

TRABAJO FINAL DE MÁSTER

Título: SEGMENTACIÓN DE RIESGOS Y MODELIZACIÓN ACTUARIAL EN IFRS17

Autoría: Núria Hom Serra

Tutoría: Mercedes Ayuso Gutiérrez



**UNIVERSITAT DE
BARCELONA**

Facultat d'Economia
i Empresa

**Màster
de Ciències
Actuarials
i Financeres**

Facultad de Economía y Empresa
Universidad de Barcelona

Trabajo Final de Máster
Máster en Ciencias Actuariales y Financieras

SEGMENTACIÓN DE RIESGOS Y MODELIZACIÓN ACTUARIAL EN IFRS17

Autoría: NÚRIA HOM SERRA

Tutoría: MERCEDES AYUSO GUTIÉRREZ

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

Resumen:

Este trabajo trata sobre la aplicación de la normativa contable IFRS17 en un portafolio de seguros de vida, específicamente utilizando el enfoque general o de bloques. Para mayor claridad, el estudio se centra en la aplicación de esta normativa de forma detallada en un contrato de seguros de los 76.102 que hay en el portafolio. El procedimiento aplicado a este contrato es extrapolable al resto y consta de etapas bien claras: agregación de contratos de seguro, flujos de efectivo futuros, tasa de descuento, ajuste por riesgo no financiero y determinación del Margen de Servicio Contractual (CSM) o Componente de Pérdida (LC).

Palabras clave:

Clúster, riesgos, Fair Value, Future Value, Tipos de interés.

English

Title: RISK SEGMENTATION AND ACTUARIAL MODELING IN IFRS17

Summary:

This work is about the application of IFRS17 accounting normative in a life insurance portfolio, specifically using the general or block approach. For greater clarity, the study focuses on the application of this regulation in detail in one insurance contract of 76.102 in the portfolio. The procedure applied to this contract can be extrapolated to the rest and consists of very clear stages: aggregation of insurance contracts, future cash flows, discount rate, adjustment for non-financial risk and determination of the Contractual Service Margin (CSM) or Loss Component (LC).

Key words:

Clúster, risks, Fair Value, Future Value, Interest Rates.

ÍNDICE

1. INTRODUCCIÓN	13
2. LA CONTABILIDAD Y ANTECEDENTES DE LA NORMATIVA IFRS17	15
3. ESTADO DE LA CUESTIÓN	22
4. ESTUDIO ACTUARIAL DE LA NORMATIVA IFRS17	25
4.1. Base de datos y análisis descriptivo	26
4.2. Método general o por bloques (BBA)	33
4.1.1. <i>Agregación de contratos de seguros</i>	34
4.1.2. <i>Técnicas de Clúster</i>	35
4.1.1.1. <i>K means</i>	36
4.1.1.2. <i>K Medoids</i>	41
4.1.1.3. <i>Agrupación jerárquica</i>	44
4.1.3. <i>Estimación de los flujos de efectivo</i>	49
4.1.3.1. <i>Estimación de la prima</i>	49
4.1.3.2. <i>Estimación de las provisiones</i>	55
4.1.3.5. <i>Estimación de las provisiones para el 2022 bajo IFRS17 e IFRS4</i>	57
4.1.3.6. <i>Estimación de las provisiones para el 2023 bajo IFRS17 e IFRS4</i>	59
4.1.4. <i>Tasa de descuento</i>	61
4.1.5. <i>Ajuste de riesgo por riesgo no financiero</i>	65
4.1.6. <i>Margen de servicio contractual</i>	67
5. CONCLUSIONES	71
6. AGRADECIMIENTOS	74
7. ABREVIATURAS	75
8. BIBLIOGRAFÍA	77
9. ANEXOS	83
9.1. Anexos contabilidad y antecedentes de la normativa IFRS17	83
9.2. Anexos análisis descriptivo	83
9.2.1. <i>Portafolio general</i>	84
9.2.2. <i>Contratos agrupados por fecha efectiva del contrato de seguro</i>	86
9.2. Anexos agregación de contratos de seguro	95
9.2.1. <i>Anexos K-Means</i>	95
9.2.1.1. <i>Método de la Silueta</i>	95
9.2.1.2. <i>Estadísticos de las distancias calculados . Algoritmo K-Means</i>	96
9.2.1.3. <i>Resultados del clúster K-Means sin datos estandarizados</i>	98

9.2.2. Anexos K-Medoids.....	100
9.2.3. Algoritmo jerárquico aglomerativo.....	101
9.2.4. Fuzzy clustering C-means	102
9.2.5. Density-Based Spatial Clustering of Applications with Noise (DBSCAN).....	104
9.3. Anexos estimación de flujos de efectivo.....	108
9.3.1. Estimación de las provisiones genéricas para el 2022 bajo IFRS17 e IFRS4 (asegurado no jubilado en el 2022)	108
9.3.2. Estimación de las provisiones genéricas para el 2023 bajo IFRS17 e IFRS4 (asegurado no jubilado en el 2023)	112
9.4. Anexos tasa de descuento.....	115
9.4.1. Nelson-Siegel.....	115
9.4.2. Tipo de interés EIOPA	117
9.5. Anexos Fair Value	121
9.5.1. EUROSAP 2022	121
9.5.2. EUROSAP 2023	124
9.5.3. EUROSAP ENERO 2022	128
9.5.4. BONO ALEMÁN 2022	131
9.5.5. BONO ALEMÁN 2023	135
9.6. Anexos resumen conferencia asistida	138
9.7. Anexos scripts en R elaborados para el trabajo	141
9.8. Anexos scripts en Python elaborados para el trabajo.....	150
9.3.1. Análisis descriptivo	150
9.3.2. Técnicas Clúster.....	154
9.3.3. Flujos de efectivo	190

ÍNDICE DE TABLAS

Tabla 1: Características más importantes de la NIIF 4.....	19
Tabla 2: Características más importantes de la IFRS 17.	19
Tabla 3: Diferencias entre NIIF 4 y IFRS 17.	20
Tabla 4: Principales diferencias entre Solvencia II y la IFRS17.....	21
Tabla 5: Variables y descripción de variables del Portafolio de asegurados.	27
Tabla 6: Estadísticos básicos del Portafolio de asegurados.....	28

Tabla 7: Estadísticos básicos de los datos de los contratos de seguro cuya fecha efectiva es el año 2000.....	28
Tabla 8: Frecuencias absolutas.....	30
Tabla 9: Contraste Lilliefors (Kolmogorv – Smirnov modificado) para las variables Age_Actuarial y Capital.....	33
Tabla 10: correlación de Kendall y Spearman de los datos.....	33
Tabla 11: Número de asegurados por año.....	34
Tabla 12: Coordenadas centroides mediante el algoritmo KMeans.....	39
Tabla 13: Principales estadísticos de los datos de cada clúster hecho con K-Means.....	41
Tabla 14: Coordenadas centroides mediante el algoritmo K-medoids.....	42
Tabla 15: Principales estadísticos de cada clúster hecho con K-Medoids.....	44
Tabla 16: Principales estadísticos de cada clúster hecho con el algoritmo Jerárquico Aglomerativo.....	48
Tabla 17: Datos del primer asegurado del portafolio.....	50
Tabla 18: Estimación de la Prima pura, Prima de inventario y Prima de tarifa.....	55
Tabla 19: Parámetros optimizados Nelson - Siegel - Svensson.....	63
Tabla 20: Matriz de correlaciones de Riesgo de suscripción Vida.....	66
Tabla 21: Cálculo SCR's de Suscripción Vida para el año 2022.....	66
Tabla 22: Cálculo SCR's de Suscripción Vida para el año 2023.....	66
Tabla 23: Ajuste por riesgo no financiero.....	67
Tabla 24: Parámetros Nelson Siegel Svensson para EUROSAP 2022 y 2023.....	68
Tabla 25: Fair Value obtenido con la curva de.....	68
Tabla 26: Parámetros Nelson Siegel Svensson para Bono Alemán 2022 y 2023.....	68
Tabla 27: Fair Value obtenido con la curva de.....	68
Tabla 28: Indicación de si hay LC o CSM en 2022 y 2023 mediante el EUROSAP.....	69
Tabla 29: Indicación de si existe LC o CSM en 2022 y 2023 mediante el Bono Alemán.....	69
Tabla 30: Indicación de si existe LC o CSM en 2022 y 2023 mediante el EUROSAP ponderado al 4%.....	70
Tabla 31: Indicación de si existe LC o CSM en 2022 y 2023 mediante el BONO ALEMÁN ponderado al 4%.....	70
Tabla 32: Indicación de si existe LC o CSM en 2022 y 2023 mediante EUROSAP de enero de 2022.....	70
Tabla 33: Dotación de provisión.....	73
Tabla 34: Listado NIF.....	83
Tabla 35: Frecuencias absolutas.....	84

Tabla 36: Frecuencias relativas.	85
Tabla 37: Tabla contingencia entre Género y Estación.	86
Tabla 38: Tabla contingencia entre estación y mes de renovación.	86
Tabla 39: Estadísticos básicos de los datos de los contratos de seguro cuya fecha efectiva es el año 2001.....	87
Tabla 40: Estadísticos básicos de los datos de los contratos de seguro cuya fecha efectiva es el año 2002.....	88
Tabla 41: Estadísticos básicos de los datos de los contratos de seguro cuya fecha efectiva es el año 2003.....	89
Tabla 42: Estadísticos básicos de los datos de los contratos de seguro cuya 90	
Tabla 43: Estadísticos básicos de los datos de los contratos de seguro cuya fecha efectiva es el año 2005.....	91
Tabla 44: Estadísticos básicos de los datos de los contratos de seguro cuya fecha efectiva es el año 2006.....	92
Tabla 45: Estadísticos básicos de los datos de los contratos de seguro cuya fecha efectiva es el año 2007.....	93
Tabla 46: Estadísticos básicos de los datos de los contratos de seguro cuya fecha efectiva es el año 2008.....	94
Tabla 47: Estadísticos básicos de los datos de los contratos de seguro cuya fecha efectiva es el año 2009.....	95
Tabla 48: Estadísticos de las distancias de cada dato al centroide más cercano.....	97
Tabla 49: Coordenadas centroide aplicando K-Means sin datos estandarizados.	99
Tabla 50: Principales estadísticos de los clústeres hechos con K-Means.....	100
Tabla 51: Estadísticos de las distancias de cada dato al centroide más cercano.	101
Tabla 52: Tipos de enlace en el Algoritmo Jerárquico Aglomerativo.....	102
Tabla 53: Grados de pertenencia de cada observación al clúster en el algoritmo Fuzzy C-Means.	104
Tabla 54: Estadísticos descriptivos calculados sobre los grados de pertenencia en algoritmo Fuzzy C-Means.	104
Tabla 55: Resultados DBSCAN (distancias).	108
Tabla 56: Datos asegurado que en 2022 y 2023.....	109
Tabla 57: Prima pura, prima de inventario y 109	
Tabla 58: Parámetros optimizados 116	
Tabla 59: Estructura temporal de tipos de interés y estructura obtenida por Nelson Siegel 2023.	116

Tabla 60: Tipos de interés EIOPA diciembre 2022 y 2023.....	120
Tabla 61: EUROSAP 2022.....	124
Tabla 62: EUROSAP 2023.....	127
Tabla 63: EUROSAP ENERO 2022.	131
Tabla 64: BONO ALEMÁN 2022.....	135
Tabla 65: BONO ALEMÁN 2023.....	138

ÍNDICE DE GRÁFICOS

Gráfico 1: Gráfico de dispersión variables capital y edad actuarial para el año 2000.	31
Gráfico 2: Q-Q Plot de la variable Age_Actuarial.	32
Gráfico 3: Q-Q Plot de la variable Capital.	32
Gráfico 4: Método del Codo.....	37
Gráfico 5: Algoritmo k-means y centroides.	39
Gráfico 6: Algoritmo k-medoids y centroides.....	42
Gráfico 7: Clúster algoritmo jerárquico aglomerativo.	46
Gráfico 8: Clúster algoritmo jerárquico aglomerativo con tres grupos.....	46
Gráfico 9: Estructura temporal de tipos de interés y estructura obtenida por Nelson Siegel para año 2023.....	64
Gráfico 10: Estructura temporal de tipos de interés y estructura obtenida por Nelson Siegel para año 2022.	64
Gráfico 11: Gráfico de dispersión variables capital y edad actuarial para el año 2000.....	87
Gráfico 12: Gráfico de dispersión variables capital y edad actuarial para el año 2002.....	88
Gráfico 13: Gráfico de dispersión variables capital y edad actuarial para el año 2003.....	89
Gráfico 14: Gráfico de dispersión variables capital y edad actuarial para el año 2004.....	90
Gráfico 15: Gráfico de dispersión variables capital y edad actuarial para el año 2005.....	91
Gráfico 16: Gráfico de dispersión variables capital y edad actuarial para el año 2006.....	92
Gráfico 17: Gráfico de dispersión variables capital y edad actuarial para año 2007.	93
Gráfico 18: Gráfico de dispersión variables capital y edad actuarial para el año 2008.....	94
Gráfico 19: Gráfico de dispersión variables capital y edad actuarial para el año 2009.....	95
Gráfico 20: Resultados Índice Silhouette.....	96
Gráfico 21: Gráfico K-Means sin datos estandarizados.	98
Gráfico 22: Gráfico de dispersión Fuzzy C-Means.....	103

Gráfico 23: Gráfico C-Means con elipses de confianza al 95%.....	103
Gráfico 24: 3NN distance DBSCAN.	107
Gráfico 25: Resultados DBSCAN.....	107
Gráfico 26: Estructura temporal de tipos de interés y estructura obtenida por Nelson Siegel 2022.....	117

ÍNDICE DE ILUSTRACIONES

Ilustración 1: Cálculo CSM/LC.....	69
Ilustración 2: Border Point, Core Point y Outlier o Noise.	105
Ilustración 3: Directly density-reachable.	106
Ilustración 4: Density reachable.....	106

ÍNDICE DE EXPRESIONES

Expresión 1: Definición de disimilitud entre datos.....	35
Expresión 2: Distancia Euclídea entre dos individuos para el conjunto de p variables analizadas.	36
Expresión 3: Cálculo suma de las distancias al cuadrado de cada punto y el centroide del clúster al que pertenece k=3. Fuente: Elaboración propia.	37
Expresión 4: Asignación de la observación a un clúster mediante el algoritmo K-Means.....	38
Expresión 5: Cálculo de los centroides.....	38
Expresión 6: Potencial incremento del error resultante de la unión de dos grupos.....	45
Expresión 7: Ecuación de equilibrio prima de tarifa.....	51
Expresión 8: Valor Actual Actuarial de la prima de tarifa.....	51
Expresión 9: Reembolso de primas y seguro vitalicio.	51
Expresión 10: Gastos de gestión interna.	52
Expresión 11: Gastos de gestión externa.	52
Expresión 12: Sustitución a la ecuación de equilibrio para obtener la Prima de Tarifa.	52
Expresión 13: Renta Actuarial de los cobros de primas.	52
Expresión 14: Cláusula reembolso de primas.	53

Expresión 15: Parte del seguro de fallecimiento.....	53
Expresión 16: Distribución de los gastos de gestión interna.	54
Expresión 17: Determinación de la Prima de Tarifa.	54
Expresión 18: Sustitución a la ecuación de equilibrio para obtener la Prima de Inventario.	54
Expresión 19: Determinación de la Prima de Inventario.	54
Expresión 20: Ecuación de equilibrio de la Prima Pura.....	55
Expresión 21: Sustitución a la ecuación de equilibrio para obtener la Prima Pura.....	55
Expresión 22: Determinación de la Prima Pura.	55
Expresión 23: Desequilibrio que se produce a favor de las prestaciones.	55
Expresión 24: Expresión equilibrio para el cálculo de provisión.	56
Expresión 25: Determinación de la provisión matemática.....	56
Expresión 26: Aplicación de la propiedad de escindibilidad.	56
Expresión 27: Determinación de la provisión matemática habiendo aplicado la propiedad de escindibilidad.	56
Expresión 28: Cálculo provisiones a efectos contables ejercicio 2022.....	58
Expresión 29: Cálculo provisiones finales a efectos contables ejercicio 2022.	58
Expresión 30: Prestación Seguro de fallecimiento ejercicio 2022.....	58
Expresión 31: Reparto de gastos de gestión interna ejercicio 2022.	59
Expresión 32: Cálculo provisiones a efectos contables ejercicio 2023.....	60
Expresión 33: Cálculo provisiones finales a efectos contables ejercicio 2023.	60
Expresión 34: Prestación Seguro de fallecimiento ejercicio 2023.....	61
Expresión 35: Reparto de gastos de gestión interna ejercicio 2023.	61
Expresión 36: Formula de Nelson Siegel Svensson.	62
Expresión 37: Formula SCR-suscripción de vida.....	66
Expresión 38: Cálculo provisiones finales a efectos contables ejercicio 2022 (asegurado no jubilado en esta fecha). Fuente: Elaboración propia.....	110
Expresión 39: primas de reembolso 2022 (asegurado no jubilado en esta fecha).	110
Expresión 40: Primas consolidadas 2022 (asegurado no jubilado en esta fecha).....	110
Expresión 41: Prestación seguro de fallecimiento 2022 (asegurado no jubilado en esta fecha).	111
Expresión 42: Reparto gastos de gestión interna 2022 (asegurado no jubilado en esta fecha).	111
Expresión 43: Renta para el cálculo de gastos de gestión interna 2022 (asegurado no jubilado en esta fecha).	111

Expresión 44: Pago de primas por parte del asegurado a la aseguradora 2022 (asegurado no jubilado en esta fecha). Fuente: Elaboración propia.....	112
Expresión 45: Cálculo provisiones finales a efectos contables ejercicio 2023 (asegurado no jubilado en esta fecha). Fuente: Elaboración propia.....	113
Expresión 46: Reembolso de primas 2023 (asegurado no jubilado en esta fecha).....	113
Expresión 47: Primas consolidadas 2023 (asegurado no jubilado en esta fecha).....	113
Expresión 48: Parte del seguro de fallecimiento 2023 (asegurado no jubilado en esta fecha).	114
Expresión 49: Reparto de gastos 2023 (asegurado no jubilado en esta fecha).....	114
Expresión 50: Renta para el cálculo de los gastos de gestión interna 2023 (asegurado no jubilado en esta fecha). Fuente: Elaboración propia.....	114
Expresión 51: Pago de primas por parte del asegurado a la aseguradora 2023 (asegurado no jubilado en esta fecha). Fuente: Elaboración propia.....	115
Expresión 52: Fórmula Nelson-Siegel.	115

1. INTRODUCCIÓN

Este trabajo se centra en el estudio actuarial de la normativa IFRS17. Introducida por el International Accounting Standards Board (IASB), representa una de las reformas más significativas en la contabilidad de seguros en los últimos años. Establecida para reemplazar la normativa IFRS4, IFRS17 tiene como objetivo proporcionar una base consistente para la contabilidad enfocada a la parte del pasivo del balance de los contratos de seguros, mejorando la comparabilidad y transparencia de los informes financieros de las empresas del sector. Esta nueva normativa aborda las debilidades de su predecesora, proporcionando un marco más riguroso y detallado para la medición y presentación de los contratos de seguros.

Dentro la normativa IFRS17, la cual entró en vigor en enero de 2023, se encuentran tres enfoques: el método general de bloques, el enfoque de asignación de prima o PAA o “Premium Allocation Approach” y el modelo de tarifa variable o VFS o “Variable Fee Approach”. Este trabajo se centra en la aplicación de la normativa IFRS17 a un portafolio de seguros de vida entera con una temporalidad superior al año y, por tanto, debe aplicarse el método general o de bloques. Este método se caracteriza por el cálculo de los flujos de efectivo futuros, la tasa de descuento, el ajuste por riesgo no financiero o Risk Adjustment (RA) y el Margen de Servicio Contractual (CSM) o Loss Component (LS).

Durante la formación del Máster se me ofreció la oportunidad de realizar este trabajo sobre la temática de la normativa IFRS17 desde el punto de vista actuarial ya que estaba cursando el Máster en Ciencias Actuariales y Financieras y debido a que había cursado anteriormente el Máster de Contabilidad y Fiscalidad en la misma Universidad. Esta base sólida en contabilidad y fiscalidad proporcionó un contexto ideal para abordar la complejidad de la normativa IFRS17, y aunque inicialmente era conocedora de la existencia de las Normas Internacionales de Contabilidad, dicha normativa en especial era totalmente desconocida por mí.

A lo largo de este trabajo, se analizará la aplicación práctica de IFRS17 desde el punto de vista actuarial en un portafolio de seguros de vida. Este análisis se realizará mediante la aplicación detallada de la normativa a un contrato específico dentro de un portafolio de 76.102 contratos. Al focalizarse en un contrato particular, se pretende proporcionar una visión clara y comprensible del proceso. El procedimiento detallado para este contrato puede extrapolarse al resto de portafolio, ilustrando de forma práctica la aplicación de cada etapa en el enfoque general.

La realización de esta investigación presentó una serie de desafíos, retos y obstáculos significativos que enriquecieron el proceso y aportaron un valor adicional a los resultados obtenidos. Uno de los primeros desafíos fue el análisis en profundidad de la normativa contable IFRS17, lo que demandó una labor investigadora exhaustiva desde el principio. La necesidad de explorar y comprender conceptos fundamentales y avanzados requirió la búsqueda activa de fuentes primarias y secundarias como son la consulta con expertos en la materia, la asistencia a conferencias¹, informes y demás documentos relacionados con la normativa y la revisión de literatura especializada.

¹ Ver 9.6. Anexos resumen conferencia asistida.- Pág. 137.

Otro de los desafíos fue encontrar una base de datos apropiada para realizar el estudio, lo cual me llevó a varios días de búsqueda, puesto que en primer lugar desconocía como y donde buscarlas, y además muchas de ellas existían barreras económicas y eran de carácter privado. En este caso finalmente conseguí encontrar un portafolio de vida que era real, y que pertenecía a una de las bibliotecas de investigación de una Universidad de Estados Unidos, a la cual estoy muy agradecida.

A medida que avanzaba en la investigación, se hizo evidente la necesidad de adoptar un enfoque autodidacta. Este proceso de aprendizaje continuo hacía que me surgieran constantemente nuevas preguntas sobre diversos aspectos del tema que permitieron desarrollar una comprensión más profunda en esta área de estudio, aunque también implicó un camino de ensayo y error. Algunos de estos errores fueron metodológicos, como la selección inapropiada de técnicas de análisis. Esta tarea no ha sido sencilla ya que la normativa en algunos casos deja elegir libremente al actuario el método a aplicar siendo a su juicio el que decide cuál es el más óptimo en cada situación. Otra dificultad surgida ha sido la interpretación de la normativa lo que ha requerido una revisión constante del trabajo, esfuerzo y constancia con el fin de que fuera correcta.

La gestión del tiempo también ha sido un reto considerable, especialmente al ser un proyecto de duración corta de tres meses, y que además se ha compaginado con las prácticas del Máster a la vez, las cuales estaban vinculadas con el ámbito actuarial pero no tenían relación alguna con los aspectos de este trabajo. Sin embargo, estas prácticas me permitieron observar el mundo real y rodearme de personas expertas a quienes podía consultar cuando surgían dudas sobre algún aspecto de la normativa y a las cuáles estoy muy agradecida.

Juntamente con el tiempo, otro obstáculo ha sido la extensión del trabajo, lo que ha llevado a simplificar algunas partes. Aun así, este trabajo logra ofrecer una visión general, ilustrativa y clara de la aplicación de la normativa IFRS17 y, en particular, de la obtención del balance de transición, sobre un portafolio de vida. Así pues, este trabajo se enfoca a los aspectos fundamentales de la normativa y aunque se hayan tenido que llevar a cabo estas simplificaciones, el trabajo mantiene su integridad y utilidad como una herramienta valiosa para aquellos interesados en comprender de una forma simple la aplicación de la normativa IFRS17 y obtención del balance de transición en el contexto específico de un portafolio de vida.

Este trabajo se ha estructurado en cuatro apartados. El primer apartado ofrece un repaso detallado sobre la normativa IFRS17, además se compara esta normativa con NIIF4 y Solvencia II. El segundo apartado se centra en realizar una revisión bibliográfica sobre IFRS17 que incluye los principales estudios que se han llevado a cabo sobre el tema. El tercer apartado trata sobre el análisis descriptivo del portafolio de seguros mediante gráficos de dispersión, estadísticos más importantes y correlaciones. El cuarto apartado trata sobre la aplicación del Método general o de bloques. Este último apartado además está formado en varios subapartados que de forma simplificada equivalen a la agregación de contratos de seguro mediante técnicas de clúster, el cálculo de la prima y la estimación de flujos de efectivo futuros bajo IFRS17 y NIIF4, la determinación de la tasa de descuento y la obtención del Margen de Servicio Contractual (CSM) o Loss Component (LC) según corresponda.

Finalmente, es necesario indicar que para realizar este trabajo se han utilizado diversos programas informáticos para asegurar un análisis riguroso y preciso. En particular se ha empleado Python² durante todo el trabajo y en concreto se han utilizado las bibliotecas Numpy, Pandas, Scikit-learn, Scikit-learn-extra, Skfuzzy, Matplotlib, Scipy y Seaborn. Además, también se ha utilizado Excel para la comprobación de rentas y RStudio y SPSS Statistics para la comprobación de la parte de análisis descriptivo y técnicas clúster. En el caso de RStudio³ se han utilizado específicamente los paquetes flexclust, mclust, clusterSim, gg dendro, ppclust, factoextra, dplyr, cluster, fclust, psych, lattice, caret, dbscan, stats, tidyverse, ggplot2, dlyr y readxl.

Con esto, se pretende no solo demostrar la aplicabilidad y beneficios de IFRS17, sino también ofrecer una guía práctica para su implementación efectiva en el sector asegurador.

2. LA CONTABILIDAD Y ANTECEDENTES DE LA NORMATIVA

IFRS17

Desde tiempos ancestrales, la contabilidad ha desempeñado un papel fundamental en las civilizaciones al proporcionar un medio para el registro, la gestión y el control de recursos en las transacciones económicas y el patrimonio. A lo largo de los siglos, esta disciplina ha ido evolucionando desde los primeros métodos rudimentarios de registro en tablillas de arcilla hasta llegar a los sistemas contables computarizados modernos permitiendo una gestión más eficiente y precisa de la información financiera en el mundo contemporáneo. De la misma forma, la contabilidad, se ha ido adaptando debido a la creciente complejidad de las operaciones comerciales y financieras. En este sentido, en un mundo globalizado las Normas Internacionales de Información Financiera (NIIF)⁴ han desempeñado un papel crucial estableciendo un conjunto común de reglas aceptadas a nivel internacional, permitiendo la comparabilidad de los estados financieros de las empresas independientemente de la jurisdicción.

En consonancia con la importancia de la contabilidad y su evolución, a nivel legislativo en España se encuentra en el Art. 83.1. de la Ley 20/2015, de 14 de julio, de ordenación, supervisión y solvencia de las entidades aseguradoras y reaseguradoras que establece que la contabilidad de este tipo de entidades se registrará por sus normas específicas, y en ausencia de estas, por las normas establecidas en el Código de Comercio, en el Plan General de Contabilidad y en las demás disposiciones de la legislación mercantil en materia contable. En este sentido, el Art. 25 del Código de Comercio señala la obligación del empresario de llevar a cabo una contabilidad ordenada y adecuada a la actividad de la empresa con el fin de poder realizar un seguimiento cronológico de sus operaciones, así como la elaboración periódica de balances e inventarios. Asimismo, el Art. 83.3 de la Ley 20/2015, de 14 de julio, de ordenación, supervisión y solvencia de las entidades aseguradoras y reaseguradoras establece la existencia de un Plan General Contable

² Ver 9.8. Anexos scripts en Python elaborados para el trabajo.- Pág. 149.

³ Ver 9.7. Anexos scripts en R elaborados para el trabajo.- Pág. 139.

⁴ Ver 9.1. Anexos contabilidad y antecedentes de la normativa IFRS17 – Pág. 82.

específico para las entidades de seguros, otorgando al Ministerio de Economía y Competitividad con previo informe del Instituto de Contabilidad y Auditoría de Cuentas y de la Junta Consultiva de Seguros y Fondos de Pensiones la potestad de poder realizar modificaciones y dictaminar normas complementarias. En este mismo artículo, se hace especial hincapié a que el Ministro de Economía y Competitividad podrá encomendar a la Dirección General de Seguros y Fondos de Pensiones el desarrollo de normativa específica de este tipo de entidades, así como la adecuación de las normas internacionales de información financiera aplicables siempre que haya realizado un informe previo por parte del Instituto de Contabilidad y Auditoría de Cuentas y de la Junta Consultiva de Seguros y Fondos de Pensiones.

Las Normas Internacionales de Información Financiera (NIIF) o International Financial Reporting Standards (IFRS)⁵ y las Normas Internacionales de Contabilidad (NIC) son un conjunto de normas que establecen los requisitos de reconocimiento, medición, presentación e información a revelar que se refieren a las transacciones y sucesos económicos que son importantes en los estados financieros con propósito de información general.⁶ En la relación entre la normativa IFRS y el contexto previamente mencionado sobre la obligación de llevar a cabo la contabilidad de forma adecuada por parte de las entidades aseguradoras, se encuentra la necesidad de estas empresas de adherirse a los estándares contables internacionales, garantizando cumplir con los principios contables⁷ de fiabilidad, relevancia, consistencia y comprensibilidad. Tal y como puede verse en la página web IFRS Foundation y tal y como indican Jousef et al. (2021) dichas normas pretenden aportar transparencia, rendición de cuentas y eficiencia. La transparencia facilita una mejor comparabilidad internacional aportando además mayor calidad en la información financiera, permitiendo a inversores y participantes del mercado mejorar la toma de decisiones. La rendición de cuentas por parte de las empresas se realiza con el fin de justificar sus decisiones y acciones ante aquellos a quienes afectan o a quienes les han cedido recursos o responsabilidades. Y, por último, la eficiencia, permitiendo a inversores y empresas entender mejor las oportunidades y riesgos siguiendo un conjunto de normas contables confiables y globalmente aceptadas.

Las Normas Internacionales de Información Financiera están formadas por las Normas Internacionales de Contabilidad (NIC), las Normas Internacionales de Información Financiera (NIIF) propiamente dichas, y las Interpretaciones sobre NIC y NIIF (Vázquez y Díaz, 2013). De cada normativa se señalan: los objetivos de emisión, partidas a las que aplica, reglas de valoración, momentos de reconocimiento y revelaciones en estados financieros.

Además de remarcar los objetivos de las NIIF es importante destacar que toda esta normativa es aplicable a cualquier país del mundo y el hecho de asumirlas o no dependerá de la propia soberanía. También es necesario incidir que puede aplicarla cualquier tipo de empresa independientemente de su tamaño.

El organismo que emite las NIIF es el Consejo de Normas Internacionales de Contabilidad o *International Accounting Standards Board* (IASB). Desde el 3 de noviembre de 2021

⁵ Hasta el 2001 se les había llamado como IAS.

⁶ Según Norma Internacional de Contabilidad 1. Presentación de estados financieros. (Párrafo 3).

⁷ Los principios contables a cumplir son: devengo, uniformidad, prudencia, no compensación, importancia relativa, empresa en funcionamiento.

también es responsable la *International Sustainability Standards Board (ISSB)*, la cual emite estándares de información relacionados con la sostenibilidad en los mercados de capital globales. Esto permite, igual que las IFRS, que esta información sea comparable a nivel global y permite a las empresas proporcionar información completa sobre los riesgos y oportunidades relacionados con la sostenibilidad evitando la fragmentación actual de los estándares y requisitos voluntarios relacionados con este tema.

Este trabajo se centra en la normativa IFRS17 relacionada con la valoración de los contratos de seguros y que entró en vigor al 1 de enero de 2023 (Yousuf et al., 2021). Las NIIF 17 se han elaborado con el fin de reemplazar la normativa NIIF 4 con la finalidad de eliminar las inconsistencias, armonizar y comparar los estados financieros, así como representar fielmente los contratos de seguro.

Tanto las NIIF 4 como las IFRS17 persiguen señalar cómo deben valorarse e informarse los derechos y obligaciones de los contratos de seguros. A nivel contable tiene especial relevancia ya que existen las llamadas asimetrías contables (Pérez et al., 2018) que aparecen cuando los criterios de valoración de los pasivos que derivan de un contrato de seguro difieren de los criterios de valoración de los activos. Ambas normativas comparten la misma definición de contrato de seguro el cual definen como:

“Un contrato en el que una de las partes (el emisor) acepta un riesgo de seguro significativo de la otra parte (el tomador de la póliza), acordando compensar al tomador de la póliza si ocurre el evento futuro incierto especificado (el evento asegurado) que afecta de forma adversa al tomador.”

Las principales características de ambas normativas aparecen en las Tablas 1 y 2, respectivamente. Las diferencias entre ambas normativas se han sintetizado en la Tabla 3.

NIIF 4

- Carácter opcional de la disociación de la componente de seguro y componente de depósito. Solo se separarán si la compañía las puede valorar por separado.
 - Se aplica la NIIF 4 al componente de seguro.
 - Se aplica la NIIF 9 al componente de depósito.
- Prueba de adecuación de los pasivos:
 - 1r caso: La entidad cumple con los requisitos mínimos para aplicar la prueba de adecuación de los pasivos:
 - Actuarialmente: deben realizarse las estimaciones actuales de los flujos de efectivo derivados de contratos y de todo aquello relacionado con ello.
 - Contablemente: Se compara el resultado obtenido con el importe en libros de los pasivos derivados de los contratos de seguros y, en caso de ser inadecuado, la diferencia se reconoce en el resultado del ejercicio.
 - 2n caso: La entidad no requiere la aplicación de la prueba de adecuación de los pasivos:
 - Contablemente: Se coge el importe en libros de los pasivos y se minoran con los costes de adquisición diferidos y los activos intangibles relacionados. El importe obtenido se compara con el importe en libros que se requeriría en caso de que los pasivos

- derivados de contratos de seguros relevantes estuvieran dentro del alcance de la NIC 37.
 - Actuarialmente: la comparación debe hacerse con una cartera de contratos de seguros que tengan riesgos similares y se gestionen conjuntamente.
- Deterioro del valor de los activos derivados de contratos de reaseguro:
 - Contablemente: Se reconocerá como una pérdida en el resultado del ejercicio.
- Cambios en las políticas contables en relación a los contratos de seguro:
 - Tipos de interés actuales de mercado. No es obligatorio.
 - Actuarialmente: Recalcular los pasivos teniendo en cuenta los tipos de interés de mercado vigentes. Introducción de otras estimaciones e hipótesis.
 - Se permite continuar con determinadas prácticas existentes relacionadas con:
 - Actuarialmente: valorar los pasivos derivados de contratos de seguros sin descontar los importes.
 - Actuarialmente: valorar derechos contractuales relacionados con futuras comisiones de gestión de inversiones por un importe que exceda al valor razonable de las mismas.
 - Contablemente: Políticas contables no uniformes en los contratos de seguro de las dependientes.
 - No eliminar el exceso de prudencia.
 - Márgenes de inversión futuros. Con el fin de que los estados financieros sean más relevantes y no menos fiables se propone:
Actuarialmente:
 - Estimaciones e hipótesis actuales
 - Ajuste razonable para reflejar el riesgo y la incertidumbre
 - Valoraciones que reflejen tanto el valor intrínseco como el valor temporal de las opciones y garantías implícitas de los contratos
 - Tipo de descuento de mercado actual
 - Contabilidad tácita
Consiste en que las pérdidas o ganancias relacionadas con los activos afecten de igual manera a los pasivos de los contratos de seguros, a los costes de adquisición diferidos y los activos intangibles.
- Método de superposición
Es opcional
Contabilidad:
 - Reclasificar los resultados a otro resultado global. El importe de la reclasificación será la diferencia entre el importe de los activos financieros aplicando NIIF 9 y el importe de los activos financieros aplicando NIC 39.
 - En la norma se indica qué activos financieros son susceptibles a aplicarse esta normativa, efectos y como debe llevarse a cabo.
- La normativa indica en relación a la información a revelar que debe indicarse
Actuarialmente:
 - la forma y métodos aplicados para la gestión del riesgo.
 - Hace especial énfasis en el riesgo de crédito, el riesgo de liquidez y al riesgo de mercado.
 - También debe indicarse los métodos e hipótesis utilizados para el análisis de sensibilidad.

- Importe, calendario e incertidumbre de los flujos de efectivo de la aseguradora.
- Desviación entre los siniestros reales y los estimados.
- Cambios en la valoración de activos y pasivos.
- Exposición al riesgo.

Contablemente:

- Informar de las pérdidas y ganancias.

Tabla 1: Características más importantes de la NIIF 4.

Fuente: Elaboración propia.

IFRS17

- Separar del contrato de seguro los derivados implícitos, componentes de inversión, futuras transmisiones de bienes o servicios. No se separarán los componentes de inversión si tiene las características de participación discrecional o servicios propios del contrato.
- Se agregarán los contratos según riesgos similares. La norma prevé que contratos de diferentes líneas de productos tienen diferentes riesgos. También deben estar en grupos distintos los contratos que se han emitido con más de un año de diferencia. A partir de grupos se pueden generar subdivisiones y la composición de los grupos no se reevaluará. Además, dentro de cada portafolio, los contratos pueden dividirse en: contratos onerosos, contratos con baja probabilidad de convertirse en onerosos y resto de contratos. Se considera que un contrato es oneroso en el momento inicial cuando los flujos de pago son superiores a los flujos de cobro.
- Existen tres métodos para la valoración de los pasivos: el método general de bloques, el enfoque de asignación de prima o PAA o “Premium Allocation Approach” y el modelo de tarifa variable o VFS o “Variable Fee Approach”.
 - Método general de bloques: deben realizarse las estimaciones de flujos futuros, calcular la tasa de descuento (enfoque Bottom-up y enfoque Top-down), el ajuste por riesgo y el margen de servicio contractual.
 - Enfoque asignación de prima: se pretende aplicar el modelo general para los pasivos por siniestros incurridos y el enfoque de asignación de primas para los pasivos por cobertura pendiente.
 - Modelo de tarifa variable: aquellos en los que el asegurado tiene derecho a participar de la rentabilidad de un conjunto de activos.
- Para la transición a la nueva normativa debe aplicarse:
 - El método retrospectivo:
Debe hacerse un recalcu desde el origen de la póliza. Dicha información no siempre se tiene por lo que a veces es imposible aplicarlo.
 - El método retrospectivo modificado:
Se pretende lograr un resultado cercano al método retrospectivo utilizando, por tanto, el menor número de modificaciones usando toda la información razonable y sustentable disponible.
 - El método Fair Value
Deberá utilizar la información de mercado en la fecha de transición.
En este método debe aplicarse lo previsto en IFRS 13.

Tabla 2: Características más importantes de la IFRS 17.

Fuente: Elaboración propia.

NIIF 4	IFRS 17
<ul style="list-style-type: none"> • Opcionalidad de a disociación del componente seguro y componente depósito • Prueba de adecuación de pasivos • Cambios de las políticas contables en relación a los contratos de seguro. En particular, la contabilidad tácita. • Método de la superposición. 	<ul style="list-style-type: none"> • Separación del contrato de seguros⁸ los derivados implícitos⁹, componentes de inversión, futuras transmisiones de bienes o servicios. • Agregación de contratos con riesgos similares. • Para la transmisión a la nueva normativa: el método retrospectivo, el método retrospectivo modificado y el método Fair Value. • Tres métodos: el método general de bloques, el enfoque de asignación de prima y el modelo de tarifa variable.

Tabla 3: Diferencias entre NIIF 4 y IFRS 17.
Fuente: Elaboración propia.

Como puede verse, IFRS 17 introduce un enfoque más detallado y preciso sobre la valoración de los contratos de seguros, lo cual requiere la aplicación de técnicas actuariales avanzadas y, en donde, por tanto, el actuario juega un papel muy importante. De hecho, en las IFRS17 los principios contables reconocen las pérdidas y ganancias teniendo en cuenta la variable tiempo (Lu, 2023). Estas técnicas permiten estimar de forma más precisa los flujos de efectivo futuros y evaluar riesgos (Pérez et al., 2018) asociados con los contratos de seguros.¹⁰ Básicamente,

- Los objetivos, políticas y procesos para la gestión de riesgos (García et al., 2019).
- La sensibilidad del riesgo de seguro.
- Las concentraciones del riesgo de seguro.
- Las desviaciones respecto las estimaciones de flujos.
- La exposición al riesgo como puede ser la derivada de inversiones (Cabedo y Tirado, 2003).

En los siguientes apartados se explorará en profundidad el impacto del componente de segmentación de riesgos y actuarial en IFRS 17, así como los desafíos y consideraciones prácticas asociadas a su implementación.

Finalmente, en el contexto asegurador existe otra normativa importante: Solvencia II. Se pueden identificar numerosas diferencias entre IFRS17 y Solvencia II, siendo una de las más notorias a simple vista el objetivo que persiguen. Mientras IFRS17 se centra en mostrar la imagen fiel de la situación financiera de la empresa desde la perspectiva contable, Solvencia II tiene como principal objetivo garantizar la solvencia financiera de las aseguradoras a través de la evaluación del capital basado en riesgo, abarcando aspectos

⁸ La separación de un derivado implícito de su contrato recibe el nombre de “bifurcación”.

⁹ Son productos dentro de los que hay un contrato de un instrumento derivado. Este instrumento derivado afecta a los flujos de efectivo futuros esperados.

¹⁰ Como puede ser, por ejemplo, medir el riesgo de que los contratos de seguro se vuelvan onerosos.

relacionados con la gestión de riesgos, el gobierno corporativo y la revelación de información financiera de las compañías de seguros. Otra diferencia principal es que Solvencia II es una normativa específica de la Unión Europea (UE) que se aplica a las compañías de seguros de los estados miembros de la UE y del Espacio Económico Europeo (EEE). Explorando más a fondo los detalles de cada normativa, se pueden apreciar diferencias más específicas tales como (Tabla 4):

SOLVENCIA II	IFRS17
Reconocimiento de ganancias esperadas	
Se reconocen en la reserva de reconciliación	Se reconocen en el margen contractual de servicios o CSM distribuyéndose a lo largo del tiempo
Flujos de efectivo	
Se basa en el “valor de transferencia”	Se basa en los llamados “flujos de cumplimiento”
Reaseguro	
Los pagos del reaseguro son incluidos en el cálculo bruto de la mejor estimación y en las proyecciones de los flujos de efectivo en las provisiones técnicas	Se tratan por separado los contratos de reaseguro y los flujos de efectivo de los contratos de seguro
Gastos	
Se asignan a las provisiones técnicas individuales	Son directamente atribuibles a un portafolio
Agrupamiento	
Los contratos se agrupan por grupos homogéneos (HRG)	Los contratos se agrupan por riesgos similares y por cohortes anuales (agruparlos por año de emisión o inicio de cobertura). Además, se dividen en: <ul style="list-style-type: none"> • Contratos onerosos al reconocimiento inicial. • Contratos sin posibilidad significativa de volverse onerosos • Otros contratos
Tasa de descuento	
Utiliza las tasas de interés libres de riesgo. Considera ajustes de volatilidad y “matching adjustment”	Dos enfoques: <ul style="list-style-type: none"> • Bottom-up • Top-down
Ajuste por riesgo	
Es el capital necesario para cubrir las obligaciones	Equivale a la compensación que refleja la incertidumbre de los flujos futuros no financieros

Tabla 4: Principales diferencias entre Solvencia II y la IFRS17.
Elaboración propia.

3. ESTADO DE LA CUESTIÓN

Prima et al. (2024) proponen para la clasificación de riesgos asociados con pólizas la aplicación de algoritmos como CLARA (Clustering Large Application), Extreme Gradient Boosting, el Support Vector Machine y el Adaptive Boosting. Los autores remarcan la importancia de un análisis de riesgos preciso ya que influye significativamente en la liberación del Margén de Servicio Contractual o Contractual Service Margin (CSM). Para la determinación del número óptimo de clústeres en CLARA utilizan la estadística de brecha. A través de los métodos Extreme Gradient Boosting, Support Vector Machine y Adaptive Boosting, se evalúa la precisión y el rendimiento de cada método. Otro de los temas que tratan en su artículo es la preparación de los datos mediante técnicas de preprocesamiento, como la eliminación de duplicados, el manejo de desequilibrios entre clases utilizando las técnicas Random Over Sampling (ROS), y la normalización de las reclamaciones totales.

En su estudio, Winkler et al. (2021), ofrecen una revisión exhaustiva de los desafíos clave asociados con la implementación de la normativa IFRS17. Entre estos desafíos destacan la división de las carteras de seguros según diferentes criterios como líneas de negocio, la rentabilidad esperada y el año de suscripción, manteniendo esta división hasta la expiración del contrato. Además de destacar la dificultad de identificar y separar componentes vinculados con los riesgos del seguro de aquellos que son independientes de ellos. La normativa también genera cierta volatilidad, particularmente debido a la agrupación de contratos de seguros, la falta de compensación de los portafolios onerosos y no onerosos, y la sensibilidad de las tasas de descuento. En relación a la no compensación de los portafolios onerosos y no onerosos debe reconocerse una pérdida de forma inmediata. Asimismo, la sensibilidad de las tasas de descuento utilizadas para el cálculo del valor presente de los flujos de efectivo futuros y la liberación de reservas de beneficios contractuales (CSM) a menudo se ve afectada por los cambios que se producen en los mercados. También puede resultar un desafío la elección de la curva de rendimiento adecuada y la realización de simulaciones estocásticas para distribuciones asimétricas.

En el mismo artículo se hace referencia a que la amortización del Margén de Servicio Contractual (CSM) mediante las llamadas “Unidades de cobertura” plantea otro desafío, dado que las métricas están poco definidas para su determinación. Se sugiere que dichas métricas podrían basarse en la suma total asegurada, primas, etc. no requiriendo, por tanto, cálculos actuariales. Del mismo modo, el Ajuste de Riesgo (AR) carece de orientación precisa, aunque se requiere revelar el cuantil representativo indicando la aversión al riesgo. Se señala que a menudo las empresas basan el AR en cálculos de Solvencia II. Por último, se destaca el desafío tecnológico que supone la aplicación de IFRS17, teniendo que analizar datos de forma granular y detallada, guardándolas durante largos periodos de tiempo, realizar cálculos estocásticos complejos, así como la ejecución de simulaciones.

Yousuf et al. (2021) realizan un estudio profundo tratando la normativa IFRS17 enfocándose en dos aspectos, i) el impacto que tiene el uso de diferentes curvas de rendimiento promedio ponderado en el CSM y, ii) el impacto que tiene el CSM en el retorno sobre el patrimonio. Contablemente, si el contrato de seguros puede generar ganancias para la aseguradora se reconoce como un activo (es una responsabilidad

negativa), y en el caso de que se considere que puede generar pérdidas debe reconocerse como un pasivo (responsabilidad positiva). Al establecerse una pérdida de un contrato de seguros no puede determinarse con un CSM negativo, sino que debe reconocerse como una pérdida en pérdidas y ganancias. Los autores ponen valor a la asimetría contable producida del contrato de seguros como una ganancia diferida en el tiempo, y que solo se reconocerá como pérdida o ganancia (derivada del contrato de seguros) cuando se presten los servicios. En este sentido,

- CSM: es el exceso de los inflows por encima de los outflows
- Loss component: es el déficit de los inflows comparado con los outflows.

También remarcan que el CSM solo puede calcularse bajo los enfoques Modelo de Medición General (GMM) y el enfoque Variable Fee Approach (VFA). Finalmente, con el fin de reconocer las ganancias de cada período deben establecerse el número de unidades de cobertura.

El Margen de Servicio Contractual (CSM) refleja las pérdidas y ganancias esperadas de un determinado grupo de contratos de seguros. Este Margen de Servicio Contractual debe revaluarse en cada periodo contable ya que debe incorporarse la nueva información que se va teniendo (Palmborg et al., 2021; Yousuf et al., 2021). Teniendo en cuenta la normativa, el CSM se puede definir como el valor presente del valor esperado de los cash inflows menos el valor presente del valor esperado de los cash outflows teniendo en cuenta el ajuste por riesgo no financiero. A modo de ejemplo, Lu (2022) estudia el Margen de Servicio Contractual analizando el mecanismo de reparto que distribuye las ganancias y riesgos entre la compañía de seguros y el asegurado en un contrato de vida cuyo valor de la póliza depende del tipo de interés del mercado financiero. El procedimiento que lleva a cabo para realizar el análisis implica calcular los valores de cuenta inicial de ingresos por inversiones, reclamaciones de fallecimiento y pólizas de cancelación en tablas detalladas. Además, el único factor estocástico asumido es la tasa de rendimiento de la inversión, suponiendo un movimiento Browniano geométrico. Luego, utiliza varias simulaciones de Monte Carlo para generar escenarios distintos de tasas de retorno de la inversión y porciones distintas de participación, calcula el Valor en Riesgo (VaR) y, finalmente, evalúa el rendimiento ajustado al riesgo de los productos de seguros y medidas de rendimiento tanto para la aseguradora como para el titular de la póliza. Es importante que exista un equilibrio entre la distribución de ganancias y riesgos entre el asegurado y la aseguradora con el fin de beneficiarse mutuamente.

Chevallier et al. (2018), proponen un modelo para modelizar el riesgo basado en la duración y convexidad y el riesgo biométrico, ambos en el contexto de un portafolio de seguros de vida. Los autores recuerdan que en el Best Estimate of Liabilities (BEL) debe tenerse en cuenta ciertas hipótesis, entre las cuales las que se basan en la mortalidad. Por tanto, el modelo se basa en dos partes principales, una que describe la sensibilidad del BEL a las tasas de interés basadas en la convexidad/duración, y la otra parte basada en el riesgo biométrico que estima el requerimiento de capital de solvencia (SCR) medido con un VaR al 99.5%. Para la agregación de los dos modelos y, en concreto, para la evaluación del riesgo proponen el coeficiente de variación, la asimetría y la curtosis y, para la determinación del nivel de confianza del riesgo de reserva utilizan dos aproximaciones:

Cornish-Fisher (Normal Power) y Bohman-Esscher. Para la agregación de la volatilidad y la asimetría proponen los polinomios de Fleishman.

Pedra et al. (2022) en su artículo desarrollan un método para que las aseguradoras evalúen el ajuste de riesgo (RA) requerido por IFRS17 basado en la Teoría de Riesgo Colectivo y simulaciones de Monte Carlo. Este método proporciona directamente el RA para cada pasivo relacionado con un grupo de contratos de seguros: coberturas restantes y reclamaciones incurridas. Los resultados muestran que, para carteras de gran escala, el Teorema Central del Límite se cumple, lo que permite aproximar las distribuciones utilizadas para evaluar el RA a la distribución Normal. Además, los valores obtenidos son relativamente pequeños lo que concuerda con la Ley de los Grandes Números. Para validar empíricamente su enfoque, los autores utilizan una base de datos que contiene información detallada sobre reclamaciones de seguros de automóviles ocurridas en 2020. Por tanto, el objetivo era estimar la distribución de la variable aleatoria número de siniestros incurridos en un año y para ello trabajaron con una muestra importante de datos y de un periodo largo de tiempo, pero no reflejaba el comportamiento del asegurado ni la siniestralidad actual. Con el fin de solucionar este problema en lugar de estimar la distribución de probabilidad de la variable aleatoria número de siniestros incurridos en un año, estiman la distribución de probabilidad del número de siniestros diarios (N_{daily}) y acaban seleccionando de forma aleatoria un total de 366 observaciones de esta distribución N_{daily} para seguir con el análisis. Seguidamente, aplican la distribución binomial negativa para modelar N_{daily} y determinar sus parámetros. Además, estiman la distribución del valor presente de las reclamaciones individuales (X_i) utilizando la distribución Lognormal. Con estas distribuciones, simulan la distribución empírica del valor presente agregado de las reclamaciones (S) mediante el método Monte Carlo. Luego, se calculan los factores de carga para evaluar el ajuste de riesgo (RA) de cada provisión técnica relacionada con los contratos de seguros (coberturas restantes y reclamaciones incurridas), utilizando diferentes niveles de confianza y medidas de riesgo como VaR y CTE. Estos factores de carga representan la cantidad adicional necesaria para cubrir el riesgo y se aplican a las primas no devengadas y al valor presente esperado de las reclamaciones incurridas para calcular el RA asociado con cada provisión técnica. Los resultados muestran que los factores de carga son relativamente pequeños y se mantienen dentro de un rango aceptable según la Ley de los Grandes Números. Además, según IFRS17, el RA debe evaluarse para cada grupo de contratos de seguros con riesgos similares, considerando también los efectos de diversificación entre diferentes carteras de seguros.

También England et al. (2019) continúan explorando enfoques específicos para llevar a cabo la estimación del ajuste por riesgo para reflejar la incertidumbre inherente de los flujos de efectivo futuros de los contratos de seguros. Se examinan exhaustivamente dos enfoques principales: el enfoque del costo de capital y la aplicación de medidas de riesgo a la distribución de los flujos de efectivo futuros descontados. Se discuten meticulosamente varios métodos, como VaR, TVaR y PHT, destacando sus ventajas y limitaciones en términos de reconocimiento de asimetría y extremos, así como su coherencia como medidas de riesgo. Además, se analiza cómo el enfoque del costo de capital, aunque común en Solvencia II, puede plantear desafíos en el contexto de IFRS17, donde la valoración se basa en el valor esperado descontado de todos los flujos de efectivo futuros. Se enfatiza la elección de la tolerancia al riesgo y se exploran métodos para

calcular los requisitos de capital, utilizando distribuciones simuladas de flujos de efectivo futuros.

Longoni (2019) lleva a cabo un estudio sobre el impacto que tiene el estándar contable IFRS17 en el valor de mercado de las aseguradoras, comparando aquellas que siguen este estándar con aquellas que aplican U.S. GAAP. En concreto, se utiliza el enfoque de Khan et al. (2018) y se lleva a cabo un análisis univariado formado por 16 eventos calculando el rendimiento anormal acumulado (CAR) de tres días alrededor de cada evento que afecta la probabilidad de que IFRS17 se apruebe. El experimento está formado por un grupo de tratamiento formado por 303 aseguradoras de 64 países diferentes, y un grupo de control formado por 124 aseguradoras estadounidenses cuyas acciones cotizan en la Bolsa de Nueva York o en el mercado NASDAQ, y que aplican GAAP. Se encontró que alrededor de cuatro eventos, los CARs fueron significativamente negativos, lo que sugiere que la introducción de IFRS17 podría reducir el valor de las empresas. Sin embargo, también encontró que el mercado tiene percepciones mixtas ya que algunos eventos mostraron retornos anormales positivos. También se llevaron pruebas de robustez, incluyendo cambios en la ventana de tiempo del evento, los cuales confirmaban los resultados. Se realizó un análisis de regresión lineal para estudiar cómo ciertas características de las empresas, como su tamaño, la asimetría de la información, los costos de agencia, las oportunidades de crecimiento y la rentabilidad, pueden afectar a los retornos anormales de las aseguradoras alrededor de los eventos de introducción de IFRS 17. Mediante el análisis de sección cruzada encontraron que el tamaño de la empresa y las oportunidades de crecimiento están negativamente relacionados con los CARs, lo que sugiere que las empresas grandes y con menos oportunidades de crecimiento experimentan retornos anormales más negativos. En resumen, el estudio sugiere que la introducción de IFRS17 está asociada con retornos anormales negativos para las aseguradoras, lo que indica que los accionistas no aprecian la valoración de las obligaciones de los contratos de seguros al valor actual.

No obstante, Dahiyat y Owais (2020), presentan otro análisis centrado en el impacto del mismo estándar contable evaluando la calidad de los informes financieros de las compañías de seguros jordanas. Este estudio emplea un enfoque exploratorio seguido de un enfoque descriptivo inferencial. Se recopilaron datos a través de 120 cuestionarios de más de 24 compañías de seguros analizando la validez y fiabilidad a través del alfa de Cronbach, con el fin de medir la estabilidad de la consistencia interna de los párrafos del cuestionario del estudio. También se utilizó el Test de Kolmogorov-Smirnov para evaluar la normalidad de los datos recopilados, obteniendo que estos seguían una distribución Normal. Asimismo, se calcularon la media, la desviación estándar, el rango, el peso relativo y la importancia relativa para comprender la dirección de opiniones. Los resultados demostraron un impacto significativo de la aplicación de IFRS17 en la calidad de los informes financieros, particularmente en la comparabilidad, representación fiel y relevancia.

4. ESTUDIO ACTUARIAL DE LA NORMATIVA IFRS17

Tal y como se ha mencionado anteriormente, en el ámbito de la normativa IFRS17, se visualizan tres enfoques distintos para medir las obligaciones derivadas de los contratos

de seguros: el método general o por bloques, el enfoque de asignación de primas y el modelo de tarifa variable. En este apartado se lleva a cabo una implementación práctica de metodologías actuariales conforme la normativa IFRS17, específicamente centrándose en el método general o por bloques. Para ello, se elige una cartera de seguros de vida ahorro como objeto de análisis, dado que, al ser seguros con una duración superior al año, este enfoque resulta idóneo. El apartado se centra en la aplicación de diversas metodologías actuariales para la estimación y cálculo de elementos tales como la agregación de contratos de seguros, la evaluación de flujos de efectivo, la tasa de descuento, el ajuste por riesgo hasta alcanzar la determinación del CSM o Loss Component según proceda. Todo este proceso tiene como propósito fundamental la medición de las responsabilidades originadas por los contratos de seguro. Antes de proceder a ello, se realizará a continuación un análisis descriptivo y de calidad de los datos, como parte de un estudio exhaustivo y riguroso.

4.1. Base de datos y análisis descriptivo

La base de datos original contiene información detallada de 76.102 asegurados, distribuida en un total de 15 variables distintas. Para una mejor comprensión y análisis de esta información, se presenta a continuación la Tabla 5 que describe de manera específica la definición de cada variable. Esta disposición pretende facilitar la interpretación y el estudio de los datos contenidos en la base de datos, permitiendo una exploración más detallada y precisa de la cartera de asegurados y sus características asociadas.

Nombre original de la variable	Descripción
ID	Código de identificación único del asegurado (un número entero).
Gender (Género)	Género del asegurado (M=masculino, F=femenino).
Birth_Date (Fecha de nacimiento)	Fecha de nacimiento del asegurado (DD/MM/AAAA).
Effective_Date (Fecha efectiva)	Fecha de inicio efectiva de la relación contractual entre la compañía de seguros y el asegurado (DD/MM/AAAA).
Capital	Capital en riesgo en caso de fallecimiento (en €).
Renewal_Date (Fecha de renovación)	Fecha de renovación durante el año del contrato de seguro (DD/MM/AAAA).
Age (Años)	Edad exacta del asegurado, medida en años, en la fecha de renovación (un número real). El tiempo transcurrido entre la fecha de nacimiento y la fecha de renovación.
T	Edad fraccional en años del asegurado en el momento de la Fecha de Renovación. Número de años transcurridos desde su última fecha de cumpleaños hasta la fecha de renovación (un número real entre 0 y 1).

Age_Actuarial	Edad del asegurado, medida en años enteros, obtenida aproximando su edad exacta al entero más cercano (número entero).
Date_Birthday	Fecha de cumpleaños en 2009 (DD/MM/AAAA).
X	Tiempo transcurrido en años entre el inicio del año (0:00 AM del 1 de enero) y el momento de la Fecha de Renovación (un número real entre 0 y 1).
R	Trimestre estacional (1=1T, 2=2T, 3=3T, 4=4T). Trimestre de edad del asegurado en el momento de la renovación.
S	Mes (1=Enero, 2=Febrero, 3=Marzo, 4=Abril, 5=Mayo, 6=Junio, 7=Julio, 8=Agosto, 9=Septiembre, 10=Octubre, 11=Noviembre, 12=Diciembre).
Age_actuarial_quarter	Edad del asegurado, medida en años, después de aproximar su edad exacta al trimestre de edad más cercano (un número real).
Month	Mes de renovación (1=Enero, 2=Febrero, 3=Marzo, 4=Abril, 5=Mayo, 6=Junio, 7=Julio, 8=Agosto, 9=Septiembre, 10=Octubre, 11=Noviembre, 12=Diciembre).

Tabla 5: Variables y descripción de variables del Portafolio de asegurados.

Fuente: Elaboración propia.

Del total de las 15 variables en primer lugar se observa el identificador único (ID) asignado a cada asegurado no aportando información sustancial para el análisis y se considera una variable nominal sin relevancia para el estudio.

A continuación, se pueden observar diversas variables cualitativas cuya finalidad es categorizar los datos, como son el sexo, la fecha de nacimiento del asegurado, la fecha efectiva del contrato, la fecha de renovación, la fecha de cumpleaños, el trimestre de edad del asegurado en el momento de la renovación, el trimestre estacional y el mes de renovación.

Finalmente, pueden verse variables numéricas. En concreto son variables de este tipo el capital en riesgo en caso de fallecimiento, la edad exacta del asegurado medida en años en la fecha de renovación, la edad fraccional en años del asegurado en el momento de la fecha de renovación, y la edad del asegurado medida en años después de aproximar su edad exacta al trimestre de edad más cercano.

Para las variables numéricas se pueden calcular los descriptivos básicos (medidas de tendencia central) como son la media, el valor mínimo y máximo, los cuartiles y la desviación estándar. Estos descriptivos básicos se calcularán, primer de todo, para todo el Portafolio que comprende contratos con fechas efectivas que se extienden desde el año 2000 hasta el 2009 ofreciendo el lector una visión global de la base de datos. Estos mismos descriptivos se calcularán para aquellos contratos de seguros cuya fecha efectiva es del año 2000 ya que el estudio realizado en este trabajo se enfoca especialmente en este grupo de contratos de seguro. Seguidamente pueden verse los resultados obtenidos de los descriptivos en las Tablas 6 y 7 para el portafolio general y portafolio con contratos cuya fecha efectiva es el año 2000 que siguen:

Todo el Portafolio						
Estadístico	Capital (€)	Age (años)	T	Age Actuarial (años)	x	Age_actuarial_quarter (años)
Media	91.452,18	44,482	0,498	44,478	0,509	44,358
Mínimo	7.000	17,892	0	18	0	17
25%	54.000	38,59	0,248	39	0,241	38,5
50%	80.000	45,144	0,494	45	0,493	45
75%	110.000	51,959	0,75	52	0,797	51,75
Máximo	3.010.000	78,845	0,999	79	0,997	78,75
Desviación Estándar	64.302,20	10,836	0,289	10,841	0,304	10,84

Tabla 6: Estadísticos básicos del Portafolio de asegurados.
Fuente: Elaboración propia.

Año 2000						
Estadístico	Capital	Age	T	Age Actuarial	x	Age_actuarial_quarter
Media	88.722,71	48,75	0,493	48,732	0,666	48,622
Mínimo	7.000	18,66	0	19,00	0,178	18,75
25%	52.000	42,81	0,241	43,00	0,471	42,75
50%	75.000	48,66	0,482	49,00	0,71	48,5
75%	110.000	55,33	0,746	55,00	0,893	55,25
Máximo	605.000	78,84	0,999	79,00	0,997	78,75
Desviación Estándar	59.419,98	9,52	0,291	9,519	0,238	9,52

Tabla 7: Estadísticos básicos de los datos de los contratos de seguro cuya fecha efectiva es el año 2000.
Fuente: Elaboración propia.

Observando las Tablas 6 y 7 si se comparan los resultados obtenidos de los descriptivos básicos del portafolio general¹¹ y el grupo específico de contratos cuya fecha efectiva es el año 2.000 se puede destacar:

- La media de la variable Capital (€) es más baja en contratos del año 2.000 que en el portafolio general, siendo 91.452,18€ y 88.722,71€ respectivamente. La dispersión es mayor en el portafolio general (64.302,20€ frente a 59.419,98€). El valor máximo de capital es significativamente mayor en el portafolio general (3.010.000€ frente 605.000€) mientras que ambos grupos comparten el mismo valor mínimo de capital (7.000€). Los cuartiles muestran una distribución más amplia en el portafolio general, con una mediana ligeramente mayor (80.000€ en el portafolio general frente a 75.000€ en el grupo del año 2.000).
- En el grupo de contratos del año 2000, las variables de edad (Age, Age Actuarial y Age Actuarial Quarter) muestran valores medios más bajos y menos variabilidad que en el portafolio general (alrededor de 45 años vs. 49 años y 9,52 años vs. 10,84 años respectivamente). Los cuartiles en el portafolio general reflejan una

¹¹ Ver 9.2. Anexos análisis descriptivo – Pág. 83.

distribución más amplia, con una mediana ligeramente superior (alrededor de 49 años vs. 45 años). Los valores del cuartil 25% están aproximadamente en 39 años y el cuartil 75% en 52 años en el portafolio general, mientras que en el grupo del año 2000, estos cuartiles se sitúan en 43 años y 55 años respectivamente.

- Para la variable T en el grupo de contratos del año 2000, la media es de aproximadamente 0,493, con una desviación estándar de alrededor de 0,291. Los valores mínimo y máximo son 0 y 0,999 respectivamente, con cuartiles en 0,241 y 0,746. En contraste, en el portafolio general, la media de T es de aproximadamente 0,498, con una desviación estándar de alrededor de 0,289. Los valores mínimo y máximo de T son 0¹² y 0,999¹³ respectivamente, con cuartiles en 0,248 y 0,75. Esto sugiere que, aunque la distribución de la fracción de edad en ambos grupos es similar en términos de rango, los cuartiles en el grupo del año 2000 están ligeramente desplazados hacia valores más altos. La variable T representa la fracción de año transcurrida desde el último cumpleaños del asegurado hasta la fecha de renovación. Observando los resultados de los estadísticos más importantes de destaca que no existe una tendencia significativa hacia la renovación en una parte específica del año, es decir, las fechas de renovación de las pólizas en relación con los cumpleaños de los asegurados están distribuidas de forma equilibrada a lo largo del año.
- Para la variable x, en el grupo del año 2000, la media es de aproximadamente 0,666, con una desviación estándar de alrededor de 0,238. Los valores mínimo y máximo son 0,178 y 0,997 respectivamente, con cuartiles en 0,471 y 0,893. En contraste, en el portafolio general, la media de x es de aproximadamente 0,509, con una desviación estándar de aproximadamente 0,304. Los valores mínimo y máximo de x son 0 y 0,997 respectivamente, con cuartiles en 0,241 y 0,797. Esto indica que la distribución del tiempo transcurrido en el año hasta la renovación es similar en términos de rango en ambos grupos, pero los cuartiles en el grupo del año 2000 están ligeramente desplazados hacia valores más altos. Respecto a la variable x, la cual nos indica el tiempo transcurrido en años entre el inicio del año y el momento de renovación permite llegar a la conclusión de que las fechas de renovación están distribuidas de forma equilibrada a lo largo del año.

Para las variables categóricas pueden observarse como estadísticos descriptivos la frecuencia absoluta y relativa (Tabla 8) de cada categoría, lo que proporciona una visión detallada de la prevalencia de cada grupo.

¹² El hecho de existir un valor mínimo de 0 indica que hay algún asegurado en el que la fecha de renovación coincide con su cumpleaños.

¹³ El valor máximo de 0,999 indica que hay asegurados cuya fecha de renovación está casi un año después de su último cumpleaños.

		Todos los años		Año 2000	
		Frecuencias absolutas	Frecuencias relativas	Frecuencias absolutas	Frecuencias relativas
Sexo	M	47,652	0,626	1091	0,557
	F	28,450	0,374	869	0,443
Ageing quarter(1	18,903	0,248	505	0,258
	2	19,307	0,254	496	0,253
	3	19,003	0,250	511	0,261
	4	18,889	0,248	448	0,229
Seasonal quarter	1	20,323	0,267	98	0,050
	2	19,023	0,250	467	0,238
	3	14,888	0,197	529	0,270
	4	21,868	0,287	866	0,442
Renewal Month (Month)	1	5,988	0,079	0	0
	2	6,155	0,081	0	0
	3	7,314	0,096	81	0,041
	4	6,089	0,080	130	0,066
	5	6,114	0,080	155	0,079
	6	6,735	0,089	166	0,084
	7	6,437	0,085	222	0,113
	8	4,368	0,057	139	0,071
	9	4,517	0,060	185	0,094
	10	5,679	0,075	201	0,103
	11	4,517	0,093	335	0,171
	12	4,368	0,126	346	0,177

Tabla 8: Frecuencias absolutas.
Fuente: Elaboración propia.

Si se visualizan la Tabla 8 en el análisis del Portafolio general, se observa que la mayoría de los contratos están asociados con hombres (47,652 contratos, o el 62,6%), mientras que las mujeres representan 28.450 contratos (37,4%). Respecto al trimestre de envejecimiento, los contratos están distribuidos de manera relativamente equilibrada entre los cuatro trimestres, con una ligera predominancia en el segundo trimestre (19,307 contratos, o el 25,4%). Por otro lado, la distribución por trimestre estacional muestra una concentración significativa en el cuarto trimestre (21,868 contratos, o el 28,7%), siendo el primer trimestre el que tiene menor concentración. Al observar los meses de renovación, se nota una mayor concentración en el tercer trimestre (julio a septiembre) y el menor número de renovaciones en agosto y noviembre.

En el análisis del grupo formado por los contratos del año 2000, la proporción de hombres y mujeres es más equilibrada comparada con la del Portafolio general, con hombres representando 1.091 contratos (55,7%) y mujeres 869 contratos (44,3%). En cuanto a los trimestres de envejecimiento, la distribución es también relativamente equilibrada, con una ligera preponderancia en el tercer trimestre (511 contratos, o el 26,1%). Sin embargo, se observa una menor concentración de contratos en el cuarto trimestre en comparación

con el Portafolio general. La distribución por trimestres estacionales muestra una concentración significativa en el cuarto trimestre (866 contratos, o el 44,2%), lo cual es consistente con la tendencia observada en el Portafolio general. Para los meses de renovación en el año 2000, se observa una mayor concentración en noviembre y diciembre, con 335 contratos (17,1%) y 346 contratos (17,7%) respectivamente, mientras que no hubo renovaciones en los meses de enero y febrero.

Con el fin de comprender la estructura y distribución de los datos y centrandolo el análisis a grupo formado por los contratos cuya fecha efectiva es el año 2000, se presenta la visualización de los mismos en un gráfico de dispersión (Gráfico 1). El gráfico de dispersión es una herramienta visual que permite examinar la relación entre dos variables representando cada punto como una observación en el espacio de dos variables. La posición del punto está determinada por los valores de las dos variables que estamos comparando.

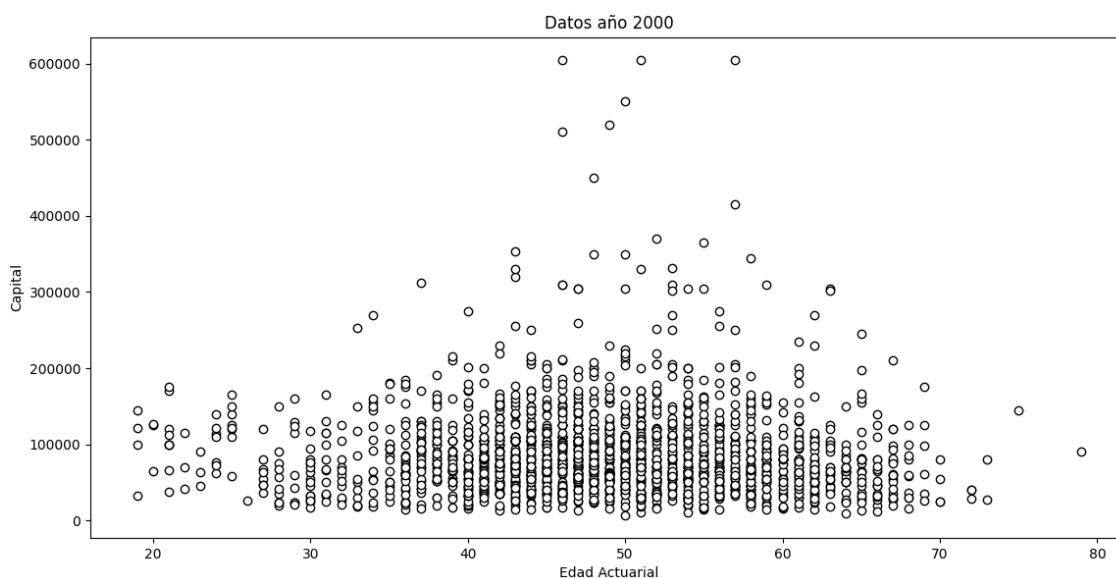


Gráfico 1: Gráfico de dispersión variables capital y edad actuarial para el año 2000.
Fuente: Elaboración propia.

El Gráfico 1 muestra el gráfico de dispersión que relaciona la variable Capital (eje Y) y la variable Edad Actuarial (eje X). Se puede observar un agrupamiento de puntos en los capitales bajos situados entre 0 y 200.000 €. También se observa un agrupamiento de puntos entre la edad actuarial 30 y 70. Por tanto, se puede concluir que gran parte de los asegurados tienen entre 30 y 70 años de edad actuarial y con capitales que se sitúan entre 0 y 200.000 €. Además, pueden observarse outliers, cuyo capital está por encima de 400.000 €.

Para investigar si existe alguna relación entre Capital y Age Actuarial se calcula la correlación que es una medida estadística que mide cuanto relacionadas están dos variables. La correlación varía entre -1 y 1: 1 indica una correlación positiva perfecta¹⁴,

¹⁴ Correlación positiva perfecta : a medida que una variable aumenta, la otra también lo hace.

-1 indica una correlación negativa perfecta¹⁵, y 0 no hay relación entre las dos variables. Los test más conocidos para calcular correlaciones son el test de Kendall¹⁶, el test de Spearman¹⁷ y el test de Pearson¹⁸.

Morales y Rodríguez (2016) indican que antes de realizarse el análisis de correlación debe evaluarse la Normalidad de las variables del estudio. Aplicar técnicas basadas en la Normalidad sin verificar previamente esta condición puede llevar a conclusiones erróneas (Delgado de la Torre, 2007). Esta normalidad puede evaluarse a través de métodos gráficos y contrastes de hipótesis. A través de métodos gráficos puede realizarse mediante el qqplot o gráfico Quantile-Quantile. Por lo que se refiere a los contrastes para evaluar si los datos se distribuyen según una Normal existen varios tipos de test entre los cuales se destacan la prueba de Kolmogorov-Smirnov¹⁹, el test de Lilliefors (Kolmogorv-Smirnov modificado), Anderson Darling²⁰, test bondad de ajuste de la Chi Cuadrado²¹ y Shapiro-Wilk²². Finalmente se decide aplicar los Tests Lilliefors (Kolmogorv-Smirnov modificado) ya que Age_Actuarial se trata de una variable discreta con empates y Capital es una variable continua pero también contiene algunos empates.

Seguidamente se realiza en los Gráficos 2 y 3 el análisis gráfico de la Normalidad a través de Q-QPlot o Quantile-Quantile. Este gráfico permite comparar la distribución de datos con una teórica. En este contexto, se elige comparar los datos con la distribución Normal.

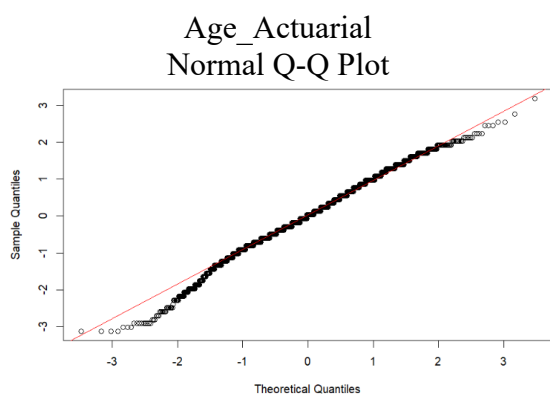


Gráfico 2: Q-Q Plot de la variable Age_Actuarial.
Fuente: Elaboración propia.

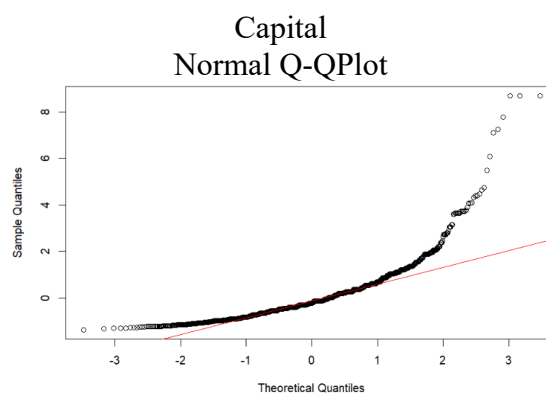


Gráfico 3: Q-Q Plot de la variable Capital.
Fuente: Elaboración propia.

Al examinar el Gráfico 2 se observa que la variable Age_Actuarial tiene un ajuste aceptable en el centro del gráfico, sin embargo, existen desviaciones en las colas. De igual manera sucede en el Gráfico 3 relacionado con la variable Capital, cuyo ajuste a la Normal

¹⁵ Correlación negativa perfecta: a medida que una variable aumenta, la otra disminuye.

¹⁶ Kendall: se aplica en variables cualitativas y aquellas las cuales no siguen una distribución normal. Se trata en una prueba no paramétrica basada en rangos basada en el conteo de concordancias y discordancias entre pares e puntos, es decir, no se basa en datos originales. Se realizan un total de $n(n-1)/2$ posibles comparaciones de n pares de datos. Todo ello lo hace robusto ante valores atípicos.

¹⁷ Spearman: también se basa en rangos de los valores, pero en general se aplica en variables cuantitativas que no siguen un comportamiento normal. Igual que Kendall, es una medida no paramétrica. Para calcular Spearman los datos son en primer lugar transformados en rangos y luego se aplica la fórmula sobre dichos rangos lo que hace que sea robusto ante valores atípicos.

¹⁸ Pearson: muy parecido al de Spearman. La principal diferencia entre ambos es que en el caso del coeficiente de Pearson se calcula a partir de los valores originales de las variables y se asume que las variables están distribuidas Normalmente.

¹⁹ El test de Kolmogorov Smirnov puede llevar a conclusiones erróneas si hay empates.

²⁰ El test de Anderson Darling está diseñado básicamente para datos continuos.

²¹ El test de Chi Cuadrado, en este caso sería necesario agrupar los datos en intervalos.

²² Se descarta la aplicación del contraste de Shapiro-Wilk ya que es para muestras de tamaño pequeño ($n \leq 30$) mientras que para el año 2000 en este portafolio hay 1.960 asegurados (Delgado de la Torre, 2007).

es peor, observando desviaciones en ambas colas y en concreto, se observa una desviación muy significativa en la cola derecha del gráfico.

Tal y como se ha comentado anteriormente, para saber si los datos siguen una distribución Normal se procede a calcularse el contraste estadístico Lilliefors (Kolmogorv-Smirnov modificado) destinado a muestras de cualquier tamaño, aplicable tanto en variables continuas como discretas y también adecuado en caso de que en las variables contengan empates (Tabla 9). La hipótesis nula para este contraste es que los datos siguen una distribución Normal (Delgado de la Torre, 2007).

Age Actuarial		Capital	
Lilliefors (Kolmogorv-Smirnov modificado)		Lilliefors (Kolmogorv-Smirnov modificado)	
D	p-Valor	D	p-Valor
0,039199	2,169e-07	0,12698	<2,2e-16

Tabla 9: Contraste Lilliefors (Kolmogorv – Smirnov modificado) para las variables Age_Actuarial y Capital.
Fuente: Elaboración propia.

Los resultados de los test Lilliefors (Kolmogorv-Smirnov modificado) presentados en la Tabla 9 si se elige el nivel de significación 0,05 confirman que las dos variables no siguen una distribución Normal ya que el p-valor para ambas variables es inferior a este valor. Con todo ello, se concluye que solo tiene sentido aplicar la correlación Tau kendall y correlación Spearman ya que son los dos asumen no Normalidad en los datos.

Variables estudio: Capital y Age Actuarial Año 2000			
Correlación Tau Kendall		Correlación Spearman	
Statistic	p-Valor	Statistic	p-Valor
-0,041	0,008	-0,061	0,007

Tabla 10: correlación de Kendall y Spearman de los datos de los contratos de seguro cuya fecha efectiva es el año 2000.
Fuente: Elaboración propia.

Los resultados mostrados en la Tabla 10 permiten concluir que la correlación entre Capital y Age_Actuarial en el año 2000 es débilmente negativa (-0,041 para Tau de Kendall y -0,061 para Spearman) pero estadísticamente significativa (p-valores de 0,008 y 0,007 respectivamente), indicando una ligera tendencia a que el capital disminuya con el aumento de la edad actuarial.

4.2. Método general o por bloques (BBA)

En este apartado, se emplea el método general o por bloques previsto en la normativa de IFRS17 desde una perspectiva estadística. Este enfoque se utiliza específicamente en

contratos que tienen una duración superior a un año. Por consiguiente, se ha considerado que la mejor manera de aplicar este método es a través de un ejemplo práctico que involucra una cartera compuesta por contratos de vida, los cuales suelen tener una duración superior. En este ejemplo ilustrativo, tal y como se comentará más adelante, el beneficiario recibe un pago establecido cuando el asegurado fallece.

4.1.1. Agregación de contratos de seguros

En primer lugar, la normativa IFRS17, específicamente en el párrafo 14, indica que los contratos de seguros deben ser agrupados en carteras y administrados de forma conjunta considerando riesgos similares, es decir, riesgos homogéneos. Para ilustrar este punto, la normativa indica a modo de ejemplo, que contratos de la misma línea de producto tienen riesgos similares. Por consiguiente, se ha seleccionado como línea de producto un seguro de vida, en el que, tal y como se ha comentado anteriormente, el beneficiario cobrará un capital en caso de fallecimiento del asegurado.

Tal y como establece la normativa (párrafo 22) se ha llevado a cabo una clasificación previa según el año en el que el contrato de seguro ha sido emitido. Por tanto, los datos del *dataset* han sido agrupados por años desde 2000 hasta el 2009, según el año de emisión del contrato tal y como puede verse en la Tabla 11.

Año	Nº de datos
2000	1960
2001	3248
2002	3875
2003	5128
2004	5831
2005	8231
2006	9700
2007	11899
2008	15046
2009	11184
	76102

Tabla 11: Número de asegurados por año.
Fuente: Elaboración propia.

Al tratarse de contratos de vida, se seleccionan como las dos variables más importantes a estudiar la edad del asegurado y el capital. A partir de estas dos variables se propone aplicar técnicas de clúster con el fin de realizar los grupos de contratos con riesgos similares que prevé la normativa IFRS17. Debido a la extensión del trabajo se llevará a cabo el análisis de clúster detallado solo para el grupo de asegurados cuya fecha efectiva del contrato del seguro es del año 2000. Aun así, es importante destacar que el procedimiento aplicado en este apartado puede extrapolarse para el resto de contratos de seguros cuya fecha efectiva es de otros años distintos al 2000.

4.1.2. Técnicas de Clúster

Con el fin de clasificar los contratos con riesgos similares se pueden utilizar varias técnicas de análisis de clúster de riesgos. Murphy (2012) definió clúster como un procedimiento estadístico multivariante que reorganiza un conjunto de datos en grupos homogéneos llamados clústeres. También se puede definir como *clustering* la clasificación no supervisada de patrones (observaciones, elementos de datos o vectores de características) en grupos o clusters (Jain et al., 1999; Møllersen et al., 2016; Dalmaijer et al., 2022; Liao, 2005; Devale y Vulkarni 2012; Oyewole et al., 2022). El aprendizaje no supervisado se caracteriza por no tener etiquetas o categorías predefinidas, es decir, el algoritmo de clúster busca patrones y similitudes intrínsecas en los datos para agruparlos en clústeres sin información previa sobre las categorías. Según Murphy (2012) una vez obtenida la matriz de similitud por ejemplo llamada S, se puede transformarla en una matriz de disimilitud aplicando cualquier función monótonamente decreciente, por ejemplo, $D = \max(S) - S$. Una de las formas estadísticas más comunes de definir la disimilitud entre datos es:

$$\Delta(x_i, x_{i'}) = \sum_{j=1}^D \Delta_j(x_{ij}, x_{i'j})$$

Expresión 1: Definición de disimilitud entre datos.
Fuente: Murphy (2012).

En esta fórmula, $\Delta(x_i, x_{i'})$ representa la disimilitud total entre los puntos de datos x_i y $x_{i'}$. La disimilitud se calcula como la suma de las disimilitudes individuales desde $j=1$ hasta $j=D$, donde D es el número total de dimensiones de los datos (variables). La parte de dentro del sumatorio, $\Delta_j(x_{ij}, x_{i'j})$ mide la diferencia entre las coordenadas x_{ij} y $x_{i'j}$.

Una vez decidido que se aplicarán técnicas de clúster para la agrupación de los contratos de seguros en riesgos similares, se ha investigado sobre que técnicas serían más óptimas para la normativa estudiada. De forma resumida, se ha observado que los algoritmos de *clustering* pueden clasificarse en jerárquicos, particionales, basados en densidad²³, basados en cuadrícula, espectrales y difusos²⁴. Siendo el objetivo la clasificación de cada contrato a un solo grupo, finalmente se opta para la aplicación del algoritmo K-Means, el algoritmo K-Medoids y el algoritmo jerárquico aglomerativo.

Además, en la aplicación de los algoritmos de clúster se pueden utilizar muchos tipos de medidas de asociación entre variables y entre individuos (asegurados).²⁵ En este trabajo, después de realizar numerosas pruebas con el fin de que los grupos fueran lo más equilibrados posible en relación al número de asegurados perteneciente a cada uno de

²³ Ver 9.2.5. Density-Based Spatial Clustering of Applications with Noise (DBSCAN). – Pág. 103.

²⁴ Ver 9.2.4. Fuzzy clustering C-means. - Pág. 101.

²⁵ Algunas de las medidas de asociación entre variables más comunes es el coeficiente de correlación y en el caso de datos dicotómicos la medida de Russell y Rao y la medida de Jaccard. En el caso de las medidas de asociación entre individuos son comunes la distancia euclídea, de Minkowski y de Mahalanobis y las distancias derivadas de la distancia chi-cuadrado (Jain, A.K., et al., 1999).

ellos, se ha decidido aplicar la distancia Euclídea y en concreto, para el algoritmo jerárquico aglomerativo, el enlace de Ward.

$$\text{Distancia Euclídea} = \sqrt{\sum_{i=1}^p (Y_{i1} - Y_{i2})^2}$$

Expresión 2: Distancia Euclídea entre dos individuos para el conjunto de p variables analizadas.

Por lo que se refiere a los algoritmos de clúster que finalmente se desean aplicar, pueden clasificarse según el output de salida en métodos no jerárquicos o *flat clustering* o *partitional clustering* como es el caso de K-means y K-Medoids y, en métodos jerárquicos o *hierarchical clustering* en el caso del algoritmo jerárquico aglomerativo. Una de las diferencias más relevantes entre ambos tipos, y que es conveniente remarcar, es que los algoritmos de clustering jerárquico son deterministas, es decir, no es necesario indicar el número de clústeres (K) ²⁶ mientras que los algoritmos de *flat clustering* o *partitional clustering* son sensibles al número de clústeres (K).

Finalmente, es necesario remarcar que antes de aplicar cualquier metodología se ha llevado a cabo la estandarización de los datos²⁷ ya que, si bien no es algo imprescindible, se hace muy recomendable con el fin de que las variables con escalas y medidas más grandes no influyan de forma desproporcionada a las medidas de distancia utilizadas para el cálculo de la similitud entre observaciones.

4.1.1.1. K means

En primer lugar, se ha elegido aplicar el algoritmo de clúster K-means siendo el algoritmo de aprendizaje no supervisado más conocido (Sinaga et al., 2020; Kiermayer et al., 2019) cuyo output es de tipo *flat clustering* o *partitional clustering*. En este tipo de algoritmo es necesario determinar el número más óptimo de clústeres antes de su ejecución (Sekula et al., 2017; Rodriguez et al., 2019; Baidari y Patil, 2020). Para la determinación del número de clústeres k pueden usarse varias metodologías como son el método del codo, los criterios de información, la teoría de la información y el método de la silueta y validación cruzada, entre otros (Kodinariya y Prashant, 2013). En este trabajo se ha decidido utilizar la técnica del método del codo o *elbow method* por ser el más simple, ágil y común a aplicar en este tipo de algoritmo (Syakur et al., 2018).

El método del codo o *Elbow method* consiste en seleccionar de entrada un rango de número de clústeres posibles k . En cada etapa del bucle los centroides irán cambiando de posición intentando estar lo más alejados posible entre sí, asociando cada dato al centroide más cercano mediante la ejecución del algoritmo *k-means* (Hardy, 1996; Saputra, 2019). Por último, en cada iteración de la ejecución se calcula la suma de las distancias al cuadrado entre cada punto de datos y el centroide del clúster correspondiente (WCSS). Si se seleccionan $j=3$ clústeres, la ecuación para calcular el WCSS sería la que se muestra en la Expresión 3 que sigue:

²⁶ La variables introducidas por el usuario reciben el nombre de hiperparámetro.

²⁷ Ver 9.2.1. Anexos K-Means – Pág. 94.

$$WCSS = \sum_{P_i \text{ cluster } 1} \text{distancia}(P_i, C_1)^2 + \sum_{P_i \text{ cluster } 2} \text{distancia}(P_i, C_2)^2 + \sum_{P_i \text{ cluster } 3} \text{distancia}(P_i, C_3)^2$$

Expresión 3: Cálculo suma de las distancias al cuadrado de cada punto y el centroide del clúster al que pertenece $k=3$.
Fuente: Elaboración propia.

Realizando los cálculos se obtiene el Gráfico 4 que se presenta a continuación:

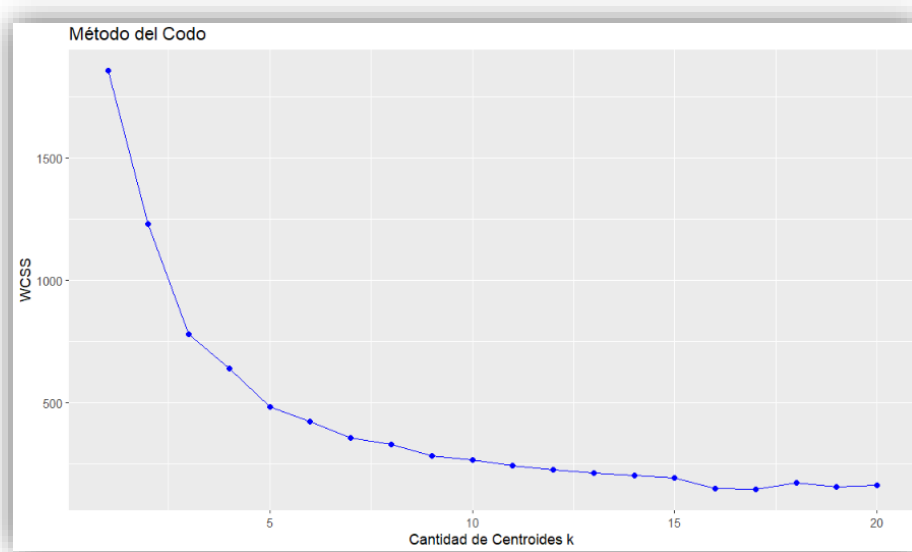


Gráfico 4: Método del Codo.
Fuente: Elaboración propia.

Tal y como se puede ver en el Gráfico 4 a medida que k aumenta el valor WCSS disminuye, ya que los datos están más cerca de los centroides. Sin embargo, para $k=3$ el WCSS deja de disminuir significativamente. En este punto se considera que se ha alcanzado un equilibrio entre la precisión de los clústeres y la simplicidad del modelo (Nainggolan et al., 2019 y Cui, 2020). Por tanto, $k=3$ se considerará el punto del “codo” y un buen candidato par el número óptimo de clústeres en el algoritmo *K-means*.

Una vez determinado el número de clústeres ya puede aplicarse el algoritmo *K-Means*. El algoritmo *K-Means* permite encontrar patrones o estructuras intrínsecas en el conjunto de los datos a través del cálculo de distancias (proximidad de las observaciones) y en IFRS17 puede usarse como método de clasificación de los asegurados que tengan características similares en relación a las dos variables seleccionadas para estudiar en grupos (Khan y Ahmad, 2004). El algoritmo *K-Means* programado realiza los siguientes pasos (Shutaywi, 2021):

- 1) Selección aleatoria de los asegurados, tantos como clústeres se hayan definido.
- 2) Asignación de cada asegurado al punto más cercano.

$$z_n^* = \operatorname{argmin}_k \|x_n - \mu_k\|_2^2$$

Expresión 4: Asignación de la observación a un clúster mediante el algoritmo K-Means.

Fuente: Murphy (2022).

Donde:

- z_n^* indica qué centroide pertenece cada asegurado.
 - La función argmin que encuentra el valor k que minimiza la distancia entre el punto x_n , y el centroide μ_k .
 - La parte $\|x_n - \mu_k\|_2^2$ calcula la distancia euclidiana al cuadrado entre x_n y μ_k , lo que implica la suma de diferencias al cuadrado entre las coordenadas de ambos puntos.
- 3) Determinación de los centroides de cada uno de los grupos creados mediante el cálculo de la media de todos los puntos asignados a cada clúster.

$$\mu_n = \frac{1}{N_k} \sum_{n:z_n=k} x_n$$

Expresión 5: Cálculo de los centroides.

Fuente: Murphy (2022).

Donde:

- μ_n es el centroide del clúster n , que representa el “centro” o la ubicación media de todos los puntos asignados a ese clúster.
 - $\frac{1}{N_k}$ sirve para calcular el promedio ponderador de los puntos para determinar el centroide.
 - $\sum_{n:z_n=k} x_n$ es la suma de las coordenadas de todos los puntos asignados al clúster k para calcular su centroide.
- 4) Reasignación de los asegurados a los nuevos centroides.
5) Se realizan las reasignaciones tantas veces como sea necesario.

Seguidamente se procede a aplicar el algoritmo a los datos correspondientes al año 2000 utilizando la biblioteca de Python *Sklearn*, la cual, en el caso de no indicarse nada en concreto, utiliza la distancia euclídea por defecto para el cálculo de distancias de proximidad entre los datos. Antes de la ejecución del algoritmo, tal y como se ha mencionado anteriormente, debe incluirse el hiperparámetro relacionado con el número óptimo de clústeres, determinado previamente en el subapartado anterior, y que es igual a 3. De este modo, al aplicar el algoritmo *K-Means* con $K=3$, se ha generado el siguiente Gráfico 5:

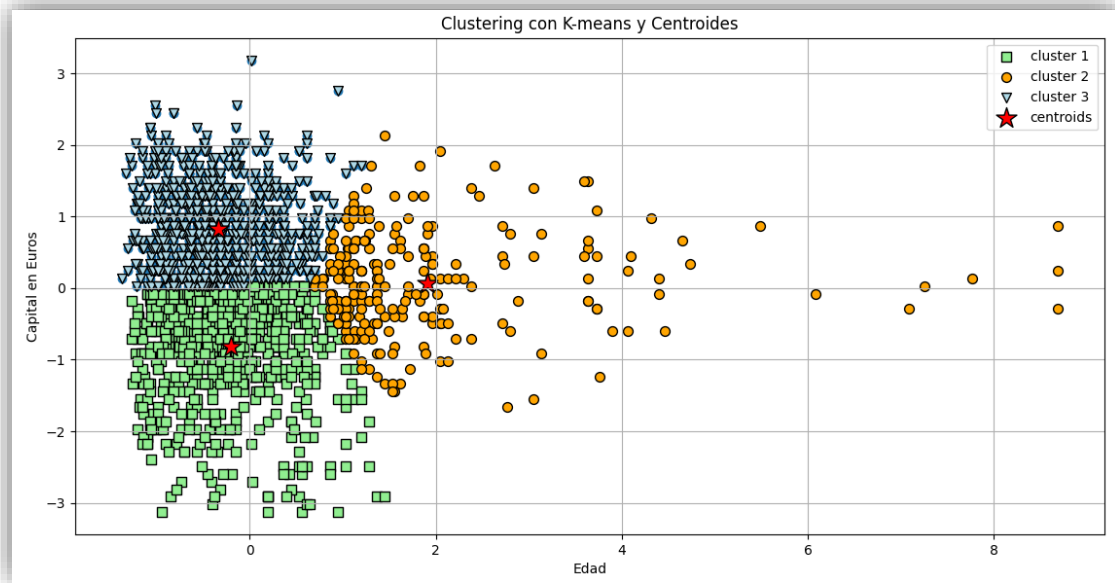


Gráfico 5: Algoritmo k-means y centroides.
Elaboración propia.

En el gráfico representado, se observa la agrupación de los datos en tres centroides. El primer clúster contiene un total de 876 datos, el segundo clúster cuenta con 243 observaciones mientras que el tercer clúster tiene 841 observaciones. Estos grupos se visualizan de forma claramente delineada en el gráfico, cada uno representado con un color único para una visualización más intuitiva.

Entre las observaciones del clúster 2 se pueden observar varios outliers. Si bien en un análisis de clúster sería necesario tratar aisladamente estos outliers y volver hacer el análisis de clúster, en el caso de IFRS17 no debe considerarse ya que el objetivo es que todos los asegurados formen parte de algún clúster.

Las coordenadas precisas de los centroides de cada clúster aparecen en la Tabla 12:

Clúster	Capital	Age Actuarial
1	-0,207	-0,814
2	1,905	0,070
3	-0,335	0,827

Tabla 12: Coordenadas centroides mediante el algoritmo KMeans.
Elaboración propia.

Además, es posible calcular los estadísticos más relevantes en relación a los datos y los centroides de cada clúster. Estos estadísticos proporcionan una visión completa de la distribución y distancia de los datos respecto a los centroides. Los estadísticos que se calcularán son la media, la mediana, la desviación estándar, la varianza, los cuartiles y percentiles, la asimetría y la curtosis para una comprensión exhaustiva de la estructura y dispersión de los de los datos pertenecientes en cada grupo (Tabla 13).

Clúster 1	Edad	Media	40,99	Curtosis	1,23
		Mediana	43,0	Primer cuartil (Q1) Percentil 25	38,0
		Varianza	42,16	Primer cuartil (Q3) Percentil 75	46,0
		Desv. estándar	6,493	Valor máximo	49
		Asimetría	-1,261	Valor mínimo	19
	Capital	Media	76.423,11	Curtosis	-0,58
		Mediana	74.750,0	Primer cuartil (Q1) Percentil 25	51.000,0
		Varianza	1.068.506.243,14	Tercer cuartil (Q3) Percentil 75	102.000,0
		Desv. Estándar	32.688,013	Valor máximo	175.000,0
		Asimetría	0,227	Valor mínimo	13.000,0
Clúster 2	Edad	Media	49,4	Curtosis	-0,05
		Mediana	49,0	Primer cuartil (Q1) Percentil 25	45,0
		Varianza	43,84	Tercer cuartil (Q3) Percentil 75	54,0
		Desv. Estándar	6,62	Valor máximo	69
		Asimetría	0,17	Valor mínimo	33
	Capital	Media	201.891,93	Curtosis	8,65
		Mediana	171.000,0	Primer cuartil (Q1) Percentil 25	155.000,0
		Varianza	6.671.219.582,15	Tercer cuartil (Q3) Percentil 75	210.000,0
		Desv. Estándar	81.677,53	Valor máximo	605.000,0
		Asimetría	2,7015	Valor mínimo	130.000,0
Edad	Media	56,61	Curtosis	-0,18	
	Mediana	56,0	Primer cuartil (Q1) Percentil 25	52,0	
	Varianza	29,89	Tercer cuartil (Q3) Percentil 75	61,0	

Clúster 3		Desv. Estándar	5,47	Valor máximo	79
		Asimetría	0,63	Valor mínimo	49
	Capital	Media	68.834,87	Curtosis	-0,55
		Mediana	65.000,0	Primer cuartil (Q1) Percentil 25	45.000,0
		Varianza	930.159.252,29	Tercer cuartil (Q3) Percentil 75	90.000,0
		Desv. Estándar	30.498,51	Valor máximo	160.000,0
		Asimetría	0,35	Valor mínimo	7.000,0

Tabla 13: Principales estadísticos de los datos de cada clúster hecho con K-Means.
Fuente: Elaboración propia.

La Tabla 13 muestra los descriptivos básicos de las variables Edad Actuarial y Capital de cada uno de los clústeres permitiendo llegar a las siguientes conclusiones:

- Respecto la variable Edad, el clúster 1 muestra las medias y medianas más bajas (40,99 y 43 respectivamente) mientras que el clúster 3 tiene las más altas (56,61 y 56 respectivamente). La varianza menor de los tres clústeres es la del Clúster 3 y la mayor es la del clúster 2. Existe una asimetría negativa de esta variable en el clúster 1 con un sesgo de la distribución hacia la izquierda. El resto de clústeres presentan una asimetría positiva o hacia la derecha. Respecto el clúster 1 también es necesario indicar que hay una curtosis positiva lo que indica que los datos están dispersos respecto a la media.
- En cuanto a la variable Capital, el clúster 3 muestra las medias y medianas más bajas (68.834,87 y 65.000 respectivamente) mientras que el clúster 2 es donde hay la media y mediana más altas (201.891,93 y 171.0000 respectivamente). Las varianzas todas son muy altas, siendo la mayor la del clúster 2 y la menor la del clúster 3. Los tres clústeres tienen asimetrías positivas. Destaca el clúster 2, con una curtosis muy elevada, lo que indica que los resultados están dispersos respecto a la media.

4.1.1.2. K Medoids

Otro de los algoritmos que se propone aplicar y se considera apropiado para IFRS17 y la agrupación de contratos según riesgos similares es K-Medoids²⁸. Este algoritmo a diferencia de K-Means, que se basa en medias, utiliza el valor más central en un clúster como representante lo que hace que no se vea tan influenciado por valores atípicos (Kaufmann y Rousseeuw, 1987). El algoritmo K-Medoids funciona de la siguiente forma:

- 1) Se seleccionan los k medoides iniciales de forma aleatoria. La diferencia básica entre K-Means y K Medoids es que el primero consiste en seleccionar centroides virtuales mientras que K Medoids se basa en que el medoide es un dato real.

²⁸ Ver 9.2.2. Anexos K-Medoids- Pág. 99.

- 2) Se asigna cada observación al cluster que tenga el medoide más cercano.
- 3) Se evalúa cambiar el medoide, que es un dato real, por otro medoide que también es un dato real de forma aleatoria.
- 4) Reasignación de los asegurados a los nuevos centroides.
- 5) Se realizan las reasignaciones tantas veces hasta que no existan mejoras significativas del error cuadrático (SSE), o se satisfaga algún criterio predefinido que indique que el algoritmo ha convergido.

Aplicando el algoritmo K Medoids a los datos se obtiene el Gráfico 6.

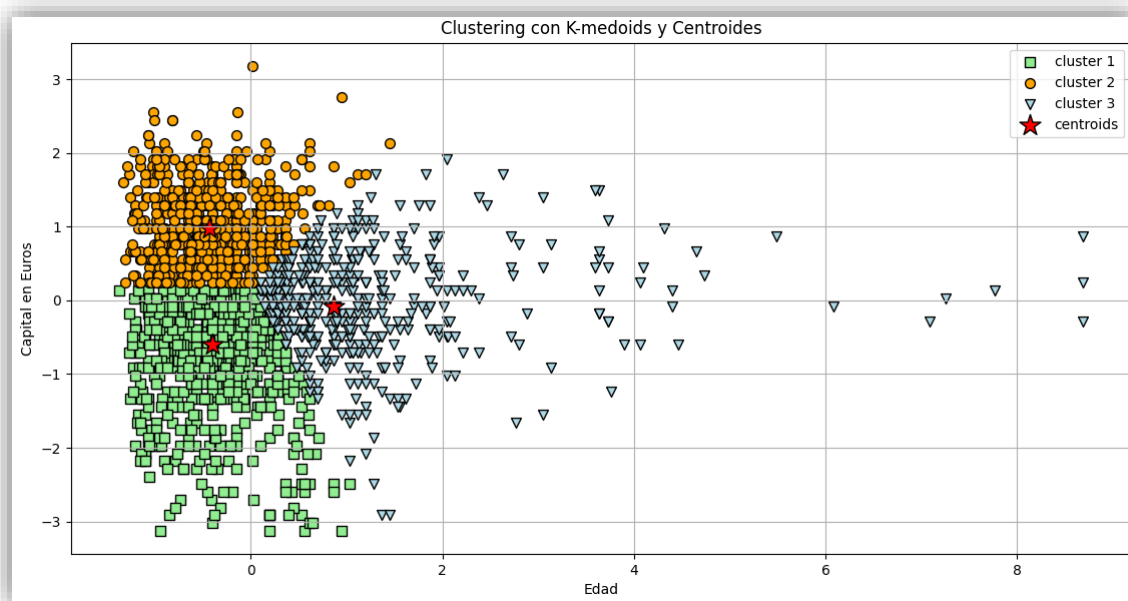


Gráfico 6: Algoritmo k-medoids y centroides.
Fuente: Elaboración propia.

El número de observaciones del clúster número 1 es de 814, para el clúster número 2 es de 650 observaciones y para el clúster número 3 es de 496 observaciones. Las coordenadas obtenidas de los medoides son las siguientes (Tabla 14):

Clúster	Capital	Age_Actuarial
1	-0,399	-0,602
2	-0,433	0,974
3	0,863	-0,077

Tabla 14: Coordenadas centroides mediante el algoritmo K-medoids.
Elaboración propia.

Además, es posible calcular los estadísticos más relevantes en relación a los datos y los medoides de cada clúster. Estos estadísticos proporcionan una visión completa de la distribución y distancia de los datos respecto a los medoides. Los estadísticos que se calcularán son la media, la mediana, la desviación estándar, la varianza, los cuartiles y

percentiles, la asimetría y la curtosis de los datos pertenecientes en cada clúster (Tabla 15).

Clúster 1	Edad	Media	41,37	Curtosis	0,94
		Mediana	43,0	Primer cuartil (Q1) Percentil 25	38,0
		Varianza	46,21	Primer cuartil (Q3) Percentil 75	46,0
		Desv. estándar	6,8	Valor máximo	50
		Asimetría	-1,12	Valor mínimo	19
	Capital	Media	65.763,16	Curtosis	-0,42
		Mediana	65.000,0	Primer cuartil (Q1) Percentil 25	45.250,0
		Varianza	692.088.031,12	Tercer cuartil (Q3) Percentil 75	82.350,0
		Desv. Estándar	26.307,57	Valor máximo	150.000,0
		Asimetría	0,22	Valor mínimo	7.000,0
Clúster 2	Edad	Media	58,22	Curtosis	-0.14
		Mediana	58,0	Primer cuartil (Q1) Percentil 25	54,0
		Varianza	25,97	Tercer cuartil (Q3) Percentil 75	62,0
		Desv. Estándar	5,1	Valor máximo	79
		Asimetría	0,569	Valor mínimo	51
	Capital	Media	63.428,96	Curtosis	0,19
		Mediana	62.000,0	Primer cuartil (Q1) Percentil 25	41.523,0
		Varianza	752.913.924,12	Tercer cuartil (Q3) Percentil 75	80.000,0
		Desv. Estándar	27.439,28	Valor máximo	175.000,0
		Asimetría	0,46	Valor mínimo	9.500,0
Edad	Media	48,38	Curtosis	1,5	
	Mediana	49,0	Primer cuartil (Q1) Percentil 25	45,0	

Clúster 3		Varianza	41,0947	Tercer cuartil (Q3) Percentil 75	52,0
		Desv. Estándar	6,41	Valor máximo	67
		Asimetría	-0,48	Valor mínimo	21
	Capital	Media	159.549,34	Curtosis	12,95
		Mediana	140.000,0	Primer cuartil (Q1) Percentil 25	117.750,0
		Varianza	5.102.556.582,03	Tercer cuartil (Q3) Percentil 75	170.000,0
		Desv. Estándar	71.432,18	Valor máximo	605.000,0
		Asimetría	3,09	Valor mínimo	95.000,0

Tabla 15: Principales estadísticos de cada clúster hecho con K-Medoids.
Fuente: Elaboración propia.

La Tabla 15 muestra los descriptivos básicos de las variables Edad Actuarial y Capital de cada uno de los clústeres permitiendo llegar a las siguientes conclusiones:

- Sobre la variable Edad se concluye que el clúster 1 muestra las medias y medianas más bajas (41,37 y 43 respectivamente) mientras que el clúster 2 tiene las más altas (58,22 y 58 respectivamente). La varianza menor de los tres clústeres es la del clúster 2 y la mayor es la del clúster 1. Se observa una asimetría negativa de esta variable en los clústeres 1 y 3, y asimetría positiva en el clúster 2. Además, el clúster 2 tiene una curtosis negativa lo que sugiere poca dispersión respecto a la media.
- Al examinar los datos relativos a la variable Capital se concluye que el clúster 3 muestra las medias y medianas más altas (159.549,34 y 140.000 respectivamente) mientras que el clúster 2 exhibe las cifras más bajas, con una media de 63.428,96 y una mediana de 62.000. El clúster 3 presenta una varianza considerablemente mayor en comparación con los otros clústeres. Sobresale el alto grado de asimetría en el clúster 3, acompañado de una curtosis positiva notablemente alta.

4.1.1.3. Agrupación jerárquica

Entre los dos tipos más comunes de algoritmos de agrupación jerárquica están el algoritmo jerárquico aglomerativo²⁹ y el algoritmo jerárquico disociativo. Se puede definir como clúster aglomerativo aquél que empieza con todos los datos y se forman los grupos hasta al final del proceso en que todos los casos están en un mismo conglomerado. Por otro lado, el análisis de clúster disociativo consiste en realizar el proceso de forma inversa, es decir, empieza con un conglomerado formado por todos los grupos y a partir de ahí se forman grupos pequeños. El output obtenido del algoritmo aglomerativo es una estructura de árbol de clasificación llamada dendograma en la que pueden visualizarse de

²⁹ Ver 9.2.3. Algoritmo jerárquico aglomerativo – Pág. 100.

forma ascendente (métodos aglomerativos o HAC) o descendente (métodos disociativos) las secuencias de fusiones o divisiones de grupos según las similitudes detectadas hasta llegar a un solo grupo que contiene de forma fusionada todos los clústeres (Balcan, 2014).

Tal y como se ha mencionado anteriormente se ha elegido aplicar en este trabajo el algoritmo jerárquico aglomerativo ya que es el que se considera más oportuno para la agrupación de contratos de seguros bajo IFRS17. Este enfoque considera inicialmente que cada contrato es un cluster individual y a medida que se avanza el proceso de agrupamiento, los clústeres individuales se fusionan gradualmente. En concreto el algoritmo se basa en los siguientes pasos:

- 1) El cálculo de una matriz de distancias. En este trabajo se ha utilizado la distancia Euclídea.
- 2) Se agrupan los individuos cuya distancia es menor y se vuelve a calcular la matriz de distancias.
- 3) El algoritmo termina cuando:
 - a. Se haya formado un solo grupo.
 - b. Se haya alcanzado un número de grupos determinado.
 - c. Cuando mediante un contraste de significación ya no existen razones para continuar debido a que las observaciones más similares ya no son suficientemente homogéneas.

Se ha aplicado el algoritmo jerárquico aglomerativo a los datos referentes al año 2000 a través de la biblioteca *sklearn* de Python, la cual permite realizar el análisis de clústeres jerárquico eligiendo el método a utilizar para llevar a cabo las diferentes uniones en los diferentes niveles (Johnson, 1967). Así los diferentes métodos que se pueden emplear son el enlace completo, el enlace promedio, el enlace simple, o el enlace de Ward (Everitt, 2011).

Finalmente, se elige el enlace de Ward por ser el que agrupa los contratos de seguros de manera más equilibrada. El enlace de Ward consiste en unir los dos clústeres que tengan un menor aumento en la suma total de los cuadrados de las diferencias de cada clúster (suma de cuadrados de errores, es decir, la distancia Euclídea). Específicamente, consiste en observar la variabilidad de dentro de cada clúster a través de las distancias de cada observación al centroide de cada grupo.

$$\Delta E_{rt} = \frac{n_r n_t}{n_r + n_t} \sum_{j=1}^n (m_j^r - m_j^t)^2$$

Expresión 6: Potencial incremento del error resultante de la unión de dos grupos.

Fuente: Gallardo (2021).

Donde:

- ΔE_{rt} es el incremento potencial de la suma total de los errores que surge de la unión del grupo r y t.
- n_r y n_t es el número de individuos que forman parte del grupo r y del grupo t respectivamente.

- m_j^r y m_j^t son las coordenadas de cada centroide del clúster r y del clúster t respectivamente, en la dimensión j .

Así, aplicando el método de aglomeración jerárquica, y como método de enlace Ward y la distancia euclídea, se obtiene el siguiente (Gráfico 7):

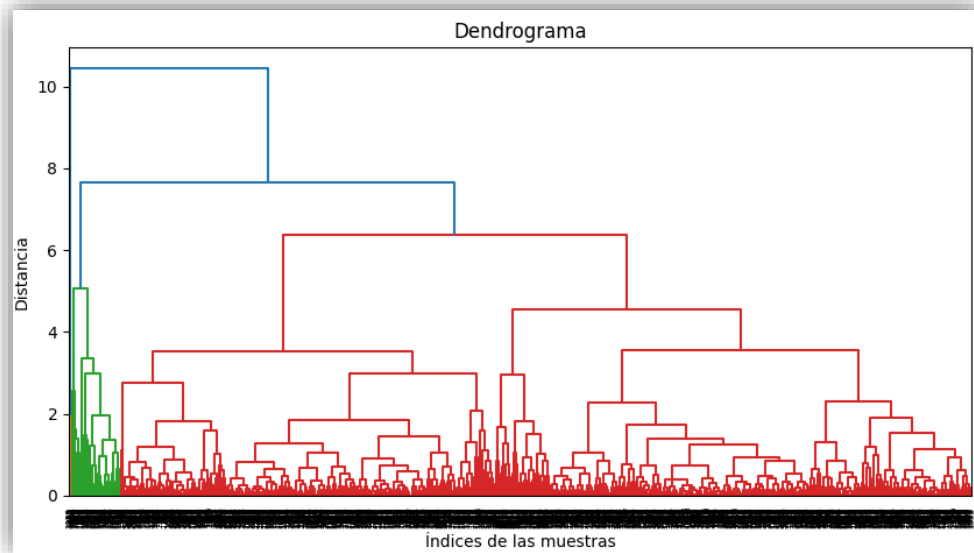


Gráfico 7: Clúster algoritmo jerárquico aglomerativo.
Fuente: Elaboración propia.

Teniendo en cuenta que anteriormente se ha visto que con el método del codo el número óptimo de centroides es 3, se puede volver a ejecutar el algoritmo con el fin de tener los resultados de una forma más clara (Gráfico 8):

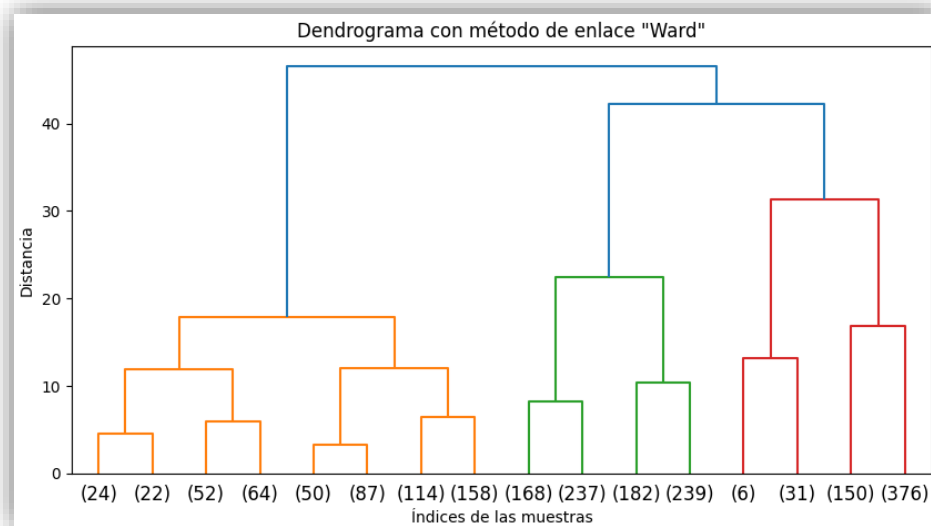


Gráfico 8: Clúster algoritmo jerárquico aglomerativo con tres grupos.
Fuente: Elaboración propia.

El número de observaciones del grupo 1 es de 571 observaciones, para el grupo número 2 es de 826 observaciones y para el grupo número 3 es de 563 observaciones

Además, es posible calcular los estadísticos más relevantes en relación a los datos y los centroides de cada clúster. Los estadísticos que se calcularán son la media, la mediana, la desviación estándar, la varianza, los cuartiles y percentiles, la asimetría y la curtosis para los datos de cada clúster (Tabla 16).

Clúster 1	Edad	Media	38,03	Curtosis	0,75
		Mediana	40,0	Primer cuartil (Q1) Percentil 25	35,0
		Varianza	38,1	Primer cuartil (Q3) Percentil 75	43,0
		Desv. estándar	6,17	Valor máximo	46
		Asimetría	-1,18	Valor mínimo	19
	Capital	Media	72.177,77	Curtosis	-0,42
		Mediana	70.000,0	Primer cuartil (Q1) Percentil 25	45.000,0
		Varianza	1.132.397.648,56	Tercer cuartil (Q3) Percentil 75	95.250,0
		Desv. Estándar	33.651,12	Valor máximo	175000,0
		Asimetría	0,39	Valor mínimo	13.000,0
Clúster 2	Edad	Media	55,46	Curtosis	-0,51
		Mediana	55,0	Primer cuartil (Q1) Percentil 25	50,0
		Varianza	41,36	Tercer cuartil (Q3) Percentil 75	60,0
		Desv. Estándar	6,43	Valor máximo	79
		Asimetría	0,28	Valor mínimo	44
	Capital	Media	59.035,01	Curtosis	-0,13
		Mediana	60.000,0	Primer cuartil (Q1) Percentil 25	41.523,0
		Varianza	510.661.706,61	Tercer cuartil (Q3) Percentil 75	71.402,83
		Desv. Estándar	22.597,83	Valor máximo	145.000,0
		Asimetría	0,23	Valor mínimo	7.000,0

Clúster 3	Edad	Media	49,72	Curtosis	1,10
		Mediana	49,0	Primer cuartil (Q1) Percentil 25	46,0
		Varianza	32,54	Tercer cuartil (Q3) Percentil 75	53,0
		Desv. Estándar	5,70	Valor máximo	69
		Asimetría	0,54	Valor mínimo	33
	Capital	Media	149.058,77	Curtosis	12,41
		Mediana	130.000,0	Primer cuartil (Q1) Percentil 25	105.000,0
		Varianza	5.176.604.625,42	Tercer cuartil (Q3) Percentil 75	165.000,0
		Desv. Estándar	71.948,63	Valor máximo	605.000,0
		Asimetría	2,95	Valor mínimo	75.000,0

Tabla 16: Principales estadísticos de cada clúster hecho con el algoritmo Jerárquico Aglomerativo.
Fuente: Elaboración propia.

La Tabla 16 muestra los descriptivos básicos de las variables Edad Actuarial y Capital de cada uno de los clústeres permitiendo llegar a las siguientes conclusiones:

- Sobre la variable Edad se concluye que el clúster 1 tiene las medias y medianas más bajas (38,08 y 40 respectivamente), mientras que el clúster 2 tiene las más altas (55,46 y 55 respectivamente). El clúster 2 es el que tiene la varianza más alta y el clúster 3 es el que tiene la varianza más baja. Se observa una asimetría negativa de esta variable en el clúster 1 y positiva en los clústeres 2 y 3. Además, el clúster 2 tiene una curtosis negativa lo que sugiere poca dispersión respecto a la media. En general, tanto la asimetría como la curtosis de todos los clústeres son bajas.
- Al analizar los datos relativos a la variable Capital se concluye que el clúster 3 muestra las medias y medianas más altas (149.058,77 y 130.000 respectivamente) mientras que el clúster 2 exhibe las cifras más bajas, con una media de 59.035,01 y una mediana de 60.000. El clúster 3 es el que presenta una varianza mayor y el clúster 2 es el que tiene una varianza menor. considerablemente mayor en comparación con los otros clústeres. Tanto la asimetría como la curtosis del clúster 3 son positivas y elevadas. Respecto a la curtosis, es negativa tanto para el clúster 2 como para el clúster 1. Las asimetrías son positivas para los tres clústeres.

Finalmente, después de proponer tres metodologías para la agrupación de contratos de seguros por riesgos y características similares se opta por el algoritmo k-means por su simplicidad y eficiencia computacional permitiendo a las empresas aseguradoras manejar grandes volúmenes de datos de manera efectiva. Además, ofrece una clara interpretación de los resultados pudiendo identificar los centroides representativos de cada clúster de forma fácil.

4.1.3. Estimación de los flujos de efectivo

Una vez se han agrupado los contratos de seguro por riesgos similares, se procede a realizar la estimación de los flujos de efectivo. En primer lugar, será necesario determinar las primas. Para ello, se aplicarán las tablas biométricas GKMF95 suizas no generacionales Unisex debido a que eran las que aplicaban las entidades aseguradoras en las fechas efectivas de los contratos de seguros que hay en el portafolio estudiado.

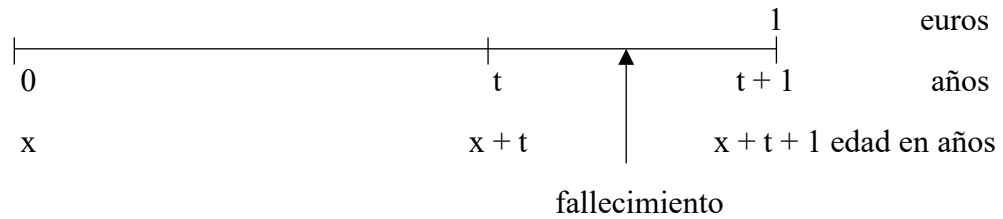
Con el fin de obtener el balance de transición la normativa establece tres metodologías para calcular el balance de transición: el método retrospectivo, el método retrospectivo modificado y el método Fair Value. En este trabajo se elige aplicar el último método, y por tanto, se calcularán las provisiones según las normativas IFRS4 e IFRS17 en los ejercicios financieros a 31 de diciembre de 2022 y 31 de diciembre de 2023.

Para el cálculo de las provisiones para los ejercicios 2022 y 2023 bajo IFRS 17 y IFRS 4, se emplearán las tablas PASEM 2020 de primer orden. En el caso de IFRS 17, se utilizará la curva de tipos de interés de la EIOPA de diciembre de 2022 para el cálculo de las provisiones del año 2022, y la curva de tipos de interés de la EIOPA de diciembre de 2023 para el cálculo de las provisiones del año 2023. Por otro lado, en el caso de IFRS 4, el tipo de interés será fijo y se especificará en el subapartados que vienen a continuación. Es importante señalar que en este apartado, en relación a IFRS 17, aún no se aplicarán ninguna hipótesis. Por tanto, las expresiones de las rentas no diferirán entre IFRS 17 y IFRS 4, siendo la única diferencia el tipo de interés. Para simplificar, se anotará de forma genérica como "I" en las expresiones de las rentas de este apartado en ambos casos.

4.1.3.1. Estimación de la prima

En primer lugar, se estimarán los cobros (primas) que pagará el asegurado. Para ello se calcularán la prima de tarifa, la prima de inventario y la prima pura. La prima pura, destinada exclusivamente a cubrir los costos directos del contrato, como las prestaciones, se calcula normalmente considerando la siniestralidad y el interés, incorporando a menudo un recargo implícito de seguridad. Por su parte, la prima de inventario incluye tanto las prestaciones como los gastos internos de gestión asociados, como sueldos, costes contables, de recursos humanos, de tecnología, apoyo informático, de depreciación, alquileres y mantenimiento (párrafo B65). Finalmente, la prima de tarifa o comercial abarca todos los costos del contrato, incluyendo prestaciones y gastos de gestión tanto internos como externos, siendo estos últimos las comisiones pagadas a los agentes de la compañía.

En el contexto del análisis de flujos de efectivo en el seguro de vida es importante destacar que estos flujos representan tanto cobros como pagos para la compañía aseguradora y se distribuyen a lo largo del tiempo. Sin embargo, para realizar una comparación adecuada con el valor presente, es necesario descontar estos flujos. En el caso específico del seguro de vida abordado en este estudio, se establece que el capital estipulado en el contrato de seguro será cobrado como beneficio al finalizar el año del fallecimiento del asegurado, tal y como se puede ver en el eje temporal que se muestra a continuación.



Además, este seguro de vida se caracteriza por:

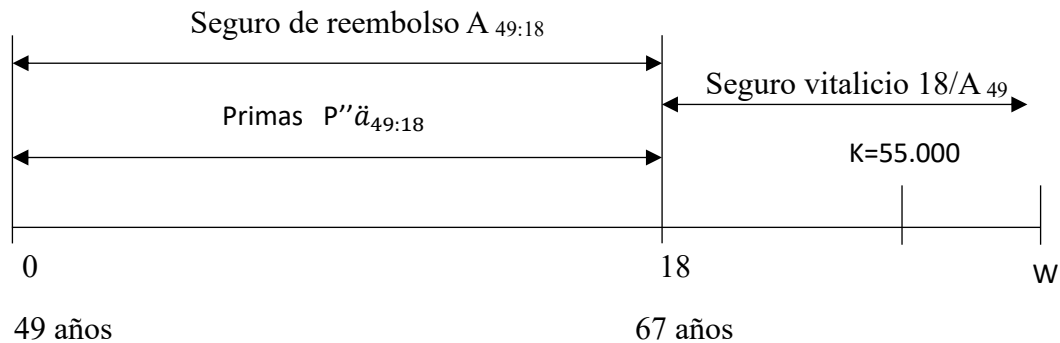
- Las primas son constantes, anuales, pagaderas mientras viva el asegurado y hasta la edad de jubilación (67 años).
- El contrato incluye además una cláusula de contraseguro de forma que, si el asegurado fallece antes de la edad de jubilación, se reembolsarán las primas pagadas hasta entonces al final del año de fallecimiento del asegurado.
- Se incluyen los recargos de 0,01% anual sobre el valor actual actuarial de las primas de tarifa para cubrir los gastos de gestión interna (ggi) y de un 3% sobre la tarifa para los gastos de gestión externa (gge).
- El tipo de interés técnico utilizado en esta operación es de 2%.

Así, si se toma por ejemplo el primer asegurado de la cartera de seguros correspondiente al primer contrato del clúster 3, se dispone de la siguiente información detallada (Tabla 17):

Nombre original de la variable	Descripción
ID	1
Gender (Género)	M
Birth_Date (Fecha de nacimiento)	10/07/1960
Effective_Date (Fecha efectiva)	08/03/2000
Capital	55000€
Renewal_Date (Fecha de renovación)	08/03/2009
Age (Años)	48,659822
T	0,65982204
Age_Actuarial	49
Date_Birthday	10/07/2009
X	0,18082192
R	4
S	1
Age actuarial quarter	48,75
Month	3

Tabla 17: Datos del primer asegurado del portafolio.
Fuente: Elaboración propia.

El esquema temporal en este caso sería el siguiente:



La ecuación de equilibrio es esencial para calcular la prima tarifa. Esta ecuación asegura que los ingresos por primas sean suficientes para cubrir los costos de siniestros y demás gastos, así como se logra un cierto equilibrio financiero. La expresión de la ecuación de equilibrio de la prima de tarifa sería la siguiente:

$$VAA_0(P'') = VAA_0(\text{prestaciones}) + VAA_0(\text{ggi}) + VAA_0(\text{gge}).$$

*Expresión 7: Ecuación de equilibrio prima de tarifa.
Fuente: Elaboración propia.*

El valor actual actuarial de la prima de tarifa equivale a la prima de tarifa descontada, es decir, al valor presente de los flujos de pagos futuros de primas de tarifa de un contrato de seguro.

$$VAA_0(P'') = P'' \ddot{a}_{49:18}$$

*Expresión 8: Valor Actual Actuarial de la prima de tarifa.
Fuente: Elaboración propia.*

El valor actual actuarial de las prestaciones que representa el valor presente de todos los pagos que la aseguradora espera realizar en el futuro, se puede calcular como la suma del reembolso de primas esperado y el seguro vitalicio, ambos descontados.

$$VAA_0(\text{Prestaciones}) = P'' A_{49:18} + 55000 18/A_{49}$$

*Expresión 9: Reembolso de primas y seguro vitalicio.
Fuente: Elaboración propia.*

El valor actual actuarial de los gastos de gestión interna se calculará teniendo en cuenta el valor actual actuarial de la prima de tarifa. Concretamente, consisten en un recargo del

0,01% anual sobre el valor actual actuarial de las primas de tarifa y se distribuirán a lo largo de la vida del contrato de seguro.

$$VAA_0 (ggi) = 0,0001 P'' \ddot{a}_{49:18} \ddot{a}_{49}$$

Expresión 10: Gastos de gestión interna.

Fuente: Elaboración propia.

El valor actual actuarial de los gastos de gestión externa se calculará teniendo en cuenta el valor actual actuarial de las primas. Específicamente, se aplicará un recargo del 3% sobre el valor actual actuarial de la prima de tarifa. Tal y como se ha indicado anteriormente estos gastos representan los costos asociados a la administración y gestión externa del contrato de seguro.

$$VAA_0 (gge) = 0,03 P'' \ddot{a}_{49:18}$$

Expresión 11: Gastos de gestión externa.

Fuente: Elaboración propia.

Para determinar la prima de tarifa (P'') de este contrato de seguros se procede a sustituir los valores correspondientes en la ecuación de equilibrio original [1] obteniéndose la siguiente expresión:

$$P'' \ddot{a}_{49:18} = P'' A_{49:18} + 55000 \cdot 18 / A_{49} + 0,0001 P'' \ddot{a}_{49:18} \ddot{a}_{49} + 0,03 P'' \ddot{a}_{49:18}$$

Expresión 12: Sustitución a la ecuación de equilibrio para obtener la Prima de Tarifa.

Fuente: Elaboración propia.

Para entender cada término de esta expresión, es importante desglosar y analizar las rentas actuariales asociadas. La renta actuarial de los cobros de primas describe los pagos realizados por el asegurado desde la edad efectiva del contrato de seguro, que es a los 49 años, hasta su jubilación a los 67 años. En este contexto, esta renta actuarial representa un período temporal de 18 años, calculado restando la edad de jubilación (67 años) menos la edad efectiva del contrato de seguro (49 años), lo que nos da una temporalidad de 18 años.

$$\ddot{a}_{49:18} = P'' \sum_{t=0}^{(18+0)*1-1} (1 + 0,02)^{-t} {}_tP_{49}$$

Expresión 13: Renta Actuarial de los cobros de primas.

Fuente: Elaboración propia.

La simbología empleada en la expresión refleja la naturaleza de los cobros de primas en el contexto del contrato de seguro. El uso de la letra “a” en minúscula y los dos puntos encima denotan que estas primas son prepagables y están asociadas con la supervivencia del asegurado. El subíndice indica el inicio del contrato a los 49 años, seguido de la duración temporal de 18 años. La expresión $a_{49:18}$ representa el valor presente actuarial de

los cobros de primas, descontados al tipo de interés determinista, considerando la supervivencia del asegurado año tras año hasta su jubilación.

Continuando con la comprensión detallada de los compromisos financieros derivados del contrato de seguro, es necesario desglosar la cláusula de reembolso de primas en caso de fallecimiento del asegurado prevista en el contrato de seguro. En concreto esta cláusula establece el reembolso de primas pagadas por el asegurado al beneficiario en caso de que el asegurado fallezca entre los 49 años, edad en la que entra en efecto el contrato de seguro, y la edad de jubilación, fijada en 67 años. Examinando esta cláusula con mayor detalle y para comprenderla en términos más específicos y claros, se puede considerar la siguiente expresión:

$$A_{49:18} = P'' \sum_{t=0}^{(18+0)*1-1} (1 + 0,02)^{-(t+1)} {}_tq_{49}$$

Expresión 14: Cláusula reembolso de primas.

Fuente: Elaboración propia.

La expresión $A_{49:18}$ represente el valor presente actuarial asociado con la mortalidad en el contexto del seguro. El subíndice indica el periodo que transcurre desde la edad efectiva del contrato de seguro (49 años) hasta su edad de jubilación (temporalidad de 18 años). En términos prácticos, este valor es el resultado del sumatorio de las primas pagadas por el asegurado desde los 49 años hasta la edad de jubilación, descontando las primas al tipo de interés determinista y vencido, simbolizado con $t+1$ y vinculándolas con las probabilidades de fallecimiento del asegurado, en particular, se trata de la probabilidad de que el asegurado haya sobrevivido hasta el año correspondiente y fallezca en ese último periodo teniendo en cuenta las probabilidades de supervivencia y mortalidad en cada año intermedio.

Seguidamente, se analiza la parte del contrato de seguro que hace referencia al seguro vitalicio, donde se establece un capital especificado en el contrato que el beneficiario recibirá en caso de fallecimiento del asegurado, siempre y cuando éste haya llegado vivo al momento de la jubilación fijada a los 67 años. En el caso particular de este asegurado, el capital estipulado en el contrato asciende a 55.000 euros. La expresión matemática actuarial asociada a este seguro vitalicio se expresa de la siguiente manera:

$${}_{18/}A_{49} = \sum_{t=18}^{(125-49)*1-1} 55000 (1 + 0,02)^{-(t+1)} {}_tq_{49}$$

Expresión 15: Parte del seguro de fallecimiento.

Fuente: Elaboración propia.

La prestación del seguro de fallecimiento comienza a partir de la edad de jubilación, con un diferimiento de dieciocho años, indicado por el subíndice "18" antes de A_{49} . Se escribe la A mayúscula debido a que dicha renta está vinculada al fallecimiento del asegurado. El valor actual actuarial se calcula como la suma desde la edad de jubilación hasta el infinito actuarial final de la tabla biométrica GKM95 unisex, que es de 125 años para los hombres como es el caso de este asegurado. Se resta la edad del asegurado en la fecha del contrato, menos 1 año, para considerar el último año de las tablas biométricas, luego se presume

que no vive nadie. Se descuenta el capital especificado en el contrato, en este caso, 55.000 euros al tipo de interés determinista, y se ajusta al año vencido, ya que la prestación la cobra el beneficiario al final del año en que ha fallecido el asegurado.

Por último, también se describirá la renta necesaria para distribuir los gastos de gestión internos durante toda la vigencia del contrato, es decir, desde la fecha efectiva del contrato, que es cuando el asegurado empieza a pagar las primas hasta que fallece.

$$\ddot{a}_{49} = \sum_{t=0}^{(125-49)*1-1} 55000 (1 + 0,02)^{-t} {}_tP_{49}$$

Expresión 16: Distribución de los gastos de gestión interna.

Fuente: Elaboración propia.

Esta renta necesaria para el reparto de gastos de gestión interna se calculará como un sumatorio que abarca desde el inicio del pago de primas por parte del asegurado hasta el infinito actuarial de las tablas descontando la edad del asegurado en la fecha efectiva del contrato, menos uno, del capital estipulado para este asegurado. Este cálculo también toma en consideración el tipo de interés establecido y las probabilidades de supervivencia.

A partir de la Expresión 12 se puede hallar la prima de tarifa P'' que paga el asegurado desde la fecha efectiva del contrato hasta la edad de jubilación y que incorpora los gastos internos y los gastos externos obteniéndose:

$$P'' = \frac{55000 \text{ 18}/A_{49}}{\ddot{a}_{49:18} - 0,0001 \ddot{a}_{49:18} \ddot{a}_{49} - 0,03 \ddot{a}_{49:18} - A_{49:18}}$$

Expresión 17: Determinación de la Prima de Tarifa.

Fuente: Elaboración propia.

Seguidamente, puede hallarse también la prima de inventario utilizando la expresión [6] y eliminando los gastos de gestión externa lo que lleva a:

$$P' \ddot{a}_{49:18} = P'' A_{49:18} + 55000 \text{ 18}/A_{49} + 0,0001 P'' \ddot{a}_{49:18} \ddot{a}_{49}$$

Expresión 18: Sustitución a la ecuación de equilibrio para obtener la Prima de Inventario.

Fuente: Elaboración propia.

Y hallando P' se obtiene:

$$P' = \frac{P'' A_{49:18} + 55000 \text{ 18}/A_{49} + 0,0001 P'' \ddot{a}_{49:18} \ddot{a}_{49}}{\ddot{a}_{49:18}}$$

Expresión 19: Determinación de la Prima de Inventario.

Fuente: Elaboración propia.

Finalmente puede obtenerse la prima pura como el valor actual actuarial de las prestaciones sin considerar ni gastos de gestión interna ni gastos de gestión externa.

$$VAA_0 (P) = VAA_0 (\text{Prestaciones})$$

Expresión 20: Ecuación de equilibrio de la Prima Pura.

Fuente: Elaboración propia.

$$P\ddot{a}_{49:18} = 55000 \cdot 18/A_{49} + 55000 A_{49:18}$$

Expresión 21: Sustitución a la ecuación de equilibrio para obtener la Prima Pura.

Fuente: Elaboración propia.

$$P = \frac{55000 \cdot 18/A_{49} + 55000 A_{49:18}}{\ddot{a}_{49:18}}$$

Expresión 22: Determinación de la Prima Pura.

Fuente: Elaboración propia.

Prima pura	Prima de inventario	Prima de tarifa
1671,4799	1686,0177	1738,6303

Tabla 18: Estimación de la Prima pura, Prima de inventario y Prima de tarifa.

Fuente: Elaboración propia.

4.1.3.2. Estimación de las provisiones

Una vez determinadas las primas deben calcularse las obligaciones que tiene la empresa aseguradora derivadas del contrato de seguro. De acuerdo al Art. 38 del Real Decreto 2486/1998, de 20 de noviembre, deben incluirse en el cálculo de las provisiones todas las obligaciones por los contratos de seguro teniendo en cuenta la información existente al cierre del ejercicio. Por tanto, calcular las obligaciones del contrato de seguro será equivalente al cálculo de las provisiones técnicas, y en el caso de los seguros de vida,³⁰ la provisión matemática.

La provisión matemática de un seguro de vida debe calcularse como la equivalencia actuarial resultante, es decir, igualando el valor esperado de las primas recibidas con el valor esperado de las prestaciones. Inicialmente, cuando se calcula la prima pura, la ecuación de equilibrio es el Valor actual actuarial de las primas puras es igual al Valor actual actuarial de las prestaciones.

Por el contrario, en un momento posterior k , no se cumple la igualdad anterior, es decir, la operación se encuentra desequilibrada. Existen dos metodologías para el cálculo de provisiones matemáticas, siendo la primera analizando el futuro (método prospectivo) y la segunda analizando el pasado de la operación (método retrospectivo).

El primer método nombrado en el párrafo anterior ha sido el prospectivo. Así, si solo se tienen en cuenta en este momento las prestaciones y primas pendientes a partir de k , la operación se desequilibra a favor de las prestaciones:

$$VAA_k \left(\begin{matrix} \text{primas puras} \\ \text{futuras} \end{matrix} \right) < VAA_k \left(\begin{matrix} \text{prestaciones} \\ \text{futuras} \end{matrix} \right).$$

Expresión 23: Desequilibrio que se produce a favor de las prestaciones.

Fuente: Elaboración propia.

³⁰ En el caso de los seguros de vida con duración igual o inferior al año se habla de la provisión de primas no consumidas PPNC y, en su caso, la provisión de riesgos en curso.

Con el fin de que la operación se encuentre equilibrada en un momento posterior k , será necesario añadir una cantidad, la cual recibe el nombre de provisión matemática $V(x;k)$, donde x será la edad del asegurado en el momento de la contratación y k el momento de cálculo de la provisión.

$$VAA_k \left(\begin{matrix} \text{primas puras} \\ \text{futuras} \end{matrix} \right) + V(x;k) = VAA_k \left(\begin{matrix} \text{prestaciones} \\ \text{futuras} \end{matrix} \right)$$

*Expresión 24: Expresión equilibrio para el cálculo de provisión.
Fuente: Elaboración propia.*

Aislando $V(x;k)$ se obtiene:

$$V(x;k) = VAA_k \left(\begin{matrix} \text{prestaciones} \\ \text{futuras} \end{matrix} \right) - VAA_k \left(\begin{matrix} \text{primas puras} \\ \text{futuras} \end{matrix} \right)$$

*Expresión 25: Determinación de la provisión matemática.
Fuente: Elaboración propia.*

Por otro lado, en el método retrospectivo se analiza el pasado de la operación. Así, partiendo de la ecuación de equilibrio inicial, se aplica la propiedad de la escindibilidad del valor actual actuarial de los capitales diferidos.

$$VAA_0 (\text{primas puras hasta } k) + {}_kE_x \cdot VAA_k (\text{primas puras a partir de } k) =$$

$$VAA_0 (\text{prestaciones hasta } k) + {}_kE_x \cdot VAA_k (\text{prestaciones a partir de } k).$$

*Expresión 26: Aplicación de la propiedad de escindibilidad.
Fuente: Elaboración propia.*

Agrupando y aislando $V(x;k)$ se obtiene:

$$V(x;k) = [VAA_0 (\text{primas puras hasta } k) - VAA_0 (\text{prestaciones hasta } k)] \cdot \frac{1}{{}_kE_x}$$

*Expresión 27: Determinación de la provisión matemática habiendo aplicado la propiedad de escindibilidad.
Fuente: Elaboración propia.*

El cálculo de las provisiones con los dos enfoques debe coincidir, por lo que en este trabajo se ha decidido aplicar el método prospectivo.

Antes de ponerse a realizar cualquier tipo de cálculo debe acudir al Real Decreto 2486/1998, de 20 de noviembre, Reglamento de Ordenación y Supervisión de los Seguros

Privados, el cual indica las consideraciones que deben tenerse en cuenta con el cálculo de provisiones de seguros de vida a efectos contables:

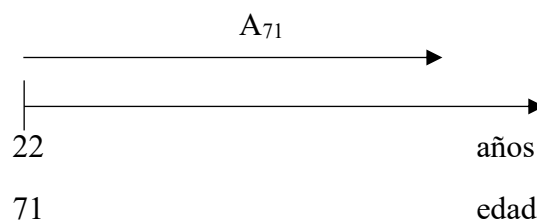
- El cálculo debe realizarse **póliza a póliza**.
- Aplicando un **método prospectivo**, salvo que no fuera posible por las características del contrato considerado o se demuestre que las provisiones obtenidas sobre la base de un método retrospectivo no son inferiores a las que resultarían de la utilización de un método prospectivo.
- La base de cálculo de esta provisión será la **prima de inventario** devengada en el ejercicio. La provisión de seguros de vida se calculará teniendo en cuenta los recargos de gestión interna (de administración) previstos en las bases técnicas.
- Tabla de mortalidad: tablas biométricas de primer orden diferenciando por género.

Finalmente, es necesario incidir que por lo que se refiere a los tipos de interés aplicados serán:

- Provisiones bajo IFRS 17 año 2022 e utilizará la curva de tipo de interés de la EIOPA del diciembre de 2022.
- Provisiones bajo IFRS 17 año 2023 e utilizará la curva de tipo de interés de la EIOPA del diciembre de 2023.
- Provisiones bajo IFRS 4 año 2022 y 2023 se utilizará el tipo de interés fijo del 2%.

4.1.3.5. Estimación de las provisiones para el 2022 bajo IFRS17 e IFRS4

Para el caso del año 2022 también es necesario revisar el esquema temporal actualizado de los flujos. Si bien el asegurado tenía 49 años en el año 2000, fecha efectiva del contrato, para el año 2022, la edad del asegurado sería de 71 años, superando así la edad de jubilación estipulada en 67 años. Igual que en el caso anterior el asegurado ya habrá alcanzado la edad de jubilación³¹ y la empresa aseguradora dejará de recibir primas y también se habrá agotado el periodo de reembolso de primas que recibiría el beneficiario y que se había estipulado en el contrato de seguros para el caso del que el asegurado falleciera entre la fecha efectiva del contrato y la edad de jubilación. En este punto, solo permanecerá la parte correspondiente al seguro de fallecimiento, que entró en vigor cuando el asegurado cumplió 67 años.



Observando el nuevo esquema temporal si se quieren calcular las provisiones a efectos contables que necesita la empresa aseguradora en 2022 se deberán realizar los siguientes cálculos:

³¹ Ver 9.3.1. Estimación de las provisiones genéricas para el 2022 bajo IFRS17 e IFRS4 (asegurado no jubilado en el 2022).- Pág. 107.

Para $k=22$ años,

$$V'(49;22) = VAA_{22} (\text{prestaciones futuras}) + VAA_{22} (\text{ggi futuros}) + VAA_{22} (\text{gge futuros}) - VAA_{22} (\text{primas tarifa futuras})$$

*Expresión 28: Cálculo provisiones a efectos contables ejercicio 2022.
Fuente: Elaboración propia.*

En este caso, tal y como se ha indicado anteriormente VAA_{22} (gge futuros) y VAA_{22} (primas tarifa futuras) serán nulos quedando la expresión reducida a:

$$V'(49;22) = VAA_{22} (\text{prestaciones futuras}) + VAA_{22} (\text{ggi futuros})$$

$$V'(49;22) = 55000 A_{71} + (0,0001 P'' \ddot{a}_{49:18}) \ddot{a}_{71}$$

*Expresión 29: Cálculo provisiones finales a efectos contables ejercicio 2022.
Fuente: Elaboración propia.*

En este punto del análisis, surge la necesidad de calcular dos nuevas rentas que surgen a partir del cambio en la edad del asegurado. La primera corresponde a la parte del seguro de fallecimiento, representada por A_{71} y la segunda se relaciona con la distribución de los gastos de gestión interna a lo largo de la vida del contrato de seguro, denotada como \ddot{a}_{71} . Además, se recupera la renta $\ddot{a}_{49:18}$, la cual se utiliza exclusivamente para calcular los gastos de gestión interna como un porcentaje del 0,0001 del valor actual actuarial de la prima de tarifa.

A continuación, se detalla la renta correspondiente al seguro de fallecimiento en el momento en el que el asegurado alcanza los 71 años de edad. Esta renta se calcula a través de un sumatorio que va desde 0, dado que no hay diferimiento debido a que el seguro de fallecimiento ya está vigente desde que el asegurado alcanzó la jubilación a los 67 años, hasta el infinito actuarial marcado por las tablas PASEM2020 No Rel, es decir, 109 años menos la edad del asegurado que es 71 años multiplicado por uno indicando una periodicidad anual y se resta uno ya que en el último año se asume que todos los individuos han fallecido. En este cálculo, se considera que el capital a pagar es de 55.000 euros en este caso, descontado al tipo de interés fijado. Se descuenta, por tanto, a $t+1$, indicándose que el capital lo cobrará el beneficiario de forma vencida, es decir, el beneficiario recibirá el capital al final del año en que fallezca el asegurado. Además, se tienen en cuenta las probabilidades de fallecimiento del asegurado (Ayuso, 2007), vinculándolas con las probabilidades de supervivencia hasta el año correspondiente y mortalidad en cada año intermedio.

$$A_{71} = \sum_{t=0}^{(109-71)*1-1} (1+I)^{-(t+1)} {}_tq_{71}$$

*Expresión 30: Prestación Seguro de fallecimiento ejercicio 2022.
Fuente: Elaboración propia.*

La otra renta que requiere atención detallada es aquella relacionada con la distribución de los gastos de gestión interna a lo largo de la duración del contrato. Este cálculo se realiza mediante un sumatorio que se inicia en el momento en el que el asegurado alcanza los 71 años, marcado como $t=0$, y se extiende hasta el infinito actuarial, es decir, 109 años menos

la edad actual de 71 años, multiplicado por la periodicidad anual. Se resta uno para considerar que en el último año se presume que no hay individuos vivos. En este proceso, se descuenta el tipo de interés y se tienen en cuenta las probabilidades de supervivencia, lo que permite evaluar con precisión los gastos de gestión interna asociados al contrato de seguro a lo largo del tiempo.

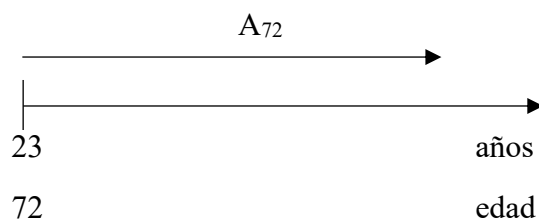
$$\ddot{a}_{71} = \sum_{t=0}^{(109-71)*1-1} (1+I)^{-t} {}_t p_{71}$$

Expresión 31: Reparto de gastos de gestión interna ejercicio 2022.
Fuente: Elaboración propia.

Para el enfoque de cálculo de provisiones bajo IFRS17 considerando la curva de interés de EIOPA de diciembre de 2022 y sin considerar aún hipótesis se obtiene una provisión de 33.211,84€. Bajo el enfoque IFRS4 para el año 2022 con el tipo interés fijo del 2%, la provisión asciende a 39.185,6 € reflejando las obligaciones financieras de la empresa aseguradora en el presente año.

4.1.3.6. Estimación de las provisiones para el 2023 bajo IFRS17 e IFRS4

Para determinar las provisiones requeridas por la empresa aseguradora en el año 2023, es crucial revisar el esquema temporal actualizado de los flujos. Tomando como ejemplo el caso en el que el asegurado tenía 49 años en el año 2000, fecha efectiva del contrato, para el año 2023, la edad del asegurado sería de 72 años, superando así la edad de jubilación³² estipulada en 67 años. Este cambio en la edad del asegurado impacta de manera significativa en el cálculo de las provisiones necesarias. A partir de la edad de jubilación, la empresa aseguradora dejará de recibir primas y también se habrá agotado el periodo de reembolso de primas que recibiría el beneficiario y que se había estipulado en el contrato de seguros para el caso del que el asegurado falleciera entre la fecha efectiva del contrato y la edad de jubilación. En este punto, solo permanecerá la parte correspondiente al seguro de fallecimiento, que entró en vigor cuando el asegurado cumplió 67 años.



³² Ver 9.3.2. Estimación de las provisiones genéricas para el 2023 bajo IFRS17 e IFRS4 (asegurado no jubilado en el 2023).- Pág. 111.

Observando el nuevo esquema temporal si se quieren calcular las provisiones a efectos contables que necesita la empresa aseguradora en 2023 se deberán realizar los siguientes cálculos:

Para $k=23$ años,

$$V'(49;23)=VAA_{23}(\text{prestaciones futuras}) + VAA_{23}(\text{ggi futuros}) + VAA_{23}(\text{gge futuros}) - VAA_{23}(\text{primas tarifa futuras})$$

Expresión 32: Cálculo provisiones a efectos contables ejercicio 2023.

Fuente: Elaboración propia.

En este caso, tal y como se ha indicado anteriormente VAA_{23} (gge futuros) y VAA_{23} (primas tarifa futuras) serán nulos quedando la expresión reducida a:

$$V'(49;23)=VAA_{23}(\text{prestaciones futuras}) + VAA_{23}(\text{ggi futuros}) \\ = V'(49;23)=55000 A_{72} + (0,0001 P'' \ddot{a}_{49:18}) \ddot{a}_{72}$$

Expresión 33: Cálculo provisiones finales a efectos contables ejercicio 2023.

Fuente: Elaboración propia.

En este punto del análisis, surge la necesidad de calcular dos nuevas rentas que surgen a partir del cambio en la edad del asegurado. La primera corresponde a la parte del seguro de fallecimiento, representada por A_{72} y la segunda se relaciona con la distribución de los gastos de gestión interna a lo largo de la vida del contrato de seguro, denotada como \ddot{a}_{72} . Además, se recupera la renta $\ddot{a}_{49:18}$, la cual se utiliza exclusivamente para calcular los gastos de gestión interna como un porcentaje del 0,0001 del valor actual actuarial de la prima de tarifa.

A continuación, se detalla la renta correspondiente al seguro de fallecimiento en el momento en el que el asegurado alcanza los 72 años de edad. Esta renta se calcula a través de un sumatorio que va desde 0, dado que no hay diferimiento debido a que el seguro de fallecimiento ya está vigente desde que el asegurado alcanzó la jubilación a los 67 años, hasta el infinito actuarial marcado por las tablas PASEM2020, siendo el mismo 109 años para los hombres y 125 para las mujeres, es decir, en este caso 109 años menos la edad del asegurado que es 72 años multiplicado por uno indicando una periodicidad anual y se resta uno ya que en el último año se asume que todos los individuos han fallecido. En este cálculo, se considera que el capital a pagar es de 55.000 euros en este caso, descontado al tipo de interés fijado. Se descuenta, por tanto, a $t+1$, indicándose que el capital lo cobrará el beneficiario de forma vencida, es decir, el beneficiario recibirá el capital al final del año en que fallezca el asegurado. Además, se tienen en cuenta las probabilidades de fallecimiento del asegurado, vinculándolas con las probabilidades de supervivencia hasta el año correspondiente y mortalidad en cada año intermedio.

$$A_{72} = \sum_{t=0}^{(109-72)*1-1} (1 + I)^{-(t+1)} {}_tq_{72}$$

Expresión 34: Prestación Seguro de fallecimiento ejercicio 2023.
Fuente: Elaboración propia.

La otra renta que requiere atención detallada es aquella relacionada con la distribución de los gastos de gestión interna a lo largo de la duración del contrato. Este cálculo se realiza mediante un sumatorio que se inicia en el momento en el que el asegurado alcanza los 72 años, marcado como $t=0$, y se extiende hasta el infinito actuarial, es decir, 109 años menos la edad actual de 72 años, multiplicado por la periodicidad anual. Se resta uno para considerar que en el último año se presume que no hay individuos vivos. En este proceso, se descuenta el tipo de interés y se tienen en cuenta las probabilidades de supervivencia, lo que permite evaluar con precisión los gastos de gestión interna asociados al contrato de seguro a lo largo del tiempo.

$$\ddot{a}_{72} = \sum_{t=0}^{(109-72)*1-1} (1 + I)^{-t} {}_tp_{72}$$

Expresión 35: Reparto de gastos de gestión interna ejercicio 2023.
Fuente: Elaboración propia.

Tras completar todos los cálculos mencionados, la provisión matemática para el año 2023 necesaria bajo el enfoque IFRS17 considerando la curva de tipo interés de EIOPA de diciembre de 2023 y sin considerar hipótesis asciende a 39.763,17 € reflejando las obligaciones financieras de la empresa aseguradora en el presente año. En el caso de IFRS 4 bajo el tipo de interés fijo del 2% se obtiene una provisión para el 2023 de 36.103,59€.

4.1.4. Tasa de descuento

Para la determinación del CSM bajo IFRS17 es necesario determinar la curva de tipos de interés. De acuerdo al párrafo 36 de la normativa, los flujos de efectivo futuros deben ajustarse teniendo en cuenta el valor temporal del dinero y los riesgos financieros relacionados con estos flujos de efectivo, si estos no se incluyen en las estimaciones.

Si bien la normativa no refleja ninguna metodología en concreto para el cálculo de la tasa de descuento, sí hace referencia a dos enfoques para su obtención: el enfoque ascendente o *bottom up* y el enfoque descendente o *top down*. Para el presente trabajo se ha decidido aplicar el enfoque ascendente o *bottom up*. Además, con el fin de que los tipos de interés sean el máximo coherentes a los de mercado, se propone aplicar aquellos que ha presentado EIOPA como tasas libres de riesgo y que ya se utilizan para Solvencia II. En concreto son necesarias las tasas libres de riesgo publicadas para el diciembre de 2022 y diciembre de 2023. Con el fin de cumplir con el párrafo B75 que hace referencia a que los tipos deben reflejar la variabilidad de los activos subyacentes se eligen las tasas libres de riesgos con ajuste de volatilidad. Es importante destacar que existen alternativas, como por ejemplo, la posibilidad de partir de la deuda del Estado para construir la curva de tipos de interés. La deuda del Estado carece de riesgo de insolvencia y a menudo de problemas de iliquidez, teniendo además grandes volúmenes de información. Por ejemplo, en el caso

de los bonos cupón cero de la Reserva Federal de Estados Unidos se tienen unos quinientos datos diarios, teniendo según la agencia de rating Moody's Investors Service la calificación AAA y según la agencia de rating S&P Global Ratings AA+. Los bonos cupón cero se caracterizan por no pagar intereses periódicos (cupones) sino que el inversor recibe el rendimiento al final del plazo.

La estructura temporal de tipo de interés representa como las tasas de interés varían a diferentes plazos, es decir, relaciona el tiempo que hay hasta el vencimiento de las obligaciones o bonos y los rendimientos. Esta puede ser representada mediante la función de descuento, los tipos de interés implícitos y los tipos de interés al contado. Sí, podrías usar "vencimientos anuales" en lugar de "vencimientos a 150 años" para hacerlo más específico y claro. Aquí tienes una versión revisada:

En el caso presentado, donde los cobros y pagos son anuales y la curva de interés proporcionada por EIOPA³³ abarca diferentes vencimientos anuales, no sería necesario llevar a cabo la modelización de la ETTI. Esto se debe a que podemos calcular los flujos con el tipo de interés establecido en cada momento. Sin embargo, si se eligiera una curva de interés que no incluyera todos los tipos de interés para los diferentes vencimientos como se verá en el siguiente apartado en el caso del EUROSAP y el BONO ALEMÁN, o si los flujos de cobros y pagos no fueran anuales, sería necesario realizar la modelización de dicha curva. En este trabajo, aunque no sea necesario en este apartado concreto, se procede a la modelización para mostrar al lector como debería llevarse a cabo en el caso de que fuera necesario. Una de las metodologías más usadas por los Bancos Centrales para modelizar la curva de tipos de interés es la propuesta por Nelson-Siegel (1987) y Svensson (1994). Su enfoque paramétrico se basa en la idea de que los tipos de interés deben ser asintóticos ya que los agentes del mercado no pueden distinguir diferencias de tipos de interés para plazos muy lejanos.

Siguiendo lo que realizan los Bancos Centrales se aplica la modelización de la curva de tipo de interés mediante la metodología de Nelson-Siegel³⁴ y Nelson-Siegel y Svensson. Aunque se han llevado a cabo ambas se opta por elegir la segunda, ya que es la que es la que tiene el error cuadrático menor.

Modelo de Nelson Siegel Svensson:

$$Y(t) = \beta_0 + \beta_1 \left(\frac{1 - e^{-\frac{t}{r_1}}}{\frac{t}{r_1}} \right) + \beta_2 \left(\frac{1 - e^{-\frac{t}{r_1}}}{\frac{t}{r_1}} - e^{-\frac{t}{r_1}} \right) + \beta_3 \left(\frac{1 - e^{-\frac{t}{r_2}}}{\frac{t}{r_2}} - e^{-\frac{t}{r_2}} \right)$$

Expresión 36: Formula de Nelson Siegel Svensson.

Fuente: Elaboración propia.

$Y(t)$: es la tasa de interés en el tiempo t .

β_0 : es el componente de tasa de interés a largo plazo o nivel de la curva. Representa la tasa de interés a largo plazo que se espera que prevalezca en la economía.

³³ Ver 9.4.2. Tipo de interés EIOPA.- Pág. 115.

³⁴ Ver 9.4.1. Nelson-Siegel.- Pág. 114.

β_1 : es el componente de pendiente de la curva. Representa la pendiente de la curva de tasas de interés a medida que cambian los plazos.

τ_1 : es el parámetro de suavizado asociado al componente de pendiente. Controla la rapidez de ajuste la pendiente de la curva.

B_2 : es el componente de curvatura de la curva de tasas de interés.

B_3 : es el componente de torsión de la curva. Representa la torsión o la curvatura adicional de la curva de tasas de interés.

τ_2 : es el parámetro de suavizado asociado al componente de torsión. Controla la rapidez de ajuste de la torsión de la curva.

Con las siguientes restricciones:

$$\beta_0 > 0$$

$$\beta_0 + \beta_1 > 0$$

$$\tau_1 > 0$$

$$\tau_2 > 0$$

Usando el comando de Solver que proporciona Excel para minimizar el error cuadrático medio, se obtienen los siguientes valores para cada parámetro de la fórmula (Tabla 19):

Nelson-Siegel-Svensson		
Parámetro	2023	2022
beta0	0,034	0,035
beta1	-0,002	-0,002
beta2	-0,274	-0,256
beta3	0,247	0,248
tau1	8,041	7,805
tau2	8,023	7,186

Tabla 19: Parámetros optimizados
Nelson - Siegel - Svensson.
Fuente: Elaboración propia.

Y se obtiene los Gráficos 9 y 10:

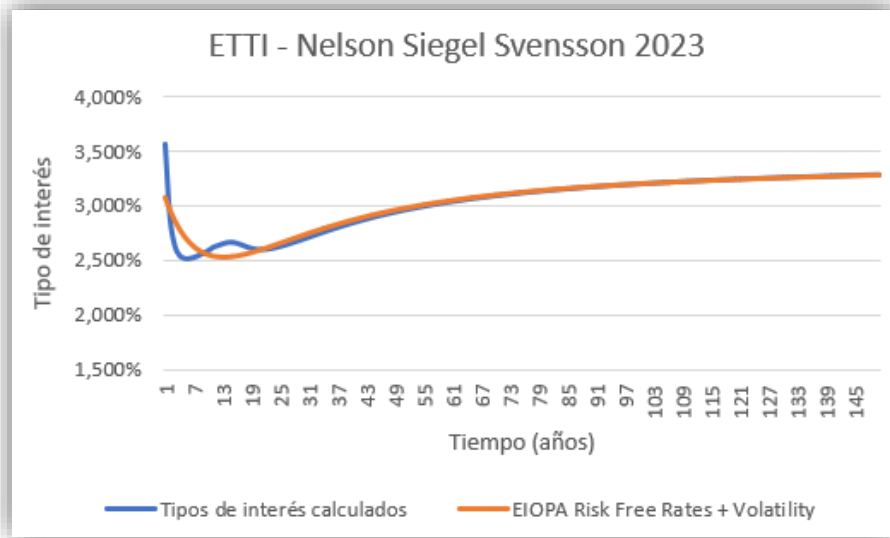


Gráfico 9: Estructura temporal de tipos de interés y estructura obtenida por Nelson Siegel para año 2023.
Fuente: Elaboración propia.

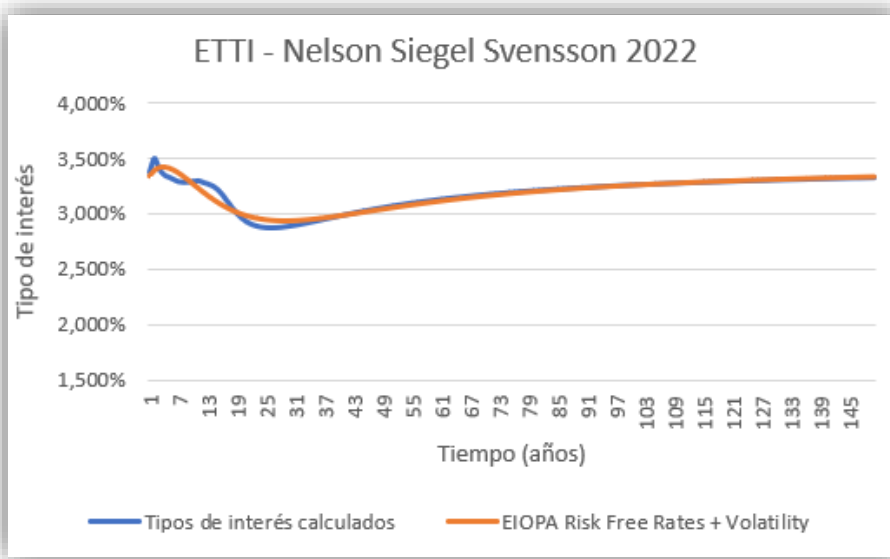


Gráfico 10: Estructura temporal de tipos de interés y estructura obtenida por Nelson Siegel para año 2022.
Fuente: Elaboración propia.

La modelización de la curva permitiría por tanto, la estimación de las tasas de interés a diferentes plazos mediante la optimización de parámetros que mejor se ajustan a los datos del mercado.

4.1.5. Ajuste de riesgo por riesgo no financiero

Una vez estos flujos han sido descontados, tal y como se indica en la normativa en los párrafos B86 a B92 deben ajustarse mediante el “ajuste por riesgo no financiero”³⁵ el cual incluye la incertidumbre sobre el importe no relacionado con el riesgo financiero (B89) y el calendario de los flujos de efectivo (Jiang, 2020). Dicha suma da lugar a los “flujos de efectivo para el cumplimiento de los compromisos del contrato de seguro”.

La normativa IFRS17 no indica un método concreto a seguir para el cálculo del ajuste de riesgo por riesgo no financiero. En este trabajo se ha decidido calcular el ajuste por riesgo no financiero a través del método de coste de capital previsto en la normativa. La normativa no indica exactamente qué riesgos deben tener en cuenta. En este trabajo se recurre al Art. 14 de la Ley 20/2015, de 14 de julio, de Ordenación, Supervisión y Solvencia de las entidades aseguradoras en el que se hace referencia a varios tipos de riesgos como el riesgo de suscripción, el riesgo de mercado, el riesgo de crédito, el riesgo operacional, el riesgo de liquidez y el riesgo de concentración (Jiang, 2020). Finalmente, para el cálculo del ajuste de riesgo por riesgo no financiero se elige aplicar el riesgo de suscripción. De acuerdo a la Ley 20/2015 el riesgo de suscripción se define como el riesgo de pérdida o de modificación adversa del valor de los compromisos derivados de la actividad aseguradora, debido a la inadecuación de las hipótesis de tarificación y constitución de provisiones. En concreto, se han realizado los cálculos como si se calculara el SCR del riesgo de suscripción para el ramo de vida previsto en Solvencia II aplicando las siguientes hipótesis (Alonso et al., 2008):

- Hipótesis riesgo de mortalidad: se basa en la premisa de que la tasa de mortalidad supera las expectativas previstas. Podría definirse como el riesgo de pérdida o de modificación adversa del valor de los compromisos contraídos en virtud de los seguros suscritos debido a variaciones en el nivel, la tendencia o la volatilidad de las tasas de mortalidad. Consiste en un incremento del 15% en las tasas de mortalidad para cada rango de edad y póliza suscrita.
- Hipótesis riesgo de gastos: se basa en la desviación de los gastos asociados aplicando un aumento del 10% de los gastos previstos. Formalmente, podría definirse como el riesgo de pérdida o de modificación adversa del valor de los compromisos contraídos en virtud de los seguros, debido a variaciones en el nivel, la tendencia o la volatilidad de los gastos de ejecución de los contratos de seguro o de reaseguro.
- Hipótesis riesgo de catástrofe: considera la posibilidad de que un evento catastrófico puede producir un aumento de la mortalidad. Para ello se le aplica un aumento de la tasa de mortalidad del 1,5 por mil. Podría definirse de forma precisa como el riesgo de pérdida o de modificación adversa del valor de los compromisos contraídos en virtud de los seguros, debido a una notable incertidumbre en las hipótesis de tarificación y constitución de provisiones correspondientes a sucesos extremos o extraordinarios.

Por tanto, lo que se lleva a cabo en este trabajo para el cálculo del ajuste por riesgo no financiero es la evaluación del contrato de seguro estudiado frente cada tipo de riesgo

³⁵ un ajuste de riesgo muy grande puede llevar a convertir un grupo de contratos aparentemente no onerosos en onerosos.

agregándose los valores resultantes a través de la matriz de correlaciones de Solvencia II. Se usa la matriz de correlaciones con el fin de saber como se relacionan entre sí los diferentes riesgos. Aplicando los shocks de Solvencia II no es necesario el cálculo del nivel de confianza ya que el marco regulador de Solvencia II está diseñado para garantizar que el Capital de Solvencia Requerido (SCR) se determine directamente con un nivel de confianza del 99,5%. Se adjunta a continuación la fórmula para este caso del SCR-suscripción de vida y la matriz de correlaciones (Tabla 20).

$$\text{SCR-suscripción de vida} = \sqrt{\sum_{i,j,k} \text{Corr}_{i,j,k} \times \text{SCR}_i \times \text{SCR}_j \times \text{SCR}_k}$$

Expresión 37: Fórmula SCR-suscripción de vida.
Fuente: Elaboración propia.

i \ j	Mortalidad	Gastos del seguro de vida	Catástrofe
Mortalidad	1	0,25	0,25
Gastos del seguro de vida	0,25	1	0,25
Catástrofe	0,25	0,25	1

Tabla 20: Matriz de correlaciones de Riesgo de suscripción Vida.
Fuente: Reglamento Delegado.

Así partiendo de los flujos de efectivo calculados anteriormente (provisiones) mediante las tablas PASEM 2020 para el primer asegurado del clúster tres aplicando el interés de EIOPA se obtiene los siguientes resultados mostrados en las Tablas 21 y 22:

Año 2022		
SCR Riesgo de mortalidad	SCR Riesgo de gastos	SCR Riesgo catastrófico
34.062,08	33.215,35	33.220,96

Tabla 21: Cálculo SCR's de Suscripción Vida para el año 2022.
Fuente: Elaboración propia.

Año 2023		
SCR Riesgo de mortalidad	SCR Riesgo de gastos	SCR Riesgo catastrófico
36.995,48	36.107,1422	36.113,203

Tabla 22: Cálculo SCR's de Suscripción Vida para el año 2023.
Fuente: Elaboración propia.

Seguidamente, se realizan las diferencias entre la provisión obtenida a través de las tablas PASEM 2020 y los diferentes SCR's y se aplica la matriz de correlaciones de Solvencia

II obteniéndose un SCR-Suscripción Vida de 895,24 para el ejercicio 2023 y un SCR-Suscripción Vida de 853,47 para el ejercicio 2022.

$$\text{Año 2023: } \frac{895,24}{36103,59} = 2,48\%$$

$$\text{Año 2022: } \frac{853,47}{33211,83} = 2,57\%$$

Posteriormente, se suman los flujos de efectivo futuros hasta la edad de 109 años (final Tablas PASEM 2020) ponderados por el SCR suscripción vida frente las provisiones calculadas mediante Tablas PASEM 2020 y tipo de interés de EIOPA. Dicho resultado, bajo Solvencia II, se aplica el factor CoC del 6% obteniéndose la cuantía necesaria para cubrir los riesgos futuros de los contratos de seguros. finalmente el ajuste por riesgo no financiero mediante el método del Coste por Capital (Tabla 23).

Ajuste por riesgo no financiero	
Año 2022	Año 2023
3.721,32	3.513,95

Tabla 23: Ajuste por riesgo no financiero.
Fuente: Elaboración propia.

4.1.6. Margen de servicio contractual

El margen de servicio contractual indica las pérdidas y ganancias esperadas de un grupo determinado de contratos de seguros (Lu R., 2022).

Una vez calculado el ajuste por riesgo no financiero se calcula la provisión matemática por Fair Value es decir, se vuelven a calcular las provisiones mediante Tablas PASSEM 2020 pero en este caso se utiliza el tipo de interés de mercado, en este caso, se elige el EUROSAP a 1 año.

Para el caso del EUROSAP, existen tipos de interés en diciembre de 2022³⁶ y 2023³⁷ a 1 año, 2 años, 3 años, 4 años, 5 años, 7 años, 10 años, 15 años, 20 años y 30 años. Para el caso que se está estudiando interesa obtener los tipos de interés año tras año para poder calcular el Fair Value. Para ello, se vuelve acudir a la fórmula de Nelson-Siegel-Svensson (Expresión 36) con sus respectivas restricciones y se optimizan los parámetros mediante Solver obteniéndose los parámetros mostrados en la Tabla 24:

³⁶ Ver 9.5.1. EUROSAP 2022.- Pág. 119.

³⁷ Ver 9.5.2. EUROSAP 2023.- Pág. 123.

Segmentación de riesgos y modelización actuarial en IFRS17

	b0	b1	b2	b3	tau1	tau2
2023	0,026	0,003	-2,020	1,863	78,092	72,468
2022	0,025	0,020	-0,348	0,338	0,839	0,822

Tabla 24: Parámetros Nelson Siegel Svesson para EUROSWAP 2022 y 2023.

Fuente: Elaboración propia.

El resultado obtenido del Fair Value utilizando el tipo de interés EUROSWAP se muestra en la Tabla 25.

Fair Value - EUROSWAP	
Año 2022	Año 2023
36.273,29	36.463,13

Tabla 25: Fair Value obtenido con la curva de tipo de interés EUROSWAP.

Fuente: Elaboración propia.

Si en lugar de elegirse el EUROSWAP se eligiera la deuda alemana del año 2022³⁸ y 2023³⁹ se obtendrían los siguientes valores de los parámetros (Tabla 26):

	b0	b1	b2	b3	tau1	tau2
2023	0,026	0,003	-1,005	0,980	3,973	3,993
2022	0,012	0,019	-0,941	0,960	6,644	6,866

Tabla 26: Parámetros Nelson Siegel Svesson para Bono Alemán 2022 y 2023.

Fuente: Elaboración propia.

El resultado obtenido del Fair Value utilizando el tipo de interés BONO ALEMÁN se muestra en la Tabla 27.

Fair Value – BONO ALEMÁN	
Año 2022	Año 2023
35.871,62	38.558,77

Tabla 27: Fair Value obtenido con la curva de tipo de interés Bono Alemán.

Fuente: Elaboración propia.

Finalmente, se obtiene el CSM o el Loss Component (LC) por diferencia entre el resultado obtenido entre Fair Value y el justo por riesgo no financiero más el Future Value (Ilustración 1). Dicha diferencia recibe el nombre de CSM cuando el Fair Value es

³⁸ Ver 9.5.4. BONO ALEMÁN 2022.- Pág. 130.

³⁹ Ver 9.5.5. BONO ALEMÁN 2023.- Pág. 133.

superior a la suma del ajuste por riesgo no financiero más el Future Value. La diferencia recibe el nombre de Loss Component cuando el ajuste por riesgo no financiero más el Future Value es superior al Fair Value.

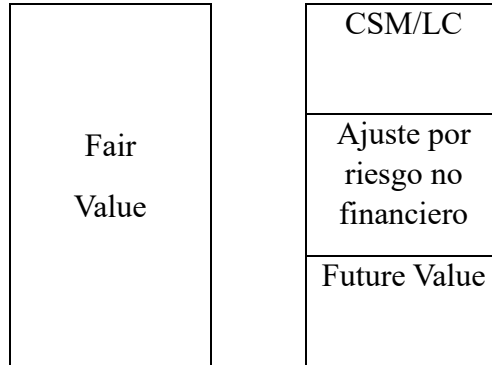


Ilustración 1: Cálculo CSM/LC.
Fuente: Elaboración propia.

El CSM obtenido según si el Fair Value ha sido calculado a través de la curva de tipos EUROSWAP o a través del BONO ALEMÁN se muestra en la Tabla 28 que sigue:

	Future Value (FV)	Risk Adjustment (RA)	FV + RA	EUROSWAP	LC/CSM
2022	33.211,83	3.721,32	36.933,15	36.273,29	LC
2023	36.103,59	3.513,95	39.617,54	36.463,13	LC

Tabla 28: Indicación de si hay LC o CSM en 2022 y 2023 mediante el EUROSWAP.
Fuente: Elaboración propia.

Tal y como se observa en la Tabla 28 para el ejercicio 2022 si se utiliza para el cálculo del Fair Value la curva de tipo de interés EUROSWAP existe un Loss Component ya que la suma del Future Value y del Risk Adjustment es superior al Fair Value calculado con la curva EUROSWAP.

Por otro lado, para el ejercicio 2023 si se utiliza para el cálculo del Fair Value la curva de tipo de interés EUROSWAP existe un Loss Component ya que la suma del Future Value y del Risk Adjustment es superior al Fair Value calculado con la curva EUROSWAP.

	Future Value (FV)	Risk Adjustment (RA)	FV + RA	BONO ALEMÁN	LC/CSM
2022	33.211,83	3.721,32	36.933,15	35.871,62	LC
2023	36.103,59	3.513,95	39.617,54	38.558,77	LC

Tabla 29: Indicación de si existe LC o CSM en 2022 y 2023 mediante el Bono Alemán.
Fuente: Elaboración propia.

Tal y como se observa en la Tabla 29 para el ejercicio 2022 si se utiliza para el cálculo del Fair Value la curva de tipo de interés BONO ALEMÁN existe un Loss Component ya que la suma del Future Value y del Risk Adjustment es superior al Fair Value calculado con la curva BONO ALEMÁN.

Asimismo, si se observa la Tabla 29 para el ejercicio 2023 si se utiliza para el cálculo del Fair Value la curva de tipo de interés BONO ALEMÁN existe un Loss Component ya que la suma del Future Value y del Risk Adjustment es superior al Fair Value calculado con la curva BONO ALEMÁN.

Ahora bien, si el resultado del sumatorio de los flujos de efectivo futuros hasta la edad de 109 años (final Tablas PASEM 2020) ponderados por el SCR suscripción vida frente las provisiones calculadas mediante Tablas PASEM 2020 y tipo de interés de EIOPA y dicho resultado se pondera con un 4% en lugar del 6% como se ha hecho anteriormente se obtienen los siguientes resultados mostrados en las Tablas 30 y 31:

	Future Value (FV)	Risk Adjustment (RA)	FV + RA	EUROSWAP	LC/CSM
2022	33.211,83	2.480,88	35.692,71	36.273,29	CSM
2023	36.103,59	2.342,63	38446,22	36.463,13	LC

Tabla 30: Indicación de si existe LC o CSM en 2022 y 2023 mediante el EUROSWAP ponderado al 4%.

Fuente: Elaboración propia.

	Future Value (FV)	Risk Adjustment (RA)	FV + RA	BONO ALEMÁN	LC/CSM
2022	33.211,83	2.480,88	35.692,71	35.871,62	CSM
2023	36.103,59	2.342,63	38446,22	38.558,77	CSM

Tabla 31: Indicación de si existe LC o CSM en 2022 y 2023 mediante el BONO ALEMÁN ponderado al 4%.

Fuente: Elaboración propia.

Por otro lado, para los ejercicios 2022 y 2023 si se utiliza para el cálculo del Fair Value la curva de tipo de interés EUROSWAP pero en lugar de utilizar el de diciembre 2022 y 2023 se utiliza para ambos ejercicios la de enero⁴⁰ de 2022 bajo la hipótesis que se mantiene constante para ambos años los resultados serían los que se muestran en la Tabla 32 que sigue:

	Future Value (FV)	Risk Adjustment (RA)	FV + RA	EUROSWAP	LC/CSM
2022	33.211,83	3.721,32	36.933,15	50.289,13	CSM
2023	36.103,59	3.513,95	39.617,54	50.563,34	CSM

Tabla 32: Indicación de si existe LC o CSM en 2022 y 2023 mediante EUROSWAP de enero de 2022.

Fuente: Elaboración propia.

⁴⁰ Ver 9.5.3. EUROSWAP ENERO 2022.- Pág. 126.

5. CONCLUSIONES

La normativa IFRS17 es una normativa contable que entró en vigor en enero de 2023 y representa un hito significativo en la contabilidad de seguros, estableciendo un marco uniforme y transparente que mejora la comparabilidad y transparencia de los estados financieros. Es una normativa que se centra en la parte del pasivo del balance estableciendo un conjunto de principios con el fin de medir mejor las obligaciones derivadas de los contratos de seguros. Esto es especialmente importante en los seguros de vida, sobre el cual trata este trabajo, puesto que suelen ser contratos de larga duración siendo necesario la valoración oportuna de riesgos e incertidumbre asociados.

La normativa establece tres enfoques: Building Block Approach (BBA), Premium Allocation Approach (PPA) y Variable Fee Approach (VFA). El primer enfoque es de aplicación obligatoria en rentas vitalicias, seguros de ahorro a largo plazo, seguros de vida temporales superiores a un año y productos de vida entera. El segundo enfoque es de carácter opcional y se aplica en seguros de No Vida y Vida riesgo anuales renovables. El tercer enfoque es de aplicación obligatoria en productos Unit Linked y seguros de ahorro con participación en beneficio (sustanciales).

Este trabajo se centra en el estudio desde el punto de vista actuarial de la normativa IFRS17 aplicada a un seguro de vida entera caracterizado por:

- el asegurado paga primas hasta la edad de jubilación (67 años)
- el asegurado si fallece antes de los 67 años existe un seguro de reembolso de primas
- si el asegurado fallece después de los 67 años, el beneficiario del recibe una cuantía estipulada en antelación.

En este sentido, de acuerdo a las características mencionadas de este tipo de producto se ha hecho necesario aplicar el enfoque general o de bloques utilizado. De un portafolio de vida formado por 67.102 contratos se ha seleccionado un contrato a partir del cual se hace todo el análisis detallado y permite ilustrar cómo los principios de IFRS17 se pueden llevar a la práctica de manera detallada pudiéndose extrapolar estos procedimientos a otros contratos del portafolio. De esta manera, permite al lector obtener una visión clara y estructurada de los pasos necesarios para cumplir con los requisitos normativos.

Llevar a cabo este trabajo sobre IFRS17 aplicada en un contrato de seguros de vida me ha permitido llegar a varias conclusiones.

Una de las primeras observaciones destacadas ha sido la utilidad de la implementación de las técnicas de clúster una vez los contratos de seguros han sido agrupados por cohortes y por producto con el fin de obtener grupos formados por contratos de seguros con riesgos similares y que sean el máximo homogéneos posible. En este sentido se observa que no debe llevarse a cabo un análisis de clúster tradicional quitando los outliers sino que deben elegirse aquellas técnicas de clustering que resultan poder ser óptimas en este contexto, es decir, aplicar aquellas técnicas que asignan cada asegurado a un solo clúster como son los de tipo particional o los de tipo jerárquico. Estos últimos debe tenerse en cuenta que

pueden surgir problemas durante la ejecución del ordenador en el caso de que existan un número elevado de datos, por lo que se aconseja más la utilización de los primeros.

En el contexto de análisis de clúster es fundamental señalar la importancia que tiene la selección de variables al inicio. En este sentido, en este trabajo al tratarse de un seguro vida se han considerado que aquellas variables más relevantes para hacer las agrupaciones eran la edad de asegurado y el capital. Otro de los aspectos observados ha sido que la estandarización de los datos hace que sea más difícil interpretar los resultados del clúster, pero por otro lado fomenta una distribución más equilibrada de los grupos obtenidos mediante clustering en los contratos de seguros agrupados.

Por otro lado, es esencial destacar que bajo IFRS17 el cálculo del Future Value debe hacerse mediante las tablas biométricas en vigor al año que se están calculando o en el caso de que la compañía tenga las suyas propias también pueden ser utilizadas. Además, bajo normativa IFRS17 debe utilizarse la curva de tipos de interés que se haya construido para descontar los flujos. En este trabajo debido a la extensión marcada no se ha centrado en la construcción de dicha curva, sino que se ha elegido aplicar la que ya se utiliza en Solvencia II proporcionada por EIOPA en diciembre de 2022 y 2023 con la volatilidad incluida.

La normativa IFRS17 deja la libertad de aplicar la metodología que se crea conveniente para el cálculo del Risk Adjustment sugiriendo el VaR, TVaR y el CoC así como el nivel de confianza. En este trabajo se elige el método del CoC y siguiendo Solvencia II se replican las hipótesis del riesgo de suscripción de vida: choque de mortalidad, choque de gastos y choque de catástrofe. Finalmente, se aplica la matriz de correlaciones y se calcula el porcentaje que representa el valor obtenido sobre el Future Value. Este porcentaje se aplicará sobre los flujos de efectivo futuros obteniéndose finalmente el Risk Adjustment. Bajo la metodología del Coste de Capital (CoC) prevista en la normativa se obtiene el Risk Adjustment total sumando todos los Risk Adjustments obtenidos para cada año ponderados los cuales se aplica bajo Solvencia II un porcentaje entre 5% y 8%. En un primer caso se elige aplicar 4% y luego se decide aplicar 6%. Mediante la aplicación del 4% sobre el sumatorio de Risk Adjustments individuales ha llevado la obtención de CSM mientras que la aplicación del 6% sobre el sumatorio de Risk Adjustments individuales ha llevado la obtención de Loss Component lo que sugiere que la aplicación de un porcentaje u otro puede hacer variar significativamente los resultados de si se obtiene CSM o Loss Coponent.

De igual manera la obtención de CSM o Loss Component está directamente vinculada a la curva de tipo de interés elegida para calcular el Fair Value previsto en IFRS13. El Fair Value consiste en calcular cuánto valdrían a valor de mercado mis pasivos. En este trabajo se ha elegido para hacer la valoración la curva de tipo de interés de mercado del EUROSWAP y del BONO ALEMÁN las cuales ya incorporan las primas de iliquidez correspondientes. Cuando se ha utilizado el EUROSWAP y el BONO ALEMÁN de diciembre 2022 y 2023 se ha obtenido Loss Component, pero cuando se ha utilizado el EUROSWAP de enero 2022 se ha obtenido CSM para ambos ejercicios lo que sugiere que también la elección de la curva de tipos de interés puede hacer variar mucho el resultado del Fair Value y el resultado global según si es CSM o Loss Component.

En el trabajo también se ha calculado el balance bajo IFRS4. La Dirección General de Seguros y Fondos de Pensiones establece para cada año un tipo de interés máximo a aplicar para el cálculo de provisiones. En el caso del ejercicio 2022 este tipo de interés era del 0,46% y para el ejercicio 2023 este tipo de interés era del 1,88%. Las provisiones bajo IFRS4 han sido calculadas en primer lugar para estos dos ejercicios al tipo de interés fijo establecido al contrato de seguro del 2%. De acuerdo IFRS4 y el Real Decreto 2486/1998, de 20 de noviembre, por el que se aprueba el Reglamento de Ordenación y Supervisión de los Seguros Privados debe llevarse a cabo la dotación de provisiones con cuantía igual a la diferencia entre los resultados obtenidos entre los dos tipos de interés (Tabla 33).

	Provisión calculada a través de IFRS4 con tipo interés 2%	Provisión calculada con el tipo de interés marcado por la DGSFP	Dotación de provisión
2022	39.185,60	50.764,47	11.578,87
2023	39.763,17	40.512,53	749,35

Tabla 33: Dotación de provisión.
Fuente: Elaboración propia.

Finalmente, como líneas futuras de investigación se sugiere ampliar el estudio de la normativa IFRS17 a otros tipos de productos de seguro explorando cómo se aplica y se adapta a diferentes situaciones. Además, sería beneficioso evaluar el impacto de su implementación, analizando la transparencia y comparabilidad de los estados financieros de las empresas de seguros, así como la gestión del riesgo. Debido a que el tema de este trabajo es muy extenso, no se ha tenido en cuenta factores como la inflación y la caída de cartera dejando otras posibles líneas futuras abiertas. Además, se podrían realizar otras hipótesis, la modificación de las planteadas así como el uso de otras metodologías diferentes a las aplicadas.

A través de este trabajo, se ha evidenciado la complejidad y el rigor que exige la aplicación de esta normativa. Se ha dedicado una cantidad de tiempo considerable a este proyecto, y aunque se hayan realizado algunas simplificaciones ofrece una visión general práctica de la aplicación real de IFRS17 en un contrato de vida entera a través del enfoque general o de bloques. Este trabajo se ha llevado a cabo con una dedicación excepcional, ilusión y coraje en el que durante el proceso se han tenido que enfrentar momentos de dificultad y obstáculos que han requerido un esfuerzo adicional para superarlos con éxito. Aprender sin conocimientos previos sobre el tema, ha sido un desafío significativo que da un gran valor tanto en este trabajo como a nivel personal y profesional aprendiendo a buscar soluciones. A nivel de conocimientos, además, me ha brindado una gran oportunidad de conocer la normativa proporcionándome una base sólida estando preparada para la aplicación práctica en situaciones reales que de bien seguro será de gran utilidad a mi carrera profesional.

6. AGRADECIMIENTOS

Quiero expresar mi más sincera gratitud a todas las personas que han hecho posible la realización de este trabajo.

En primer lugar, a mis padres, para creer en mí y brindarme las herramientas necesarias para llegar hasta aquí.

También quiero expresar mi agradecimiento a mi tutora, la Dra. Mercedes Ayuso por su guía, apoyo y dedicación.

Finalmente, quiero dar las gracias a mis compañeros de prácticas, y especialmente a Fani, por sus valiosos consejos y orientación.

A la Universidad y a todos aquellos que me han formado durante mi vida académica.

7. ABREVIATURAS

IASB: *International Accounting Standards Board*

IFRS: *Financial Reporting Standards*

PAA: *Premium Allocation Approach*

VFA: *Variable Fee Approach*

RA: *Risk Adjustment*

LS: *Loss Component*

CSM: *Contractual Service Margin*

NIIF: Normas Internacionales de Contabilidad Financiera

NIC: Normas Internacionales de Contabilidad

IAS: *International Accounting Standards*

ISSB: *International Sustainability Standards Board*

UE: Unión Europea

EEE: Espacio Económico Europeo

CLARA: *Clustering Large Application*

ROS: *Random Over Sampling*

AR: Ajuste de Riesgo

LC: *Loss Component*

GMM: Modelo de medición general

VaR: *Value at Risk*

CTE: *Conditional Tail Expected*

CAR: *Cumulative Abnormal Return*

US GAAP: *Generally accepted accounting principles*

NASDAQ: *National Association of Securities Dealers Automated Quotations*

HAC: *Hierarchical Agglomerative Clustering*

UPGMA: *Unweighted Pair Group Method with Arithmetic Mean*

GKMF95: *Generationen-Kohorten-Mischungstafel 1995*

GGI: Gastos de gestión interna

GGE: Gastos de gestión externa

VAA: Valor Actual Actuarial

No Rel: No Relacionadas

EIOPA: *European Insurance and Occupational Pensions Authority*

SCR: *Solvency Capital Requirement*

TVaR: *Tail Value at Risk*

8. BIBLIOGRAFÍA

Legislación

Real Decreto de 22 de agosto de 1885 por el que se publica el Código de Comercio. (en línea). Consultada el 11/05/2024. Disponible en:

<https://www.boe.es/buscar/act.php?id=BOE-A-1885-6627>

Ley 20/2015, de 14 de julio, de ordenación, supervisión y solvencia de las entidades aseguradoras y reaseguradoras. (en línea). Consultada el 5/04/2024. Disponible en:

<https://www.boe.es/buscar/act.php?id=BOE-A-2015-7897>

Normativa IFRS17. Consultada el 10/02/2024. Disponible en:

<https://www.icac.gob.es/sites/default/files/2022-01/IFRS%2017%20Diciembre%2021-redacci%C3%B3n%201-1-23.pdf>

Libros consultados

Normas Internacionales de Contabilidad (teoría)

Mallo C. y Pulido A., (2006). *Normas Internacionales de Contabilidad (NIIF). Las Normas Internacionales de Información Financiera*. Ed. Thomson. Madrid (España).

Vázquez N. y Díaz M. Prólogo por Cano J.A. y revisión técnica por Oliveras E., (2013). *Normas Internacionales de Información Financiera (NIIF-IFRS). Principios y ejercicios avanzados de contabilidad global*. Ed. Profit. Barcelona (España).

Selfridge I., (2018). *Manual of Accounting IFRS 2021 – Volume 2*. Ed. PricewaterhouseCoopers LLP. CPI Group.

Estadística actuarial vida

Ayuso, M., Corrales, H., Guillén, M., Martín, A.P., Rojo, J., (2007). *Estadística actuarial vida*. Publicacions i Edicions UB.

Test Normalidad y correlaciones

Delgado de la Torre. (2007). *Probabilidad y estad.* Ed. Delta. Madrid (España)

Técnicas actuariales en Solvencia II

Alonso P. e Albarrán I., (2008). *Análisis del riesgo en seguros en el marco de Solvencia II: Técnicas estadísticas avanzadas Monte Carlo y Bootstrapping*. Ed. Fundación Mapfre. Madrid (España).

Técnicas y algoritmos de Clúster

Kaufman L., Rousseeuw P.J., (1987). *Clustering by Means of Medoids*. Ed. Faculty of Mathematics and Informatics. Delft (Países Bajos).

Everitt B.S., Landau S., Leese M. y Stahl D. (2011). *Wiley series in probability and statistics*. King's College London (UK).

Murphy, K. P. (2022). *Probabilistic machine learning: an introduction*. Ed. The MIT Press. Cambridge, Massachusetts.

Murphy, K. P. (2012). *Machine Learning: a probabilistic perspective*. Ed. Massachusetts Institute of Technology. Massachusetts (United States of America).

Géron A., (2019) Segunda edición. *Aprende Machine Learning con Scikit-Learn, Keras y TensorFlow*. Ed. Anaya Multimedia. Madrid (España).

Raschka S. y Mirjalili V., (2019) Segunda edición en español. *Python Machine Learning*. Ed. Marcombo. Madrid (España).

IFRS9, 17, Basilea III y Solvencia II

Pérez, J., et al., 2018. *Banca y Seguros. Capital y Contabilidad. IFRS9, 17, Basilea III y Solvencia II*. Ed. Marcial Pons. Madrid (España).

Páginas web consultadas

Diferencias entre Correlación de Spearman, Tau de Kendall y Pearson (en línea). Consultada el 11/05/2024. Disponible en:

<http://www.postgradovipi.50webs.com/archivos/agrollania/2016/agro8.pdf>

Tipo de interés EUROSWP (en línea). Consultada el 19/05/2024. Disponible en:

https://cliente bancario.bde.es/pcb/es/menuhorizontal/productoservici/relacionados/tipo_sinteres/guiatextual/tiposinteresrefe/tabla_tipos_referencia_oficiales_mercado_hipotecario.html

Tipos de interés del Bono Alemán (en línea). Consultada el 19/05/2024. Disponible en:

<https://www.bundesbank.de/en/statistics/money-and-capital-markets/interest-rates-and-yields/daily-yields-of-current-federal-securities-772220>

IASB. Consultada el 22/02/2024. Disponible en:

<https://www.ifrs.org/content/dam/ifrs/about-us/who-we-are/who-we-are-spanish-v2.pdf>

Artículos consultados

Baidari I. y Patil C., (2020). *A Criterion for Deciding the Number of Clusters in a Dataset Based on Data Depth*. Vietnam journal of Computer Science. Vol. 07, No. 04, pp. 417-431 (2020).

Balcan M.F., Liang Y., Gupta P., (2014). *Robust Hierarchical Clustering*. Journal of Machine Learning Research 15 (2014) 4011-4051.

Cabedo, J.M. y Tirado, J.M., (2003). *The disclosure of risk in financial statements*. Accounting Forum 28 (2004) 181-200.

Chevallier F., Dal Moro E., Krvavych Y. y Rudenko I., (2018). *Probability of Sufficiency of the Risk Margin for Life Companies Under IFRS17*. SSRN: <https://ssrn.com/abstract=3192502>.

Dahiyat A. y Owais W., (2020). *The expected impact of applying IFRS (17) insurance contracts on the quality of financial reports*. Accounting 7 (2021) 581–590.

Dalmajjer, E.S., Nord, C.L. y Astle D.E., (2022). *Statistical power cluster analysis*. BMC Bioinformatics.

Devale, A.B. y Kulkarni, R.V., (2012). *Applications of data mining techniques in life insurance*. International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.2, No.4.

England P.D., Verrall R.J. y Wüthrich M.V., (2019). *On the lifetime and one-year views of reserve risk, with application to IFRS17 and Solvency II risk margins*. Insurance: Mathematics and Economics 85 (2019) 74–88.

García D., Martínez F.J., y Antón M., (2019). *Principales riesgos que afectan a las empresas*. Revista de Contabilidad y Dirección Vol. 28, año 2019, pp. 11-26.

Hardy A., (1996). *On the number of clusters*. Computational Statistics & Data Analysis 23 (1996) 83-96.

Jain, A.K., Murty M.N., y Flynn, P.J., (1999). *Data Clustering: A Review*. ACM Computing Surveys, 31, 264-323.

Khan U., Li B., Rajgopal S., y Venkatachalam M., (2014). *Do the FASB's Standards Add* Kiermayer M. y Weiß (2019). *Grouping of Contracts in Insurance using Neural Networks*. Scandinavian Actuarial Journal. Volume 2021. 2021 – Issue 4.

Kodinariya T. y Makwana P., (2013). *Review on Determining of Cluster in K-means Clustering*. International Journal of Advance Research in Computer Science and Management Studies. Volume 1, Issue 6, November 2013.

Liao, T.W., (2005). *Clustering of Time-Series Data—A Survey*. Pattern Recognition, 38, 1857-1874.

Longoni P., (2019). *IFRS17 Insurance Contracts and Firm Value*. SSRN: <https://ssrn.com/abstract=3589560> y <http://dx.doi.org/10.2139/ssrn.3589560>.

Lu R., Wang J.J., y Pan S.H., (2022). *A Study on the Risk and Return Sharing Mechanism for Interest-sensitive Life Insurance Products under IFRS17 Accounting*. *Advances in Decision Sciences*, 2022, vol. 26, issue Special, 40-52.

Møllersen K., Dhar S.S., y Godtlielsen F., (2016). *On Data-Independent Properties for Density-Based Dissimilarity Measures in Hybrid Clustering*. *Applied Mathematics*, Vol.7 No.15.

Nainggolan R., Perangin-angin R., Simarmata E., Tarigan F.A., (2019). *Improved the Performance of the K-Means Cluster Using the Sum of Squared Error (SSE) optimized by using the Elbow Method*. *J. Phys.: Conf. Ser.* 1361 012015.

Owais W.O., y Dahiyat A.A., (2021). *Readiness and Challenges for Applying IFRS17 (Insurance Contracts): The Case of Jordanian Insurance Companies*. *Journal of Asian Finance, Economics and Business* Vol 8 No 3 (2021) 0277–0286.

Oyewole, G.J. y Thopil, G.A., (2022). *Data clustering: application and trends*. *Artificial Intelligence Review* 56(1).

Palmborg L., Lindholm M. y Lindskog F., (2021). *Financial position and performance in IFRS17*. *Scandinavian Actuarial Journal*, Volume 2021, 2021 – Issue 3.

Pedra T., Heitor C. y Da Rocha C., (2022). *Assessing Risk Adjustment for Non-Life Insurance under IFRS 17 (Short Version)*. *R. Bras. Risco e Seg.*, Rio de Janeiro, v. 16, n. 28, Ed. Especial p. 1-12, Jan. 2022/Dez. 2022.

Prima A. y Teguh D., (2024). *Classifying Risks On Motor Insurance Policies For IFRS 17 Implementation In General Insurance Companies*. *ITM Web Conf.* Volume 58, 2024. The 6th IndoMS International Conference on Mathematics and Applications (The 6th IICMA 2023).

Rodriguez M.Z., Comin C.H., Casanova D., Bruno O.M., Amancio D.R., Costa L.dF., Rodrigues F.A., (2019). *Clustering algorithms: A comparative approach*. *PLoS ONE*: <https://doi.org/10.1371/journal.pone.0210236>

Saputra D.M., Saputra D. y Oswari L., (2019). *Effect of Distance Metrics in Determining K-Value in KMeans Clustering Using Elbow and Silhouette Method*. Advances in Intelligent Systems Research, volume 172 Sriwijaya International Conference on Information Technology and Its Applications (SICONIAN 2019).

Sekula M., Datta S., Datta S., (2017). *optCluster: an R package for determining the optimal clustering algorithm*. Bioinformatics. 2017 Mar 31;13(3): 101-103.

Khan U., Li B., Rajgopal S. y Venkatachalam M., (2017). *Do the FASB's Standards Add Shareholder Value?*. The accounting Review, Forthcoming, Columbia Business School Research Paper No. 17-82.

Shutaywi M. y Kachouie N. N., (2021). *Silhouette Analysis for Performance Evaluation in Machine Learning with Applications to Clustering*. Entropy 2021, 23, 759.

Sinaga K.P. y Yang M., (2020). *Unsupervised K-Means Clustering Algorithm*. Open Access Journal. IEEE Access. ACCESS.2020_2988796.

Syakur M.A., Klotimah B.K., Rochman B.D. (2018). *Integration K-Means Clustering Method and Elbow Method For Identification of The Best Customer Profile Cluster*. IOP Conf. Series: Materials Science and Engineering 336 (2018) 012017.

Winkler M. y Kansal S. (2021). *Actuarial Challenges and IFRS17*. SSRN: <https://ssrn.com/abstract=3670808> y <http://dx.doi.org/10.2139/ssrn.3670808>.

Yousuf W., Stansfield J., Malde K., Mirin N., Walton R., Thorpe B., Thorpe J., Iftode C., Tan L., Dyble R., Pelsser A., Ghosh A., Qin W., Berry T. y Er C., (2021). *The IFRS 17 contractual service margin: a life insurance perspective*. British Actuarial Journal (2021), Vol. 26, e2, pp. 1–10.

9. ANEXOS

9.1. Anexos contabilidad y antecedentes de la normativa IFRS17

En este Anexo 9.1. se incluye la Tabla 34 en la que se recogen las Normas Internacionales de Información Financiera (NIIF) vigentes. Esta normativa es emitida por la Junta de Normas Internacionales de Contabilidad (IASB). La normativa IFRS17 está enfocada en la parte de Pasivo del balance e interacciona en numerosos aspectos con NIIF 9, relacionado con la parte de Activo, con la NIIF 13 en la valoración a Valor Razonable y la NIIF 4 de la contabilidad nacional.

NIIF IFRS	Nombre
NIIF 1	Adopción por primera vez de las Normas Internacionales de Información Financiera.
NIIF 2	Pago basado en acciones
NIIF 3	Combinaciones de negocios
NIIF 4	Contratos de seguro
NIIF 5	Activos no corrientes mantenidos para la venta y operaciones discontinuadas
NIIF 6	Exploración y evaluación de recursos minerales
NIIF 7	Instrumentos financieros: información a revelar
NIIF 8	Segmentos operativos
NIIF 9	Instrumentos financieros
NIIF 10	Estados financieros consolidados
NIIF 11	Acuerdos conjuntos
NIIF 12	Información a revelar sobre participaciones en otras entidades
NIIF 13	Medición del valor razonable
NIIF 14	Cuentas de diferimientos de actividades reguladas
NIIF 15	Ingresos procedentes de contratos con clientes
NIIF 16	Arrendamientos
NIIF 17	Contratos de seguros (sustituye la NIIF 4)

Tabla 34: Listado NIF.

Fuente: Elaboración propia.

9.2. Anexos análisis descriptivo

En este Anexo 9.2. relacionado con el análisis descriptivo de la base datos se presentan dos apartados diferenciados. El primer apartado, trata sobre un estudio profundizado de la base de datos general, en el cual se presentan las frecuencias absolutas y relativas de los datos. El segundo apartado, se muestran los gráficos de dispersión de los contratos de seguro agrupados por año de fecha efectiva del contrato, así como los estadísticos descriptivos básicos para cada grupo.

9.2.1. Portafolio general

Para las variables categóricas pueden observarse como estadísticos descriptivos la frecuencia absoluta (Tabla 35) y relativa (Tabla 36) de cada categoría, lo que proporciona una visión detallada de la prevalencia de cada grupo.

Frecuencias absolutas	Sexo	M	47,652
		F	28,450
	Ageing quarter(r)	1	18,903
		2	19,307
		3	19,003
		4	18,889
	Seasonal quarter (S)	1	20,323
		2	19,023
		3	14,888
		4	21,868
	Renewal Month (Month)	1	5,988
		2	6,155
		3	7,314
		4	6,089
		5	6,114
		6	6,735
7		6,437	
8		4,368	
9		4,517	
10		5,679	
11		4,517	
12		4,368	

Tabla 35: Frecuencias absolutas.
Fuente: Elaboración propia.

Frecuencias relativas	Sexo	M	0,626
		F	0,374
	Ageing quarter (r)	1	0,248
		2	0,254
		3	0,250
		4	0,248
	Seasonal quarter (S)	1	0,267
		2	0,250
		3	0,197
		4	0,287
	Renewal Month (Month)	1	0,079
		2	0,081
		3	0,096
		4	0,080
		5	0,080
		6	0,089
7		0,085	
8		0,057	
9		0,060	
10		0,075	
11		0,093	
12		0,126	

Tabla 36: Frecuencias relativas.
Fuente: Elaboración propia.

Si se visualizan la Tabla 36 correspondientes a las frecuencias absolutas y relativas de las variables Sexo, Ageing quarter, Seasonal quarter y Renewal Month del portafolio se destaca que por lo que se refiere el Sexo hay más hombres(62,6%) que mujeres (37,4%) equivaliendo a 47.652 y 28.450 asegurados respectivamente. La variable Ageing quarter que representa el trimestre del año en el que se encuentra el asegurado en el momento de renovación de la póliza las observaciones están equidistribuidas entre los diferentes momentos del año 1Q, 2Q, 3Q y 4Q. En el caso de las variable Seasonal quarter que indica el trimestre estacional en el que ocurre la renovación de la póliza también parece ser que las observaciones se encuentran distribuidas equitativamente, siendo ligeramente inferior las pertenecientes al verano. En relación a la variable Renewal Month el mes que hay más renovaciones es el marzo y el que hay menos renovaciones es el agosto y diciembre.

En el contexto de este estudio sobre el seguro de vida ahorro, las tablas de contingencia son esenciales para examinar las relaciones entre variables categóricas. Estas tablas permiten visualizar la distribución conjunta de las dos categorías de una variable en relación con las categorías de la otra variable. Con la finalidad de conocer mejor las

posibles interacciones entre variables se han realizado las tablas de contingencias que se muestran a continuación (Tablas 37 i 38).

	Invierno	Primavera	Verano	Otoño
Mujer	7495	7108	5590	8257
Hombre	12828	11915	9298	13611

Tabla 37: Tabla contingencia entre Género y Estación.
Fuente: Elaboración propia.

Observando la Tabla 37 se concluye que, tanto por las mujeres como los hombres, la estación que hay más renovaciones es en otoño y la estación que hay menos renovaciones es el verano.

	Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
Inv	5988	6155	7314	866								
Prim				5223	6114	6735	951					
Ver							5486	4368	4517	517		
Oto										5162	7089	9617

Tabla 38: Tabla contingencia entre estación y mes de renovación.
Fuente: Elaboración propia.

Los resultados que aparecen en la Tabla 38 muestran diferencias en las renovaciones a lo largo del año, con una alta concentración en invierno, particularmente en enero (5988) y febrero (6155), y en otoño, especialmente en noviembre (7089) y diciembre (9617). La primavera presenta un pico en marzo (7314) y un incremento gradual en los meses siguientes, mientras que el verano registra niveles moderados de renovaciones, siendo julio (4368) y agosto (4517) los meses con menor actividad. Se puede concluir que los inviernos y otoños son periodos de mayor actividad de renovación, y veranos y primaveras tienen volúmenes más variables y generalmente menores.

9.2.2. Contratos agrupados por fecha efectiva del contrato de seguro

En este subapartado 9.1.2. tal y como se ha explicado anteriormente se muestran los gráficos de dispersión y los estadísticos descriptivos para los asegurados agrupados por fecha efectiva del contrato de seguro. En general, se observa en cada fecha efectiva del contrato de seguro outliers. El Capital y la variables relacionadas con la edad (Age, Age_Actuarial y Age_Actuarial_quarter) tienen medias parecidas.

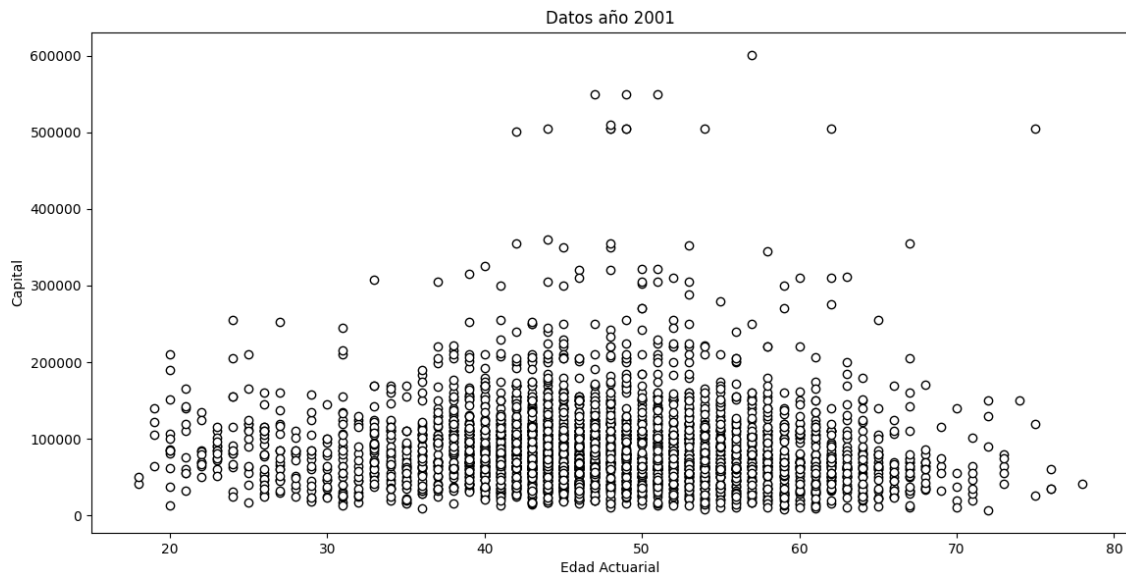


Gráfico 11: Gráfico de dispersión variables capital y edad actuarial para el año 2000.
Fuente: Elaboración propia.

Estadístico	Año 2001					
	Capital	Age	T	Age Actuarial	x	Age_actuarial_quarter
Media	89856,418	47,4188	0,495	47,407	0,54	47,29
Mínimo	7000	17,91	0	18,00	0	17,00
25%	55000	41,691	0,237	42,00	0,304	41,50
50%	80000	46,942	0,479	47,00	0,529	46,75
75%	110000	53,975	0,755	54,00	0,805	53,75
Máximo	601000	77,941	0,999	78,00	0,997	77,00
Desviación Estándar	58038,183	9,764	0,293	9,769	0,292	9,77

Tabla 39: Estadísticos básicos de los datos de los contratos de seguro cuya fecha efectiva es el año 2001.
Fuente: Elaboración propia.

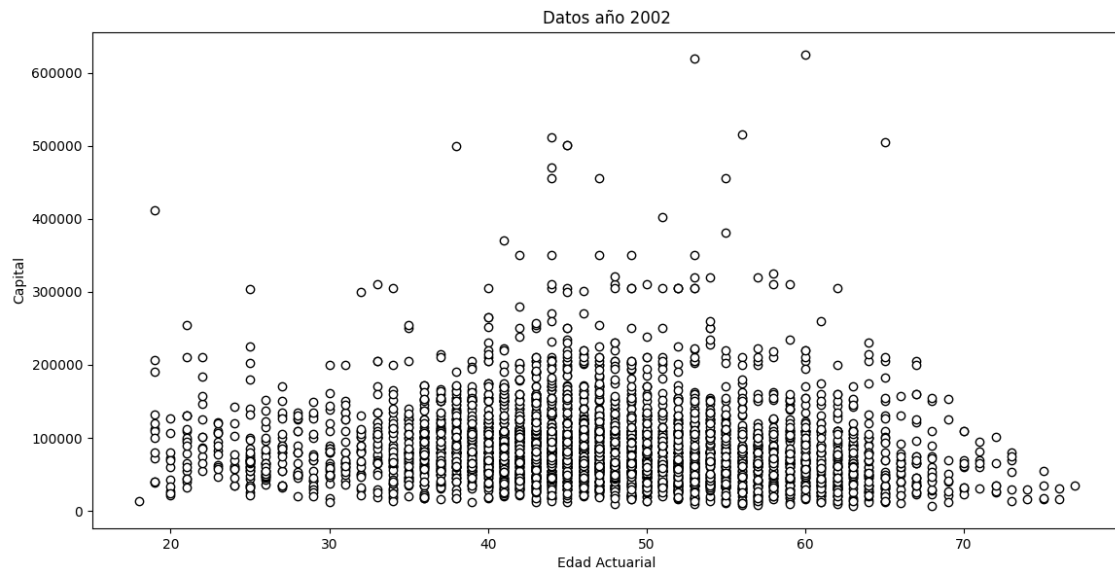


Gráfico 12: Gráfico de dispersión variables capital y edad actuarial para el año 2002.
Fuente: Elaboración propia.

Estadístico	Año 2002					
	Capital	Age	T	Age Actuarial	x	Age_actuarial_quarter
Media	88.884,268	47,071	0,485	47,061	0,548	46,961
Mínimo	7000	18,409	0	18,00	0	18,00
25%	52.000	41,131	0,241	41,00	0,279	41,00
50%	80.000	46,470	0,476	46,00	0,548	46,25
75%	110.000	53,848	0,729	54,00	0,844	53,75
Máximo	625.000	77,068	0,999	77,00	0,997	77,00
Desviación Estándar	56.916,156	10,024	0,286	10,024	0,306	10,02

Tabla 40: Estadísticos básicos de los datos de los contratos de seguro cuya fecha efectiva es el año 2002.
Fuente: Elaboración propia.

Segmentación de riesgos y modelización actuarial en IFRS17

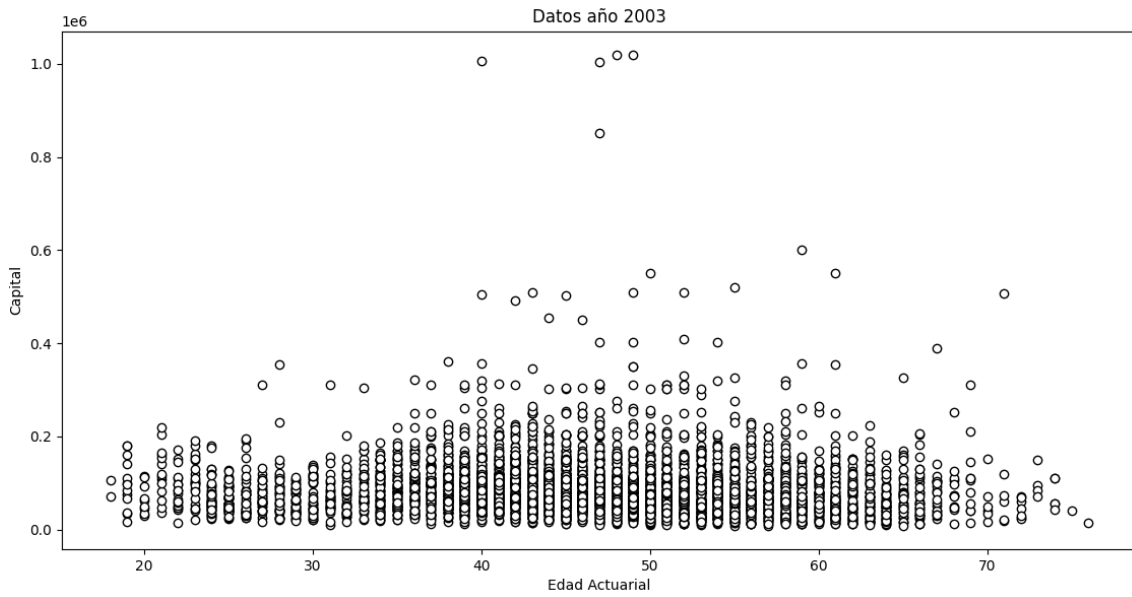


Gráfico 13: Gráfico de dispersión variables capital y edad actuarial para el año 2003.
Fuente: Elaboración propia.

Año 2003						
Estadístico	Capital	Age	T	Age Actuarial	x	Age_actuarial_quarter
Media	88.425,23	46,410	0,496	46,409	0,540	46,29
Mínimo	8.000	18,363	0	18,00	0	18,00
25%	5.2000	40,207	0,246	40,00	0,285	40,00
50%	7.8000	46,497	0,494	46,00	0,526	46,25
75%	110.000	53,282	0,744	53,00	0,833	53,25
Máximo	1.020.000	76,268	0,999	76,00	0,997	76,00
Desviación Estándar	6.2308,23	10,004	0,289	10,007	0,303	10,01

Tabla 41: Estadísticos básicos de los datos de los contratos de seguro cuya fecha efectiva es el año 2003.
Fuente: Elaboración propia.

Segmentación de riesgos y modelización actuarial en IFRS17

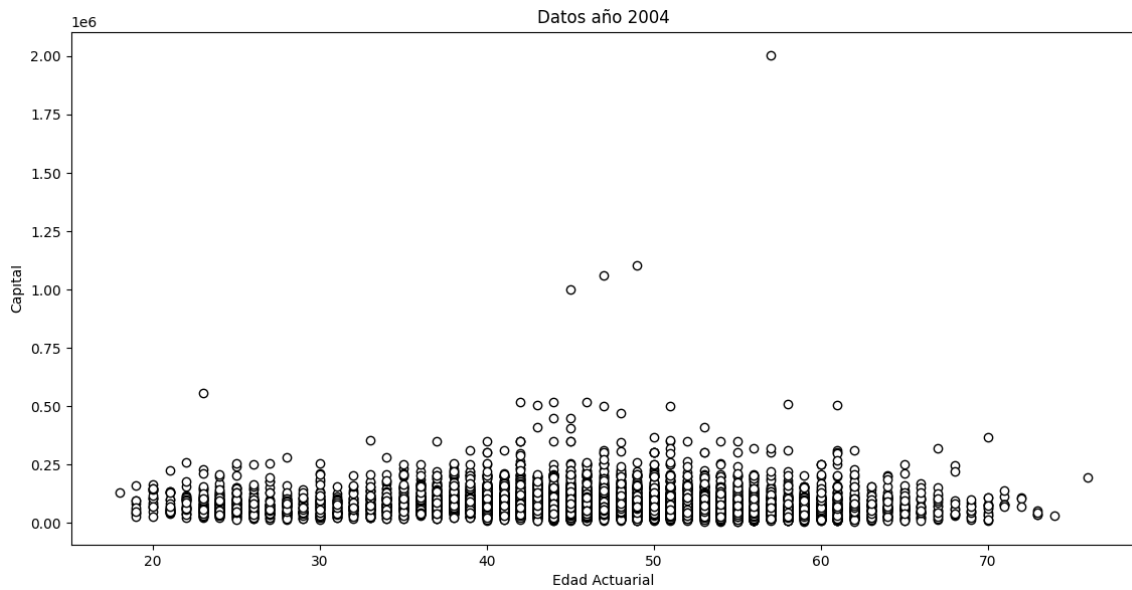


Gráfico 14: Gráfico de dispersión variables capital y edad actuarial para el año 2004.
Fuente: Elaboración propia.

Año 2004						
Estadístico	Capital	Age	T	Age Actuarial	x	Age_actuarial_quarter
Media	8.991105e+04	45.750	0.497	46.409	0.559	45.628
Mínimo	7.000000e+03	18.248	0	18.00	0	18.00
25%	5.200000e+04	40.140	0.250	40.00	0.282	40.00
50%	8.000000e+04	45.878	0.491	46.00	0.570	45.75
75%	1.100000e+05	52.631	0.743	53.00	0.860	52.50
Máximo	2.005000e+06	75.904	0,999	76.00	0.997	75.00
Desviación Estándar	6.466545e+04	10.158	0.287	10.007	0.315	10.164

Tabla 42: Estadísticos básicos de los datos de los contratos de seguro cuya fecha efectiva es el año 2004.
Fuente: Elaboración propia.

Segmentación de riesgos y modelización actuarial en IFRS17

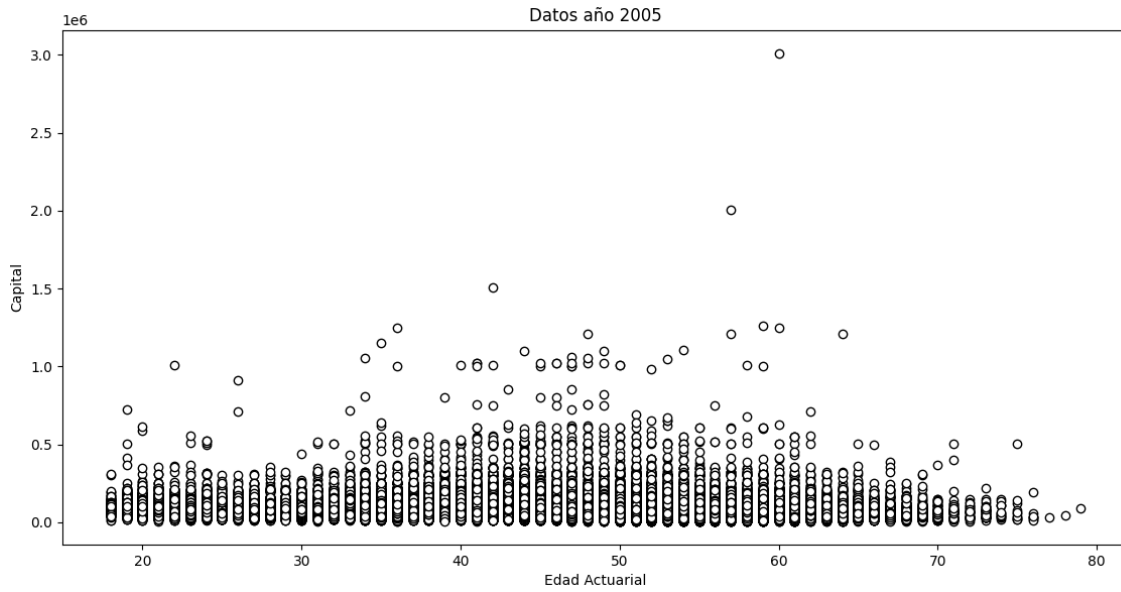


Gráfico 15: Gráfico de dispersión variables capital y edad actuarial para el año 2005.
Fuente: Elaboración propia.

Año 2005						
Estadístico	Capital	Age	T	Age Actuarial	x	Age_actuarial_quarter
Media	91.452,18	44,482	0,498	44,478	0,509	44,358
Mínimo	7.000,00	17,892	0,000	18,000	0,000	17,000
25%	54.000,00	38,590	0,248	39,000	0,241	38,500
50%	80.000,00	45,144	0,494	45,000	0,493	45,000
75%	110.000,0	51,959	0,749	52,000	0,797	51,750
Máximo	3.010.000	78,845	0,999	79,000	0,997	78,750
Desviación Estándar	64.302,20	10,836	0,289	10,841	0,305	10,840

Tabla 43: Estadísticos básicos de los datos de los contratos de seguro cuya fecha efectiva es el año 2005.
Fuente: Elaboración propia.

Segmentación de riesgos y modelización actuarial en IFRS17

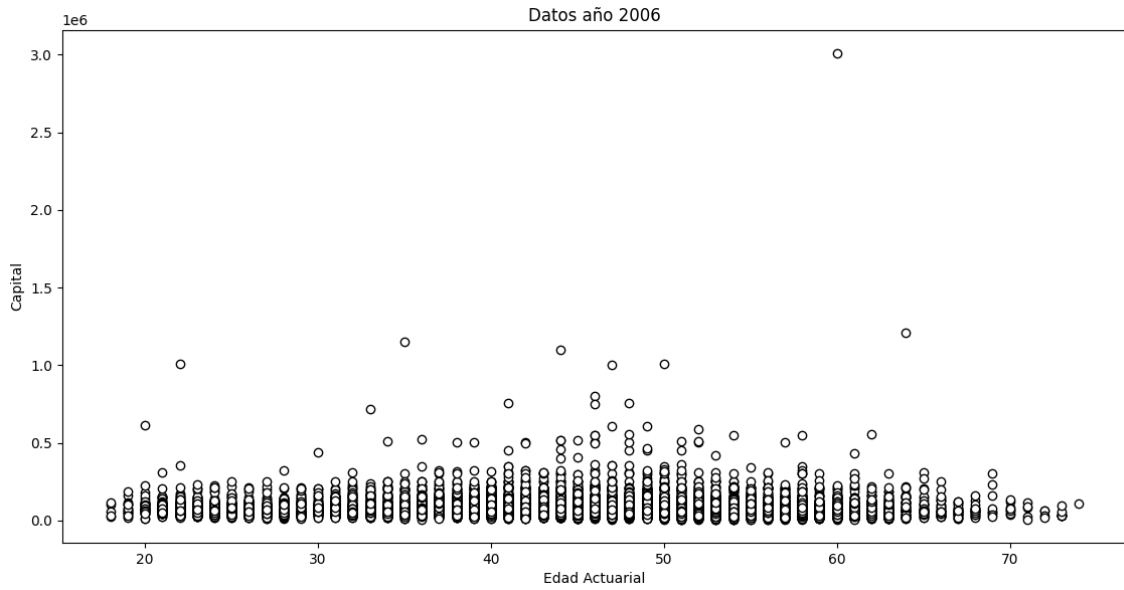


Gráfico 16: Gráfico de dispersión variables capital y edad actuarial para el año 2006.
Fuente: Elaboración propia.

Año 2006						
Estadístico	Capital	Age	T	Age Actuarial	x	Age_actuarial_quarter
Media	90.909,61	44,335	0,502	44,334	0,524	44,206
Mínimo	7.000,00	18,034	0,000	18,000	0,000	18,000
25%	54.000,00	38,433	0,252	39,000	0,249	38,250
50%	80.000,00	45,107	0,502	45,000	0,518	45,000
75%	110.000,0	51,803	0,759	52,000	0,814	51,750
Máximo	3.010.000	73,845	0,999	79,000	0,997	73,750
Desviación Estándar	71.055,58	10,715	0,290	10,723	0,304	10,719

Tabla 44: Estadísticos básicos de los datos de los contratos de seguro cuya fecha efectiva es el año 2006.
Fuente: Elaboración.

Segmentación de riesgos y modelización actuarial en IFRS17

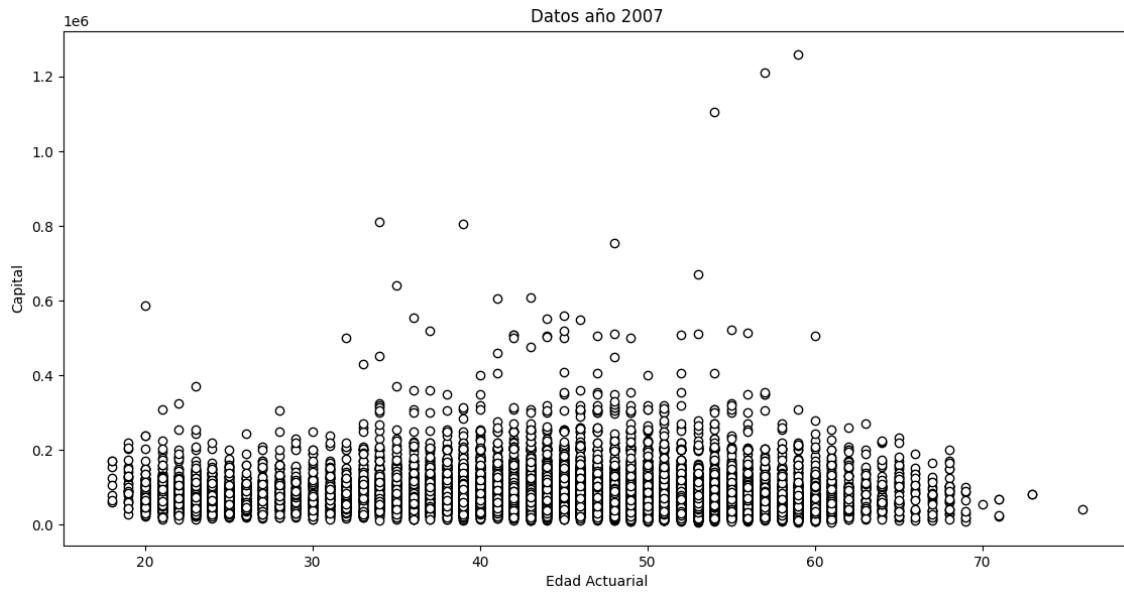


Gráfico 17: Gráfico de dispersión variables capital y edad actuarial para año 2007.
Fuente: Elaboración propia.

Año 2007						
Estadístico	Capital	Age	T	Age Actuarial	x	Age_actuarial_quarter
Media	90.909,61	44,335	0,502	44,334	0,524	44,206
Mínimo	7.000,00	18,034	0,000	18,000	0,000	18,000
25%	52.000,00	38,433	0,251	38,000	0,249	38,250
50%	80.000,00	45,107	0,502	45,000	0,518	45,000
75%	110.000,0	51,803	0,759	52,000	0,814	51,750
Máximo	3.010.000	73,845	0,999	74,000	0,997	73,750
Desviación Estándar	71.055,58	10,715	0,290	10,723	0,304	10,719

Tabla 45: Estadísticos básicos de los datos de los contratos de seguro cuya fecha efectiva es el año 2007.
Fuente: Elaboración propia.

Segmentación de riesgos y modelización actuarial en IFRS17

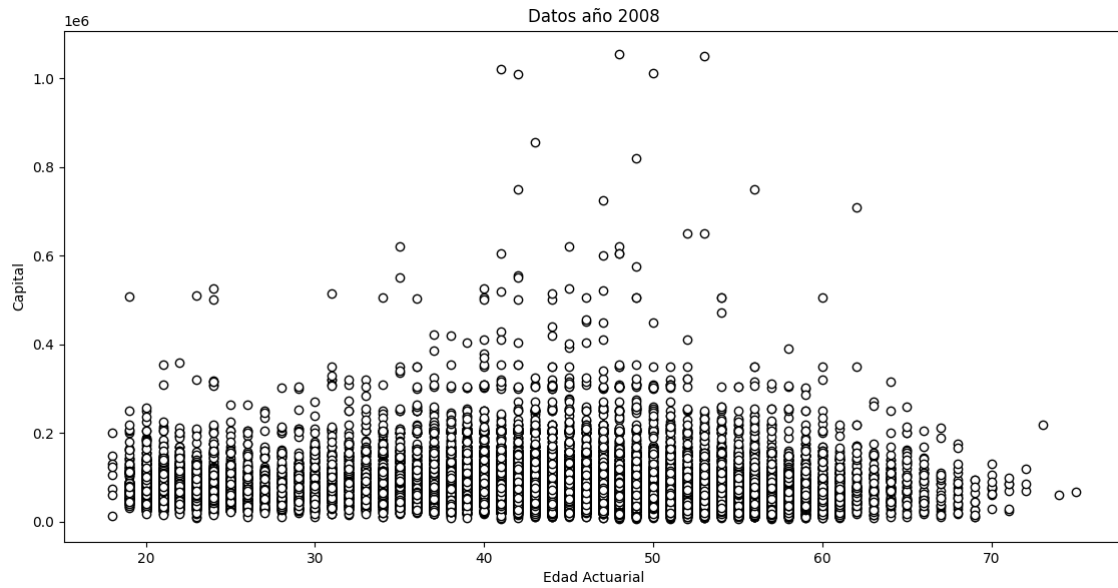


Gráfico 18: Gráfico de dispersión variables capital y edad actuarial para el año 2008.
Fuente: Elaboración propia.

Año 2008						
Estadístico	Capital	Age	T	Age Actuarial	x	Age_actuarial_quarter
Media	9.228944e+04	43.019079	0.496881	43.0159	0.531394	42.897
Mínimo	7.000000e+03	18.017796	0.000000	18.0000	0.0000	18.000
25%	5.500000e+04	36.681040	0.245722	37.0000	0.222603	36.500
50%	8.000000e+04	44.131417	0.493498	44.0000	0.534247	44.000
75%	1.120000e+05	50.893224	0.747433	51.0000	0.841096	50.750
Máximo	1.055000e+06	74.962355	0.999316	75.0000	0.997260	74.250
Desviación Estándar	6.200504e+04	11.120369	0.288612	11.1276	0.320114	11.128

Tabla 46: Estadísticos básicos de los datos de los contratos de seguro cuya fecha efectiva es el año 2008.
Fuente: Elaboración propia.

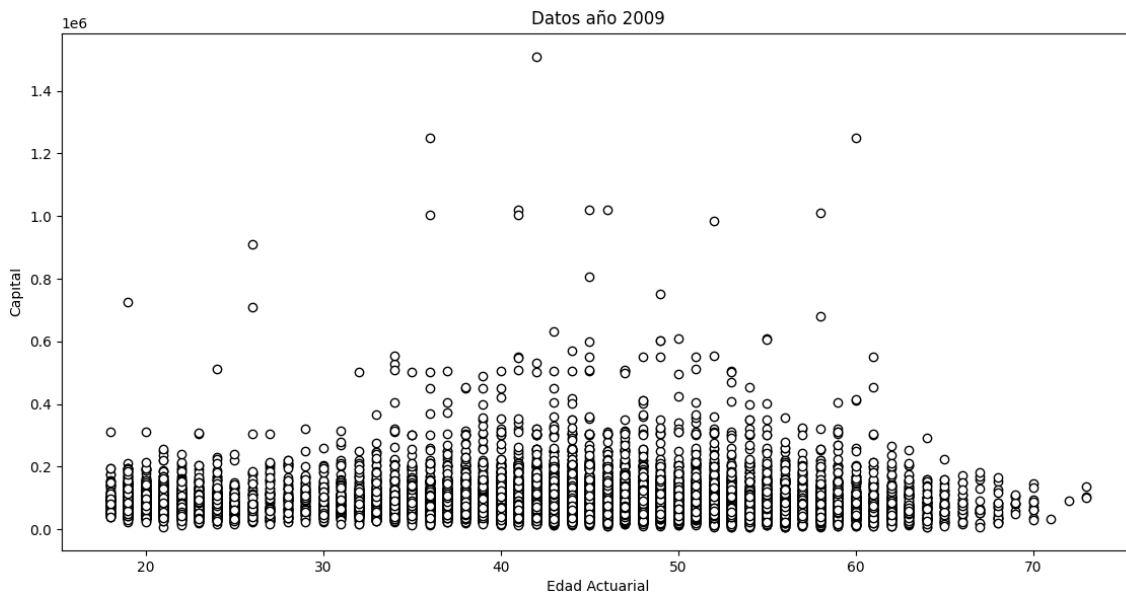


Gráfico 19: Gráfico de dispersión variables capital y edad actuarial para el año 2009.
Fuente: Elaboración propia.

Año 2009						
Estadístico	Capital	Age	T	Age Actuarial	x	Age_actuarial_quarter
Media	9.228944e+04	43.019079	0.496881	43.0159	0.531394	42.897
Mínimo	7.000000e+03	18.017796	0.000000	18.0000	0.0000	18.000
25%	5.500000e+04	36.681040	0.245722	37.0000	0.222603	36.500
50%	8.000000e+04	44.131417	0.493498	44.0000	0.534247	44.000
75%	1.120000e+05	50.893224	0.747433	51.0000	0.841096	50.750
Máximo	1.055000e+06	74.962355	0.999316	75.0000	0.997260	74.250
Desviación Estándar	6.200504e+04	11.120369	0.288612	11.1276	0.320114	11.128

Tabla 47: Estadísticos básicos de los datos de los contratos de seguro cuya fecha efectiva es el año 2009.
Fuente: Elaboración propia.

9.2. Anexos agregación de contratos de seguro

9.2.1. Anexos K-Means

9.2.1.1. Método de la Silueta

El objetivo de la aplicación de técnicas de clúster es obtener grupos que sean el máximo distinto entre ellos y que dentro de cada grupo las observaciones sean el máximo homogéneas posible. Todo ello se consigue si, se elige el número óptimo de clústeres, especialmente en aquellas técnicas de clúster en que el actuario debe elegir el número de clústeres apropiado. Elegir el número de clústeres, no es una tarea sencilla, juntamente con el método del codo, el método de la silueta es otro de los más populares (Saputra D., et al., 2019). Este consiste en calcular la puntuación de cada punto de datos mediante la siguiente formula:

$$\text{Silhouette Score} = \frac{b-a}{\max(a,b)}$$

Donde,

a : es la distancia promedio del punto a otros puntos en el mismo clúster. Cuanto más pequeña mejor.

B : es la distancia promedio del punto a clúster más cercano al que no pertenece. Cuanto más grande sea b mejor.

Por tanto, mide la similitud de un objeto a su propio clúster (cohesión) en comparación con otros clústeres (separación). Así se obtiene el siguiente gráfico:

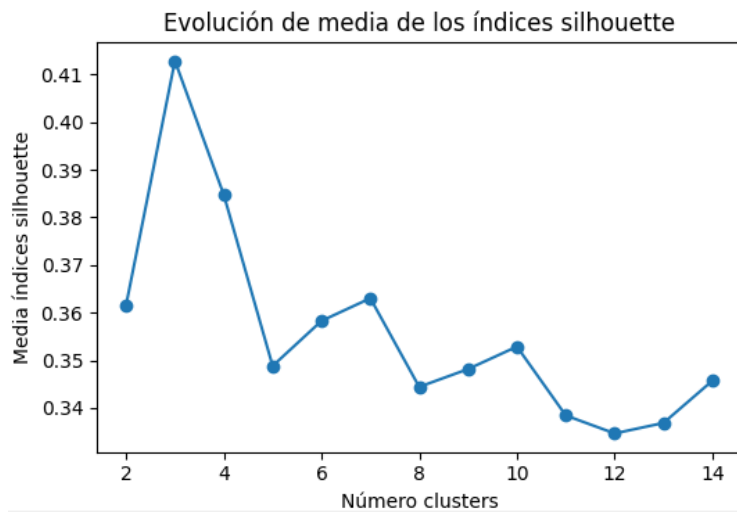


Gráfico 20: Resultados Índice Silhouette.
Fuente: Elaboración propia.

El valor medio de los índices Silhouette se maximiza con 3 clústeres indicando que en este valor máximo los puntos de dentro de cada clúster son más similares entre sí y mejor separados respecto a los de los otros clústeres. Se puede observar que se llega a las mismas conclusiones que aplicando el método del Codo lo que confirma los resultados.

9.2.1.2. Estadísticos de las distancias calculados . Algoritmo K-Means

También pueden calcularse para la distancia entre los datos y los centroides a los que pertenecen lo cual es crucial para comprender cómo se distribuyen espacialmente los datos en relación con los centroides permitiendo evaluar la cohesión y estructura de los grupos y la eficacia del proceso de agrupamiento. Una vez calculadas estas distancias pueden calcularse los estadísticos descriptivos básicos de las mismas (Tabla 48).

	Rango	2,670
	Media	0,759
	Mediana	0,703

Clúster 1	Varianza	0,193
	Desviación estándar	0,439
	Asimetría	1,330
	Curtosis	2,538
	Primer cuartil (Q1)	0,479
	Percentil 25	
	Tercer cuartil (Q3)	0,956
	Percentil 75	
	Rango	6,760
	Media	1,187
	Mediana	0,990
Clúster 2	Varianza	0,965
	Desviación estándar	0,982
	Asimetría	3,545
	Curtosis	15,603
	Primer cuartil (Q1)	0,725
	Percentil 25	
	Tercer cuartil (Q3)	1,233
	Percentil 75	
	Rango	2,334
	Media	0,698
	Mediana	0,696
Clúster 3	Varianza	0,106
	Desviación estándar	0,326
	Asimetría	0,663
	Curtosis	1,481
	Primer cuartil (Q1)	0,467
	Percentil 25	
	Tercer cuartil (Q3)	0,880
	Percentil 75	

Tabla 48: Estadísticos de las distancias de cada dato al centroide más cercano.

Fuente: Elaboración propia.

9.2.1.3. Resultados del clúster K-Means sin datos estandarizados

Estandarizar los datos antes de aplicar el clúster k Means cuando se tienen variables con escalas distintas no es algo imprescindible pero sí recomendable. Aún así la no estandarización permite que la interpretación de los resultados sea más fácil. Mantener la diversidad de escalas también puede permitir captar mejor la variabilidad de los datos. La estandarización siempre puede hacer perder información. Si se ejecuta el algoritmo K Means sin estandarizar los datos se obtienen el Gráfico 21.

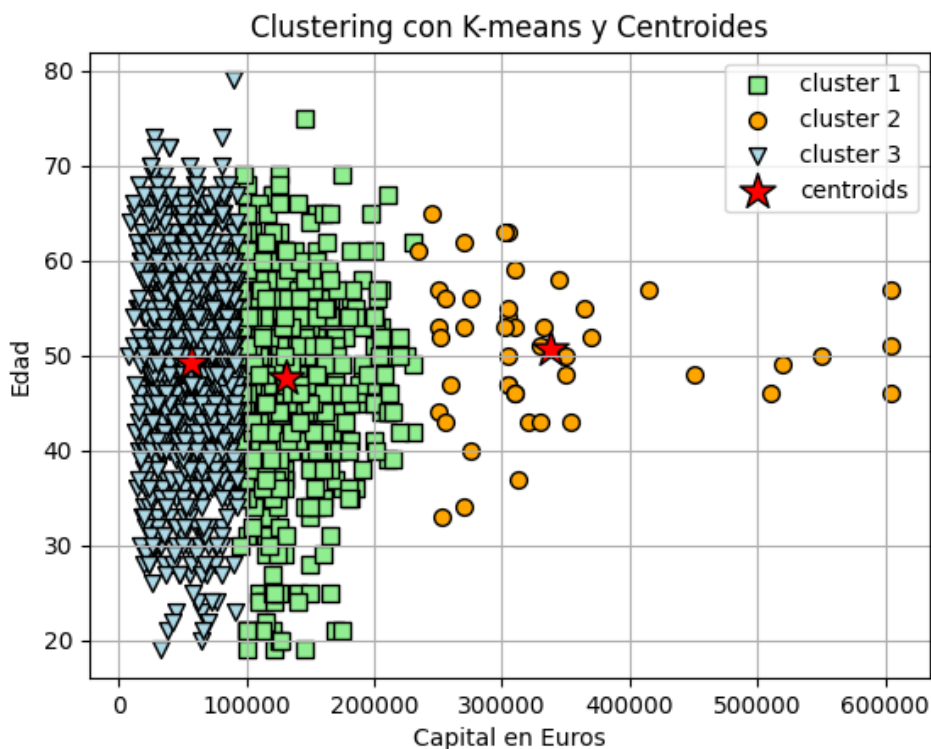


Gráfico 21: Gráfico K-Means sin datos estandarizados.
Fuente: Elaboración propia.

Después de ejecutar el algoritmo K Means se obtiene que para el primer clúster pertenecen 668 contratos de seguro, para el segundo clúster 46 contratos de seguro y para el tercer clúster aparecen 1.246 contratos de seguro. En relación con la aplicación de K Means habiendo estandarizado los datos se concluye que el número de observaciones está distribuido de una forma más equilibrada entre los clústeres.

De la misma forma pueden obtenerse las coordenadas de los centroides de cada Clúster (Tabla 49) y los estadísticos descriptivos básicos de los datos pertenecientes a cada clúster (Tabla 50).

Clúster	Capital	Age Actuarial
1	130641,24	47,7
2	338173,91	50,63
3	57040,25	49,22

Tabla 49: Coordenadas centroide aplicando K-Means sin datos estandarizados.
Fuente: Elaboración propia.

Clúster 1	Edad	Media	47,7	Curtosis	0,825
		Mediana	48,0	Primer cuartil (Q1) Percentil 25	43,0
		Varianza	84,43	Primer cuartil (Q3) Percentil 75	54,0
		Desv. estándar	9,19	Valor máximo	19
		Asimetría	-0,491	Valor mínimo	19
	Capital	Media	130641,24	Curtosis	0,361
		Mediana	120767,5	Primer cuartil (Q1) Percentil 25	105.000,0
		Varianza	976237959,22	Tercer cuartil (Q3) Percentil 75	105.000,0
		Desv. Estándar	31244,81	Valor máximo	230.000,0
		Asimetría	1,058	Valor mínimo	94.000,0
Clúster 2	Edad	Media	50,63	Curtosis	-0,131
		Mediana	51,0	Primer cuartil (Q1) Percentil 25	46,0
		Varianza	53,06	Tercer cuartil (Q3) Percentil 75	55,75
		Desv. Estándar	6,621	Valor máximo	65
		Asimetría	-0,269	Valor mínimo	33
	Capital	Media	338.173,91	Curtosis	8,653
		Mediana	307.500,0	Primer cuartil (Q1) Percentil 25	270.000,0
		Varianza	9.885.187.145,56	Tercer cuartil (Q3) Percentil 75	350.000,0
		Desv. Estándar	99.424,28	Valor máximo	605000,0
		Asimetría	1,565	Valor mínimo	235.000,0
	Media	49,215	Curtosis	-0,162	

Clúster 3	Edad	Mediana	49,0	Primer cuartil (Q1) Percentil 25	43,0
		Varianza	94,297	Tercer cuartil (Q3) Percentil 75	56,0
		Desv. Estándar	9,711	Valor máximo	79
		Asimetría	-0,169	Valor mínimo	19
	Capital	Media	57.040,26	Curtosis	-0,915
		Mediana	60.000,0	Primer cuartil (Q1) Percentil 25	40.000,0
		Varianza	419.725.887,06	Tercer cuartil (Q3) Percentil 75	73.000,0
		Desv. Estándar	20.487,212	Valor máximo	93.000,0
		Asimetría	-0,230	Valor mínimo	7.000,0

Tabla 50: Principales estadísticos de los clústeres hechos con K-Means.
Fuente: Elaboración propia.

9.2.2. Anexos K-Medoids

Igual que se ha realizado con K Means, se puede calcular la distancia entre los datos y los centroides y luego una vez obtenidas estas distancias, pueden calcularse los estadísticos descriptivos básicos (Taula 51).

Clúster 1	Rango	2,859
	Media	0,706
	Mediana	0,613
	Varianza	0,237
	Desviación estándar	0,486
	Asimetría	1,756
	Curtosis	3,455
	Primer cuartil (Q1)	0,420
	Percentil 25	
	Tercer cuartil (Q3)	0,806
	Percentil 75	

Clúster 2	Rango	2,258
	Media	0,632
	Mediana	0,631
	Varianza	0,101
	Desviación estándar	0,318
	Asimetría	1,072
	Curtosis	3,019
	Primer cuartil (Q1)	0,422
	Percentil 25	
	Tercer cuartil (Q3)	0,791
Clúster 3	Percentil 75	
	Rango	7,885
	Media	0,998
	Mediana	0,679
	Varianza	1,013
	Desviación estándar	1,007
	Asimetría	3,692
	Curtosis	17,914
	Primer cuartil (Q1)	0,518
	Percentil 25	
Tercer cuartil (Q3)	1,090	
Percentil 75		

Tabla 51: Estadísticos de las distancias de cada dato al centroide más cercano.
Fuente: Elaboración propia.

9.2.3. Algoritmo jerárquico aglomerativo

Se muestra a continuación la Tabla 52 en la que pueden verse distintos métodos de enlace que ofrecen diferentes criterios para fusionar grupos en el algoritmo jerárquico aglomerativo.

Método	Python	Definición	Coefficiente de correlación
Enlace completo	“complete”	Calcula la distancia máxima entre todos los pares de puntos pertenecientes a diferentes grupos.	0,011
Enlace promedio o UPGMA⁴¹	“average”	Calcula la distancia promedio entre todos los pares de puntos pertenecientes a diferentes grupos.	0,005
Enlace simple	“single”	Calcula la distancia mínima entre todos los pares de puntos pertenecientes a diferentes grupos.	-0,001
Enlace de Ward	“ward”	Consiste en minimizar la varianza intracluster al fusionar los grupos.	-0,104

Tabla 52: Tipos de enlace en el Algoritmo Jerárquico Aglomerativo.
Fuente: Elaboración propia.

9.2.4. Fuzzy clustering C-means

Este método fue propuesto por Dunn (1973) para mejorar el algoritmo K Means. Una de las principales características de este algoritmo es que los datos pueden pertenecer a múltiples clústeres. En concreto se basa en probabilidades de pertenencia de cada dato a un clúster (Raschka, 2019). Además, no existe un único centro para cada clúster, sino que utiliza varios centros difusos, que se calculan mediante la media ponderada de los datos. Otra característica es que se basa en la minimización de la suma de las distancias al cuadrado ponderadas entre los puntos de los datos y los centros de los clústeres.

En el caso aquí presente se ha elegido $K=3$ siguiendo los resultados obtenidos en el Método del Codo y el Método de la Silueta. Mediante una función de asignación se asignarán los diferentes datos a los subconjuntos difusos.

Después de aplicar el algoritmo con $K=3$ clústeres en el portafolio que contiene los contratos de seguro con fecha efectiva del año 2000 se han obtenido como calidad del ajuste 1059,113 con 86 iteraciones.

⁴¹ UPGMA: “Unweighted Pair Group Method with Arithmetic Mean”.

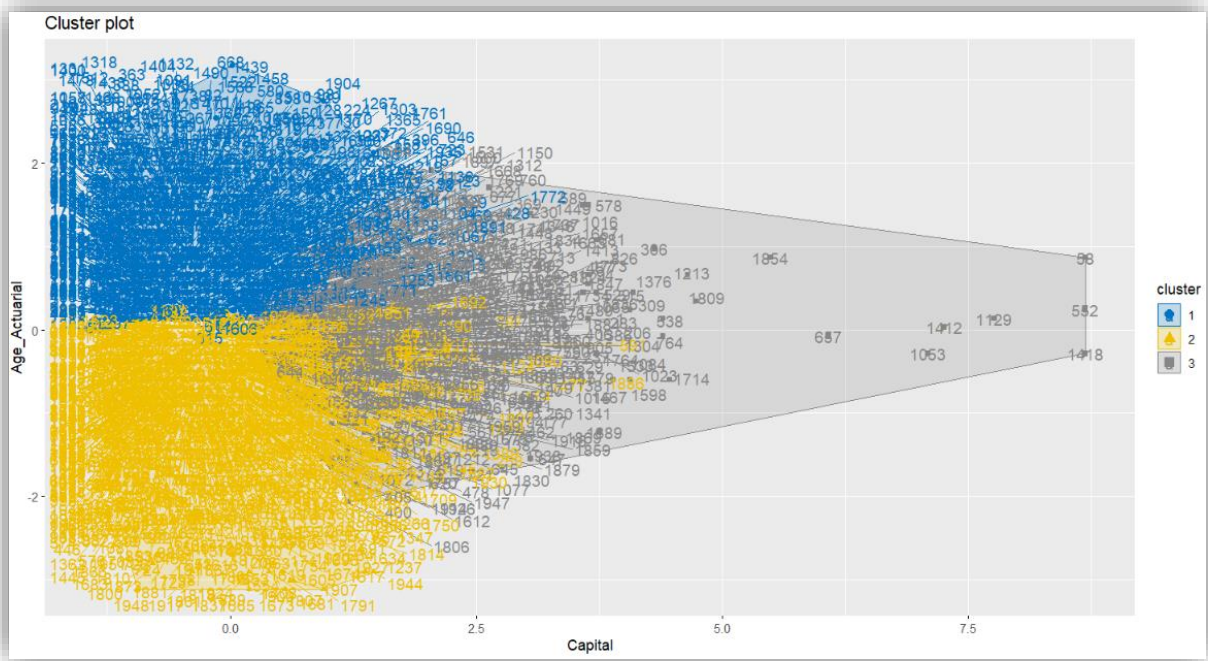


Gráfico 22: Gráfico de dispersión Fuzzy C-Means.
Fuente: Elaboración propia.

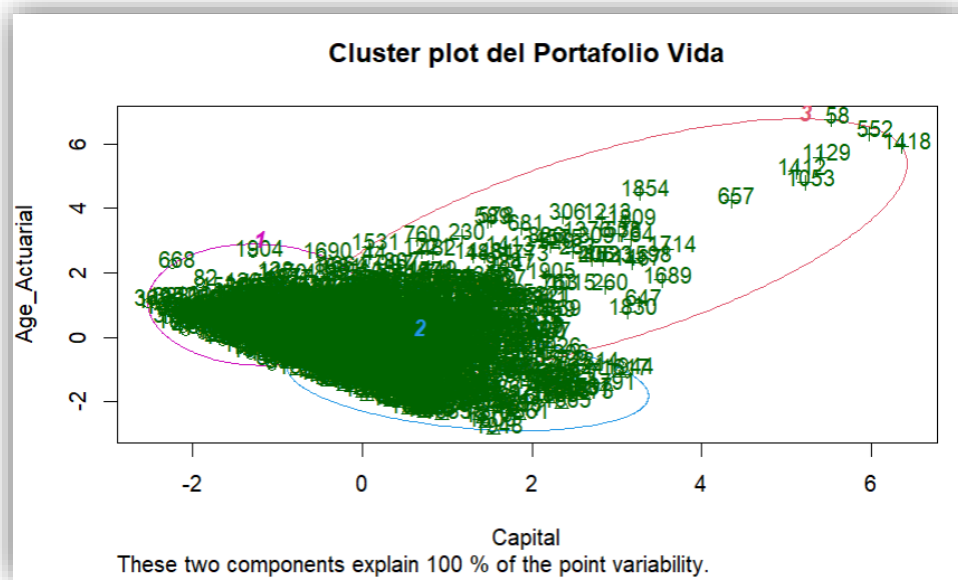


Gráfico 23: Gráfico C-Means con elipses de confianza al 95%.
Fuente: Elaboración propia.

Es importante destacar que tanto en el Gráfico 22 como el Gráfico 23 se observan outliers y que en un análisis de Clúster deberían excluir y volver a aplicar el algoritmo.

Seguidamente se muestran como sería la matriz que contine los porcentajes de pertenencia de cada observación al clúster. Se muestran en la Tabla 53 estos porcentajes para las 5 primeras observaciones y las últimas 5 observaciones:

Observación	Clúster 1	Clúster 2	Clúster 3
1	0,213	0,363	0,423
2	0,077	0,845	0,078
3	0,012	0,974	0,013
4	0,513	0,155	0,332
5	0,718	0,124	0,158
...
1955	0,062	0,150	0,787
1956	0,097	0,611	0,292
1957	0,047	0,882	0,070
1958	0,147	0,749	0,103
1959	0,061	0,160	0,779

Tabla 53: Grados de pertenencia de cada observación al clúster en el algoritmo Fuzzy C-Means.
Fuente: Elaboración propia.

Asimismo, puede obtenerse los estadísticos descriptivos calculados sobre los grados de pertenencia que indican la probabilidad de que un determinado punto pertenezca a un clúster:

	Número observaciones	Min	Q1	Media	Mediana	Q3	Max
Clúster 1	728	0,367	0,658	0,764	0,788	0,894	0,999
Clúster 2	817	0,389	0,593	0,731	0,729	0,880	0,999
Clúster 3	414	0,352	0,533	0,671	0,666	0,801	0,999

Tabla 54: Estadísticos descriptivos calculados sobre los grados de pertenencia en algoritmo Fuzzy C-Means.
Fuente: Elaboración propia.

9.2.5. Density-Based Spatial Clustering of Applications with Noise (DBSCAN)

Los métodos de agrupamiento basados en la densidad consisten en que cada clúster es una área donde hay una alta concentración de puntos y estas áreas están separadas por regiones donde hay una concentración baja de datos. Este método se caracteriza por tener la capacidad de manejar datos ruidosos los cuales no se asignan a ningún grupo en particular, sino que se consideran valores atípicos o ruido. Uno de los algoritmos de este tipo que se ha convertido en una herramienta importante en la minería de datos y la analítica siendo hoy uno de los más usados para la comunidad científica es el *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN). Una de las principales ventajas que tiene DBSCAN es que es capaz de encontrar cualquier tipo de clúster, incluso aquellos que no tienen forma circular (Raschka, 2019). La principal desventaja que tiene DBSCAN es que es muy sensible a los parámetros Epsilon y Puntos Mínimos (MinPts) los cuales se definirán en que consisten a continuación.

En el algoritmo DBSCAN deben destacarse:

1) Los parámetros de entrada que deben proporcionarse para aplicar este algoritmo son ε (épsilon) y minPts (observaciones mínimas).

- Epsilon (Eps): define un radio (radio ε) alrededor de cada punto de datos. En concreto, marca una distancia máxima en la que el algoritmo buscará otros puntos cercanos alrededor de un punto dado, los cuales se considerarán vecinos de este. Matemáticamente, se define como:

$$N_{\varepsilon}(p) = \{q \in D \mid d(p,q) < \varepsilon\}$$

Donde “D” es una medida de distancia y ε es un número real positivo.

- minPts: es el número mínimo de puntos que deben estar dentro del radio épsilon de un punto en particular para formar una región densa.

2) El algoritmo clasifica los puntos en:

- Punto núcleo (Core Point): si el número de puntos del radio épsilon es mayor o igual a minPts. El punto núcleo pertenece siempre a una región densa.
- Punto borde o Border Point: es aquél que tiene menos puntos mínimos (minPts) dentro de épsilon (Eps) pero se encuentra en la vecindad de un punto núcleo, es decir, la distancia entre ambos puntos es menor o igual al radio épsilon.
- Un punto de ruido, es cualquier punto que no es ni punto núcleo ni punto borde.

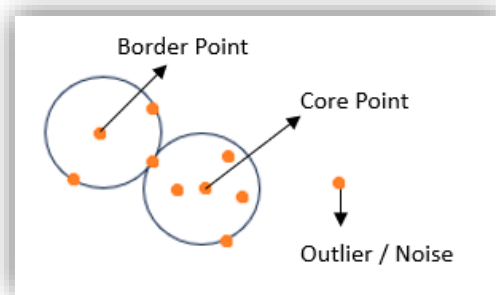


Ilustración 2: Border Point, Core Point y Outlier o Noise.
Fuente: Elaboración propia.

También es necesario definir los siguientes conceptos con los que se basa el algoritmo (Charu y Chandan, 2014):

- Borde por densidad (Directly density-reachable):
Un punto p es directamente alcanzable por densidad por un punto q respecto Epsilon y MintPts si:
 - p se encuentra dentro del radio Epsilon de q .

$p \in \text{Neps}(q)$

- Si q cumple la Core condition, es decir, que $|\text{Neps}(q)| \geq \text{MinPts}$

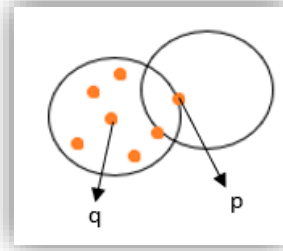


Ilustración 3: Directly density-reachable.
Fuente: Elaboración propia.

- Alcanzable por densidad (Density reachable)

Un punto p es alcanzable por densidad por el punto q respecto Epsilon y MinPts si existe una cadena de puntos $p_1, p_2, \dots, p_n, p_1 = q, p_n = p$ entonces p_{i+1} es directamente alcanzable por densidad por p_i .

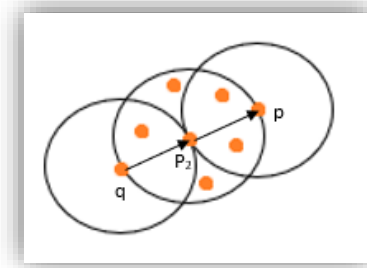


Ilustración 4: Density reachable.
Fuente: Elaboración propia.

- Densidad de puntos conectados (Density-connected): dos puntos p y q son puntos de conexión de densidad respecto Eps y MinPts si existe otro punto, por ejemplo O , a partir del cual p y q son alcanzables por densidad. El algoritmo realiza los siguientes pasos (Han, 2016):
 - Todos los puntos son “no visitados”.
 - Se selecciona de forma arbitraria un punto p que se marca como “visitado”.
 - Respecto el punto p teniendo en cuenta el valor Eps y MinPts:
 - o Si p contiene al menos MinPts en la vecindad de epsilon. Si no los contiene, se marca como un punto ruido.
 - Finalmente, el proceso se reinicia con un nuevo punto.

Tal y como se acaba de comentar es necesario entrar en el algoritmo el nivel de epsilon. Para ello se cogen los datos del portafolio que contiene los contratos de seguro cuya fecha efectiva es el año 2000 y se realiza el gráfico de los tres vecinos más cercanos en relación a la distribución de distancias entre los puntos de los datos:

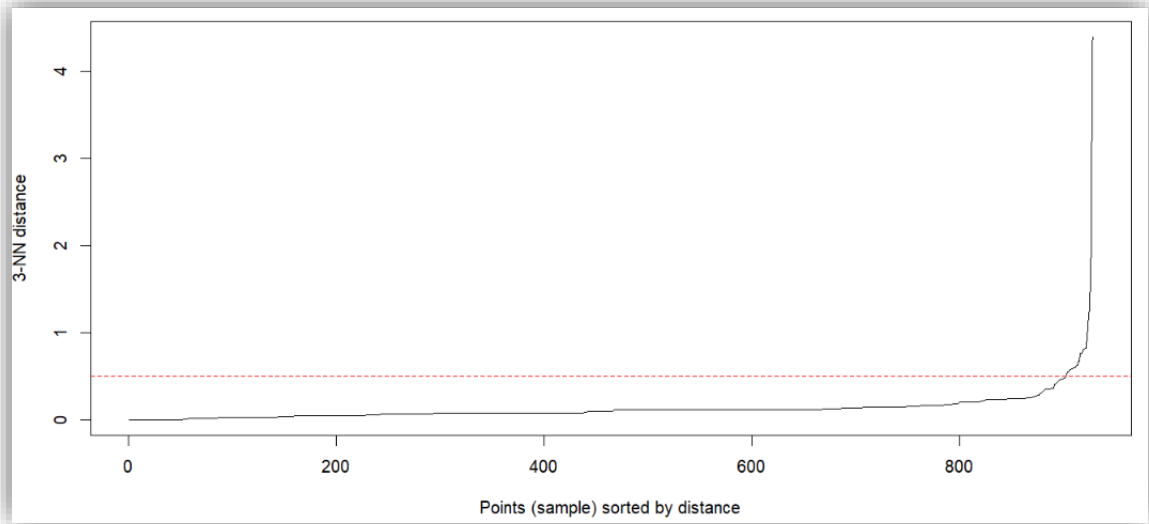


Gráfico 24: 3NN distance DBSCAN.
Fuente: Elaboración propia.

Y se obtiene el Gráfico 25:

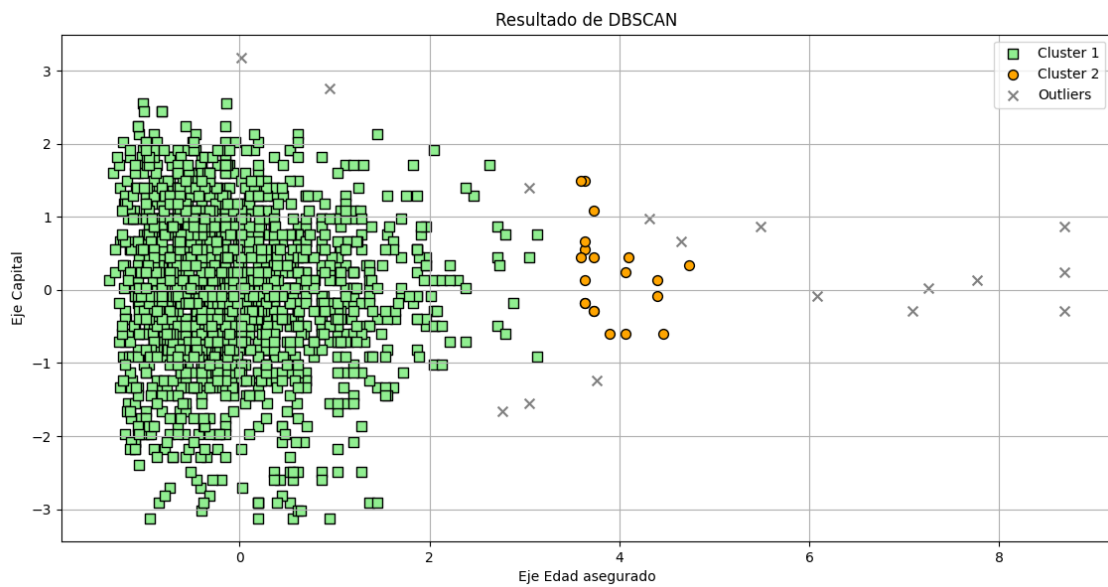


Gráfico 25: Resultados DBSCAN.
Fuente: Elaboración propia.

El algoritmo DBSCAN indica como resultado que existen 2 clústeres, con 16 puntos de ruido que no pertenecen a ningún clúster. En el primer clúster estará formado por 1924 datos y el segundo por 20 datos.

Número de clústeres	2
Nº de observaciones de cada clúster	Clúster 1: 1.924 observaciones Clúster 2: 20 observaciones
Distancia máxima entre dos puntos dentro de un clúster	Clúster 1: 6,013 Clúster 2: 2,277
Distancia promedio entre todos los puntos dentro de un clúster	Clúster 1: 1,548 Clúster 2: 0,891
Distancia mediana entre todos los puntos dentro de un clúster	Clúster 1: 1,420 Clúster 2: 0,824
Distancia promedio entre los puntos de diferentes clústeres	4,161
Distancia promedio dentro de los clústeres	1,219

Tabla 55: Resultados DBSCAN (distancias).
Elaboración propia.

9.3. Anexos estimación de flujos de efectivo

9.3.1. Estimación de las provisiones genéricas para el 2022 bajo IFRS17 e IFRS4 (asegurado no jubilado en el 2022)

También sería interesante analizar la provisión necesaria de la compañía aseguradora para un asegurado que en los ejercicios 2022 y 2023 no se haya jubilado aun ya que existirán flujos de primas futuros que deberá desembolsar. Por ejemplo, sería el caso del octavo asegurado del portafolio inicial correspondiente al tercer asegurado del primer clúster. Los datos originales para este asegurado son:

Nombre original de la variable	Descripción
ID	8
Gender (Género)	M
Birth_Date (Fecha de nacimiento)	28/03/2009
Effective_Date (Fecha efectiva)	11/03/2000
Capital	16000€
Renewal_Date (Fecha de renovación)	11/03/2009
Age (Años)	39,9534565
T	0,95345654
Age Actuarial	40
Date_Birthday	28/03/1969
X	0,1890411
R	1
S	1
Age actuarial quarter	39
Month	3

Tabla 56: Datos asegurado que en 2022 y 2023 aún tiene que jubilarse.

Fuente: Elaboración propia.

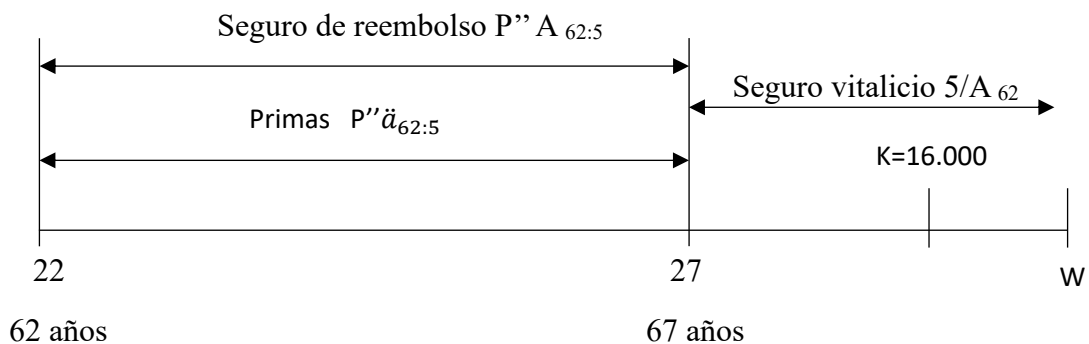
Si se replican los cálculos hechos para obtener la prima conforme los datos de este asegurado se obtiene:

Prima tarifa	232,45
Prima inventario	225,44
Prima pura	224,14

Tabla 57: Prima pura, prima de inventario y prima de tarifa del asegurado que en 2022 y 2023 aún tiene que jubilarse.

Fuente: Elaboración propia.

El esquema temporal que se debe seguir es el siguiente para el año 2022:



Observando el nuevo esquema temporal si se quieren calcular las provisiones a efectos contables que necesita la empresa aseguradora en 2022 se deberán realizar los siguientes cálculos:

Para $k=22$ años,

$$V'(40;22) = VAA_{22}(\text{prestacions futures}) + VAA_{22}(\text{ggi futuros}) - VAA_{22}(\text{primes de inventario futures}) = P'' A_{62:5} + P'' 22 A_{62:5} + 16000 \cdot 5/A_{62} + 0,0001 P'' \ddot{a}_{40:27} \ddot{a}_{62} - P' \ddot{a}_{62:5}$$

*Expresión 38: Cálculo provisiones finales a efectos contables ejercicio 2022 (asegurado no jubilado en esta fecha).
Fuente: Elaboración propia.*

En primer lugar, el termino $P'' A_{62:5}$ hace referencia a la cláusula del contrato de seguro que trata sobre las primes de reembolso que cobraría el beneficiario en caso de que el asegurado falleciera entre 62 años y 67 años.

$$P'' A_{62:5} = P'' \sum_{t=0}^{(5+0)*1-1} (1 + 0,02)^{-(t+1)} {}_tq_{62}$$

*Expresión 39: primas de reembolso 2022 (asegurado no jubilado en esta fecha).
Fuente: Elaboración propia.*

La expresión $A_{62:5}$ representa el valor presente actuarial asociado con la mortalidad en el contexto del seguro. El subíndice indica el periodo que transcurre desde el 2022 cuando el asegurado tiene 62 años hasta su edad de jubilación (temporalidad de 5 años). En términos prácticos, este valor es el resultado del sumatorio de las primas pagadas por el asegurado desde los 62 años hasta la edad de jubilación, descontando las primas al tipo de interés determinista y vencido, simbolizado con $t+1$ y vinculándolas con las probabilidades de fallecimiento del asegurado, en particular, se trata de la probabilidad de que el asegurado haya sobrevivido hasta el año correspondiente y fallezca en ese último período teniendo en cuenta las probabilidades de supervivencia y mortalidad en cada año intermedio.

A continuación, se detalla a la renta correspondiente a las primas consolidadas, es decir, aquellas que ha pagado el asegurado desde la fecha efectiva del contrato de seguro (año 2000) hasta el año 2022 y que en el caso de que falleciera desde la actualidad (año 2022) hasta la edad de jubilación tendría derecho a cobrarlas el beneficiario del contrato de seguro. La expresión es la misma que la anterior pero debe ser multiplicada por 22 ya que son las primas que habrá pagado el asegurado durante estos dos 22 años que hay entre la fecha efectiva del contrato de seguro y el año 2022.

$$22 P'' A_{62:5} = P'' 22 \sum_{t=0}^{(5+0)*1-1} (1 + 0,02)^{-(t+1)} {}_tq_{62}$$

*Expresión 40: Primas consolidadas 2022 (asegurado no jubilado en esta fecha).
Fuente: Elaboración propia.*

La prestación del seguro de fallecimiento comienza a partir de la edad de jubilación, con un diferimiento de cinco años, indicado por el subíndice “5” antes de A_{62} . Se escribe la A mayúscula debido a que dicha renta está vinculada al fallecimiento del asegurado. El valor actual actuarial se calcula como la suma desde la edad de jubilación hasta el infinito actuarial final de la tabla biométrica PASEM 2020 No Rel, que es de 109 años para los hombres como es el caso de este asegurado. Se resta la edad del asegurado en la fecha del contrato, menos 1 año, para considerar el último año de las tablas biométricas, luego se presume que no vive nadie. Se descuenta el capital especificado en el contrato, en este caso, 16.000 euros al tipo de interés determinista, y se ajusta al año vencido, ya que la prestación la cobra el beneficiario al final del año en que ha fallecido el asegurado.

$$16000 \text{ } 5/A_{62} = \sum_{t=5}^{(109-62)*1-1} (1 + 0,02)^{-(t+1)} \text{ } {}_tq_{62}$$

*Expresión 41: Prestación seguro de fallecimiento 2022 (asegurado no jubilado en esta fecha).
Fuente: Elaboración propia.*

El termino $0,0001 P'' \ddot{a}_{40:27} \ddot{a}_{62}$ hace referencia, como en los casos anteriores estudiados al reparto de gastos de gestión interna. En primer lugar, el término \ddot{a}_{62} hace referencia a la renta necesaria para el reparto de gastos de gestión interna desde el año 2022 hasta el infinito actuarial de las tablas descontando la edad del asegurado en la fecha efectiva del contrato, menos uno, del capital estipulado para este asegurado. Este cálculo también toma en consideración el tipo de interés establecido y las probabilidades de supervivencia.

$$\ddot{a}_{62} = \sum_{t=0}^{(125-62)*1-1} (1 + 0,02)^{-t} \text{ } {}_tP_{62}$$

*Expresión 42: Reparto gastos de gestión interna 2022 (asegurado no jubilado en esta fecha).
Fuente: Elaboración propia.*

También en el reparto de gastos aparece el término $\ddot{a}_{40:27}$ que sirve para el cálculo de gastos de gestión interna basados en 0,01% sobre la prima de inventario P' . El detalle de este término sería el siguiente:

$$\ddot{a}_{40:27} = P'' \sum_{t=0}^{(27+0)*1-1} (1 + 0,02)^{-t} \text{ } {}_tP_{40}$$

*Expresión 43: Renta para el cálculo de gastos de gestión interna 2022 (asegurado no jubilado en esta fecha).
Fuente: Elaboración propia.*

Las primas al ser ligadas a la supervivencia del asegurado y prepagables se indican con la letra “a” en minúscula y los dos puntos encima. El subíndice indica el inicio del contrato a los 40 años, seguido de la duración temporal de 27 años que es todo el periodo que pagará primas. La expresión $a_{49:18}$ representa el valor presente actuarial de los cobros de

primas, descontados al tipo de interés determinista, considerando la supervivencia del asegurado año tras año hasta su jubilación.

La renta actuarial de los cobros de primas en este caso, describe los pagos que le quedan para realizar el asegurado desde el año 2022 hasta que alcance su jubilación a los 67 años. En este contexto, esta renta actuarial representa un período temporal de 5 años, calculado restando la edad de jubilación (67 años) menos la edad efectiva del contrato de seguro (62 años), lo que nos da una temporalidad de 5 años.

$$\ddot{a}_{62:5} = P'' \sum_{t=0}^{(5+0)*1-1} (1 + 0,02)^{-t} {}_tP_{62}$$

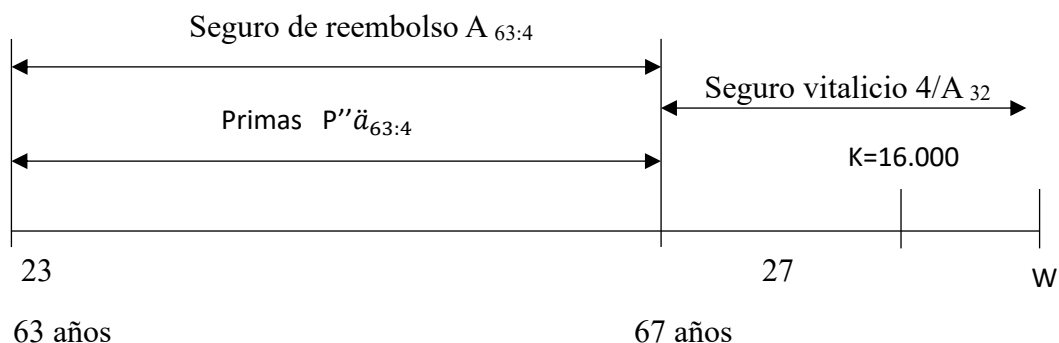
Expresión 44: Pago de primas por parte del asegurado a la aseguradora 2022 (asegurado no jubilado en esta fecha).
Fuente: Elaboración propia.

La simbología empleada en la expresión refleja la naturaleza de los cobros de primas en el contexto del contrato de seguro. El uso de la letra “a” en minúscula y los dos puntos encima denotan que estas primas son prepagables y están asociadas con la supervivencia del asegurado. El subíndice indica el inicio del contrato a los 49 años, seguido de la duración temporal de 18 años. La expresión $a_{49:18}$ representa el valor presente actuarial de los cobros de primas, descontados al tipo de interés determinista, considerando la supervivencia del asegurado año tras año hasta su jubilación.

Si se sustituyen los valores de estas rentas descritas a la Expresión 38 se obtienen que bajo IFRS4 la provisión debe ser 9.943,78 € y bajo IFRS17 la provisión debe ser de 8.489 €.

9.3.2. Estimación de las provisiones genéricas para el 2023 bajo IFRS17 e IFRS4 (asegurado no jubilado en el 2023)

Para el año 2023 debe rehacerse el esquema temporal:



Observando el nuevo esquema temporal si se quieren calcular las provisiones a efectos contables que necesita la empresa aseguradora en 2023 se deberán realizar los siguientes cálculos:

Para $k=23$ años,

$$V'(40;23) = VAA_{23}(\text{prestacions futures}) + VAA_{23}(\text{ggi futuros}) - VAA_{23}(\text{primes de inventario futures}) = P'' A_{63:4} + P'' 22 A_{63:4} + 16000 5/A_{63} + 0,0001 P'' \ddot{a}_{40:27} \ddot{a}_{63} - P' \ddot{a}_{63:4}$$

Expresión 45: Cálculo provisiones finales a efectos contables ejercicio 2023 (asegurado no jubilado en esta fecha).
Fuente: Elaboración propia.

En primer lugar, el termino $P'' A_{63:4}$ hace referencia a la cláusula del contrato de seguro que trata sobre las primes de reembolso que cobraría el beneficiario en caso de que el asegurado falleciera entre 63 años y 67 años.

$$P'' A_{63:4} = P'' \sum_{t=0}^{(4+0)*1-1} (1 + 0,02)^{-(t+1)} {}_tq_{63}$$

Expresión 46: Reembolso de primes 2023 (asegurado no jubilado en esta fecha).
Fuente: Elaboración propia.

La expresión $A_{62:5}$ representa el valor presente actuarial asociado con la mortalidad en el contexto del seguro. El subíndice indica el periodo que transcurre desde el 2023 cuando el asegurado tiene 63 años hasta su edad de jubilación (temporalidad de 4 años). En términos prácticos, este valor es el resultado del sumatorio de las primes pagadas por el asegurado desde los 63 años hasta la edad de jubilación, descontando las primes al tipo de interés determinista y vencido, simbolizado con $t+1$ y vinculándolas con las probabilidades de fallecimiento del asegurado, en particular, se trata de la probabilidad de que el asegurado haya sobrevivido hasta el año correspondiente y fallezca en ese último período teniendo en cuenta las probabilidades de supervivencia y mortalidad en cada año intermedio.

A continuación, se detalla a la renta correspondiente a las primes consolidadas, es decir, aquellas que ha pagado el asegurado desde la fecha efectiva del contrato de seguro (año 2000) hasta el año 2023 y que en el caso de que falleciera desde la actualidad (año 2023) hasta la edad de jubilación tendría derecho a cobrarlas el beneficiario del contrato de seguro. La expresión es la misma que la anterior, pero debe ser multiplicada por 23 ya que son las primes que habrá pagado el asegurado durante estos 23 años que hay entre la fecha efectiva del contrato de seguro y el año 2023.

$$23 P'' A_{63:4} = P'' 23 \sum_{t=0}^{(4+0)*1-1} (1 + 0,02)^{-(t+1)} {}_tq_{63}$$

Expresión 47: Primes consolidadas 2023 (asegurado no jubilado en esta fecha).
Fuente: Elaboración propia.

La prestación del seguro de fallecimiento comienza a partir de la edad de jubilación, con un diferimiento de cinco años, indicado por el subíndice “4” antes de A_{63} . Se escribe la A mayúscula debido a que dicha renta está vinculada al fallecimiento del asegurado. El valor actual actuarial se calcula como la suma desde la edad de jubilación hasta el infinito actuarial final de la tabla biométrica PASEM 2020 No Rel, que es de 109 años para los hombres como es el caso de este asegurado. Se resta la edad del asegurado en la fecha del contrato, menos 1 año, para considerar el último año de las tablas biométricas, luego se presume que no vive nadie. Se descuenta el capital especificado en el contrato, en este caso, 16.000 euros al tipo de interés determinista, y se ajusta al año vencido, ya que la prestación la cobra el beneficiario al final del año en que ha fallecido el asegurado.

$$16000 \cdot 4/A_{63} = \sum_{t=4}^{(109-63)*1-1} (1 + 0,02)^{-(t+1)} \cdot {}_tq_{63}$$

*Expresión 48: Parte del seguro de fallecimiento 2023 (asegurado no jubilado en esta fecha).
Fuente: Elaboración propia.*

El término $0,0001 P'' \ddot{a}_{40:27} \ddot{a}_{63}$ hace referencia, como en los casos anteriores estudiados al reparto de gastos de gestión interna. En primer lugar, el término \ddot{a}_{63} hace referencia a la renta necesaria para el reparto de gastos de gestión interna desde el año 2023 hasta el infinito actuarial de las tablas descontando la edad del asegurado en la fecha efectiva del contrato, menos uno, del capital estipulado para este asegurado. Este cálculo también toma en consideración el tipo de interés establecido y las probabilidades de supervivencia.

$$\ddot{a}_{63} = \sum_{t=0}^{(125-63)*1-1} (1 + 0,02)^{-t} \cdot {}_tP_{63}$$

*Expresión 49: Reparto de gastos 2023 (asegurado no jubilado en esta fecha).
Fuente: Elaboración propia.*

También en el reparto de gastos aparece el término $\ddot{a}_{40:27}$ que sirve para el cálculo de gastos de gestión interna basados en 0,01% sobre la prima de inventario P' . El detalle de este término sería el siguiente:

$$\ddot{a}_{40:27} = P'' \sum_{t=0}^{(27+0)*1-1} (1 + 0,02)^{-t} \cdot {}_tP_{40}$$

*Expresión 50: Renta para el cálculo de los gastos de gestión interna 2023 (asegurado no jubilado en esta fecha).
Fuente: Elaboración propia.*

Las primas al ser ligadas a la supervivencia del asegurado y prepagables se indican con la letra “a” en minúscula y los dos puntos encima. El subíndice indica el inicio del contrato a los 40 años, seguido de la duración temporal de 27 años que es todo el periodo que pagará primas. La expresión $a_{49:18}$ representa el valor presente actuarial de los cobros de primas, descontados al tipo de interés determinista, considerando la supervivencia del asegurado año tras año hasta su jubilación.

La renta actuarial de los cobros de primas en este caso, describe los pagos que le quedan para realizar el asegurado desde el año 2022 hasta que alcance su jubilación a los 67 años. En este contexto, esta renta actuarial representa un período temporal de 4 años, calculado restando la edad de jubilación (67 años) menos la edad efectiva del contrato de seguro (63 años), lo que nos da una temporalidad de 4 años.

$$\ddot{a}_{63:4} = P \sum_{t=0}^{(4+0)*1-1} (1 + 0,02)^{-t} {}_tP_{63}$$

Expresión 51: Pago de primas por parte del asegurado a la aseguradora 2023 (asegurado no jubilado en esta fecha).
Fuente: Elaboración propia.

La simbología empleada en la expresión refleja la naturaleza de los cobros de primas en el contexto del contrato de seguro. El uso de la letra “a” en minúscula y los dos puntos encima denotan que estas primas son prepagables y están asociadas con la supervivencia del asegurado. El subíndice indica el inicio del contrato a los 40 años, seguido de la duración temporal de 4 años. La expresión $a_{49:18}$ representa el valor presente actuarial de los cobros de primas, descontados al tipo de interés determinista, considerando la supervivencia del asegurado año tras año hasta su jubilación.

Si se sustituyen los valores de estas rentas descritas a la Expresión 45 se obtiene que la provisión bajo IFRS17 debe ser de 17.712,45€ y bajo IFRS4 debe ser de 9.513,18€.

9.4. Anexos tasa de descuento

9.4.1. Nelson-Siegel

En este trabajo se han probado dos métodos para modelizar la estructura temporal de tipo de interés en el Anexo 8.3. se muestra la aplicación del Modelo de Nelson-Siegel

Modelo de Nelson – Siegel:

$$Y(t) = \beta_0 + \beta_1 \left(\frac{1 - e^{-\frac{t}{r_1}}}{\frac{t}{r_1}} \right) + \beta_2 \left(\frac{1 - e^{-\frac{t}{r_1}}}{\frac{t}{r_1}} - e^{-\frac{t}{r_1}} \right)$$

Expresión 52: Fórmula Nelson-Siegel.
Fuente: Elaboración propia.

$Y(t)$: es la tasa de interés en el tiempo t .

β_0 : es el componente de tasa de interés a largo plazo o nivel de la curva. Representa la tasa de interés a largo plazo que se espera que prevalezca en la economía.

β_1 : es el componente de pendiente de la curva. Representa la pendiente de la curva de tasas de interés a medida que cambian los plazos.

r_1 : es el parámetro de suavizado asociado al componente de pendiente. Controla la rapidez de ajuste la pendiente de la curva.

Nelson-Siegel		
Parámetro	Año 2023	Año 2022
beta0	0,031	0,032
beta1	0,239	0,015
beta2	-0,302	0,019
lambda1	2,322	12,792

Tabla 58: Parámetros optimizados
Nelson – Siegel.
Fuente: Elaboración propia

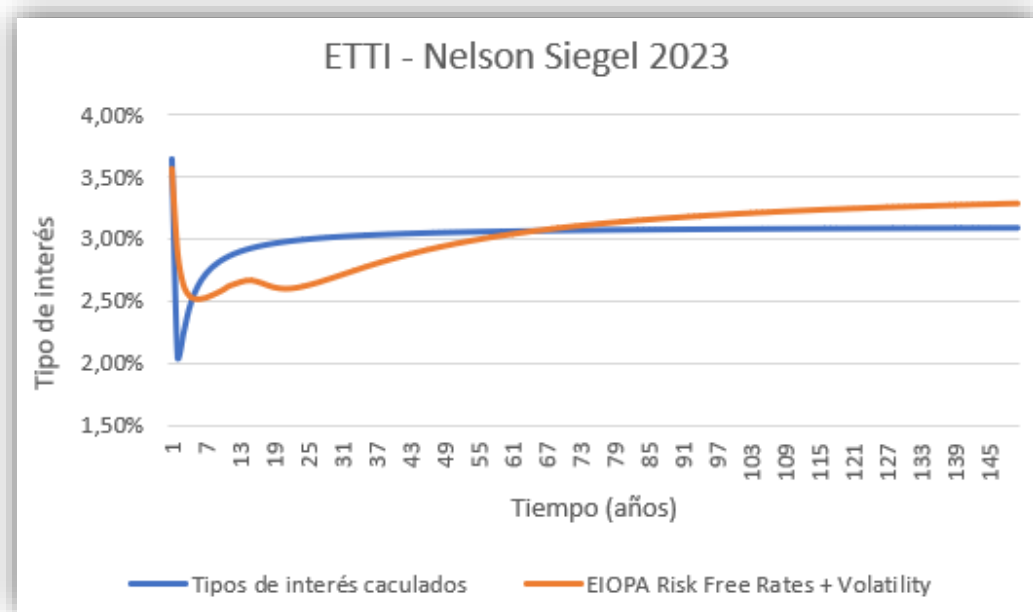


Tabla 59: Estructura temporal de tipos de interés y estructura obtenida por Nelson Siegel 2023.
Fuente: Elaboración propia.

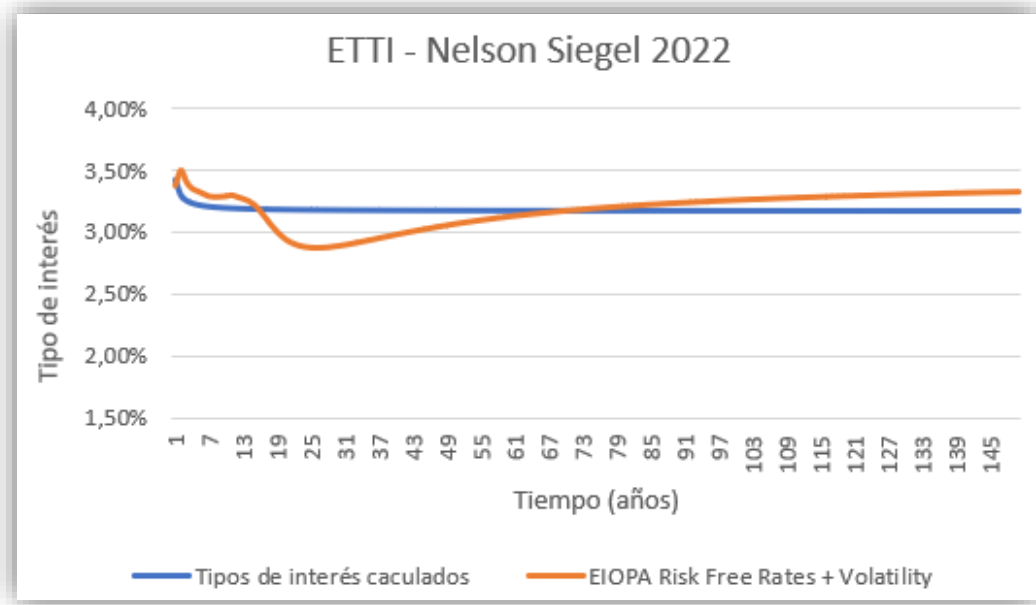


Gráfico 26: Estructura temporal de tipos de interés y estructura obtenida por Nelson Siegel 2022.
Fuente: Elaboración propia.

9.4.2. Tipo de interés EIOPA

EIOPA risk free rates + volatility Diciembre 2023	EIOPA risk free rates + volatility Diciembre 2022
3,557%	3,366%
2,890%	3,485%
2,639%	3,393%
2,550%	3,342%
2,523%	3,321%
2,521%	3,300%
2,531%	3,281%
2,549%	3,276%
2,570%	3,278%
2,593%	3,282%
2,623%	3,290%
2,639%	3,275%
2,655%	3,261%
2,667%	3,243%
2,670%	3,212%
2,660%	3,164%
2,645%	3,106%
2,628%	3,049%

Segmentación de riesgos y modelización actuarial en IFRS17

2,614%	2,997%
2,606%	2,955%
2,603%	2,923%
2,605%	2,901%
2,611%	2,886%
2,620%	2,877%
2,631%	2,872%
2,643%	2,871%
2,656%	2,872%
2,671%	2,875%
2,685%	2,881%
2,700%	2,887%
2,715%	2,894%
2,730%	2,902%
2,745%	2,911%
2,760%	2,920%
2,775%	2,929%
2,789%	2,938%
2,803%	2,948%
2,817%	2,957%
2,830%	2,966%
2,843%	2,976%
2,855%	2,985%
2,868%	2,994%
2,879%	3,002%
2,891%	3,011%
2,902%	3,019%
2,913%	3,027%
2,923%	3,035%
2,933%	3,043%
2,943%	3,050%
2,952%	3,058%
2,962%	3,065%
2,970%	3,072%
2,979%	3,078%
2,987%	3,085%
2,995%	3,091%
3,003%	3,097%
3,011%	3,103%
3,018%	3,109%
3,025%	3,114%
3,032%	3,120%
3,039%	3,125%
3,045%	3,130%
3,051%	3,135%

Segmentación de riesgos y modelización actuarial en IFRS17

3,058%	3,140%
3,064%	3,144%
3,069%	3,149%
3,075%	3,153%
3,080%	3,158%
3,086%	3,162%
3,091%	3,166%
3,096%	3,170%
3,101%	3,174%
3,105%	3,177%
3,110%	3,181%
3,114%	3,185%
3,119%	3,188%
3,123%	3,192%
3,127%	3,195%
3,131%	3,198%
3,135%	3,201%
3,139%	3,204%
3,143%	3,207%
3,147%	3,210%
3,150%	3,213%
3,154%	3,216%
3,157%	3,218%
3,161%	3,221%
3,164%	3,224%
3,167%	3,226%
3,170%	3,229%
3,173%	3,231%
3,176%	3,233%
3,179%	3,236%
3,182%	3,238%
3,185%	3,240%
3,188%	3,243%
3,190%	3,245%
3,193%	3,247%
3,196%	3,249%
3,198%	3,251%
3,201%	3,253%
3,203%	3,255%
3,205%	3,257%
3,208%	3,258%
3,210%	3,260%
3,212%	3,262%
3,215%	3,264%
3,217%	3,266%

Segmentación de riesgos y modelización actuarial en IFRS17

3,219%	3,267%
3,221%	3,269%
3,223%	3,271%
3,225%	3,272%
3,227%	3,274%
3,229%	3,275%
3,231%	3,277%
3,233%	3,278%
3,235%	3,280%
3,236%	3,281%
3,238%	3,283%
3,240%	3,284%
3,242%	3,285%
3,243%	3,287%
3,245%	3,288%
3,247%	3,289%
3,248%	3,291%
3,250%	3,292%
3,252%	3,293%
3,253%	3,294%
3,255%	3,296%
3,256%	3,297%
3,258%	3,298%
3,259%	3,299%
3,261%	3,300%
3,262%	3,301%
3,263%	3,302%
3,265%	3,303%
3,266%	3,305%
3,267%	3,306%
3,269%	3,307%
3,270%	3,308%
3,271%	3,309%
3,273%	3,310%
3,274%	3,311%
3,275%	3,312%
3,276%	3,313%
3,277%	3,314%
3,279%	3,314%
3,280%	3,315%
3,281%	3,316%
3,282%	3,317%

Tabla 60: Tipos de interés EIOPA diciembre 2022 y 2023.
Fuente: Elaboración propia.

9.5. Anexos Fair Value

9.5.1. EUROSAP 2022

Maturity	EUROSAP	EUROSAP (%)	EUROSAP CALCULADO	ERROR
1	2,942	0,02942	0,029578909	2,5252E-08
2	3,058	0,03058	0,029453338	1,26937E-06
3	2,951	0,02951	0,029313276	3,87005E-08
4	2,877	0,02877	0,029159184	1,51464E-07
5	2,836	0,02836	0,028991515	3,98811E-07
6		0	0,02881071	
7	2,788	0,02788	0,028617203	5,43469E-07
8		0	0,028411416	
9		0	0,028193764	
10	2,787	0,02787	0,02796465	8,95858E-09
11		0	0,027724471	
12		0	0,027473613	
13		0	0,027212456	
14		0	0,026941369	
15	2,737	0,02737	0,026660713	5,03087E-07
16		0	0,026370844	
17		0	0,026072106	
18		0	0,025764837	
19		0	0,025449368	
20	2,532	0,02532	0,025126021	3,76278E-08
21		0	0,024795112	
22		0	0,024456949	
23		0	0,024111832	
24		0	0,023760056	
25		0	0,023401909	
26		0	0,02303767	
27		0	0,022667614	
28		0	0,022292008	
29		0	0,021911114	
30	2,131	0,02131	0,021525186	4,63052E-08
31		0	0,021134474	
32		0	0,020739221	
33		0	0,020339663	
34		0	0,019936033	
35		0	0,019528555	
36		0	0,019117451	
37		0	0,018702935	
38		0	0,018285218	
39		0	0,017864503	
40		0	0,017440989	

Segmentación de riesgos y modelización actuarial en IFRS17

41		0	0,017014872	
42		0	0,016586341	
43		0	0,01615558	
44		0	0,015722769	
45		0	0,015288085	
46		0	0,014851697	
47		0	0,014413773	
48		0	0,013974473	
49		0	0,013533958	
50		0	0,013092379	
51		0	0,012649887	
52		0	0,012206628	
53		0	0,011762743	
54		0	0,01131837	
55		0	0,010873644	
56		0	0,010428694	
57		0	0,009983649	
58		0	0,00953863	
59		0	0,009093759	
60		0	0,008649151	
61		0	0,00820492	
62		0	0,007761176	
63		0	0,007318025	
64		0	0,006875572	
65		0	0,006433916	
66		0	0,005993157	
67		0	0,005553387	
68		0	0,0051147	
69		0	0,004677184	
70		0	0,004240925	
71		0	0,003806008	
72		0	0,003372513	
73		0	0,002940518	
74		0	0,0025101	
75		0	0,002081333	
76		0	0,001654286	
77		0	0,001229029	
78		0	0,000805628	
79		0	0,000384148	
80		0	-3,53505E-05	
81		0	-0,000452807	
82		0	-0,000868163	
83		0	-0,001281364	
84		0	-0,001692356	
85		0	-0,002101087	

Segmentación de riesgos y modelización actuarial en IFRS17

86		0	-0,002507509	
87		0	-0,002911572	
88		0	-0,003313231	
89		0	-0,003712442	
90		0	-0,004109162	
91		0	-0,004503352	
92		0	-0,004894971	
93		0	-0,005283982	
94		0	-0,00567035	
95		0	-0,006054041	
96		0	-0,006435021	
97		0	-0,006813261	
98		0	-0,007188729	
99		0	-0,007561398	
100		0	-0,007931241	
101		0	-0,008298233	
102		0	-0,008662349	
103		0	-0,009023566	
104		0	-0,009381863	
105		0	-0,00973722	
106		0	-0,010089617	
107		0	-0,010439036	
108		0	-0,010785461	
109		0	-0,011128876	
110		0	-0,011469266	
111		0	-0,011806618	
112		0	-0,012140918	
113		0	-0,012472157	
114		0	-0,012800322	
115		0	-0,013125406	
116		0	-0,013447398	
117		0	-0,013766291	
118		0	-0,014082079	
119		0	-0,014394755	
120		0	-0,014704315	
121		0	-0,015010754	
122		0	-0,015314069	
123		0	-0,015614257	
124		0	-0,015911316	
125		0	-0,016205245	
126		0	-0,016496043	
127		0	-0,016783711	
128		0	-0,01706825	
129		0	-0,01734966	
130		0	-0,017627945	

131		0	-0,017903108	
132		0	-0,01817515	
133		0	-0,018444078	
134		0	-0,018709895	
135		0	-0,018972606	
136		0	-0,019232217	
137		0	-0,019488735	
138		0	-0,019742165	
139		0	-0,019992516	
140		0	-0,020239795	
141		0	-0,02048401	
142		0	-0,02072517	
143		0	-0,020963283	
144		0	-0,02119836	
145		0	-0,02143041	
146		0	-0,021659444	
147		0	-0,021885471	
148		0	-0,022108504	
149		0	-0,022328553	
150		0	-0,02254563	
				3,02304E-06

Tabla 61: EUROSAP 2022.

Fuente: Elaboración propia.

9.5.2. EUROSAP 2023

Maturity	EUROSAP	EUROSAP (%)	EUROSAP CALCULADO	ERROR
1	3,489	0,03489	0,034962197	5,21242E-09
2	2,995	0,02995	0,02971984	5,29734E-08
3	2,743	0,02743	0,027465127	1,23391E-09
4	2,634	0,02634	0,026530775	3,63952E-08
5	2,589	0,02589	0,026134701	5,98784E-08
6		0	0,025952471	
7	2,572	0,02572	0,025856674	1,86799E-08
8		0	0,0257981	
9		0	0,025757416	
10	2,608	0,02608	0,025726599	1,24892E-07
11		0	0,025701986	
12		0	0,025681679	
13		0	0,025664565	
14		0	0,025649919	
15	2,665	0,02665	0,025637233	1,0257E-06
16		0	0,025626135	
17		0	0,025616344	

Segmentación de riesgos y modelización actuarial en IFRS17

18		0	0,025607641	
19		0	0,025599853	
20	2,606	0,02606	0,025592845	2,18234E-07
21		0	0,025586504	
22		0	0,02558074	
23		0	0,025575477	
24		0	0,025570652	
25		0	0,025566214	
26		0	0,025562117	
27		0	0,025558323	
28		0	0,0255548	
29		0	0,025551521	
30	2,414	0,02414	0,025548459	1,98376E-06
31		0	0,025545596	
32		0	0,025542911	
33		0	0,025540389	
34		0	0,025538016	
35		0	0,025535778	
36		0	0,025533664	
37		0	0,025531665	
38		0	0,025529771	
39		0	0,025527974	
40		0	0,025526267	
41		0	0,025524643	
42		0	0,025523096	
43		0	0,025521622	
44		0	0,025520214	
45		0	0,025518869	
46		0	0,025517582	
47		0	0,025516351	
48		0	0,02551517	
49		0	0,025514038	
50		0	0,025512951	
51		0	0,025511907	
52		0	0,025510902	
53		0	0,025509936	
54		0	0,025509005	
55		0	0,025508109	
56		0	0,025507244	
57		0	0,02550641	
58		0	0,025505604	
59		0	0,025504826	
60		0	0,025504074	
61		0	0,025503346	
62		0	0,025502642	

Segmentación de riesgos y modelización actuarial en IFRS17

63		0	0,02550196	
64		0	0,0255013	
65		0	0,025500659	
66		0	0,025500039	
67		0	0,025499436	
68		0	0,025498852	
69		0	0,025498284	
70		0	0,025497733	
71		0	0,025497197	
72		0	0,025496676	
73		0	0,025496169	
74		0	0,025495676	
75		0	0,025495197	
76		0	0,025494729	
77		0	0,025494274	
78		0	0,025493831	
79		0	0,025493399	
80		0	0,025492977	
81		0	0,025492566	
82		0	0,025492165	
83		0	0,025491774	
84		0	0,025491392	
85		0	0,025491019	
86		0	0,025490655	
87		0	0,025490299	
88		0	0,025489951	
89		0	0,025489611	
90		0	0,025489279	
91		0	0,025488953	
92		0	0,025488635	
93		0	0,025488324	
94		0	0,025488019	
95		0	0,025487721	
96		0	0,025487429	
97		0	0,025487143	
98		0	0,025486863	
99		0	0,025486588	
100		0	0,025486319	
101		0	0,025486056	
102		0	0,025485797	
103		0	0,025485544	
104		0	0,025485295	
105		0	0,025485051	
106		0	0,025484812	
107		0	0,025484577	

Segmentación de riesgos y modelización actuarial en IFRS17

108		0	0,025484347	
109		0	0,025484121	
110		0	0,025483898	
111		0	0,02548368	
112		0	0,025483466	
113		0	0,025483256	
114		0	0,025483049	
115		0	0,025482846	
116		0	0,025482646	
117		0	0,02548245	
118		0	0,025482257	
119		0	0,025482067	
120		0	0,025481881	
121		0	0,025481697	
122		0	0,025481517	
123		0	0,02548134	
124		0	0,025481165	
125		0	0,025480993	
126		0	0,025480824	
127		0	0,025480658	
128		0	0,025480494	
129		0	0,025480333	
130		0	0,025480174	
131		0	0,025480017	
132		0	0,025479863	
133		0	0,025479712	
134		0	0,025479562	
135		0	0,025479415	
136		0	0,02547927	
137		0	0,025479127	
138		0	0,025478986	
139		0	0,025478847	
140		0	0,02547871	
141		0	0,025478576	
142		0	0,025478443	
143		0	0,025478311	
144		0	0,025478182	
145		0	0,025478055	
146		0	0,025477929	
147		0	0,025477805	
148		0	0,025477682	
149		0	0,025477561	
150		0	0,025477442	
				3,52695E-06

Tabla 62: EUROSAP 2023.

Fuente: Elaboración propia.

9.5.3. EUROSAP ENERO 2022

Maturity	EUROSAP	EUROSAP (%)	EUROSAP CALCULADO	ERROR
1	-0,482	-0,00482	-0,004844857	6,17866E-10
2	-0,243	-0,00243	-0,002318139	1,25129E-08
3	-0,077	-0,00077	-0,000922276	2,31881E-08
4	0,028	0,00028	0,000184519	9,11662E-09
5	0,098	0,00098	0,001119809	1,95466E-08
6		0	0,001894351	
7	0,215	0,00215	0,002522971	1,39108E-07
8		0	0,003029537	
9		0	0,003439134	
10	0,377	0,00377	0,003773518	1,23791E-11
11		0	0,00404992	
12		0	0,004281413	
13		0	0,004477762	
14		0	0,004646249	
15	0,543	0,00543	0,004792346	4,06603E-07
16		0	0,00492021	
17		0	0,005033043	
18		0	0,005133343	
19		0	0,005223088	
20	0,581	0,00581	0,005303858	2,5618E-07
21		0	0,005376936	
22		0	0,005443371	
23		0	0,005504029	
24		0	0,005559632	
25		0	0,005610787	
26		0	0,005658007	
27		0	0,005701729	
28		0	0,005742328	
29		0	0,005780128	
30	0,5	0,005	0,005815407	6,64888E-07
31		0	0,00584841	
32		0	0,00587935	
33		0	0,005908416	
34		0	0,005935771	
35		0	0,005961564	
36		0	0,005985923	
37		0	0,006008966	

Segmentación de riesgos y modelización actuarial en IFRS17

38		0	0,006030796	
39		0	0,006051506	
40		0	0,006071181	
41		0	0,006089897	
42		0	0,006107721	
43		0	0,006124716	
44		0	0,006140938	
45		0	0,006156439	
46		0	0,006171267	
47		0	0,006185464	
48		0	0,006199069	
49		0	0,006212118	
50		0	0,006224646	
51		0	0,006236682	
52		0	0,006248256	
53		0	0,006259393	
54		0	0,006270117	
55		0	0,006280451	
56		0	0,006290417	
57		0	0,006300032	
58		0	0,006309316	
59		0	0,006318286	
60		0	0,006326956	
61		0	0,006335342	
62		0	0,006343457	
63		0	0,006351315	
64		0	0,006358928	
65		0	0,006366306	
66		0	0,00637346	
67		0	0,006380401	
68		0	0,006387138	
69		0	0,00639368	
70		0	0,006400034	
71		0	0,00640621	
72		0	0,006412214	
73		0	0,006418054	
74		0	0,006423735	
75		0	0,006429266	
76		0	0,00643465	
77		0	0,006439895	
78		0	0,006445006	
79		0	0,006449987	
80		0	0,006454843	
81		0	0,00645958	
82		0	0,006464201	

Segmentación de riesgos y modelización actuarial en IFRS17

83		0	0,00646871	
84		0	0,006473113	
85		0	0,006477411	
86		0	0,00648161	
87		0	0,006485712	
88		0	0,006489721	
89		0	0,00649364	
90		0	0,006497472	
91		0	0,00650122	
92		0	0,006504886	
93		0	0,006508473	
94		0	0,006511984	
95		0	0,006515421	
96		0	0,006518787	
97		0	0,006522083	
98		0	0,006525312	
99		0	0,006528475	
100		0	0,006531575	
101		0	0,006534614	
102		0	0,006537594	
103		0	0,006540515	
104		0	0,00654338	
105		0	0,006546191	
106		0	0,006548949	
107		0	0,006551655	
108		0	0,006554311	
109		0	0,006556918	
110		0	0,006559478	
111		0	0,006561992	
112		0	0,006564461	
113		0	0,006566886	
114		0	0,006569269	
115		0	0,00657161	
116		0	0,006573911	
117		0	0,006576172	
118		0	0,006578395	
119		0	0,006580581	
120		0	0,00658273	
121		0	0,006584844	
122		0	0,006586923	
123		0	0,006588969	
124		0	0,006590981	
125		0	0,006592961	
126		0	0,00659491	
127		0	0,006596828	

128		0	0,006598716	
129		0	0,006600575	
130		0	0,006602405	
131		0	0,006604208	
132		0	0,006605983	
133		0	0,006607731	
134		0	0,006609453	
135		0	0,00661115	
136		0	0,006612821	
137		0	0,006614469	
138		0	0,006616092	
139		0	0,006617692	
140		0	0,00661927	
141		0	0,006620824	
142		0	0,006622357	
143		0	0,006623869	
144		0	0,006625359	
145		0	0,006626829	
146		0	0,006628279	
147		0	0,006629709	
148		0	0,00663112	
149		0	0,006632512	
150		0	0,006633885	
				1,53177E-06

Tabla 63: EUROSAP ENERO 2022.

Fuente: Elaboración propia.

9.5.4. BONO ALEMÁN 2022

Maturity	EUROSAP	BONO ALEMÁN (%)	BONO ALEMÁN CALCULADO	ERROR
1		0	0,028643854	
2	2,71	0,0271	0,027166525	4,42553E-09
3		0	0,0261697	
4		0	0,025535752	
5	2,53	0,0253	0,025171062	1,66249E-08
6		0	0,025001598	
7	2,49	0,0249	0,02496924	4,79424E-09
8		0	0,025028768	
9		0	0,025145364	
10	2,51	0,0251	0,025292575	3,70853E-08
11		0	0,025450633	

Segmentación de riesgos y modelización actuarial en IFRS17

12		0	0,025605082	
13		0	0,025745668	
14		0	0,025865424	
15	2,61	0,0261	0,025959935	1,96182E-08
16		0	0,026026746	
17		0	0,026064886	
18		0	0,026074483	
19		0	0,026056462	
20		0	0,026012309	
21		0	0,025943881	
22		0	0,025853261	
23		0	0,025742646	
24		0	0,025614266	
25		0	0,025470318	
26		0	0,025312925	
27		0	0,025144101	
28		0	0,024965734	
29		0	0,024779572	
30	2,46	0,0246	0,024587217	1,63405E-10
31		0	0,024390124	
32		0	0,024189604	
33		0	0,02398683	
34		0	0,023782841	
35		0	0,023578555	
36		0	0,02337477	
37		0	0,023172182	
38		0	0,022971386	
39		0	0,022772889	
40		0	0,02257712	
41		0	0,022384434	
42		0	0,022195121	
43		0	0,022009417	
44		0	0,021827503	
45		0	0,02164952	
46		0	0,021475568	
47		0	0,021305711	
48		0	0,021139987	
49		0	0,020978408	
50		0	0,020820963	
51		0	0,020667624	
52		0	0,020518347	
53		0	0,020373074	
54		0	0,02023174	
55		0	0,020094267	
56		0	0,019960574	

Segmentación de riesgos y modelización actuarial en IFRS17

57		0	0,019830572	
58		0	0,019704169	
59		0	0,01958127	
60		0	0,019461778	
61		0	0,019345596	
62		0	0,019232625	
63		0	0,019122768	
64		0	0,019015927	
65		0	0,018912006	
66		0	0,018810912	
67		0	0,01871255	
68		0	0,018616832	
69		0	0,018523668	
70		0	0,018432972	
71		0	0,018344661	
72		0	0,018258654	
73		0	0,018174872	
74		0	0,01809324	
75		0	0,018013683	
76		0	0,017936131	
77		0	0,017860516	
78		0	0,017786772	
79		0	0,017714835	
80		0	0,017644643	
81		0	0,017576139	
82		0	0,017509266	
83		0	0,017443969	
84		0	0,017380195	
85		0	0,017317896	
86		0	0,017257021	
87		0	0,017197525	
88		0	0,017139363	
89		0	0,017082493	
90		0	0,017026872	
91		0	0,016972462	
92		0	0,016919224	
93		0	0,016867121	
94		0	0,01681612	
95		0	0,016766184	
96		0	0,016717283	
97		0	0,016669385	
98		0	0,01662246	
99		0	0,016576478	
100		0	0,016531413	
101		0	0,016487236	

Segmentación de riesgos y modelización actuarial en IFRS17

102		0	0,016443924	
103		0	0,01640145	
104		0	0,01635979	
105		0	0,016318922	
106		0	0,016278824	
107		0	0,016239474	
108		0	0,016200852	
109		0	0,016162937	
110		0	0,01612571	
111		0	0,016089154	
112		0	0,016053249	
113		0	0,01601798	
114		0	0,015983328	
115		0	0,015949279	
116		0	0,015915817	
117		0	0,015882926	
118		0	0,015850592	
119		0	0,015818802	
120		0	0,015787541	
121		0	0,015756796	
122		0	0,015726556	
123		0	0,015696807	
124		0	0,015667537	
125		0	0,015638736	
126		0	0,015610392	
127		0	0,015582494	
128		0	0,015555032	
129		0	0,015527996	
130		0	0,015501376	
131		0	0,015475162	
132		0	0,015449345	
133		0	0,015423916	
134		0	0,015398867	
135		0	0,015374189	
136		0	0,015349873	
137		0	0,015325913	
138		0	0,0153023	
139		0	0,015279027	
140		0	0,015256086	
141		0	0,01523347	
142		0	0,015211173	
143		0	0,015189188	
144		0	0,015167508	
145		0	0,015146128	
146		0	0,01512504	

147		0	0,015104239	
148		0	0,015083719	
149		0	0,015063474	
150		0	0,0150435	
				8,27116E-08

Tabla 64: BONO ALEMÁN 2022.

Fuente: Elaboración propia.

9.5.5. BONO ALEMÁN 2023

Maturity	EUROSWAP	EUROSWAP (%)	EUROSWAP CALCULADO	ERROR
1		0	0,025350142	
2	2,39	0,0239	0,023003611	8,03513E-07
3		0	0,021466472	
4		0	0,020508668	
5	1,94	0,0194	0,019961187	3,14931E-07
6		0	0,019700427	
7	1,93	0,0193	0,019636481	1,1322E-07
8		0	0,019704374	
9		0	0,019857511	
10	2,02	0,0202	0,020062798	1,88244E-08
11		0	0,020297015	
12		0	0,020544124	
13		0	0,020793279	
14		0	0,021037357	
15	2,23	0,0223	0,021271886	1,05702E-06
16		0	0,021494253	
17		0	0,021703137	
18		0	0,021898091	
19		0	0,022079246	
20		0	0,022247103	
21		0	0,022402382	
22		0	0,022545919	
23		0	0,022678598	
24		0	0,022801305	
25		0	0,022914894	
26		0	0,023020176	
27		0	0,023117903	
28		0	0,023208767	
29		0	0,023293399	
30	2,26	0,0226	0,02337237	5,96556E-07
31		0	0,023446198	
32		0	0,023515345	
33		0	0,023580228	

Segmentación de riesgos y modelización actuarial en IFRS17

34		0	0,023641221	
35		0	0,023698656	
36		0	0,023752835	
37		0	0,023804026	
38		0	0,02385247	
39		0	0,023898383	
40		0	0,023941961	
41		0	0,023983379	
42		0	0,024022795	
43		0	0,024060354	
44		0	0,024096185	
45		0	0,024130407	
46		0	0,024163127	
47		0	0,024194442	
48		0	0,024224443	
49		0	0,024253212	
50		0	0,024280823	
51		0	0,024307346	
52		0	0,024332845	
53		0	0,024357378	
54		0	0,024380999	
55		0	0,024403759	
56		0	0,024425704	
57		0	0,024446878	
58		0	0,02446732	
59		0	0,024487069	
60		0	0,024506158	
61		0	0,024524621	
62		0	0,024542488	
63		0	0,024559787	
64		0	0,024576545	
65		0	0,024592787	
66		0	0,024608537	
67		0	0,024623817	
68		0	0,024638647	
69		0	0,024653047	
70		0	0,024667036	
71		0	0,02468063	
72		0	0,024693847	
73		0	0,024706701	
74		0	0,024719208	
75		0	0,024731382	
76		0	0,024743235	
77		0	0,024754781	
78		0	0,02476603	

Segmentación de riesgos y modelización actuarial en IFRS17

79		0	0,024776994	
80		0	0,024787685	
81		0	0,024798111	
82		0	0,024808283	
83		0	0,02481821	
84		0	0,024827901	
85		0	0,024837363	
86		0	0,024846606	
87		0	0,024855636	
88		0	0,024864461	
89		0	0,024873087	
90		0	0,024881522	
91		0	0,024889771	
92		0	0,024897841	
93		0	0,024905738	
94		0	0,024913467	
95		0	0,024921032	
96		0	0,024928441	
97		0	0,024935696	
98		0	0,024942803	
99		0	0,024949767	
100		0	0,024956592	
101		0	0,024963281	
102		0	0,024969839	
103		0	0,02497627	
104		0	0,024982577	
105		0	0,024988764	
106		0	0,024994835	
107		0	0,025000792	
108		0	0,025006638	
109		0	0,025012377	
110		0	0,025018012	
111		0	0,025023546	
112		0	0,02502898	
113		0	0,025034319	
114		0	0,025039564	
115		0	0,025044717	
116		0	0,025049782	
117		0	0,02505476	
118		0	0,025059654	
119		0	0,025064465	
120		0	0,025069196	
121		0	0,025073849	
122		0	0,025078426	
123		0	0,025082929	

124		0	0,025087358	
125		0	0,025091717	
126		0	0,025096007	
127		0	0,025100229	
128		0	0,025104385	
129		0	0,025108477	
130		0	0,025112506	
131		0	0,025116473	
132		0	0,02512038	
133		0	0,025124229	
134		0	0,02512802	
135		0	0,025131754	
136		0	0,025135434	
137		0	0,02513906	
138		0	0,025142634	
139		0	0,025146156	
140		0	0,025149628	
141		0	0,025153051	
142		0	0,025156425	
143		0	0,025159752	
144		0	0,025163033	
145		0	0,025166269	
146		0	0,025169461	
147		0	0,025172609	
148		0	0,025175714	
149		0	0,025178778	
150		0	0,025181801	
				2,90406E-06

Tabla 65: BONO ALEMÁN 2023.
Fuente: Elaboración propia.

9.6. Anexos resumen conferencia asistida

Conferencia: Seminario 14 marzo sobre IFRS 17

Ponente: Don Antonio Sau Ortega

Día: 14/3/2024

Observación: Se escriben en negrita las preguntas que he hecho al ponente. *En cursiva las respuestas suyas.*

La normativa IFRS17 traslada las hipótesis al mundo de la Contabilidad.

Solvencia II no tiene relación con la contabilidad de las empresas.

Visión general:

- Si se coge la vida de una póliza, el resultado será el mismo ahora que en IFRS17. Lo que cambia es como deben valorarse y contabilizarse los pasivos por contratos de seguro, pero el resultado de lo que va a ganar una compañía de seguros, al final de la vida de una póliza no va a cambiar (el acumulado).
- IFRS17 afecta a como valoramos y contabilizamos los pasivos por contratos de seguros.
- Es una normativa basada en principios.
- Entró en vigor el 1 enero de 2023.
- Aplica a las compañías que cotizan y aquellas que consolidan bajo Normas Internacionales.
- La DGSFP tiene que coger la normativa general y transponerla a la local.

Clasificación de contratos: un punto importante es tener separados los contratos de seguro y de reaseguro.

- La normativa define como Portfolio los riesgos similares que se gestionan de forma conjunta.
- Indica que deben separarse en relación a la rentabilidad-> separar las onerosas y las rentables.
- Límites del contrato: hasta que momento hay que valorar los pasivos. Ejemplo de un seguro de autos-> después de un año la compañía puede modificar la prima y el asegurado decidir que no quiere este seguro (cancelación del contrato)-> límite del contrato= 1año. Los límites del contrato definidos en Solvencia II no son los mismos que los de IFRS17.

Métodos:

- Método general basado en bloques
- Variable Fee Approach
- Método de asignación de prima

Transición: pasar de valorar de IFRS4 a IFRS17. Métodos para hacer la transición:

- Método retrospectivo
- Método retrospectivo modificado
- Enfoque Fair Value

Hay dos tipos de pasivos en las compañías de seguros:

- El pasivo que tienen que dotar por siniestros que todavía no han ocurrido.
- El pasivo que tienen que dotar cuando ocurre un siniestro.

Método General

- Estimación de flujos de entrada y salida-> El BEL (Best Estimate Liabilities) se hace con hipótesis
- El CSM es el beneficio futuro que esperamos para un contrato.
- El Risk Adjustment es como un “colchón” para si se desvían mis hipótesis.

- Si estimo que voy a perder hay que dotar la pérdida a Pérdidas y Ganancias. Esta pérdida luego se revierte.

Los resultados deben ser los mismos en contabilidad local que en IFRS17. Existen algunas diferencias:

- Se dota el pasivo aunque no hayas cobrado la prima (IFRS4).
- Tienes que dotar pasivo cuando cobras prima (IFRS17).
- Contabilidad local hay provisión matemática y PPNC. -> pasivo por cobertura restante.

Variable Fee Approach-> se aplica básicamente a productos Unit Linked. -> en este tipo de productos hay cobro de comisión para gestión de activos.

En IFRS17 -> como ingreso tengo la liberación del pasivo (siniestros y gastos esperados).

PREGUNTAS QUE HE REALIZADO AL PONENTE:

Pregunta 1: ¿Cuándo se trata de un producto de vida en el que hay una parte que se invierte en un fondo de inversión, qué método aplicarías?

Método general con matices previstos en la normativa.

Pregunta 2: ¿Se continúan generando asimetrías contables en IFRS17?

Sí, y bastante importantes. El típico producto de ahorro preguntado antes, en la contabilidad local, el cálculo de la provisión matemática es de los más sencillos que existen: cobro prima de 100 -1 de gastos – prima de riesgo y todo capitalizado queda 102. La provisión matemática sería de 102. En IFRS17 ya no es así, hay que capitalizar flujos, calcular tasas de descuento que cambian, etc.

Pregunta 3: ¿Como se libera el CSM?

La normativa dice que tienes que tener en cuenta el servicio que das al asegurado.

- *A grandes rasgos se está utilizando para vida riesgo la suma asegurada.*
- *Vida Ahorro normalmente se usa el valor del fondo*
- *Productos de renta se usa el importe de la renta.*

Esto se llama el patrón de liberación del CSM, es decir, se va liberando este beneficio futuro a la cuenta de resultados.

9.7. Anexos scripts en R elaborados para el trabajo

```
install.packages("tidyverse")
library(tidyverse)
install.packages("dplyr")
library(dplyr)
library(ggplot2)
install.packages("readxl")
library(readxl)
Portlofio_vida <-
read_excel("G:/Esta_es_mi_carpeta2/Excels/datos_filtrados_2
000.xlsx")
maximo_valor <- max(Portlofio_vida$Capital[-1])

str(Portlofio_vida)
summary(Portlofio_vida)
library(ggplot2)
print(ggplot() + geom_point(aes(x = Age_Actuarial , y =
Capital), data = Portlofio_vida, alpha = 0.5) +
ggtitle('Conjunto de Datos'))
insurance.scale<-data.frame(scale((Portlofio_vida[-
1,c(5,9)])))
set.seed(1234)
wcss <- vector()
for(i in 1:20){
  wcss[i] <- sum(kmeans(insurance.scale, i)$withinss)
}
print(ggplot() + geom_point(aes(x = 1:20, y = wcss), color
= 'blue') +
  geom_line(aes(x = 1:20, y = wcss), color = 'blue') +
  ggtitle("Método del Codo") +
  xlab('Cantidad de Centroides k') +
  ylab('WCSS'))
#Para seleccionar el valor óptimo de k, se escoje
#entonces ese punto en donde ya no se dejan de producir
#variaciones importantes del valor de WCSS al aumentar k.

#Evaluaremos k=3.
set.seed(1234)
kmeans <- kmeans(insurance.scale, 3, iter.max = 1000,
nstart = 10)

#prueba
kmeans$cluster
centroid<-kmeans$centers
euclidean_distance <- function(point, centroid) {
  sqrt(sum((point - centroid)^2))
}
```

```
}
distances_to_centroids <- numeric(length(kmeans$cluster))
for (i in 1:length(kmeans$cluster))
{
  if (kmeans$cluster[i]==1) {
    distances_to_centroids[i] <-
euclidean_distance(insurance.scale[i, ], centroid[1,])}
  if (kmeans$cluster[i]==2) {
    distances_to_centroids[i] <-
euclidean_distance(insurance.scale[i, ], centroid[2,])}
  if (kmeans$cluster[i]==3) {
    distances_to_centroids[i] <-
euclidean_distance(insurance.scale[i, ], centroid[3,])}
}

distances_to_centroids
indices_mayores <- which(distances_to_centroids > 5)
indices_mayores
kmeans$cluster[58]
distances_to_centroids[58]
kmeans$cluster[306]
distances_to_centroids[306]
kmeans$cluster[552]
distances_to_centroids[552]
kmeans$cluster[657]
distances_to_centroids[657]
kmeans$cluster[668]#este no
distances_to_centroids[668]
kmeans$cluster[1053]
distances_to_centroids[1053]
kmeans$cluster[1129]
distances_to_centroids[1129]
kmeans$cluster[1213]
distances_to_centroids[1213]
kmeans$cluster[1412]
distances_to_centroids[1412]
kmeans$cluster[1418]
distances_to_centroids[1418]
kmeans$cluster[1809]
distances_to_centroids[1809]
kmeans$cluster[1854]
distances_to_centroids[1854]
#iter.max son el máximo de iteraciones a aplicar al
algoritmo.
#nstart es la cantidad de conjuntos de centroides que
emplea internamente el mismo para ejecutar sus cálculos.
```

```
insurance.scale$cluster <- kmeans$cluster
print(ggplot() + geom_point(aes(x = Age_Actuarial, y =
Capital, color = cluster), data = insurance.scale, size =
2) +
  scale_colour_gradientn(colours=rainbow(4)) +
  geom_point(aes(x = kmeans$centers[, 1], y =
kmeans$centers[, 2]), color = 'black', size = 3) +
  ggtitle('Clusters de Datos con k = 3 / K-Medios') +
  xlab('Edad asegurado') + ylab('Capital'))
#Alternativa
print(ggplot() + geom_point(aes(x = Age_Actuarial, y =
Capital, color = cluster), data = insurance.scale, size =
2) +
  scale_colour_gradientn(colours=rainbow(4)) +
  ggtitle('Clusters de Datos con k = 13 / K-Medios')
+
  xlab('X') + ylab('Y'))
#Con 4 centroides
set.seed(1234)
kmeans <- kmeans(insurance.scale, 14, iter.max = 1000,
nstart = 10)
insurance.scale$cluster <- kmeans$cluster
print(ggplot() + geom_point(aes(x = Age_Actuarial, y =
Capital, color = cluster), data = insurance.scale, size =
2) +
  scale_colour_gradientn(colours=rainbow(4)) +
  geom_point(aes(x = kmeans$centers[, 1], y =
kmeans$centers[, 2]), color = 'black', size = 3) +
  ggtitle('Clusters de Datos con k = 14/ K-Medios') +
  xlab('X') + ylab('Y'))
#Con 5 centroides
set.seed(1234)
kmeans <- kmeans(insurance.scale, 15, iter.max = 1000,
nstart = 10)
insurance.scale$cluster <- kmeans$cluster
print(ggplot() + geom_point(aes(x = Age_Actuarial, y =
Capital, color = cluster), data = insurance.scale, size =
2) +
  scale_colour_gradientn(colours=rainbow(4)) +
  geom_point(aes(x = kmeans$centers[, 1], y =
kmeans$centers[, 2]), color = 'black', size = 3) +
  ggtitle('Clusters de Datos con k = 15 / K-Medios')
+
  xlab('X') + ylab('Y'))
```

```
install.packages("fpc")
library(fpc)
clustering_indices <- cluster.stats(dist(insurance.scale),
kmeans$cluster)
print(clustering_indices)
#índice de Dunn
dunn_index <- clustering_indices$dunn
cat("Dunn Index:", dunn_index, "\n")
#Silhouette
clustering_indices$avg.silwidth
#Calinki
clustering_indices$ch
#Entropía
clustering_indices$entropy
#Índice de Dunnmodificado
clustering_indices$dunn2
#Índice S
clustering_indices$sindex
# Calculate the Davies-Bouldin Index
db_index <- clusterSim::index.DB(insurance.scale,
kmeans$cluster)
print(db_index)
install.packages("flexclust")
library(flexclust)
install.packages("mclust")
library(mclust)
install.packages("clusterSim")
library(clusterSim)
library(cluster)
install.packages("hopkins")
library(hopkins)
library(tidytext)
library(scatterplot3d)
library(tidyverse)

#PCA
# Perform PCA and extract results
pca <- prcomp(insurance.scale, scale. = TRUE)
pca_results <- factoextra::get_pca(pca)
plot(pca)
summary(pca)
##AGRUPAMIENTO JERÁRQUICO
#es un método de agrupamiento que basa su funcionamiento
#en encontrar jerarquías en los datos de entrada a
#partir de generar grupos basados en la cercanía o
#semejanza de tales datos. En el caso aglomerativo, se
#empiezan calculando los puntos de los datos de entrada
```



```
#que estén más cercanos y se crea un grupo entre ellos.
#Luego se calculan los siguientes pares más cercanos y
#de manera ascendente se van generando grupos de clases
#que, de manera visual podrán observarse a partir de la
#construcción de un Dendrograma.
install.packages("ggdendro")
library(ggdendro)
insurance.scale<-data.frame(scale((Portafolio_Vida_breve[-
1,c(3,4)])))
dendrogram <- hclust(dist(insurance.scale, method =
'euclidean'), method = 'ward.D')
dendrogram2 <- hclust(dist(insurance.scale, method =
'euclidean'), method = 'average')
ggdendrogram(dendrogram2, rotate = FALSE, labels = FALSE,
theme_dendro = TRUE) +
  labs(title = "Dendrograma")
# Crea los clusters cortando el dendrograma en el nivel
deseado (en este caso, 3 clusters)
cutree_clusters <- cutree(dendrogram2, k = 3)

# Imprime las asignaciones de cada elemento a un cluster
print(cutree_clusters)

# Combinar las asignaciones de clusters con tus datos
originales
clustered_data <- cbind(insurance.scale, Cluster =
cutree_clusters)

# Crear un gráfico de dispersión
ggplot(clustered_data, aes(x = Age_Actuarial, y = Capital,
color = factor(Cluster))) +
  geom_point() +
  labs(x = "Edad asegurado", y = "Capital", title =
"Gráfico de Dispersión de Datos Agrupados en Clusters") +
  scale_color_discrete(name = "Cluster")

mat_dist<-dist(insurance.scale,method="euclidean")
hc_euclidea_complete <- hclust(d = mat_dist, method =
"complete")
cor(x = mat_dist, cophenetic(hc_euclidea_complete))
hc_euclidea_average <- hclust(d = mat_dist, method =
"average")
cor(x = mat_dist, cophenetic(hc_euclidea_average))
hc_euclidea_single <- hclust(d = mat_dist, method =
"single")
cor(x = mat_dist, cophenetic(hc_euclidea_single))
```

```
hc_euclidea_ward.D2 <- hclust(d = mat_dist, method =
"ward.D2")
cor(x = mat_dist, cophenetic(hc_euclidea_ward.D2))
hc_euclidea_median <- hclust(d = mat_dist, method =
"median")
cor(x = mat_dist, cophenetic(hc_euclidea_median))
hc_euclidea_centroid <- hclust(d = mat_dist, method =
"centroid")
cor(x = mat_dist, cophenetic(hc_euclidea_centroid))
hc_euclidea_mcquitty <- hclust(d = mat_dist, method =
"mcquitty")
cor(x = mat_dist, cophenetic(hc_euclidea_mcquitty))
#índice de Dunn
dunn_index <- clustering_indices$dunn
cat("Dunn Index:", dunn_index, "\n")
# Calculate the Davies-Bouldin Index
db_index <- clusterSim::index.DB(insurance.scale,
kmeans$cluster)
print(db_index)
class(insurance.scale)
#Clustering-Fuzzy C Means Clustering
install.packages("ppclust")
require(ppclust)
library(ppclust)
install.packages("factoextra")
library(factoextra)
install.packages("dplyr")
library(dplyr)
install.packages("cluster")
library(cluster)
install.packages("fclust")
library(fclust)
install.packages("psych")
library(psych)
res.fcm<-fcm(insurance.scale,centers = 3)
res.fcm$func.val#Esto me indica la calidad del clustering
#será mejor cuanto menor sea el valor

res.fcm$iter
#número de iteraciones que realizó el algoritmo de
clustering difuso (Fuzzy C-Means) antes de converger hacia
una solución.
#cuántas veces el algoritmo ajustó los centros
#de los clusters y recalculó la asignación
#difusa de los puntos de datos antes de llegar
#a una solución óptima (o una solución que
#cumple con los criterios de convergencia
```

```
#predefinidos).

res.fcm$best.start
# proporciona el índice (o número) de la mejor
#configuración inicial entre las que se
#probaron.
#Fuzzy C-Means con múltiples configuraciones
#iniciales (nstart > 1), el algoritmo calcula
#la función objetivo (generalmente, la suma de
#las distancias al cuadrado ponderadas) para
#cada ejecución y luego selecciona la ejecución
#con la mejor función objetivo, es decir, la
#configuración inicial que dio lugar a la
#solución de clustering más óptima.

summary(res.fcm)
install.packages("RColorBrewer")
library(RColorBrewer)
my_palette <- colorRampPalette(viridis::viridis_pal()(13))

res.fcm2 <- ppclust2(res.fcm, "kmeans")
fviz_cluster(res.fcm2, data = insurance.scale,
              ellipse.type = "convex",
              palette = "jco",
              repel = TRUE)

res.fcm3 <- ppclust2(res.fcm, "fanny")

cluster::clusplot(scale(insurance.scale), res.fcm3$cluster,
                  main = "Cluster plot del Portafolio
Vida",
                  color=TRUE, labels = 2, lines = 2, cex=1)

#Random Forest
install.packages("randomForest")
library(caret)
library(randomForest)
install.packages("lattice")
library(lattice)

set.seed(2018)
head(Portafolio_Vida_breve)
insurance.scale<-data.frame((Portafolio_Vida_breve[-
1,1:16]))
insurance.scale$r<-factor(insurance.scale$r)
insurance.scale$s<-factor(insurance.scale$s)
```

```
insurance.scale$Month<-factor(insurance.scale$Month)
insurance.scale$Gender<-factor(insurance.scale$Gender)
insurance.scale$Class_Age<-
factor(insurance.scale$Class_Age)
training.ids<-
createDataPartition(insurance.scale$Capital,p=0.7,list = F)

mod<-randomForest(x=insurance.scale[-
1,1:15],y=insurance.scale[-
1,16],ntree=500,keep.forest=TRUE)
#por defecto se crean 500 árboles
pred<-predict(mod,insurance.scale[-training.ids,])

#Matriz de confusión
table(insurance.scale[-
training.ids,"Class_Age"],pred,dnn=c("Actual","Predicho"))

#DBSCAN
install.packages("dbscan")
library(dbscan)
install.packages("stats")
library(stats)
insurance.scale
kNNdistplot(insurance.scale,k=3)
abline(h=.5, col = "red", lty=2)
insurance.scale.2 <-
as.data.frame(sapply(insurance.scale[,1:2], as.numeric))
db<-dbscan(insurance.scale.2,eps=0.5,MinPts=4)
plot(insurance.scale.2,col=db$cluster)
points(insurance.scale.2[db$cluster==0,],pch=3,col="grey")
hullplot(insurance.scale.2,db)
#Clustering espectral
#kmeans application
set.seed(1748)
##Run Kmeans algorithm with 3 clusters
km<-kmeans(insurance.scale,3)
##Retrieve and store cluster assignments
kmeans_clusters<-km$cluster
kmeans_results<-
cbind(insurance.scale,cluster=as.factor(kmeans_clusters))
head(kmeans_results)
##Create similarity matrix of Euclidean distance between
points
S<-as.matrix(dist(insurance.scale))
##Create Degree matrix
```

```
D<-
matrix(0,nrow=nrow(insurance.scale),ncol=nrow(insurance.sca
le))
for (i in 1:nrow(insurance.scale)) {

  # Find top 10 nearest neighbors using Euclidean distance
  index <- order(S[i,])[2:11]

  # Assign value to neighbors
  D[i,][index] <- 1
}
#find mutual neighbours
D=D+t(D)
D[D==2]=1
#find degrees of vertices
degrees=colSums(D)
n=nrow(D)
##Compute Laplacian matrix
#Since k>2 clusters(3), we normalize the Laplacian matrix
laplacian = ( diag(n) - diag(degrees^(-1/2)) %*% D %*%
diag(degrees^(-1/2)) )
## Compute eigenvectors
eigenvectors = eigen(laplacian, symmetric = TRUE)
n = nrow(laplacian)
eigenvectors = eigenvectors$vector[, (n - 2):(n - 1)]
set.seed(1748)
## Run Kmeans on eigenvectors
sc = kmeans(eigenvectors, 3)

## Pull clustering results
sc_results = cbind(insurance.scale, cluster =
as.factor(sc$cluster))
head(sc_results)
#kmeans plot
kmeans_plot = ggplot(data = kmeans_results,aes(x=Capital,
y=Age_Actuarial, color = cluster)) +
  geom_point() +
  scale_color_manual(values = c('1' = "violetred2",
                                '2' = "darkorchid2",
                                '3' = "darkolivegreen2")) +
  ggtitle("K-Means Clustering") +
  theme(panel.grid.major = element_blank(),
panel.grid.minor = element_blank(),
        panel.background = element_blank(), axis.line =
element_line(colour = "black"),
```

```
    plot.title = element_text(hjust = 0.5),
legend.position="bottom")
#Spectral Clustering plot
sc_plot = ggplot(data = sc_results, aes(x=Capital,
y=Age_Actuarial, color = cluster)) +
  geom_point() +
  scale_color_manual(values = c('1' = "violetred2",
                                '2' = "darkorchid2",
                                '3' = "darkolivegreen2")) +
  ggtitle("Spectral Clustering") +
  theme(panel.grid.major = element_blank(),
panel.grid.minor = element_blank(),
        panel.background = element_blank(), axis.line =
element_line(colour = "black"),
        plot.title = element_text(hjust = 0.5),
legend.position="bottom")
install.packages("biclust")
install.packages("MASS")
library(biclust)
library(MASS)
res1 <- biclust(as.matrix(insurance.scale[,1:2]),
method=BCCC(), delta=1.5, alpha=1, number=3)
```

9.8. Anexos scripts en Python elaborados para el trabajo

9.3.1. Análisis descriptivo

```
#iMPORTO LIBRERÍAS
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

#Importación datos
ruta_archivo_excel = 'G:\Esta_es_mi_carpeta2\Portlofio
vida.xlsx'
ruta_archivo_excel_2000='G:\Esta_es_mi_carpeta2\Excels_defi
nitivos\Portafolio_vida_sin_filtro_2000.xlsx'
ruta_archivo_excel_2001 =
'G:\Esta_es_mi_carpeta2\Excels_definitivos\Portafolio_vida_
sin_filtro_2001.xlsx'
ruta_archivo_excel_2002 =
'G:\Esta_es_mi_carpeta2\Excels_definitivos\Portafolio_vida_
sin_filtro_2002.xlsx'
ruta_archivo_excel_2003 =
'G:\Esta_es_mi_carpeta2\Excels_definitivos\Portafolio_vida_
sin_filtro_2003.xlsx'
ruta_archivo_excel_2004 =
'G:\Esta_es_mi_carpeta2\Excels_definitivos\Portafolio_vida_
```

```
sin_filtro_2004.xlsx'  
ruta_archivo_excel_2005 =  
'G:\Esta_es_mi_carpeta2\Excels_definitivos\Portafolio_vida_  
sin_filtro_2005.xlsx'  
ruta_archivo_excel_2006 =  
'G:\Esta_es_mi_carpeta2\Excels_definitivos\Portafolio_vida_  
sin_filtro_2006.xlsx'  
ruta_archivo_excel_2007 =  
'G:\Esta_es_mi_carpeta2\Excels_definitivos\Portafolio_vida_  
sin_filtro_2007.xlsx'  
ruta_archivo_excel_2008 =  
'G:\Esta_es_mi_carpeta2\Excels_definitivos\Portafolio_vida_  
sin_filtro_2008.xlsx'  
ruta_archivo_excel_2009 =  
'G:\Esta_es_mi_carpeta2\Excels_definitivos\Portafolio_vida_  
sin_filtro_2009.xlsx'  
  
datos_excel = pd.read_excel(ruta_archivo_excel)  
datos_excel_2001=pd.read_excel(ruta_archivo_excel_2001)  
datos_excel_2000=pd.read_excel(ruta_archivo_excel_2000)  
datos_excel_2002=pd.read_excel(ruta_archivo_excel_2002)  
datos_excel_2003=pd.read_excel(ruta_archivo_excel_2003)  
datos_excel_2004=pd.read_excel(ruta_archivo_excel_2004)  
datos_excel_2005=pd.read_excel(ruta_archivo_excel_2005)  
datos_excel_2006=pd.read_excel(ruta_archivo_excel_2006)  
datos_excel_2007=pd.read_excel(ruta_archivo_excel_2007)  
datos_excel_2008=pd.read_excel(ruta_archivo_excel_2008)  
datos_excel_2009=pd.read_excel(ruta_archivo_excel_2009)  
pd.set_option('display.max_columns', None)  
pd.set_option('display.max_rows', None)  
datos_excel = datos_excel.iloc[:, 1:]  
datos_excel_2000 = datos_excel_2000.iloc[:, 1:]  
datos_excel_2001 = datos_excel_2001.iloc[:, 1:]  
datos_excel_2002=datos_excel_2002.iloc[:,1:]  
datos_excel_2003=datos_excel_2003.iloc[:,1:]  
datos_excel_2004=datos_excel_2004.iloc[:,1:]  
datos_excel_2005=datos_excel_2005.iloc[:,1:]  
datos_excel_2006=datos_excel_2006.iloc[:,1:]  
datos_excel_2007=datos_excel_2007.iloc[:,1:]  
datos_excel_2008=datos_excel_2008.iloc[:,1:]  
datos_excel_2009=datos_excel_2009.iloc[:,1:]  
print(datos_excel.head())  
print(datos_excel_2000.describe())  
print(datos_excel_2001.describe())  
print(datos_excel_2002.describe())  
print(datos_excel_2003.describe())  
print(datos_excel_2004.describe())  
print(datos_excel['km_nocturnos'].describe())  
print(datos_excel_2000.shape)  
  
datos_excel.info() #indica num de datos, media, std, min,
```

```

max, percentiles
#int-> variable numerica
#float64-> numero decimal con 64 bits
#category-> variables ctegóricas o string
import scipy.stats
pearson_corr_2000=scipy.stats.pearsonr(datos_excel_2000['Ca
pital'],datos_excel_2000['Age_Actuarial'])
kendall_corr_2000=
scipy.stats.kendalltau(datos_excel_2000['Capital'],datos_ex
cel_2000['Age_Actuarial'])
spearman_corr_2000=scipy.stats.spearmanr(datos_excel_2000['
Capital'],datos_excel_2000['Age_Actuarial'])
datos_excel
datos_excel_2000['Gender']=datos_excel_2000['Gender'].astyp
e('category')#cambio estos datos categorías
datos_excel_2000['Birth_Date']=datos_excel_2000['Birth_Date
'].astype('category')#cambio estos datos a categorías
datos_excel_2000['Effecitive_Date']=datos_excel_2000['Effec
itive_Date'].astype('category')
datos_excel_2000['Renewal_Date']=datos_excel_2000['Renewal_
Date'].astype('category')
datos_excel_2000['Birthday']=datos_excel_2000['Birthday'].a
stype('category')
datos_excel_2000['r']=datos_excel_2000['r'].astype('categor
y')
datos_excel_2000['s']=datos_excel_2000['s'].astype('categor
y')
datos_excel_2000['Month']=datos_excel_2000['Month'].astype(
'category')
print(datos_excel_2000.dtypes)#me indica cada dato que tipo
es.

#####FREC.
ABSOLUTAS#####
#####
frecuencia_absoluta_gender =
datos_excel_2000['Gender'].value_counts()
print(frecuencia_absoluta_gender)
frecuencia_absoluta_gender.plot(kind='bar')
plt.xlabel('Género')
plt.ylabel('Frecuencia Absoluta')
plt.title('Frecuencia Absoluta por Género')
plt.show()
len(datos_excel)
frecuencia_absoluta_gender/len(datos_excel_2000)

frecuencia_absoluta_r =
datos_excel_2000['r'].value_counts()
print(frecuencia_absoluta_r)
frecuencia_absoluta_r/len(datos_excel_2000)

```



```
frecuencia_absoluta_s =
datos_excel_2000['s'].value_counts()
print(frecuencia_absoluta_s)
frecuencia_absoluta_s/len(datos_excel_2000)

frecuencia_absoluta_Month =
datos_excel_2000['Month'].value_counts()
print(frecuencia_absoluta_Month)
frecuencia_absoluta_Month/len(datos_excel_2000)

##### TABLAS DE CONTINGENCIA
#####
tabla_contingencia_genero_edad =
pd.crosstab(datos_excel['Gender'],
datos_excel['Age_Actuarial'])
print("Tabla de Contingencia entre Género y Edad
Actuarial:")
print(tabla_contingencia_genero_edad)

tabla_contingencia_genero_estacion
=pd.crosstab(datos_excel['Gender'],datos_excel['s'])
print("\nTabla de Contingencia entre Género y Estación:")
print(tabla_contingencia_genero_estacion)

tabla_contingencia_estacion_mes =
pd.crosstab(datos_excel['s'], datos_excel['Month'])
print("\nTabla de Contingencia entre Estación y Mes de
Renovación:")
print(tabla_contingencia_estacion_mes)

tabla_contingencia_trimestre_capital =
pd.crosstab(datos_excel['r'], datos_excel['Capital'])
print("\nTabla de Contingencia entre Trimestre de
Envejecimiento y Capital:")
print(tabla_contingencia_trimestre_capital)

tabla_contingencia_mes_cumple =
pd.crosstab(datos_excel['Month'],
datos_excel['Birthday'].dt.month)
print("\nTabla de Contingencia entre Mes de Renovación y
Mes de Cumpleaños:")
print(tabla_contingencia_mes_cumple)

#####
CORRELACIONES #####
correlacion_pearson=datos.corr(method='pearson')
correlacion_kendall=datos.corr(method='kendall')
correlacion_spearman=datos.corr(method='spearman')
```

9.3.2. Técnicas Clúster

```
import numpy as np
import pandas as pd
from sklearn.datasets import make_blobs
from sklearn.ensemble import RandomForestClassifier

import matplotlib.pyplot as plt
import seaborn as sns
ruta_archivo_excel =
'G:/Esta_es_mi_carpeta2/Excels_definitivos/Portafolio_vida_
sin_filtro_2000.xlsx'
ruta_archivo_excel = 'G:/Esta_es_mi_carpeta2/Portlofio
vida.xlsx'
datos_excel = pd.read_excel(ruta_archivo_excel)
pd.set_option('display.max_columns', None)
print(datos_excel.columns)
datos_excel=datos_excel.iloc[:,[4,8]]
print(datos_excel.head())
print(datos_excel.tail())
print(datos_excel.describe())
correlacion_pearson =
datos_excel['Capital'].corr(datos_excel['Age_Actuarial'],
method='pearson')
print(correlacion_pearson)
p_valor_pearson =
datos_excel['Capital'].corr(datos_excel['Age_Actuarial'],
method='pearson')
print(p_valor_pearson)
from scipy.stats import pearsonr
correlation_coefficient,
p_value=pearsonr(datos_excel['Capital'],datos_excel['Age_Actuarial'])
from scipy.stats import kendalltau
correlation_coefficient, p_value =
kendalltau(datos_excel['Capital'],datos_excel['Age_Actuarial'])
from scipy.stats import spearmanr
correlation_coefficient, p_value =
spearmanr(datos_excel['Capital'],datos_excel['Age_Actuarial'])
datos_excel.shape
fig, ax=plt.subplots(1,1,figsize=(6,3.84))
ax.scatter(
    x=datos_excel.iloc[:,1],
    y=datos_excel.iloc[:,0],
    c='white',
    marker='o',
    edgecolor='black'
)
ax.set_title('Datos año 2000')
```

```

ax.set_xlabel('Edad Actuarial')
ax.set_ylabel('Capital')
plt.show()
#Eso para cuando no queremos estandarizarlos
# scaled_datos=datos_excel
from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
scaled_datos = scaler.fit_transform(datos_excel)
from sklearn.cluster import KMeans
distortions=[]
for i in range (1,11):
    km=KMeans(n_clusters=i,
              init='random', n_init=25, random_state=123)
    km.fit(scaled_datos)
    distortions.append(km.inertia_)
plt.plot(range(1,11),distortions,marker='o')
plt.xlabel('Number of clusters')
plt.ylabel('WCSS')
plt.title('Método del Codo')
plt.show()
scaled_datos=datos_excel
from sklearn.cluster import KMeans
modelo_kmeans=KMeans(n_clusters=3,n_init=25,random_state=123)
y_predict = modelo_kmeans.fit_predict(X=scaled_datos)
y_predict = modelo_kmeans.fit_predict(X=datos_excel)
#Aquí hago una lista con sublistas de cada asegurado
formato
#[[capital, edad, cluster],[capital,edad, cluster],...]
datos = [[capital, edad] for capital, edad in
zip(datos_excel['Capital'], datos_excel['Age_Actuarial'])]
for i in range(len(datos)):
    datos[i].append(y_predict[i])
print(datos)
scaled_datos
for i in range(len(datos)):
    datos[i].append(list(scaled_datos[i]))
print(datos)
#Coordenadas de cada clúster
modelo_kmeans.cluster_centers_
cluster_centers = modelo_kmeans.cluster_centers_.tolist()
print(cluster_centers)

#media de las distancias de los datos al centroide que
pertanecen:
from scipy.spatial.distance import euclidean
distancias_centroide1=[]
distancias_centroide2=[]
distancias_centroide3=[]
for dato in datos:
    coordenadas_dato=dato[3]

```

```
cluster_dato=dato[2]
coordenadas_centroide = cluster_centers[cluster_dato]
distancia = euclidean(coordenadas_dato,
coordenadas_centroide)
if cluster_dato == 0:
    distancias_centroide1.append(distancia)
elif cluster_dato == 1:
    distancias_centroide2.append(distancia)
elif cluster_dato == 2:
    distancias_centroide3.append(distancia)
print("Distancias para el centroide 1:",
distancias_centroide1)
print("Distancias para el centroide 2:",
distancias_centroide2)
print("Distancias para el centroide 3:",
distancias_centroide3)
datos
#Finalmente se hacen las medias
distancia_media_centroide1 = sum(distancias_centroide1) /
len(distancias_centroide1) if distancias_centroide1 else 0
distancia_media_centroide2 = sum(distancias_centroide2) /
len(distancias_centroide2) if distancias_centroide2 else 0
distancia_media_centroide3 = sum(distancias_centroide3) /
len(distancias_centroide3) if distancias_centroide3 else 0

print("Distancia media para el centroide 1:",
distancia_media_centroide1)
print("Distancia media para el centroide 2:",
distancia_media_centroide2)
print("Distancia media para el centroide 3:",
distancia_media_centroide3)

#Cálculo del rango
rango_centroide1 = max(distancias_centroide1) -
min(distancias_centroide1) if distancias_centroide1 else 0
rango_centroide2 = max(distancias_centroide2) -
min(distancias_centroide2) if distancias_centroide2 else 0
rango_centroide3 = max(distancias_centroide3) -
min(distancias_centroide3) if distancias_centroide3 else 0
print(rango_centroide1)
print(rango_centroide2)
print(rango_centroide3)

#Cálculo de percentiles
cuartil1_centroide1 = np.percentile(distancias_centroide1,
25) if distancias_centroide1 else 0
cuartil3_centroide1 = np.percentile(distancias_centroide1,
75) if distancias_centroide1 else 0
cuartil1_centroide2 = np.percentile(distancias_centroide2,
25) if distancias_centroide2 else 0
cuartil3_centroide2 = np.percentile(distancias_centroide2,
```

```
75) if distancias_centroide2 else 0
cuartil1_centroide3 = np.percentile(distancias_centroide3,
25) if distancias_centroide3 else 0
cuartil3_centroide3 = np.percentile(distancias_centroide3,
75) if distancias_centroide3 else 0

#Cálculo de la Skewness (Asimetría)
from scipy.stats import skew
skewness_centroide1 = skew(distancias_centroide1) if
distancias_centroide1 else 0
skewness_centroide2 = skew(distancias_centroide2) if
distancias_centroide2 else 0
skewness_centroide3 = skew(distancias_centroide3) if
distancias_centroide3 else 0

#Cálculo de la curtosis
from scipy.stats import kurtosis
kurtosis_centroide1 = kurtosis(distancias_centroide1) if
distancias_centroide1 else 0
kurtosis_centroide2 = kurtosis(distancias_centroide2) if
distancias_centroide2 else 0
kurtosis_centroide3 = kurtosis(distancias_centroide3) if
distancias_centroide3 else 0

#cálculo de la mediana
distancias_centroide1_ordenadas=sorted(distancias_centroide
1)
distancias_centroide2_ordenadas=sorted(distancias_centroide
2)
distancias_centroide3_ordenadas=sorted(distancias_centroide
3)
mediana_centroide1=np.median(distancias_centroide1)
mediana_centroide2=np.median(distancias_centroide2)
mediana_centroide3=np.median(distancias_centroide3)

#Cálculo de la varianza
varianza_centroide1=np.var(distancias_centroide1)
varianza_centroide2=np.var(distancias_centroide2)
varianza_centroide3=np.var(distancias_centroide3)
des_est_centroide1=np.std(distancias_centroide1)
des_est_centroide2=np.std(distancias_centroide2)
des_est_centroide3=np.std(distancias_centroide3)

#####
datos_cluster1_edad = []
datos_cluster2_edad = []
datos_cluster3_edad = []
datos_cluster1_capital = []
datos_cluster2_capital = []
datos_cluster3_capital = []
```

```
# Separar los datos por cluster
for dato in datos:
    if dato[2] == 0:
        datos_cluster1_edad.append(dato[1])
        datos_cluster1_capital.append(dato[0])
    elif dato[2] == 1:
        datos_cluster2_edad.append(dato[1])
        datos_cluster2_capital.append(dato[0])
    elif dato[2] == 2:
        datos_cluster3_edad.append(dato[1])
        datos_cluster3_capital.append(dato[0])

# Calcular estadísticos para edades y capitales de cada
cluster
#Clúster 1
#Edad
media_cluster1_edad = np.mean(datos_cluster1_edad)
varianza_cluster1_edad = np.var(datos_cluster1_edad)
np.std(datos_cluster1_edad)
mediana_cluster1_edad = np.median(datos_cluster1_edad)
cuartil25_cluster1_edad =
np.percentile(datos_cluster1_edad, 25)
cuartil75_cluster1_edad =
np.percentile(datos_cluster1_edad, 75)
maximo_cluster1_edad = np.max(datos_cluster1_edad)
minimo_cluster1_edad = np.min(datos_cluster1_edad)
from scipy.stats import skew
asimetria_cluster1_edad = skew(datos_cluster1_edad)
from scipy.stats import kurtosis
curtosis_cluster1_edad = kurtosis(datos_cluster1_edad)
#Capital
media_cluster1_capital = np.mean(datos_cluster1_capital)
varianza_cluster1_capital = np.var(datos_cluster1_capital)
np.std(varianza_cluster1_capital)
mediana_cluster1_capital =
np.median(datos_cluster1_capital)
cuartil25_cluster1_capital =
np.percentile(datos_cluster1_capital, 25)
cuartil75_cluster1_capital =
np.percentile(datos_cluster1_capital, 75)
maximo_cluster1_capital = np.max(datos_cluster1_capital)
minimo_cluster1_capital = np.min(datos_cluster1_capital)
asimetria_cluster1_capital = skew(datos_cluster1_capital)
curtosis_cluster1_capital =
kurtosis(datos_cluster1_capital)
#Segundo clúster
#Edad
media_cluster2_edad = np.mean(datos_cluster2_edad)
varianza_cluster2_edad = np.var(datos_cluster2_edad)
mediana_cluster2_edad = np.median(datos_cluster2_edad)
cuartil25_cluster2_edad =
```

```
np.percentile(datos_cluster2_edad, 25)
cuartil75_cluster2_edad =
np.percentile(datos_cluster2_edad, 75)
maximo_cluster2_edad = np.max(datos_cluster2_edad)
minimo_cluster2_edad = np.min(datos_cluster2_edad)
asimetria_cluster2_edad = skew(datos_cluster2_edad)
curtosis_cluster2_edad = kurtosis(datos_cluster2_edad)
#Capital
media_cluster2_capital = np.mean(datos_cluster2_capital)
varianza_cluster2_capital = np.var(datos_cluster2_capital)
mediana_cluster2_capital =
np.median(datos_cluster2_capital)
cuartil25_cluster2_capital =
np.percentile(datos_cluster2_capital, 25)
cuartil75_cluster2_capital =
np.percentile(datos_cluster2_capital, 75)
maximo_cluster2_capital = np.max(datos_cluster2_capital)
minimo_cluster2_capital = np.min(datos_cluster2_capital)
asimetria_cluster2_capital = skew(datos_cluster2_capital)
curtosis_cluster2_capital =
kurtosis(datos_cluster2_capital)
#Cluster 3
#Edad
media_cluster3_edad = np.mean(datos_cluster3_edad)
varianza_cluster3_edad = np.var(datos_cluster3_edad)
mediana_cluster3_edad = np.median(datos_cluster3_edad)
cuartil25_cluster3_edad =
np.percentile(datos_cluster3_edad, 25)
cuartil75_cluster3_edad =
np.percentile(datos_cluster3_edad, 75)
maximo_cluster3_edad = np.max(datos_cluster3_edad)
minimo_cluster3_edad = np.min(datos_cluster3_edad)
asimetria_cluster3_edad = skew(datos_cluster3_edad)
curtosis_cluster3_edad = kurtosis(datos_cluster3_edad)
#Capital
media_cluster3_capital = np.mean(datos_cluster3_capital)
varianza_cluster3_capital = np.var(datos_cluster3_capital)
mediana_cluster3_capital =
np.median(datos_cluster3_capital)
cuartil25_cluster3_capital =
np.percentile(datos_cluster3_capital, 25)
cuartil75_cluster3_capital =
np.percentile(datos_cluster3_capital, 75)
maximo_cluster3_capital = np.max(datos_cluster3_capital)
minimo_cluster3_capital = np.min(datos_cluster3_capital)
asimetria_cluster3_capital = skew(datos_cluster3_capital)
curtosis_cluster3_capital =
kurtosis(datos_cluster3_capital)
#####
#Calculo el número de datos dentro cada clúster
asignaciones=modelo_kmeans.labels_
```

```
asignaciones
conteo_por_cluster = np.bincount(asignaciones)
conteo_por_cluster

#Distancia máxima entre dos puntos dentro de un mismo
clúster
cluster_id = 0
puntos_cluster = scaled_datos[asignaciones == cluster_id]
from scipy.spatial.distance import pdist,squareform
distancias = pdist(puntos_cluster)
distancias_cuadrada = squareform(distancias)
distancia_maxima = np.max(distancias_cuadrada)
print("Distancia máxima dentro del clúster:",
distancia_maxima)

cluster_id = 1
puntos_cluster = scaled_datos[asignaciones == cluster_id]
from scipy.spatial.distance import pdist,squareform
distancias = pdist(puntos_cluster)
distancias_cuadrada = squareform(distancias)
distancia_maxima = np.max(distancias_cuadrada)
print("Distancia máxima dentro del clúster:",
distancia_maxima)

cluster_id = 2
puntos_cluster = scaled_datos[asignaciones == cluster_id]
from scipy.spatial.distance import pdist,squareform
distancias = pdist(puntos_cluster)
distancias_cuadrada = squareform(distancias)
distancia_maxima = np.max(distancias_cuadrada)
print("Distancia máxima dentro del clúster:",
distancia_maxima)

#Distancia promedio entre todos los puntos dentro de un
clúster
cluster_id = 0
puntos_cluster = scaled_datos[asignaciones == cluster_id]
distancias = pdist(puntos_cluster)
distancia_promedio = np.mean(distancias)
distancia_mediana=np.median(distancias)
print("Distancia promedio dentro del clúster:",
distancia_promedio)
print("Distancia mediana dentro del clúster:",
distancia_mediana)

cluster_id = 1
puntos_cluster = scaled_datos[asignaciones == cluster_id]
distancias = pdist(puntos_cluster)
distancia_promedio = np.mean(distancias)
distancia_mediana=np.median(distancias)
print("Distancia promedio dentro del clúster:",
```



```
distancia_promedio)
print("Distancia mediana dentro del clúster:",
distancia_mediana)

cluster_id = 2
puntos_cluster = scaled_datos[asignaciones == cluster_id]
distancias = pdist(puntos_cluster)
distancia_promedio = np.mean(distancias)
distancia_mediana=np.median(distancias)
print("Distancia promedio dentro del clúster:",
distancia_promedio)
print("Distancia mediana dentro del clúster:",
distancia_mediana)

#Separación entre clústeres
centroides = modelo_kmeans.cluster_centers_
distancias_centroides = np.linalg.norm(centroides[:,
np.newaxis] - centroides, axis=-1)
separacion_entre_clusteres =
np.min(distancias_centroides[distancias_centroides > 0])
print("Separación entre clústeres:",
separacion_entre_clusteres)
separacion_maxima_entre_clusteres =
np.max(distancias_centroides[distancias_centroides > 0])
print("Separación máxima entre clústeres:",
separacion_maxima_entre_clusteres)

#Distancia promedio entre los puntos de diferentes
clústeres
from scipy.spatial.distance import cdist
distancias_promedio_entre_clusteres = 0
total_pares = 0
for i in range(len(centroides)):
    for j in range(i+1, len(centroides)):
        puntos_cluster_1 = scaled_datos[asignaciones == i]
        puntos_cluster_2 = scaled_datos[asignaciones == j]
        distancias = cdist(puntos_cluster_1,
puntos_cluster_2)
        distancias_promedio_entre_clusteres +=
np.mean(distancias)
        total_pares += 1
distancia_promedio_entre_clusteres =
distancias_promedio_entre_clusteres / total_pares
print("Distancia promedio entre los puntos de diferentes
clústeres:", distancia_promedio_entre_clusteres)

#Distancia promedio dentro de los clústeres
distancia_promedio_dentro_clusteres = 0
total_clusteres = len(centroides)
for i in range(total_clusteres):
    puntos_cluster = scaled_datos[asignaciones == i]
```

```

    distancias = pdist(puntos_cluster)
    distancia_promedio_dentro_clusteres +=
np.mean(distancias)
distancia_promedio_dentro_clusteres =
distancia_promedio_dentro_clusteres / total_clusteres
print("Distancia promedio dentro de los clústeres:",
distancia_promedio_dentro_clusteres)

plt.scatter(scaled_datos[:, 0], scaled_datos[:, 1],
c=y_predict, cmap='viridis')
plt.show()
#Gráfico alternativo k means
#Es para cuando no escalirizo los datos
cluster_0 = scaled_datos[y_predict == 0]
cluster_1 = scaled_datos[y_predict == 1]
cluster_2 = scaled_datos[y_predict == 2]

plt.scatter(cluster_0.iloc[:, 0], cluster_0.iloc[:, 1],
s=50, c='lightgreen', marker='s', edgecolor='black',
label='cluster 1')
plt.scatter(cluster_1.iloc[:, 0], cluster_1.iloc[:, 1],
s=50, c='orange', marker='o', edgecolor='black',
label='cluster 2')
plt.scatter(cluster_2.iloc[:, 0], cluster_2.iloc[:, 1],
s=50, c='lightblue', marker='v', edgecolor='black',
label='cluster 3')

# Para los centroides, asumiendo que
`modelo_kmeans.cluster_centers_` es un array de NumPy
plt.scatter(modelo_kmeans.cluster_centers_[:, 0],
modelo_kmeans.cluster_centers_[:, 1], s=250, marker='*',
c='red', edgecolor='black', label='centroids')

plt.title('Clustering con K-means y Centroides')
plt.xlabel('Capital en Euros')
plt.ylabel('Edad')
plt.legend(scatterpoints=1)
plt.grid()
plt.show()

#Gráfico KMeans
plt.scatter(scaled_datos[y_predict==0,0],
scaled_datos[y_predict==0,1],
s=50, c='lightgreen',
marker='s',edgecolor='black',
label='cluster 1')
plt.scatter(scaled_datos[y_predict==1,0],
scaled_datos[y_predict==1,1],
s=50, c='orange',
marker='o',edgecolor='black',
label='cluster 2')

```

```

plt.scatter(scaled_datos[y_predict==2,0],
            scaled_datos[y_predict==2,1],
            s=50, c='lightblue',
            marker='v',edgecolor='black',
            label='cluster 3')
plt.scatter(modelo_kmeans.cluster_centers_[:,0],
            modelo_kmeans.cluster_centers_[:,1],
            s=250, marker='*',
            c='red', edgecolor='black',
            label='centroids')
plt.title('Clustering con K-means y Centroides')
plt.xlabel('Capital en Euros')
plt.ylabel('Edad')
plt.legend(scatterpoints=1)
plt.grid()
plt.show()

#####K-
medoids#####
from sklearn_extra.cluster import KMedoids

scaled_datos
kmedoids=KMedoids(n_clusters=3,
random_state=0).fit(scaled_datos)
kmedoids.labels_
kmedoids.cluster_centers_
kmedoids.inertia_
y_predict = kmedoids.predict(X=scaled_datos)
plt.scatter(scaled_datos[:,0],scaled_datos[:,1])

#Aquí hago una lista con sublistas de cada asegurado
formato
#[[capital, edad, cluster],[capital,edad, cluster],...]
datos_kmedoids = [[capital, edad] for capital, edad in
zip(datos_excel['Capital'], datos_excel['Age_Actuarial'])]
for i in range(len(datos_kmedoids)):
    datos_kmedoids[i].append(y_predict[i])
print(datos_kmedoids)
scaled_datos
for i in range(len(datos_kmedoids)):
    datos_kmedoids[i].append(list(scaled_datos[i]))
print(datos_kmedoids)

cluster_centers = kmedoids.cluster_centers_.tolist()
print(cluster_centers)

from scipy.spatial.distance import euclidean
# Media de las distancias de los datos al centroide que
pertenecen:
distancias_centroide1 = []
distancias_centroide2 = []

```

```
distancias_centroide3 = []

for dato in datos_kmedoids:
    coordenadas_dato = dato[3]
    cluster_dato = dato[2]
    coordenadas_centroide = cluster_centers[cluster_dato]
    distancia = euclidean(coordenadas_dato,
    coordenadas_centroide)
    if cluster_dato == 0:
        distancias_centroide1.append(distancia)
    elif cluster_dato == 1:
        distancias_centroide2.append(distancia)
    elif cluster_dato == 2:
        distancias_centroide3.append(distancia)

print("Distancias para el centroide 1:",
distancias_centroide1)
print("Distancias para el centroide 2:",
distancias_centroide2)
print("Distancias para el centroide 3:",
distancias_centroide3)

# Finalmente se calculan las medias de las distancias
distancia_media_centroide1 = sum(distancias_centroide1) /
len(distancias_centroide1) if distancias_centroide1 else 0
distancia_media_centroide2 = sum(distancias_centroide2) /
len(distancias_centroide2) if distancias_centroide2 else 0
distancia_media_centroide3 = sum(distancias_centroide3) /
len(distancias_centroide3) if distancias_centroide3 else 0

print("Distancia media para el centroide 1:",
distancia_media_centroide1)
print("Distancia media para el centroide 2:",
distancia_media_centroide2)
print("Distancia media para el centroide 3:",
distancia_media_centroide3)

# Cálculo del rango
rango_centroide1 = max(distancias_centroide1) -
min(distancias_centroide1) if distancias_centroide1 else 0
rango_centroide2 = max(distancias_centroide2) -
min(distancias_centroide2) if distancias_centroide2 else 0
rango_centroide3 = max(distancias_centroide3) -
min(distancias_centroide3) if distancias_centroide3 else 0
print("Rango para el centroide 1:", rango_centroide1)
print("Rango para el centroide 2:", rango_centroide2)
print("Rango para el centroide 3:", rango_centroide3)

# Cálculo de percentiles
cuartil1_centroide1 = np.percentile(distancias_centroide1,
25) if distancias_centroide1 else 0
```

```
cuartil3_centroide1 = np.percentile(distancias_centroide1,
75) if distancias_centroide1 else 0
cuartil1_centroide2 = np.percentile(distancias_centroide2,
25) if distancias_centroide2 else 0
cuartil3_centroide2 = np.percentile(distancias_centroide2,
75) if distancias_centroide2 else 0
cuartil1_centroide3 = np.percentile(distancias_centroide3,
25) if distancias_centroide3 else 0
cuartil3_centroide3 = np.percentile(distancias_centroide3,
75) if distancias_centroide3 else 0
print("Cuartil 1 para el centroide 1:",
cuartil1_centroide1)
print("Cuartil 3 para el centroide 1:",
cuartil3_centroide1)
print("Cuartil 1 para el centroide 2:",
cuartil1_centroide2)
print("Cuartil 3 para el centroide 2:",
cuartil3_centroide2)
print("Cuartil 1 para el centroide 3:",
cuartil1_centroide3)
print("Cuartil 3 para el centroide 3:",
cuartil3_centroide3)

# Cálculo de la Skewness (Asimetría)
from scipy.stats import skew
skewness_centroide1 = skew(distancias_centroide1) if
distancias_centroide1 else 0
skewness_centroide2 = skew(distancias_centroide2) if
distancias_centroide2 else 0
skewness_centroide3 = skew(distancias_centroide3) if
distancias_centroide3 else 0
print("Skewness para el centroide 1:", skewness_centroide1)
print("Skewness para el centroide 2:", skewness_centroide2)
print("Skewness para el centroide 3:", skewness_centroide3)

# Cálculo de la curtosis
from scipy.stats import kurtosis
kurtosis_centroide1 = kurtosis(distancias_centroide1) if
distancias_centroide1 else 0
kurtosis_centroide2 = kurtosis(distancias_centroide2) if
distancias_centroide2 else 0
kurtosis_centroide3 = kurtosis(distancias_centroide3) if
distancias_centroide3 else 0
print("Curtosis para el centroide 1:", kurtosis_centroide1)
print("Curtosis para el centroide 2:", kurtosis_centroide2)
print("Curtosis para el centroide 3:", kurtosis_centroide3)

# Cálculo de la mediana
mediana_centroide1 =
np.median(sorted(distancias_centroide1)) if
```

```
distancias_centroide1 else 0
mediana_centroide2 =
np.median(sorted(distancias_centroide2)) if
distancias_centroide2 else 0
mediana_centroide3 =
np.median(sorted(distancias_centroide3)) if
distancias_centroide3 else 0
print("Mediana para el centroide 1:", mediana_centroide1)
print("Mediana para el centroide 2:", mediana_centroide2)
print("Mediana para el centroide 3:", mediana_centroide3)

# Cálculo de la varianza
varianza_centroide1 = np.var(distancias_centroide1) if
distancias_centroide1 else 0
varianza_centroide2 = np.var(distancias_centroide2) if
distancias_centroide2 else 0
varianza_centroide3 = np.var(distancias_centroide3) if
distancias_centroide3 else 0
print("Varianza para el centroide 1:", varianza_centroide1)
print("Varianza para el centroide 2:", varianza_centroide2)
print("Varianza para el centroide 3:", varianza_centroide3)

# Cálculo de la desviación estándar
des_est_centroide1 = np.std(distancias_centroide1) if
distancias_centroide1 else 0
des_est_centroide2 = np.std(distancias_centroide2) if
distancias_centroide2 else 0
des_est_centroide3 = np.std(distancias_centroide3) if
distancias_centroide3 else 0
print("Desviación estándar para el centroide 1:",
des_est_centroide1)
print("Desviación estándar para el centroide 2:",
des_est_centroide2)
print("Desviación estándar para el centroide 3:",
des_est_centroide3)

#####
datos_cluster1_edad = []
datos_cluster2_edad = []
datos_cluster3_edad = []
datos_cluster1_capital = []
datos_cluster2_capital = []
datos_cluster3_capital = []

# Separar los datos por cluster
for dato in datos_kmedoids:
    if dato[2] == 0:
        datos_cluster1_edad.append(dato[1])
        datos_cluster1_capital.append(dato[0])
    elif dato[2] == 1:
        datos_cluster2_edad.append(dato[1])
```

```
        datos_cluster2_capital.append(dato[0])
    elif dato[2] == 2:
        datos_cluster3_edad.append(dato[1])
        datos_cluster3_capital.append(dato[0])

# Calcular estadísticos para edades y capitales de cada
cluster
# Clúster 1
# Edad
media_cluster1_edad = np.mean(datos_cluster1_edad)
varianza_cluster1_edad = np.var(datos_cluster1_edad)
mediana_cluster1_edad = np.median(datos_cluster1_edad)
cuartil25_cluster1_edad =
np.percentile(datos_cluster1_edad, 25)
cuartil75_cluster1_edad =
np.percentile(datos_cluster1_edad, 75)
maximo_cluster1_edad = np.max(datos_cluster1_edad)
minimo_cluster1_edad = np.min(datos_cluster1_edad)
asimetria_cluster1_edad = skew(datos_cluster1_edad)
curtosis_cluster1_edad = kurtosis(datos_cluster1_edad)

# Capital
media_cluster1_capital = np.mean(datos_cluster1_capital)
varianza_cluster1_capital = np.var(datos_cluster1_capital)
mediana_cluster1_capital =
np.median(sorted(datos_cluster1_capital))
cuartil25_cluster1_capital =
np.percentile(datos_cluster1_capital, 25)
cuartil75_cluster1_capital =
np.percentile(datos_cluster1_capital, 75)
maximo_cluster1_capital = np.max(datos_cluster1_capital)
minimo_cluster1_capital = np.min(datos_cluster1_capital)
asimetria_cluster1_capital = skew(datos_cluster1_capital)
curtosis_cluster1_capital =
kurtosis(datos_cluster1_capital)

# Segundo clúster
# Edad
media_cluster2_edad = np.mean(datos_cluster2_edad)
varianza_cluster2_edad = np.var(datos_cluster2_edad)
mediana_cluster2_edad = np.median(datos_cluster2_edad)
cuartil25_cluster2_edad =
np.percentile(datos_cluster2_edad, 25)
cuartil75_cluster2_edad =
np.percentile(datos_cluster2_edad, 75)
maximo_cluster2_edad = np.max(datos_cluster2_edad)
minimo_cluster2_edad = np.min(datos_cluster2_edad)
asimetria_cluster2_edad = skew(datos_cluster2_edad)
curtosis_cluster2_edad = kurtosis(datos_cluster2_edad)

# Capital
```

```
media_cluster2_capital = np.mean(datos_cluster2_capital)
varianza_cluster2_capital = np.var(datos_cluster2_capital)
mediana_cluster2_capital =
np.median(datos_cluster2_capital)
cuartil25_cluster2_capital =
np.percentile(datos_cluster2_capital, 25)
cuartil75_cluster2_capital =
np.percentile(datos_cluster2_capital, 75)
maximo_cluster2_capital = np.max(datos_cluster2_capital)
minimo_cluster2_capital = np.min(datos_cluster2_capital)
asimetria_cluster2_capital = skew(datos_cluster2_capital)
curtosis_cluster2_capital =
kurtosis(datos_cluster2_capital)
```

```
# Cluster 3
```

```
# Edad
```

```
media_cluster3_edad = np.mean(datos_cluster3_edad)
varianza_cluster3_edad = np.var(datos_cluster3_edad)
mediana_cluster3_edad = np.median(datos_cluster3_edad)
cuartil25_cluster3_edad =
np.percentile(datos_cluster3_edad, 25)
cuartil75_cluster3_edad =
np.percentile(datos_cluster3_edad, 75)
maximo_cluster3_edad = np.max(datos_cluster3_edad)
minimo_cluster3_edad = np.min(datos_cluster3_edad)
asimetria_cluster3_edad = skew(datos_cluster3_edad)
curtosis_cluster3_edad = kurtosis(datos_cluster3_edad)
```

```
# Capital
```

```
media_cluster3_capital = np.mean(datos_cluster3_capital)
varianza_cluster3_capital = np.var(datos_cluster3_capital)
mediana_cluster3_capital =
np.median(datos_cluster3_capital)
cuartil25_cluster3_capital =
np.percentile(datos_cluster3_capital, 25)
cuartil75_cluster3_capital =
np.percentile(datos_cluster3_capital, 75)
maximo_cluster3_capital = np.max(datos_cluster3_capital)
minimo_cluster3_capital = np.min(datos_cluster3_capital)
asimetria_cluster3_capital = skew(datos_cluster3_capital)
curtosis_cluster3_capital =
kurtosis(datos_cluster3_capital)
```

```
#distancia máxima entre dos puntos dentro de un mismo
clúster
```

```
cluster_id = 0
puntos_cluster = scaled_datos[kmedoids.labels_ ==
cluster_id]
distancias = pdist(puntos_cluster)
distancias_cuadrada = squareform(distancias)
distancia_maxima = np.max(distancias_cuadrada)
```



```
print("Distancia máxima dentro del clúster:",
      distancia_maxima)

cluster_id = 1
puntos_cluster = scaled_datos[kmedoids.labels_ ==
                               cluster_id]
distancias = pdist(puntos_cluster)
distancias_cuadrada = squareform(distancias)
distancia_maxima = np.max(distancias_cuadrada)
print("Distancia máxima dentro del clúster:",
      distancia_maxima)

cluster_id = 2
puntos_cluster = scaled_datos[kmedoids.labels_ ==
                               cluster_id]
distancias = pdist(puntos_cluster)
distancias_cuadrada = squareform(distancias)
distancia_maxima = np.max(distancias_cuadrada)
print("Distancia máxima dentro del clúster:",
      distancia_maxima)

#distancia promedio y distancia mediana
cluster_id = 0
puntos_cluster = scaled_datos[kmedoids.labels_ ==
                               cluster_id]
distancias = pdist(puntos_cluster)
distancia_promedio = np.mean(distancias)
distancia_mediana = np.median(distancias)
print("Distancia promedio dentro del clúster:",
      distancia_promedio)
print("Distancia mediana dentro del clúster:",
      distancia_mediana)

cluster_id = 1
puntos_cluster = scaled_datos[kmedoids.labels_ ==
                               cluster_id]
distancias = pdist(puntos_cluster)
distancia_promedio = np.mean(distancias)
distancia_mediana = np.median(distancias)
print("Distancia promedio dentro del clúster:",
      distancia_promedio)
print("Distancia mediana dentro del clúster:",
      distancia_mediana)

cluster_id = 2
puntos_cluster = scaled_datos[kmedoids.labels_ ==
                               cluster_id]
distancias = pdist(puntos_cluster)
distancia_promedio = np.mean(distancias)
distancia_mediana = np.median(distancias)
print("Distancia promedio dentro del clúster:",
```

```
distancia_promedio)
print("Distancia mediana dentro del clúster:",
distancia_mediana)

#Separación entre clústeres
centroides = scaled_datos[kmedoids.medoid_indices_]
distancias_centroides = np.linalg.norm(centroides[:,
np.newaxis] - centroides, axis=-1)
separacion_entre_clusteres =
np.min(distancias_centroides[distancias_centroides > 0])
print("Separación entre clústeres:",
separacion_entre_clusteres)
separacion_maxima_entre_clusteres =
np.max(distancias_centroides[distancias_centroides > 0])
print("Separación máxima entre clústeres:",
separacion_maxima_entre_clusteres)

# Distancia promedio entre los puntos de diferentes
clústeres
distancias_promedio_entre_clusteres = 0
total_pares = 0
for i in range(len(centroides)):
    for j in range(i+1, len(centroides)):
        puntos_cluster_1 = scaled_datos[kmedoids.labels_ ==
i]
        puntos_cluster_2 = scaled_datos[kmedoids.labels_ ==
j]
        distancias = cdist(puntos_cluster_1,
puntos_cluster_2)
        distancias_promedio_entre_clusteres +=
np.mean(distancias)
        total_pares += 1
distancia_promedio_entre_clusteres =
distancias_promedio_entre_clusteres / total_pares
print("Distancia promedio entre los puntos de diferentes
clústeres:", distancia_promedio_entre_clusteres)

# Distancia promedio dentro de los clústeres
distancia_promedio_dentro_clusteres = 0
total_clusteres = len(centroides)
for i in range(total_clusteres):
    puntos_cluster = scaled_datos[kmedoids.labels_ == i]
    distancias = pdist(puntos_cluster)
    distancia_promedio_dentro_clusteres +=
np.mean(distancias)
distancia_promedio_dentro_clusteres =
distancia_promedio_dentro_clusteres / total_clusteres
print("Distancia promedio dentro de los clústeres:",
distancia_promedio_dentro_clusteres)

#Gráfico K-medoids
```

```
import numpy as np

def distancia_euclidiana(x1, x2):
    return np.sqrt(np.sum((x1 - x2) ** 2))

def k_medoids(datos, k, max_iter=100):
    # Inicialización aleatoria de los medoids
    indices_medoids = np.random.choice(len(datos), size=k,
replace=False)
    medoids = datos[indices_medoids]

    for _ in range(max_iter):
        # Asignación de puntos al medoid más cercano
        asignaciones = []
        for punto in datos:
            distancias = [distancia_euclidiana(punto,
medoid) for medoid in medoids]
            cluster_asignado = np.argmin(distancias)
            asignaciones.append(cluster_asignado)

        # Actualización de los medoids
        for i in range(k):
            cluster_i_indices =
np.where(np.array(asignaciones) == i)[0]
            distancias_cluster_i = []
            for j in cluster_i_indices:
                distancia_total =
np.sum([distancia_euclidiana(datos[j], datos[m]) for m in
cluster_i_indices])

            distancias_cluster_i.append(distancia_total)
            nuevo_medoid_index =
cluster_i_indices[np.argmin(distancias_cluster_i)]
            medoids[i] = datos[nuevo_medoid_index]

    return medoids, asignaciones

# Ejemplo de uso
datos = np.random.rand(100, 2) # Datos aleatorios
k = 3 # Número de clusters
medoids, asignaciones = k_medoids(datos, k)
medoids, asignaciones=k_medoids(scaled_datos, 3)
print("Medoids:")
print(medoids)
print("Asignaciones:")
print(asignaciones)
import matplotlib.pyplot as plt
plt.scatter(scaled_datos[y_predict==0, 0],
```

```

        scaled_datos[y_predict==0,1],
        s=50, c='lightgreen',
        marker='s',edgecolor='black',
        label='cluster 1')
plt.scatter(scaled_datos[y_predict==1,0],
            scaled_datos[y_predict==1,1],
            s=50, c='orange',
            marker='o',edgecolor='black',
            label='cluster 2')
plt.scatter(scaled_datos[y_predict==2,0],
            scaled_datos[y_predict==2,1],
            s=50, c='lightblue',
            marker='v',edgecolor='black',
            label='cluster 3')
plt.scatter(kmedoids.cluster_centers_[:,0],
            kmedoids.cluster_centers_[:,1],
            s=250, marker='*',
            c='red', edgecolor='black',
            label='centroids')
plt.title('Clustering con K-medoids y Centroides')
plt.xlabel('Edad')
plt.ylabel('Capital en Euros')
plt.legend(scatterpoints=1)
plt.grid()
plt.show()

#Evaluación
y_predict
modelo_kmeans
modelo_kmeans=KMeans(n_clusters=3,n_init=25,random_state=12
3)
modelo_kmeans.fit(X=scaled_datos)

y_predict = modelo_kmeans.predict(X=scaled_datos)
from matplotlib import cm
from sklearn.metrics import silhouette_samples
cluster_labels=np.unique(y_predict)
n_clusters=cluster_labels.shape[0]
silhouette_vals=silhouette_samples(scaled_datos,y_predict,m
etric='euclidean')
y_ax_lower, y_ax_upper=0,0
yticks=[]
for i,c in enumerate(cluster_labels):
    c_silhouette_vals=silhouette_vals[y_predict==c]
    c_silhouette_vals.sort()
    y_ax_upper += len(c_silhouette_vals)
    color=cm.jet(float(i)/n_clusters)
    plt.barh(range(y_ax_lower,y_ax_upper),
             c_silhouette_vals,
             height=1.0,
             edgecolor='none',

```

```

        color=color)
    yticks.append((y_ax_lower+y_ax_upper)/2.)
    y_ax_lower += len(c_silhouette_vals)
    silhouette_avg=np.mean(silhouette_vals)
    plt.axvline(silhouette_avg,
                color="red",
                linestyle="--")
    plt.yticks(yticks,cluster_labels+1)
    plt.ylabel('Cluster')
    plt.xlabel('Silhouette coefficient')
    plt.show

#dos centroides
from sklearn.cluster import KMeans
modelo_kmeans=KMeans(n_clusters=4,n_init=25,random_state=12
3)
modelo_kmeans.fit(X=scaled_datos)

y_predict = modelo_kmeans.predict(X=scaled_datos)
plt.scatter()

plt.scatter(scaled_datos[y_predict==0,0],
            scaled_datos[y_predict==0,1],
            s=50, c='lightgreen',
            marker='s',edgecolor='black',
            label='cluster 1')
plt.scatter(scaled_datos[y_predict==1,0],
            scaled_datos[y_predict==1,1],
            s=50, c='orange',
            marker='o',edgecolor='black',
            label='cluster 2')
plt.scatter(scaled_datos[y_predict==2,0],
            scaled_datos[y_predict==2,1],
            s=50, c='green',
            marker='v',edgecolor='black',
            label='cluster 3')
plt.scatter(scaled_datos[y_predict==3,0],
            scaled_datos[y_predict==3,1],
            s=50, c='bluelight',
            marker='b',edgecolor='black',
            label='cluster 4')
plt.scatter(modelo_kmeans.cluster_centers_[:,0],
            modelo_kmeans.cluster_centers_[:,1],
            s=250, marker='*',
            c='red', edgecolor='black',
            label='centroids')
plt.title('Clustering con K-means y Centroides')
plt.xlabel('Edad')
plt.ylabel('Capital en Euros')
plt.legend(scatterpoints=1)

```

```

plt.grid()
plt.show()

#Evaluación
y_predict
modelo_kmeans

from matplotlib import cm
from sklearn.metrics import silhouette_samples
cluster_labels=np.unique(y_predict)
n_clusters=cluster_labels.shape[0]
silhouette_vals=silhouette_samples(scaled_datos,y_predict,m
etric='euclidean')
y_ax_lower, y_ax_upper=0,0
yticks=[]
for i,c in enumerate(cluster_labels):
    c_silhouette_vals=silhouette_vals[y_predict==c]
    c_silhouette_vals.sort()
    y_ax_upper += len(c_silhouette_vals)
    color=cm.jet(float(i)/n_clusters)
    plt.barh(range(y_ax_lower,y_ax_upper),
             c_silhouette_vals,
             height=1.0,
             edgecolor='none',
             color=color)
    yticks.append((y_ax_lower+y_ax_upper)/2.)
    y_ax_lower += len(c_silhouette_vals)
silhouette_avg=np.mean(silhouette_vals)
plt.axvline(silhouette_avg,
            color="red",
            linestyle="--")
plt.yticks(yticks,cluster_labels+1)
plt.ylabel('Cluster')
plt.xlabel('Silhouette coefficient')
plt.show

#índice Silhouette
from sklearn.metrics import silhouette_score
range_n_clusters = range(2, 15)
valores_medios_silhouette = []

for n_clusters in range_n_clusters:
    modelo_kmeans = KMeans(
        n_clusters=n_clusters,
        n_init=20,
        random_state=123
    )
    cluster_labels =
modelo_kmeans.fit_predict(scaled_datos)
    silhouette_avg = silhouette_score(scaled_datos,
cluster_labels)

```

```

valores_medios_silhouette.append(silhouette_avg)

fig, ax = plt.subplots(1, 1, figsize=(6, 3.84))
ax.plot(range_n_clusters, valores_medios_silhouette,
marker='o')
ax.set_title("Evolución de media de los índices
silhouette")
ax.set_xlabel('Número clusters')
ax.set_ylabel('Media índices silhouette')
plt.show()

#detección de outliers
scaled_datos
Q1 = np.percentile(scaled_datos, 25)
Q3 = np.percentile(scaled_datos, 75)
IQR = Q3 - Q1
umbral_superior = Q3 + 4 * IQR
umbral_inferior = Q1 - 4 * IQR
outliers = scaled_datos[(scaled_datos < umbral_inferior) |
(scaled_datos > umbral_superior)]
print("Outliers:", outliers)

lista=scaled_datos.tolist()
def listar(l):
    n=0
    for elem in lista:
        if elem[0]==-0.23100331 or elem[1]==-0.23100331:
            return elem
    return None

m=listar(lista)
print(m)

#DBSCAN
from sklearn.cluster import DBSCAN
modelo_dbscan = DBSCAN(eps=0.5, min_samples=4)

# Ajustar el modelo a tus datos escalados
y_predict = modelo_dbscan.fit_predict(X=scaled_datos)

plt.scatter(scaled_datos[y_predict == 0, 0],
scaled_datos[y_predict == 0, 1], s=50, c='lightgreen',
marker='s', edgecolor='black', label='Cluster 1')
plt.scatter(scaled_datos[y_predict == 1, 0],
scaled_datos[y_predict == 1, 1], s=50, c='orange',
marker='o', edgecolor='black', label='Cluster 2')
plt.scatter(scaled_datos[y_predict == -1, 0],
scaled_datos[y_predict == -1, 1], s=50, c='gray',
marker='x', edgecolor='black', label='Outliers')
plt.legend(scatterpoints=1)
plt.title('Resultado de DBSCAN')

```

```
plt.xlabel('Eje Edad asegurado')
plt.ylabel('Eje Capital')
plt.grid()
plt.show()

modelo_dbscan.fit(X=scaled_datos)
labels = modelo_dbscan.labels_

fig, ax = plt.subplots(1, 1, figsize=(4.5, 4.5))

ax.scatter(
    x = scaled_datos[:, 0],
    y = scaled_datos[:, 1],
    c = labels,
    marker = 'o',
    edgecolor = 'black'
)

# Los outliers se identifican con el label -1
ax.scatter(
    x = scaled_datos[labels == -1, 0],
    y = scaled_datos[labels == -1, 1],
    c = 'red',
    marker = 'o',
    edgecolor = 'black',
    label = 'outliers'
)

ax.legend()
ax.set_title('Clusterings generados por DBSCAN')
plt.show()

from sklearn.datasets import make_moons
X, y = make_moons(n_samples=1960, noise=0.05, random_state=0)
from sklearn.cluster import DBSCAN
db=DBSCAN(eps=0.5, min_samples=4, metric='euclidean')
y_db=db.fit_predict(scaled_datos)
plt.scatter(scaled_datos[y_db==0, 0],
            scaled_datos[y_db==0, 1],
            c='lightblue',
            edgecolor='black',
            marker='o',
            s=40,
            label='cluster 1')
plt.scatter(scaled_datos[y_db==1, 0],
            scaled_datos[y_db==1, 1],
            c='red', edgecolor='black',
            marker='s',
            s=40,
```



```

        label='cluster 2')
plt.legend()
plt.show()

##Convex Cluster Hulls
import numpy as np
import matplotlib.pyplot as plt
from sklearn.cluster import DBSCAN
from sklearn.preprocessing import StandardScaler
from scipy.spatial import ConvexHull
modelo_dbscan = DBSCAN(eps=0.5, min_samples=4)
y_predict = modelo_dbscan.fit_predict(X=scaled_datos)
fig, ax = plt.subplots(1, 1, figsize=(4.5, 4.5))

for cluster_label in np.unique(y_predict):
    if cluster_label == -1: # Ignorar los puntos
        clasificados como ruido
        continue

    # Obtener los puntos correspondientes al cluster actual
    cluster_points = scaled_datos[y_predict ==
cluster_label]

    # Calcular el límite convexo para los puntos del
cluster
    hull = ConvexHull(cluster_points)

    # Dibujar los puntos del cluster
    plt.scatter(cluster_points[:, 0], cluster_points[:, 1],
s=50, label=f'Cluster {cluster_label}')

    # Dibujar el límite convexo del cluster
    for simplex in hull.simplices:
        plt.plot(cluster_points[simplex, 0],
cluster_points[simplex, 1], 'k-', lw=2)

# Dibujar los puntos clasificados como ruido
outliers = scaled_datos[y_predict == -1]
plt.scatter(outliers[:, 0], outliers[:, 1], s=50, c='gray',
marker='x', label='Outliers')

# Establecer el título y etiquetas de los ejes
plt.title('Resultado de DBSCAN con Convex Cluster Hulls')
plt.xlabel('Eje Edad asegurado')
plt.ylabel('Eje Capital')
plt.legend()
plt.grid()
plt.show()

###cluster hulls alternativo
import numpy as np

```

```
import matplotlib.pyplot as plt
from sklearn.cluster import DBSCAN
from sklearn.preprocessing import StandardScaler
from scipy.spatial import ConvexHull

# Crear y ajustar el modelo DBSCAN a tus datos escalados
modelo_dbscan = DBSCAN(eps=0.5, min_samples=4)
y_predict = modelo_dbscan.fit_predict(X=scaled_datos)

# Crear un gráfico de dispersión para visualizar los
clusters y los límites convexos
plt.figure(figsize=(8, 6))

# Iterar sobre cada cluster
for cluster_label in np.unique(y_predict):
    if cluster_label == -1: # Ignorar los puntos
        clasificados como ruido
            continue

    # Obtener los puntos correspondientes al cluster actual
    cluster_points = scaled_datos[y_predict ==
cluster_label]

    # Calcular el límite convexo para los puntos del
cluster
    hull = ConvexHull(cluster_points)

    # Dibujar el límite convexo del cluster
    for simplex in hull.simplices:
        plt.plot(cluster_points[simplex, 0],
cluster_points[simplex, 1], 'k-', lw=2)

# Dibujar las áreas de los clusters con diferentes colores
for cluster_label, color in zip(np.unique(y_predict),
['lightgreen', 'orange', 'lightblue', 'yellow']):
    if cluster_label == -1: # Ignorar los puntos
        clasificados como ruido
            continue

    # Obtener los puntos correspondientes al cluster actual
    cluster_points = scaled_datos[y_predict ==
cluster_label]

    # Dibujar los puntos del cluster con el color
correspondiente
    plt.scatter(cluster_points[:, 0], cluster_points[:, 1],
s=50, c=color, label=f'Cluster {cluster_label}')

# Dibujar los puntos clasificados como ruido
outliers = scaled_datos[y_predict == -1]
plt.scatter(outliers[:, 0], outliers[:, 1], s=50, c='gray',
```

```
marker='x', label='Outliers')

# Establecer el título y etiquetas de los ejes
plt.title('Resultado de DBSCAN con Convex Cluster Hulls y
Áreas de Cluster')
plt.xlabel('Eje Edad asegurado')
plt.ylabel('Eje Capital')
plt.legend()
plt.grid()
plt.show()

#métricas
from sklearn.cluster import DBSCAN
from sklearn.preprocessing import StandardScaler
from scipy.spatial.distance import pdist, squareform
import numpy as np

# Ajustar el modelo DBSCAN a tus datos escalados
modelo_dbscan = DBSCAN(eps=0.5, min_samples=4)
y_predict = modelo_dbscan.fit_predict(X=scaled_datos)

# Etiquetas de cluster asignadas por DBSCAN (-1 para
outliers)
etiquetas = modelo_dbscan.labels_

# Calcula el número de datos dentro de cada clúster
conteo_por_cluster = np.bincount(etiquetas[etiquetas != -
1])

# Calcula la distancia máxima entre dos puntos dentro de
cada clúster
for cluster_id in np.unique(etiquetas[etiquetas != -1]):
    puntos_cluster = scaled_datos[etiquetas == cluster_id]
    distancias = pdist(puntos_cluster)
    distancia_maxima = np.max(squareform(distancias))
    print(f"Distancia máxima dentro del clúster
{cluster_id}:", distancia_maxima)

# Calcula la distancia promedio entre todos los puntos
dentro de cada clúster
for cluster_id in np.unique(etiquetas[etiquetas != -1]):
    puntos_cluster = scaled_datos[etiquetas == cluster_id]
    distancias = pdist(puntos_cluster)
    distancia_promedio = np.mean(distancias)
    distancia_mediana = np.median(distancias)
    print(f"Distancia promedio dentro del clúster
{cluster_id}:", distancia_promedio)
    print(f"Distancia mediana dentro del clúster
{cluster_id}:", distancia_mediana)

# Separación entre clústeres (puede no tener sentido en
```

```

DBSCAN)
if len(np.unique(etiquetas)) > 1:
    centroides = []
    for cluster_id in np.unique(etiquetas[etiquetas != -
1]):
        puntos_cluster = scaled_datos[etiquetas ==
cluster_id]
        centroide = np.mean(puntos_cluster, axis=0)
        centroides.append(centroide)
        centroides = np.array(centroides)
        distancias_centroides = np.linalg.norm(centroides[:,
np.newaxis] - centroides, axis=-1)
        separacion_entre_clusteres_min =
np.min(distancias_centroides[distancias_centroides > 0])
        print("Separación mínima entre clústeres:",
separacion_entre_clusteres_min)

#Separación máxima entre clústeres
if len(np.unique(etiquetas)) > 1:
    centroides = []
    for cluster_id in np.unique(etiquetas[etiquetas != -
1]):
        puntos_cluster = scaled_datos[etiquetas ==
cluster_id]
        centroide = np.mean(puntos_cluster, axis=0)
        centroides.append(centroide)
        centroides = np.array(centroides)
        distancias_centroides = np.linalg.norm(centroides[:,
np.newaxis] - centroides, axis=-1)

        # Excluimos las distancias diagonales y las distancias
negativas (debido a que np.linalg.norm puede devolver
valores negativos debido a errores de redondeo)
        distancias_validas =
distancias_centroides[(distancias_centroides > 0) &
(distancias_centroides != np.nan)]

        # Calculamos la separación mínima entre los clusters
        separacion_entre_clusteres_min =
np.min(distancias_validas)
        print("Separación mínima entre clústeres:",
separacion_entre_clusteres_min)

#la máxima
if len(np.unique(etiquetas)) > 1:
    centroides = []
    for cluster_id in np.unique(etiquetas[etiquetas != -
1]):
        puntos_cluster = scaled_datos[etiquetas ==
cluster_id]
        centroide = np.mean(puntos_cluster, axis=0)

```

```

centroides.append(centroide)
centroides = np.array(centroides)
distancias_centroides = np.linalg.norm(centroides[:,
np.newaxis] - centroides, axis=-1)

# Excluimos las distancias diagonales y las distancias
negativas (debido a que np.linalg.norm puede devolver
valores negativos debido a errores de redondeo)
distancias_validas =
distancias_centroides[(distancias_centroides > 0) &
(distancias_centroides != np.nan)]

# Calculamos la separación máxima entre los clusters
separacion_entre_clusteres_max =
np.max(distancias_validas)
print("Separación máxima entre clústeres:",
separacion_entre_clusteres_max)
# Distancia promedio entre los puntos de diferentes
clústeres (puede no tener sentido en DBSCAN)
from scipy.spatial.distance import cdist
if len(np.unique(etiquetas)) > 1:
    distancias_promedio_entre_clusteres = 0
    total_pares = 0
    for i in range(len(centroides)):
        for j in range(i+1, len(centroides)):
            puntos_cluster_1 = scaled_datos[etiquetas == i]
            puntos_cluster_2 = scaled_datos[etiquetas == j]
            distancias = cdist(puntos_cluster_1,
puntos_cluster_2)
            distancias_promedio_entre_clusteres +=
np.mean(distancias)
            total_pares += 1
    distancia_promedio_entre_clusteres =
distancias_promedio_entre_clusteres / total_pares
    print("Distancia promedio entre los puntos de
diferentes clústeres:", distancia_promedio_entre_clusteres)

# Distancia promedio dentro de los clústeres (puede no
tener sentido en DBSCAN)
distancia_promedio_dentro_clusteres = 0
total_clusteres = len(np.unique(etiquetas[etiquetas != -
1]))
for cluster_id in np.unique(etiquetas[etiquetas != -1]):
    puntos_cluster = scaled_datos[etiquetas == cluster_id]
    distancias = pdist(puntos_cluster)
    distancia_promedio_dentro_clusteres +=
np.mean(distancias)
distancia_promedio_dentro_clusteres =
distancia_promedio_dentro_clusteres / total_clusteres
print("Distancia promedio dentro de los clústeres:",
distancia_promedio_dentro_clusteres)

```

```
#Jerárquico
import numpy as np
#paso 1: construir la matriz de distancias
datos_excel
scaled_datos
variables=['Age_Actuarial','Capital']
df=pd.DataFrame(scaled_datos,columns=variables)
from scipy.spatial.distance import pdist, squareform
row_dist=pd.DataFrame(squareform(pdist(df,metric='euclidean')
))
row_dist
from scipy.cluster.hierarchy import linkage, dendrogram
row_clusters=linkage(pdist(df,
metric='euclidean'),method='complete')
num_clusters = 3
from scipy.cluster.hierarchy import linkage, dendrogram,
cut_tree
cluster_labels = cut_tree(row_clusters,
n_clusters=num_clusters)
cluster_info = pd.DataFrame(cluster_labels,
columns=['Cluster'], index=df.index)
from scipy.cluster.hierarchy import fcluster
cluster_labels = fcluster(row_clusters, t=num_clusters,
criterion='maxclust')
from scipy.stats import f_oneway
row_dendr = dendrogram(row_clusters, truncate_mode='lastp',
p=num_clusters)
plt.tight_layout()
plt.ylabel('Euclidean distance')
plt.show()

#Evaluar clusteres
import numpy as np
import pandas as pd
from scipy.spatial.distance import pdist, squareform
from scipy.cluster.hierarchy import linkage, dendrogram,
fcluster

# Obtener las etiquetas de los clusters
cluster_labels = fcluster(row_clusters, t=num_clusters,
criterion='maxclust')

# Añadir las etiquetas de los clusters al DataFrame
df['Cluster'] = cluster_labels

# 1. Número de observaciones de cada grupo
num_obs_por_grupo = df['Cluster'].value_counts()

# 2. Distancia máxima entre dos puntos dentro de un clúster
dist_max_por_cluster = df.groupby('Cluster').apply(lambda
```

```
x: pdist(x.iloc[:, :-1]).max())

# 3. Distancia promedio entre todos los puntos dentro de un
clúster
dist_prom_por_cluster = df.groupby('Cluster').apply(lambda
x: pdist(x.iloc[:, :-1]).mean())

# 4. Distancia mediana entre todos los puntos dentro de un
clúster
dist_mediana_por_cluster =
df.groupby('Cluster').apply(lambda x:
np.median(pdist(x.iloc[:, :-1])))

# 5. Separación mínima entre clústeres
separacion_minima = linkage(df.iloc[:, :-1],
method='single').min()

# 6. Separación máxima entre clústeres
separacion_maxima = linkage(df.iloc[:, :-1],
method='single').max()

# 7. Distancia promedio entre los puntos de diferentes
clústeres
dist_prom_intercluster = pdist(df.iloc[:, :-1]).mean()

# 8. Distancia promedio dentro de los clústeres
dist_prom_intracluster = df.groupby('Cluster').apply(lambda
x: pdist(x.iloc[:, :-1]).mean())
#fin
cluster_stats = df.groupby(cluster_labels).agg(['mean',
'std', 'count']).T
cluster_stats

df_rowclust=df.iloc[row_dendr['leaves'][:, :-1]]

from sklearn.metrics import silhouette_score,
davies_bouldin_score

# Evaluar la calidad de los clusters para diferentes
números
for num_clusters in range(2, 10):
    cluster_labels = fcluster(row_clusters, t=num_clusters,
criterion='maxclust')
    silhouette = silhouette_score(df, cluster_labels)
    db_index = davies_bouldin_score(df, cluster_labels)
    print(f"Clusters: {num_clusters}, Silhouette Score:
{silhouette}, Davies-Bouldin Index: {db_index}")
```

```

from scipy.cluster.hierarchy import dendrogram
row_dendr=dendrogram(row_clusters)
plt.tight_layout()
plt.ylabel('Euclidean distance')
plt.show()

#####ALGORITMO
ALGOMERATIVO #####
from sklearn.cluster import AgglomerativeClustering
from scipy.stats import spearmanr
n_clusters = 3
metodos_enlace = ['complete', 'average', 'single', 'ward']
for metodo_enlace in metodos_enlace:
    modelo_aglomerativo =
AgglomerativeClustering(n_clusters=n_clusters,
linkage=metodo_enlace)
    etiquetas =
modelo_aglomerativo.fit_predict(scaled_datos)
    coeficiente_correlacion, _ = spearmanr(etiquetas,
range(len(etiquetas)))
    print(f"Coeficiente de correlación para el método de
enlace '{metodo_enlace}': {coeficiente_correlacion}")
len(scaled_datos)
from scipy.cluster.hierarchy import dendrogram, linkage,
fcluster
metodo_enlace = 'ward'
matriz_enlace = linkage(scaled_datos, method=metodo_enlace)
plt.figure(figsize=(10, 5))
dendrogram(matriz_enlace,truncate_mode='level',p=3)
plt.title('Dendrograma con método de enlace "Ward"')
plt.xlabel('Índices de las muestras')
plt.ylabel('Distancia')
plt.show()

# Podar el dendrograma para obtener solo 3 clusters
num_clusters_deseados = 3
clusters = fcluster(matriz_enlace, num_clusters_deseados,
criterion='maxclust')
print(clusters)
np.bincount(clusters)
len(clusters)

datos_cluster = [[capital, edad] for capital, edad in
zip(datos_excel['Capital'], datos_excel['Age_Actuarial'])]
for i in range(len(datos_cluster)):
    datos_cluster[i].append(clusters[i])
print(datos_cluster)
for i in range(len(datos_cluster)):
    datos_cluster[i].append(list(scaled_datos[i]))
print(datos_cluster)

```



```
datos_cluster1_capital = []
datos_cluster1_edad = []
datos_cluster2_capital = []
datos_cluster2_edad = []
datos_cluster3_capital = []
datos_cluster3_edad = []

for dato in datos_cluster:
    if dato[2] == 1:
        datos_cluster1_edad.append(dato[1])
        datos_cluster1_capital.append(dato[0])
    elif dato[2] == 2:
        datos_cluster2_edad.append(dato[1])
        datos_cluster2_capital.append(dato[0])
    elif dato[2] == 3:
        datos_cluster3_edad.append(dato[1])
        datos_cluster3_capital.append(dato[0])

# Calcular estadísticos para edades y capitales de cada
cluster
#Clúster 1
#Edad
media_cluster1_edad = np.mean(datos_cluster1_edad)
varianza_cluster1_edad = np.var(datos_cluster1_edad)
np.std(datos_cluster1_edad)
mediana_cluster1_edad = np.median(datos_cluster1_edad)
cuartil25_cluster1_edad =
np.percentile(datos_cluster1_edad, 25)
cuartil75_cluster1_edad =
np.percentile(datos_cluster1_edad, 75)
maximo_cluster1_edad = np.max(datos_cluster1_edad)
minimo_cluster1_edad = np.min(datos_cluster1_edad)
asimetria_cluster1_edad = skew(datos_cluster1_edad)
curtosis_cluster1_edad = kurtosis(datos_cluster1_edad)

#Capital
media_cluster1_capital = np.mean(datos_cluster1_capital)
varianza_cluster1_capital = np.var(datos_cluster1_capital)
np.std(varianza_cluster1_capital)
mediana_cluster1_capital =
np.median(datos_cluster1_capital)
cuartil25_cluster1_capital =
np.percentile(datos_cluster1_capital, 25)
cuartil75_cluster1_capital =
np.percentile(datos_cluster1_capital, 75)
maximo_cluster1_capital = np.max(datos_cluster1_capital)
minimo_cluster1_capital = np.min(datos_cluster1_capital)
asimetria_cluster1_capital = skew(datos_cluster1_capital)
curtosis_cluster1_capital =
kurtosis(datos_cluster1_capital)
```

```
#Segundo clúster
#Edad
media_cluster2_edad = np.mean(datos_cluster2_edad)
varianza_cluster2_edad = np.var(datos_cluster2_edad)
mediana_cluster2_edad = np.median(datos_cluster2_edad)
cuartil25_cluster2_edad =
np.percentile(datos_cluster2_edad, 25)
cuartil75_cluster2_edad =
np.percentile(datos_cluster2_edad, 75)
maximo_cluster2_edad = np.max(datos_cluster2_edad)
minimo_cluster2_edad = np.min(datos_cluster2_edad)
asimetria_cluster2_edad = skew(datos_cluster2_edad)
curtosis_cluster2_edad = kurtosis(datos_cluster2_edad)
#Capital
media_cluster2_capital = np.mean(datos_cluster2_capital)
varianza_cluster2_capital = np.var(datos_cluster2_capital)
mediana_cluster2_capital =
np.median(datos_cluster2_capital)
cuartil25_cluster2_capital =
np.percentile(datos_cluster2_capital, 25)
cuartil75_cluster2_capital =
np.percentile(datos_cluster2_capital, 75)
maximo_cluster2_capital = np.max(datos_cluster2_capital)
minimo_cluster2_capital = np.min(datos_cluster2_capital)
asimetria_cluster2_capital = skew(datos_cluster2_capital)
curtosis_cluster2_capital =
kurtosis(datos_cluster2_capital)

#Cluster 3
#Edad
media_cluster3_edad = np.mean(datos_cluster3_edad)
varianza_cluster3_edad = np.var(datos_cluster3_edad)
mediana_cluster3_edad = np.median(datos_cluster3_edad)
cuartil25_cluster3_edad =
np.percentile(datos_cluster3_edad, 25)
cuartil75_cluster3_edad =
np.percentile(datos_cluster3_edad, 75)
maximo_cluster3_edad = np.max(datos_cluster3_edad)
minimo_cluster3_edad = np.min(datos_cluster3_edad)
asimetria_cluster3_edad = skew(datos_cluster3_edad)
curtosis_cluster3_edad = kurtosis(datos_cluster3_edad)
#Capital
media_cluster3_capital = np.mean(datos_cluster3_capital)
varianza_cluster3_capital = np.var(datos_cluster3_capital)
mediana_cluster3_capital =
np.median(datos_cluster3_capital)
cuartil25_cluster3_capital =
np.percentile(datos_cluster3_capital, 25)
cuartil75_cluster3_capital =
np.percentile(datos_cluster3_capital, 75)
```

```
maximo_cluster3_capital = np.max(datos_cluster3_capital)
minimo_cluster3_capital = np.min(datos_cluster3_capital)
asimetria_cluster3_capital = skew(datos_cluster3_capital)
curtosis_cluster3_capital =
kurtosis(datos_cluster3_capital)

# Imprime los clusters obtenidos
print("Clusters obtenidos después del corte del
dendrograma:")
print(clusters)

plt.figure(figsize=(8, 6))
colores = ['r', 'g', 'b']

for cluster_id in range(1, 4): # Los clusters están
numerados del 1 al 3
    plt.scatter(scaled_datos[clusters == cluster_id, 0],
                scaled_datos[clusters == cluster_id, 1],
                c=colores[cluster_id - 1], label=f'Cluster
{cluster_id}')

plt.title('Gráfico de dispersión algoritmo jerárquico
aglomerativo')
plt.xlabel('Edad')
plt.ylabel('Capital en Euros')
plt.legend()
plt.grid(True)
plt.show()
#####Hasta aquí algoritmo
jerárquico aglomerativo####
#####CLUSTERING C
MEANS#####
import skfuzzy as fuzz
n_clusters = 3
m = 2
error = 0.005
centros, pertenencias, u0, distancia, iteracion, fpc =
fuzz.cluster.cmeans(
    scaled_datos.T, n_clusters, m, error=error,
maxiter=1000, init=None)

#-----
#Empieza Random Forest
```

```

ruta_archivo_excel =
'G:/Esta_es_mi_carpeta2/Portafolio_Random_Forest.xlsm'
datos_excel = pd.read_excel(ruta_archivo_excel)
print(datos_excel.columns)
datos_excel=datos_excel.iloc[:, [2,16]]
X1=datos_excel.iloc[:, [0]].values
y1=datos_excel.iloc[:, [1]].values
print('Etiquetas de clase:', np.unique(y1))
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test
=train_test_split(X1,y1,test_size=0.2,

random_state=1,

stratify=y1)
from sklearn.ensemble import RandomForestClassifier
classifier_rf=RandomForestClassifier(random_state=42,
criterion='gini', max_depth=5,
n_estimators=100,
oob_score=True)
classifier_rf.fit(X_train,y_train.ravel())
classifier_rf.oob_score_
#tuning hyperparameters
rf = RandomForestClassifier(random_state=42, n_jobs=-1)
params = {
    'max_depth': [2,3,5,10,20],
    'n_estimators': [10,25,30,50,100,200]
}
from sklearn.model_selection import GridSearchCV
grid_search = GridSearchCV(estimator=rf,
param_grid=params,
cv = 4,
n_jobs=-1, verbose=1,
scoring="accuracy")
grid_search.fit(X_train, y_train)
grid_search.best_score_
rf_best = grid_search.best_estimator_
rf_best
from sklearn.tree import plot_tree
plt.figure(figsize=(50,40))
plot_tree(rf_best.estimators_[2], feature_names =
X_train,class_names=['Disease', "No Disease"],filled=True);

#Random Forest Balanceado
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from imlearn.over_sampling import SMOTE
from sklearn.preprocessing import LabelEncoder
from sklearn.linear_model import LogisticRegression

```

```
from sklearn.tree import DecisionTreeClassifier
from sklearn.tree import resample
from sklearn.model_selection import train_test_split,
cross_val_score, GridSearchCV
from sklearn import neighbors
from sklearn.metrics import confusion_matrix,
precision_score, recall_score

#classifier
from imblearn.ensemble import
BalancedRandomForestClassifier
from sklearn.metrics import classification_report
from sklearn.metrics import confusion_matrix
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X1, y1,
test_size=0.2, random_state=1)

model=BalancedRandomForestClassifier(n_estimators=100,
random_state=0)
model.fit(X_train, y_train)
y_pred_rf=model.predict(X_test)
print("Training Accuracy: ", model.score(X_train, y_train))
print('Testing Accuracy: ', model.score(X_test, y_test))
cr=classification_report(y_test,y_pred_rf)
print(cr)
plt.rcParams['figure.figsize']=(5,5)
cm=confusion_matrix(y_test, y_pred_rf)
sns.heatmap(cm,annot=True,cmap='spring')
plt.show()

#XGBoost
import xgboost as xgb
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score,
classification_report

model=xgb.XGBClassifier(objective='multi:softprob',num_clas
s=2,max_depth=5,
                        n_estimators=100)
X,y=datos_excel.Age_Actuarial,datos_excel.Class_Capital
X_train, X_test, y_train, y_test
=train_test_split(datos_excel.iloc[:,
0],datos_excel.iloc[:,1],

test_size=0.2,random_state=42)
model.fit(X_train,y_train)
predictions=model.predict(X_test)
accuracy=accuracy_score(y_test,predictions)
print("Accuracy:", accuracy)
print("\nClassification Report:")
print(classification_report(y_test,predictionprint("Accurac
```

```
y:", accuracy)
print("\nClassification Report:")
print(classification_report(y_test,predictions,
target_names=data.target_names))
```

9.3.3. Flujos de efectivo

```
import numpy as np
import pandas as pd
from sklearn.datasets import make_blobs
from sklearn.ensemble import RandomForestClassifier

import matplotlib.pyplot as plt
import seaborn as sns

ruta_archivo_excel =
'G:/Esta_es_mi_carpeta2/Excels_definitivos/Portafolio_vida_
sin_filtro_2000.xlsx'
datos_excel1 = pd.read_excel(ruta_archivo_excel)
print(datos_excel1.columns)
datos_excel=datos_excel1.iloc[:,[4,8]]
print(datos_excel.head())
print(datos_excel.tail())
print(datos_excel.describe())
datos
#Para el primer asegurado
datos[0]
#Diferimiento
edad_actuarial=datos[0][1]
m=67-edad_actuarial
h=1
t_final=(m+n)*h-1
t=(m*h)
capital=datos[0][0]
tipo_interes=0.02

lista_de_filas = PER_2o_orden.values.tolist()
print(lista_de_filas)

PER_nueva_produccion=
pd.read_excel('G:\Esta_es_mi_carpeta2\Excels_definitivos\Pa
ra_los_contratos_de
nueva_produccion_Tablas_PERM_2000.xlsx')
PER_contratos_en_cartera=pd.read_excel('G:\Esta_es_mi_carpe
ta2\Excels_definitivos\Para_los_contratos_en_cartera_Tablas
_PERM_2000.xlsx')
GKMF95_Unisex=pd.read_excel('G:\Esta_es_mi_carpeta2\Excels_
definitivos\GKMF95_Unisex.xlsx')
(GKMF95_Unisex.iloc[0:111]).tail()
GKMF95_Unisex=GKMF95_Unisex.iloc[0:111]
```

```
PASEM_2020=pd.read_excel('G:\Esta_es_mi_carpeta2\Excels_definitivos\PASEM_2020.xlsx')
(PASEM_2020.iloc[0:95]).tail()
PASEM_2020=PASEM_2020.iloc[0:95]
x=49
GKMF95 =
pd.read_excel(r"G:\Esta_es_mi_carpeta2\Excels_definitivos\GKMF95.xls", engine='xlrd')
GKMF95= GKMF95.iloc[0:111]
datos_excel
datos_excel1
datos

datos = [[identificador, sexo, capital, edad] for
identificador, sexo, capital, edad in
zip(datos_excel1['ID'], datos_excel1['Gender'], datos_excel1[
'Capital'], datos_excel1['Age_Actuarial'])]
for i in range(len(datos)):
    datos[i].append(y_predict[i])
print(datos)
PER_contratos_en_cartera =
pd.read_excel('G:\Esta_es_mi_carpeta2\Excels_definitivos\Para_los_contratos_en_cartera_Tablas_PERM_2000.xlsx')

datos_PER_contratos_en_cartera = [[nacimiento, hombres,
mujeres, hombres_mejora, mujeres_mejora] for nacimiento,
hombres, mujeres, hombres_mejora, mujeres_mejora in
zip(PER_contratos_en_cartera['Nacimiento'],
PER_contratos_en_cartera['Hombres'],
PER_contratos_en_cartera['Mujeres'],
PER_contratos_en_cartera['Factor_mejora_Hombres'],
PER_contratos_en_cartera['Factor_mejora_Mujeres'])]

datos_GKMF95 = [[edad, q_hombres, q_mujeres] for edad,
q_hombres, q_mujeres in zip(GKMF95['EDAD'], GKMF95['qxGKM-
95'], GKMF95['qyGKF-95'])]

datos_GKMF95_Unisex = [[edad, q] for edad, q in
zip(GKMF95_Unisex['EDAD'], GKMF95_Unisex['qxGKM-
95Unisex'])]

datos_PASEM_2020 = [[edad, q_hombres, q_mujeres] for edad,
q_hombres, q_mujeres in zip(PASEM_2020['EDAD'],
PASEM_2020['Rel_qx_mil_hombres'],
PASEM_2020['Rel_qx_mil_mujeres'])]

# Crea un diccionario para almacenar los datos
datos_dict_GKMF95 = {}

# Itera sobre las sublistas de datos_GKMF95 y agrega cada
```

```
sublista al diccionario con la edad como clave
for dato in datos_GKMF95:
    edad = dato[0]
    datos_dict_GKMF95[edad] = {
        'Hombres': dato[1],
        'Mujeres': dato[2]
    }

# Muestra el diccionario
print(datos_dict_GKMF95)

# Para acceder a un determinado valor de una palabra clave
del diccionario
edad_45_hombres = datos_dict_GKMF95[45]['Hombres']
print("Cantidad de hombres de 45 años:", edad_45_hombres)

# Crea un diccionario para almacenar los datos
datos_dict_GKMF95_Unisex = {}

# Itera sobre las sublistas de datos_GKMF95_Unisex y agrega
cada sublista al diccionario con la edad como clave
for dato in datos_GKMF95_Unisex:
    edad = dato[0]
    datos_dict_GKMF95_Unisex[edad] = {
        'Probabilidad': dato[1]
    }

# Muestra el diccionario
print(datos_dict_GKMF95_Unisex)

# Para acceder a un determinado valor de una palabra clave
del diccionario
probabilidad_50_anios =
datos_dict_GKMF95_Unisex[50]['Probabilidad']
print("Probabilidad a los 50 años:", probabilidad_50_anios)

# Crea un diccionario para almacenar los datos
datos_dict_PASEM_2020 = {}

# Itera sobre las sublistas de datos_PASEM_2020 y agrega
cada sublista al diccionario con la edad como clave
for dato in datos_PASEM_2020:
    edad = dato[0]
    datos_dict_PASEM_2020[edad] = {
        'Hombres': dato[1],
        'Mujeres': dato[2]
    }

# Muestra el diccionario
print(datos_dict_PASEM_2020)
```



```
GKMF95_Unisex=pd.read_excel('G:\Esta_es_mi_carpeta2\Excels_
definitivos\GKMF95_Unisex.xlsx')
(GKMF95_Unisex.iloc[0:111]).tail()
GKMF95_Unisex=GKMF95_Unisex.iloc[0:111]
PASEM_2020=pd.read_excel('G:\Esta_es_mi_carpeta2\Excels_def
initivos\PASEM_2020.xlsx')
(PASEM_2020.iloc[0:95]).tail()
PASEM_2020=PASEM_2020.iloc[0:95]
GKMF95 =
pd.read_excel(r"G:\Esta_es_mi_carpeta2\Excels_definitivos\G
KMF95.xls", engine='xlrd')
GKMF95= GKMF95.iloc[0:111]
#####LISTA DE ASEGURADOS:
[[identificado, sexo, capital,
edad],...]#####
datos = [[identificador, sexo,capital, edad] for
identificador, sexo,capital, edad in
zip(datos_excel1['ID'],datos_excel1['Gender'],datos_excel1[
'Capital'], datos_excel1['Age_Actuarial'])]
for i in range(len(datos)):
    datos[i].append(y_predict[i])
print(datos)

datos_cluster0_M = []
datos_cluster1_M = []
datos_cluster2_M = []
datos_cluster0_F = []
datos_cluster1_F = []
datos_cluster2_F = []

for dato in datos:
    genero = dato[1]
    cluster = dato[4]
    if cluster == 0:
        if genero == 'M':
            datos_cluster0_M.append(dato)
        else:
            datos_cluster0_F.append(dato)
    elif cluster == 1:
        if genero == 'M':
            datos_cluster1_M.append(dato)
        else:
            datos_cluster1_F.append(dato)
    elif cluster == 2:
        if genero == 'M':
            datos_cluster2_M.append(dato)
        else:
            datos_cluster2_F.append(dato)

print("Datos en cluster 0, Hombres:", datos_cluster0_M)
```

```

print("Datos en cluster 0, Mujeres:", datos_cluster0_F)
print("Datos en cluster 1, Hombres:", datos_cluster1_M)
print("Datos en cluster 1, Mujeres:", datos_cluster1_F)
print("Datos en cluster 2, Hombres:", datos_cluster2_M)
print("Datos en cluster 2, Mujeres:", datos_cluster2_F)

diccionario_asegurados = {}
for asegurado in datos_cluster2_M:
    id_asegurado = asegurado[0]
    edad_original = asegurado[3]
    lista_probabilidades = []
    productorio_probabilidades = 1
    for edad in range(edad_original, 67):
        probabilidad_edad =
datos_dict_GKMF95_Unisex[edad]['Probabilidad']
        lista_probabilidades.append((1-
probabilidad_edad) * productorio_probabilidades)
        productorio_probabilidades *= (1-
probabilidad_edad)
    diccionario_asegurados[id_asegurado] =
lista_probabilidades
print(diccionario_asegurados)

diccionario_suma_final = {}
for id_asegurado, lista_probabilidades in
diccionario_asegurados.items():
    t_min = 1
    suma_final = 1
    for probabilidad in lista_probabilidades:
        suma_final += ((1 + 0.02) ** (-t_min)) *
probabilidad
        t_min += 1
    diccionario_suma_final[id_asegurado] = suma_final
print(diccionario_suma_final)

diccionario_suma_final = {}
for id_asegurado, lista_probabilidades in
diccionario_asegurados.items():
    edad_original = datos_cluster2_M[id_asegurado][3] #
Obtener la edad original del asegurado
    suma_final =
datos_dict_GKMF95_Unisex[edad_original]['Probabilidad'] *
(1 + 0.02) ** 1 # Primer término
    edad = edad_original + 1
    i=1
    for probabilidad in (lista_probabilidades):
        suma_final += probabilidad *
datos_dict_GKMF95_Unisex[edad]['Probabilidad'] * (1 + 0.02)
** -(i + 1)

```

```

        edad+=1
        i+=1
        diccionario_suma_final[id_asegurado] = suma_final

print(diccionario_suma_final)

diccionario_suma_final2 = {}
posicion=0
for id_asegurado, lista_probabilidades in
diccionario_asegurados.items():
    edad_original = datos_cluster2_M[posicion][3] #
Obtener la edad original del asegurado
    suma_final =
datos_dict_GKMF95_Unisex[edad_original]['Probabilidad'] *
(1 + 0.02) ** 1 # Primer término
    edad = edad_original + 1
    i = 1
    for probabilidad in lista_probabilidades:
        suma_final += probabilidad *
datos_dict_GKMF95_Unisex[edad]['Probabilidad'] * (1 + 0.02)
** -(i + 1)
        edad += 1
        i += 1
        diccionario_suma_final2[id_asegurado] = suma_final
    posicion=posicion+1
print(diccionario_suma_final2)

diccionario_suma_total = {}
for id_asegurado, lista_prob_acumuladas in
diccionario_prob_acumuladas.items():
    suma_total = (1 + 0.02) ** -(0)
    i=1
    for elem in lista_prob_acumuladas:
        suma_total += (1 + 0.02) ** -(i) * elem
        i=i+1
    diccionario_suma_total[id_asegurado] = suma_total
print(diccionario_suma_total)

diccionario_nueva_suma_acumulada = {}
for asegurado in datos_cluster2_M:
    id_asegurado = asegurado[0] # ID del asegurado
    edad_actual = asegurado[3] # Edad actual del asegurado
    t_min = 67 - edad_actual
    suma_acumulada = 0
    probab_mortalidad=67
    lista_prob_acumuladas =
diccionario_prob_acumuladas[id_asegurado]
    for t in range(t_min, len(lista_prob_acumuladas)):
        prob_acumulada_t = lista_prob_acumuladas[t]

```

```
        suma_acumulada += (1 + 0.02) ** -(t + 1) *
prob_acumulada_t*datos_dict_GKMF95_Unisex[probab_mortalidad
][['Probabilidad']]
        probab_mortalidad=probab_mortalidad+1
        diccionario_nueva_suma_acumulada[id_asegurado] =
suma_acumulada

print(diccionario_nueva_suma_acumulada)
```