El histograma de accidentes mensuales, muestra un sesgo a la izquierda posiblemente explicado por la baja cantidad de accidentes en el último año.

Para analizar más a detalle el comportamiento de la serie de datos, se utiliza el método de descomposición de series de tiempo, el cual permite analizar la tendencia, estacionalidad y el factor aleatorio que compone la serie de datos. Se utiliza el modelo aditivo para realizar la descomposición ya que el efecto estacional no para decrecer al momento que lo hace la tendencia.

La figura 6.2 muestra los componentes de la serie.

Como se señaló anteriormente, los 3 años en conjunto presentan una tendencia a la baja, la cual es estimada mediante el uso de medias móviles.

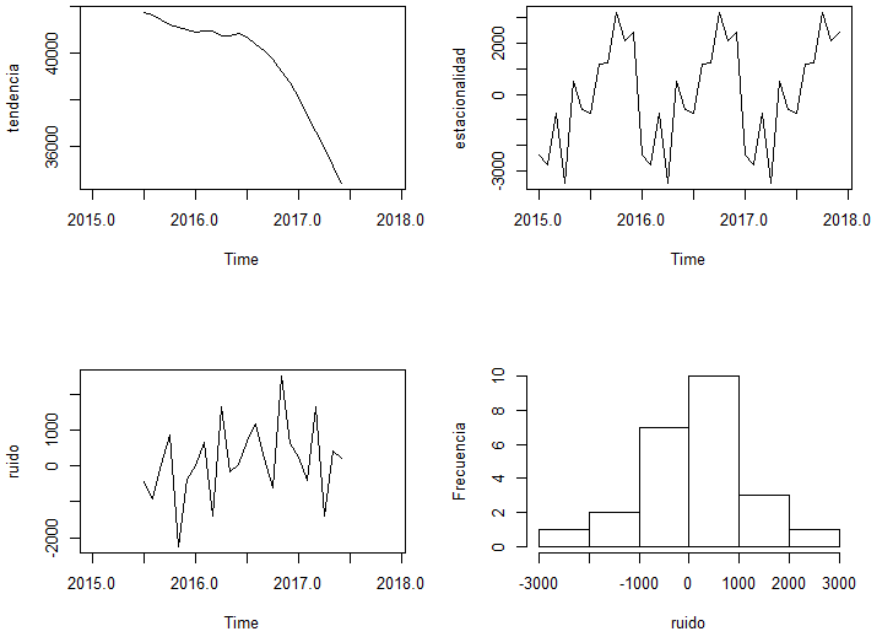


Figura 6.2: Descomposición de la serie

Respecto a la estacionalidad de la serie, conforme al año avanza tienden a presentarse más accidentes viales. En la siguiente sección se muestran los valores numéricos del factor de estacionalidad para hacer más clara la diferencia entre los meses del año.

El histograma del componente aleatorio presenta una forma similar a la distribución normal, aunque no es evidencia suficiente para considerar este componente como ruido blanco.

Es posible observar la relación entre tendencia y estacionalidad de manera más detallada, la gráfica 6.3 muestra dicha relación.

Los boxplot del efecto estacional, muestra que la mediana de los 4 primeros meses del año es similar, dado que abril presenta una observación con un valor muy por debajo del resto su factor de estacionalidad es menor (dado que se

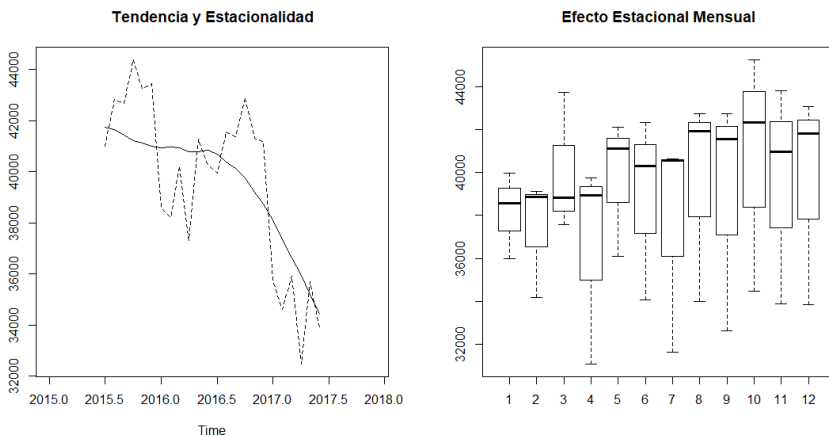


Figura 6.3: Tendencia y Estacionalidad

utiliza la media y no la mediana para su cálculo), caso contrario a marzo, el cual su factor de estacionalidad es mayor debido a la observación cuyo valor presenta la segunda mayor cantidad de percances en el año. Otros casos similares al mes de abril son los meses de julio y septiembre.

6.2. Pronóstico de accidentes

Para realizar el pronóstico de percances viales se utiliza la metodología de Holt-Winters. Dicho método consiste en ajustar una curva a la serie original utilizando una triple suavización mediante medias móviles, utilizando los valores de esa curva divididos entre la serie original se estiman los factores de estacionalidad, los cuales son normalizados para sumar 12 (dado que en este análisis el horizonte temporal son los meses).

El pronóstico se obtiene al predecir la tendencia obtenida de los MCO para un periodo adelante y agregando el factor de estacionalidad del mes correspondiente al periodo. Al siguiente periodo se estima nuevamente una regresión lineal por el método de MCO considerando el valor predicho del periodo anterior y se agrega el factor de estacionalidad, y así sucesivamente hasta obtener el pronóstico de los 12 meses que conforman el año 2018.

Las gráficas 6.4 y 6.5 muestran los resultados de la metodología Holt-Winters.

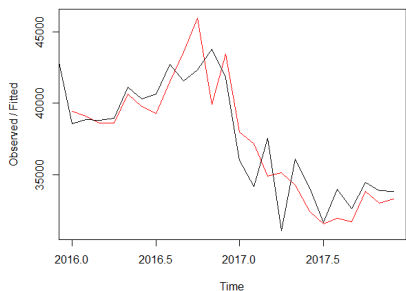


Figura 6.4: Ajuste Holt-Winters

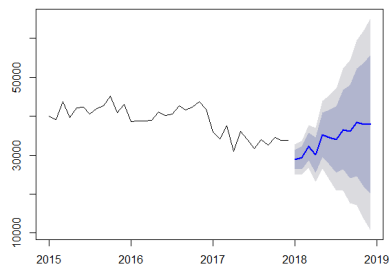


Figura 6.5: Pronóstico

El pronóstico obtenido mantiene coherencia con lo observado en la descomposición de la serie, un aumento paulatino de la cantidad de accidentes conforme aumenta el año, aunque el 2018 no llegara a los niveles de los años 2015 y 2016, mantendrá cantidades similares a las observadas en 2017.

Es importante señalar que, en el gráfico del pronóstico, debido a las escalas del eje y, la tendencia a la baja de los accidentes no es muy notoria como en las figuras anteriormente mostradas, lo que nos indica que, aunque hubo una disminución en percances, no fueron drásticos podría parecer.

Las tablas 6.6 y 6.7 muestran el pronóstico numérico, los intervalos de confianza y los factores de estacionalidad.

Mes	Pronostico	Inferior	Superior
1	28,874	25,029	32,719
2	29,308	24,913	33,703
3	32,235	26,803	37,667
4	30,060	23,160	36,961
5	35,186	26,472	43,900
6	34,594	23,792	45,396
7	33,994	20,872	47,115
8	36,567	20,926	52,208
9	36,081	17,741	54,421
10	38,379	17,175	59,582
11	37,856	13,636	62,076
12	37,923	10,544	65,303

Figura 6.6: Intervalos de Confianza

Mes	Factor
s1	- 3,381
s2	- 3,300
s3	- 726
s4	- 3,254
s5	1,518
s6	573
s7	- 381
s8	1,839
s9	1,000
s10	2,944
s11	2,068
s12	1,782

Figura 6.7: Factores de Estacionalidad

Los intervalos son calculados para un nivel de confianza del 95%. Los factores se presentan como la desviación que tiene cada mes, en cuestión de cantidad de percances, a la tendencia actual que se presenta. Como observamos en la descomposición de la serie, abril es el mes con menor cantidad de accidentes mientras que octubre y noviembre son los meses con mayor cantidad.

La gráfica 6.8 muestra los residuales del modelo de pronóstico.

El sesgo presentado por el histograma se debe a los pronósticos realizados para el segundo semestre del año 2016, donde el modelo no se ajustó de buena manera.

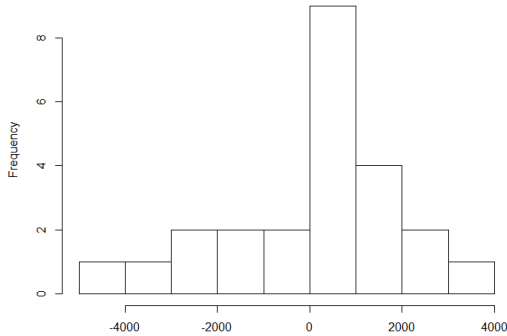


Figura 6.8: Residuales

6.3. Pronósticos por Cluster

En esta sección se realiza la distribución de los accidentes a nivel nacional por cada uno de los cluster estimados en el capítulo 2. La distribución se efectúa bajo el supuesto que la proporción de accidentes de cada región observada en el año 2017, se mantiene constante para el año siguiente.

Debido a lo anterior, el comportamiento de accidentes en las regiones no difiere en función de la dinámica solo en la magnitud de los percances.

La gráfica 6.9 muestra los resultados.

El cluster 1,3 y 4 son los que presentaran mayor cantidad de accidentes, siendo el último el de mayor cantidad. Respecto el cluster 5, presenta alrededor de 3 veces menos accidentes que el cluster 4, y la región 2 casi 9 veces menos.

Retomando lo observado en el capítulo 2, el comportamiento de las regiones, respecto la cantidad de accidentes reportados, difiere a los atendidos por la aseguradora. El caso más significativo es el cluster 5 (Nuevo León), el cual, según datos de la INEGI reporta la mayor cantidad de accidentes, sin embargo; para la aseguradora representa la segunda región con menor cantidad de percances.

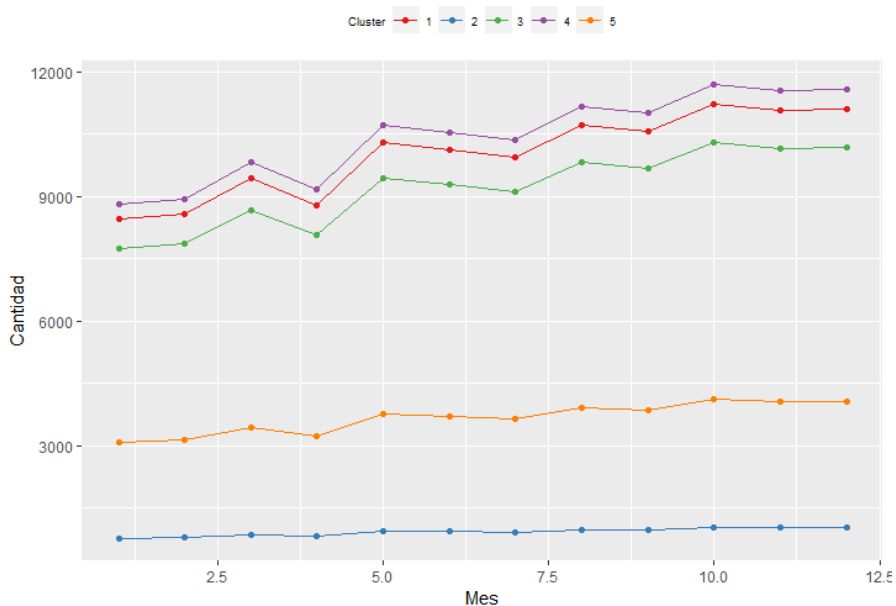


Figura 6.9: Pronósticos por Cluster

Posiblemente, lo anterior se deba a la baja cantidad de asegurados que tiene AXXA en esa región. Otro caso importante es el cluster 1 (Aguascalientes, Colima, Querétaro, etc.) el cual, a nivel nacional es el segundo más bajo, y para la aseguradora es el segundo con mayor cantidad. El efecto es el inverso del cluster anterior, posiblemente la cantidad de asegurados sea muy grande en esta región.

Las demás regiones tienen un comportamiento similar a lo reportado por la INEGI.

6.4. Accidentes Esperados vs Costos

En esta última sección, se realiza la relación entre el pronóstico de accidentes por región y los costos que implican para la aseguradora, con el fin de establecer una estrategia clara de administración del capital humano y los recursos financieros.

A continuación, se presenta el análisis por cada una de las regiones.

Región 1. En accidentes reportados por la INEGI, esta región resalta por tener los niveles más bajos, caso contrario a lo reportado por la aseguradora, ya que esta región es la segunda con mayor cantidad de accidentes esperados, sin embargo; si índice de costos es muy bajo y considerando que es la región con mayor cantidad de Estados, hace de la región 1 un nicho importante de mercado.

La baja cantidad de accidentes totales, los bajos costos, la cantidad de Estados que lo componen hacen de esta región una de las principales que debe considerar la aseguradora para extender su participación de mercado.

La participación en esta región implicaría una baja reserva monetaria y sin incurrir a gastos de logística, solo se debe prever un punto importante, dado que en cantidad de autos en circulación ocupa el penúltimo lugar, la falta de demanda o la competencia pueden afectar la expansión del negocio de la aseguradora.

Región 2. Referente a accidentes, tanto los reportados por la INEGI y los reportados por la aseguradora, esta región presenta los niveles más bajos, al igual que la cantidad de autos en circulación. Representa el segundo costo más bajo para la aseguradora, por lo que en características debe ser una región rentable para AXXA.

Se deben considerar los inconvenientes que presenta, el más significativo es el tamaño de la región, solo es compuesta por dos estados (Chiapas y Tabasco) y ocupa el último lugar en cantidad de autos en circulación, por lo que el nivel de demanda de seguros en esta región es un punto crítico para su posible rentabilidad. Una estrategia de difusión de la importancia de los seguros de automóviles para fomentar el crecimiento de la demanda podría hacer, en el mediano plazo, de esta región una de las más rentables para la aseguradora.

Región 3. Sus características de percances viales la convierten en una región de riesgo, presentan un alto nivel de accidentes reportados por la INEGI y por la aseguradora, además tienen el mayor índice de muertes por accidentes; considerando que no resaltan en cantidad de autos en circulación, la región 3 es la que presenta mayor cantidad de siniestros per cápita.

Es importante mencionar, que los costos para la aseguradora en esta región son ligeramente más altos que en la región 2 por lo que, a pesar de sus características, no representan un mercado inviable para AXXA. Las condicio-

nes que presentan los estados pertenecientes a esta región posiblemente estén más conscientes de la necesidad de la contratación de un seguro, por lo que la demanda debe ser alta y los costos para la aseguradora relativamente bajos.

Aunque el índice de costos de esta región no sea alto, los gastos por reservas y capital humano para hacer frente a los continuos siniestros que se presentan podrían disminuir la rentabilidad esperada de la región.

Región 4. Está compuesta por CDMX y el Estado de México y ocupa el primer lugar en los principales rubros, cantidad de accidentes esperados, costo para la aseguradora y autos en circulación. Existe una gran demanda por seguros pero probablemente mucha competencia, se debe tener una reserva importante de recursos financieros y un uso intensivo del capital humano.

A pesar de los grandes costos, en la región 4 siempre existirá demanda por los servicios de la aseguradora, por lo que es importante tener presencia en la región.

Región 5. Se compone únicamente del Estado de Nuevo León, comparando la cantidad de accidentes que reporta la INEGI con los de la aseguradora, se puede deducir que AXXA no tiene muchos asegurados en ese Estado. Es la región con el segundo costo más alto para la aseguradora. Representa una región de muchos accidentes por lo que aumentar el número de asegurados podría aumentar el riesgo y por ende las reservas monetarias necesarias para cubrir los siniestros.

Capítulo 7

Conclusiones Generales

1. La creación de agrupaciones de estados con base a las características de percances viales permite identificar similitudes entre los mismos, y más importante crear estrategias comerciales y financieras para maximizar la rentabilidad. La ventaja de la creación de grupos mediante variables de accidentes es que la locación demográfica no es una limitante. Se deben agregar otras variables de interés para aseguradora para verificar si tienen un impacto en la conformación de los cluster.
2. Los costos son una parte fundamental para la creación de una estrategia financiera, la incorporación de datos monetarios a la estimación del índice de costos haría mucha más efectivo su uso. El índice se puede tratar de manera individual para cada Estado, para identificar dentro de las regiones cuales son los Estados de mayor costo para la optimización de procesos que permitan la minimización del gasto.
3. Prever la cantidad de percances que existirán en el futuro es algo necesario para el funcionamiento eficaz de la aseguradora, una forma de optimizar los pronósticos es hacerlo de manera individual para cada una de las regiones, de esta manera tendrán su propia dinámica y se podrá hacer una administración de los recursos más eficiente.
4. En términos generales, la interacción entre la identificación de regiones, creación del índice de costos y el pronóstico de percances viales da un panorama detallado de la situación a la que se enfrenta la aseguradora en el mercado, y cuáles pueden ser sus estrategias para la consecución de objetivos monetarios, de servicio y de control interno.

Capítulo 8

Bibliografía

1. García, Jesús & Molina, José M. (2018). *Ciencia de Datos. Técnicas analíticas y aprendizaje estadístico*. Alfaomega Grupo Editor.
2. Marqués Asensio, Felicidad (2015). *Técnicas de investigación de mercados*. Alfaomega Grupo Editor.
3. Aldas Manzano, Joaquín & Jimenez, Ezequiel Uriel (2017) *Análisis Multivariante aplicado con R*. Paraninfo
4. Consejo Nacional de Población (2005) *Metodología de estimación del índice de marginación*.