

Máster Universitario en Ciencias Actuariales y
Financieras 2016-2018

Trabajo Fin de Máster

**“Técnicas de pricing avanzadas para Seguros
Colectivos de Vida (Modelos GLM y
Credibilidad)”**

Paula Rivas de las Heras

Trabajo tutorizado por:

José Miguel Rodríguez-Pardo del Castillo

Jesús Ramón Simón del Potro

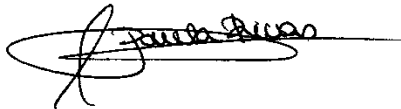
Esta tesis es propiedad del autor. No está permitida la reproducción total o parcial de este documento sin mencionar su fuente. El contenido de este documento es de exclusiva responsabilidad del autor, quien declara que no se ha incurrido en plagio y que la totalidad de referencias a otros autores han sido expresadas en el texto.

En caso de obtener una calificación igual o superior a 8.0 Notable, autorizo la publicación de este trabajo en el centro de Documentación de la Fundación Mapfre.

Sí, autorizo a su publicación.

No, desestimo su publicación.

Firmado:

A handwritten signature in black ink, appearing to read 'Javier Ruiz', is written over a horizontal line. The signature is stylized and somewhat cursive.

La modelización de riesgos en el mercado asegurador está cada vez más enfocada al uso de modelos predictivos, especialmente en Seguros No Vida. Sin embargo, es una técnica poco habitual en Seguros de Vida pues, para la tarificación de estos seguros se emplean las Tablas de Mortalidad.

En esta tesis se analizan métodos avanzados de tarificación para Seguros Colectivos de Vida, en concreto, se estudian Modelos Lineales Generalizados para variables dependientes limitadas, como son los modelos Probit y Logit, y los Modelos de Credibilidad de distribución libre, Bühlmann y Bühlmann-Straub.

Se realiza una aplicación práctica de estos modelos, utilizando la base de datos real de la cartera de seguros Temporales Anuales Renovables con garantía de fallecimiento de una Aseguradora de Vida. Para la elaboración de estos modelos se estudian variables biológicas como la edad y sexo de los asegurados, y otras variables relacionadas con la empresa tomadora del seguro, como la actividad de la misma y su categorización en función del número de asegurados.

Se concluye que, en este tipo de seguros, la edad del asegurado y el tamaño de la empresa no son variables de riesgo significativas para determinar la probabilidad de fallecimiento, en un modelo que explica, aproximadamente, el 95% de la siniestralidad real.

***Palabras clave:** Modelos predictivos, Modelos Lineales Generalizados, Credibilidad, Seguros de Vida.*

***Keywords:** Predictive models, Generalized Linear Models, Credibility, Life Insurance.*

Índice

Índice	7
1. Introducción al Seguro de Vida Colectivo:	9
2. Vida Riesgo Colectivos. Seguro Temporal Anual Renovable:	11
3. Mercados imperfectos e información asimétrica:	12
3.1. Riesgo Moral:	12
3.2. Selección adversa:	13
4. Tarificación Seguros de Vida Riesgo Colectivos:	13
5. Modelos de regresión para variables dependientes limitadas:	17
5.1. Modelo de Regresión Lineal: Mínimos Cuadrados Ordinarios.	18
5.1.1. Supuestos Clásicos de los Estimadores de Mínimos Cuadrados Ordinarios:	18
5.1.2. Limitaciones del Modelo Mínimos Cuadrados Ordinarios:	19
5.1.3. Errores heterocedásticos:	20
5.1.4. Test de Errores de Especificación: Test Reset de Ramsey.	21
5.1.5. Test de Heterocedasticidad: White.	21
5.1.6. Contraste de Heterocedasticidad: Breusch-Pagan	22
5.1.7. Contraste de normalidad: Jarque-Bera.....	22
5.1.8. Contraste de normalidad: Shapiro-Wilk.....	23
5.2. Modelos Lineales Generalizados: Probit y Logit.	24
5.2.1. Modelo Logit	24
5.2.2. Modelo Probit	25
5.3. Significatividad de los coeficientes:	25
5.3.1. Significatividad individual de los coeficientes:.....	25
5.3.2. Significatividad de los coeficientes: Estadístico Wald.....	26
5.3.1. Significatividad conjunta de los coeficientes: Test de Lagrange.	27
6. Introducción a la Teoría de la Credibilidad:	27
6.1. Modelo de Bühlmann:	28
6.2. Modelo de Bühlmann-Straub:	31
7. Aplicación práctica de Modelos Lineales Generalizados y Modelos de Credibilidad a la Tarificación de Seguros de Vida:.....	34
7.1. Aplicación de Modelos de Regresión a la Tarificación de Seguros de Vida:	35

7.1.1. Modelo de Regresión 1: Mínimos Cuadrados Ordinarios.	36
7.1.2. Modelo de Regresión 2: Probit.	38
7.1.3. Modelo de Regresión: Modelo Logit.	41
7.1.4. Prueba de aciertos:.....	43
7.1.5. Prima de riesgo individual:	45
7.1. Aplicación de Modelos de Credibilidad a la Tarificación de Seguros de Vida: .	46
7.1.1. Programación: Modelos Bühlmann y Bühlmann-Straub.	46
7.1.2. Prueba de aciertos.....	52
7.1.3. Resultados: Modelos Bühlmann y Bühlmann-Straub.	56
8. Comparativa: Modelos Lineales Generalizados, Modelos de Credibilidad y Modelo de Tarificación Tradicional.	58
9. Gestión de cartera: Modelos Lineales Generalizados y Modelos de Credibilidad....	62
10. Conclusiones.....	63
11. Próximas líneas de investigación:	64
12. Bibliografía:.....	65

1. Introducción al Seguro de Vida Colectivo:

Un contrato de seguro es aquel por el que se establece la transferencia total o parcial, a otra entidad, de las consecuencias económicas producidas por determinadas contingencias objeto de cobertura por este contrato. El contrato de seguro se materializa en una póliza y su precio se denomina prima de seguro.

Las personas, generalmente, son adversas al riesgo, es decir, presentan actitud de rechazo hacia el riesgo. El concepto de aversión al riesgo es comúnmente empleado en la teoría microeconómica, rama en la que se estudian tres actitudes frente al riesgo en base a la función de utilidad esperada, de manera que cada individuo será adverso, neutral o amante del riesgo si su función de utilidad esperada es cóncava, afín o convexa, respectivamente.

Para minimizar el riesgo existen numerosos instrumentos que ofrecen seguridad y, el contrato de seguro es uno de ellos. Al tratarse de un mecanismo de transferencia del riesgo, entendiendo el riesgo como un suceso previsible que ocurre de forma accidental o fortuita pudiendo generar pérdida o daño. Aparece aquí el concepto de siniestro, que puede definirse como el suceso cuya ocurrencia da lugar al pago de la prestación, en los términos establecidos en la póliza de seguro. La persona física o jurídica que recibe la prestación es el beneficiario que, generalmente, coincide con la figura de asegurado salvo en los seguros de vida en los que la figura de beneficiario y asegurado no siempre coincide, como es el caso de los siniestros causados por fallecimiento.

Por lo tanto, las entidades aseguradoras asumen los riesgos a cambio de una prima, cuyo importe debe destinarse a un fondo que garantice el pago de los siniestros objeto de cobertura del seguro, debe además cubrir los gastos derivados de la actividad industrial y debe generar beneficios como remuneración al capital invertido en la empresa. Las entidades de seguro cuya cartera está compuesta por un número de asegurados lo suficientemente grande, son capaces de predecir la ocurrencia de los siniestros y obtener beneficios a partir de las primas captadas, mediante una correcta tarificación y gestión del riesgo.

En el mundo actuarial los seguros se dividen en dos clasificaciones, Vida y No Vida. En este trabajo el objeto de estudio son los Seguros de Vida.

Los Seguros de Vida Colectivos son seguros dirigidos a un conjunto de individuos que mantienen un vínculo profesional. En estos seguros el Tomador es una empresa o asociación y el colectivo asegurado mantiene un vínculo con el Tomador. Los seguros colectivos, por tanto, dan cobertura individual a todo un colectivo.

En los Seguros de Vida, se distinguen cinco figuras: Aseguradora, Tomador, Asegurado, Beneficiario y Figura de Intervención, en los supuestos en que la póliza este intervenida,

ya sea por un corredor, agente, delegado... A diferencia de los Seguros de Vida Individuales, en los Seguros de Vida Colectivos la figura de Tomador no coincide con la de Asegurado, si no que el Tomador media en interés del grupo cuyos miembros pueden o no mantener relación laboral con el Tomador, e instrumentar o no la póliza un Compromiso por Pensiones.

En cuanto a la finalidad de cobertura de los Seguros de Vida, se pueden diferenciar dos grupos, los seguros de Vida Riesgo y los seguros de Vida Ahorro. Los Seguros de Vida Riesgo cubren, como garantía principal, el fallecimiento del asegurado ya sea fallecimiento natural o accidental, siempre y cuando no haya sido producido por algún riesgo excluido por la póliza. Por otro lado, los Seguros de Vida Ahorro, estos seguros garantizan el pago de una prestación al asegurado alcanzada determinada edad. Por último, destacar los seguros Mixtos, que combinan ambas contingencias.

A continuación, se describen las distintas tipologías de seguro:

Seguro de Capital Diferido: Seguro que garantiza el pago de la prestación si el asegurado sobrevive en el transcurso de “n” años, a contar desde la fecha de emisión de la póliza.

Seguro Diferido “m” años: Seguro que garantiza el pago de la prestación si el asegurado fallece transcurridos “m” años, a contar desde la emisión de la póliza.

Seguro Temporal “n” años: Seguro que garantiza el pago de la prestación en caso de fallecimiento del asegurado durante los “n” años de duración del seguro.

Seguro Mixto: Seguro que garantiza el pago de la prestación ya sea después del fallecimiento del asegurado o de la supervivencia del mismo si vive “n” años, lo que ocurra primero.

Seguro Vida Entera: Seguro que garantiza el pago de la prestación en caso de fallecimiento del asegurado. Este seguro finaliza con el pago de la prestación.

Tanto el Seguro Mixto como el Seguro Vida entera, a diferencia del resto de seguros anteriormente descritos, siempre generan el pago de una prestación.

En los seguros colectivos, a excepción de los Seguros Vida Entera, la edad de cese suele estar en torno a los 65-70 años, es decir, se limita a edad laboral.

Esta tesis se centra en el estudio de métodos de tarificación avanzados para los Seguros de Vida Riesgo Colectivos, más concretamente, en los Seguros Temporales Anuales Renovables (TAR), que es la modalidad más comercializada. Atendiendo a los últimos datos de ICEA publicados en mayo de 2018 sobre el mercado de seguros colectivos de vida, el 81,83% del total de asegurados de Vida Colectivo figuran en Seguros de Vida Riesgo y, en cuanto a volumen de primas netas emitidas hasta 31 de marzo, el 62,06% corresponden a esta modalidad de seguro.

En el siguiente apartado se detallan las diferentes peculiaridades de estos Seguros, pues se adaptan a distintas duraciones de cobertura, duración del pago de primas, forma de cobro de la prestación, coberturas aseguradas, gestión de la póliza...

2. Vida Riesgo Colectivos. Seguro Temporal Anual Renovable:

El Seguro Temporal Anual Renovable es un Seguro de Vida Temporal que garantiza el pago de una prestación a los beneficiarios, en caso de fallecimiento del asegurado durante la vigencia del contrato de seguro. Este seguro renueva anualmente de forma automática, es decir, llegado el vencimiento del seguro, salvo notificación expresa de las partes, el seguro se renueva por una anualidad más. En cuanto a los plazos para comunicar la no renovación, se estará en lo dispuesto en el Artículo 22 de la Ley de Contrato de Seguro, que establece un período mínimo de un mes antes de la finalización del contrato, en caso de que quién se oponga a la prórroga sea el tomador, y amplía a dos meses este plazo en caso de que quién se oponga sea la entidad aseguradora.

Como se ha indicado anteriormente, el Seguro Temporal Anual Renovable, garantiza el pago de la prestación en caso de fallecimiento del asegurado. No obstante, los Seguros de Vida Riesgo, también contemplan otras garantías como las Invalideces Permanentes, en cualquiera de sus grados.

Generalmente, los seguros de Vida Riesgo Colectivos se desarrollan mediante una gestión nominada del colectivo, es decir, para dar cobertura al colectivo, la aseguradora debe disponer de información de cada uno de los asegurados, Nombre, Apellidos y DNI. En ocasiones, aunque no es muy común en Seguros Colectivos, se solicita información adicional como declaraciones de salud, requisitos médicos... Mediante esta gestión de la póliza la aseguradora puede disponer en cada momento de información sobre el colectivo, respecto a edades, garantías aseguradas, capitales... Disponer de esta información permite, además de una mejor valoración del riesgo, generar una importante base de datos para la aseguradora que puede aplicar a la tarificación de estos seguros.

Además de la gestión nominada, también existe la gestión innominada de las pólizas, que está enfocada a colectivos con un elevado número de asegurados o empresas con un alto grado de rotación del colectivo.

Por último, destacar la gestión de pólizas mediante convenio. Esta gestión esta enfocada a dar cobertura a convenios públicos, de manera que se agrupa al colectivo asociado a dicho convenio, permitiendo una actualización automática de las garantías y capitales asegurados establecidos por el convenio.

En cuanto a la forma de cobro de la prestación, este tipo de seguro permite el cobro de la prestación en forma de capital o de rentas.

3. Mercados imperfectos e información asimétrica:

En mercados que no son perfectos, el actuario puede no disponer de toda la información necesaria para una correcta valoración del riesgo, lo que supone una gran problemática en el proceso de tarificación.

En mercados imperfectos, con información asimétrica, aparecen los problemas de riesgo moral¹ y selección adversa², que son términos aplicables tanto en el mercado de seguro, como en la teoría microeconómica, la teoría macroeconómica...

Este concepto tiene sus orígenes en Inglaterra, donde las aseguradoras empleaban este término para referirse al riesgo que éstas asumían al no disponer de información perfecta sobre los asegurados y no poder, por lo tanto, realizar una correcta valoración de dicho riesgo. Posteriormente, Adam Smith introdujo este concepto en el campo de la economía.

3.1. Riesgo Moral:

El riesgo moral se da en situaciones en las que, una vez celebrado el contrato de seguro, el comportamiento del asegurado cambia y asume más riesgos, dado que el contrato de seguro puede desincentivar la prevención. En consecuencia, la probabilidad de que se produzca la contingencia cubierta por la póliza aumenta, generando efectos negativos para la entidad de seguro. Por este motivo, las compañías aseguradoras han empleado diferentes instrumentos para eliminar este riesgo, como: franquicias, planes de bonificaciones, exclusión de riesgos...

Estos mecanismos para incentivar la prevención son útiles para los Seguros No Vida, en los que el riesgo moral supone una grave problemática para las aseguradoras. En cuanto a los Seguros de Vida, que cubren contingencias relacionadas con la supervivencia, fallecimiento e invalideces y, por lo tanto, bienes insustituibles como son la vida o la salud, el riesgo moral supone un riesgo de menor relevancia.

1 RIESGO MORAL: Concepto que describe la situación en que el comportamiento del agente no es verificable o situación en que el agente recibe información privada, una vez iniciado el contrato.

2 SELECCIÓN ADVERSA: Concepto que describe la situación en que una de las partes contratantes está menos informada, no pudiendo realizar una correcta valoración de lo que oferta la parte contraria.

3.2. Selección adversa:

La información asimétrica lleva también a un proceso de selección adversa, de manera que, asegurados que presentan mayor probabilidad de ocurrencia de la contingencia cubierta por la póliza son, a su vez, más propensos a contratar el seguro. Esto supone un riesgo de gran importancia para las entidades aseguradoras que, al no disponer de información suficiente para valorar correctamente el riesgo a asegurar, establecen una prima de seguro inferior al riesgo que soportan.

Por otro lado, considerando que no se dispone de información perfecta, también puede suceder que asegurados que presentan una exposición al riesgo inferior a la de la prima ofertada no contraten los seguros, pues pueden no estar dispuestos a soportar primas superiores a su nivel de riesgo. Si las entidades de seguro pudieran disponer de información perfecta, podrían realizar una correcta tarificación, de manera que, la prima del seguro se ajustase perfectamente al riesgo soportado por la póliza de seguro.

Por tanto, la selección adversa o antiselección, se refiere a la tendencia que presentan individuos con mayor exposición al riesgo a contratar seguros, lo que supone un riesgo de gran relevancia en los Seguros de Vida. Un método para eliminar este riesgo consiste en llevar a cabo un proceso de selección de riesgos antes de formalizar definitivamente el contrato de seguro, tal y como se procede en los Seguros de Vida Individuales. En cuanto a los Seguros de Vida Colectivos, cabe destacar que este problema se elimina en gran parte, pues la mayoría de seguros instrumentan un Compromiso por Pensiones asegurando a todos los empleados o una colectividad definida, con unas garantías y capitales especificadas en este compromiso.

4. Tarificación Seguros de Vida Riesgo Colectivos:

La Tarificación de los Seguros consiste en el cálculo de una prima justa y suficiente para cubrir, al menos, el valor esperado de los siniestros y el importe de los gastos derivados de la gestión de estos contratos.

Este proceso de Tarificación, comienza con el cálculo de la denominada “prima pura” que se corresponde con el valor esperado del riesgo, matemáticamente:

$$P = E(S)$$

Siendo P la prima pura y $E(S)$ el valor esperado de los siniestros. Por lo tanto, la prima pura puede definirse como el valor esperado actual actuarial del riesgo que se le atribuye al asegurado, en función de técnicas estadísticas y del tipo de interés técnico que corresponda. Aplicando al proceso de Tarificación la Ley de los Grandes Números, es decir, el conjunto de teoremas que explican que, a medida que se repite un experimento,

la probabilidad de dicho suceso se aproxima a su frecuencia relativa; la siniestralidad media se aproxima a la prima pura.

Además de la prima pura, en el proceso de tarificación se distingue, la prima de inventario, que se determina recargando la prima pura con los gastos de administración de la entidad aseguradora, y la prima neta o de tarifa, que incorpora, además, los gastos de gestión externa. Matemáticamente,

$$PRIMA\ NETA = \frac{P}{1 - ggi - gge}$$

Siendo, *ggi* los gastos de gestión interna sobre la prima, *gge* los gastos de gestión externa sobre la prima, y *P* la prima pura determinada anteriormente.

Una vez determinada la prima de tarifa o prima neta, se calcula la prima total, que incorpora, además, los impuestos y recargos legales correspondientes, es decir, CCS, LEA e IPS.

En el proceso de Tarificación, cabe destacar la Selección de Riesgos, es decir, el proceso de análisis y valoración del riesgo individual del colectivo a asegurar, que sirve de apoyo al proceso de tarificación descrito anteriormente. La Tarificación tradicional de los Seguros de Vida, por ejemplo, se apoya en Tablas de Mortalidad cuyas probabilidades de mortalidad y supervivencia se determinan en función de la edad y sexo. En el proceso de Selección de Riesgos se analizan, además, otros factores, como pueden ser, el estado de salud del individuo, profesión, actividades deportivas... Es decir, realiza un reconocimiento previo del asegurado para evitar la antiselección.

Durante este proceso, se detecta la existencia de riesgos agravados y, mediante la aplicación de sobreprimas o exclusiones del riesgo, se trata de aproximar la prima al valor esperado de los siniestros. Cabe destacar, entre el resto de tipos de riesgo, los riesgos considerados catastróficos, es decir, aquellas empresas que, por la actividad que desempeñan, pueden generar el pago de varios siniestros como consecuencia de un mismo evento siniestral. En este tipo de riesgos, además, del análisis del Departamento de Selección de Riesgos, entra la actuación del Reaseguro.

Por lo tanto, el proceso de Tarificación se complementa con el proceso de Selección de Riesgos. Las sobreprimas pueden aplicarse sobre el capital o sobre la prima y se pueden distinguir dos tipos:

Ocupacionales: Recargos por el tipo de actividad que desempeña el asegurado.

Médicas: Se establecen en función del estado de salud del asegurado.

El proceso de selección de riesgos, trata de ajustar la siniestralidad real a la siniestralidad esperada. Por este motivo, es un aspecto de gran importancia para las entidades de seguro pues, tiene consecuencias en los resultados del negocio.

En los Seguros de Vida Riesgo Colectivos se solicitan menos requisitos de selección que en los Seguros Vida Riesgo Individuales, ya que, por ejemplo, cuando se trata de colectivos con un elevado número de asegurados, solicitar requisitos de selección a cada uno de ellos no siempre sería posible. Además, habitualmente en las pólizas colectivas, la colectividad asegurable está definida, por lo que se elimina en gran parte el problema de selección adversa. Sin embargo, este problema si existe en pólizas colectivas de adhesión voluntaria o pólizas que no instrumentan compromiso por pensiones, es decir, pólizas en las que la colectividad asegurable, capitales y garantías no están definidas y, por este motivo, para este tipo de pólizas en concreto, se hace más necesario el proceso de selección de riesgos.

La Tarificación de los Seguros, trata de proyectar el comportamiento futuro de la mortalidad de una población, dejando a un lado la experiencia propia de cada Entidad Aseguradora, el instrumento del que se dispone para medir este fenómeno es la tabla de mortalidad. La tarificación tradicional de los Seguros de Vida se basa en el uso de Tablas de Mortalidad para modelizar la probabilidad de supervivencia o fallecimiento de un asegurado o un colectivo. Además, en la tarificación de los Seguros de Vida Riesgo Colectivos, se tiene en cuenta la siniestralidad ocurrida en las ultimas anualidades de seguro, diferencia con respecto a la Tarificación de los Seguros de Vida Riesgo Individuales.

Aunque, de manera general, se habla de tablas de mortalidad, debe realizarse una distinción entre las tablas de mortalidad y las tablas de supervivencia. Las tablas de mortalidad contienen el número de individuos fallecidos a cada edad, partiendo de un colectivo inicial de "L" individuos, con edad de inicio preestablecida. Por otro lado, las tablas de supervivencia contienen el número de individuos vivos a cada edad, partiendo de un colectivo inicial de "L" individuos, con edad de inicio preestablecida. Por lo tanto, las tablas de mortalidad pueden obtenerse a partir de tablas de supervivencia.

Se definen las siguientes variables:

$l_x \rightarrow$ *Número de personas vivas a la edad "x"*

$p_x \rightarrow$ *Probabilidad anual de supervivencia de una persona de edad "x"*

$q_x \rightarrow$ *Probabilidad anual de fallecimiento de una persona de edad "x"*

Matemáticamente, los tantos anuales de mortalidad, q_x , y supervivencia, p_x , se calculan mediante las siguientes ecuaciones,

$$p_x = l_{x+1}/l_x$$

$$q_x = 1 - p_x = 1 - l_{x+1}/l_x$$

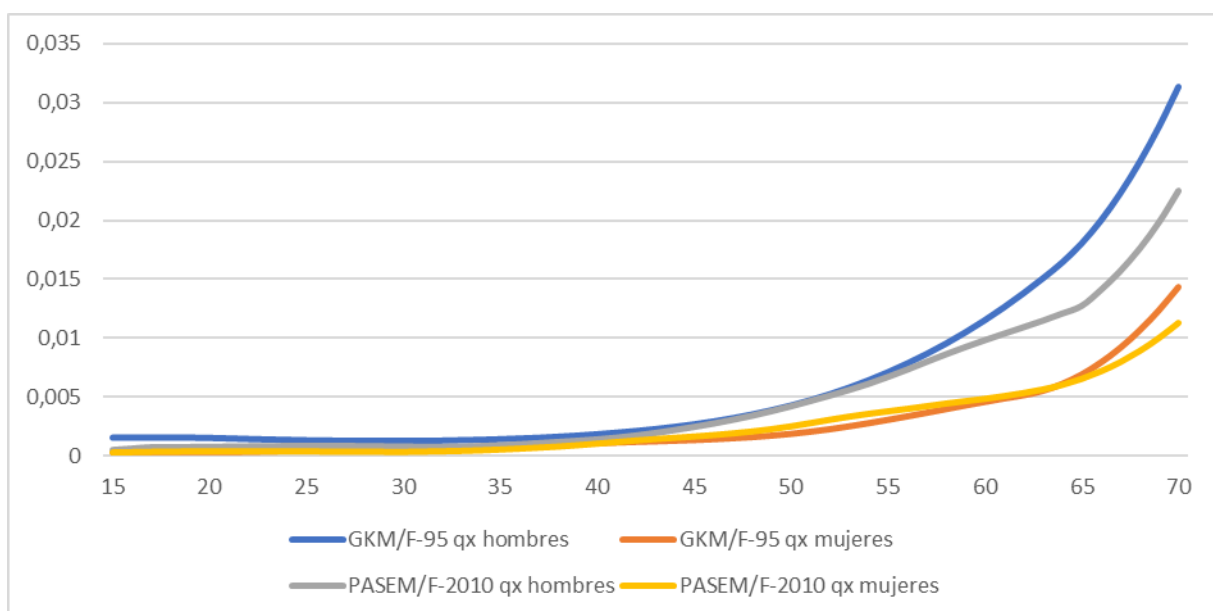
Las tablas biométricas de mayor uso en los seguros de vida son las Tablas de Mortalidad PASEM/F-2010 y, para la supervivencia, las Tablas PERMF-2000 P. A partir del día 1 de enero de 2013, para el cálculo de las primas de seguro de nueva suscripción con

garantía de fallecimiento se emplean las Tablas de Mortalidad PASEM/F-2010, que sustituyen a las Tablas de GKM/F-95 que venían utilizándose. Cabe destacar que estas tablas se han elaborado a partir de datos de población asegurada.

Las Tablas de Mortalidad GKM/F son tablas de experiencia suiza “GruppenKapitalversicherung – Männer / Frauen”. En concreto, las Tablas GKM/F-95 se elaboraron a partir de datos correspondientes al período 1986-90 de 3.800.000 varones y 1.540.000 mujeres. La edad máxima de esta tabla son los 120 años para el caso de los varones y 126 años para las mujeres. En estas tablas se observa un aumento en la probabilidad de fallecimiento en el tramo de los 15 a los 17 años, explicado por los accidentes, este fenómeno es conocido como “joroba de la mortalidad”.

A diferencia de las Tablas GKM/F, la Tabla de Mortalidad PASEM/F-2010 se han elaborado con muestras de la población asegurada española, en concreto, a partir de las carteras de fallecimiento de entidades españolas durante el período de observación 2003-2007.

Realizando una comparativa de la Tabla GKM/F-95 con la PASEM/F-2010, se puede observar que, para edades hasta los 40 años, aproximadamente, la probabilidad de fallecimiento es inferior en la Tabla PASEM/F-2010. Por otro lado, para el tramo de edades de los 42 a los 63, se verifica que las probabilidades de fallecimiento de mujeres son más elevadas que en la GKF-95. A continuación se muestra un gráfico comparativo.



Fuente: Elaboración propia. Este gráfico muestra la comparativa entre las curvas de las Tablas GKM/F-95 y PASEM/F-2010.

En cuanto a la diferenciación por género, en diciembre de 2014 se aprobó la Directiva 2004/113/CE que establece el principio de igualdad de trato entre hombres y mujeres en el acceso a bienes y servicios y su suministro. El Artículo 5 de esta Directiva, regula el uso del género como factor actuarial para el cálculo de las primas y prestaciones de seguro. En el ámbito de seguros, esta Directiva establece que el género no puede dar lugar a diferencias ni las primas ni en las prestaciones, por lo tanto, el sexo no puede utilizarse como factor actuarial para tarificación de los seguros. No obstante, se permite el uso del género para el cálculo de las primas y prestaciones a nivel agregado, siempre y cuando, el uso de este factor no genere diferencias a nivel individual.

La experiencia en el mercado asegurador confirma la tarificación por diferenciación de sexos en Seguros Colectivos de Vida.

5. Modelos de regresión para variables dependientes limitadas:

En esta parte de la tesis se estudian Modelo Lineales Generalizados (GLM) para tarificar los Seguros de Vida Colectivos, de manera que la variable dependiente, también denominada variable explicada, respuesta o predicha, es la probabilidad de fallecimiento con temporalidad un año de un individuo. Esta variable tiene alcance limitado, pues solo puede tomar valores positivos y finitos, comprendidos entre 0 y 1.

Generalmente, las variables dependientes limitadas como tipos de interés, precios... no requieren un modelo econométrico especializado, sin embargo, variables dependientes como las probabilidades que están restringidas a valores no negativos y no superiores a 1, es posible que requieran de modelos especiales pues, aunque el Modelo de Regresión Lineal no es inapropiado para modelar este tipo de variables, presenta inconvenientes, el principal es que resulta muy difícil obtener conclusiones ceteris paribus de cómo afecta cada variable dependiente. Estos modelos parten de la premisa de que los factores que afectan a la variable dependiente no están correlacionados con ninguna variable explicativa y, en ocasiones, no es un supuesto realista. Otro inconveniente del modelo lineal es que las probabilidades ajustadas pueden alcanzar valores superiores a 1 o inferiores a 0.

Por lo tanto, se plantean modelos en los que Y_i es la estimación de la probabilidad de que ocurra el suceso $Y_i = 1$.

Los Modelos Lineales Generalizados (GLM) están formados por tres componentes, una variable respuesta (componente aleatorio), variables explicativas (componente sistemática) y una función enlace.

Entre las variables explicativas existe una relación lineal. Sin embargo, esto no siempre ocurre entre el componente sistemático (variables explicativas), y el componente

aleatorio (variable dependiente). La función enlace $g(\cdot)$ es la relación entre estos dos componentes, de manera que, $g(\cdot)$ puede dar lugar a modelos en los que ambos componentes tienen una relación lineal, como en el modelo clásico, cuando toma valor μ , y a modelos en los que dicha relación no es lineal, como el Modelo Probit o el Modelo Logit, cuando $g(\cdot)$ es la función de densidad de la normal estándar o función logística, respectivamente.

5.1. Modelo de Regresión Lineal: Mínimos Cuadrados Ordinarios.

En el caso que se está estudiando la variable dependiente es binaria y se regresa como una función de variables explicativas x_i .

$$Y_i = \beta_0 + x_i\beta + \varepsilon_i$$

Los coeficientes del Modelo de Regresión Lineal recogen el efecto sobre la probabilidad de que se cumpla el suceso $Y_i = 1$, de una variación unitaria en cada variable explicativa. La relación entre las variables explicativas, x_i , y la variable respuesta, y_i es lineal.

En los modelos con variable dependiente dicotómica, bajo el supuesto de esperanza del error igual a cero $E(\varepsilon = 0)$, el valor esperado de la variable explicativa es:

$$E(y_i) = 0 \cdot P_i + 1 \cdot (1 - P_i) = P_i$$

Por lo tanto,

$$E(y_i|x_i) = P_i$$

Obteniendo la siguiente restricción, que implica que no puedan obtenerse valores superiores a 1 ni inferiores a 0.

$$0 \leq E(Y_i|X_i) \leq 1$$

Esta limitación que presentan los modelos con variable dependiente binaria, presenta dificultades a la hora de regresar empleando un Modelo de Regresión Lineal.

5.1.1. Supuestos Clásicos de los Estimadores de Mínimos Cuadrados Ordinarios:

En este apartado se describen los supuestos clásicos establecidos por el Teorema de Gauss-Márkov, para los modelos lineales:

El primer supuesto hace referencia a la Linealidad de los parámetros. Este supuesto establece la existencia de una relación lineal entre los parámetros y la variable

dependiente. En caso de no existir dicha relación lineal, se dice que hay un error de especificación en el modelo.

El segundo supuesto establece que las observaciones del modelo deben ser seleccionadas mediante un muestreo aleatorio simple. Esta técnica garantiza que todas las observaciones tienen la misma probabilidad de ser seleccionados y, en consecuencia, se garantiza una muestra representativa de la población, pues, el único error que puede producirse en el modelo se debe al azar.

El tercer supuesto es la independencia del error condicional a las variables explicativas del modelo, matemáticamente, debe satisfacerse la siguiente condición,

$$E(\varepsilon|x_1, \dots, x_n) = 0$$

El cuarto supuesto es el supuesto de variación muestral en la variable independiente, que requiere la existencia de variaciones de la muestra en las variables x_i .

Por último, el quinto supuesto, el supuesto de homocedasticidad condicional, es decir, que la esperanza del error condicionada a las variables explicativas es constante,

$$Var(\varepsilon|x_1, \dots, x_n) = \sigma^2$$

Conforme a estos cinco supuestos clásicos, se pueden definir las propiedades de los estimadores, es decir, de los $\hat{\beta}_i$. De manera que, si se cumplen los cuatro primeros supuestos se puede afirmar que los estimadores del Modelo de Mínimos Cuadrados Ordinarios son estimadores lineales insesgados. El cumplimiento del último supuesto, implica, además, que los estimadores sean óptimos, es decir, de mínima varianza. Existe una nomenclatura comúnmente empleada cuando se cumplen los cinco supuestos, caso en el que se define a los estimadores como ELIO, es decir, Estimadores Lineales, Insesgados y Óptimos.

5.1.2. Limitaciones del Modelo Mínimos Cuadrados Ordinarios:

Los modelos de probabilidad estimados mediante Modelos de Regresión Lineal miden la probabilidad condicional de que un suceso, en este caso, la probabilidad de fallecimiento, ocurra. Esta probabilidad se mide a través de la esperanza de la variable dependiente, condicionada a las variables explicativas. Dado que las probabilidades son variables limitadas que toman valores comprendidos entre 0 y 1, el término $E(y|x)$ debe estar limitado por dichos valores, realizando estimaciones por Mínimos Cuadrados Ordinarios (MCO), no puede garantizarse con total certeza que, para ciertos valores de variables explicativas, siempre se obtengan valores de la variable dependiente dentro de dicho rango. Este es el principal inconveniente de este tipo de Modelos, que pueden obtenerse valores de probabilidad negativos o mayores al 100%.

Otro inconveniente que presentan las estimaciones por MCO, es que los errores no siguen una distribución Normal. No obstante, esta restricción no invalida el modelo, pues los estimadores continuarían siendo estimadores lineales, insesgados y óptimos, pero pierden eficiencia entre los estimadores insesgados.

El último inconveniente a destacar, es la heterocedasticidad del error, pues la varianza del error no es constante, por lo tanto, aunque los estimadores sean insesgados, hay presencia de heterocedasticidad.

5.1.3. Errores heterocedásticos:

Tal y como se ha desarrollado en los anteriores apartados, uno de los supuestos clásicos en los que se basa el Modelo de Regresión Lineal es el supuesto de que la varianza del error condicional a los regresores del modelo es constante y, por lo tanto, supone que los errores son homocedásticos.

En este tipo de modelo que se estudia en este apartado de la tesis, los errores no son constantes, es decir, no se cumple la siguiente expresión:

$$\text{Var}(\varepsilon_i | x_i) = \sigma^2$$

Por lo tanto, aunque los estimadores del modelo estimado por MCO sean consistentes e insesgados, ni los errores estándar, ni los estimadores de la varianza lo son. Cuando esto ocurre, las pruebas de hipótesis quedan invalidadas.

Una posible solución es ajustar el modelo introduciendo errores estándar robustos a la heterocedasticidad. A continuación, se define matemáticamente el error estándar robusto a la heterocedasticidad para el estimador $\hat{\beta}_j$:

$$\sqrt{\widehat{\text{Var}}(\hat{\beta}_j)} = \left(\frac{\sum_{i=1}^n \hat{r}_{ij}^2 \hat{\varepsilon}_i^2}{SCT^2} \right)^{\frac{1}{2}}$$

Donde, $SCT = \sum (y_i - \bar{y})^2$, \hat{r}_{ij} es el residuo i -ésimo de regresar x_j sobre el resto de variables explicativas en un modelo de regresión múltiple, y $\hat{\varepsilon}_i$ los errores estándar del modelo inicial.

Incorporando estos errores al modelo, se pueden realizar las pruebas de hipótesis T-Student, Estadístico F...

5.1.4. Test de Errores de Especificación: Test Reset de Ramsey.

El Test Reset de Ramsey es un test de errores de especificación. Mediante la ejecución de este test se puede detectar si la forma funcional propuesta es incorrecta, es decir, permite contrastar la linealidad del modelo.

En primer lugar, se regresa el modelo de Mínimos Cuadrados Ordinarios,

$$Y_i = \beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki} + \epsilon_i$$

Posteriormente, se estima el siguiente modelo ampliado,

$$Y_i = \beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki} + \alpha_2 \hat{Y}^2 + \alpha_3 \hat{Y}^3 + \dots + \epsilon_i$$

Una vez regresados estos modelos, se contrasta si las potencias incluidas en el modelo ampliado se anulan, es decir, se contrasta si son significativas. El contraste de Ramsey realiza una prueba de significación de estos coeficientes y, en caso de ser significativos, entonces el modelo presenta de errores de especificación. El estadístico que se utiliza para realizar este contraste sigue una distribución $F_{q,n-q-k}$.

5.1.5. Test de Heterocedasticidad: White.

El Test White es una prueba general para detectar la existencia de heterocedasticidad en un modelo de regresión lineal. Esta prueba no requiere establecer supuestos ni determinar las variables explicativas de la heterocedasticidad. El procedimiento que lleva a cabo este test es el siguiente:

En primer lugar, se estima el modelo mediante Mínimos Cuadrados Ordinarios y se calcula la serie de errores.

$$Y_i = \beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki} + \epsilon_i$$

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \dots + \hat{\beta}_k X_{ki}$$

$$e = Y - \hat{Y}$$

Se regresan los residuos del modelo al cuadrado a partir de los regresores del modelo, estos al cuadrado y el producto de cada dos regresores del modelo. De este modelo se obtiene R^2 :

$$e_i^2 = \delta_0 + \delta_1 x_1 + \dots + \delta_k x_k + \gamma_1 x_1^2 + \dots + \gamma_k x_k^2 + \eta_1 x_1 x_2 + \eta_1 x_1 x_k + \dots + \epsilon_i$$

A continuación, se lleva a cabo el siguiente contraste:

$$H_0: \sigma_i^2 = \sigma^2 \rightarrow \text{Homocedasticidad.}$$

$$H_1: \sigma_i^2 \neq \sigma^2 \rightarrow \text{Heterocedasticidad.}$$

$$nR^2 \sim \chi_{k-1}^2$$

Donde k indica el número de variables explicativas del último modelo.

5.1.6. Contraste de Heterocedasticidad: Breusch-Pagan

Al igual que con el Test de White, con el Test de Breusch-Pagan se contrasta la existencia o no de heterocedasticidad.

En primer lugar, se estima el modelo mediante Mínimos Cuadrados Ordinarios y se calcula la serie de errores.

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki} + \epsilon_i \\ \hat{Y}_i &= \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \dots + \hat{\beta}_k X_{ki} \\ e &= Y - \hat{Y} \end{aligned}$$

Se obtienen los errores al cuadrado del modelo anteriormente indicado y se estima la siguiente regresión:

$$\epsilon^2 = \delta_0 + \delta_1 x_1 + \dots + \delta_k x_k + e$$

Posteriormente, se obtiene el R^2 de este modelo y se calcula el test estadístico del siguiente contraste:

$$\begin{aligned} H_0: \sigma_i^2 &= \sigma^2 \rightarrow \text{Homocedasticidad.} \\ H_1: \sigma_i^2 &\neq \sigma^2 \rightarrow \text{Heterocedasticidad.} \end{aligned}$$

$$LM = nR^2 \sim \chi_{k-1}^2$$

Donde k indica el número de variables explicativas del último modelo.

5.1.7. Contraste de normalidad: Jarque-Bera

La prueba Jarque-Bera testea si una serie de datos, en este caso, la serie de los errores del Modelo de Mínimos Cuadrados Ordinarios, presentan una asimetría y curtosis tal que se asemeja a la distribución normal.

El coeficiente de asimetría se define mediante la siguiente expresión matemática:

$$S = \frac{\sum_{t=1}^n (x_t - \bar{x})^3 / n}{(\sum_{t=1}^n (x_t - \bar{x})^2 / n)^{3/2}}$$

El coeficiente de Curtosis, se define

$$C = \frac{\sum_{t=1}^n (x_t - \bar{x})^4 / n}{(\sum_{t=1}^n (x_t - \bar{x})^2 / n)^2}$$

A partir de la combinación de los dos coeficientes descritos anteriormente, es decir, el coeficiente de asimetría y el coeficiente de curtosis, se determina el estadístico de Jarque-Bera:

$$JB = (n - k + 1)/6 \left[S^2 + \frac{(C - 3)^2}{4} \right]$$

El contraste que se efectúa es el siguiente:

$$\begin{aligned} H_0: & \text{La muestra se distribuye normalmente.} \\ H_1: & \text{La muestra no se distribuye normalmente.} \end{aligned}$$

5.1.8. Contraste de normalidad: Shapiro-Wilk

Al igual que el test anterior, el Test de Shapiro-Wilk contrasta la normalidad de, en este caso, los residuos del Modelo estimado por Mínimos Cuadrados Ordinarios. Para llevar a cabo este test, se ordena la muestra de menor a mayor obteniendo el siguiente vector muestral:

$$x_{(1)}, x_{(2)}, \dots, x_{(n)}$$

Se calcula la varianza muestral de la anterior muestra S^2 y se calcula el estadístico de contraste siguiente:

$$W = \frac{1}{nS^2} \left(\sum_{j=1}^h a_{jn} (x_{n-i+1} + x_i) \right)^2$$

Donde h es $n/2$ si n es par y toma valor $(n - 1)/2$ si n es impar. En cuanto al sumatorio $\sum_{j=1}^h a_{jn} (x_{n-i+1} + x_i)$, se obtiene a partir del vector de la muestra ordenado, calculando las diferencias entre los valores extremos y corregidos por unos coeficientes determinados por Shapiro-Wilk.

El contraste que se efectúa es el siguiente:

$$\begin{aligned} H_0: & \text{La muestra se distribuye normalmente.} \\ H_1: & \text{La muestra no se distribuye normalmente.} \end{aligned}$$

Además de los contrastes anteriormente descritos, existen numerosos contrastes adicionales que permiten testear la normalidad y heterocedasticidad de los residuos del modelo que se está estudiando.

5.2. Modelos Lineales Generalizados: Probit y Logit.

Las limitaciones que presentan los Modelos de Regresión Lineal pueden subsanarse mediante modelos de respuesta limitada como el Modelo Logit y el Modelo Probit, que serán objeto de estudio en esta tesis. La variable dependiente limitada hace que el modelo deba asumir la restricción siguiente:

$$0 \leq (E(Y|X)) \leq 1$$

Donde X es el vector de variables dependientes, también denominadas, variables explicativas, de control o predictoras.

5.2.1. Modelo Logit

Cabe destacar que existen dos clasificaciones del Modelo Logit según del número de alternativas que se estudian. En este caso, el modelo que se plantea es un Modelo Logit binario y no de respuesta múltiple ya que se estudian dos posibles alternativas.

Este modelo logit utiliza como función de estimación la función logística, en lugar de la función lineal como el Modelo de Regresión Lineal.

$$E(Y|X) = 1 \cdot P(y = 1|x) + 0 \cdot (1 - P(y = 1|x)) = \Lambda(z)$$

La función $\Lambda(z)$ es una función que toma valores comprendidos en el intervalo $[0,1]$, pues la variable dependiente limitada, como se ha indicado en el anterior apartado, hace que el modelo deba asumir $0 \leq (E(Y|X)) \leq 1$. La función $\Lambda(z)$, es una función logística, conocida como función enlace:

$$\Lambda(z) = \frac{e^z}{1 + e^z} = \frac{1}{1 + e^{-z}}$$

Donde, $z = x'\beta$, siendo x es el vector de variables independientes y β es el vector de parámetros. Por lo tanto, se obtiene la siguiente ecuación

$$E(Y|X) = P(y = 1|x) = \Lambda(z) = \frac{1}{1 + e^{-x'\beta}}$$

Por otro lado,

$$1 - \Lambda(z) = \frac{1}{1 + e^z}$$

Relacionando los términos $\Lambda(z)$ y $1 - \Lambda(z)$, se tiene que:

$$\frac{\Lambda(z)}{1 - \Lambda(z)} = \frac{1 + e^z}{1 + e^{-z}} = e^z$$

En el Modelo Logit los errores siguen una distribución logística.

5.2.2. Modelo Probit

Al igual que en el Modelo Logit, se pueden distinguir dos clasificaciones según del número de alternativas que se estudian. En este caso, se estudia el Modelo Probit ordenado.

En el Modelo Probit se utiliza como función de estimación la función de distribución acumulada de la normal estándar, es decir

$$E(Y|X) = 1 \cdot P(y = 1|x) + 0 \cdot (1 - P(y = 1|x)) = \Phi(z)$$

La función $\Phi(z)$ es una función que toma valores comprendidos en el intervalo $[0,1]$, pues la variable dependiente limitada, como se ha indicado en el anterior apartado, hace que el modelo deba asumir $0 \leq (E(Y|X)) \leq 1$. La función $\Phi(z)$, es la función de densidad de la normal estándar:

$$\Phi(z) = \int_{-\infty}^z \phi(v)dv = (2\pi)^{-1/2} e^{(-z^2/2)}$$

Donde, $z = x'\beta$, siendo x es el vector de variables independientes y β es el vector de parámetros. Por lo tanto, se obtiene la siguiente ecuación

$$E(Y|X) = P(y = 1|x) = \Phi(z) = (2\pi)^{-1/2} e^{(-z^2/2)}$$

En el Modelo Probit supone que los errores siguen una distribución normal.

5.3. Significatividad de los coeficientes:

Los contrastes paramétricos tienen como finalidad comprobar si se cumplen o no las suposiciones planteadas en el modelo sobre la población, todo ello a partir de unos datos muestrales.

Para realizar un contraste se requiere establecer tres elementos. El primer elemento es la Hipótesis Nula, esto es, el parámetro que se desea contrastar. Otro elemento es un contraste es el estadístico, es decir, una función derivada de un conjunto de datos muestrales que estima características de una población. Por último, una regla de rechazo, para poder determinar si se rechaza o acepta la Hipótesis Nula.

5.3.1. Significatividad individual de los coeficientes:

En este tipo de contrastes la hipótesis nula se plantea sobre un único coeficiente β . A continuación, se definen los tres elementos necesarios para plantear el contraste.

Hipótesis Nula:

$$H_0: \beta_j = 0$$

Estadístico *T – student*:

$$t = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}$$

Se establecen un nivel de significación α , definiendo las siguientes reglas de aceptación o rechazo:

$$t = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \in \left[-t_{n-1, 1-\frac{\alpha}{2}}, t_{n-1, 1-\frac{\alpha}{2}} \right] \rightarrow \text{Se acepta la Hipótesis Nula } H_0.$$

$$t = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \notin \left[-t_{n-1, 1-\frac{\alpha}{2}}, t_{n-1, 1-\frac{\alpha}{2}} \right] \rightarrow \text{Se rechaza la Hipótesis Nula } H_0.$$

5.3.2. Significatividad de los coeficientes: Estadístico Wald.

Partiendo de un parámetro de interés β , este Test compara la estimación de máxima verosimilitud $\hat{\beta}$ de dicho parámetro con un valor propuesto β_0 ; todo ello bajo la hipótesis de que la diferencia tipificada entre la estimación del parámetro por máxima verosimilitud y el valor propuesto sigue, de manera aproximada, una distribución normal.

El contraste que se prueba mediante el estadístico de Wald es el siguiente:

$$H_0: \beta_j = 0$$

Se define,

$$wald = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}$$

donde $SE(\hat{\beta}_j)$ denota el error estándar de la estimación de máxima verosimilitud.

Este estadístico de Wald $\frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}$ se compara con una distribución normal $N(0,1)$ o, de manera equivalente, $\left(\frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}\right)^2$ se compara con una distribución Chi-cuadrado.

Este modelo permite, además de contrastar la significación del parámetro β_j , contrastar la significación conjunta de los coeficientes del modelo, siendo las hipótesis,

$$H_0: \beta_1 = 0, \beta_2 = 0, \dots, \beta_k = 0$$

$$H_1: \forall \beta_j \neq 0$$

5.3.1. Significatividad conjunta de los coeficientes: Test de Lagrange.

Al igual que el anterior test, el Test de Lagrange permite contrastar la significatividad conjunta de las variables explicativas del modelo que se está estudiando. El test del multiplicador de Lagrange exige estimar, además del modelo no restringido, es decir, el modelo que se está analizando, el modelo restringido, es decir, un modelo en el que se eliminan las variables cuya significatividad se quiere testear. En este aspecto se diferencia del Estadístico de Wald, pues este solo exige estimar el modelo no restringido.

Este test se basa en el cálculo de la diferencia entre la log-verosimilitud del modelo no restringido y la log-verosimilitud del modelo restringido.

$$LR = 2 \cdot (l_{ur} - l_r)$$

Este estadístico se distribuye mediante una Chi-cuadrado de q grados de libertad. Es decir,

$$LR \sim \chi_q^2$$

donde q es el número de variables restringidas.

El contraste que se prueba es el siguiente:

$$\begin{aligned} H_0: \beta_1 = 0, \beta_2 = 0, \dots, \beta_k = 0 \\ H_1: \forall \beta_j \neq 0 \end{aligned}$$

6. Introducción a la Teoría de la Credibilidad:

La teoría de la credibilidad es una rama de ciencia actuarial enfocada al cálculo de la prima de riesgo individual, es decir, el cálculo de una prima para cada grupo de riesgo atendiendo a la información histórica de siniestralidad. Dicha prima de riesgo individual atiende a la experiencia de la aseguradora sobre un grupo de riesgo, así como a información externa a dicho riesgo concreto. Por lo tanto, la Teoría de la Credibilidad permite a las compañías aseguradoras utilizar la información de siniestralidad pasada, es decir, su experiencia, ya sea de la cartera global de los Seguros de Vida o de una Modalidad concreta, para estimar los siniestros que se producirán en el futuro y aplicar esta estimación a la Tarificación de los Seguros.

La Teoría de la Credibilidad asigna un peso denominado Z o peso de credibilidad a la experiencia observada del colectivo y un peso $1 - Z$ a la experiencia externa a este colectivo, para la determinación de la prima de riesgo. Matemáticamente,

$$P_j = (1 - Z) \cdot P_0 + Z \cdot \hat{P}_j$$

Donde, P_0 se conoce como la prima a priori (información a priori), \hat{P}_j como experiencia de siniestralidad del riesgo j observada y P_j como la prima a posteriori que surge de la combinación de las anteriores, ponderadas según los pesos de credibilidad obtenidos. Por lo tanto, de la ecuación anterior se observa la relación entre la Teoría de la Credibilidad y el Teorema de Bayes³, al combinar información a priori e información muestral.

En ocasiones las aseguradoras no disponen de información suficiente respecto a un colectivo concreto y utilizan información adicional externa a dicho colectivo para una mejor estimación de las tarifas a aplicar, esto sucede cuando el peso credibilidad, es decir Z , es inferior a la unidad. Existen varias teorías para hallar el peso de credibilidad, como el Modelo de Bühlmann, Modelo de Bühlmann-Straub, Modelo de Jewell, Modelo de Hachemeister... En esta tesis se estudia la Teoría de la Credibilidad de Bühlmann y Bühlmann-Straub, que son modelos de credibilidad de distribución libre.

6.1. Modelo de Bühlmann:

El Modelo de Bühlmann es una teoría de credibilidad de distribución libre que surge en el año 1967. Este Modelo de Credibilidad se basa en el método de mínimos cuadrados ordinarios para la determinación de la prima de riesgo, en concreto, el objetivo de este método es la obtención una prima lineal que se ajuste a la experiencia que tiene la compañía sobre cada riesgo. Por lo tanto, este modelo no precisa realizar hipótesis respecto a las funciones de distribución de los parámetros de riesgo ni de los riesgos individuales.

Este modelo parte de dos hipótesis, la primera supone que las pólizas son independientes e idénticamente distribuidas y, por lo tanto, supone la existencia de independencia entre riesgos. La segunda supone homogeneidad temporal e independencia dentro de cada riesgo, al considerar que la experiencia de siniestralidad de cada póliza y periodo temporal, condicionada al parámetro de riesgo asignado a cada póliza, son variables independientes e idénticamente distribuidas. Como conclusión a estas hipótesis, el Modelo de Bühlmann supone homogeneidad en el tiempo e independencia dentro de cada póliza y entre pólizas.

A continuación, se presentan y definen las variables de partida que se utilizan para la elaboración de este modelo,

3 TEOREMA DE BAYES: Procedimiento estadístico mediante el cual se determina la probabilidad de ocurrencia de un determinado suceso a partir de la información de ocurrencia de acontecimientos anteriores.

$s \rightarrow$ Variable que indica el periodo temporal y, por lo tanto, presenta valores comprendidos entre 0 y t , siendo t el número total de períodos observados.

$j \rightarrow$ Variable que indica el grupo de riesgo y, por lo tanto, presenta valores comprendidos entre 0 y k , siendo k el número total de grupos de riesgo.

$X_{js} \rightarrow$ Variable que indica la experiencia respecto a la frecuencia siniestral del riesgo j –ésimo en el período s –ésimo.

$\theta_j \rightarrow$ Parámetro de riesgo que describe las características que presenta el riesgo j . Este parámetro, generalmente, es desconocido, no puede observarse.

A partir de estas variables, se calcula la media esperada de los siniestros, la varianza entre las primas de riesgo individuales y la dispersión de los siniestros de la cartera, estas variables se denominan parámetros estructurales, y su estimación permite determinar el factor de credibilidad Z y así establecer la prima de riesgo individual estimada, $\hat{\mu}(\theta_j)$.

La media esperada de los siniestros, puede definirse como la prima de riesgo colectiva y, por lo tanto, equivale a $E[X_{js}] = E[\mu(\theta_j)]$.

Se denota a a la varianza existente entre las primas de riesgo individuales y, por lo tanto, puede definirse como $Var[E[X_{js}|\theta_j]] = Var[\mu(\theta_j)]$. Este parámetro mide el grado de heterogeneidad de la cartera.

Por último, la dispersión o varianza total de los siniestros de la cartera, se define como $E[Var[X_{js}|\theta_j]] = E[\sigma^2(\theta_j)]$.

A continuación, se describen, mediante expresiones matemáticas, los estimadores insesgados de estos parámetros estructurales,

Estimador insesgado de la media poblacional,

$$\hat{m} = \frac{1}{k} \cdot \sum_{j=1}^k \sum_{s=1}^t X_{js}/t$$

Estimador insesgado de la varianza entre las primas de riesgo individuales,

$$\hat{a} = \frac{1}{k-1} \cdot \sum_{j=1}^k (M_j - m)^2 - \frac{1}{t} \cdot \hat{S}^2$$

Estimador insesgado de la dispersión total esperada de los siniestros de la cartera:

$$\hat{S}^2 = \frac{1}{k} \cdot \sum_{j=1}^k \frac{t}{t-1} \sum_{s=1}^t (X_{js} - M_j)^2$$

El objetivo es obtener el factor de credibilidad Z y la mejor estimación de credibilidad de la prima de riesgo individual, $\hat{\mu}(\theta_j)$, que se asigna a cada asegurado del grupo de

riesgo j . En el Modelo de Bühlmann, esta prima se estima mediante una aproximación lineal, buscando los valores $c_{j0}, c_{j1}, \dots, c_{jt}$ que minimizan la siguiente expresión,

$$E \left[\left(\hat{\mu}(\theta_j) - \left(c_{j0} + \sum_{s=1}^t c_{js} \cdot X_{js} \right) \right)^2 \right]$$

Teniendo en cuenta las hipótesis que contempla el Modelo de Bühlmann, se resuelve este problema de minimización y se obtiene la siguiente ecuación, que corresponde al estimador de credibilidad de la prima de riesgo individual, esto es, $\hat{\mu}(\theta_j)$

$$\hat{\mu}(\theta_j) = c_{j0} + \sum_{s=1}^t c_{js} \cdot X_{js} = c_{j0} + c \cdot \sum_{s=1}^t X_{js}$$

En la resolución a este problema de minimización se determinan las expresiones matemáticas de c_{j0} y c .

$$c = \frac{a}{S^2 + a \cdot t} = \frac{Z}{t}$$

$$c_{j0} = m \cdot (1 - t \cdot c) = m \cdot [1 - Z]$$

El peso de credibilidad Z , se define mediante la siguiente expresión matemática,

$$Z = \frac{a \cdot t}{S^2 + a \cdot t}$$

Este factor o peso de credibilidad toma valores comprendidos entre 0 y 1 y, como puede observarse a partir de esta expresión, se acerca a la unidad cuando se tienen altos valores de t , altos valores de a y bajos valores de S^2 .

Sustituyendo en la ecuación anterior de $\hat{\mu}(\theta_j)$ las expresiones matemáticas de c_{j0} y c , se obtiene la ecuación final del estimador de credibilidad de la prima de riesgo individual,

$$\hat{\mu}(\theta_j) = m \cdot [1 - Z] + \frac{Z}{t} \cdot \sum_{s=1}^t X_{js}$$

Que puede expresarse de la siguiente manera,

$$\hat{\mu}(\theta_j) = m \cdot [1 - Z] + Z \cdot \bar{X}_j$$

En definitiva, en el Modelo de Bühlmann el estimador de credibilidad de la prima de riesgo individual puede definirse como la suma de la experiencia siniestral del colectivo con peso de credibilidad Z y la media poblacional de la cartera global con peso $1 - Z$.

A partir de la expresión final de $\hat{\mu}(\theta_j)$, es decir, del estimador de credibilidad de la prima de riesgo individual, y la expresión final del factor o peso de credibilidad Z , puede observarse que la prima de riesgo individual estimada se aproxima a la experiencia

siniestral de cada grupo de riesgo, a medida que la aseguradora tiene mayor información sobre el riesgo pues, en este caso, el peso de credibilidad se aproxima a la unidad y, por lo tanto, \bar{X}_j se considera el mejor estimador lineal de $\mu(\theta_j)$. También puede observarse que esto ocurre cuando hay heterogeneidad dentro de la cartera o cuando hay poca dispersión en la siniestralidad individual.

En este proceso de determinación de la prima no se han realizado hipótesis sobre el tipo de distribución, es por ello que el Modelo de Bühlmann se conoce como modelo de credibilidad de distribución libre.

Cabe destacar la importancia de determinar adecuadamente el valor de Z , pues, no se conocen los valores reales a y S^2 , si no que son valores estimados y, de no coincidir con los reales, no se resuelve adecuadamente el problema de minimización anteriormente planteado. Un valor incorrectamente estimado de Z implica que la prima de riesgo individual se aleje de la prima que correspondería dadas las características de dicho grupo de riesgo.

Las críticas que presenta el Modelo de Bühlmann derivan de la “rigidez” de sus hipótesis fundamentales. Este modelo no tiene en cuenta si hay pólizas que suponen un mayor impacto para la aseguradora dado que trabaja con observaciones no ponderadas. Por otro lado, supone homogeneidad temporal, por lo que el modelo queda invalidado cuando se tratan datos que están afectados por la inflación o una tendencia. En el caso de Seguros de Vida que se estudia en esta tesis, se trabaja con datos deflactados que no presentan tendencia.

6.2. Modelo de Bühlmann-Straub:

El Modelo de Bühlmann-Straub es una ampliación del Modelo de Bühlmann y, por lo tanto, se incorpora a los modelos de teoría de credibilidad de distribución libre al no ser necesaria la realización de hipótesis respecto a las funciones de distribución. La novedad que incorpora este modelo es la ponderación de las observaciones, aspecto que permite ampliar su campo de aplicación al no ser necesaria la consideración de que todas las pólizas o riesgos provocan el mismo impacto en la compañía.

Cabe destacar, con respecto al modelo anterior, que la varianza deja de ser homogénea, pues depende de los pesos asignados a cada riesgo y, en consecuencia, se asigna un factor o peso de credibilidad a cada riesgo, es decir, Z ya no es constante para toda la cartera.

Este modelo parte de las mismas hipótesis que el Modelo de Bühlmann, supone que las pólizas son independientes e idénticamente distribuidas y, por lo tanto, hay independencia entre de riesgos. Además, supone también que hay homogeneidad temporal e independencia dentro de cada riesgo.

Las variables de partida que se utilizan para la elaboración de este modelo, son las que a continuación se indican,

$s \rightarrow$ Variable que indica el periodo temporal y, por lo tanto, presenta valores comprendidos entre 0 y t , siendo t el número total de períodos observados.

$j \rightarrow$ Variable que indica el grupo de riesgo y, por lo tanto, presenta valores comprendidos entre 0 y k , siendo k el número total de grupos homogéneos de riesgo.

$X_{js} \rightarrow$ Variable que indica la experiencia respecto a la frecuencia siniestral del riesgo j –ésimo en el período s –ésimo.

$\theta_j \rightarrow$ Parámetro de riesgo que describe las características que presenta el riesgo j . Este parámetro, generalmente, es desconocido, no puede observarse.

$w_{js} \rightarrow$ Parámetro que indica las ponderaciones asignadas a cada grupo de riesgo y periodo temporal.

Para llevar a cabo este modelo, se calculan los parámetros estructurales descritos en el Modelo de Bühlmann, es decir, $E[\mu(\theta_j)]$, $Var[\mu(\theta_j)]$ y $E[\sigma^2(\theta_j)]$. En primer lugar, a partir la estimación de $Var[\mu(\theta_j)]$ y $E[\sigma^2(\theta_j)]$, se determinan los pesos de credibilidad, z_j . Una vez determinados los pesos, se puede calcular el estimador de la media poblacional. Por último, se calcula la prima de riesgo individual estimada, $\hat{\mu}(\theta_j)$.

Atendiendo a la ponderación de las observaciones, se calcula: la media aritmética ponderada,

$$X_{jw} = \sum_{s=1}^t \frac{w_{js} \cdot X_{js}}{w_j}$$

Y la media aritmética de las medias individuales,

$$X_{ww} = \sum_{j=1}^k \frac{X_{jw} \cdot w_j}{w}$$

Donde,

$$w_j = \sum_{s=1}^t w_{js}$$

$$w = \sum_{j=1}^k w_j$$

A continuación, se describen, los estimadores insesgados de estos parámetros estructurales.

Estimador insesgado de la varianza entre las primas de riesgo individuales,

$$\hat{a} = \frac{1}{c} \cdot \left(\sum_{j=1}^k \frac{w_j}{w} \cdot (X_{jw} - X_{ww})^2 - (k-1) \cdot \frac{\hat{S}^2}{w} \right)$$

Estimador insesgado de la dispersión esperada de los siniestros de la cartera:

$$\hat{S}^2 = \frac{1}{k} \cdot \sum_{j=1}^k \frac{t}{t-1} \sum_{s=1}^t w_{js} \cdot (X_{js} - X_{jw})^2$$

Los factores o pesos de credibilidad, se calculan mediante la siguiente fórmula,

$$z_j = \frac{a \cdot w_j}{S^2 + a \cdot w_j}$$

$$Z = \sum_{j=1}^k z_j$$

Estimador insesgado de la media poblacional,

$$\hat{m} = \sum_{j=1}^k \frac{z_j}{Z} \cdot M_j$$

Por último, se calcula la prima de riesgo individual,

$$\hat{\mu}(\theta_j) = m \cdot [1 - z_j] + z_j \cdot X_{jw}$$

La interpretación de la prima de riesgo individual calculada mediante el Modelo de Bühlmann-Straub es similar a la interpretación que se realizó del Modelo Bühlmann. El estimador de credibilidad de la prima de riesgo individual equivale a la suma de la experiencia siniestral del colectivo ponderada con peso de credibilidad z_j y la media poblacional de la cartera global con peso $1 - z_j$. La única diferencia de interpretación con respecto al Modelo Bühlmann es que los pesos de credibilidad varían y, por lo tanto, los pesos que se le asignan a la experiencia siniestral ponderada del colectivo concreto y a la media poblacional varían en función del grupo de riesgo.

A partir de la expresión final de $\hat{\mu}(\theta_j)$ y de los factores o pesos de credibilidad Z , se observa que la prima de riesgo individual estimada se aproxima a la experiencia siniestral ponderada de cada grupo de riesgo conforme aumenta el peso de ponderación asignado a dicho grupo, es decir, w_j , pues la aseguradora tiene mayor información sobre el riesgo y el peso de credibilidad se aproxima a la unidad. También puede observarse que esto ocurre cuando hay heterogeneidad dentro de la cartera y cuando hay poca dispersión en la siniestralidad individual.

El Modelo de Bühlmann-Straub soluciona una de las críticas que recibe el Modelo de Bühlmann al introducir las observaciones ponderadas. No obstante, mantiene el supuesto de homogeneidad temporal, por lo únicamente es válido cuando se trabaja con datos deflactados que no presentan tendencia.

7. Aplicación práctica de Modelos Lineales Generalizados y Modelos de Credibilidad a la Tarificación de Seguros de Vida:

En esta parte de la tesis se realiza una aplicación práctica de Modelos de Regresión y Modelos de Credibilidad para tarificar Seguros de Vida Riesgo Colectivo, en concreto, Seguros Temporales Anuales Renovables (TAR). Para esta aplicación práctica se han obtenido datos de la cartera de una Aseguradora de Vida, eliminando datos confidenciales y personales para dar cumplimiento a la Ley Protección de Datos de Carácter Personal.

Se comienza el estudio con los datos de la cartera de seguros TAR, con garantía de fallecimiento, y un ámbito temporal que comprende los últimos cinco años completos. Esta muestra contiene 1.252.378 observaciones. Los datos que se utilizan para el análisis son, la edad actuarial del asegurado, el sexo del asegurado, la Clasificación Nacional de Actividades Económicas establecida en 2009 (CNAE) de la empresa tomadora del Seguro, y la clasificación por número de empleados de la empresa tomadora del Seguro.

La primera variable indica la edad actuarial del individuo, es decir, su edad en la fecha más próxima al cumpleaños, pues esta es la edad que se tiene en cuenta para suscribir un seguro. La segunda variable es una variable dicotómica que toma valor 1 cuando el asegurado es varón y valor 0 en caso contrario. La tercera variable es también una variable dicotómica que indica si la empresa tomadora del seguro es una gran empresa, según la clasificación oficial por número de trabajadores, en cuyo caso la variable toma valor 1, o si es una pyme, en cuyo caso toma valor 0. La cuarta variable indica el nivel de riesgo asociado a la actividad profesional, en este caso, de la empresa. A continuación, se explica con mayor detalle cómo se ha llevado a cabo la elaboración de esta cuarta variable.

La clasificación de las actividades económicas es una clasificación oficial que tiene como objetivo favorecer la elaboración de estadísticas nacionales, por lo que todas las empresas tienen asignado un código de CNAE. La CNAE-2009 se estructura en cuatro niveles: Sección, División, Grupo y Clase. La Sección está compuesta por 21 códigos alfabéticos de 1 dígito, la División por 88 códigos numéricos de 2 dígitos, el Grupo por 272 códigos numéricos de 3 dígitos y la Clase por 629 códigos numéricos de 4 dígitos. Para generar la variable de nivel de riesgo asociado a la actividad profesional se ha tenido en cuenta el primer nivel de esta estructura, es decir, el código alfabético de 1 dígito, y la experiencia de la aseguradora para clasificar estos códigos en 4 niveles de riesgo según cuartiles, atendiendo a la exposición al riesgo.

Una vez generadas estas variables, se procede a dividir la muestra completa en dos submuestras que contienen, cada una, el 50%. Esta división se realiza de manera

aleatoria con el objetivo de regresar el modelo con una de estas submuestras, a la que se denomina “muestra modelo”, y aplicarlo a la otra submuestra, a la que se denominará “muestra testeo”, para comprobar el porcentaje de acierto de este modelo. Para llevar a cabo esta división aleatoria se crea un script en Matlab⁴, ejecutando comandos que garantizan la aleatoriedad de la muestra. Este script puede verse en el *Anexo 1* de esta tesis.

7.1. Aplicación de Modelos de Regresión a la Tarificación de Seguros de Vida:

En esta parte de la tesis se realiza una aplicación práctica de Modelos de Regresión a la Tarificación de Seguros de Vida Riesgo Colectivos, proponiendo tres modelos: Modelo de Mínimos Cuadrados Ordinarios, Modelo Probit y Modelo Logit. El Modelo de Mínimos Cuadrados Ordinarios se rechaza al no cumplir con supuestos clásicos, tales como la homocedasticidad, normalidad de los residuos, linealidad del modelo... Sin embargo, se aceptan los otros dos modelos planteados (Logit y Probit), concluyendo que, entre las variables planteadas, sexo, edad, gran empresa y nivel de riesgo asociado a la actividad profesional, las variables que son significativas para estimar la probabilidad de fallecimiento de los asegurados son el sexo y el nivel de riesgo asociado a la actividad profesional. No siendo significativa la variable edad ni la variable gran empresa, para determinar la probabilidad de fallecimiento.

Finalmente, conforme a los resultados obtenidos en el análisis de estos dos modelos, se opta por la tarificación de los Seguros de Vida Colectivos Temporales Anuales Renovables, mediante las variables sexo y nivel de riesgo asociado a la actividad profesional. Los resultados que se obtienen de estos modelos son muy similares, a pesar de que sus funciones enlace obedecen a distribuciones distintas.

Para la elaboración de los Modelos de Regresión se elabora un script en Gretl⁵. En este script se ejecuta el Modelo de Mínimos Cuadrados Ordinarios, Modelo Probit y Modelo Logit, se realizan contrastes de significación individual y conjunta, test de homogeneidad, heterocedasticidad, normalidad de los residuos... Además, se analizan los estadísticos descriptivos de las variables contenidas en esta la base de datos. Este script puede verse en el *Anexo 2* de esta tesis.

4 MATLAB es una herramienta de software matemático que permite, entre sus aplicaciones básicas, tratamiento de matrices, algoritmos, creación de interfaces, representación de variables... Además, dispone de herramientas adicionales que amplían sus funciones y ámbito de aplicación.

5 GRETL es un software econométrico de libre distribución que se utiliza para realizar análisis estadísticos y modelos de regresión.

7.1.1. Modelo de Regresión 1: Mínimos Cuadrados Ordinarios.

A partir de la “muestra modelo”, se lleva a cabo un modelo de regresión de Mínimos Cuadrados Ordinarios para estimar la probabilidad de fallecimiento a partir de las variables: edad, sexo, gran empresa y nivel de riesgo asociado a la actividad profesional. Los resultados obtenidos, en cuanto a la significatividad de estas variables para determinar la probabilidad de fallecimiento, han sido los siguientes:

EDAD	SEXO	GRAN EMPRESA	NIVEL DE RIESGO
***	***		***

** Se rechaza la Hipótesis Nula a un nivel de confianza del 90%*
*** Se rechaza la Hipótesis Nula a un nivel de confianza del 95%*
**** Se rechaza la Hipótesis Nula a un nivel de confianza del 99%*

Fuente: Elaboración propia. En esta tabla se muestran los resultados del contraste de significatividad de los coeficientes correspondientes a las variables edad, sexo, gran empresa y nivel de riesgo, en el Modelo de Mínimos Cuadrados Ordinarios propuesto.

La edad del asegurado, su sexo y el nivel de riesgo asociado a la actividad profesional afectan de forma significativa a la probabilidad de fallecimiento. Este resultado se obtiene realizando contrastes conjuntos de significación para cada grupo de variables relacionadas con la edad, sexo y el nivel de riesgo, respectivamente. En concreto se puede afirmar que estas variables son significativas a un nivel de significación del 1%. En cuanto a la variable dicotómica gran empresa, se concluye que no es una variable significativa para determinar la probabilidad de fallecimiento de un individuo.

Una vez regresado el modelo, se prueba la validez del mismo. En primer lugar, se realizan test de heterocedasticidad del error, para comprobar si la varianza del error es o no constante. Los resultados que se obtienen tras realizar estos contrastes son los siguientes,

```

? modtest --white --quiet #test de heteroscedasticidad de White

Contraste de heterocedasticidad de White

Estadístico de contraste: TR^2 = 1457.818299,
con valor p = P(Chi-cuadrado(330) > 1457.818299) = 0.000000

? modtest --breusch-pagan --quiet #test de heteroscedasticidad de Breusch - \
Pagan

Contraste de heterocedasticidad de Breusch-Pagan

Estadístico de contraste: LM = 537351.042042,
con valor p = P(Chi-cuadrado(58) > 537351.042042) = 0.000000

```

Se rechaza la Hipótesis Nula, la varianza del error no es constante y, por lo tanto, hay presencia de heterocedasticidad. Por este motivo, en el modelo planteado se han incorporado errores robustos a la heterocedasticidad, con el objetivo de poder realizar contrastes de hipótesis válidas.

En segundo lugar, se llevan a cabo contrastes de Normalidad de los residuos de este modelo, en concreto, los Test de Jarque Bera y Shapiro–Wilk, ya explicados anteriormente en esta tesis. A continuación, se muestran los resultados obtenidos tras realizar estos contrastes de Normalidad:

```

? normtest uhat --jbera #normalidad de los residuos
Contraste de normalidad de uhat:
Contraste de Jarque-Bera = 2.61932e+010, con valor p 0

? normtest uhat --swilk #normalidad de los residuos
Contraste de normalidad de uhat:
W de Shapiro-Wilk = 0.0201564, con valor p 1.39405e-229

```

Como se puede observar, el p-valor de estos contrastes es muy pequeño, prácticamente cero, por lo que se rechaza la Hipótesis Nula de Normalidad de los residuos. A pesar de esta restricción, el modelo no queda invalidado, pues los estimadores continuarían siendo estimadores lineales, insesgados y óptimos, pero pierden eficiencia entre los estimadores insesgados.

En tercer lugar, se contrasta si existe error de especificación en este modelo. Para ello se lleva a cabo el Test Reset de Ramsey, cuyos resultados se muestran a continuación,

```

? reset --quiet #test linealidad

Contraste de especificación RESET (cuadrados y cubos)
Estadístico de contraste: F = 24.069966,
con valor p = P(F(2,626133) > 24.07) = 3.52e-011

```

Se comprueba que existe un error de especificación en el modelo,

Además de las limitaciones ya comentadas, este modelo no garantiza que las probabilidades estén comprendidas entre 0 y 1. Finalmente, atendiendo a todos los contrastes realizados, se opta por rechazar la validez del Modelo de Mínimos Cuadrados Ordinarios para estimar las probabilidades de fallecimiento y, por lo tanto, se rechaza para la Tarificación de Seguros de Vida Colectivos, que es el objetivo de esta tesis.

7.1.2. Modelo de Regresión 2: Probit.

Siguiendo el mismo procedimiento para este modelo, se parte de la “muestra modelo” para estimar la probabilidad de fallecimiento a partir de las variables: edad, sexo, gran empresa y nivel de riesgo asociado a la actividad profesional, en este caso, el modelo a regresar es un Modelo Probit. Los resultados obtenidos, en cuanto a la significatividad de estas variables para determinar la probabilidad de fallecimiento, han sido los siguientes:

EDAD	SEXO	GRAN EMPRESA	NIVEL DE RIESGO
	***	**	***

* Se rechaza la Hipótesis Nula a un nivel de confianza del 90%
 ** Se rechaza la Hipótesis Nula a un nivel de confianza del 95%
 *** Se rechaza la Hipótesis Nula a un nivel de confianza del 99%

Fuente: Elaboración propia. En esta tabla se muestran los resultados del contraste de significatividad de los coeficientes correspondientes a las variables edad, sexo, gran empresa y nivel de riesgo, en el Modelo Probit propuesto.

A diferencia del Modelo de Mínimos Cuadrados Ordinarios comentado en el anterior punto, las variables relacionadas con la edad del asegurado no son significativas de manera conjunta en este modelo de regresión, no siendo tampoco significativas de manera individual. En cuanto a la variable gran empresa, no es una variable significativa para explicar la probabilidad de fallecimiento a un nivel de significación del 1%. Sin embargo, considerando un nivel de significación del 10% o del 5%, si es significativa y, por lo tanto, se rechaza la Hipótesis Nula. En cuanto a las variables relacionadas con el nivel de riesgo y la variable sexo, se obtiene como resultado que estas son significativas al 10%, al 5% y al 1% de nivel de significación. Una vez analizados los coeficientes y su nivel de significación, se contrasta la Normalidad de los Residuos de este Modelo Probit. Como ya se ha comentado en la parte teórica de este modelo, supone que los residuos se distribuyen normalmente, y el test lo corrobora.

Tras los resultados obtenidos, se decide eliminar del modelo de Tarificación la variable edad, pues, no es significativa. Por lo tanto, se decide regresar nuevamente un Modelo Probit, ahora con las siguientes variables explicativas: sexo, gran empresa y nivel de riesgo asociado a la actividad profesional, obteniendo los siguientes resultados de significatividad:

SEXO	GRAN EMPRESA	NIVEL DE RIESGO
***		***
* Se rechaza la Hipótesis Nula a un nivel de confianza del 90%		
** Se rechaza la Hipótesis Nula a un nivel de confianza del 95%		
*** Se rechaza la Hipótesis Nula a un nivel de confianza del 99%		

Fuente: Elaboración propia. En esta tabla se muestran los resultados del contraste de significatividad de los coeficientes correspondientes a las variables sexo, gran empresa y nivel de riesgo, en el Modelo Probit propuesto.

Tanto las variables relacionadas con el nivel de riesgo, como la variable sexo, son significativas. Sin embargo, ahora la variable gran empresa deja de ser significativa, al 10%, al 5% y al 1%. En cuanto al Test de Normalidad de los Residuos, se confirma que los errores de este modelo se distribuyen normalmente.

Dado que la variable gran empresa no es significativa, se decide eliminar esta variable del modelo de regresión y, por lo tanto, se regresa nuevamente un Modelo Probit, ahora, con las siguientes variables: sexo y gran empresa. Los resultados obtenidos respecto a la significatividad de los coeficientes son los siguientes,

SEXO	NIVEL DE RIESGO
***	***
* Se rechaza la Hipótesis Nula a un nivel de confianza del 90%	
** Se rechaza la Hipótesis Nula a un nivel de confianza del 95%	
*** Se rechaza la Hipótesis Nula a un nivel de confianza del 99%	

Fuente: Elaboración propia. En esta tabla se muestran los resultados del contraste de significatividad de los coeficientes correspondientes a las variables sexo y nivel de riesgo, en el Modelo Probit propuesto.

Se concluye que estas variables, sexo y niveles de riesgo, son significativas. Además, de estos dos contrastes, se contrasta si estas variables son significativas conjuntamente, es decir, si las variables explicativas del modelo son significativas de manera conjunta,

$$H_0: \beta_1 = 0, \beta_2 = 0, \dots, \beta_k = 0$$

$$H_1: \forall \beta_j \neq 0$$

Los resultados del Test de Wald,

```
Hipótesis nula: los parámetros de regresión son cero para las variables
SEXO, DNIVELDERIESGO_2, DNIVELDERIESGO_3, DNIVELDERIESGO_4
Estadístico de contraste: F robusto(4, 626189) = 43.4605, valor p 1.58458e-036
```

Los resultados del Test de Lagrange,

```
LR=101.5555
pvalue=8.8614243e-023
```

Se concluye que las variables explicativas de este modelo son significativas de manera conjunta.

Por otro lado, se realizan contrastes de significación individuales para contrastar las variables del grupo nivel de riesgo, es decir, “nivel de riesgo 2”, “nivel de riesgo 3” y “nivel de riesgo 4”. Los resultados obtenidos para estos contrastes de significación individual son los siguientes:

NIVEL DE RIESGO 2	NIVEL DE RIESGO 3	NIVEL DE RIESGO 4
***	***	***
<i>* Se rechaza la Hipótesis Nula a un nivel de confianza del 90%</i>		
<i>** Se rechaza la Hipótesis Nula a un nivel de confianza del 95%</i>		
<i>*** Se rechaza la Hipótesis Nula a un nivel de confianza del 99%</i>		

Fuente: Elaboración propia. En esta tabla se muestran los resultados del contraste de significatividad de los coeficientes nivel de riesgo 2, nivel de riesgo 3 y nivel de riesgo 4, en el Modelo Probit propuesto.

Las variables relacionadas con el nivel de riesgo asociado a la actividad de la empresa, es decir, las variables dicotómicas “nivel de riesgo 2”, “nivel de riesgo 3” y nivel de riesgo 4” son significativas al 10%, 5% y 1%.

Se realiza el Test de Normalidad de los Residuos de este modelo, obteniendo el siguiente resultado

```
? modtest --normality
Contraste de normalidad de los residuos -
Hipótesis nula: el error se distribuye normalmente
Estadístico de contraste: Chi-cuadrado(2) = 3.73285
con valor p = 0.154676
```

Se concluye que los residuos de este modelo planteado se distribuyen normalmente y se opta por escoger este modelo como el Modelo Probit a aplicar a la Tarificación de Seguros Colectivos Temporales Anuales Renovables, posteriormente, se probará su nivel de acierto con la “muestra testeo”.

7.1.3. Modelo de Regresión: Modelo Logit.

Al igual que en los modelos anteriores, se parte de la “muestra modelo” para estimar la probabilidad de fallecimiento a partir de las variables: edad, sexo, gran empresa y nivel de riesgo asociado a la actividad profesional, en este caso, el modelo a regresar es un Modelo Logit. A continuación, se muestran los resultados obtenidos en cuanto a la significatividad de estas variables para determinar la probabilidad de fallecimiento:

EDAD	SEXO	GRAN EMPRESA	NIVEL DE RIESGO
	***	**	***

* Se rechaza la Hipótesis Nula a un nivel de confianza del 90%

** Se rechaza la Hipótesis Nula a un nivel de confianza del 95%

*** Se rechaza la Hipótesis Nula a un nivel de confianza del 99%

Fuente: Elaboración propia. En esta tabla se muestran los resultados del contraste de significatividad de los coeficientes correspondientes a las variables edad, sexo, gran empresa y nivel de riesgo, en el Modelo Logit propuesto.

Dado que la variable sexo no es significativa, se opta por regresar de nuevo el modelo eliminando esta variable. Por lo tanto, se regresa un modelo con las variables relacionadas con el nivel de riesgo asociado a la actividad profesional y el sexo del asegurado, ya que estas son significativas al 10%, 5% y 1%, y con la variable gran empresa que, aunque no es significativa al 1%, si lo es al 5% y al 10%.

Los resultados obtenidos, en cuanto a significatividad de las variables en el nuevo modelo regresado que omite la variable edad, son los siguientes:

SEXO	GRAN EMPRESA	NIVEL DE RIESGO
***		***

* Se rechaza la Hipótesis Nula a un nivel de confianza del 90%

** Se rechaza la Hipótesis Nula a un nivel de confianza del 95%

*** Se rechaza la Hipótesis Nula a un nivel de confianza del 99%

Fuente: Elaboración propia. En esta tabla se muestran los resultados del contraste de significatividad de los coeficientes correspondientes a las variables sexo, gran empresa y nivel de riesgo, en el Modelo Logit propuesto.

Las variables relacionadas con el nivel de riesgo asociado a la actividad profesional, así como la variable sexo, son significativas al 10%, 5% y 1%. Sin embargo, en este modelo

se concluye que la variable gran empresa no es significativa por lo que se opta por regresar de nuevo otro modelo omitiendo esta variable.

Los resultados obtenidos, en cuanto a significatividad, en este Modelo Logit que tiene como variables explicativas los niveles de riesgo asociados a la actividad profesional y el sexo del asegurado, son los siguientes:

SEXO	NIVEL DE RIESGO
***	***
<i>* Se rechaza la Hipótesis Nula a un nivel de confianza del 90%</i>	
<i>** Se rechaza la Hipótesis Nula a un nivel de confianza del 95%</i>	
<i>*** Se rechaza la Hipótesis Nula a un nivel de confianza del 99%</i>	

Fuente: Elaboración propia. En esta tabla se muestran los resultados del contraste de significatividad de los coeficientes correspondientes a las variables sexo y nivel de riesgo, en el Modelo Logit propuesto.

Las variables relacionadas con el nivel de riesgo asociado a la actividad profesional, así como la variable sexo del asegurado, son significativas al 10%, 5% y 1%. Se realiza, además, contrastes de significación conjunta de todas las variables explicativas, esto es,

$$H_0: \beta_1 = 0, \beta_2 = 0, \dots, \beta_k = 0$$

$$H_1: \forall \beta_j \neq 0$$

Los resultados del Test de Wald,

```
Hipótesis nula: los parámetros de regresión son cero para las variables
SEXO, DNIVELDERIESGO_2, DNIVELDERIESGO_3, DNIVELDERIESGO_4
Estadístico de contraste: F robusto(4, 626189) = 43.2497, valor p 2.40347e-036
```

Los resultados del Test de Lagrange,

```
LR=190.53489
pvalue=4.678041e-041
```

Se concluye que todas las variables explicativas del modelo son significativas de manera conjunta.

Por otro lado, se prueba la significatividad de cada variable explicativa a nivel individual, obteniendo los siguientes resultados,

NIVEL DE RIESGO 2 NIVEL DE RIESGO 3 NIVEL DE RIESGO 4

* Se rechaza la Hipótesis Nula a un nivel de confianza del 90%

** Se rechaza la Hipótesis Nula a un nivel de confianza del 95%

*** Se rechaza la Hipótesis Nula a un nivel de confianza del 99%

Fuente: Elaboración propia. En esta tabla se muestran los resultados del contraste de significatividad de los coeficientes nivel de riesgo 2, nivel de riesgo 3 y nivel de riesgo 4, en el Modelo Logit propuesto.

Las variables “nivel de riesgo 2”, “nivel de riesgo 3” y nivel de riesgo 4”, son significativas al 10%, 5% y 1%.

Se opta por escoger este último modelo como Modelo Logit a aplicar a la Tarificación de Seguros Colectivos Temporales Anuales Renovables, posteriormente, se probará su nivel de acierto con la submuestra que no se ha utilizado para elaborar este modelo.

Como puede observarse, los resultados obtenidos en el Modelo Probit y Logit son muy similares.

7.1.4. Prueba de aciertos:

En esta parte de la tesis se pretende probar el porcentaje de acierto del Modelo Probit y del Modelo Logit, respectivamente, para ello se utiliza la “muestra testeó”. Hasta ahora se ha empleado una submuestra aleatoria de la base de datos completa para la elaboración de los Modelos de Regresión, denominada “muestra modelo”, ahora se utiliza la otra submuestra restante para probar el modelo propuesto, denominada “muestra testeó”.

Para realizar esta prueba se elabora un script en Matlab que calcula, utilizando los coeficientes de los modelos de regresión, las probabilidades de fallecimiento de cada asegurado y estima el número de siniestros que se esperan en dicho colectivo. Además de probar los Modelos sobre el 100% de la “muestra testeó”, se prueba sobre el 90%, 80%, ...,10% de dicha muestra, escogiendo estas muestras de manera aleatoria y repitiendo este proceso 1.000 veces. El script de Matlab que se ha desarrollado para ejecutar esta prueba puede verse en el *Anexo 3* de esta tesis.

El objetivo es observar el porcentaje de error del Modelo. Este porcentaje que se calcula de la siguiente manera:

$$\% \text{ error} = \left| 1 - \frac{\text{número de siniestros estimados en el colectivo}}{\text{número siniestros reales en el colectivo}} \right|$$

A continuación, se muestran los porcentajes de error obtenidos al aplicar los Modelos Probit y Logit a las 1.000 muestras obtenidas aleatoriamente para el 90%, 80%, ...,10% de la “muestra testeo”.

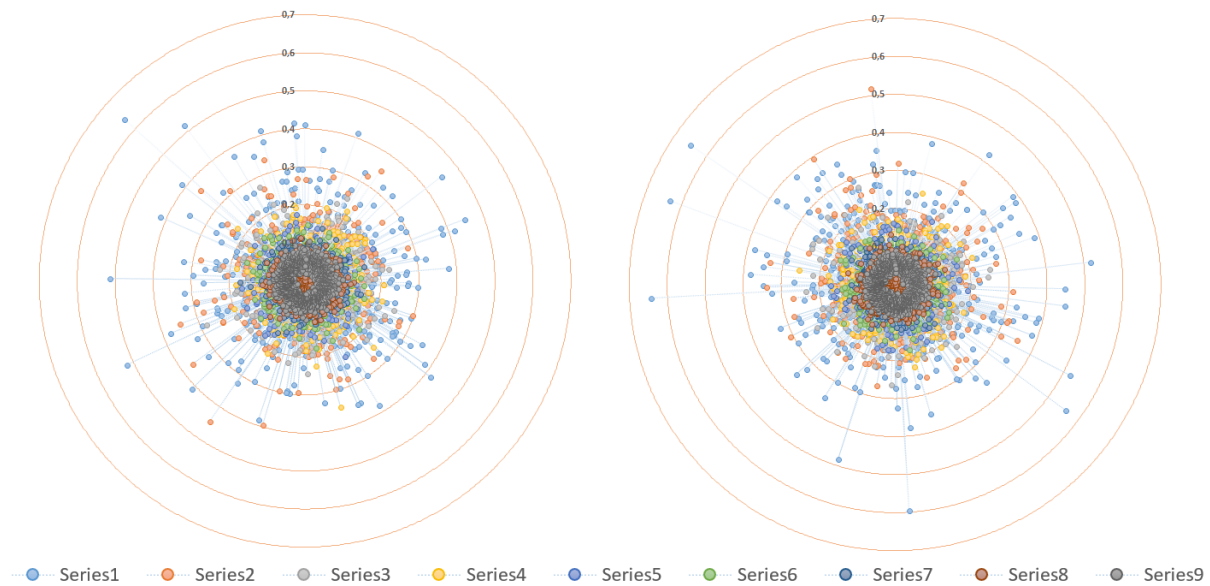
APLICACIÓN MODELO PROBIT										
PORCENTAJE DE LA SUBMUESTRA										
		10%	20%	30%	40%	50%	60%	70%	80%	90%
% ERROR	MÁX	63,93%	44,65%	28,56%	34,32%	22,35%	18,82%	17,27%	13,08%	10,61%
	PROM	11,42%	8,35%	7,06%	6,69%	6,09%	5,83%	5,70%	5,67%	5,71%
	MÍN	0,06%	0,11%	0,07%	0,03%	0,01%	0,01%	0,08%	0,16%	2,11%

Fuente: Elaboración propia. Esta tabla muestra los porcentajes máximos, promedios y mínimos de error del Modelo Probit propuesto con respecto a la siniestralidad real observada para cada una de las 9.000 muestras aleatorias seleccionadas.

APLICACIÓN MODELO LOGIT										
PORCENTAJE DE LA SUBMUESTRA										
		10%	20%	30%	40%	50%	60%	70%	80%	90%
% ERROR	MÁX	65,31%	52,02%	27,57%	25,62%	19,25%	17,48%	15,11%	13,63%	10,84%
	PROM	11,47%	8,39%	6,99%	6,75%	6,11%	5,99%	5,69%	5,78%	5,62%
	MÍN	0,04%	0,07%	0,00%	0,00%	0,00%	0,01%	0,08%	0,01%	1,96%

Fuente: Elaboración propia. Esta tabla muestra los porcentajes máximos, promedios y mínimos de error del Modelo Logit propuesto con respecto a la siniestralidad real observada para cada una de las 9.000 muestras aleatorias seleccionadas.

Por otro lado, probando los modelos sobre el 100% de la “muestra testeo”, se obtiene que, tanto el Modelo Probit como el Modelo Logit predicen, aproximadamente, el 95% de la frecuencia siniestral real de la “muestra testeo”. Estos modelos tienen un alto porcentaje de acierto cuando el colectivo está formado por un elevado número de asegurados, sin embargo, a medida que disminuye el tamaño de la muestra, el porcentaje de error aumenta. A continuación, se muestra gráficamente como aumenta el porcentaje de error conforme disminuye el tamaño de la muestra.



Fuente: Elaboración propia. Estos gráficos representan el porcentaje de errores que se obtienen en los Modelos Lineales Generalizados propuestos, es decir, Modelo Probit y Modelo Logit. El gráfico de la izquierda corresponde al Modelo Probit y el de la derecha al Modelo Logit. Cada observación muestra el porcentaje de error del modelo con respecto a la siniestralidad real observada. Cada serie se ha probado sobre 1.000 colectivos seleccionados de manera aleatoria, de manera que cada colectivo de la Serie 1 contiene un 10% de la “muestra testeo”, la Serie 2 un 20%, la Serie 3 un 30%, la Serie 4 un 40%, la Serie 5 un 50%, la Serie 6 un 60%, la Serie 7 un 70%, la Serie 8 un 80% y la Serie 9 un 90%.

Como conclusión, los Modelos de Regresión estimados, aplicados a esta cartera de Seguros de Vida Riesgo Colectivos Temporales Anuales Renovables, explican la siniestralidad observada con un porcentaje de acierto entorno al 95%, en un modelo en el que la edad de los asegurados no es una variable de riesgo significativa.

7.1.5. Prima de riesgo individual:

Se realiza una comparativa entre las primas de riesgo individuales que se obtienen mediante la aplicación de cada uno de los modelos de regresión seleccionados, es decir, el Modelo Probit y el Modelo Logit.

	VARON	MUJER
NIVEL DE RIESGO 1		
NIVEL DE RIESGO 2		
NIVEL DE RIESGO 3		
NIVEL DE RIESGO 4		

** Diferencia de tasas de un 5%, ** Diferencia de tasas de un 10%,
*** Diferencia de tasas de un 15%, **** Diferencia de tasas de un 20%*

Fuente: Elaboración propia. Esta tabla muestra las diferencias entre las tarifas obtenidas en el Modelo Probit y el Modelo Logit para cada uno de los grupos de riesgo estudiados.

Como puede observarse en esta tabla, las diferencias entre las primas de riesgo del Modelo Probit y del Modelo Logit son inferiores al 5%.

7.1. Aplicación de Modelos de Credibilidad a la Tarificación de Seguros de Vida:

En esta parte de la tesis se realiza una aplicación práctica de dos Modelos de Credibilidad: Modelo de Bühlmann y Bühlmann-Straub. En primer lugar, se explica la aplicación que se ha creado en Visual Basic for Applications⁶ para la ejecución de ambos modelos y, posteriormente, se comparan los resultados obtenidos.

7.1.1. Programación: Modelos Bühlmann y Bühlmann-Straub.

En primer lugar, para ejecutar la macro, se introduce la base de datos que contiene información respecto al sexo, la edad del asegurado, si el asegurado trabaja en una gran empresa o una pyme y el nivel de riesgo asociado a la actividad profesional de la empresa en la que trabaja, en este caso, la empresa tomadora del seguro.

Una vez introducidos los datos, se debe realizar una clasificación de estas observaciones en grupos de riesgo. Existen técnicas para realizar la clasificación por riesgos, por ejemplo, el análisis de conglomerados, también conocido como análisis de clúster, que es una técnica multivariante que tiene como objetivo agrupar variables o elementos en grupos lo más homogéneos posibles, manteniendo heterogeneidad con el resto de grupos.

⁶ VISUAL BASIC FOR APPLICATIONS es un lenguaje de programación de Microsoft que permite desarrollar aplicaciones para Windows. Este lenguaje está integrado en aplicativos como Word, Excel, Access...

En este caso concreto, se programan dos posibles clasificaciones de estas observaciones: categorización por edad y sexo (con el objetivo de realizar una comparativa con el modelo tradicional de tarificación) y categorización por nivel de riesgo y sexo (variables de riesgo según indican los Modelos Logit y Probit anteriormente estimados). En concreto la clasificación por grupos que se ha realizado es la siguiente:

Grupos de riesgo categorizados por sexo y nivel de riesgo asociado a la actividad profesional de la empresa:

GRUPO DE RIESGO	CATEGORIZACIÓN	GRUPO DE RIESGO	CATEGORIZACIÓN
Grupo 1	Sexo: Varón. Nivel de Riesgo: 1	Grupo 5	Sexo: Mujer Nivel de Riesgo: 1
Grupo 2	Sexo: Varón Nivel de Riesgo: 2	Grupo 6	Sexo: Mujer Nivel de Riesgo: 2
Grupo 3	Sexo: Varón Nivel de Riesgo: 3	Grupo 7	Sexo: Mujer Nivel de Riesgo: 3
Grupo 4	Sexo: Varón Nivel de Riesgo: 4	Grupo 8	Sexo: Mujer Nivel de Riesgo: 4

Fuente: Elaboración propia. En esta tabla se describen los requisitos establecidos para la clasificación por grupos atendiendo al sexo y al nivel de riesgo asociado a la actividad profesional de la empresa.

Grupos de riesgo categorizados por edad y sexo:

GRUPO DE RIESGO	CATEGORIZACIÓN	GRUPO DE RIESGO	CATEGORIZACIÓN
Grupo 1	Sexo: Varón Edad: Entre 16 y 20	Grupo 12	Sexo: Mujer Edad: Entre 16 y 20
Grupo 2	Sexo: Varón Edad: Entre 21 y 25	Grupo 13	Sexo: Mujer Edad: Entre 21 y 25
Grupo 3	Sexo: Varón Edad: Entre 26 y 30	Grupo 14	Sexo: Mujer Edad: Entre 26 y 30
Grupo 4	Sexo: Varón Edad: Entre 31 y 35	Grupo 15	Sexo: Mujer Edad: Entre 31 y 35
Grupo 5	Sexo: Varón Edad: Entre 36 y 40	Grupo 16	Sexo: Mujer Edad: Entre 36 y 40

Grupo 6	Sexo: Varón Edad: Entre 41 y 45	Grupo 17	Sexo: Mujer Edad: Entre 41 y 45
Grupo 7	Sexo: Varón Edad: Entre 46 y 50	Grupo 18	Sexo: Mujer Edad: Entre 46 y 50
Grupo 8	Sexo: Varón Edad: Entre 51 y 55	Grupo 19	Sexo: Mujer Edad: Entre 51 y 55
Grupo 9	Sexo: Varón Edad: Entre 56 y 60	Grupo 20	Sexo: Mujer Edad: Entre 56 y 60
Grupo 10	Sexo: Varón Edad: Entre 61 y 65	Grupo 21	Sexo: Mujer Edad: Entre 61 y 65
Grupo 11	Sexo: Varón Edad: Entre 66 y 70	Grupo 22	Sexo: Mujer Edad: Entre 66 y 70

Fuente: Elaboración propia. En esta tabla se describen los requisitos establecidos para la clasificación por grupos atendiendo a la edad del asegurado y a su sexo.

Una vez decidida la clasificación por grupos a realizar, se ejecuta la macro. A continuación, se muestra el primer formulario que aparece al ejecutar esta aplicación.

CREDIBILIDAD BÜHLMANN Y BÜHLMANN STRAUB ×

CLASIFICACION DE GRUPOS

EDAD Y SEXO NIVEL DE RIESGO Y SEXO

RESULTADOS CREDIBILIDAD:

TARIFICACIÓN COLECTIVO:

(*) Introducir la Base de Datos que se utilizará para realizar los Modelos de Credibilidad en la Hoja 2 de este Excel.
 (**) Introducir la Base de Datos del Colectivo a tarificar en la Hoja 4 de este Excel.

Fuente: Elaboración Propia. Formulario creado en Visual Basic a partir del cual se pueden ejecutar cuatro formularios, dos de ellos permiten observar los resultados obtenidos en el Modelo de Bühlmann y Bühlmann-Straub, respectivamente, y los otros dos formularios permiten tarificar a un nuevo colectivo por estos modelos.

Los dos primeros botones que aparecen a la izquierda ejecutan formularios que muestran resultados de estos modelos tales como, los pesos de credibilidad, parámetros estructurales, tarifas... Los otros dos botones que aparecen a la derecha, ejecutan formularios que permiten tarificar a un nuevo colectivo mediante estos modelos.

Suponiendo que se desea ejecutar el modelo de Bühlmann y observar sus resultados, se debe hacer clic el botón “Bühlmann”,

Número de períodos:

Número de grupos:

RESULTADOS:

Factor de heterogeneidad estimado:

Media estimada:

Varianza estimada:

Factor de credibilidad estimado:

EXPUESTOS POR AÑO Y GRUPO:

	AÑO 1	AÑO 2	AÑC
GRUPO 1			
GRUPO 2			
GRUPO 3			
GRUPO 4			
GRUPO 5			
GRUPO 6			
GRUPO 7			
GRUPO 8			
GRUPO 9			
GRUPO 10			
GRUPO 11			
GRUPO 12			
GRUPO 13			
GRUPO 14			
GRUPO 15			
GRUPO 16			

SINIESTROS POR AÑO Y GRUPO:

	AÑO 1	AÑO 2	AÑC
GRUPO 1			
GRUPO 2			
GRUPO 3			
GRUPO 4			
GRUPO 5			
GRUPO 6			
GRUPO 7			
GRUPO 8			
GRUPO 9			
GRUPO 10			
GRUPO 11			
GRUPO 12			
GRUPO 13			
GRUPO 14			
GRUPO 15			
GRUPO 16			

TARIFAS POR GRUPO:

TARIFAS
GRUPO 1
GRUPO 2
GRUPO 3
GRUPO 4
GRUPO 5
GRUPO 6
GRUPO 7
GRUPO 8
GRUPO 9
GRUPO 10
GRUPO 11
GRUPO 12
GRUPO 13
GRUPO 14
GRUPO 15
GRUPO 16
GRUPO 17
GRUPO 18
GRUPO 19
GRUPO 20
GRUPO 21
GRUPO 22

Fuente: Elaboración Propia. Formulario creado en Visual Basic que muestra los resultados que se obtienen realizando el Modelo de Credibilidad de Bühlmann, con clasificación de riesgos en función del sexo y la edad.

Número de períodos:

Número de grupos:

RESULTADOS:

Factor de heterogeneidad estimado:

Media estimada:

Varianza estimada:

Factor de credibilidad estimado:

EXPUESTOS POR AÑO Y GRUPO:

	AÑO 1	AÑO 2	AÑC
GRUPO 1			
GRUPO 2			
GRUPO 3			
GRUPO 4			
GRUPO 5			
GRUPO 6			
GRUPO 7			
GRUPO 8			

SINIESTROS POR AÑO Y GRUPO:

	AÑO 1	AÑO 2	AÑC
GRUPO 1			
GRUPO 2			
GRUPO 3			
GRUPO 4			
GRUPO 5			
GRUPO 6			
GRUPO 7			
GRUPO 8			

TARIFAS POR GRUPO:

TARIFAS
GRUPO 1
GRUPO 2
GRUPO 3
GRUPO 4
GRUPO 5
GRUPO 6
GRUPO 7
GRUPO 8

Fuente: Elaboración Propia. Formulario creado en Visual Basic que muestra los resultados que se obtienen realizando el Modelo de Credibilidad de Bühlmann, con clasificación de riesgos en función del nivel de riesgo y el sexo.

Este formulario muestra la experiencia de la aseguradora respecto al número de expuestos por año y grupo, el número de siniestros por año y grupo, número de períodos de estudio, número de grupos según la clasificación escogida, factor de heterogeneidad

estimado, media estimada, varianza estimada, el factor de credibilidad estimado y las tarifas por grupo.

La aplicación también permite determinar la prima de un nuevo colectivo aplicando las tarifas obtenidas mediante el Modelo de Bühlmann, para ello se clica el botón “Tarificación Credibilidad de Bühlmann”.

VARIABLE CLASIFICATORIA	SEXO	CAPITAL ASEGURADO	PRIMA PURA	PRIMA NETA
-------------------------	------	-------------------	------------	------------

Fuente: Elaboración Propia. Formulario creado en Visual Basic que permite tarificar a un nuevo colectivo mediante el Modelo de Credibilidad de Bühlmann.

A continuación, deben introducirse las bases técnicas que se aplicarán para la tarificación de dicho colectivo, es decir, el tipo de interés técnico, gastos de gestión externa, gastos de gestión interna, y recargo sobre la prima pura que se desea aplicar. Una vez introducidos estos datos, se clica el botón “Tarificar Colectivo” para mostrar las primas desglosadas a nivel asegurado, la prima pura total recargada y la prima neta total.

El proceso es el mismo para ejecutar el Modelo de Bühlman-Straub, si se clica el botón “Bühlman-Straub”, se muestran los siguientes resultados,

Número de períodos:

Número de grupos:

RESULTADOS:

Factor de heterogeneidad estimado:

Media estimada:

Varianza estimada:

Factor de credibilidad estimado:

EXPUESTOS POR AÑO Y GRUPO:

	AÑO 1	AÑO 2	AÑO 3	AÑO 4
GRUPO 1				
GRUPO 2				
GRUPO 3				
GRUPO 4				
GRUPO 5				
GRUPO 6				
GRUPO 7				
GRUPO 8				
GRUPO 9				
GRUPO 10				
GRUPO 11				
GRUPO 12				
GRUPO 13				
GRUPO 14				
GRUPO 15				
GRUPO 16				

TARIFAS POR GRUPO:

	TARIFAS
GRUPO 1	
GRUPO 2	
GRUPO 3	
GRUPO 4	
GRUPO 5	
GRUPO 6	
GRUPO 7	
GRUPO 8	
GRUPO 9	
GRUPO 10	
GRUPO 11	
GRUPO 12	
GRUPO 13	
GRUPO 14	
GRUPO 15	
GRUPO 16	
GRUPO 17	
GRUPO 18	
GRUPO 19	
GRUPO 20	
GRUPO 21	
GRUPO 22	

FACTOR DE CREDIBILIDAD POR GRUPO:

	FACTOR DE CREDIBILIDAD
GRUPO 1	
GRUPO 2	
GRUPO 3	
GRUPO 4	
GRUPO 5	
GRUPO 6	
GRUPO 7	
GRUPO 8	
GRUPO 9	
GRUPO 10	
GRUPO 11	
GRUPO 12	
GRUPO 13	
GRUPO 14	
GRUPO 15	
GRUPO 16	
GRUPO 17	

SINIESTROS POR AÑO Y GRUPO:

	AÑO 1	AÑO 2	AÑO 3	AÑO 4
GRUPO 1				
GRUPO 2				
GRUPO 3				
GRUPO 4				
GRUPO 5				
GRUPO 6				
GRUPO 7				
GRUPO 8				
GRUPO 9				
GRUPO 10				
GRUPO 11				
GRUPO 12				
GRUPO 13				
GRUPO 14				
GRUPO 15				
GRUPO 16				

Fuente: Elaboración Propia. Formulario creado en Visual Basic que muestra los resultados que se obtienen realizando el Modelo de Credibilidad de Bühlmann-Straub, con clasificación de riesgos en función del sexo y la edad.

Número de períodos:

Número de grupos:

RESULTADOS:

Factor de heterogeneidad estimado:

Media estimada:

Varianza estimada:

Factor de credibilidad estimado:

EXPUESTOS POR AÑO Y GRUPO:

	AÑO 1	AÑO 2	AÑO 3	AÑO 4
GRUPO 1				
GRUPO 2				
GRUPO 3				
GRUPO 4				
GRUPO 5				
GRUPO 6				
GRUPO 7				
GRUPO 8				

TARIFAS POR GRUPO:

	TARIFAS
GRUPO 1	
GRUPO 2	
GRUPO 3	
GRUPO 4	
GRUPO 5	
GRUPO 6	
GRUPO 7	
GRUPO 8	

FACTOR DE CREDIBILIDAD POR GRUPO:

	FACTOR DE CREDIBILIDAD
GRUPO 1	
GRUPO 2	
GRUPO 3	
GRUPO 4	
GRUPO 5	
GRUPO 6	
GRUPO 7	
GRUPO 8	

SINIESTROS POR AÑO Y GRUPO:

	AÑO 1	AÑO 2	AÑO 3	AÑO 4
GRUPO 1				
GRUPO 2				
GRUPO 3				
GRUPO 4				
GRUPO 5				
GRUPO 6				
GRUPO 7				
GRUPO 8				

Fuente: Elaboración Propia. Formulario creado en Visual Basic que muestra los resultados que se obtienen realizando el Modelo de Credibilidad de Bühlmann-Straub, con clasificación de riesgos en función del nivel de riesgo y el sexo.

Se puede observar el número de expuestos por año y grupo, el número de siniestros por año y grupo, número de períodos de estudio, número de grupos según la clasificación

escogida, factor de heterogeneidad estimado, media estimada, varianza estimada, los factores de credibilidad estimados por grupos y las tarifas por grupo.

Siguiendo el mismo procedimiento que en el Modelo de Bühlmann, es posible tarificar un colectivo mediante el Modelo de Bühlmann-Straub, para ello se clica el botón “Tarificación Credibilidad de Bühlmann-Straub”.

Fuente: Elaboración Propia. Formulario creado en Visual Basic que permite tarificar a un nuevo colectivo mediante el Modelo de Credibilidad de Bühlmann-Straub.

El código que se ha creado para ejecutar esta aplicación puede verse en el *Anexo 4* de esta tesis.

7.1.2. Prueba de aciertos

En esta parte de la tesis se pretende probar el porcentaje de acierto de los Modelos de Credibilidad de Bühlmann y Bühlmann-Straub, en cuanto a número de siniestros esperados. Para ello se utiliza la “muestra testeó”.

Para realizar esta prueba se elabora un script en Matlab que, a partir de las primas de riesgo individuales obtenidas en los modelos de credibilidad, calcula el número de siniestros que se esperan en cada colectivo. Además de probar los Modelos sobre el 100% de la “muestra testeó”, se prueba sobre el 90%, 80%, ...,10% de la misma, escogida de manera aleatoria, repitiendo este proceso 1.000 veces. El script de Matlab que se ha desarrollado para ejecutar esta prueba puede verse en el *Anexo 5* de esta tesis.

El objetivo es observar el porcentaje de fallos de estos modelos, el cual se obtiene mediante la siguiente ecuación,

$$\% \text{ error} = \left| 1 - \frac{\text{número de siniestros estimados en el colectivo}}{\text{número siniestros reales en el colectivo}} \right|$$

A continuación, se muestran los porcentajes de error máximos, mínimos y promedios obtenidos al aplicar los Modelos Probit y Logit a las 1.000 muestras obtenidas aleatoriamente para el 90%, 80%, ...,10% de la “muestra testeo”.

APLICACIÓN MODELO BÜHLMANN
EDAD Y SEXO

		PORCENTAJE DE LA SUBMUESTRA								
		10%	20%	30%	40%	50%	60%	70%	80%	90%
% ERROR	MÁX	65,01%	49,98%	35,60%	31,73%	27,20%	24,67%	23,27%	20,33%	17,58%
	PROM	15,31%	12,60%	12,34%	12,09%	11,98%	12,09%	11,94%	11,93%	12,02%
	MÍN	0,00%	0,01%	0,03%	0,15%	0,04%	1,59%	1,12%	5,34%	8,07%

Fuente: Elaboración propia. Esta tabla muestra los porcentajes máximos, promedios y mínimos de error del Modelo de Bühlmann con categorización de riesgos en función de la edad y sexo del asegurado con respecto a la siniestralidad real observada para cada una de las muestras aleatorias seleccionadas.

APLICACIÓN MODELO BÜHLMANN-STRAUB
EDAD Y SEXO

		PORCENTAJE DE LA SUBMUESTRA								
		10%	20%	30%	40%	50%	60%	70%	80%	90%
% ERROR	MÁX	68,03%	40,98%	32,39%	23,08%	23,79%	16,53%	15,51%	14,24%	10,01%
	PROM	11,92%	8,44%	7,59%	6,58%	5,97%	5,77%	5,50%	5,52%	5,54%
	MÍN	0,03%	0,00%	0,00%	0,04%	0,00%	0,00%	0,01%	0,10%	1,00%

Fuente: Elaboración propia. Esta tabla muestra los porcentajes máximos, promedios y mínimos de error del Modelo de Bühlmann-Straub con categorización de riesgos en función de la edad y sexo del asegurado con respecto a la siniestralidad real observada para cada una de las muestras aleatorias seleccionadas.

APLICACIÓN MODELO BÜHLMANN
NIVEL DE RIESGO Y SEXO

		PORCENTAJE DE LA SUBMUESTRA								
		10%	20%	30%	40%	50%	60%	70%	80%	90%
% ERROR	MÁX	60,10%	54,09%	32,67%	29,38%	21,95%	22,77%	16,55%	12,95%	11,47%
	PROM	12,38%	8,74%	7,54%	7,01%	6,47%	6,23%	6,10%	6,06%	5,94%
	MÍN	0,21%	0,02%	0,02%	0,02%	0,01%	0,01%	0,01%	0,05%	2,15%

Fuente: Elaboración propia. Esta tabla muestra los porcentajes máximos, promedios y mínimos de error del Modelo de Bühlmann con categorización de riesgos en función del nivel de riesgo asociado a la actividad de la empresa y sexo del asegurado con respecto a la siniestralidad real observada para cada una de las muestras aleatorias seleccionadas.

APLICACIÓN MODELO BÜHLMANN-STRAUB
NIVEL DE RIESGO Y SEXO

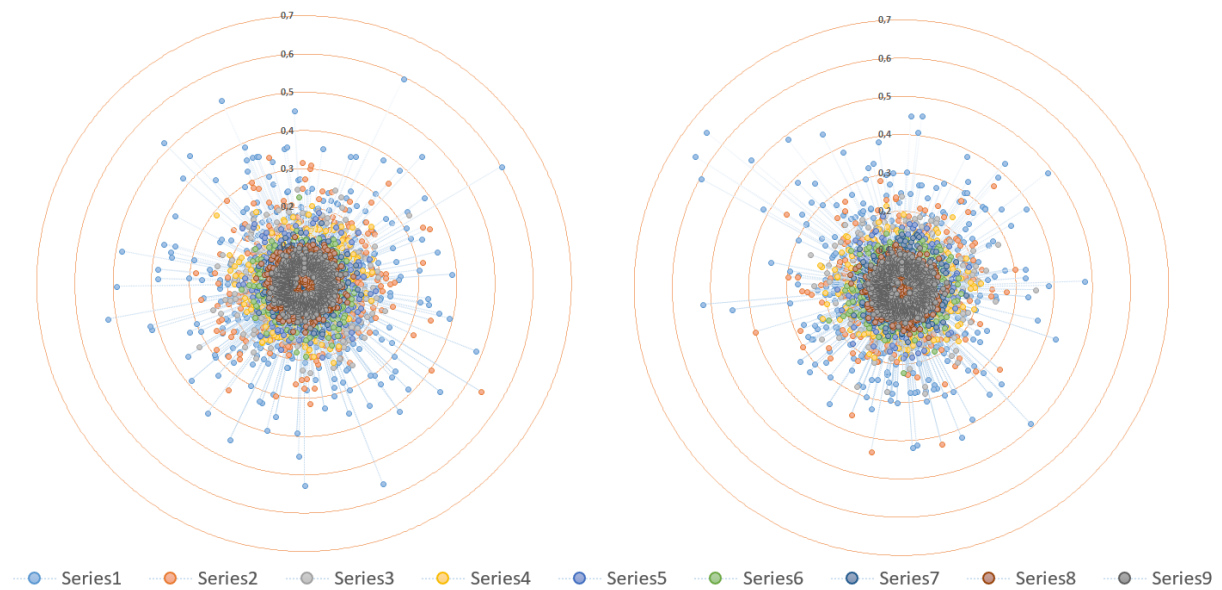
		PORCENTAJE DE LA SUBMUESTRA								
		10%	20%	30%	40%	50%	60%	70%	80%	90%
% ERROR	MÁX	65,40%	43,57%	34,99%	22,99%	19,52%	22,02%	16,87%	14,61%	11,70%
	PROM	11,80%	8,14%	7,53%	7,08%	6,35%	6,26%	6,23%	6,11%	6,07%
	MÍN	0,31%	0,00%	0,01%	0,01%	0,00%	0,08%	0,07%	0,47%	1,76%

Fuente: Elaboración propia. Esta tabla muestra los porcentajes máximos, promedios y mínimos de error del Modelo de Bühlmann-Straub con categorización de riesgos en función del nivel de riesgo asociado a la actividad de la empresa y sexo del asegurado con respecto a la siniestralidad real observada para cada una de las muestras aleatorias seleccionadas.

A continuación, se muestran gráficamente cada uno de los porcentajes de error obtenidos al aplicar los Modelos de Bühlmann y Bühlmann-Straub, respectivamente.



Fuente: Elaboración propia. Estos gráficos representan el porcentaje de errores que se obtienen en los Modelos de Credibilidad con grupos clasificados en función del sexo y de la edad del asegurado. El gráfico de la izquierda corresponde al Modelo de Credibilidad de Bühlmann y el de la derecha al Modelo de Credibilidad de Bühlmann-Straub. Cada observación muestra el porcentaje de error del modelo con respecto a la siniestralidad real observada. Cada serie se ha probado sobre 1.000 colectivos seleccionados de manera aleatoria, de manera que cada colectivo de la Serie 1 contiene un 10% de la “muestra testeo”, la Serie 2 un 20%, la Serie 3 un 30%, la Serie 4 un 40%, la Serie 5 un 50%, la Serie 6 un 60%, la Serie 7 un 70%, la Serie 8 un 80% y la Serie 9 un 90%.



Fuente: Elaboración propia. Estos gráficos representan el porcentaje de errores que se obtienen en los Modelos de Credibilidad con grupos clasificados en función del nivel de riesgo asociado a la actividad de la empresa y del sexo del asegurado. El gráfico de la izquierda corresponde al Modelo de Credibilidad de Bühlmann y el de la derecha al Modelo de Credibilidad de Bühlmann-Straub. Cada observación muestra el porcentaje de error del modelo con respecto a la siniestralidad real observada. Cada serie se ha probado sobre 1.000 colectivos seleccionados de manera aleatoria, de manera que cada colectivo de la Serie 1 contiene un 10% de la “muestra testeó”, la Serie 2 un 20%, la Serie 3 un 30%, la Serie 4 un 40%, la Serie 5 un 50%, la Serie 6 un 60%, la Serie 7 un 70%, la Serie 8 un 80% y la Serie 9 un 90%.

Los Modelos de Credibilidad estimados, aplicados a la “muestra testeó” de Seguros de Vida Riesgo Colectivos Temporales Anuales Renovables, explican la siniestralidad observada con un porcentaje de acierto entorno al 95%.

7.1.3. Resultados: Modelos Bühlmann y Bühlmann-Straub.

En este apartado se comparan los resultados obtenidos en el Modelo de Bühlmann y Bühlmann-Straub. Se muestran tablas en las que se comparan las tarifas obtenidas mediante cada uno de los modelos y se compara la bondad de los modelos para cada una de las categorizaciones atendiendo a los parámetros estructurales y los pesos de credibilidad obtenidos.

A continuación, se muestra una comparativa de las tasas obtenidas en el Modelo de Bühlmann y Bühlmann-Straub en la categorización por sexo y edad,

GRUPO DE RIESGO	DIFERENCIA TASA	GRUPO DE RIESGO	DIFERENCIA TASA
Grupo 1	****	Grupo 12	****
Grupo 2	****	Grupo 13	*
Grupo 3	****	Grupo 14	****
Grupo 4	****	Grupo 15	****
Grupo 5	****	Grupo 16	****
Grupo 6	****	Grupo 17	****
Grupo 7	*	Grupo 18	****
Grupo 8		Grupo 19	*
Grupo 9	*	Grupo 20	
Grupo 10	*	Grupo 21	
Grupo 11	***	Grupo 22	**

** Diferencia de tasas de un 5%, ** Diferencia de tasas de un 10%,*

**** Diferencia de tasas de un 15%, **** Diferencia de tasas de un 20%*

Fuente: Elaboración propia. Esta tabla muestra las diferencias entre las tarifas obtenidas en el Modelo de Credibilidad de Bühlmann y Bühlmann-Straub para cada uno de los grupos de riesgo estudiados.

Seguidamente, se muestra una comparativa de las tasas obtenidas en el Modelo de Bühlmann y Bühlmann-Straub en la categorización por sexo y riesgo asociado a la actividad profesional,

GRUPO DE RIESGO	DIFERENCIA TASA	GRUPO DE RIESGO	DIFERENCIA TASA
Grupo 1		Grupo 5	
Grupo 2		Grupo 6	
Grupo 3		Grupo 7	
Grupo 4		Grupo 8	

** Diferencia de tasas de un 5%, ** Diferencia de tasas de un 10%,*

**** Diferencia de tasas de un 15%, **** Diferencia de tasas de un 20%*

Fuente: Elaboración propia. Esta tabla muestra las diferencias entre las tarifas obtenidas en el Modelo de Credibilidad de Bühlmann y Bühlmann-Straub para cada uno de los grupos de riesgo estudiados.

A diferencia de los modelos de credibilidad con categorización en función del sexo y la edad del asegurado, las tasas obtenidas en los Modelo Bühlmann y Bühlmann-Straub con categorización de riesgos en función de la actividad de la empresa y sexo del asegurado son muy similares, pues no se observan diferencias entre las primas de riesgo individuales obtenidas en cada modelo superiores al 5%.

Atendiendo a la categorización en función del nivel de riesgo asociado a la actividad de la empresa y el sexo del asegurado, se obtiene que ambos modelos presentan valores mínimos de dispersión. Comparando el estimador de dicho parámetro, es decir s^2 , se observa que este es inferior en el Modelo de Bühlmann que en el Modelo de Bühlmann-Straub. Atendiendo a los pesos de credibilidad, en ambos modelos se obtienen pesos de credibilidad elevados y muy similares entre ellos, sin embargo, son algo superiores en el Modelo de Bühlmann, por lo que este favorece en mayor medida la fijación de la prima de riesgo atendiendo a la experiencia de siniestralidad individual.

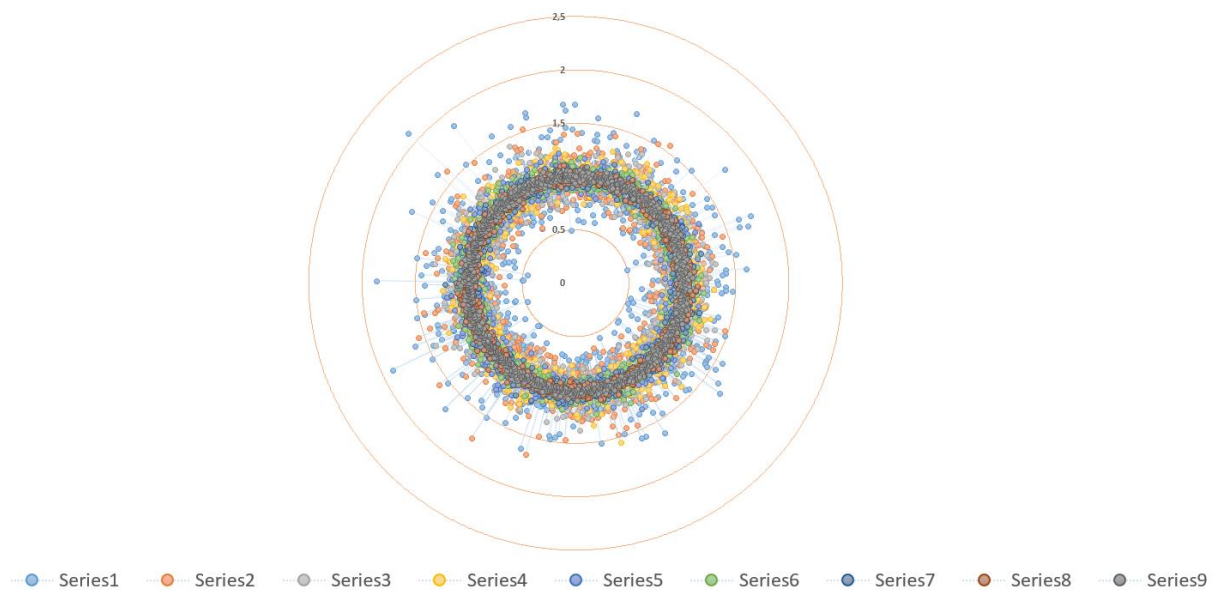
Sin embargo, para los modelos de credibilidad propuestos con categorización en función de la edad y el sexo del asegurado, se obtiene que, aunque ambos modelos presentan valores mínimos de dispersión, estos son superiores a los valores obtenidos en los modelos anteriormente comentados, es decir, en los modelos que atienden a una categorización en función del sexo y nivel de riesgo asociado a la actividad profesional. Comparando el estimador del parámetro estructural S^2 , se observa que es inferior en el Modelo de Bühlmann. Atendiendo a los pesos de credibilidad, se observa que son considerablemente inferiores en el Modelo de Bühlmann-Straub, por lo que este modelo no favorece la fijación de la prima de riesgo atendiendo a la experiencia de siniestralidad individual.

Como puede observarse en la prueba de aciertos del anterior apartado, el porcentaje de errores obtenidos en el modelo de Bühlmann con categorización en función de la edad y sexo del asegurado, es muy superior al del resto de modelos y, el modelo de Bühlmann-Straub no favorece la fijación de la prima según la experiencia de siniestralidad individual, por lo que se concluye que los modelos de credibilidad estimados que presentan mejores resultados son los modelos de credibilidad con categorización en función del sexo del asegurado y el nivel de riesgo asociado a la actividad de la empresa.

8. Comparativa: Modelos Lineales Generalizados, Modelos de Credibilidad y Modelo de Tarificación Tradicional.

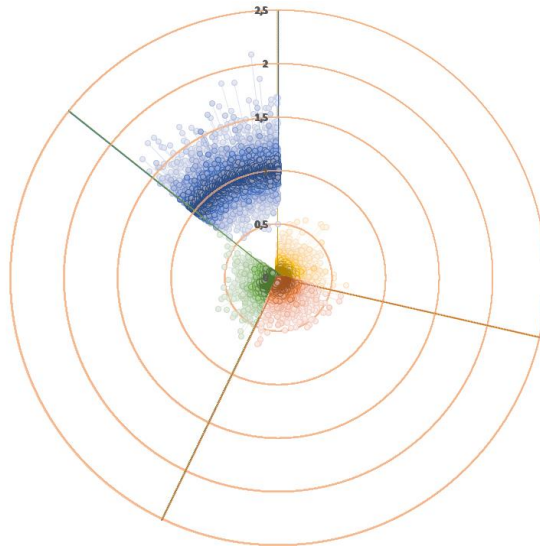
En esta parte de la tesis se realiza una comparativa entre los Modelos Lineales Generalizados, los Modelos de Credibilidad y el Modelo de Tarificación Tradicional. Para ello se lleva a cabo la prueba de aciertos anteriormente comentada, aplicando el

Modelo de Tarificación Tradicional, es decir, la tarificación empleando la Tabla de Mortalidad PASEM/F-2010, eliminando los recargos que presenta esta tabla. Para ello se considera que la compañía tarifica este tipo de seguros a un 80% de la PASEM/F-2010 sin recargos. Dicho porcentaje empleado no es un dato real y, por lo tanto, no guarda ninguna relación con la aseguradora de la que se ha obtenido la base de datos, si no que es una hipótesis que se realiza en esta tesis. El script de Matlab que se ha desarrollado para ejecutar esta prueba puede verse en el *Anexo 6* de esta tesis. Los resultados obtenidos han sido los siguientes,



Fuente: Elaboración propia. Este gráfico representa el porcentaje de errores que se obtiene en el Modelo de Tarificación Tradicional, empleando la Tabla de Mortalidad PASEM/F-2010, sin recargos. Cada observación muestra el porcentaje de error del modelo con respecto a la siniestralidad real observada. Cada serie se ha probado sobre 1.000 colectivos seleccionados de manera aleatoria, de manera que cada colectivo de la Serie 1 contiene un 10% de la “muestra testeó”, la Serie 2 un 20%, la Serie 3 un 30%, la Serie 4 un 40%, la Serie 5 un 50%, la Serie 6 un 60%, la Serie 7 un 70%, la Serie 8 un 80% y la Serie 9 un 90%.

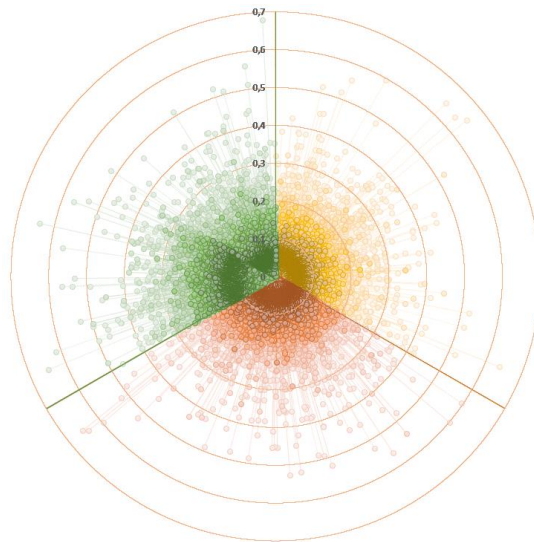
A continuación, se muestra en un solo gráfico los resultados obtenidos en todos los modelos propuestos, incluyendo este último,



Fuente: Elaboración propia. Este gráfico representa los porcentajes de errores que se obtienen en cada uno de los modelos estudiados, es decir, Modelo de Tarificación Tradicional, en color azul, Modelos de Credibilidad con clasificación de riesgos en función del sexo y la edad, en color verde, Modelos de Credibilidad con clasificación de riesgos en función del sexo y el nivel de riesgo asociado a la actividad de la empresa, en color rojo, y Modelos Lineales Generalizados, en color amarillo. Cada observación muestra el porcentaje de error del modelo con respecto a la siniestralidad real observada. Cuando se observa un color menos intenso de cada uno de los errores representados, la muestra escogida de manera aleatoria contiene una proporción inferior de la “muestra testeó”.

Como puede observarse en este gráfico, hay un error de modelo cuando se emplean las tablas de mortalidad para tarificar este tipo de seguros que se están estudiando, lo que hace preciso aplicar modelos de tarificación basados en la experiencia de la compañía, como los modelos propuestos de Credibilidad y GLM.

En cuanto a estos modelos basados en la experiencia, se adjunta un gráfico comparativo,



Fuente: Elaboración propia. Este gráfico representa el porcentaje de errores que se obtiene en los Modelos de Credibilidad con clasificación de riesgos en función del sexo y la edad, en color verde, Modelos de Credibilidad con clasificación de riesgos en función del sexo y el nivel de riesgo asociado a la actividad de la empresa, en color rojo, y Modelos Lineales Generalizados, en color amarillo. Cada observación muestra el porcentaje de error del modelo con respecto a la siniestralidad real observada. Conforme se aclara el color de cada uno de los errores representados, se indica que la muestra escogida de manera aleatoria contiene una proporción inferior de la “muestra testeó”.

Comparando estos tres modelos basados en la experiencia, y atendiendo a los estimadores de los parámetros estructurales, en los modelos de credibilidad, se concluye que el modelo de credibilidad con categorización en función de la edad y el sexo del asegurado, no favorece la determinación de la prima de riesgo basada en la experiencia de siniestralidad individual, en el caso del Modelo de Bühlmann-Straub y, en el caso del Modelo de Bühlmann, que es el que presenta menor dispersión en la siniestralidad individual, se obtienen porcentajes de error superiores al resto de modelos basados en la experiencia que se plantean en esta tesis.

Los modelos de tarificación basados en el nivel de riesgo asociado a la actividad de la empresa y el sexo del asegurado son los que mayor porcentaje de aciertos presentan, así como mayores pesos de confianza en el caso de los modelos de credibilidad. Aunque estos modelos presentan resultados muy similares, se observan porcentajes de error inferiores en los Modelos Probit y Logit.

VARIABLES DE RIESGO NIVEL DE RIESGO Y SEXO		
TRAMOS % ERROR	GLM	CREDIBILIDAD
[0%,5%]	7.098	6.344
[5%,10%]	7.571	8.014
[10%,15%]	2.044	2.259
[15%,20%]	777	793
[20%,25%]	248	300
[25%,30%]	142	121
[30%,35%]	48	73
[35%,40%]	44	50
[40%,45%]	14	17
[45%,50%]	2	16
[50%,55%]	6	7
[55%,60%]	2	3
[60%,65%]	3	2
[65%,70%]	1	1

Fuente: Elaboración propia. Esta tabla muestra el porcentaje de error que se ha obtenido en la aplicación de los modelos de Credibilidad y GLM a los distintos colectivos generados de manera aleatoria a partir de la “muestra testeo”.

Por último, considerando estos modelos basados en la experiencia se obtiene que, atendiendo al criterio de juicio experto de mercado, la tasa pura de colectivos de fallecimiento equivale al 60% de la Tabla de Mortalidad PASEM/F-2010 sin recargos. Destacar que, la tarificación de fallecimientos en Seguros Temporales Anuales Renovables Colectivos, a dicho porcentaje de la tabla, no incorporaría recargos por margen de beneficios.

9. Gestión de cartera: Modelos Lineales Generalizados y Modelos de Credibilidad.

En esta parte de la tesis se comenta una posible gestión adaptada a los modelos planteados basados en la experiencia individual de siniestralidad.

Tal y como se ha ido comentando en estos últimos apartados de la tesis, los modelos basados en la experiencia propia de la aseguradora predicen un alto porcentaje de la siniestralidad real, sin embargo, conforme disminuye el tamaño de la cartera, el porcentaje de errores aumenta. Por lo tanto, estos modelos de tarificación son eficientes cuando se aplican a nivel cartera y no a nivel individual de cada póliza.

La gestión de la cartera desde el punto de vista de la tarificación de seguros basada en estos modelos, debe dividir la cartera global de la aseguradora en tantas subcarteras

como grupos de riesgo se hayan determinado, de manera que, el montante de primas de cada subcartera deba soportar la siniestralidad real que se derive de dicho riesgo.

Este sistema de tarificación basado en la experiencia propia de la compañía debe realimentarse conforme se adquiere más información, de manera que las primas de riesgo individuales se ajusten al riesgo soportado, pues el comportamiento de siniestralidad para cada grupo puede variar en el tiempo.

Por último, como se ha comentado, los Seguros Temporales Anuales Renovables, que son objeto de estudio de esta tesis, pueden contemplar garantías adicionales al fallecimiento, por lo que se recomienda realizar estos modelos para cada una de estas garantías, de manera que, cada garantía y riesgo soporten la siniestralidad que se derive de dicha subcartera.

10. Conclusiones

Las entidades aseguradoras almacenan un gran volumen de datos que debe ser empleado para la modelización de riesgos. Esto ya ocurre en Seguros No Vida donde, para la fijación de la prima de seguro, se llevan a cabo métodos estadísticos más sofisticados como GLM, Redes neuronales, credibilidad, árboles de decisión... Sin embargo, el desarrollo de estas técnicas no es común en la tarificación de Seguros de Vida, donde la determinación de la prima atiende especialmente a características biológicas como la edad y, en ocasiones, el sexo del asegurado.

En el negocio de Seguros de Vida Individuales, se pueden determinar numerosas variables de riesgo, pues, dado que el Tomador del Seguro es el propio y único asegurado la solicitud de información es menos costosa que en Seguros Colectivos, donde se asegura a toda una colectividad, por lo tanto, se pueden obtener numerosas variables que pueden ser determinantes del riesgo como, por ejemplo, ámbito geográfico, redes sociales, nivel educativo, estado civil, hábitos saludables...

La selección de riesgos que se aplica en estos seguros permite disponer de bases de datos formadas por numerosas variables que, aunque se utilizan para una valoración individual del riesgo, su aplicación no se extiende a la elaboración técnicas o modelos estadísticos sofisticados.

A pesar de que la información que se puede obtener en Seguros Individuales sería muy costosa de adquirir en Seguros Colectivos, pueden estudiarse otras posibles variables de riesgo, como, por ejemplo, el tamaño de la empresa, su ámbito de actividad profesional, el canal comercial, figuras de intervención...

Los resultados obtenidos en los modelos estudiados concluyen que, la utilización de modelos estadísticos más sofisticados, basados en la experiencia propia de la

aseguradora, se ajustan adecuadamente a la siniestralidad real de la cartera, por lo que se concluye que estas técnicas son adecuadas para la modelización de riesgos, en este caso, del riesgo de fallecimiento.

La tendencia en el mercado asegurador de Vida debe dirigirse al uso de variables socioeconómicas, biológicas e, incluso, propias del producto contratado para la modelización del riesgo, atendiendo a la experiencia de cada compañía y al uso de técnicas predictivas. El uso de tablas de mortalidad se limita a la valoración individual de los asegurados que componen el colectivo. Los modelos propuestos van más allá de una valoración individual de cada asegurado al incorporar a la tarificación variables que dependen de la actividad de la empresa, en lugar de la actividad concreta que realiza el asegurado pues, en estas variables ya se incluye la distribución de profesiones.

En la aplicación de estos métodos se destaca la importancia de disponer de una base de datos de cada colectividad asegurable y datos de siniestralidad de la misma fiable y estructurada para un correcto análisis de la información almacenada por la compañía. Además, estos modelos deben ser realimentados con información actualizada, garantizando que las primas obtenidas se ajusten adecuadamente a cada riesgo.

En definitiva, en base a los positivos resultados obtenidos en esta tesis, se concluye que el mercado asegurador de Vida Riesgo Colectivo debe reestructurar las bases de su tarificación, dando mayor peso a una tarificación basada en modelos estadísticos más sofisticados elaborados mediante bases de datos de la propia compañía.

11. Próximas líneas de investigación:

En cuanto a los modelos de credibilidad, para aseguradoras con reaseguro de exceso de pérdida (*excess of loss reinsurance*), sería interesante estudiar los modelos de credibilidad incorporando capitales asegurados y distinguiendo la parte a cargo de la cedente de la parte reasegurada.

En cuanto a los Modelos Lineales Generalizados, se plantea el estudio de otro tipo de modelos como, por ejemplo, modelos de conteo. Además, la incorporación de más variables a estos modelos como los canales de intervención, ámbito geográfico de la empresa... y la incorporación de interacciones entre las variables.

Por último, se propone la búsqueda de una posible unión entre ambos modelos, es decir, entre los Modelos Lineales Generalizados y los Modelos de Credibilidad.

12. Bibliografía:

- [1] Goulet, V. Principles and Application of Credibility Theory. Concordia University Canada. Journal of Actuarial Practice Vol. 6; 1998.
- [2] Howard C. Mahler y Curtis Gary Dean. “Credibility”. Chapter 8 Foundations of Casualty Actuarial Sciences, 4th edition; 2001.
- [3] Mora, R. Microeconomía Cuantitativa. Grado en Economía. Universidad Carlos III de Madrid. Getafe; 2016.
- [4] Moreno Muñoz M, Ramos Burgoa, L. Aplicación de Modelos de Credibilidad para el Cálculo de Primas en el Seguro de Automóviles. Comisión Nacional de Seguros y Finanzas; 2003.
- [5] Peña Sánchez, I. Aplicación práctica de Modelos de Credibilidad en la tarificación de Seguros de Salud. Fundación Mapfre. Nº 108; 2010.
- [6] Pons Cardell, M. La Teoría de la Credibilidad y su Aplicación a los Seguros Colectivos. Universidad de Barcelona; 1991.
- [7] Rodríguez-Pardo del Castillo, J. Modelos predictivos aplicados al seguro de vida. Gerencia de riesgos y seguros, Nº. 114; 2012, págs. 24-35.
- [8] Tablas de Mortalidad de la Población Asegurada Española. Investigación Cooperativa entre Entidades Aseguradoras (ICEA) y Munich Re.
- [9] Usabel, M. Modelos de Supervivencia. Máster en Ciencias Actuariales y Financieras. Universidad Carlos III de Madrid. Puerta de Toledo, 2017.
- [10] Usabel, M. Seguros de Vida. Máster en Ciencias Actuariales y Financieras. Universidad Carlos III de Madrid. Puerta de Toledo, 2017.
- [11] Wooldridge, J. 4ed. Introducción la Econometría. Un enfoque moderno. Michigan State University.

Anexo 1: Separación aleatoria de la base de datos.

Programación en Matlab.

```

1  %SEPARACIÓN ALEATORIA DE LA MUESTRA:
2
3  %Se convierte la tabla importada en un vector:
4  data=table2array(BBDD);
5  %Vector que indica el número de filas y columnas de la Base de datos:
6  dimfilas=size(data);
7
8  %transpose: Devuelve la traspuesta de una matriz.
9  %sort: Ordena los valores en orden ascendente.
10 %randperm(n,k): devuelve un vector de fila que contiene
11 %k enteros únicos seleccionados aleatoriamente de 1 a n inclusive.
12 r=transpose(sort(randperm(dimfilas(1,1),floor(dimfilas(1,1)*0.5))));
13
14 %Se proporciona dimension al vector que será la muestra a aplicar en el
15 %modelo y el vector que contiene la muestra sobre la que se probará el
16 %modelo.
17 vector_modelo=zeros(floor(dimfilas(1,1)*0.5),dimfilas(1,2));
18 vector_aplicacion=zeros(floor(dimfilas(1,1)*0.5),dimfilas(1,2));
19 j=1
20 k=1
21
22 %Se generan los vectores
23 for i=1:dimfilas(1,1)
24     if i==r(j,1)
25         vector_modelo(j,1)=data(i,1);%siniestro
26         vector_modelo(j,2)=data(i,2);%edad
27         vector_modelo(j,3)=data(i,3);%sexo
28         vector_modelo(j,4)=data(i,4);%gran_empresa
29         vector_modelo(j,5)=data(i,5);%nivel_riesgo
30         if j==floor(dimfilas(1,1)*0.5)
31             j=j;
32         else
33             j=j+1;
34         end
35     else
36         vector_aplicacion(k,1)=data(i,1);%siniestro
37         vector_aplicacion(k,2)=data(i,2);%edad
38         vector_aplicacion(k,3)=data(i,3);%sexo
39         vector_aplicacion(k,4)=data(i,4);%gran_empresa
40         vector_aplicacion(k,5)=data(i,5);%nivel_riesgo
41         if k==floor(dimfilas(1,1)*0.5)
42             k=k;
43         else
44             k=k+1;
45         end
46     end
47 end
48

```

Anexo 2: Modelos Lineales Generalizados (GLM).

Programación en Gretl.

```

1  # Estadísticos descriptivos de las variables independientes:
2  summary EDAD SEXO GRAN_EMPRESA NIVELDERIESGO
3  # Frecuencias de las variables independientes:
4  freq EDAD
5  freq SEXO
6  freq GRAN_EMPRESA
7  freq NIVELDERIESGO
8  # Estadísticos descriptivos de la variable dependiente:
9  summary SINIESTRO
10 # Frecuencias de la variable dependiente:
11 freq EDAD
12 # Estadísticos descriptivos de las variables independientes.
13 # Separando muestras cuando SINIESTRO=0 y cuando SINIESTRO=1:
14 summary EDAD --by=SINIESTRO
15 summary SEXO --by=SINIESTRO
16 summary GRAN_EMPRESA --by=SINIESTRO
17 summary NIVELDERIESGO --by=SINIESTRO
18
19 # MODELOS:
20 #Además de variables dummy por cada edad, se crean 6 variables dummy con tramos de
21 10 años de edad.
22
23 loop i=1..6
24     genr edad$i=(EDAD>((i*10)) && (EDAD<=(10+(i*10))) #EDAD==$i
25 endloop
26
27 #MCO:
28 ols SINIESTRO const dummify(EDAD) dummify(SEXO) dummify(GRAN_EMPRESA)
29 dummify(NIVELDERIESGO) --robust
30 genr uhat=$uhat
31 omit dummify(EDAD) --quiet
32 smpl --full
33 omit dummify(SEXO) --quiet
34 smpl --full
35 omit dummify(GRAN_EMPRESA) --quiet
36 smpl --full
37 omit dummify(NIVELDERIESGO) --quiet
38 smpl --full
39 ols SINIESTRO const dummify(EDAD) dummify(SEXO) dummify(GRAN_EMPRESA)
40 dummify(NIVELDERIESGO) --robust
41 reset --quiet #test linealidad
42 modtest --white --quiet #test de heteroscedasticidad de White
43 modtest --breusch-pagan --quiet #test de heteroscedasticidad de Breusch - Pagan
44 normtest uhat --jbera #normalidad de los residuos
45 normtest uhat --swilk #normalidad de los residuos
46
47 ols SINIESTRO const edad2 edad3 edad4 edad5 edad6 dummify(SEXO)
48 dummify(GRAN_EMPRESA) dummify(NIVELDERIESGO) --robust
49 genr uhat=$uhat
50 omit edad2 edad3 edad4 edad5 edad6 --quiet
51 smpl --full
52 omit dummify(SEXO) --quiet
53 smpl --full
54 omit dummify(GRAN_EMPRESA) --quiet
55 smpl --full
56 omit dummify(NIVELDERIESGO) --quiet
57 smpl --full
58 ols SINIESTRO const edad2 edad3 edad4 edad5 edad6 dummify(SEXO)
59 dummify(GRAN_EMPRESA) dummify(NIVELDERIESGO) --robust
60 reset --quiet #test linealidad
61 modtest --white --quiet #test de heteroscedasticidad de White
62 modtest --breusch-pagan --quiet #test de heteroscedasticidad de Breusch - Pagan
63 normtest uhat --jbera #normalidad de los residuos
64 normtest uhat --swilk #normalidad de los residuos
65
66 #PROBIT:#####
67 #####
68
69 probit SINIESTRO const dummify(EDAD) dummify(SEXO) dummify(GRAN_EMPRESA)
70 dummify(NIVELDERIESGO) --robust --verbose --p-values
71
72 # Se estudia la significatividad conjunta utilizando los métodos: Wald y
73 Multiplicador de Lagrange

```

```

66 # Hipótesis Nula del Contraste: Las variables edad no son significativas de
manera conjunta.
67 # Hipótesis Alternativa del Contraste: Las variables edad son significativas
de manera conjunta.
68 #WALD:
69 probit SINIESTRO const dummify(EDAD) dummify(SEXO) dummify(GRAN_EMPRESA)
dummify(NIVELDERIESGO) --robust --quiet
70 omit dummify(EDAD) --wald
71 #LAGRANGE:
72 # Verosimilitud del modelo sin restringir:
73 probit SINIESTRO const dummify(EDAD) dummify(SEXO) dummify(GRAN_EMPRESA)
dummify(NIVELDERIESGO) --robust --quiet
74 scalar lur=$lnl
75 # Verosimilitud del modelo restringido:
76 probit SINIESTRO const dummify(SEXO) dummify(GRAN_EMPRESA)
dummify(NIVELDERIESGO) --robust --quiet
77 scalar lr=$lnl
78 #Estadístico de Lagrange:
79 # Formado por la log-verosimilitud del modelo no restringido menos el
restringido)
80 scalar LR=2*(lur-lr)
81 #Comparando LR con su valor crítico en una X^2
82 scalar pvalor=pvalue(X, k-1, LR) #Número de restricciones - 1 = grados de
libertad.
83 # Mostrar resultado en pantalla:
84 printf "LR=%.8g\n pvalue=%.8g", LR, pvalor
85 probit SINIESTRO const dummify(EDAD) dummify(SEXO) dummify(GRAN_EMPRESA)
dummify(NIVELDERIESGO) --robust --verbose --p-values
86 # Hipótesis Nula del Contraste: Las variables nivel de riesgo no son
significativas de manera conjunta.
87 # Hipótesis Alternativa del Contraste: Las variables nivel de riesgo son
significativas de manera conjunta.
88 #WALD:
89 probit SINIESTRO const dummify(EDAD) dummify(SEXO) dummify(GRAN_EMPRESA)
dummify(NIVELDERIESGO) --robust --quiet
90 omit dummify(NIVELDERIESGO) --wald
91 #LAGRANGE:
92 # Verosimilitud del modelo sin restringir:
93 probit SINIESTRO const dummify(EDAD) dummify(SEXO) dummify(GRAN_EMPRESA)
dummify(NIVELDERIESGO) --robust --quiet
94 scalar lur=$lnl
95 # Verosimilitud del modelo restringido:
96 probit SINIESTRO const dummify(EDAD) dummify(SEXO) dummify(GRAN_EMPRESA)
--robust --quiet
97 scalar lr=$lnl
98 #Estadístico de Lagrange:
99 # Formado por la log-verosimilitud del modelo no restringido menos el
restringido.
100 scalar LR=2*(lur-lr)
101 #Comparando LR con su valor crítico en una X^2
102 scalar pvalor=pvalue(X, k-1, LR) #Número de restricciones menos 1=grados de
libertad.
103 # Mostrar resultado en pantalla:
104 printf "LR=%.8g\n pvalue=%.8g", LR, pvalor
105
106 #CONTRASTE DE NORMALIDAD DE LOS RESIDUOS:
107 probit SINIESTRO const dummify(EDAD) dummify(SEXO) dummify(GRAN_EMPRESA)
dummify(NIVELDERIESGO) --robust --quiet
108 modtest --normality
109
110 probit SINIESTRO const edad2 edad3 edad4 edad5 edad6 dummify(SEXO)
dummify(GRAN_EMPRESA) dummify(NIVELDERIESGO) --robust --verbose --p-values
111
112 # Se estudia la significatividad conjunta utilizando los métodos: Wald y
Multiplicador de Lagrange
113 # Hipótesis Nula del Contraste: Las variables edad no son significativas de
manera conjunta.
114 # Hipótesis Alternativa del Contraste: Las variables edad son significativas
de manera conjunta.
115 #WALD:
116 probit SINIESTRO const edad2 edad3 edad4 edad5 edad6 dummify(SEXO)
dummify(GRAN_EMPRESA) dummify(NIVELDERIESGO) --robust --quiet
117 omit edad2 edad3 edad4 edad5 edad6 --wald

```

```

118     #LAGRANGE:
119     # Verosimilitud del modelo sin restringir:
120     probit SINIESTRO const edad2 edad3 edad4 edad5 edad6 dummify(SEXO)
dummify(GRAN_EMPRESA) dummify(NIVELDERIESGO) --robust --quiet
121     scalar lur=$lnl
122     # Verosimilitud del modelo restringido:
123     probit SINIESTRO const dummify(SEXO) dummify(GRAN_EMPRESA)
dummify(NIVELDERIESGO) --robust --quiet
124     scalar lr=$lnl
125     #Estadístico de Lagrange:
126     # Formado por la log-verosimilitud del modelo no restringido menos el
restringido)
127     scalar LR=2*(lur-lr)
128     #Comparando LR con su valor crítico en una X^2
129     scalar pvalor=pvalue(X, k-1, LR) #Número de restricciones - 1 = grados de
libertad.
130     # Mostrar resultado en pantalla:
131     printf "LR=%.8g\n pvalue=%.8g", LR, pvalor
132     probit SINIESTRO const edad2 edad3 edad4 edad5 edad6 dummify(SEXO)
dummify(GRAN_EMPRESA) dummify(NIVELDERIESGO) --robust --verbose --p-values
133 # Hipótesis Nula del Contraste: Las variables nivel de riesgo no son
significativas de manera conjunta.
134 # Hipótesis Alternativa del Contraste: Las variables nivel de riesgo son
significativas de manera conjunta.
135     #WALD:
136     probit SINIESTRO const edad2 edad3 edad4 edad5 edad6 dummify(SEXO)
dummify(GRAN_EMPRESA) dummify(NIVELDERIESGO) --robust --quiet
137     omit dummify(NIVELDERIESGO) --wald
138     #LAGRANGE:
139     # Verosimilitud del modelo sin restringir:
140     probit SINIESTRO const edad2 edad3 edad4 edad5 edad6 dummify(SEXO)
dummify(GRAN_EMPRESA) dummify(NIVELDERIESGO) --robust --quiet
141     scalar lur=$lnl
142     # Verosimilitud del modelo restringido:
143     probit SINIESTRO const edad2 edad3 edad4 edad5 edad6 dummify(SEXO)
dummify(GRAN_EMPRESA) --robust --quiet
144     scalar lr=$lnl
145     #Estadístico de Lagrange:
146     # Formado por la log-verosimilitud del modelo no restringido menos el
restringido)
147     scalar LR=2*(lur-lr)
148     #Comparando LR con su valor crítico en una X^2
149     scalar pvalor=pvalue(X, k-1, LR) #Número de restricciones menos 1=grados de
libertad.
150     # Mostrar resultado en pantalla:
151     printf "LR=%.8g\n pvalue=%.8g", LR, pvalor
152
153     #CONTRASTE DE NORMALIDAD DE LOS RESIDUOS:
154     probit SINIESTRO const edad2 edad3 edad4 edad5 edad6 dummify(SEXO)
dummify(GRAN_EMPRESA) dummify(NIVELDERIESGO) --robust --quiet
155     modtest --normality
156
157
158     probit SINIESTRO const dummify(SEXO) dummify(GRAN_EMPRESA) dummify(NIVELDERIESGO)
--robust --verbose --p-values
159
160     # Hipótesis Nula del Contraste: Las variables de niveles de riesgo no son
significativas de manera conjunta.
161     # Hipótesis Alternativa del Contraste: Las variables de niveles de riesgo
son significativas de manera conjunta.
162     #WALD:
163     probit SINIESTRO const dummify(SEXO) dummify(GRAN_EMPRESA)
dummify(NIVELDERIESGO) --robust --quiet
164     omit dummify(NIVELDERIESGO) --wald
165     #LAGRANGE:
166     # Verosimilitud del modelo sin restringir:
167     probit SINIESTRO const dummify(SEXO) dummify(GRAN_EMPRESA)
dummify(NIVELDERIESGO) --robust --quiet
168     scalar lur=$lnl
169     # Verosimilitud del modelo restringido:
170     probit SINIESTRO const dummify(SEXO) dummify(GRAN_EMPRESA) --robust --quiet
171     scalar lr=$lnl
172     #Estadístico de Lagrange:

```

```

173     # Formado por la log-verosimilitud del modelo no restringido menos el
174     restringido)
175     scalar LR=2*(lur-lr)
176     #Comparando LR con su valor crítico en una X^2
177     scalar pvalor=pvalue(X, k-1, LR) #Número de restricciones menos 1=grados de
178     libertad.
179     # Mostrar resultado en pantalla:
180     printf "LR=%.8g\n pvalue=%.8g", LR, pvalor
181
182     #CONTRASTE DE NORMALIDAD DE LOS RESIDUOS:
183     probit SINIESTRO const dummify(SEXO) dummify(GRAN_EMPRESA)
184     dummify(NIVELDERIESGO) --robust --quiet
185     modtest --normality
186
187     probit SINIESTRO const dummify(SEXO) dummify(NIVELDERIESGO) --robust --verbose
188     --p-values
189
190     # Hipótesis Nula del Contraste: Las variables de niveles de riesgo no son
191     significativas de manera conjunta.
192     # Hipótesis Alternativa del Contraste: Las variables de niveles de riesgo
193     son significativas de manera conjunta.
194     #WALD:
195     probit SINIESTRO const dummify(SEXO) dummify(NIVELDERIESGO) --robust
196     --quiet
197     omit dummify(NIVELDERIESGO) --wald
198     #LAGRANGE:
199     # Verosimilitud del modelo sin restringir:
200     probit SINIESTRO const dummify(SEXO) dummify(NIVELDERIESGO) --robust
201     --quiet
202     scalar lur=$lnl
203     # Verosimilitud del modelo restringido:
204     probit SINIESTRO const dummify(SEXO) --robust --quiet
205     scalar lr=$lnl
206     #Estadístico de Lagrange:
207     # Formado por la log-verosimilitud del modelo no restringido menos el
208     restringido)
209     scalar LR=2*(lur-lr)
210     #Comparando LR con su valor crítico en una X^2
211     scalar pvalor=pvalue(X, k-1, LR) #Número de restricciones menos 1=grados de
212     libertad.
213     # Mostrar resultado en pantalla:
214     printf "LR=%.8g\n pvalue=%.8g", LR, pvalor
215     # Hipótesis Nula del Contraste: Las variables explicativas no son
216     significativas de manera conjunta.
217     # Hipótesis Alternativa del Contraste: Las variables explicativas son
218     significativas de manera conjunta.
219     #WALD:
220     probit SINIESTRO const dummify(SEXO) dummify(NIVELDERIESGO) --robust
221     --quiet
222     omit dummify(SEXO) dummify(NIVELDERIESGO) --wald
223     #LAGRANGE:
224     # Verosimilitud del modelo sin restringir:
225     probit SINIESTRO const dummify(SEXO) dummify(NIVELDERIESGO) --robust
226     --quiet
227     scalar lur=$lnl
228     # Verosimilitud del modelo restringido:
229     probit SINIESTRO const --robust --quiet
230     scalar lr=$lnl
231     #Estadístico de Lagrange:
232     # Formado por la log-verosimilitud del modelo no restringido menos el
233     restringido)
234     scalar LR=2*(lur-lr)
235     #Comparando LR con su valor crítico en una X^2
236     scalar pvalor=pvalue(X, k-1, LR) #Número de restricciones menos 1=grados de
237     libertad.
238     # Mostrar resultado en pantalla:
239     printf "LR=%.8g\n pvalue=%.8g", LR, pvalor
240
241     #CONTRASTE DE NORMALIDAD DE LOS RESIDUOS:
242     probit SINIESTRO const dummify(SEXO) dummify(NIVELDERIESGO) --robust
243     --quiet

```



```

229         modtest --normality
230 #####
231 #####
232
233 #LOGIT:#####
234 #####
235
236 logit SINIESTRO const dummify(EDAD) dummify(SEXO) dummify(GRAN_EMPRESA)
dummify(NIVELDERIESGO) --robust --verbose --p-values
237
238 # Se estudia la significatividad conjunta utilizando los métodos: Wald y
Multiplicador de Lagrange
239     # Hipótesis Nula del Contraste: Las variables edad no son significativas de
manera conjunta.
240     # Hipótesis Alternativa del Contraste: Las variables edad son significativas
de manera conjunta.
241     #WALD:
242     logit SINIESTRO const dummify(EDAD) dummify(SEXO) dummify(GRAN_EMPRESA)
dummify(NIVELDERIESGO) --robust --quiet
243     omit dummify(EDAD) --wald
244     #LAGRANGE:
245     # Verosimilitud del modelo sin restringir:
246     logit SINIESTRO const dummify(EDAD) dummify(SEXO) dummify(GRAN_EMPRESA)
dummify(NIVELDERIESGO) --robust --quiet
247     scalar lur=$lnl
248     # Verosimilitud del modelo restringido:
249     logit SINIESTRO const dummify(SEXO) dummify(GRAN_EMPRESA)
dummify(NIVELDERIESGO) --robust --quiet
250     scalar lr=$lnl
251     #Estadístico de Lagrange:
252     # Formado por la log-verosimilitud del modelo no restringido menos el
restringido)
253     scalar LR=2*(lur-lr)
254     #Comparando LR con su valor crítico en una X^2
255     scalar pvalor=pvalue(X, k-1, LR) #Número de restricciones - 1 = grados de
libertad.
256     # Mostrar resultado en pantalla:
257     printf "LR=%.8g\n pvalue=%.8g", LR, pvalor
258     logit SINIESTRO const dummify(EDAD) dummify(SEXO) dummify(GRAN_EMPRESA)
dummify(NIVELDERIESGO) --robust --verbose --p-values
259     # Hipótesis Nula del Contraste: Las variables nivel de riesgo no son
significativas de manera conjunta.
260     # Hipótesis Alternativa del Contraste: Las variables nivel de riesgo son
significativas de manera conjunta.
261     #WALD:
262     logit SINIESTRO const dummify(EDAD) dummify(SEXO) dummify(GRAN_EMPRESA)
dummify(NIVELDERIESGO) --robust --quiet
263     omit dummify(NIVELDERIESGO) --wald
264     #LAGRANGE:
265     # Verosimilitud del modelo sin restringir:
266     logit SINIESTRO const dummify(EDAD) dummify(SEXO) dummify(GRAN_EMPRESA)
dummify(NIVELDERIESGO) --robust --quiet
267     scalar lur=$lnl
268     # Verosimilitud del modelo restringido:
269     logit SINIESTRO const dummify(EDAD) dummify(SEXO) dummify(GRAN_EMPRESA)
--robust --quiet
270     scalar lr=$lnl
271     #Estadístico de Lagrange:
272     # Formado por la log-verosimilitud del modelo no restringido menos el
restringido)
273     scalar LR=2*(lur-lr)
274     #Comparando LR con su valor crítico en una X^2
275     scalar pvalor=pvalue(X, k-1, LR) #Número de restricciones menos 1=grados de
libertad.
276     # Mostrar resultado en pantalla:
277     printf "LR=%.8g\n pvalue=%.8g", LR, pvalor
278
279 logit SINIESTRO const edad2 edad3 edad4 edad5 edad6 dummify(SEXO)
dummify(GRAN_EMPRESA) dummify(NIVELDERIESGO) --robust --verbose --p-values
280
281 # Se estudia la significatividad conjunta utilizando los métodos: Wald y

```

```

Multiplicador de Lagrange
282 # Hipótesis Nula del Contraste: Las variables edad no son significativas de
manera conjunta.
283 # Hipótesis Alternativa del Contraste: Las variables edad son significativas
de manera conjunta.
284 #WALD:
285 logit SINIESTRO const edad2 edad3 edad4 edad5 edad6 dummify(SEXO)
dummify(GRAN_EMPRESA) dummify(NIVELDERIESGO) --robust --quiet
286 omit edad2 edad3 edad4 edad5 edad6 --wald
287 #LAGRANGE:
288 # Verosimilitud del modelo sin restringir:
289 logit SINIESTRO const edad2 edad3 edad4 edad5 edad6 dummify(SEXO)
dummify(GRAN_EMPRESA) dummify(NIVELDERIESGO) --robust --quiet
290 scalar lur=$lnl
291 # Verosimilitud del modelo restringido:
292 logit SINIESTRO const dummify(SEXO) dummify(GRAN_EMPRESA)
dummify(NIVELDERIESGO) --robust --quiet
293 scalar lr=$lnl
294 #Estadístico de Lagrange:
295 # Formado por la log-verosimilitud del modelo no restringido menos el
restringido)
296 scalar LR=2*(lur-lr)
297 #Comparando LR con su valor crítico en una X^2
298 scalar pvalor=pvalue(X, k-1, LR) #Número de restricciones - 1 = grados de
libertad.
299 # Mostrar resultado en pantalla:
300 printf "LR=%.8g\n pvalue=%.8g", LR, pvalor
301 logit SINIESTRO const edad2 edad3 edad4 edad5 edad6 dummify(SEXO)
dummify(GRAN_EMPRESA) dummify(NIVELDERIESGO) --robust --verbose --p-values
302 # Hipótesis Nula del Contraste: Las variables nivel de riesgo no son
significativas de manera conjunta.
303 # Hipótesis Alternativa del Contraste: Las variables nivel de riesgo son
significativas de manera conjunta.
304 #WALD:
305 logit SINIESTRO const edad2 edad3 edad4 edad5 edad6 dummify(SEXO)
dummify(GRAN_EMPRESA) dummify(NIVELDERIESGO) --robust --quiet
306 omit dummify(NIVELDERIESGO) --wald
307 #LAGRANGE:
308 # Verosimilitud del modelo sin restringir:
309 logit SINIESTRO const edad2 edad3 edad4 edad5 edad6 dummify(SEXO)
dummify(GRAN_EMPRESA) dummify(NIVELDERIESGO) --robust --quiet
310 scalar lur=$lnl
311 # Verosimilitud del modelo restringido:
312 logit SINIESTRO const edad2 edad3 edad4 edad5 edad6 dummify(SEXO)
dummify(GRAN_EMPRESA) --robust --quiet
313 scalar lr=$lnl
314 #Estadístico de Lagrange:
315 # Formado por la log-verosimilitud del modelo no restringido menos el
restringido)
316 scalar LR=2*(lur-lr)
317 #Comparando LR con su valor crítico en una X^2
318 scalar pvalor=pvalue(X, k-1, LR) #Número de restricciones menos 1=grados de
libertad.
319 # Mostrar resultado en pantalla:
320 printf "LR=%.8g\n pvalue=%.8g", LR, pvalor
321
322
323 logit SINIESTRO const dummify(SEXO) dummify(GRAN_EMPRESA) dummify(NIVELDERIESGO)
--robust --verbose --p-values
324
325 # Hipótesis Nula del Contraste: Las variables de niveles de riesgo no son
significativas de manera conjunta.
326 # Hipótesis Alternativa del Contraste: Las variables de niveles de riesgo
son significativas de manera conjunta.
327 #WALD:
328 logit SINIESTRO const dummify(SEXO) dummify(GRAN_EMPRESA)
dummify(NIVELDERIESGO) --robust --quiet
329 omit dummify(NIVELDERIESGO) --wald
330 #LAGRANGE:
331 # Verosimilitud del modelo sin restringir:
332 logit SINIESTRO const dummify(SEXO) dummify(GRAN_EMPRESA)
dummify(NIVELDERIESGO) --robust --quiet
333 scalar lur=$lnl

```

```

334 # Verosimilitud del modelo restringido:
335 logit SINIESTRO const dummify(SEXO) dummify(GRAN_EMPRESA) --robust --quiet
336 scalar lr=$lnl
337 #Estadístico de Lagrange:
338 # Formado por la log-verosimilitud del modelo no restringido menos el
restringido)
339 scalar LR=2*(lur-lr)
340 #Comparando LR con su valor crítico en una X^2
341 scalar pvalor=pvalue(X, k-1, LR) #Número de restricciones menos 1=grados de
libertad.
342 # Mostrar resultado en pantalla:
343 printf "LR=%.8g\n pvalue=%.8g", LR, pvalor
344
345
346
347 logit SINIESTRO const dummify(SEXO) dummify(NIVELDERIESGO) --robust --verbose
--p-values
348
349 # Hipótesis Nula del Contraste: Las variables de niveles de riesgo no son
significativas de manera conjunta.
350 # Hipótesis Alternativa del Contraste: Las variables de niveles de riesgo
son significativas de manera conjunta.
351 #WALD:
352 logit SINIESTRO const dummify(SEXO) dummify(NIVELDERIESGO) --robust --quiet
353 omit dummify(NIVELDERIESGO) --wald
354 #LAGRANGE:
355 # Verosimilitud del modelo sin restringir:
356 logit SINIESTRO const dummify(SEXO) dummify(NIVELDERIESGO) --robust --quiet
357 scalar lur=$lnl
358 # Verosimilitud del modelo restringido:
359 logit SINIESTRO const dummify(SEXO) --robust --quiet
360 scalar lr=$lnl
361 #Estadístico de Lagrange:
362 # Formado por la log-verosimilitud del modelo no restringido menos el
restringido)
363 scalar LR=2*(lur-lr)
364 #Comparando LR con su valor crítico en una X^2
365 scalar pvalor=pvalue(X, k-1, LR) #Número de restricciones menos 1=grados de
libertad.
366 # Mostrar resultado en pantalla:
367 printf "LR=%.8g\n pvalue=%.8g", LR, pvalor
368 # Hipótesis Nula del Contraste: Las variables explicativas no son
significativas de manera conjunta.
369 # Hipótesis Alternativa del Contraste: Las variables explicativas son
significativas de manera conjunta.
370 #WALD:
371 logit SINIESTRO const dummify(SEXO) dummify(NIVELDERIESGO) --robust --quiet
372 omit dummify(SEXO) dummify(NIVELDERIESGO) --wald
373 #LAGRANGE:
374 # Verosimilitud del modelo sin restringir:
375 logit SINIESTRO const dummify(SEXO) dummify(NIVELDERIESGO) --robust --quiet
376 scalar lur=$lnl
377 # Verosimilitud del modelo restringido:
378 logit SINIESTRO const --robust --quiet
379 scalar lr=$lnl
380 #Estadístico de Lagrange:
381 # Formado por la log-verosimilitud del modelo no restringido menos el
restringido)
382 scalar LR=2*(lur-lr)
383 #Comparando LR con su valor crítico en una X^2
384 scalar pvalor=pvalue(X, k-1, LR) #Número de restricciones menos 1=grados de
libertad.
385 # Mostrar resultado en pantalla:
386 printf "LR=%.8g\n pvalue=%.8g", LR, pvalor
387
388
389 #####
#####

```

Anexo 3: Prueba de aciertos.
Modelos Lineales Generalizados
(GLM).

Programación en Matlab.

```

1  %PRUEBA ACIERTOS MODELOS LINEALES GENERALIZADOS: PROBIT Y LOGIT.
2
3  %Se convierte la tabla importada en un vector:
4  data=table2array(APLICACION);
5  %Vector que indica el número de filas y columnas de la Base de datos:
6  dimfilas=size(data);
7  num_bucles=1000
8  %Se proporciona dimension al vector de siniestros estimados.
9  siniestros_estimados=zeros(num_bucles,9);
10
11  %se indica si se ejecutará el modelo probit o logit.
12  probit=0;
13  logit=1-probit;
14
15  %Se proporciona valor a los coeficientes obtenidos en el modelo.
16  if probit==1
17      b_const=0 %beta 0
18      b_sexo=0 %beta sexo
19      b_gran_empresa=0 %beta gran empresa
20      b_nivel2=0 %beta nivel de riesgo 2
21      b_nivel3=0 %beta nivel de reiso 3
22      b_nivel4=0 %beta nivel de riesgo 4
23  end
24  %Se proporciona valor a los coeficientes obtenidos en el modelo.
25  if logit==1
26      b_const=0 %beta 0
27      b_sexo=0 %beta sexo
28      b_gran_empresa=0 %beta gran empresa
29      b_nivel2=0 %beta nivel de riesgo 2
30      b_nivel3=0 %beta nivel de riesgo 3
31      b_nivel4=0 %beta nivel de riesgo 4
32  end
33
34  for x=1:9
35      proporcion=x/10 %proporciones que se van seleccionar de la muestra.
36          %10%,20%,30%,40%,50%,60%,70%,80%,90%
37      %A continuación se indica el número de observaciones a seleccionar.
38      proporcion_aplicacion=floor(fix(dimfilas(1,1)*(proporcion)));
39      z=1;
40      for z=1:1000 %se ejecutan los 1.000 bucles
41          j=1;
42          %Se proporciona dimension a los vectores:
43          vector_aplicacion=zeros(proporcion_aplicacion,dimfilas(1,2));
44          probabilidad_estimada=zeros(proporcion_aplicacion,1);
45          siniestros=zeros(proporcion_aplicacion,1);
46          %transpose: Devuelve la traspuesta de una matriz.
47          %sort: Ordena los valores en orden ascendente.
48          %randperm(n,k): devuelve un vector de fila que contiene
49          %k enteros únicos seleccionados aleatoriamente de 1 a n inclusive.
50          r=transpose(sort(randperm(dimfilas(1,1),proporcion_aplicacion)));
51          for i=1:dimfilas(1,1) %El bucle pasa por todas las observaciones.
52              if i==r(j,1) %Este bucle pasa unicamente por las observaciones
53                  %que se han escogido de manera aleatoria
54                  vector_aplicacion(j,1)=data(i,1);%siniestro
55                  siniestros(j,1)=data(i,1);%siniestro
56                  vector_aplicacion(j,2)=data(i,2);%edad
57                  vector_aplicacion(j,3)=data(i,3);%sexo
58                  vector_aplicacion(j,4)=data(i,4);%gran_empresa
59                  vector_aplicacion(j,5)=data(i,5);%nivel_riesgo
60                  if data(i,5)==2 %Si el nivel de riesgo es 2
61                      b_riesgo=b_nivel2;%se capta el coeficiente
62                  elseif data(i,5)==3 %Si el nivel de riesgo es 3
63                      b_riesgo=b_nivel3;%se capta el coeficiente
64                  elseif data(i,5)==4 %Si el nivel de riesgo es 4
65                      b_riesgo=b_nivel4;%se capta el coeficiente
66                  else %Si el nivel de riesgo es 1
67                      b_riesgo=0;%no se capta coeficiente
68                  end
69                  %Se estima la probabilidad de fallecimiento:
70
71                  y_estimada(j,1)=b_const+(data(i,3)*b_sexo)+(data(i,4)*b_gran_empresa)+b_riesgo
72                  ;
73                  if probit==1

```

```

72         %funcion enlace
73         probabilidad_estimada(j,1)=normcdf(y_estimada(j,1));
74     end
75     if logit==1
76         %funcion enlace
77
78         probabilidad_estimada(j,1)=exp(y_estimada(j,1))/(1+exp(y_estim
79         ada(j,1)));
80     end
81     if j==proporcion_aplicacion
82         break;
83     end
84     end
85     %vector siniestros estimados:
86     siniestros_estimados(z,x)=sum(probabilidad_estimada);
87     %vector siniestros observados:
88     siniestros_observados(z,x)=sum(siniestros);
89     %vector proporcion siniestros estimados sobre los observados.
90
91     proporcion_estimador_observados(z,x)=siniestros_estimados(z,x)/siniestros_obse
92     rvados(z,x);
93 end
94 end

```

Anexo 4: Modelos de Credibilidad, Bühlmann y Bühlmann-Straub.

Programación en Visual Basic.

```

1  ' Se define:
2  'n: número de asegurados.
3  't: número de años.
4  'k: numero de grupos.
5
6  'Subrutina vectores_grupos:
7  'INPUTS: Vectores de los años objeto de estudio. k número de grupos. t número de
8  años.
9  'OUTPUTS:
10  'expuestos_año_grupo() Vector que contiene el número de expuestos por año y
11  grupo.
12  'siniestros_año_grupo() Vector que contiene el número de siniestros por año
13  y grupo.
14  Sub vectores_grupos(ByRef bbdd() As Integer, ByRef expuestos_año_grupo() As Long,
15  ByRef siniestros_año_grupo() As Long, _
16  ByRef clasificación As Integer, ByRef k As Integer, ByRef t As Integer)
17
18  'A continuación se genera el vector que indica el grupo de riesgo a que
19  pertenece el asegurado para el año.
20  Dim grupo_expuestos() As Integer
21  Dim grupo_siniestros() As Integer
22  'Se genera el vector que asigna un grupo a cada individuo.
23  asignacion_grupo bbdd(), grupo_expuestos(), grupo_siniestros(), clasificacion,
24  k, t
25
26  Dim i As Long 'contador de bucles.
27  Dim l As Long 'contador de bucles.
28  Dim j As Long 'contador de bucles.
29
30  ReDim expuestos_año_grupo(1 To k, 1 To t)
31  ReDim siniestros_año_grupo(1 To k, 1 To t)
32  For j = 1 To t 'Contador de años
33  For l = 1 To k 'Contador número de grupos.
34  siniestros_año_grupo(l, j) = 0 'Se parte del valor 0.
35  expuestos_año_grupo(l, j) = 0 'Se parte del valor 0.
36  For i = 1 To UBound(grupo_expuestos()) 'Contador número de asegurados.
37  If grupo_siniestros(i, j) = 1 Then
38  siniestros_año_grupo(l, j) = siniestros_año_grupo(l, j) + 1
39  'Se crea el vector.
40  End If
41  If grupo_expuestos(i, j) = 1 Then
42  expuestos_año_grupo(l, j) = expuestos_año_grupo(l, j) + 1
43  'Se crea el vector.
44  End If
45  Next i
46  Next l
47  Next j
48  End Sub
49
50  'Subrutina asignacion_grupo:
51  'INPUTS: Vector bbdd, contiene las variables edad, sexo y siniestro. k número de
52  grupos. t número de años.
53  'OUTPUTS:
54  'grupo_expuestos() vector que asigna un grupo de riesgo a cada individuo.
55  'grupo_siniestro() vector que asigna un grupo de riesgo a cada individuo que
56  ha causado siniestro.
57  Sub asignacion_grupo(ByRef bbdd() As Integer, ByRef grupo_expuestos() As Integer,
58  ByRef grupo_siniestros() As Integer, _
59  ByRef clasificacion As Integer, ByRef k As Integer, ByRef t As Integer)
60
61  ReDim grupo_expuestos(1 To UBound(bbdd()), 1 To t) 'Número de filas=Número
62  asegurados.
63
64  'Número de columnas=Número
65  de años.
66
67  Dim i As Long 'contador de bucles.
68  Dim j As Long 'contador de bucles.
69  Dim l As Long 'contador de bucles.
70  Dim w As Long 'contrador de bucles.
71
72  If clasificacion = 1 Then
73  For w = 3 To t * 3 Step 3

```



```

61         For i = 1 To UBound(bbdd()) 'Bucle número asegurado.
62             For j = 1 To (k / 2) 'Bucle número grupo asignado
63                 If bbdd(i, w - 1) = 1 Then 'Si el sexo del asegurado es Varón.
64                     grupo_expuestos(i, w / 3) = 0
65                     For l = 15 To 65 Step 5 'Bucle edad grupo.
66                         If bbdd(i, w - 2) > l Then
67                             grupo_expuestos(i, w / 3) = grupo_expuestos(i, w
68                                 / 3) + 1
69                             End If
70                         Next l
71                     Else
72                         If bbdd(i, w - 2) = 0 Then
73                             Else
74                                 grupo_expuestos(i, w / 3) = k / 2
75                                 For l = 15 To 65 Step 5 'Bucle edad grupo.
76                                     If bbdd(i, w - 2) > l Then
77                                         grupo_expuestos(i, w / 3) =
78                                             grupo_expuestos(i, w / 3) + 1
79                                         End If
80                                     Next l
81                                 End If
82                             End If
83                         Next j
84                     Next i
85                 Next w
86             ElseIf clasificacion = 2 Then
87                 For w = 3 To t * 3 Step 3
88                     For i = 1 To UBound(bbdd()) 'Bucle número asegurado.
89                         For j = 1 To (k / 2) 'Bucle número grupo asignado
90                             If bbdd(i, w - 1) = 1 Then 'Si el sexo del asegurado es Varón.
91                                 grupo_expuestos(i, w / 3) = bbdd(i, w - 2)
92                             Else
93                                 If bbdd(i, w - 2) = 0 Then
94                                     Else
95                                         grupo_expuestos(i, w / 3) = k / 2 + bbdd(i, w - 2)
96                                         End If
97                                     End If
98                                 End If
99                             Next j
100                             Next i
101                             Next w
102                         End If
103                     ReDim grupo_siniestros(1 To UBound(bbdd()), 1 To t) 'Número de filas=Número
104                         asegurado.
105
106                         'Número de columnas=Número
107                         de años.
108
109                     For w = 3 To t * 3 Step 3
110                         For i = 1 To UBound(bbdd())
111                             grupo_siniestros(i, w / 3) = bbdd(i, w) * grupo_expuestos(i, w / 3)
112                         Next i
113                     Next w
114                 End Sub
115
116 'Subrutina Modelo buhlmann:
117 'INPUTS:
118 ' Base de datos: bbdd().
119 ' Número de grupos: k
120 ' Número de años: t
121 'OUTPUTS:
122 ' Estimador de la media, mediante el Modelo buhlmann
123 ' Estimador de la varianza, mediante el Modelo buhlmann
124 ' Estimador del factor de heterogeneidad, mediante el Modelo buhlmann
125 ' Estimador del factor de credibilidad, mediante el Modelo buhlmann
126 ' Tarifas Modelo Buhlmann
127 Sub Modelo_buhlmann(ByRef bbdd() As Integer, ByRef clasificacion As Integer, ByRef k
128 As Integer, _
129 ByRef t As Integer, ByRef estimador_media As Double, ByRef estimador_varianza As
130 Double, _
131 ByRef estimador_factor_heterogeneidad As Double, ByRef factor_credibilidad As
132 Double, _
133 ByRef tarifas_buhlmann() As Double)
134
135 ' Se definen las variables output que se van a obtener de cada subrutina.
136 Dim expuestos_año_grupo() As Long

```

```

127 Dim siniestros_año_grupo() As Long
128
129 vectores_grupos bdd(), expuestos_año_grupo(), siniestros_año_grupo(),
    clasificacion, k, t
130
131 ReDim proporcion_siniestros_expuestos(1 To k, 1 To t) As Double
132 ReDim promedio_proporcion(1 To k) As Double
133 Dim i As Long 'contador de bucles.
134 Dim j As Long 'contador de bucles.
135
136 For l = 1 To k
137
138     For j = 1 To t
139         If siniestros_año_grupo(l, j) = 0 Or expuestos_año_grupo(l, j) = 0
140             Then
141                 proporcion_siniestros_expuestos(l, j) = 0
142             Else
143                 proporcion_siniestros_expuestos(l, j) = siniestros_año_grupo(l,
144                     j) / expuestos_año_grupo(l, j)
145             End If
146         promedio_proporcion(l) = promedio_proporcion(l) +
147             proporcion_siniestros_expuestos(l, j) / t
148     Next j
149
150     estimador_media = estimador_media + promedio_proporcion(l) / k ' Se crea
151     el estimador de la media.
152 Next l
153
154 For l = 1 To k
155     For j = 1 To t
156         ' Se crea el estimador de la variación.
157         estimador_varianza = estimador_varianza +
158             ((proporcion_siniestros_expuestos(l, j) - promedio_proporcion(l)) ^
159             2) / ((t - 1) * k)
160     Next j
161     ' Se crea el estimador del factor de heterogeneidad.
162     estimador_factor_heterogeneidad = estimador_factor_heterogeneidad +
163         ((promedio_proporcion(l) - estimador_media) ^ 2)
164 Next l
165
166 estimador_factor_heterogeneidad = (estimador_factor_heterogeneidad / (k - 1)) -
167 (estimador_varianza / t)
168
169 ' Se calcula el factor de credibilidad Z:
170 factor_credibilidad = (estimador_factor_heterogeneidad * t) /
171 (estimador_varianza + (estimador_factor_heterogeneidad * t))
172
173 'TARIFAS MODELO CREDIBILIDAD Bühlmann:
174 ReDim tarifas_buhlmann(1 To k) As Double
175 'Se limita la Colectividad asegurable a asegurados con edades comprendidas
176 entre los 17 y 70 años.
177 For i = 1 To k
178     ' Se crean las tarifas del modelo Bühlmann.
179     tarifas_buhlmann(i) = ((1 - factor_credibilidad) * estimador_media) +
180         (factor_credibilidad * promedio_proporcion(i))
181 Next i
182
183 End Sub
184
185 'Subrutina Modelo buhlmann Straub:
186 'INPUTS:
187     ' Base de datos: bdd().
188     ' Número de grupos: k
189     ' Número de años: t
190     ' Clasificacion de grupos.
191 'OUTPUTS:
192     ' Estimador de la media, mediante el Modelo buhlmann Straub
193     ' Estimador de la varianza, mediante el Modelo buhlmann Straub
194     ' Estimador del factor de heterogeneidad, mediante el Modelo buhlmann Straub
195     ' Estimador del factor de credibilidad, mediante el Modelo buhlmann Straub

```

```

188     ' Tarifas Modelo Buhlmann Straub.
189 Sub Modelo_buhlmann_Straub(ByRef bbdd() As Integer, ByRef clasificacion As Integer,
ByRef k As Integer, ByRef t As Integer, _
190     ByRef estimador_media As Double, ByRef estimador_varianza As Double, _
191     ByRef estimador_factor_heterogeneidad As Double, ByRef
factor_credibilidad_grupo() As Double, _
192     ByRef factor_credibilidad As Double, ByRef tarifas_buhlmann_Straub() As Double)
193
194     ' Se definen las variables output que se van a obtener de cada subrutina.
195 Dim expuestos_año_grupo() As Long
196 Dim siniestros_año_grupo() As Long
197
198     vectores_grupos bbdd(), expuestos_año_grupo(), siniestros_año_grupo(),
clasificacion, k, t
199
200 ReDim proporcion_siniestros_expuestos(1 To k, 1 To t) As Double
201 ReDim promedio_proporcion(1 To k) As Double
202 ReDim factor_ponderacion_expuestos_año_grupo(1 To k, 1 To t) As Double
203 ReDim factor_ponderacion_expuestos_grupo(1 To k) As Double
204 ReDim factor_ponderacion_grupo(1 To k) As Double
205 Dim suma_factores_ponderacion As Double
206 ReDim siniestralidad_media_ponderada_grupo(1 To k) As Double
207 ReDim factor_credibilidad_grupo(1 To k) As Double
208 Dim i As Long 'contador de bucles.
209 Dim j As Long 'contador de bucles.
210
211     For l = 1 To k
212         For j = 1 To t
213             If siniestros_año_grupo(l, j) = 0 Or expuestos_año_grupo(l, j) = 0
Then
214                 proporcion_siniestros_expuestos(l, j) = 0
215             Else
216                 proporcion_siniestros_expuestos(l, j) = siniestros_año_grupo(l,
j) / expuestos_año_grupo(l, j)
217             End If
218             promedio_proporcion(l) = promedio_proporcion(l) +
(promencion_siniestros_expuestos(l, j) / t)
219             factor_ponderacion_expuestos_año_grupo(l, j) =
expuestos_año_grupo(l, j) 'Factor de ponderación de cada grupo en
cada año.
220             factor_ponderacion_expuestos_grupo(l) =
factor_ponderacion_expuestos_grupo(l) +
factor_ponderacion_expuestos_año_grupo(l, j)
221             factor_ponderacion_grupo(l) = factor_ponderacion_grupo(l) +
(proporcion_siniestros_expuestos(l, j) *
factor_ponderacion_expuestos_año_grupo(l, j))
222         Next j
223
224         suma_factores_ponderacion = suma_factores_ponderacion +
factor_ponderacion_expuestos_grupo(l)
225
226         If factor_ponderacion_expuestos_grupo(l) = 0 Or
factor_ponderacion_grupo(l) = 0 Then
227             siniestralidad_media_ponderada_grupo(l) = 0 'Se evita la
indeterminación n/0, forzando el valor 0.
228         Else
229             siniestralidad_media_ponderada_grupo(l) =
factor_ponderacion_grupo(l) / factor_ponderacion_expuestos_grupo(l)
230         End If
231     Next l
232
233     For l = 1 To k
234         siniestralidad_media_ponderada = siniestralidad_media_ponderada +
(factor_ponderacion_expuestos_grupo(l) * promedio_proporcion(l) /
suma_factores_ponderacion)
235     Next l
236
237     For l = 1 To k
238         For j = 1 To t
239             ' Se crea el estimador de la varianza.
240             estimador_varianza = estimador_varianza +

```

```

                (factor_ponderacion_expuestos_año_grupo(l, j) *
                ((proporcion_siniestros_expuestos(l, j) -
                siniestralidad_media_ponderada_grupo(l) ^ 2)) / (k * (t - 1))
243 Next j
244 sumatorio = sumatorio + ((factor_ponderacion_expuestos_grupo(l) /
                suma_factores_ponderacion) * ((siniestralidad_media_ponderada_grupo(l) -
                siniestralidad_media_ponderada) ^ 2))
245 divisor = divisor + (factor_ponderacion_expuestos_grupo(l) * (1 -
                factor_ponderacion_expuestos_grupo(l) / suma_factores_ponderacion) /
                suma_factores_ponderacion)
246 Next l
247
248 ' Se crea el estimador del factor de heterogeneidad.
249 estimador_factor_heterogeneidad = (sumatorio - ((k - 1) *
                (estimador_varianza / suma_factores_ponderacion))) / divisor
250
251 For l = 1 To k
252     factor_credibilidad_grupo(l) = estimador_factor_heterogeneidad *
                factor_ponderacion_expuestos_grupo(l) / (estimador_varianza +
                (estimador_factor_heterogeneidad * factor_ponderacion_expuestos_grupo(l)))
253     factor_credibilidad = factor_credibilidad + factor_credibilidad_grupo(l)
254     estimador_media = estimador_media + (factor_credibilidad_grupo(l) *
                promedio_proporcion(l))
255 Next l
256
257 'Se crea el estimador de la media.
258 estimador_media = estimador_media / factor_credibilidad
259
260 ReDim tarifas_buhlmann_Straub(1 To k)
261 For l = 1 To k
262     'Se crean las tarifas del modelo Bühlmann-Straub
263     tarifas_buhlmann_Straub(l) = ((1 - factor_credibilidad_grupo(l)) *
                estimador_media) + (factor_credibilidad_grupo(l) *
                siniestralidad_media_ponderada_grupo(l))
264 Next l
265 End Sub
266
267
268
269
270 '.....'ARRAY
271 'Subrutina que transforma un rango en un vector.
272 'Input:
273 'rango: Rango Excel.
274 'Output:
275 'nombre(): Array.
276 Public Sub RangetoArray_integer(ByRef nombre() As Integer, ByVal rango As Range)
277     Dim inter As Variant
278     Dim i As Long
279     inter = rango
280     ReDim nombre(1 To UBound(inter, 1))
281     'Conversión:
282     For i = 1 To UBound(inter, 1)
283         nombre(i) = inter(i, 1)
284     Next i
285 End Sub
286
287
288

```

```
1 'Inicio de los formularios:
2
3 Private Sub cb_boton_Click()
4     'Ejecuta el formulario de bühlmann
5     buhlmann_formulario.Show
6 End Sub
7
8 Private Sub cb_tarificacion_boton_Click()
9     'Ejecuta el formulario para tarificar a un colectivo usando el modelo de bühlmann
10    tarificacion_buhlmann.Show
11 End Sub
12
13 Private Sub cbs_boton_Click()
14     'Ejecuta el formulario de bühlmann-straub
15     buhlmann_straub_formulario.Show
16 End Sub
17
18 Private Sub cbs_tarificacion_boton_Click()
19     'Ejecuta el formulario para tarificar a un colectivo usando el modelo de
20     bühlmann-straub
21     tarificacion_buhlmann_straub.Show
22 End Sub
23
24 Private Sub edadsexoboton_Click()
25     Hoja1.Cells(1, 16) = 1
26     Hoja5.Cells(1, 4) = 1
27 End Sub
28
29 Private Sub riesgosexoboton_Click()
30     Hoja1.Cells(1, 16) = 2
31     Hoja5.Cells(1, 4) = 2
32 End Sub
33
34 Public Sub UserForm_initialize()
35     'Al ejecutar el formulario aparece por defecto la clasificación por sexo y edad.
36     edadsexoboton.Value = True
37 End Sub
38
```

```

1  ' Formulario de Bühlmann
2
3  Public Sub UserForm_initialize()
4
5      ' Clasificación grupos:
6      'Esta macro contempla dos posibles divisiones de grupo:
7      'Según edad y sexo, o según nivel de riesgo y sexo.
8      'No obstante, la macro permite añadir otros requisitos.
9      Dim k As Integer ' Número de grupos
10     Dim clasificacion As Integer
11
12     If Credibilidad_Inicio.edadsexoboton.Value = True Then
13         clasificacion = 1 'Clasificación por sexo y edad.
14         k = 22 'Número de grupos.
15     ElseIf Credibilidad_Inicio.riesgosexoboton.Value = True Then
16         clasificacion = 2 'Clasificación por nivel de riesgo y sexo.
17         k = 8 'Número de grupos.
18     End If
19
20     k_text.Value = k 'Se representa en el formulario el número de grupos de riesgo
21     estudiados.
22
23     Dim i As Long 'contador de bucles.
24     Dim t As Integer ' Número de años
25     t = 5
26     t_text.Value = t 'Se representa en el formulario el número de años estudiados.
27     Dim n As Long
28     n = Hoja1.Range(Hoja1.Cells(3, 1), Hoja1.Cells(3, 1).End(xlDown)).Rows.Count
29     'Número de filas introducidas
30
31     '(equivalente número
32     de asegurados)
33
34     'Se captura la base de datos introducida en la Hoja1 de Excel.
35     'Requisitos: 3 variables por año.
36     'Variable 1: edad. ó 'Variable 1: nivel de riesgo.
37     'Variable 2: sexo.
38     'Variable 3: siniestro.
39     Dim bbdd() As Integer
40     ReDim bbdd(1 To n, 1 To t * 3)
41     For j = 1 To t * 3
42         Dim rango_seleccion As Range
43         Dim vector_seleccion() As Integer
44         Set rango_seleccion = Hoja1.Range(Hoja1.Cells(3, j),
45         Hoja1.Cells(3, j).End(xlDown))
46         Conversor.RangeToArray_integer vector_seleccion(),
47         rango_seleccion 'Se transforma en un Vector.
48         For i = 1 To n
49             bbdd(i, j) = vector_seleccion(i)
50         Next i
51     Next j
52
53     ' Se definen las variables output que se van a obtener de cada subrutina.
54     Dim expuestos_año_grupo() As Long 'Vector que representa el número de expuestos
55     por año y grupo.
56     Dim siniestros_año_grupo() As Long 'Vector que representa el número de
57     siniestros por año y grupo.
58
59     'Subrutina que, introduciendo el vector bbdd, clasificación, el número de grupos
60     y años, proporciona:
61     ' Vector que contiene el número de expuestos por año y grupo.
62     ' Vector que contiene el número de siniestros por año y grupo.
63     vectores_grupos bbdd(), expuestos_año_grupo(), siniestros_año_grupo(),
64     clasificacion, k, t
65
66     ' Se definen las variables output que se van a obtener de la subrutina
67     Modelo_buhlmann.
68     Dim estimador_media As Double
69     Dim estimador_varianza As Double
70     Dim estimador_factor_heterogeneidad As Double
71     Dim factor_credibilidad As Double
72     Dim tarifas_buhlmann() As Double

```

```

64
65 'Subrutina que, introduciendo:
66     'bbdd()
67     'Número de grupos de riesgo.
68     'Número de años de estudio.
69 'Proporciona los siguientes outputs:
70     'Estimador de la media por el Modelo buhlmann.
71     'Estimador de la varianza por el Modelo buhlmann.
72     'Estimador del factor de heterogeneidad por el Modelo buhlmann.
73     'Factor de credibilidad por el Modelo buhlmann.
74     'Tarifas Modelo buhlmann.
75 Modelo_buhlmann bbdd(), clasificacion, k, t, estimador_media,
estimador_varianza, _
76 estimador_factor_heterogeneidad, factor_credibilidad, tarifas_buhlmann()
77
78 'Representación de los resultados en el Formulario:
79     fheterog_text.Value = estimador_factor_heterogeneidad
80     media_text.Value = estimador_media
81     varianza_text.Value = estimador_varianza
82     fcred_text.Value = factor_credibilidad
83
84 'Representación del Vector que muestra el número de expuestos por año y grupo.:
85 Dim expuestos_año_grupo_respresentar() As String 'Se crea un vector con formato
texto
86
87                                     'para poder visualizar qué indica
88                                     cada variable.
89 ReDim expuestos_año_grupo_respresentar(1 To k + 1, 1 To t + 1)
90
91     For i = 1 To k
92         For j = 1 To t
93             expuestos_año_grupo_respresentar(i + 1, j + 1) =
expuestos_año_grupo(i, j)
94             expuestos_año_grupo_respresentar(1, j + 1) = "AÑO " & j
95         Next j
96         expuestos_año_grupo_respresentar(i + 1, 1) = "GRUPO " & i
97     Next i
98
99 expuestos_año_grupo_list.List = expuestos_año_grupo_respresentar()
100 'Representación en cuadro de lista.
101
102
103
104 'Representación del Vector que muestra el número de siniestros por año y grupo:
105 Dim siniestros_año_grupo_respresentar() As String 'Se crea un vector con formato
texto
106
107                                     'para poder visualizar qué indica
108                                     cada variable.
109 ReDim siniestros_año_grupo_respresentar(1 To k + 1, 1 To t + 1)
110
111     For i = 1 To k
112         For j = 1 To t
113             siniestros_año_grupo_respresentar(i + 1, j + 1) =
siniestros_año_grupo(i, j)
114             siniestros_año_grupo_respresentar(1, j + 1) = "AÑO " & j
115         Next j
116         siniestros_año_grupo_respresentar(i + 1, 1) = "GRUPO " & i
117     Next i
118
119 siniestros_año_grupo_list.List = siniestros_año_grupo_respresentar()
120 'Representación en cuadro de lista.
121
122
123
124 'Representación del Vector que muestra las tarifas de cada grupo:
125 Dim tarifas_buhlmann_representar() As String 'Se crea un vector con formato texto
126
127                                     'para poder visualizar qué indica
128                                     cada variable.
129 ReDim tarifas_buhlmann_representar(1 To k + 1, 1 To 2)
130
131     For i = 1 To k
132         tarifas_buhlmann_representar(i + 1, 2) = tarifas_buhlmann(i)
133         tarifas_buhlmann_representar(1, 2) = "TARIFAS"
134         tarifas_buhlmann_representar(i + 1, 1) = "GRUPO " & i
135     Next i
136
137 tarifas_list.List = tarifas_buhlmann_representar() 'Representación en cuadro de

```

lista.

127

128 End Sub

129

130

131


```

1  ' Formulario de Bühlmann-Straub
2
3  Public Sub UserForm_initialize()
4
5      ' Clasificación grupos:
6      'Esta macro contempla dos posibles divisiones de grupo:
7      'Según edad y sexo, o según nivel de riesgo y sexo.
8      'No obstante, la macro permite añadir otros requisitos.
9      Dim k As Integer ' Número de grupos
10     Dim clasificacion As Integer
11
12
13     If Credibilidad_Inicio.edadsexoboton.Value = True Then
14         clasificacion = 1 'Clasificación por sexo y edad.
15         k = 22 'Número de grupos.
16     ElseIf Credibilidad_Inicio.riesgosexoboton.Value = True Then
17         clasificacion = 2 'Clasificación por nivel de riesgo y sexo.
18         k = 8 'Número de grupos.
19     End If
20
21     k_text.Value = k 'Se representa en el formulario el número de grupos de riesgo
22     estudiados.
23
24     Dim i As Long 'contador de bucles.
25     Dim t As Integer ' Número de años
26     t = 5
27     t_text.Value = t 'Se representa en el formulario el número de años estudiados.
28     Dim n As Long
29     n = Hoja1.Range(Hoja1.Cells(3, 1), Hoja1.Cells(3, 1).End(xlDown)).Rows.Count
30     'Número de filas introducidas
31     '(equivalente número
32     de asegurados)
33
34     'Se captura la base de datos introducida en la Hoja1 de Excel.
35     'Requisitos: 3 variables por año.
36     'Variable 1: edad. ó 'Variable 1: nivel de riesgo.
37     'Variable 2: sexo.
38     'Variable 3: siniestro.
39     Dim bbdd() As Integer
40     ReDim bbdd(1 To n, 1 To t * 3)
41     For j = 1 To t * 3
42         Dim rango_seleccion As Range
43         Dim vector_seleccion() As Integer
44         Set rango_seleccion = Hoja1.Range(Hoja1.Cells(3, j),
45         Hoja1.Cells(3, j).End(xlDown))
46         Conversor.RangeToArray_integer_vector_seleccion(),
47         rango_seleccion 'Se transforma en un Vector.
48         For i = 1 To n
49             bbdd(i, j) = vector_seleccion(i)
50         Next i
51     Next j
52
53     ' Se definen las variables output que se van a obtener de cada subrutina.
54     Dim expuestos_año_grupo() As Long 'Vector que representa el número de expuestos
55     por año y grupo.
56     Dim siniestros_año_grupo() As Long 'Vector que representa el número de
57     siniestros por año y grupo.
58
59     'Subrutina que, introduciendo el vector bbdd, clasificación, el número de grupos
60     y años, proporciona:
61     ' Vector que contiene el número de expuestos por año y grupo.
62     ' Vector que contiene el número de siniestros por año y grupo.
63     vectores_grupos bbdd(), expuestos_año_grupo(), siniestros_año_grupo(),
64     clasificacion, k, t
65
66     ' Se definen las variables output que se van a obtener de la subrutina
67     Modelo_buhlmann_Straub.
68     Dim estimador_media_Straub As Double
69     Dim estimador_varianza_Straub As Double
70     Dim estimador_factor_heterogeneidad_Straub As Double
71     Dim factor_credibilidad_Straub As Double

```

```

64     Dim factor_credibilidad_grupo_Straub() As Double
65     Dim tarifas_buhlmann_Straub() As Double
66
67     'Subrutina que, introduciendo:
68     'bbdd()
69     'Número de grupos de riesgo.
70     'Número de años de estudio.
71     'Clasificación.
72     'Proporciona los siguientes outputs:
73     'Estimador de la media por el Modelo buhlmann Straub.
74     'Estimador de la varianza por el Modelo buhlmann Straub.
75     'Estimador del factor de heterogeneidad por el Modelo buhlmann Straub.
76     'Factor de credibilidad por el Modelo buhlmann Straub.
77     'Vector del Factor de credibilidad por el Modelo buhlmann Straub por grupo
de riesgo.
78     'Tarifas buhlmann Straub.
79     Modelo_buhlmann_Straub bbdd(), clasificacion, k, t, estimador_media_Straub,
estimador_varianza_Straub,
80     estimador_factor_heterogeneidad_Straub, factor_credibilidad_grupo_Straub(),
81     factor_credibilidad_Straub, tarifas_buhlmann_Straub()
82
83     'Representación de los resultados en el Formulario:
84     fheterog_text.Value = estimador_factor_heterogeneidad_Straub
85     media_text.Value = estimador_media_Straub
86     varianza_text.Value = estimador_varianza_Straub
87     fcred_text.Value = factor_credibilidad_Straub
88
89     'Representación del Vector que muestra el número de expuestos por año y grupo.:
90     Dim expuestos_año_grupo_respresentar() As String 'Se crea un vector con formato
texto
91
92                                     'para poder visualizar qué indica
93                                     cada variable.
94     ReDim expuestos_año_grupo_respresentar(1 To k + 1, 1 To t + 1)
95
96     For i = 1 To k
97         For j = 1 To t
98             expuestos_año_grupo_respresentar(i + 1, j + 1) =
expuestos_año_grupo(i, j)
99             expuestos_año_grupo_respresentar(1, j + 1) = "AÑO " & j
100         Next j
101         expuestos_año_grupo_respresentar(i + 1, 1) = "GRUPO " & i
102     Next i
103
104     expuestos_año_grupo_list.List = expuestos_año_grupo_respresentar()
105     'Representación en cuadro de lista.
106
107     'Representación del Vector que muestra el número de siniestros por año y grupo:
108     Dim siniestros_año_grupo_respresentar() As String 'Se crea un vector con formato
texto
109
110                                     'para poder visualizar qué indica
111                                     cada variable.
112     ReDim siniestros_año_grupo_respresentar(1 To k + 1, 1 To t + 1)
113
114     For i = 1 To k
115         For j = 1 To t
116             siniestros_año_grupo_respresentar(i + 1, j + 1) =
siniestros_año_grupo(i, j)
117             siniestros_año_grupo_respresentar(1, j + 1) = "AÑO " & j
118         Next j
119         siniestros_año_grupo_respresentar(i + 1, 1) = "GRUPO " & i
120     Next i
121
122     siniestros_año_grupo_list.List = siniestros_año_grupo_respresentar()
123     'Representación en cuadro de lista.
124
125     'Representación del Vector que muestra el factor de credibilidad de cada grupo:
126     Dim factor_credibilidad_grupo_Straub_representar() As String 'Se crea un vector
con formato texto
127
128                                     'para poder visualizar qué indica
129                                     cada variable.

```

```

125 ReDim factor_credibilidad_grupo_Straub_representar(1 To k + 1, 1 To 2)
126
127 For i = 1 To k
128     factor_credibilidad_grupo_Straub_representar(i + 1, 2) =
129         factor_credibilidad_grupo_Straub(i)
130     factor_credibilidad_grupo_Straub_representar(1, 2) = "FACTOR DE
131     CREDIBILIDAD"
132     factor_credibilidad_grupo_Straub_representar(i + 1, 1) = "GRUPO " & i
133 Next i
134 factor_credibilidad_grupo_list.List =
135     factor_credibilidad_grupo_Straub_representar() 'Representación en cuadro de lista.
136
137 'Representación del Vector que muestra las tarifas de cada grupo:
138 Dim tarifas_buhlmann_Straub_representar() As String 'Se crea un vector con
139     formato texto
140     'para poder visualizar qué indica
141     cada variable.
142 ReDim tarifas_buhlmann_Straub_representar(1 To k + 1, 1 To 2)
143
144 For i = 1 To k
145     tarifas_buhlmann_Straub_representar(i + 1, 2) = tarifas_buhlmann_Straub(i)
146     tarifas_buhlmann_Straub_representar(1, 2) = "TARIFAS"
147     tarifas_buhlmann_Straub_representar(i + 1, 1) = "GRUPO " & i
148 Next i
149 tarifas_list.List = tarifas_buhlmann_Straub_representar() 'Representación en
150     cuadro de lista.
151 End Sub

```

```

1  ' Formulario Tarificación Bühlmann:
2
3  Private Sub cb_boton_Click()
4  ' Clasificación grupos:
5      'Esta macro contempla dos posibles divisiones de grupo:
6          'Según edad y sexo, o según nivel de riesgo y sexo.
7          'No obstante, la macro permite añadir otros requisitos.
8      Dim k As Integer ' Número de grupos
9      Dim clasificacion As Integer
10
11     If Credibilidad_Inicio.edadsexoboton.Value = True Then
12         clasificacion = 1
13         k = 22
14     ElseIf Credibilidad_Inicio.riesgosexoboton.Value = True Then
15         clasificacion = 2
16         k = 8
17     End If
18
19
20     Dim i As Long 'contador de bucles.
21     Dim t As Integer ' Número de años
22     t = 5
23     Dim n As Long
24     n = Hoja1.Range(Hoja1.Cells(3, 1), Hoja1.Cells(3, 1).End(xlDown)).Rows.Count
25     'Número de filas introducidas
26
27                                     '(equivalente número
28                                     de asegurados)
29
30     'Se captura la base de datos que sirve para la ejecución del Modelo.
31     'Requisitos: 3 variables por año.
32         'Variable 1: edad. ó 'Variable 1: nivel de riesgo.
33         'Variable 2: sexo.
34         'Variable 3: siniestro.
35     Dim bbdd() As Integer
36     ReDim bbdd(1 To n, 1 To t * 3)
37     For j = 1 To t * 3
38         Dim rango_seleccion As Range
39         Dim vector_seleccion() As Integer
40         Set rango_seleccion = Hoja1.Range(Hoja1.Cells(3, j),
41         Hoja1.Cells(3, j).End(xlDown))
42         Conversor.RangetoArray_integer vector_seleccion(),
43         rango_seleccion 'Se transforma en un Vector.
44         For i = 1 To n
45             bbdd(i, j) = vector_seleccion(i)
46         Next i
47     Next j
48
49     ' Se definen las variables output que se van a obtener de la subrutina
50     Modelo_buhlmann.
51     Dim estimador_media As Double
52     Dim estimador_varianza As Double
53     Dim estimador_factor_heterogeneidad As Double
54     Dim factor_credibilidad As Double
55     Dim tarifas_buhlmann() As Double
56     Dim expuestos_año_grupo() As Long 'Vector que representa el número de
57     expuestos por año y grupo.
58     Dim siniestros_año_grupo() As Long 'Vector que representa el número de
59     siniestros por año y grupo.
60
61     'Subrutina que, introduciendo:
62     'bbdd()
63     'Número de grupos de riesgo.
64     'Número de años de estudio.
65     'Proporciona los siguientes outputs:
66     'Estimador de la media por el Modelo buhlmann.
67     'Estimador de la varianza por el Modelo buhlmann.
68     'Estimador del factor de heterogeneidad por el Modelo buhlmann.
69     'Factor de credibilidad por el Modelo buhlmann.
70     'Tarifas Modelo buhlmann.
71     Modelo_buhlmann bbdd(), clasificacion, k, t, estimador_media,
72     estimador_varianza, _
73     estimador_factor_heterogeneidad, factor_credibilidad, tarifas_buhlmann()

```

```

66 'Se captura la base de datos del Colectivo a tarificar.
67 'Requisitos: 3 variables por año.
68 'Variable 1: edad. ó 'Variable 1: nivel de riesgo.
69 'Variable 2: sexo.
70 'Variable 3: capital asegurado.
71 Dim bbdd_tarificar() As Integer
72 Dim n2 As Integer
73 n2 = Hoja5.Range(Hoja5.Cells(3, 1), Hoja5.Cells(3,
74 1).End(xlDown)).Rows.Count 'Número de filas introducidas
                                                                    '(equivalente número
                                                                    se asegurados)
75 ReDim bbdd_tarificar(1 To n2, 1 To 3)
76 For j = 1 To 3
77 Dim rango_seleccion2 As Range
78 Dim vector_seleccion2() As Integer
79 Set rango_seleccion2 = Hoja5.Range(Hoja5.Cells(3, j), Hoja5.Cells(3,
80 j).End(xlDown))
81 Conversor.RangetoArray_integer vector_seleccion2(), rango_seleccion2
82 'Se transforma en un Vector.
83 For i = 1 To UBound(vector_seleccion2())
84 bbdd_tarificar(i, j) = vector_seleccion2(i)
85 Next i
86 Next j
87 'Se ejecuta la subrutina que asigna el número de grupo a cada asegurado.
88 Dim grupo_colectividad() As Integer
89 Dim vector_novalido() As Integer ' Vector que no tiene aplicación para este
90 caso.
91 asignacion_grupo bbdd_tarificar(), grupo_colectividad(),
92 vector_novalido(), clasificacion, k, 1
93 'Una vez asignado el grupo al que pertenece cada asegurado, se lleva a cabo la
94 tarificación del Seguro TAR.
95 'Se definen las variables:
96 Dim tasa() As Double 'Tasas obtenidas con el modelo.
97 ReDim tasa(1 To UBound(bbdd_tarificar()))
98 Dim prima_pura() As Double 'Prima pura.
99 ReDim prima_pura(1 To UBound(bbdd_tarificar()))
100 Dim prima_pura_recargada() As Double ' Prima pura con recargo.
101 ReDim prima_pura_recargada(1 To UBound(bbdd_tarificar()))
102 Dim prima_pura_total As Double 'Prima pura total.
103 Dim prima_neta() As Double 'Prima neta.
104 ReDim prima_neta(1 To UBound(bbdd_tarificar()))
105 Dim prima_neta_total As Double 'Prima neta total.
106
107 'Variables introducidas por el usuario:
108 Dim ggi As Double 'Gastos de gestión interna.
109 ggi = ggi_text.Value / 100
110 Dim gge As Double 'Gastos de gestión externa.
111 gge = gge_text.Value / 100
112 Dim it As Double 'Tipo de interés técnico.
113 it = it_text.Value / 100
114 Dim recargo As Double 'Recargo.
115 recargo = recargo_text.Value / 100
116
117 For i = 1 To UBound(bbdd_tarificar()) ' Se definen los valores de las
118 variables.
119 tasa(i) = tarifas_buhlmann(grupo_colectividad(i, 1))
120 prima_pura(i) = tasa(i) * bbdd_tarificar(i, 3) / ((1 + it) ^ (1 / 2))
121 prima_pura_recargada(i) = prima_pura(i) * (1 + recargo)
122 prima_pura_total = prima_pura_total + prima_pura_recargada(i)
123 prima_neta(i) = prima_pura_recargada(i) / (1 - ggi - gge)
124 prima_neta_total = prima_neta_total + prima_neta(i)
125 Next i
126
127 'Representación del Vector que muestra las tres variables del asegurado, la
128 prima pura y la prima neta:
129 Dim tarificacion_representar() As String 'Se crea un vector con formato texto
130 'para poder visualizar qué indica
131 cada variable.
132 ReDim tarificacion_representar(1 To UBound(bbdd_tarificar()) + 1, 1 To 5)
133
134 For i = 1 To UBound(grupo_colectividad()) 'Se representa el colectivo y las
135 primas.
136 tarificacion_representar(i + 1, 1) = bbdd_tarificar(i, 1)

```

```
128         tarificacion_representar(i + 1, 2) = bbdd_tarificar(i, 2)
129         tarificacion_representar(i + 1, 3) = bbdd_tarificar(i, 3)
130         tarificacion_representar(i + 1, 4) = prima_pura_recargada(i)
131         tarificacion_representar(i + 1, 5) = prima_neta(i)
132     Next i
133     tarificacion_representar(1, 1) = "VARIABLE CLASIFICATORIA"
134     tarificacion_representar(1, 2) = "SEXO"
135     tarificacion_representar(1, 3) = "CAPITAL ASEGURADO"
136     tarificacion_representar(1, 4) = "PRIMA PURA"
137     tarificacion_representar(1, 5) = "PRIMA NETA"
138     tarificacion_representar_list.List = tarificacion_representar() 'Representación
en cuadro de lista.
139     prima_pura_total_text.Value = prima_pura_total ' Se muestra la prima pura total.
140     prima_neta_total_text.Value = prima_neta_total ' Se muestra la prima neta total.
141
142 End Sub
143
144
```

```

1  ' Formulario Tarificación Bühlmann-Straub:
2
3  Private Sub cbs_boton_Click()
4  ' Clasificación grupos:
5      'Esta macro contempla dos posibles divisiones de grupo:
6          'Según edad y sexo, o según nivel de riesgo y sexo.
7          'No obstante, la macro permite añadir otros requisitos.
8  Dim k As Integer ' Número de grupos
9  Dim clasificacion As Integer
10
11     If Credibilidad_Inicio.edadsexoboton.Value = True Then
12         clasificacion = 1
13         k = 22
14     ElseIf Credibilidad_Inicio.riesgosexoboton.Value = True Then
15         clasificacion = 2
16         k = 8
17     End If
18
19
20 Dim i As Long 'contador de bucles.
21 Dim t As Integer ' Número de años
22     t = 5
23 Dim n As Long
24     n = Hoja1.Range(Hoja1.Cells(3, 1), Hoja1.Cells(3, 1).End(xlDown)).Rows.Count
25     'Número de filas introducidas
26
27                                     '(equivalente número
28                                     de asegurados)
29
30 'Se captura la base de datos que sirve para la ejecución del Modelo.
31 'Requisitos: 3 variables por año.
32     'Variable 1: edad. ó 'Variable 1: nivel de riesgo.
33     'Variable 2: sexo.
34     'Variable 3: siniestro.
35 Dim bbdd() As Integer
36 ReDim bbdd(1 To n, 1 To t * 3)
37 For j = 1 To t * 3
38     Dim rango_seleccion As Range
39     Dim vector_seleccion() As Integer
40     Set rango_seleccion = Hoja1.Range(Hoja1.Cells(3, j),
41     Hoja1.Cells(3, j).End(xlDown))
42     Conversor.RangetoArray_integer vector_seleccion(),
43     rango_seleccion 'Se transforma en un Vector.
44     For i = 1 To n
45         bbdd(i, j) = vector_seleccion(i)
46     Next i
47 Next j
48
49 ' Se definen las variables output que se van a obtener de la subrutina
50 Modelo_buhlmann_Straub.
51 Dim estimador_media_Straub As Double
52 Dim estimador_varianza_Straub As Double
53 Dim estimador_factor_heterogeneidad_Straub As Double
54 Dim factor_credibilidad_Straub As Double
55 Dim factor_credibilidad_grupo_Straub() As Double
56 Dim tarifas_buhlmann_Straub() As Double
57
58 'Subrutina que, introduciendo:
59     'bbdd()
60     'Número de grupos de riesgo.
61     'Número de años de estudio.
62     'Clasificación.
63 'Proporciona los siguientes outputs:
64     'Estimador de la media por el Modelo buhlmann Straub.
65     'Estimador de la varianza por el Modelo buhlmann Straub.
66     'Estimador del factor de heterogeneidad por el Modelo buhlmann Straub.
67     'Factor de credibilidad por el Modelo buhlmann Straub.
68     'Vector del Factor de credibilidad por el Modelo buhlmann Straub por grupo
69     de riesgo.
70     'Tarifas buhlmann Straub.
71 Modelo_buhlmann_Straub bbdd(), clasificacion, k, t, estimador_media_Straub,
72 estimador_varianza_Straub, _
73 estimador_factor_heterogeneidad_Straub, factor_credibilidad_grupo_Straub(), _
74 factor_credibilidad_Straub, tarifas_buhlmann_Straub()

```

```

67
68 'Se captura la base de datos del Colectivo a tarificar.
69 'Requisitos: 3 variables por año.
70 'Variable 1: edad. ó 'Variable 1: nivel de riesgo.
71 'Variable 2: sexo.
72 'Variable 3: capital asegurado.
73 Dim bbdd_tarificar() As Integer
74 Dim n2 As Integer
75 n2 = Hoja5.Range(Hoja5.Cells(3, 1), Hoja5.Cells(3,
1).End(xlDown)).Rows.Count 'Número de filas introducidas
76
' (equivalente número
se asegurados)
77 ReDim bbdd_tarificar(1 To n2, 1 To 3)
78 For j = 1 To 3
79 Dim rango_seleccion2 As Range
80 Dim vector_seleccion2() As Integer
81 Set rango_seleccion2 = Hoja5.Range(Hoja5.Cells(3, j), Hoja5.Cells(3,
j).End(xlDown))
82 Conversor.RangetoArray_integer vector_seleccion2(), rango_seleccion2
83 'Se transforma en un Vector.
84 For i = 1 To UBound(vector_seleccion2())
85 bbdd_tarificar(i, j) = vector_seleccion2(i)
86 Next i
87 Next j
88 'Se ejecuta la subrutina que asigna el número de grupo a cada asegurado.
89 Dim grupo_colectividad() As Integer
90 Dim vector_novalido() As Integer ' Vector que no tiene aplicación para este
caso.
91 asignacion_grupo bbdd_tarificar(), grupo_colectividad(),
vector_novalido(), clasificacion, k, 1
92 'Una vez asignado el grupo al que pertenece cada asegurado, se lleva a cabo la
tarificación del Seguro TAR.
93 'Se definen las variables:
94 Dim tasa() As Double ' Tasa obtenida mediante el modelo.
95 ReDim tasa(1 To UBound(bbdd_tarificar()))
96 Dim prima_pura() As Double 'Prima pura.
97 ReDim prima_pura(1 To UBound(bbdd_tarificar()))
98 Dim prima_pura_recargada() As Double 'Prima pura recargada.
99 ReDim prima_pura_recargada(1 To UBound(bbdd_tarificar()))
100 Dim prima_pura_total As Double 'Prima pura Total.
101 Dim prima_neta() As Double ' Prima neta Total.
102 ReDim prima_neta(1 To UBound(bbdd_tarificar())) 'Prima neta.
103 Dim prima_neta_total As Double
104
'Variables introducidas por el usuario:
105 Dim ggi As Double 'Gastos de gestión interna.
106 ggi = ggi_text.Value / 100
107 Dim gge As Double 'Gastos de gestión externa.
108 gge = gge_text.Value / 100
109 Dim it As Double 'Tipo de interés técnico.
110 it = it_text.Value / 100
111 Dim recargo As Double ' Recargo.
112 recargo = recargo_text.Value / 100
113
114 For i = 1 To UBound(bbdd_tarificar()) ' Se define el valor de estas variables.
115 tasa(i) = tarifas_buhlmann_Straub(grupo_colectividad(i, 1))
116 prima_pura(i) = tasa(i) * bbdd_tarificar(i, 3) / ((1 + it) ^ (1 / 2))
117 prima_pura_recargada(i) = prima_pura(i) * (1 + recargo)
118 prima_pura_total = prima_pura_total + prima_pura_recargada(i)
119 prima_neta(i) = prima_pura_recargada(i) / (1 - ggi - gge)
120 prima_neta_total = prima_neta_total + prima_neta(i)
121 Next i
122
123 'Representación del Vector que muestra las tres variables del asegurado, la
prima pura y la prima neta:
124 Dim tarificacion_representar() As String 'Se crea un vector con formato texto
125 'para poder visualizar qué indica
cada variable.
126 ReDim tarificacion_representar(1 To UBound(bbdd_tarificar()) + 1, 1 To 5)
127
128 For i = 1 To UBound(grupo_colectividad()) ' Se representa el colectivo y las
primas.
129 tarificacion_representar(i + 1, 1) = bbdd_tarificar(i, 1)

```



```
130         tarificacion_representar(i + 1, 2) = bbdd_tarificar(i, 2)
131         tarificacion_representar(i + 1, 3) = bbdd_tarificar(i, 3)
132         tarificacion_representar(i + 1, 4) = prima_pura_recargada(i)
133         tarificacion_representar(i + 1, 5) = prima_neta(i)
134     Next i
135     tarificacion_representar(1, 1) = "VARIABLE CLASIFICATORIA"
136     tarificacion_representar(1, 2) = "SEXO"
137     tarificacion_representar(1, 3) = "CAPITAL ASEGURADO"
138     tarificacion_representar(1, 4) = "PRIMA PURA"
139     tarificacion_representar(1, 5) = "PRIMA NETA"
140     tarificacion_representar_list.List = tarificacion_representar() 'Representación
141     en cuadro de lista.
142     prima_pura_total_text.Value = prima_pura_total
143     prima_neta_total_text.Value = prima_neta_total
144 End Sub
```

Anexo 5: Prueba de aciertos.
Modelos de Credibilidad,
Bühlmann y Bühlmann-Straub.

Programación en Matlab.

```

1  %PRUEBA ACIERTOS MODELOS DE CREDIBILIDAD: BÜHLMANN Y BÜHLMANN-STRAUB - NIVEL DE
  RIESGO Y SEXO.
2
3  %Se convierte la tabla importada en un vector:
4  data=table2array(APLICACION);
5  %Vector que indica el número de filas y columnas de la Base de datos:
6  dimfilas=size(data);
7  num_bucles=1000
8  %Se proporciona dimension al vector de siniestros estimados.
9  siniestros_estimados=zeros(num_bucles,9);
10
11 %se indica si se ejecutará Bühlmann o Bühlmann-Straub.
12 buhlmann=1;
13 buhlmannstraub=1-buhlmann;
14
15 %Se definen las tarifas obtenidas en del modelo.
16 if buhlmann==1
17     tarifa1=0;
18     tarifa2=0;
19     tarifa3=0;
20     tarifa4=0;
21     tarifa5=0;
22     tarifa6=0;
23     tarifa7=0;
24     tarifa8=0;
25 end
26 %Se definen las tarifas obtenidas en del modelo.
27 if buhlmannstraub==1
28     tarifa1=0;
29     tarifa2=0;
30     tarifa3=0;
31     tarifa4=0;
32     tarifa5=0;
33     tarifa6=0;
34     tarifa7=0;
35     tarifa8=0;
36 end
37
38 for x=1:9
39     proporcion=x/10 %proporciones que se van seleccionar de la muestra.
40                     %10%,20%,30%,40%,50%,60%,70%,80%,90%
41     %A continuación se indica el número de observaciones a seleccionar.
42     proporcion_aplicacion=floor(fix(dimfilas(1,1)*(proporcion)));
43     z=1;
44     for z=1:1000 %se ejecutan los 1.000 bucles
45         j=1;
46         %Se proporciona dimension a los vectores:
47         vector_aplicacion=zeros(proporcion_aplicacion,dimfilas(1,2));
48         prob_estimada=zeros(proporcion_aplicacion,1);
49         siniestros=zeros(proporcion_aplicacion,1);
50         %transpose: Devuelve la traspuesta de una matriz.
51         %sort: Ordena los valores en orden ascendente.
52         %randperm(n,k): devuelve un vector de fila que contiene
53         %k enteros únicos seleccionados aleatoriamente de 1 a n inclusive.
54         r=transpose(sort(randperm(dimfilas(1,1),proporcion_aplicacion)));
55         for i=1:dimfilas(1,1) %El bucle pasa por todas las observaciones.
56             if i==r(j,1) %Este bucle pasa unicamente por las observaciones
57                 %que se han escogido de manera aleatoria
58                 vector_aplicacion(j,1)=data(i,1);%siniestro
59                 siniestros(j,1)=data(i,1);%siniestro
60                 vector_aplicacion(j,2)=data(i,2);%edad
61                 vector_aplicacion(j,3)=data(i,3);%sexo
62                 vector_aplicacion(j,4)=data(i,4);%gran_empresa
63                 vector_aplicacion(j,5)=data(i,5);%nivel_riesgo
64                 if data(i,3)==1 %Si es varón.
65                     if data(i,5)==1 %Si es nivel de riesgo 1
66                         prob_estimada(j,1)=tarifa1;
67                     elseif data(i,5)==2 %Si es nivel de riesgo 2
68                         prob_estimada(j,1)=tarifa2;
69                     elseif data(i,5)==3 %Si es nivel de riesgo 3
70                         prob_estimada(j,1)=tarifa3;
71                     elseif data(i,5)==4 %Si es nivel de riesgo 4
72                         prob_estimada(j,1)=tarifa4;

```

```

73         end
74     else %Si es mujer.
75         if data(i,5)==1 %Si es nivel de riesgo 1
76             prob_estimada(j,1)=tarifa5;
77         elseif data(i,5)==2 %Si es nivel de riesgo 2
78             prob_estimada(j,1)=tarifa6;
79         elseif data(i,5)==3 %Si es nivel de riesgo 3
80             prob_estimada(j,1)=tarifa7;
81         elseif data(i,5)==4 %Si es nivel de riesgo 4
82             prob_estimada(j,1)=tarifa8;
83         end
84     end
85
86     if j==proporcion_aplicacion
87         break;
88     end
89     j=j+1;
90 end
91 end
92 %vector siniestros estimados:
93 siniestros_estimados(z,x)=sum(prob_estimada);
94 %vector siniestros observados:
95 siniestros_observados(z,x)=sum(siniestros);
96 %vector proporcion siniestros estimados sobre los observados.
97
98     proporcion_estimador_observados(z,x)=siniestros_estimados(z,x)/siniestros_obs
99     rvados(z,x);
100 end
end

```

```

1  %PRUEBA ACIERTOS MODELOS DE CREDIBILIDAD: BÜHLMANN Y BÜHLMANN-STRAUB - EDAD Y SEXO.
2
3  %Se convierte la tabla importada en un vector:
4  data=table2array(APLICACION);
5  %Vector que indica el número de filas y columnas de la Base de datos:
6  dimfilas=size(data);
7  num_bucles=1000
8  %Se proporciona dimension al vector de siniestros estimados.
9  siniestros_estimados=zeros(num_bucles,9);
10
11  %se indica si se ejecutará Bühlmann o Bühlmann-Straub.
12  buhlmann=0;
13  buhlmannstraub=1-buhlmann;
14
15  %Se definen las tarifas obtenidas en del modelo.
16  if buhlmann==1
17      tarifa1=0;
18      tarifa2=0;
19      tarifa3=0;
20      tarifa4=0;
21      tarifa5=0;
22      tarifa6=0;
23      tarifa7=0;
24      tarifa8=0;
25      tarifa9=0;
26      tarifa10=0;
27      tarifa11=0;
28      tarifa12=0;
29      tarifa13=0;
30      tarifa14=0;
31      tarifa15=0;
32      tarifa16=0;
33      tarifa17=0;
34      tarifa18=0;
35      tarifa19=0;
36      tarifa20=0;
37      tarifa21=0;
38      tarifa22=0;
39  end
40  %Se definen las tarifas obtenidas en del modelo.
41  if buhlmannstraub==1
42      tarifa1=0;
43      tarifa2=0;
44      tarifa3=0;
45      tarifa4=0;
46      tarifa5=0;
47      tarifa6=0;
48      tarifa7=0;
49      tarifa8=0;
50      tarifa9=0;
51      tarifa10=0;
52      tarifa11=0;
53      tarifa12=0;
54      tarifa13=0;
55      tarifa14=0;
56      tarifa15=0;
57      tarifa16=0;
58      tarifa17=0;
59      tarifa18=0;
60      tarifa19=0;
61      tarifa20=0;
62      tarifa21=0;
63      tarifa22=0;
64  end
65
66  for x=1:9
67      proporcion=x/10 %proporciones que se van seleccionar de la muestra.
68          %10%,20%,30%,40%,50%,60%,70%,80%,90%
69      %A continuación se indica el número de observaciones a seleccionar.
70      proporcion_aplicacion=floor(fix(dimfilas(1,1)*(proporcion)));
71      z=1;
72      for z=1:1000 %se ejecutan los 1.000 bucles
73          j=1;

```

```

74 %Se proporciona dimension a los vectores:
75 vector_aplicacion=zeros(proporcion_aplicacion,dimfilas(1,2));
76 prob_estimada=zeros(proporcion_aplicacion,1);
77 siniestros=zeros(proporcion_aplicacion,1);
78 %transpose: Devuelve la traspuesta de una matriz.
79 %sort: Ordena los valores en orden ascendente.
80 %randperm(n,k): devuelve un vector de fila que contiene
81 %k enteros únicos seleccionados aleatoriamente de 1 a n inclusive.
82 r=transpose(sort(randperm(dimfilas(1,1),proporcion_aplicacion)));
83 for i=1:dimfilas(1,1) %El bucle pasa por todas las observaciones.
84     if i==r(j,1) %Este bucle pasa unicamente por las observaciones
85         %que se han escogido de manera aleatoria
86         vector_aplicacion(j,1)=data(i,1);%siniestro
87         siniestros(j,1)=data(i,1);%siniestro
88         vector_aplicacion(j,2)=data(i,2);%edad
89         vector_aplicacion(j,3)=data(i,3);%sexo
90         vector_aplicacion(j,4)=data(i,4);%gran_empresa
91         vector_aplicacion(j,5)=data(i,5);%nivel_riesgo
92         if data(i,3)==1 %Si es varón.
93             if (data(i,2) >= 16) && (data(i,2) <= 20)
94                 prob_estimada(j,1)=tarifa1;
95             elseif (data(i,2) >= 21) && (data(i,2) <= 25)
96                 prob_estimada(j,1)=tarifa2;
97             elseif (data(i,2) >= 26) && (data(i,2) <= 30)
98                 prob_estimada(j,1)=tarifa3;
99             elseif (data(i,2) >= 31) && (data(i,2) <= 35)
100                 prob_estimada(j,1)=tarifa4;
101             elseif (data(i,2) >= 36) && (data(i,2) <= 40)
102                 prob_estimada(j,1)=tarifa5;
103             elseif (data(i,2) >= 41) && (data(i,2) <= 45)
104                 prob_estimada(j,1)=tarifa6;
105             elseif (data(i,2) >= 46) && (data(i,2) <= 50)
106                 prob_estimada(j,1)=tarifa7;
107             elseif (data(i,2) >= 51) && (data(i,2) <= 55)
108                 prob_estimada(j,1)=tarifa8;
109             elseif (data(i,2) >= 56) && (data(i,2) <= 60)
110                 prob_estimada(j,1)=tarifa9;
111             elseif (data(i,2) >= 61) && (data(i,2) <= 65)
112                 prob_estimada(j,1)=tarifa10;
113             elseif (data(i,2) >= 66) && (data(i,2) <= 70)
114                 prob_estimada(j,1)=tarifa11;
115             end
116         else %Si es mujer.
117             if (data(i,2) >= 16) && (data(i,2) <= 20)
118                 prob_estimada(j,1)=tarifa12;
119             elseif (data(i,2) >= 21) && (data(i,2) <= 25)
120                 prob_estimada(j,1)=tarifa13;
121             elseif (data(i,2) >= 26) && (data(i,2) <= 30)
122                 prob_estimada(j,1)=tarifa14;
123             elseif (data(i,2) >= 31) && (data(i,2) <= 35)
124                 prob_estimada(j,1)=tarifa15;
125             elseif (data(i,2) >= 36) && (data(i,2) <= 40)
126                 prob_estimada(j,1)=tarifa16;
127             elseif (data(i,2) >= 41) && (data(i,2) <= 45)
128                 prob_estimada(j,1)=tarifa17;
129             elseif (data(i,2) >= 46) && (data(i,2) <= 50)
130                 prob_estimada(j,1)=tarifa18;
131             elseif (data(i,2) >= 51) && (data(i,2) <= 55)
132                 prob_estimada(j,1)=tarifa19;
133             elseif (data(i,2) >= 56) && (data(i,2) <= 60)
134                 prob_estimada(j,1)=tarifa20;
135             elseif (data(i,2) >= 61) && (data(i,2) <= 65)
136                 prob_estimada(j,1)=tarifa21;
137             elseif (data(i,2) >= 66) && (data(i,2) <= 70)
138                 prob_estimada(j,1)=tarifa22;
139             end
140         end
141     end
142     if j==proporcion_aplicacion
143         break;
144     end
145     j=j+1;
146 end

```

```
147     end
148     %vector siniestros estimados:
149     siniestros_estimados(z,x)=sum(prob_estimada);
150     %vector siniestros observados:
151     siniestros_observados(z,x)=sum(siniestros);
152     %vector proporcion siniestros estimados sobre los observados.
153
154     proporcion_estimador_observados(z,x)=siniestros_estimados(z,x)/siniestros_obse
155     rvados(z,x);
156     end
157     end
```

Anexo 6: Prueba de aciertos. Tabla de Mortalidad PASEM/F-2000.

Programación en Matlab.


```

1  %PRUEBA ACIERTOS TABLA DE MORTALIDAD PASEM/F-2010.
2
3  %Se convierte la tabla importada en un vector:
4  data=table2array(APLICACION);
5  %Se convierte la tabla importada en un vector:
6  qx=0.8*table2array(PASEMF2010);
7  %Vector que indica el número de filas y columnas de la Base de datos:
8  dimfilas=size(data);
9  num_bucles=1000;
10 %Se proporciona dimension al vector de siniestros estimados.
11 siniestros_estimados=zeros(num_bucles,9);
12
13 for x=1:9
14     proporcion=x/10 %proporciones que se van seleccionar de la muestra.
15                     %10%,20%,30%,40%,50%,60%,70%,80%,90%
16     %A continuación se indica el número de observaciones a seleccionar.
17     proporcion_aplicacion=floor(fix(dimfilas(1,1)*(proporcion)));
18     z=1;
19     for z=1:1000 %se ejecutan los 1.000 bucles
20         j=1;
21         %Se proporciona dimension a los vectores:
22         vector_aplicacion=zeros(proporcion_aplicacion,dimfilas(1,2));
23         prob_estimada=zeros(proporcion_aplicacion,1);
24         siniestros=zeros(proporcion_aplicacion,1);
25         %transpose: Devuelve la traspuesta de una matriz.
26         %sort: Ordena los valores en orden ascendente.
27         %randperm(n,k): devuelve un vector de fila que contiene
28         %k enteros únicos seleccionados aleatoriamente de 1 a n inclusive.
29         r=transpose(sort(randperm(dimfilas(1,1),proporcion_aplicacion)));
30         for i=1:dimfilas(1,1) %El bucle pasa por todas las observaciones.
31             if i==r(j,1) %Este bucle pasa unicamente por las observaciones
32                 %que se han escogido de manera aleatoria
33                 vector_aplicacion(j,1)=data(i,1);%siniestro
34                 siniestros(j,1)=data(i,1);%siniestro
35                 vector_aplicacion(j,2)=data(i,2);%edad
36                 vector_aplicacion(j,3)=data(i,3);%sexo
37                 vector_aplicacion(j,4)=data(i,4);%gran_empresa
38                 vector_aplicacion(j,5)=data(i,5);%nivel_riesgo
39                 prob_estimada(j,1)=qx(data(i,2)+1,3-data(i,3));
40                 if j==proporcion_aplicacion
41                     break;
42                 end
43                 j=j+1;
44             end
45         end
46         %vector siniestros estimados:
47         siniestros_estimados(z,x)=sum(prob_estimada);
48         %vector siniestros observados:
49         siniestros_observados(z,x)=sum(siniestros);
50         %vector proporcion siniestros estimados sobre los observados.
51
52         proporcion_estimador_observados(z,x)=siniestros_estimados(z,x)/siniestros_obs
53         rvados(z,x);
54     end
55 end

```