

Metodología propuesta para la suavización de la estructura de la mortalidad: el modelo estadístico no paramétrico Spline frente a la técnica demográfica de la mano alzada¹

Luis Alberto Meza Santa Cruz²

RESUMEN

Por décadas en algunos países latinoamericanos se ha utilizado la “técnica demográfica de mano alzada”, para la suavización de las curvas correspondientes a los componentes del crecimiento de la población, así como en la suavización de la estructura de la mortalidad, etc.

En los lustros iniciales de la segunda mitad del siglo XX la falta de medios tecnológicos hizo imposible la utilización de modelos más sofisticados para la suavización de la estructura de la mortalidad, específicamente la gráfica de las tasas de mortalidad por sexo y edades.

Al llegar la década de los '80 del siglo pasado, el descubrimiento de la computadora hizo posible retomar los estudios de los modelos matemáticos y adicionarle la incertidumbre de cuál sería el valor verdadero de una variable estudiada, lo que condujo a los modelos estadísticos-matemáticos.

Ya Schoenberg (1964) y Reinsche (1967) habían propuesto una formulación moderna del modelo no paramétrico Spline, que mucho antes había sido también propuesto con unas variantes por Whittaker (1923), para finalmente Grace Wahba (1990) presentar su implementación en la Estadística³.

Al comparar las gráficas resultantes de la aplicación de las dos metodologías, se observa mayor precisión en la gráfica que corresponde al modelo Spline.

Se concluye que la utilización del modelo Spline en la suavización de las estructuras de mortalidad sería la más adecuada e incluso se dispone de programas para, o que contienen, el modelo Spline (R, MatLab, JMP, entre otros).

INTRODUCCIÓN

Como es bien sabido, la estructura de la mortalidad por sexo y edad, es la base para la construcción de las tablas de mortalidad. En las últimas décadas en algunos países latinoamericanos se ha mejorado el registro y procesamiento de los hechos vitales, entre ellos las defunciones ocurridas en un año determinado, disminuyendo el porcentaje de omisión, que en el siglo pasado, llegaba o pasaba del 50 por ciento. Ello hace plausible, suponiendo que la omisión no sea diferencial por sexo y edad, que la estructura obtenida de esta información incompleta de defunciones, no variaría en el supuesto de tener el registro completo de las mismas.

Una vez obtenida una o dos estructuras de mortalidad representativas en el tiempo, se puede construir una serie de tablas de mortalidad, aplicando las técnicas y metodología existentes para la elaboración de las mismas, a la vez que proyectarlas al futuro. Las tablas de vida o de mortalidad por sexo, proporcionan las relaciones de sobrevivencia por sexo y edad, que son un insumo necesario para realizar las estimaciones y proyecciones de población de los

¹ Trabajo presentado en el VI Congreso de la Asociación Peruana de Demografía y Población, realizado en Lima-Perú, del 12 al 15 de agosto de 2014.

² Asociación Peruana de Demografía y Población. lamezasc@gmail.com.

³ Eubank, R.L. (1999). Nonparametric regression and Spline smoothing (2th ed). New York: Marcel Dekker, Inc.

diferentes países: nacional y áreas mayores (llámese departamentos, provincias, estados federados, etc.).

De la precisión obtenida en la elaboración adecuada de una estructura de mortalidad por sexo y edad, se podrá obtener tablas de mortalidad representativas del comportamiento de las muertes en las áreas en estudio, y que al ser elaboradas, sean representativas del real comportamiento de la mortalidad, con una mayor precisión y menor sesgo.

En el siglo pasado, en los países latinoamericanos fue práctica común la utilización de tablas modelo de mortalidad, entre ellas las de Coale & Demeny, Naciones Unidas, Latinoamericana, etc. con las que mediante el uso de algunos parámetros propios de cada país o área mayor (tasa de mortalidad infantil, segmentos de tasas de mortalidad adulta proporcionada por alguna encuesta especializada, etc.), se podía determinar una estructura modelo que le sería representativa, o también suponiendo que la mortalidad se comportaba de acuerdo a las características de mortalidad de alguna de las cuatro familias de las tablas modelo de Coale & Demeny.

También como es sabido, las referidas tablas modelo de mortalidad (Coale & Demeny, Naciones Unidas, Latinoamericana, etc.) fueron construidas con información retrospectiva de defunciones del siglo XIX correspondientes mayormente países europeos y primera mitad del siglo XX, cuando las enfermedades y características de la mortalidad fueron muy diferentes al presente. Pasando a ser tablas inadecuadas por no decir obsoletas.

Ante esta realidad, en el 2001 durante la revisión de las estimaciones y proyecciones de población del Perú se decidió trabajar con las defunciones correspondientes a las estadísticas vitales de nuestro país, y lograr con ello tablas propias de mortalidad para cada quinquenio. Hasta las proyecciones anteriores se habían utilizado mayormente las tablas modelo de Coale & Demeny modelos Sur y/u Oeste.

Por décadas en la mayoría de países latinoamericanos se ha utilizado una combinación de medias móviles con el método demográfico de la mano alzada, este último que era más un procedimiento práctico que un método sustentado teóricamente, para la suavización de las curvas correspondientes a la estructura de la mortalidad por sexo.

Dada la poca robustez de este procedimiento, se comenzó a buscar alternativas más sólidas, para lograr el suavizado adecuado y aproximadamente preciso de las estructuras de mortalidad por sexo. Lógicamente se pensó en un modelo matemático, cuya teoría sea comprensible y su desarrollo fácilmente aplicable, a los variados conjuntos de datos de defunciones por sexo y edad.

En esta búsqueda se encuentra que el modelo no paramétrico Spline, tiene un parámetro que permite lograr una suavización en toda la extensión de la curva, a través de un polinomio de tercer grado, cuyo proceso de ajuste se acerca al método de mínimos cuadrados, y se logra una aproximación a un suavizado óptimo del conjunto de datos, solo que en este caso la suma de los errores al cuadrado no dará cero, ya que es una regresión no paramétrica, pero cuanto mejor sea el ajuste, esta suma de cuadrados será cercana a cero.

En los lustros iniciales de la segunda mitad del siglo XX la falta de medios tecnológicos hizo imposible la utilización de modelos más sofisticados para la suavización de la estructura de la mortalidad por sexo y edad.

Al llegar la década de los '80 del siglo pasado, el descubrimiento de la computadora hizo posible retomar los estudios de los modelos matemáticos, y adicionarle la incertidumbre de cuál sería el valor verdadero de una variable estudiada, lo que condujo a los modelos estadísticos-matemáticos.

OBJETIVOS

Objetivo general

Mostrar que el modelo de regresión no paramétrico Spline, se adecúa muy bien para la suavización de la estructura de la mortalidad por sexo y edad a datos de la Región de Lima, Perú.

Objetivos específicos

- a. Estudiar el modelo de regresión no paramétrico Spline.
- b. Aplicar el modelo de regresión no paramétrico Spline a datos de mortalidad de la Región de Lima, Perú para obtener la estructura de la mortalidad por sexo y edad.

METODOLOGÍA

El conocimiento e interés por las estructuras de la mortalidad en las poblaciones humanas, se remonta al siglo XVII.

Fue John Graunt quién en base a los Bills of Mortality, que eran boletines que se publicaban semanalmente en la ciudad de Londres desde comienzos del siglo XVII, y que contenía la relación de las defunciones (a veces los nacimientos) registrados en las diferentes Parroquias de la ciudad de Londres, publica en el año de 1662 *“Natural and political observations mentioned in a following index, and made upon the Bills of mortality, with reference to the government, religion, trade, growth, air, diseases and the several changes of the said city”*, y en el que se encuentran las primeras estructuras de mortalidad por sexo y edad. Se tiene conocimiento que fue el primero que aplicó el Análisis Exploratorio de Datos y calcula de forma simple las tasas de mortalidad por sexo y edad.

Posteriormente en el siglo XIX, exactamente en 1825, Gompertz construyó funciones matemáticas para las tasas de mortalidad en las edades adultas (de 45 a más años), para dicha elaboración se basó en funciones exponenciales.

En 1955, por iniciativa de V.G. Valaoras se publican las primeras series modernas de Tablas Modelo de Mortalidad de las Naciones Unidas, “se basa en un conjunto de 158 tablas de vida observadas para cada sexo”...“se construyeron bajo el supuesto que el valor de cada $5q_x$ ⁴ es una función cuadrática del valor q anterior”...“Como los coeficientes de las ecuaciones cuadráticas que relacionan cada valor $5q_x$ con su predecesor no se conocían a priori, tuvieron que estimarse en base a datos observados. Se recurrió a la regresión para estimar esos coeficientes con los 158 patrones de mortalidad disponibles para cada sexo”⁵.

En 1959 Ledermann y Breas mediante el Análisis Factorial determinan 5 factores que explicaban la variabilidad en 154 tablas de mortalidad observadas. “El primero y más importante se refiere al nivel general de la mortalidad; el segundo refleja la relación entre la mortalidad en la niñez y adulta; el tercero está relacionado con el patrón de la mortalidad en las edades avanzadas; mientras que el cuarto va asociado con los patrones de mortalidad por debajo de los cinco años y, por último, el quinto refleja las diferencias entre la mortalidad masculina y la femenina en las edades comprendidas entre los 5 y 70 años.”⁶

Coale y Demeny, en 1966, presenta sus cuatro familias de Tablas Modelo de Mortalidad (Norte, Sur, Este y Oeste), obtenidas en base a “coeficientes de las ecuaciones lineales que relacionaban los valores ${}_nq_x$ con e_{10} (esperanza de vida) a los 10 años, y de aquellas que relacionaban los valores de $\log_{10}({}_nq_x)$ con e_{10} , se estimaron utilizando la regresión por

⁴ Probabilidad de morir entre la edad x y la $x+5$ en una tabla de vida.

⁵ NN UU. Manual X Técnicas Indirectas de Estimación Demográfica. Departamento de Asuntos Económicos y sociales Internacionales, Estudios de Población, N° 81. Naciones Unidas, Nueva York, 1986.

⁶ NN UU. Manual X. Op. Cit.

mínimos cuadrados”⁷ que realizan sobre 192 tablas de mortalidad observadas (seleccionadas de un total de 326); 39 de ellas correspondían al siglo XIX, y 69 a después de la Segunda Guerra Mundial. Encontrándose sobre-representada la experiencia occidental (Europa, América del Norte, Australia y Nueva Zelanda con un total de 176 tablas; 3 de Israel, 6 de Japón, 3 de Taiwán y 4 de la población blanca de Sudáfrica.

En 1968 Brass y colegas, obtienen “un modelo que brinda un mayor grado de flexibilidad”...”mejor conocido como sistema logito. Brass intentó relacionar matemáticamente dos tablas de vida diferentes. Descubrió que una determinada transformación de las probabilidades de sobrevivir hasta la edad x (valores de $l(x)$ en términos de la tabla de vida) hacía que la relación entre las correspondientes probabilidades de las distintas tablas de vida resultase aproximadamente lineal”⁸.

En el Perú no se cuenta con modelos teóricos que representen adecuadamente la tendencia de la mortalidad en nuestro país, dada las bondades del Spline amerita utilizarla para obtener estructuras de mortalidad por sexo y edad que reflejen la realidad de la mortalidad peruana, teniendo en cuenta que hasta la actualidad no existen trabajos preliminares en Perú respecto a la utilización de este modelo específicamente. A finales de la década de los años 60 del siglo pasado se construyeron tablas de mortalidad para el Perú utilizando el modelo logito, posteriormente se utilizaron las tablas modelo de Coale y Demeny.

A nivel mundial los modelos de mortalidad elaborados son de la década de los años 60 del siglo pasado y para lo cual utilizaron el modelo logito en base a un modelo estándar de mortalidad, o regresión por mínimos cuadrados, obteniéndose en ambos casos modelo de mortalidad con información del siglo XIX y primera mitad del siglo XX, y que no representan la mortalidad actual, ya que además de no pertenecer a nuestra región, no reflejan la situación de salud que en su época si predominaba.

Ya Schoenberg (1964) y Reinsche (1967) habían propuesto una formulación moderna del modelo no paramétrico Spline, que mucho antes había sido también propuesto con unas variantes por Whittaker (1923), para finalmente Grace Wahba (1990) presentar su implementación en la Estadística⁹.

Whittaker con un procedimiento algo engorroso, logra una aproximación al modelo Spline que permite la suavización de la estructura de la mortalidad, obstáculo que no se presenta cuando se utiliza el modelo no paramétrico Spline utilizando un polinomio de tercer grado, cuya parte teórica pasamos a describir.

Espacio de los polinomios Spline

Propiedades básicas

1)

Sea $[a, b]$ un intervalo cerrado de extremos finitos,

y sea $\Delta = \{x_1, x_2, \dots, x_k\}$ con $a = x_0 < x_1 < x_2 < \dots < x_k < x_{k+1} = b$

una partición de $k + 1$ sub intervalos

$I_i = [x_i, x_{i+1})$, donde $i = 0, 1, 2, \dots, k - 1$ e $I_k = [x_k, x_{k+1}]$

⁷ NN UU. Manual X, Op. Cit.

⁸ NN UU. Manual X. Op. Cit.

⁹ Eubank, R.L. (1999). Nonparametric regression and Spline smoothing (2th ed). New York: Marcel Dekker, Inc.

2)

Sea m un entero positivo, y sea $M = (m_1, m_2, \dots, m_k)$ un vector de enteros (exponente de los términos del polinomio) con $1 \leq m_i \leq m, i = 1, 2, \dots, k$

Definición

Al espacio

$$\delta(P_m; M; \Delta) = \{s: \text{existen polinomios } s_0, \dots, s_k \text{ en } P_m \text{ tal que } s(x) = s_i(x)$$

para x perteneciente a $I_i, i = 0, 1, \dots, k$, y

$$D_j s_{i-1}(x_i) = D_j s_i(x_i) \text{ para } j = 0, 1, \dots, m-1-m_i, i = 1, \dots, k\}$$

Llamamos espacio de polinomios Spline de orden m con nodos x_1, x_2, \dots, x_k de multiplicidad (número de veces que se repite) m_1, m_2, \dots, m_k .

M es el vector multiplicidad, que controla la naturaleza del espacio $\delta^m(P_m; M; \Delta)$ mediante el suavizamiento de los Spline en los nodos (intersecciones).

Si $m_i = m$ el promedio de las dos secciones de polinomios adyacentes s_{i-1} y s_i en los intervalos adyacentes al nodo x_i son inconexos y pueden dar un salto discontinuo en x_i .

Si $m_i < m$ se logra que las dos secciones de polinomios se unan suavemente, de tal forma que el Spline s y sus primeras $m-1-m_i$ derivadas también son continuas en dicho nodo.

Por ejemplo, sea el espacio de polinomios de orden m

$$P_m = \left\{ p(x) : p(x) = \sum_{i=1}^m c_i x^{i-1}, c_1 \dots c_m, x \in \mathfrak{R} \right\}$$

Para $m=4$

$$\begin{aligned} p(x) &= \sum_{i=1}^4 c_i x^{i-1} = c_1 x^{1-1} + c_2 x^{2-1} + c_3 x^{3-1} + c_4 x^{4-1} = \\ &= c_1 x^0 + c_2 x^1 + c_3 x^2 + c_4 x^3 \end{aligned}$$

Según la segunda propiedad básica del espacio de polinomios Spline

$$M = (m_1, m_2, \dots, m_k) = (m_1, m_2, m_3, m_4) = (1, 2, 3, 4)$$

$$\Leftrightarrow k = 1, 2, 3, 4$$

$$1 \leq m_i \leq m, i = 1, 2, \dots, k, \text{ que en este caso } k = 4$$

$$1 \leq m_1 = 1 \leq m = 4 \quad \Rightarrow \quad 1 \leq 1 \leq 4$$

$$1 \leq m_2 = 2 \leq m = 4 \quad \Rightarrow \quad 1 \leq 2 \leq 4$$

$$1 \leq m_3 = 3 \leq m = 4 \quad \Rightarrow \quad 1 \leq 3 \leq 4$$

$$1 \leq m_4 = 4 \leq m = 4 \quad \Rightarrow \quad 1 \leq 4 \leq 4$$

Por la definición del espacio de polinomios Spline, si $m = 4$

$$\delta(P_4; M; \Delta) = \{s: \text{existen polinomios } s_0, \dots, s_k \text{ en } P_4 \text{ tal que } s(x) = s_i(x)$$

para x perteneciente a $I_i, i = 0, 1, \dots, k$, y

$$D_j s_{i-1}(x_i) = D_j s_i(x_i) \text{ para } j = 0, 1, \dots, m-1-m_i, i = 1, \dots, k\}$$

Como $M = (m_1, m_2, \dots, m_k) = (m_1, m_2, m_3, m_4) = (1, 2, 3, 4) \Rightarrow i = 1, 2, 3, 4$

Si $j = m - 1 - m_i$ con $i = 1, 2, 3, 4$

Para $i = 1 \Rightarrow j = 4 - 1 - m_1 = 4 - 1 - 1 = 4 - 2 = 2$

Para $i = 2 \Rightarrow j = 4 - 1 - m_2 = 4 - 1 - 2 = 4 - 3 = 1$

Para $i = 3 \Rightarrow j = 4 - 1 - m_3 = 4 - 1 - 3 = 4 - 4 = 0$

Para $i = 4 \Rightarrow j = 4 - 1 - m_4 = 4 - 1 - 4 = 4 - 5 = -1$

$\Rightarrow D^0 s_{i-1}(x_i) = D^0 s_i(x_i)$, la derivada en 0 es la misma función pero continua.

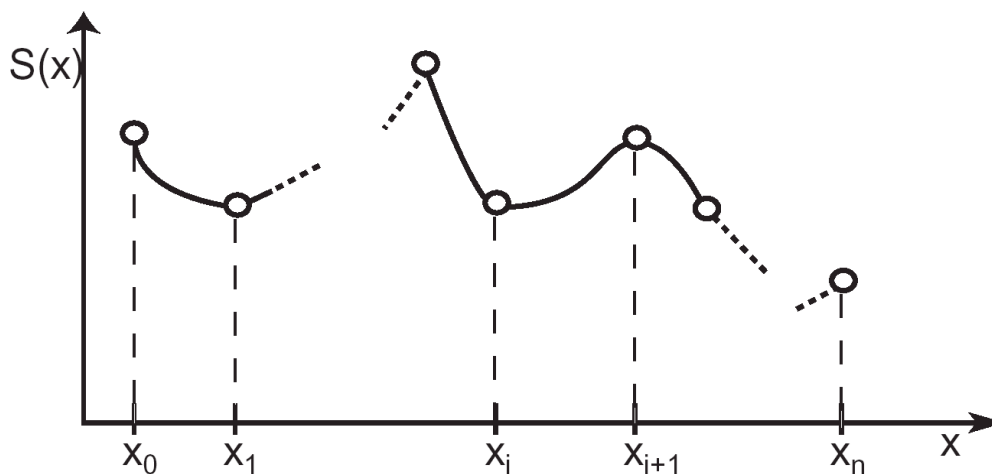
$\Rightarrow D^1 s_{i-1}(x_i) = D^1 s_i(x_i)$, la primera derivada suaviza la función, no puede hacer ángulo.

$\Rightarrow D^2 s_{i-1}(x_i) = D^2 s_i(x_i)$, en la segunda derivada la función mantiene el sentido de la concavidad.

Además por la propiedad 1) el polinomio Spline está definido en cada uno de los intervalos. Todo ello permite controlar el suavizamiento en los nodos (intersecciones), siendo esta las propiedades básicas del polinomio Spline, y que sirven para suavizar las curvas no lineales.

El polinomio Spline

Dados $n+1$ puntos $(x_0; f(x_0)), (x_1; f(x_1)), \dots, (x_n; f(x_n))$



Con $x_0 < x_1 < \dots < x_n$ y la función f , definidos en el intervalo $[a, b]$ que tiene nodos x_0, x_1, \dots, x_n , se aproximará la función f en cada subintervalo $[x_j; x_{j+1}]$, $j = 0, 1, \dots, n-1$, siendo el polinomio correspondiente

$$P_j(x) = c_1 + c_2(x - x_j) + c_3(x - x_j)^2 + c_4(x - x_j)^3,$$

con $j = 0, 1, \dots, n-1$

El modelo de regresión no paramétrico “Spline”

“El análisis de regresión permite construir modelos matemáticos que estudian la relación existente entre una variable dependiente y una o más variables independientes. Estos modelos se utilizan para estimar respuesta de valores futuros no observados de la o las variables independientes.

En el caso simple cuando ambas variable la dependiente Y , y la independiente X , son escalares, dadas las observaciones (x_i, y_i) para $i = 1, \dots, n$, se relaciona un modelo de regresión para dichas variables de la siguiente forma:

$$y_i = f(x_i) + \varepsilon_i, \quad \text{para } i = 1, \dots, n \quad (1)$$

Donde f es la función de regresión y ε_i son los errores aleatorios independientes con media cero y varianza común σ^2 . El objetivo del análisis de regresión es construir un modelo para f y realizar la estimación en base a los datos observados.

A menudo f es no lineal en x . Un enfoque frecuente es tratarla como una relación no lineal aproximando f a un polinomio de orden m

$$f(x) = \beta_0 + \beta_1 x + \dots + \beta_{m-1} x^{m-1} \quad (2)$$

En general, un modelo de regresión paramétrico asume que la forma de f es conocida excepto para un número finito de parámetros desconocidos. La forma específica de f puede provenir de la teoría conocida y/o aproximaciones a mecanismos bajo algunos supuestos simplificados. Los supuestos pueden ser demasiado restrictivos y la aproximación puede ser demasiado bruta para algunas aplicaciones. Un modelo inapropiado conduce a un sesgo sistemático y conclusiones engañosas. En la práctica, siempre se debe comprobar la forma supuesta para la función f .

Esto a menudo es difícil, a veces imposible, de obtener una forma funcional para f .

Un modelo de regresión no paramétrico no supone una forma predeterminada. En lugar de ello se hace supuestos de las propiedades cualitativas de f . Por ejemplo, uno puede estar dispuesto a asumir que f se suaviza y que no se reducirá a una forma específica con un número finito de parámetros. Más bien, en general conduce a algunos grupos de funciones infinito dimensionales.

La idea básica de la regresión no paramétrica es dejar que los datos hablen por sí mismos. Es dejar que la data decida que función se ajusta mejor sin imponer una forma específica de f . Por consiguiente los métodos no paramétricos son en general más flexibles. Ellos pueden descubrir estructuras en los datos que de lo contrario se perderían.

La técnica de regresión no paramétrica puede ser aplicada en diferentes partes del análisis de regresión: exploración de datos, construcción de modelos, pruebas de modelos paramétricos, y diagnósticos. De hecho, la suavización Spline es una herramienta potente y versátil para la construcción de modelos estadísticos en la explotación de estructuras de datos.

El polinomio (2) es un modelo global que suaviza las variaciones puntuales de un conjunto de datos. Las observaciones individuales pueden tener influencia negativa sobre los extremos del conjunto de datos.

Estas variaciones lleva a oscilaciones en ambos extremos del rango en el polinomio de ajuste. Una solución para superar esta limitación es la utilización de polinomios por tramos o trozos, que es la idea central de los polinomios Spline.

Sea $a < t_1 < \dots < t_k < b$ puntos fijos llamados nodos. Sea $t_0 = a$ y $t_{k+1} = b$. en términos generales los polinomios Spline son polinomios por trozos unidos suavemente en sus nodos o extremos. Formalmente, un polinomio Spline de orden r es una función real valorada sobre $[a, b]$, $f(t)$, de tal manera que:

- (i) f es un polinomio por trozos de orden r sobre $[t_i, t_{i+1}]$, $i = 0, 1, \dots, k$;
- (ii) f tiene $r-2$ derivadas continuas y la derivada $r-1$ (segunda derivada si $r=3$) es una función escalonada con saltos en los nudos.

Ahora consideremos a las órdenes representadas como $r = 2m$. La función f es un polinomio Spline natural de orden $2m$ si, en adición a (i) y (ii), satisface las condiciones naturales límite,

$$(iii) \quad f^{(j)}(a) = f^{(j)}(b) = 0, \quad j = m, \dots, 2m-1$$

Las condiciones naturales de contorno implican que f es un polinomio de orden m en los sub-intervalos extremos $[a, t_1]$ y $[t_k, b]$, denotando la función del espacio de polinomios naturales Spline de orden $2m$ con nodos t_1, \dots, t_k como

$$N S^{2m}(t_1, \dots, t_k)$$

Una aproximación conocida como regresión Spline, es acercar f utilizando un polinomio Spline o un polinomio natural Spline. Para conseguir una buena aproximación, se necesita decidir el número y ubicación de los nodos. Acá se cubre un enfoque diferente conocido como suavización Spline. Se empieza con un espacio modelo definido para f y se introduce una penalidad para evitar el exceso de ajuste. Luego se describe esta aproximación para los polinomios Spline.

Consideremos el modelo de regresión (1). Suponemos que f es un modelo suavizado. Específicamente, se asume que

$$f \in W_2^m[a, b]$$

donde el espacio Sobolev¹⁰

$$W_2^m[a, b] = \left\{ f : f, f', \dots, f^{(m-1)} \text{ son absolutamente continuas, } \int_a^b (f^{(m)})^2 dx < \infty \right\} \quad (3)$$

Para cualquier $a \leq x \leq b$, el Teorema de Taylor afirma que

$$f(x) = \underbrace{\sum_{v=0}^{m-1} \frac{f^{(v)}(a)}{v!} (x-a)^v}_{\text{polinomio de orden } m} + \underbrace{\int_a^x \frac{(x-u)^{m-1}}{(m-1)!} f^{(m)}(u) du}_{\text{Rem}(x)} \quad (4)$$

Está claro que el modelo de regresión polinomial (2) ignora el término del resto del polinomio $\text{Rem}(x)$ suponiendo que es insignificante. A menudo, en la práctica es difícil verificar esta

¹⁰ Los espacios de Sobolev hacen el papel de derivadas fraccionarias.

hipótesis. La idea detrás del suavizamiento Spline está en dejar que la data genere que tan grande debe ser $Rem(x)$. Desde que $W_2^m[a,b]$ es un espacio finito dimensional, un ajuste directo de f minimizando por mínimos cuadrados (LS)

$$\frac{1}{n} \sum_{i=1}^n (y_i - f(x_i))^2 \quad (5)$$

conduce a la interpolación. Por lo tanto, es necesario cierto control sobre $Rem(x)$. Un enfoque básico es controlar hasta qué punto se puede permitir a f alejarse del modelo del polinomio.

Bajo reglas apropiadas una medida de distancia entre f y los polinomios es

$\int_a^b (f^{(m)})^2 dx$. Entonces es razonable estimar f minimizando los mínimos cuadrados (LS) bajo la restricción

$$\int_a^b (f^{(m)})^2 dx \leq \rho \quad (6)$$

para una constante ρ . Introduciendo el multiplicador Lagrange, el problema de minimización restringida (5) y (6) son equivalentes a minimizar los mínimos cuadrados penalizados (PLS):

$$\frac{1}{n} \sum_{i=1}^n (y_i - f(x_i))^2 + \lambda \int_a^b (f^{(m)})^2 dx \quad (2.7)$$

O sea el polinomio Spline se referirá a la solución del PLS (7) en el espacio modelo $W_2^m[a,b]$, un Spline cúbico es un caso especial de los polinomios Spline con $m=2$. La medida de la

aspereza de la función f , $\int_a^b (f^{(m)})^2 dx$ está referida como una penalidad de aspereza. Es obvio que no hay penalidad para polinomios de orden menor o igual que m . El parámetro de suavización λ equilibra el intercambio entre la bondad de ajuste por el LS y la medida de aspereza por $\int_a^b (f^{(m)})^2 dx$ „11.

El parámetro λ

Si hacemos $[a,b] = [0,1]$ podemos reemplazar 7 por:

$$n^{-1} \sum_{i=1}^n (y_i - f(t_i))^2 + \lambda \int_0^1 f^{(m)}(t)^2 dt, \quad \lambda > 0, \quad f \in W_2^m[0,1] \quad (8)$$

donde el parámetro λ en la ecuación regula el equilibrio entre el suavizamiento y la bondad de ajuste y, por esta razón se suele hacer referencia como el parámetro de suavizamiento. Cuando el valor de λ es grande (es decir cercana a 1) el suavizamiento es casi una recta, y los estimadores potenciales con m derivadas muy grandes son penalizados. En el caso límite de

$$\lambda = \infty \text{ (o } q = 1)$$

¹¹ Wang, Yuedong. Smoothing Splines, Methods and Applications. Monographs on Statistics and Applied Probability 121. Taylor & Francis Group, LLC 2011.

se produce un polinomio de regresión de orden m o $m-1$ grados. A la inversa, cuando el valor de λ es pequeño, le corresponde una mejor bondad de ajuste, con $\lambda = q = 0$ se tiene un estimador que interpola todos los datos.

Se quiere ajustar un conjunto de datos a una función que refleje la indispensable característica de los datos pero conservando algún grado de suavizamiento.

Una medida natural de suavizamiento asociado con una función

$$f \in W_2^m[0,1] \text{ es } \int_0^1 f^{(m)}(t)^2 dt$$

Mientras que una medida estándar de bondad de ajuste a los datos es la media residual de la suma de los $n-1$ cuadrados

$$n^{-1} \sum_{i=1}^n (y_i - f(t_i))^2$$

Así una valoración de la calidad de un candidato estimador f es proporcionado por la suma convexa

$$(1-q) n^{-1} \sum_{i=1}^n (y_i - f(t_i))^2 + q \int_0^1 f^{(m)}(t)^2 dt$$

Para algunos $0 < q < 1$, un estimador óptimo podría ser obtenido reduciendo al mínimo la función

$$W_2^m[0,1]$$

al establecer

$$\lambda = \frac{q}{(1-q)}$$

Esta se vuelve equivalente a la estimación de μ por la función $\mu\lambda$ que la minimiza.

Determinación de los coeficientes de la función Spline

El término “Spline” se refiere a un conjunto de funciones de diversa graduación, todas ellas utilizadas para la interpolación de datos o suavizamiento de curvas.

Existen varios tipos de funciones “Spline” entre ellas las lineales, cuadráticas, cúbicas, y de mayor grado. Pero los que según la experiencia demostrada en las diferentes investigaciones y aplicaciones que se realizan y para los que sirven con un alto grado de suavización son los “Spline” cúbicos o de tercer grado.

“Las funciones para la interpolación por Spline normalmente se determinan como minimizadores de la aspereza sometidas a una serie de restricciones”.

También se la identifica como una herramienta para la interpolación polinómicas por trozos. Donde dados $n+1$ puntos $(x_0;f(x_0)), (x_1;f(x_1)), \dots, (x_n;f(x_n))$ se puede lograr una curva suavizada que se adapte al conjunto de dichos pares.

La función Spline en cada subintervalo $(x_i; x_{i+1})$, es la siguiente

$$p_k(x) = c_{1k} + c_{2k}(x-x_k) + c_{3k}(x-x_k)^2 + c_{4k}(x-x_k)^3 ; k = 0; 1; \dots; n-1$$

conteniendo cuatro coeficientes $c_{1k}, c_{2k}, c_{3k}, c_{4k}$.

Estos coeficientes son despejados, de los n polinomios de grado menor o igual que tres.

Un método de interpolación cuyo proceso se ha generalizado con el soporte computacional (aunque anteriormente se realizaba en forma manual aunque tediosa), es el de la interpolación Spline cúbico. Al igual que en otros métodos de interpolación este se ajusta a un polinomio cúbico por secciones, con algunas ventajas respecto a las otras, y cuya forma es:

$$y = c_1 + c_2x + c_3x^2 + c_4x^3$$

aplicado a una sección de los datos.

Sin embargo, con los Spline cúbicos, uno logra engranar la relación de una sección del polinomio Spline cúbico con su siguiente sección, de tal forma que la pendiente del límite superior del primer polinomio debe coincidir con la pendiente del límite inferior del siguiente polinomio, además de tener una suavización que es característica de los polinomios Spline.

Esta propiedad de los polinomios Spline por secciones nos permite encontrar un sistema lineal de ecuaciones que tienen solución, lo que nos procura hallar un conjunto de coeficientes para cada una de las secciones del polinomio Spline.

Se comienza por suponer que se tiene una colección de puntos x_1, x_2, \dots, x_n ordenados y a lo largo de una curva continua. A cada uno de estos puntos se asocia algún $y_i = f(x_i)$. A raíz de las derivaciones hechas por Johnson y Percy (2000), y, Burden y Faires (1993) se divide esta secuencia continua en "i" intervalos.

En cada intervalo el objetivo es ajustar un polinomio cúbico, sea $h_i = x_{i+1} - x_i$, es decir, h_i es la diferencia entre dos puntos x_i sucesivos perteneciente a dos intervalos sucesivos. En el i-ésimo intervalo, se desea ajustar un polinomio de la forma

$$y = c_{4i}(x - x_i)^3 + c_{3i}(x - x_i)^2 + c_{2i}(x - x_i) + c_{1i}$$

donde x_i es el primer valor de x en el i-ésimo intervalo. Recordemos que, para ajustarse a un polinomio de tercer orden, el intervalo debe contener al menos cuatro puntos.

El objetivo en este punto es encontrar soluciones para $c_{1i}, c_{2i}, c_{3i}, c_{4i}$, en el i-ésimo intervalo. Se procederá a desarrollar estas soluciones, la escritura de cada coeficiente, tanto como sea posible, en términos de valores observados x_i y y_i .

En el extremo inferior del intervalo, el polinomio es simple, es sólo

$$y = c_{4i}(x_i - x_i)^3 + c_{3i}(x_i - x_i)^2 + c_{2i}(x_i - x_i) + c_{1i} \Rightarrow y = c_{1i}$$

En el extremo superior del intervalo, el polinomio es

$$y = c_{4i}(x - x_i)^3 + c_{3i}(x - x_i)^2 + c_{2i}(x - x_i) + c_{1i}, \quad \text{sea } h_i = x - x_i \Rightarrow$$

$$y = c_{4i}(h_i)^3 + c_{3i}(h_i)^2 + c_{2i}(h_i) + c_{1i}$$

Tomamos la primera y segunda derivada de este polinomio, y obtenemos

$$\frac{dy}{dx} = 3c_{4i}(h_i)^2 + 2c_{3i}h_i + c_{2i}$$

y

$$\frac{d^2y}{dx^2} = 6c_{4i}(h_i) + 2c_{3i}$$

Una vez más después de las derivaciones hechas por Jhonson y Percy, y, Burden y Faires, se escriben los coeficientes en términos de la segunda derivada en los extremos del intervalo. Así, en el extremo inferior del intervalo de orden i ,

$$S_i = \left(\frac{d^2 y}{dx^2} \right)_i = 6c_{4i}(x_i - x_i) + 2c_{3i} = 2c_{3i} \Rightarrow c_{3i} = \frac{S_i}{2}$$

y en el extremo superior del intervalo de orden i ,

$$S_{i+1} = \left(\frac{d^2 y}{dx^2} \right)_{i+1} = 6c_{4i}(x_{i+1} - x_i) + 2c_{3i} = 6c_{4i}h_i + 2c_{3i}$$

como, $S_i = 2c_{3i}$, se sustituye en la ecuación del extremo superior del intervalo la ecuación correspondiente al extremo inferior del intervalo, y se tiene:

$$S_{i+1} = 6c_{4i}h_i + S_i$$

Se resuelve para c_{4i} y se obtiene:

$$c_{4i} = \frac{S_{i+1} - S_i}{6h_i}$$

Ahora se sustituye c_{4i} , c_{3i} , c_{1i} en la ecuación del extremo superior del intervalo, y se obtiene:

$$y_{i+1} = \frac{S_{i+1} - S_i}{6h_i} (h_i)^3 + \frac{S_i}{2} h_i^2 + c_{2i} h_i + y_i$$

Por último, se despeja para c_{2i} :

$$\begin{aligned}
& y_{i+1} - \left[\frac{S_{i+1} - S_i}{6h_i} h_i^3 \right] - \left[\frac{S_i}{2} h_i^2 \right] - y_i = c_{2i} h_i \\
& \left[\frac{1}{h_i} y_{i+1} - y_i \right] - \left\{ \left[\frac{1}{h_i} \frac{S_{i+1} - S_i}{6h_i} h_i^3 \right] - \left[\frac{1}{h_i} \frac{S_i}{2} h_i^2 \right] \right\} = c_{2i} \\
& \left[\frac{y_{i+1} - y_i}{h_i} \right] - \left\{ \left[\frac{S_{i+1} - S_i}{6h_i^2} h_i^3 \right] + \left[\frac{S_i}{2h_i} h_i^2 \right] \right\} = c_{2i} \\
& \left[\frac{y_{i+1} - y_i}{h_i} \right] - \left[\frac{S_{i+1}h_i - S_i h_i}{6} + \frac{S_i}{2} h_i \right] = c_{2i} \\
& \text{se toma m. c. m.} \\
& \left[\frac{y_{i+1} - y_i}{h_i} \right] - \left[\frac{S_{i+1}h_i - S_i h_i + 3S_i h_i}{6} \right] = c_{2i} \\
& \left[\frac{y_{i+1} - y_i}{h_i} \right] - \left[\frac{S_{i+1}h_i + 2S_i h_i}{6} \right] = c_{2i} \\
& \left[\frac{y_{i+1} - y_i}{h_i} \right] - \left[\frac{2h_i S_i + h_i S_{i+1}}{6} \right] = c_{2i} \\
& c_{2i} = \frac{y_{i+1} - y_i}{h_i} - \frac{2h_i S_i + h_i S_{i+1}}{6}
\end{aligned}$$

Estas sustituciones han dado ecuaciones para c_{1i} , c_{2i} , c_{3i} , c_{4i} y, en el intervalo de orden i , en el que estas constantes se expresan en términos de valores conocidos (y_i , y_{i+1} , y h_i) y aún desconocidos como la primera derivada (S_i 's).

Para encontrar las primeras derivadas, se utiliza la condición que las pendientes de dos polinomios sucesivos son iguales en su punto común. Usando la definición de la derivada

$$\begin{aligned}
\left(\frac{dy}{dx} \right)_i &= c_{2i} \quad y \\
\left(\frac{dy}{dx} \right)_{i-1} &= 3c_{4i} h_{i-1}^2 + 2c_{3i} h_{i-1} + c_{2(i-1)}
\end{aligned}$$

Se igualan

$$c_{2i} = 3c_{4(i-1)} h_{i-1}^2 + 2c_{3(i-1)} h_{i-1} + c_{2(i-1)}$$

Ahora se sustituye para los valores c_{2i} , c_{4i} , c_{3i} y $c_{2(i-1)}$, y se resuelve la ecuación

$$h_{i-1} S_{i-1} + (2h_i + 2h_i) S_i + h_i S_{i+1} = 6 \left(\frac{y_{i+1} - y_i}{h_i} - \frac{y_i - y_{i-1}}{h_{i-1}} \right)$$

Esta relación tiene valores conocidos y_i , y_{i+1} , h_i , h_{i-1} , y valores desconocidos S_i , S_{i-1} , y S_{i+1} .

Ahora bien, esta relación contiene valores conocidos y_i , y_{i+1} , h_i , h_{i-1} , y las incógnitas de S_i , S_{i-1} y S_{i+1} . Mediante la combinación de todas las ecuaciones implícitas, para todos los intervalos i , se construye el sistema de la siguiente forma:

$$\begin{bmatrix} h_1 & 2(h_1 + h_2) & h_2 & & & & & & & \\ & h_2 & 2(h_2 + h_3) & h_3 & & & & & & \\ & & h_3 & 2(h_3 + h_4) & h_4 & & & & & \\ & & & \ddots & \ddots & \ddots & & & & \\ & & & & \ddots & \ddots & \ddots & & & \\ & & & & & h_{n-2} & 2(h_{n-2} + h_{n-1}) & & & \\ & & & & & & & h_{n-1} & & \end{bmatrix} \begin{bmatrix} S_1 \\ S_2 \\ S_3 \\ \vdots \\ \vdots \\ S_{n-1} \\ S_n \end{bmatrix} = 6 \begin{bmatrix} \frac{y_3 - y_2}{h_2} - \frac{y_2 - y_1}{h_1} \\ \frac{y_4 - y_3}{h_3} - \frac{y_3 - y_2}{h_2} \\ \frac{y_5 - y_4}{h_4} - \frac{y_4 - y_3}{h_3} \\ \vdots \\ \vdots \\ \frac{y_n - y_{n-1}}{h_{n-1}} - \frac{y_{n-1} - y_{n-2}}{h_{n-2}} \end{bmatrix}$$

El sistema lineal en la ecuación anterior contiene $n-2$ ecuaciones y n incógnitas. Dos ecuaciones se necesitan más para hacer única la solución. Si se aplican los valores finales $S_1 = S_n = 0$ (lo que implica que el polinomio es plano en los extremos inferior y superior), se puede resolver este sistema de ecuaciones para todos los S_i 's. La aplicación de estas dos condiciones de contorno elimina efectivamente dos columnas, la primera y la última, en la matriz, y se crea el sistema:

$$\begin{bmatrix} 2(h_1 + h_2) & h_2 & & & & & & & & \\ & h_2 & 2(h_2 + h_3) & h_3 & & & & & & \\ & & h_3 & 2(h_3 + h_4) & h_4 & & & & & \\ & & & \ddots & \ddots & \ddots & & & & \\ & & & & \ddots & \ddots & \ddots & & & \\ & & & & & h_{n-2} & 2(h_{n-2} + h_{n-1}) & & & \end{bmatrix} \begin{bmatrix} S_2 \\ S_3 \\ \vdots \\ \vdots \\ S_{n-1} \end{bmatrix}$$

$$= 6 \begin{bmatrix} \frac{y_3 - y_2}{h_2} - \frac{y_2 - y_1}{h_1} \\ \frac{y_4 - y_3}{h_3} - \frac{y_3 - y_2}{h_2} \\ \frac{y_5 - y_4}{h_4} - \frac{y_4 - y_3}{h_3} \\ \vdots \\ \vdots \\ \frac{y_n - y_{n-1}}{h_{n-1}} - \frac{y_{n-1} - y_{n-2}}{h_{n-2}} \end{bmatrix}$$

Esta última ecuación es el sistema solución para resolver los valores desconocidos $S_2 \dots S_{n-1}$.

MATERIALES Y MÉTODOS

Las estructuras de mortalidad por sexo y edad deben reflejar una mortalidad que sea representativa de la realidad del país y en base a la data de la que se dispone especialmente las defunciones de las estadísticas vitales, porque al no considerar con precisión las muertes, se puede modifica sustancialmente las proyecciones de población.

El Spline es muy flexible en la determinación de curvas no lineales a través de funciones cuadráticas o cúbicas que se generan para un conjunto de datos del que se disponga.

Para la aplicación del Spline se dispone de información proveniente de las estadísticas vitales de defunciones del Ministerio de Salud y los datos de población correspondientes al último Censo Nacional de Población 2007 del INEI.

Datos y Software

Los datos de Perú con los que se trabaja corresponden a una parte del mismo, concretamente al Departamento o Región de Lima, como un ejemplo de cómo se puede estructurar la mortalidad de las áreas mayores o Regiones del país.

El insumo que se necesita para la aplicación del Spline es la tasa de mortalidad por sexo y edades simples, la cual es el resultado de dividir las defunciones de cada una de las edades entre la población total de la misma edad.

Se ha tomado las bases de datos de las defunciones ocurridas en la Región Lima por sexo y edades simples (0 a 95 y más años de edad), para los años 2006 y 2008, información proporcionada por el Ministerio de Salud (MINSA) y que en el caso del departamento de Lima reúne a un buen volumen (aproximadamente el 25 por ciento) de las defunciones ocurridas en el Perú.

Asimismo, se ha tomado la información de la población de la Región Lima por sexo y edades simples (0 a 95 y más años de edad) correspondiente al Censo Nacional de Población del año 2007 publicado por el Instituto Nacional de Estadística e Informática (INEI).

Aplicación del Modelo de Regresión No Paramétrico “Spline”

Para viabilizar la aplicación del modelo Spline se ha hecho uso de la versión 10 de prueba por 30 días del software John's Macintosh Project (JMP) del grupo SAS, que es un programa que ayuda a hacer pruebas y análisis estadísticos; la cual tiene entre otras opciones las bondades de poder modelar rápidamente con ayuda del mouse, la curva que más se ajusta al conjunto de datos con los que se está trabajando, a su vez se va mostrando el coeficiente de determinación R^2 , el valor del parámetro de suavización λ , así como la suma de cuadrados residuales.

Una vez que se tiene las defunciones y la población censada, ambas por sexo y edades simples, en el caso de las defunciones se procede a calcular un promedio de las defunciones del 2006 y 2008, para centrarlas en el 2007, en el caso del Censo de Población del 2007, como este se realizó el 21 de octubre del 2007, con una tasa de crecimiento intercensal 1993-2007, se procede a retroceder la población al 30 de junio del 2007, para que de esta manera quede centrada la población correspondiente al año 2007.

A continuación se presenta la información cruda inicial correspondiente a las defunciones y a la población censada, y las primeras operaciones que se han realizado en ella.

CUADRO N° 1
REGIÓN LIMA: DEFUNCIONES 2006, 2008 Y PROMEDIO CENTRADO AL 2007

Edad	Defunciones 2006			Defunciones 2008			Defunciones promedio 2007		
	Total	Hombre	Mujer	Total	Hombre	Mujer	Total	Hombre	Mujer
Total	21551	11446	10105	28714	15012	13702	25183	13254	11929
0	951	560	391	1004	547	457	978	554	424
1	94	48	46	92	48	44	93	48	45
2	41	21	20	55	26	29	49	24	25
3	23	14	9	39	25	14	32	20	12
4	22	16	6	32	21	11	28	19	9
5	19	12	7	26	14	12	23	13	10
6	24	14	10	22	14	8	23	14	9
7	16	7	9	23	14	9	20	11	9
8	29	16	13	33	15	18	32	16	16
9	23	17	6	24	17	7	24	17	7
10	21	16	5	25	14	11	23	15	8
11	13	6	7	27	16	11	20	11	9
12	20	13	7	25	15	10	23	14	9
13	23	14	9	28	17	11	26	16	10
14	28	16	12	36	23	13	33	20	13
15	24	17	7	38	20	18	32	19	13
16	40	27	13	46	28	18	44	28	16
17	39	19	20	58	35	23	49	27	22
18	54	33	21	58	26	32	57	30	27
19	37	25	12	78	46	32	58	36	22
20	62	34	28	97	58	39	80	46	34
21	62	42	20	84	50	34	73	46	27
22	70	46	24	76	52	24	73	49	24
23	65	46	19	76	51	25	71	49	22
24	78	45	33	100	60	40	90	53	37
25	78	50	28	122	93	29	101	72	29
26	63	41	22	106	64	42	85	53	32
27	76	49	27	118	83	35	97	66	31
28	86	59	27	100	69	31	93	64	29
29	99	62	37	102	62	40	101	62	39
30	82	62	20	128	90	38	105	76	29
31	90	56	34	109	78	31	100	67	33
32	92	69	23	93	58	35	93	64	29
33	87	53	34	122	82	40	105	68	37
34	109	80	29	118	71	47	114	76	38
35	107	67	40	132	86	46	120	77	43
36	104	68	36	120	74	46	112	71	41
37	82	52	30	126	81	45	105	67	38
38	103	66	37	142	73	69	123	70	53
39	101	67	34	166	93	73	134	80	54
40	117	72	45	170	96	74	144	84	60
41	118	64	54	137	71	66	128	68	60
42	107	56	51	192	103	89	150	80	70
43	121	62	59	163	87	76	143	75	68
44	121	72	49	181	102	79	151	87	64
45	141	75	66	200	112	88	171	94	77

Continúa ...

CUADRO N° 1
REGIÓN LIMA: DEFUNCIONES 2006, 2008 Y PROMEDIO CENTRADO AL 2007

Conclusión.

Edad	Defunciones 2006			Defunciones 2008			Defunciones promedio 2007		
	Total	Hombre	Mujer	Total	Hombre	Mujer	Total	Hombre	Mujer
46	139	72	67	179	89	90	160	81	79
47	140	77	63	181	94	87	161	86	75
48	134	71	63	206	112	94	171	92	79
49	160	83	77	226	123	103	193	103	90
50	194	102	92	274	146	128	234	124	110
51	186	95	91	239	119	120	213	107	106
52	187	100	87	279	145	134	234	123	111
53	182	101	81	274	157	117	228	129	99
54	209	110	99	221	105	116	216	108	108
55	190	120	70	242	131	111	217	126	91
56	215	110	105	273	141	132	245	126	119
57	211	106	105	300	150	150	256	128	128
58	228	137	91	278	127	151	253	132	121
59	251	148	103	309	171	138	281	160	121
60	306	174	132	371	198	173	339	186	153
61	236	117	119	351	178	173	294	148	146
62	250	120	130	369	190	179	310	155	155
63	260	136	124	341	180	161	301	158	143
64	283	155	128	381	219	162	332	187	145
65	374	193	181	390	213	177	382	203	179
66	307	174	133	361	221	140	335	198	137
67	306	184	122	436	268	168	371	226	145
68	377	214	163	448	250	198	413	232	181
69	384	225	159	500	279	221	442	252	190
70	449	257	192	577	314	263	514	286	228
71	403	222	181	525	278	247	464	250	214
72	445	240	205	554	312	242	500	276	224
73	430	224	206	561	299	262	496	262	234
74	448	240	208	613	342	271	531	291	240
75	477	267	210	660	357	303	569	312	257
76	521	266	255	604	333	271	563	300	263
77	493	282	211	667	350	317	580	316	264
78	583	319	264	753	403	350	668	361	307
79	485	264	221	768	389	379	627	327	300
80	591	308	283	731	372	359	661	340	321
81	532	275	257	765	412	353	649	344	305
82	564	312	252	750	395	355	658	354	304
83	535	272	263	680	359	321	608	316	292
84	516	246	270	715	360	355	616	303	313
85	563	267	296	724	376	348	644	322	322
86	508	246	262	656	335	321	583	291	292
87	402	193	209	695	347	348	549	270	279
88	401	183	218	603	270	333	503	227	276
89	371	172	199	518	244	274	445	208	237
90	349	154	195	497	224	273	423	189	234
91	325	143	182	456	188	268	391	166	225
92	283	104	179	397	171	226	341	138	203
93	252	94	158	407	142	265	330	118	212
94	212	85	127	277	93	184	245	89	156
95y+	742	261	481	1083	361	722	915	312	603

CUADRO N° 2

REGIÓN LIMA: POBLACIÓN CENSADA EL 21 DE OCTUBRE DEL 2007 Y POBLACIÓN CENSADA RETROCEDIDA AL 30 DE JUNIO DEL 2007

Edad	Población censada el 21 de octubre del 2007			Población censada retrocedida al 30 de junio del 2007			Edad	Población censada el 21 de octubre del 2007			Población censada retrocedida al 30 de junio del 2007		
	Total	Hombre	Mujer	Total	Hombre	Mujer		Total	Hombre	Mujer	Total	Hombre	Mujer
Total	8445211	4139686	4305525	8394200	4114683	4279517	46	88035	42208	45827	87503	41953	45550
0	132652	67663	64989	131850	67254	64596	47	101166	48862	52304	100555	48567	51988
1	137083	70271	66812	136255	69847	66408	48	89718	42608	47110	89176	42351	46825
2	151267	76778	74489	150353	76314	74039	49	80595	38390	42205	80108	38158	41950
3	150196	76635	73561	149289	76172	73117	50	91904	42708	49196	91349	42450	48899
4	142609	73147	69462	141747	72705	69042	51	67598	31891	35707	67189	31698	35491
5	135737	69455	66282	134917	69035	65882	52	84974	40595	44379	84461	40350	44111
6	133386	68022	65364	132580	67611	64969	53	79509	37523	41986	79028	37296	41732
7	139824	71624	68200	138979	71191	67788	54	75243	35369	39874	74788	35155	39633
8	141106	72079	69027	140254	71644	68610	55	67186	32016	35170	66781	31823	34958
9	135291	69230	66061	134474	68812	65662	56	63342	30622	32720	62959	30437	32522
10	148102	75333	72769	147207	74878	72329	57	67119	32094	35025	66713	31900	34813
11	146706	74551	72155	145820	74101	71719	58	57907	27768	30139	57557	27600	29957
12	152154	77192	74962	151235	76726	74509	59	52550	25618	26932	52232	25463	26769
13	147214	74318	72896	146325	73869	72456	60	63542	30101	33441	63158	29919	33239
14	152495	76430	76065	151574	75968	75606	61	40494	20083	20411	40250	19962	20288
15	156445	77060	79385	155500	76595	78905	62	50576	24629	25947	50270	24480	25790
16	145761	71290	74471	144880	70859	74021	63	47192	22792	24400	46907	22654	24253
17	154252	76437	77815	153320	75975	77345	64	43293	21145	22148	43031	21017	22014
18	169019	84247	84772	167998	83738	84260	65	48687	23505	25182	48393	23363	25030
19	173687	85141	88546	172640	84628	88012	66	34887	17007	17880	34676	16904	17772
20	178023	87446	90577	176950	86919	90031	67	41664	20020	21644	41412	19899	21513
21	155328	76143	79185	154390	75683	78707	68	34593	16283	18310	34384	16185	18199
22	167932	83445	84487	166918	82941	83977	69	30258	15228	15030	30075	15136	14939
23	166517	81450	85067	165511	80958	84553	70	38095	17936	20159	37865	17828	20037
24	169238	83295	85943	168216	82792	85424	71	24418	12196	12222	24270	12122	12148
25	167682	82049	85633	166669	81553	85116	72	31330	14881	16449	31141	14791	16350
26	153880	74731	79149	152951	74280	78671	73	27554	13260	14294	27388	13180	14208
27	163787	81015	82772	162798	80526	82272	74	26812	12433	14379	26650	12358	14292
28	152011	74553	77458	151093	74103	76990	75	28671	13336	15335	28497	13255	15242
29	144113	69639	74474	143242	69218	74024	76	22133	10655	11478	22000	10591	11409
30	166588	81102	85486	165582	80612	84970	77	23270	11045	12225	23129	10978	12151
31	133281	64399	68882	132476	64010	68466	78	21811	10177	11634	21680	10116	11564
32	147552	71523	76029	146661	71091	75570	79	17145	8230	8915	17041	8180	8861
33	143455	70647	72808	142588	70220	72368	80	20338	8965	11373	20215	8911	11304
34	130233	63408	66825	129446	63025	66421	81	12316	5801	6515	12242	5766	6476
35	131965	63584	68381	131168	63200	67968	82	14105	6374	7731	14019	6335	7684
36	122011	58324	63687	121274	57972	63302	83	11640	5218	6422	11569	5186	6383
37	133981	64377	69604	133172	63988	69184	84	11062	4894	6168	10995	4864	6131
38	123238	58862	64376	122493	58506	63987	85	10774	4513	6261	10709	4486	6223
39	115609	55660	59949	114911	55324	59587	86	8731	3824	4907	8678	3801	4877
40	128881	61631	67250	128103	61259	66844	87	8471	3513	4958	8420	3492	4928
41	96264	46500	49764	95682	46219	49463	88	5643	2438	3205	5609	2423	3186
42	121072	58795	62277	120341	58440	61901	89	4941	2062	2879	4912	2050	2862
43	103858	49712	54146	103231	49412	53819	90	4619	1734	2885	4592	1724	2868
44	96374	46877	49497	95792	46594	49198	91	2476	1033	1443	2461	1027	1434
45	95597	45217	50380	95020	44944	50076	92	2812	1061	1751	2795	1055	1740
							93	2344	849	1495	2330	844	1486
							94	1887	708	1179	1876	704	1172
							95+y+	6325	2198	4127	6287	2185	4102

Tanto las defunciones como la población censada adolecen de una omisión, sea de registro y/o envío a las oficinas del Ministerio de Salud en todo el país en el caso de las defunciones, sea de no empadronamiento de las personas en el caso del Censo Nacional de Población.

El demógrafo matemático William Brass creó una metodología para poder calcular el volumen de las defunciones omitidas, para la que usó la estructura de la población censada por sexo y edad. El porcentaje de omisión obtenido de la aplicación de esta metodología, sirve para que se complete las defunciones faltantes.

Una vez obtenido el porcentaje de omisión en las defunciones, se procede a aplicarse a las defunciones registradas por edad y sexo, para obtener el total de las defunciones, las cuáles serán el numerador para el cálculo de las tasas de mortalidad por sexo y edad.

Con respecto a la población censada, paralelamente al Censo o dentro de los días siguientes al mismo se realiza una encuesta post-censal cuya finalidad es calcular la población y viviendas, que no han sido empadronadas o registradas el día del Censo. Esta información conduce a la

obtención de un porcentaje de omisión de la población que no fue censada. Dicho porcentaje asume que no hay diferenciales respecto al sexo y edad, con dicho porcentaje se calcula la población omitida el día del Censo, que para el caso de la Región Lima fue 1.4%. Finalmente la data estará lista para proceder al cálculo de la tasa de mortalidad por sexo y para cada una de las edades comprendidas entre 0 y 95 años, utilizando la siguiente fórmula:

$$\text{Tasa de mortalidad de la edad } x_i = \frac{\text{Número de defunciones, de la edad } x}{\text{Población total a mitad de año, de la edad } x}$$

donde $x = 0, 1, 2, \dots, 95$

Se procede a calcular para luego importar los logaritmos de las tasas de mortalidad por sexo y edad al software JMP versión 10, luego de ingresar al módulo: Spline cúbico, se obtiene una gráfica bivariada de la Edad y Tasa de Mortalidad por Edad. También se muestra su correspondiente λ , R^2 y la suma de errores al cuadrado.

El valor de λ es mayor que cero y puede variar de acuerdo a la amplitud entre el mínimo y máximo valor en estudio. Cuando el valor de λ se aproxima a cero el ajuste se hace más flexible y curvado, cuando λ aumenta el ajuste se hace más rígido (menos curvado) aproximándose a una línea recta, cuando llega al límite superior cercano al valor máximo del estudio.

El software permite determinar el mejor ajuste para la curva del diagrama de dispersión, moviendo el mouse en la barra deslizadora que se encuentra debajo de la gráfica y del valor de lambda (λ). Se pueda observar a simple vista la mejor gráfica para los valores esperados de Y en X .

El informe que se obtiene del proceso, correspondiente a cada suavizamiento del Spline, contiene no sólo los coeficientes correspondientes al polinomio de cada sección del Spline, sino también el R^2 y la suma de cuadrados residuales de todo el Spline. Se pueden realizar varias pruebas hasta lograr la suavización más conveniente de la curva.

Luego se hacen las comparaciones tanto del R^2 como de la suma de cuadrados residuales, para decidir por la curva con el mejor ajuste Spline.

Los resultados de la aplicación del modelo Spline a los logaritmos de las tasas de mortalidad por sexo y edad se presentan en el siguiente cuadro, seguido de sus gráficas correspondientes, finalmente se presentan los cuadros de los antilogaritmos de las tasas trabajadas, para tener las tasas de mortalidad por sexo y edad ajustadas con el Spline.

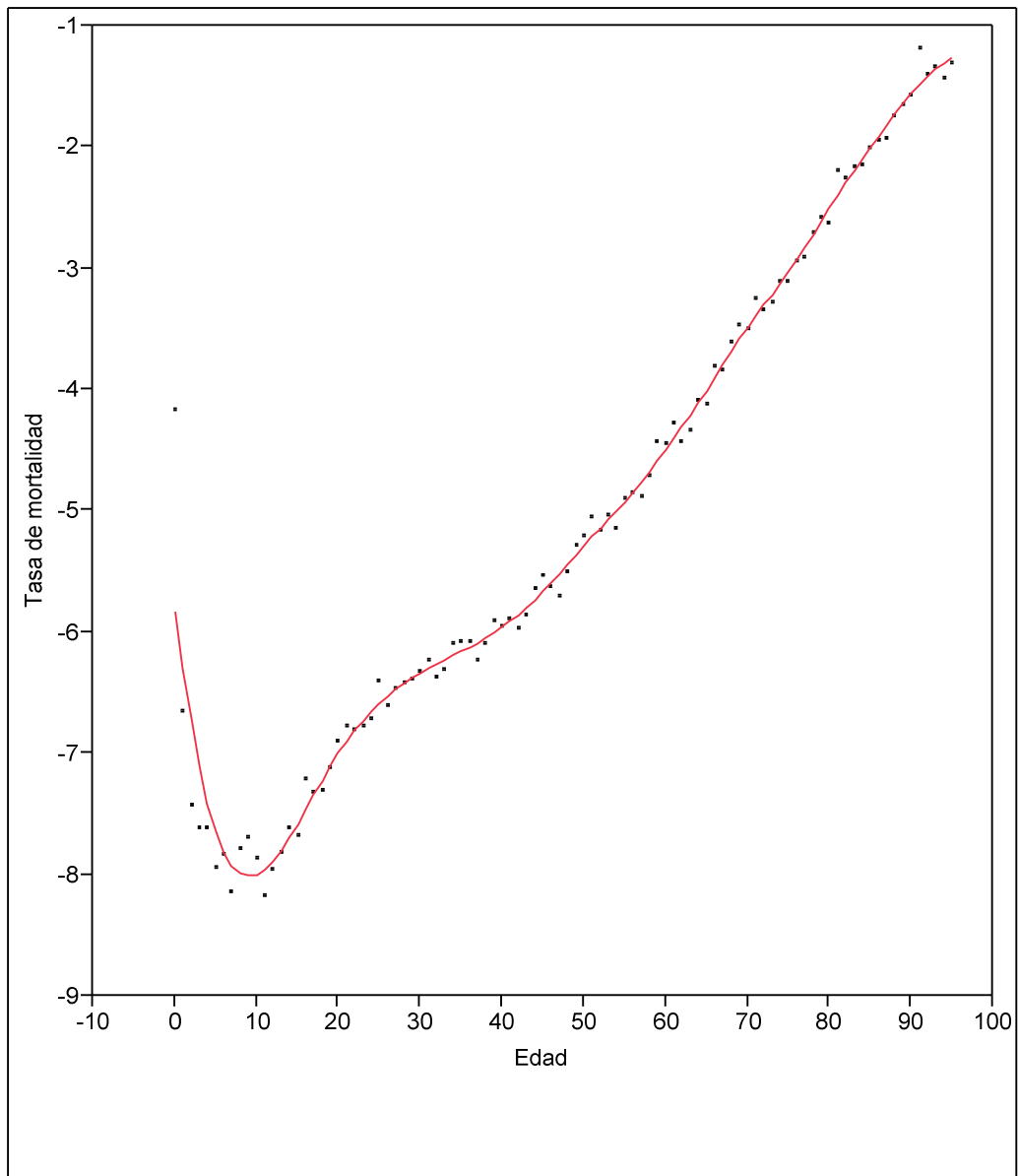
CUADRO Nº 3
REGIÓN LIMA: RESULTADOS DE APLICAR EL MODELO SPLINE, LN DE TASA DE MORTALIDAD, RESIDUALES Y PREDICTORES SPLINE POR EDAD,
HOMBRES 2007

Edad	Ln de tasa de mortalidad	Predictor	Residual	Predictor Spline	Edad	Ln de tasa de mortalidad	Predictor	Residual	Predictor Spline
0	-4,1690	-5,8233	1,6543	-5,8233	46	-5,6215	-5,5943	-0,0271	-5,5943
1	-6,6573	-6,2903	-0,3670	-6,2903	47	-5,7049	-5,5216	-0,1833	-5,5216
2	-7,4281	-6,7255	-0,7026	-6,7255	48	-5,5031	-5,4448	-0,0582	-5,4448
3	-7,6173	-7,1041	-0,5131	-7,1041	49	-5,2855	-5,3660	0,0805	-5,3660
4	-7,6247	-7,4140	-0,2108	-7,4140	50	-5,2064	-5,2884	0,0820	-5,2884
5	-7,9376	-7,6525	-0,2851	-7,6525	51	-5,0600	-5,2139	0,1540	-5,2139
6	-7,8398	-7,8229	-0,0169	-7,8229	52	-5,1641	-5,1428	-0,0213	-5,1428
7	-8,1427	-7,9326	-0,2101	-7,9326	53	-5,0354	-5,0729	0,0375	-5,0729
8	-7,7924	-7,9911	0,1988	-7,9911	54	-5,1537	-5,0018	-0,1519	-5,0018
9	-7,6875	-8,0102	0,3227	-8,0102	55	-4,9014	-4,9270	0,0256	-4,9270
10	-7,8704	-7,9983	0,1279	-7,9983	56	-4,8569	-4,8480	-0,0089	-4,8480
11	-8,1828	-7,9589	-0,2239	-7,9589	57	-4,8873	-4,7644	-0,1229	-4,7644
12	-7,9663	-7,8930	-0,0733	-7,8930	58	-4,7142	-4,6762	-0,0380	-4,6762
13	-7,8230	-7,8047	-0,0182	-7,8047	59	-4,4388	-4,5854	0,1466	-4,5854
14	-7,6146	-7,6999	0,0853	-7,6999	60	-4,4511	-4,4942	0,0431	-4,4942
15	-7,6769	-7,5844	-0,0925	-7,5844	61	-4,2738	-4,4029	0,1291	-4,4029
16	-7,2123	-7,4635	0,2513	-7,4635	62	-4,4328	-4,3103	-0,1225	-4,3103
17	-7,3204	-7,3427	0,0223	-7,3427	63	-4,3351	-4,2138	-0,1213	-4,2138
18	-7,3065	-7,2240	-0,0824	-7,2240	64	-4,0922	-4,1123	0,0201	-4,1123
19	-7,1260	-7,1090	-0,0170	-7,1090	65	-4,1146	-4,0069	-0,1077	-4,0069
20	-6,9095	-6,9999	0,0904	-6,9999	66	-3,8172	-3,8988	0,0816	-3,8988
21	-6,7711	-6,8996	0,1285	-6,8996	67	-3,8488	-3,7907	-0,0581	-3,7907
22	-6,8074	-6,8090	0,0016	-6,8090	68	-3,6146	-3,6844	0,0698	-3,6844
23	-6,7832	-6,7271	-0,0560	-6,7271	69	-3,4652	-3,5823	0,1172	-3,5823
24	-6,7231	-6,6529	-0,0701	-6,6529	70	-3,5018	-3,4855	-0,0163	-3,4855
25	-6,4031	-6,5861	0,1830	-6,5861	71	-3,2514	-3,3934	0,1419	-3,3934
26	-6,6146	-6,5269	-0,0877	-6,5269	72	-3,3506	-3,3046	-0,0460	-3,3046
27	-6,4741	-6,4737	-0,0005	-6,4737	73	-3,2879	-3,2163	-0,0716	-3,2163
28	-6,4233	-6,4254	0,0021	-6,4254	74	-3,1190	-3,1259	0,0070	-3,1259
29	-6,3884	-6,3813	-0,0071	-6,3813	75	-3,1193	-3,0319	-0,0874	-3,0319
30	-6,3348	-6,3406	0,0058	-6,3406	76	-2,9344	-2,9333	-0,0011	-2,9333
31	-6,2289	-6,3025	0,0737	-6,3025	77	-2,9175	-2,8301	-0,0874	-2,8301
32	-6,3818	-6,2661	-0,1157	-6,2661	78	-2,7022	-2,7231	0,0209	-2,7231
33	-6,3137	-6,2297	-0,0840	-6,2297	79	-2,5890	-2,6140	0,0251	-2,6140
34	-6,0887	-6,1932	0,1045	-6,1932	80	-2,6368	-2,5047	-0,1322	-2,5047
35	-6,0777	-6,1578	0,0800	-6,1578	81	-2,1892	-2,3967	0,2075	-2,3967
36	-6,0765	-6,1231	0,0466	-6,1231	82	-2,2532	-2,2930	0,0397	-2,2930
37	-6,2285	-6,0874	-0,1411	-6,0874	83	-2,1676	-2,1937	0,0261	-2,1937
38	-6,1006	-6,0484	-0,0523	-6,0484	84	-2,1459	-2,0979	-0,0480	-2,0979
39	-5,9112	-6,0056	0,0944	-6,0056	85	-2,0045	-2,0042	-0,0003	-2,0042
40	-5,9618	-5,9597	-0,0021	-5,9597	86	-1,9400	-1,9119	-0,0281	-1,9119
41	-5,8954	-5,9099	0,0145	-5,9099	87	-1,9302	-1,8204	-0,1098	-1,8204
42	-5,9660	-5,8554	-0,1106	-5,8554	88	-1,7383	-1,7299	-0,0083	-1,7299
43	-5,8592	-5,7955	-0,0637	-5,7955	89	-1,6582	-1,6423	-0,0159	-1,6423
44	-5,6513	-5,7310	0,0796	-5,7310	90	-1,5801	-1,5599	-0,0202	-1,5599
45	-5,5399	-5,6637	0,1238	-5,6637	91	-1,1932	-1,4853	0,2921	-1,4853
					92	-1,4033	-1,4205	0,0172	-1,4205
					93	-1,3362	-1,3636	0,0274	-1,3636
					94	-1,4410	-1,3115	-0,1295	-1,3115
					95y+	-1,3166	-1,2609	-0,0557	-1,2609

Cuadro Nº 4
REGIÓN LIMA: RESULTADOS DE APLICAR EL MODELO SPLINE, LN DE TASA DE MORTALIDAD, RESIDUALES Y PREDICTORES SPLINE POR EDAD,
MUJERES 2007

Edad	Ln de tasa de mortalidad	Predictor	Residual	Predictor Spline	Edad	Ln de tasa de mortalidad	Predictor	Residual	Predictor Spline
0	-4,3261	-5,8906	1,5644	-5,8906	46	-5,6593	-5,7449	0,0857	-5,7449
1	-6,5959	-6,4005	-0,1953	-6,4005	47	-5,8424	-5,6669	-0,1755	-5,6669
2	-7,2946	-6,8799	-0,4147	-6,8799	48	-5,6869	-5,5882	-0,0987	-5,5882
3	-7,9950	-7,3019	-0,6931	-7,3019	49	-5,4434	-5,5094	0,0660	-5,5094
4	-8,2662	-7,6485	-0,6176	-7,6485	50	-5,3955	-5,4326	0,0371	-5,4326
5	-8,1140	-7,9142	-0,1998	-7,9142	51	-5,1159	-5,3594	0,2435	-5,3594
6	-8,2054	-8,1047	-0,1007	-8,1047	52	-5,2836	-5,2898	0,0062	-5,2898
7	-8,2479	-8,2308	-0,0171	-8,2308	53	-5,3449	-5,2203	-0,1246	-5,2203
8	-7,6538	-8,3054	0,6516	-8,3054	54	-5,2034	-5,1471	-0,0563	-5,1471
9	-8,4673	-8,3397	-0,1277	-8,3397	55	-5,2502	-5,0681	-0,1821	-5,0681
10	-8,4305	-8,3367	-0,0938	-8,3367	56	-4,9107	-4,9831	0,0723	-4,9831
11	-8,3042	-8,2993	-0,0049	-8,2993	57	-4,9073	-4,8940	-0,0133	-4,8940
12	-8,3424	-8,2322	-0,1102	-8,2322	58	-4,8122	-4,8028	-0,0094	-4,8028
13	-8,2091	-8,1406	-0,0685	-8,1406	59	-4,6997	-4,7111	0,0114	-4,7111
14	-7,9516	-8,0316	0,0801	-8,0316	60	-4,6826	-4,6209	-0,0617	-4,6209
15	-7,9943	-7,9132	-0,0810	-7,9132	61	-4,2348	-4,5341	0,2993	-4,5341
16	-7,7297	-7,7930	0,0633	-7,7930	62	-4,4129	-4,4524	0,0394	-4,4524
17	-7,4635	-7,6789	0,2154	-7,6789	63	-4,4336	-4,3736	-0,0600	-4,3736
18	-7,3484	-7,5778	0,2294	-7,5778	64	-4,3232	-4,2944	-0,0288	-4,2944
19	-7,5927	-7,4926	-0,1000	-7,4926	65	-4,2393	-4,2119	-0,0274	-4,2119
20	-7,1879	-7,4229	0,2350	-7,4229	66	-4,1647	-4,1241	-0,0406	-4,1241
21	-7,2803	-7,3678	0,0876	-7,3678	67	-4,3001	-4,0295	-0,2706	-4,0295
22	-7,4606	-7,3237	-0,1368	-7,3237	68	-3,9097	-3,9283	0,0186	-3,9283
23	-7,5526	-7,2852	-0,2674	-7,2852	69	-3,6648	-3,8242	0,1594	-3,8242
24	-7,0387	-7,2494	0,2107	-7,2494	70	-3,7753	-3,7210	-0,0542	-3,7210
25	-7,2883	-7,2168	-0,0715	-7,2168	71	-3,3390	-3,6208	0,2817	-3,6208
26	-7,1127	-7,1861	0,0734	-7,1861	72	-3,5914	-3,5245	-0,0669	-3,5245
27	-7,1887	-7,1561	-0,0326	-7,1561	73	-3,4061	-3,4300	0,0240	-3,4300
28	-7,1880	-7,1250	-0,0630	-7,1250	74	-3,3872	-3,3350	-0,0522	-3,3350
29	-6,8442	-7,0912	0,2470	-7,0912	75	-3,3825	-3,2371	-0,1454	-3,2371
30	-7,2866	-7,0535	-0,2331	-7,0535	76	-3,0703	-3,1352	0,0649	-3,1352
31	-6,9435	-7,0082	0,0647	-7,0082	77	-3,1296	-3,0300	-0,0995	-3,0300
32	-7,1694	-6,9538	-0,2156	-6,9538	78	-2,9288	-2,9223	-0,0065	-2,9223
33	-6,8729	-6,8892	0,0163	-6,8892	79	-2,6852	-2,8137	0,1286	-2,8137
34	-6,7612	-6,8161	0,0550	-6,8161	80	-2,8609	-2,7062	-0,1547	-2,7062
35	-6,6636	-6,7366	0,0731	-6,7366	81	-2,3554	-2,6003	0,2449	-2,6003
36	-6,6390	-6,6516	0,0127	-6,6516	82	-2,5297	-2,4975	-0,0322	-2,4975
37	-6,8019	-6,5609	-0,2410	-6,5609	83	-2,3853	-2,3966	0,0112	-2,3966
38	-6,3984	-6,4645	0,0661	-6,4645	84	-2,2769	-2,2959	0,0190	-2,2959
39	-6,3088	-6,3656	0,0568	-6,3656	85	-2,2624	-2,1937	-0,0687	-2,1937
40	-6,3120	-6,2670	-0,0451	-6,2670	86	-2,1161	-2,0883	-0,0278	-2,0883
41	-6,0109	-6,1707	0,1598	-6,1707	87	-2,1712	-1,9787	-0,1924	-1,9787
42	-6,0845	-6,0785	-0,0060	-6,0785	88	-1,7455	-1,8655	0,1200	-1,8655
43	-5,9730	-5,9905	0,0175	-5,9905	89	-1,7914	-1,7511	-0,0403	-1,7511
44	-5,9425	-5,9060	-0,0366	-5,9060	90	-1,8060	-1,6375	-0,1684	-1,6375
45	-5,7792	-5,8242	0,0451	-5,8242	91	-1,1530	-1,5272	0,3742	-1,5272
					92	-1,4476	-1,4238	-0,0238	-1,4238
					93	-1,2471	-1,3267	0,0795	-1,3267
					94	-1,3157	-1,2339	-0,0818	-1,2339
					95+	-1,2177	-1,1429	-0,0748	-1,1429

GRÁFICO Nº 1
REGIÓN LIMA: TASAS DE MORTALIDAD MASCULINAS
SUAVIZADAS CON SPLINE, 2007

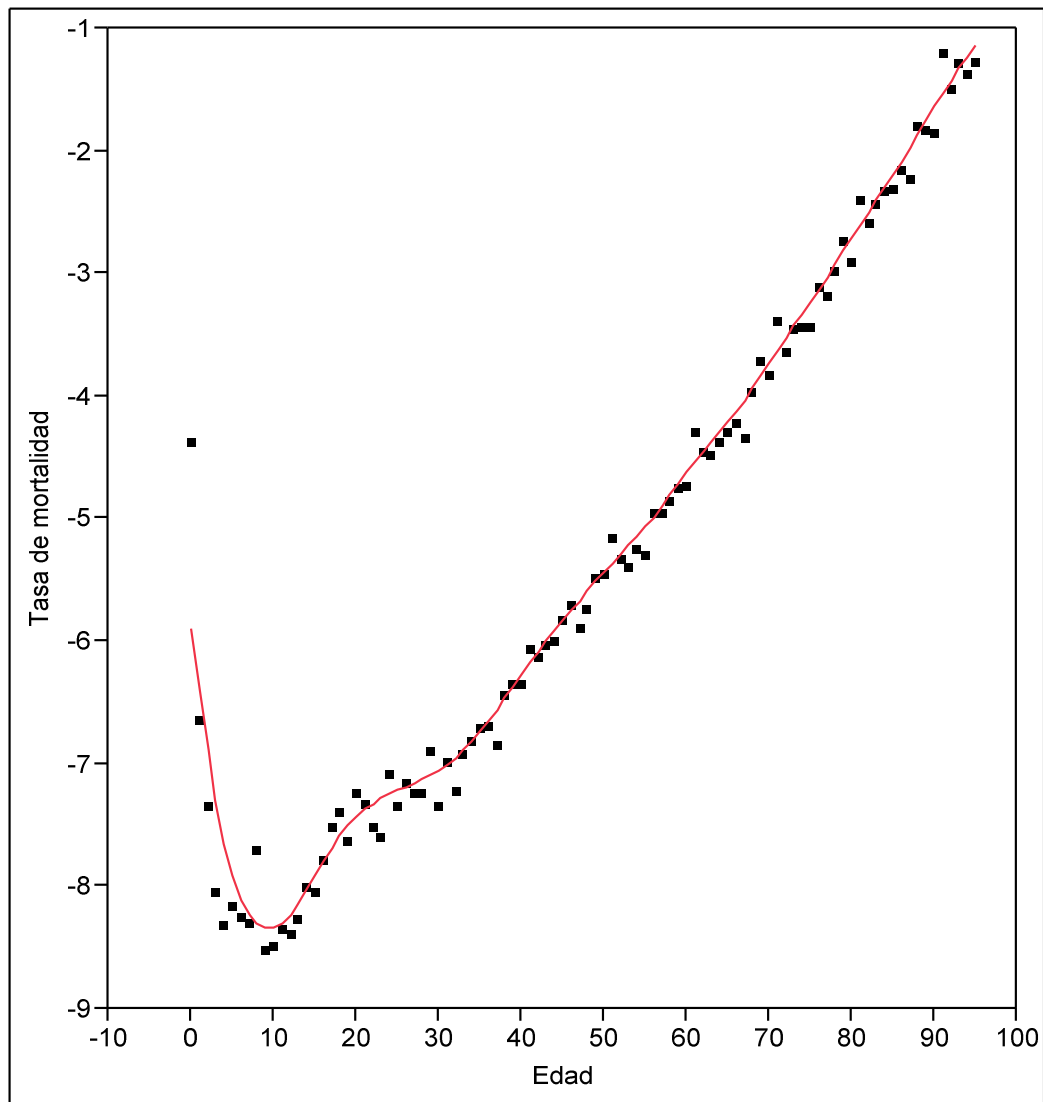


— Smoothing Spline Fit, lambda=100

Smoothing Spline Fit, lambda=100

R-Square 0,987954
Sum of Squares Error 4,746199

GRÁFICO Nº 2
REGIÓN LIMA: TASAS DE MORTALIDAD FEMENINAS
SUAVIZADAS CON SPLINE, 2007



— Smoothing Spline Fit, lambda=100

Smoothing Spline Fit, lambda=100

R-Square 0,987526
Sum of Squares Error 5,491795

Cuadro N° 5
REGIÓN LIMA: TASAS DE MORTALIDAD SUAVIZADAS CON SPLINE POR SEXO Y EDAD,
2007

Edad	Hombre	Mujer	Edad	Hombre	Mujer
0	0,0030	0,0028	46	0,0037	0,0032
1	0,0019	0,0017	47	0,0040	0,0035
2	0,0012	0,0010	48	0,0043	0,0037
3	0,0008	0,0007	49	0,0047	0,0040
4	0,0006	0,0005	50	0,0050	0,0044
5	0,0005	0,0004	51	0,0054	0,0047
6	0,0004	0,0003	52	0,0058	0,0050
7	0,0004	0,0003	53	0,0063	0,0054
8	0,0003	0,0002	54	0,0067	0,0058
9	0,0003	0,0002	55	0,0072	0,0063
10	0,0003	0,0002	56	0,0078	0,0069
11	0,0003	0,0002	57	0,0085	0,0075
12	0,0004	0,0003	58	0,0093	0,0082
13	0,0004	0,0003	59	0,0102	0,0090
14	0,0005	0,0003	60	0,0112	0,0098
15	0,0005	0,0004	61	0,0122	0,0107
16	0,0006	0,0004	62	0,0134	0,0117
17	0,0006	0,0005	63	0,0148	0,0126
18	0,0007	0,0005	64	0,0164	0,0136
19	0,0008	0,0006	65	0,0182	0,0148
20	0,0009	0,0006	66	0,0203	0,0162
21	0,0010	0,0006	67	0,0226	0,0178
22	0,0011	0,0007	68	0,0251	0,0197
23	0,0012	0,0007	69	0,0278	0,0218
24	0,0013	0,0007	70	0,0306	0,0242
25	0,0014	0,0007	71	0,0336	0,0268
26	0,0015	0,0008	72	0,0367	0,0295
27	0,0015	0,0008	73	0,0401	0,0324
28	0,0016	0,0008	74	0,0439	0,0356
29	0,0017	0,0008	75	0,0482	0,0393
30	0,0018	0,0009	76	0,0532	0,0435
31	0,0018	0,0009	77	0,0590	0,0483
32	0,0019	0,0010	78	0,0657	0,0538
33	0,0020	0,0010	79	0,0732	0,0600
34	0,0020	0,0011	80	0,0817	0,0668
35	0,0021	0,0012	81	0,0910	0,0743
36	0,0022	0,0013	82	0,1010	0,0823
37	0,0023	0,0014	83	0,1115	0,0910
38	0,0024	0,0016	84	0,1227	0,1007
39	0,0025	0,0017	85	0,1348	0,1115
40	0,0026	0,0019	86	0,1478	0,1239
41	0,0027	0,0021	87	0,1620	0,1382
42	0,0029	0,0023	88	0,1773	0,1548
43	0,0030	0,0025	89	0,1935	0,1736
44	0,0032	0,0027	90	0,2102	0,1945
45	0,0035	0,0030	91	0,2264	0,2171
			92	0,2416	0,2408
			93	0,2557	0,2654
			94	0,2694	0,2912
			95y+	0,2834	0,3189

Como puede apreciarse en las gráficas correspondientes a hombres y mujeres, las tasas de mortalidad por sexo y edad suavizadas con el Spline (curva modelo), muestran que el ajuste es bastante bueno, para el conjunto inicial de datos, con un R cuadrado muy próximo a 1 y suma de cuadrados de los errores 4.75 en los hombres y 5.49 en las mujeres.

Se puede asegurar que los modelos obtenidos por sexo se ajustan a su correspondiente diagrama de dispersión.

Aplicación de medias móviles y “método de la mano alzada”

Una forma de obtener la estructura de la mortalidad por sexo y edad era combinando medias móviles con el método demográfico de la mano alzada, método que era más un procedimiento práctico que un método sustentado teóricamente, con el cual se “lograba” suavizar las curvas correspondientes a la estructura de la mortalidad por sexo y edad.

En el Cuadro N° 6 se presentan los cálculos correspondientes a la aplicación de medias móviles de 5 pivotes, y los ajustes obtenidos por la aplicación de la mano alzada.

Cuadro Nº 6

DEPARTAMENTO DE LIMA: TASA DE MORTALIDAD SIN SUAVIZAR, SUAVIZADA MEDIA MOVIL 5 Y SUAVIZADO CON MANO ALZADA, POR SEXO Y EDAD, 2007

Edad	Sin suavizar		Suavizada media móvil 5		Suavizada MM5 con mano alzada		Edad	Sin suavizar		Suavizada media móvil 5		Suavizada MM5 con mano alzada	
	Hombre	Mujer	Hombre	Mujer	Hombre	Mujer		Hombre	Mujer	Hombre	Mujer	Hombre	Mujer
0	0,0155	0,0132	0,0155	0,0132	0,01547	0,01322	46	0,0036	0,0035	0,0037	0,0031	0,00369	0,00310
1	0,0013	0,0014	0,0013	0,0014	0,00650	0,00650	47	0,0033	0,0029	0,0040	0,0034	0,00400	0,00344
2	0,0006	0,0007	0,0037	0,0032	0,00250	0,00317	48	0,0041	0,0034	0,0043	0,0037	0,00430	0,00373
3	0,0005	0,0003	0,0006	0,0006	0,00064	0,00059	49	0,0051	0,0043	0,0049	0,0042	0,00470	0,00400
4	0,0005	0,0003	0,0005	0,0004	0,00047	0,00041	50	0,0055	0,0045	0,0053	0,0047	0,00510	0,00430
5	0,0004	0,0003	0,0004	0,0003	0,00040	0,00034	51	0,0063	0,0060	0,0058	0,0049	0,00550	0,00470
6	0,0004	0,0003	0,0004	0,0003	0,00039	0,00031	52	0,0057	0,0051	0,0060	0,0052	0,00580	0,00500
7	0,0003	0,0003	0,0004	0,0003	0,00038	0,00029	53	0,0065	0,0048	0,0064	0,0053	0,00630	0,00550
8	0,0004	0,0005	0,0004	0,0003	0,00038	0,00028	54	0,0058	0,0055	0,0066	0,0056	0,00670	0,00590
9	0,0005	0,0002	0,0004	0,0003	0,00037	0,00027	55	0,0074	0,0052	0,0070	0,0061	0,00750	0,00640
10	0,0004	0,0002	0,0004	0,0003	0,00037	0,00026	56	0,0078	0,0074	0,0075	0,0067	0,00820	0,00690
11	0,0003	0,0002	0,0004	0,0002	0,00037	0,00027	57	0,0075	0,0074	0,0087	0,0074	0,00880	0,00745
12	0,0003	0,0002	0,0004	0,0003	0,00038	0,00027	58	0,0090	0,0081	0,0096	0,0082	0,00970	0,00825
13	0,0004	0,0003	0,0004	0,0003	0,00043	0,00029	59	0,0118	0,0091	0,0108	0,0097	0,01078	0,00930
14	0,0005	0,0004	0,0005	0,0003	0,00049	0,00033	60	0,0117	0,0093	0,0117	0,0106	0,01170	0,00980
15	0,0005	0,0003	0,0006	0,0004	0,00055	0,00039	61	0,0139	0,0145	0,0125	0,0114	0,01300	0,01100
16	0,0007	0,0004	0,0006	0,0005	0,00061	0,00045	62	0,0119	0,0121	0,0135	0,0122	0,01400	0,01220
17	0,0007	0,0006	0,0007	0,0005	0,00070	0,00050	63	0,0131	0,0119	0,0144	0,0132	0,01600	0,01300
18	0,0007	0,0006	0,0008	0,0006	0,00077	0,00056	64	0,0167	0,0133	0,0160	0,0134	0,01750	0,01400
19	0,0008	0,0005	0,0009	0,0006	0,00086	0,00059	65	0,0163	0,0144	0,0179	0,0137	0,02000	0,01500
20	0,0010	0,0008	0,0009	0,0006	0,00095	0,00062	66	0,0220	0,0155	0,0207	0,0154	0,02200	0,01650
21	0,0011	0,0007	0,0010	0,0006	0,00104	0,00065	67	0,0213	0,0136	0,0236	0,0178	0,02400	0,01750
22	0,0011	0,0006	0,0011	0,0007	0,00112	0,00068	68	0,0269	0,0200	0,0263	0,0195	0,02700	0,01900
23	0,0011	0,0005	0,0012	0,0007	0,00120	0,00070	69	0,0313	0,0256	0,0297	0,0235	0,02900	0,02100
24	0,0012	0,0009	0,0013	0,0007	0,00129	0,00073	70	0,0301	0,0229	0,0324	0,0263	0,03242	0,02300
25	0,0017	0,0007	0,0014	0,0007	0,00138	0,00076	71	0,0387	0,0355	0,0345	0,0289	0,03650	0,02600
26	0,0013	0,0008	0,0015	0,0008	0,00142	0,00078	72	0,0351	0,0276	0,0371	0,0306	0,03850	0,02800
27	0,0015	0,0008	0,0016	0,0008	0,00150	0,00080	73	0,0373	0,0332	0,0399	0,0328	0,04300	0,03100
28	0,0016	0,0008	0,0016	0,0008	0,00159	0,00081	74	0,0442	0,0338	0,0428	0,0350	0,04550	0,03498
29	0,0017	0,0011	0,0017	0,0008	0,00165	0,00083	75	0,0442	0,0340	0,0466	0,0382	0,05000	0,03822
30	0,0018	0,0007	0,0017	0,0008	0,00175	0,00087	76	0,0532	0,0464	0,0525	0,0423	0,05600	0,04227
31	0,0020	0,0010	0,0018	0,0009	0,00179	0,00090	77	0,0541	0,0437	0,0587	0,0492	0,06000	0,04916
32	0,0017	0,0008	0,0019	0,0009	0,00190	0,00096	78	0,0671	0,0535	0,0642	0,0538	0,06800	0,05381
33	0,0018	0,0010	0,0020	0,0010	0,00198	0,00104	79	0,0751	0,0682	0,0760	0,0635	0,07596	0,06200
34	0,0023	0,0012	0,0021	0,0011	0,00202	0,00111	80	0,0716	0,0572	0,0862	0,0707	0,08616	0,06900
35	0,0023	0,0013	0,0021	0,0012	0,00213	0,00118	81	0,1120	0,0949	0,0956	0,0784	0,09564	0,07500
36	0,0023	0,0013	0,0022	0,0013	0,00218	0,00130	82	0,1051	0,0797	0,1040	0,0853	0,10800	0,08400
37	0,0020	0,0011	0,0023	0,0014	0,00230	0,00144	83	0,1144	0,0921	0,1166	0,0947	0,11400	0,09466
38	0,0022	0,0017	0,0024	0,0015	0,00236	0,00160	84	0,1170	0,1026	0,1230	0,0998	0,12300	0,10510
39	0,0027	0,0018	0,0025	0,0018	0,00245	0,00177	85	0,1347	0,1041	0,1310	0,1067	0,14000	0,11500
40	0,0026	0,0018	0,0026	0,0020	0,00257	0,00190	86	0,1437	0,1205	0,1433	0,1232	0,15000	0,13000
41	0,0028	0,0025	0,0027	0,0022	0,00269	0,00218	87	0,1451	0,1140	0,1580	0,1360	0,17000	0,14000
42	0,0026	0,0023	0,0029	0,0023	0,00285	0,00234	88	0,1758	0,1746	0,1722	0,1480	0,18800	0,15500
43	0,0029	0,0025	0,0031	0,0026	0,00305	0,00250	89	0,1905	0,1667	0,2041	0,1871	0,20413	0,17000
44	0,0035	0,0026	0,0033	0,0028	0,00320	0,00270	90	0,2059	0,1643	0,2243	0,2113	0,22426	0,19000
45	0,0039	0,0031	0,0034	0,0029	0,00345	0,00293	91	0,3033	0,3157	0,2417	0,2338	0,24167	0,21000
							92	0,2458	0,2351	0,2509	0,2541	0,26000	0,23000
							93	0,2629	0,2873	0,2633	0,2805	0,29000	0,25000
							94	0,2367	0,2683	0,2367	0,2683	0,32000	0,28000
							95y+	0,2681	0,2959	0,2681	0,2959	0,35000	0,30500

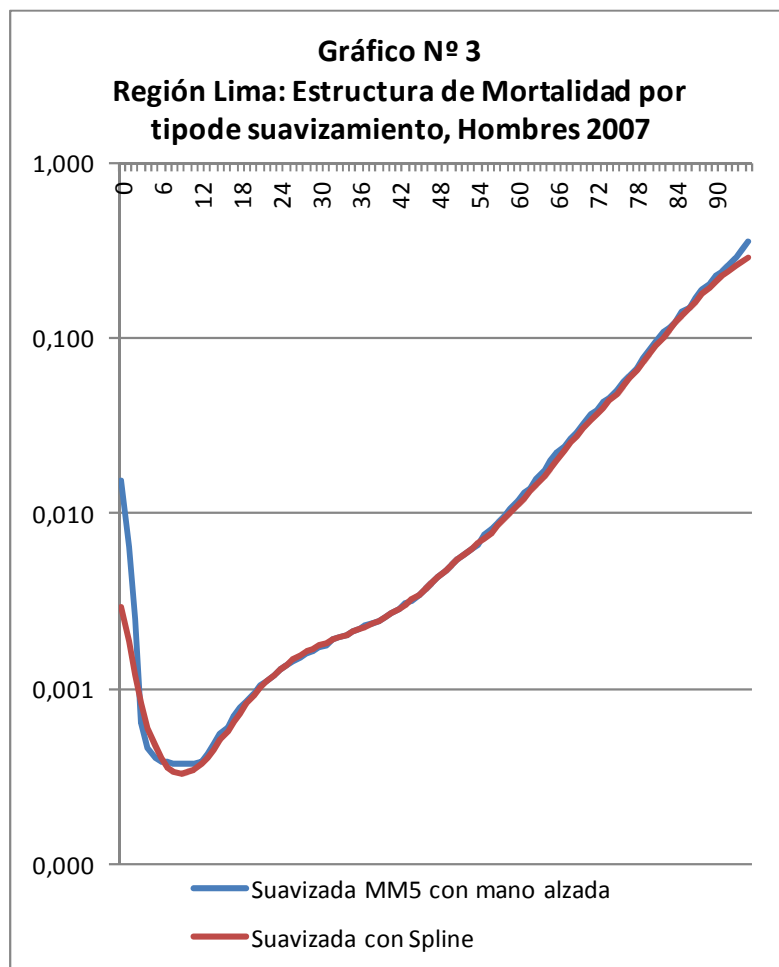
¿Se logra un mejor ajuste con la aplicación del Spline?

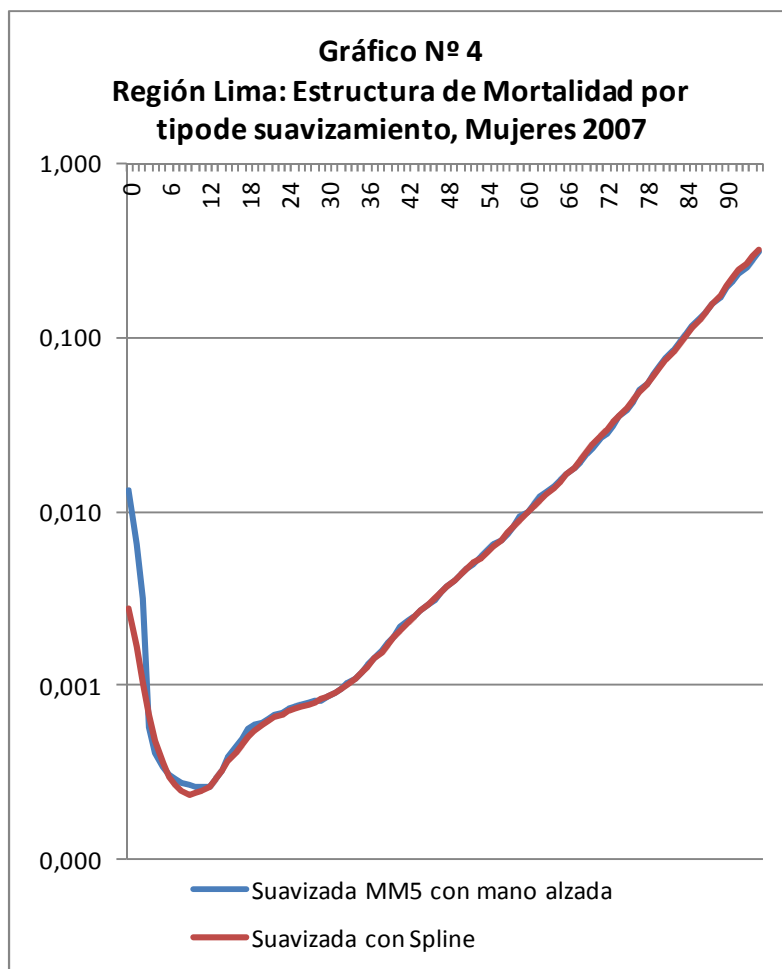
En base a lo observado en las Gráficas Nos. 3 y 4, podemos afirmar que las estructuras de muerte obtenidas por la aplicación del Spline son más precisas que el método de la mano alzada, al mostrar el comportamiento de la estructura de mortalidad por sexo y edad, con lo cual queda sentado que se está utilizando un modelo estadístico-matemático que permite obtener estructuras de muerte suavizadas y más robustas.

En cada uno de los gráficos, se muestra tanto la curva suavizada por el método práctico de medias móviles y mano alzada, así como la curva suavizada por el Spline, para hombres y mujeres.

En el caso concreto de las presentes aplicaciones, se prueba que el método de la mano alzada es aceptable, y que mediante el Spline se puede mejorar la precisión de la estructura de mortalidad.

Del análisis de las dos últimas gráficas se aprecia que en la gráfica correspondiente a hombres hay una ligera sobreestimación de muertes masculinas en la curva resultante de la aplicación del método de la mano alzada, tanto entre las edades 5 a 20 años como a partir de los 50 años. En la gráfica correspondiente a las muertes femeninas estaría indicando que ambas suavizaciones son aceptables, aunque la curva del método de la mano alzada indicaría una sobreestimación de las muertes femeninas entre los 5 y 20 años, e intercaladamente sobre y subestimación de muertes a partir de los 20 años.





Conclusiones

1. La estadística matemática tiene una serie de técnicas y métodos que permiten precisar más adecuadamente los comportamientos de los fenómenos demográficos, esto conlleva el conocimiento demográfico adecuado de las poblaciones, para poder adaptar los diversos modelos a las realidades concretas.
2. El Spline, conocido también por algunos investigadores como polinomio cúbico, interpolador cúbico, polinomio por segmentos, aunque como se demostró el desarrollo de su teoría no es tan simple como el polinomio en sí, permite modelar las estructuras de mortalidad por sexo y edad, curva que una vez suavizada es la base para la construcción de las tablas de mortalidad o tablas de vida demográficas.
3. La técnica de la mano alzada, cuando el investigador tiene experiencia y conoce la procedencia de los datos con los que trabaja, se aproximan bastante bien a lo que se puede obtener con un modelo estadístico-matemático, caso contrario puede que no se logre una curva que se aproxime a la obtenida por el Spline.
4. En vista de todo lo realizado en la presente investigación, y además dado el reemplazo de generación de demógrafos, se recomienda por ser mas precisa la utilización del Spline para la suavización de la estructura de mortalidad por sexo y edad, a partir de los 5 hasta los 95 y más años de edad, por la precisión de los métodos estadístico-matemáticos; de la edad 0 a los 4 años todavía será necesario la utilización de métodos demográficos específicos, y seguir estudiando alternativas estadístico-matemáticas.

REFERENCIAS BIBLIOGRÁFICAS

- Asís López, E.H. (2010). *Métodos numéricos con MatLab*. Lima: Fondo Editorial Universidad de Ciencias y Humanidades.
- Burden, R.L., Faires, J.D. (1985). *Análisis numérico*. México: Grupo Editorial Iberoamérica.
- Daniel, W.W. (1974). *Applied nonparametric statistics*. Boston: Houghton Mifflin Company.
- Eubank, R.L. (1999). *Nonparametric regression and Spline smoothing* (2th ed). New York: Marcel Dekker, Inc.
- Harrison, D. & Rubinfeld, D.L. (1978). *Boston Housing Data*. Hedonic prices and the demand for clean air. Boston: Orange. Retrieved from <http://orange.biolab.si/doc/datasets/housing.htm>
- INEI. Perú (2010). *Situación y perspectivas de la mortalidad por sexo y grupos de edad, nacional y por departamentos, 1990-2025. (metodología y tablas de mortalidad)*. Lima: Imprenta INEI.
- Ipanaqué, R., Urbina, R.T. y Correa, S.B. (1998). *B-Splines con Mathematica 5.1*. Piura: Universidad de Piura.
- Meza Santa Cruz, L.A. (2013). *Regresión no paramétrica utilizando Spline para la suavización de la estructura de la mortalidad en el Perú*. Tesis para optar el título de Licenciado en Estadística.
- Montgomery, D.C., Peck, E.A. y Vining, G.G (2004). *Introducción al análisis de regresión lineal* (primera reimpresión). México: Compañía Editorial Continental.
- NN UU. (1986). *Manual X, Técnicas indirectas de estimación demográfica*. Departamento de Asuntos Económicos y Sociales Internacionales. Estudios de Población N° 81. ST/ESA/SER.A/81. Nueva York.
- Ortega, A. (1987). *Tablas de mortalidad*. San José, Costa Rica: Centro Latinoamericano de Demografía (CELADE).
- Schumaker, L.L. (2007). *Spline functions: basic theory* (3th ed.). Cambridge: Cambridge University Press, Third edition, 2007.
- Siegel, J.S. & Swanson, D.A. (2008). *The methods and materials of demography* (2th ed.). Bingley, United Kingdom: Emerald Group Publishing Ltd.
- Simonoff, J.S.(1996). *Smoothing methods in statistics*. New York: Springer.
- Wang, Y. (2011). *Smoothing Splines, methods and applications*. Boca Raton: Taylor & Francis Group, LLC.