

ACTUARIOS TRABAJANDO

Modelación empírica de la mortalidad en México mediante el método de suavizamiento polinómico por nodos

Erick Morales Bonilla y Denise Gómez Hernández

DOI: <https://doi.org/10.64806/cnmx.actstrab.18/20250908023>



Modelación empírica de la mortalidad en México mediante el método de suavizamiento polinómico por nodos

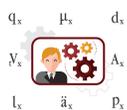
Erick Morales Bonilla* y Denise Gómez Hernández*,¹

*Universidad Autónoma de Querétaro

RESUMEN Este estudio presenta una estimación empírica de la mortalidad en México mediante el uso del método de suavizamiento polinómico por nodos, aplicado a registros de defunciones de 2012 a 2021. A través de una depuración de más de 7.5 millones de datos del INEGI, se construyeron tablas de vida diferenciadas por sexo, utilizando modelos de suavizamiento para estimar funciones de supervivencia y fuerza de mortalidad con mayor precisión que los modelos tradicionales. Los resultados muestran diferencias significativas respecto a las tablas oficiales EMSSAH-09 y EMSSAM-09 utilizadas por la Comisión Nacional de Seguros y Fianzas, detectadas mediante la prueba Kolmogorov-Smirnov. La investigación sugiere que estas tablas podrían subestimar la mortalidad actual. Además, se destaca la flexibilidad del modelo propuesto, la ventaja del procesamiento computacional moderno y su potencial para aplicarse a subgrupos específicos de población. No obstante, se reconoce como limitación la selección arbitraria del periodo de análisis y la posibilidad de incorporar otras metodologías o niveles de desagregación. Finalmente, el estudio aporta una herramienta metodológica accesible para generar estimaciones demográficas más precisas, promoviendo así la mejora continua de la práctica actuarial en México.

Palabras clave

Tablas de vida, Suavizamiento por splines, Modelación empírica, Seguridad social.



Derechos reservados © 2025 por el Colegio Nacional de Actuarios
Última actualización del manuscrito: 4 de junio de 2025

Artículo formateado por Kristian Frich Villaseñor, y comunicado por José Daniel López Barrientos. La visión y comentarios contenidos en este documento corresponden exclusivamente a su autor y, por tanto, son ajenos a cualquier entidad pública o privada, incluyendo aquella para la que actualmente colabora. El CONAC no recomienda ni respalda el uso de la información proporcionada en este estudio. El CONAC no ofrece ninguna garantía, expresa o implícita, ni representación de ningún tipo y no asume ninguna responsabilidad en relación con el uso o mal uso de este trabajo.

1. INTRODUCCIÓN

El comportamiento de la mortalidad es un factor socioeconómico clave que permea en prácticamente todos los sectores de una población. De este comportamiento se derivan las probabilidades de muerte que, a su vez, fungen como la piedra angular para los cálculos actuariales. Uno de los primeros intentos por modelar las probabilidades de fallecimiento fueron las tablas de mortalidad. De acuerdo a [Forfar \(2004\)](#), algunos ejemplos de las leyes más utilizadas son las de De Moivre, Gompertz y Makeham. 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.

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 [Comisión Nacional de Seguros y Fianzas \(2022\)](#) 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.

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 contramedidas operan en diversas dimensiones, abarcando desde lo microeconómico, 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. Para [Contreras y Márquez \(2018\)](#), 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.

Es así que, la estabilidad financiera global depende en gran medida del correcto funcionamiento de este contexto, 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 este 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.

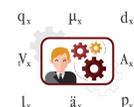
En otras palabras, el uso de los modelos actuales implica, de manera implícita, la aceptación del riesgo de no capturar completamente el verdadero comportamiento de las variables en estudio. No obstante, gracias a los avances en la capacidad de procesamiento de las máquinas modernas, muchos de los cálculos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

Es por todo lo expuesto anteriormente, que el objetivo de este trabajo es estimar las probabilidades de muerte sobre la población de México mediante un modelo de suavizamiento polinómico por nodos (*Smoothing Splines*) que pueda ser contrastado con los métodos actuales de forma numérica. (Vea el trabajo de [Rice y Rosenblatt \(1983\)](#).)

2. ANTECEDENTES

2.1. Modelos de Suavizamiento

De acuerdo a [James et al. \(2013\)](#), 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. 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.



¹Correo electrónico: denise.gomez@uaq.edu.mx

El análisis de regresión es una herramienta estadística que permite estudiar la relación entre dos o más variables, y así, poder inferir resultados. Según [Stanton \(2001\)](#), 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. De acuerdo a [Estepa et al. \(2012\)](#), a pesar de no utilizar los conocimientos estadísticos del momento, desarrollados por otros académicos como Gauss y Laplace, Galton introduce el concepto de correlación para auxiliarse en el estudio de la genética hereditaria, utilizando métodos estadísticos sencillos y empíricos.

En el estudio de [Stanton \(2001\)](#) se detalla cómo, a causa de la carencia de rigor matemático en el trabajo de Galton, Karl Pearson respaldó y formalizó 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, Pearson publica su primer texto sobre el tema, titulado *Mathematical contributions to the theory of evolution*.

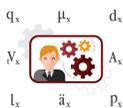
Es así como Isaac Jacob Schönberg propone la metodología de Suavizamiento por *splines* (vea la obra de [Rice y Rosenblatt \(1983\)](#)), la cual consiste en encontrar una relación funcional entre una variable objetivo Y y una variable predictora X que asemeje adecuadamente el comportamiento de la variable objetivo. Es decir, se desea una función cualquiera $g(x)$ que logre un error cuadrático $\sum_i (y_i - g(x_i))^2$ mínimo.

No obstante, sin imponer restricciones adicionales a $g(x)$, se puede proponer una función que imite en cada punto las observaciones y_i , minimizando el error cuadrático medio, pero ocasionando un sobre ajuste. Así, según [James et al. \(2013\)](#), la manera de considerar también la variabilidad de $g(x)$ es mediante un componente extra, $\lambda \cdot \int_{-\infty}^{\infty} g''(t)^2 dt$, el cual penaliza y regula la flexibilidad de la función $g(x)$ encontrada. 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, vea (1).

$$\min \left\{ \sum_i (y_i - g(x_i))^2 + \lambda \cdot \int_{-\infty}^{\infty} g''(t)^2 dt \right\}. \quad (1)$$

Así, con el 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 Suavizamiento por Splines, al descubrir que la función $g(x)$ que minimiza el cálculo es precisamente un *Spline Natural Cúbico* (de tercer orden) con n nodos, donde n denota el número de observaciones en la variable X (vea el trabajo de [Pollock \(1999\)](#)).

Como especifica [Varol et al. \(2020\)](#), un *Spline Natural Cúbico* es el resultado de una Regresión Polinómica por Partes



modificada. Es decir, si se desea ajustar un polinomio de tercer grado, 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 cuatro coeficientes de regresión, para un total $4(k + 1)$ coeficientes de regresión en el modelo. Esta estructura de modelo recibe el nombre de “Regresión Polinómica por Partes” y su forma se puede ver en (2).

Aunque el modelo (2) goza de una gran flexibilidad, la curva ajustada resultante suele ser discontinua en los nodos, ocasionando un sobreajuste 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 tercer orden con k nodos, es una regresión polinómica por partes en donde la curva, y sus primeras dos derivadas son continuas en todos los nodos (vea el trabajo de [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, [James et al. \(2013\)](#) comenta que se acostumbra a utilizar uno de tercer grado, y en su lugar aumentar o disminuir la cantidad de nodos utilizados para partir la variable X , variando así la flexibilidad del ajuste.

Finalmente, se obtiene un “Spline Natural Cúbico” al modificar los extremos del ajuste de un Spline de Regresión Polinómico; en el cual, en lugar de utilizar un polinomio en el primer y último segmento de X , se ajusta una línea. De acuerdo con [James et al. \(2013\)](#), este cambio permite corregir la divergencia que exhiben las funciones polinomiales en sus extremos.

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 (3), la cual se puede reescribir mediante las llamadas “funciones base” (citadas por [Neter y Wasserman \(1997\)](#)) que permiten modelar relaciones no lineales entre X y Y , compare con (4). La expresión (4), a su vez, se puede reescribir como (5) con la igualdad $X_j = b_j(X) = X^j$.

La expresión (5) corresponde a la estructura clásica de un modelo de Regresión Lineal Múltiple. [Neter y Wasserman \(1997\)](#) comentan las condiciones siguientes sobre esta ecuación:

- La variable ε es aleatoria, tal que

$$E(\varepsilon) = 0 \text{ y } \text{Var}(\varepsilon) = \sigma^2$$

es desconocida.

$$Y = \begin{cases} \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \varepsilon, & x < \xi_1, \\ (\beta_0 + \beta_4) + (\beta_1 + \beta_5)x + (\beta_2 + \beta_6)x^2 + (\beta_3 + \beta_7)x^3 + \varepsilon, & \xi_1 \leq x < \xi_2, \\ \vdots \\ (\beta_0 + \beta_{4k}) + (\beta_1 + \beta_{4k+1})x + \cdots + (\beta_3 + \beta_{4k+3})x^3 + \varepsilon, & \xi_k \leq x, \end{cases} \quad (2)$$

- Las variables X_p no son variables aleatorias.
- Cada observación de ε es independiente entre sí, o bien,

$$\text{Cov}(\varepsilon_i, \varepsilon_j) = 0 \text{ para toda } i \neq j.$$

- El conjunto de variables predictoras $X_{i,1}, \dots, X_{i,p}$ son linealmente independientes.

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \cdots + \beta_p X_i^p + \varepsilon_i, \quad (3)$$

$$Y_i = \beta_0 + \beta_1 b_1(X_i) + \cdots + \beta_p b_p(X_i) + \varepsilon_i, \quad (4)$$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \varepsilon. \quad (5)$$

Estos supuestos permiten obtener el valor esperado de la variable objetivo, mostrado en (6).

$$\mathbb{E}(Y_i) = \beta_0 + \beta_1 X_{i,1} + \cdots + \beta_p X_{i,p}. \quad (6)$$

La estimación $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \cdots + \hat{\beta}_p X_p$ busca encontrar aquellos valores de $\hat{\beta}_1, \dots, \hat{\beta}_p$ 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 (7):

$$\min_{\beta_0, \dots, \beta_p} \left\{ \sum (y_i - \hat{y}_i)^2 \right\}, \quad (7)$$

con $\sum (y_i - \hat{y}_i)^2 = \sum (y_i - \beta_0 - \beta_1 x_{i,1} - \cdots - \beta_p x_{i,p})^2$. Dada la creciente complejidad de notación, seguimos a [Sengupta \(2001\)](#) y utilizamos notación matricial para desarrollar los resultados del modelo. Si se tienen m observaciones, a cada Y_i con $i = 1, \dots, m$ se le puede asociar el resultado de una multiplicación de matrices, que puede

$$\text{de reescribirse como } Y = X\beta + \varepsilon, \text{ donde } Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_m \end{pmatrix},$$

$$X = \begin{pmatrix} 1 & X_{1,1} & X_{1,2} & \cdots & X_{1,p} \\ 1 & X_{2,1} & X_{2,2} & \cdots & X_{2,p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{m,1} & X_{m,2} & \cdots & X_{m,p} \end{pmatrix}, \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} \text{ y}$$

$$\varepsilon = \begin{pmatrix} \varepsilon_0 \\ \varepsilon_1 \\ \vdots \\ \varepsilon_m \end{pmatrix}. \text{ Así, siguiendo a } \text{Sengupta (2001)}, \text{ sabemos que}$$

los valores de la matriz de estimadores de los coeficientes de regresión, definida en (8), alcanzan un mínimo cuando se obtienen de (9).

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_p \end{pmatrix}, \quad (8)$$

$$\hat{\beta} = (X^T X)^{-1} X^T Y, \quad (9)$$

donde X^T denota la matriz transpuesta de X , y $(X^T X)^{-1}$ es la matriz inversa de $X^T X$. Note que esta expresión permite reescribir la matriz de estimaciones como en (10), donde se suele referir a $H = X(X^T X)^{-1} X^T$ como la matriz de proyección.

$$\hat{Y} = X\hat{\beta} = X(X^T X)^{-1} X^T Y = HY. \quad (10)$$

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 (1), es mediante una metodología de selección de modelo



llamada *Leave One Out Cross Validation* (LOOCV), citado por [James et al. \(2013\)](#), que consiste en construir el modelo excluyendo un par de observaciones (x_i, y_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 a continuación.

$$ECM_{LOOCV} = \frac{1}{n} \sum_i ECM_i. \quad (11)$$

Sin embargo, [James et al. \(2013\)](#) demuestran 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 H , h_i , como se ve aquí:

$$ECM_{LOOCV} = \frac{1}{n} \sum_i \left(\frac{y_i - \hat{y}_i}{1 - h_i} \right)^2. \quad (12)$$

[James et al. \(2013\)](#) proponen una malla de valores para λ , donde el valor elegido será aquel que alcance el menor ECM_{LOOCV} de entre todos los demás propuestos.

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, que a su vez es emitida por la 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” (vea el trabajo de la [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. El artículo titulado “Comparativo de Metodologías para el cálculo de Tablas de Mortalidad de Pensiones”, liberado por la [Comisión Nacional de Seguros y Fianzas \(2022\)](#) 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. 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 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. Así, en este modelo, para cada edad $x = 0, 1, \dots, 110$ se definen tres variables, que están relacionadas entre sí mediante la expresión

$$E_x = P_{x+1} + d_x, \quad (13)$$

donde P_{x+1} representa el número de pólizas o de expuestos a edad $x + 1$ al final del año, d_x es el número de muertes de edad x observadas durante el año, y E_x es el número de personas expuestas en edad x al principio del año. La ecuación (13) permite calcular, desde un punto de vista frecuentista, la tasa bruta de mortalidad observada, de la forma expuesta en $q_x = \frac{d_x}{E_x}$.

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. Además, siguiendo a [Ramírez et al. \(2000\)](#), se hace uso de la función de enlace logit, que describe la relación entre las variables q_x y Y_i :

$$Y_i = \ln \left(\frac{\hat{q}_x}{1 - \hat{q}_x} \right) = \beta_0 + \beta_1 X_i + \varepsilon_i \quad (14)$$

Un modelo alternativo que proponen [Ramírez et al. \(2000\)](#) es el de Lee-Carter, el cual se describe como una combinación de un modelo extrapolativo y series de tiempo, lo cual se logra mediante un ajuste que se muestra a continuación.

$$q_{x,t} = \exp(a_x + \beta_x \cdot k_t + \varepsilon_{x,t}). \quad (15)$$

En la ecuación (15), t denota el tiempo y x_t la diferencia de cada cohorte generacional [Ramírez et al. \(2000\)](#). Es así como se obtiene:

- 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.

El trabajo de [Ramírez et al. \(2000\)](#) 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, 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. Por ejemplo, se puede consultar una estimación sobre la fuerza de mortalidad de la población de Mongolia, mediante ajustes de *Splines* en el trabajo de [Oirov et al. \(2021\)](#). Este procedimiento de *Splines* 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 $c = (c_1, c_2, \dots, c_k)$ y de manera similar, al vector de coeficientes de regresión como $\theta = (\theta_1, \theta_2, \dots, \theta_{k+n})$ donde n expresa el orden del *Spline*. De esta manera, cada *Spline* se puede definir como una función de estos dos vectores con $g(x, \theta, c)$.

Como la fuerza de mortalidad $\mu(x)$ debe satisfacer que $\mu(x) \geq 0$, para todo valor de x , [Oirov et al. \(2021\)](#) utilizan el logaritmo natural de esta función con $\ln(\mu(x))$, cuya imagen corresponde a todos los valores reales. Finalmente, el planteamiento del modelo se muestra en (16), donde se iguala la fuerza de mortalidad al *Spline* obtenido por máxima verosimilitud.

$$\ln(\mu(x)) = g(x, \theta, c), \quad (16)$$

o lo que es equivalente a $\mu(x) = \exp(g(x, \theta, c))$ para resolver la ecuación (16). A diferencia del *Spline* obtenido de minimizar (11), [Oirov et al. \(2021\)](#) utilizan una prueba χ^2 de bondad de ajuste modificada para obtener la forma del *Spline* ajustado. Es posible relacionar $\mu(x)$ con la función $F_X(x) = \Pr(X \leq x)$ para todo $x \geq 0$, mediante la expresión $F_X(x) = 1 - \exp[-\int_0^x \mu(t) dt]$ [Bowers et al. \(1986\)](#), de manera que $F_X(x)$ denota a la función de distribución obtenida de calcular $\mu(x) = \exp(g(x, \theta, c))$. La hipótesis nula de la prueba de bondad de ajuste se convierte en la igualdad (17).

$$H_0 : F_X(x) = F_X(x, \theta), \quad (17)$$

donde $F_X(x)$ es la función de distribución verdadera de la vida denotada como X usada por [Oirov et al. \(2021\)](#). Los valores asignados a los vectores θ y 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 en

$$\hat{F}_X(x) = \frac{n_1 + n_2 + \dots + n_x}{N}, \quad x = 1, 2, \dots, 100, \quad (18)$$

donde n_x se define como el número de muertes ocurridas entre las edades $[x-1, x]$ para $x = 1, 2, \dots, 100$, y $N = \sum n_x$.

Esta función empírica después se utiliza para obtener otra estimación para la fuerza de mortalidad de la forma dada a continuación.

$$\hat{\mu}(x) = \frac{\hat{F}_X(x) - \hat{F}_X(x-1)}{1 - \hat{F}_X(x)}, \quad x = 1, 2, \dots, 100. \quad (19)$$

La expresión (19) nos da una estimación para la fuerza de mortalidad para un individuo de edad x .

3. MÉTODOS Y TÉCNICAS

Para lograr el objetivo de este trabajo, que es estimar las probabilidades de muerte sobre la población de México mediante un modelo de suavizamiento polinómico por nodos, se definen a continuación las técnicas y métodos cuantitativos que se basan principalmente en la definición de las variables objeto de estudio, el universo y la muestra estadística. Posteriormente, se presentan los resultados en la sección 4, derivados de la aplicación de estas técnicas y métodos.

3.1. Definición de variables

Para determinar la probabilidad de muerte de la población en México y de acuerdo a la revisión de la literatura de este trabajo, se definen como variables independientes y dependientes las que se describen a continuación.

Variables Independientes: año de registro de muerte, r ; sexo de nacimiento, z puede ser “masculino” o “femenino”.

Variable Dependiente: número de sobrevivientes a edad discreta, k , con defunción registrada el año r y sexo z ; representada por $L(k)_{r,z}$.

Covariables:

- $n_{r,z}$: Número de registros disponibles que pertenecen a la cohorte de defunciones registradas para el año r y sexo z .
- $X_{i,r,z}$: Edad de muerte para el individuo i , con año registrado r y sexo z .
- (x) : Edad del individuo.
- $\hat{l}(x)_{r,z}$: Función suavizada del número de sobrevivientes a edad (x) para el año registrado r y sexo z , y que representa el suavizamiento de la función $L(k)_{r,z}$ para la edad (x) mediante *Splines*.
- $D(k)_{r,z}$: Cuento de los individuos que fallecen con edad alcanzada k , donde $k = 1, \dots, \max_i(X_{i,r,z})$.
- $\hat{s}(x)_{r,z}$: Función estimada de supervivencia a edad (x) , para el año registrado r y sexo z , a través de $\hat{l}(x)_{r,z}$.
- $\hat{\mu}(x)_{r,z}$: Fuerza de mortalidad a edad (x) , para el año registrado r y sexo z , estimada a través de $\hat{s}(x)_{r,z}$.

Para el cálculo de las variables $L(k)_{r,z}$ y $D(k)_{r,z}$ se definieron las fórmulas (20) y (21), de acuerdo con la [Comisión Nacional](#)



de Seguros y Fianzas (2022).

$$L(0)_{r,z} = n_{r,z} \quad (20)$$

$$D(k)_{r,z} = \sum_{i=1}^{n_{r,z}} \mathbb{1}(k \leq X_{i,r,z} < k+1). \quad (21)$$

Se estableció además que

$$L(0)_{r,z} = L(k-1)_{r,z} - D(k-1)_{r,z},$$

lo que permitió obtener de forma iterativa a $L(k)_{r,z}$. Para garantizar que se cumplieran las propiedades de una función de supervivencia, se optó por suavizar la función $L(k)_{r,z}$ mediante un *Spline* natural cúbico, obteniendo así la función continua $\hat{l}(x)_{r,z}$ con la expresión siguiente

$$\hat{l}(x)_{r,z} := \min_g \left\{ \sum_{k=1}^{\max(X_{i,r,z})} (L(k)_{r,z} - g(k)_{r,z})^2 + \lambda \cdot \int_{-\infty}^{\infty} g''(t)_{r,z}^2 dt \right\}. \quad (22)$$

La función de supervivencia se estima con la expresión (23), de acuerdo con Bowers *et al.* (1986).

$$\hat{s}(x)_{r,z} = \frac{\hat{l}(x)_{r,z}}{\hat{l}(0)_{r,z}}. \quad (23)$$

Así, 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 (23) como se muestra a continuación.

$$\frac{d}{dx} \hat{s}(x)_{r,z} = \frac{d}{dx} \left(\frac{\hat{l}(x)_{r,z}}{\hat{l}(0)_{r,z}} \right). \quad (24)$$

Finalmente, (23) y (24) ofrecen una expresión equivalente para la fuerza de mortalidad estimada en términos de la función suavizada $\hat{l}(x)_{r,z}$, mostrada a continuación.

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

3.2. Definición del universo

Este estudio se basa en datos históricos de la población general de México, restringiendo el análisis a individuos cuyo fallecimiento fue registrado en el periodo 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.



Asimismo, no se discriminó la muestra por ninguna causa de defunción ni por características sociales específicas. Como se expuso anteriormente, las probabilidades de muerte utilizadas en la matemática actuarial incorporan una amplia gama de factores sociodemográficos característicos de la población en su conjunto y no están limitadas a causas particulares de defunción.

3.3. 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 que presentaban algún tipo de problema en la captura, reduciendo así el tamaño final de la muestra a 6,865,137 registros.

4. RESULTADOS Y DISCUSIÓN

4.1. 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 1. Dado que la muestra fue discriminada en dos subgrupos principales según el sexo, se presentarán las características de cada subgrupo por separado.

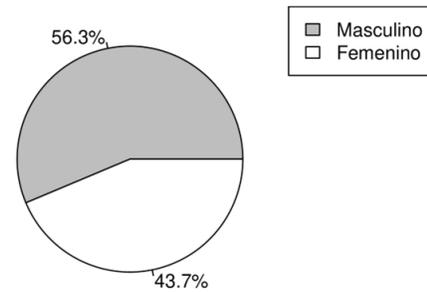
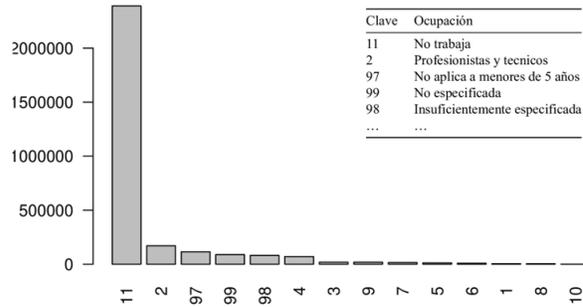


Figura 1 Composición de la muestra. Elaboración propia.

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

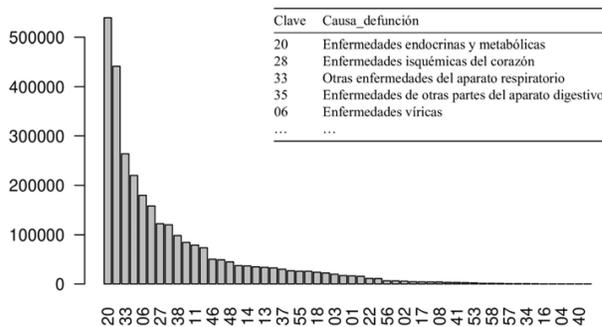
laboral estuvo mayormente conformada por mujeres sin ocupación laboral, seguido de aquellas con perfiles profesionales y técnicos, y finalmente, se encontraron individuos cuya ocupación no fue especificada. Esto está representado en la figura 2.

Figura 2 Distribución de ocupación laboral, género femenino. Elaboración propia.



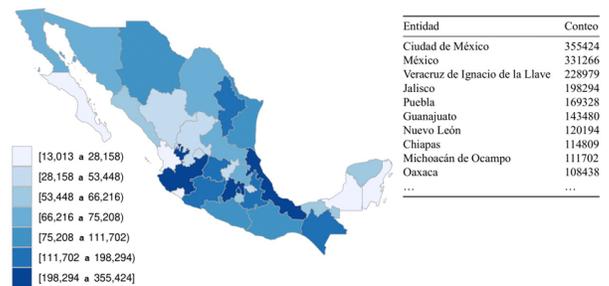
Respecto a la causa de fallecimiento para el género femenino, se identificaron como principales razones las enfermedades endocrinas y metabólicas, seguidas de las enfermedades isquémicas del corazón y las enfermedades del aparato respiratorio. La jerarquía completa de estas causas puede visualizarse en el histograma presentado en la figura 3.

Figura 3 Distribución de causas de fallecimiento, género femenino. Elaboración propia.



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 figura 4.

Figura 4 Entidad de ocurrencia, género femenino. Elaboración propia.



En relación con el 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. Asimismo, 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, 6 y 7 respectivamente.

Figura 5 Distribución de ocupación laboral, género masculino. Elaboración propia.

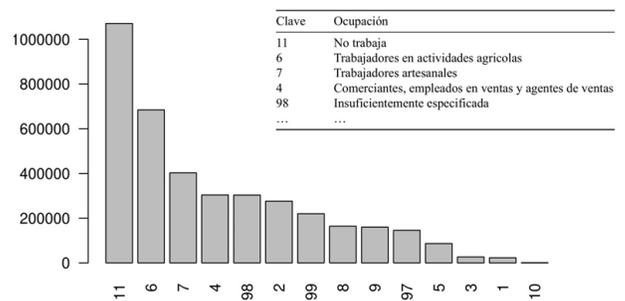


Figura 6 Distribución de causas de fallecimiento, género masculino. Elaboración propia.

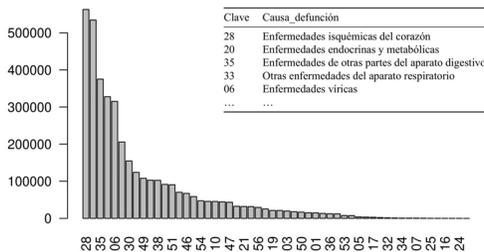
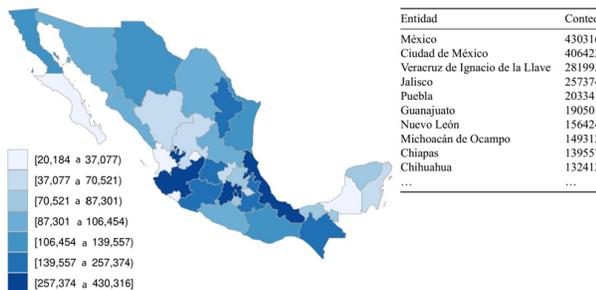


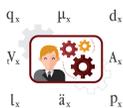
Figura 7 Entidad de ocurrencia, género masculino. Elaboración propia.



4.2. Cálculo de variables

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. 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.

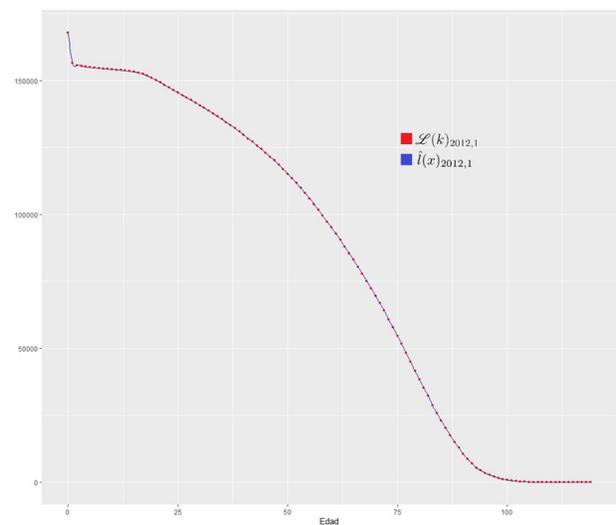
De acuerdo a la definición previa de las variables, se obtiene entonces que el año de registro toma los valores $r = 2012, 2013, \dots, 2021$. Además, el sexo de nacimiento se define con los valores $z = 1, 2$ (donde 1 representa el sexo masculino y 2 el femenino). A continuación, se llevó a cabo un 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 formación final de la base de datos.



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.

Los resultados para el cálculo de la variable $\hat{l}(x)_{r,z}$, y que en específico en este trabajo se obtiene la función continua $\hat{l}(x)_{2021,1}$ al solucionar la expresión (22), se muestran en la figura 8.

Figura 8 Función $\hat{l}(x)_{2021,1}$ suavizada por Splines con el método natural cúbico. Elaboración propia.



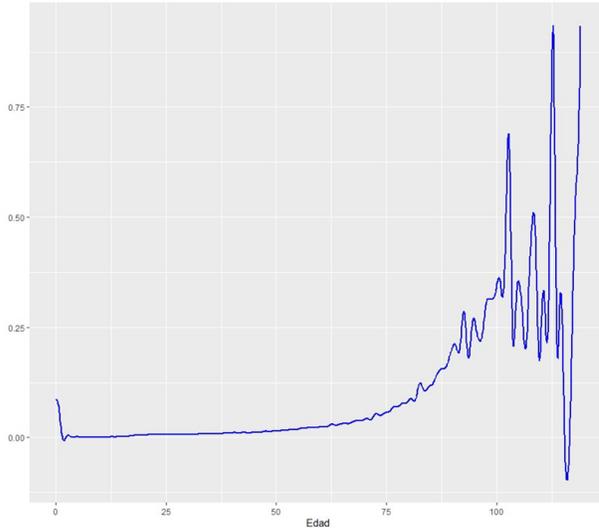
Entre los valores de la función $\hat{l}(x)_{2021,1}$, se encuentra el valor calculado para λ , obtenido mediante la metodología LOOCV expuesta en las ecuaciones (11)-(12), el cual resultó igual a $\lambda_{LOOCV} = 1.87365 \times 10^{-16}$, lo que equivale a la expresión siguiente.

$$\frac{1}{n} \sum_i (y_i - \hat{l}(x)_{2012,1,\lambda_{LOOCV}})^2 \leq \frac{1}{n} \sum_i (y_i - \hat{l}(x)_{2012,1,\lambda_j})^2,$$

para toda $\lambda_j \neq \lambda_{LOOCV}$. Por otro lado, se muestra la función $\hat{\mu}(x)_{2021,1}$, obtenida mediante splines por suavizamiento en la figura 9, cuyo procedimiento es completamente análogo para todos los años de registro disponibles y es posible de generalizar, ciclar y almacenar.

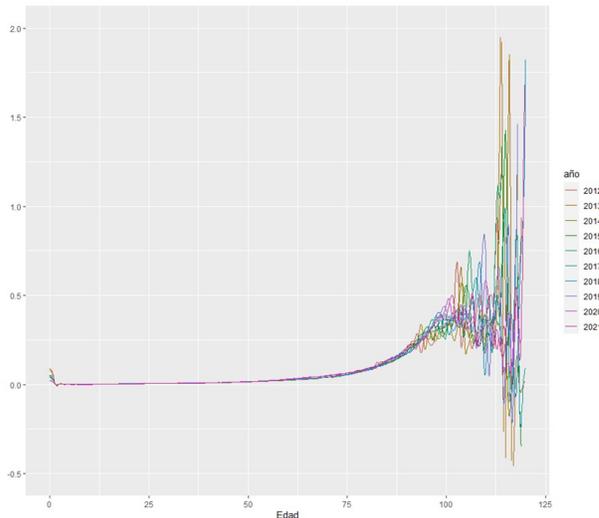
Al calcular y obtener las funciones de fuerza de mortalidad estimada para cada año de registro, se hizo evidente la presencia

Figura 9 Fuerza de mortalidad estimada 2012, género masculino. Elaboración propia.



de una considerable variabilidad entre las estimaciones. Estas diferencias se representan gráficamente en la figura 10.

Figura 10 Funciones de fuerza de mortalidad estimadas, género masculino. Elaboración propia.



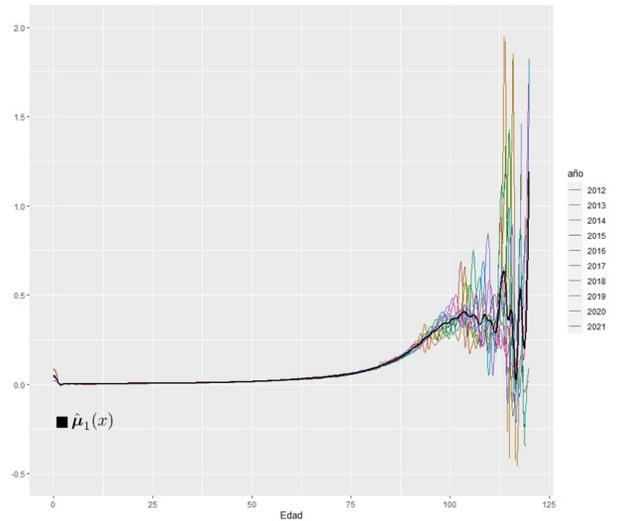
Con el fin de disminuir esta variabilidad y de incorporar

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 siguiente.

$$\hat{\mu}_z(x) := \min_g \left\{ \sum_{r=2012}^{2021} \sum_i (\hat{l}(i)_{r,z} - g_z(i))^2 + \lambda \cdot \int_{-\infty}^{\infty} g_z''(x)^2 dx \right\}.$$

El procesamiento de la información permite obtener la función $\hat{\mu}_z(x)$, que representa la fuerza de mortalidad estimada para el sexo masculino. Esta función logró incorporar los comportamientos observados en todos los años analizados, además de presentar una baja variabilidad, lo cual se visualiza en la figura 11. A partir de esta función es posible extraer y calcular todas las variables necesarias para construir una tabla de vida que englobe el comportamiento de mortalidad encontrado.

Figura 11 Función final de fuerza de mortalidad estimada, género masculino. Elaboración propia.



Construir la función final de fuerza de mortalidad estimada para el sexo femenino $\hat{\mu}_z(x)$, involucró un proceso semejante, sustituyendo el valor de la variable $z = 2$ en la programación previamente mencionada. En la figura 12 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 12 Función final de fuerza de mortalidad estimada, género femenino. Elaboración propia.

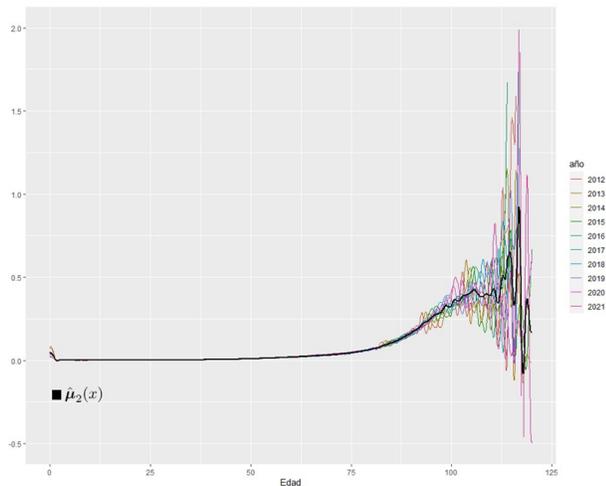
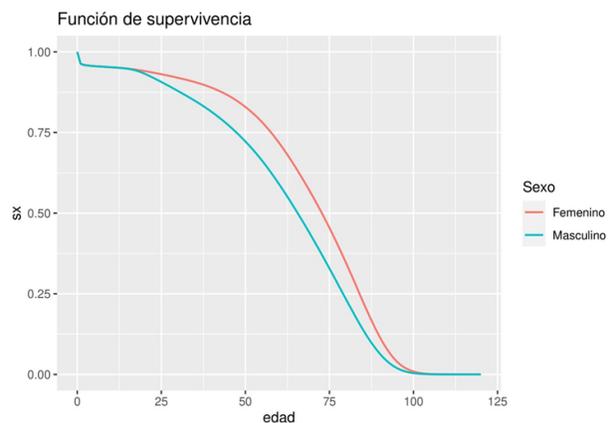


Figura 13 Función de supervivencia obtenida por sexos. Elaboración propia.



4.3. Cálculo de tabla de mortalidad

Obtenidas las funciones de fuerza de mortalidad, μ_z para el sexo masculino y μ_z para el femenino, fue posible derivar todas las funciones asociadas al uso de probabilidades de muerte como se muestra a continuación:

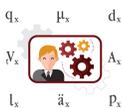
$$\exp\left(-\int_0^n \mu(x) dx\right) = \frac{s(n)}{s(0)} = \frac{s(n)}{1} = s(n),$$

para $n > 0$. Aunque es posible derivar una expresión exacta para la integral $\int_0^n \mu(x) dx$, esto requeriría integrar todos los nodos que integran la función suavizada, lo cual resulta impreciso. En su lugar, la integral se puede aproximar mediante un método numérico. Ambas funciones se pueden ver en la figura 13.

De forma intrínseca se incorporan la totalidad de los datos disponibles en los registros, al obtener $\hat{s}(x)$ a partir de $\mu(x)$. Por último, al definir $l_0 = 100,000$ se logró construir las tablas de mortalidad mostradas en la tablas 1-3.

4.4. Comparación empírica con otras tablas de mortalidad

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 determinan que existen diferencias en una función específica, estas divergencias también se



manifestarán en todas las demás funciones, ya que todas están estrechamente relacionadas. Para realizar esta comparación se decidió 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 edad, a través de los valores de q_x indicados en las tablas EMSSAH-09 y EMSSAM-09. Para lograr esto, primero seguimos a [Bowers et al. \(1986\)](#) y obtuvimos la expresiones siguientes:

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

$$s(x+1) = (1 - q_x) \cdot s(x).$$

Los resultados se encuentran en la tablas 4-6.

La figura 14 muestra la divergencia entre las estimaciones de las tablas EMSSAH-09 y EMSSAM-09, y las tablas 4-6. La diferencia es evidente.

No obstante, se decidió emplear la prueba de bondad de ajuste Kolmogorov-Smirnov, en la cual se contrasta la hipótesis nula de que una muestra aleatoria dada sigue alguna distribución de probabilidad teórica propuesta. Es decir, si denotamos por $F_{X,CNSF}$ a la función de distribución exhibida por las tablas de vida EMSSAH-09 y EMSSAM-09 para $z = 1, 2$ respectivamente, y $X_{z,t}$ 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

■ **Tabla 1** Tabla de vida construida por sexos para las edades de 0 a 51 años. Elaboración propia.

Edad	lx_masculino	qx_masculino	lx_femenino	qx_femenino
0	100,000.00	0.0369221	100,000.00	0.0365917
1	96,307.79	0.0041892	96,340.83	0.0037477
2	95,904.34	0.0011124	95,979.77	0.0016054
3	95,797.66	0.0013370	95,825.68	0.0012818
4	95,669.58	0.0008940	95,702.86	0.0009646
5	95,584.05	0.0008195	95,610.55	0.0008345
6	95,505.72	0.0007434	95,530.76	0.0007294
7	95,434.73	0.0007058	95,461.08	0.0007303
8	95,367.36	0.0006871	95,391.37	0.0006623
9	95,301.84	0.0006973	95,328.19	0.0007109
10	95,235.38	0.0007393	95,260.42	0.0007269
11	95,164.98	0.0007743	95,191.17	0.0007350
12	95,091.29	0.0009015	95,121.21	0.0008834
13	95,005.56	0.0010481	95,037.18	0.0009832
14	94,905.98	0.0014205	94,943.74	0.0011023
15	94,771.17	0.0018854	94,839.08	0.0013679
16	94,592.49	0.0025115	94,709.34	0.0015422
17	94,354.93	0.0032742	94,563.28	0.0016632
18	94,045.99	0.0039775	94,406.01	0.0017779
19	93,671.93	0.0045653	94,238.17	0.0018719
20	93,244.29	0.0050643	94,061.77	0.0019842
21	92,772.07	0.0053656	93,875.13	0.0020246
22	92,274.28	0.0056489	93,685.07	0.0020752
23	91,753.03	0.0058099	93,490.65	0.0021532
24	91,219.96	0.0059359	93,289.35	0.0022460
25	90,678.48	0.0060943	93,079.82	0.0023201
26	90,125.86	0.0061528	92,863.87	0.0023430
27	89,571.34	0.0063921	92,646.29	0.0025013
28	88,998.79	0.0064473	92,414.55	0.0024539
29	88,424.99	0.0065476	92,187.77	0.0025692
30	87,846.02	0.0066609	91,950.93	0.0026446
31	87,260.88	0.0067352	91,707.75	0.0026972
32	86,673.17	0.0070018	91,460.40	0.0028768
33	86,066.30	0.0071114	91,197.29	0.0030503
34	85,454.25	0.0072646	90,919.11	0.0031161
35	84,833.45	0.0076423	90,635.80	0.0033605
36	84,185.13	0.0079028	90,331.22	0.0036439
37	83,519.83	0.0082713	90,002.06	0.0038359
38	82,829.01	0.0085611	89,656.83	0.0041865
39	82,119.90	0.0090544	89,281.48	0.0044889
40	81,376.36	0.0094831	88,880.70	0.0047577
41	80,604.66	0.0098241	88,457.83	0.0051768
42	79,812.78	0.0104818	87,999.90	0.0055620
43	78,976.21	0.0107961	87,510.44	0.0059660
44	78,123.57	0.0114153	86,988.35	0.0065190
45	77,231.77	0.0120134	86,421.28	0.0069738
46	76,303.95	0.0126178	85,818.59	0.0075300
47	75,341.16	0.0132653	85,172.38	0.0082634
48	74,341.74	0.0139652	84,468.56	0.0088682
49	73,303.54	0.0145694	83,719.48	0.0094174
50	72,235.55	0.0153249	82,931.05	0.0100656
51	71,128.55	0.0161575	82,096.30	0.0109119

supone que la distribución empírica de $X_{z,t} \sim F_{X,CNSF}$. Esto es equivalente a plantear las hipótesis expuestas en (25).

$$H_0 : X_{z,t} = F_{X,CNSF} \text{ vs. } H_1 : X_{z,t} \neq F_{X,CNSF}. \quad (25)$$

■ **Tabla 2** Tabla de vida construida por sexos para las edades de 52 a 86 años. Elaboración propia.

Edad	lx_masculino	qx_masculino	lx_femenino	qx_femenino
52	69,979.29	0.0171651	81,200.48	0.0116797
53	68,778.09	0.0181549	80,252.08	0.0126291
54	67,529.42	0.0191893	79,238.57	0.0136426
55	66,233.58	0.0204448	78,157.55	0.0145905
56	64,879.45	0.0215715	77,017.20	0.0155112
57	63,479.90	0.0227467	75,822.57	0.0166609
58	62,035.94	0.0239268	74,559.30	0.0179429
59	60,551.62	0.0250289	73,221.49	0.0190115
60	59,036.07	0.0264452	71,829.44	0.0201270
61	57,474.85	0.0275716	70,383.73	0.0212363
62	55,890.18	0.0292141	68,889.04	0.0225864
63	54,257.40	0.0302514	67,333.09	0.0238802
64	52,616.04	0.0320283	65,725.16	0.0251233
65	50,930.84	0.0332071	64,073.93	0.0263373
66	49,239.57	0.0350492	62,386.40	0.0276423
67	47,513.76	0.0366005	60,661.89	0.0288565
68	45,774.73	0.0390402	58,911.40	0.0307837
69	43,987.68	0.0411069	57,097.89	0.0319295
70	42,179.48	0.0433218	55,274.78	0.0340647
71	40,352.19	0.0453834	53,391.86	0.0357733
72	38,520.87	0.0487046	51,481.86	0.0384665
73	36,644.73	0.0517126	49,501.53	0.0407687
74	34,749.73	0.0546877	47,483.42	0.0430342
75	32,849.35	0.0582211	45,440.01	0.0463478
76	30,936.82	0.0625121	43,333.97	0.0497419
77	29,002.90	0.0675412	41,178.45	0.0535261
78	27,044.01	0.0723180	38,974.33	0.0582527
79	25,088.24	0.0771699	36,703.97	0.0622437
80	23,152.18	0.0832827	34,419.38	0.0683659
81	21,224.00	0.0890135	32,066.27	0.0726628
82	19,334.78	0.0981228	29,736.24	0.0814168
83	17,437.60	0.1055157	27,315.21	0.0896849
84	15,597.66	0.1140285	24,865.45	0.0963752
85	13,819.08	0.1220585	22,469.04	0.1037292
86	12,132.34	0.1310469	20,138.34	0.1127017

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 $(0, 1)$ para simular valores aleatorios equiprobables entre 0 y 1. Estos valores se introducen posteriormente en alguna función de distribución inversa, en este caso $F_z^{-1}(x)$, para obtener realizaciones aleatorias de X_z . Esto es:

$$X_{z,i} = F_z^{-1}(U_i), \text{ para } i = 1, \dots, n. \quad (26)$$

El tamaño de la muestra aleatoria n_z dependerá de la cantidad de valores extraídos de los valores aleatorios con distribución uniforme U_i . Es importante enfatizar que, mientras $F_{X,CNSF}(x)$ representa la distribución teórica de la variable aleatoria X_z obtenida mediante suavizado 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 (26) y (19). Esto es:

$$D := \max_x |F_{X,CNSF}(x) - \hat{F}_{X_z}(x)| \quad (27)$$



■ **Tabla 3** Tabla de vida construida por sexos para las edades de 77 a 120 años. Elaboración propia.

Edad	lx_masculino	qx_masculino	lx_femenino	qx_femenino
87	10,542.44	0.1410209	17,868.72	0.1243636
88	9,055.73	0.1530408	15,646.50	0.1355771
89	7,669.84	0.1653049	13,525.19	0.1476404
90	6,401.98	0.1802720	11,528.33	0.1601642
91	5,247.88	0.1922704	9,681.90	0.1712760
92	4,238.87	0.2068379	8,023.62	0.1916315
93	3,362.11	0.2249165	6,486.04	0.2047162
94	2,605.92	0.2368292	5,158.25	0.2182004
95	1,988.76	0.2517885	4,032.71	0.2338731
96	1,488.01	0.2627052	3,089.57	0.2436604
97	1,097.10	0.2771772	2,336.77	0.2557088
98	793.01	0.2874340	1,739.23	0.2772323
99	565.07	0.2904527	1,257.06	0.2760963
100	400.95	0.3048427	909.99	0.2997517
101	278.72	0.3076645	637.22	0.3013631
102	192.97	0.3197835	445.19	0.3147910
103	131.26	0.3326217	305.05	0.3249809
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

Figura 14 Comparativa de funciones de supervivencia. Elaboración propia.

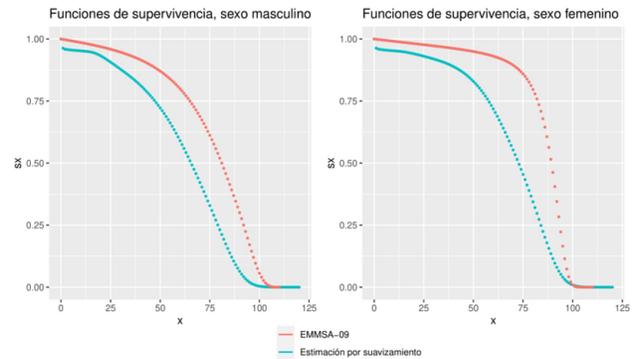
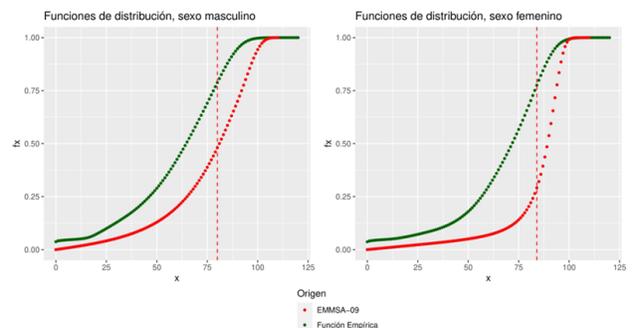
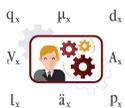


Figura 15 Contraste de las funciones de distribución empírica y teórica de la prueba. Elaboración propia.



Según Rohatgi y Saleh (2015), el estadístico de prueba expresado por (27) mide la discrepancia máxima observada entre las dos distribuciones, y para un nivel de significancia α de la prueba del 99 %, se rechaza la hipótesis nula si se cumple la condición $D > D_{0.01} = 0.0163$. La visualización de la prueba se presenta en la figura 15. El valor calculado del estadístico fue de $D = 0.3706448$ para el subgrupo masculino y de $D = 0.48562075$ para el subgrupo femenino. En ambos casos, se rechaza la hipótesis nula aún con un nivel de significancia del 99 %. Por lo tanto, se concluye que las estimaciones resultantes de un suavizamiento sobre las probabilidades de muerte de la población de México provienen 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.



■ **Tabla 4** Tablas de vida EMMSA-09 para las edades de 0 a 51 años. Elaboración propia.

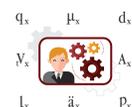
Edad	Hombres qx	Mujeres qx	Hombres sx	Mujeres sx
0	0.00138	0.00092	1.0000000	1.0000000
1	0.00138	0.00092	0.9986200	0.9990800
2	0.00139	0.00092	0.9972419	0.9981608
3	0.00140	0.00092	0.9958557	0.9972425
4	0.00142	0.00092	0.9944615	0.9963251
5	0.00143	0.00092	0.9930494	0.9954085
6	0.00145	0.00092	0.9916293	0.9944927
7	0.00147	0.00092	0.9901915	0.9935777
8	0.00149	0.00092	0.9887359	0.9926637
9	0.00152	0.00092	0.9872627	0.9917504
10	0.00155	0.00092	0.9857620	0.9908380
11	0.00158	0.00092	0.9842341	0.9899264
12	0.00161	0.00092	0.9826790	0.9890157
13	0.00165	0.00092	0.9810969	0.9881058
14	0.00168	0.00092	0.9794781	0.9871967
15	0.00172	0.00092	0.9778326	0.9862885
16	0.00177	0.00092	0.9761507	0.9853811
17	0.00181	0.00092	0.9744229	0.9844746
18	0.00186	0.00093	0.9726592	0.9835689
19	0.00191	0.00093	0.9708501	0.9826541
20	0.00197	0.00093	0.9689957	0.9817403
21	0.00202	0.00093	0.9670868	0.9808273
22	0.00209	0.00094	0.9651333	0.9799151
23	0.00215	0.00094	0.9631162	0.9789940
24	0.00222	0.00095	0.9610455	0.9780737
25	0.00230	0.00095	0.9589120	0.9771445
26	0.00237	0.00096	0.9567065	0.9762163
27	0.00246	0.00096	0.9544391	0.9752791
28	0.00254	0.00097	0.9520911	0.9743428
29	0.00264	0.00098	0.9496728	0.9733977
30	0.00274	0.00099	0.9471657	0.9724438
31	0.00284	0.00100	0.9445705	0.9714811
32	0.00295	0.00101	0.9418879	0.9705096
33	0.00307	0.00102	0.9391093	0.9695294
34	0.00319	0.00104	0.9362262	0.9685404
35	0.00332	0.00105	0.9332397	0.9675332
36	0.00346	0.00107	0.9301413	0.9665173
37	0.00361	0.00109	0.9269230	0.9654831
38	0.00377	0.00111	0.9235768	0.9644307
39	0.00393	0.00113	0.9200950	0.9633602
40	0.00411	0.00116	0.9164790	0.9622716
41	0.00430	0.00118	0.9127123	0.9611554
42	0.00450	0.00121	0.9087876	0.9600212
43	0.00471	0.00124	0.9046981	0.9588596
44	0.00493	0.00128	0.9004369	0.9576706
45	0.00517	0.00132	0.8959978	0.9564448
46	0.00542	0.00136	0.8913655	0.9551823
47	0.00569	0.00141	0.8865343	0.9538832
48	0.00598	0.00146	0.8814899	0.9525382
49	0.00629	0.00151	0.8762186	0.9511475
50	0.00661	0.00158	0.8707072	0.9497113
51	0.00696	0.00164	0.8649518	0.9482107

■ **Tabla 5** Tablas de vida EMMSA-09 para las edades de 52 a 86 años. Elaboración propia.

Edad	Hombres qx	Mujeres qx	Hombres sx	Mujeres sx
52	0.00733	0.00172	0.8589317	0.9466557
53	0.00772	0.00180	0.8526358	0.9450274
54	0.00814	0.00189	0.8460534	0.9433264
55	0.00859	0.00199	0.8391665	0.9415435
56	0.00906	0.00211	0.8319581	0.9396698
57	0.00957	0.00223	0.8244205	0.9376871
58	0.01011	0.00237	0.8165308	0.9355961
59	0.01069	0.00253	0.8082757	0.9333787
60	0.01131	0.00270	0.7996352	0.9310173
61	0.01198	0.00290	0.7905914	0.9285035
62	0.01268	0.00312	0.7811201	0.9258109
63	0.01344	0.00336	0.7712155	0.9229223
64	0.01425	0.00364	0.7608504	0.9198213
65	0.01512	0.00396	0.7500082	0.9164732
66	0.01604	0.00432	0.7386681	0.9128439
67	0.01703	0.00473	0.7268199	0.9089004
68	0.01810	0.00520	0.7144421	0.9046013
69	0.01924	0.00574	0.7015107	0.8998974
70	0.02045	0.00636	0.6880137	0.8947320
71	0.02176	0.00707	0.6739438	0.8890415
72	0.02316	0.00790	0.6592788	0.8827560
73	0.02467	0.00886	0.6440099	0.8757822
74	0.02628	0.00998	0.6281221	0.8680228
75	0.02801	0.01130	0.6116151	0.8593599
76	0.02986	0.01285	0.5944838	0.8496492
77	0.03185	0.01467	0.5767325	0.8387312
78	0.03399	0.01683	0.5583635	0.8264270
79	0.03629	0.01940	0.5393848	0.8125182
80	0.03875	0.02247	0.5198105	0.7967554
81	0.04139	0.02614	0.4996678	0.7788523
82	0.04423	0.03056	0.4789866	0.7584931
83	0.04728	0.03588	0.4578010	0.7353135
84	0.05055	0.04233	0.4361562	0.7089305
85	0.05406	0.05014	0.4141085	0.6789214
86	0.05783	0.05964	0.3917218	0.6448803

estimaciones sustancialmente diferentes a las dictaminadas por la CNSF, sino que también se identificó evidencia que sugiere una posible subestimación 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 la figura 15.

Es de resaltar que no fue necesario implementar restricciones explícitas para la forma del modelo, como aquellas expuestas en las expresiones (14) y (15). Además, la metodología propuesta ofrece una aproximación más precisa para funciones que involucran cálculos infinitesimales, como la función de fuerza



5. CONCLUSIONES

Tras el análisis de los resultados, no solo se determinó que el procedimiento de suavizamiento por *splines* proporciona

■ **Tabla 6** Tablas de vida EMMSA-09 para las edades de 87 a 110 años. Elaboración propia.

Edad	Hombres qx	Mujeres qx	Hombres sx	Mujeres sx
87	0.06187	0.07121	0.3690685	0.6064197
88	0.06621	0.08530	0.3462342	0.5632365
89	0.07087	0.10245	0.3233101	0.5151924
90	0.08147	0.12327	0.3003971	0.4624110
91	0.09207	0.14846	0.2759237	0.4054096
92	0.10439	0.17874	0.2505194	0.3452225
93	0.11835	0.21478	0.2243677	0.2835174
94	0.13418	0.25716	0.1978138	0.2226235
95	0.15212	0.30616	0.1712711	0.1653737
96	0.17247	0.36163	0.1452174	0.1147429
97	0.19554	0.42286	0.1201717	0.0732484
98	0.22170	0.48842	0.0966734	0.0422746
99	0.25135	0.55626	0.0752409	0.0216268
100	0.28497	0.62390	0.0563291	0.0095967
101	0.32309	0.68873	0.0402770	0.0036093
102	0.36630	0.74844	0.0272639	0.0011235
103	0.41530	0.80133	0.0172771	0.0002826
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

de mortalidad, en contraste con enfoques discretos como el de (19).

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 federativa u otras estructuras variables, con el objetivo de construir modelos estadísticos de mortalidad más detallados por factores 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 $\hat{l}(x)$. 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

$$q_x = \mu_x d_x$$

$$N_x = A_x$$

$$l_x = \ddot{a}_x p_x$$

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 por 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.

REFERENCIAS

- Bertrand, C., F. Ernest, y H. Hao, 2009 *Springer Series in Statistics*. Springer.
- Booth, H. y L. Tickle, 2008 Mortality modelling and forecasting: a review of methods. *Annals of Actuarial Science* **3**: 3–43.
- Bowers, N. L., H. U. Gerber, J. C. Hickman, D. A. Jones, y C. J. Nesbitt, 1986 Actuarial mathematics. En *Mathematics of Life Insurance*, The Society of Actuaries.
- Comisión Nacional de Seguros y Fianzas, 2022 Comparativo de metodologías para el cálculo de tablas de mortalidad de pensiones. Reporte técnico, Comisión Nacional de Seguros y Fianzas.
- Comisión Nacional de Seguros y Fianzas, 2023 Circular Única de seguros y fianzas. Reporte técnico, Comisión Nacional de Seguros y Fianzas.
- Contreras, F. H. y F. P. Márquez, 2018 Modelo mexicano de supervisión basado en riesgos tipo solvencia 2. Reporte técnico 171, Comisión Nacional de Seguros y Fianzas.
- Estepa, A., M. Gea, G. Cañadas, y J. Contreras, 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.
- Forfar, D. O., 2004 Mortality laws. En *Encyclopedia of Actuarial Science*, John Wiley & Sons, Ltd.
- James, G., D. Witten, T. Hastie, y R. Tibshirani, 2013 *An Introduction to Statistical Learning*, volumen 103. Springer New York.
- Neter, J. y W. Wasserman, 1997 *Applied Linear Regression Models*, volumen 39. McGraw-Hill.
- Oirov, T., G. Terbish, y N. Dorj, 2021 B-spline estimation for force of mortality. *Mathematics and Statistics* **9**: 736–743.
- Pollock, D. S. G., 1999 Smoothing with cubic splines. En *Handbook of Time Series Analysis, Signal Processing, and Dynamics*, pp. 293–322, Elsevier.
- Ramírez, M. M., A. M. M. Gómez, y E. M. Torres, 2000 Tablas de mortalidad cnsf 2000-i y cnsf 2000-g serie documentos de

trabajo. Reporte técnico 80, Comisión Nacional de Seguros y Fianzas.

Rice, J. y M. Rosenblatt, 1983 Smoothing splines: Regression, derivatives and deconvolution. *The Annals of Statistics* **11**: 1403–1433.

Rohatgi, V. K. y A. K. M. E. Saleh, 2015 *An Introduction to Probability and Statistics*. Wiley.

Sengupta, D., 2001 Linear models in statistics. *Journal of the American Statistical Association* **96**: 1138–1138.

Stanton, J. M., 2001 Galton, pearson, and the peas: A brief history of linear regression for statistics instructors. *Journal of Statistics Education* **9**.

Varol, B., I. Kurt Omurlu, y M. Türe, 2020 Comparison of piecewise regression and polynomial regression analyses in health and simulation data sets. *Süleyman Demirel Üniversitesi Sağlık Bilimleri Dergisi* **11**: 144–151.



Erick Morales Bonilla es licenciado en Actuaría por la Universidad Autónoma de Querétaro, donde se graduó en 2023. Cuenta con certificaciones de la Society of Actuaries, habiendo aprobado los exámenes P (Probabilidad) y FM (Matemáticas Financieras). Actualmente se desempeña como Especialista de Datos Senior en Traxión, donde colabora en el desarrollo de infraestructura

de datos, la implementación de modelos de optimización e inteligencia artificial, así como en la generación de reportes de inteligencia de negocios.

Entre 2024 y 2025 trabajó como Científico de Datos en Turing, participando en un proyecto conjunto con Google para el entrenamiento del modelo de lenguaje Gemini, enfocado en tareas de análisis, modelado y evaluación de resultados en ciencia de datos.

De 2022 a 2024 formó parte de Accumetrica Quantitative Research, donde lideró el diseño de pipelines de datos, la creación de bases relacionales y el desarrollo de soluciones de inteligencia de negocio con enfoque en análisis de mercado.



Denise Gómez Hernández se doctoró en Ciencias Actuariales por la Bayes Business School (antes Cass Business School). También es maestra en Ciencias Actuariales por la Herriot Watt University. Denise estudió Actuaría en la Universidad de las Américas Puebla. Desde 2008 y hasta la actualidad, se desempeña como profesora e investigadora de tiempo completo en la Universidad

Autónoma de Querétaro.

Da clases de modelos actuariales para el seguro de vida, teoría del interés y pensiones privadas a nivel licenciatura y de Métodos Cuantitativos a nivel posgrado. Es miembro del Sistema Nacional de Investigadoras e Investigadores Nivel II y es líder del Cuerpo Académico “Gestión Tecnológica y Financiera”.

A nivel profesional, trabajó como consultora actuarial en Towers Perrin, Sungard y la OCDE; y como jefa de unidad en el área de riesgos financieros en CONSAR. Sus líneas de investigación incluyen modelos actuariales en pensiones y análisis de indicadores financieros.

