Máster en Ciencias Actuariales y Financieras

2024-2025

Trabajo Fin de Máster

"Modelización de la Mortalidad con Enfoques Actuariales y de Inteligencia Artificial"

Alex Caro Martínez

Tutor/es

Raquel Pérez Calderón

Lugar y fecha de presentación prevista

RESUMEN

El presente trabajo tiene como objetivo predecir la tasa de mortalidad aplicada a las primas de seguros, utilizando tanto modelos tradicionales como técnicas de inteligencia artificial. En primer lugar, se llevará a cabo un análisis basado en el modelo de Lee-Carter, ampliamente utilizado en el ámbito actuarial para la proyección de tasas de mortalidad. Este modelo permitirá comprender su estructura, componentes y aplicación práctica en la estimación de la longevidad.

Posteriormente, se incorporarán métodos basados en inteligencia artificial para calcular y analizar las tasas de mortalidad desde un enfoque más innovador y con mayor proyección de futuro. En particular, se aplicarán técnicas de aprendizaje profundo con el fin de optimizar la capacidad predictiva de los modelos utilizados.
Finalmente, se realizará una comparación entre los enfoques tradicionales y los modelos de inteligencia artificial para evaluar su eficiencia predictiva, identificar fortalezas y debilidades, y analizar su aplicabilidad en el contexto asegurador. Este trabajo pretende así ofrecer una visión integral del potencial de los modelos clásicos y modernos en la predicción de la mortalidad, con especial énfasis en su impacto sobre la tarificación de productos de seguros.

Palabras Clave: Mortalidad, seguros de vida, Lee-Carter, inteligencia artificial, aprendizaje profundo.

DEDICATORIA

A mis padres, José y María, por su amor incondicional, su ejemplo de constancia y apoyo en cada paso del camino. Este trabajo también es vuestro.

ÍNDICE GENERAL

ÍNDICE DE FIGURAS

1. INTRODUCCIÓN

1.1 PLANTEAMIENTO DEL PROBLEMA

El envejecimiento de la población es un fenómeno global que está generando transformaciones de gran envergadura en diversos sectores de la sociedad. La industria aseguradora y los fondos de pensiones no son una excepción. Los avances en el campo de la medicina, las mejoras continuas en la calidad de vida de la población y el aumento de la disponibilidad de herramientas tecnológicas para la gestión y eficiencia de los sistemas de salud, han contribuido de manera decisiva a una mayor esperanza de vida. Este fenómeno ha cambiado la estructura demográfica en todos los países, generando tanto oportunidades como desafíos significativos en distintos ámbitos de la economía y la sociedad.

Uno de los principales retos derivados de este fenómeno es la incertidumbre que genera la evolución de la longevidad, particularmente para las aseguradoras. En el caso de las aseguradoras, que ofrecen distintos productos financieros, un incremento inesperado de la esperanza de vida podría poner en peligro su estabilidad financiera y solvencia. Esto ocurre porque, al ofrecer este tipo de productos, las aseguradoras están sujetas a la incertidumbre de la longevidad de sus asegurados, ya que, si estos viven más tiempo del anticipado, la aseguradora se vería obligada a hacer pagos durante un periodo más largo al inicialmente proyectado.

El riesgo asociado al envejecimiento y la longevidad tiene implicaciones profundas en la sostenibilidad de los productos financieros que las aseguradoras ofrecen, particularmente aquellos que implican pagos periódicos a lo largo de la vida del beneficiario, como las rentas vitalicias. Estos productos son fundamentales para las personas mayores, ya que les proporcionan seguridad financiera durante la jubilación. Sin embargo, la efectividad y sostenibilidad de los distintos productos dependen directamente de una estimación precisa de la mortalidad de los asegurados. Si las aseguradoras no pueden prever con precisión cuánto tiempo vivirán sus clientes, cualquier desviación en las expectativas de longevidad podría resultar en desajustes significativos en la planificación financiera, especialmente en cuanto a las reservas matemáticas necesarias para cubrir los pagos a largo plazo. Dado que los contratos de un horizonte temporal extenso, incluso pequeñas variaciones en la esperanza de vida pueden generar desajustes financieros considerables, lo que podría poner en riesgo la solvencia de las aseguradoras.

Este riesgo de longevidad ha emergido como uno de los principales desafíos a abordar en la modelización actuarial de las rentas vitalicias. Es por ello por lo que, las aseguradoras deben contar con modelos actuariales robustos y bien fundamentados para estimar la mortalidad y la esperanza de vida, y así mitigar el riesgo derivado de una longevidad mayor a la prevista. Los modelos actuariales deben ser lo suficientemente flexibles como para incorporar las incertidumbres y los cambios en los factores demográficos y de salud de la población, lo cual es esencial para garantizar que las aseguradoras puedan cumplir con sus compromisos a largo plazo.

Desde una perspectiva regulatoria, el impacto del riesgo de longevidad ha sido ampliamente reconocido en las normativas internacionales más relevantes, como Solvencia II e IFRS 17. Bajo el marco de Solvencia II, que regula las aseguradoras en la Unión Europea, se exige que las entidades calculen su *Solvency Capital Requirement* (SCR), teniendo en cuenta escenarios adversos relacionados con la longevidad. Específicamente, las aseguradoras deben prever y mantener un capital adicional para hacer frente a aumentos inesperados de la esperanza de vida de los asegurados, lo que implica un desafío en términos de gestión del riesgo y de planificación financiera. De este modo, las aseguradoras están obligadas a realizar análisis actuariales detallados y a incorporar márgenes de seguridad para mitigar el riesgo que el aumento de la longevidad representa.

Por otro lado, la normativa IFRS 17, que establece los principios contables para las aseguradoras, también reconoce la importancia de la estimación precisa de la longevidad. En este sentido, cualquier error en la estimación de la longevidad puede tener consecuencias graves, no solo en la estabilidad financiera de la aseguradora, sino también en la forma en que se presentan sus estados financieros. Una estimación incorrecta de la esperanza de vida puede alterar de manera significativa los informes contables, afectando la confianza de los inversores, analistas y reguladores.

La regulación en torno al riesgo de longevidad subraya, por tanto, la necesidad de contar con una adecuada gestión y modelización actuarial, que permita a las aseguradoras y fondos de pensiones enfrentar los posibles desafíos que surjan con la evolución de la esperanza de vida. Esto es especialmente relevante en un contexto donde las expectativas de vida continúan aumentando debido a los avances médicos y a la mejora de los estilos de vida.

Ante estos desafíos, es fundamental el desarrollo de metodologías avanzadas y adaptadas para evaluar el impacto del riesgo de longevidad en la sostenibilidad de los productos de renta vitalicia. Estas metodologías deben ser capaces de evaluar cómo las variaciones en la longevidad afectan no solo la suficiencia de las reservas matemáticas, sino también los requerimientos de capital establecidos por la normativa vigente. Las aseguradoras deben estar preparadas para ajustar sus modelos de negocio, sus estrategias de inversión y sus políticas de precios en función de los riesgos asociados a la longevidad.

La capacidad de adaptación y de anticiparse a estos cambios será determinante en el futuro de la industria aseguradora. Solo mediante el uso de herramientas actuariales sofisticadas y un enfoque de gestión integral del riesgo, las aseguradoras podrán garantizar la viabilidad y la sostenibilidad de los productos financieros que ofrecen, como las rentas vitalicias, y cumplir con los compromisos asumidos con sus clientes a largo plazo. Esto también contribuirá a la estabilidad del sistema financiero y a la protección de los derechos de los asegurados, que se verán beneficiados por un modelo de seguros más robusto, flexible y adaptado a la realidad demográfica del futuro.

1.2 MOTIVACIÓN DEL ESTUDIO

El riesgo de longevidad es uno de los principales desafíos en la gestión actuarial, el problema central, radica en que las aseguradoras basan sus cálculos en tablas de mortalidad, que muestran las estimaciones de longevidad basadas en datos históricos y tendencias observadas. Sin embargo, si la esperanza de vida futura supera lo previsto en estas tablas, las reservas técnicas pueden resultar insuficientes para cubrir los pagos de las rentas, afectando la estabilidad del sistema asegurador. En este contexto, el presente estudio se justifica por la necesidad de evaluar cómo distintos escenarios de estrés en la longevidad pueden afectar las reservas y la solvencia de una aseguradora que comercializa rentas vitalicias.

Uno de los mayores desafíos a los que se enfrentan los actuarios es la modelización precisa de la longevidad futura. Tradicionalmente, la proyección de la mortalidad se ha basado en modelos paramétricos y estocásticos que buscan capturar tendencias históricas y extrapolarlas hacia el futuro. Sin embargo, para abordar estos desafíos, es crucial que

las aseguradoras implementen metodologías de estrés en sus modelos actuariales que permitan evaluar el impacto de variaciones inesperadas en la longevidad.
Este estudio busca contribuir con este pretexto, proporcionando una metodología para evaluar el efecto del estrés de longevidad en la suficiencia de reservas y en la solvencia de las aseguradoras. Al utilizar enfoques tanto deterministas como estocásticos, se podrá obtener una visión integral de cómo distintos escenarios de longevidad pueden afectar la estabilidad financiera de una entidad aseguradora.

1.3 OBJETIVOS DEL ESTUDIO

El objetivo principal, es cuantificar el impacto de escenarios de estrés en tablas de mortalidad, mediante sus distintas formas de estudio, desde los enfoques deterministas (aumento en la esperanza de vida) como en modelos estocásticos (Lee-Carter); hasta sus formas más dinámicas, para poder evaluar las tasas de mortalidad con la utilización de la Inteligencia Artificial.

2. MARCO TEÓRICO SOBRE EL LEE-CARTER

El modelo Lee-Carter (Carter y Lee, 1992) no solo ha sido únicamente un hito en la modelización de la mortalidad, también se ha consolidado como una base fundamental para el análisis actuarial y la planificación de pensiones y seguros.

Desde una perspectiva más amplia, se puede ver como un modelo de descomposición de series temporales aplicado a las tasas de mortalidad. A través de la descomposición, el modelo permite identificar y separar los efectos de la edad y el tiempo sobre la mortalidad, lo que facilita la predicción de la mortalidad futura.

2.1.1 CONTEXTO HISTÓRICO Y DESARROLLO DEL MODELO

La predicción de la mortalidad ha sido un desafío central en el ámbito de la ciencia actuarial y la salud pública, siendo esencial para el desarrollo y la ejecución de políticas en áreas cruciales como los seguros, las pensiones y los sistemas de salud. La estimación precisa de las tasas de mortalidad resulta clave para evaluar la sostenibilidad a largo plazo de los sistemas de pensiones y seguros, así como para una adecuada planificación de los recursos en el sector de la salud pública.

En los primeros enfoques de modelización de la mortalidad, se adoptaban métodos deterministas que consistían principalmente en proyecciones simples. Estas técnicas no contemplaban las fluctuaciones ni los cambios no lineales en las tasas de mortalidad, limitándose a proyectar las tasas a partir de promedios históricos sin integrar la evolución dinámica de los factores que afectan a estas tasas. Los modelos tradicionales, como los basados en cohortes y extrapolaciones lineales o no lineales, tenían la capacidad de hacer proyecciones a futuro, pero no podían reflejar adecuadamente la complejidad del comportamiento de la mortalidad a lo largo del tiempo ni las variaciones asociadas a factores socioeconómicos, avances médicos o cambios en el estilo de vida de la población.

En este contexto, el modelo Lee-Carter, desarrollado en la década de 1990, supuso un avance considerable en la predicción de la mortalidad. Este modelo introdujo una descomposición estadística innovadora que permitió captar de forma más precisa la complejidad dinámica de las tasas de mortalidad. La principal fortaleza del modelo reside en su capacidad para descomponer las tasas de mortalidad en tres componentes fundamentales. En primer lugar, el componente de nivel de mortalidad por edad modela las tasas de mortalidad específicas para cada grupo de edad, permitiendo observar cómo cambian a medida que la población envejece.
En segundo lugar, el componente de tendencia temporal captura las variaciones de las tasas de mortalidad a lo largo del tiempo, lo que facilita modelar las mejoras o el deterioro de las condiciones de salud de la población debido a factores como los avances médicos, los cambios en los hábitos de vida o las políticas de salud pública. Por último, el componente de variabilidad aleatoria refleja las fluctuaciones impredecibles en las tasas de mortalidad, causadas por eventos inesperados como epidemias, desastres naturales u otros factores aleatorios que afectan a la población de manera imprevisible.

El modelo Lee-Carter ha sido ampliamente adoptado debido a su capacidad para ofrecer proyecciones de mortalidad más realistas y ajustadas a la evolución temporal. Al separar los tres componentes esenciales, el modelo facilita una estimación más precisa y flexible de las tasas de mortalidad futuras, lo que es fundamental para la gestión de los riesgos en sectores como el asegurador y el de pensiones. Su implementación ha mejorado la precisión en la valoración de los pasivos en las aseguradoras, así como en la formulación de políticas de pensiones que dependen de las proyecciones de la esperanza de vida de los individuos.

2.1.2. EL COMPONENTE DE EDAD Y TIEMPO

Uno de los aspectos clave del modelo Lee-Carter es cómo maneja las diferencias por edad y la evolución temporal de las tasas de mortalidad. Las tasas de mortalidad no son constantes a lo largo del tiempo, ni homogéneas entre los diferentes grupos de edad. El modelo captura la variabilidad de las tasas de mortalidad, mediante la descomposición en dos componentes principales:

Componente de edad (αx): Este término refleja el nivel de mortalidad para cada grupo de edad en un periodo base. Debido a que la mortalidad suele aumentar con la edad, este término captura la estructura fundamental de la mortalidad por edad.

Componente temporal (Kt): Este componente captura la tasa de mortalidad como han cambiado a lo largo del tiempo para todas las edades. En términos prácticos, Kt puede reflejar mejoras en la salud pública o avances médicos. Es común que Kt muestre una tendencia decreciente a lo largo del tiempo, dada la mejora en la esperanza de vida global debido a los avances médicos y las condiciones de vida.

El modelo, entonces, no solo predice las tasas de mortalidad en un año determinado, sino que también descompone la mortalidad en estos dos componentes fundamentales, lo que permite modelar el comportamiento de la mortalidad de forma más robusta.

2.1.3. ESTIMACIÓN Y PROYECCIÓN DE LA MORTALIDAD FUTURA

El modelo Lee-Carter se enfoca en predecir la mortalidad futura. Para ello, utiliza la estimación de los parámetros αx , βx y Kt por medio de las series temporales históricas. Esto se hace utilizando técnicas estadísticas, especialmente la descomposición en valores singulares (*SVD*), que es un enfoque robusto para identificar las características más significativas de los datos.

Cuando se proyectan tasas de mortalidad futuras, se asume que los coeficientes de edad (αx) y los coeficientes de edad ajustados (βx) permanecen constantes a lo largo del tiempo. El parámetro más dinámico es (Kt) , que representa la tendencia general y se proyecta hacia el futuro.

Las proyecciones de mortalidad futuras son entonces una combinación de estos tres componentes.

2.1.4. IMPLICACIONES ACTUARIALES Y APLICACIONES

El modelo Lee-Carter tiene una relevancia significativa en la planificación de seguros y pensiones, especialmente en la predicción de la mortalidad, un factor fundamental para garantizar la solvencia y la sostenibilidad financiera de las instituciones que gestionan estos productos. Las proyecciones precisas de las tasas de mortalidad son esenciales para calcular de manera adecuada las reservas de seguros y para determinar las primas de los seguros de vida, asegurando que las aseguradoras puedan cumplir con sus obligaciones a largo plazo frente a los asegurados.

Al modelar de manera eficiente las tasas de mortalidad futuras, las aseguradoras pueden diseñar productos que reflejen con mayor precisión los riesgos asociados con la longevidad. Esto les permite ajustar las primas de los seguros de vida según el comportamiento proyectado de la mortalidad, teniendo en cuenta factores como los avances en la medicina, los cambios en el estilo de vida de la población, y las políticas de salud pública que podrían afectar la esperanza de vida. El modelo Lee-Carter ofrece una herramienta valiosa para predecir cómo evolucionarán las tasas de mortalidad a lo largo del tiempo, lo que es especialmente relevante para los seguros de vida, que se basan en la duración de la vida de los asegurados para determinar los pagos futuros.

En el ámbito de las pensiones, el modelo Lee-Carter también tiene un papel crucial. Las instituciones financieras que gestionan fondos de pensiones se benefician enormemente de la capacidad del modelo para prever la mortalidad futura. Este aspecto es fundamental para garantizar la sostenibilidad de los fondos, ya que permite a las entidades calcular las reservas técnicas necesarias para cubrir los pagos de pensiones durante los años venideros. En particular, el modelo ayuda a estimar la probabilidad de que las personas vivan más tiempo de lo esperado, lo que tiene un impacto directo en la cantidad de fondos necesarios para asegurar que los pagos de pensión se mantengan durante toda la vida del beneficiario.

Por lo tanto, la aplicación del modelo Lee-Carter en la planificación de seguros y pensiones no solo contribuye a una mayor precisión en la estimación de las reservas, sino que también mejora la capacidad de las aseguradoras y las instituciones financieras para gestionar los riesgos relacionados con la longevidad, protegiendo así tanto a las entidades como a los asegurados. Esto permite un diseño de productos más ajustado a las realidades demográficas y un mejor manejo de las expectativas futuras sobre la mortalidad, lo cual es esencial para la estabilidad y sostenibilidad a largo plazo de los sistemas de seguros y pensiones.

2.1.5. LIMITACIONES Y MEJORAS DEL MODELO

Aunque el modelo Lee-Carter ha demostrado ser una herramienta eficaz y ampliamente utilizada para la proyección de tasas de mortalidad, presenta algunas limitaciones importantes que deben ser consideradas al aplicarlo en la práctica. Una de las principales limitaciones es que el modelo asume que las tendencias de mortalidad por edad se mantendrán constantes a lo largo del tiempo. Esta suposición puede no reflejar adecuadamente la realidad, ya que las tasas de mortalidad por edad pueden cambiar debido a una serie de factores que no están completamente capturados en el modelo. Entre estos factores se incluyen los avances en la medicina, las mejoras en los hábitos de vida de la población, así como cambios en las políticas sociales y sanitarias, que podrían influir significativamente en la mortalidad en distintas edades.

Además, el modelo Lee-Carter se basa en una estructura lineal, lo que implica que no tiene la capacidad de modelar dinámicas complejas y no lineales que pueden ser relevantes para ciertas poblaciones o en periodos específicos de tiempo. En la práctica, las tasas de mortalidad pueden verse influenciadas por factores que siguen patrones no lineales, como crisis de salud pública, fenómenos epidemiológicos inesperados o fluctuaciones en las condiciones sociales y económicas que impactan de manera no uniforme a la población.

Para abordar estas limitaciones, se han propuesto diversas extensiones del modelo Lee-Carter que permiten capturar aspectos más complejos de la mortalidad. Una de estas extensiones es el modelo de Lee-Carter con cohortes, que incorpora la variabilidad de las tasas de mortalidad en función de cohortes específicas de edad, lo que permite un enfoque más detallado en las transiciones entre diferentes grupos de edad. Otra extensión es el modelo de efectos aleatorios, que introduce una mayor flexibilidad al permitir que las tasas de mortalidad varíen de manera más dinámica entre diferentes grupos de población, incorporando la incertidumbre y los factores imprevistos que podrían afectar a la mortalidad en ciertos contextos.

Además, la incorporación de factores estacionales en el modelo Lee-Carter podría mejorar su capacidad para capturar variaciones en la mortalidad que dependen de estaciones del año, condiciones climáticas o episodios de enfermedades estacionales, como la gripe, que pueden afectar a diferentes grupos de edad de manera desigual. Estas extensiones permiten un ajuste más preciso a las realidades cambiantes de las poblaciones y mejoran la capacidad del modelo para anticipar las tendencias de mortalidad de manera más fiel a los cambios sociales, económicos y médicos.

En resumen, aunque el modelo Lee-Carter ofrece una base sólida para la proyección de tasas de mortalidad, es importante reconocer sus limitaciones y considerar las posibles mejoras a través de extensiones que aborden la complejidad de las dinámicas no lineales y los factores externos que pueden influir en las tasas de mortalidad a lo largo del tiempo. Esto permitirá una mejor adaptación a los cambios y proporcionará estimaciones más precisas y relevantes para la toma de decisiones en el ámbito actuarial y de salud pública.

2.2. DESCRIPCIÓN MATEMÁTICA DEL MODELO LEE-CARTER

2.2.1 ECUACIÓN PRINCIPAL

El modelo Lee-Carter describe la tasa de mortalidad en función de la edad y el tiempo mediante la siguiente ecuación:

$$m(x,t) = e\text{^}(a\_x + k\_t * b\_x + \varepsilon\_\{x,t\}) \quad (1)$$

O en su versión logarítmica:

$$\ln(m(x,t)) = a\_x + k\_t * b\_x + \varepsilon\_\{x,t\} \quad (2)$$

2.2.2. DESCRIPCIÓN DE LOS PARÁMETROS

Donde:

- *m(x,t)* es la tasa de mortalidad a la edad x y el tiempo t.
- *a_x* es un parámetro que depende solo de la edad x, es decir, es un parámetro estático que refleja la tasa de mortalidad en una edad específica a través del tiempo.

- b_x es un índice que refleja cuánto varía la mortalidad con la edad. Este índice puede ser negativo, lo que indica una disminución en la mortalidad con la edad.
- k_t es el parámetro temporal que depende solo del tiempo. Es un parámetro dinámico que cambia a lo largo del tiempo y modela la evolución temporal de la mortalidad.
- $\varepsilon_{x,t}$ es un término de error, que representa la aleatoriedad no explicada por los parámetros anteriores.

Ecuación para k_t

El modelo Lee-Carter supone que el parámetro k_t, que captura las variaciones de la mortalidad a lo largo del tiempo, sigue un proceso de paseo aleatorio con deriva:

$$k_t = k_{t-1} + \delta + \varepsilon_t \quad (3)$$

Donde δ es una constante que representa la deriva, y ε_t es un término de error que sigue una distribución normal.

Desviación Estándar

Cuando se realiza la predicción de k_t, la desviación estándar aumenta con el tiempo, y se calcula como:

$$SE = \sigma \sqrt{t} \quad (4)$$

Esto implica que, a medida que avanzamos en el tiempo, la incertidumbre sobre la predicción aumenta, lo cual es típico en los modelos de mortalidad debido a la incertidumbre inherente a la predicción de fenómenos a largo plazo.

2.2.3. RESTRICCIONES DEL MODELO

Para poder ajustar el modelo, se imponen dos restricciones clave:

La suma de los b_x debe ser igual a 1:

$$\sum_{i=1}^{n} b_i = 1 \quad (5)$$

La suma de los k_t debe ser igual a 0:

$$\sum_{j=1}^{n} k_t = 0 \quad (6)$$

Estimación de los Parámetros del Modelo

El objetivo del modelo Lee-Carter es estimar los parámetros a_x, b_x, y k_t. Para ello, se minimiza la siguiente función de error:

$$f(a, b, k) = \sum\_\{x, t\}(\ln(m(x, t)) - a\_x - b\_x * k\_t)\hat{}2 \quad (7)$$

La minimización de esta función permite obtener las estimaciones óptimas de los parámetros a_x, b_x, y k_t.

Estimación de a_x

Para obtener a_x, se resuelve la siguiente ecuación bajo la restricción mencionada:

$$a\_x = (1/n) \sum\_\{t = t1\}\hat{}\{n\}\ln(m(x, t)) \quad (8)$$

Esto indica que a_x es simplemente el promedio logarítmico de las tasas de mortalidad a lo largo del tiempo para cada edad x. Este parámetro describe cómo cambia la mortalidad para una edad dada.

2.2.4. DESCOMPOSICIÓN EN VALORES SINGULARES (*SVD*)

Para obtener las estimaciones de b_x y k_t, se utiliza un método conocido como Descomposición en Valores Singulares (*SVD*). Este método es útil para factorizar matrices y se usa en el contexto del modelo Lee-Carter para resolver el sistema de ecuaciones.

La matriz A se define como:

$$A = \ln(m(x, t) - \hat{a}\_x) \quad (9)$$

La descomposición en valores singulares de la matriz A se da como:

$$A = U \Sigma V\hat{}T \quad (10)$$

Donde:

U es una matriz ortogonal de dimensiones m × n.

Σ es una matriz diagonal que contiene los valores singulares.

V es una matriz ortogonal de dimensiones n × n.

La descomposición en valores singulares permite obtener las estimaciones de b_x y k_t como:

$$SVD(A) = U \, \Sigma \, V^\wedge T = \hat{b}\_x * \hat{k}\_t \quad (11)$$

2.3. MODELOS DE MORTALIDAD: APC Y LC-BIFACTORIAL

2.3.1. MODELO AGE-PERIOD-COHORT (APC)

Enfoque utilizado en análisis demográficos y epidemiológicos para descomponer la variabilidad de una variable de interés (como tasas de mortalidad o prevalencia de enfermedades) en tres componentes:

La fórmula general del modelo Age-Period-Cohort (APC) en su versión logit es la siguiente:

$$logit(p\_\{x,t\}\ ) = \alpha\_x + \beta\_t + \gamma\_c \quad (12)$$

Donde:

- p_{x,t} es la probabilidad de la variable dependiente (por ejemplo, la tasa de mortalidad) para la edad x y el período t.
- α_x es el efecto de la edad para la edad x.
- β_t es el efecto del período para el período t.
- γ_c es el efecto de la cohorte, que está asociado con la cohorte de personas nacidas en el período c, que se relaciona con la edad x y el período t.

Este modelo descompone los efectos observados de la variable dependiente en tres componentes:

- Edad (Age): Efectos relacionados con la edad de los individuos.
- Período (Period): Efectos que afectan a todos los individuos en un momento determinado (por ejemplo, eventos históricos o cambios en políticas).

- Cohorte (Cohort): Efectos relacionados con los grupos de individuos nacidos en un mismo período de tiempo (cohorte), que experimentan las mismas condiciones durante su vida.

2.3.2. MODELO LC-BIFACTORIAL (LC-BIFACTORIAL)

El modelo LC-Bifactorial es una extensión del modelo Lee-Carter tradicional. La principal diferencia radica en la introducción de restricciones adicionales sobre los parámetros del modelo para mejorar el ajuste y la capacidad de predicción.

Diferencias principales:

Restricciones adicionales: El modelo LC-BIFACTORIAL introduce restricciones específicas en los parámetros b(x) y k(t) que no están presentes en el modelo LC básico.

La estructura general del modelo LC-BIFACTORIAL es muy similar a la del modelo Lee-Carter, pero con los ajustes de restricciones mencionados. El modelo básico sigue siendo:

$$\log\left(q(x,t)\right) = a(x) + b(x) * k(t) \quad (13)$$

Estas restricciones buscan controlar o suavizar el ajuste de los parámetros y asegurar que los resultados sean más estables o realistas.

Restricción sobre b(x): Se impone una condición para que la suma de los parámetros b(x) sea igual a 1. Esta restricción garantiza que la contribución de b(x) no sea arbitraria y que esté más alineada con las tendencias observadas en los datos.

Restricción sobre k(t): Se ajusta el parámetro k(t) para que su suma sea cero, lo que implica que no haya una tendencia neta hacia el aumento o disminución de la mortalidad a lo largo del tiempo, sino que los efectos temporales estén equilibrados en torno a un valor medio.

Respecto al modelo Lee Carter, el modelo LC-BIFACTORIAL aporta mayor control sobre los parámetros; las restricciones impuestas en el modelo permiten tener un mayor control sobre la influencia de los efectos de la edad y el periodo de mortalidad; mejor ajuste a datos complejos, adaptarse mejor a los datos de mortalidad, en concreto, a las fluctuaciones significativas de la tasa de mortalidad, y por último, estabilidad en el ajuste de los parámetros, lo que permite mayor robustez.

2.4. LEYES DE MORTALIDAD ESTÁTICAS

2.4.1. LEY DE MOIVRE

La Ley de Moivre es una de las leyes más antiguas utilizadas para modelar la mortalidad en función de la edad. Esta ley describe la función de supervivencia en términos de una relación simple y lineal entre la edad y la mortalidad.

Este modelo, sin embargo, no toma en cuenta muchos factores dinámicos, como las variaciones en la mortalidad causadas por enfermedades, cambios en el estilo de vida, o avances médicos, lo que limita su aplicabilidad a ciertos contextos.

Fórmula de la Ley de Moivre:

$$lx = l\_0\,(1 - x/w) \quad (14)$$

- lx es el número de individuos que sobreviven hasta la edad x.
- l_0 es el número inicial de individuos en el grupo (típicamente, la población al nacer).
- x es la edad en años.
- w es una constante que determina la pendiente de la mortalidad; a menudo se interpreta como la edad máxima alcanzable o el valor límite para la mortalidad en la población.

Interpretación:

- Supervivencia con la Edad: La Ley de Moivre establece que la probabilidad de supervivencia lx disminuye linealmente con la edad x, alcanzando su límite en la edad w. En otras palabras, la mortalidad aumenta a medida que la persona envejece, de acuerdo con una tasa constante por cada año de vida.
- Edad Máxima: La constante w representa la edad máxima que una persona puede alcanzar según el modelo. A medida que x se aproxima a w, el número de sobrevivientes lx tiende a ser cero, lo que significa que, según este modelo, todos los individuos mueren a la edad w.

Limitaciones:

- La Ley de Moivre es una simplificación que asume una mortalidad lineal. Sin embargo, los datos reales muestran que la mortalidad no siempre sigue un patrón lineal y suele acelerarse con la edad en etapas más avanzadas.
- No captura la variedad de la mortalidad en función de otros factores como el género, el comportamiento social, la clase social, o el entorno.

2.4.2. LEY DE DORMOY (VERSIÓN 1)

La Ley de Dormoy (Versión 1) es una extensión de la Ley de Moivre, y se utiliza para modelar la mortalidad de manera más flexible. La principal diferencia con la Ley de Moivre es que introduce una constante de reducción de la mortalidad, denotada como S. Esta constante ajusta la tasa de mortalidad a lo largo de la edad x, reflejando una disminución en la mortalidad a medida que envejece la población.
Dicha ley puede usarse para modelar la mortalidad ajustada de una población bajo la suposición de que la mortalidad disminuye a medida que los individuos envejecen.
Fórmula de la Ley de Dormoy (Versión 1):

$$lx = l0 \cdot S\^{}x$$

Interpretación:
La Ley de Dormoy (Versión 1) asume que la mortalidad disminuye a una tasa constante según la constante S a medida que avanza la edad.
Si S es mayor que 1, la mortalidad aumenta con la edad; si es menor que 1, la mortalidad disminuye con la edad.

2.4.3. LEY DE DORMOY (VERSIÓN 2)

La Ley de Dormoy (Versión 2) ajusta la fórmula de la versión anterior, incorporando una segunda constante S2, lo que permite modelar una tasa de mortalidad que disminuye de forma más compleja con la edad.

Esta versión introduce una dependencia cuadrática de la edad, lo que hace que la mortalidad disminuya más rápidamente en edades avanzadas, también es útil, para simular la mortalidad de poblaciones que experimentan una disminución acelerada de la tasa de mortalidad a medida que avanzan en la edad.

Fórmula de la Ley de Dormoy (Versión 2):

$$lx = l0 \cdot S1\^{}x * S2\^{}(x\^{}2\ )$$

Interpretación:
La fórmula introduce un término cuadrático S2^(x^2), lo que implica que el efecto de la

edad en la mortalidad es más significativo en edades mayores. Este modelo es útil cuando se desea una reducción de la mortalidad que disminuye a un ritmo acelerado conforme se incrementa la edad.

2.4.4. LEY DE SHANG

La Ley de Shang es una ley de mortalidad que se utiliza para modelar la mortalidad con un enfoque que incorpora un factor adicional b.

Esta ley tiene una estructura matemática más compleja y se aplica a modelos donde la mortalidad se ve afectada por la variabilidad en la población o en los individuos.

Fórmula de la Ley de Shang:

$$lx = l0 \,/\, (1 - b\text{^}w) \cdot (b\text{^}x - b\text{^}w)$$

Interpretación:

- La Ley de Shang introduce un factor b que permite un mayor control sobre el comportamiento de la mortalidad, lo que la hace más flexible que otros modelos.
- La fórmula refleja un incremento o decremento exponencial de la mortalidad dependiendo del valor de b y su influencia en la edad x.

2.4.5. LEY DE GOMPERTZ

La Ley de Gompertz es un modelo clásico en la teoría de la mortalidad que describe el aumento exponencial de la mortalidad con la edad. Este modelo es ampliamente utilizado en actuarialismo y biología para modelar la mortalidad de individuos y tiene una forma matemática que refleja un crecimiento exponencial acelerado de la mortalidad a medida que los individuos envejecen.

Fórmula de la Ley de Gompertz:

$$lx = l0 \cdot g\text{^}(C\text{^}x - 1)$$

Interpretación:

La Ley de Gompertz establece que la mortalidad aumenta de manera exponencial con la edad, y lo hace con una tasa que se acelera conforme los individuos envejecen.
Este modelo refleja un comportamiento no lineal en el que la mortalidad se incrementacon el paso del tiempo de manera mucho más pronunciada en edades avanzadas.

2.5. MACHING LEARNING

El Machine Learning, se enmarca dentro del campo de la Inteligencia Artificial y engloba, una serie de técnicas y algoritmos utilizados para modelar patrones en los datos.

Existen diferentes enfoques dentro del Machine Learning, siendo los más comunes el aprendizaje supervisado, aprendizaje no supervisado y aprendizaje semisupervisado.

En el aprendizaje supervisado, el modelo se entrena con datos etiquetados, lo que significa que, para cada entrada, se tiene un valor de salida o respuesta conocido. En este contexto, los algoritmos buscan aprender la relación entre las variables de entrada y la variable de salida. Un ejemplo clásico de este tipo de aprendizaje es la regresión lineal, cuyo objetivo es predecir un valor numérico, y la regresión logística, que se utiliza cuando la variable de respuesta es categórica.

El aprendizaje no supervisado, por otro lado, trabaja con datos en los que no se conocen las etiquetas, es decir, no se sabe si la variable de salida es numérica o categórica.

El objetivo de este enfoque es identificar patrones o estructuras subyacentes en los datos. Un ejemplo es el Análisis de Componentes Principales (PCA) (Hotelling, 1933), que busca reducir la dimensionalidad de los datos representando la información en un menor número de variables construidas como combinaciones lineales de las variables originales.

Otro algoritmo destacado es el clustering o análisis de conglomerados, que agrupa los datos en función de características similares (Peña et al., 2013).

Finalmente, el aprendizaje semisupervisado es una combinación de ambos enfoques. Este tipo de aprendizaje emplea tanto datos etiquetados como no etiquetados, lo que permite construir modelos más robustos aprovechando los datos disponibles de manera más eficiente.

2.5.1. REGRESIÓN LINEAL

La regresión lineal es un algoritmo dentro del campo del Machine Learning y corresponde a un tipo de aprendizaje supervisado, en el que las variables están etiquetadas.

Para este trabajo, y considerando la definición de la tasa central de mortalidad, que corresponde al número de defunciones ocurridas para una edad determinada en una población, se utilizará la regresión lineal.

A diferencia del método de Lee-Carter, donde no se podía aplicar regresión debido a la falta de variables explicativas, en este caso, se tratará de explicar el valor de m_x a partir de algunas variables explicativas.

2.5.2. REGRESIÓN LINEAL SIMPLE

Este tipo de modelo establece una relación entre una variable de respuesta y una sola variable explicativa, siguiendo la siguiente ecuación (Montgomery, 2002):

$$y\_i = \beta\hat{}0 + \beta\hat{}1\, x\_i + \varepsilon\_i$$

Donde:

- y_i es la variable de respuesta,
- x_i es la variable explicativa,
- β_0 y β_1 son los coeficientes de la ecuación de la recta (conocidos como coeficientes de la regresión)
- ε_i es el término de error.

El objetivo es estimar los valores de β_0 y β_1 mediante el método de mínimos cuadrados, que consiste en minimizar la diferencia entre los valores estimados y los reales. Esto se logra mediante la siguiente función:

$$S(\beta\hat{}0,\beta\hat{}1) = \sum(y\_i - \beta\hat{}0 - \beta\hat{}1\, x\_i)\hat{}2$$

Al derivar la función respecto a β_0 y β_1, obtenemos las siguientes ecuaciones que nos permiten encontrar los valores de los coeficientes:

$$S/(\partial\beta\hat{}0) = -2\sum(y\_i - \beta\hat{}0 - \beta\hat{}1\, x\_i) = 0$$
$$\partial S/(\partial\beta\hat{}1) = -2\sum(y\_i - \beta\hat{}0 - \beta\hat{}1\, x\_i)\, x\_i = 0$$

Al resolver estas ecuaciones, obtenemos los valores de β_0 y β_1:

$$\hat{\beta}\hat{}0 = \bar{y} - \hat{\beta}\hat{}1\, \bar{x}$$

$$\hat{\beta}\hat{}1 = [\sum(y\_i\, x\_i) - (\sum y\_i)(\sum x\_i)/\, n]/\, [x\sum\_i\hat{}2 - (\sum x\_i)\hat{}2/\, n]$$

Donde:

- $\bar{y}$ es el promedio de la variable de respuesta,
- $\bar{x}$ es el promedio de la variable explicativa.

2.5.3. REGRESIÓN LINEAL MÚLTIPLE

Una vez explicada la regresión lineal simple, se detallará cómo se calculan los coeficientes para más dimensiones. Debido al mayor número de coeficientes, se realiza un análisis de su significatividad. Además, dado que habrá más variables explicativas, se comprobará si existen problemas de multicolinealidad y el método necesario para detectarlos y mejorar el modelo.
Para este tipo de regresiones, se emplea también el método de mínimos cuadrados, pero las ecuaciones son las siguientes (Aldás Manzano y Uriel Jiménez, 2017):

$$y\_i = \beta\hat{}0 + \beta\hat{}1\, x\hat{}1\, i + \beta\hat{}2\, x\hat{}2\, i + \dots + \beta_k\, x\_ki + \varepsilon\_i$$

La función de mínimos cuadrados es:

$$S(\beta\text{^}0, \beta\text{^}1, \dots, \beta_k) = \sum(\varepsilon\_i\ )\text{^}2 = \sum(y\_i - \beta\text{^}0 - \sum\beta_k\, x\_ijk\ )\text{^}2$$

El proceso de derivación y minimización se realiza de manera similar al caso de la regresión lineal simple, utilizando notación matricial:

$$y = X\beta + \varepsilon$$

Donde:

- y es un vector n × 1,
- X es una matriz n × k
- β es un vector k × 1.

La función de mínimos cuadrados en términos matriciales es:

$$S(\beta) = y\text{^}'\, y - 2\beta\text{^}'\, X\text{^}'\, y + \beta\text{^}'\, X\text{^}'\, X\beta$$

Al derivar respecto a los coeficientes e igualar a cero, obtenemos:

$$X\text{^}'\, X\beta\hat{} = X\text{^}'\, y$$

Finalmente, para resolver esta ecuación, se multiplica por la inversa de X'X a ambos lados, obteniendo el vector de coeficientes:

$$\beta\hat{} = (X\text{^}'\, X)\text{^}(-1)\, X\text{^}'\, y$$

2.6. TEORÍA DE XGBOOST (EXTREME GRADIENT BOOSTING)

Extreme Gradient Boosting es una librería optimizada que implementa el algoritmo de Gradient Boosting, un potente método de aprendizaje automático supervisado utilizado en tareas de regresión y clasificación. Su principal ventaja radica en su velocidad y precisión, lo que lo ha convertido en una de las herramientas más populares en el ámbito de la ciencia de datos. A continuación, se detallan los conceptos clave que fundamentan XGBoost.

2.6.1. INTRODUCCIÓN A GRADIENT BOOSTING

Para comprender XGBoost, es fundamental entender primero los conceptos de boosting y gradient boosting, que son la base de este algoritmo.

2.6.2. BOOSTING:

- Boosting es una técnica de ensamblaje que combina múltiples modelos débiles (modelos que, por sí solos, tienen un rendimiento bajo, como los árboles de decisión) para crear un modelo fuerte y preciso.
- La principal característica del boosting es que los modelos se entrenan de manera secuencial, es decir, cada nuevo modelo intenta corregir los errores cometidos por los modelos anteriores. Este proceso se conoce como aprendizaje secuencial.
- En cada iteración, el modelo posterior se enfoca en predecir los errores o residuos del modelo anterior. De esta manera, el modelo va aprendiendo de los errores previos y mejora progresivamente su rendimiento en cada paso.
- El objetivo del boosting es reducir los errores del modelo global y mejorar la capacidad de predicción al combinar de manera efectiva modelos simples que individualmente tienen un bajo desempeño.

2.6.3. GRADIENT BOOSTING:

Gradient Boosting es una variante de boosting en la que los modelos (típicamente árboles de decisión) se entrenan de manera secuencial para minimizar una función de pérdida.

A diferencia de otros métodos de boosting, donde cada modelo puede predecir los errores directamente, en gradient boosting se ajusta un modelo a los residuos de los modelos anteriores utilizando gradiente descendente.
El gradiente descendente es una técnica de optimización que se utiliza para minimizar la función de pérdida, ajustando de manera iterativa los parámetros del modelo en la dirección que reduce los errores.
La idea central de gradient boosting es utilizar el gradiente de la función de pérdida para ajustar el siguiente modelo de manera que se mejore el rendimiento general. Este enfoque

hace que el modelo sea eficiente al enfocarse en los errores cometidos en las iteraciones anteriores y mejorando el ajuste en cada paso.

2.6.4. APLICACIÓN DE XGBOOST

En el ámbito de las ciencias actuariales, XGBoost se aplica principalmente para modelizar variables continuas, como la mortalidad, que es fundamental para la fijación de primas de seguros, la estimación de reservas de seguros y la planificación de fondos de pensiones. Las tasas de mortalidad ajustadas (*qxt*) y otras métricas actuariales son esenciales para estos cálculos, y XGBoost ofrece una forma eficiente de predecirlas con alta precisión.

Uno de los usos más comunes de XGBoost es la predicción de tasas de mortalidad a partir de características como la edad, el sexo, el año de nacimiento, entre otros factores. Para modelar la mortalidad, las aseguradoras deben tener en cuenta cómo evoluciona la esperanza de vida y las tasas de mortalidad de las personas a lo largo del tiempo.

¿Cómo se implementa XGBoost en este contexto?

Variables de Entrada:
En el contexto actuarial, las variables de entrada (características o *features*) pueden incluir:

- **Edad** del asegurado.
- **Año** de referencia.
- **Factores de salud** como la tasa de mortalidad bruta (*mxt*), el nivel de ingresos, o el historial de enfermedades.

Tasa de Mortalidad Ajustada (*qxt*):
El objetivo del modelo es predecir la tasa de mortalidad ajustada (***qxt***) en función de estas variables. Esta tasa se calcula teniendo en cuenta la probabilidad de fallecimiento de un grupo de personas en función de sus características (por ejemplo, edad, género, condiciones de salud).

Entrenamiento del Modelo:

XGBoost entrenará secuencialmente una serie de árboles de decisión, cada uno de los cuales ajustará los errores (residuos) cometidos por los árboles anteriores, utilizando el gradiente descendente para minimizar la función de pérdida. Cada árbol en el modelo tiene un impacto mayor en la predicción al corregir los errores acumulados por los modelos anteriores.

Optimización:

XGBoost optimiza el proceso de entrenamiento a través de diversas mejoras, como la **regularización** (para evitar el sobreajuste), el uso de **muestras aleatorias** (para mejorar la generalización del modelo) y la **paralelización** (que permite entrenar el modelo más rápidamente).

Predicción:

Una vez entrenado, el modelo de XGBoost realiza predicciones sobre las tasas de mortalidad ajustadas para el conjunto de prueba, utilizando las mismas variables de entrada. Estas predicciones pueden ser comparadas con las tasas reales de mortalidad para evaluar la precisión del modelo mediante métricas como el Error Cuadrático Medio (RMSE) y el R^2.

2.6.5. BENEFICIOS DE XGBOOST EN ESTE CONTEXTO:

Precisión y Eficiencia:

XGBoost es capaz de manejar grandes volúmenes de datos con alta precisión y rapidez, lo que es esencial en la modelización actuarial, donde se manejan grandes bases de datos de asegurados.

Flexibilidad:

XGBoost puede manejar tanto datos lineales como no lineales, lo que es útil para modelar relaciones complejas entre las características y las tasas de mortalidad. Esto es importante, ya que la mortalidad no sigue patrones estrictamente lineales y puede estar influenciada por múltiples factores interrelacionados.

Reducción de Overfitting:

A través de su mecanismo de regularización, XGBoost reduce el riesgo de sobreajuste (overfitting), lo cual es crucial en el análisis actuarial, ya que un modelo sobreajustado podría no generalizar bien a nuevos datos.

Interpretabilidad:
Aunque XGBoost es un modelo complejo, los resultados pueden ser interpretados a través de herramientas de importancia de variables que permiten entender cómo influyen las diferentes características en las predicciones de mortalidad. Esto es fundamental para las aseguradoras que requieren transparencia en los modelos utilizados para fijar las primas y las reservas.

Mejor Toma de Decisiones Actuariales:
Con modelos más precisos, las aseguradoras pueden ajustar las primas y las reservas de manera más precisa, lo que mejora la estabilidad financiera de las compañías de seguros y permite ofrecer productos más adaptados a las necesidades de los asegurados.

XGBoost, al ser una implementación optimizada del algoritmo de gradient boosting, proporciona una herramienta poderosa y eficiente para la modelización de tasas de mortalidad ajustadas en el campo actuarial. Gracias a su capacidad para manejar datos complejos y no lineales, y su eficiencia en el entrenamiento, se convierte en una herramienta clave en la predicción de variables críticas para la fijación de primas de seguros, la estimación de reservas y la planificación de fondos de pensiones. Su uso en el ámbito actuarial contribuye a mejorar la precisión en la gestión del riesgo y la sostenibilidad de los productos aseguradores.

2.7. GRADUACIÓN KERNEL Y ESTIMACIÓN DE COEFICIENTES EN MODELOS ACTUARIALES DE MORTALIDAD

2.7.1. GRADUACIÓN KERNEL

Es una técnica ampliamente utilizada en el campo actuarial para suavizar datos, como las tasas de mortalidad observadas, a fin de obtener estimaciones más precisas de las tendencias subyacentes. A continuación, se presentan los principales estimadores utilizados en este proceso y su aplicación práctica.

2.7.2. ESTIMADOR DE NADARAYA-WATSON

Esta función asigna mayor peso a las observaciones cercanas a la edad de interés, lo que permite que las estimaciones sean más sensibles a los datos cercanos en lugar de a los valores extremos. El método se basa en el concepto de ancho de banda (bandwidth), el cual controla la extensión del suavizado. Un ancho de banda más grande implica un suavizado más pronunciado, mientras que un ancho de banda pequeño permite que el modelo se ajuste más a los datos locales.

Este estimador minimiza los cuadrados medios locales, lo que ayuda a ajustar las tasas de mortalidad de manera precisa a lo largo de las edades, reflejando con mayor exactitud las variaciones que pueden no ser evidentes en los datos brutos.

2.7.3. ESTIMADOR DE COPAS-HABERMAN

En lugar de minimizar los cuadrados medios locales, este estimador minimiza la log-verosimilitud local binomial, lo que lo hace más robusto frente a fluctuaciones en los datos. Este enfoque es particularmente útil en contextos donde la precisión de las tasas de mortalidad es crucial, ya que permite una estimación más exacta en presencia de ruido o datos atípicos. La principal ventaja de este método es que puede proporcionar estimaciones más estables, especialmente cuando se enfrentan a variaciones significativas en los datos.

2.7.4. APLICACIÓN DE LA GRADUACIÓN KERNEL

En la práctica, la graduación kernel puede considerarse una media móvil ponderado, ya que da mayor peso a las observaciones cercanas, suavizando los datos y ajustándolos de manera que las tasas de mortalidad sean más representativas a lo largo de diferentes edades. Este enfoque mejora la precisión del modelo, permitiendo que las predicciones sean más fiables y ajustadas a la realidad observada.

2.7.5. PROBLEMAS DEL ESTIMADOR KERNEL

A pesar de su flexibilidad y efectividad, el estimador kernel presenta algunas limitaciones. En particular, tiende a infraestimar las tasas de mortalidad en edades altas, donde la curva de mortalidad es típicamente cóncava, y sobrestima las tasas de mortalidad en edades bajas, donde la curva es más convexa.

Estos problemas surgen debido a la forma en que la función kernel asigna peso a las observaciones cercanas. Para mitigar estos problemas, se puede ajustar el ancho de banda o utilizar técnicas adicionales de suavización que permitan un mejor ajuste a las características específicas de los datos.

2.7.6. ESTIMACIÓN DE LOS COEFICIENTES MEDIANTE EL MÉTODO DE MÍNIMOS CUADRADOS

El proceso de estimación de los coeficientes en un modelo de regresión es crucial para entender cómo las variables explicativas afectan a la variable dependiente. Una vez realizado el ajuste del modelo de regresión, es necesario llevar a cabo una serie de verificaciones para asegurar que la estimación se ha realizado correctamente.

Los errores en la predicción, definidos como la diferencia entre los valores observados y los valores ajustados, deben seguir una distribución normal, es decir, tener una media igual a cero y una varianza constante en todo el conjunto de datos. Además, debe existir independencia entre los errores, es decir, los errores deben no estar correlacionados entre sí.

Esto se considera un supuesto fundamental para muchos modelos de regresión.
Para comprobar que los residuos siguen una distribución normal con media cero y varianza constante, se aplican diferentes pruebas de bondad de ajuste, como el test de Kolmogorov-Smirnov (Massey Jr, 1951).

Este test evalúa la distancia entre la función de distribución empírica de los residuos y la función de distribución esperada bajo una distribución normal. Si la distancia entre ambas distribuciones es pequeña, no se rechaza la hipótesis nula de normalidad. En caso contrario, si la distancia es significativa, se rechaza la hipótesis de normalidad de los residuos, lo que indica que el modelo puede no ser adecuado.

Además, se debe verificar si los residuos presentan independencia. Esto es importante para asegurarse de que no haya patrones de autocorrelación no capturados por el modelo.

Para ello, se utiliza el test de Ljung-Box (Ljung y Box, 1978), que evalúa si existe autocorrelación en los residuos a diferentes rezagos. Si los residuos son independientes, el test no debería rechazar la hipótesis nula de independencia.
Por último, es necesario comprobar si la media de los residuos es igual a cero. Esta verificación es fundamental porque, si los residuos tienen una media distinta de cero, el modelo podría estar sesgado.

Para ello, se realiza un contraste de hipótesis bilateral bajo la suposición de que los residuos provienen de una distribución normal con media cero. Se rechazaría la hipótesis nula si el estadístico de prueba ZZZ se encuentra en la región de rechazo, es decir:

$$| Z | \geq Z\alpha/2, siendo\ \alpha\ el\ nivel\ de\ significancia.$$

Este tipo de contraste nos permite determinar si el modelo está correctamente especificado y si no existe sesgo en la estimación de los coeficientes.

2.7.7. ANÁLISIS DE LOS COEFICIENTES DE REGRESIÓN

Una vez estimados los coeficientes de regresión, se deben realizar análisis estadísticos para evaluar la significatividad de cada uno de ellos. Para ello, se emplea el estadístico t, que permite realizar una prueba de hipótesis sobre la relevancia de los coeficientes en el modelo.

La hipótesis nula en este caso es que el coeficiente es igual a cero, lo que indica que la variable explicativa correspondiente no tiene un efecto significativo sobre la variable dependiente.

El estadístico t0 se calcula dividiendo el valor estimado del coeficiente β^1 entre su error estándar SE(β^1):

$$t0 = \beta\hat{}1/SE(\beta\hat{}1\ )$$

En el caso de una regresión lineal simple, se realiza un contraste de hipótesis bilateral. La hipótesis nula es que el coeficiente es igual a cero, lo que implicaría que la variable explicativa no tiene ningún efecto sobre la variable dependiente. La hipótesis alternativa es que el coeficiente es distinto de cero, lo que indicaría que la variable explicativa tiene un efecto significativo.

Las hipótesis se formulan de la siguiente manera:

$$H0{:}\ \beta j\ =\ 0\ (\text{sin}\ significancia\ )$$

$$H1{:}\ \beta j\ \neq\ 0\ (con\ significancia)$$

Para que se rechace la hipótesis nula, el valor absoluto del estadístico t0 debe ser mayor que el valor crítico tα/2, n−2, que depende del nivel de significancia α y del número de observaciones en el conjunto de datos.

En una regresión lineal múltiple, el análisis de los coeficientes se realiza a través de un contraste de hipótesis global. La hipótesis nula en este caso es que todos los coeficientes de la regresión son cero, lo que implicaría que ninguna de las variables explicativas tiene un efecto significativo sobre la variable dependiente.

La hipótesis alternativa es que al menos uno de los coeficientes es distinto de cero. Para evaluar esta hipótesis, se utiliza el estadístico F0, que se calcula de la siguiente manera:

$$F0\ = ((\Sigma i = 1n\ (\hat{y}i\ -\ \bar{y}\ )\hat{}2/\ k))/\ (\Sigma i = 1n\ (yi\ -\ \hat{y}i)\hat{}2/\ (n\ -\ k\ -\ 1))$$

Donde:

- ŷi es el valor predicho para la observación i
- ȳ es el valor promedio de la variable dependiente.

El valor F0 se compara con el valor crítico Fα, k, n−k−1, y si F0 es mayor que este valor, se rechaza la hipótesis nula y se concluye que al menos uno de los coeficientes es significativamente distinto de cero.

2.7.8. MULTICOLINEALIDAD EN REGRESIONES MÚLTIPLES

Un problema común en las regresiones lineales múltiples es la “multicolinealidad”, que se presenta cuando las variables explicativas están altamente correlacionadas entre sí.

Esto puede generar estimaciones imprecisas de los coeficientes y afectar la fiabilidad del modelo. Cuando existe multicolinealidad, los coeficientes de regresión pueden volverse inestables, lo que dificulta la interpretación del modelo y puede inflar los errores estándar de los coeficientes.

Para detectar la multicolinealidad, se utiliza el VIF (Variance Inflation Factor), un parámetro que mide la inflación de la varianza de los coeficientes debido a la colinealidad entre las variables explicativas.

El VIF se calcula para cada variable explicativa en el modelo y se define como:

$$VIFj = 1 / (1 - Rj\text{^}2)$$

donde Rj^2 es el coeficiente de determinación de la variable j frente al resto de las variables explicativas.

Un valor alto de VIF (mayor a 5 o 10) indica que la variable j está causando problemas de multicolinealidad en el modelo. En estos casos, se recomienda eliminar la variable con el mayor VIF y volver a estimar el modelo.

Además, los problemas de multicolinealidad pueden dificultar la interpretación del modelo y aumentar el error estándar de los coeficientes, lo que afecta la precisión de las predicciones.

Para corregir este problema, se puede eliminar una de las variables correlacionadas o aplicar técnicas de regularización como “Ridge Regression” o “Lasso Regression”, que penalizan la magnitud de los coeficientes y ayudan a reducir la multicolinealidad.

Finalmente, se evalúa la calidad del ajuste del modelo utilizando el “R^2”, que indica qué tan bien el modelo explica la variabilidad de la variable dependiente.

Un valor alto de R^2 sugiere que el modelo ajusta bien los datos, mientras que un valor bajo sugiere que el modelo no captura adecuadamente las relaciones entre las variables explicativas y la variable dependiente.

2.7.8. ANÁLISIS DE INFLUENCIA Y DIAGNÓSTICO DE MODELOS

Una técnica esencial en la validación de modelos estadísticos, particularmente en el contexto de modelos de mortalidad, donde se busca garantizar que las predicciones no estén distorsionadas por valores atípicos o puntos de datos que tengan un impacto desproporcionado en los resultados.
Este análisis ayuda a evaluar si ciertos datos están ejerciendo una influencia demasiado fuerte sobre los coeficientes estimados del modelo y, por ende, sobre las predicciones generadas.

2.7.9. ANÁLISIS DE INFLUENCIA

Tiene como objetivo identificar aquellas observaciones en los datos que tienen un impacto excesivo en la estimación de los coeficientes del modelo. Esto es crucial en la modelización actuarial de mortalidad, donde datos atípicos o errores de medición pueden distorsionar los resultados, llevando a conclusiones incorrectas o poco fiables.

Una observación es influyente si, al eliminarla o modificarla, se producen cambios significativos en los coeficientes del modelo y, por lo tanto, en los resultados y las predicciones. Este tipo de observaciones pueden ser outliers (valores extremos) o leverage points (puntos que tienen una posición especial en el espacio de las variables predictoras y pueden arrastrar la línea de regresión o la curva de mortalidad).

2.7.10. HERRAMIENTAS COMUNES PARA EL ANÁLISIS DE INFLUENCIA

Existen varias herramientas y técnicas para llevar a cabo un análisis de influencia en modelos estadísticos, siendo dos de las más utilizadas el diagrama de residuos y el cociente de influencia.

2.7.11. DIAGRAMA DE RESIDUOS

Es una herramienta visual que permite identificar patrones en los residuos (la diferencia entre los valores observados y las predicciones del modelo). En un modelo bien ajustado, los residuos deben ser distribuidos aleatoriamente alrededor de cero sin mostrar ninguna tendencia sistemática.

- Si los residuos de un punto de datos están significativamente alejados del valor cero, esto indica que ese punto podría ser un outlier que está ejerciendo una influencia excesiva en el modelo.

- Un patrón sistemático en los residuos podría indicar que el modelo no está capturando adecuadamente ciertas relaciones en los datos, lo que sugiere la necesidad de ajustar el modelo o incluso de considerar nuevas variables.

3. BASE DE DATOS

Para la realización de este trabajo se ha utilizado la base de datos pública de Human Mortality Database. Esta base de datos ofrece información detallada y actualizada sobre la mortalidad de la población española desde 1908 hasta 2020, abarcando tanto datos de mortalidad general como específicos para distintos grupos de edad y sexo.

La principal motivación es desarrollar una metodología común para el cálculo de las tasas de mortalidad, empleando las tasas específicas de mortalidad, como mx, correspondientes a cada grupo de edad. En este caso, se han considerado las tasas de mortalidad tanto para hombres como para mujeres, lo que permite un análisis más profundo de la mortalidad en la población española a lo largo del tiempo.

A partir de estos datos, se empleará el modelo Lee-Carter, utilizando el logaritmo de la tasa de mortalidad mx para analizar las tendencias de mortalidad y su evolución a lo largo de los años. Este modelo se ha complementado con técnicas de regresión lineal múltiple, que permiten comparar las predicciones generadas por los modelos de series temporales con las obtenidas mediante regresión.

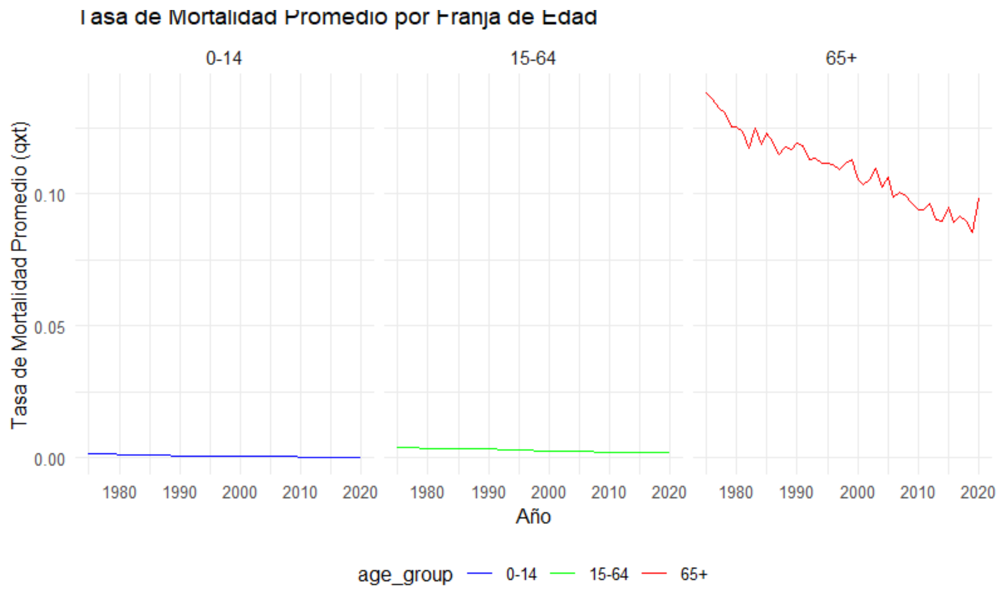
Por otro lado, la calidad del dato es un aspecto fundamental en cualquier análisis predictivo, especialmente cuando se trabajan con bases de datos históricas como la Human Mortality Database. Esta base de datos proporciona información detallada sobre la mortalidad de la población española desde 1975 hasta 2020, lo que permite realizar estudios exhaustivos sobre las tasas de mortalidad. Sin embargo, es importante considerar varios factores que pueden afectar la calidad de los datos.

En primer lugar, la cobertura temporal es clave; la base de datos cubre un periodo largo, pero ciertos años históricos, como aquellos afectados por guerras o pandemias, pueden introducir fluctuaciones que no representan la tendencia general de mortalidad. Para abordar esto, se ha decidido descartar años específicos que podrían generar volatilidad en los resultados.
Otro aspecto relevante es la homogeneidad de los datos, ya que, a lo largo del tiempo, los avances médicos y los cambios demográficos pueden hacer que los datos de mortalidad no sean directamente comparables entre diferentes periodos. Además, los cambios en las metodologías de recopilación de datos también pueden introducir cierta inconsistencia, especialmente en los primeros años del siglo XX.

En cuanto a la fiabilidad, la Human Mortality Database es ampliamente reconocida por su rigor en la recopilación y tratamiento de los datos. Sin embargo, siempre existe el riesgo de errores de transcripción o omisiones, que deben ser tenidos en cuenta durante el análisis.

Además, la calidad de las tasas de mortalidad puede verse afectada por definiciones de edad que varían a lo largo del tiempo, lo que puede hacer que las comparaciones no sean completamente consistentes entre diferentes periodos. Para mejorar el modelo predictivo, se han utilizado transformaciones de los datos, como el logaritmo de las tasas de mortalidad, lo que ayuda a mejorar la linealidad de los datos y su ajuste al modelo.

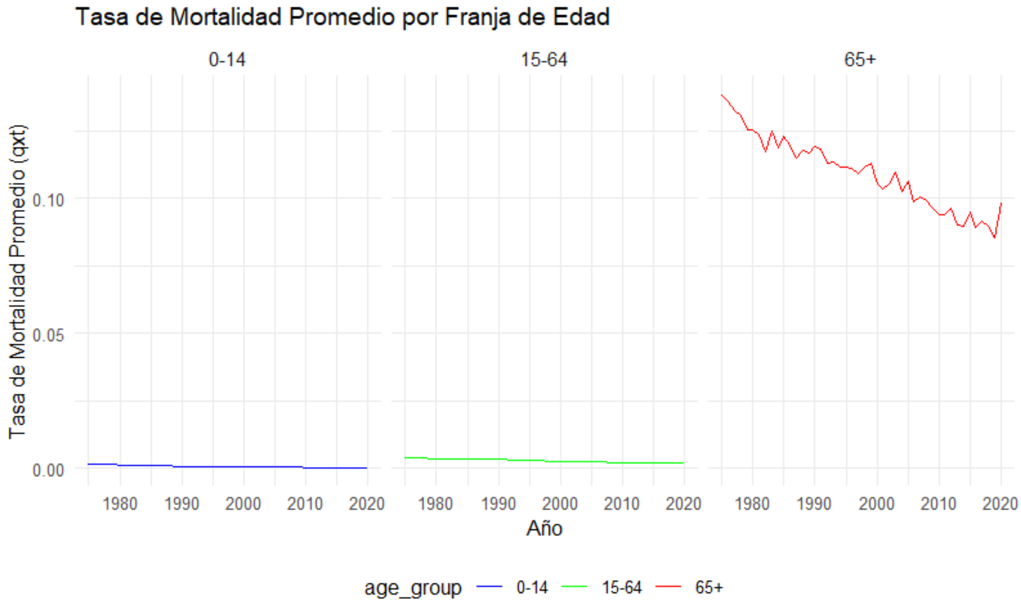
Figura: 1 Evolución de la tasa de mortalidad promedio por franjas de edad en España (1975-2020)



Fuente: Elaboración Propia 1

El análisis gráfico evidencia que la tasa de mortalidad en los grupos de edad infantil (0-14 años) y adulta joven a mediana edad (15-64 años) se mantiene en niveles consistentemente bajos a lo largo del tiempo, lo que refleja el impacto positivo de las políticas de salud pública, avances médicos y mejoras en las condiciones de vida. En estos grupos, la relación con el PIB per cápita no muestra un patrón significativo, lo que sugiere que las causas de mortalidad en estas franjas de edad están más vinculadas a factores no directamente dependientes del desarrollo económico, tales como enfermedades congénitas o accidentes.

Figura: 2 Tasa de Mortalidad Promedio por Franja de Edad



Fuente: Elaboración Propia 2

Por otro lado, en la población de 65 años o más, se observa una clara tendencia descendente en la tasa de mortalidad a medida que aumenta el PIB per cápita, indicando que el crecimiento económico favorece el acceso a mejores servicios de salud, tecnologías médicas y cuidados geriátricos, contribuyendo a la prolongación de la esperanza de vida en esta franja. La evolución temporal de este grupo también refleja una reducción sostenida de la mortalidad, con un ligero incremento en años recientes, posiblemente asociado a eventos sanitarios excepcionales como la pandemia de COVID-19.

Estos resultados subrayan la importancia de considerar el contexto económico y las mejoras en el sistema sanitario como factores clave en la reducción de la mortalidad en las edades avanzadas. La diferenciación por grupos etarios permite enfocar con mayor precisión las políticas y modelos actuariales orientados a optimizar la gestión del riesgo y la planificación de recursos en salud.

En resumen, la utilización de esta base de datos proporciona una sólida fundamentación para el análisis actuarial y epidemiológico desarrollado en este TFM, aportando evidencia empírica sobre las tendencias demográficas y económicas que condicionan la mortalidad, y sirviendo como punto de partida para futuras investigaciones y modelos predictivos.

4. METODOLOGÍA

En este código se emplean una combinación de métodos estadísticos tradicionales y técnicas avanzadas para modelizar las tasas de mortalidad, utilizando tanto modelos clásicos como enfoques modernos como el aprendizaje automático y la inteligencia artificial. Para la predicción de las tasas de mortalidad, se utiliza el modelo de Lee-Carter y el modelo Age-Period-Cohort (APC), que son modelos clásicos de mortalidad ampliamente conocidos por su capacidad para describir la evolución temporal de las tasas de mortalidad.

La metodología empleada para ajustar estos modelos sigue una estructura de descomposición de las tasas de mortalidad en componentes clave, como el nivel de mortalidad por edad, la tendencia temporal y la variabilidad aleatoria, lo que permite una modelización precisa y flexible. En el caso del modelo Lee-Carter, se ajusta mediante una descomposición logit, donde se utiliza un proceso iterativo para ajustar las tasas de mortalidad por edad y por periodo temporal, así como para proyectar las tasas de mortalidad futuras utilizando series temporales.

Por otro lado, se utiliza un enfoque de modelización bifactorial (Lee-Carter bifactorial), el cual añade una dimensión extra al incorporar un componente de cohorte, permitiendo una mayor precisión en el modelado de las tasas de mortalidad en función de las cohortes de edad.

Además, se implementan diversas leyes de mortalidad clásicas (Moivre, Dormoy, Shang, Gompertz) para comparar cómo cada una modela las tasas de mortalidad y qué tan bien se ajustan a los datos reales de mortalidad.

Para la parte de aprendizaje automático, se entrenan modelos de XGBoost, una técnica avanzada de machine learning, para predecir las tasas de mortalidad y se realiza una evaluación comparativa con los modelos estadísticos tradicionales (Lee-Carter, APC). El modelo de XGBoost se entrena utilizando los datos de mortalidad, dividiendo los datos en conjuntos de entrenamiento y prueba para evaluar su capacidad predictiva. Se calculan métricas como el RMSE (Root Mean Squared Error) y R^2 para evaluar la precisión de las predicciones.

La metodología de evaluación de los modelos se complementa con una serie de gráficos de residuos, análisis de normalidad y autocorrelación, y el diagnóstico de leverage e influencia utilizando el cociente de Cook, para asegurar que los modelos sean robustos y que las predicciones sean fiables.

Finalmente, los resultados de los modelos tradicionales y los modelos de aprendizaje automático se comparan utilizando RMSE, R^2, y gráficos de comparación, para determinar cuál de los métodos proporciona las mejores predicciones y si el uso de técnicas avanzadas como XGBoost y redes neuronales mejora las predicciones de los modelos tradicionales. El enfoque sigue una metodología rigurosa de validación y ajuste, con el objetivo de seleccionar el mejor modelo predictivo para las tasas de mortalidad a lo largo del tiempo.

5. RESULTADOS

El presente proyecto se ha desarrollado utilizando el lenguaje de programación R, junto con su entorno de desarrollo integrado, RStudio. Esta elección responde a la amplia aceptación y uso de R en ámbitos científicos y académicos, especialmente en disciplinas que demandan un manejo exhaustivo de datos y un análisis estadístico riguroso.

R destaca por ser una herramienta especializada en la manipulación, análisis y visualización de datos, lo que la posiciona como una opción idónea para realizar cálculos complejos y procesar grandes volúmenes de información de forma eficiente. Su sint*ax*is se caracteriza por ser intuitiva y accesible, facilitando la ejecución de tareas analíticas tanto a usuarios noveles como a expertos en la materia.
Además, R cuenta con una comunidad activa y en constante crecimiento, que aporta numerosos paquetes y librerías especializadas. Estos recursos permiten extender las funcionalidades básicas del lenguaje, abarcando desde operaciones estadísticas sencillas hasta la implementación de modelos avanzados de aprendizaje automático y simulaciones complejas.

Por su parte, RStudio complementa las capacidades de R al proporcionar un entorno amigable y herramientas integradas que optimizan la experiencia del usuario. Entre sus características destacan un editor de código eficiente, una consola interactiva, utilidades para depuración y opciones para la visualización de resultados, facilitando así un desarrollo más ágil y ordenado.

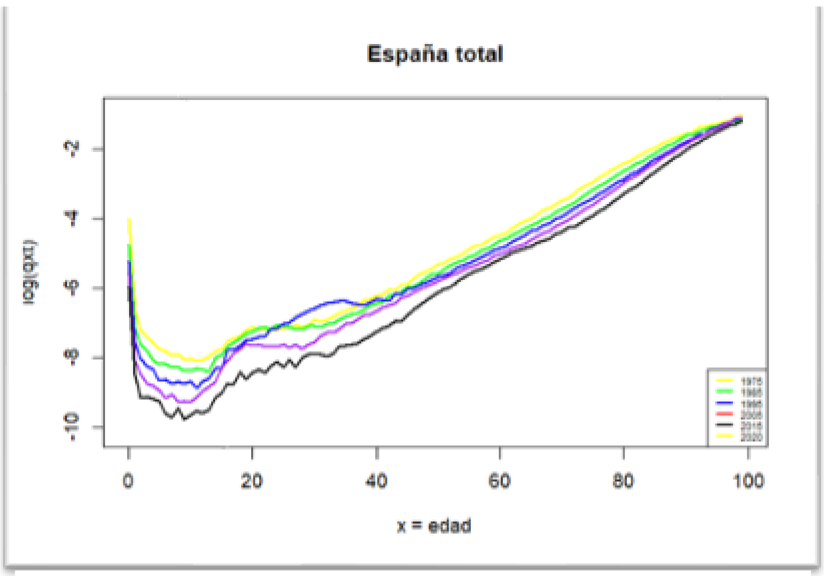
Adicionalmente, la posibilidad de combinar scripts, gráficos y reportes en un único entorno convierte a RStudio en una plataforma robusta para la realización de análisis reproducibles y la generación sistemática de informes.

Estas cualidades posicionan a R y RStudio como las herramientas preferentes en proyectos que requieren análisis detallados y rigurosos de datos, siendo su aplicación habitual en áreas como estadística, bioinformática, economía y, en el caso que nos ocupa, modelización actuarial y análisis de datos vinculados al sector asegurador.

5.1. ANÁLISIS DE LA TASA DE MORTALIDAD ESPAÑOLA

Evolución de la Tasa de Mortalidad en España: Análisis por Edad y Año
El gráfico presenta la evolución de la tasa de mortalidad para la población de España según la edad (x = edad), durante los años seleccionados (1975, 1985, 1995, 2005, 2015 y 2020). La tasa de mortalidad se representa mediante log(qxt), donde qxt es la tasa de mortalidad específica por edad y año, y log(qxt) es su transformación logarítmica.

Figura: 3 Evolución de la Mortalidad por Edad en España (1975-2020)



Fuente: Elaboración Propia 3

Estructura del Gráfico:

Eje X (Edad): Rango de 0 a 100 años, lo que permite observar cómo varía la mortalidad en todas las edades, desde los más jóvenes hasta los más mayores.

Eje Y (log(*qxt*)): Muestra el logaritmo de la tasa de mortalidad. La escala logarítmica es comúnmente utilizada en estudios de mortalidad, ya que suaviza la variabilidad de las tasas y facilita la comparación entre diferentes edades y años.

Curvas de los diferentes años: Cada curva corresponde a un año específico, representado con diferentes colores:

- 1975: Verde
- 1985: Amarillo
- 1995: Azul
- 2005: Morado
- 2015: Rojo
- 2020: Negro

Análisis de las Curvas:

1. Comportamiento General de las Curvas de Mortalidad

Las curvas analizadas revelan una tendencia consistente de incremento en la mortalidad conforme avanza la edad, un patrón habitual en los estudios demográficos y epidemiológicos. Este fenómeno se refleja en la progresiva elevación de las tasas de mortalidad a medida que la población envejece, lo que se observa en el gráfico mediante el ascenso de las curvas hacia edades mayores.

2. Mortalidad en las Edades Tempranas (0 a 20 años)

En el tramo de edades más jóvenes, especialmente en la primera infancia y adolescencia, las curvas permanecen muy próximas entre sí, mostrando tasas de mortalidad bajas y con escasas variaciones a lo largo del tiempo. Esta estabilidad sugiere que la mortalidad infantil y juvenil no ha experimentado cambios significativos, lo que podría atribuirse a que los factores determinantes en estas edades son predominantemente biológicos y menos influenciados por modificaciones en la atención sanitaria o avances tecnológicos.

3. Mortalidad en Edades Intermedias (20 a 60 años)

Durante la adultez temprana y media, las curvas comienzan a mostrar una separación más visible, aunque las diferencias no alcanzan la magnitud observada en edades avanzadas. Este comportamiento podría reflejar el impacto gradual de factores sociales, económicos y médicos que afectan la mortalidad en este grupo etario.

La tasa de mortalidad aumenta de forma progresiva, aunque es probable que los avances en medicina y las mejoras en el estilo de vida hayan contribuido a mitigar este aumento a lo largo de las diferentes décadas.

4. Mortalidad en Edades Avanzadas (60 a 100 años)

En el segmento de población de mayor edad, a partir de los 60 años, las discrepancias entre las curvas correspondientes a distintos años se vuelven más evidentes. Se observa un aumento acelerado en la tasa de mortalidad a partir de los 70 u 80 años, especialmente en los años más recientes, como 2020.

Esto puede interpretarse como una consecuencia de la mejora en la longevidad debido a los avances en la salud pública y la medicina. Sin embargo, el hecho de que las curvas recientes se ubiquen por encima de las de años anteriores podría indicar un incremento en la mortalidad en edades avanzadas, posiblemente relacionado con el envejecimiento poblacional, la mayor prevalencia de enfermedades crónicas asociadas a la edad y la influencia de eventos extraordinarios, como la pandemia de COVID-19.

5. Tendencias Temporales en la Mortalidad

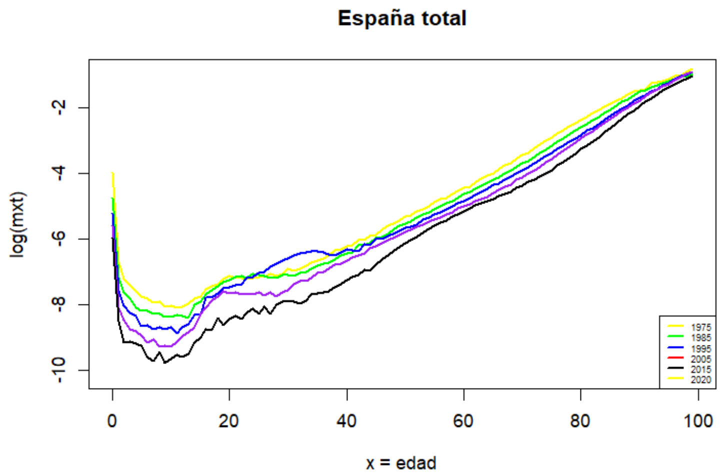
El análisis de las curvas a lo largo de diferentes décadas muestra una progresión en la tasa de mortalidad, con un patrón de aumento especialmente notable en las cohortes más recientes (2015 y 2020) en comparación con décadas anteriores (1975, 1985, 1995). Este fenómeno puede atribuirse a factores demográficos y epidemiológicos, tales como el envejecimiento de la población, la incidencia creciente de enfermedades crónicas y la presión sobre los sistemas de salud debido a crisis sanitarias globales.

Por otro lado, la relativa proximidad entre las curvas correspondientes a los años recientes sugiere una estabilización en las tasas de mortalidad durante los últimos años, posiblemente reflejando los efectos positivos de las políticas de salud pública y los programas de prevención implementados para mejorar la calidad y esperanza de vida.

5.2. ANÁLISIS COMPARATIVO DE LAS TASAS DE MORTALIDAD: ENFOQUES PUNTUALES Y PROMEDIADOS

Este gráfico es similar al anterior, pero en este caso muestra la evolución de la tasa de mortalidad central (log(mx)) para la población total de España, en lugar de la tasa de mortalidad específica por edad (log(qxt)).

Figura: 4 Curvas logarítmicas de mortalidad según edad en España para distintos años



Fuente: Elaboración Propia 4

El análisis de la mortalidad en la población española se puede abordar desde diversas perspectivas, y uno de los elementos clave para entender las variaciones en las tasas de mortalidad a lo largo del tiempo es la elección de las tasas de mortalidad que se utilizan.

En este caso, se presentan dos gráficos que muestran diferentes enfoques para representar la mortalidad.

El primer gráfico representa el logaritmo de la tasa de mortalidad específica por edad y año (log(*qxt*)), donde *qxt* se refiere a la probabilidad de muerte para una persona de edad x en un año determinado. Esta tasa refleja la mortalidad puntual para cada año y grupo de edad, proporcionando una visión precisa de las fluctuaciones anuales en la mortalidad.

Dado que los datos son anuales, es probable que los valores de mortalidad presenten picos y caídas debido a factores excepcionales, como pandemias, crisis sanitarias o cambios políticos, lo que introduce mayor variabilidad en los resultados.

Por otro lado, el segundo gráfico muestra el logaritmo de la tasa de mortalidad central (log(mx)), donde mx representa la tasa de mortalidad promediada durante un periodo de tiempo, generalmente en intervalos de cinco o diez años. Este enfoque suaviza las fluctuaciones anuales, proporcionando una visión más estable y general de la mortalidad. Al promediar los datos, la tasa de mortalidad central tiende a reflejar mejor las tendencias a largo plazo y es menos susceptible a las variaciones anuales causadas por eventos extraordinarios.

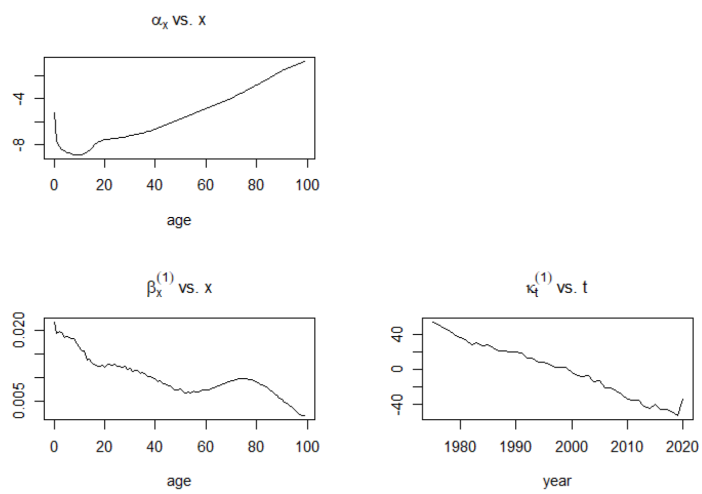
La principal diferencia entre estos dos gráficos radica en cómo las tasas de mortalidad son calculadas. Mientras que el primer gráfico se enfoca en los cambios anuales y muestra una visión más detallada, el segundo gráfico proporciona una perspectiva a largo plazo, promediando los datos para reducir el ruido generado por fluctuaciones puntuales.

Ambos métodos son útiles en el estudio de la mortalidad, pero ofrecen enfoques complementarios para entender los cambios en la mortalidad a lo largo del tiempo, especialmente cuando se quiere observar la evolución de la mortalidad a lo largo de varias décadas.

5.3. ANÁLISIS DEL MODELO LEE- CARTER

En este análisis se ha utilizado el modelo Lee-Carter para modelizar las tasas de mortalidad a partir de los datos históricos de mortalidad, abarcando el período desde 1975 hasta 2020. El modelo ha permitido ajustar la mortalidad específica por edad, generando estimaciones de las tasas de mortalidad para cada año y edad. A continuación, se detallan los principales resultados obtenidos a partir del modelo y los gráficos generados, así como los diagnósticos de calidad del ajuste.

Figura: 5 Componentes del modelo de Lee-Carter para mortalidad: perfil por edad y evolución temporal



Fuente: Elaboración Propia 5

Funciones Ajustadas del Modelo Lee-Carter

Los gráficos obtenidos muestran las funciones ajustadas para los parámetros ***ax***, ***bx*** y ***Kt*** del modelo Lee-Carter:

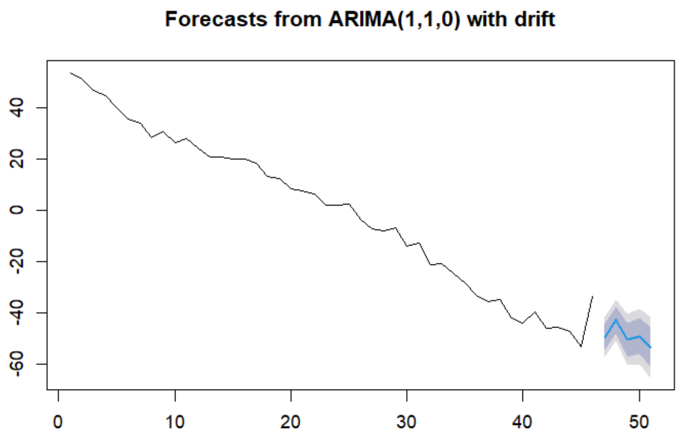
- *ax* vs. x: El gráfico de la función *ax* frente a la edad muestra cómo cambia la mortalidad específica por edad en función de la edad. Este parámetro es el componente de nivel de mortalidad que describe las tasas de mortalidad básicas para cada edad.

- *bx* vs. x: El gráfico de *bx* frente a la edad ilustra la evolución del componente de edad del modelo, es decir, cómo la mortalidad cambia a lo largo del tiempo según el parámetro de edad. Se observa una tendencia a la disminución en las tasas de mortalidad para ciertas edades con el tiempo.

- *Kt* vs. t: El gráfico de *Kt* frente al tiempo refleja la evolución de la tendencia temporal de la mortalidad, que está relacionada con los cambios en las tasas de mortalidad a lo largo de los años. Se observa una disminución en las tasas de mortalidad a lo largo de los años debido a los avances en medicina, salud pública y otras intervenciones.

Proyección de las Tasas de Mortalidad

Figura: 6 Pronóstico de la serie temporal mediante modelo ARIMA(1,1,0)



Fuente: Elaboración Propia 6

Se ha utilizado el modelo ajustado para realizar una proyección de las tasas de mortalidad. Para ello, se aplicó el modelo ARIMA(1,1,0) sobre el componente temporal *Kt*, con el objetivo de prever las tendencias futuras de la mortalidad en la población.

Gráfico de las proyecciones ARIMA: El gráfico de proyecciones generadas por el modelo ARIMA muestra cómo se espera que evolucione la mortalidad en los próximos años, destacando la tendencia decreciente en las tasas de mortalidad proyectadas.

Diagnóstico de Ajuste del Modelo

Se realizaron varias pruebas para evaluar la calidad del ajuste del modelo. A continuación, se presentan los resultados clave:

- **Prueba de Normalidad (Shapiro-Wilk)**: Los residuos del modelo fueron sometidos a la prueba de normalidad de Shapiro-Wilk, con el siguiente resultado:
    - **Valor W** = 0.68798
    - **Valor p** = 2.854e-13

Un valor p extremadamente bajo, indica que los residuos no siguen una distribución normal, lo cual es un indicio de que el modelo podría no estar capturando completamente algunas estructuras o patrones de los datos. Sin embargo, es importante tener en cuenta que la normalidad de los residuos no es una condición estricta para modelos de series temporales.

- **Prueba de Autocorrelación (Box-Ljung):** La prueba de autocorrelación de Ljung-Box se utilizó para verificar si los residuos presentan autocorrelación. El resultado fue:

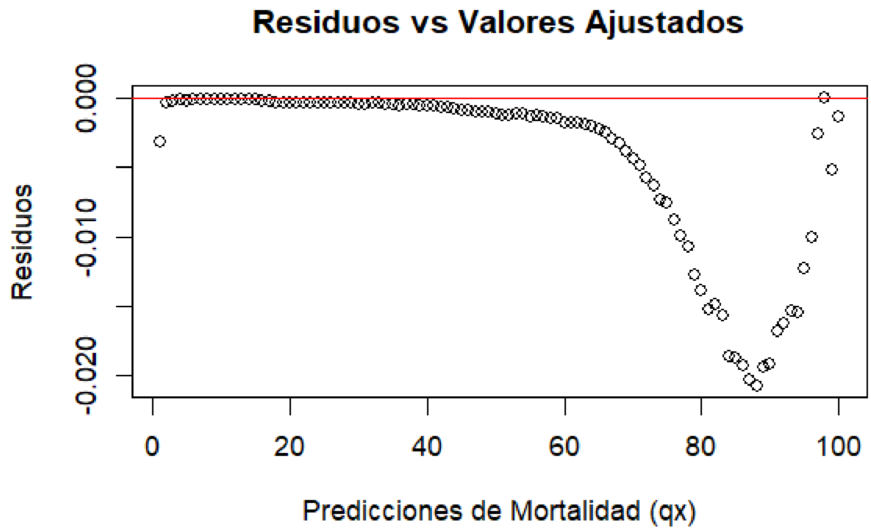
    - **Valor p** < 2.2e-16

Este valor p extremadamente bajo sugiere que los residuos sí están autocorrelacionados, lo que indica que el modelo podría no haber capturado completamente las dependencias temporales. Es recomendable revisar si el modelo debe incorporar efectos adicionales o si se deben ajustar los parámetros para mejor capturar las correlaciones a lo largo del tiempo.

Gráficos de Diagnóstico del Residuos

Residuos vs. Valores Ajustados: El gráfico de residuos frente a valores ajustados muestra la relación entre los errores del modelo y las predicciones realizadas sobre la mortalidad (*qx*). En general, los residuos se mantienen cercanos a cero para la mayoría de los valores predichos, lo que indica un ajuste adecuado del modelo en esos rangos.

Figura: 7 Diagnóstico del Residuos Modelo Lee-Carter



Fuente: Elaboración Propia 7

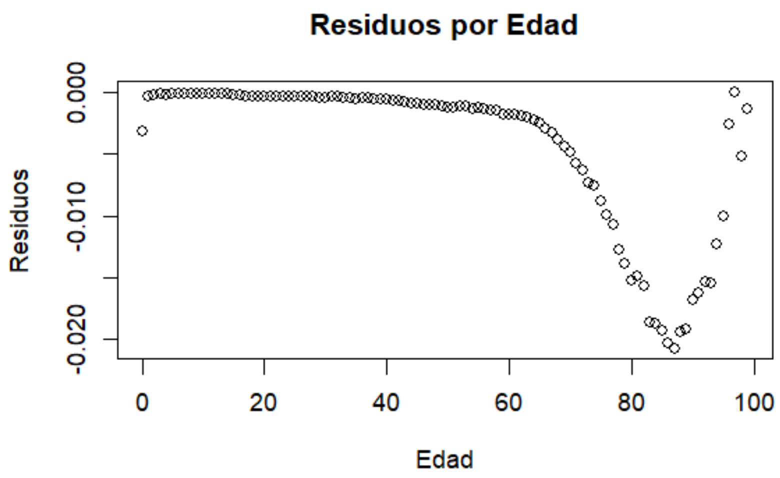
Sin embargo, para los valores más altos de mortalidad predicha, se observa una tendencia clara y sistemática de residuos negativos, lo que implica que el modelo tiende a sobreestimar la mortalidad en esos casos. Esta desviación sugiere que el modelo presenta un sesgo en la zona de valores extremos, lo que puede ser indicativo de una falta de adecuación del modelo para capturar completamente la dinámica de mortalidad en esos niveles. Por tanto, es necesario considerar posibles mejoras en la especificación del modelo, ya sea incorporando transformaciones, variables adicionales o utilizando enfoques alternativos que permitan un mejor ajuste para toda la gama de valores predichos.

Residuos por Edad: El gráfico muestra los residuos de un modelo en función de la variable "Edad". En el eje horizontal se encuentra la edad (rango de 0 a 100), y en el eje vertical los residuos correspondientes.

Se observa que, para edades menores a aproximadamente 75 años, los residuos se mantienen muy cercanos a cero, indicando que el modelo ajusta bastante bien para ese rango de edad. Sin embargo, a partir de los 75 años, los residuos empiezan a ser negativos y con una pendiente descendente, alcanzando valores mínimos alrededor de 90 años. Luego, cerca de los 100 años, los residuos vuelven a acercarse a cero, aunque con mayor dispersión.

Este comportamiento puede indicar que el modelo tiende a subestimar el valor para las edades avanzadas (residuos negativos significan que la predicción es mayor que el valor observado). Además, la mayor dispersión cerca de los 100 años puede deberse a menor cantidad de datos en esas edades o mayor variabilidad natural.

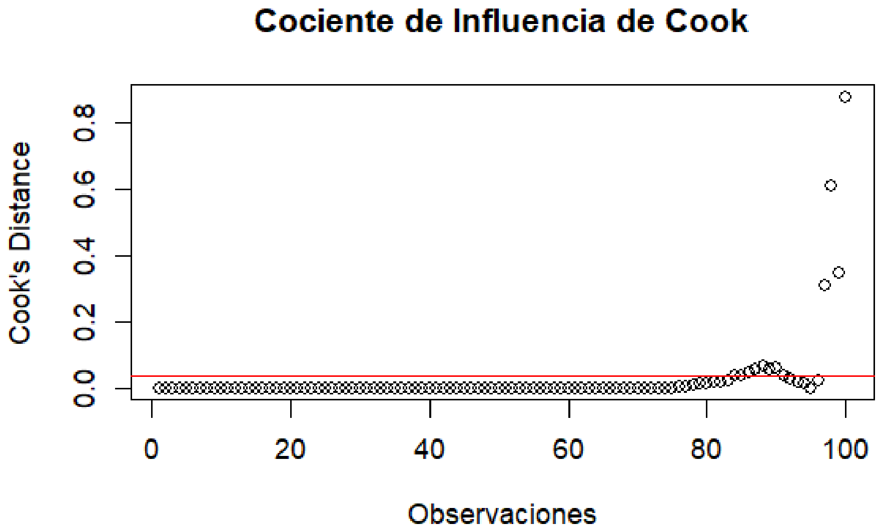
Cociente de Influencia de Cook: El gráfico del Cociente de Influencia de Cook muestra la influencia que tienen las observaciones individuales sobre el ajuste del modelo. En el eje horizontal se representan las observaciones ordenadas, mientras que en el eje vertical



se mide la distancia de Cook, que cuantifica el impacto que tiene eliminar cada observación sobre los coeficientes estimados.

En este caso, la mayoría de las observaciones presentan valores de influencia cercanos a cero, lo que indica que no afectan significativamente el modelo. Sin embargo, se identifican varias observaciones, ubicadas hacia el final del conjunto de datos, con valores elevados de distancia de Cook, superando el umbral marcado por la línea roja.

Esto indica que dichas observaciones son potencialmente influyentes y pueden estar afectando de manera considerable la estimación del modelo, por lo que su análisis y posible tratamiento resulta importante para asegurar la robustez y validez del ajuste realizado.



Conclusiones del Diagnóstico y Ajuste del Modelo

El modelo de Lee-Carter ajustado a las tasas de mortalidad de la población española muestra un buen ajuste general, pero con áreas de mejora en términos de normalidad de los residuos y autocorrelación. Aunque los residuos parecen tener una distribución relativamente aleatoria, las pruebas realizadas indican que el modelo no ha capturado completamente la autocorrelación en los datos, lo que podría sugerir que los componentes temporales (*Kt*) necesitan ser modelizados de manera más compleja. Además, la presencia de puntos influyentes señala que ciertas observaciones deben ser analizadas más detalladamente.

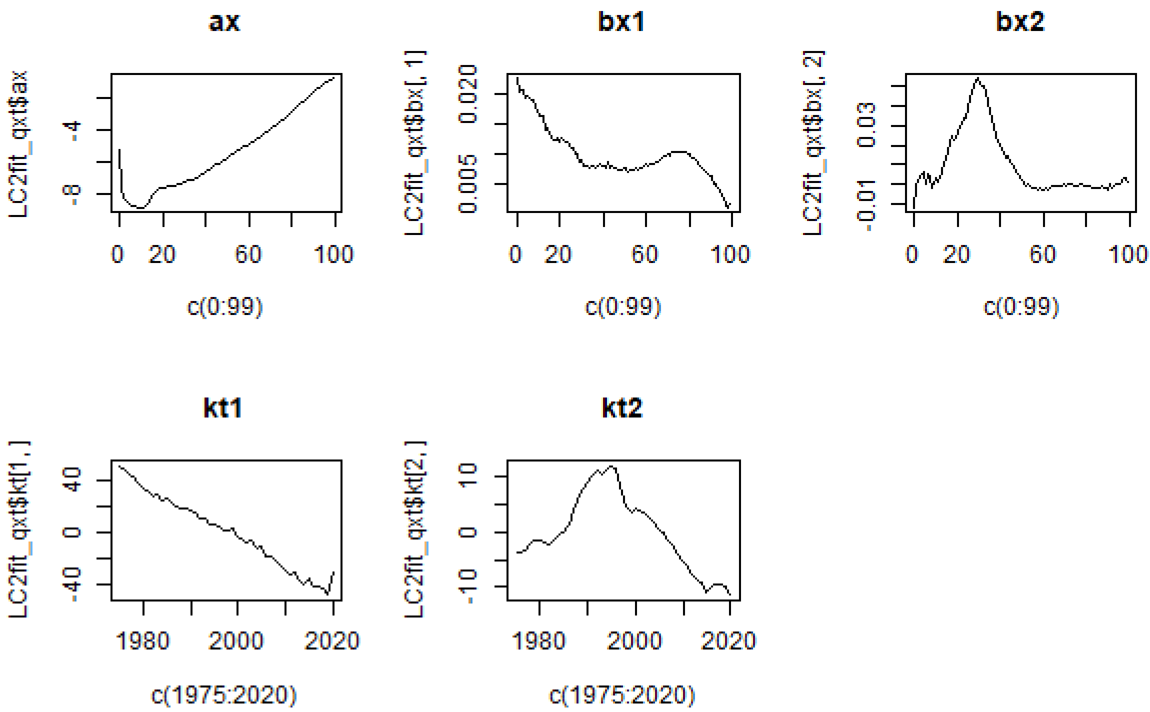
Para optimizar el modelo, se sugiere ajustar los parámetros o emplear modelos más avanzados que permitan captar con mayor precisión la temporalidad y las interacciones presentes en los datos. Además, es fundamental analizar en profundidad los puntos de influencia detectados para determinar su impacto en la precisión del modelo. Asimismo, resulta pertinente evaluar la aplicación de transformaciones adicionales a los datos con el fin de corregir posibles desviaciones de normalidad en los residuos. Este proceso de validación y ajuste constituye una etapa clave para garantizar que el modelo empleado para la predicción de la mortalidad en la población española sea sólido y fiable.

5.4. ANÁLISIS LEE-CARTER BIFACTORIAL

Los gráficos proporcionados corresponden a los resultados del ajuste del modelo de mortalidad de Lee-Carter bifactorial (LC2) aplicado a los datos de mortalidad. A continuación, se presenta una interpretación detallada de cada uno de los gráficos:

Gráficos de las funciones

Los primeros gráficos muestran las funciones *ax* (mortalidad por edad), *bx*1 y *bx*2 (coeficientes de la mortalidad por edad y el componente temporal), y *Kt* (tendencia temporal). Estos coeficientes son fundamentales en el modelo de Lee-Carter, ya que describen cómo evoluciona la mortalidad a lo largo de las edades y los años.



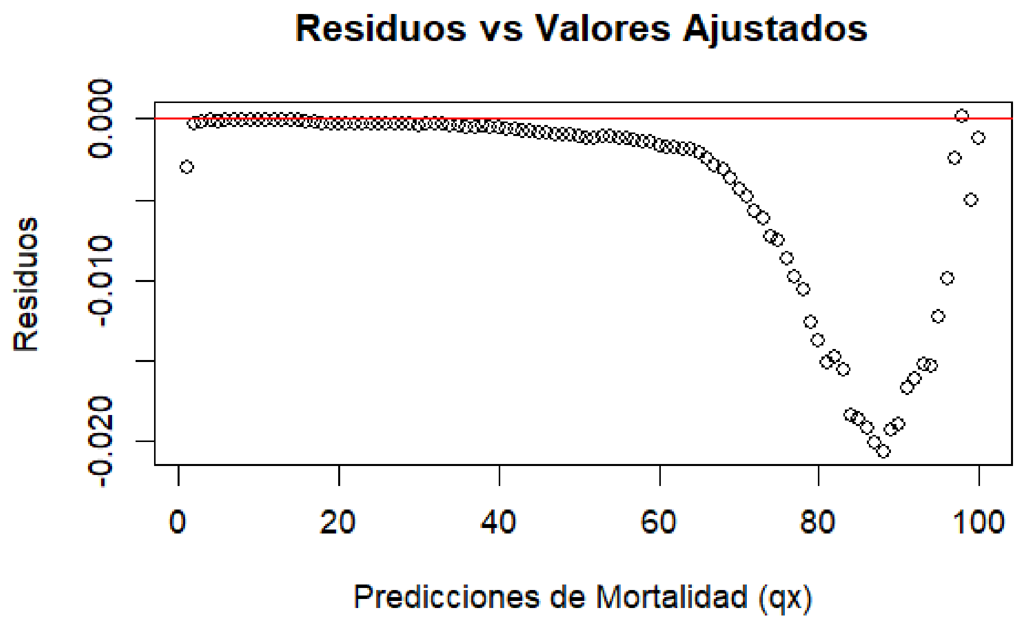
El gráfico que representa la función *ax* frente a la edad muestra de manera clara cómo las tasas de mortalidad aumentan de forma progresiva conforme avanza la edad. Este comportamiento es coherente con lo que se espera desde un punto de vista demográfico y actuarial, ya que la mortalidad suele ser baja durante la infancia y juventud, aumentando considerablemente en edades avanzadas debido al deterioro fisiológico y la mayor prevalencia de enfermedades.

Por otro lado, el gráfico correspondiente al coeficiente $bx1$ frente a la edad refleja el primer componente asociado a la variación de la mortalidad según la edad y el tiempo. Se observa una tendencia decreciente en este coeficiente conforme aumenta la edad, lo que implica que el efecto que la edad tiene sobre la mortalidad no es constante a lo largo del tiempo, sino que se modula. Este comportamiento sugiere que la influencia de ciertos factores relacionados con la edad puede estar disminuyendo o cambiando su impacto en distintas cohortes o periodos históricos.

De manera análoga, el gráfico que muestra el segundo componente $bx2$ en función de la edad revela una variabilidad adicional en las tasas de mortalidad vinculada a otro patrón temporal. Esta curva evidencia cómo, además del primer componente, existen otras dinámicas que afectan la mortalidad según diferentes rangos de edad, posiblemente asociadas a cambios en factores socioeconómicos, avances médicos o condiciones ambientales que varían a lo largo del tiempo y afectan de forma diferencial a ciertos grupos de edad.

Finalmente, el gráfico del parámetro temporal Kt en función de los años expone la evolución temporal general de las tasas de mortalidad. Se puede apreciar una tendencia descendente en este parámetro a lo largo del tiempo, lo cual es indicativo de una reducción sostenida en las tasas de mortalidad. Esta tendencia se puede interpretar como el resultado de mejoras en la salud pública, avances en la medicina, mejores condiciones de vida y mayor longevidad en la población. Los valores negativos en este gráfico son consistentes con esta interpretación, evidenciando una disminución en los riesgos de mortalidad general, lo que representa un aspecto positivo dentro del análisis.

Gráfico de Residuales vs. Valores Ajustados:



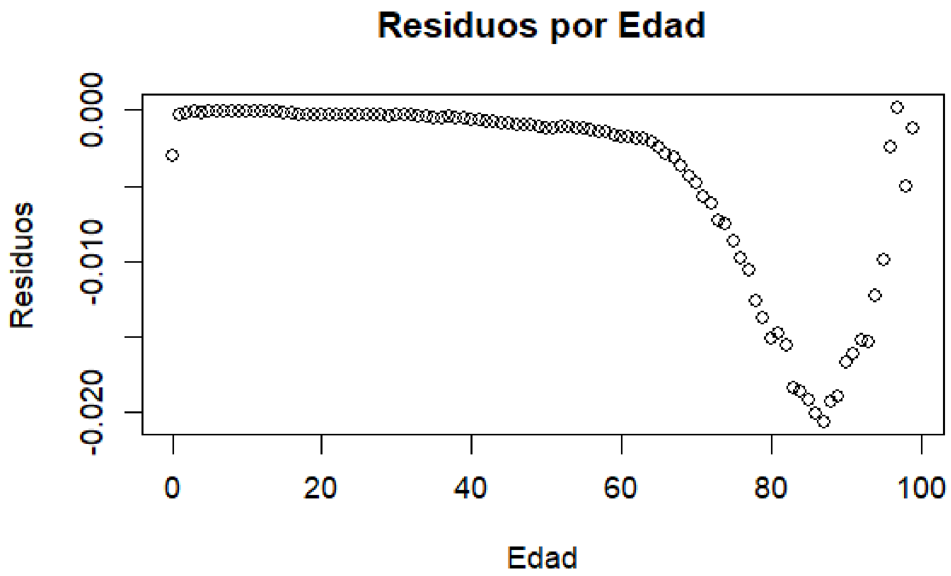
El gráfico presentado ilustra la relación entre los residuales, es decir, las diferencias entre los valores observados y los valores ajustados del parámetro de mortalidad qx y los valores ajustados del mismo parámetro. Este tipo de gráfico es fundamental para evaluar la calidad del ajuste del modelo empleado en la estimación de la mortalidad.

En el análisis del gráfico se observa que los puntos se distribuyen de forma bastante homogénea a lo largo del eje de los valores ajustados, lo que indica una ausencia de patrones sistemáticos en los residuos. Sin embargo, se identifican algunos residuales tanto negativos como positivos en las zonas correspondientes a valores ajustados más elevados, lo que podría estar señalando una ligera subestimación o sobreestimación del modelo en determinados rangos de mortalidad.

Es importante destacar que, en general, la magnitud de los residuos es pequeña para la mayoría de las observaciones, lo que sugiere que el modelo proporciona un ajuste adecuado a los datos en la mayoría de los casos. No obstante, la existencia de ciertas desviaciones puntuales pone de manifiesto que el modelo podría beneficiarse de una revisión o ajuste adicional para mejorar la precisión en esos casos específicos.

Gráfico de Residuales por Edad:

El gráfico muestra la distribución de los residuos en función de la edad, lo cual es un análisis esencial para identificar posibles sesgos o patrones sistemáticos en el ajuste del modelo a lo largo de las diferentes edades.

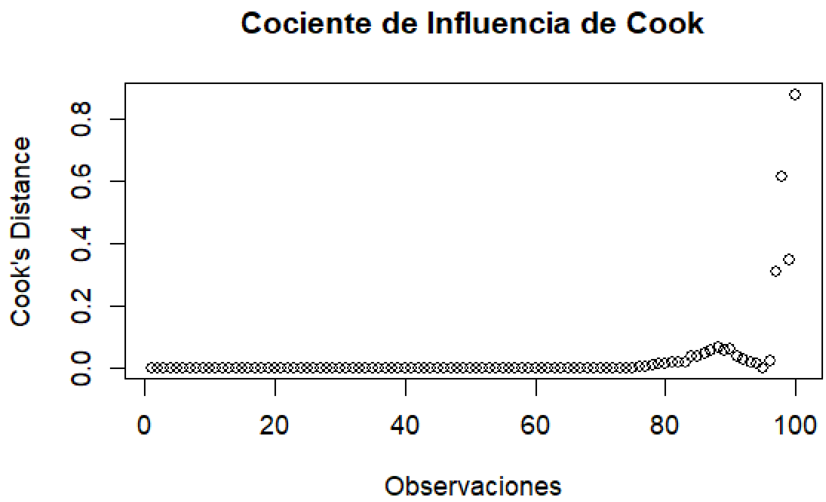


Al observar el gráfico, se aprecia que los residuos se concentran mayoritariamente alrededor de cero en las edades más bajas, lo que indica que el modelo logra un buen ajuste en estos tramos de edad. Esto sugiere que las estimaciones de las tasas de mortalidad para los grupos etarios jóvenes son precisas y fiables.

No obstante, conforme la edad avanza, se observa un aumento en la dispersión de los residuos, con mayores desviaciones respecto a cero. Este patrón puede indicar que el modelo presenta mayores dificultades para capturar la dinámica real de la mortalidad en edades avanzadas. Es posible que factores no contemplados o la mayor variabilidad biológica y social en edades mayores influyan en esta menor precisión del ajuste.

En consecuencia, estos resultados señalan la necesidad de considerar ajustes adicionales o modelos alternativos que permitan mejorar la estimación de la mortalidad en los rangos de edad más altos, donde el modelo actual muestra limitaciones.

Gráfico de la Distancia de Cook:



El gráfico de la distancia de Cook representa la influencia que cada observación ejerce sobre los coeficientes estimados en el modelo. Este tipo de gráfico es fundamental para identificar observaciones atípicas o influyentes, que pueden tener un impacto desproporcionado y potencialmente distorsionar el ajuste global del modelo.

El análisis del gráfico revela que existen varias observaciones con valores elevados de distancia de Cook, lo que indica que ciertos puntos de datos poseen una influencia significativa en la estimación de los parámetros. Particularmente, se observa que estas observaciones influyentes suelen corresponder a edades avanzadas, lo que sugiere que en estas zonas la información puede estar sesgada o presentar mayor variabilidad.

La presencia de estos puntos atípicos justifica un examen detallado para determinar si representan errores en la recogida de datos, valores extremos legítimos o si, por el contrario, deben ser excluidos o tratados de forma diferenciada en el modelo. Esta evaluación es clave para asegurar la robustez y validez del ajuste estadístico y evitar que dichas observaciones influyan indebidamente en las conclusiones del análisis.

El gráfico que representa la función *ax* frente a la edad muestra cómo las tasas de mortalidad aumentan progresivamente conforme avanza la edad. Este comportamiento es esperado, dado que la mortalidad suele ser baja en edades tempranas y tiende a incrementarse en edades avanzadas debido a factores fisiológicos y epidemiológicos asociados al envejecimiento.

En conclusión, el modelo de Lee-Carter muestra un desempeño adecuado en general, especialmente para edades jóvenes, aunque presenta limitaciones en la estimación para edades mayores. La identificación de residuos y observaciones atípicas señala la necesidad de revisar estos puntos para optimizar el modelo.

Las tendencias temporales reflejan mejoras en la mortalidad a lo largo del tiempo, pero las anomalías en edades avanzadas sugieren que podrían incorporarse ajustes o variables adicionales para capturar mejor las fluctuaciones en estos grupos.

En conjunto, el análisis de residuos y de la influencia de observaciones es fundamental para validar y mejorar la precisión y robustez del modelo, facilitando proyecciones de mortalidad más confiables.

5.5. ANÁLISIS DE LOS RESULTADOS UTILIZANDO EL MODELO APC

En este análisis, se ha utilizado el modelo Age-Period-Cohort (APC) para modelizar las tasas de mortalidad (qx) en función de tres componentes principales: la edad (Age), el periodo (Period) y la cohorte (Cohort). Este enfoque tiene como objetivo descomponer las tasas de mortalidad en sus efectos relativos para cada uno de estos factores. A través de la implementación de este modelo, se han obtenido diversos gráficos y análisis de residuos, los cuales permiten evaluar la validez y precisión del modelo ajustado.

Resultados de los Análisis de Residuos

El gráfico que representa los residuos en función de los valores ajustados muestra una tendencia negativa conforme aumentan las predicciones de mortalidad qx. Sin embargo, se observa que los residuos no se distribuyen aleatoriamente alrededor del cero, lo que indica que el modelo podría no estar capturando adecuadamente ciertas relaciones subyacentes en los datos, particularmente para las predicciones de mortalidad más elevadas.

En cuanto a la distribución de los residuos por edad, se evidencia que estos son relativamente pequeños en las edades jóvenes, pero aumentan de forma significativa en las edades avanzadas. Esta tendencia sugiere que el modelo enfrenta dificultades para ajustar con precisión las tasas de mortalidad en la población de mayor edad, fenómeno que es habitual debido a la mayor variabilidad y complejidad de las tasas de mortalidad en estos rangos etarios.

El análisis de la distancia de Cook revela una alta influencia de determinadas observaciones, principalmente aquellas correspondientes a edades avanzadas. Las observaciones con valores elevados en esta métrica ejercen un impacto considerable en la estimación de los parámetros del modelo. Por lo tanto, es fundamental revisar estos puntos atípicos para garantizar la robustez y fiabilidad del modelo.

Pruebas de Normalidad y Autocorrelación

La prueba de normalidad de Shapiro-Wilk aplicada a los residuos indica que éstos no siguen una distribución normal (p-valor $< 2.2e-16$). Esta desviación de la normalidad constituye una limitación, dado que la suposición de normalidad de los residuos es fundamental en muchos enfoques estadísticos y afecta la validez de ciertas inferencias.

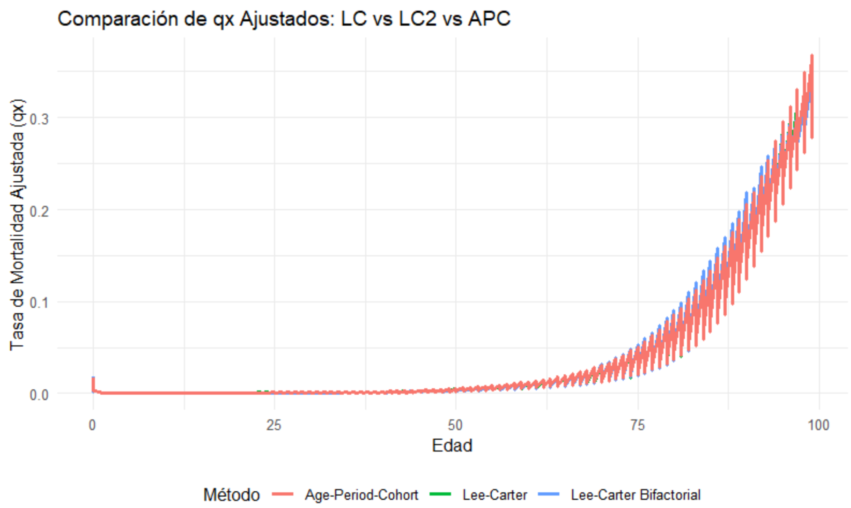
Por otro lado, la prueba de autocorrelación de Box-Ljung evidencia que los residuos presentan autocorrelación significativa (p-valor $< 2.2e-16$), lo que sugiere que el modelo no ha capturado adecuadamente las dependencias temporales presentes en los datos. Esto señala la posible necesidad de incorporar efectos dinámicos adicionales o adoptar técnicas más complejas de modelado de series temporales para mejorar la precisión predictiva.
En síntesis, el modelo APC aplicado a las tasas de mortalidad ha permitido descomponer y analizar de manera detallada los efectos relacionados con la edad, el periodo y la cohorte.

Sin embargo, los análisis residuales evidencian limitaciones importantes: la no normalidad y la autocorrelación de los residuos, junto con la influencia marcada de ciertas observaciones atípicas, indican que el modelo requiere ajustes adicionales para optimizar su capacidad predictiva.

Estos hallazgos enfatizan la necesidad de una revisión exhaustiva del modelo, incluyendo la posible exclusión o tratamiento especial de observaciones influyentes y la consideración de nuevas variables o estructuras dinámicas. De este modo, se podrá mejorar la precisión, robustez y aplicabilidad del modelo en la proyección de las tasas de mortalidad.

5.6. Interpretación del Gráfico de Comparación de Modelos de Mortalidad Ajustados: Lee-Carter (LC) vs Lee-Carter Bifactorial (LC2) vs Age-Period-Cohort (APC)

El gráfico generado muestra la comparación de las tasas de mortalidad ajustadas (qx) obtenidas a partir de tres modelos distintos: Lee-Carter (LC), Lee-Carter Bifactorial (LC2) y Age-Period-Cohort (APC). Los modelos se aplicaron a un conjunto de datos de mortalidad que abarca 46 años (1975-2020) para 100 grupos de edad (0-99 años).



El gráfico representa la relación entre la edad y las tasas de mortalidad ajustadas qx, con el eje horizontal mostrando la edad de los individuos desde 0 hasta 100 años, y el eje vertical reflejando la probabilidad de fallecimiento a cada edad, que varía entre 0 y 0.35. Se observa un aumento esperado en las tasas de mortalidad conforme la edad avanza.

Se presentan tres curvas correspondientes a distintos modelos: el modelo Lee-Carter (LC) en rojo, que descompone la mortalidad en función de la edad, el periodo y la cohorte; el modelo Lee-Carter Bifactorial (LC2) en azul, que añade una interacción entre edad y periodo; y el modelo Age-Period-Cohort (APC) en verde, que también descompone la mortalidad según estos tres factores, pero con una metodología distinta.

El modelo LC exhibe una forma típica, con tasas bajas en edades tempranas y un incremento progresivo con la edad, ajustándose bien a la tendencia general, pero mostrando menor precisión en las edades avanzadas donde la mortalidad aumenta de forma más abrupta.

Por su parte, el modelo LC2 refleja una mayor variabilidad en las tasas ajustadas para edades avanzadas, capturando mejor las fluctuaciones gracias a la interacción edad-periodo. Finalmente, el modelo APC sigue el patrón general de aumento de mortalidad con la edad, aunque presenta diferencias sutiles en los grupos de edad media (20-60 años), debido a su mayor flexibilidad para modelar efectos específicos de cohortes y periodos.

En conjunto, el análisis de estas curvas permite observar cómo cada modelo maneja la complejidad de la mortalidad en función de la edad y el tiempo, destacando que las versiones extendidas como LC2 y APC aportan una mejor captura de las variaciones en edades avanzadas y en grupos etarios específicos.

5.7. INTERPRETACIÓN DE RESULTADOS RMSE ENTRE MODELOS LEE-CARTER, LEE-CARTER BIFACTORIAL Y AGE-PERIOD-COHORT

Análisis Comparativo de Modelos mediante RMSE

Para complementar la comparación entre los modelos de mortalidad ajustada Lee-Carter (LC), Lee-Carter Bifactorial (LC2) y Age-Period-Cohort (APC), se ha calculado la métrica RMSE (Root Mean Squared Error), la cual mide la diferencia promedio entre los valores predichos por cada modelo y los valores observados en los datos reales. A continuación, se presentan y analizan los resultados obtenidos:

RMSE entre Lee-Carter y Lee-Carter Bifactorial:

El RMSE calculado entre el modelo Lee-Carter clásico y su versión bifactorial es

0,0009644892. Este valor reducido indica que la diferencia promedio entre las tasas de mortalidad ajustadas por ambos modelos es mínima, sugiriendo que la extensión bifactorial introduce solo una ligera mejora en el ajuste sin generar discrepancias significativas respecto al modelo original.

RMSE entre Lee-Carter y Age-Period-Cohort:

El RMSE entre Lee-Carter y el modelo Age-Period-Cohort es 0,004698728, valor superior al anterior. Esto revela una mayor diferencia en el ajuste de las tasas de mortalidad entre estos dos modelos, lo que puede atribuirse a la distinta forma en que el modelo APC descompone la mortalidad en edad, periodo y cohorte, y la metodología alternativa empleada para capturar dichos efectos.

RMSE entre Lee-Carter Bifactorial y Age-Period-Cohort:

El RMSE entre el modelo Lee-Carter Bifactorial y APC es 0,004701918, muy próximo al obtenido entre Lee-Carter y APC. Esto sugiere que el modelo APC se comporta de manera similar al Lee-Carter Bifactorial en términos de discrepancias con respecto a las predicciones observadas, indicando que ambas variantes presentan un ajuste ligeramente diferente al modelo Lee-Carter clásico.

Resumen de la Comparación

Los modelos Lee-Carter y Lee-Carter Bifactorial presentan un ajuste muy similar, reflejado en un RMSE muy bajo (0.00096), lo que implica que la inclusión del término bifactorial no aporta una mejora sustancial en la precisión del modelo respecto al modelo base.

Por su parte, el modelo Age-Period-Cohort exhibe un RMSE mayor en comparación con ambos modelos Lee-Carter, lo que indica un ajuste algo menos preciso. Sin embargo, la diferencia en RMSE no es lo suficientemente significativa como para descartar el modelo APC, sino que destaca su validez como alternativa metodológica, aportando una perspectiva diferente al análisis de mortalidad.

El análisis comparativo basado en el RMSE revela que los modelos Lee-Carter y Lee-Carter Bifactorial ofrecen un ajuste muy similar y cercano a los datos observados, con diferencias mínimas entre ellos. Por otro lado, el modelo Age-Period-Cohort, aunque presenta un ajuste ligeramente menos preciso, sigue siendo una opción válida debido a su enfoque más flexible y su capacidad para capturar efectos específicos de edad, periodo y cohorte.

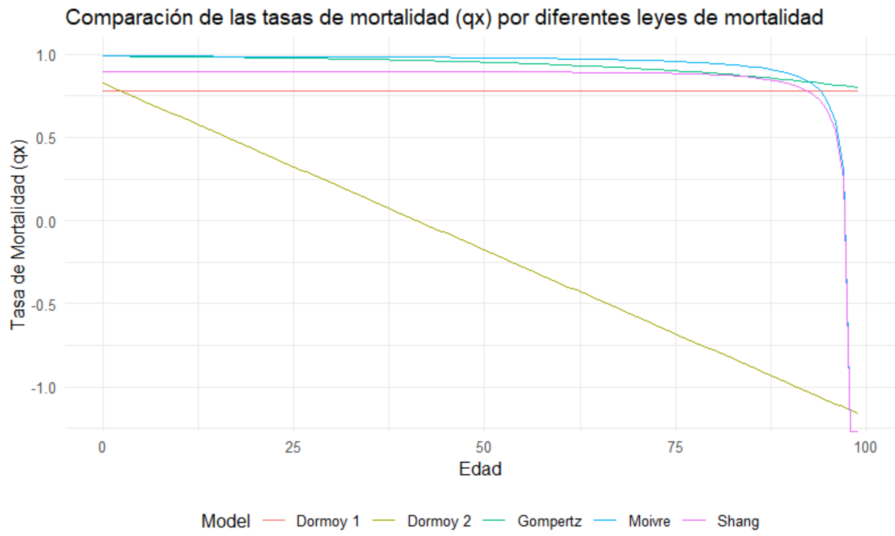
En términos de precisión pura, el modelo Lee-Carter Bifactorial muestra una leve mejora respecto al modelo clásico de Lee-Carter, situándose como el modelo con mejor ajuste entre los tres. No obstante, la elección del modelo más adecuado también debe considerar la complejidad, interpretabilidad y objetivos específicos del análisis.

Por tanto, si se prioriza la simplicidad con buen ajuste, Lee-Carter clásico es suficiente; mientras que, si se requiere capturar interacciones más complejas, Lee-Carter Bifactorial o APC pueden ser preferibles.

5.8. ANÁLISIS COMPARATIVO DE LAS TASAS DE MORTALIDAD AJUSTADAS (*QX*) POR DIFERENTES LEYES DE MORTALIDAD

Tras el análisis previo realizado con el modelo Lee-Carter, se ha incorporado un estudio comparativo de diversas leyes clásicas de mortalidad. Este enfoque permite examinar el comportamiento y las características de la mortalidad ajustada bajo distintos métodos, proporcionando una visión más amplia y detallada de las diferentes formulaciones utilizadas en la modelización actuarial.

El gráfico que se presenta a continuación compara las tasas de mortalidad ajustadas qx obtenidas mediante diferentes leyes de mortalidad: Moivre, Dormoy (1), Dormoy (2), Shang y Gompertz. A continuación, se describen las principales observaciones y conclusiones derivadas del análisis.



Leyes de Dormoy:

La Ley de Dormoy (1) se caracteriza por mantener una tasa de mortalidad constante a lo largo de las distintas edades, lo que sugiere una mortalidad relativamente estable en la población sin un incremento notable asociado al envejecimiento.

En contraste, la Ley de Dormoy (2) incorpora una componente adicional en su formulación que provoca un aumento gradual en la tasa de mortalidad conforme avanza la edad, especialmente en las etapas más avanzadas. Esto refleja una mayor sensibilidad al factor edad, aunque con un comportamiento más moderado en comparación con otras leyes como Gompertz.

Ley de Moivre:

La Ley de Moivre se distingue por una tasa de mortalidad que disminuye considerablemente en las primeras edades y mantiene esta tendencia de reducción hasta alcanzar un valor cercano a uno en edades muy avanzadas. Este patrón sugiere una evolución de la mortalidad menos abrupta que en otros modelos, resaltando un aumento en la mortalidad de forma progresiva, pero no tan pronunciada.

Ley de Shang:

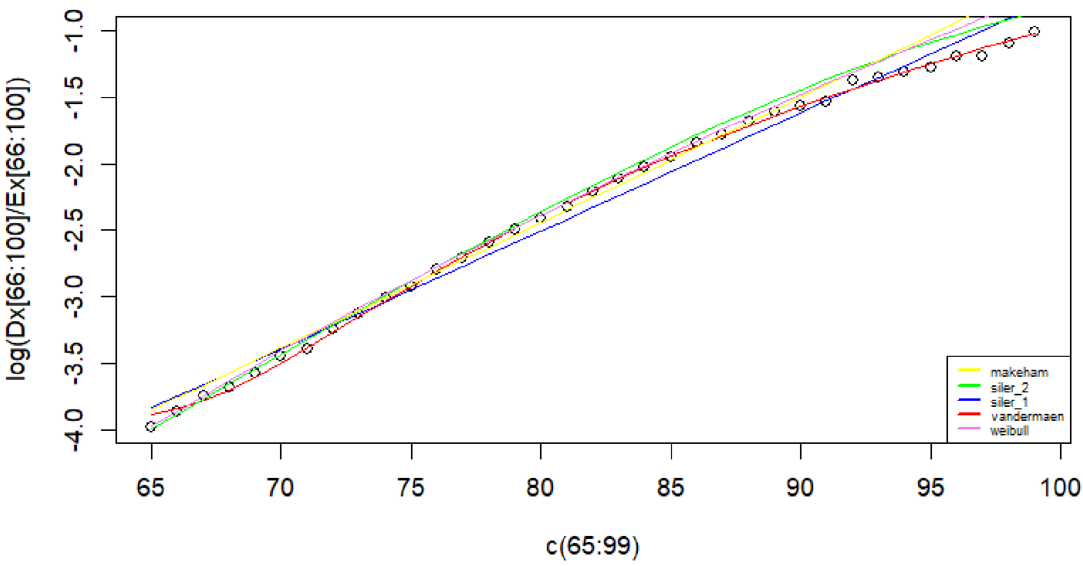
La Ley de Shang presenta una curva dinámica, en la que la mortalidad disminuye progresivamente en las edades iniciales para luego experimentar un incremento abrupto en las edades avanzadas. Este modelo incorpora un factor de modificación basado en parámetros ajustables, lo que le confiere flexibilidad para adaptarse a escenarios complejos o situaciones donde la vulnerabilidad puede variar significativamente.

Ley de Gompertz:

La Ley de Gompertz exhibe un comportamiento coherente con las observaciones empíricas sobre mortalidad en poblaciones envejecidas, caracterizándose por un aumento exponencial de la tasa de mortalidad a partir de cierta edad. Este crecimiento pronunciado refleja la aceleración en el riesgo de muerte conforme la edad avanza, posicionándose como un modelo fundamental en el análisis actuarial y demográfico.

5.9. ANÁLISIS DE AJUSTE DE LEYES DE MORTALIDAD EN EL MODELO ACTUARIAL

El gráfico que se presenta muestra una comparación entre diferentes leyes de mortalidad aplicadas a los datos de mortalidad de la población española en edades avanzadas (65-99 años). En este análisis, se ajustan varias leyes clásicas de mortalidad utilizando datos de defunciones (Dx) y población expuesta (Ex) para evaluar cómo diferentes modelos ajustan las tasas de mortalidad observadas.



Modelos Ajustados