

TRABAJO FINAL DE MÁSTER

Título: El seguro basado en el uso: modelización del número de siniestros con datos simulados

Autoría: Manar Bouanani Boubakra

Tutoría: Ana María Pérez-Marín

Curso académico: 2020-2021

Facultad de Economía y Empresa

Universidad de Barcelona

Trabajo Final de Máster

Máster en Ciencias Actuariales y Financieras

**EL SEGURO BASADO EN EL
USO: MODELIZACIÓN DEL
NÚMERO DE SINIESTROS
CON DATOS SIMULADOS**

Autoría: Manar Bouanani Boubakra

Tutoría: Ana María Pérez-Marín

“El contenido de este documento es de exclusiva responsabilidad del autor, quien declara que no ha incurrido en plagio y que la totalidad de referencias a otros autores han sido expresadas en el texto”.

RESUMEN

En el presente trabajo final de máster se explora mediante una revisión bibliográfica profunda el seguro basado en el uso, con la finalidad de conocer y entender cuál es el efecto de las variables telemáticas en el riesgo siniestral. En este trabajo también se realiza una aplicación práctica, con datos simulados, para ilustrar una modelización del número de siniestros incluyendo variables telemáticas, mediante un modelo de Poisson con *offset* clásico y un modelo de Poisson con *offset* generalizado, en los que se analiza la relación del kilometraje respecto la frecuencia de siniestros, verificando una relación no lineal.

PALABRAS CLAVE

Seguro de automóvil, seguro basado en el uso, factores de riesgo, kilometraje, regresión de Poisson con offset

ABSTRACT

The topic of the master thesis is about usage-based insurance which is explored by a literature review in order to know and understand the effect of telematic variables on accident risk. Furthermore, a practical application is realized with simulated data to illustrate a modelling of number of claims including telematic variables by using a Poisson model with classic offset and a Poisson model with generalized offset, in which the relationship between the mileage and the frequency of claims is analysed to verify its nonlinear relationship.

KEYWORDS

Automobile insurance, usage-based insurance, risk factors, mileage, offset Poisson regression

Índice

1. Introducción.....	1
I. REVISIÓN BIBLIOGRÁFICA	3
2. El seguro basado en el uso	3
2.1. Modalidades de producto	3
2.2. Ventajas.....	5
2.3. Inconvenientes	7
3. Variables de riesgo	8
4. Factores de riesgo	9
4.1. Tradicionales	10
4.1.1. Género.....	10
4.1.2. Edad.....	14
4.1.3. Experiencia del conductor.....	15
4.1.4. Antigüedad del vehículo.....	16
4.1.5. Potencia del motor.....	17
4.1.6. Lugar aparcamiento	17
4.2. Telemáticos	18
4.2.1. Kilometraje	18
4.2.2. Conducción urbana	20
4.2.3. Conducción nocturna	21
4.2.4. Conducción con exceso de velocidad.....	23
II. APLICACIÓN PRÁCTICA	25
5. Modelos lineales generalizados	25
6. Datos	27
6.1. Análisis descriptivo.....	28
7. Estimación y resultados	36
7.1. Modelo de regresión de Poisson.....	36
7.2. Modelo de regresión de Poisson con <i>offset</i> clásico	36
7.3. Modelo de regresión de Poisson con <i>offset</i> generalizado	38
7.4. Comparativa de modelos	39
8. Conclusiones	44
9. Bibliografía	46
10. Anexos	50
10.1. Cuadros resumen artículos	50
10.2. Código aplicación práctica.....	76

Índice de figuras

Figura 1. Resumen de la distribución de las variables.....	27
Figura 2. Distribución asegurados con exceso de velocidad.....	28
Figura 3. Distribución asegurados conducción urbana.....	29
Figura 4. Distribución asegurados antigüedad carnet.....	30
Figura 5. Frecuencia absoluta género del asegurado.....	31
Figura 6. Medidas descriptivas número de siniestros.....	32
Figura 7. Histograma número de siniestros.....	33
Figura 8. Histograma y Boxplot Kilometraje.....	35
Figura 9. Medidas descriptivas kilometraje.....	35
Figura 10. Estimación modelo de regresión de Poisson con <i>offset</i> clásico.....	37
Figura 11. Estimación modelo de regresión de Poisson con <i>offset</i> generalizado.....	38
Figura 12. Valores observados y predichos con el modelo de regresión de Poisson con <i>offset</i> clásico.....	40
Figura 13. Valores observados y predichos con el modelo de regresión de Poisson con <i>offset</i> generalizado.....	41

Índice de tablas

Tabla 1. Clasificación modelos de regresión según la variable dependiente.....	25
Tabla 2. Funciones enlace según modelo.....	26
Tabla 3. Descripción de las variables de la base de datos.....	27
Tabla 4. Frecuencias exceso de velocidad.....	28
Tabla 5. Ocurrencia de siniestro según exceso de velocidad.....	28
Tabla 6. Frecuencias conducción urbana.....	29
Tabla 7. Ocurrencia de siniestro según conducción urbana.....	29
Tabla 8. Frecuencias antigüedad carnet.....	30
Tabla 9. Ocurrencia de siniestro según antigüedad de carnet.....	30
Tabla 10. Frecuencias género.....	31
Tabla 11. Ocurrencia de siniestro según género.....	31
Tabla 12. Análisis descriptivo según el género.....	31
Tabla 13. Frecuencias número de siniestros.....	33
Tabla 14. Número de siniestros según exceso de velocidad.....	33
Tabla 15. Número de siniestros según conducción urbana.....	33
Tabla 16. Número de siniestros según antigüedad de carnet.....	34
Tabla 17. Número de siniestros según género.....	34
Tabla 18. Kilometraje según número de siniestros.....	35
Tabla 19. Comparativa bondad del ajuste entre modelos de Poisson.....	39
Tabla 20. Interpretación de coeficientes significativos modelo de regresión de Poisson con <i>offset</i> generalizado.....	42

1. Introducción

Los aseguradores deben vigilar el ofrecer primas precisas, a fin de evitar la selección adversa. Por consiguiente, el cálculo de las primas mediante los sistemas de tarificación debe resultar en primas equitativas, en base del riesgo que suponga cada asegurado, solventes, es decir, que sean suficientes permitiendo a las compañías aseguradoras una estabilidad y rentabilidad a largo plazo, y solidarias, repartiendo el riesgo y prima total, los asegurados con un riesgo menor pagaran más de lo correspondido, de esta forma a los que incorporan más riesgo se les podrá reducir la prima (Boj et al, 2021).

Para conseguir una equidad actuarial se deben considerar y medir los factores de riesgo observables que tengan un mayor efecto en la siniestralidad y que mejoren su explicación, así como su predicción (principalmente en la ocurrencia de siniestro, número de siniestros o su coste). Es de elevada importancia una correcta selección de los factores de riesgo puesto que se usan para distinguir y clasificar riesgos de los asegurados, con el fin de realizar una correcta tarificación (Boj et al, 2021).

En el seguro de automóvil tradicionalmente los factores de riesgo han sido características, tanto cuantitativas como cualitativas, personales del asegurado (edad, antigüedad de carnet, zona en la que se vive, etc.) o del objeto asegurado (valor del vehículo, potencia del vehículo, tipo de vehículo, cantidad de asientos, antigüedad del vehículo, etc.), si bien estos factores de riesgo tienen relación con la siniestralidad, actualmente existen otros factores con una relación directa con los siniestros, que tienen una causalidad probada en los accidentes, siendo ésta más fuerte que las tradicionales, estos factores son los factores telemáticos, (kilometraje, porcentajes de conducción nocturna, urbana y con exceso de velocidad, aceleración y frenadas bruscas, etc.). Con el desarrollo de las tecnologías actuales enfocadas en los vehículos, es posible medir el comportamiento de conducción y el uso que se hace del vehículo, permitiendo una mayor equidad y mejor precisión de los precios a ofrecer en el sector del seguro de automóviles (Huang & Meng, 2019).

El seguro basado en el uso es un tipo de seguro cada vez más extendido en el sector automovilístico, siendo más comercializado entre perfiles jóvenes, con esquemas de *Pay-As-You-Drive* (PAYD) y *Pay-How-You-Drive* (PHYD), donde la prima se determina en base al tiempo, distancia recorrida (Kilometraje) y el comportamiento del conductor (cuándo y dónde se conduce). Gracias a los dispositivos telemáticos que se instalan en el propio vehículo o mediante una aplicación en el móvil, las aseguradoras pueden recoger información muy relevante que les permite realizar tarifas más acordes al perfil de cada persona. Con esta modalidad de seguro, aquellos conductores que usen menos el vehículo y los que tengan una mejor conducción se van a ver beneficiados pudiendo pagar una prima menor.

Aunque este trabajo se centra en el seguro basado en el uso en los automóviles, este tipo de seguro también se extiende, en menor medida, a otros productos como el seguro de hogar, el de vida *Pay-As-You-Live* (PAYL), así como hay la posibilidad de extenderlo a otros productos, como el seguro de accidentes laborales (Alcañiz et al, 2014).

La **motivación** del trabajo es la de conocer y dar a conocer el seguro basado en el uso (UBI), un seguro que se lleva desarrollando mucho en los últimos años a nivel global, según datos de *PTOLEMUS Consulting Group*, en 2018 se alcanzaron casi 25 millones

de pólizas UBI activas, con 372 programas UBI activos en 58 países. La previsión de su tamaño en el futuro es positiva, en un informe de mercado publicado por *Allied Market Research* se pone de manifiesto que “El tamaño del mercado global de seguros basado en el uso se valoró en 28,7 mil millones de dólares en 2019 y se prevé que alcance los 149,2 mil millones de dólares en 2027”. Si bien es cierto que el UBI lleva mucho más recorrido en países como Estados Unidos, Reino Unido, Italia, China y Francia, en los últimos años se está desarrollando más y se está volviendo más conocido también en España, siendo MAPFRE la empresa pionera con el seguro YCAR, otras empresas del mercado asegurador español han introducido también esta modalidad de seguro como, GENERALI con el seguro Pago como conduzco, VERTI con el seguro CuentaKms y LÍNEA DIRECTA con la medida Registra tus KMs. Sin duda la situación vivida debido al covid-19, en la que la circulación ha sido menor, ha generado un mayor interés por este tipo de seguros tanto por parte de compañías como de conductores, siendo esta una motivación más para la realización del trabajo.

Los **objetivos** principales de este estudio son:

Por una parte, la realización de una revisión bibliográfica profunda, según los estudios elaborados hasta ahora, de la temática del seguro basado en el uso, con el fin de conocer y entender cuál es el efecto de las variables telemáticas en el riesgo asociado a los siniestros. Por otra parte, mediante la realización de una aplicación práctica con datos simulados, se pretende ilustrar como realizar la modelización del número de siniestros cuando hay variables telemáticas, con especial atención a la relación del kilometraje respecto la frecuencia siniestral, con el propósito de determinar si existe una relación lineal o no.

El presente trabajo se **estructura** de la siguiente forma:

Un primer bloque, en el que se realiza la revisión bibliográfica, formado por los apartados 2 al 4, en los que se explica en detalle el seguro basado en el uso, y se exponen los hallazgos de los distintos artículos revisados, respecto los factores de riesgo tanto tradicionales como telemáticos.

Un segundo bloque, en el que se lleva a cabo una aplicación práctica, formado por los apartados 5 al 7, utilizando datos simulados, se muestra primero un análisis descriptivo, seguidamente se realiza la estimación de los dos modelos utilizados, modelo de regresión de Poisson con *offset* clásico y modelo de regresión de Poisson con *offset* generalizado, se exponen los resultados efectuando una comparativa de ambos modelos.

En el apartado 8, una vez finalizada la revisión bibliográfica y la aplicación práctica, se presentan las conclusiones extraídas del trabajo elaborado.

Finalmente, en el apartado 9 se detalla la bibliografía utilizada, y en el apartado 10 se muestran los anexos del trabajo, formados por los cuadros resumen de los artículos académicos, en los que se describen los datos, la metodología y conclusiones de cada uno de ellos, así como se adjuntan los códigos empleados para la aplicación práctica en R.

I. REVISIÓN BIBLIOGRÁFICA

En este primer bloque, se expondrá de forma estructurada la revisión bibliográfica realizada, en los anexos se pueden encontrar los cuadros resumen de los artículos académicos mencionados a lo largo del bloque, agrupados en ocho grupos según su temática.

2. El seguro basado en el uso

El seguro basado en el uso, llamado en inglés *Usage-Based Insurance* (UBI), es un seguro que cada vez tiene más atención por parte de las compañías aseguradoras, así como de los conductores. Mediante la instalación de dispositivos telemáticos (principalmente el sistema GPS) en los vehículos de los asegurados (desde fábrica o una vez comprados) por parte de los particulares o las propias compañías de seguros al ofrecer el seguro, junto el consentimiento del asegurado, se obtienen amplios datos telemáticos de manera individualizada, permitiendo la medición de los patrones de conducción, así como el uso del vehículo (kilometraje) de una forma fiable y detallada, pudiendo tener en cuenta otros factores de riesgo en la tarificación a parte de los factores tradicionales. Con los seguros basados en el uso los asegurados pagan según cuanto conduzcan y como conduzcan, de esta forma aquellos asegurados que no conduzcan con frecuencia y tengan una conducción segura, van a pagar menos por su póliza por su menor exposición al riesgo.

2.1. Modalidades de producto

Pay-As-You-Drive (PAYD)

El seguro basado en el uso más conocido y comercializado es el seguro *Pay-As-You-Drive* (PAYD), este seguro mide la exposición al riesgo en la distancia recorrida, la prima se calcula principalmente en base del kilometraje, siendo la prima más elevada cuando la exposición al riesgo es mayor, por lo tanto, cuando se recorra un mayor número de kilómetros.

Tipologías sistemas de tarificación PAYD

El estudio realizado por William Vickrey en 1968 llamado “*Automobile accidents, tort law, externalities, and insurance: an economist's critique*” (Vickrey, 1968) fue de los primeros estudios en el que se critica el sistema de precios de suma global en el seguro de automóvil y se promueve un sistema de precios en los seguros basado en la distancia.

Los inicios del seguro basado en el kilometraje empiezan con un método llamado ***Pay-at-the-Pump (PATP)***, en el que se pagaba según el consumo de combustible para el vehículo. Habiendo una relación entre el kilometraje y el consumo de combustible, este método presenta varias similitudes con el sistema de precios en base a la distancia recorrida. El método PATP no fue muy extendido, algunas de las críticas que se le hizo son que, el hecho de que se recargue a quien más consume puede llevar a que los conductores busquen conducir vehículos más eficientes, pero no por ello reducirán su

distancia anual recorrida, y que creaba una incertidumbre de los ingresos en las compañías aseguradoras. También se propuso un sistema de precios con el nombre de “*insured tires*” en el que una compañía aseguradora asociada se identificase con los neumáticos de alguna forma, responsabilizándose de indemnizar a las víctimas en el caso de siniestro del vehículo con esos neumáticos midiendo el uso en términos de desgaste de los neumáticos, este sistema también fue criticado puesto que tampoco tenía en cuenta la distancia que se recorría realmente y no era capaz al igual que el PATP de diferenciar entre conductores buenos y los de más riesgo (Boucher et al, 2013; Ayuso et al, 2016a; Ayuso et al, 2016b; Tselentis et al, 2016).

Posteriormente en el año 2011, Todd Litman realiza en el artículo “*Pay-As-You-Drive Insurance Recommendations for Implementation*” una clasificación de los sistemas de tarificación PAYD (Boucher et al, 2013; Alcañiz et al, 2014; Tselentis et al, 2016):

- ***Mileage Rate Factor (MRF)***: este primer sistema tiene en consideración el kilometraje anual de cada asegurado en la tarificación, aquellos con un menor kilometraje pagan una menor prima, la crítica a este sistema es que el kilometraje anual es el declarado por parte del asegurado, este kilometraje esperado no se verifica ni ajusta una vez finalizada la póliza, por lo que en numerosas ocasiones el kilometraje es subestimado.
- ***Per-Mile-Premium (PMP)***: este segundo sistema tiene un precio por unidad de distancia que se recorre (coma la milla), teniendo en cuenta además los factores tradicionalmente usados, la crítica a este sistema es que la medición de la distancia de cada vehículo se efectúa mediante un cuentakilómetros, utilizado para verificar los kilómetros conducidos respecto los que indicó el asegurado en el inicio de la póliza, aunque haya una verificación el cuentakilómetros puede ser manipulado fraudulentamente por los asegurados.
- ***GPS-Based Pricing***: el último sistema de precios surge gracias al desarrollo tecnológico, en la actualidad se puede medir y controlar de forma objetiva el uso del vehículo mediante la instalación de dispositivos GPS, contando siempre con el consentimiento del asegurado es posible fijar una tarifa teniendo en cuenta el kilometraje, las vías y horas de conducción, así como el exceso de velocidad entre otros factores. En la actualidad, el sistema de tarificación del seguro PAYD que más se usa entre las compañías aseguradoras es este. El inconveniente de este sistema es la preocupación de los asegurados sobre la privacidad, así como el marco legal.

Hay que mencionar que a veces el seguro PAYD es utilizado como sinónimo para el UBI ya que se introducen también algunos patrones de conducción (porcentaje de conducción nocturna, urbana y conducción por exceso de velocidad), aunque en su esencia el seguro PAYD tiene en cuenta el kilometraje (principalmente) o el tiempo durante el que se conduce.

Pay-How-You-Drive (PHYD)

Este tipo de seguro también es muy conocido dentro de los seguros UBI, aunque comercializado en menor medida que el PAYD, con el desarrollo de la tecnología ha sido posible tener en cuenta otras variables para el cálculo de la prima, en este seguro se tiene en cuenta el comportamiento del asegurado durante su conducción, midiendo los excesos

de velocidad, los frenados y aceleraciones bruscas, y las curvas pronunciadas principalmente.

Como también pasaba en el PAYD, es habitual, que con el seguro PHYD se recoja otra información, a la ya mencionada, sobre el tipo de vía en la que se conduce (urbana/no urbana) y durante que horas se conduce (diurnas/nocturnas), y también complementarse con el kilometraje.

PHYD respecto PAYD tiene en cuenta más variables que indican el comportamiento de conducción, permitiendo una mayor precisión en las primas al tener una evaluación del riesgo más precisa, teniendo así una mejor estimación del riesgo que comporta cada asegurado. También los asegurados con PHYD pueden ser más conscientes de su comportamiento y como consecuencia mejorar su conducción lo que se traduciría en menos reclamaciones que cubrir por parte de las aseguradoras, por otro lado, la implementación del seguro PAYD es mucho más sencilla al tener menos variables de riesgo, también el periodo necesario de desarrollo y verificación es menor al de PHYD (Alcañiz et al, 2014; Tselentis et al, 2016).

Manage-How-You-Drive (MHYD)

Es de los seguros basados en el uso más novedoso y menos comercializado hasta el momento, combina un sistema de tarificación como el de PHYD, junto alertas y sugerencias en tiempo real, se da un *feedback* más rápido mejorando el servicio al cliente y consiguiendo una mejor protección en el sector, previniendo situaciones de riesgo mediante la información continuada que se ofrece a los asegurados sobre su comportamiento (excesos de velocidad, aceleraciones y frenados bruscos, giros repentinos, recomendaciones al conducir muchas horas, etc.), además también se muestra dicha información al asegurado de manera dinámica, influyendo así al modo de conducción del asegurado. MHYD va un paso más allá en el procesamiento y aplicación de los datos que el PHYD, pero esto implica también una mayor complejidad (Sun et al, 2020; Guillén et al, 2021).

2.2. Ventajas

Las ventajas que se obtiene con los seguros basados en el uso para automóviles son amplias tanto para asegurados como para las compañías aseguradoras, incluso para la sociedad:

Una de las principales ventajas es la mejora en la conducta de conducción, reduciéndose así su exposición al riesgo y reduciendo aquellos comportamientos que tienen asociados una mayor probabilidad de siniestro y pudiéndose obtener una reducción en su prima. Con estas mejoras, por ejemplo, la reducción de la conducción con exceso de velocidad resulta en una reducción de la siniestralidad, mejorando así la seguridad vial. También el seguro es más asequible para los distintos perfiles de conductores al ofrecer primas menores por una conducción más segura y un menor uso del vehículo, dando lugar a una reducción del uso de subsidios cruzados¹, se puede contribuir a una reducción de los vehículos que circulan sin un seguro (Bolderdijk et al, 2011; Ayuso et al, 2014a; Ayuso et al, 2016a; Tselentis et al, 2016; Huang & Meng, 2019).

¹ “El subsidio cruzado es una estrategia que consiste en elevar el precio de un producto o servicio a un grupo de clientes. De esa forma, la empresa genera un excedente que le permite reducir la tarifa del mismo bien para otro segmento económico” (Westreicher)

Estos seguros permiten a las empresas una clasificación del riesgo más efectiva midiendo con una mayor precisión la exposición al riesgo de cada asegurado, obteniendo una mejor equidad actuarial en las primas que las determinadas por segmentaciones tradicionales, sobre todo en el grupo de conductores más jóvenes que no presentan una homogeneidad en su conducción y el riesgo siniestral. La empresa también puede verse beneficiada en su imagen, al ser un seguro más personalizado con una clara orientación al cliente y con ventajas sociales que se detallan a continuación, los potenciales clientes harán una asociación positiva con la compañía aseguradora (Ayuso et al, 2016a; Ayuso et al, 2016b).

El seguro basado en el uso tiene importantes ventajas también para el conjunto de la sociedad, con la incentivación de un menor uso del vehículo y una mejora de la conducta, se contribuye a una reducción de la contaminación y del consumo de energía, así como de la congestión del tráfico y el gasto público para el mantenimiento de las carreteras (Bolderdijk et al, 2011; Ayuso et al, 2014a; Ayuso et al, 2016a; Tselentis et al, 2016; Huang & Meng, 2019).

Mejoras en la conducta de los asegurados

Los distintos registradores de datos en vehículos son una gran herramienta para recoger información y dar un *feedback* a los conductores de sus patrones de conducción registrados, puesto que las personas que son conscientes de que son registradas suelen modificar su comportamiento hacia uno más seguro, también es de utilidad para poder revisar su conducción después de un evento (como un accidente o casi accidentes). Wouters et al. (2000) realizan un análisis, de una muestra de conductores en la que su comportamiento de conducción quedó registrado, con conocimiento de estos y sabiendo que recibirán un *feedback*, con la finalidad de ver si resultaba en una reducción de los siniestros, los resultados del análisis concluyeron que en promedio los accidentes se redujeron en un 20%.

El seguro PAYD, haciendo uso de los datos que puede ofrecer el instrumento GPS, es otra alternativa a las ya convencionales herramientas que buscan reducir la velocidad. Con el PAYD los conductores tienen un incentivo financiero, una reducción de la prima, al no conducir con exceso de velocidad, siendo el PAYD una herramienta de aplicación que no se limita a zonas concretas (como radares y cámaras) sino que aplica de forma continuada a la conducción. Con el objetivo de ver si existe una reducción en el exceso de velocidad con este seguro, Bolderdijk et al. (2011) realizan un estudio de 141 asegurados de menos de 30 años, ofreciendo un descuento en la prima y *feedback* del comportamiento recogido mediante un dispositivo GPS, una vez finalizado el periodo de estudio la estimación de la reducción del exceso de velocidad con el sistema PAYD es de un 14%, habiendo tenido un mayor efecto en la reducción de los excesos de velocidad en las carreteras de 50 km/h, 80 km/h y 100 km/h. La reducción en las carreteras de 30km/h es menor puesto que el exceso de velocidad de un 6% (equivalente a 31,8km/h) o más, es muy fácil de producirse y a veces se produce de manera involuntaria aun teniendo en mente el incentivo, el exceso de velocidad en las carreteras de 120km/h son menos frecuentes así que el hecho de introducir un incentivo no reduce en gran medida estos excesos. Siendo la mayoría de los accidentes en las carreteras que el PAYD tiene más efecto, este seguro puede ser una muy buena medida para una mejora en la seguridad vial.

Si bien se da una reducción del exceso de velocidad en las dos fases de intervención que se ofrece un descuento, una vez acaban estas, aunque los conductores siguen estando monitorizados el exceso de velocidad aumenta entre la fase de pre-medición y post-medición, posiblemente por una pérdida de conciencia de que se está registrando su conducción. Complementariamente también se intenta comprobar si hay una reducción en el kilometraje y la conducción nocturna los fines de semana ofreciendo otro descuento a parte, los resultados en este estudio no demuestran que el seguro PAYD provoque una reducción en el kilometraje ni en la conducción nocturna, siendo los posibles motivos que el incentivo era menor que el ofrecido por la reducción con exceso de velocidad, también una reducción del kilometraje y conducción nocturna implica una mayor planificación y dificultad.

Ellison et al. (2015) muestran, realizando un estudio del seguro PAYD en conductores de Sídney, que hay una reducción del riesgo de exceso de velocidad, esta reducción se da tanto en una fase con incentivo y *feedback*, como en una con solo *feedback*, demostrando así que se produce una reducción simplemente recibiendo información sobre el exceso de velocidad siendo esta mayor cuanto más frecuentemente se reciba la información, aunque también muestra que se obtiene una mayor reducción cuando se incluyen incentivos financieros. Otro punto interesante que señalan los autores es que los conductores que tenían un comportamiento menos peligro antes de los incentivos y la información de conducción son los que han tenido una mayor reducción del exceso de velocidad, y comentan que para los conductores con más riesgo tal vez se deberían de poner sanciones, en vez de un incentivo, puesto que no se da un gran cambio en ellos.

2.3. Inconvenientes

El principal inconveniente que con el que se encuentra el seguro basado en el uso es como valoran los asegurados la pérdida de privacidad:

Las compañías aseguradoras asumen el riesgo de los conductores a cambio de una prima ex ante, con el seguro basado en el uso se llegan a dar precios más justos ya que se basan en un uso real y comprobado, además puede reducir las pérdidas incurridas de las aseguradoras al poder hacer estimaciones de riesgo más precisas y poner en disposición de los asegurados un *feedback* de su comportamiento. Con las tecnologías del internet de las cosas (IoT), las aseguradoras pueden acceder a una gran cantidad de datos privados de los asegurados, en (Derikx et al, 2016) se estudia si se puede compensar la pérdida de privacidad que sufren conductores con un seguro basado en el uso y de qué forma, haciendo una distinción entre tipos de privacidad. Se expone como los participantes, en la elegibilidad del seguro de auto basado en el uso, le dan mayor importancia al descuento ofrecido, con un peso del 65%, mientras que los demás atributos (registro del comportamiento en carretera, ofertas de seguros adicionales, publicidad de terceros y el registro del kilometraje) tienen un peso similar entre el 10% y 7%. Se concluye que hay preferencia a una póliza convencional a una basada en el uso, es necesaria una compensación de 9,54€ al mes para cambiarse de modalidad, por otro lado los participantes sí que estarían dispuestos a renunciar a las distintas privacidades, por una compensación, siendo el *buy-off* más elevado para la privacidad referente al comportamiento y acciones, la privacidad de datos e imágenes si la usan terceros para publicidad y por último la privacidad de localización y espacio, la única privacidad que

los asegurados consideran de forma positiva cederla es la de datos e imágenes que use la propia compañía para anuncios y ofertas más personalizados.

Hoy en día la tecnología permite a través de un monitoreo digital, el conocer el comportamiento de los asegurados de forma transparente, reduciendo las posibles asimetrías de información, ofreciendo primas más eficientes, en (Biener et al, 2019) se analiza qué supone introducir este tipo de contratos, en los que una parte de la privacidad del conductor queda al descubierto. Llevando a cabo un experimento en el que los participantes están monitorizados teniendo que compartir datos de su ubicación, el resultado que se obtiene es que su umbral de *Willingness To Pay* (WTP) por una póliza de seguro se reduce entre un 25% y 50%, aunque el uso de *transparency contracts* reduce los costes que se derivan del riesgo moral, por lo que es necesario valorar los beneficios de este tipo de contratos. El bienestar de los conductores se puede aumentar cuando los beneficios de una prima más baja, por una mejora en el comportamiento de conducción es mayor al esfuerzo realizado para mejorarlo.

3. Variables de riesgo

En una amplia parte de los artículos revisados la ocurrencia de siniestro y la frecuencia siniestral ha sido de gran interés de estudio. (Alcañiz et al, 2014; Ayuso et al, 2014a; Pesantez-Narvaez et al, 2019) centran sus estudios en la ocurrencia de siniestros, (Ayuso et al. 2014b; Ayuso et al, 2016a; Ayuso et al, 2016b) tienen como variable respuesta el tiempo y distancia transcurrida al primer accidente, (Boucher et al, 2013; Ayuso et al, 2018; Guillén et al, 2019) estudian el número de siniestros en los asegurados.

Actualmente los distintos instrumentos telemáticos que hay disponible permiten el estudio de otras variables indicadoras de riesgo que guardan relación con los siniestros, pudiendo ser de interés para cuantificar el riesgo o analizar perfiles de riesgo entre conductores:

En (Pérez-Marín et al, 2019) y (Guillén et al, 2021) se estudia el efecto que tienen los distintos factores de riesgo en los kilómetros recorridos a velocidades superiores al límite permitido. Teniendo en cuenta que cuantos más kilómetros recorridos por encima del límite permitido el riesgo de accidente aumenta, es posible identificar conducciones peligrosas mediante la información telemática que nos muestra el comportamiento específico de cada conductor, Pérez-Marín et al. (2019) identifican en bajo, medio y alto riesgo, a los conductores en función del percentil condicional estimado dado un número fijo de kilómetros conducidos con exceso de velocidad y presentando un valor distinto en las demás variables.

Hoy en día existe la posibilidad de obtener datos telemáticos de alta frecuencia, utilizando datos recogidos segundo a segundo, mediante GPS, sobre velocidades, aceleración, frenado y cambios de dirección. En (Gao & Wüthrich, 2019) sobre una muestra de conductores de automóviles, haciendo uso de *Convolutional Neural Networks* (*ConvNets*) profundas y utilizando estos datos de alta frecuencia se consigue clasificar correctamente viajes individuales a sus conductores, información que podría ser de interés para las aseguradoras en la clasificación de riesgo, ya que permite conocer información específica de cada conductor, además de ser de utilidad cuando un mismo vehículo es

conducido por varias personas, así la clasificación de los asegurados puede ser más precisa. Estos datos podrían ser beneficiosos para analizar aquellos viajes que provoquen más reclamaciones, así como estudiar la probabilidad de que ocurra un siniestro, pudiendo clasificar los viajes que tengan más peligro.

En el estudio realizado con conductores griegos (Guillén et al, 2020), se remarca la importancia de conocer los *near-miss events*, casi-accidentes (en adelante), que son aquellos accidentes que no acabaron sucediendo por poco. El estudio de estos casi-accidentes aporta información de valor tanto para aseguradoras como a las autoridades de tráfico, para la prevención de accidentes futuros e identificación de conductas con riesgo, pudiendo conocer qué factores de riesgo son más relevantes, en qué tipología de casi-accidentes tienen incidencia y en qué medida.

Otro estudio con datos de vehículos de China (Sun et al, 2020) utiliza dos variables que son consideradas de utilidad para el estudio del comportamiento con riesgo, en concreto utilizan el número de frenos, que puede indicar que se estaba conduciendo a una alta velocidad y ha sido necesario frenar ante alguna situación con peligrosidad, y el acelerador, ya que una frecuencia alta de aceleraciones bruscas es más frecuente en los conductores con más riesgo. Se pone de manifiesto que puede ser de interés el estudio de estas variables para aquellas aseguradoras que no disponen de datos de siniestros, pero sí que tienen datos *Internet of Vehicles (IoV)*. También se muestra que es posible generar una clasificación de los conductores según su riesgo de conducción en buenos conductores y conductores con riesgo, mediante la predicción y el valor observado de la variable acelerador, por lo que sin información de los siniestros es posible tener información de comportamientos en la conducción con riesgo.

4. Factores de riesgo

En la actualidad es posible medir factores de riesgos distintos a los ya muy conocidos tradicionales, numerosos artículos de los que forman parte de la revisión bibliográfica revelan la importancia de la inclusión de variables telemáticas:

En (Alcañiz et al, 2014; Ayuso et al, 2014a; Pesantez-Narvaez et al, 2019) se evidencia que las variables telemáticas son muy relevantes en explicar la probabilidad de que ocurra un siniestro con y sin culpa. Las variables telemáticas también son muy relevantes en la explicación de la distancia y el tiempo transcurrido al primer accidente (Ayuso et al, 2014b; Ayuso et al, 2016a; Ayuso et al, 2016b). Así como lo son para el número de siniestros, en (Boucher et al, 2013) el kilometraje se muestra relevante en la explicación de cuatro categorías de siniestros (daños materiales y corporales, con y sin culpa), en (Ayuso et al, 2018) y posteriormente en (Guillén et al, 2019) se logra un mejor ajuste incluyendo ambos tipos de variables (tradicionales y telemáticas) para siniestros con y sin culpa, y en el caso que se use solo un tipo de variables, se obtiene un mejor ajuste incluyendo solamente las variables telemáticas.

Hay una distinguida relación entre las variables telemáticas y el riesgo asociado a la conducción con independencia del modelo utilizado para su medición (*Logistic Regression, Support Vector Machine, Random Forests, XGBoost, Neural Networks y Poisson Regression (logarithm of annual mileage offset)*), además con la inclusión de

variables telemáticas se mejora la predicción de la probabilidad de riesgo y frecuencia siniestral (Huang & Meng, 2019).

Las variables telemáticas también son muy relevantes en los kilómetros recorridos a velocidades superiores al límite permitido (Pérez-Marín et al, 2019; Guillén et al, 2021). Guillén et al. (2020) muestran que para los tres tipos de casi-accidentes (aceleración, frenado y curvas), son factores de riesgo relevantes ambos tipos de variables, tradicionales y telemáticas, y se obtiene un mejor ajuste en las estimaciones del modelo de regresión binomial negativa introduciendo ambas variables.

En los siguientes apartados veremos para cada factor de riesgo, tanto tradicionales (género, edad, experiencia del conductor, antigüedad del vehículo, potencia del motor y lugar de aparcamiento), como telemáticos (kilometraje, conducción urbana, conducción nocturna y conducción con exceso de velocidad), su efecto en las distintas variables de riesgo mencionadas en el apartado anterior.

4.1. Tradicionales

Seguidamente veremos aquellos factores que han sido utilizados más ampliamente en la segmentación de riesgos y tarificación de primas del seguro automovilístico, los factores tradicionales.

4.1.1. Género

El primer factor de riesgo tradicional que se expone es el género, actualmente no se permite una tarificación diferenciada en género por la implementación de la Directiva Europea 2004/113/CE, una vez el Tribunal de Justicia de la Unión Europea invalidó la excepción al principio de igualdad de trato entre hombres y mujeres en el acceso a bienes y servicios así como su suministro, excepción en la que aquellos estados miembros que considerasen el factor de género como determinante en el riesgo podían hacer distinciones en las primas y prestaciones. Así, desde el 21 de diciembre de 2012 se deben cobrar las mismas primas, independientemente del género, si las demás características son iguales, aún con evidencias de que los hombres tienen unos patrones de conducción de más riesgo como se mostrará a continuación mediante la revisión bibliográfica:

Varios artículos llegan a la misma conclusión, los hombres presentan comportamientos con más riesgo en la conducción. En (Ayuso et al, 2014b) tanto para los conductores principiantes (antigüedad de carnet menor a 1 año) y conductores más experimentados (con antigüedad de carnet mayor a 1 año), los hombres son los que tienen más kilómetros por día recorridos, un mayor porcentaje de conducción urbana, un mayor porcentaje de conducción nocturna y un mayor porcentaje de conducción con exceso de velocidad. En (Ayuso et al, 2014a) el número de kilómetros recorridos anualmente, el porcentaje de kilómetros con conducción nocturna y el porcentaje de kilómetros con exceso de velocidad es superior para los hombres comparando tanto la muestra total de estudio como las submuestras con y sin siniestros. El porcentaje de conducción urbana es superior en mujeres en el total y en la submuestra sin siniestros, en la submuestra con siniestros es superior en los hombres. Posteriormente en los artículos (Ayuso et al, 2016a) y (Ayuso et al, 2016b) la muestra de estudio presenta que los hombres tienen un mayor porcentaje de kilómetros conducidos en zonas urbana, un mayor porcentaje de kilómetros

conducidos de noche y un mayor porcentaje de kilómetros conducidos por encima del límite de velocidad.

En (Ayuso et al, 2014b) la submuestra de conductores principiantes deja ver que, los hombres tienen el primer accidente en menor tiempo que las mujeres, entre mujeres principiantes aquellas con mayor porcentaje de conducción nocturna tienen el primer accidente en menor tiempo, mientras que para los hombres la conducción nocturna no es una variable relevante. Para los conductores experimentados el aumento del porcentaje de exceso de velocidad (con efecto cuadrático positivo²) reduce más el tiempo al primer accidente en los hombres. Respecto a la distancia recorrida al primer accidente, para los conductores experimentados, las mujeres recorren menos kilómetros antes del primer accidente, en los conductores principiantes la variable género no presenta un coeficiente significativo.

Estudiando la ocurrencia o no de siniestro (Ayuso et al 2014a), para ambos géneros se da que un aumento de la antigüedad de carnet conlleva una menor probabilidad de tener un siniestro, mientras que un aumento de la potencia del vehículo, el aumento del número de kilómetros recorridos y la conducción por zonas urbanas hacen aumentar la probabilidad de tener un siniestro. Las diferencias encontradas entre géneros son que, al aumentar la antigüedad del vehículo la probabilidad de siniestro aumenta y se da una menor probabilidad de siniestro con un exceso de velocidad entre 0-2% frente a un exceso de velocidad mayor al 12% solo en los hombres. Mientras que, solo para las mujeres el hecho de tener un porcentaje de conducción nocturna sea superior al 30% hace aumentar la probabilidad de tener un siniestro en ellas. Los parámetros que acompañan las variables de edad y lugar de aparcamiento no son significativos para ambos géneros.

Los hombres recorren más kilómetros que las mujeres hasta tener el primer accidente con culpa. Cuanta más experiencia mayor es la distancia al primer siniestro con culpa, mientras que a mayor antigüedad del vehículo y porcentaje de kilómetros recorridos en zonas urbanas más se reduce la distancia en ambos sexos, con efecto cuadrático positivo en la conducción urbana de mujeres. Cuanto mayor es el porcentaje de kilómetros por encima de los límites (con efecto cuadrático positivo) más se reduce la distancia en los hombres, pero no es significativo en las mujeres. Y a mayor porcentaje de kilómetros conducidos de noche más se reduce la distancia en las mujeres, pero no es significativo en los hombres (Ayuso et al, 2016a).

Siguiendo en el estudio (Ayuso et al, 2016b), se estiman dos modelos Weibull separando la muestra por hombres y mujeres, en ellos el efecto de las variables explicativas presenta similitudes y diferencias entre géneros: Cuanta más experiencia mayor es el tiempo transcurrido al primer siniestro, mientras que, a mayor antigüedad del vehículo, kilómetros recorridos en zonas urbanas (con efecto cuadrático positivo) y mayor intensidad de conducción menor es el tiempo transcurrido en ambos sexos. Cuantos más kilómetros por encima de los límites (con efecto cuadrático positivo) más se reduce el tiempo transcurrido para los hombres, pero no es significativo en las mujeres y a mayores kilómetros conducidos de noche más se reduce el tiempo transcurrido en las mujeres, pero no es significativo en los hombres. Finalmente, en el estudio, se forman cuatro grupos de riesgo (bajo, medio, alto y sin descuento), según los kilómetros recorridos al año, los

² La interpretación del efecto cuadrático será el mismo para las distintas variables y estudios en los que aparece, este efecto captura la no linealidad del factor de riesgo, siendo el coeficiente positivo (negativo), indica que el efecto, ya sea en el tiempo, distancia al siniestro o número de siniestros, no es tan negativo(positivo) cuando el factor de riesgo presenta un valor alto a uno bajo al que hubiese tenido sin este efecto cuadrático.

kilómetros de conducción nocturna, el porcentaje de conducción urbana y el exceso de velocidad. En ellos no se muestran grandes diferencias entre hombre y mujeres, aunque los años esperados que han de transcurrir hasta el primer siniestro con culpa son menores en las mujeres en todos los grupos de riesgo menos en el de sin descuento. Las probabilidades de que el primer accidente con culpa se de en el primer año son similares entre hombres y mujeres, aunque las mujeres presentan unos valores un poco más altos en el grupo bajo, medio y alto, y más bajo en el sin descuento.

En (Ayuso et al, 2016b) se estiman dos modelos Weibull uno primero que no incluye la variable explicativa kilómetros recorridos por día, y uno segundo que si la incluye, en el primer modelo el coeficiente de la variable género sí es significativo, mostrando que se espera que el tiempo transcurrido al primer accidente con culpa sea más reducido en los hombres, pero en la estimación del segundo modelo, el coeficiente de la variable género deja de ser significativo, y sí lo es el de la variable kilómetros por día, siendo negativo, por lo que cuanto mayor es la intensidad de conducción menor es el tiempo esperado al primer accidente con culpa. Hecho que puede indicar que las disparidades entre hombres y mujeres están mayoritariamente explicadas por una mayor conducción de los hombres, lo que conlleva a una mayor exposición al riesgo de sufrir un accidente. En (Alcañiz et al, 2014) la muestra presenta un porcentaje levemente superior de hombres en los asegurados con siniestro, importante recalcar que en este estudio la variable género no se muestra relevante en la explicación de la ocurrencia o no de los siniestros y en este caso la regresión logística estimada también incluye el kilometraje (número de kilómetros anuales recorridos).

Si el asegurado es hombre se espera un mayor número de siniestros sin culpa de daños materiales y un menor número de siniestros sin culpa de daños corporales (Boucher et al, 2013).

No hay diferencias significativas en el género en términos de porcentajes de siniestralidad, pero los hombres presentan un porcentaje más elevado de siniestros. Teniendo en cuenta una muestra de asegurados con siniestros con culpa y sin culpa, cuando se tiene en cuenta ambas variables (tradicionales y telemáticas), la variable género no es relevante, aunque si solo se tienen en cuenta los siniestros con culpa el ser hombre presenta un efecto negativo en el número de siniestros, por lo que se espera un menor número de siniestros. Cuando el kilometraje es introducido como una variable más, sino que es introducido como un *offset*, el parámetro que acompaña al género es significativo para la estimación de siniestros con y sin culpa, con un efecto negativo, en la estimación de siniestros solo con culpa, el parámetro que acompaña al género no es significativo (Ayuso et al, 2018).

Los conductores con siniestros son en su mayoría hombres, si el conductor es hombre el número de siniestros (con y sin culpa) esperados es menor, no se muestra relevante al tener en cuenta solo los siniestros con culpa, es una de las variables menos importantes al determinar el riesgo de siniestro, aunque en la estimación del número de siniestros también se incluye el kilometraje como *offset* (Guillén et al, 2019).

El género no muestra ser una variable relevante en la explicación de ninguno de los tres casi-accidentes (aceleración, frenado y curvas) del estudio de Guillén et al. (2020).

Si el conductor es hombre se espera que el número de kilómetros con exceso de velocidad sea mayor, si se analizan los kilómetros recorridos a velocidad superiores al

límite mediante una regresión por cuantiles³ también se espera un mayor número si el conductor es hombre en todos los percentiles, excepto el percentil 99% que el género no presenta un parámetro significativo (Pérez-Marín et al, 2019). En un estudio posterior de Guillén et al. (2021), el género muestra el mismo efecto, los hombres conducen más kilómetros con exceso de velocidad, aunque la variable género en la regresión por cuantiles, no presenta un parámetro significativo en los percentiles 97,5% y 99%, para los demás (50%, 75%, 90% y 95%) el hecho de ser hombre supone un aumento de los kilómetros en exceso de velocidad en estos percentiles, a modo de ejemplo en el artículo se calculan los percentiles para hombres y mujeres, conduciendo 10.000, 20.000 y 30.000 kilómetros por año de los cuales 2.000 se encuentran por encima del límite permitido, los resultados muestran un percentil para los hombres del 90%,54% y 29% respectivamente y para las mujeres del 92%,57% y 30%, también se hace el mismo cálculo para 3.000 kilómetros por encima del límite permitido y los resultados son para hombres el percentil 98%, 74% y 44%, y para mujeres el 98%, 75% y 45%, estos ejemplos muestran que para los mismos kilómetros anuales recorridos y kilómetros con exceso de velocidad, la posición del percentil es superior para las mujeres.

Las conclusiones a las que se llegan en estos estudios son de elevada importancia puesto que como se apuntaba en el primer párrafo del subapartado actualmente no se permite una tarificación diferenciada por el género, aunque la revisión bibliográfica indica que los hombres tienen unos patrones de conducción de más riesgo, lo que reforzaría la idea de que los hombres deberían de pagar unas primas más elevadas por el seguro de automóvil.

Ante el condicionante de no poder diferenciar las tarifas, en el seguro de automóvil se da una mayor facilidad en comparación con los seguros de vida para encontrar factores de riesgo que compensen la eliminación del género como factor clasificatorio, en este nuevo paradigma las aseguradoras disponen de la posibilidad de introducir entre sus productos el seguro basado en el uso, el cual permite tener en cuenta otras variables que pueden recoger las diferencias que se dan entre géneros, reduciendo así el impacto de la imposibilidad de tener en cuenta esta variable como factor de riesgo en la tarificación y pudiendo mantener una equidad actuarial. Estos nuevos factores recogen el uso que haga el asegurado del vehículo, así como su comportamiento en la conducción, factores que son claves para medir el riesgo de los asegurados, hombres y mujeres con un mismo comportamiento y uso del vehículo (porcentaje de conducción urbana, nocturna, con exceso de velocidad y los kilómetros recorridos anuales) no se muestran con grandes diferencias respecto al riesgo de sufrir un siniestro, sino que es muy parecido.

Aquellos conductores que conducen de forma ocasional y tienen un comportamiento de conducción con menos riesgo, serán los grandes beneficiados de este seguro, de esta forma no solo se solventa en gran medida la imposibilidad de tener en cuenta el género como factor de riesgo sino que aquellos hombres que se veían perjudicados por el hecho de pertenecer a este género pero que, presentan un bajo kilometraje o que su comportamiento en la conducción es el adecuado, podrían ver su prima reducida, por otra parte las empresas aseguradoras pueden identificar aquellas mujeres aseguradas que por su conducción presentan un mayor riesgo y adjudicarles una prima superior.

³ Se estima una regresión por cuantiles debido que la variable dependiente, el número de kilómetros con exceso de velocidad es muy asimétrica por la derecha.

Importante comentar por qué no se solventa en su totalidad la imposibilidad de tener en cuenta el género en la tarificación, y es porqué aun teniendo en cuenta otros factores, hemos visto mediante la revisión bibliográfica que hay factores de riesgo que son relevantes para explicar las distintas variables dependientes para ambos géneros, pero otros factores tan solo se muestran relevantes en uno de los géneros, los hombres recorren más kilómetros antes de tener el primer siniestro, también hemos visto que un aumento de la conducción con exceso de velocidad reduce la distancia y tiempo al primer accidente así como aumenta la probabilidad de tener un siniestro en los hombres pero no en mujeres, y por último que un aumento de la conducción nocturna reduce la distancia y tiempo al primer accidente así como aumenta la probabilidad de tener un siniestro en las mujeres y no en los hombres, pero para ninguno de los factores se permite una diferenciación a la hora de determinar distintos umbrales y el descuento por un menor kilometraje, exceso de velocidad o menor conducción nocturna.

4.1.2. Edad

A continuación, veremos qué efecto presenta en las variables de riesgo el segundo factor tradicional, la edad del asegurado:

Los asegurados con siniestro son un poco más jóvenes respecto los que no tienen siniestros, aunque la variable edad no se muestra relevante en explicar la probabilidad de tener o no un siniestro, independientemente si se separan las muestras por género o no (Alcañiz et al, 2014) y (Ayuso et al, 2014a).

La edad media es menor en los conductores con menos experiencia, aunque su efecto en la distancia y tiempo transcurrido hasta el primer accidente no se muestra relevante tanto en los conductores experimentados como los principiantes (Ayuso et al, 2014b). Tampoco es una variable relevante para explicar la distancia y tiempo transcurrido al primer siniestro con culpa (Ayuso et al. 2016a) y (Ayuso et al, 2016b) también con muestras separadas por género. Aunque posteriormente en (Pesantez-Narvaez et al, 2019) este factor de riesgo resulta ser el segundo más importante para la predicción de siniestro con culpa según el modelo de regresión logística simple del estudio, la variable edad presenta un coeficiente negativo significativo, que indicaría que, a mayor edad, menor es la probabilidad de tener un accidente con culpa.

Se espera un número mayor de siniestros en aquellos conductores con menos de 25 años en las cuatro categorías de siniestros (daños materiales con culpa, siniestros con daños corporales con culpa, siniestros con daños materiales sin culpa, siniestros con daños corporales sin culpa). También se espera un mayor número de siniestros sin culpa con daños materiales en los asegurados entre 25 y 30 años (Boucher et al, 2013).

Los conductores sin siniestro presentan una media de edad levemente mayor, la variable edad se muestra relevante en la explicación de los siniestros con culpa reduciendo el número de siniestros esperados (con un efecto cuadrático positivo) no se muestra una variable relevante al tener en cuenta todos los siniestros (con y sin culpa) (Ayuso et al, 2018) y (Guillén et al, 2019).

La edad de los conductores no se muestra como una variable relevante para explicar el número de kilómetros con exceso de velocidad en el modelo de regresión lineal utilizado en (Pérez-Marín et al, 2019), además al estudiar el posible efecto de los factores

de riesgo en los kilómetros recorridos a velocidades superiores al límite por percentiles, la variable edad, en todos los percentiles, no presenta ningún parámetro significativo. Posteriormente en otro estudio (Guillén et al, 2021) mediante una regresión gamma tampoco se muestra relevante la variable edad con la misma variable respuesta, y en el modelo de regresión por cuantiles, el parámetro que acompaña a la variable edad solo es significativo en los percentiles 95% y 97,5%, siendo este positivo, por lo que, a mayor edad, estos dos percentiles del número de kilómetros con exceso de velocidad aumentan.

La variable edad en el estudio de Guillén et al. (2020) muestra una relación positiva con los tres casi-accidentes, a mayor edad mayor es el número esperado de tener un casi-accidente de aceleración, frenado o curvas, por lo que el riesgo de tener un casi-accidente es mayor en aquellos conductores con más edad.

Ante los hallazgos de la no relevancia de la variable edad en algunos estudios, hay que tener en cuenta dos puntos importantes, uno de ellos es que la mayoría de las muestras utilizadas en los diferentes artículos estaban formadas por conductores jóvenes con un seguro UBI, por lo que puede ser que el parámetro que acompaña a esta variable no sea significativo porque el rango de edades no es muy grande como para que se recoja el efecto que un aumento de esta pueda conllevar en los siniestros o los kilómetros con exceso de velocidad. Otro punto importante es que, en las distintas estimaciones realizadas para las variables explicativas de estudio de siniestralidad, generalmente se incluía también la variable antigüedad de carnet, y esta tiene una alta correlación con la edad, la edad media es menor en los conductores con menos experiencia. No obstante, también puede ser que la variable edad no sea una variable relevante en los estudios comentados.

4.1.3. Experiencia del conductor

El factor de riesgo experiencia del conductor, recogido como años con permiso de conducción (antigüedad de carnet), se muestra muy relevante para explicar el riesgo de un asegurado:

Los asegurados con siniestro tienen menos experiencia respecto los que no tienen siniestros, el efecto que tiene un aumento de la antigüedad de carnet es el de disminuir la probabilidad de tener un siniestro, también separando la muestra por género (Alcañiz et al, 2014) y (Ayuso et al, 2014a).

El aumento de la experiencia aumenta la distancia y tiempo al primer accidente en los conductores más experimentados, no hay este afecto en los principiantes (Ayuso et al, 2014b). Analizando solo los siniestros con culpa, los asegurados culpables tienen menos experiencia (Pesantez-Narvaez et al, 2019), además cuanto más experiencia mayor es la distancia y tiempo transcurrido al primer siniestro con culpa, también separando la muestra entre géneros, (Ayuso et al, 2016a) y (Ayuso et al, 2016b).

Los conductores sin siniestro tienen más experiencia. Con una mayor experiencia se espera un menor número de siniestros, tanto teniendo en cuenta todos los siniestros como solo los siniestros con culpa (Ayuso et al, 2018) y (Guillén et al, 2019). Adicionalmente en (Guillén et al, 2019) se determina que la antigüedad de carnet es la segunda variable más importante para determinar el riesgo de siniestro.

La experiencia en la conducción tiene un efecto negativo respecto los casi-accidentes en curvas, a mayor experiencia de conducción, el número esperado de casi-accidente en curvas se reduce (Guillén et al, 2020).

Una vez realizada la revisión bibliográfica, podemos determinar que, a mayor experiencia menor es el riesgo esperado, la experiencia del conductor es un factor de riesgo que se ha mostrado muy relevante para explicar y predecir la ocurrencia de siniestro, la distancia y tiempo al primer siniestro, el número de siniestros, así como los casi-accidentes.

4.1.4. Antigüedad del vehículo

El factor de riesgo antigüedad del vehículo, se muestra relevante para explicar el riesgo de un asegurado:

Referente a la antigüedad del vehículo no se encuentran diferencias significativas entre asegurados con y sin siniestros, además la antigüedad del vehículo no muestra ser una variable relevante en la probabilidad de tener un siniestro en los artículos (Alcañiz et al, 2014) y (Pesantez-Narvaez et al, 2019). Analizando muestras separadas entre hombres y mujeres, la antigüedad del vehículo es relevante solo en hombres, un aumento de esta aumenta la probabilidad de tener un siniestro (Ayuso et al, 2014a).

Separando la muestra de estudio entre principiantes y experimentados, la antigüedad del vehículo es mayor en los conductores con menos experiencia, pero solo para los conductores experimentados a mayor antigüedad del vehículo la distancia recorrida al primer accidente se reduce (Ayuso et al, 2014b). En dos estudios posteriores (Ayuso et al, 2016a) y (Ayuso et al, 2016b), se muestra que a mayor antigüedad del vehículo menor es la distancia y el tiempo transcurrido al primer accidente con culpa, también con la muestra separada entre hombres y mujeres.

Se esperan menos siniestros de daños corporales sin culpa en aquellos conductores con una antigüedad del vehículo menor a dos años (Boucher et al, 2013). No hay diferencias significativas entre asegurados con y sin siniestros en la antigüedad del vehículo, se espera un mayor número de siniestros cuanto mayor sea antigüedad del vehículo, tanto teniendo en cuenta todos los siniestros como solo los siniestros con culpa (Ayuso et al, 2018). Los conductores sin siniestros presentan una antigüedad del vehículo un poco mayor, sin ser significativa la diferencia con los asegurados con siniestro, es una de las variables menos importantes al determinar el riesgo, el número de siniestros esperados aumenta al aumentar la antigüedad del vehículo, tanto con los siniestros con culpa como sin culpa (Guillén et al, 2019).

La antigüedad del vehículo también muestra ser importante en los casi-accidentes, concretamente, una mayor antigüedad del vehículo reduce el número esperado de casi-accidente de frenado y aceleración, posiblemente por las características técnicas que tienen los vehículos con mayor antigüedad (Guillén et al, 2020).

Una vez concluida la revisión bibliográfica, se ha observado que no se encuentran diferencias significativas entre asegurados con y sin siniestros en la antigüedad del vehículo, aunque si se muestra que un aumento de la antigüedad del vehículo implica un mayor riesgo esperado, siendo relevante para explicar y predecir la ocurrencia de

siniestro, la distancia y tiempo al primer siniestro, el número de siniestros, así como los casi-accidentes.

4.1.5. Potencia del motor

El factor de riesgo potencia del motor, no es un factor que haya estado introducido en demasía en los diferentes estudios, pero sí que se muestra relevante para explicar el riesgo de un asegurado:

La potencia del vehículo es superior en aquellos asegurados con siniestros, además se muestra que una mayor potencia del vehículo aumenta la probabilidad de tener un siniestro, también separando la muestra por género, (Alcañiz et al, 2014) y (Ayuso et al, 2014a).

Los conductores con siniestros presentan vehículos con mayor potencia. Con un aumento de la potencia se espera un mayor número de siniestros teniendo en cuenta los siniestros con culpa y sin culpa (Ayuso et al, 2018). Los conductores con siniestros tienen una mayor potencia del vehículo, el número de siniestros esperados aumenta al aumentar la potencia del vehículo, tanto con los siniestros con culpa como sin culpa (Guillén et al, 2019).

En (Guillén et al, 2020) se sacan otras conclusiones de interés, a mayor potencia del vehículo mayor es el número esperado de casi-accidentes de aceleración y se espera un menor número en casi-accidentes de curvas.

Tal y como se ha visto en la revisión bibliográfica, la potencia del conductor es un factor de riesgo relevante para explicar y predecir la ocurrencia y el número de siniestros, así como los casi-accidentes, a mayor potencia mayor es el riesgo esperado.

4.1.6. Lugar aparcamiento

El factor de riesgo zona de aparcamiento, no es un factor que haya estado ampliamente introducido en los diferentes estudios, debido a que no es una variable muy relevante para explicar el riesgo de un asegurado:

Dos artículos (Alcañiz et al, 2014) y (Ayuso et al, 2014a) incluyen el lugar de aparcamiento como variable independiente dicotómica (si está aparcado en garaje o no), teniendo como variable explicativa la ocurrencia o no de siniestro, ambos artículos llegan a la misma conclusión, el lugar de aparcamiento no muestra ser una variable relevante en la probabilidad de tener un siniestro, independientemente si se separa la muestra de estudio por género o no.

Se espera un mayor número de siniestros de daños materiales sin culpa en los asegurados que tienen el coche aparcado en garaje privado (Boucher et al, 2013). El lugar de aparcamiento no es relevante en ninguna de las estimaciones del número de siniestros con culpa y sin culpa o solo con culpa (Ayuso et al, 2018). En (Guillén et al, 2019) se identifica el lugar de aparcamiento como una de las variables menos importantes al determinar el riesgo de siniestro, no es una variable relevante para el número de siniestros

con o sin culpa, aunque en la muestra los conductores con siniestros presentan un mayor porcentaje de vehículos en garaje hay que tener en cuenta que los vehículos con mayor valor y potencia se guardan en garaje privado con mayor frecuencia, y como hemos comentado en el apartado anterior se espera un mayor número de siniestros a mayor potencia.

En (Guillén et al, 2020) el hecho de estacionar el vehículo en garaje por la noche reduce el número esperado de casi-accidentes por frenado y aceleración.

La revisión bibliográfica nos llevaría a concluir que, como se apuntaba en el principio, el lugar de aparcamiento no es una variable relevante para explicar o predecir la ocurrencia o el número de siniestros.

4.2. Telemáticos

Seguidamente veremos en detalle aquellos factores más novedosos en la clasificación de riesgos y tarificación de primas del seguro automovilístico, los factores telemáticos.

4.2.1. Kilometraje

El primer factor de riesgo telemático que se expone es el kilometraje, con la revisión bibliográfica se ha querido no solo ver qué efecto tiene en las variables de riesgo, sino también examinar la relación que tiene el kilometraje y la frecuencia siniestral:

El kilometraje anual recorrido es superior en aquellos asegurados con siniestros, un aumento del kilometraje aumenta la probabilidad de tener un siniestro (Alcañiz et al, 2014). El porcentaje de kilómetros por día es mayor en los conductores con más experiencia (Ayuso et al, 2014b), cuanto mayor es la intensidad de conducción menor es el tiempo transcurrido al primer accidente con culpa (Ayuso et al, 2016b).

Como se indicaba en el apartado referente al género, separando las muestras entre hombres y mujeres, el kilometraje es un factor relevante en la siniestralidad para ambos, un aumento del número de kilómetros recorridos anualmente aumenta la probabilidad de tener un siniestro (Ayuso et al, 2014a) y una mayor intensidad de conducción también implica un menor tiempo transcurrido al primer accidente con culpa en ambos sexos (Ayuso et al, 2016b).

Los asegurados con los mismos kilómetros presentan una frecuencia de siniestros similares, representando una relación lineal entre el número de siniestros y el kilometraje, la estimación de los valores predichos respecto de los observados muestra que se subestima la frecuencia de siniestros de los asegurados con menos de 10.000 kilómetros anuales y sobreestima los siniestros de aquellos con más de 10.000 kilómetros anuales. Realizando una generalización del modelo de regresión de Poisson *offset* para recoger la no linealidad se demuestra que los siniestros no aumentan de forma lineal al aumentar el número de kilómetros anuales, sino que, para las cuatro categorías de accidentes (daños materiales y corporales, con y sin culpa), hay una influencia de la experiencia del conductor y otros elementos que proporcionan una mayor seguridad, esta influencia compensaría el mayor riesgo de siniestro por conducir más kilómetros (Boucher et al, 2013).

Los conductores con siniestros presentan un mayor porcentaje de kilómetros anuales. El kilometraje es el factor de riesgo que tiene un mayor efecto en el número de siniestros tanto teniendo en cuenta el conjunto (con y sin culpa), como solo los siniestros con culpa, el efecto es positivo, se espera un aumento del número de los siniestros con un aumento del número de kilómetros anuales recorridos (Ayuso et al, 2018).

Analizando el kilometraje respecto los siniestros, de los conductores que se sitúan en el primer cuartil de la distancia recorrida, han recorrido una mayor distancia aquellos conductores que han tenido siniestro respecto los que no han tenido, así mismo sucede con el segundo y el tercer cuartil. Con la estimación de un modelo de Zero Inflado de Poisson con *offset* se concluye que, cuanto más distancia se recorre, la proporción de excesos de cero siniestros es mayor. No hay una relación proporcional entre el número de siniestros y el kilometraje, puesto que el aumento del riesgo por una mayor distancia recorrida queda en parte compensado por la experiencia, como se comentaba en un párrafo anterior el porcentaje de kilómetros por día es mayor en los conductores con más experiencia, así entre aquellos que conducen largos kilómetros y reportan cero siniestros también se encuentran buenos conductores (Guillén et al, 2019).

El kilometraje anual recorrido es superior en aquellos asegurados con siniestros con culpa respecto los que no tienen siniestros. Este factor de riesgo es el tercero más importante para la predicción de siniestro con culpa según la regresión logística simple en la muestra estudiada en (Pesantez-Narvaez et al, 2019). En (Huang & Meng, 2019) de los cinco modelos de probabilidad de riesgo de riesgo, la variable kilómetros anuales se sitúa en todos los modelos entre las ocho primeras con más importancia de treinta y siete variables.

La importancia del kilometraje como factor de riesgo es relevante no solo en los siniestros, también es de importancia en otras variables, en (Sun et al, 2020) se muestra que existe una relación entre el kilometraje y la aceleración, así como el kilometraje y el frenado. En (Pérez-Marín et al, 2019) se muestra que con un mayor número de kilómetros recorridos se espera que el número de kilómetros con exceso de velocidad sea mayor, además esta relación es positiva en todos los percentiles del número de kilómetros con exceso de velocidad, por lo que a más kilómetros los distintos percentiles de los kilómetros en exceso de velocidad aumentan. En (Guillén et al, 2021) se llegan a las mismas conclusiones, con el añadido que se muestra que existe una relación exponencial entre la distancia total recorrida y los kilómetros recorridos con exceso de velocidad, por lo que a medida que se recorre una mayor distancia el aumento del número de kilómetros con exceso de velocidad es mayor.

La revisión bibliográfica muestra de forma clara que un mayor kilometraje supone un mayor riesgo. El kilometraje es un factor de riesgo que se muestra muy relevante para explicar y predecir la ocurrencia de siniestro, la distancia y tiempo al primer siniestro, el número de siniestros, así como es relevante en otras variables de riesgo, como la aceleración, frenado y el exceso de velocidad. También se ha podido determinar que la relación entre el kilometraje y la frecuencia de siniestros no es lineal.

4.2.2. Conducción urbana

El siguiente factor de riesgo telemático revisado, es la conducción en vías urbanas, los datos más recientes disponibles por la dirección general de tráfico muestran que los accidentes por víctimas para el año 2019 fueron 37.339 para las vías interurbanas (autopistas, autovías y convencionales principalmente), mientras que el número de víctimas para las vías urbanas asciende a 66.741.

Con estos datos, y como veremos a continuación se pone de manifiesto que la conducción urbana es un factor de riesgo muy relevante para explicar el riesgo de un asegurado:

El porcentaje de conducción urbana es superior en aquellos asegurados con siniestros. El aumento del porcentaje de conducción en vías urbanas aumenta la probabilidad de tener un siniestro (Alcañiz et al, 2014).

El porcentaje de conducción urbana es superior en los conductores con menos experiencia, para estos conductores, el aumentar el porcentaje de conducción por vías urbana (superior a un 25%) conlleva que el tiempo al primer accidente se reduzca, para los conductores principiantes el aumentar la conducción por vías urbana (con efecto cuadrático positivo) reduce la distancia esperada al primer accidente. Para los conductores experimentados, al aumentar la conducción por vías urbanas (con efecto cuadrático positivo), el tiempo y la distancia al primer accidente se reduce (Ayuso et al, 2014b).

Mirando la probabilidad de sufrir un accidente durante el primer año y la probabilidad de sufrir un accidente durante los primeros 10.000 km, la probabilidad de ambos aumenta a medida que el porcentaje de conducción urbana aumenta por grupos (menor del 15% de conducción urbana, 15% hasta menos del 25%, e igual o mayor a un 25% de conducción urbana) (Ayuso et al, 2014b).

Si se separan las muestras entre hombres y mujeres, como se ha comentado en el apartado referente en el género, el factor de conducción urbana es relevante en ambos géneros, un aumento de la conducción por zonas urbanas aumenta la probabilidad de tener un siniestro tanto en hombres como mujeres (Ayuso et al, 2014a). A mayor porcentaje de kilómetros recorridos en zonas urbanas (efecto cuadrático en mujeres) más se reduce la distancia en ambos sexos (Ayuso et al, 2016a) y a mayor porcentaje de conducción en vías urbanas (con efecto cuadrático positivo) menor es el tiempo al primer accidente con culpa para ambos (Ayuso et al, 2016b).

Los conductores con siniestros presentan un mayor porcentaje de kilómetros recorridos por vías urbanas, al aumentar el porcentaje de conducción urbana aumenta el número de siniestros tanto teniendo en cuenta todos los siniestros como solo los siniestros con culpa (Ayuso et al, 2018) y (Guillén et al, 2019), adicionalmente en (Guillén et al, 2019), se indica que la conducción por vías urbanas es la variable con más importancia al determinar el riesgo de siniestro.

El porcentaje de conducción urbana es superior en aquellos asegurados con siniestros con culpa respecto los que no tienen siniestros. A más kilómetros recorridos en áreas urbanas, mayor es la probabilidad de tener un accidente con culpa (Pesantez-

Narvaez et al, 2019). Tanto el método de XGBoost (*TreeBooster*) como la regresión logística simple consideran este factor de los tres más importantes en la predicción de los siniestros con culpa, el segundo para el primer método y el primero para el segundo método (Pesantez-Narvaez et al, 2019).

Con un aumento del porcentaje de kilómetros recorridos en vías urbanas se espera un menor número de kilómetros recorridos con exceso de velocidad, esta relación se muestra negativa en los distintos percentiles, cuanto mayor es el porcentaje de kilómetros con conducción urbana más disminuye el cuantil del número de kilómetros con exceso de velocidad (Pérez-Marín et al, 2019) y (Guillén et al, 2021).

Cuanto mayor sea el porcentaje de kilómetros de conducción urbana, se espera un mayor número de casi-accidentes de frenado, esto es debido al tráfico de las ciudades donde es presente una mayor densidad de coches. Este factor de riesgo telemático es el único relevante en este tipo de casi-accidentes (Guillén et al, 2020).

A lo largo de la revisión bibliográfica se muestra que, a mayor conducción urbana mayor es el riesgo asociado al asegurado. Es un factor de riesgo que se muestra muy relevante para explicar y predecir la ocurrencia de siniestro, la distancia y tiempo al primer siniestro, el número de siniestros, así como los casi-accidentes.

4.2.3. Conducción nocturna

El siguiente factor de riesgo telemático, es la conducción nocturna, factor que nos permite ver si el riesgo aumenta según las horas en las que se conduce:

El porcentaje de conducción nocturna media es superior en aquellos asegurados con siniestros, si bien los que presentan una conducción nocturna de entre el 11-20%, tienen menos probabilidad de tener un siniestro que los que presentan un porcentaje más bajo al 11% (categoría base), los que presentan más del 20% de conducción nocturna no muestran una diferencia significativa respecto la categoría base en la probabilidad de siniestro, estos hallazgos nos podrían indicar que hay una mayor siniestralidad nocturna en aquellos asegurados que conducen de forma ocasional por la noche (Alcañiz et al, 2014).

El porcentaje de conducción nocturna es superior en los conductores con menos experiencia, aunque no hay diferencias significativas entre los principiantes y los experimentados en términos de porcentaje. Tanto para los conductores experimentados como principiantes, al aumentar el porcentaje de conducción nocturna, el tiempo al primer accidente se reduce, aunque como se ha comentado en el apartado referente al género en el caso de los principiantes el efecto es relevante solo en mujeres (Ayuso et al, 2014b). A mayor porcentaje de conducción nocturna la distancia al primer accidente con culpa es menor (Ayuso et al, 2016b).

Mirando la probabilidad de sufrir un accidente durante el primer año y la probabilidad de sufrir un accidente durante los primeros 10.000 km, la probabilidad de ambos aumenta a medida que el porcentaje de conducción nocturna aumenta por grupos (menor del 5% de conducción nocturna, 5% hasta menos del 10%, e igual o mayor a un 10% de conducción nocturna) (Ayuso et al, 2014b).

Si se separan las muestras entre hombres y mujeres, como se ha comentado en el apartado referente en el género, la conducción nocturna es un factor relevante en las mujeres, un porcentaje de conducción nocturna superior al 30% hace aumentar la probabilidad de tener un siniestro (Ayuso et al, 2014a). Además, a mayor porcentaje de kilómetros conducidos de noche más se reduce la distancia y el tiempo al primer accidente con culpa solo en las mujeres (Ayuso et al, 2016a) y (Ayuso et al, 2016b).

Los conductores con siniestros presentan un mayor porcentaje de kilómetros recorridos de noche. Cuando se tiene en cuenta todos los siniestros (con culpa y sin culpa) la conducción nocturna muestra un efecto cuadrático positivo indicando que, cuando se haya conducido un elevado porcentaje de conducción nocturna esta tendrá un efecto positivo, así que se esperará un número mayor de siniestros. Cuando se tienen en cuenta solo los siniestros con culpa la conducción nocturna no se muestra relevante en la explicación de su frecuencia (Ayuso et al, 2018).

Los conductores con siniestros presentan un mayor porcentaje de kilómetros recorridos por la noche. La conducción nocturna es una de las cuatro variables con menos importancia al determinar el riesgo de siniestro, no muestra ser una variable relevante en la explicación del número de siniestros esperados (con y sin culpa) o solo con culpa (Guillén et al, 2019).

El porcentaje de kilómetros recorridos nocturnamente es ligeramente superior en los asegurados sin siniestro respecto los que tienen algún siniestro con culpa. Este factor de riesgo es el tercero más importante para la predicción de siniestro con culpa según el método XGBoost (*TreeBooster*) (Pesantez-Narvaez et al, 2019).

Con un mayor porcentaje de conducción nocturna se espera que el número de kilómetros con exceso de velocidad sea mayor, mirando la relación por percentiles, los kilómetros recorridos a velocidades superiores al límite permitido dependen positivamente del porcentaje de conducción nocturna en los percentiles 50%, 75% y 90%, no es así para los más elevados (97.5% y 99%) (Pérez-Marín et al, 2019). En el estudio posterior (Guillén et al, 2021) la conducción nocturna, tan solo tiene un efecto en el percentil 50%, por lo que al aumentar el porcentaje de conducción nocturna la mediana de los kilómetros recorridos con exceso de velocidad aumenta.

Con un mayor porcentaje de kilómetros con conducción nocturna se espera un menor número de casi-accidentes en curvas, siendo este factor de riesgo telemático el único relevante en este tipo de casi-accidentes, la posible explicación a esta esperada reducción de casi-accidentes en curvas por la noche, es que la conducción durante el día puede ser más brusca (Guillén et al, 2020).

La conducción nocturna resulta ser un factor generalmente relevante en explicar y predecir la ocurrencia de siniestro, la distancia y tiempo al primer siniestro, el número de siniestros, así como los casi-accidente. Aunque hay que apuntar que se muestran diferencias entre géneros al separar las muestras y no siempre se muestra relevante al explicar la variable dependiente, como en el caso del número de siniestros o el exceso de velocidad.

4.2.4. Conducción con exceso de velocidad

El factor de riesgo, exceso de velocidad, es sin duda uno de los factores que suponen un mayor riesgo para los conductores, la Comisión Europea advierte que este es uno de los problemas que más afecta a la seguridad vial “El exceso de velocidad es uno de los mayores problemas de seguridad vial y un factor determinante en alrededor del 30% de los accidentes mortales de tráfico. Asimismo, aumenta sustancialmente el riesgo de sufrir un accidente” (Exceso de velocidad, párrafo 1).

En la revisión bibliográfica veremos que queda recogido de forma clara este mayor riesgo:

El porcentaje de conducción con exceso de velocidad es superior en aquellos asegurados con siniestros. Los conductores que presentan un exceso de velocidad de hasta el 2% tienen menor probabilidad de tener un siniestro respecto a los que tienen un exceso superior al 11% (categoría base), los excesos de velocidad entre el 3-5% y el 6-11% no presentan un parámetro significativo, por lo que no se esperan diferencias significativas en las probabilidades de tener un siniestro (Alcañiz et al, 2014). El porcentaje de kilómetros en exceso de velocidad es mayor en los conductores con más experiencia, además para estos, el aumento del porcentaje de exceso de velocidad (con efecto cuadrático positivo) reduce el tiempo y la distancia recorrida al primer accidente. En el caso de los conductores principiantes al aumentar el porcentaje de conducción con exceso de velocidad (con efecto cuadrático positivo) conlleva también que el tiempo al primer accidente se reduzca (Ayuso et al, 2014b). A mayor porcentaje de conducción con exceso de velocidad (con efecto cuadrático positivo), menor es la distancia al primer accidente con culpa (Ayuso et al, 2016b).

Un punto interesante a destacar del artículo (Ayuso et al, 2014b) es que al estudiar la probabilidad de tener un primer accidente en los primeros 10.000 kilómetros conducidos tanto para los menos experimentados como para los más experimentados, la probabilidad es menor a medida que aumentan los kilómetros con exceso de velocidad en los distintos grupos (menos del 5% de exceso de velocidad, 5% hasta menos del 10%, más o igual del 10% de exceso de velocidad), estos resultados se dan porque los conductores que tienen un mayor porcentaje de conducción por encima del límite también son los que conducen más kilómetros por día, adquiriendo así más experiencia, y utilizan vías más seguras como las autopistas, compensando así el aumento de la probabilidad de tener un siniestro por el aumento del exceso de velocidad.

Si se separan las muestras entre hombres y mujeres, como se ha comentado en el apartado referente en el género, el factor que recoge el exceso de velocidad solo se muestra relevante para los hombres. Los hombres tienen una menor probabilidad de siniestro si presentan un exceso de velocidad entre 0-2% frente a un exceso de velocidad mayor al 12% (Ayuso et al, 2014a), además a mayor es el porcentaje de kilómetros conducidos por encima de los límites (con efecto cuadrático positivo) más se reduce la distancia y el tiempo al primer accidente con culpa para ellos (Ayuso et al, 2016a) y (Ayuso et al, 2016b).

Los conductores con siniestros presentan un mayor porcentaje de kilómetros recorridos con exceso de velocidad, con el aumento del porcentaje de kilómetros con exceso de velocidad (con efecto cuadrático negativo) se espera un mayor número de siniestros, tanto teniendo en cuenta todos los siniestros como solo los siniestros con culpa

(Ayuso et al, 2018) y (Guillén et al, 2019). Además, en (Guillén et al, 2019) se especifica que es uno de los tres factores con más importancia para determinar el riesgo de siniestro.

El porcentaje de conducción con exceso de velocidad es superior en aquellos asegurados con algún siniestro con culpa respecto los que no tienen siniestros. Este factor de riesgo es el más importante para la predicción de siniestro con culpa según el método XGBoost (*TreeBooster*) (Pesantez-Narvaez et al, 2019). La velocidad es la variable que tiene una mayor influencia en el riesgo de conducción (entendido como frenados y aceleraciones bruscas), cuanto más elevada, el riesgo es más elevado, el conductor debe acelerar más, las RPM (media de revoluciones por minuto) son más elevadas y es más necesario hacer frenadas (Sun et al, 2020).

Un mayor porcentaje de kilómetros conducidos con exceso de velocidad llevan a un mayor número esperado de casi-accidentes de aceleración. Siendo este factor de riesgo telemático, el único relevante en este tipo de casi-accidentes, el parámetro de la regresión binomial negativa asociado a esta variable deja ver la trascendencia que tiene el exceso de velocidad en este tipo de casi-accidentes, siendo el parámetro de 5,63, con un aumento del 1% del porcentaje de kilómetros conducidos con exceso de velocidad se espera que el número de casi-accidentes de aceleración aumente un 6% aproximadamente (Guillén et al, 2020).

A lo largo de la revisión bibliográfica se muestra que, un mayor porcentaje de conducción con exceso de velocidad implica un mayor riesgo, el exceso de velocidad es relevante en la explicación y predicción de la ocurrencia de siniestro, la distancia y tiempo al primer siniestro (con diferencias entre géneros al separar las muestras), el número de siniestros, así como los casi-accidente.

II. APLICACIÓN PRÁCTICA

En este segundo bloque, de aplicación práctica, se hará uso de datos simulados de asegurados con pólizas *Pay-As-You-Drive* con la finalidad de, por un lado, ver qué efecto tienen las variables tradicionales y telemáticas en la frecuencia de siniestro, y, por otro lado, identificar la relación entre el kilometraje y la frecuencia de siniestros. En el anexo se puede encontrar el código utilizado en R.

5. Modelos lineales generalizados

Los modelos lineales generalizados son ampliamente utilizados por las empresas aseguradoras, mediante ellos se puede explicar y predecir variables muy importantes, como el número de siniestros la ocurrencia de siniestros o el coste de siniestros.

Los modelos lineales generalizados están compuestos por tres componentes:

- Componente aleatoria, que identifica la variable respuesta (dependiente) aleatoria Y que se distribuye según una ley de probabilidad.

Tabla 1. Clasificación modelos de regresión según la variable dependiente

Variable dependiente	Modelo de regresión
Cuantitativa continua	Lineal múltiple, Supervivencia, Tobit
Cuantitativa continua positiva	Gamma, Lognormal
Cuantitativa discreta	Poisson, Poisson mixto, Binomial Negativa
Cuantitativa temporal	Cox
Cualitativa dicotómica	Logit, probit
Cualitativa ordinal	Ordinal
Cualitativa nominal	Multinomial

Fuente: *Elaboración propia*

- Componente sistemática o predictor lineal (η), formada por la combinación de los factores de riesgo (X) y sus parámetros (β):

$$\eta_i = \sum_{j=1}^k X_{ij}\beta_j = \beta_1 + \beta_2 X_{i2} + \dots + \beta_k X_{ik} \quad \forall_i = 1, \dots, n$$

- Función de ligadura, enlace o link $g(\cdot)$, que se encarga de relacionar la esperanza de la variable respuesta (dependiente) con el predictor lineal:

$$\eta_i = g(\mu_i) \quad , \quad \forall_i = 1, \dots, n$$

Esta función enlace es diferente según el modelo GLM utilizado y es esencial para poder realizar una correcta interpretación de los coeficientes estimados:

- Al estimar un modelo de regresión lineal, podemos ver e interpretar de forma directa los coeficientes de la estimación, ya que existe una relación lineal entre

la combinación lineal ($x'\beta$) y el valor esperado de la variable dependiente. Al realizar las derivadas parciales de la ecuación ($y = (x'\beta + \varepsilon)$) lo que obtenemos es que, los efectos marginales son iguales al parámetro que acompaña a cada variable:

$$\frac{\partial y}{\partial x_j} = \beta_j$$

En este caso particular la función enlace $g(\cdot)$ es la identidad: $\eta_i = \mu_i$

- A diferencia del modelo de regresión lineal, en el resto de los modelos GLM que no son lineales sus coeficientes estimados no se pueden interpretar directamente, en primera instancia tan solo podemos valorar su signo, esto sucede porque, aunque existe una combinación lineal entre las variables y los coeficientes, esta combinación lineal, no se relaciona con el valor esperado de la variable respuesta linealmente:

$$\frac{\partial h(x\beta)}{\partial x_j} = \beta_j h'(x\beta)$$

Siendo $h(\cdot)$ la inversa de la función enlace

Así cada modelo tiene su función de enlace, a continuación, se detallan algunas de ellas:

Tabla 2. Funciones enlace según modelo

Modelo	Función enlace	$\eta_i = h'(\mu_i)$ $= g(\mu_i)$
Supervivencia	Logaritmo	$\log(\mu_i)$
Poisson	Logaritmo	$\log(\mu_i)$
Logit	Inversa distribución logística	$\log\left(\frac{\mu_i}{1 - \mu_i}\right)$
Probit	Inversa distribución normal	$\Phi^{-1}(\mu_i)$
Gamma	Inversa	μ_i^{-1}

Fuente: Elaboración propia

La estimación de los parámetros más utilizada en los modelos lineales generalizados es la estimación por máxima verosimilitud (MV), con la que se encuentran aquellos coeficientes estimados que aumentan la probabilidad de replicar la muestra original.

Para la formulación de este apartado han sido utilizados los manuales:

Curso de Econometría Actuarial y Análisis de la supervivencia en seguros, Modelos Lineales Generalizados (Ayuso, M & Bolancé, C, Curso 2020/2021).

Asignatura Modelos estadísticos aplicados del Máster de Ciencias Actariales y Financieras, Modelos lineales generalizados (tarificación a priori). (Alemany, R; Ayuso, M. & Bolancé, C, Curso 2020/2021).

Como veremos a continuación, la variable respuesta que se utilizará en las estimaciones es una variable cuantitativa discreta no negativa, por consiguiente, el método de regresión que utilizaremos para explicar y predecir esta variable será el de Poisson.

6. Datos

La base de datos utilizada, “datostelemat.csv”, contiene datos telemáticos ficticios de conductores con pólizas PAYD anuales. Aunque hubiese sido deseable trabajar con datos reales, estos datos han tenido de ser simulados por mi tutora, Ana María Pérez-Marín, debido a la imposibilidad de obtener datos con variables telemáticas reales.

La base de datos de estudio está formada por 9.225 pólizas y 6 variables. A continuación, se detallan las variables junto su descripción:

Tabla 3. Descripción de las variables de la base de datos

Variable	Descripción
Sex	Género del asegurado
Excess	Excesos de velocidad por encima del límite permitido: No=excesos en promedio por debajo del 5% de kilómetros recorridos Sí=excesos en promedio por encima del 5% de kilómetros recorridos
Urban	Conducción por vías urbanas: No=conducción por vía urbana por debajo del 25% de kilómetros recorridos Sí=conducción por vía urbana por encima del 25% de kilómetros recorridos
Km	Kilometraje anual
Y	Número de siniestros durante el último año
Antigc	Antigüedad del carnet de conducir: 1=menos de dos años, 2=dos años o más

Fuente: Elaboración propia

Podemos visualizar en la Tabla 3, que disponemos de dos variables cuantitativas, el kilometraje (km) y el número de siniestros (y), también disponemos de cuatro variables cualitativas, el género (sex), el exceso de velocidad (excess), la conducción por vías urbanas (urban) y la antigüedad del carnet de conducir (antigc).

En la preparación de los datos, se ha revisado la base de datos en búsqueda de correcciones, inconsistencias o errores, y no se encontraron, tampoco se han encontrado valores *missings* en las variables.

Figura 1. Resumen de la distribución de las variables

sex	excess	urban	y	km	antigc
Mujer :4621	No:4585	No:5615	Min. :0.0000	Min. : 1210	1:4679
Hombre:4604	Sí:4640	Sí:3610	1st Qu.:0.0000	1st Qu.: 7105	2:4546
			Median :0.0000	Median :10496	
			Mean :0.1528	Mean :11971	
			3rd Qu.:0.0000	3rd Qu.:15547	
			Max. :4.0000	Max. :33950	

Fuente: Elaboración propia

En la Figura 1, se muestra una primera inspección de los datos, haciendo una estadística descriptiva básica de cada una de las variables descritas, con el fin de obtener más información inicial, la cual veremos en detalle en el apartado siguiente.

6.1. Análisis descriptivo

Variable exceso de velocidad

Figura 2. Distribución asegurados con exceso de velocidad

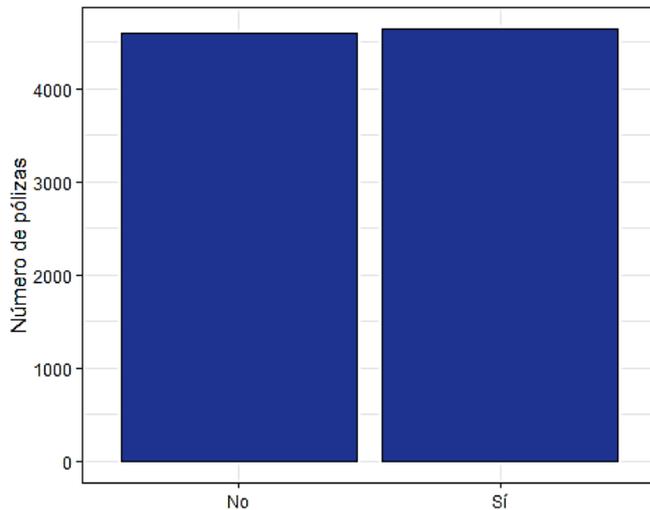


Tabla 4. Frecuencias exceso de velocidad

Excess	Frecuencia absoluta	Frecuencia relativa
No	4585	49,7%
Sí	4640	50,3%

Fuente: Elaboración propia

Fuente: Elaboración propia

Podemos observar en la Figura 2, como se distribuyen los asegurados según el grupo de exceso de velocidad al que pertenecen, vemos que el conjunto de los asegurados en el grupo Sí (excesos en promedio por encima del 5% de kilómetros recorridos) es mayor al grupo No (excesos en promedio por debajo del 5% de kilómetros recorridos). En la Tabla 4, podemos ver los valores exactos, precisamente el 50,3% de los asegurados se encuentran en el grupo Sí, y un 49,7% de los asegurados en el No.

Tabla 5. Ocurrencia de siniestro según exceso de velocidad

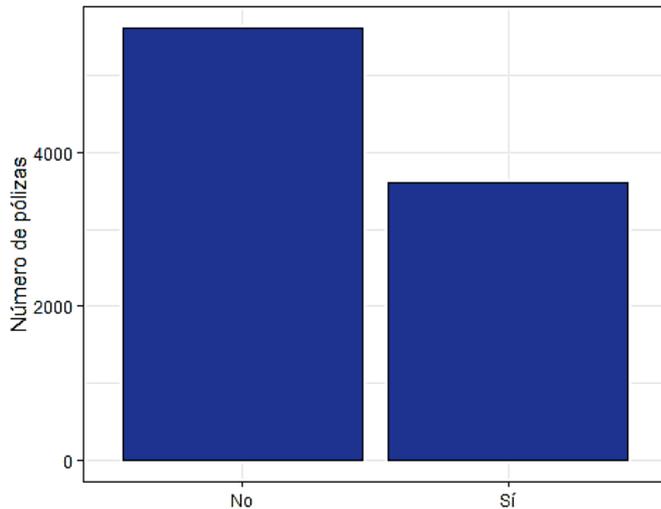
Ocurrencia de siniestro	No (N=4585)	Sí (N=4640)
Sin siniestro	3987 (87.0%)	3951 (85.2%)
Siniestro	598 (13.0%)	689 (14.8%)

Fuente: Elaboración propia

En la Tabla 5, podemos observar que de los asegurados que de promedio han tenido excesos por debajo del 5% de kilómetros recorridos (No, en la tabla), el porcentaje de los que han tenido siniestros ha sido de un 13.0%, menor al que presentan los asegurados con excesos en promedio por encima del 5% de kilómetros recorridos (Sí, en la Tabla) que han tenido siniestros un 14,8% de ellos.

Variable conducción urbana

Figura 3. Distribución asegurados conducción urbana



Fuente: Elaboración propia

Tabla 6. Frecuencias conducción urbana

Urban	Frecuencia absoluta	Frecuencia relativa
No	5615	60,9%
Sí	3610	39,1%

Fuente: Elaboración propia

Podemos observar en la Figura 3, como se distribuyen los asegurados según el grupo de conducción urbana al que pertenecen, vemos que el conjunto de los asegurados en el grupo No (conducción por vía urbana por debajo del 25% de kilómetros recorridos) es mayor al grupo Sí (conducción por vía urbana por encima del 25% de kilómetros recorridos). En la Tabla 6, podemos ver los valores exactos, precisamente el 60,9% de los asegurados se encuentran en el grupo No, y un 39,1% de los asegurados en el No.

Tabla 7. Ocurrencia de siniestro según conducción urbana

Ocurrencia de siniestro	No (N=5615)	Sí (N=3610)
Sin siniestro	4869 (86.7%)	3069 (85.0%)
Siniestro	746 (13.3%)	541 (15.0%)

Fuente: Elaboración propia

En la Tabla 7, podemos observar que de los asegurados que han tenido conducción por vías urbanas por debajo del 25% de kilómetros recorridos (No, en la tabla), el porcentaje de los que han tenido siniestros ha sido de un 13.3%, menor al que presentan los asegurados con una conducción por vía urbana por encima del 25% de kilómetros recorridos (Sí, en la Tabla) que han sido un 15.0% de ellos.

Variable antigüedad de carnet

Figura 4. Distribución asegurados antigüedad carnet

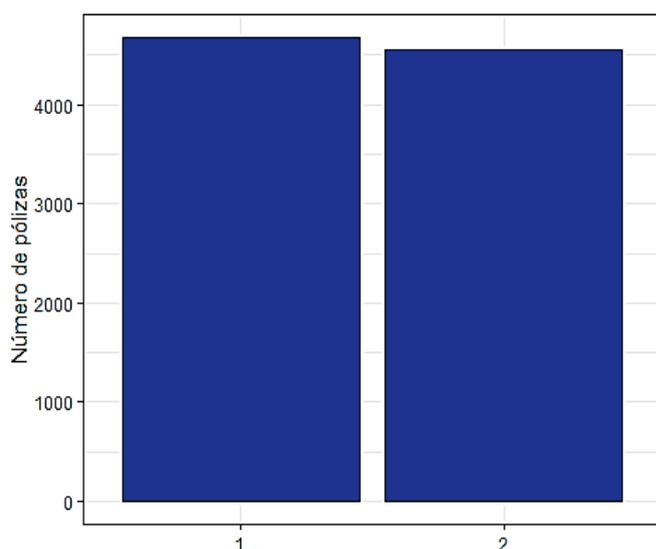


Tabla 8. Frecuencias antigüedad carnet

Antigüedad	Frecuencia absoluta	Frecuencia relativa
1	4679	50,7%
2	4546	49,3%

Fuente: Elaboración propia

Fuente: Elaboración propia

Podemos observar en la Figura 4 como se distribuyen los asegurados según el grupo de antigüedad de carnet al que pertenecen, vemos que el conjunto de los asegurados en el grupo 1 (antigüedad de carnet de menos de dos años) es mayor al grupo 2 (antigüedad de carnet de dos años o más). En la Tabla 8, podemos ver los valores exactos, precisamente el 50,7% de los asegurados se encuentran en el grupo 1, y un 49,3% de los asegurados en el 2.

Tabla 9. Ocurrencia de siniestro según antigüedad de carnet

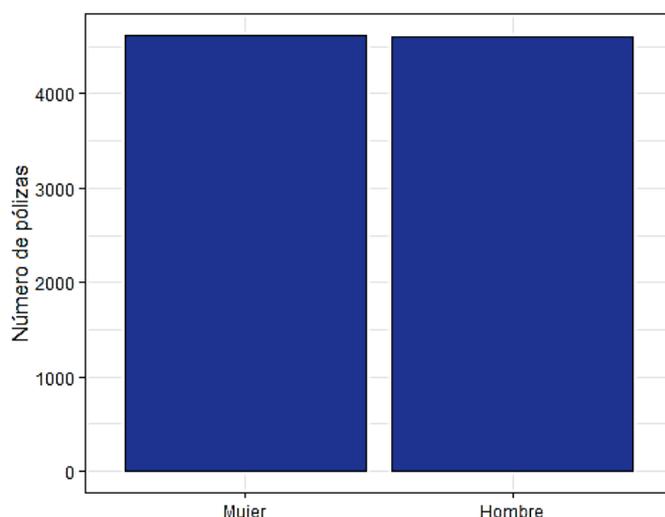
Ocurrencia de siniestro	1 (N=4679)	2 (N=4546)
Sin siniestro	3992 (85.3%)	3946 (86.8%)
Siniestro	687 (14.7%)	600 (13.2%)

Fuente: Elaboración propia

En la Tabla 9, podemos observar que de los asegurados con una antigüedad de carnet de menos de dos años (1, en la tabla), el porcentaje con siniestro ha sido de un 14.7%, mayor al que presentan los asegurados con antigüedad de carnet de dos años o más (2, en la Tabla) que es del 13,2% de ellos.

Variable género

Figura 5. Frecuencia absoluta género del asegurado



Fuente: Elaboración propia

Tabla 10. Frecuencias género

Sex	Frecuencia absoluta	Frecuencia relativa
Mujer	4621	50,1%
Hombre	4604	49,9%

Fuente: Elaboración propia

Podemos observar en la Figura 5, como se distribuyen los asegurados según su género, así vemos que los datos son muy balanceados en base al género, aunque muy ligeramente hay un mayor porcentaje de mujeres. En la Tabla 10, podemos ver los valores exactos, precisamente el 50,1% de los asegurados son mujeres, y un 49,9% de los asegurados son hombres.

Tabla 11. Ocurrencia de siniestro según género

Ocurrencia de siniestro	Mujer (N=4621)	Hombre (N=4604)
Sin siniestro	4033 (87.3%)	3905 (84.8%)
Siniestro	588 (12.7%)	699 (15.2%)

Fuente: Elaboración propia

En la Tabla 11, podemos observar que en las mujeres el porcentaje que ha tenido siniestro ha sido de un 12.7%, menor al de los hombres que es del 15,2%.

Tabla 12. Análisis descriptivo según el género

	Mujer (N=4621)	Hombre (N=4604)
Urban		
No	2811 (60.8%)	2804 (60.9%)
Sí	1810 (39.2%)	1800 (39.1%)
Y		
Mean (SD)	0.139 (0.380)	0.167 (0.414)
Median [Min, Max]	0 [0, 4.00]	0 [0, 3.00]
Km		
Mean (SD)	11700 (6490)	12300 (6450)
Median [Min, Max]	10200 [1210, 33900]	10800 [1630, 34000]

Antigc		
1	2316 (50.1%)	2363 (51.3%)
2	2305 (49.9%)	2241 (48.7%)
Excess		
No	2748 (59.5%)	1837 (39.9%)
Sí	1873 (40.5%)	2767 (60.1%)

Fuente: Elaboración propia

La Tabla 12, nos permite ver las diferencias entre hombres y mujeres, referentes a su experiencia, comportamiento de conducción, así como el uso del vehículo. Si analizamos, primeramente, la experiencia de conducción (antigc), vemos que los hombres se muestran ligeramente con menor experiencia, un 48,7% de los hombres tienen una antigüedad de carnet de dos años o más, mientras que para las mujeres es del 49,9%.

Referente al exceso de velocidad (excess) en la conducción vemos que, hay un mayor porcentaje de hombres que han tenido excesos en promedio por encima del 5% de kilómetros recorridos, concretamente un 60,1% de los hombres, mientras que en el caso de las mujeres es un 40,5% del total de mujeres. En la conducción urbana (urban) vemos que los porcentajes son muy similares, tan solo hay una diferencia de un 0,1 punto porcentual, las mujeres que tienen una conducción por vía urbana por encima del 25% de kilómetros recorridos es del 39,2% y los hombres un 39,1%.

En el número de siniestros (y) los hombres presentan un número medio de siniestros mayor que las mujeres (0,167 y 0,139, respectivamente), aunque sí que hay que comentar que en las mujeres se llega hasta los 4 siniestros. Finalmente, en el kilometraje (km) vemos que de media los hombres han conducido más kilómetros anuales exactamente 12.300 km/anuales, mientras que las mujeres han conducido 11.700 km/anuales, también vemos que el máximo de conducción de los hombres es mayor se llega a los 34.000 km/anuales.

Variable número de siniestros

El número de siniestros será nuestra variable dependiente, objeto de análisis y predicción en este presente trabajo. A continuación, vamos a analizar en detalle esta variable:

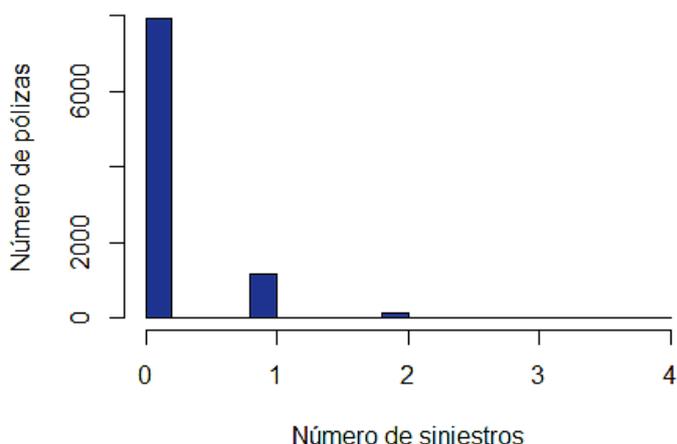
En la Figura 6, se muestran las distintas medidas descriptivas de la variable número de siniestros (y), podemos observar que, del cuartil 1 al cuartil 3 el valor de la variable es 0, lo que nos indica que al menos un 75% de los asegurados no han tenido siniestros, también podemos ver que el máximo de siniestros que han tenido los asegurados ha sido de 4 siniestros.

Si nos fijamos en la asimetría vemos que, esta presenta un valor de 2,68, por lo que la distribución será asimétrica hacia la derecha (positiva), por último, comentar el coeficiente de curtosis, siendo el valor de 10,59 nos indica que es una distribución leptocúrtica.

Figura 6. Medidas descriptivas número de siniestros

	y
media	0.15
mediana	0.00
var	0.16
desv.est	0.40
mín	0.00
Q1	0.00
Q2	0.00
Q3	0.00
máx	4.00
rango	4.00
IC	0.00
asimetría	2.68
curtosis	10.59
coeficiente_variación	2.60

Fuente: Elaboración propia

Figura 7. Histograma número de siniestros

Fuente: Elaboración propia

Podemos observar en la Figura 7, como se distribuyen los asegurados según el número de siniestros, así, vemos que la gran mayoría de asegurados de forma muy destacada han tenido cero siniestros, y los que han tenido siniestros en su mayoría han tenido solo un siniestro. En la Tabla 13 podemos ver los valores exactos, en concreto el 86% no han tenido siniestros, del 14% restante el 12,7% han tenido solo un siniestro.

Tabla 13. Frecuencias número de siniestros

Y	Frecuencia absoluta	Frecuencia relativa
0	7938	(86.05%)
1	1171	(12.69%)
2	110	(1.19%)
3	5.00	(0.05%)
4	1.00	(0.01%)

Fuente: Elaboración propia

Tabla 14. Número de siniestros según exceso de velocidad

Y	No (N=4585)	Sí (N=4640)
0	3987 (86.96%)	3951 (85.15%)
1	554 (12.08%)	617 (13.30%)
2	42 (0.92%)	68 (1.47%)
3	1 (0.02%)	4 (0.09%)
4	1 (0.02%)	0 (0.00%)

Fuente: Elaboración propia

En la Tabla 14, podemos ver como se distribuye el número de siniestros según los grupos de excesos de velocidad, vemos que la distribución es similar entre los asegurados que de promedio han tenido excesos por debajo del 5% de kilómetros recorridos (No, en la tabla) y los asegurados con excesos en promedio por encima del 5% de kilómetros recorridos (Sí, en la Tabla), a mayor número de siniestro menor el porcentaje. Si nos fijamos en los asegurados que han tenido siniestros, hay un porcentaje más elevado de asegurados con uno, dos y tres siniestros entre aquellos que tienen un mayor porcentaje de kilómetros con exceso de velocidad (Sí, en la Tabla).

Tabla 15. Número de siniestros según conducción urbana

Y	No (N=5615)	Sí (N=3610)
0	4869 (86.71%)	3069 (85.01%)
1	677 (12.06%)	494 (13.68%)
2	66 (1.18%)	44 (1.22%)
3	3 (0.05%)	2 (0.06%)
4	0 (0.00%)	1 (0.03%)

Fuente: Elaboración propia

En la Tabla 15, podemos ver como se distribuye el número de siniestros según los grupos de conducción por vías urbanas, vemos que la distribución es similar entre los asegurados con conducción por vía urbana por debajo del 25% de kilómetros recorridos (No, en la tabla) y los con conducción por vía urbana por encima del 25% de kilómetros recorridos (Sí, en la Tabla), a mayor número de siniestro menor el porcentaje. Si nos fijamos en los asegurados que han tenido siniestros, hay un porcentaje más elevado de asegurados con un siniestro entre aquellos que tienen una mayor conducción por vías urbanas (Sí, en la Tabla), en los de dos siniestros el porcentaje es más elevado en los de menor conducción.

Tabla 16. Número de siniestros según antigüedad de carnet

Y	1 (N=4679)	2 (N=4546)
0	3992 (85.32%)	3946 (86.80%)
1	623 (13.31%)	548 (12.05%)
2	59 (1.26%)	51 (1.12%)
3	5 (0.11%)	0 (0,00%)
4	0 (0,00%)	1 (0.02%)

Fuente: Elaboración propia

En la Tabla 16, podemos ver como se distribuye el número de siniestros según los grupos de antigüedad de carnet, vemos que la distribución es similar entre los asegurados con antigüedad de carnet de menos de dos años (No, en la tabla) y los con antigüedad de carnet de dos años o más (Sí, en la Tabla), a mayor número de siniestro menor el porcentaje, a excepción de los asegurados con tres o cuatro siniestros del grupo 2. Si nos fijamos en los asegurados que han tenido siniestros, hay un porcentaje más elevado de asegurados con uno, dos y tres siniestros entre aquellos que pertenecen al grupo 1.

Tabla 17. Número de siniestros según género

Y	Mujer (N=4621)	Hombre (N=4604)
0	4033 (87.28%)	3905 (84.82%)
1	538 (11.64%)	633 (13.75%)
2	48.0 (1.04%)	62 (1.35%)
3	1.00 (0.02%)	4 (0.09%)
4	1.00 (0.02%)	0 (0.00%)

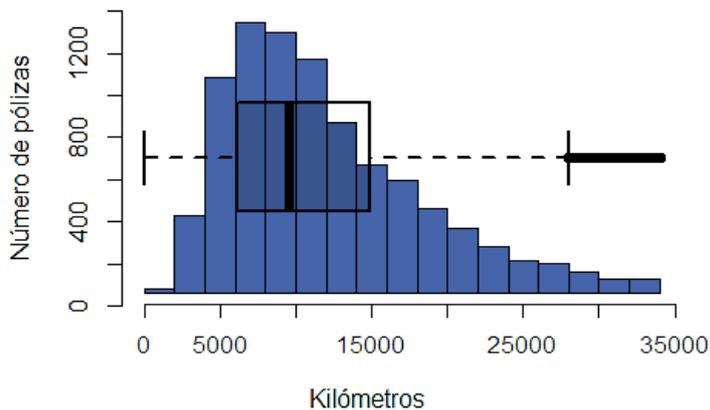
Fuente: Elaboración propia

En la Tabla 17, podemos ver como se distribuye el número de siniestros según el género, y vemos que esta distribución es similar a mayor número de siniestro menor el porcentaje. Si nos fijamos en los asegurados que han tenido siniestros, hay un porcentaje más elevado de asegurados con uno, dos y tres siniestros entre los hombres.

Variable Kilometraje

Un objetivo importante de esta aplicación práctica consiste en determinar qué relación existe entre el kilometraje y la frecuencia de siniestros. A continuación, vamos a analizar en detalle esta variable:

Figura 8. Histograma y Boxplot Kilometraje



Fuente: Elaboración propia

Figura 9. Medidas descriptivas kilometraje

	km
media	11971.39
mediana	10496.00
var	41973465.68
desv.est	6478.69
mín	1210.00
Q1	7105.00
Q2	10496.00
Q3	15547.00
máx	33950.00
rango	32740.00
IC	8442.00
asimetría	1.00
curtosis	3.62
coeficiente_variación	0.54

Fuente: Elaboración propia

En la Figura 8, se muestra el histograma y Boxplot del kilometraje, podemos complementar su interpretación con la Figura 9, que muestra las distintas medidas descriptivas de esta variable. Observamos que, el valor del primer cuartil, aquel que recoge el 25% de las observaciones es de 7.105 km/anales, y el tercer cuartil es de 15.547 km/anales, así la mayoría de los asegurados presentan un kilometraje menor a los 15.000 km/anales. Los valores atípicos se encuentran representados en la Figura 8 como puntos negros fuera de los “bigotes” de la caja, vemos que hay un elevado número y se han mantenido la totalidad de ellos.

Si nos fijamos en la asimetría vemos que, esta presenta un valor de 1, por lo que la distribución es asimétrica hacia la derecha (positiva), por último, comentar el coeficiente de curtosis, siendo el valor de 3,62 nos indica que es una distribución leptocúrtica, tiene una gran concentración de los valores en torno la mediana que es de 10.496 (la línea negra vertical, dibujada dentro de la caja del Boxplot).

Tabla 18. Kilometraje según número de siniestros

Km	0 (N=7938)	1 (N=1171)	2 (N=110)	3 (N=5)	4 (N=1)
Mean (SD)	11800 (6400)	12900 (6770)	14700 (7680)	16300 (7540)	12300 (NA)
Median	10400	11300	12900	13900	12300
[Min, Max]	[1210, 34000]	[2130, 33800]	[3720, 33900]	[8450, 27200]	[12300, 12300]

Fuente: Elaboración propia

La Tabla 18, nos ofrece una primera información sobre la relación entre el kilometraje y el número de siniestros. Si nos fijamos en la media del kilometraje vemos que esta

aumenta a medida que aumenta el número de siniestros, a excepción del único asegurado con cuatro siniestros, que presenta un menor kilometraje. Si nos fijamos en la mediana que es más robusta que la media y no le afectan los valores extremos, vemos la misma dinámica que en la media, también vemos que el kilometraje en el caso de cuatro siniestros es menor a la mediana de los asegurados que presentan dos siniestros.

Por último, mirando los valores mínimos y máximos del kilometraje por número de siniestros, encontramos que si bien los que no han tenido siniestro presentan el menor valor de kilometraje, también se encuentra el asegurado con la mayor distancia recorrida anualmente. Además, si nos fijamos en los que han tenido tres siniestros, el asegurado con máximo kilometraje (27.200 km/anuales) es menor al máximo de los siniestros 1 (33.800 km/anuales) y 2 (33.900 km/anuales), estos números nos podrían dejar ver que no necesariamente por tener el doble de distancia recorrida se vaya a tener el doble de siniestros.

7. Estimación y resultados

7.1. Modelo de regresión de Poisson

Como se ha comentado anteriormente, la variable que queremos explicar es el número de siniestros que se han tenido durante el último año (Y_i), esta sigue una distribución de Poisson con parámetro λ_i :

$$Y_i \approx \text{Poisson}(\lambda_i) \quad \forall i = 1, \dots, n$$

Aplicando la función de enlace del modelo de regresión de Poisson, el logaritmo, tenemos que:

$$E(Y_i) = \lambda_i = \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik})$$

Donde λ_i es el número esperado de siniestros para cada asegurado, que dependerá de sus características de riesgo.

Siendo el cálculo de las probabilidades:

$$\Pr(Y_i = k) = \frac{e^{-\lambda_i} \lambda_i^k}{k!}$$

7.2. Modelo de regresión de Poisson con *offset* clásico

A continuación, se especifica la formulación del modelo de Poisson con *offset* clásico, extraída del artículo “*Pay-as-you-drive insurance: the effect of the kilometers on the risk of accident*” de (Boucher et al, 2013):

$$\lambda_i = t_i \times \exp(x_i' \beta) = \exp(x_i' \beta + \log(t_i)) \quad (1)$$

Donde t_i es el tiempo en el que el conductor de la póliza i ha estado asegurado $i = 1, \dots, I$

De manera tradicional, se ha utilizado como medida de exposición al riesgo del conductor, el tiempo en el que el conductor está asegurado (t_i). La forma de introducir esta exposición en el modelo de Poisson ha sido multiplicando el valor esperado de siniestros por la exposición, así en la estimación del modelo se introduce como variable *offset* la exposición. Con esta especificación, se espera que un conductor que solo va a estar medio

año asegurado tenga la mitad de siniestro respecto de un conductor que ha estado todo el año asegurado, presentando las demás características similares.

De la ecuación (1) pasamos a la ecuación (2), midiendo la exposición al riesgo como kilómetros recorridos:

$$\lambda_i = Km_i \times \exp(x'_i\beta) = \exp(x'_i\beta + \log(Km_i)) \quad (2)$$

Donde Km_i es el número de kilómetros conducidos por la póliza i , $i = 1, \dots, I$

Con esta especificación, se espera que un conductor que recorra la mitad de los kilómetros respecto otro conductor, presentando unos valores similares en el resto de los factores de riesgo, tenga la mitad de siniestro que el segundo conductor.

Si aplicamos esta interpretación (el doble de kilómetros, el doble de riesgo) a la prima, tenemos que, “Si alguien conduce 10.000 km en un año, debería espera tener una prima que es el 50% más barata que un conductor similar que conduce 20.000 km” (Boucher et al, 2013).

Con el modelo de Poisson con *offset* clásico lo que estamos especificando es una relación directa y lineal entre el kilometraje y la frecuencia de siniestros, así, con un aumento del kilometraje tenemos un aumento de la frecuencia de los siniestros proporcional.

A continuación, se muestra en la figura 9 la estimación del modelo por máxima verosimilitud, realizada con R, en el que la **variable dependiente** es el número de siniestros (y), las **variables independientes** son el género (sex), la conducción por vías urbanas ($urban$), la conducción con exceso de velocidad ($excess$) y la antigüedad del vehículo ($antigc$), y la **variable offset** es el logaritmo de los kilómetros anuales ($\log km$):

Figura 10. Estimación modelo de regresión de Poisson con *offset* clásico

```
Call:
glm(formula = y ~ sex + urban + excess + antigc, family = poisson(link = "log"),
     data = datos, offset = logkm)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.0480  -0.6043  -0.4800  -0.3641   4.3747

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -11.38656   0.05714 -199.286 <2e-16 ***
sexHombre    0.10744   0.05447   1.972  0.0486 *
urbanSi      0.12118   0.05397   2.245  0.0248 *
excessSi     0.13392   0.05444   2.460  0.0139 *
antigc2     -0.12127   0.05342  -2.270  0.0232 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 5748.1 on 9224 degrees of freedom
Residual deviance: 5725.7 on 9220 degrees of freedom
AIC: 8383.4

Number of Fisher Scoring iterations: 6
```

Estimación por máxima verosimilitud. Categorías de referencia: sex (mujer), urban (No), excess (No), antigc (1).
Fuente: Elaboración propia

En la figura 9, podemos observar que todos los coeficientes se muestran significativos al 5% (el p -value es menor al 0,05, se rechaza la hipótesis nula: $\beta_k=0$).

La interpretación de los coeficientes se realizará una vez hecha la comparativa de los dos modelos, y escogido el que tiene un mejor ajuste.

7.3. Modelo de regresión de Poisson con *offset* generalizado

A continuación, se especifica la formulación del modelo de Poisson con *offset* generalizado, extraída del artículo “*Pay-as-you-drive insurance: the effect of the kilometers on the risk of accident*” de (Boucher et al, 2013):

$$\lambda_i = Km_i \times \exp(x'_i\beta) = Km \times \exp(x'_i\beta + \log(Km_i)) \quad (3)$$

Generalizando el modelo de Poisson con *offset*, se puede expresar como:

$$\lambda_i = \exp(x'_i\beta + c \times \log(Km)) = \exp(x'_i\beta) \times Km^c \quad (4)$$

$$\begin{cases} c = 1 \rightarrow \exp(x'_i\beta) \times Km \\ c \neq 1 \rightarrow \exp(x'_i\beta) \times Km^c \end{cases}$$

Cuando el parámetro *c* es igual a 1, el modelo converge al modelo de regresión de Poisson con *offset* clásico. Si el parámetro *c* es diferente a 1, el modelo de generalización es un caso particular del modelo de regresión de Poisson en el que se introduce el logaritmo de la variable que recoge el kilometraje como una variable independiente más.

A continuación, se muestra en la Figura 10 la estimación del modelo por máxima verosimilitud, realizada con R, en el que la **variable dependiente** es el número de siniestros (*y*), y las **variables independientes** son el género (*sex*), la conducción por vías urbanas (*urban*), la conducción con exceso de velocidad (*excess*), la antigüedad del vehículo (*antigc*), y el logaritmo de los kilómetros anuales (*logkm*):

Figura 11. Estimación modelo de regresión de Poisson con *offset* generalizado

```
Call:
glm(formula = y ~ sex + urban + excess + antigc + logkm, family = poisson,
     data = datos)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.7691  -0.5829  -0.5302  -0.4716   4.3698

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -5.36474    0.46648  -11.500  < 2e-16 ***
sexHombre     0.13698    0.05459   2.510   0.0121 *
urbanSí       0.12000    0.05397   2.224   0.0262 *
excessSí      0.13004    0.05448   2.387   0.0170 *
antigc2      -0.11635    0.05341  -2.178   0.0294 *
logkm         0.36063    0.04964   7.264  3.75e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 5645.9  on 9224  degrees of freedom
Residual deviance: 5565.3  on 9219  degrees of freedom
AIC: 8225

Number of Fisher Scoring iterations: 6
```

Estimación por máxima verosimilitud. Categorías de referencia: *sex* (mujer), *urban* (No), *excess* (No), *antigc* (1).
Fuente: Elaboración propia

En la Figura 11, podemos observar que todos los coeficientes que acompañan las variables independientes se muestran significativos al 5% (el *p-value* es menor al 0,05, se rechaza la hipótesis nula: $\beta_k=0$), en el caso de la variable logkm, el coeficiente llega a ser significativo al 0,1%.

7.4.Comparativa de modelos

Comparativa bondad del ajuste

Antes de proceder con la comparativa de bondad del ajuste, se comprueba la hipótesis del modelo, que en el caso de la regresión de Poisson es que hay una igualdad entre la media y la varianza de la variable dependiente. Los valores obtenidos de estas dos medidas para el número de siniestros son 0,1528455 y 0,1579020 respectivamente, la igualdad no es total pero no parece que haya mucha sobre dispersión. Realizando un test de sobre dispersión para los dos modelos estimados, se muestra una dispersión de 1,045049 para el modelo de regresión de Poisson con *offset* clásico y de 1,023435 para el modelo de Poisson con *offset* generalizado.

En la siguiente tabla veremos qué modelo tiene mejor ajuste:

Tabla 19. Comparativa bondad del ajuste entre modelos de Poisson

Modulo	AIC	BIC	<i>Deviance</i>	Pearson
<i>Offset</i> clásico	8383.431	8419.080	5725.699	1.240851
<i>Offset</i> generalizado	8225.039	8267.817	5565.306	1.022850

Fuente: Elaboración propia

La Tabla 19, recoge cuatro medidas de bondad del ajuste para los dos modelos estimados, para las cuatro medidas cuanto menor sea el valor mejor será el ajuste.

Podemos observar que las dos primeras medidas, el Criterio de Información de Akaike (AIC) y el Criterio de Información Bayesiano (BIC), presentan ambas un valor menor en el modelo de Poisson con *offset* generalizado, por lo tanto, presenta un mejor ajuste.

La tercera medida utilizada para comparar la bondad del ajuste es la Devianza (*Deviance*), esta es una medida específica para los modelos generalizados, vemos que los valores vuelven a ser menores para el modelo de Poisson con *offset* generalizado. Hay que apuntar que adicionalmente, se ha hecho una comprobación de si la devianza se encuentra por debajo del valor crítico al 5% para una Chi-Cuadrado, para ver si hay un buen ajuste, y en ambos modelos el *p-value* es mayor al 5%, por lo que no se rechaza la hipótesis nula de un buen ajuste.

Por último, con la cuarta medida podemos ver cómo afecta la sobre dispersión de cada modelo. Para el modelo de Poisson con *offset* clásico se observa que la varianza es un 24,09% mayor que la media, mientras que para el modelo de Poisson con *offset* generalizado es mucho menor un 2,29%, hay un mejor ajuste en el segundo modelo.

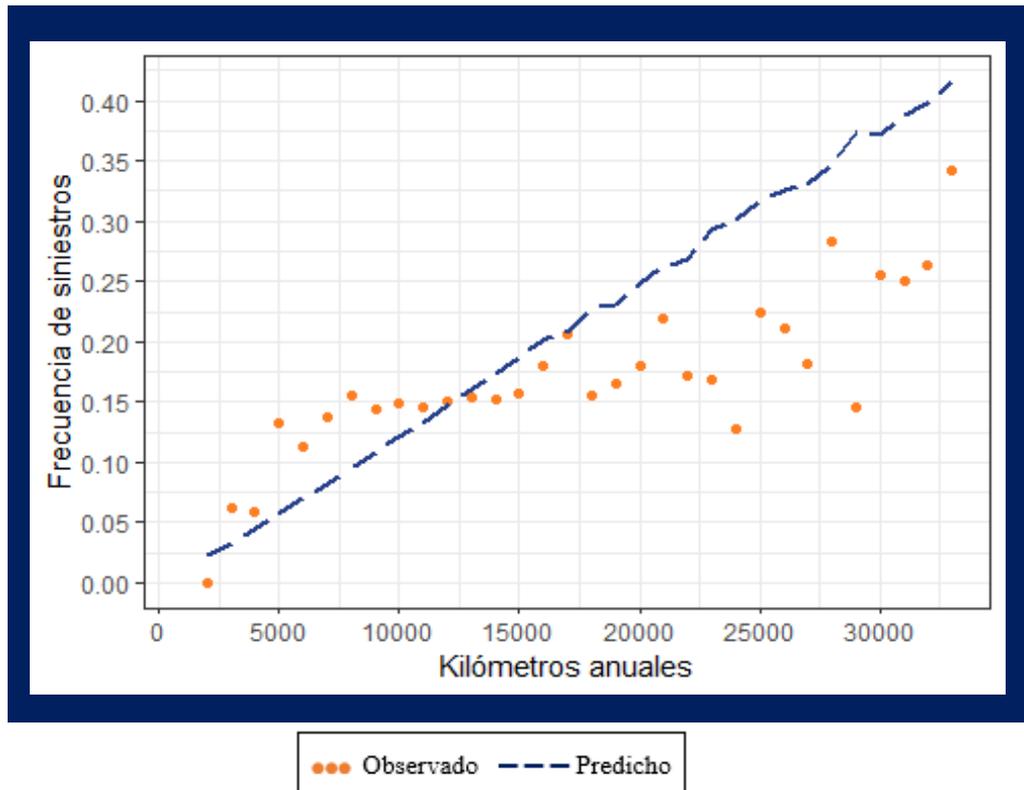
Seguidamente veremos reflejado el mejor ajuste obtenido con una comparativa grafica d ellos valores observados con las predicciones de cada modelo.

Comparativa valores observados y predichos

Una vez estimados los coeficientes de los dos modelos, se ha llevado a cabo una predicción de la frecuencia de siniestros (λ_i) para todas las observaciones.

A continuación, veremos gráficamente la comparación de los valores observados y los predichos para los dos modelos, con el propósito de visualizar como es la relación entre el kilometraje y el número de siniestros, y como se ajustan los modelos estimados.

Figura 12. Valores observados y predichos con el modelo de regresión de Poisson con *offset* clásico



*Frecuencia media agrupando los asegurados en intervalos de 1000 km/anuales.
Fuente: Elaboración propia*

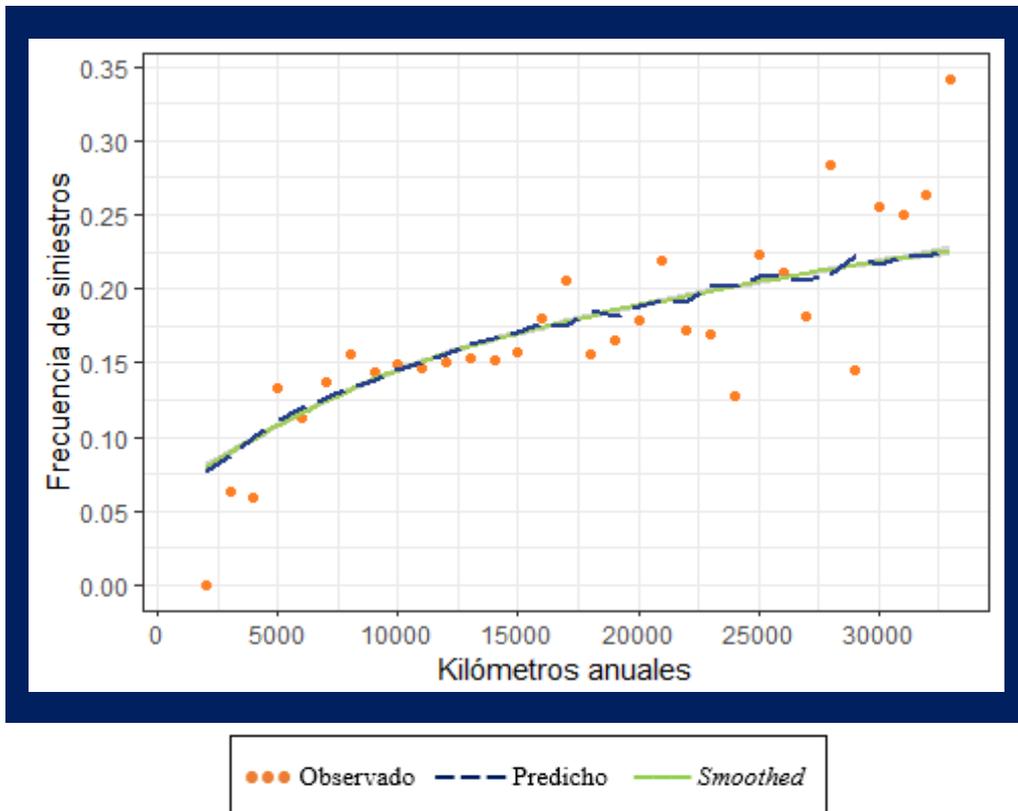
En la Figura 11, podemos ver que las frecuencias de siniestros observadas (puntos en naranja) no se ajustan bien a las predichas por el modelo de regresión de Poisson con *offset* clásico (línea discontinua en azul).

Las predicciones se han realizado con todos los factores de riesgo detallados en la estimación, por ello la línea no es perfectamente lineal, tiene en cuenta el efecto de las demás variables a parte del *offset* clásico, aun teniendo en cuenta todas las variables se ve claramente la relación que establece el modelo mediante la ecuación (2), a mayor kilometraje mayor es la frecuencia de siniestro de manera proporcional, manteniendo unas características similares.

Se puede observar claramente que la línea discontinua subestima las frecuencias de siniestros de los conductores con menos de 12.500, por el contrario, sobrestima la frecuencia de siniestros en aquellos conductores que presentan una conducción mayor a los 12.500 kilómetros anuales.

Por lo tanto, podríamos concluir que, en los datos utilizados no hay una relación lineal entre el kilometraje y el número esperado de siniestros tal y como se especifica en el modelo de regresión de Poisson con *offset* clásico.

Figura 13. Valores observados y predichos con el modelo de regresión de Poisson con *offset* generalizado



Frecuencia media agrupando los asegurados en intervalos de 1000 km/anuales.

Fuente: Elaboración propia

En la Figura 12, podemos ver que las frecuencias de siniestros observadas (puntos en naranja) tienen un buen ajuste en comparación a las predichas por el modelo de regresión de Poisson con *offset* generalizado (línea discontinua en azul), también podemos ver el buen ajuste en la línea suavizada de los valores predichos (línea continua en verde).

Como en el modelo anterior las predicciones se han realizado con todos los factores de riesgo detallados en la estimación, por ello el trazado de la línea discontinua no se muestra perfectamente regular, tiene en cuenta el efecto de las demás variables a parte del *offset* generalizado. Aun teniendo en cuenta todas las variables se ve claramente la relación que establece el modelo mediante la ecuación (4), a mayor kilometraje la frecuencia de siniestro aumenta, pero no proporcionalmente.

Podemos observar una pendiente más pronunciada cuando el kilometraje es menor, que se reduce a medida que se aumenta el kilometraje, y parece que en el tramo final vuelve a haber una mayor pendiente.

También se puede observar en el gráfico que, aproximadamente, cuando los kilómetros conducidos superan los 22.500, las frecuencias se muestran más heterogéneas entre

asegurados con kilometrajes cercanos respecto los que presentan un kilometraje inferior, esta mayor heterogeneidad se da porque no hay tantas observaciones en los kilometrajes más elevados, lo que dificulta el ajuste.

Una vez hecha la comparativa entre valores observados y predichos para el segundo modelo, podemos reafirmar la conclusión que se apuntaba en el primer modelo. En los datos de estudio no hay una relación lineal entre el kilometraje y el número esperado de siniestros tal y como se especifica en el modelo de regresión de Poisson con *offset* clásico, sino que el aumento de exposición al riesgo por un mayor kilometraje queda compensado por la experiencia de conducción y otros elementos que aportan una conducción más segura, como el uso de vehículos más seguros y nuevos, así como una mayor conducción por vías más seguras, aquellas que no son urbanas. La relación entre el kilometraje y el número de siniestros se recoge de forma más adecuada mediante el modelo de regresión de Poisson con *offset* generalizado.

Tabla 20. Interpretación de coeficientes significativos modelo de regresión de Poisson con *offset* generalizado

Variable	$\exp(\beta_k)$
sexHombre	1.1468078
urbanSí	1.1275010
excessSí	1.1388770
antigc2	0.8901624

No se incluye la variable logkm.

Fuente: Elaboración propia

En la Tabla 20, si nos fijamos primero en las variables tradicionales, vemos que, para el género, el número de siniestros esperados para los hombres es un 14,68% superior respecto las mujeres. En el caso de la antigüedad de carnet, para aquellos asegurados con más experiencia, dos años o más, se espera que el número de siniestros sea un 10,98% menor respecto los más principiantes con antigüedad de carnet menor a dos años.

En cuanto a las variables telemáticas, si nos fijamos en la variable de conducción urbana vemos que, el número medio de siniestros es un 12,75% mayor para los asegurados que conducen por encima del 25% de kilómetros por vías urbanas respecto los que conducen menos del 25% en estas vías. El exceso de velocidad muestra que, el número medio de siniestros esperados es un 13,89% mayor para los conductores con excesos en promedio por encima del 5% de kilómetros recorridos, respecto aquellos que presentan excesos en promedio por debajo del 5% de kilómetros recorridos.

Para realizar una completa interpretación del parámetro que acompaña la variable logaritmo del kilometraje, recuperamos la ecuación (4):

$$\lambda_i = \exp(x'_i\beta + c \times \log(Km)) = \exp(x'_i\beta) \times Km^c$$

Como se comentaba en la Figura 12, la relación entre el kilometraje y la frecuencia siniestral del modelo de regresión de Poisson con *offset* generalizado no es lineal, sino que el aumento de la exposición al riesgo por recorrer una mayor distancia queda

compensado por la experiencia de conducción y otros elementos que aportan una conducción más segura, esta compensación vendría recogida en el parámetro c .

Cuando el parámetro c es igual a 1 (modelo de regresión de Poisson con *offset* clásico) o tiende a 1, la relación entre el kilometraje y la frecuencia de siniestros es lineal, no hay una compensación del riesgo, por lo que para una persona que conduzca el doble de los kilómetros respecto otra, teniendo ambas personas características similares, se esperaría que la que conduce el doble de kilómetros tenga el doble de los siniestros, así como el riesgo es el doble, su prima también debería serlo.

En cambio, si el valor tiende a 0, el aumento del riesgo por un aumento de los kilómetros recorridos queda compensado en su totalidad por el efecto de la experiencia del conductor, un uso de vehículos más seguros, así como conducción por vías más seguras, entre otros posibles factores que aporten una mayor seguridad. En este caso, el número de siniestros esperados no aumenta al aumentar el kilometraje, por consiguiente, la prima que pagaría la persona que recorre el doble de kilómetros sería la misma que la que conduce la mitad de los kilómetros.

En la Figura 10, se puede observar que el parámetro c estimado tiene un valor de 0,36, bastante alejado del 1, por lo que habrá una elevada compensación del aumento del riesgo debido a un aumento del kilometraje. Siguiendo con el ejemplo de la persona que conduce el doble de kilómetros en la que se espera el doble de siniestros, al elevar el kilometraje al parámetro c , el efecto proporcional de esperar el doble de siniestros disminuye quedando en $2^{0,36}$, por lo que en vez de tener que pagar el doble de la prima se pagará un 1,28 de la prima del asegurado con la mitad de los kilómetros, lo que implica una gran reducción respecto a tener que pagar el doble.

Finalizada la aplicación práctica, se ha podido comprobar la existencia de un efecto en el riesgo siniestral tanto de los factores de riesgo telemáticos, como los factores tradicionales. Con un aumento de los kilómetros anuales, del exceso de velocidad, de la conducción por vías urbanas, una menor experiencia y si el asegurado es hombre, el riesgo es mayor. Además, se ha podido visualizar la no linealidad entre el kilometraje y la frecuencia siniestral.

8. Conclusiones

Uno de los principales objetivos de este trabajo era el de examinar el efecto de las variables telemáticas en el riesgo que puede aportar un conductor, a lo largo de la revisión bibliográfica, se ha podido verificar que los factores telemáticos tienen una gran influencia en la siniestralidad, y que la inclusión de estos en los modelos que buscan explicar la ocurrencia de siniestros, el número de siniestros o el tiempo y distancia al primer siniestro, consiguen un mejor ajuste incluyendo las variables telemáticas. Se ha podido comprobar que una mayor conducción por vías urbanas, con exceso de velocidad, una conducción nocturna y una mayor distancia recorrida implican un mayor riesgo. Por lo que se concluye que las variables telemáticas son muy relevantes para explicar y predecir la ocurrencia de siniestro, la distancia y tiempo al primer siniestro, el número de siniestros, en consecuencia, en la medida que fuese posible tendrían de ser factores tenidos en cuenta en la tarificación.

Otro de los objetivos principales de este trabajo, era realizar una modelización del número de siniestros con variables telemáticas y el analizar cuál era la relación del kilometraje respecto la frecuencia siniestral. Los resultados que se han obtenido en la aplicación práctica están en consonancia con los hallazgos de la revisión bibliográfica, una mayor conducción con exceso de velocidad y por vías urbanas implica un mayor número medio de siniestros, por otro lado, una mayor experiencia supone un menor riesgo respecto el número de siniestros, en el caso del género con la base de datos utilizada se espera un mayor número de siniestros si el conductor es hombre, mientras que en la bibliografía o bien no se muestra significativo o generalmente al ser hombre se espera un menor número de siniestros, aunque sí que se coincide en que el hombre tiene un comportamiento con más riesgo.

Con la estimación de los modelos de regresión de Poisson con *offset* clásico y con *offset* generalizado se ha podido ejemplificar la relación del kilometraje y la frecuencia siniestral, que también era objeto de análisis en este trabajo. Se ha podido comprobar que un mayor kilometraje representa un mayor riesgo, pero que la frecuencia siniestral no aumenta proporcionalmente al aumento del kilometraje, como se especifica en el modelo de Poisson con *offset* clásico, sino que hay un efecto de la experiencia, del vehículo que se usa, las vías en las que se conduce, que hace que aunque el kilometraje sea mayor y por lo tanto haya una mayor exposición al riesgo, esta estará compensada en cierta medida por los elementos que proporcionen mayor seguridad en la conducción, tal y como se recoge en el modelo de Poisson con *offset* generalizado, por lo tanto para los datos estudiados se concluye que no hay una relación lineal entre el kilometraje y la frecuencia de siniestros, tal y como se apuntaba en la revisión bibliográfica.

La principal limitación del estudio práctico ha sido la imposibilidad de obtener datos telemáticos reales, aunque con los datos simulados se ha podido realizar un estudio completo, lo deseado era una modelización con datos reales. También hubiese sido oportuno la aplicación de otros enfoques propuestos para la introducción del kilometraje en la modelización del número de siniestros, como la introducción de un modelo ZIP (Guillen et al 2013), o un modelo GAM (Boucher et al, 2017) y (Boucher & Turcotte, 2020), entre otros, los cuales no han podido ser abordados en el presente trabajo, pero serán tenidos en cuenta para futuras investigaciones.

En una sociedad, incierta y cambiante, en la que por factores externos puede verse reducido el uso del automóvil, el seguro basado en el uso es un seguro con mucho potencial, acompañándose del desarrollo digital, es un seguro con amplias ventajas para los conductores, las compañías aseguradoras y la sociedad, siendo especialmente relevante la disminución del riesgo siniestral y las mejoras en la conducción que se pueden lograr. También se consigue una mejor medición del riesgo de los conductores, permitiendo una mayor equidad y mejor precisión en las tarifas del seguro de automóviles, siendo de especial interés para aquellos jóvenes que tienen un menor kilometraje anual y tienen un comportamiento de conducción seguro, ya que podrán ver su prima reducida considerablemente respecto los seguros tradicionales.

9. Bibliografía

- Alcañiz, M., Ayuso, M., & Pérez-Marin, A.M. (2014). El seguro basado en el uso. *Gerencia de riesgos y seguros*, 120.
- Aleman, R; Ayuso, M. & Bolancé, C. (Curso 2020/2021). Modelos lineales generalizados (tarificación a priori). *Asignatura Modelos estadísticos aplicados del Máster de Ciencias Actuariales y Financieras*. Departamento de Econometría, Estadística y Economía Aplicada, Universidad de Barcelona
- Ayuso, M & Bolancé, C. (Curso 2020/2021). Curso de Econometría Actuarial y Análisis de la supervivencia en seguros, Modelos Lineales Generalizados. *Asignatura Modelos estadísticos aplicados del Máster de Ciencias Actuariales y Financieras*. Departamento de Econometría, Estadística y Economía Aplicada, Universidad de Barcelona
- Ayuso, M., Guillén, M. and Pérez-Marin, A.M. (2014a). Los hábitos de conducción al volante según el género en los seguros pay-as-you-drive o usage based. *Anales Del Instituto de Actuarios Españoles* 20(3), 17-32.
- Ayuso, M., Guillén, M. & Pérez-Marin, A.M. (2014b). Time and distance to first accident and driving patterns of young drivers with pay-as-you-drive insurance. *Accident Analysis and Prevention*, 73, 125-131. doi:10.1016/j.aap.2014.08.017
- Ayuso, M., Guillén, M. & Pérez-Marin, A.M. (2016a). Using GPS data to analyse the distance travelled to the first accident at fault in pay-as-you-drive insurance. *Transportation Research Part C:Emerging Technologies*, 68, 160-167. doi:10.1016/j.trc.2016.04.004
- Ayuso, M., Guillén, M., & Pérez-Marin, A. (2016b). Telematics and Gender Discrimination: Some Usage-Based Evidence on Whether Men's Risk of Accidents Differs from Women's. *Risks*, 4(2), 1-10. doi:10.3390/risks4020010
- Ayuso, M., Guillén, M. & Nielsen, J.P. (2018). Improving automobile insurance ratemaking using telematics: incorporating mileage and driver behaviour data. *Transportation*, 46(3), 735-752. doi: 10.1007/s11116-018-9890-7
- Biener, C., Eling, M. & Lehmann, M. (2019). Balancing the desire for privacy against the desire to hedge risk. *Journal of Economic Behavior & Organization*, 180, 608-620. doi: 10.1016/j.jebo.2020.03.007
- Bolderdijk, J.W., Knockaert, J., Steg, E.M. & Verhoef, E.T. (2011). Effects of Pay-As-You-Drive vehicle insurance on young drivers' speed choice: results of a Dutch field experiment. *Accident Analysis and Prevention*, 43(3), 1181-1186. doi:10.1016/j.aap.2010.12.032
- Boucher, J.P., Pérez-Marin, A.M. & Santolino, M. (2013). Pay-as-you-drive insurance: the effect of the kilometers on the risk of accident. *Anales Del Instituto de Actuarios Españoles*, 19(3), 135-154.

- Boucher, J.-P., Côté, S., & Guillén, M. (2017). Exposure as Duration and Distance in Telematics Motor Insurance Using Generalized Additive Models. *Risks*, 5(4), 54. doi:10.3390/risks5040054
- Boucher, J.-P., & Turcotte, R. (2020). A Longitudinal Analysis of the Impact of Distance Driven on the Probability of Car Accidents. *Risks*, 8(3), 91. doi:10.3390/risks8030091
- Boj, E; Claramunt, M.M. & Costa, T. (2021). Tarificación y provisiones (Tercera edición). *Asignatura Matemática Actuarial del Máster de Ciencias Actuariales y Financieras*. Departamento de Matemática Económica, Financiera y Actuarial, Universidad de Barcelona (UB).
- Diario Oficial de la Unión Europea (2004) *Directiva del Consejo 2004/113/CE, de 13 de diciembre de 2004, por la que se aplica el principio de igualdad de trato entre hombres y mujeres al acceso a bienes y servicios y su suministro*.
- Diario Oficial de la Unión Europea (2011) *Sentencia del Tribunal de Justicia (Gran Sala) de 1 de marzo de 2011 (petición de decisión prejudicial planteada por la Cour constitutionnelle — Bélgica) — Association belge des Consommateurs Test-Achats ASBL, Yann van Vugt, Charles Basselier/Conseil des ministres (Asunto C-236/09)*.
- Derikx, S., Reuver, M. & de Kroesen, M. (2016). Can privacy concerns for insurance of connected cars be compensated?. *Electronic Markets*, 26(1), 73-81. doi:10.1007/s12525-015-0211-0
- Eling, M. & Kraft, M. (2020). The impact of telematics on the insurability of risks. *Journal of Risk Finance*, 21(2), 77-109. doi:10.1108/JRF-07-2019-0129
- Ellison, A.B., Bliemer, M.C.J. & Greaves, S.P. (2015). Evaluating changes in driver behaviour: a risk profiling approach. *Accident Analysis and Prevention*, 75, 298-309. doi:10.1016/j.aap.2014.12.018
- Gao, G., & Wüthrich, M. (2019). Convolutional Neural Network Classification of Telematics Car Driving Data. *Risks*, 7(1), 6. doi:10.3390/risks7010006
- Guillén, M., Nielsen, J. P., Ayuso, M. & Pérez-Marín, A. M. (2019). The Use of Telematics Devices to Improve Automobile Insurance Rates. *Risk Analysis*, 39(3), 662-672. doi:10.1111/risa.13172.
- Guillén, M., Nielsen, J.P., Pérez-Marín, A.M. & Elpidorou, V. (2020). Can automobile insurance telematics predict the risk of near-miss events?. *North American Actuarial Journal*, 24(1), 141–152. doi:10.1080/10920277.2019.1627221
- Guillén, M.; Pérez-Marín, A. M. & Alcañiz, M. (2021). Percentile reference charts for speeding based on telematics information. *Accident Analysis and Prevention*, 150. doi: 10.1016/j.aap.2020.105865

- Guillén, M & Fernández, D. (Curso 2020/2021). Modelización y análisis, Machine Learning. *Postgrado, Data Analytics for social science*, Universidad de Barcelona
- Huang, Y. & Meng, S. (2019). Automobile insurance classification ratemaking based on telematics driving data. *Decision Support Systems*, 127. doi:10.1016/j.dss.2019.113156
- Litman, T. (2011). Pay-As-You-Drive Insurance: Recommendations for Implementation. *Victoria Transport Policy Institute*.
- Pérez-Marín, A. M., Guillén, M., Alcañiz, M., & Bermúdez, L. (2019). Quantile Regression with Telematics Information to Assess the Risk of Driving above the Posted Speed Limit. *Risks*, 7(3), 80. doi:10.3390/risks7030080
- Pesantez-Narvaez, J., Guillén, M., & Alcañiz, M. (2019). Predicting Motor Insurance Claims Using Telematics Data—XGBoost versus Logistic Regression. *Risks*, 7(2), 70. doi:10.3390/risks7020070
- Sun, S., Bi, J., Guillén, M., & Pérez-Marín, A. M. (2020). Assessing Driving Risk Using Internet of Vehicles Data: An Analysis Based on Generalized Linear Models. *Sensors*, 20(9), 2712. doi:10.3390/s20092712
- Tselentis, D.I., Yannis, G. & Vlahogianni, E.I. (2016). Innovative insurance schemes: pay as/how you drive. *Transportation Research Procedia*, 14, 362-371. doi:10.1016/j.trpro.2016.05.088
- Vickrey, W. (1968). Automobile accidents, tort law, externalities, and insurance: an economist's critique. *Law and Contemporary Problems*, 33, 464-487.
- Wouters, P. & Bos, J. (2000). Traffic accident reduction by monitoring driver behaviour with in-car data recorders. *Accident Analysis and Prevention*, 32(5), 643-650. doi:10.1016/S0001-4575(99)00095-0

Webgrafía

- Álvarez.R (2021). ¿Por qué hay que pagar el seguro si no se usa el coche?. LA VANGUARDIA. Recuperado de <https://www.lavanguardia.com/economia/bolsillo/20210430/7134915/seguro-coche-pagar-poliza-companias-coronavirus.html>
- Exceso de velocidad. *Comisión Europea*. Recuperado de https://ec.europa.eu/transport/road_safety/topics/behaviour/speeding_es
- Manage How You Drive. *Amodo*. Recuperado de <https://www.amodo.eu/en/blog/mhyd-and-tbyd-next-generation-of-ubi-product-models>
- PHYD and PAYD. *Amodo*. Recuperado de <https://www.amodo.eu/en/blog/phyd-and-payd-ubi-most-popular-product-models>

Tablas estadísticas. Grupo 7. Accidentes y víctimas en función de la vía 2019. *Dirección General de Tráfico*. Recuperado de <https://www.dgt.es/es/seguridad-vial/estadisticas-e-indicadores/accidentes-30dias/tablas-estadisticas/>

UBI Infographic 2018. *PTOLEMUS Consulting group*. Recuperado de <https://www.ptolemus.com/ubi-infographic-2018/>

Usage-Based Insurance Market Outlook- 2027. *Allied Market Research*. Recuperado de <https://www.alliedmarketresearch.com/usage-based-insurance-market>

USAGE-BASED INSURANCE PROGRAMS. *Buy auto insurance*. Recuperado de <https://www.buyautoinsurance.com/usage-based-insurance/>

Westreicher, G. Subsidio Cruzado. *Economipedia*. Recuperado de <https://economipedia.com/definiciones/subsidio-cruzado.html>

10. Anexos

10.1. Cuadros resumen artículos

El formato de resumen de los artículos académicos ha sido inspirado en la lista de artículos académicos que realizan Eling, M. & Kraft, M. (2020).

En este presente trabajo se exponen recuadros de forma individual para cada artículo académico⁴, agrupando estos en 8 grupos (Ocurrencia o no de siniestros; Kilómetros recorridos a velocidades superiores al límite permitido; Distancia y tiempo transcurrido al primer accidente; Aceleración, frenado y curvas; Frecuencia siniestral con *offset*; Mejoras en la conducta mediante seguimiento; Privacidad;Otros). Detallando en una fila, los datos y metodología, y en una segunda fila, las conclusiones. En la parte superior de los recuadros se detallan los autores, el año, el título y la revista de publicación.

Ocurrencia o no de siniestros

Tabla 1. Alcañiz, M., Ayuso, M., & Pérez-Marín, A.M. (2014). El seguro basado en el uso. *Gerencia de riesgos y seguros*, 120.

Datos y Metodología	<ul style="list-style-type: none">• Muestra de 25.014 conductores jóvenes, con una póliza PAYD durante el año 2011, en una compañía española.• Comparativa conductores sin siniestros (82,4%) y con siniestros (17,6%) de la muestra.• Variable respuesta binaria: Ocurrencia (Y=1) o no (Y=0) de siniestro.• Variables independientes: Tradicionales (edad, género, antigüedad de carnet, antigüedad del vehículo, potencia del vehículo y lugar de aparcamiento). Telemáticas (núm. km anuales recorridos, núm. km nocturnos, % km nocturnos, % de exceso de velocidad y % km en vías urbana).• Modelo de regresión logística simple.
Conclusiones	<ul style="list-style-type: none">• Sin siniestros vs con siniestros: Hay un porcentaje levemente superior de hombres y vehículos aparcados en garaje en los asegurados con siniestro, estos también son un poco más jóvenes y tienen menos experiencia.

⁴ No se incluyen los artículos de (Boucher et al, 2017), (Boucher & Turcotte, 2020), (Litman, 2011) y (Vickrey, 1968).

	<p>La potencia del vehículo, el kilometraje anual recorrido, el porcentaje de conducción urbana y nocturna, así como el porcentaje de conducción con exceso de velocidad es superior en aquellos asegurados con siniestros.</p> <p>Referente a la antigüedad del vehículo no se encuentran diferencias entre asegurados con y sin siniestros.</p> <ul style="list-style-type: none"> • Estimación modelo de regresión logística simple: <p>El aumento de la experiencia disminuye la probabilidad de tener un siniestro. Una mayor potencia del vehículo, el aumento del kilometraje y el aumento del porcentaje de conducción en vías urbanas aumentan la probabilidad de tener un siniestro.</p> <p>Los que presentan una conducción nocturna de entre el 11-20%, tienen menos probabilidad de tener un siniestro que los que presentan un porcentaje más bajo al 11% (categoría base).</p> <p>Los que presentan un exceso de velocidad de hasta el 2% tienen menor probabilidad de tener un siniestro respecto a los que tienen un exceso superior al 11% (categoría base). Los excesos de velocidad entre el 3-5% y el 6-11% no presentan un parámetro significativo.</p> <p>Los parámetros de las variables género, lugar de aparcamiento, antigüedad del vehículo no son significativos, tampoco lo son los de los grupos de edad, por una posible correlación con la antigüedad de carnet.</p>
--	--

Fuente: Elaboración propia

Tabla 2. Ayuso, M., Guillén, M. and Pérez-Marín, A.M. (2014a). Los hábitos de conducción al volante según el género en los seguros pay-as-you-drive o usage based. *Anales Del Instituto de Actuarios Españoles*, 20(3), 17-32.

Datos y Metodología	<ul style="list-style-type: none"> • Muestra de 25.014 asegurados, jóvenes, con una póliza PAYD durante el año 2011, en una compañía española. • Comparativa de hombres (49%) y mujeres (51%) de la muestra. • Variable respuesta binaria: Ocurrencia (Y=1) o no (Y=0) de siniestro. • Variables independientes: Tradicionales (edad, antigüedad de carnet, antigüedad del vehículo, potencia
----------------------------	---

	<p>del vehículo y lugar de aparcamiento). Telemáticas (número de km anuales, % km recorridos nocturnamente, % km recorridos en zonas urbanas y % km con exceso de velocidad).</p> <ul style="list-style-type: none"> • Modelo de regresión logística simple.
<p>Conclusiones</p>	<ul style="list-style-type: none"> • Comparativa de género y sus submuestras: El número de kilómetros recorridos, el porcentaje de kilómetros con conducción nocturna y con exceso de velocidad es superior para los hombres comparando tanto en el total como las submuestras con y sin siniestros. El porcentaje de conducción urbana es superior en mujeres en el total y en la submuestra sin siniestros, en la submuestra con siniestros es superior en los hombres. • Estimación regresión logística simple: Se dan similitudes entre géneros: Para ambos géneros se da que un aumento de la antigüedad de carnet conlleva una menor probabilidad de tener un siniestro, mientras que un aumento de la potencia del vehículo, el aumento del número de kilómetros recorridos y la conducción por zonas urbanas hacen aumentar la probabilidad de tener un siniestro. Diferencias entre géneros: Al aumentar la antigüedad del vehículo la probabilidad de siniestro aumenta y se da una menor probabilidad de siniestro con un exceso de velocidad entre 0-2% frente a un exceso de velocidad mayor al 12% solo en los hombres. Mientras que solo para las mujeres el hecho que el porcentaje de conducción nocturna sea superior al 30% hace aumentar la probabilidad de tener un siniestro en ellas. Los parámetros que acompañan las variables de edad y lugar de aparcamiento no son significativos para ambos géneros.

Fuente: Elaboración propia

Tabla 3. Pesantez-Narvaez, J., Guillén, M., & Alcañiz, M. (2019). Predicting Motor Insurance Claims Using Telematics Data—XGBoost versus Logistic Regression. *Risks*, 7(2), 70. doi:10.3390/risks7020070

<p>Datos y Metodología</p>	<ul style="list-style-type: none"> • Muestra de 2.767 conductores, con menos de 30 años, con una póliza (PAYD) en una compañía de seguro española. Durante el 1 de enero al 31 de diciembre de 2011 se recopiló información de estos conductores mediante GPS.
-----------------------------------	---

	<ul style="list-style-type: none"> • Variable respuesta binaria: Ocurrencia (Y=1) o no (Y=0) de siniestro con culpa. • Variables independientes: Tradicionales (edad, género, antigüedad del vehículo y experiencia). Telemáticas (núm. km totales, % km recorridos en áreas urbanas, % km recorridos de noche y % km con exceso de velocidad). • Métodos utilizados: Regresión logística simple y XGBoost (<i>tree booster</i> y <i>linear booster</i>)
<p style="text-align: center;">Conclusiones</p>	<ul style="list-style-type: none"> • Sin siniestros vs con siniestro con culpa: Hay un porcentaje levemente superior de hombres en los asegurados con siniestro, estos también son un poco más jóvenes y tienen menos experiencia. El kilometraje anual recorrido, el porcentaje de conducción urbana y el porcentaje de conducción con exceso de velocidad es superior en aquellos asegurados con siniestros. El porcentaje de kilómetros recorridos nocturnamente es ligeramente superior en los asegurados sin siniestro. • El modelo de regresión logística simple muestra que: A mayor edad, menor es la probabilidad de tener un accidente con culpa y a más kilómetros recorridos en áreas urbanas, mayor es probabilidad de tener un accidente con culpa. • XGBoost vs Regresión logística simple: Las tres variables con más importancia en la predicción según el modelo en orden son: para la regresión logística simple (porcentaje de kilómetros recorridos en áreas urbanas, la edad y los kilómetros totales) y para el XGBoost-<i>TreeBooster</i> (el porcentaje de kilómetros con exceso de velocidad, el porcentaje de kilómetros recorridos en áreas urbanas y el porcentaje de kilómetros recorridos nocturnamente). • El modelo XGBoost se ha de ajustar mediante mayores esfuerzos (regularización), para obtener un rendimiento predictivo que sea similar al de la regresión logística. También el modelo XGBoost presenta una mayor dificultad de interpretación.

Fuente: Elaboración propia

Kilómetros recorridos a velocidades superiores al límite permitido

Tabla 4. Pérez-Marín, A. M., Guillén, M., Alcañiz, M., & Bermúdez, L. (2019). Quantile Regression with Telematics Information to Assess the Risk of Driving above the Posted Speed Limit. *Risks*, 7(3), 80. doi:10.3390/risks7030080

<p>Datos y Metodología</p>	<ul style="list-style-type: none"> • Muestra de 9.614 conductores, entre 18 y 35 años, cubiertos por el seguro UBI en 2010. • Variable respuesta: Núm. de km recorridos a velocidades superiores al límite permitido. • Variables independientes: Tradicionales (edad y género). Telemáticas (núm. km totales, % km recorridos en vías urbanas y % km recorridos de noche). • Modelos: Modelo de regresión lineal (MRL) y modelo de regresión por cuantiles, en los percentiles 50%, 75%, 90%, 95%, 97.5% y 99%.
<p>Conclusiones</p>	<ul style="list-style-type: none"> • La estimación del MRL muestra que: Con un mayor número de kilómetros recorridos, un mayor porcentaje de conducción nocturna y si el conductor es hombre se espera que el número de kilómetros con exceso de velocidad sea mayor. Con un aumento del porcentaje de kilómetros recorridos en vías urbanas se espera un menor número de kilómetros recorridos con exceso de velocidad. • Estimación modelo de regresión por cuantiles: Los kilómetros recorridos a velocidades superiores al límite permitido dependen positivamente de la distancia total recorrida y el porcentaje de conducción nocturna (excepto el percentil 97.5% y 99%), dependen negativamente del porcentaje de conducción urbana. Se espera un mayor número de kilómetros recorridos a velocidad superiores al límite si el conductor es hombre (excepto el percentil 99%). La edad no presenta ningún parámetro significativo en todos los percentiles. A mayor es el nivel de percentil mayor es el efecto marginal de las covariables a la variable respuesta. • Se pueden identificar conducciones peligrosas mediante la información telemática que nos muestra el comportamiento específico de cada conductor.

Fuente: Elaboración propia

Tabla 5. Guillén, M.; Pérez-Marín, A. M. & Alcañiz, M. (2021). Percentile reference charts for speeding based on telematics information. *Accident Analysis and Prevention*, 150. doi:10.1016/j.aap.2020.105865

<p>Datos y Metodología</p>	<ul style="list-style-type: none"> • Muestra de 9.585 conductores, entre 18 y 35 años, cubiertos por una póliza PHYD en una compañía española en el año 2010. • Variable respuesta: número de km recorridos a velocidades superiores al límite permitido. • Variables independientes: Tradicionales (edad y género). Telemáticas (núm. km totales, % km recorridos en vías urbanas y % km recorridos de noche). • Modelos: modelo de regresión gamma y modelo de regresión por cuantiles, en los percentiles 50%, 75%, 90%, 95%, 97.5% y 99%.
<p>Conclusiones</p>	<ul style="list-style-type: none"> • En la estimación del modelo gamma se observa que: <p>Cuanto más kilómetros recorridos mayor es el número de kilómetros conducidos con exceso de velocidad esperado y a mayor porcentaje de kilómetros conducidos en vías urbanas menor son los kilómetros esperados con exceso de velocidad esperados. Por lo referente al género, se espera que los hombres conduzcan más kilómetros con exceso de velocidad. Cuanto mayor es la conducción nocturna, mayor es el número de kilómetros esperados con exceso de velocidad (con una significación al 10%).</p> • Estimación del modelo de regresión por cuantiles: <p>A más kilómetros el cuantil de los kilómetros en exceso de velocidad aumenta, en el caso del porcentaje de conducción urbana su efecto es negativo, cuanto mayor es el porcentaje más disminuye el cuantil del número de kilómetros con exceso de velocidad. Estas dos variables presentan un parámetro significativo para todos los percentiles.</p> <p>El porcentaje de conducción nocturna tan solo presenta su parámetro significativo en el percentil 50%, por lo que al aumentar este porcentaje la mediana de kilómetros recorridos con exceso de velocidad aumenta. En el caso de la edad el parámetro solo es significativo en los percentiles 95% y 97,5% y muestra que, a mayor edad, el percentil del número de kilómetros con exceso de velocidad aumenta.</p>

	<p>En el caso de la variable género, a excepción de los percentiles 97,5% y 99%, el hecho de ser hombre supone un aumento del cuantil.</p> <p>A mayor es el nivel de percentil mayor es el efecto de las variables independientes (cuando su parámetro es significativo) a la variable respuesta.</p> <ul style="list-style-type: none"> • Existe una relación exponencial entre la distancia total recorrida y los kilómetros recorridos con exceso de velocidad.
--	---

Fuente: Elaboración propia

Distancia y tiempo transcurrido al primer accidente

Tabla 6. Ayuso, M., Guillén, M. & Pérez-Marin, A.M. (2014b). Time and distance to first accident and driving patterns of young drivers with pay-as-you-drive insurance. *Accident Analysis and Prevention*, 73, 125-131. doi:10.1016/j.aap.2014.08.017

Datos y Metodología	<ul style="list-style-type: none"> • Muestra de 15.940 conductores, menores de 30 años, cubiertos con una póliza PAYD, con una aseguradora española, el año 2009. Los patrones de conducción se registran a través de GPS hasta el 31 de diciembre de 2011. • Comparativa entre conductores principiantes con antigüedad de carnet menor a 1 año (17,5%) y conductores experimentados con antigüedad de carnet mayor a 1 año (82,5%). • Variables respuesta: Tiempo y distancia al primer accidente. • Variables independientes: Tradicionales (Género, edad, antigüedad del vehículo y experiencia). Telemáticas (% km en conducción urbana, % km en conducción nocturna y % km en conducción con exceso de velocidad). • Método de análisis de supervivencia: Modelo de regresión de Weibull.
Conclusiones	<ul style="list-style-type: none"> • Conductores principiantes vs experimentados: La edad media es menor y la antigüedad del vehículo es mayor en los conductores con menos experiencia.

	<p>El porcentaje de conducción urbana es superior en los conductores con menos experiencia, el porcentaje de kilómetros en exceso de velocidad y kilómetros por día es mayor en los conductores con más experiencia.</p> <p>No hay diferencias significativas para el porcentaje de conducción nocturna.</p> <p>Para ambos grupos los hombres presentan comportamientos con más riesgo en la conducción (más kilómetros por día recorridos, un mayor porcentaje de conducción urbana, nocturna y con exceso de velocidad).</p> <ul style="list-style-type: none">• Tiempo al primer accidente: <p>Los conductores principiantes tienen el primer accidente antes que los conductores experimentados.</p> <p>Mediante la regresión de Weibull se observa que:</p> <p>Para los conductores principiantes, al aumentar la conducción por vías urbana (superior al 25%) y el porcentaje de conducción con exceso de velocidad (con efecto cuadrático positivo) conlleva que el tiempo al primer accidente se reduzca. Los hombres tienen el primer accidente en menor tiempo que las mujeres, las mujeres con mayor porcentaje de conducción nocturna tienen el primer accidente en menor tiempo, mientras que para los hombres la conducción nocturna no tiene un efecto relevante.</p> <p>Para los conductores experimentados, al aumentar la conducción por vías urbanas (con efecto cuadrático positivo) y el porcentaje de conducción nocturna, el tiempo al primer accidente se reduce. El aumento del porcentaje de exceso de velocidad (con efecto cuadrático positivo) reduce el tiempo al primer accidente, siendo esta reducción mayor para los hombres. El aumento de la experiencia aumenta el tiempo al primer accidente.</p> <ul style="list-style-type: none">• Distancia al primer accidente: <p>Los conductores principiantes recorren una distancia menor antes del primer accidente que los conductores experimentados.</p> <p>Mediante la regresión de Weibull se observa que:</p>
--	---

	<p>Para los conductores principiantes, el aumentar la conducción por vías urbana (con efecto cuadrático positivo) reduce la distancia esperada al primer accidente.</p> <p>Para los conductores experimentados, el aumento de la experiencia aumenta la distancia recorrida al primer accidente y las mujeres recorren menos kilómetros antes del primer accidente. Al aumentar la conducción por vías urbanas y el porcentaje de exceso de velocidad (con efecto cuadrático en ambas variables) y a mayor antigüedad del vehículo la distancia recorrida al primer accidente se reduce.</p>
--	--

Fuente: Elaboración propia

Tabla 7. Ayuso, M., Guillén, M. & Pérez-Marin, A.M. (2016a). Using GPS data to analyse the distance travelled to the first accident at fault in pay-as-you-drive insurance. <i>Transportation Research Part C:Emerging Technologies</i> , 68, 160-167. doi:10.1016/j.trc.2016.04.004	
Datos y Metodología	<ul style="list-style-type: none"> • Muestra de 8.198 conductores, menores de 30 años, con una póliza PAYD en España el año 2009. Los patrones de conducción se registran a través de GPS hasta el 31 de diciembre de 2011. • Comparativa de hombres (54.68%) y mujeres (45.32%) de la muestra. • Variable respuesta: Distancia recorrida hasta el primer accidente con culpa. • Variables independientes: Tradicionales (edad, antigüedad del vehículo y experiencia). Telemáticas (% km recorridos en zonas urbanas, % km recorridos de noche y % km recorridos por encima de los límites permitidos). • Método de análisis de supervivencia: Modelo de regresión de Weibull.
Conclusiones	<ul style="list-style-type: none"> • Los hombres tienen comportamientos con más riesgo en la conducción (mayor porcentaje de kilómetros conducidos en zonas urbana, de noche y por encima del límite de velocidad) que las mujeres. • Los hombres recorren más kilómetros que las mujeres hasta tener el primer accidente con culpa.

	<ul style="list-style-type: none"> Las estimaciones de la regresión de Weibull muestran que: <p>El efecto de las covariables presenta similitudes y diferencias entre géneros.</p> <p>Cuanta más experiencia mayor es la distancia al primer siniestro con culpa, mientras que a mayor antigüedad del vehículo y porcentaje de kilómetros recorridos en zonas urbanas (efecto cuadrático en mujeres) más se reduce la distancia en ambos sexos.</p> <p>Cuanto mayor el porcentaje de kilómetros por encima de los límites (con efecto cuadrático positivo) más se reduce la distancia en los hombres, pero no es significativo en las mujeres. Y a mayor porcentaje de kilómetros conducidos de noche más se reduce la distancia en las mujeres, pero no es significativo en los hombres.</p>
--	--

Fuente: Elaboración propia

Tabla 8. Ayuso, M., Guillén, M., & Pérez-Marín, A. (2016b). Telematics and Gender Discrimination: Some Usage-Based Evidence on Whether Men's Risk of Accidents Differs from Women's. <i>Risks</i> , 4(2), 1-10. doi:10.3390/risks4020010	
Datos y Metodología	<ul style="list-style-type: none"> Muestra de 8.198 conductores, menores de 30 años, con una póliza PAYD en España el año 2009. Los patrones de conducción se registran a través de GPS hasta el 31 de diciembre de 2011. Comparativa de hombres (54.68%) y mujeres (45.32%) de la muestra. Variable respuesta: Tiempo transcurrido hasta el primer accidente con culpa. Variabes independientes: Tradicionales (género, edad, antigüedad del vehículo y experiencia). Telemáticas (km/día, % km recorridos en zonas urbanas, % km recorridos de noche y % km recorridos por encima de los límites permitidos). Método de análisis de supervivencia: Modelo de regresión de Weibull. Se forman 4 grupos de riesgo (bajo, medio, alto y sin descuento), según los km recorridos al año, los km de conducción nocturna, el % de conducción urbana y el exceso de velocidad.

Conclusiones

- Comportamiento conducción en el género:

Los hombres tienen comportamientos con más riesgo en la conducción (mayor porcentaje de kilómetros conducidos en zonas urbana, de noche y por encima del límite de velocidad) que las mujeres. Además, en el caso de los hombres se observa un porcentaje superior al de las mujeres en el grupo de riesgo más alto (sin descuento).

- Estimación modelos Weibull ambos géneros:

En la estimación del modelo 1 (sin la variable explicativa kilómetros por día) se muestra que, el tiempo transcurrido esperado al primer accidente es más reducido en los hombres. A mayor porcentaje de conducción en vías urbanas y en exceso de velocidad, menor es el tiempo al primer accidente con culpa (ambas variables con efecto cuadrático positivo). A mayor porcentaje de conducción nocturna también es menor el tiempo al primer accidente con culpa. Cuanta más experiencia mayor es el tiempo transcurrido al primer siniestro con culpa, mientras que a mayor antigüedad del vehículo menor es el tiempo.

En la estimación del modelo 2 (introduciendo la variable explicativa kilómetros por día), el parámetro de la variable hombre deja de ser significativo, mientras que cuanto mayor es la intensidad de conducción menor es el tiempo transcurrido al primer accidente. La interpretación de las demás variables independientes es la misma que la del modelo 1.

De los dos modelos, el segundo es el que mejor se ajusta.

- Estimación modelos Weibull para hombres y mujeres:

El efecto de las variables explicativas presenta similitudes y diferencias entre géneros: Cuanta más experiencia mayor es el tiempo transcurrido al primer siniestro, mientras que a mayor antigüedad del vehículo y kilómetros recorridos en zonas urbanas (con efecto cuadrático positivo) y mayor intensidad de conducción menor es el tiempo transcurrido en ambos sexos.

Cuanto más kilómetros por encima de los límites (con efecto cuadrático positivo) más se reduce el tiempo transcurrido para los hombres, pero no es significativo en las mujeres y a más kilómetros conducidos de noche más se reduce el tiempo transcurrido en las mujeres, pero no es significativo en los hombres.

	<ul style="list-style-type: none"> • Grupos de riesgo: <p>En las estimaciones de los años esperados que han de transcurrir hasta el primer siniestro con culpa para los 4 grupos de riesgo, manteniendo los mismos valores límites de las variables que forman cada grupo, no se muestran gran diferencia entre hombre y mujeres, si bien los años esperados son menores en las mujeres en todos los grupos de riesgo menos en el de sin descuento.</p> <p>Las probabilidades de que el primer accidente con culpa se de en el primer año de cobertura para los grupos de riesgo, muestran que son muy altas en los grupos con más riesgo, en el grupo sin descuento un 30% y en el grupo de alto riesgo un 14%. Las probabilidades entre hombres y mujeres son similares, aunque las mujeres presentan unos valores un poco más altos en el grupo bajo, medio y alto, y más bajo en el sin descuento.</p>
--	---

Fuente: Elaboración propia

Aceleración, frenado y curvas

Tabla 9. Guillén, M., Nielsen, J.P., Pérez-Marín, A.M. & Elpidorou, V. (2020). Can automobile insurance telematics predict the risk of near-miss events?. <i>North American Actuarial Journal</i> , 24(1), 141–152. doi:10.1080/10920277.2019.1627221	
Datos y Metodología	<ul style="list-style-type: none"> • Estudio con una muestra de 157 conductores griegos que, durante un periodo de 8 semanas, entre los años 2016-2017, se recopiló su información telemática. • Variables respuesta: Eventos de casi-accidentes (<i>near-miss events</i>) de tres tipos (aceleración, frenado y curvas). • Variables independientes: Tradicionales (género, edad, experiencia del conductor, antigüedad del vehículo, potencia del motor y lugar aparcamiento). Telemáticas (% de km de conducción urbana, % km de conducción nocturna y % de km con exceso de velocidad). • Modelo: Regresión Binomial Negativa (BN).
Conclusiones	<ul style="list-style-type: none"> • Para los tres casi-accidentes, son factores de riesgo relevantes ambos tipos de variables, tradicionales y telemáticas.

	<ul style="list-style-type: none"> • El impacto de los factores de riesgo difiere según la tipología del <i>near-miss events</i>: <p>En las variables tradicionales, cuanto más edad mayor es el número esperado de los tres <i>near-miss events</i>. A mayor experiencia de conducción se reduce el número esperado de evento en curvas. A mayor potencia del vehículo mayor es el número esperado de evento de aceleración y un menor número en el evento de curvas. Tanto para una mayor antigüedad del vehículo como el estacionamiento del vehículo en parking, se espera un menor número de eventos de frenado y aceleración.</p> <p>En las variables telemáticas, se espera un mayor número de eventos de frenado con la conducción urbana, con la conducción nocturna un menor número esperado de eventos en curvas, y el exceso de velocidad un mayor número de eventos de aceleración.</p>
--	---

Fuente: Elaboración propia

Tabla 10. Sun, S., Bi, J., Guillén, M., & Pérez-Marín, A. M. (2020). Assessing Driving Risk Using Internet of Vehicles Data: An Analysis Based on Generalized Linear Models. <i>Sensors</i> , 20(9), 2712. doi:10.3390/s20092712	
Datos y Metodología	<ul style="list-style-type: none"> • Datos de 253 de vehículos en China, recogidos mediante una caja telemática (T-box), obteniendo información para 62 variables. • Variables dependientes: Frenado y acelerador. • Variables independientes telemáticas: Distancia (distancia de conducción acumulada en km), velocidad (km/h), RPM (media de revoluciones por minuto), rango de conducción (unidades geográficas). Frenado y acelerador como variables independientes cuando no son dependientes. • Modelos: Regresión por Mínimos Cuadrados Ordinarios (OLS) y Regresión Logística (LR).
Conclusiones	<ul style="list-style-type: none"> • Comparativa de variables dependientes: <p>La variable que tiene una mayor influencia en el riesgo de conducción es la velocidad, cuanto más elevada el riesgo es más elevado (el conductor debe acelerar más, las RPM son más elevadas y es más necesario hacer frenadas).</p>

	<p>Hay una relación existente entre el kilometraje y la aceleración y el kilometraje y el frenado, aunque diferente en las dos variables dependientes.</p> <p>Tanto el R^2 ajustado (OLS) como el Pseudo R^2 (LR) es superior en las regresiones con la variable dependiente acelerador, siendo esta la variable dependiente de estudio final.</p> <ul style="list-style-type: none"> • Clasificación de los conductores: <p>Es posible generar una clasificación de los conductores según su riesgo de conducción en buenos conductores y conductores con riesgo, mediante la predicción y el valor observado de la variable acelerador (Bueno = 0, Mediano = mediana, Malo = valor máximo). Para una mejor predicción se deberían de usar los dos modelos (OLS, LR).</p> <p>Sin información de los siniestros es posible tener información de comportamientos en la conducción con riesgo.</p>
--	---

Fuente: Elaboración propia

Frecuencia siniestral con *offset*

Tabla 11. Boucher, J.P., Pérez-Marín, A.M. & Santolino, M. (2013). Pay-as-you-drive insurance: the effect of the kilometers on the risk of accident. *Anales Del Instituto de Actuarios Españoles*, 19(3), 135-154.

Datos y Metodología	<ul style="list-style-type: none"> • Muestra de 25.014 asegurados, menores de 40 años, con una póliza PAYD durante el año 2011, en una compañía española. Los patrones de conducción se registran a través de GPS. • Variables respuesta: Cuatro categorías de siniestros. Siniestros con daños materiales con culpa, siniestros con daños corporales con culpa, siniestros con daños materiales sin culpa, siniestros con daños corporales sin culpa. • Variables independientes: Edad, antigüedad del vehículo, género y lugar de aparcamiento. • Modelos: Modelo de regresión de Poisson con los kilómetros anuales como <i>offset</i> sin covariables y modelo de regresión de Poisson con los kilómetros anuales como <i>offset</i> generalizado.
----------------------------	--

Conclusiones

- Comparativa entre categorías de siniestros:

Tanto en los siniestros de daños materiales como el de daños corporales, no se encuentran diferencias relevantes entre los siniestros con culpa y sin culpa.

Se han reportado pocos siniestros, más del 90% de los asegurados no presentan siniestros de daños materiales y en el caso de los siniestros de lesiones corporales el porcentaje es aún más elevado un 98%. Del porcentaje restante para ambos tipos de daños, la mayoría de los asegurados tienen solo un siniestro.

- Los asegurados con los mismos kilómetros presentan una frecuencia de siniestros similares.
- Modelo de regresión de Poisson km como *offset*:

La predicción con el *offset* sin ninguna covariable que representa una relación lineal entre el kilometraje y el número esperado de siniestros subestima la frecuencia de siniestros de los asegurados con menos de 10.000 km/año y sobreestima los siniestros de aquellos con más de 10.000 km/año.

- Generalización del modelo de regresión de Poisson *offset*:

Se espera un número mayor de siniestros en aquellos conductores con menos de 25 años en las cuatro categorías. También se espera un mayor número de siniestros sin culpa con daños materiales en los asegurados entre 25 y 30 años.

Si el asegurado es hombre se espera un mayor número de siniestros sin culpa de daños materiales y un menor número de siniestros sin culpa de daños corporales, para el lugar de aparcamiento se espera un mayor número de siniestros de daños materiales sin culpa en los asegurados que tienen el coche aparcado en garaje privado.

Se esperan menos siniestros de daños corporales sin culpa en aquellos conductores con una antigüedad del vehículo menor a dos años.

Se demuestra que los siniestros no aumentan de forma lineal al aumentar el número de kilómetros anuales, sino que, para las cuatro categorías de accidentes, hay una influencia de la experiencia del conductor y otros elementos (vehículos más nuevos, conducción por vías más seguras).

Fuente: Elaboración propia

Tabla 12. Ayuso, M., Guillén, M. & Nielsen, J.P. (2018). Improving automobile insurance ratemaking using telematics: incorporating mileage and driver behaviour data. *Transportation*, 46(3), 735-752. doi: 10.1007/s11116-018-9890-7

<p>Datos y Metodología</p>	<ul style="list-style-type: none"> • Muestra de 25.014 asegurados, menores de 40 años, con una póliza PAYD durante el año 2011, en una compañía española. Los patrones de conducción se registran a través de GPS. • Variable respuesta: Número de siniestros. • Variables independientes: Tradicionales (edad, género, antigüedad de carnet, antigüedad del vehículo, potencia del vehículo, lugar aparcamiento). Telemáticas (km anuales, % km recorridos de noche, % km recorridos en zonas urbanas y % km recorridos con exceso de velocidad.) • Modelos: Modelo clásico de Poisson y modelo de Poisson con <i>offset</i> el kilometraje. • Comparativa cuatro estimaciones: Variables no telemática, variables telemáticas, ambos tipos de variables (telemáticas y no telemáticas), y telemáticas con un <i>offset</i> (log de la predicción del modelo con variables no-telemáticas).
<p>Conclusiones</p>	<ul style="list-style-type: none"> • Asegurados sin siniestros vs con siniestros: <p>Los conductores sin siniestro presentan una media de edad levemente mayor y antigüedad de carnet mayor.</p> <p>Los conductores con siniestros presentan vehículos con mayor potencia, un mayor porcentaje de kilómetros anuales, porcentaje de kilómetros recorridos de noche, porcentaje de kilómetros recorridos en zonas urbanas y porcentaje de kilómetros recorridos con exceso de velocidad.</p> <p>No hay diferencias significativas en la antigüedad del vehículo ni en el género, pero los hombres presentan un porcentaje más elevado de siniestros.</p> <ul style="list-style-type: none"> • Modelo clásico de Poisson siniestros con y sin culpa: <p>Se obtiene un mejor ajuste incluyendo ambos tipos de variables, y en el caso que se use solo un tipo se obtiene un mejor ajuste incluyendo solo las variables telemáticas.</p>

Todas las variables telemáticas presentan un parámetro significativo en la estimación de la predicción del número de siniestros con las variables no-telemáticas incluidas como *offset*. El número esperado de siniestros aumenta con un aumento del porcentaje de conducción por vías urbanas y un aumento del número de kilómetros anuales recorridos, también aumenta con el aumento del porcentaje de kilómetros con exceso de velocidad (con efecto cuadrático negativo) y la conducción nocturna solo con efecto cuadrático positivo.

Las variables tradicionales presentan un efecto diferente al contemplar ambos tipos de variables. En aquellas estimaciones que incluyen las variables telemáticas, la variable género no es relevante. Con un aumento de la potencia se espera un mayor número de siniestros y con una mayor experiencia un menor número de siniestros en todas las estimaciones, mientras que, la variable antigüedad del vehículo es relevante solo en las estimaciones que incluyen las variables telemáticas, esperando un mayor número de siniestros si esta aumenta. Las variables edad y lugar de aparcamiento no son relevantes en ninguna de las estimaciones. Los kilómetros anuales es el factor de riesgo que tiene un mayor efecto en el número de siniestros.

- Modelo clásico de Poisson siniestros con culpa:

Como en el modelo anterior, se obtiene un mejor ajuste incluyendo ambos tipos de variables, y en el caso que se use solo un tipo se obtiene un mejor ajuste incluyendo solo las variables telemáticas. La variable que tiene un mayor efecto en el número de siniestros son los kilómetros anuales. Las diferencias relevantes respecto el anterior modelo son que, el ser hombre presenta un efecto negativo en el número de siniestros, la edad sí que se muestra relevante reduciendo el número de siniestros esperados (con un efecto cuadrático positivo) y la conducción nocturna no se muestra relevante en la explicación de los siniestros con culpa.

- Modelo de Poisson con *offset* kilometraje siniestros con y sin culpa:

Se obtiene un mejor ajuste incluyendo ambos tipos de variables, y en el caso que se use solo un tipo se obtiene un mejor ajuste incluyendo solo las variables telemáticas. El efecto de las variables no es dispar al obtenido en el modelo clásico, con la diferencia que el parámetro que acompaña al género es significativo, siendo este negativo se espera un menor número de accidentes si es hombre.

	<ul style="list-style-type: none"> • Modelo de Poisson con <i>offset</i> kilometraje siniestros con culpa: <p>Como en los demás modelos, se obtiene un mejor ajuste incluyendo ambos tipos de variables, y en el caso que se use solo un tipo se obtiene un mejor ajuste incluyendo solo las variables telemáticas. El efecto de las variables no es dispar al obtenido en el modelo clásico de siniestros con culpa, con la diferencia que el parámetro que acompaña al género no es significativo.</p> <ul style="list-style-type: none"> • La predicción del número esperado de siniestros en comparación con el valor observado muestra que se dan los mejores resultados cuando se tienen en cuenta ambos tipos de variables en los dos modelos.
--	---

Fuente: Elaboración propia

Tabla 13. Guillén, M., Nielsen, J. P., Ayuso, M. & Pérez-Marín, A. M. (2019). The Use of Telematics Devices to Improve Automobile Insurance Rates. <i>Risk Analysis</i> , 39(3), 662-672. doi:10.1111/risa.13172.	
Datos y Metodología	<ul style="list-style-type: none"> • Muestra de 25.014 asegurados, con una edad máxima de 37 años, con una póliza PAYD durante el año 2011, en una compañía española. Los patrones de conducción se registran a través de GPS. • Variable dependiente: Número de siniestros. • Variables independientes: Tradicionales (edad, género, antigüedad de carnet, antigüedad del vehículo, potencia del vehículo, lugar de aparcamiento). Telemáticas (km anuales, % km recorridos de noche, % km recorridos con exceso de velocidad % km recorridos en zonas urbanas). • Modelos: Modelo de Poisson Inflado de Ceros (ZIP) con <i>offset</i>. • Comparativa tres estimaciones: Variables no telemáticas, variables telemáticas, y ambos tipos de variables (telemáticas y no telemáticas).
Conclusiones	<ul style="list-style-type: none"> • Asegurados sin siniestros vs con siniestros: <p>De los conductores que se sitúan en el primer cuartil de la variable distancia recorrida, han recorrido una mayor distancia aquellos conductores que han tenido siniestro, así mismo sucede con el segundo y el tercer cuartil.</p>

Los conductores con siniestros son en su mayoría hombres, tienen una mayor potencia del vehículo y un mayor porcentaje de sus vehículos están aparcados. También presentan un mayor porcentaje de kilómetros recorridos por la noche, por vías urbana y con exceso de velocidad.

Los conductores sin siniestros presentan una edad ligeramente mayor, más experiencia, y mayor antigüedad del vehículo (sin ser significativa la diferencia con los asegurados con siniestro).

- Modelo de Poisson Zero-Inflado con *offset* con siniestros con y sin culpa:

Se obtiene un mejor ajuste incluyendo ambos tipos de variables, y en el caso que se use solo un tipo se obtiene un mejor ajuste incluyendo solo las variables telemáticas.

Son variables relevantes y con un efecto positivo (aumenta el número esperado de siniestros al aumentar la variable), la antigüedad del vehículo y la potencia del vehículo, así como el porcentaje de kilómetros conducidos por encima del límite de velocidad (con efecto cuadrático negativo) y por zonas urbanas. Son variables relevantes y con un efecto negativo (se reduce el número esperado de siniestros al aumentar la variable) si el conductor es hombre, y con la antigüedad de carnet.

No tienen parámetros significativos las variables edad, lugar de aparcamiento y el porcentaje de conducción de nocturna.

En la parte cero inflado, el coeficiente que acompaña la variable kilómetros anuales es positivo indicando que, cuanta más distancia se recorre, la proporción de excesos de cero siniestros es mayor.

- Modelo de Poisson Zero-Inflado con *offset* siniestros con culpa:

Se obtiene un mejor ajuste incluyendo ambos tipos de variables, y en el caso que se use solo un tipo se obtiene un mejor ajuste incluyendo solo las variables telemáticas.

Son variables relevantes y con un efecto positivo (aumenta el número esperado de siniestros al aumentar la variable), la antigüedad del vehículo, la potencia del vehículo, el porcentaje de kilómetros conducidos por encima del límite de velocidad (con efecto cuadrático negativo) y por zonas urbanas. Son variables relevantes y con un efecto negativo (se

	<p>reduce el número esperado de siniestros al aumentar la variable), la edad (con efecto cuadrático positivo), y la antigüedad de carnet.</p> <p>No tienen parámetros significativos el género, el lugar de aparcamiento y el porcentaje de conducción nocturna.</p> <ul style="list-style-type: none"> • Importancia de las variables en el riesgo de siniestro: <p>Las tres variables con más importancia al determinar el riesgo de siniestro son de más a menos, el porcentaje de conducción urbana, la antigüedad de carnet y el porcentaje de conducción con exceso de velocidad. Las cuatro variables con menos importancia al determinar el riesgo de siniestro son, la antigüedad el vehículo, el género, el porcentaje de conducción nocturna y el lugar de aparcamiento.</p> <ul style="list-style-type: none"> • No hay una relación proporcional entre el número de siniestros y el kilometraje, se encuentran buenos conductores entre aquellos que conducen largos kilómetros y reportan cero siniestros. • Se obtiene una mejor bondad del ajuste con los modelos ZIP en Poisson con <i>offset</i> que un Poisson clásico con <i>offset</i>.
--	---

Fuente: Elaboración propia

Mejoras en la conducta mediante seguimiento

Tabla 14. Wouters, P. & Bos, J. (2000). Traffic accident reduction by monitoring driver behaviour with in-car data recorders. <i>Accident Analysis and Prevention</i> , 32(5), 643-650. doi:10.1016/S0001-4575(99)00095-0	
Datos y Metodología	<ul style="list-style-type: none"> • Estudio formado por 11 flotas distintas, con más de 840 vehículos, entre ellos 270 vehículos estaban equipados con un registrador, del total de vehículos se formaron 7 grupos. • Se usan dos tipos de registradores: ‘<i>Accident Data Recorder</i>’ (ADR), que recoge datos de antes y durante un accidente o incidente y ‘<i>Journey Data Recorder</i>’ (JDR), que recoge horarios, velocidad media, aceleraciones y desaceleraciones rápidas, consumo de combustible, etc.
Conclusiones	<ul style="list-style-type: none"> • El análisis muestra que el registro de los datos del comportamiento del conductor, con conocimiento de

	<p>estos y sabiendo que recibirán un <i>feedback</i>, resultó en una reducción promedio del 20% de accidentes.</p> <ul style="list-style-type: none"> • El efecto de seguridad del registro fue diferentes en los 7 grupos de vehículos del estudio.
--	---

Fuente: Elaboración propia

Tabla 15. Bolderdijk, J.W., Knockaert, J., Steg, E.M. & Verhoef, E.T. (2011). Effects of Pay-As-You-Drive vehicle insurance on young drivers' speed choice: results of a Dutch field experiment. <i>Accident Analysis and Prevention</i> , 43(3), 1181-1186. doi:10.1016/j.aap.2010.12.032	
Datos y Metodología	<ul style="list-style-type: none"> • Experimento con 141 participantes, de menos de 30 años, clientes de cinco compañías de seguros holandesas. Se les hace un seguimiento mediante GPS de a dónde conducían, la velocidad a la que iban, la hora de conducción y los km. • El experimento se divide en cuatro fases, de 2 meses cada una: Pre-medición, Intervención 1, Intervención 2 y Post-medición. • Los participantes se dividen en dos grupos: Grupo de Incentivos (n=100) con acceso a feedback, además el conductor podía recibir un total de 200€ de descuento en la prima de la póliza, 100€ en la fase de Intervención 1 y 100€ en la fase Intervención 2. Grupo de control (n=41) sin acceso a feedback, independientemente del comportamiento en la conducción recibían 200€ en la fase post-medición.
Conclusiones	<ul style="list-style-type: none"> • El porcentaje de conducción con exceso de velocidad aumentó en ambos grupos entre la fase de pre-medición y post-medición: Este aumento fue constante en el grupo de control durante las cuatro fases, mientras que en el grupo de incentivos se reduce en las dos fases de descuento y aumenta en la fase post-medición, posiblemente por una pérdida de conciencia de que se está registrando. • PAYD tiene un mayor efecto en la reducción de los excesos de velocidad en las carreteras de 50 km/h, 80 km/h y 100 km/h. • Estimación de una reducción del 14% del exceso de velocidad con el sistema PAYD.

	<ul style="list-style-type: none"> • En este experimento no queda demostrado que el PAYD provoque una reducción en el kilometraje ni en la conducción nocturna.
--	--

Fuente: Elaboración propia

Tabla 16. Ellison, A.B., Bliemer, M.C.J. & Greaves, S.P. (2015). Evaluating changes in driver behaviour: a risk profiling approach. *Accident Analysis and Prevention*, 75, 298-309. doi:10.1016/j.aap.2014.12.018.

Datos y Metodología	<ul style="list-style-type: none"> • Estudio sobre el seguro PAYD en Sídney, con 148 conductores de los cuales 106 completan el estudio y presentan valores sin inconsistencias. • El estudio se divide en dos fases: Una “anterior” con una duración de cinco semanas: Sin incentivos y sin <i>feedback</i> (no se da información a los conductores del exceso de velocidad). Una “posterior” con una duración de cinco semanas: <i>After 1</i>, con incentivos y <i>feedback</i>. <i>After 2</i>, sin incentivos y <i>feedback</i>. • Hay tres niveles jerárquicos que incluyen diferentes variables independientes: El primero es el <i>TSI (Temporal and Spatial Identifiers)</i> con las variables: Límite de velocidad, zona escolar, lluvia, hora del día, fin de semana y número de pasajeros. El segundo, <i>Driver</i> (características del conductor y vehículo) con las variables: Transmisión del vehículo (Vehicle transmisión), carrocería del vehículo, año del modelo del vehículo género y edad. <i>El tercero, Estudy fase</i> (referentes al estudio) con las variables: Hizo dinero, incentivo inicial, fase, inicios de sesión. • Variables respuesta: Exceso de velocidad, aceleración, frenado y el riesgo total. • Modelos de riesgo multinivel.
----------------------------	---

Conclusiones	<ul style="list-style-type: none"> • Se reduce el riesgo de exceso de velocidad entre la fase “anterior” y “posterior”, esta reducción es significativa tanto en el <i>after 1</i> como en el <i>after 2</i>, por lo que el hecho de que los conductores pudieran consultar los datos referentes al exceso de velocidad reduce en gran medida el riesgo de exceso de velocidad. Aunque se obtiene una mayor reducción cuando también se incluyen incentivos financieros. • Aquellos conductores que inicialmente tenían un incentivo mayor, por su mayor riesgo en la conducción, menor fue su reducción en la fase “posterior” respecto la “anterior”. Se da una mayor mejora en el comportamiento del exceso de velocidad en aquellos conductores que presentaban un comportamiento menos peligroso en la fase “anterior”.
---------------------	--

Fuente: Elaboración propia

Privacidad

Tabla 17. Derikx, S., Reuver, M. & de Kroesen, M. (2016). Can privacy concerns for insurance of connected cars be compensated?. *Electronic Markets*, 26(1), 73-81. doi:10.1007/s12525-015-0211-0

Datos y Metodología	<ul style="list-style-type: none"> • Experimento de elección discreta en el que se encuesta a propietarios de vehículos holandeses en octubre de 2014, en el que se obtuvieron 55 cuestionarios completados válidos. • Tres tipos de privacidad con sus atributos: La privacidad de la ubicación y el espacio con el atributo registro del kilometraje. La privacidad del comportamiento y acciones con el atributo registro del comportamiento en carretera. Y la privacidad de datos e imagen con los atributos ofertas de seguros adicionales y publicidad de terceros. • Dos niveles para cada atributo: Uno que invade la privacidad y otro que no supone ninguna invasión a la privacidad. • Se realiza un análisis conjunto y en su realización se añade un último atributo, de ahorro relativo del consumidor, con los niveles 0€, 10€ o 20€ de descuento.
----------------------------	--

Conclusiones	<ul style="list-style-type: none"> • Los encuestados, de los cinco atributos en la elegibilidad del seguro de auto basado en el uso, le dan la mayor importancia al ahorro relativo del consumidor, es decir, el descuento ofrecido, con un peso del 65%. Los demás atributos tienen un peso similar entre el 10% y 7%, de mayor a menor: registro del comportamiento en carretera, ofertas de seguros adicionales, publicidad de terceros y el registro del kilometraje. • Los participantes del análisis prefieren la póliza actual (convencional) a una basada en el uso. Siendo la desutilidad de escoger el seguro UBI de 1,21, en términos de <i>buy-off</i>, los participantes cambiarían a esta modalidad de seguro por una compensación de 9,54€ al mes. • Por una compensación mensual los participantes estarían dispuestos a renunciar a la privacidad, los valores del <i>buy-off</i> para los 3 tipos de privacidad son similares, aunque es más elevado para la privacidad referente al comportamiento y acciones. En el caso de la privacidad de datos e imágenes se considera positiva su pérdida en el atributo de ofertas adicionales por parte de la aseguradora, pero negativa si la van a usar terceros para publicidad. • Para pasarse al seguro UBI los que conducen más kilómetros necesitan de una mayor compensación que los que conducen menos y los que tienen una educación superior requieren de una menor compensación frente los que presentan una baja educación. Las muestras son muy pequeñas por lo que los resultados no son extrapolables a la población.
---------------------	--

Fuente: Elaboración propia

Tabla 18. Biener, C., Eling, M. & Lehmann, M. (2019). Balancing the desire for privacy against the desire to hedge risk. *Journal of Economic Behavior & Organization*, 180, 608-620. doi: 10.1016/j.jebo.2020.03.007

Datos y Metodología	<ul style="list-style-type: none"> • Experimento de laboratorio con incentivos, con un total de 485 participantes que consienten compartir datos de su ubicación mediante la aplicación WhatsApp, durante 1 o 8 horas. • Se cuantifica el efecto que tiene el monitoreo digital determinando el umbral de <i>Willingness To Pay</i> (WTP) en el que los participantes aceptan renunciar a la privacidad.
----------------------------	--

Conclusiones	<ul style="list-style-type: none"> • Hay una reducción del 25% y 50% en el WTP por una póliza de seguros al introducir un monitoreo digital, perdiendo así privacidad por parte de los participantes. Este hecho implica una reducción del potencial beneficio del asegurador. • El uso de <i>transparency contracts</i> reduce los costes que se derivan del riesgo moral.
---------------------	---

Fuente: Elaboración propia

Otros

Tabla 19. Tselentis, D.I., Yannis, G. & Vlahogianni, E.I. (2016). Innovative insurance schemes: pay as/how you drive. *Transportation Research Procedia*, 14, 362-371. doi:10.1016/j.trpro.2016.05.088

Metodología	<ul style="list-style-type: none"> • Recopilación de estudios científicos relevantes sobre el seguro basado en el uso (UBI). • Se exponen las distintas ventajas del seguro UBI frente un seguro convencional. • Detalla mediante artículos los indicadores de riesgo de los seguros Pay-as-you-drive (PAYD) y Pay-how-you-drive (PHYD), así como una comparativa de ambos.
--------------------	--

Fuente: Elaboración propia

Tabla 20. Gao, G., & Wüthrich, M. (2019). Convolutional Neural Network Classification of Telematics Car Driving Data. *Risks*, 7(1), 6. doi:10.3390/risks7010006

Datos y Metodología	<ul style="list-style-type: none"> • Muestra de 416 conductores de automóviles, de los cuales se recogen datos de ubicación GPS, segundo por segundo, de sus viajes individuales. • Se realizan estadísticas descriptivas sobre velocidades, aceleración, frenado y cambios de dirección. • Uso de <i>Convolutional Neural Networks</i> (ConvNets) profundas para clasificar correctamente viajes individuales a sus conductores.
----------------------------	--

Conclusiones	<ul style="list-style-type: none"> • Las ConvNets profundas asignan con éxito los viajes individuales a sus conductores correspondientes, más de un 75% son correctamente clasificados. • La asignación de viajes mediante ConvNets puede ser muy útil cuando un mismo vehículo es conducido por varias personas, y también cuando la misma persona conduce en distintas condiciones.
---------------------	---

Fuente: Elaboración propia

Tabla 21. Huang, Y. & Meng, S. (2019). Automobile insurance classification ratemaking based on telematics driving data. *Decision Support Systems*, 127. doi:10.1016/j.dss.2019.113156

Datos y Metodología	<ul style="list-style-type: none"> • Muestra de 2.151 pólizas de una compañía de seguros de Propiedad y Accidentes en China. Se recopila información de la conducción mediante datos telemáticos, entre el 26 de abril del 2015 y el 8 de julio de 2017. • Variables tradicionales: Edad, género, precio del vehículo, antigüedad del vehículo, cantidad de asientos, coeficientes bonus-malus y número de siniestros. • Variables telemáticas: Kilometraje, velocidad, ruta de viaje, tiempo de viaje, frecuencia cambios de carril, estacionariedad de la conducción, giros bruscos, aceleraciones y frenos repentinos. • Modelos de probabilidad de riesgo: <i>Logistic Regression (LR)</i>, <i>Support Vector Machine (SVM)</i>, <i>Random Forests (RF)</i>, <i>XGBoost</i> y <i>Neural Networks (NN)</i>. Modelo de frecuencia de siniestros: <i>Poisson Regression (logarithm of annual mileage offset)</i>. Método de aumento de datos: <i>Data augmentation method of variable (binning)</i>.
Conclusiones	<ul style="list-style-type: none"> • Con independencia del modelo utilizado, con las variables telemáticas se mejora la predicción. Hay una notable relación entre las variables telemáticas y el riesgo asociado a la conducción. • Utilizar variables agrupadas (<i>binning</i>) es favorable en los GLM, con su uso también se mejora la interpretación de los modelos de machine learning, facilitando así una clasificación actuarial.

Fuente: Elaboración propia

10.2. Código aplicación práctica

A continuación, se detalla el código R utilizado:

```
rm(list =ls())

#Carga de Librerías
library(table1)
library(knitr)
library(ggplot2)
library(AER)
library(xtable)
library(dplyr)
library(Countr)
library(plyr)
library(AER)
library(effects)
library(psc1)
library(MASS)
library(moments)
library(modeest)
library(flextable)

#Carga de Los datos
setwd("D:/manar/Documents/TFM/Práctica")
datos <- read.csv("datostelemat.csv")
datos<-as.data.frame(datos)
names(datos)
n.var <-names(datos)

#La base de datos "datostelemat.csv" contiene datos telemáticos fictic
ios de conductores con pólizas PAYD. Esta base de datos está formada p
or:
nrow(datos) #pólizas
ncol(datos) #variables
#Comprobación de una correcta carga de Los datos
#Visualización de las 6 primeras filas
head(datos)
#Visualización de las 6 últimas filas
tail(datos)
#Descripción variables
kable(data.frame(Variables= n.var, Descripción= c("Género","Excesos de
velocidad por encima del límite permitido: Sí=excesos en promedio por
encima del 5% de kilómetros recorridos, No=excesos en promedio por deb
ajo del 5% de kilómetros recorridos","Conducción por vía urbana: Sí=co
nducción por vía urbana por encima del 25% de kilómetros recorridos, N
o=conducción por vía urbana por debajo del 25% de kilómetros recorrido
s","Número de siniestros durante el último año","Kilometraje anual", "
Antigüedad del carnet de conducir: 1=menos de dos años, 2=dos años o m
ás"))))

#Valores missing
sum(is.na(datos))
```

```

v.p <- names(datos)[!complete.cases(t(datos))]
v.p
#Tipo de variable estadística de cada variable
#Son variables cuantitativas :
toString(n.var[c(4,5)])
#Son variables cualitativas :
toString(n.var[c(1,2,3,5)])

#Asignación por R
res <- sapply(datos,class)
kable(data.frame(Variables=names(res),Clase=as.vector(res)))

#Cambio asignación variables
# Cambiamos la variable sex a factor
datos$sex <- factor(datos$sex, levels=c("woman","man"), labels=c("Mujer", "Hombre"))

# Cambiamos la variable excess a factor
datos$excess <- factor(datos$excess, levels=c("no","yes"), labels=c("No", "Sí"))

# Cambiamos la variable urban a factor
datos$urban <- factor(datos$urban, levels=c("no","yes"), labels=c("No", "Sí"))

# Cambiamos la variable antigc a factor
datos$antigc <- factor(datos$antigc, levels=c("1","2"))

#Visualizamos que la asignación sea la correcta:
res <- sapply(datos,class)
kable(data.frame(Variables=names(res),Clase=as.vector(res)))
#Resumen de la distribución de las variables
summary(datos)
#Estadísticas descriptivas
indic <- cbind(datos$y,"datos$km")
colnames(indic)<-c("y","km")

estad.basic <- function(x){
  est <- cbind(mean(x),median(x),var(x),sd(x),t(quantile(x)), diff(range(x)),IQR(x),skewness(x),kurtosis(x), sd(x)/abs(mean(x)))
  return(round(est ,2))}

options(warn=0)

est.total <- apply(indic,2,estad.basic)
rownames(est.total)<- c("media","mediana","var","desv.est","mín","Q1",
"Q2","Q3","máx","rango","IC","asimetría","curtosis", "coeficiente_variación")
print(est.total)

#Moda variable sex
mfv(datos$sex)

```

```

#Moda variable y
mfv(datos$y)
#Moda variable urban
mfv(datos$urban)
#Moda variable excess
mfv(datos$excess)
#Moda variable antigc
mfv(datos$antigc)
#Moda variable km
mfv(datos$km)

#Gráficos
#Variables cualitativas
#Género
a<-ggplot(datos)+
  geom_bar(aes(x = sex,fill=sex),colour="black",show.legend = FALSE)+
  scale_fill_manual(values=c("royalblue4","royalblue4"))+
  labs(x="" , y = "Número de pólizas")
a+theme_bw()
#Conducción urbana
b<-ggplot(datos)+
  geom_bar(aes(x = urban,fill=urban),colour="black",show.legend = FALSE)+
  scale_fill_manual(values=c("royalblue4","royalblue4"))+
  labs(x="" , y = "Número de pólizas")
b+theme_bw()
#Antigüedad de carnet
c<-ggplot(datos)+
  geom_bar(aes(x = antigc,fill=antigc),colour="black",show.legend = FALSE)+
  scale_fill_manual(values=c("royalblue4","royalblue4"))+
  labs(x="" , y = "Número de pólizas")
c+theme_bw()
#Exceso de velocidad
d<-ggplot(datos)+
  geom_bar(aes(x = excess,fill=excess),colour="black",show.legend = FALSE)+
  scale_fill_manual(values=c("royalblue4","royalblue4"))+
  labs(x="" , y = "Número de pólizas")
d+theme_bw()

#Variables cuantitativas
#Histograma y Boxplot kilometraje
hist(datos$km, probability = TRUE, ylab="Número de pólizas",xlab="Kilómetros anuales", main = "",
      col = "royalblue4", axes = FALSE)
axis(1) # Añade el eje horizontal
par(new = TRUE)
boxplot(datos$km, horizontal = TRUE, axes = FALSE,
        lwd = 2, col = rgb(0, 0, 0, alpha = 0.2))
#Histograma número de siniestros
hist(datos$y,main="",ylab="Número de pólizas",xlab="Número de siniestros",col="royalblue4")

```

```

# Valores outliers para la variable km
vars.cuantitativas <- c(5)
for(i in vars.cuantitativas ) {
  indices <- which(datos[,i] %in% boxplot.stats(datos[,i])$out)
  cat(names(datos)[i],"obs. outliers:", toString(indices), "\n" )
}

#Tablas
#Según ocurrencia de siniestro
y2<-ifelse(datos$y>0,"Siniestro","No siniestro")
table1(~y2|excess, data=datos,topclass="Rtable1-grid Rtable1-shade Rtable1-times")
table1(~y2|urban, data=datos,topclass="Rtable1-grid Rtable1-shade Rtable1-times")
table1(~y2|antigc, data=datos,topclass="Rtable1-grid Rtable1-shade Rtable1-times")
table1(~y2|sex, data=datos,topclass="Rtable1-grid Rtable1-shade Rtable1-times")

#Frecuencias absolutas y relativas
table1(~ excess, data=datos,topclass="Rtable1-grid Rtable1-shade Rtable1-times")
table1(~ urban, data=datos,topclass="Rtable1-grid Rtable1-shade Rtable1-times")
table1(~ antigc, data=datos,topclass="Rtable1-grid Rtable1-shade Rtable1-times")
table1(~ sex, data=datos,topclass="Rtable1-grid Rtable1-shade Rtable1-times")
table1(~ factor(y), data=datos,topclass="Rtable1-grid Rtable1-shade Rtable1-times")
table1(~ km, data=datos,topclass="Rtable1-grid Rtable1-shade Rtable1-times")

#Según el número de sinestros
table1(~ factor(y)|factor(excess), data=datos,topclass="Rtable1-grid Rtable1-shade Rtable1-times")
table1(~ factor(y)|factor(urban), data=datos,topclass="Rtable1-grid Rtable1-shade Rtable1-times")
table1(~ factor(y)|factor(antigc), data=datos,topclass="Rtable1-grid Rtable1-shade Rtable1-times")
table1(~ factor(y)|factor(sex), data=datos,topclass="Rtable1-grid Rtable1-shade Rtable1-times")
table1(~ km|factor(y), data=datos,topclass="Rtable1-grid Rtable1-shade Rtable1-times")

#Según el género
table1(~ urban + y+ km+ antigc + excess|sex, data=datos,topclass="Rtable1-grid Rtable1-shade Rtable1-times")
# Estimación Modelos
#Validación hipótesis Poisson
r <- c(mean(datos$y), var(datos$y))
c(mean=r[1], var=r[2], ratio=r[2]/r[1])

#Poisson con offset clásico

```

```

logkm<-log(datos$km)
Poi_clas<-glm(y~sex+urban+excess+antigc, data =datos, family=poisson(link="log"),offset=logkm)
summary(Poi_clas)
AIC(Poi_clas)

#Test de sobre o infra dispersión
dispersiontest(Poi_clas)

qchisq(0.95, df.residual(Poi_clas))
deviance(Poi_clas)
pchisq(deviance(Poi_clas),df.residual(Poi_clas),lower.tail=FALSE)

pr <- residuals(Poi_clas,"pearson")
phi1 <- sum(pr^2)/df.residual(Poi_clas)
round(c(phi1,sqrt(phi1)),4)

#Poisson con ofsset generalizado
lnkm<-log(datos$km)
Poi_gen<-glm(y~sex+urban+excess+antigc+logkm, data =datos, family=poisson)
summary(Poi_gen)

#Test de sobre o infra dispersión
dispersiontest(Poi_gen)

qchisq(0.95, df.residual(Poi_gen))
deviance(Poi_gen)
pchisq(deviance(Poi_gen),df.residual(Poi_gen),lower.tail=FALSE)

pr <- residuals(Poi_gen,"pearson")
phi2 <- sum(pr^2)/df.residual(Poi_gen)
round(c(phi2,sqrt(phi2)),4)

#Comparación bondad del ajuste entre modelos
ic <- data.frame(Modelo = c("Offset clásico", "Offset generalizado"),
AIC = c(AIC(Poi_clas), AIC(Poi_gen)), BIC = c(BIC(Poi_clas), BIC(Poi_gen)),
Deviance = c(deviance(Poi_clas),deviance(Poi_gen)), Pearson=c(phi1,phi2))
kable(data.frame(ic))

#Gráficos observado vs predicción
int <- seq(0,33950,by=1000)
y1<-c(tapply(datos$y,cut(datos$km,breaks=int),mean))
km1<-c(seq(0,33950,by=1000))
y2<-c(tapply(predict(Poi_clas,datos, type="response"),cut(datos$km,breaks=int),mean))
km2<-km1[-1]
y3<-c(tapply(predict(Poi_gen,datos, type="response"),cut(datos$km,breaks=int),mean))
data<-data.frame(km2,y1,y2,y3)

#Poisson con offset clásico
Poi_cl<-ggplot(data, aes(km2,y1,colour="Observado"))+

```

```

geom_point(aes(),size=1.5)+
geom_line(aes(y=y2,colour="Predicho"),size=0.75,linetype = "longdash
")+
labs(x="Kilómetros anuales" , y = "Frecuencia de siniestros")+
scale_color_manual(name="",values = c("Observado"="chocolate1","Pred
icho"="royalblue4"))
Poi_cl<-Poi_cl+theme_bw()
Poi_cl<-Poi_cl+theme(legend.position="bottom",legend.background = elem
ent_rect(colour = 'black', fill = 'white', linetype='solid'))
Poi_cl+ scale_y_continuous(breaks=seq(0, 0.4,0.05))+ scale_x_continuou
s(breaks=seq(0, 34000,5000))

#Poisson con ofsset generalizado
Poi_gn<-ggplot(data, aes(km2))+
geom_point(aes(y=y1,colour="Observado"),size=1.5,show.legend=FALSE)+
geom_smooth(aes(y=y3,colour="Smoothed")) +
labs(x="Kilómetros anuales" , y = "Frecuencia de siniestros")+
scale_color_manual(name="",values = c("Observado"="chocolate1","Pred
icho"="royalblue4","Smoothed"="darkolivegreen3"))
Poi_gn<-Poi_gn+theme_bw()
Poi_gn<-Poi_gn+theme(legend.position="bottom",legend.background = elem
ent_rect(colour = 'black', fill = 'white', linetype='solid'))
Poi_gn<-Poi_gn+ scale_y_continuous(breaks=seq(0, 0.4,0.05))+ scale_x_c
ontinuous(breaks=seq(0, 34000,5000))
Poi_gn+geom_line(aes(y=y3,colour="Predicho"),size=0.75,linetype = "lon
gdash")

#Interpretación de Los coeficientes
exp(coef(Poi_gen)[c(2,3,4,5)])
#
round((1-exp(coef(Poi_gen)[2]))*100,2)
#
round((1-exp(coef(Poi_gen)[3]))*100,2)
#
round((1-exp(coef(Poi_gen)[4]))*100,2)
#
round((1-exp(coef(Poi_gen)[5]))*100,2)

```