



Universidad Autónoma de Querétaro
Facultad de Contaduría y Administración
Licenciatura en Actuaría

ESTIMACIÓN DE LA MORTALIDAD MEDIANTE MODELOS DE SUAVIZAMIENTO

Tesis Individual

Que como parte de los requisitos para obtener el grado de
Licenciado en Actuaría

Presenta:
Erick Morales Bonilla

Dirigido por:
Dra. Denise Gómez Hernández

Dra. Denise Gómez Hernández
Presidente

Firma

Dr. Michael Demmler
Secretario

Firma

Dr. Humberto Banda Ortiz
Vocal

Firma

Dr. Ignacio Almaraz Rodríguez
Suplente

Firma

Mtro. Roberto Yoan Castillo Diéguez
Suplente

Firma

D.G.T.I Martin Vivanco
Vargas
Director de la Facultad

C.P. Omar Bautista Hernández
Secretario Académico de la Facultad

Centro Universitario
Querétaro, Qro.
Junio de 2023
México



Dirección General de Bibliotecas y Servicios Digitales
de Información



Estimación de la mortalidad mediante modelos de
suavizamiento

por

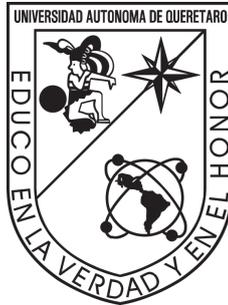
Erick Morales Bonilla

se distribuye bajo una [Licencia Creative Commons
Atribución-NoComercial-SinDerivadas 4.0 Internacional](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Clave RI: CALIN-287612

UNIVERSIDAD AUTÓNOMA DE QUERÉTARO

Facultad de Contaduría y Administración



**ESTIMACIÓN DE LA MORTALIDAD
MEDIANTE MODELOS DE SUAVIZAMIENTO**

Tesis para obtener el grado de

A C T U A R I O

Presenta

Erick Morales Bonilla

Dirigida por

Dra. Denise Gómez Hernández

Resumen

En este estudio, se abordan las limitaciones de las técnicas actuariales, al basarse en métodos de estimación que restringen la forma de funciones asociadas a la mortalidad. El uso de estos modelos conlleva el riesgo de no capturar completamente el verdadero comportamiento de las variables en estudio. Aprovechando los avances en la capacidad de procesamiento de las máquinas modernas, se explora la alternativa de modelar las probabilidades de muerte en la población de México mediante un enfoque de suavizamiento polinómico por nodos, también referido como Smoothing Splines. Los datos utilizados para entrenar el modelo abarcan el período de 2012 a 2021, obtenidos a partir de los registros de mortalidad general incluidos en el subsistema de información demográfica y social del Instituto Nacional de Estadística y Geografía (INEGI). Se contrastaron los resultados con la la Experiencia Demográfica de Mortalidad para Activos “EMSSA-09”, proporcionada por la Comisión Nacional de Seguros y Fianzas (CNSF). Los resultados obtenidos revelan que el procedimiento de suavizamiento por splines proporciona estimaciones sustancialmente diferentes a las dictaminadas por la CNSF, evidenciando una posible sobre estimación de mortalidad por parte de esta última.

Palabras clave: **Mortalidad, Estimación, Smoothing Splines, Suavizamiento polinómico por nodos, Fuerza de Mortalidad**

Abstract

This study addresses the limitations of actuarial techniques, which rely on estimation methods that constrain the form of functions associated with mortality. The use of these models carries the risk of not fully capturing the true behavior of the variables under study. Leveraging advancements in the processing power of modern machines, we explore the alternative of modeling death probabilities in the population of Mexico through Smoothing Splines. The data used to train the model covers the period from 2012 to 2021, obtained from the general mortality records included in the demographic and social information subsystem of the National Institute of Statistics and Geography (INEGI). The results were contrasted with the Demographic Experience of Mortality for Assets .^{EM}SSA-09,” provided by the National Insurance and Bonding Commission (CNSF). The findings reveal that the splines smoothing procedure provides substantially different estimates compared to those determined by the CNSF, indicating a possible overestimation of mortality by the latter.

Keywords: **Mortality, Estimation, Smoothing Splines, Force of Mortality**

Agradecimientos

Agradezco profundamente a mi directora de tesis, la Dra. Denise Gómez Hernández, cuyo constante apoyo y guía hicieron posible este trabajo. Su amabilidad y compromiso impactaron fuertemente en mi desarrollo profesional, y estoy sumamente agradecido de haber tenido la oportunidad de aprender de ella . Así mismo, agradezco sinceramente a mis sinodales, quienes dedicaron su tiempo y esfuerzo para asegurar la calidad de este escrito.

Me gustaría agradecer especialmente a mi madre, quien me impulsó a escribir este trabajo en primer lugar y siempre ha sido mi principal apoyo y fuente de motivación. Todo lo que soy y aspiro a ser se lo debo a ella.

A mi familia, les agradezco por su amor incondicional, y a mis amigos, por su apoyo constante y compañía a lo largo de este camino. Finalmente, quiero expresar mi gratitud a la Universidad Autónoma de Querétaro y a la Facultad de Contaduría y Administración, por proporcionar un entorno propicio para la investigación y el aprendizaje continuo.

Índice general

1. Introducción	1
1.1. Justificación	2
1.2. Planteamiento del problema	3
1.3. Objetivos	4
1.4. Hipótesis	5
2. Marco Teórico	6
2.1. Antecedentes	6
2.1.1. Matemáticas Actuariales	6
2.1.2. Modelos de Suavizamiento	10
2.2. Investigaciones Relacionadas	16
3. Características de la Investigación	21
3.1. Justificación	21
3.2. Planteamiento del problema	22
3.3. Objetivos	23
3.4. Definición del universo	23
3.5. Tamaño y tipo de la muestra	24
3.6. Definición de variables	24
3.7. Hipótesis	26
4. Metodología	27
4.1. Diseño de la Investigación	27
4.2. Tipo de estudio	27

4.3. Instrumento	28
4.4. Procesamiento de la información	30
5. Resultados y discusión	41
5.1. Características de la muestra	41
5.2. Resultados de las variables	45
5.3. Comprobación de hipótesis	48
5.4. Propuestas y Conclusiones	52
Bibliografía	55

Capítulo 1

Introducción

El comportamiento de la mortalidad es un factor socio económico clave, que permea en prácticamente todos los sectores de una población. De este comportamiento se derivan las probabilidades de muerte que, a su vez, funcionan como la piedra angular para los cálculos actuariales. Uno de los primeros intentos por modelar las probabilidades de fallecimiento fueron las tablas de mortalidad. Aunque algunos autores (Escuder Bueno y Escuder Vallés, 2009), consideran que no es posible dar con el origen exacto de la primera tabla de mortalidad, diversas fuentes proponen que la primera tabla “exacta” se creó en 1662 por John Graunt. No fue hasta un par de siglos después que se introdujeron las leyes de mortalidad. Estas leyes combinan los modelos actuariales de incertidumbre con fundamentos epidemiológicos, biológicos y demográficos. Algunos ejemplos de las leyes más utilizadas son las de De Moivre, Gompertz, y Makeham (Forfar, 2004b). Sin embargo, según (Booth y Tickle, 2008), aunque estas propuestas se acercan más a un rigor matemático de forma simple, aún contienen cierto grado de subjetividad. Por esta razón, a medida que la teoría estadística continuó desarrollándose, se comenzaron a proponer modelos estocásticos modernos que pudieran explicar la relación entre las defunciones y otras variables demográficas de una manera más sustentada en la observación de las poblaciones.

La importancia de estimar correctamente la mortalidad se ve reflejada directamente

en el cálculo de las pensiones, ya sea en el contexto privado, sindical, o de seguridad social. A causa de la insostenibilidad de los sistemas de reparto para las pensiones, en 1995 el Instituto Mexicano del Seguro Social (IMSS) optó por reformar el esquema de pensiones en México, instaurando un esquema de capitalización individual. Las pensiones tienen un alto impacto bilateral; por un lado, generan un considerable gasto para los organismos responsables de administrarlas, y por otro lado, determinan en gran parte la calidad de vida que tendrán sus beneficiarios cuando se encuentren en incapacidad laboral. Las principales variables que afectan la magnitud final de una pensión son la esperanza de vida, los rendimientos del ahorro acumulado, el porcentaje de contribución y la edad de retiro, entre otras (Villareal y Macías, 2020). En sus inicios, la construcción de una tabla de mortalidad era realizada mediante censos y encuestas dirigidas a un grupo específico de la población, con el propósito de registrar las defunciones del grupo. Sin embargo, dada la complejidad de los movimientos demográficos y la impracticabilidad de actualizar constantemente las tablas en uso, se comenzaron a buscar alternativas de construcción para las tablas. La Comisión Nacional de Seguros y Fianzas (CNSF), organismo responsable de aprobar y supervisar los seguros de pensiones en México, ha comenzado a incorporar modelos logísticos en la estimación y desarrollo de tablas de mortalidad para el sector asegurador (Ramírez et al., 2000). En el documento “Comparativo de Metodologías para el cálculo de Tablas de Mortalidad de Pensiones”, la CNSF también comparte otros procedimientos recomendados para pronosticar el comportamiento de la mortalidad, como el modelo de Lee-Carter, o un enfoque de series temporales mediante modelos autorregresivos (Comisión Nacional de Seguros y Fianzas, 2022).

1.1. Justificación

Desde la instauración de la profesión actuarial, se han desarrollado incontables productos y mecanismos de transferencia de riesgo con el propósito de aminorar las repercusiones económicas que los individuos enfrentan debido a la incerti-

dumbre de la mortalidad. Estas contra medidas operan en diversas dimensiones, abarcando desde lo micro económico, mediante seguros de vida por ejemplo, hasta lo macroeconómico, interviniendo en la seguridad social mediante los sistemas de pensiones. Por otro lado, este riesgo tiene una importancia igual o mayor para la entidad financiera que decide absorberlo, pues ésta debe cerciorarse de tener las facultades apropiadas de cumplir con las responsabilidades adquiridas, así como de contar con las técnicas actuariales adecuadas para tarificar, generar reservas, e incluso reasegurarse de ser necesario. La relevancia de este punto se puede ver reflejada en la reciente implementación del proyecto Solvencia II, una iniciativa para revisar y modernizar la regulación de los sistemas financieros aseguradores a nivel internacional (Contreras y Márquez, 2018).

Es así que la estabilidad financiera global depende, en gran medida, del correcto funcionamiento de esta contextura, y uno de sus pilares fundamentales es precisamente el entendimiento y uso preciso del comportamiento de la mortalidad. A medida que se desarrollan las bases matemáticas que dan lugar a la ciencia actuarial, las metodologías utilizadas para estimar la mortalidad también han cambiado y evolucionado a lo largo del tiempo. Razón por la cual, es importante estimular y contribuir a éste avance, explorando las posibles alternativas que existan para refinar estos procedimientos de estimación. En tanto que se incorporen técnicas de estimación más precisas, la capacidad de los actuarios para brindar seguridad a los asegurados y a las aseguradoras también aumentará significativamente.

1.2. Planteamiento del problema

Un modelo matemático es una representación simplificada de un proceso o fenómeno real; a pesar de ser de utilidad para comprender y analizar el comportamiento de dicho proceso, no lo asemejan completamente. Suele ser necesario ceder cierto nivel de “realismo”, a través de supuestos y simplificaciones, con el objetivo de obtener un sistema más sencillo y manejable. Como se verá en futuras secciones, la matemática actuarial no está exenta de estos sacrificios.

Las técnicas actuariales modernas suelen comenzar los métodos de estimación de probabilidades de muerte, restringiendo la forma de alguna de las funciones clásicas asociadas a la mortalidad, comúnmente la fuerza de mortalidad $\mu(x)$. Asemejando la ley de mortalidad de Gompertz-Makeham, en donde se asume que $\mu(x) = A + Bc^x$ (Bowers et al., 1986), se suele suponer que $\mu(x)$ tiene la forma de alguna función $h(x, \alpha_1, \dots, \alpha_p)$. El siguiente paso es entonces estimar las mejores propuestas para los parámetros utilizados en los supuestos, basándose en alguna metodología de optimización. Naturalmente, la flexibilidad del modelo dependerá en gran medida de la definición de h , y de la cantidad de grados de libertad p . Establecer de forma previa la forma de las funciones que se van a estimar permite agilizar significativamente la obtención de valores estimados, aunque esto conlleva una disminución en la flexibilidad del modelo.

En otras palabras, el uso de estos modelos implica, de manera implícita, la aceptación del riesgo de no capturar completamente el verdadero comportamiento de las variables en estudio. No obstante, gracias a los avances en la capacidad de procesamiento de las máquinas modernas, muchos de los cálculos necesarios para las estimaciones ahora pueden completarse en cuestión de segundos, lo que hace que el sacrificio previo resulte obsoleto. Este razonamiento conduce a la siguiente pregunta de investigación;

¿Los modelos más flexibles capturan con mayor precisión el comportamiento de la mortalidad demográfica en comparación con las metodologías clásicas?

1.3. Objetivos

Objetivo general:

Estimar las probabilidades de muerte sobre la población de México mediante un modelo de suavizamiento polinómico por nodos (Smoothing Splines).

Objetivos específicos:

- Definir las variables relacionadas a la matemática actuarial necesarias para

estimar probabilidades de muerte

- Determinar la metodología apropiada para estimar y construir tablas de vida
- Revisar los fundamentos y condiciones necesarias para utilizar un modelo de suavizamiento
- Contrastar los resultados de la estimación con otras metodologías utilizadas

1.4. Hipótesis

Las estimaciones resultantes de un suavizamiento sobre las probabilidades de muerte de la población de México capturan de un comportamiento diferente respecto a las estimaciones exhibidas en las tasas de mortalidad de activos para la seguridad social mexicana EMSSAH-09 y EMSSAM-09.

Capítulo 2

Marco Teórico

2.1. Antecedentes

2.1.1. Matemáticas Actuariales

Las matemáticas actuariales surgen como respuesta a las demandas de la sociedad por obtener seguridad financiera, y se constituyen como una rama de las ciencias matemáticas, combinando conocimientos probabilísticos, estadísticos y económicos. Aunque todavía no se conocían con este nombre, los primeros trabajos que incorporaron la ciencia actuarial se remontan al siglo 17, cuando Edmund Halley publicó en 1693 un texto donde expone como calcular anualidades de vida, seguido por otros académicos como De Moivre, James Dodson, Friedrich Gauss, entre otros (Forfar, 2004a). Sin embargo, no fue hasta 1848 que se estableció formalmente la profesión actuarial con la fundación del “Institute of Actuaries” en Reino Unido, responsable de supervisar que los candidatos tuvieran fundamentos en álgebra, probabilidad, interés compuesto, contabilidad, contingencias de vida, y uso de tablas de mortalidad.

En México, la técnica actuarial llega hasta 1899, con la creación de “La Anglo Mexicana de Seguros”, la primer compañía aseguradora en el país (Colegio Nacional de Actuarios, 2021). No obstante, ésta y las subsiguientes compañías de seguros

se sustentaban exclusivamente del trabajo de actuarios extranjeros. Así como Inglaterra, México funda el Instituto Mexicano de Actuarios (IMA) en 1937, y en 1947 se autoriza el primer plan de estudios para la carrera de Actuario en la Universidad Autónoma de México (UNAM).

Dentro de la matemática actuarial, se define la variable aleatoria continua X , como la edad de muerte en años de un recién nacido escogido de la población de forma arbitraria. Y se suelen denotar la función de distribución de X y la función de supervivencia como se ve en (2.1) y (2.2) respectivamente (Bowers et al., 1986).

$$F_X(x) = Pr(X \leq x) \quad x \geq 0 \quad (2.1)$$

$$s(x) = 1 - F_X(x) = Pr(X > x) \quad x \geq 0. \quad (2.2)$$

Sin embargo, se suele trabajar sobre una transformada de X , $T(x) = X - x$, que denota el tiempo restante de vida para un individuo de edad (x). Entonces, se definen las ecuaciones (2.3) y (2.4) (Bowers et al., 1986)

$${}_tq_x = Pr(T(x) \leq t) \quad t \geq 0 \quad (2.3)$$

$${}_tp_x = 1 - {}_tq_x = Pr(T(x) > t) \quad t \geq 0 \quad (2.4)$$

tal que, ${}_tq_x$ es la probabilidad de que un individuo de edad (x) fallezca dentro de los siguientes t años, y de forma similar, ${}_tp_x$ represente la probabilidad de (x) alcance la edad $x + t$. Aunque la notación utilizada para definir estas variables puede variar dependiendo el autor, se utilizarán aquellas notaciones propuestas por (Perryman, 1949), que han sido adoptadas por la “Society of Actuaries” en Estados Unidos, y el “Institute of Actuaries” en Inglaterra.

Ciertamente, existe una estrecha relación entre $T(x)$ y X , pues cuando se asume que el hecho de que X sobreviva a edad (x) no aporta ninguna información de mortalidad adicional a la que se tiene hoy sobre (x) (por ejemplo alguna enfermedad), entonces ambas variables se relacionan mediante la expresión (2.5).

$${}_tp_x = Pr(X > x + t | X > x) = \frac{s(x + t)}{s(x)}. \quad (2.5)$$

La última variable esencial que se define es la fuerza de mortalidad $\mu(x)$ (Bowers et al., 1986), expresada inicialmente como un límite en (2.6), pero pudiendo

simplificarse hasta obtener la expresión (2.7).

$$\begin{aligned}
 \mu(x) &= \lim_{h \rightarrow 0} \frac{Pr(x < X \leq x + h | X > x)}{h} & (2.6) \\
 &= \lim_{h \rightarrow 0} \frac{Pr(x < X \leq x + h)}{h \cdot Pr(X > x)} \\
 &= \lim_{h \rightarrow 0} \frac{F_X(x + h) - F_X(x)}{h \cdot (1 - F_X(x))} \\
 &= \frac{f_X(x)}{1 - F_X(x)} = \frac{-s'(x)}{s(x)}. & (2.7)
 \end{aligned}$$

De (2.6), se puede interpretar a $\mu(x)$ como la tasa que expresa la probabilidad de que un individuo muera un instante después del tiempo x , respecto a la duración del instante h . La igualdad (2.7) es equivalente a $\mu(y) = \frac{d}{dy} \ln(s(y))$, que se puede reescribir como la ecuación (2.8).

$$-\mu(y) dy = d \ln(s(y)). \quad (2.8)$$

Integrar ambos lados de (2.8) desde x hasta $x + n$ y aplicar la función exponencial al resultado, da lugar a la ecuación (2.9) (Bowers et al., 1986).

$${}_n p_x = \frac{s(x + n)}{s(x)} = \exp\left(- \int_x^{x+n} \mu(y) dy\right) \quad (2.9)$$

Otro tema fundamental en el ámbito de las matemáticas actuariales son las tablas de vida, las cuales ilustran la relación entre la probabilidad de fallecimiento de un individuo y su edad. Autores como Ortega (Ortega, 1987), proponen a John Graunt como el primer académico en formalizar las ideas sobre las tablas de vida, en su libro de 1662, "Natural and Political Observations made upon the Bills of Mortality", donde incluyó a 100 personas para su construcción. Aunque el siguiente avance relacionado a las tablas fue proporcionado por Edmond Halley en 1693, fue hasta 1815 que el actuario Joshua Milne incorporó herramientas actuariales modernas para elaborar una tabla de mortalidad para la ciudad de Carlisle. Con la llegada de la profesión actaurial a América Latina, también se comenzaron a desarrollar tablas de vida estimadas para cada país. En México, la CNSF reglamenta el uso de la tabla "EMSSA-09" para el cálculo de pensiones y seguros, según la disposición 14.2.6 de la Circular Única de Seguros y Fianzas.

Mediante las variables mencionadas previamente es que se puede definir la estructura apropiada que se le da a una tabla de mortalidad. El número de recién nacidos que se utiliza para construir la tabla se expresa con l_0 . Así mismo, el comportamiento de la mortalidad de cada recién nacido estará entonces gobernado por $s(x)$ (o bien $F(x)$). Se busca contabilizar el número de sobrevivientes para cada edad (x), para esto, se puede diferenciar a cada recién nacido mediante un índice, dígase, $j = 1, 2, \dots, l_0$. Si se define una variable indicadora como en la ecuación (2.10)

$$I_{j,x} = \begin{cases} 1, & \text{si la vida } j \text{ sobrevive a edad } x \\ 0, & \text{en otro caso} \end{cases} \quad (2.10)$$

entonces la suma reflejada en (2.11),

$$\mathcal{L}(x) := \sum_{j=1}^{l_0} I_{j,x} \quad (2.11)$$

expresa precisamente el número de sobrevivientes a edad (x). Además, al depender I_j de la supervivencia del individuo j a edad (x), su probabilidad es función de $s(x)$, y por lo tanto $E(I_{j,x}) = s(x)$. Si finalmente, se define la variable l_x como la esperanza de $E(\mathcal{L}(x))$, ésta se puede expresar como se muestra en (2.12).

$$l_x := E[u(x)] = \sum_{j=1}^{l_0} E[I_{j,x}] = l_0 \cdot s(x). \quad (2.12)$$

También es de utilidad contabilizar las defunciones ocurridas en intervalos de edades específicos, de forma que ${}_n\mathcal{D}_x$ se utiliza para expresar el número de muertes entre las edades (x) y ($x + n$). Análogo a las definiciones anteriores, si se define ${}_n d_x$ como la esperanza de ${}_n\mathcal{D}_x$, entonces también es posible reescribirla como la ecuación (2.13).

$$\begin{aligned} {}_n d_x &:= E[{}_n\mathcal{D}_x] = l_0 \cdot [s(x) - s(x + n)] \\ &= l_x - l_{x+n} \end{aligned} \quad (2.13)$$

Aunque las variables incluidas en una tabla de vida pueden variar, (Bowers et al., 1986) incluye en sus tablas didácticas las variables d_x, l_x , y $1000 \cdot q_x$, tabuladas

por cada año de edad (nótese que cuando $n = 1$ en ${}_nq_x$, y ${}_nd_x$, este prefijo se suele omitir). Por otro lado, la “EMSSA-09” solamente incluye la variable q_x tabulada para cada año de edad (Comisión Nacional de Seguros y Fianzas, 2023).

Figura 2.1: Tabla de Vida Ilustrativa. Adaptada de Bowers et al., 1986[p.675]

Age	l_x	d_x	1,000 q_x
0	100 000.00	2 042.1700	20.4217
1	97 957.83	131.5672	1.3431
2	97 826.26	119.7100	1.2237
3	97 706.55	109.8124	1.1239
4	97 596.74	101.7056	1.0421
5	97 495.03	95.2526	0.9770
6	97 399.78	90.2799	0.9269
7	97 309.50	86.6444	0.8904
8	97 222.86	84.1950	0.8660
9	97 138.66	82.7816	0.8522

Si se establece la edad $x = 5$, entonces la tabla refleja, al inicio del año 5 de edad, una cantidad esperada de sobrevivientes de $l_5 = 97,495.03$, una cifra de defunciones esperadas durante este año de $d_5 = 95.2526$ y una probabilidad de fallecer entre el año 5 y 6 de $q_5 = \frac{0.977}{1000}$.

2.1.2. Modelos de Suavizamiento

Los modelos de suavizamiento surgen como respuesta al deseo de modelar de manera precisa el comportamiento de una variable en estudio, logrando una imitación adecuada de sus patrones mientras se eliminan las perturbaciones que podrían considerarse como ruido (James et al., 2013). A pesar de que los análisis de regresión eran el enfoque preferido para establecer relaciones funcionales entre variables, se aspiraba a que el enfoque de suavizamiento incorporara una aproximación más flexible, dando lugar a metodologías innovadoras. No obstante, como se verá a continuación, se descubre que existe una relación estrecha entre los modelos de regresión lineal y el planteamiento de suavizamiento, siendo este último, un caso particular del primero.

El análisis de regresión es una herramienta estadística que permite estudiar la relación entre dos o más variables, y así, poder inferir resultados. Se suele atribuir

el desarrollo de esta rama de la estadística a Karl Pearson, sin embargo, el concepto de regresión fue ideado originalmente por Francis Galton (Stanton, 2001). A pesar de no utilizar los conocimientos estadísticos del momento, desarrollados por otros académicos como Guass y Laplace, Galton introduce el concepto de correlación a mediados del siglo 19 para auxiliarse en el estudio de la genética hereditaria, utilizando métodos estadísticos sencillos y empíricos (Estepa et al., 2012).

A causa de la carencia de rigor matemático en el trabajo de Galton, Karl Pearson se encargó de respaldar y formalizar sus ideas, ayudándose a su vez de otros académicos como Auguste Bravais, Adrien-Marie Legendre, entre otros, para desarrollar los aspectos algebraicos del análisis de regresión. Finalmente, en 1896, Pearson publica su primer texto sobre el tema, titulado “Mathematical contributions to the theory of evolution” (Stanton, 2001).

Después de más de 50 años, Isaac Jacob Schoenberg propone la metodología de Suavizamiento por “Splines” en 1964 (Rice y Rosenblatt, 1983).

$$\min_g \left\{ \sum (y_i - g(x_i))^2 + \lambda \cdot \int_{-\infty}^{\infty} g''(t)^2 dt \right\}. \quad (2.14)$$

Este planteamiento consiste en encontrar una relación funcional entre la variable objetivo Y , y una variable predictora X que asemeje adecuadamente el comportamiento de Y . Es decir, se desea una función cualquiera $g(x)$ que logre un error cuadrático $\sum_i (y_i - g(x_i))^2$ pequeño. No obstante, sin imponer restricciones adicionales a g , se puede proponer una función que imite en cada punto las observaciones y_i , minimizando el error cuadrático medio pero ocasionando un sobre ajuste. La manera de considerar también la variabilidad de g es mediante un componente extra, $\lambda \cdot \int_{-\infty}^{\infty} g''(t)^2 dt$, el cual penaliza y regula la flexibilidad de la función g encontrada (James et al., 2013). A medida que se incrementa el valor de λ , mayor es la penalización que se le da a cada ajuste dependiendo de qué tan “suave” es. De esta forma, se obtiene la expresión (2.14).

Es decir, del espacio de funciones doblemente diferenciables, se está buscando aquella que logre obtener el mejor equilibrio entre imitar la variable Y y exhibir un comportamiento de baja variabilidad.

A este modelo se le bautiza como Suvaziamiento por Splines, al descubrir que la función g que minimiza el cálculo es precisamente un Spline Natural Cúbico (de orden 3) con n nodos, donde n denota el número de observaciones en la variable X (Pollock, 1999).

Un Spline Natural Cúbico es el resultado de una Regresión Polinómica por Partes modificada. Supóngase que se desea ajustar un polinomio de grado 3, a cada segmento de la variable X , dividida en $k + 1$ intervalos (o equivalentemente, k nodos ubicados en ξ_i , para $i = 1, \dots, k$). Entonces se deberán de considerar $k + 1$ diferentes polinomios, cada uno con 4 coeficientes de regresión, para un total de $(4) \cdot (k + 1)$ coeficientes de regresión en el modelo (Adresi et al., 2020).

Esta estructura de modelo recibe el nombre de “Regresión Polinómica por Partes” (Adresi et al., 2020), y su forma se puede ver en la igualdad (2.15).

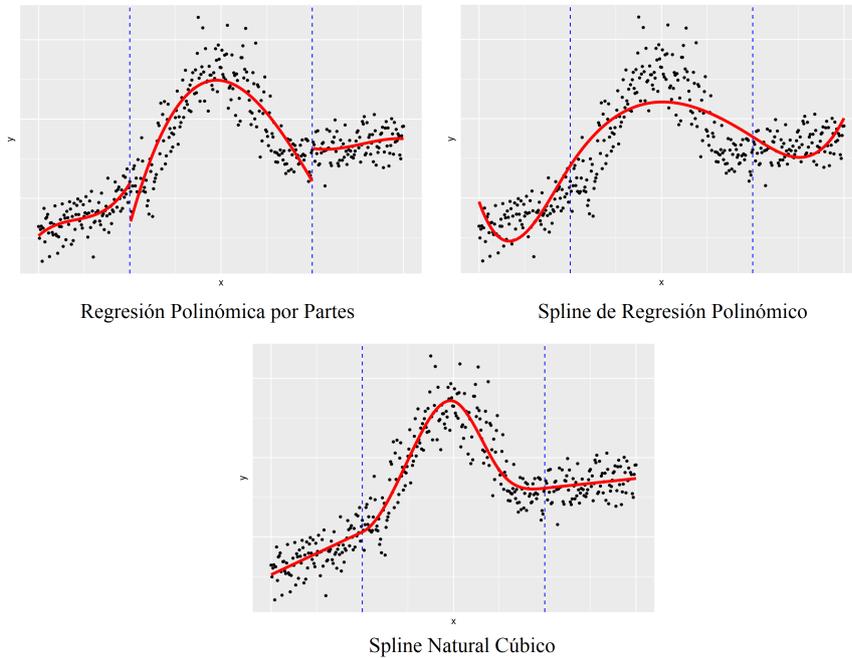
$$Y = \begin{cases} \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \epsilon, & x < \xi_1 \\ (\beta_0 + \beta_4) + (\beta_1 + \beta_5)x + \dots + (\beta_3 + \beta_7)x^3 + \epsilon, & \xi_1 \leq x < \xi_2 \\ \dots & \\ (\beta_0 + \beta_{4k}) + (\beta_1 + \beta_{4k+1})x + \dots + (\beta_3 + \beta_{4k+3})x^3 + \epsilon, & \xi_k \leq x \end{cases} \quad (2.15)$$

Aunque este modelo goza de una gran flexibilidad, la curva ajustada resultante suele ser discontinua en los nodos, ocasionando un sobre ajuste en el modelo. Para alcanzar un balance entre estos dos escenarios se añaden restricciones que garanticen una curva “suave”. Un “Spline de Regresión Polinómico” de orden 3 con k nodos, es una regresión polinómica por partes en donde la curva, y sus primeras 2 derivadas son continuas en todos los nodos (Neter y Wasserman, 1997). Esta última condición no solo garantiza una curva “suave”, sino que también disminuye la cantidad de parámetros por estimar a $k + 3$. Aunque es posible generalizar el orden del polinomio que se ajusta en cada sección, se acostumbra utilizar uno de grado 3, y en su lugar aumentar o disminuir la cantidad de nodos

utilizados para particionar la variable X , variando así la flexibilidad del ajuste (James et al., 2013).

Finalmente, se obtiene un “Spline Natural Cúbico” al modificar los extremos del ajuste de un Spline de Regresión Polinómico; en lugar de utilizar un polinomio en el primer y último segmento de X , se ajusta una línea. Este cambio permite corregir la divergencia que exhiben las funciones polinomiales en sus extremos (James et al., 2013). Se puede visualizar las diferencias de estos modelos en las gráficas [2.2].

Figura 2.2: Comparativa de modelos de ajuste.



La manera de obtener los valores para los coeficientes de regresión es completamente análoga al planteamiento de Regresión Lineal Múltiple, pues se ajusta una relación lineal en los extremos de X , y en las secciones restantes, se ajusta una relación polinomial de la forma expuesta en (2.16).

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \cdots + \beta_p X_i^p + \epsilon_i. \quad (2.16)$$

La ecuación (2.16) se puede reescribir mediante las llamadas “funciones base” (Neter y Wasserman, 1997) $b_j(X)$, las cuales permiten modelar relaciones no

lineales entre Y y X como se muestra en (2.17)

$$Y_i = \beta_0 + \beta_1 b_1(X_i) + \cdots + \beta_p b_p(X_i) + \epsilon_i. \quad (2.17)$$

Si ahora se define $X_j = b_j(X) = X^j$, entonces se obtiene la ecuación (2.18), la cual corresponde a la estructura clásica de un modelo de Regresión Lineal Múltiple, cuya teoría y metodología se repasará brevemente.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \epsilon. \quad (2.18)$$

Se suele suponer (Neter y Wasserman, 1997) que :

- La variable ϵ es aleatoria, tal que $E[\epsilon] = 0$ y $Var[\epsilon] = \sigma^2$ desconocida.
- Las variables X_p no son variables aleatorias.
- Cada observación de ϵ es independiente entre sí, o bien, $Cov[\epsilon_i, \epsilon_j] = 0$, para toda $i \neq j$.
- El conjunto de variables predictoras $X_{i,1}, \dots, X_{i,p}$, es linealmente independiente.

Estos supuestos permiten obtener rápidamente el valor esperado de la variable objetivo, mostrado en la ecuación (2.19).

$$E[Y_i] = \beta_0 + \beta_1 X_{i,1} + \cdots + \beta_p X_{i,p}. \quad (2.19)$$

Apoyándose de este resultado, se define la estimación $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \cdots + \hat{\beta}_p X_p$, y se busca encontrar aquellos valores de $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p$ que minimicen la variación entre la estimación propuesta y el valor real de cada observación en la variable respuesta. Este objetivo es equivalente a plantear la expresión (2.20).

$$\min_{\hat{\beta}_0, \dots, \hat{\beta}_p} \left\{ \sum (y_i - \hat{y}_i)^2 = \sum (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i,1} - \cdots - \hat{\beta}_p x_{i,p})^2 \right\}. \quad (2.20)$$

Dada la creciente complejidad de notación, se frecuenta utilizar notación matricial para desarrollar los resultados del modelo (Sengupta, 2001). Si se tienen m observaciones, a cada Y_i , $i = 1, \dots, m$ se le puede asociar el resultado de una multiplicación de matrices, como se ve en la igualdad (2.21).

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_m \end{pmatrix} = \begin{pmatrix} 1 & X_{1,1} & X_{1,2} & \dots & X_{1,p} \\ 1 & X_{2,1} & X_{2,2} & \dots & X_{2,p} \\ \dots & & & & \\ 1 & X_{m,1} & X_{m,2} & \dots & X_{m,p} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \epsilon_0 \\ \epsilon_1 \\ \dots \\ \epsilon_m \end{pmatrix} \quad (2.21)$$

Si se utilizan las variables \mathbf{Y} , \mathbf{X} , $\boldsymbol{\beta}$ y $\boldsymbol{\epsilon}$ para denotar las respectivas matrices de (2.21), entonces esta expresión es equivalente a la ecuación (2.22).

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (2.22)$$

Es demostrable que los valores de la matriz de estimadores de los coeficientes de regresión, definida como se ve en (2.23),

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \dots \\ \hat{\beta}_p \end{pmatrix} \quad (2.23)$$

alcanzan un mínimo en (2.20) (Sengupta, 2001) cuando se obtienen de la ecuación (2.24).

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} \quad (2.24)$$

Donde \mathbf{X}^T se refiere a la matriz transpuesta de \mathbf{X} y el operador $(\mathbf{A})^{-1}$ denota la matriz inversa de \mathbf{A} . Notar que esta expresión permite reescribir la matriz de estimaciones como la expresión (2.25), donde se suele referir a $\mathbf{H} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$ como la matriz de proyección (Sengupta, 2001).

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} = \mathbf{H}\mathbf{Y} \quad (2.25)$$

Este proceso de minimización se repite para obtener los diferentes coeficientes estimados en cada intervalo de X . En el caso de un Spline Natural Cúbico, se debe de resolver este problema de minimización sujeto a las condiciones de continuidad sobre las primeras dos derivadas en cada nodo.

Por otra parte, una de las maneras en las que se puede encontrar el mejor valor para el parámetro de suavizamiento λ en (2.14), es mediante una metodología

de selección de modelo llamada “Leave One Out Cross Validation” (LOOCV) (James et al., 2013), que consiste en construir el modelo excluyendo un par de observaciones (y_i, x_i) , y después utilizar estos valores excluidos para obtener el error cuadrático medio del modelo $ECM_i = (y_i - g_i(x_i))^2$, repitiendo este proceso para cada par de observaciones disponibles. El error cuadrático medio total resultante de esta metodología será el promedio de todos los errores cuadráticos medios individuales, cuyo cálculo se expone en la ecuación (2.26).

$$ECM_{LOOCV} = \frac{1}{n} \sum_i ECM_i. \quad (2.26)$$

Sin embargo, se demuestra en (James et al., 2013), que en el caso de la Regresión Lineal Múltiple se puede obtener este valor por medio de los valores de la traza de la matriz de proyección \mathbf{H} , h_i , como se ve en la expresión (2.27).

$$ECM_{LOOCV} = \frac{1}{n} \sum_i^n \left(\frac{y_i - \hat{y}_i}{1 - h_i} \right)^2 \quad (2.27)$$

Se suele proponer una malla de valores para λ , donde el valor escogido será aquel que alcance el menor ECM_{LOOCV} de entre todos los demás propuestos (James et al., 2013).

2.2. Investigaciones Relacionadas

Los estándares de la práctica actuarial en México para aquellas metodologías relacionadas con cualquier producto de vida y no vida se encuentran plasmados en la Circular Única de Seguros y Fianzas (CUSF), que a su vez es emitida por la Comisión Nacional de Seguros y Fianzas (CNSF). En ella se pueden encontrar lineamientos para el cálculo y valuación de productos de vida relacionados con la mortalidad, destacando en diversos artículos el uso de la Experiencia Demográfica de Mortalidad para Activos “EMSSA-09” (Comisión Nacional de Seguros y Fianzas, 2023), una serie de tablas de vida anexadas segmentadas por género, capacidad y situación laboral, entre otros.

A pesar de incorporar metodologías para la proyección de tasas de mortalidad en años futuros (basándose en aquellas del 2009), no se mencionan los procedimientos de estimación para las tasas base expuestas a lo largo del anexo. En 2022, la CNSF liberó a través de su revista “Actualidad en Seguros y Fianzas” un artículo titulado “Comparativo de Metodologías para el cálculo de Tablas de Mortalidad de Pensiones”, donde se menciona que el modelo utilizado por la CNSF en publicaciones anteriores para estimar la mortalidad mexicana es el de Análisis Bayesiano de Modelos de Regresión Logística (Comisión Nacional de Seguros y Fianzas, 2022).

Según redacta el artículo, el fin último del modelo de regresión es obtener una estimación para las tasas de mortalidad q_x . Los datos utilizados para entrenar el modelo provienen de la experiencia en defunciones del registro de pólizas de la CNSF, aunque no se hace mención específica del producto o programa al cuál están suscritos dichos individuos.

Para cada edad $x = 0, 1, \dots, 110$, se definen tres variables (Comisión Nacional de Seguros y Fianzas, 2022),

- $P_{x+1} :=$ Pólizas o expuestos a edad $x + 1$ al final del año.
- $d_x :=$ Muertes de edad x observadas durante el año.
- $E_x :=$ Expuestos de edad x al principio del año.

Estas se encuentran relacionadas mediante la expresión (2.28),

$$E_x = P_{x+1} + d_x \tag{2.28}$$

que permite calcular, desde un punto de vista frecuentista, la tasa bruta de mortalidad observada, de la forma vista en (2.29).

$$q_x = \frac{d_x}{E_x} \tag{2.29}$$

La relación no lineal que se observa entre la edad x y la tasa bruta de mortalidad calculada q_x sustenta el uso de un modelo de regresión para explicar el comportamiento entre ambas variables. Dado que los valores de q_x se encuentran entre

0 y 1, es necesario utilizar el enfoque de un modelo de Regresión Logística, que pertenece al planteamiento de Regresión Lineal Generalizada. Se menciona el uso de la función de enlace Logit, que describe la relación mostrada en (2.30) (Ramírez et al., 2000).

$$Y_i = \ln \left(\frac{\hat{q}_x}{1 - \hat{q}_x} \right) = \beta_0 + \beta_1 X_i + \epsilon_i. \quad (2.30)$$

Un modelo alternativo que se propone en el documento es el de Lee-Carter, el cual se describe como una combinación de un modelo extrapolativo y series de tiempo. Esto se logra buscando el ajuste reflejado en (2.31), (Ramírez et al., 2000)

$$q_{x_t} = \exp(\alpha_x + \beta_x \cdot k_t + \epsilon_{x_t}) \quad (2.31)$$

donde t denota el tiempo y x_t diferencia cada cohorte generacional, entonces (Ramírez et al., 2000)

- α_x es un parámetro independiente del tiempo, e indica aquellos comportamientos que permanecen constantes en la mortalidad.
- β_x regula el peso que recibe el cambio de la tendencia de mortalidad a lo largo del tiempo, k_t .
- k_t es el parámetro que expresa las desviaciones en el comportamiento de la mortalidad en el tiempo para cada edad específica.

Se menciona que es posible encontrar una solución al planteamiento usando descomposición en valores singulares, siempre y cuando se impongan además las restricciones $\sum \beta_x = 1$ y $\sum k_t = 0$. Por otro lado, al ser k_t dependiente sólo del tiempo, se suele analizar mediante modelos de series temporales, específicamente, modelos Autorregresivos Integrados de Medias Móviles (ARIMA) (Ramírez et al., 2000), los cuales permiten estudiar tendencias no estacionarias en el tiempo.

Pese a no ser una metodología mencionada en las regulaciones de la CNSF, también existen investigaciones enfocadas en las aplicaciones de los Splines sobre la estimación de la mortalidad. Se puede consultar en (Oirov et al., 2021), una estimación sobre la fuerza de mortalidad de la población de Mongolia, mediante ajustes de Splines. Se debe mencionar sin embargo, que este procedimiento no es el mismo

que el de un modelo de suavizamiento, pues la cantidad de nodos, su ubicación, y los coeficientes de regresión son obtenidos mediante máxima verosimilitud. Se denota al vector que contiene los k nodos como $\mathbf{c} = (c_1, c_2, \dots, c_k)$ y de manera similar, al vector de coeficientes de regresión como $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_{k+n})$, donde d expresa el orden del Spline. De esta manera, cada Spline se puede definir como una función de estos dos vectores, $g(x, \boldsymbol{\theta}, \mathbf{c})$.

Como la fuerza de mortalidad $\mu(x)$, debe satisfacer que $\mu(x) \geq 0$, para todo valor de x , se utiliza el logaritmo natural de ésta, $\ln(\mu(x))$, cuya imagen corresponde a todos los valores reales. Finalmente, el planteamiento del modelo se muestra en (2.32), donde se iguala la fuerza de mortalidad al spline obtenido por máxima verosimilitud (Oirov et al., 2021),

$$\ln(\mu(x)) = g(x, \boldsymbol{\theta}, \mathbf{c}) \quad (2.32)$$

o equivalentemente, $\mu(x) = \exp\{g(x, \boldsymbol{\theta}, \mathbf{c})\}$.

A diferencia del Spline obtenido de minimizar (2.26), Oirov utiliza una prueba ji-cuadrada de bondad de ajuste modificada para obtener la forma del Spline ajustado. Es posible relacionar $\mu(x)$ con la función $F_X(x)$ mencionada en (2.1), mediante la expresión $F_X(x) = 1 - \exp[-\int_0^x \mu(t)dt]$ (Bowers et al., 1986), de manera que $F(x, \boldsymbol{\theta})$ denota a la función de distribución obtenida de calcular $\mu(x) = \exp\{g(x, \boldsymbol{\theta}, \mathbf{c})\}$. La hipótesis nula de la prueba de bondad de ajuste se convierte en la igualdad (2.33),

$$H_0 : F_X(x) = F(x, \boldsymbol{\theta}), \quad (2.33)$$

donde $F_X(x)$ es la función de distribución verdadera de la vida X (Oirov et al., 2021). Los valores asignados a los vectores $\boldsymbol{\theta}$ y \mathbf{c} serán aquellos que maximicen la probabilidad de aceptar H_0 .

Para comparar los valores estimados de $\mu(x)$, se utiliza la función de distribución empírica $\hat{F}_X(x)$, definida como se ve en la ecuación (2.34) para la cual se define $n_x, x = 1, 2, 3, \dots, 100$ como el número de muertes ocurridas entre las edades

$[x - 1, x]$, y $N = \sum_x n_x$.

$$\hat{F}_X(x) = \frac{n_1 + n_2 + \cdots + n_x}{N}, x = 1, 2, \dots, 100. \quad (2.34)$$

Esta función empírica después se utiliza para obtener otra estimación para la fuerza de mortalidad de la forma (2.35),

$$\mu(\hat{x}) = \frac{\hat{F}_X(x) - \hat{F}_X(x - 1)}{1 - \hat{F}_X(x)}, x = 1, 2, \dots, 100, \quad (2.35)$$

que asemeja (2.6) utilizando $h = 1$.

Capítulo 3

Características de la Investigación

3.1. Justificación

Desde la instauración de la profesión actuarial, se han desarrollado incontables productos y mecanismos de transferencia de riesgo con el propósito de aminorar las repercusiones económicas que los individuos enfrentan debido a la incertidumbre de la mortalidad. Estas contra medidas operan en diversas dimensiones, abarcando desde lo micro económico, mediante seguros de vida por ejemplo, hasta lo macroeconómico, interviniendo en la seguridad social mediante los sistemas de pensiones. Por otro lado, este riesgo tiene una importancia igual o mayor para la entidad financiera que decide absorberlo, pues ésta debe cerciorarse de tener las facultades apropiadas de cumplir con las responsabilidades adquiridas, así como de contar con las técnicas actuariales adecuadas para tarificar, generar reservas, e incluso reasegurarse de ser necesario. La relevancia de este punto se puede ver reflejada en la reciente implementación del proyecto Solvencia II, una iniciativa para revisar y modernizar la regulación de los sistemas financieros aseguradores a nivel internacional (Contreras y Márquez, 2018).

Es así que la estabilidad financiera global depende, en gran medida, del correcto funcionamiento de esta contextura, y uno de sus pilares fundamentales es precisamente el entendimiento y uso preciso del comportamiento de la mortalidad. A

medida que se desarrollan las bases matemáticas que dan lugar a la ciencia actuarial, las metodologías utilizadas para estimar la mortalidad también han cambiado y evolucionado a lo largo del tiempo. Razón por la cual, es importante estimular y contribuir a éste avance, explorando las posibles alternativas que existan para refinar estos procedimientos de estimación. En tanto que se incorporen técnicas de estimación más precisas, la capacidad de los actuarios para brindar seguridad a los asegurados y a las aseguradoras también aumentará significativamente.

3.2. Planteamiento del problema

Un modelo matemático es una representación simplificada de un proceso o fenómeno real; a pesar de ser de utilidad para comprender y analizar el comportamiento de dicho proceso, no lo asemejan completamente. Suele ser necesario ceder cierto nivel de “realismo”, a través de supuestos y simplificaciones, con el objetivo de obtener un sistema más sencillo y manejable. Como se ha evidenciado en secciones anteriores, la matemática actuarial no está exenta de estos sacrificios.

Las técnicas actuariales modernas suelen comenzar los métodos de estimación de probabilidades de muerte, restringiendo la forma de alguna de las funciones clásicas asociadas a la mortalidad, comúnmente la fuerza de mortalidad $\mu(x)$. Asemejando la ley de mortalidad de Gompertz-Makeham, en donde se asume que $\mu(x) = A + Bc^x$ (Bowers et al., 1986), se suele suponer que $\mu(x)$ tiene la forma de alguna función $h(x, \alpha_1, \dots, \alpha_p)$. El siguiente paso es entonces estimar las mejores propuestas para los parámetros utilizados en los supuestos, basándose en alguna metodología de optimización. Naturalmente, la flexibilidad del modelo dependerá en gran medida de la definición de h , y de la cantidad de grados de libertad p . Establecer de forma previa la forma de las funciones que se van a estimar permite agilizar significativamente la obtención de valores estimados, aunque esto conlleva una disminución en la flexibilidad del modelo.

En otras palabras, el uso de estos modelos implica, de manera implícita, la aceptación del riesgo de no capturar completamente el verdadero comportamiento de

las variables en estudio. No obstante, gracias a los avances en la capacidad de procesamiento de las máquinas modernas, muchos de los cálculos necesarios para las estimaciones ahora pueden completarse en cuestión de segundos, lo que hace que el sacrificio previo resulte obsoleto. Este razonamiento conduce a la siguiente pregunta de investigación;

¿Los modelos más flexibles capturan con mayor precisión el comportamiento de la mortalidad demográfica en comparación con las metodologías clásicas?

3.3. Objetivos

Objetivo general:

Estimar las probabilidades de muerte sobre la población de México mediante un modelo de suavizamiento polinómico por nodos (Smoothing Splines).

Objetivos específicos:

- Definir las variables relacionadas a la matemática actuarial necesarias para estimar probabilidades de muerte
- Determinar la metodología apropiada para estimar y construir tablas de vida
- Revisar los fundamentos y condiciones necesarias para utilizar un modelo de suavizamiento
- Contrastar los resultados de la estimación con otras metodologías utilizadas

3.4. Definición del universo

Este estudio se basó en datos históricos de la población general de México, restringiendo el análisis a individuos cuyo fallecimiento fue registrado en el período comprendido entre 2012 y 2021. En búsqueda de obtener un comportamiento generalizado a nivel nacional y mantener la posibilidad de comparación con

referencias existentes, se admitieron registros provenientes de las 32 entidades federativas de México.

Así mismo, no se discriminó la muestra por ninguna causa de defunción ni por características sociales específicas, exceptuando el sexo de nacimiento. Como se expuso en el marco teórico de la presente investigación, las probabilidades de muerte utilizadas en la matemática actuarial incorporan una amplia gama de factores socio-demográficos característicos de la población en su conjunto y no están limitadas a causas de defunción particulares.

3.5. Tamaño y tipo de la muestra

La base de datos estuvo compuesta por un total de 7, 583, 482 registros de defunción en el periodo mencionado. Las selecciones de los registros son independientes entre sí y representativas de la población, sin embargo, se realizaron esfuerzos para mejorar la integridad de los datos mediante la eliminación de registros corruptos o incompletos. Los datos obtenidos de censos suelen incorporar errores de captura debido a factores como la imprecisión humana o canales de comunicación ineficientes para obtener estos datos. Para garantizar la calidad y confiabilidad de la información, se realizó un proceso de depuración de datos para identificar y eliminar los registros problemáticos, reduciendo así el tamaño final de la muestra a 6, 865, 137 registros.

3.6. Definición de variables

- **Variable Independiente:** Año de registro de muerte (r). Se define como el número natural que representa el año en el cual fue registrado el fallecimiento para cada individuo.
- **Variable Independiente:** Sexo de nacimiento (z). Se define como el género biológico asignado al momento de nacer para cada individuo, “Masculino” o

“Femenino”.

- **Variable Dependiente:** Número de sobrevivientes a edad discreta k , con defunción registrada el año r y sexo z ($\mathcal{L}(k)_{r,z}$). Se define como el número de individuos que alcanzan la edad discreta k , con sexo z y cuya defunción fue registrada el año r .

Covariables:

- Número de registros, para el año r con sexo z ($n_{r,z}$). Número entero, que representa el número de registros disponibles que pertenezcan a la cohorte de defunciones registradas el año r y que tengan sexo z .
- Edad de muerte, para el individuo i , con año registrado r y sexo z ($X_{i,r,z}$). Número real mayor o igual a 0, que representa la edad alcanzada en años del individuo $i = 1, \dots, n_{r,z}$, con sexo z y cuya defunción fue registrada el año r .
- Edad (x). Se define como el número real que representa los posibles valores continuos que puede tomar la variable Edad de muerte $X_{i,r,z}$.
- Edad discreta (k). Se define como el número entero que representa los posibles valores discretos que puede tomar la variable Edad de muerte $X_{i,r,z}$.
- Función suavizada del número de sobrevivientes a edad x , para el año registrado r y sexo z ($\hat{l}(x)_{r,z}$). Se define como el suavizamiento de la función $\mathcal{L}(k)_{r,z}$ para la edad x , estimada mediante Suavizamiento por Splines.
- Función estimada de supervivencia a edad x , para el año registrado r y sexo z ($\hat{s}(x)_{r,z}$). Se define como la función de supervivencia para la edad x , estimada a través de $\hat{l}(x)_{r,z}$.
- Fuerza de mortalidad a edad x , para el año registrado r y sexo z ($\hat{\mu}(x)_{r,z}$). Se define como la fuerza de mortalidad para la edad x , estimada a través de $\hat{s}(x)_{r,z}$.

3.7. Hipótesis

Las estimaciones resultantes de un suavizamiento sobre las probabilidades de muerte de la población de México capturan de un comportamiento diferente respecto a las estimaciones exhibidas en las tasas de mortalidad de activos para la seguridad social mexicana EMSSAH-09 y EMSSAM-09.

Capítulo 4

Metodología

4.1. Diseño de la Investigación

De acuerdo con (Sampieri Hernández, 2014), una investigación cuantitativa se considera como no experimental cuando se observan escenarios ya existentes, no provocados por el estudio que se está realizando. Al utilizar las estadísticas de defunciones generales proporcionadas por el INEGI, no fue posible manipular las variables independientes que pudieran haber influido en el registro de muertes obtenido. Las relaciones y comportamientos identificados son aquellos que surgieron en el contexto natural de las variables, por lo cual esta investigación se considera de este tipo, cuantitativa no experimental.

4.2. Tipo de estudio

A diferencia de los estudios transeccionales, en donde se busca analizar información proveniente de un sólo punto temporal (Sampieri Hernández, 2014), la muestra de esta investigación incorporó datos recopilados desde el año 2012 hasta el año 2021. Aunque el enfoque principal de este estudio no era identificar una tendencia en el cambio anual del comportamiento de la mortalidad, se consideraron todos los comportamientos observados en esta ventana temporal. Admitir estos registros

permitió obtener resultados estadísticamente robustos, que reflejan la evolución de la mortalidad a lo largo del tiempo. Esta recolección de datos a lo largo de diferentes períodos, caracterizó el diseño del estudio como uno de tipo longitudinal. (Sampieri Hernández, 2014).

4.3. Instrumento

La totalidad de la muestra se obtuvo mediante la inclusión de los registros de mortalidad general correspondientes al período de tiempo entre los años 2012 y 2021. Estos datos forman parte del subsistema de información demográfica y social, dentro del apartado de estadísticas de mortalidad, perteneciente al INEGI, cuya fuente de origen se puede consultar en (Instituto Nacional de Estadística y Geografía, 2021). En conformidad con la Norma Técnica para el Acceso y Publicación de Datos Abiertos de la Información Estadística y Geográfica de Interés Nacional (Instituto Nacional de Estadística y Geografía, 2014), el investigador responsable se comprometió a utilizar estos registros de manera ética y respetuosa, garantizando la confidencialidad y la privacidad de los datos cuando correspondió.

Según lo detallado en la Síntesis Metodológica para la Estadística de Defunciones Generales (Instituto Nacional de Estadística y Geografía, 2014), el método de captación de datos utilizado por la INEGI fue el aprovechamiento de registros administrativos pertenecientes al sistema de Registros Civil. El proceso comienza con la emisión del certificado de defunción, autorizado por la Secretaría de Salud de México, en el momento del fallecimiento de cada individuo. Posteriormente, interviene el Ministerio Público mexicano para investigar las características particulares de muerte y transcribirlas al “cuaderno de defunciones”. Es así, que el certificado de defunción fue la fuente principal de información para los registros, y en caso de presentar datos parciales, se complementó con el cuaderno de defunciones.

A partir de estos documentos, se extrajo un total de 59 variables plasmadas en un archivo (Instituto Nacional de Estadística y Geografía, 2021). Del mismo modo, se crea, para cada año de registro, un diccionario de datos que permitió consultar

el tipo de dato y el significado de cada variable. A continuación se mencionan las principales variables de interés para el estudio:

- **Entidad de Registro** (Cadena de texto comprendida entre los caracteres “01,02,...,32” que identifica a cada estado de la República Mexicana en donde se registró el fallecimiento del individuo.)
- **Entidad de ocurrencia** (Cadena de texto comprendida entre los caracteres “01,02,...,32,99” que identifica a cada estado de la República Mexicana en donde ocurrió el fallecimiento del individuo. El texto “99” denota una entidad no especificada.)
- **Causa de la defunción (lista mexicana)** (Cadena de texto comprendida entre los caracteres “01,02,...,61” que identifica el tipo de causa de defunción del individuo.)
- **Sexo del (la) fallecido (a)** (Número entero que toma los valores 1, 2 y 9 que identifica el sexo de nacimiento del individuo como masculino, femenino, y no especificado, correspondientemente.)
- **Día de ocurrencia** (Número entero que toma los valores 1, ..., 31, 99 que identifica el día de defunción del individuo. El número 99 denota un día no especificado.)
- **Mes de ocurrencia** (Número entero que toma los valores 1, ..., 12, 99 que identifica el mes de defunción del individuo. El número 99 denota un mes no especificado.)
- **Año de ocurrencia** (Número entero que identifica el año de defunción del individuo. El número 9999 denota un año no especificado.)
- **Día de nacimiento** (Número entero que toma los valores 1, ..., 31, 99 que identifica el día de nacimiento del individuo. El número 99 denota un día no especificado.)
- **Mes de nacimiento** (Número entero que toma los valores 1, ..., 12, 99 que identifica el mes de nacimiento del individuo. El número 99 denota un mes

no especificado.)

- **Año de nacimiento** (Número entero que identifica el año de nacimiento del individuo. El número 9999 denota un año no especificado.)
- **Ocupación del (la) fallecido (a)** (Número entero que toma los valores 1, ..., 10, 11, 97, 98, 99 que identifica el tipo de ocupación del individuo.)

La INEGI indica que la estructura seleccionada para el almacenamiento de los registros se basó en los lineamientos recomendados por la Organización de las Naciones Unidas, tal como se establece en su documento, “Principios y recomendaciones para un sistema de estadísticas vitales” (Instituto Nacional de Estadística y Geografía, 2014).

4.4. Procesamiento de la información

Los datos se procesaron a través del lenguaje de programación “R”, dada su eficiencia para manejar amplias bases de información y su compatibilidad con el tratamiento estadístico. Las librerías utilizadas para llevar a cabo los cálculos fueron “tidyverse”, “lubridate”, “dplyr”, “stats”, y “ggplot2”, las cuales pueden ser instaladas mediante la siguiente estructura de código;

```
install.packages("libreria")  
library("libreria")
```

El objetivo principal de este estudio consistió en transformar los registros de defunción en una tabla de vida para cada año de fallecimiento registrado disponible. Además, se segmentaron las tablas por género del individuo, lo que dio como resultado un total de 20 tablas distintas. Se denotaron el año de registro y el género de nacimiento con las variables $r = 2012, 2013, \dots, 2021$, y $z = 1, 2$ (donde 1 representa el sexo masculino y 2 el femenino), que a su vez, permitieron definir las variables “ $DF_{r,z}$ ”, donde se almacenó la información para cada año y sexo registrado. Para ilustrar el procedimiento, se tomará como ejemplo la base correspondiente al año 2012, filtrando exclusivamente a aquellos individuos con

sexo de nacimiento masculino, o equivalentemente, $r = 2012$ y $z = 1$, por lo que se omitirán estos sufijos temporalmente.

```
r<-2012
z <- 1
DF <- read.csv(paste("conjunto_de_datos
_defunciones_registradas_", r, ".csv", sep = ""))
```

A continuación, se llevó a cabo un proceso de filtración y limpieza de datos con el propósito de enfocar el procedimiento exclusivamente en las variables esenciales necesarias para la creación de la tabla. Durante este proceso, se eliminaron los registros que carecían de información suficiente y se aplicó el formato necesario para la manipulación de la base de datos.

```
DF<-subset(DF, select =
c("dia_ocurr", "mes_ocurr", "anio_ocur", "dia_nacim",
"mes_nacim", "anio_nacim", "sexo"))
DF<-subset(DF,
dia_ocurr!=99 & mes_ocurr!=99 & anio_ocur!=9999
& dia_nacim!=99 & mes_nacim!=99 & anio_nacim!=9999 & sexo!=9)
DF$fecha_nacim =
paste(DF$anio_nacim, DF$mes_nacim, DF$dia_nacim)
DF$fecha_ocur =
paste(DF$anio_ocur, DF$mes_ocurr, DF$dia_ocurr)
DF$fecha_nacim <- ymd(DF$fecha_nacim)
DF$fecha_ocur <- ymd(DF$fecha_ocur)
DF<-subset(DF,
!is.na(fecha_nacim) & !is.na(fecha_ocur))
DF<-subset(DF, select = c("fecha_nacim", "fecha_ocur", "sexo"))
DF$dias_vida <- DF$fecha_ocur-DF$fecha_nacim
DF$edad_años <-
DF$dias_vida %>% as.duration() %>% as.numeric("years")
DF<-subset(DF, edad_años>=0 & sexo==z)
```

El código utilizado dio como resultado la creación de la tabla mostrada en la

Figura 4.1: Base transformada 2012, género masculino.

fecha_nacim	fecha_ocur	sexo	edad_dias	edad_años
1929-01-06	2011-12-27	1	30305	82.9705681040383
1939-01-21	2011-09-29	1	26549	72.6872005475702
1925-04-25	2011-09-29	1	31568	86.4284736481862
1923-01-27	2011-12-13	1	32462	88.8761122518823
1932-01-10	2011-12-30	1	29209	79.9698836413416
...

figura [4.1], donde se muestra la fecha de nacimiento, fecha de fallecimiento, y la edad en años de cada individuo con género de nacimiento masculino y cuya defunción fue registrada en el año 2012.

La tabla resultante contenía un total de 167,948 registros después de haber sido sometida al proceso de limpieza y selección mencionados previamente. Estos registros formaron la base que se utilizó para construir la tabla de vida correspondiente al año 2012 para el género masculino. Para reorganizar estos registros en una tabla de vida, bastó crear un “Data Frame” con la estructura de años adecuada y una función que llenara dicho Data Frame con el conteo de individuos que alcanzaron la edad correspondiente. Este Data Frame se almacenó dentro de la variable “*Tabla*”, y fue generado mediante el siguiente código:

```
Tabla <-
data.frame(Año = 0:ceiling(max(DF$edad_años)-1))
Tabla$Poblacion = 0
Tabla$Defunciones = 0
```

Posteriormente, se definió la variable $n_{r,z}$, como el número total de registros para cada año y sexo, tal que $n_{2012,1} = 167,948$. A continuación, se introdujo la variable aleatoria $X_{i,r,z}$ que representa la edad de muerte en años para un individuo con año de defunción r , y género de nacimiento z , donde $i = 1, \dots, n_{r,z}$. Ambas variables permitieron definir $\mathcal{L}(0)_{r,z}$, que representa el número de población para cada año de registro, y $\mathcal{D}(k)_{r,z}$, que denota el conteo de los individuos que fallecen con edad alcanzada k , donde $k = 1, \dots, \max_i\{\lfloor X_{i,r,z} \rfloor\}$. Estos cálculos se realizaron

utilizando las ecuaciones (4.1) y (4.2), esta ultima asemejando a la expresión (2.11).

$$\mathcal{L}(0)_{r,z} = n_{r,z} \quad (4.1)$$

$$\mathcal{D}(k)_{r,z} = \sum_{i=1}^{n_{r,z}} \mathbb{1}_{(k \leq x_{i,r,z} < k+1)} \quad (4.2)$$

Usando de sustento la ecuación (2.13), se estableció además que $\mathcal{L}(k)_{r,z} := \mathcal{L}(k-1)_{r,z} - \mathcal{D}(k-1)_{r,z}$, lo que permitió obtener de forma iterativa a $\mathcal{L}(k)_{r,z}$.

La obtención de $\mathcal{D}(k)_{r,z}$, y $\mathcal{L}(k)_{r,z}$ respectivamente, se llevó acabo mediante las siguientes líneas de código:

```
f <- function(x) {
  return((DF$edad_años<(x+1) & DF$edad_años>=x) %>% sum())
}
Tabla$Defunciones <- sapJply(Tabla$Año,FUN = f)
Tabla$Poblacion[1] = nrow(DF)
for (i in 2:(max(Tabla$Año)+1)){
  Tabla$Poblacion[i] <-
  (Tabla$Poblacion[i-1]-Tabla$Defunciones[i-1]) }

```

De esta forma se construyó la tabla de vida [4.2].

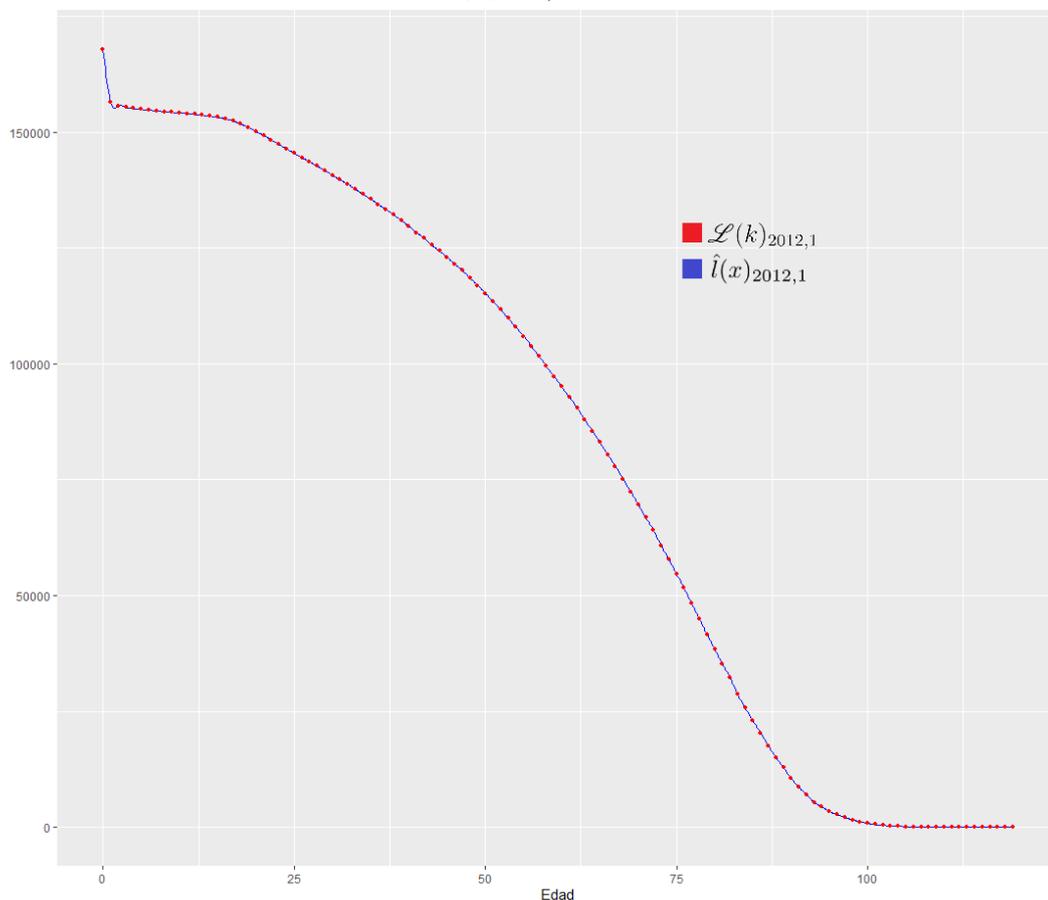
Figura 4.2: Tabla de vida construida 2012, género masculino.

Año	Poblacion	Defunciones
0	167948	11523
1	156425	708
2	155717	363
3	155354	239
4	155115	178
...
115	3	0
116	3	0
117	3	1
118	2	1
119	1	1

Para garantizar que se cumplieran las propiedades de una función de supervivencia, se optó por suavizar la función $\mathcal{L}(k)_{2012,1}$ mediante un spline natural cúbico, obteniendo así la función continua $\hat{l}(x)_{2012,1}$ al solucionar la expresión (4.3), y cuya visualización se puede ver en la figura [4.3].

$$\hat{l}(x)_{r,z} := \min_g \left\{ \sum_{k=1}^{\max_i \{ \lfloor X_{i,r,z} \rfloor \}} (\mathcal{L}(k)_{r,z} - g(k)_{r,z})^2 + \lambda \cdot \int_{-\infty}^{\infty} g''(t)_{r,z}^2 dt \right\} \quad (4.3)$$

Figura 4.3: Función $\hat{l}(x)_{2012,1}$ suavizada por Splines.



La obtención del spline fue posible gracias a la función “*smooth.spline()*” de R, la cual realiza un cálculo similar al descrito en la ecuación (2.24) para determinar la estructura de la función minimizadora. Para ello, se proporcionan las observaciones de la variable predictora y la variable objetivo en forma de vectores.

```
lx_spline <-
smooth.spline(Tabla$Año, Tabla$Poblacion, all.knots = TRUE)
```

Entre los valores que devuelve la función, se encuentra el valor calculado para λ , obtenido mediante la metodología LOOCV expuesta en la ecuación (2.26), el cual resultó igual a $\lambda_{LOOCV} = 1.87366 * 10^{-16}$, lo que equivale a la expresión (4.4).

$$\frac{1}{n} \sum_i (y_i - \hat{l}(x_i)_{2012,1,i,\lambda_{LOOCV}})^2 \leq \frac{1}{n} \sum_i (y_i - \hat{l}(x_i)_{2012,1,i,\lambda})^2, \forall \lambda \neq \lambda_{LOOCV} \quad (4.4)$$

Los valores de la función obtenida $\hat{l}(x)_{2012,1}$ se pueden extraer mediante la función “*predict()*”, que además es capaz de calcular las derivadas de la función. Al combinar esta herramienta con la ecuación (2.12), fue posible estimar la función de supervivencia de acuerdo a la expresión (4.5).

$$s(\hat{x})_{r,z} = \frac{\hat{l}(x)_{r,z}}{\hat{l}(0)_{r,z}} \quad (4.5)$$

Al ser \hat{l} una función doblemente diferenciable, y $\hat{l}(0)_{r,z}$ una constante respecto a la edad x , se puede expresar la derivada de (4.5) como se muestra en la expresión (4.6).

$$\frac{d}{dx} s(\hat{x})_{r,z} = \frac{\frac{d}{dx} \hat{l}(x)_{r,z}}{\hat{l}(0)_{r,z}} \quad (4.6)$$

Finalmente, ambas ecuaciones ofrecen una expresión equivalente para la fuerza de mortalidad estimada (similar a la igualdad (2.7)), en términos de la función suavizada $\hat{l}(x)_{r,z}$, mostrada en (4.7).

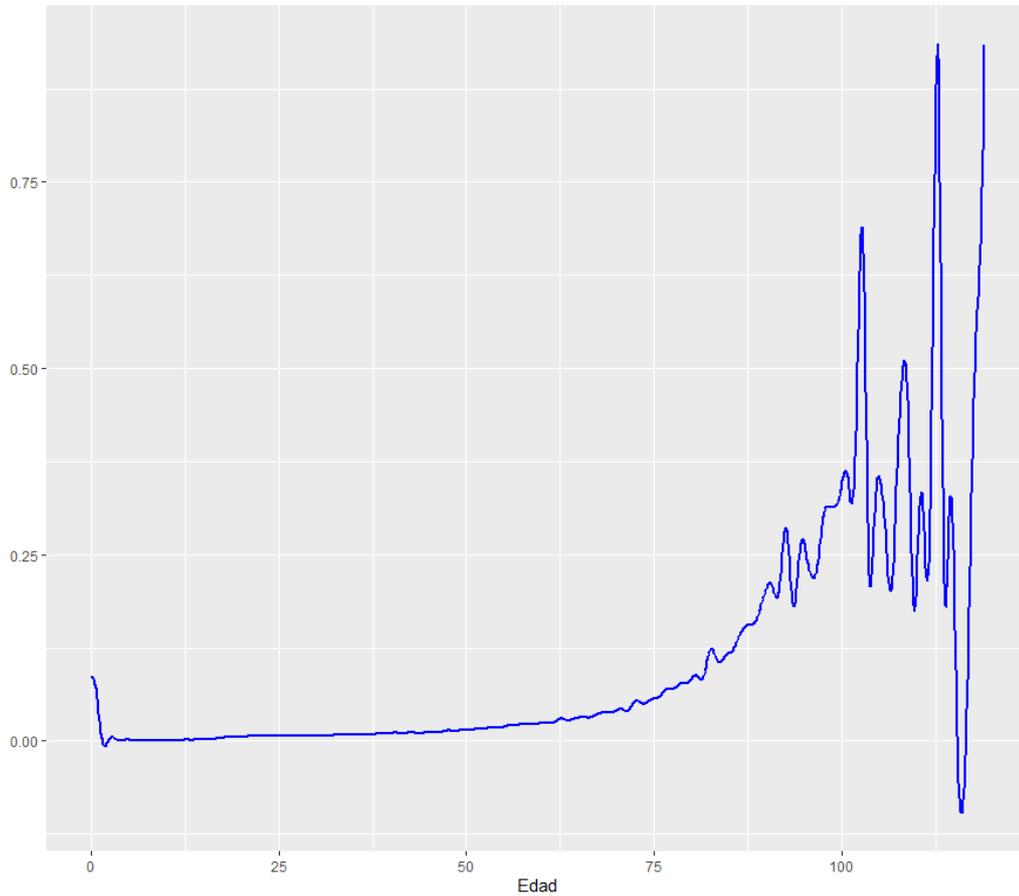
$$\mu(\hat{x})_{r,z} = \frac{-s'(\hat{x})_{r,z}}{s(\hat{x})_{r,z}} = \frac{-\frac{d}{dx} \hat{l}(x)_{r,z}}{\hat{l}(x)_{r,z}} \quad (4.7)$$

A continuación se muestra el código utilizado para calcular los valores de $\hat{l}(x)_{2012,1}$, $\hat{l}'(x)_{2012,1}$, $s(\hat{x})_{2012,1}$, $s'(\hat{x})_{2012,1}$ y $\mu(\hat{x})_{2012,1}$.

```
eje_x <- seq(0, max(Tabla$Año), by = 0.1)
lx_continua <- predict(lx_spline$fit, eje_x, deriv = 0)
lx_derivada <- predict(lx_spline$fit, eje_x, deriv = 1)
sx_continua <- lx_continua$y / Tabla$Poblacion[1]
sx_derivada <- lx_derivada$y / Tabla$Poblacion[1]
g <- function(x) { return(max(0, x)) }
sapply(sx_continua, FUN = g)
miu <- (-sx_derivada) / sx_continua
```

Por último, se muestra la función $\mu(\hat{x})_{2012,1}$ obtenida mediante splines por suavizado en la figura [4.4].

Figura 4.4: Fuerza de mortalidad estimada 2012, género masculino.



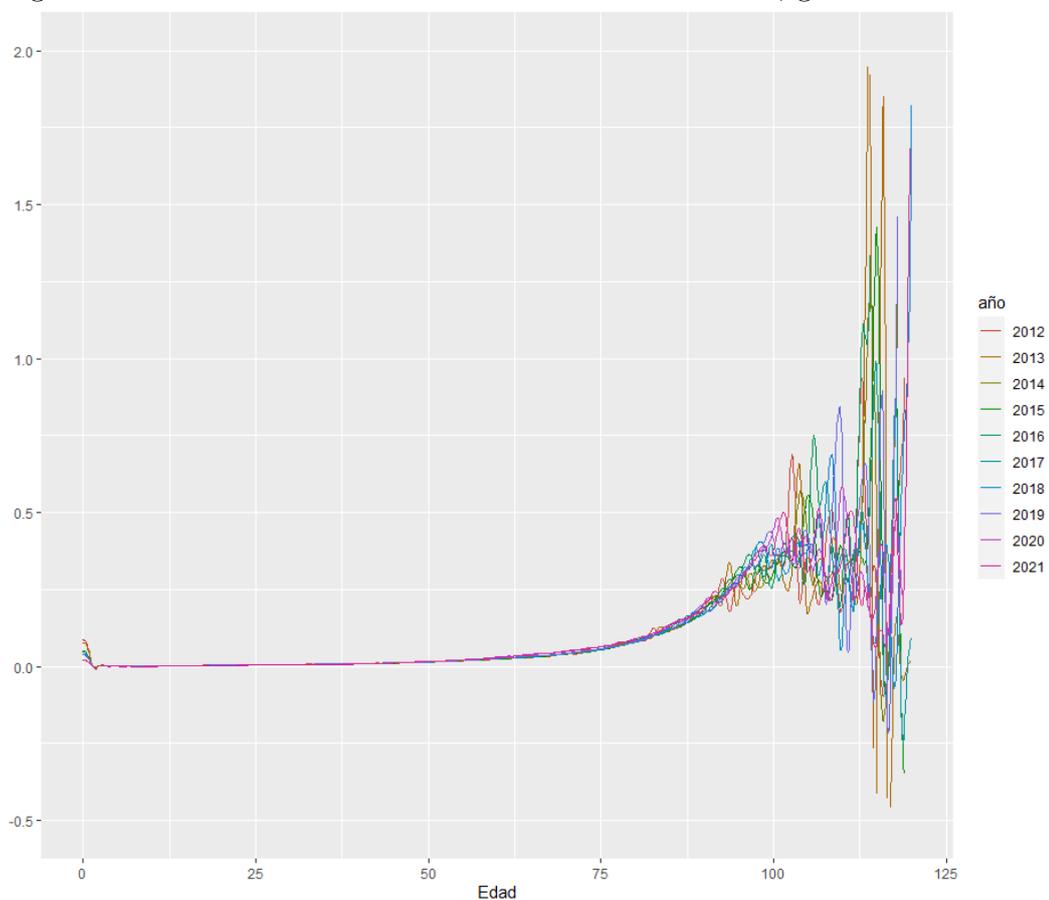
Este procedimiento es completamente análogo para todos los años de registro disponibles, y es posible de generalizar, ciclar y almacenar como se muestra a continuación:

```
for (r in 2012:2021){
  ...
  codigo <- paste("DF_", r, "_", z, "<-DF", sep = "")
  eval(parse(text=codigo))
  codigo <- paste("Tabla_", r, "_", z, "<-Tabla", sep = "")
  eval(parse(text=codigo))
  codigo <- paste("lxspline_", r, "_", z, "<-lx_spline", sep = "")
  eval(parse(text=codigo))
}
```

```
codigo <-  
paste("miu_", r, "_", z, "<-data.frame(eje_x,miu)", sep = "  
eval(parse(text=codigo))
```

Al calcular y obtener las funciones de fuerza de mortalidad estimada para cada año de registro, se hizo evidente la presencia de una considerable variabilidad entre las estimaciones. Estas diferencias se representan gráficamente en la figura [4.5].

Figura 4.5: Funciones de fuerza de mortalidad estimadas, género masculino.



Con el fin de disminuir esta variabilidad, y de incorporar los comportamientos identificados para cada año de manera simultánea, se realizó un suavizamiento final sobre todas las funciones de fuerza de mortalidad estimada, cuyo planteamiento

se expone en la ecuación (4.8).

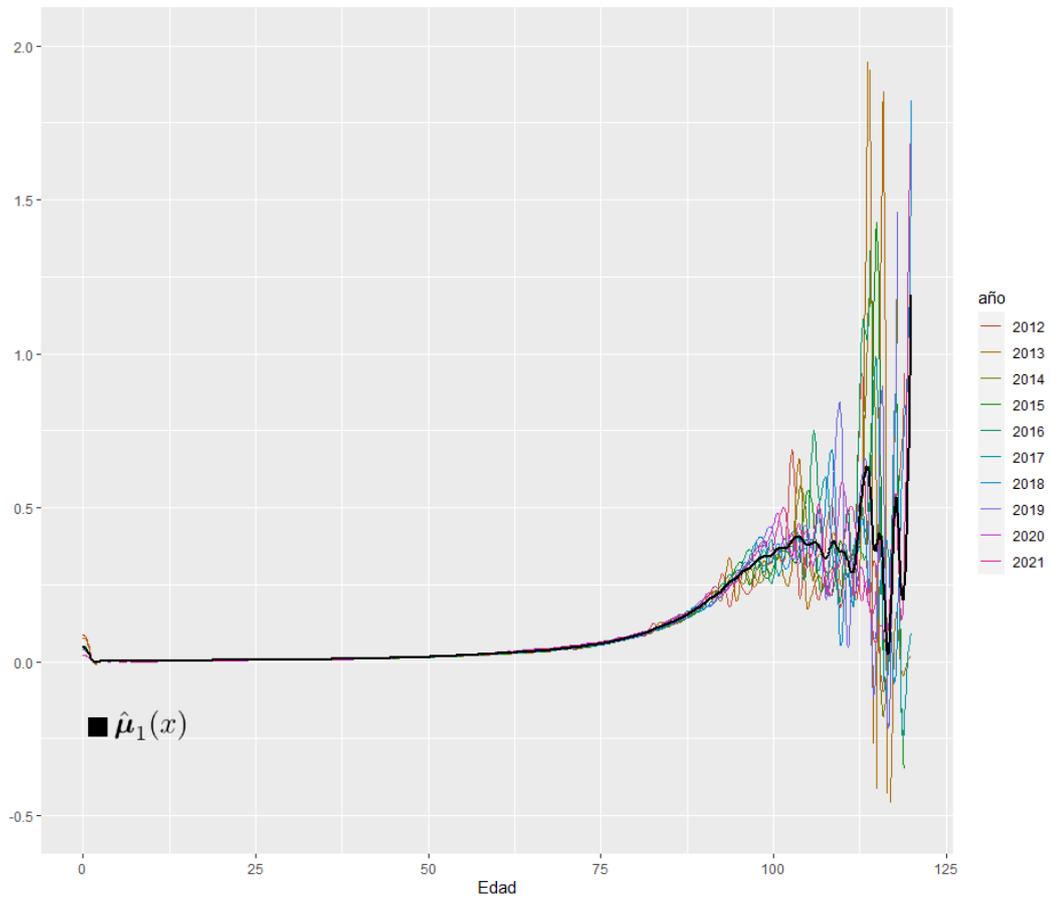
$$\hat{\mu}_z(x) := \underset{g}{\text{mín}} \left\{ \sum_{r=2012}^{2021} \sum_i (\hat{l}(i)_{r,z} - g_z(i))^2 + \lambda \cdot \int_{-\infty}^{\infty} g_z''(t)^2 dt \right\}. \quad (4.8)$$

Para estimar esta función suavizada, se creó un Data Frame que contuviera todas las realizaciones de las funciones de fuerza de mortalidad previamente estimadas i , para todos los años analizados. Estas observaciones fueron las que posteriormente se introdujeron a la función “*smooth.spline()*”, y que, mediante la función “*predict()*”, generaron una representación suavizada en una malla de valores para su visualización. Estas operaciones se llevaron a cabo mediante el siguiente código:

```
miu_total_1 <- data.frame()
for (r in 2012:2021){
  codigo <-
  paste("miu_", r, "_", z, "$año<-", r, sep = "")
  eval(parse(text=codigo))
  codigo <-
  paste("miu_total_", z, "<-rbind(miu_total_",
  z, ",miu_", r, "_", z, ") ", sep = "")
  eval(parse(text=codigo))
ajustemiu_1<- smooth.spline(miu_total_1$eje_x,
miu_total_1$miu, all.knots = TRUE)
spline_miu_1 <- predict(ajustemiu_1$fit,
seq(0, max(miu_total_1$eje_x), by = 0.1), deriv = 0)
```

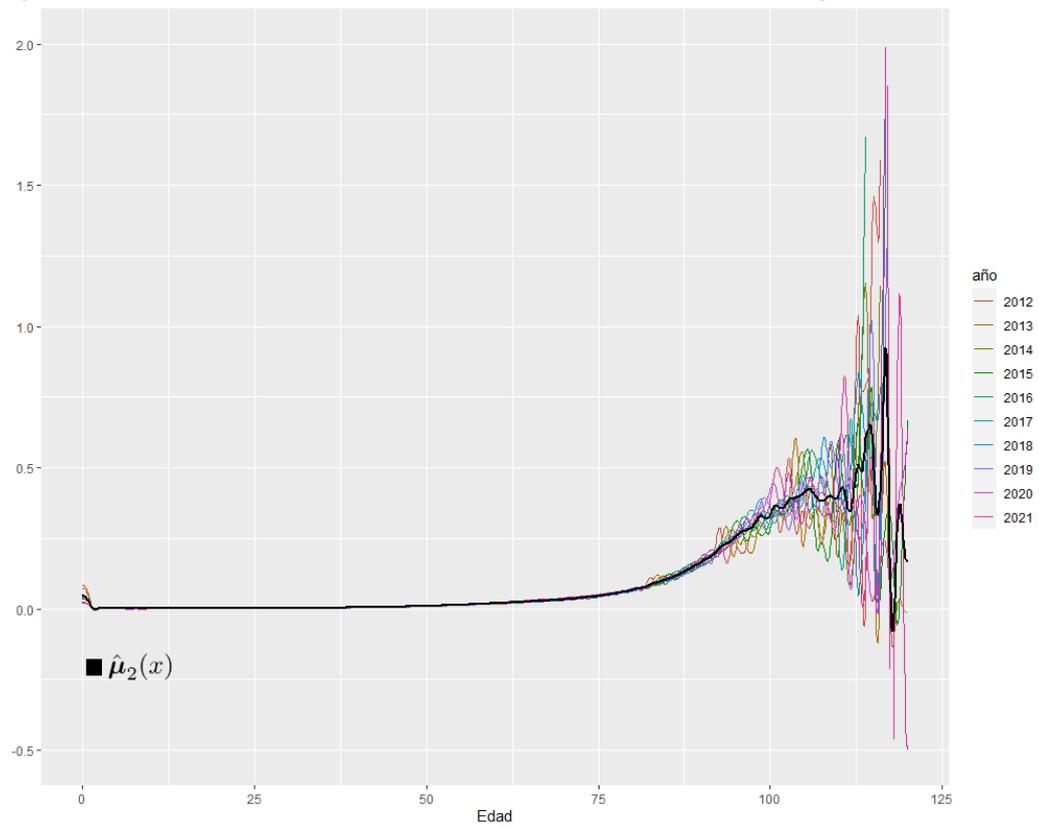
El procesamiento de la información culminó con la obtención de la función $\hat{\mu}_1(x)$, que representa la fuerza de mortalidad estimada para el sexo masculino. Esta función logró incorporar de manera exitosa los comportamientos observados en todos los años analizados, además de presentar una baja variabilidad, lo cual se visualiza en la figura [4.6]. A partir de esta función, como se detalló en el capítulo 2, es posible extraer y calcular todas las variables necesarias para construir una tabla de vida que englobe el comportamiento de mortalidad encontrado.

Figura 4.6: Función final de fuerza de mortalidad estimada, género masculino.



Construir la función final de fuerza de mortalidad estimada para el sexo femenino $\hat{\mu}_2(x)$, involucró un proceso completamente semejante, sustituyendo el valor de la variable z por $z = 2$ en la programación previamente mencionada. En la figura [4.7] se presentan los resultados de las estimaciones de la fuerza de mortalidad para cada año, así como la estimación final correspondiente para los individuos con género de nacimiento femenino.

Figura 4.7: Función final de fuerza de mortalidad estimada, género femenino.



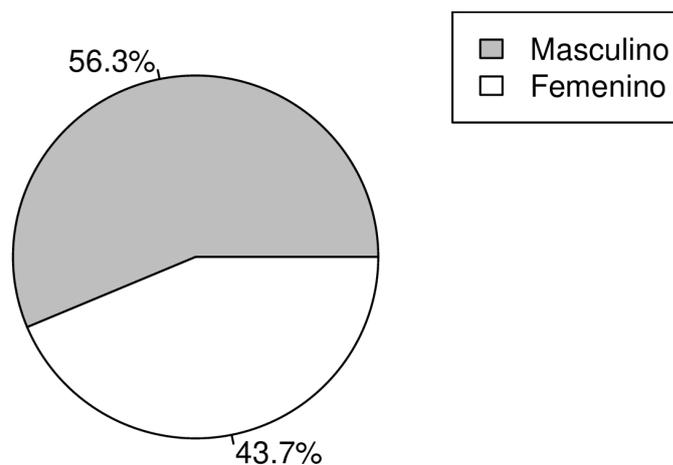
Capítulo 5

Resultados y discusión

5.1. Características de la muestra

La muestra presentó una composición de 2,998,574 individuos con sexo de nacimiento femenino, y 3,866,563 individuos con sexo de nacimiento masculino, como se puede ver en la figura [5.1]. Dado que la muestra fue discriminada en dos subgrupos principales según el sexo, se presentarán las características de cada subgrupo por separado.

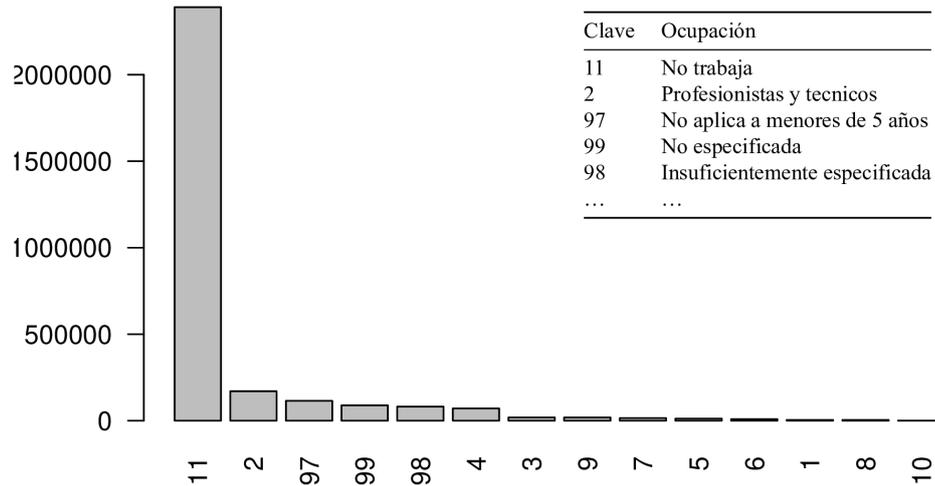
Figura 5.1: Distribución de géneros de la muestra.



En el segmento de la muestra correspondiente al género femenino, se observó que

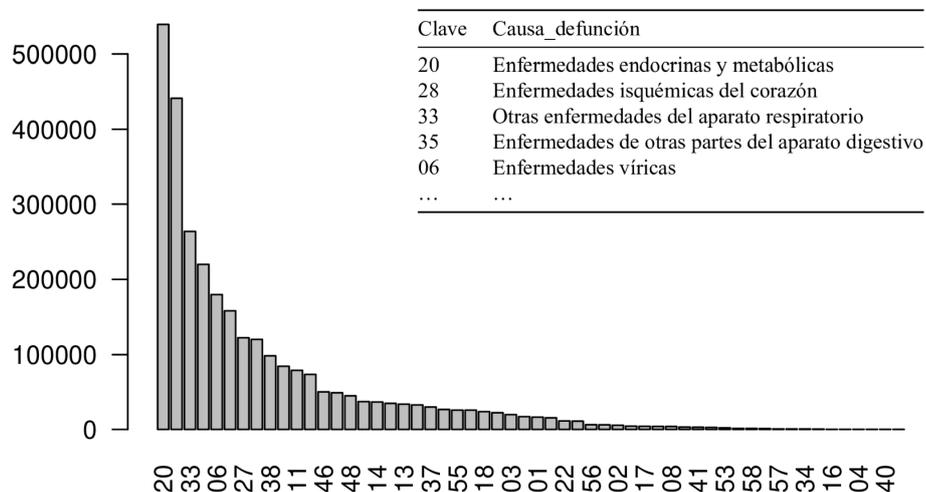
la composición del comportamiento laboral estuvo mayormente conformada por mujeres sin ocupación laboral, seguido de aquellas con perfiles profesionales y técnicos, y finalmente, se encontraron individuos cuya ocupación no fue especificada, representado en la figura [5.2].

Figura 5.2: Distribución de ocupación laboral, género femenino.



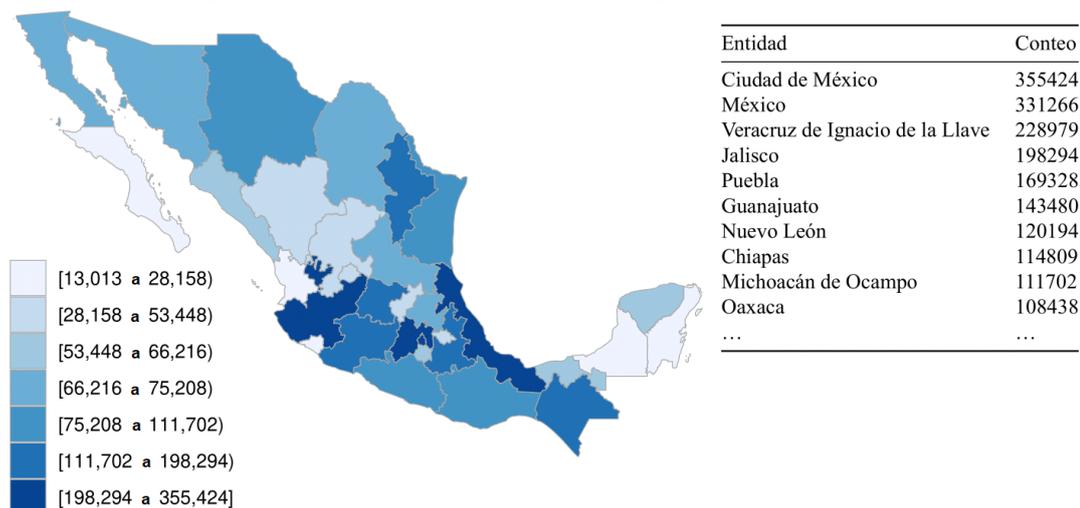
Respecto a la causa de fallecimiento dentro del mismo género, se identificaron como principales razones las enfermedades endocrinas y metabólicas, seguidas de las enfermedades isquémicas del corazón y las enfermedades del aparato respiratorio. La jerarquía completa de estas causas puede visualizarse en el histograma presentado en la figura [5.3].

Figura 5.3: Distribución de causas de fallecimiento, género femenino.



Finalmente, la mayoría de los fallecimientos registrados en el sexo femenino tuvieron lugar en la Ciudad de México, el Estado de México, Veracruz y Jalisco. Esta dinámica demográfica se representa visualmente en la ilustración [5.4].

Figura 5.4: Entidad de ocurrencia, género femenino.



En relación al subgrupo de individuos de sexo de nacimiento masculino, se observaron características similares al subgrupo femenino. Predominantemente, se encontraron individuos sin ocupación laboral, seguidos de trabajadores agrícolas y artesanales. Así mismo, las principales causas de muerte registradas fueron las enfermedades isquémicas del corazón, situando en este caso a las enfermedades endocrinas y metabólicas como la segunda causa más común, seguido de enfermedades del aparato digestivo (excluyendo enfermedades infecciosas intestinales). Similar al subgrupo anterior, la mayoría de los fallecimientos del sexo masculino fueron registrados en el Estado de México, la Ciudad de México, Veracruz y Jalisco. Las visualizaciones detalladas de esta información se presentan en las figuras [5.5],[5.6] y [5.7] respectivamente.

Figura 5.5: Distribución de ocupación laboral, género masculino.

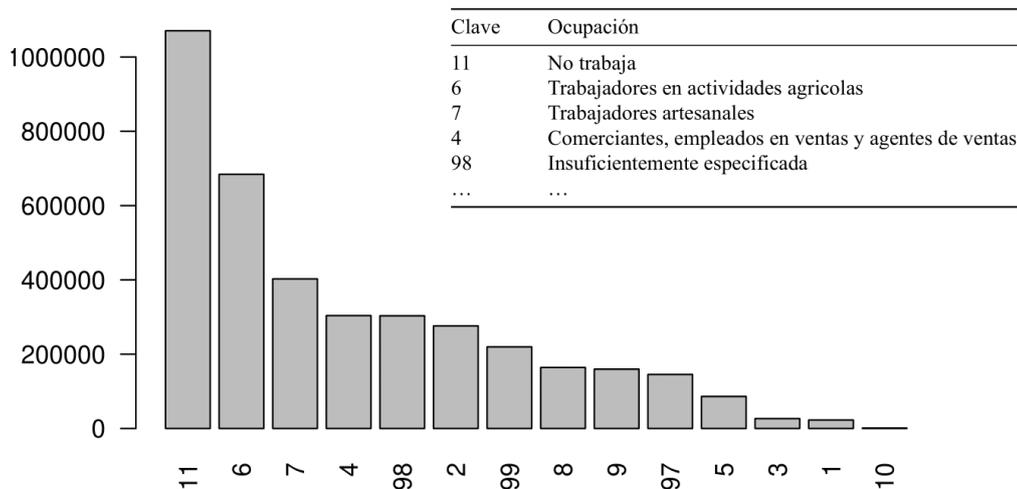


Figura 5.6: Distribución de causas de fallecimiento, género masculino.

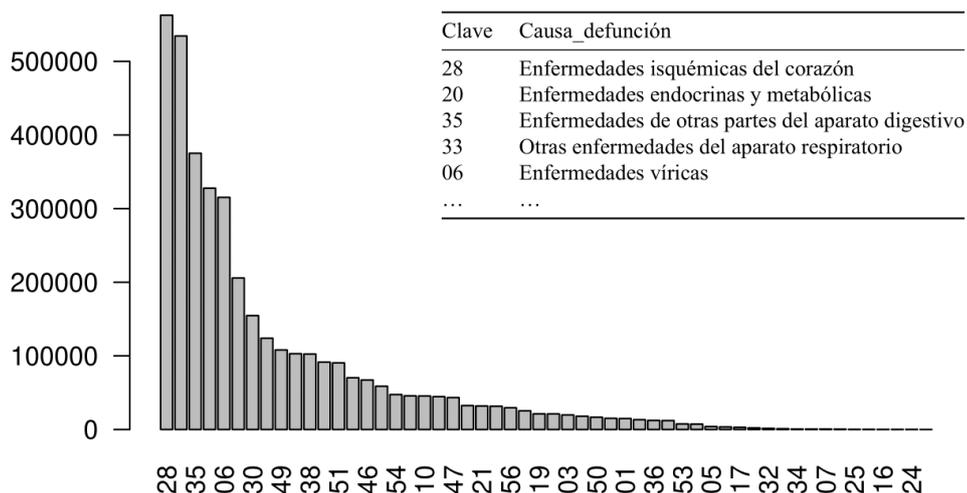
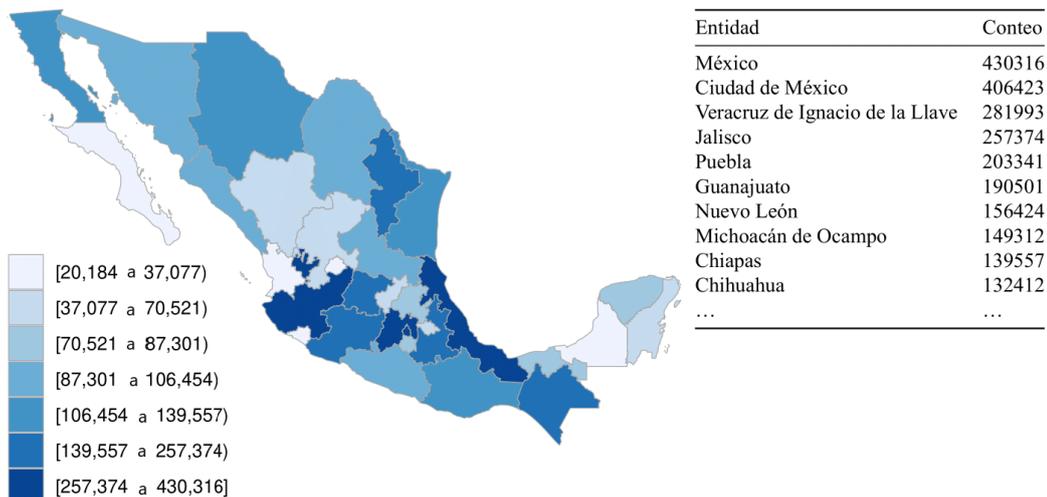


Figura 5.7: Entidad de ocurrencia, género femenino.



5.2. Resultados de las variables

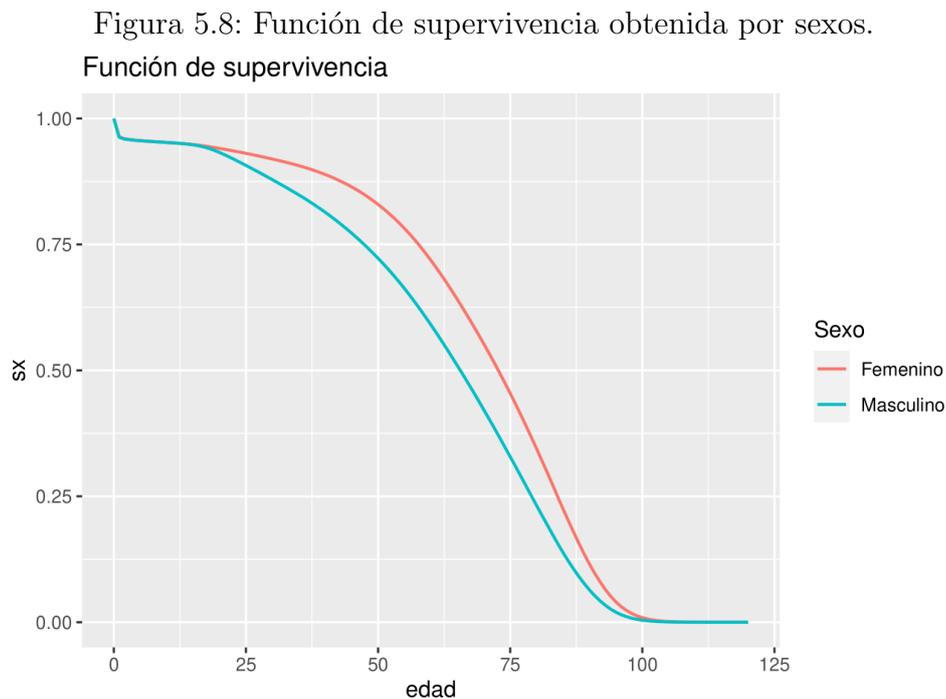
Obtenidas las funciones de fuerza de mortalidad, μ_1 para el sexo masculino, y μ_2 para el femenino, fue posible derivar todas las funciones asociadas al uso de probabilidades de muerte. Particularmente, establecer $x = 0$ en la ecuación (2.9), permitió obtener la función s_z , para $z = 1, 2$ a partir de μ_z , como se ve en (5.1).

$$\exp\left(-\int_0^n \mu(y)dy\right) = \frac{s(n)}{s(0)} = \frac{s(n)}{1} = s(n), n > 0 \quad (5.1)$$

Aunque es posible derivar una expresión exacta para la integral $\int_0^n \mu(y)dy$, esto requeriría integrar todos los nodos que integran la función suavizada, lo cual resulta sumamente impráctico. En su lugar, la integral se puede aproximar mediante algún método numérico. Se decidió utilizar la función "integrate" de R, implementada mediante el siguiente fragmento de código.

```
integral_1 <- function(x) {predict(ajustemiu_1$fit, x, deriv=0) $y}
sxspline_1 <- function(x) {exp(-integrate(integral_1, 0, x) [[1]]) }
```

Ambas funciones se pueden ver en la gráfica [5.8].



Es así que, en virtud de las expresiones (2.5) y (2.12), se incorporaron de forma intrínseca la totalidad de los datos disponibles en los registros, al obtener $\hat{l}(x)_z = \hat{s}(x)_z$ a partir de $\mu(x)_z$. Por último, al definir $l_0 := 100,000$, se logró construir las tablas de vida mostradas en las tablas [5.9].

Figura 5.9: Tabla de vida construida por sexos.

Edad	lx_masculino	qx_masculino	lx_femenino	qx_femenino	Edad	lx_masculino	qx_masculino	lx_femenino	qx_femenino
0	100,000.00	0.0369221	100,000.00	0.0365917	52	69,979.29	0.0171651	81,200.48	0.0116797
1	96,307.79	0.0041892	96,340.83	0.0037477	53	68,778.09	0.0181549	80,252.08	0.0126291
2	95,904.34	0.0011124	95,979.77	0.0016054	54	67,529.42	0.0191893	79,238.57	0.0136426
3	95,797.66	0.0013370	95,825.68	0.0012818	55	66,233.58	0.0204448	78,157.55	0.0145905
4	95,669.58	0.0008940	95,702.86	0.0009646	56	64,879.45	0.0215715	77,017.20	0.0155112
5	95,584.05	0.0008195	95,610.55	0.0008345	57	63,479.90	0.0227467	75,822.57	0.0166609
6	95,505.72	0.0007434	95,530.76	0.0007294	58	62,035.94	0.0239268	74,559.30	0.0179429
7	95,434.73	0.0007058	95,461.08	0.0007303	59	60,551.62	0.0250289	73,221.49	0.0190115
8	95,367.36	0.0006871	95,391.37	0.0006623	60	59,036.07	0.0264452	71,829.44	0.0201270
9	95,301.84	0.0006973	95,328.19	0.0007109	61	57,474.85	0.0275716	70,383.73	0.0212363
10	95,235.38	0.0007393	95,260.42	0.0007269	62	55,890.18	0.0292141	68,889.04	0.0225864
11	95,164.98	0.0007743	95,191.17	0.0007350	63	54,257.40	0.0302514	67,333.09	0.0238802
12	95,091.29	0.0009015	95,121.21	0.0008834	64	52,616.04	0.0320283	65,725.16	0.0251233
13	95,005.56	0.0010481	95,037.18	0.0009832	65	50,930.84	0.0332071	64,073.93	0.0263373
14	94,905.98	0.0014205	94,943.74	0.0011023	66	49,239.57	0.0350492	62,386.40	0.0276423
15	94,771.17	0.0018854	94,839.08	0.0013679	67	47,513.76	0.0366005	60,661.89	0.0288565
16	94,592.49	0.0025115	94,709.34	0.0015422	68	45,774.73	0.0390402	58,911.40	0.0307837
17	94,354.93	0.0032742	94,563.28	0.0016632	69	43,987.68	0.0411069	57,097.89	0.0319295
18	94,045.99	0.0039775	94,406.01	0.0017779	70	42,179.48	0.0433218	55,274.78	0.0340647
19	93,671.93	0.0045653	94,238.17	0.0018719	71	40,352.19	0.0453834	53,391.86	0.0357733
20	93,244.29	0.0050643	94,061.77	0.0019842	72	38,520.87	0.0487046	51,481.86	0.0384665
21	92,772.07	0.0053656	93,875.13	0.0020246	73	36,644.73	0.0517126	49,501.53	0.0407687
22	92,274.28	0.0056489	93,685.07	0.0020752	74	34,749.73	0.0546877	47,483.42	0.0430342
23	91,753.03	0.0058099	93,490.65	0.0021532	75	32,849.35	0.0582211	45,440.01	0.0463478
24	91,219.96	0.0059359	93,289.35	0.0022460	76	30,936.82	0.0625121	43,333.97	0.0497419
25	90,678.48	0.0060943	93,079.82	0.0023201	77	29,002.90	0.0675412	41,178.45	0.0535261
26	90,125.86	0.0061528	92,863.87	0.0023430	78	27,044.01	0.0723180	38,974.33	0.0582527
27	89,571.34	0.0063921	92,646.29	0.0025013	79	25,088.24	0.0771699	36,703.97	0.0622437
28	88,998.79	0.0064473	92,414.55	0.0024539	80	23,152.18	0.0832827	34,419.38	0.0683659
29	88,424.99	0.0065476	92,187.77	0.0025692	81	21,224.00	0.0890135	32,066.27	0.0726628
30	87,846.02	0.0066609	91,950.93	0.0026446	82	19,334.78	0.0981228	29,736.24	0.0814168
31	87,260.88	0.0067352	91,707.75	0.0026972	83	17,437.60	0.1055157	27,315.21	0.0896849
32	86,673.17	0.0070018	91,460.40	0.0028768	84	15,597.66	0.1140285	24,865.45	0.0963752
33	86,066.30	0.0071114	91,197.29	0.0030503	85	13,819.08	0.1220585	22,469.04	0.1037292
34	85,454.25	0.0072646	90,919.11	0.0031161	86	12,132.34	0.1310469	20,138.34	0.1127017
35	84,833.45	0.0076423	90,635.80	0.0033605	87	10,542.44	0.1410209	17,868.72	0.1243636
36	84,185.13	0.0079028	90,331.22	0.0036439	88	9,055.73	0.1530408	15,646.50	0.1355771
37	83,519.83	0.0082713	90,002.06	0.0038359	89	7,669.84	0.1653049	13,525.19	0.1476404
38	82,829.01	0.0085611	89,656.83	0.0041865	90	6,401.98	0.1802720	11,528.33	0.1601642
39	82,119.90	0.0090544	89,281.48	0.0044889	91	5,247.88	0.1922704	9,681.90	0.1712760
40	81,376.36	0.0094831	88,880.70	0.0047577	92	4,238.87	0.2068379	8,023.62	0.1916315
41	80,604.66	0.0098241	88,457.83	0.0051768	93	3,362.11	0.2249165	6,486.04	0.2047162
42	79,812.78	0.0104818	87,999.90	0.0055620	94	2,605.92	0.2368292	5,158.25	0.2182004
43	78,976.21	0.0107961	87,510.44	0.0059660	95	1,988.76	0.2517885	4,032.71	0.2338731
44	78,123.57	0.0114153	86,988.35	0.0065190	96	1,488.01	0.2627052	3,089.57	0.2436604
45	77,231.77	0.0120134	86,421.28	0.0069738	97	1,097.10	0.2771772	2,336.77	0.2557088
46	76,303.95	0.0126178	85,818.59	0.0075300	98	793.01	0.2874340	1,739.23	0.2723323
47	75,341.16	0.0132653	85,172.38	0.0082634	99	565.07	0.2904527	1,257.06	0.2760963
48	74,341.74	0.0139652	84,468.56	0.0088682	100	400.95	0.3048427	909.99	0.2997517
49	73,303.54	0.0145694	83,719.48	0.0094174	101	278.72	0.3076645	637.22	0.3013631
50	72,235.55	0.0153249	82,931.05	0.0100656	102	192.97	0.3197835	445.19	0.3147910
51	71,128.55	0.0161575	82,096.30	0.0109119	103	131.26	0.3326217	305.05	0.3249809

Edad	lx_masculino	qx_masculino	lx_femenino	qx_femenino
104	87.60	0.3213328	205.91	0.3317357
105	59.45	0.3181446	137.60	0.3433943
106	40.54	0.3135346	90.35	0.3315319
107	27.83	0.2898633	60.40	0.3181683
108	19.76	0.3155009	41.18	0.3271224
109	13.53	0.3043105	27.71	0.3254155
110	9.41	0.2885405	18.69	0.3427086
111	6.70	0.2616140	12.29	0.3041001
112	4.94	0.3727273	8.55	0.3752962
113	3.10	0.4579323	5.34	0.4108194
114	1.68	0.3420489	3.15	0.4664055
115	1.11	0.3144588	1.68	0.3234208
116	0.76	0.0910488	1.14	0.5050578
117	0.69	0.3387663	0.56	0.1757120
118	0.46	0.2499203	0.46	0.2108398
119	0.34	0.0369221	0.37	0.0365917
120	0.18	0.0369221	0.29	0.0365917

5.3. Comprobación de hipótesis

Para examinar y comparar los patrones de mortalidad presentados en la tabla de vida previamente obtenida, así como los reflejados en las tasas de mortalidad de los beneficiarios de la seguridad social mexicana según EMSSAH-09 y EMSSAM-09, fue suficiente analizar una de las funciones relacionadas con las probabilidades de muerte. Si se determinara que existiesen diferencias en una función específica, estas divergencias también se manifestarían en todas las demás funciones, ya que todas están estrechamente relacionadas.

Se optó por utilizar la función de supervivencia para la comparación, pues al ser el complemento de una función de distribución probabilística, se pudo evaluar si existía una diferencia entre ambas muestras mediante una prueba de bondad de ajuste.

Pese a que no se contaba con la expresión exacta de la función de supervivencia utilizada por la CNSF, fue posible derivar los valores discretos de la función evaluada en cada año de edad, a través de los valores de q_x incluidos en ambas tablas EMSSAH-09 y EMSSAM-09. Para lograr esto, primero se utilizaron las ecuaciones (2.4) y (2.5) para se obtener la expresión (5.2).

$$q_x = 1 - p_x = 1 - \frac{s(x+1)}{s(x)} \quad (5.2)$$

Además, al ser $s(0) = 1$, entonces $q_0 = 1 - s(1)$, o equivalentemente, $s(1) = 1 - q_0$. Al combinar estos resultados, se calcularon los demás valores para $s(x)$ de manera recursiva mediante la fórmula (5.3), y cuyos resultados se pueden consultar en la tabla [5.10].

$$s(x+1) = (1 - q_x) * s(x) \quad (5.3)$$

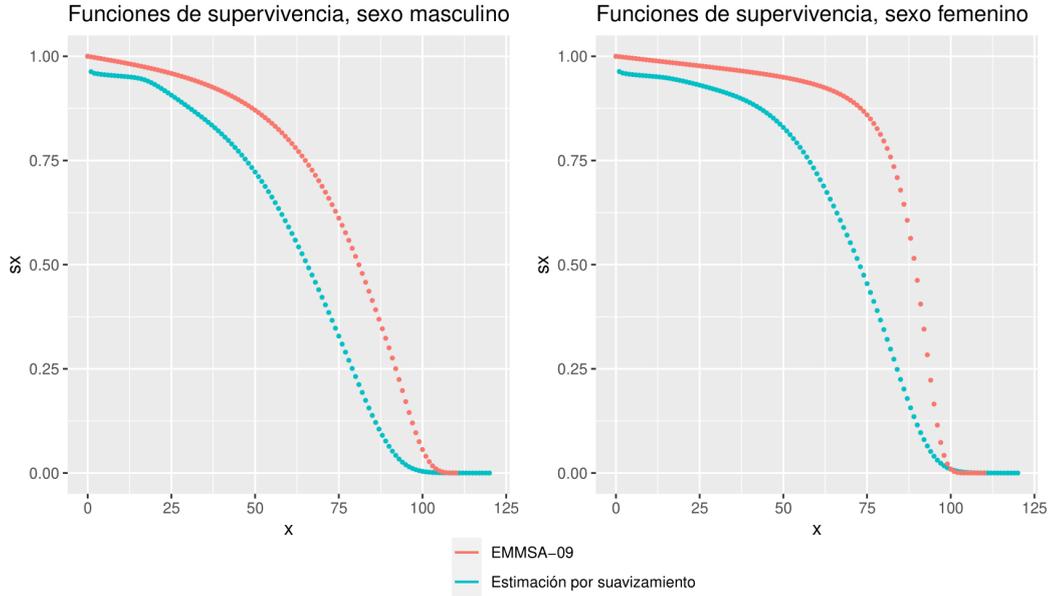
Figura 5.10: Tablas de vida EMMSA-09.

Edad	Hombres qx	Mujeres qx	Hombres sx	Mujeres sx	Edad	Hombres qx	Mujeres qx	Hombres sx	Mujeres sx
0	0.00138	0.00092	1.0000000	1.0000000	52	0.00733	0.00172	0.8589317	0.9466557
1	0.00138	0.00092	0.9986200	0.9990800	53	0.00772	0.00180	0.8526358	0.9450274
2	0.00139	0.00092	0.9972419	0.9981608	54	0.00814	0.00189	0.8460534	0.9433264
3	0.00140	0.00092	0.9958557	0.9972425	55	0.00859	0.00199	0.8391665	0.9415435
4	0.00142	0.00092	0.9944615	0.9963251	56	0.00906	0.00211	0.8319581	0.9396698
5	0.00143	0.00092	0.9930494	0.9954085	57	0.00957	0.00223	0.8244205	0.9376871
6	0.00145	0.00092	0.9916293	0.9944927	58	0.01011	0.00237	0.8165308	0.9355961
7	0.00147	0.00092	0.9901915	0.9935777	59	0.01069	0.00253	0.8082757	0.9333787
8	0.00149	0.00092	0.9887359	0.9926637	60	0.01131	0.00270	0.7996352	0.9310173
9	0.00152	0.00092	0.9872627	0.9917504	61	0.01198	0.00290	0.7905914	0.9285035
10	0.00155	0.00092	0.9857620	0.9908380	62	0.01268	0.00312	0.7811201	0.9258109
11	0.00158	0.00092	0.9842341	0.9899264	63	0.01344	0.00336	0.7712155	0.9229223
12	0.00161	0.00092	0.9826790	0.9890157	64	0.01425	0.00364	0.7608504	0.9198213
13	0.00165	0.00092	0.9810969	0.9881058	65	0.01512	0.00396	0.7500082	0.9164732
14	0.00168	0.00092	0.9794781	0.9871967	66	0.01604	0.00432	0.7386681	0.9128439
15	0.00172	0.00092	0.9778326	0.9862885	67	0.01703	0.00473	0.7268199	0.9089004
16	0.00177	0.00092	0.9761507	0.9853811	68	0.01810	0.00520	0.7144421	0.9046013
17	0.00181	0.00092	0.9744229	0.9844746	69	0.01924	0.00574	0.7015107	0.8989774
18	0.00186	0.00093	0.9726592	0.9835689	70	0.02045	0.00636	0.6880137	0.8947320
19	0.00191	0.00093	0.9708501	0.9826541	71	0.02176	0.00707	0.6739438	0.8890415
20	0.00197	0.00093	0.9689957	0.9817403	72	0.02316	0.00790	0.6592788	0.8827560
21	0.00202	0.00093	0.9670868	0.9808273	73	0.02467	0.00886	0.6440099	0.8757822
22	0.00209	0.00094	0.9651333	0.9799151	74	0.02628	0.00998	0.6281221	0.8680228
23	0.00215	0.00094	0.9631162	0.9789940	75	0.02801	0.01130	0.6116151	0.8593599
24	0.00222	0.00095	0.9610455	0.9780737	76	0.02986	0.01285	0.5944838	0.8496492
25	0.00230	0.00095	0.9589120	0.9771445	77	0.03185	0.01467	0.5767325	0.8387312
26	0.00237	0.00096	0.9567065	0.9762163	78	0.03399	0.01683	0.5583635	0.8264270
27	0.00246	0.00096	0.9544391	0.9752791	79	0.03629	0.01940	0.5393848	0.8125182
28	0.00254	0.00097	0.9520911	0.9743428	80	0.03875	0.02247	0.5198105	0.7967554
29	0.00264	0.00098	0.9496728	0.9733977	81	0.04139	0.02614	0.4996678	0.7788523
30	0.00274	0.00099	0.9471657	0.9724438	82	0.04423	0.03056	0.4789866	0.7584931
31	0.00284	0.00100	0.9445705	0.9714811	83	0.04728	0.03588	0.4578010	0.7353135
32	0.00295	0.00101	0.9418879	0.9705096	84	0.05055	0.04233	0.4361562	0.7089305
33	0.00307	0.00102	0.9391093	0.9695294	85	0.05406	0.05014	0.4141085	0.6789214
34	0.00319	0.00104	0.9362262	0.9685404	86	0.05783	0.05964	0.3917218	0.6448803
35	0.00332	0.00105	0.9332397	0.9675332	87	0.06187	0.07121	0.3690685	0.6064197
36	0.00346	0.00107	0.9301413	0.9665173	88	0.06621	0.08530	0.3462342	0.5632365
37	0.00361	0.00109	0.9269230	0.9654831	89	0.07087	0.10245	0.3233101	0.5151924
38	0.00377	0.00111	0.9235768	0.9644307	90	0.08147	0.12327	0.3003971	0.4624110
39	0.00393	0.00113	0.9200950	0.9633602	91	0.09207	0.14846	0.2759237	0.4054096
40	0.00411	0.00116	0.9164790	0.9622716	92	0.10439	0.17874	0.2505194	0.3452225
41	0.00430	0.00118	0.9127123	0.9611554	93	0.11835	0.21478	0.2243677	0.2835174
42	0.00450	0.00121	0.9087876	0.9600212	94	0.13418	0.25716	0.1978138	0.2226235
43	0.00471	0.00124	0.9046981	0.9588596	95	0.15212	0.30616	0.1712711	0.1653737
44	0.00493	0.00128	0.9004369	0.9576706	96	0.17247	0.36163	0.1452174	0.1147429
45	0.00517	0.00132	0.8959978	0.9564448	97	0.19554	0.42286	0.1201717	0.0732484
46	0.00542	0.00136	0.8913655	0.9551823	98	0.22170	0.48842	0.0966734	0.0422746
47	0.00569	0.00141	0.8865343	0.9538832	99	0.25135	0.55626	0.0752409	0.0216268
48	0.00598	0.00146	0.8814899	0.9525382	100	0.28497	0.62390	0.0563291	0.0095967
49	0.00629	0.00151	0.8762186	0.9511475	101	0.32309	0.68873	0.0402770	0.0036093
50	0.00661	0.00158	0.8707072	0.9497113	102	0.36630	0.74844	0.0272639	0.0011235
51	0.00696	0.00164	0.8649518	0.9482107	103	0.41530	0.80133	0.0172771	0.0002826

Edad	Hombres qx	Mujeres qx	Hombres sx	Mujeres sx
104	0.47085	0.84648	0.0101019	0.0000561
105	0.53383	0.88375	0.0053454	0.0000086
106	0.60523	0.91358	0.0024919	0.0000010
107	0.68618	0.93683	0.0009837	0.0000001
108	0.77796	0.95453	0.0003087	0.0000000
109	0.88202	0.96773	0.0000685	0.0000000
110	1.00000	1.00000	0.0000081	0.0000000

Desde este punto se volvió evidente la divergencia entre ambas estimaciones, al compararlas visualmente para cada subgrupo a través de la gráfica [5.11].

Figura 5.11: Comparativa de funciones de supervivencia.



No obstante, se decidió emplear la prueba de bondad de ajuste Kolmogórov-Smirnov, en la cual se contrasta la hipótesis nula de que una muestra aleatoria dada sigue alguna distribución de probabilidad teórica propuesta (Rohatgi y Saleh, 2015). Es decir, si se denota $F_{z,CNSF}$ como la función de distribución exhibida por las tablas de vida EMSSAH-09 y EMSSAM-09 para $z = 1, 2$ respectivamente, y X_z como la variable aleatoria que representa la edad de muerte de un individuo registrado en las estadísticas de defunciones generales del INEGI, entonces la hipótesis nula supone que la distribución empírica de X_z , $\hat{F}_{X,z}(x)$ (como se definió en la ecuación (2.34)) es igual a $F_{z,CNSF}$. Esto es equivalente a plantear las hipótesis expuestas en (5.4).

$$H_0 : X_z \sim F_{z,CNSF} \quad vs \quad H_1 : X_z \not\sim F_{z,CNSF} \quad (5.4)$$

Se calculó la distribución empírica de X_z mediante el método de la transformada inversa (Bertrand et al., 2009). En este enfoque, se utiliza una variable aleatoria distribuida uniformemente en el intervalo $(0, 1)$, $U \sim Unif(0, 1)$, para simular valores aleatorios equiprobables entre 0 y 1. Estos valores se introducen posteriormente en alguna función de distribución inversa, en este caso $\hat{F}_z^{-1}(x)$, para obtener realizaciones aleatorias de X_z , tal como se describe en la ecuación (5.5).

El tamaño de la muestra aleatoria n , dependerá de la cantidad valores extraídos de U .

$$X_{z,i} = \hat{F}_z^{-1}(U_i) \quad \text{para } i = 1, \dots, n \quad (5.5)$$

Es importante remarcar que, mientras $\hat{F}_z(x)$ representa la distribución teórica de la variable aleatoria X_z , obtenida mediante suavizamiento por splines, $\hat{F}_{X,z}(x)$ corresponde a la distribución empírica de X_z . Esta última se calcula a partir del muestreo aleatorio descrito en la ecuación (5.5) y la fórmula (2.34).

Se logró reunir todas las variables necesarias para realizar la prueba de hipótesis al combinar las funciones `runif()` y `ecdf()` de R. Esto permitió simular $n = 1,000,000$ realizaciones de U y construir la función de distribución empírica respectiva, como se muestra en las siguientes líneas de código.

```
u <- runif(1000000)
ecdf_m <-
ecdf(survival_1$edad[findInterval(u, 1-survival_1$sx)])
ecdf_f <-
ecdf(survival_2$edad[findInterval(u, 1-survival_2$sx)])
ecdf_m <-
data.frame(Edad = survival_1$edad,
f_empirica = ecdf_m(survival_1$edad))
ecdf_f <-
data.frame(Edad = survival_2$edad,
f_empirica = ecdf_f(survival_2$edad))
```

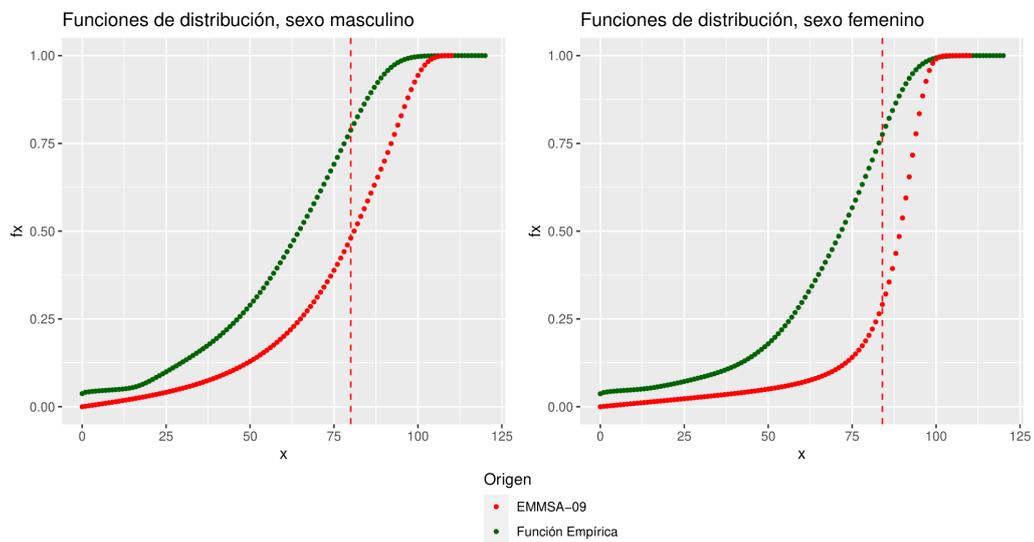
El estadístico de prueba utilizado en la prueba Kolmogórov-Smirnov se define según la expresión (5.6) (Rohatgi y Saleh, 2015). Este estadístico mide la discrepancia máxima observada entre las dos distribuciones, y para un nivel de significancia de la prueba del 99 %, se rechaza la hipótesis nula si $D > \frac{1.63}{\sqrt{n}} = 0.00163$ (Rohatgi y Saleh, 2015).

$$D := \max_x (|F_{z,CNSF}(x) - \hat{F}_{X,z}(x)|) \quad (5.6)$$

La visualización de la prueba se presenta en las gráficas [5.12]. El valor calculado del estadístico fue de 0.3076448 para el subgrupo masculino, y de 0.4862075 para

el subgrupo femenino. En ambos casos, se rechazó la hipótesis nula H_0 con un nivel de significancia del 99 %. Por lo tanto, se concluyó que las estimaciones resultantes de un suavizamiento sobre las probabilidades de muerte de la población de México capturan de un comportamiento diferente respecto a las estimaciones exhibidas en las tasas de mortalidad de activos para la seguridad social mexicana EMSSAH-09 y EMSSAM-09.

Figura 5.12: Contraste de las funciones de distribución empírica y teórica de la prueba.



5.4. Propuestas y Conclusiones

Tras el análisis de los resultados, no solo se determinó que el procedimiento de suavizamiento por splines proporciona estimaciones sustancialmente diferentes a las dictaminadas por la CNSF, sino que también se identificó evidencia que sugiere una posible sobreestimación de la mortalidad actual en las tablas de vida EMSSAH-09 y EMSSAM-09. En ambos subgrupos de sexo, el modelo suavizado exhibió probabilidades de muerte inferiores desde la edad 0 hasta los 101 años en el caso femenino y los 108 años en el masculino. Sin embargo, este fenómeno se compensó por un decremento mucho más gradual en la función de supervivencia

suavizada, a comparación de su homólogo regulatorio. Dicho distintivo se manifestó de manera más pronunciada en el subgrupo femenino, como se evidencia en las gráficas [5.3]. Este resultado respalda la tendencia general de crecimiento en la esperanza de vida, como consecuencia de los avances en tratamientos médicos.

En comparación con las técnicas alternativas vistas en la sección 2.3, no fue necesario implementar restricciones explícitas para la forma del modelo, como aquellas expuestas en las expresiones (2.30) y (2.31). Además, la metodología propuesta ofrece una aproximación más precisa para funciones que involucran cálculos infinitesimales, como la función de fuerza de mortalidad, en contraste con enfoques discretos como el de (2.35).

En cuanto al aspecto práctico, es relevante resaltar la disponibilidad de los registros de mortalidad general del INEGI para uso público. Aunque esta investigación no discrimina los datos, se abre la posibilidad de hacerlo según la ocupación, causa de muerte, entidad estatal o municipal, entre otras variables, con el objetivo de obtener estimaciones demográficas más detalladas para conjuntos de individuos específicos. Esto podría traducirse en probabilidades de muerte más precisas, sacrificando robustez estadística en aras de una estimación más especializada.

No obstante, existen desventajas importantes en el uso de este procedimiento. Por un lado, la elección de los años de registro para la ventana temporal 2012-2021, fue completamente arbitraria. Utilizar una cantidad de registros menor (conservando sólo los años más recientes por ejemplo), o mayor (de ser disponibles), hubiera proporcionado resultados diferentes. Igualmente, se eligió implementar la técnica de suavizamiento por splines sobre las funciones $l(x)$ y $\mu(x)$, sin embargo, también hubiera sido posible construir una tabla de vida que incorporase todos los registros disponibles, sin agruparlos por año de registro. Sumado a estas alternativas, también se sugiere como tema de investigación futuro explorar si las estimaciones obtenidas mediante este método son objetivamente superiores a las comparaciones realizadas, puesto que la hipótesis propuesta sólo contemplaba validar si eran diferentes.

Un objetivo implícito de esta investigación fue proporcionar una metodología

clara y flexible, accesible para todos los lectores, para obtener estimaciones de probabilidades de muerte. Se repasaron los requisitos teóricos para la aplicación de un modelo de suavizamiento por splines, con el propósito de dotar a cualquier interesado de las herramientas necesarias para personalizar este modelo según sus necesidades y objetivos de investigación. Finalmente, este estudio fue un esfuerzo para alentar a futuros investigadores de incursionar en el descubrimiento de técnicas novedosas, que fortalezcan y desarrollen la profesión actuarial en México y en el mundo.

Bibliografía

- Adresi, Y., Varol, B., Kurt Ömürlü, İ., & Türe, M. (2020). Comparison Of Piecewise Regression and Polynomial Regression Analyses In Health and Simulation Data Sets. <https://doi.org/10.22312/sdusbed.636203>
- Bertrand, C., Ernest, F., & Hao, H. (2009). *Springer Series in Statistics* (Vol. 27). <http://www.springerlink.com/index/D7X7KX6772HQ2135.pdf>
- Booth, H., & Tickle, L. (2008). Mortality Modelling and Forecasting: a Review of Methods. *Annals of Actuarial Science*, 3(1-2), 3-43. <https://doi.org/10.1017/S1748499500000440>
- Bowers, N., Gerber, H., Hickman, J., Jones, D., & Nesbit, C. (1986). *Actuarial Mathematics* (Vol. 93). <https://doi.org/10.1080/00029890.1986.11971867>
- Colegio Nacional de Actuarios. (2021). Historia de la actuaría en México De sus Orígenes hasta el 2021, 1-80. <https://dgesui.ses.sep.gob.mx/sites/default/files/2022-02/HISTORIA%20DE%20LA%20ACTUAR%20%7B%5C%27%7BI%7D%7DA%20EN%20M%20%7B%5C%27%7BE%7D%7DXICO.%20DE%20SUS%20OR%20%7B%5C%27%7BI%7D%7DGNES%20HASTA%20EL%202021.pdf>
- Comisión Nacional de Seguros y Fianzas. (2022). Comparativo de Metodologías para el cálculo de Tablas de Mortalidad de Pensiones, 14. <https://www.cnsf.gob.mx/cnsf/revista/Documentos%20compartidos/Comparativo%20de%20Metodolog%20%7B%5C%27%7Bi%7D%7Das%20para%20el%20c%27%7B%5C%27%7Ba%7D%7Dlculo%20de%20Tablas%20de%20Mortalidad%20de%20Pensiones.pdf>
- Comisión Nacional de Seguros y Fianzas. (2023). Circular Única de Seguros y Fianzas. <https://www.gob.mx/cms/uploads/attachment/file/825846/>

02%7B%5C_%7DCircular%7B%5C_%7D%7B%5C_%7Dnica%7B%5C_%7Dde%7B%5C_%7DSeguros%7B%5C_%7Dy%7B%5C_%7DFianzas%7B%5C_%7Dcompulsada%7B%5C_%7Dsin%7B%5C_%7DAnexos%7B%5C_%7D%7B%5C_%7D16-may-2023%7B%5C_%7D.pdf

Contreras, F. H., & Márquez, F. P. (2018). Modelo Mexicano de Supervisión Basado en Riesgos tipo Solvencia 2. (171).

Escuder Bueno, J., & Escuder Vallés, R. (2009). *Historia de las Tablas de Mortalidad Españolas y su Evolución*. Universidad de Huelva. <http://www.ahepe.es/Documentos/IVCongreso-Sevilla-Huelva2007/HISTORIA%20DE%20LA%20PROBABILIDAD%20Y%20LA%20ESTADISTICA%20IV.pdf>

Estepa, A., Gea, M., Cañadas, G., & Contreras, J. (2012). Algunas notas históricas sobre la correlación y regresión y su uso en el aula. *Números. Revista de Didáctica de las Matemáticas*, 81, 5-14. http://www.sinewton.org/numeros/numeros/81/Articulos%7B%5C_%7D01.pdf

Forfar, D. O. (2004a). History of Actuarial Education. En *Encyclopedia of Actuarial Science*. John Wiley & Sons, Ltd. <https://doi.org/10.1002/9780470012505.tah011>

Forfar, D. O. (2004b). Mortality Laws. En *Encyclopedia of Actuarial Science*. John Wiley & Sons, Ltd. <https://doi.org/10.1002/9780470012505.tam029>

Instituto Nacional de Estadística y Geografía. (2014). *Estadística de defunciones generales. Síntesis metodológica*.

Instituto Nacional de Estadística y Geografía. (2021). Mortalidad. https://www.inegi.org.mx/programas/mortalidad/%7B%5C#%7DDatos%7B%5C_%7Dabiertos

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning* (Vol. 103). Springer New York. <https://doi.org/10.1007/978-1-4614-7138-7>

Neter, J., & Wasserman, W. (1997). Applied Linear Regression Models. *Technometrics*, 39(3), 342-342. <https://doi.org/10.1080/00401706.1997.10485142>

- Oirov, T., Terbish, G., & Dorj, N. (2021). B-spline estimation for force of mortality. *Mathematics and Statistics*, 9(5), 736-743. <https://doi.org/10.13189/ms.2021.090513>
- Ortega, A. (1987). *La tabla de mortalidad*.
- Perryman, F. (1949). International Actuarial Notation. *Journal of the Institute of Actuaries*, 75(1), 121-129. <https://doi.org/10.1017/S0020268100012956>
- Pollock, D. (1999). Smoothing with Cubic Splines. En *Handbook of Time Series Analysis, Signal Processing, and Dynamics* (pp. 293-322). Elsevier. <https://doi.org/10.1016/B978-012560990-6/50013-0>
- Ramírez, M. M., Gómez, A. M. M., & Torres, E. M. (2000). Tablas de Mortalidad CNSF 2000-I y CNSF 2000-G Serie Documentos de Trabajo. (80), 18.
- Rice, J., & Rosenblatt, M. (1983). Smoothing Splines: Regression, Derivatives and Deconvolution. *The Annals of Statistics*, 11(1), 1403-1433. <https://doi.org/10.1214/aos/1176346065>
- Rohatgi, V. K., & Saleh, A. K. M. E. (2015). *An Introduction to Probability and Statistics*. Wiley. <https://doi.org/10.1002/9781118799635>
- Sampieri Hernández, R. (2014). *Metodología de la Investigación* (6.^a ed.). McGraw Hill Education.
- Sengupta, D. (2001). Linear Models in Statistics. *Journal of the American Statistical Association*, 96(455), 1138-1138. <https://doi.org/10.1198/jasa.2001.s414>
- Stanton, J. M. (2001). Galton, Pearson, and the Peas: A Brief History of Linear Regression for Statistics Instructors. *Journal of Statistics Education*, 9(3). <https://doi.org/10.1080/10691898.2001.11910537>
- Villareal, H., & Macías, A. (2020). El sistema de pensiones en México; Institucionalidad, gasto público y sostenibilidad financiera. *Macroeconomía del desarrollo; CEPAL, Naciones Unidas.*, 42. <https://hdl.handle.net/11362/45780>