

Costes esperados por daños corporales en seguro de automóviles e influencia en reservas

Grupo 3

Víctor Alonso Lara
David López Avakian
Sergio Obando Henao
Víctor Manuel Pérez
Miquel Trullols Salat

30 de Octubre de 2025

Índice

1. Objetivo	2
2. Estado de la cuestión: Revisión de la literatura existente	2
2.1. Número de siniestros	2
2.2. Cantidad del siniestro	3
3. Análisis metodológico escogido	5
3.1. Análisis descriptivo univariado y bivariado de la base de datos	5
3.1.1. Análisis descriptivo univariado	5
3.1.2. Análisis descriptivo bivariado	7
3.2. Modelización seleccionada y objetivos a alcanzar	7
4. Resultados	8
5. Informe ejecutivo	8
6. Anexos	8
Referencias	11

1. Objetivo

Este estudio tiene como objetivo aplicar los conocimientos en modelos estadísticos y cuantificación de riesgos al análisis de un caso práctico en el ámbito del seguro de automóviles. En particular, se centra en el estudio de los costes esperados por daños corporales y su influencia en el cálculo de reservas técnicas. Se trata de un tema de gran relevancia, especialmente en el mercado español, donde la compensación por daños corporales representa más del 60 % del coste total en el seguro de responsabilidad civil del automóvil [Santolino Prieto, 2011].

Para su desarrollo, el trabajo se estructura en dos fases:

- Revisión de la literatura existente, con especial atención a los modelos utilizados en la estimación de costes y reservas.
- Modelización del número y la cuantía de los siniestros, mediante técnicas estadísticas aplicadas a datos reales.

Antes de abordar el análisis, se presentarán dos conceptos fundamentales que permiten contextualizar el problema y establecer las bases teóricas del estudio:

- En España, el **seguro obligatorio de automóviles** está regulado por el Real Decreto Legislativo 8/2004. La ley establece que todo propietario de un vehículo a motor con estacionamiento habitual en España debe contratar y mantener un seguro que cubra la responsabilidad civil por los daños causados a personas o bienes durante la circulación del vehículo [Estado, 2004].

- Los **daños corporales** son las lesiones físicas o psíquicas sufridas por una persona en un accidente de circulación, incluyendo lesiones temporales, secuelas permanentes y fallecimiento. Su valoración e indemnización se regulan en la Ley 35/2015 mediante el Baremo oficial [Jefatura Estado, 2015].

2. Estado de la cuestión: Revisión de la literatura existente

2.1. Número de siniestros

La frecuencia de siniestros es un elemento fundamental en la estimación de los costes esperados por daños corporales en el seguro del automóvil, ya que una predicción más precisa del número de reclamaciones permite ajustar con mayor exactitud las reservas técnicas y reducir la incertidumbre asociada a los siniestros futuros. En los últimos años, distintos estudios han buscado mejorar la modelización de esta variable mediante enfoques capaces de reflejar la complejidad y las dependencias que caracterizan los datos de siniestralidad.

Los siniestros con daños corporales constituyen un subconjunto específico dentro del conjunto total de siniestros en el seguro de automóviles, dado que no todos los siniestros implican necesariamente lesiones físicas.

En este contexto, Álvarez Jareño y Muñiz Rodríguez analizan la idoneidad de las distribuciones clásicas para modelizar el número de siniestros en carteras de seguros de responsabilidad civil de automóviles. A partir del estudio de 15 carteras, los autores identifican diversas anomalías muestrales recurrentes que cuestionan la validez de la distribución de Poisson como modelo base. Entre estas anomalías destacan el contagio, la sobre-dispersión (varianza superior a la media), el inflado de ceros (frecuencia excesiva de asegurados sin siniestros), el desinflado de unos (subestimación de asegurados con un único siniestro) y la presencia de colas más pesadas (subestimación de conductores con múltiples siniestros). Estas irregularidades evidencian que el supuesto de independencia entre eventos y la igualdad entre media y

varianza del modelo de Poisson no se cumplen en la práctica [Álvarez Jareño et al., 2010].

Para abordar estas limitaciones, los autores proponen la reparametrización de distribuciones alternativas que ofrecen un mejor ajuste a los datos observados: la distribución binomial negativa, la distribución Polya-Aeppli, la distribución Poisson Inversa Gaussiana y la distribución Poisson Pascal Generalizada. Estas distribuciones permiten capturar la presencia de colas pesadas, proporcionando así una base más robusta para la modelización de la frecuencia siniestral.

Otros autores como Pechon, Trufin y Denuit (2018) analizan la frecuencia de siniestros en el seguro obligatorio de responsabilidad civil automotriz tomando al hogar como unidad de riesgo. A diferencia de los modelos que tratan cada póliza de manera independiente, los autores incorporan efectos aleatorios correlacionados a través de mezclas Poisson–LogNormal y Poisson–Gamma, con el propósito de capturar la dependencia entre los miembros de un mismo hogar y la heterogeneidad no observada. Los resultados muestran una correlación significativa entre las siniestralidades de los cónyuges, cercana al 40 %, lo que confirma la existencia de una propensión común al riesgo. Este enfoque permite afinar las estimaciones de frecuencia y, en consecuencia, mejorar la valoración de los daños corporales y la suficiencia de las provisiones técnicas [Pechon et al., 2018].

Otro enfoque a tener en cuenta es el de [Tzougas and di Cerchiara, 2023]. En su artículo, han desarrollado una clase de modelos de regresión Poisson bivariados mixtos con dispersión variable, orientados a modelizar de forma conjunta la frecuencia de reclamaciones por daños corporales y la frecuencia de reclamaciones por daños materiales en el seguro de responsabilidad civil de automóviles. Estos modelos incorporan distribuciones de mezcla para capturar la variabilidad no explicada y permiten analizar simultáneamente las dos variables correlacionadas. Además, son capaces de reflejar tanto la sobre dispersión como la correlación positiva entre ambas frecuencias, lo que representa un avance significativo en la modelización multivariada del riesgo.

La literatura revisada evidencia que comprender la frecuencia de siniestros no solo mejora la precisión en la estimación de los daños corporales, sino que también constituye un componente esencial en la gestión del riesgo y en la sostenibilidad del sistema asegurador.

2.2. Cuantía del siniestro

Los siniestros con daños corporales en el seguro de automóviles se caracterizan por una alta variabilidad en sus costes. En España, durante 2005, la mayoría de estos siniestros costaron menos de 1.500 €, pero un 0,5 % superaron los 300.000 €, y algunos casos graves, como lesiones tetrapléjicas, pueden superar el millón de euros [Santolino and Ayuso Gutiérrez, 2007].

Como primera aproximación, se realiza un análisis descriptivo de los factores que determinan el coste por daños corporales. Por ejemplo, Marter y Weisberg (1991) clasifican los siniestros de tráfico en cuatro categorías según el tipo de lesión sufrida por la víctima —esguince, fractura, contusión y herido grave— y, para cada una, comparan elementos como el coste médico total, el coste sin hospitalización, el proveedor de asistencia, la frecuencia de visitas y el período de curación [Santolino Prieto, 2011]. Santolino también subraya la relevancia del análisis descriptivo como punto de partida en su estudio sobre indemnizaciones por daños corporales en seguros de auto fijadas judicialmente en Cataluña y Aragón durante el período 2001-2003 [Santolino Prieto, 2011].

A partir del análisis descriptivo, la literatura propone modelos para estimar el coste de los siniestros. Weisberg y Derrig [Santolino Prieto, 2011] plantean el uso del modelo Tobit, adecuado para datos censurados, donde la indemnización no puede ser inferior a cero ni superar los límites legales o de póliza. La variable dependiente es la indemnización (continua y censurada), mientras que las explicativas incluyen factores dicotómicos —contratación de abogado, lesión grave, fractura, indicios de exageración— y cuantitativos —porcentaje de culpa, coste médico total, semanas de incapacidad—. Este enfoque permite estimar el impacto de factores médicos y legales sobre el logaritmo de la indemnización esperada, ajustando por censura. Los resultados muestran que la contratación de un abogado, la clasificación de la lesión como grave y la presencia de fracturas incrementan la indemnización, mientras que los indicios de exageración la reducen.

El modelo logit ordenado es una herramienta estadística adecuada para analizar variables categóricas jerárquicas, especialmente cuando las categorías tienen un orden natural. En el contexto de siniestros, este modelo permite clasificar la severidad de los eventos en distintos niveles como leve, moderado y grave, y evaluar cómo diferentes factores influyen en la probabilidad de que un siniestro pertenezca a una categoría de mayor severidad. El objetivo principal de este modelo es identificar los factores que incrementan la probabilidad de que un siniestro se clasifique en niveles superiores de pérdida, lo que resulta fundamental para la gestión del riesgo. Para abordar limitaciones del modelo logit ordenado clásico y capturar mejor la complejidad de los datos, se pueden considerar varias extensiones: ordenado mixto, ordenado heterocedástico y multinomiales [Santolino Prieto, 2011].

Santolino propone un modelo econométrico log-lineal para explicar el logaritmo de la indemnización total, incorporando variables como edad, tipo de lesión, tipo de vehículo y sexo del lesionado. Un hallazgo relevante es que ni el tipo de vehículo ni la edad del conductor resultan significativos al 10 % de nivel de confianza. Además, se observa que las mujeres reciben indemnizaciones mayores que los hombres y que, cuando el perito necesita más de una visita a la víctima, la cuantía indemnizatoria tiende a incrementarse [Santolino Prieto, 2011].

La predicción del coste de indemnización es un aspecto crítico para las compañías aseguradoras, ya que determina la capacidad de la entidad para cumplir con sus obligaciones futuras. Las aseguradoras deben disponer de reservas suficientes que garanticen la estabilidad financiera y la solvencia del ramo. Este desafío se intensifica en los siniestros corporales cuya indemnización se reclama por vía judicial, dado que la resolución suele demorarse durante meses o incluso años. En consecuencia, estos expedientes permanecen abiertos en la contabilidad de la compañía, lo que obliga a realizar provisiones adecuadas para cubrir el coste esperado. [Santolino and Ayuso Gutiérrez, 2007].

Por último, se recomienda a las aseguradoras prestar especial atención a los siniestros que superan el percentil 90-95 %, ya que representan casos atípicos con costes significativamente elevados. Estos expedientes requieren una evaluación más exhaustiva para verificar la consistencia de los gastos médicos reclamados y detectar posibles exageraciones o prácticas fraudulentas. Un análisis detallado en esta franja no solo contribuye a reducir el riesgo de sobre indemnización, sino que también permite optimizar la asignación de reservas [Weisberg and Derrig,].

3. Análisis metodológico escogido

El objetivo inicial es comprender la estructura del conjunto de datos y familiarizarnos con las variables disponibles, ya que esto constituye la base para cualquier análisis exploratorio. En el Cuadro 1 se presenta una descripción resumida de las variables correspondientes a una cartera de una aseguradora en Francia, que incluyen características del vehículo, del conductor, de la póliza y los montos asociados a los reclamos, entre otros.

Cuadro 1: Diccionario de Variables

Variable	Tipo ⁽¹⁾	Descripción
IDpol	2	Número de póliza.
ClaimNb	2	Número de siniestros.
Exposure	3	Tiempo de vigencia y exposición al riesgo, en años.
Power	1	Potencia del coche (en orden ascendente, de d a o).
CarAge	2	Antigüedad del vehículo, en años.
DriverAge	2	Edad del conductor, en años.
VehBrand	1	Marca del vehículo.
VehGas	1	Tipo de combustible: Diesel o Regular.
Density	2	Número de habitantes por km ² en la ciudad del conductor.
Region	1	Región de la póliza en Francia.
ClaimAmount	3	Costo total del reclamo.
InjuryAmount	3	Costo de compensación por lesiones corporales.
PropertyAmount	3	Costo por daños materiales.

⁽¹⁾ 1 = variable categórica, 2 = variable discreta, 3 = variable continua.

3.1. Análisis descriptivo univariado y bivariado de la base de datos

3.1.1. Análisis descriptivo univariado

Se realiza un análisis descriptivo de las variables numéricas considerando medidas de tendencia central, dispersión y forma ([Anexos - Cuadro 2](#)). Adicionalmente, para las variables Exposure, CarAge, DriverAge y Density se aplica una técnica de segmentación mediante el método de k-means ([Anexos - Figura 1](#)). Esta metodología permite clasificar los datos en conjuntos homogéneos, minimizando la variabilidad interna dentro de cada grupo y maximizando la diferencia entre ellos. Para determinar el número óptimo de clusters se utilizó el método del codo. Concluido este análisis, se procede al estudio de las variables categóricas, examinando su distribución de frecuencias y representaciones gráficas para identificar patrones relevantes. Los aspectos más significativos se detallan a continuación:

- **Número de siniestros (ClaimNb):** Presenta una media y mediana muy bajas (0.04 y 0), consecuencia de que el 96.1 % de las pólizas analizadas no registraron siniestros. Esto se ve reflejado en su alta asimetría (5.78), indicando una distribución con cola larga hacia la derecha. La curtosis, extremadamente alta (38.79), confirma una distribución leptocúrtica, muy concentrada alrededor de la media y con mayor probabilidad de valores extremos. La desviación estándar (0.22), aunque pequeña en términos absolutos, es elevada respecto a la media (0.04), lo que indica una gran variabilidad relativa: la mayoría de los valores son cero, pero existen pocos casos con valores altos que generan dispersión.
- **Tiempo de vigencia y exposición al riesgo (Exposure):** La vigencia promedio de una póliza es de aproximadamente medio año (media 0.56, mediana 0.54), con mínima va-

riabilidad (desviación estándar 0.37), lo que significa que la mayoría de las pólizas tienen duraciones similares. El coeficiente de asimetría (-0.05) indica una distribución prácticamente simétrica, semejante a una campana de Gauss. La curtosis negativa (-1.57) sugiere una distribución platicúrtica, más plana que la normal, con menor concentración en torno a la media y colas ligeras.

- **Antigüedad del vehículo (CarAge):** La edad promedio del vehículo es de 7.53 años, con una alta dispersión (desviación estándar 5.76), lo que significa que existen diferencias importantes entre vehículos nuevos y antiguos. Esta variabilidad se explica porque, aunque el 48.7 % de los autos tienen menos de 6 años, hay casos extremos (máximo 100 años). La asimetría positiva (1.21) indica predominio de autos relativamente nuevos, pero con presencia de valores altos. La curtosis (8.3) confirma una distribución leptocúrtica, concentrada alrededor de la media y con mayor probabilidad de valores extremos.
- **Edad del conductor (DriverAge):** La edad promedio del conductor es 45.32 años, con una mediana cercana (44), lo que indica una distribución equilibrada. El rango es amplio (de 18 a 99 años), y la desviación estándar (14.33) refleja una dispersión moderada: los valores están relativamente separados de la media, mostrando diversidad entre conductores jóvenes y mayores, pero sin diferencias extremas que distorsionen la distribución. A diferencia de otras variables, como la antigüedad del vehículo, no se observan valores atípicos desproporcionados. La asimetría (0.46) es baja, lo que sugiere un ligero sesgo hacia edades mayores, mientras que la curtosis (-0.3) indica una forma cercana a la normal, sin colas pronunciadas.
- **Densidad poblacional por km² en la ciudad del conductor (Density):** Presenta una media muy superior a la mediana (1987.33 vs 287), lo que indica una distribución fuertemente sesgada hacia valores altos. La gran variabilidad (desviación estándar 4779.6) y la asimetría positiva (4.13) reflejan diferencias significativas entre zonas de baja y alta densidad. La curtosis (17.72) confirma una distribución leptocúrtica, con concentración en valores bajos y presencia de colas largas hacia la derecha.
- **Costo total del reclamo (ClaimAmount):** presenta una distribución altamente desbalanceada y con valores extremos. El valor promedio es de 832,57 euros, mientras que la mediana es cero, lo que indica que más del 50 % de los registros no tienen reclamo asociado. El mínimo también es cero, confirmando la existencia de pólizas sin siniestro, mientras que el máximo alcanza los 20.368.330 euros, evidenciando la presencia de valores extremadamente altos. La desviación estándar, de 41.847 euros, refleja una gran dispersión en los datos. Además, la asimetría es de 375,9 y la curtosis de 166.362,6, lo que indica una distribución extremadamente sesgada hacia la derecha, con una concentración muy alta en valores bajos y la presencia de outliers. Para complementar este análisis, se elaboraron dos histogramas ([Anexos - Figura 2](#)), considerando únicamente las pólizas con al menos un siniestro, dado que el 96,1 % de las pólizas no registraron ningún reclamo y, por tanto, un histograma completo concentraría la masa de probabilidad en cero, ocultando el comportamiento del resto:
 - El primer histograma corresponde al 99 % de las pólizas con siniestros y revela que la gran mayoría de los reclamos son de bajo coste: el 96,9 % no supera los 50.000 euros, mientras que solo el 2,4 % se encuentra entre 50.000 y 100.000 euros, y menos del 1 % supera los 100.000 euros. El coste promedio dentro de este grupo es de 13.560 euros, lo que confirma que los siniestros de alto importe son excepcionales. El gráfico refuerza esta conclusión, mostrando una barra dominante en el primer intervalo (0–50 miles de euros) y frecuencias marginales en los tramos superiores, lo que evidencia una concentración extrema en valores bajos.

- El histograma del 1 % de las pólizas con siniestros muestra que, aunque este grupo corresponde a los casos más costosos, la concentración sigue siendo muy marcada en el primer intervalo: el 98,8 % de estos reclamos extremos se sitúan entre 161.500 y 5.000.000 euros, con un coste promedio en este tramo de 577.685 euros. Esto indica que, incluso dentro de la cola, la mayoría de los siniestros se ubican en rangos elevados pero moderados en comparación con los máximos observados, aunque siguen siendo significativamente superiores al promedio general. Los intervalos superiores, que incluyen reclamos entre 10 y 15 millones y entre 20 y 25 millones de euros, apenas representan el 0,6 % cada uno, lo que confirma que los valores más altos son casos excepcionales con un impacto potencial muy significativo en la compañía aseguradora.
- **Costo de compensación por daños corporales (InjuryAmount):** Presenta una alta concentración en valores nulos y una cola prolongada con pocos casos de gran magnitud. El promedio es 615,9 euros, mientras que la mediana y el mínimo son cero; el máximo alcanza 19.972.821 euros, reflejando reclamos excepcionales. La dispersión es elevada (desviación estándar: 40.867 euros) y los indicadores de forma son extremos: asimetría 375,25 y curtosis 165.244,5, lo que evidencia un fuerte sesgo a la derecha. Para evitar que los ceros oculten el patrón real, se elaboraron dos histogramas considerando solo pólizas con siniestros, confirmando la concentración en valores bajos y la existencia de una cola con pocos casos de gran coste. Ambos gráficos se incluyen en los anexos. ([Anexos - Figura 2](#)).
- **Costo por daños materiales (PropertyAmount):** La distribución muestra una fuerte concentración en valores nulos y casos positivos menos extremos que en *InjuryAmount*. El promedio es 216,7 euros, la mediana y el mínimo son cero, y el máximo llega a 575.508,8 euros, lo que indica reclamos significativos pero no desproporcionados. La dispersión es alta (desviación estándar: 1.554 euros) y la asimetría (129,54) y curtosis (45.682,36) confirman un sesgo marcado hacia la derecha. Para evitar que los ceros oculten el patrón real, se elaboraron dos histogramas: el primero concentra el 99 % de los casos por debajo de 12.000 euros, destacando los rangos de 8.000-10.000 euros (26,3 %), 4.000-6.000 euros (23,7 %) y 2.000-4.000 euros (22,3 %). El segundo muestra el 1 % restante, casi todo entre 10.400 y 100.000 euros, con algunos reclamos aislados de mayor coste. Aunque existen siniestros elevados, la mayoría se sitúa en valores moderados. Ambos gráficos se incluyen en los anexos. ([Anexos - Figura 2](#)).

Tras el análisis detallado de las variables numéricas, se continúa con el estudio de las variables categóricas, examinando su distribución de frecuencias y representación gráfica para identificar patrones relevantes:

- **Potencia del coche (Power):**
- **Marca del vehículo (VehBrand):**
- **Tipo de combustible (VehGas):**
- **Región de la póliza (Region):**

3.1.2. Análisis descriptivo bivariado

3.2. Modelización seleccionada y objetivos a alcanzar

Por desarrollar.

4. Resultados

Por desarrollar.

5. Informe ejecutivo

Por desarrollar.

6. Anexos

Cuadro 2: Análisis descriptivo univariado - variables numéricas

Concepto	Claim Nb	Exposure	CarAge	DriverAge	Density
Media	0,04	0,56	7,53	45,32	1.987,33
Mediana	0	0,54	7	44	287
Mínimo	0	0	0	18	2
Máximo	4	1,99	100	99	27000
Desv. Estándar	0,22	0,37	5,76	14,33	4.779,6
Asimetría	5,78	-0,05	1,21	0,46	4,13
Curtosis	38,79	-1,57	8,3	-0,3	17,72

Cuadro 2: Análisis descriptivo univariado - variables numéricas

Concepto	Claim Amount	Injury Amount	Property Amount
Media	832,57	615,9	216,67
Mediana	0	0	0
Mínimo	0	0	0
Máximo	20.368.330	19.792.821	575.508,8
Desv. Estándar	41.847	40.867,4	1.554,24
Asimetría	375,9	375,25	129,54
Curtosis	166.362,6	165.244,5	45.682,36

Figura 1: Análisis descriptivo univariado - Técnica de segmentación

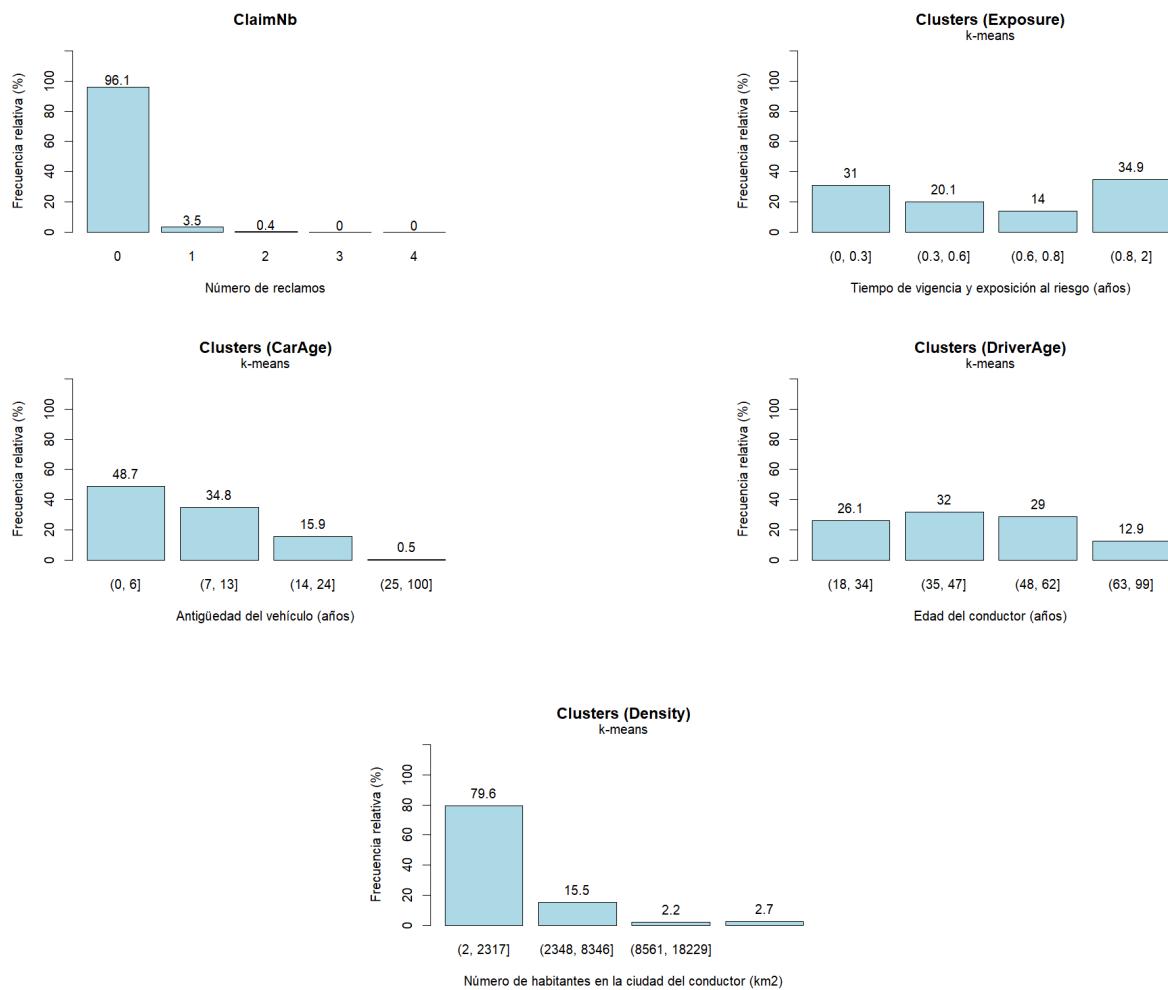
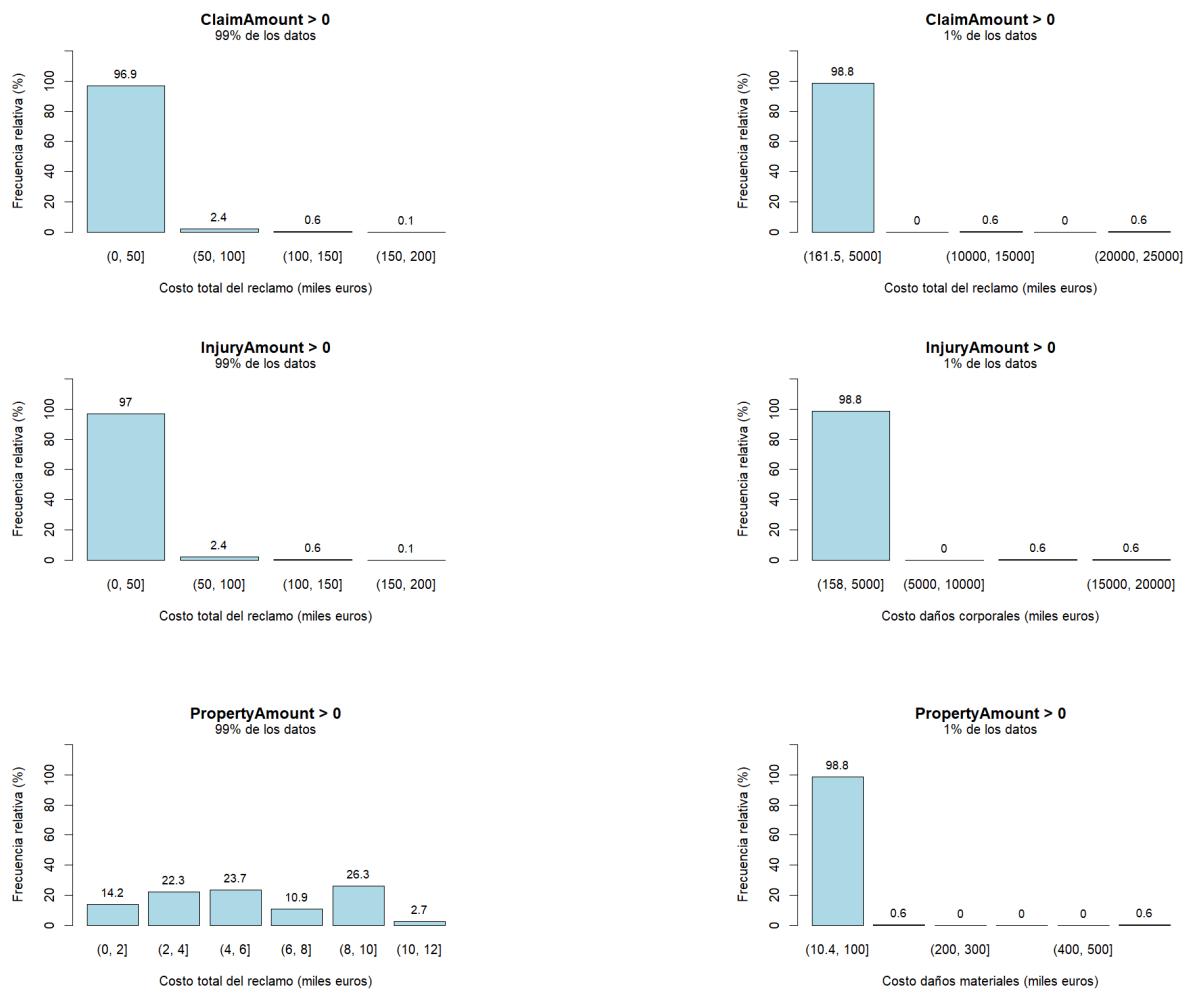


Figura 2: Análisis descriptivo univariado - Histogramas



Referencias

- [Estado, 2004] Estado, J. (2004). Boe-a-2004-18911 real decreto legislativo 8/2004, por el que se aprueba el texto refundido de la ley sobre responsabilidad civil y seguro en la circulación de vehículos a motor.
- [Jefatura Estado, 2015] Jefatura Estado (2015). Ley 35/2015, de 22 de septiembre, de reforma del sistema para la valoración de los daños y perjuicios causados a las personas en accidentes de circulación.
- [Pechon et al., 2018] Pechon, F., Trufin, J., and Denuit, M. (2018). Multivariate modelling of household claim frequencies in motor third party liability insurance. *ASTIN Bulletin: The Journal of the IAA*, 48(3):969–993.
- [Santolino and Ayuso Gutiérrez, 2007] Santolino, M. and Ayuso Gutiérrez, M. (2007). Una revisión metodológica de la valoración actuarial de los siniestros con daños corporales en el seguro del automóvil. *Anales del Instituto de Actuarios Españoles*, (13):143–172.
- [Santolino Prieto, 2011] Santolino Prieto, M. (2011). *Métodos econométricos para la valoración cualitativa y cuantitativa del daño corporal en el seguro del automóvil*. PhD thesis, Universitat de Barcelona. Book Title: Métodos econométricos para la valoración cualitativa y cuantitativa del daño corporal en el seguro del automóvil ISBN: 9788469368862.
- [Tzougas and di Cerchiara, 2023] Tzougas, G. and di Cerchiara, A. P. (2023). Bivariate Mixed Poisson Regression Models with Varying Dispersion. *North American Actuarial Journal*, 27(2):211–241.
- [Weisberg and Derrig,] Weisberg, H. I. and Derrig, R. A. Quantitative methods for detecting fraudulent automobile bodily injury claims.
- [Álvarez Jareño et al., 2010] Álvarez Jareño, J. A., Muñiz Rodríguez, P., Álvarez Jareño, J. A., and Muñiz Rodríguez, P. (2010). Reparametrización de las principales distribuciones de probabilidad en el estudio del número de siniestros debido: determinación del índice de dispersión. *Anales del Instituto de Actuarios Españoles*, 16:1–24.