

Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.

Trabajo presentado para el XXI Premio de Investigación sobre Seguros y Fianzas 2014
“Antonio Minzoni Consorti”

Act. Karina Vargas Cruz

“LEUCE”



COMISIÓN NACIONAL DE
SEGUROS Y FIANZAS

XXI Premio de Investigación
sobre Seguros y Fianzas 2014
“Antonio Minzoni Consorti”

Segundo Lugar
Categoría de Seguros

Introducción

Durante los últimos años se ha observado que en las ciudades más desarrolladas del mundo la edad máxima alcanzada de las personas se ha incrementado en forma considerable, sin que esto signifique necesariamente una mejora en la mortalidad de los jóvenes sino más bien una mejora en la mortalidad de los ancianos.

El caso de México no es la excepción ya que el país ha sufrido grandes transformaciones a través de toda su historia, entre ellas (el que nos interesa para efectos de la presente tesis) el aumento de la edad máxima alcanzada por sus habitantes (sobre todo en sus zonas urbanas).

En general se espera que la edad máxima alcanzada en nuestro país tenga una trayectoria ascendente como consecuencia del envejecimiento de la población dada la estructura poblacional actual, los avances médicos y las políticas de salud vigentes.

El efecto anterior es importante que se reconozca en las actuales tablas de mortalidad, tal que la edad máxima de dichas tablas sea un reflejo de estas mejoras en la mortalidad de ancianos y no una edad máxima arbitraria que por costumbre se establece en 100 años.

Desde la aparición de las tablas de mortalidad (o tablas de vida) se ha tenido la inquietud por parte de especialistas dedicados a diferentes áreas de estudio, de terminarlas en una edad adecuada, por ejemplo desde el punto de vista demográfico se busca el hacer una proyección correcta del tamaño y estructura de la población por edades, mientras que en la profesión actuarial se busca un cálculo justo, óptimo, competitivo y eficiente de las anualidades vitalicias y los seguros de vida.

Por esta razón durante mucho tiempo se han desarrollado algoritmos para predecir los datos de mortalidad de una población (debido a que en ocasiones resulta sumamente complicado la obtención de estos datos) y a su vez que la predicción logre que la tabla de mortalidad termine en una edad que represente lo mejor posible a la longevidad de dicha población y no se haya elegido arbitrariamente.

Algunos de estos modelos serán desarrollados y explicados con mayor detenimiento en los siguientes capítulos, pues cada uno de los métodos tienen características particulares que los vuelven importantes; sin embargo, estos algoritmos también tienen carencias, lo que obliga como especialistas a comparar las alternativas y seleccionar la(s) que cubra(n) mejor las necesidades del medio asegurador.

Se puede decir sin temor a equivocarse que los procedimientos presentados en este trabajo, no se han aplicado en México y mucho menos se ha tratado de estimar una edad máxima para una tabla de mortalidad con base en la experiencia estadística del país. El análisis de las Tablas de Mortalidad Mexicana utilizadas en la actualidad e implementar un algoritmo para estimar la edad máxima límite en la que dichas tablas deben terminar permitirán cubrir los requerimientos de seguridad social, anualidades y seguros de vida.

Dado que sería muy ambicioso tratar de trabajar con todas las tablas de mortalidad mexicanas existentes se limitará al análisis y estudio de la tabla conocida en el sector asegurador como Tabla de Mortalidad CNSF-2000-I, ya que además fue para la única que se logró conseguir las edades y probabilidades de muerte sin ajustar.

Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.

La investigación está integrada por 4 capítulos, su contenido queda descrito como:

Capítulo 1: Se explica que es una tabla de mortalidad, así como las características que debe cumplir, los tipos de construcción que existen para éstas. También se describe la simbología que se utilizará durante el desarrollo de la presente investigación.

Capítulo 2: Se realizó una recapitulación de los antecedentes a lo largo de la historia para la estimación de la edad máxima alcanzada en diferentes poblaciones del mundo, con el objetivo de aplicarse a los datos de mortalidad mexicana 2000-I para observar y concluir si alguno de los modelos podría describir de manera adecuada a la población mexicana, entre los modelos analizados se encuentra el modelo de extrapolación polinomial, el modelo de Gompertz, el modelo de Makeham, el modelo e Heligmand & Pollard el modelo de Coale-Kisker, el modelo relacional de mortalidad y por último el modelo Logit.

Capítulo 3: Después del desarrollo de los modelos anteriores se observó que la combinación de éstos ha logrado hacer cambios a favor de la población, por esta razón se aplicó el Teorema del Valor Extremo al objetivo inicial, estimar la edad máxima alcanzada de la población mexicana de manera adecuada. Se dedicó el capítulo completo a este análisis debido a que se considera un modelo innovador y se sabe que ha sido eficaz en Canadá y Japón.

Capítulo 4: Un punto importante que se resalta de esta investigación es la importancia en las aplicaciones que tiene el estimar la edad máxima de la población mexicana 2000-I, que concluye con un ejemplo práctico de los beneficios y carencias que resultan de realizar una estimación no adecuada o incluso sugerir arbitrariamente el final de dicha curva.

Finalmente se encuentran las conclusiones que generó la realización de la presente investigación.

Índice

Introducción	2
Capítulo 1 Conceptos básicos.....	7
1.1 Notación.....	9
1.2 Datos utilizados para los ajustes realizados.....	12
Capítulo 2 Antecedentes históricos para la estimación de la edad máxima alcanzada, aplicados en la Mortalidad Mexicana 2000-I	13
2.1 Extrapolación polinomial	14
2.2 Modelo de Gompertz.....	17
2.3 Modelo de Makeham.....	21
2.4 Modelo de Heligman & Pollard	24
2.5 Modelo de Coale-Kisker	29
2.6 Modelo relacional de mortalidad	35
2.6.1 Modelo Logit	36
2.6.1.1 Tabla de mortalidad CNSF 2000-I. Un caso particular	39
Capítulo 3 Aplicación del Teorema del Valor Extremo al cálculo de la edad máxima esperada de la Tabla de Mortalidad Mexicana CNSF 2000-I	44
3.1 Teoría de Extremos.....	44
3.1.1 Distribución Generalizada de Extremos (GEVD)	45
3.1.2 Distribución Generalizada de Pareto. (GPD)	46
3.2 Especificaciones del modelo.....	47
3.2.1 Estimación de parámetros	49
3.2.1.1 Máxima verosimilitud.....	49
3.2.1.2 Mínimos cuadrados.....	51
Capítulo 4 Importancia de la edad máxima en la Tabla de Mortalidad Mexicana CNSF-2000-I	55
4.1 Importancia en aplicaciones actuariales	56
4.1.1 Anualidades Contingentes.....	56
4.1.1.1 Plan de pensiones. Un caso particular de las anualidades	56
4.1.2 Seguros.....	58
Conclusiones	61
Anexos.....	66
Anexo A (Códigos de análisis)	66
A.1 Código de extrapolación polinomial en R	66
A.2 Comparación de muestras mediante la prueba Cuantil-Cuantil	66
A.3 Prueba Kolmogorov-Smirnov.....	67
A.4 Desarrollo del modelo de Coale-Ksiker.....	69

A.5	<i>Simulaciones binomiales del modelo logit, para la estimación de la curva que describe la mortalidad mexicana 2000-I</i>	70
<i>Anexos B (Observaciones de la distribución Pareto)</i>		70
B.1	<i>Distribución Pareto generalizada</i>	70
B.1.1	<i>Características para la distribución del máximo</i>	71
B.1.2	<i>Características para la distribución del mínimo</i>	71
B.1.3	<i>Casos particulares</i>	72
B.1.3.1	<i>Para el máximo</i>	72
B.1.3.2	<i>Para el mínimo</i>	72
B.2	<i>Desarrollo de la igualdad para la estimación de parámetros del modelo construido a partir de la teoría de extremo (GPD)</i>	72
B.3	<i>Simulación de la curva de mortalidad cuando el umbral es de 85 años</i>	73
<i>Anexos C (Metodologías de cálculo)</i>		74
C.1	<i>Ajuste de parámetros del modelo de Gompertz</i>	74
C.2	<i>Ajuste de parámetros del modelo de Makeham</i>	76
C.3	<i>Ajuste de parámetros del modelo de Heligman & Pollard</i>	77
C.4	<i>Ajuste de parámetros del modelo de logit</i>	79
C.5	<i>Ajuste de parámetros de la teoría de extremos</i>	81
<i>Anexos D (Tabla de mortalidad mexicana 2000-I, ajustada por diferentes metodologías)</i>		83
<i>Anexos D.1 Muestra de mortalidad mexicana CNSF-2000-I. Datos base</i>		84
<i>Anexos D.2 Tabla de mortalidad mexicana ajustada mediante el modelo de extrapolación polinomial</i>		85
<i>Anexos D.3 Tabla de mortalidad mexicana ajustada mediante el modelo de Gompertz</i>		86
<i>Anexos D.4 Tabla de mortalidad mexicana ajustada mediante el modelo de Makeham</i>		87
<i>Anexos D.5 Tabla de mortalidad mexicana ajustada mediante el modelo de Heligman & Pollard</i>		88
<i>Anexos D.6 Tabla de mortalidad mexicana ajustada mediante el modelo de Coale Kisker</i>		89
<i>Anexos D.7 Tabla de mortalidad mexicana ajustada mediante el modelo de Logit</i> ..		90
<i>Anexos D.8 Tabla de mortalidad mexicana ajustada mediante el modelo de CNSF</i> ...		91
<i>Anexos D.9 Tabla de mortalidad mexicana ajustada mediante el modelo de Valores extremos</i>		92

Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.

<i>Anexos D.10 Tabla de mortalidad mexicana ajustada mediante el modelo de Valores extremos.</i>	93
Glosario	94
Bibliografía.....	95
Referencias	97
Páginas web	97

Capítulo 1

Conceptos básicos

La herramienta estadística por excelencia para el análisis de la supervivencia, evolución del volumen y estructura demográfica de una población es la tabla de mortalidad.

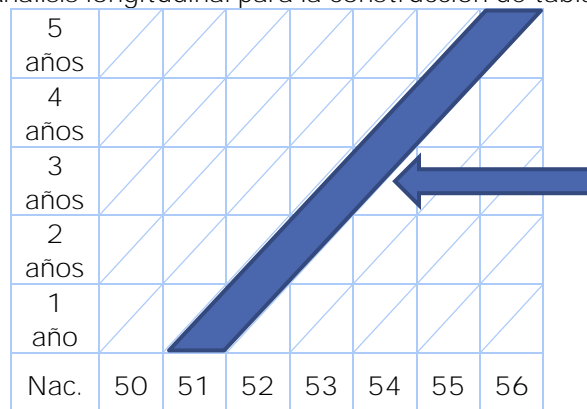
Las tablas de mortalidad son un instrumento que permite realizar análisis temporales y comparaciones especiales sobre la incidencia de algún fenómeno con independencia de la estructura por edad de las poblaciones estudiadas (INE, 2011). Una tabla de mortalidad contiene una descripción de las características más importantes de la población de un país o de una ciudad. Éstas deben cumplir las siguientes características:

1. Las tasas de fallecimiento deben ser positivas y encontrarse en el intervalo $[0,1]$, es decir son probabilidades.
2. Las probabilidades de ocurrencia de siniestro, en este caso el fallecimiento deben ser crecientes; pues se considerará que a mayor edad, mayor es la probabilidad de fallecer.
3. Se pueden construir a partir de Censos o Información Estadística de Aseguradoras (en México).

La construcción de una tabla de mortalidad se puede hacer a partir de dos tipos de análisis de la supervivencia de una población, los cuales son:

1. Análisis longitudinal (Tablas de generación). Suponen seguir la evolución a lo largo de la existencia de la cohorte para conocer cómo se comporta la mortalidad hasta llegar a la extinción de la misma, es decir la probabilidad de fallecimiento en ese momento será uno.

Figura 1.1 Análisis longitudinal para la construcción de tablas de mortalidad.



Fuente: Elaboración propia.

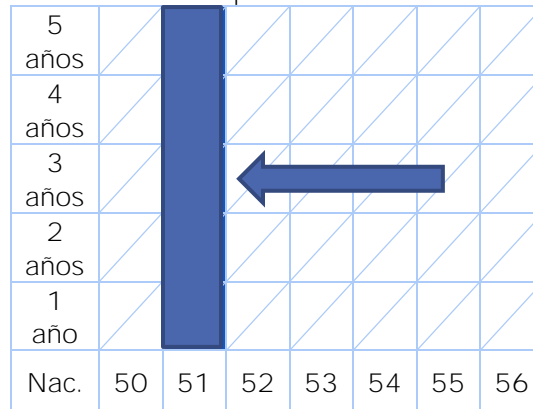
La eficiencia de estas tablas es muy limitada debido a que el análisis es extenso; sin embargo, es bueno realizar este tipo de análisis si el propósito es observar las transformaciones en salud.

2. Análisis transversal (Tablas de momento). Este modelo relaciona la mortalidad en un corto período de tiempo (generalmente toma entre uno y cuatro años, como máximo),

Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.

como una generación ficticia dada por la población existente en el momento al que están referidos los fallecimientos. La ventaja resulta de inmediato debido a que la construcción de las tablas sigue el mismo proceso que el análisis anterior, sólo que ahora la disponibilidad de datos es mayor. Este análisis permite observar el cambio de la mortalidad de un periodo a otro y es particularmente útil cuando se examinan cambios bruscos.

Figura 1.2 Análisis transversal para la construcción de tablas de mortalidad.



Fuente: Elaboración propia.

A partir de lo anterior se puede encontrar el análisis de patrones de edades avanzadas dividido en 3 representaciones, las cuales son:

1. Leyes de mortalidad. Son funciones algebraicas que representan índices de mortalidad en función de la edad que tiene el individuo (análisis transversal). Por mencionar algunas de ellas se tienen la ley de Gompertz, de Makeham y de Heligman-Pollard.
2. Modelos de tablas de mortalidad. Son tablas que resumen las observaciones de varias poblaciones a través del establecimiento de patrones de mortalidad, todos ellos relacionados con la esperanza de vida al nacer, a diferencia de las anteriores que están en función de la edad del individuo (análisis longitudinal).
3. Modelo relacional. Este modelo relaciona los 2 anteriores (análisis transversal y longitudinal). Debido a que las tablas de mortalidad son un pilar fundamental en la ciencia actuarial, a lo largo de la historia se han intentado dar formas analíticas para describir las funciones de mortalidad. Las principales razones para postular las formas analíticas son las siguientes:
 - a. Filosófica. Debido a que muchos fenómenos pueden explicarse eficientemente mediante fórmulas sencillas usando argumentos sociales, biológicos, etc.
 - b. Práctica. Es más sencillo comunicar la idea de una función que tiene pocos parámetros y no con una tabla que consta de 100 observaciones.
 - c. Facilidad de estimación. Resulta atractivo poder obtener más información de mortalidad con pocos parámetros, que la necesidad de conocer la información completa.

1.1 Notación

Para que una tabla de mortalidad sea útil es recomendable conocer la notación que se presenta en la construcción de la misma, por esta razón se explica la misma:

- x : una persona con edad x .
- X : Edad a la que un recién nacido fallece, que es una variable aleatoria continua.
- $f(x)$: Función de densidad de X , que puede ser obtenida como: $f(x) = \frac{d}{dx} F(x) = F'(x)$.
- $F(x)$: Función de distribución de X . Esta función representará la probabilidad de que un recién nacido fallezca antes de alcanzar la edad x . Se puede expresar de la siguiente forma: $F(x) = P[X > x]$ para $x \geq 0$.
- $S(x)$: Función de supervivencia. Es la probabilidad de que un recién nacido sobreviva o alcance la edad x . Es muy común encontrar la función expresada como: $S(x) = P[X > x] = 1 - P[X \leq x] = 1 - F(x)$ para $x \geq 0$.
- **$P[x < X \leq y | X > x]$** : Es la probabilidad de que un recién nacido fallezca entre las edades x y y sabiendo que sobrevive a edad x . O bien es la probabilidad de que una persona con edad x fallezca antes de alcanzar la edad $x + y$. Esta probabilidad condicional puede ser expresada en términos de la función de supervivencia y de distribución de la siguiente manera:

$$P[x < X \leq y | X > x] = \frac{F(y) - F(x)}{1 - F(x)} = \frac{1 - S(y) - 1 + S(x)}{S(x)} = \frac{S(x) - S(y)}{S(x)}$$

Para $x \geq 0$ y $y \geq 0$.

- $\mu(x)$: Fuerza de mortalidad. En términos de la función de densidad y la función de supervivencia, ésta es la probabilidad condicional de un recién nacido de morir entre las edades x y $x + \Delta x$, dado que ya sobrevivió a edad x , con $\Delta x \rightarrow 0$, es decir:

$$P[x < X \leq x + \Delta x | X > x] = \frac{F(x + \Delta x) - F(x)}{1 - F(x)} \cong \frac{f(x)}{1 - F(x)}$$

Por propiedades de $f(x)$ y $F(x)$, $\mu(x)$ es positiva.

- d_x : Número de muertes entre las edad x y $x+1$.
- E_x : Número de expuestos al riesgo entre las edad x y $x+1$. En la práctica, E_x se toma a mitad del año en cuyo caso es el número de sobrevivientes a edad $x + \frac{1}{2}$.
- l_x : Número de sobrevivientes a edad x .
- $q_x = \frac{d_x}{l_x}$: Probabilidad de muerte entre las edades x y $x+1$ y puede ser expresada como:

$$q_x = 1 - e^{-\int_0^1 \mu(x+s) ds}$$

- q'_{x-i} : Será la probabilidad estimada de muerte entre edades x y $x+1$, por la metodología i .
- $m_x = \frac{d_x}{E_x}$: Índice de mortalidad central en el intervalo $(x, x+1)$. Una manera útil de reexpresar dicha información, es considerando que los sobrevivientes a edad $x + \frac{1}{2}$, son las personas que viven a edad x menos las que mueren a edad $x + \frac{1}{2}$, es decir, $E_x = l_x - d_{x+\frac{1}{2}}$. Para el desarrollo de la presente investigación, se puede deducir entonces una relación:

$$m_x = \frac{d_x}{E_x} = \frac{d_x}{l_x - d_{x+\frac{1}{2}}} \left[\frac{\frac{1}{l_x}}{\frac{1}{l_x}} \right] = \frac{q_x}{1 - \frac{q_x}{2}}$$

De lo anterior se llega a que:

$$q_x = \frac{m_x}{1 + \frac{m_x}{2}} \quad 1.1$$

- N : Edad umbral. Es la edad a partir de la cual resulta más adecuada la aplicación de algún algoritmo de aproximación para determinar la edad máxima en la que una tabla de mortalidad debiera terminar.
- w : Edad máxima estimada. A partir de este momento $q_x = 1$ con $x \geq w$.
- $V^t = (1+k)^r$: Valor Presente. Valor de una unidad monetaria en el momento de valuación. con k la tasa de interés anual, r el número de periodos anuales.
- a_x : Anualidad vitalicia vencida. Pagos que se realizan al final del año, se efectúan mientras el individuo sobreviva. Para una unidad monetaria se calcula como:

$$a_x = \sum_{r=1}^{w-x} V^r {}_r p_x$$

- \ddot{a}_x : Anualidad vitalicia anticipada. Pagos realizados al inicio del año y se efectuarán mientras el individuo sobreviva. Para una unidad monetaria se calcula como:

$$\ddot{a}_x = \sum_{r=0}^{w-x-1} V^r {}_r p_x$$

- $a_{x:n}$: Anualidad temporal n -años Vencida. Pagos realizados al final del año, acordados por un periodo de n años (donde $0 < n \leq (w-x)$ años, con w la edad máxima estimada de fallecimiento (es decir para todo $w \leq x$, $q_x = 1$) y x la edad actual del individuo). Para una unidad monetaria se calcula como:

Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.

$$a_{x:n} = \sum_{r=1}^n V^r {}_r p_x$$

- **$\ddot{a}_{x:n}$** : Anualidad Temporal n-años Anticipada. Son pagos realizados al inicio del año, acordados por un periodo de n años (donde $0 \leq n \leq (w-x-1)$ años y w la edad máxima estimada de fallecimiento (es decir para todo $w \leq x, q_x = 1$) con x la edad actual del individuo). Para una unidad monetaria se calcula como:

$$\ddot{a}_{x:n} = \sum_{r=0}^{n-1} V^r {}_r p_x$$

- **${}_k/\ddot{a}_{x:n}$** : Anualidad anticipada temporal n -años diferida k -años. Se comporta como una anualidad anticipada temporal a n -años con la diferencia de que se inician los pagos con un desplazamiento k -años respecto a la edad x . Se calcula como:

$${}_k/\ddot{a}_{x:n} = \sum_{r=k}^{n+k-1} V^r {}_r p_x$$

- **${}_k/a_{x:n}$** : Anualidad vencida temporal n -años diferida k -años. Es una anualidad vencida temporal n -años con la diferencia de que son acordadas con un desplazamiento k -años de pago respecto a la edad x , se obtiene como:

$${}_k/a_{x:n} = \sum_{r=k+1}^{n+k} V^r {}_r p_x$$

- **A_x** : Seguro vitalicio. Es el costo de un seguro que paga una unidad monetaria en el momento de fallecer a una persona de edad x de manera vitalicia. Puede ser calculado como:

$$A_x = \sum_{r=1}^{w-x} V^r {}_{r-1} p_x q_{x+i}$$

- **$A_{x:n}$** Seguro temporal n -años. Es el costo de un seguro que paga una unidad monetaria en el momento de fallecer a una persona de edad x durante n -años y es calculado como:

$$A_{x:n} = \sum_{r=1}^{w-x-n} V^r {}_{r-1} p_x q_{x+i}$$

- **PNU** : Prima Única. Se entiende como prima única, el monto que debe pagar el asegurado en una única exhibición por el seguro que adquirió.

Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.

- **PNN:** Prima Neta Nivelada. Se entiende como prima neta nivelada, el monto que debe pagar el asegurado en un número k de exhibiciones (pactado al inicio del contrato) por el seguro que adquirió.

1.2 Datos utilizados para los ajustes realizados

Los datos con los cuales se trabajó en esta investigación corresponden a lo utilizado por Gutiérrez-Peña E:

“La operación del Sector Asegurador Mexicano durante el período 1982 a 1989. Son datos que las compañías de seguros reportaron a la Comisión Nacional de Seguros y Fianzas¹ durante ese período y que sirvieron de base para construir las tablas de mortalidad que estuvieron vigentes en México hasta 1999 (para Seguro de Vida Individual). Los datos están agregados, es decir, para cada edad tanto el número de expuestos como el de siniestros corresponde a la suma de las cifras de los ocho años.²”

Estos datos fueron consolidados en una base de datos que pasó una revisión, exploración y validación para ser utilizados en la construcción de las tablas de mortalidad 2000-I que se publicaron por la Comisión Nacional de Seguros y Fianzas (CNSF³).

Esta muestra es representativa de la población mexicana asegurada debido a lo anterior no hay que perder de vista que no incluye aquellos sectores de la población mexicana sin capacidad económica para contratar un seguro, en caso que se deseara utilizar para otros propósitos fuera del sector asegurador, habría que tener una nueva muestra representativa de dicha población.

¹ De aquí en adelante se escribirá CNSF, para hacer referencia a la Comisión Nacional de Seguros y Fianzas.

² La muestra de mortalidad mexicana para 2000-I.

³ Tabla de mortalidad mexicana 2000-I, ajustada mediante el modelo logit publicada por la CNSF.

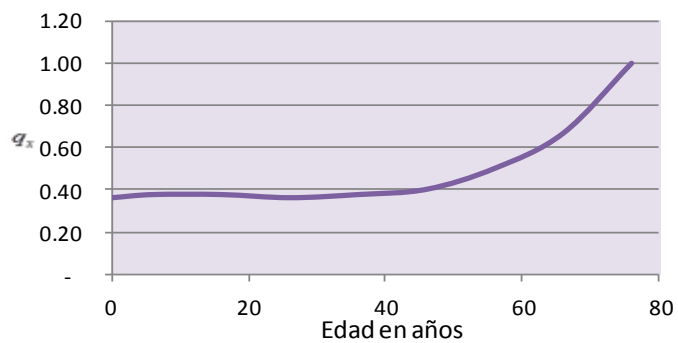
Capítulo 2

Antecedentes históricos para la estimación de la edad máxima alcanzada, aplicados en la Mortalidad Mexicana 2000-I

En 1662, John Graunt, publicó la primera tabla de mortalidad (que describía la población de Londres) en la obra titulada “Observations Natural the Bills of Mortality” (Graunt, J., 1683).

Figura 2.1 Mortalidad de Londres en 1662 según John Graunt.

X	l_x
0	100
6	64
16	40
26	25
36	16
46	10
56	6
66	3
76	1



Fuente: Elaboración propia.

Los registros de mortalidad a los que tuvo acceso John Graunt indicaban la causa de muerte y sexo de los individuos pero no su edad; a pesar de esto, la tabla se construyó presentando edades y número de individuos por cada una de ellas, provocando así desconfianza en los datos mostrados.

Sin embargo, algunos investigadores creen que las hipótesis que usó Graunt para lograr esta tabla de mortalidad fueron:

- Para establecer el número de muertes de individuos entre cero y seis años de edad contabilizó los fallecimientos cuya causa de muerte se debiera a enfermedades típicamente infantiles y agregó la mitad de fallecimientos de viruela y sarampión considerando que estas enfermedades las sufrían tanto niños como adultos.
- Para establecer el número de muertes de individuos entre los 6 y los 76 años de edad realizó una extrapolación asumiendo $\mu = 0.047$ siguiendo una ley exponencial, así que la función de supervivencia que refleja esta hipótesis, quedó expresada de la siguiente manera:

$$l_x = 64e^{-0.047(x-6)^4}$$

- Ningún individuo permanece con vida después de los 76 años de edad.

Por esta razón Graunt fue considerado el primer demógrafo de la historia.

⁴ Recordemos que la ley exponencial está dada por $l_1 = l_0 e^{k\Delta t}$ donde Δt es el incremento de tiempo.

Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.

“La primera tabla de mortalidad desarrollada de una manera lógica, la tabla de mortalidad de Halley, se publicó en 1693 y está basada en los registros de muerte y de nacimiento de la ciudad de Breslau durante los años de 1687 a 1691. Para la preparación de esta tabla se asumió que la población de Breslau había permanecido estable (por ejemplo, que el número completo de la población al igual que la edad y el género no cambiaban en muchas décadas) y esta suposición no era del todo correcta, por lo tanto, la tabla de mortalidad resultante era imprecisa” (Gil Fana, 1999).

Después de esta publicación, le siguieron un gran número de tablas de mortalidad por ejemplo:

“En 1775 Abraham De Moivre ajustó por primera vez una fórmula matemática a una tabla empírica. La tabla fue la de Halley y la fórmula ajustada fue:

$$l_x = 86 - x$$

Para las edades x , comprendidas entre los 12 y 86 años” (Gil Fana, 1999).

Durante los primeros años después de estos análisis, sólo se registraba el desarrollo para países europeos; sin embargo, a lo largo de la historia este proceso se fue ampliando para las demás ciudades del mundo. Actualmente con el desarrollo de la tecnología las tablas de mortalidad son de fácil acceso para casi todos los países del mundo y no solo eso, también existe la posibilidad que puedan ser presentadas por continente.

2.1 Extrapolación polinomial⁵

Son muchas las situaciones y los casos en que se considera adecuado hablar de extrapolación, debido a que este método sólo necesita de una serie de datos o resultados experimentales, de las que sólo se conoce una cantidad finita y para las que se necesita encontrar una Ley General que aplique al fenómeno bajo estudio, a la que se le llamará función.

La función buscada debe tener características que permitan trabajar fácilmente con los datos que forman la muestra, y finalmente ésta ayudará a interpretar y predecir de mejor manera el comportamiento de los mismos.

Un caso particular en el que se puede desarrollar un proceso de extrapolación, es precisamente el ajuste de la tabla de mortalidad, debido a que la información que se tiene es sesgada y por esta razón, al aplicar la metodología se encontrará un modelo que describa de manera clara la muestra.

Al concluir el proceso automáticamente se asigna una función, la cual nos conducirá a encontrar la edad máxima estimada en que la tabla de mortalidad debiera concluir, que es el objetivo principal de este trabajo.

⁵ León Vázquez R., Conzanzo J. (2006). Splines Cúbicos. (Valparaíso). Forsythe, G. E., Malcolm, M. A. and Moler, C. B. (1977) Computer Methods for Mathematical Computations.

Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.

A pesar de que la interpolación con polinomios es una gran herramienta teórica en Análisis Numérico, este método presenta una desventaja de suma importancia, pues cuando se tiene una muestra grande de datos a interpolar, el polinomio generalmente tiene un exponente de grado alto y suelen aparecer un gran número de oscilaciones a lo largo de la curva.

Por esta razón la estrategia que generalmente se utiliza para mitigar este tipo de errores es interpolar con polinomios de grado bajo entre datos que son consecutivos y así formar una función general a trozos, formada por todo este conjunto de polinomios.

La idea general para realizar un ajuste a los datos de una tabla de mortalidad o a cualquier función es la siguiente:

Resulta más conveniente partir los datos en subintervalos, a los cuales se les ajustará un polinomio de grado k o menor (el objetivo es que sea el de menor grado posible) y al finalizar la tarea se realizará la unión de estos ajustes, es decir, la unión de los polinomios obtenidos; teniendo presente la suavización de nodos, por tal motivo la función debe ser diferenciable $k-1$ veces en cada uno de los nodos, tal que se pueda tener la certeza de que la función que describirá el modelo es continua en los nodos interiores, y así se pueda agregar cada uno de los intervalos ajustados por un polinomio de grado bajo.

Existen 4 clasificaciones para las Extrapolicaciones cúbicas, las cuales son:

1. *No nodo.* Esta clasificación se obtiene con la condición de que los 2 primeros y últimos polinomios sean iguales.
2. *Completo.* Este Spline resulta de conocer el valor de la primera derivada en los extremos.
3. *Natural.* Se conoce la segunda derivada de ambos extremos, en caso contrario se asume que ambas son iguales a cero.
4. *Caso particular del Natural.* Se obtienen calculando la segunda derivada en los extremos, mediante la interpolación de los nodos más cercanos.

Por lo atractivo que resulta el modelo se calculó con los datos base de la tabla de mortalidad mexicana 2000-I⁶, tal como lo aplicaron Panjer y Russo y Panjer y Tan, con la experiencia de mortalidad individual canadiense en 1992 y 1995, respectivamente.

Para aplicar este modelo, se utilizó el lenguaje de programación⁷ “R” el cual cuenta con la función de extrapolación cúbica que permitió realizar el ajuste deseado, sin necesidad de desarrollar un programa de cálculo complicado⁸.

Realizando un comparativo entre los datos originales y el modelo ajustado mediante la extrapolación polinomial resulta la Figura 2.2.1.

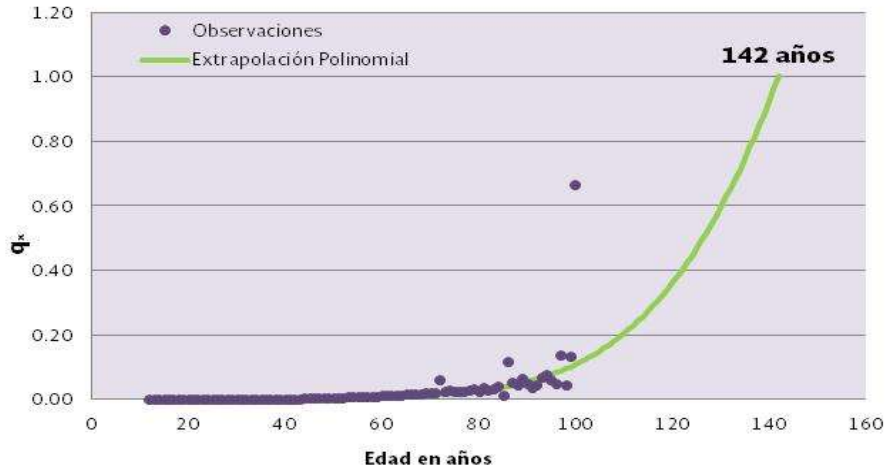
⁶ Tabla de mortalidad mexicana 2000-I ajustada mediante el método de extrapolación polinomial. Anexo D2

⁷ El lenguaje de programación es gratuito. Se puede obtener de la página <http://cran.r-project.org>.

⁸ El código que modela el método de extrapolación polinomial en “R” se encuentra en el Anexo A1

Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.

Figura 2.2.1 Método de extrapolación polinomial aplicado a los datos base de mortalidad mexicana 2000-I.



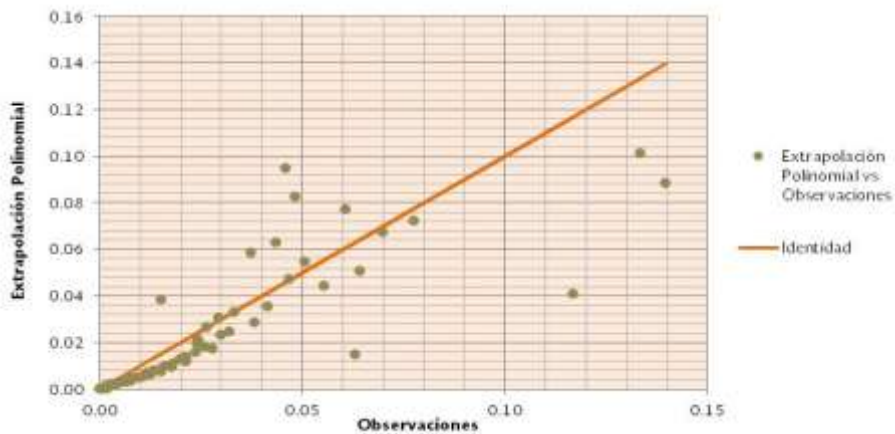
Fuente: Elaboración propia.

La curva estimada de mortalidad conduce a concluir que la edad máxima esperada para la población mexicana es de 142 años, es decir, a partir de este momento $q_x = 1$.

Así se logró hacer una estimación para la edad máxima alcanzada de la población mexicana, es decir, ha cubierto el propósito básico de esta investigación, a pesar de esto, como especialistas del tema se sabe que siempre que se realice un ajuste estadístico de un conjunto de datos, será importante realizar alguna prueba que permita decidir si el conjunto de datos se ajusta apropiadamente a la función que se propone, esta prueba es conocida como Prueba de Bondad de Ajuste.

Una forma de contrastar los datos del ajuste fue un gráfico cuantil a cuantil⁹ para analizar si adecuado para describir la mortalidad mexicana 2000-I (Figura 2.2.2).

Figura 2.2.2 Cuantil-Cuantil extrapolación polinomial vs observaciones aplicado a la mortalidad mexicana 2000-I.



Fuente: Elaboración propia.

⁹ Si se desea obtener más información de un gráfico cuantil a cuantil ver el Anexo A.2

Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.

Este análisis permitió concluir que la estimación de la curva mediante la extrapolación polinomial no es la más adecuada para edades avanzadas, debido a la gran variabilidad que existe entre los datos reales y la curva estimada, sobre todo en la cola de la curva de mortalidad que es el objetivo principal de la investigación; sin embargo, al ser un gráfico y el análisis un poco subjetivo, se buscó una prueba más consistente, que definiría si el modelo propuesto es adecuado.

Para realizar un análisis más consistente se utilizó la prueba de Kolmogorov-Smirnov¹⁰, la cual arrojó el siguiente resultado:

Observaciones vs Extrapolación Polinomial
D = 0.070554,
p-value < 2.2e-16

Por lo tanto, esta prueba rechaza la idea de que $F(x)$ queda descrita adecuadamente por $F_o(x)$, es decir, q_x no queda descrita adecuadamente por $q_{x\text{-extrapolación}}$, debido a que:

$D < 0.2067$
 $p\text{-value} > 0.838^{11}$

Por esta razón no puede ser la solución más adecuada ante el problema de la edad máxima estimada de la población.

A pesar de la existencia de mexicanos con edades muy avanzadas, éstos representan casos excepcionales y no debieran tener mucho peso al estimar la de edad máxima para la tabla de mortalidad mexicana 2000-I.

Finalmente el modelo de extrapolación polinomial ha mostrado una de sus limitantes, a pesar de haberse realizado un ajuste matemático importante, ya que logró la estimación de la edad máxima suavizando la curva que describe la mortalidad mexicana, no cubrió las características que se presentan en la vida diaria, ya que imposible imaginar personas de 142 años debido a que no se tiene conocimiento de individuos que hayan sobrevivido más de 124 años.

Esto significa que aunque la extrapolación cúbica es un método eficaz en algunas situaciones, también puede tener carencias en otros ámbitos, lo que invita a no sólo realizar ajustes matemáticos, sino ejecutar análisis de los resultados.

2.2 Modelo de Gompertz¹²

Un modelo que resulta interesante analizar dentro del ámbito actuarial en el tema de tablas de mortalidad es el modelo de Gompertz éste es un modelo demográfico publicado en 1825 por Benjamin Gompertz, el cual supone que cada individuo presenta una resistencia

¹⁰ Si se desea conocer más acerca del desarrollo de la prueba de Kolmogorov-Smirnov dirigirse al Anexo A3. Siempre que se hable en la presente investigación de la prueba Kolmogorov-Smirnov, se dará por entendido que se trabajó en la paquetería R.

¹¹ Si se desea saber la justificación de los valores para D y p-value ver el Anexo A.3

¹² Gompertz, B. (1825). On the nature of the function expressive of the law of human mortality and on a new mode of determining life contingencies. Phil. Trans. Roy. Soc. (London). Ser. A. 115: 513-585.

Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.

decreciente a las enfermedades y a fallecer por causas naturales, es decir, conforme la edad aumenta menor será la resistencia que presenta el individuo, por lo que la fuerza de mortalidad aumenta con la edad y su incremento relativo es constante. Así la función de densidad está dada por:

$$f(x) = BC^x(C - 1)e^{-\frac{BC^x}{\ln(C)}(C-1)} \quad 2.2.1$$

Con $B > 0$, $C > 1$ y $x \geq 0$.

La función de distribución está dada por:

$$F(x) = 1 - e^{-\frac{BC^x}{\ln(C)}(C-1)}$$

La función de supervivencia está dada por:

$$S(x) = e^{-\frac{BC^x}{\ln(C)}(C-1)}$$

Finalmente la fuerza de mortalidad¹³ está dada por:

$$\mu(x) = BC^x$$

Si $C = 1$ en la ley de Gompertz, la fuerza de mortalidad es constante.

Para realizar la estimación de los parámetros que describen la tabla de mortalidad mexicana 2000-I mediante el modelo de Gompertz se utilizó la técnica de mínimos cuadrados con apoyo de una hoja electrónica de cálculo (Excel)¹⁴.

Al terminar el proceso de minimización se encuentran los parámetros que ajustan el modelo de Gompertz. Para el presente ajuste los resultados fueron:

$$B = 0.000173$$

$$C = 1.067372$$

Lo cual resulta consistente si se compara con las condiciones que debe cumplir cada uno de los parámetros.

Con el resultado obtenido se realizó un gráfico que permitiera realizar una comparación de los datos observados contra los datos estimados por el modelo de Gompertz, usando la función de la fuerza de mortalidad para poder analizar la consistencia que existía entre ellos,

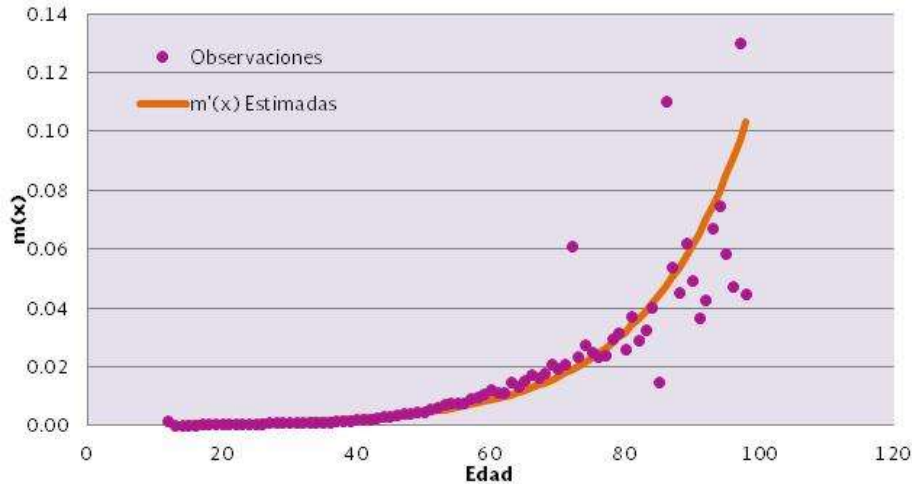
¹³ Siendo esta un caso particular de la ley de Makeham ($\mu(x) = A + BC^x$) con $A=0$. Ver Modelo de Makeham, pág.

¹⁴ Si se desea ver el cálculo realizado ir al Anexo C.1

Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.

y verificar visualmente que realmente los parámetros obtenidos corresponden al objetivo inicial del ajuste.

Figura 2.3.1 Ajuste de parámetros estimados para m_x para el modelo de Gompertz de los datos base de la población mexicana 2000-I.



Fuente: Elaboración propia.

Al observar que la curva generada podía describir a las observaciones, se realizó entonces el cálculo de q_x que puede ser reexpresado como $q_x^{15} = \frac{m_x}{1 + \frac{m_x}{2}}$.

Es importante destacar que se tiene una aproximación de $\mu(x) \cong m_x$, por lo tanto, de la ecuación anterior se puede obtener mediante un despeje algebraico un ajuste a la tabla de mortalidad mexicana, el cual queda descrito por la ecuación:

$$q_x = \frac{BC^x}{1 + \frac{BC^x}{2}}$$

Debido a que son conocidos los parámetros B y C, se calcula entonces cada probabilidad de fallecimiento (q_x) como:

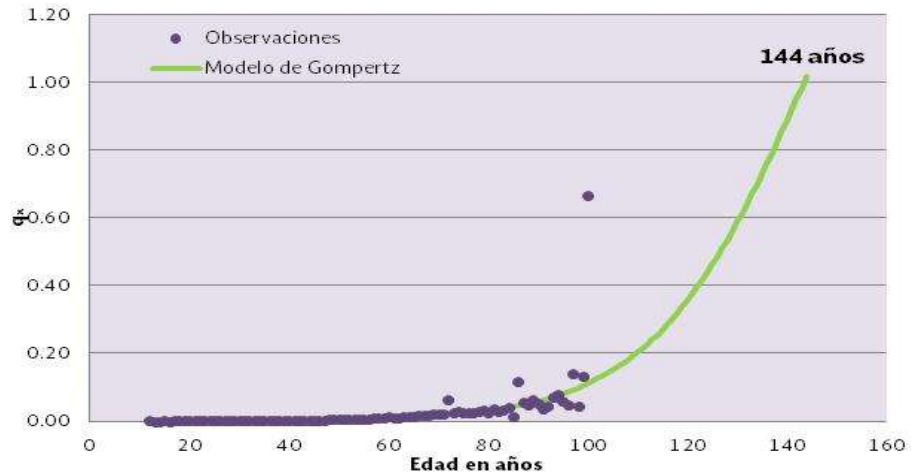
$$q_x = \frac{(0.000173)(1.067372)^x}{1 + \frac{(0.000173)(1.067372)^x}{2}}$$

Con $x \in \{1, 2, 3, \dots, 144\}$. Con el resultado anterior se construyó la curva de mortalidad que se puede observar en la figura 2.3.2.

¹⁵ Si se tiene desea verificar la igualdad ir al apartado 1.1

Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.

Figura 2.3.2 Modelo de Gompertz aplicado a los datos base de mortalidad mexicana 2000-I.

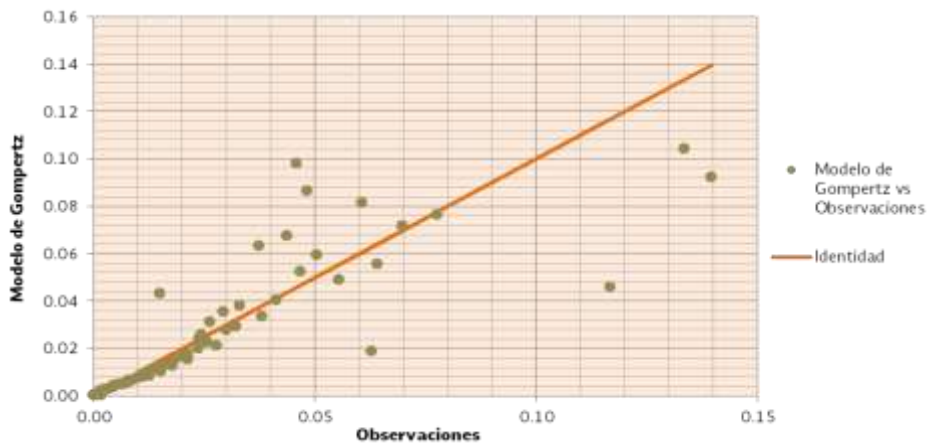


Fuente: Elaboración propia.

El modelo que propone Gompertz resultó interesante debido a que encuentra la edad máxima (a partir de este momento la $q_x = 1$) para la muestra de mortalidad mexicana 2000-I, la curva que describe la mortalidad estimada además, es una curva suave que describe la mortalidad sin cambios inesperados¹⁶.

Sin embargo, antes de aceptar esta prueba como la más adecuada para describir la mortalidad mexicana 2000-I, es necesario realizar un análisis de consistencia de la curva estimada con las observaciones. Por esta razón se realizó una prueba cuantil a cuantil que logró mostrar más a detalle la similitud y las diferencias entre la curva estimada y los datos observados.

Figura 2.3.3 Cuantil-Cuantil modelo de Gompertz vs observaciones aplicado a la mortalidad mexicana 2000-I.



Fuente: Elaboración propia.

¹⁶ Tabla de mortalidad mexicana 2000-I ajustada mediante el método de Gompertz, Anexo D.3

Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.

Aunque la curva de mortalidad ajustada presenta suavización en su estructura, es creciente y los parámetros ajustados a la curva de fuerza de mortalidad resultan atractivos para edades tempranas, al realizar el análisis cuantil a cuantil la curva ajustada a la mortalidad presentó mayor variabilidad a edades más avanzadas, por esta razón se puede concluir que no es el ajuste más adecuado para dar solución al objetivo principal. Sin embargo; para analizar la consistencia que tiene los parámetros estimados con las observaciones se tiene una base más sólida que permita tomar la decisión de aceptación o rechazo se realizó la prueba de “Kolmogorov-Smirnov” de la cual se obtuvieron los siguientes resultados:

Observaciones vs Modelo de Gompertz

$$D = 0.07058$$

$$p\text{-value} = 0.0001$$

Por lo tanto, esta prueba rechaza la idea de que q_x queda descrita adecuadamente por q_x modelo de Gompertz. Además en este modelo se puede observar que carece de coherencia con la realidad, debido a que no es factible pensar en personas de 144 años, debido a que se puede asegurar que nadie conoce una persona que haya sobrevivido a esta edad.

Por todas las razones anteriores el modelo de Gompertz se rechaza como una solución para encontrar la edad máxima de la curva de la tabla de mortalidad Mexicana 2000-I.

2.3 Modelo de Makeham¹⁷

Cuarenta y dos años después de la construcción del modelo de Gompertz, William Makeham creó un modelo demográfico con el propósito de hacer una mejora a la ley de Gompertz. El modelo de Makeham es preferentemente utilizado para edades adultas seniles, y tiene las siguientes restricciones: $-B \leq A$, $0 < B$, $1 < C$ y $0 \leq x$. Las funciones que describen el modelo son:

La función de densidad:

$$f(x) = BC^x(C-1)e^{-A-\frac{BC^x}{\ln(C)}(C-1)}$$

Con $B > 0$, $C > 1$ y $x \geq 0$.

La función de distribución está dada por:

$$F(x) = 1 - e^{-A-\frac{BC^x}{\ln(C)}(C-1)}$$

La función de supervivencia está dada por:

$$S(x) = e^{-A-\frac{BC^x}{\ln(C)}(C-1)}$$

¹⁷ Norström F.(1997). The Gompertz-Makeham distribution. Umeå University (London).

Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.

Finalmente la fuerza de mortalidad está dada por:

$$\mu(x) = A + BC^x$$

Esta función se integra por dos componentes, los cuales representan:

- **A** es el parámetro asociado al efecto de las causas de muerte independientes de la edad, por ejemplo, las epidemias, las guerras, eventos naturales, etc.
- **B** es un parámetro que es independiente a la edad del individuo, debido a que considera el deterioro natural de la gente en el tiempo.
- **C** es un parámetro que al combinarse con la edad, representara la resistencia que tiene el individuo a fallecer.

Debido a que se conoce la función de distribución se ajustaron los parámetros para describir la mortalidad mexicana 2000-I mediante el modelo de Makeham y de esta manera encontrar la edad máxima esperada.

El modelo se ajustó por mínimos cuadrados en una hoja de cálculo (Excel)¹⁸. Con las restricciones originales del modelo. Finalmente los parámetros que describen la curva de mortalidad de manera óptima, los cuales están dados por:

$$A = 0.000175228$$

$$B = 0.0001152283$$

$$C = 1.072372478$$

Al obtener el ajuste de los parámetros se continuó con la búsqueda de la edad máxima alcanzada, sin embargo al intentar resolver el problema de manera analítica se observó que:

$$q_x = 1 - e^{-A - \frac{BC^x}{\ln(C)}(C-1)}$$

Se busca entonces que la probabilidad de fallecimiento a edad x sea 1, es decir, debe ocurrir que:

$$e^{-A - \frac{BC^x}{\ln(C)}(C-1)} = 0$$

Al intentar obtener el valor de x (que será la edad máxima alcanzada), se observó que no es posible calcular el logaritmo de cero. Este problema puede ser resuelto con límites pero desafortunadamente el valor encontrado hace de la edad máxima esperada un valor que tiende a infinito.

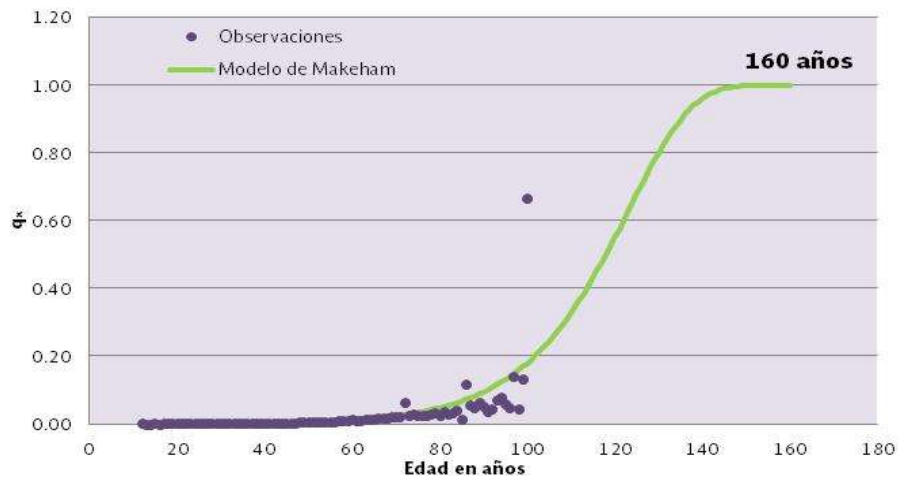
¹⁸ Si se desea ver el cálculo realizado ir al Anexo C.2

Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.

Sin embargo no es posible pensar en infinito como una solución a la curva de mortalidad mexicana 2000-I, debido a que todas las personas conocen que morirán en algún momento y justo el propósito de la presente investigación es terminar de manera adecuada la curva de mortalidad mexicana.

Se contemplaron entonces 5 decimales en el redondeo de q_x , lo que condujo a concluir la curva de mortalidad en edad 160 años, aquí la probabilidad ajustada de fallecer (q_x) es 1, es decir, se encontró el momento donde debe terminar la curva que describe la mortalidad mexicana 2000-I. En la figura 2.4.1 se puede observar el ajuste realizado mediante el modelo de Makeham y los datos observados de mortalidad mexicana 2000-I.

Figura 2.4.1 Modelo de Makeham aplicado a los datos base de mortalidad mexicana 2000-I¹⁹.



Fuente: Elaboración propia

Al ser un ajuste parecido al de Gompertz, se encuentran similitudes importantes debido a que la curva ajustada es una curva creciente y suavizada; sin embargo, al ser una curva asintótica la edad máxima variará de acuerdo a la elección de la persona que realice el ajuste porque él será el encargado de seleccionar la edad más adecuada de acuerdo a las necesidades a las que se enfrente.

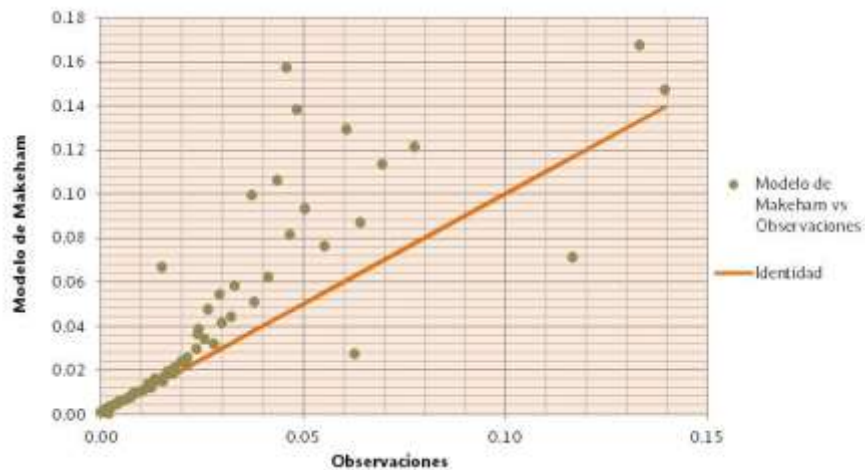
Aun cuando se obtuvo ya un ajuste de parámetros y un resultado que parece resolver la incógnita que propicia esta investigación se procede a analizar la consistencia de estos parámetros en el modelo que están representando, mediante un ajuste cuantil a cuantil.

Como ocurre con el modelo de Gompertz para edades tempranas es un modelo que logra describir adecuadamente los datos observados, en el gráfico (figura 4.2.4) se puede observar que los puntos más parecidos a la recta identidad en estas edades, (se sabe que son a edades tempranas, por ser una función creciente, es decir, a mayor edad mayor valor esperado de mortalidad); sin embargo, a mayor edad mayor es la dispersión que existe entre los datos provocando que el ajuste no sea atractivo y se obtenga un resultado contrario al buscado en la presente investigación.

¹⁹ La estimación de mortalidad mexicana 2000-I mediante el modelo de Makeham se encuentran en el Anexo D.4

Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.

Figura 2.4.2 Cuantil-Cuantil modelo de Makeham vs observaciones aplicado a la mortalidad mexicana 2000-I.



Fuente: Elaboración propia

Finalmente para asegurar que el modelo no es lo suficientemente adecuado para describir la mortalidad mexicana 2000-I se realizó también un ajuste de Kolmogorov-Smirnov.

Observaciones vs Modelo de Makeham

$D = 0.0698$

$p\text{-value} = 0.3763$

A pesar de que el test muestra un valor menor que el establecido para D , se rechaza el modelo de Makeham, al tener un $p\text{-value}$ mayor a 0.05 , lo que refleja la poca compatibilidad del modelo propuesto con la muestra de mortalidad mexicana 2000-I, además de proponer como edad máxima 160 años (aproximando a 5 decimales el resultado), debido a que en México la persona más longeva de la que se tiene registro es de 125 años (de la señora Leandra Becerra Lumbreras que nació 31-Agosto-1887, en Tamaulipas México).

Lo anterior provoca que se descarte la posibilidad de ser seleccionado como el ajuste óptimo que describe la mortalidad mexicana 2000-I y a la edad máxima estimada, Por esta razón se buscó implementar otro modelo que logre ajustarse más a las características buscadas.

2.4 Modelo de Heligman & Pollard²⁰

Durante mucho tiempo se han elaborado trabajos para desarrollar funciones para predecir los datos de mortalidad de una población en edades avanzadas. Entre estos trabajos se encuentra el realizado por Heligman y Pollard (1980), quienes crearon una función que se adapta a los cocientes de mortalidad en poblaciones ayudando así a cumplir el objetivo. Un punto que es importante resaltar es que todos los parámetros de dicha función son positivos y sólo se podrán aplicar para edades mayores a cero.

²⁰ Graça Magalhães M., Coelho E., Miguel Bravo (2006). M. MORTALITY AND LONGEVITY PROJECTIONS FOR THE OLDEST-OLD IN PORTUGAL. Statistics Portugal, University of Évora.

Aunque Heligman y Pollard realizaron varias versiones de su función, es importante destacar que todas ellas ofrecen una buena aproximación a la mortalidad en todas las edades. Esta función cumple con las características básicas de una q_x , es decir:

1. Toma valores entre cero y uno, ya que se trata de una función de probabilidad, es decir $0 \leq q_x \leq 1$.
2. Es válida para cualquier rango de edad en una tabla de mortalidad de la población.
3. Es una función continua.
4. Todos los parámetros que conforman la ecuación tienen una interpretación ya sea biológica o demográfica.
5. Tiene una flexibilidad que le permite ajustarse a una gran variedad de modelos de mortalidad.
6. Tiene muy pocos parámetros, con respecto a la cantidad de variantes que considera y el número de datos que se deben ajustar.

El desarrollo que realizaron Heligman y Pollard produjo algunas funciones que describen la mortalidad, cada una de ellas se obtuvo de la búsqueda de mejorar a la función anterior. Algunas de estas funciones son:

1. Primera Ley
 - a. Primera publicación:

$$\frac{q_x}{1 - q_x} = A^{(x+B)^C} + De^{-E[\ln(x) - \ln(F)]^2} + GH^{(x-x_0)}$$

La siguiente ecuación es equivalente a la anterior (Debón, Montes, & Sala, Graduación de datos de Mortalidad, 8-11 Abril 2003):

- b. Segunda publicación:

$$q_x = A^{(x+B)^C} + De^{-E[\ln(x) - \ln(F)]^2} + \frac{GH^x}{1 + GH^x}$$

2. Segunda Ley (buscando mejorar el ajuste en edades seniles²¹):

$$q_x = A^{(x+B)^C} + De^{-E[\ln(x) - \ln(F)]^2} + \frac{GH^x}{1 + KGH^x}$$

²¹ Edades mayores a 65 años.

3. Tercera Ley :

$$q_x = A(x+B)^C + De^{-E[\ln(x)-\ln(F)]^2} + \frac{GH^{x^K}}{1 + GH^{x^K}}$$

Esta función está compuesta por tres sumandos los cuales representan:

1. **Primer sumando:** Es una función exponencial decreciente. describe la mortalidad para edades menores o iguales a un año y muestra la capacidad del bebé para ganar inmunidad a las enfermedades y la forma en que se adapta a su entorno, por esta razón se refleja una caída. A su vez se expresa con tres parámetros:

Variable	¿Qué representa?	Intervalo de valores que toma la variable
A	Tiene un valor similar a la q_1 .	(0,1)
B	Localización de la mortalidad en el primer año de vida. Cuando $B=0$ y $q_0= 0.5$ conforme aumenta B, ceteris paribus ²² , mas próximo es q_0 de q_1	$(q_1,0.5)$ considerando que $q_1 < q_0$. En la práctica toma valores entre (0.01, 0.03)
C	Asociado al decrecimiento del índice de mortalidad para edades mayores a un años, a mayor C más rápido fue la adaptación del niño a su entorno	(0,1)

Fuente: Elaboración propia.

2. **Segundo sumando:** La función en esta parte es muy similar a la lognormal, ésta representa la mortalidad para edades mayores a un año y considera la llamada “joroba de los accidentes” (muertes por accidentes para hombres jóvenes y para el caso de las mujeres, la mortalidad materna).

Esta joroba introduce una mortalidad adicional y aparecerá como una curva que distinguirá a esta curva de mortalidad de otras curvas de mortalidad. La joroba se encuentra entre los quince y treinta años de edad, y está compuesta por tres variables, las cuales se describen a continuación:

Variable	¿Qué representa?	Intervalo de valores que toma la variable
D	Establece la severidad de la joroba. Efecto de los accidentes sobre la mortalidad.	(0,1)
E	Relacionada con la dispersión de las q_x . A mayor E menor dispersión de las q_x	$(0, \infty)$
F	Indica la localización de la joroba de los accidentes.	(15,30]

Fuente: Elaboración propia.

²² Permanece sin cambios, en las condiciones ajenas a la variable B.

Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.

3. **Tercer Sumando:** Es la ley de Gompertz. Refleja la mortalidad senil, es decir, el crecimiento geométrico de la mortalidad en edades altas, ésta tiene como base el deterioro del cuerpo con los años y está compuesta por las siguientes variables:

Variable	¿Qué representa?	Intervalo de valores que toma la variable
G	Toma valores de acuerdo a la mortalidad senil.	(0,1)
H	Tasa de crecimiento de la mortalidad senil	(0,10)
x_0	Es una edad cercana a la edad límite. Conocida como "span of life".	Valor de x_0 tal que $q_{x_0} = 0.5$

Fuente: Elaboración propia.

Es importante aclarar que los valores tomados por las variables sólo sirven para ajustar el modelo a los datos originales y no describen o explican las circunstancias biológicas o ambientales de la mortalidad.

Aunque existen variantes de la primera ecuación presentada por Heligman y Pollard, ellos consideran que la interpretación que se hizo anteriormente para los parámetros A,B,C,D,E,F,G,H, tendrá validez para cualquiera de las ecuaciones antes mencionadas. El valor de K modifica la dirección y extensión de la curvatura de mortalidad; sin embargo, esta variable es la única que actualmente no tiene una interpretación demográfica muy clara.

Para concluir la descripción de la ecuación, se realizó el ajuste las observaciones base de la tabla de mortalidad mexicana 2000-I²³.

La primera limitante encontrada surge a partir de que se carece de las observaciones que corresponden a los registros para edades menores a 12 años por lo tanto los parámetros que corresponden al primer sumando serán omitidos, es decir, se trabaja con la función:

$$\frac{q_x}{1 - q_x} = D e^{-E[\ln(x) - \ln(F)]^2} + GH^{(x-x_0)}$$

Los parámetros de la función se aproximaron mediante la función Solver con el código de optimización Simplex LP, desarrollando la teoría de mínimos cuadrados con ayuda de una hoja de cálculo (Excel).

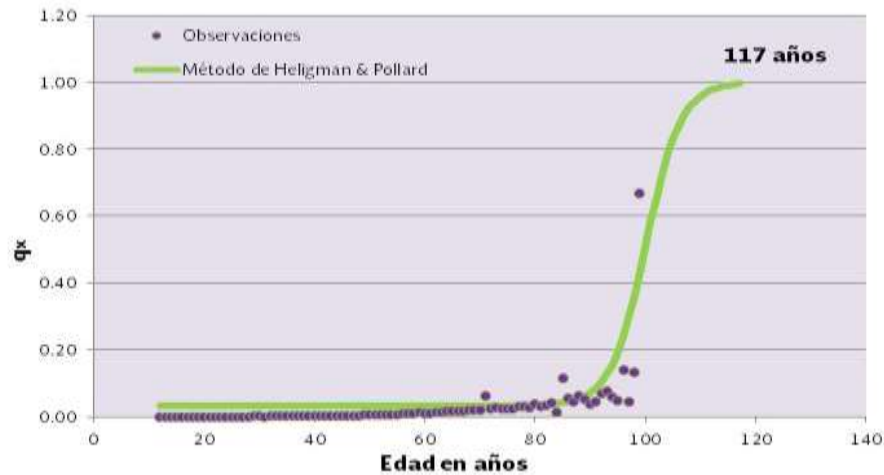
Al concluir el algoritmo se obtuvieron los siguientes parámetros para la ecuación 1.2, que queda representada en la Figura 2.5.1:

D	E	F	G	H
0.03479	0.00010	25.00298	0.99999	1.37

²³ La muestra de la tabla de mortalidad mexicana para 2000-I se encuentran en el Anexo D.1

Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.

Figura 2.5.1 Modelo de Heligman & Pollard aplicado a los datos base de mortalidad mexicana 2000-I.

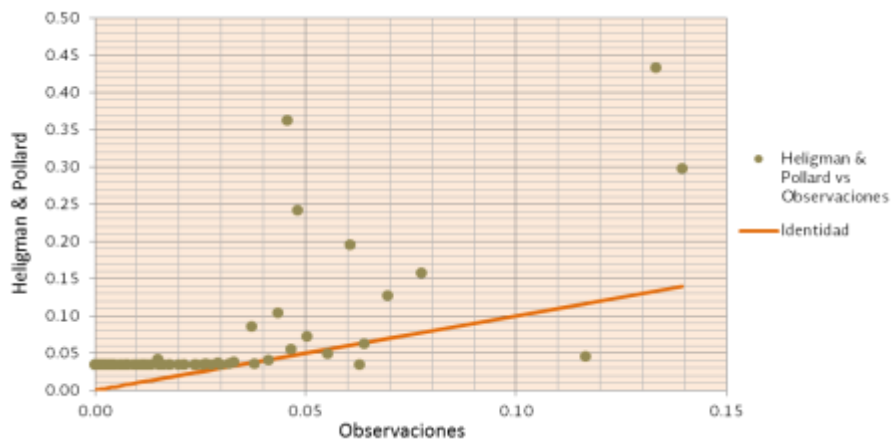


Fuente: Elaboración propia.

Mediante el modelo de Heligman y Pollard, se encontró un ajuste suavizado²⁴ así como la edad máxima de la curva de mortalidad mexicana 2000-I, donde resultó que a partir de 117 años $q_x = 1$, además respecto a los modelos anteriormente presentados es el modelo que tiene una finalización más aceptable, aun cuando es superado por un caso real.

Sin embargo, al tratarse de un modelo cuyo objetivo es el ajuste de la edad máxima alcanzada, se analizará la significancia que tiene que la curva no se ajuste de manera adecuada en la parte izquierda, además del comportamiento como función asintótica en la cola derecha de la curva así como la mayor variabilidad de los datos estimados con las observaciones para edades avanzadas.

Figura 2.5.2 Cuantil-Cuantil modelo de Heligman & Pollard vs observaciones aplicado a la mortalidad mexicana 2000-I.



Fuente: Elaboración propia.

El análisis realizado fue una prueba de cuantil a cuantil el cual se muestra en la Figura 2.5.2, ésta prueba permitió decidir que la metodología de Heligman & Pollard no es la mejor para

²⁴ Tabla de mortalidad mexicana 2000-I ajustada por el método de Heligman & Pollard se encuentra en el anexo D4

Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.

describir la mortalidad mexicana 2000-I. Además muestra también la variabilidad tan grande que tienen las observaciones de los datos estimados en edades de menores (en este caso de uno a once años).

Además resultó interesante el análisis de esta curva en todo el rango de edades debido a que al contrario de los modelos trabajados anteriormente este ajuste tiene carencias desde edades tempranas, es decir, tiene diferencias significativas para edades tempranas donde la probabilidad de fallecimiento es pequeña (menores de 12 años) y edades avanzadas (75 años en adelante).

También se realizó el test de Kolmogorov-Smirnov, se encontró lo siguiente:

Observados vs Heligman & Pollard
D = 0.31638
p-value = 0.838

Lo cual implica que la estimación mediante el modelo de Heligman & Pollard queda totalmente rechazada, al no cumplir la condición de tener un valor de $D < 0.20667$ y un $p\text{-value} > 0.05$, es imposible asegurar que la curva estimada describe de manera adecuada la mortalidad observada en la población mexicana 2000-I. Por este motivo se continúa con la aplicación de modelos que logren describir la mortalidad mexicana más precisa.

2.5 Modelo de Coale-Kisker²⁵

El método de Ansley Coale y Ellen Eliason Kisker, conocido como el método de Coale-Kisker ha sido implementado en varias ciudades desarrolladas del mundo, dando como resultado una buena aproximación del índice de mortalidad. Este algoritmo fue desarrollado en 1989 por Ansley Coale and Guang Guo y un año más tarde fue perfeccionado por Coale y Kisker.

Para aplicar este método se tendrán como base los siguientes supuestos:

- a) La ley de Gompertz supone que todos los individuos tienen una resistencia decreciente a morir por causas naturales en función de la edad, es decir, la mortalidad crece en forma constante conforme se incrementa la edad del individuo, y su crecimiento relativo es constante. Este método por el contrario supone que el crecimiento relativo $\frac{\mu'_x}{\mu_x}$ de la mortalidad no es constante.
- b) El modelo original supone la edad de 85 años como edad pivote, que es donde se presentaba la mayor variabilidad de los datos utilizados por Coale-Kisker. Sin embargo, en el caso de la base de los datos de la tabla de mortalidad mexicana es a partir de edad $80 \leq x$ donde se observa mayor variación e incremento en las tasas de mortalidad, razón por la cual se modificará a 80 años la hipótesis de Coale-Kisker.

²⁵ Graca Magallanes, M., & Coelho, E. (2006). MORTALITY AND LONGEVITY PROJECTIONS FOR THE OLDEST-OLD IN PORTUGAL. Portugal, Portugal, Portugal.

Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.

- c) El modelo original es aplicado a una muestra clasificada en hombres y mujeres: para hombres asigna $m_{110} = 1.0$ y para mujeres $m_{110} = 0.8$, dado que supone que dependiendo del género de la persona será la probabilidad de fallecimiento que se le debe asignar y debido al hecho de que hay muy pocos sobrevivientes a edades superiores a 110 años. Para el caso de la población mexicana 2000-I el índice se homogeneizará a $m_{105} = 1$ debido a que no se tiene una tabla de mortalidad para hombres y mujeres.

Se define $k(x)$ como:

$$k(x) = k(x-1) - R \text{ para } 80 \leq x \quad 2.1$$

Con $k(x) = \ln \left[\frac{m_x}{m_{x-1}} \right]$ y R una constante. Entonces se busca el valor de la constante R . Esto se hace mediante el algoritmo recursivo, que propone la ecuación 1.1, junto con la definición de $k(x)$.

De lo anterior se obtienen que R^{26} es:

$$R = \frac{26k(79) - \ln(m_{105}) + \ln(m_{79})}{231}$$

$$R = \frac{26 \ln \left(\frac{m_{79}}{m_{78}} \right) - \ln(m_{105}) + \ln(m_{79})}{231}$$

$$R = -0.004606359$$

Finalmente se podrán encontrar todos los valores de $k(x)$ de manera recursiva, a partir de la ecuación 2.1, como:

$$m_x = m_{x-1} e^{k(x)}$$

Además se conoce la igualdad 1.1, lo cual permite obtener el valor de q_x , como:

$$q_x = \frac{2m_{x-1}e^{k(x)}}{2 + m_{x-1}e^{k(x)}} \quad 2.2$$

Este resultado es particularmente importante debido a que las observaciones que se toman como base para la presente investigación generan el índice de mortalidad central a edad x (m_x) a partir de la cual se aproximan la probabilidad de fallecimiento a edad x (q_x), en este

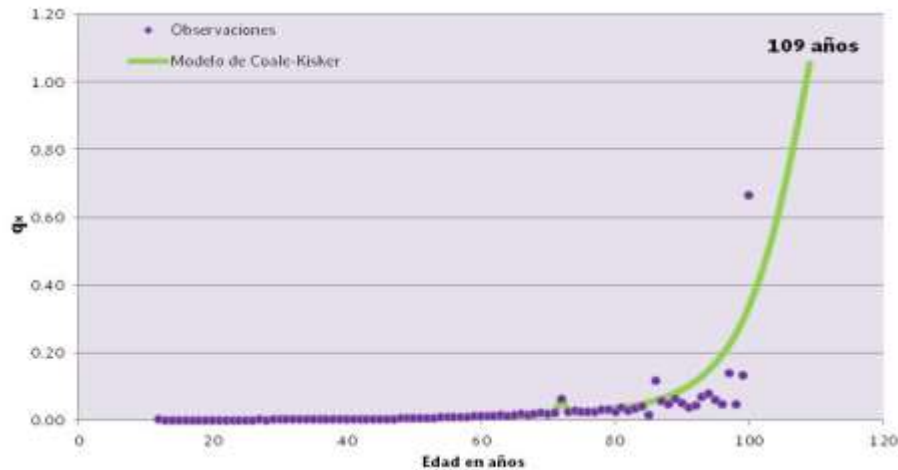
²⁶ El desarrollo del modelo elaborado por Coale-Kisker se encuentra en el Anexo A.4

Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.

caso se refleja el mismo efecto, que tiene el índice de mortalidad central sobre la probabilidad de fallecimiento.

La tabla de mortalidad mexicana 2000-I ajustada por el método de Coale-Kisker, se presenta en la Figura 2.6.1²⁷.

Figura 2.6.1 Modelo de Coale-Kisker aplicado a los datos base de mortalidad mexicana 2000-I.

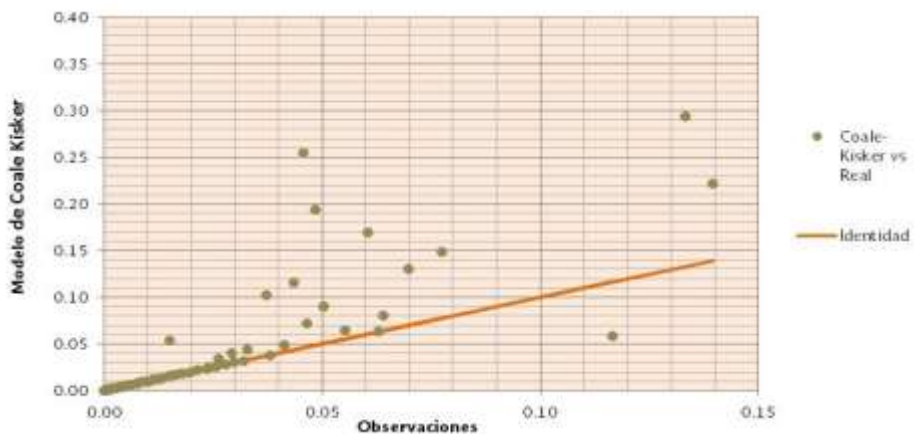


Fuente: Elaboración propia.

La conclusión anterior surgió a partir del supuesto $m_x = 1$, cuando $x = 105$ años, esto provocó que la edad máxima alcanzada por la población mexicana 2000-I sea de 109 años.

A simple vista parece ser un ajuste mejor que cualquiera de aplicados anteriormente; sin embargo, es importante realizar alguna prueba estadística que permita dar sustento a este modelo, por esta razón se realizó un análisis de cuantil a cuantil, el cual se puede observar en la Figura 2.6.2.

Figura 2.6.2 Cuantil-Cuantil modelo de Coale-Kisker vs observaciones aplicado a la mortalidad mexicana 2000-I.



Fuente: Elaboración propia.

²⁷ Tabla de mortalidad mexicana 2000-I ajustada mediante el modelo de Coale- Kisker, Anexo D6

Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.

El ajuste realizado con el modelo de Coale- Kisker presenta menor variabilidad con la recta identidad que cualquiera de los ajustes anteriores es decir, hasta el momento es el modelo que describe mejor la mortalidad mexicana 2000-I. Para realizar un análisis con más y mejores bases se procedió también a realizar el tests de Kolmogorov-Smirnov, el cual mostro los siguientes resultados:

Observados vs Coale & Kisker
D = 0.04391,
p-value = 0.8910

Esta prueba muestra un p-value más atractivo por ser más grande que 0.05, además cumple que (la distancia máxima entre las observaciones y la estimación) $D < 0.20667$, por este motivo no se rechaza el modelo de Coale-Kisker, es decir, se considera una posible para modelo para describir a q_x .

Antes de considerar que q_x es descrita de manera adecuada por $q'_{x-Coale-Kisker}$, se observó que la curva que se genera que depende de la elección (subjativa en ocasiones) de los siguientes supuestos:

- Primero el impacto que tiene el supuesto de la edad a la cual $m_x = 1$, pues el modelo de Coale-Kisker en ocasiones se vuelve un modelo que alcanza un máximo antes de que $q_x = 1$, lo anterior provocaría no cumplir el objetivo deseado, es decir, no encontrar la edad máxima estimada en caso de modificar algún supuesto.
- La elección de la edad en que $m_x = 1$, es subjativa debido a que no existe algún supuesto que induzca a pensar alguna edad en particular que se adecuara como edad optima del modelo.
- La edad x a partir de la cual la mortalidad presenta mayor variabilidad, debido a que será esta la edad en la que se comience a realizar el ajuste de mortalidad.

Por las razones anteriores se procedió a realizar algunas pruebas, modificando la edad (x) en la que $m_x = 1$, y también la edad que se consideró como la edad a partir de la cual comenzaba a presentarse mayor variabilidad en las probabilidades de fallecimiento con el propósito de observar la importancia de la elección de estos dos supuestos.

Las pruebas realizadas se observan en el Cuadro 1.1.1, cada una de ellas cuenta con los supuestos de los cuales se partió para modelar la mortalidad mexicana 2000-I y así concluir cual es la edad máxima estimada para la población mexicana.

Cuadro 1.1.1 Pruebas realizadas con modificación de supuestos aplicando el modelo de Coale-Kisker.

Gráficas	Supuestos y Edad Máxima Estimada
	<p>Supuestos: 1</p> <p>$x = 70$ años y $m_z = 1$ cuando $z = 112$ años.</p> <p>Ecuación que describe el modelo:</p> $R = \frac{43 k(69) - \ln(m_{112}) + \ln(m_{69})}{946} = 0.0036$ <p>Edad Máxima Estimada:</p> <p style="text-align: center;">Mayor a 118 años.</p>
	<p>Supuestos: 2</p> <p>$x = 75$ años y $m_z = 1$ cuando $z = 112$ años</p> <p>Ecuación que describe el modelo:</p> $R = \frac{38 k(74) - \ln(m_{112}) + \ln(m_{74})}{741} = 0.0037$ <p>Edad Máxima Estimada:</p> <p style="text-align: center;">Mayor a 124 años.</p>
	<p>Supuestos: 3</p> <p>$x = 80$ años y $m_z = 1$ cuando $z = 112$ años</p> <p>Ecuación que describe el modelo:</p> $R = \frac{33 k(79) - \ln(m_{112}) + \ln(m_{79})}{561} = 0.0021$ <p>Edad Máxima Estimada:</p> <p style="text-align: center;">117 años.</p>
	<p>Supuestos: 4</p> <p>$x = 85$ años y $m_z = 1$ cuando $z = 112$ años</p> <p>Ecuación que describe el modelo:</p> $R = \frac{28 k(84) - \ln(m_{112}) + \ln(m_{84})}{378} = 0.0085$ <p>Edad Máxima Estimada:</p> <p style="text-align: center;">Mayor a 114 años.</p>

Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.

	<p>Supuestos: $x = 70$ años y $m_z = 1$ cuando $z = 105$ años</p> <p>Ecuación que describe el modelo: $R = \frac{36k(69) - \ln(m_{105}) + \ln(m_{69})}{666} = 0.0034$</p> <p>Edad Máxima Estimada: Mayor a 125 años.</p>
	<p>Supuestos: $x = 75$ años y $m_z = 1$ cuando $z = 105$ años</p> <p>Ecuación que describe el modelo: $R = \frac{31k(74) - \ln(m_{105}) + \ln(m_{74})}{496} = 0.0032$</p> <p>Edad Máxima Estimada: 125 años.</p>
	<p>Supuestos: $x = 85$ años y $m_z = 1$ cuando $z = 105$ años</p> <p>Ecuación que describe el modelo: $R = \frac{21k(84) - \ln(m_{105}) + \ln(m_{79})}{231} = 0.0070$</p> <p>Edad Máxima Estimada: Mayor a 118 años.</p>

Fuente: Elaboración propia.

Las modificaciones que se realizaron mostraron las siguientes variaciones:

- No se alcanza en todas las curvas el momento en el que $q_x = 1$ (las curvas 1, 2, 4, 5 y 7), es decir, aun cuando la curva alcanza un máximo, este no toma el valor de uno, lo que provoca que se fracase con el objetivo inicial de esta investigación, ya que no se logró estimar la edad máxima para la población mexicana 2000-I.
- A pesar de que se encuentra la edad máxima estimada en las curvas 3 y 6 el momento a partir de la cual comienza un importante incremento en la variabilidad de los datos observados es 80 años y 75 años respectivamente, además la edad a la que $m_x = 1$, también resulto ser elegida de manera distinta 112 años y 105 años respectivamente.

Se puede concluir entonces que los resultados que se obtienen dependerán en gran medida de la edad x a la que se le asigna $m_x = 1$, así como de la edad a partir de la cual se considere que existe mayor variabilidad en las observaciones, lo que produce que el modelo tenga carencias complicadas de ajustar.

De las curvas generadas se seleccionó la que se encuentra en la Figura 2.6.2 debido a que se encontró una edad máxima para la tabla de mortalidad mexicana 2000-I más aceptable de manera intuitiva, con una curva suavizada; sin embargo, las curvas 3 y 6 del Cuadro 1.1.1 también cubren las características anteriores.

Debido a lo anterior se realizaron algunas pruebas de Kolmogorov-Smirnov, que permitieron seleccionar de manera más adecuada la curva que describe mejor la mortalidad mexicana 2000-I. Dichas pruebas están descritas en el Cuadro 1.1.2:

Cuadro 1.1.2 Contraste de pruebas Kolmogorov-Smirnov para edades máximas estimadas de la población mexicana 2000-I con el modelo de Coale-Kisker.

Edad Máxima Alcanzada	D	1-P-Value
109 años	0.1839	0.1054
117 años	0.0795	0.9435
125 años	0.1818	0.1090

Fuente: Elaboración propia.

La curva correspondiente a 109 años sobresale de las otras dos curvas (117 años y 125 años) por tener un p-value más alto, tener un valor de D bajo, además de estimar una edad máxima alcanzada más intuitiva, debido a que finaliza la curva de mortalidad en los 109 años.

Aun con la estimación que resultó más atractiva es importante continuar realizando algoritmos de ajuste que permitan no sólo la suavización de datos, sino encontrar el umbral de éstos (es decir la edad pivote a partir de la cual se comienza el análisis de los datos para determinar la edad máxima de la tabla de mortalidad), considerando el que sea más adecuado de acuerdo a la mortalidad mexicana 2000-I, debido a que este ajuste depende de ciertas hipótesis que son manipulables a criterio del responsable de ajustar la curva de mortalidad.

2.6 Modelo relacional de mortalidad²⁸

Christine L. Himes, Samuel H. Preston y Gretchen A. Condran (1994) analizaron la información de 82 poblaciones que presentaban mortalidad baja y consideraron que la mejor **manera de describir esa mortalidad, era creando una tabla de mortalidad “estándar”** para edades entre 45 y 99 años. Es importante mencionar que cada una de las 82 poblaciones examinadas, se buscó que existiera coherencia entre el número de muertes registradas entre un censo y otro y el número de personas vivas entre un censo y otro, de la misma cohorte.

El Modelo relacional construido asume que existe una relación entre la edad y la fuerza de mortalidad, que se denomina transformación lineal de fuerza. Este método presenta la característica de reducir el número de parámetros que lo conforman, lo cual lo vuelve atractivo respecto a otros modelos.

²⁸ M Mendoza Ramírez, M., Madrigal Gómez, A. M. & Martínez Torres, E., 2000 “Tablas de Mortalidad CNSF2000-I y CNSF 2000-G”, CNSF, 2000.

La construcción del modelo relacional no requiere una función en particular para que se represente la relación que existe entre la fuerza mortalidad y la edad, aunque la mayor parte del tiempo se sigue la transformación logit.

Se procede entonces a examinar detalladamente la fuerza de mortalidad, más allá de examinar las probabilidades de fallecimiento o supervivencia. Para el análisis de edades máximas generalmente se procede a linealizar la relación entre las tasas de mortalidad y la edad (Himes, C., H. Preston, S., & A. Condran, G., 1994).

2.6.1 Modelo Logit²⁹

La transformación logit, se caracteriza por estar entre los modelos de probabilidad (tomando valores en el intervalo (0,1)) no lineales y por ser una función continua en todo su dominio. Para el modelo Logit es necesario partir de las siguientes hipótesis:

1.- Considerando bajo el supuesto de independencia y homogeneidad, que el número de muertes d_x puede ser modelado como una distribución Binomial de probabilidad q_x , es decir:

$$P [d_x | E_x q_x] = \text{Binomial} (E_x, q_x)$$

Cómo se analiza un conjunto de muestras, se tiene lo siguiente:

$E_x^* = \sum_{i=1}^n E_{x_i}$, n es el número de muestras que se tienen, es decir, el total de expuestos al fallecer.

$d_x^* = \sum_{i=1}^n d_{x_i}$, n es el número de muestras que se tienen, es decir, el total de fallecimientos o su equivalente $q_x^* = \frac{d_x^*}{E_x^*}$.

Lo anterior debido a que se tienen E_x^* expuestos al riesgo y q_x^* la probabilidad que modelara cada uno de los resultados esperados. De las probabilidades de ocurrencia de fallecimiento, se obtiene el índice "odds" = $\frac{q_x}{1-q_x}$.

2.- El índice "odds" toma 3 posibles valores:

- a) $\frac{q_x}{1-q_x} = 1$, en este caso $q_x = 1/2$, lo cual significa que ambas alternativas tienen la misma probabilidad de ocurrencia.

²⁹ Medina Moral E., (2003), Modelos De Elección Discreta. España.

Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.

- b) $\frac{q_x}{1-q_x} < 1$, para esta situación se tiene que $0 \leq q_x < 1/2$, de aquí se concluye que la probabilidad de morir tiene menor probabilidad que la de sobrevivir en edad i .
- c) $\frac{q_x}{1-q_x} > 1$, para esta situación se tiene que $1 > q_x \geq 1/2$, de aquí se concluye que la probabilidad de fallecer es mayor que la probabilidad de sobrevivir.

3.- Se toman logaritmos del índice “odds”, para calcular el valor de la función logit, es decir:

$$\text{logit}(q_x) = \ln \left[\frac{q_x}{1 - q_x} \right]$$

La cual se puede expresar también como:

$$\text{logit}(q_x) = \ln \left[\frac{q_x}{1 - q_x} \right] = \alpha + \beta x$$

4.- La transformación inversa a la función logit, es la función antilogit, y es utilizada para regresar los valores probabilísticos a su estado natural.

$$e^{\text{logit}(q_x)} = \frac{q_x}{1 - q_x}$$

$$q_x = \frac{e^{\text{logit}(q_x)}}{1 + e^{\text{logit}(q_x)}}$$

$$q_x = \frac{e^{\alpha + \beta x}}{1 + e^{\alpha + \beta x}} \quad 2.6$$

Se resolvió el modelo de regresión logística y se estimaron los parámetros utilizando una hoja de cálculo (Excel)³⁰, desarrollando la teoría de mínimos cuadrados se pueden obtenerlos sin necesidad de procedimientos largos.

El valor de $q_{x-\text{modelo Logit}}$, estará descrita por los parámetros:

$$\alpha = -11.670418$$

$$\beta = 0.0991340$$

Es importante recordar que las probabilidades de fallecer entre edad 12 y 79 no sufrieron modificaciones, en la Figura 2.7.1 se puede observar la curva generada, para la ecuación 2.6.1. Finalmente se ha encontrado un ajuste para la cola de la curva de mortalidad mexicana 2000-I dando como resultado una $q_x = 1$ para $x = 172$ años, es decir se logró estimar la edad máxima que alcanza la población mexicana, que es el propósito original.

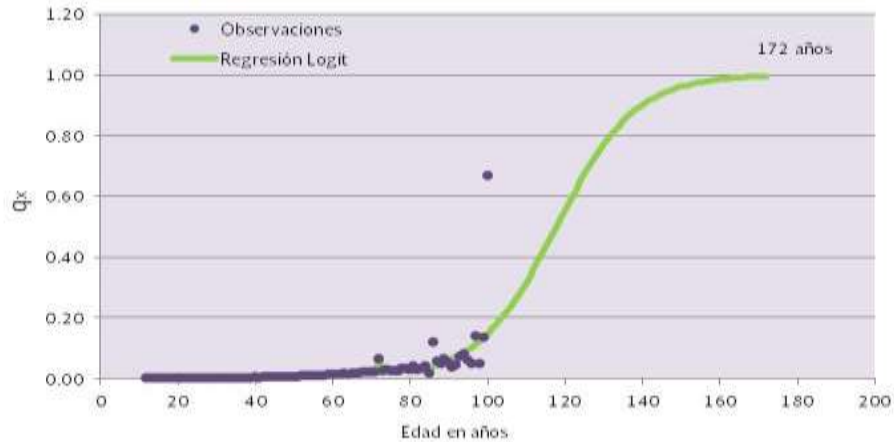
³⁰ Si se desea ver el desarrollo dirigirse al Anexo C.4

Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.

Este método aun cuando suaviza la curva de mortalidad mexicana 2000-I propone una edad máxima muy elevada, lo que sugiere realizar una prueba de bondad de ajuste para decidir si el modelo describe adecuadamente la realidad.

La prueba inicial que se decidió realizar fue una de cuantil a cuantil, este método por la construcción que presenta permite observar en la Figura 2.7.2 que ajusta de manera adecuada los datos centrales; sin embargo, olvida el ajuste de los datos de la cola de la distribución, lo cual se descarta como apoyo al propósito de esta investigación.³¹

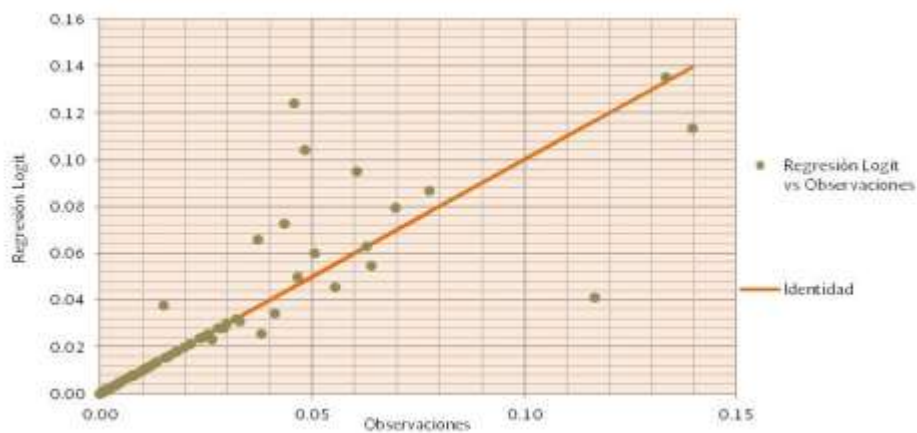
Figura 2.7.1 Modelo logit aplicado a los datos base de mortalidad mexicana 2000-I.



Fuente: Elaboración propia.

Finalmente como se había pensado antes de realizar la prueba cuantil a cuantil se puede observar que las diferencias entre observaciones con los datos estimados por la regresión logit para edades avanzadas es mayor, a pesar de que se ajusta más que cualquiera de los análisis anteriores a edades más tempranas, se descarta como un modelo óptimo para darle solución a encontrar la edad máxima alcanzada por la población mexicana 2000-I, al proponer que la edad máxima alcanzada por la población mexicana es de 172 años.

Figura 2.7.2 Cuantil-Cuantil regresión logit vs observaciones aplicado a la mortalidad mexicana 2000-I.



Fuente: Elaboración propia.

³¹ Tabla de mortalidad mexicana 2000-I ajustada mediante el modelo logit, se encuentra en el Anexo D.7

2.6.1.1 Tabla de mortalidad CNSF 2000-I. Un caso particular³²

En México gracias a la información del sector asegurador y la Comisión Nacional de Seguros y Fianzas, se han elaborado y desarrollado algoritmos para la construcción de las tablas de mortalidad mexicana.

Es importante mencionar que independientemente de la relevancia que tiene dentro del sector actuarial, el problema de ajustar una tabla de mortalidad de cualquier país o ciudad en el mundo es esencialmente, un algoritmo estadístico.

Básicamente el algoritmo realizado es el siguiente:

- Bajo el supuesto de independencia y homogeneidad, establece que el número de muertes d_x puede ser modelado como una distribución Binomial de probabilidad q_x , es decir:

$$P [d_x | E_x q_x] = \text{Binomial} (E_x, q_x)$$

Es decir, si se exponen E_x individuos al riesgo de fallecer, el propósito será pronosticar d_x^* la cantidad de fallecimientos o su equivalente $q_x^* = \frac{d_x^*}{E_x}$. El desarrollo de dicho método para el caso de Seguro de Vida individual es ajustar la muestra con un modelo de regresión logística dado por:

$$Y = \ln \left[\frac{q_x}{1 - q_x} \right] = \alpha + \beta x$$

Donde x es la única variable independiente del modelo. De lo anterior, se obtienen las modas de las distribuciones finales de los parámetros en el modelo:

$$\begin{aligned} \alpha &= -9.146130 \text{ (0.035469)} \\ \beta &= 0.074355 \text{ (0.000700)} \end{aligned}$$

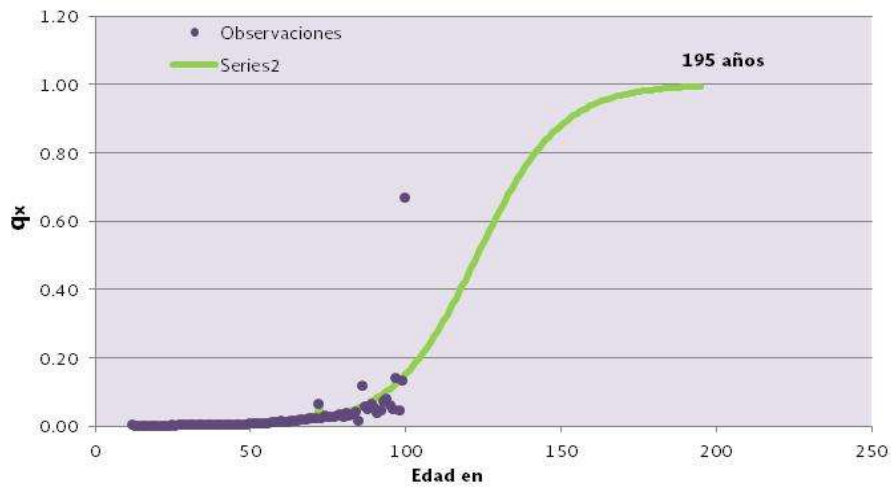
En la Figura 2.7.1 se muestra el ajuste logit con los parámetros estimados por la Comisión Nacional de Seguros y Fianzas, que actualmente rigen a la tabla de mortalidad mexicana, que concluyen en edad 100 años.

El ajuste que emite la CNSF sólo lo realiza hasta edad 99, debido a que se considera que nadie rebasa edad 100 años, es decir se define $q_{100} = 1$. Sin embargo, al trazar el ajuste sugerido para la muestra (Figura 2.7.3), encuentra que la curva alcanza una edad máxima para la población mexicana 2000-I muy elevada (195 años), lo que lleva a pensar que este ajuste no refleja las características de mortalidad de la población mexicana de la manera más adecuada. Además al estimar que la edad máxima alcanzada supera 125 años, que es el registro de la persona más longeva en México, se descarta como solución al objetivo original de la presente investigación.

³² M Mendoza Ramírez, M., Madrigal Gómez, A. M. & Martínez Torres, E., 2000 "Tablas de Mortalidad CNSF2000-I y CNSF 2000-G", CNSF, 2000.

Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.

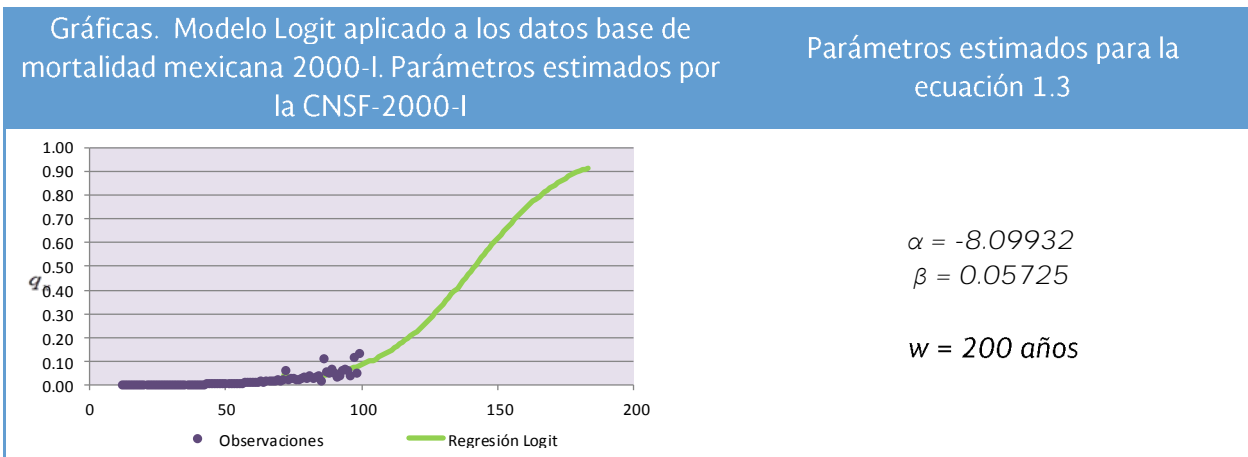
Figura 2.7.3 Modelo logit aplicado a los datos base de mortalidad mexicana 2000-I. Parámetros estimados por la CNSF-2000-I³³.



Fuente: Elaboración propia.

Lo importante del método, es que permite realizar simulaciones estocásticas³⁴, dando lugar a resultados más precisos. En el Cuadro 1.2 se muestran algunas de las simulaciones realizadas³⁵, los parámetros que ajustan y suavizan la curva que describe la mortalidad mexicana CNSF-2000-I, así como la edad máxima estimada mediante este algoritmo.

Cuadro 1.2 Simulaciones estocásticas de la mortalidad mexicana 2000-I.

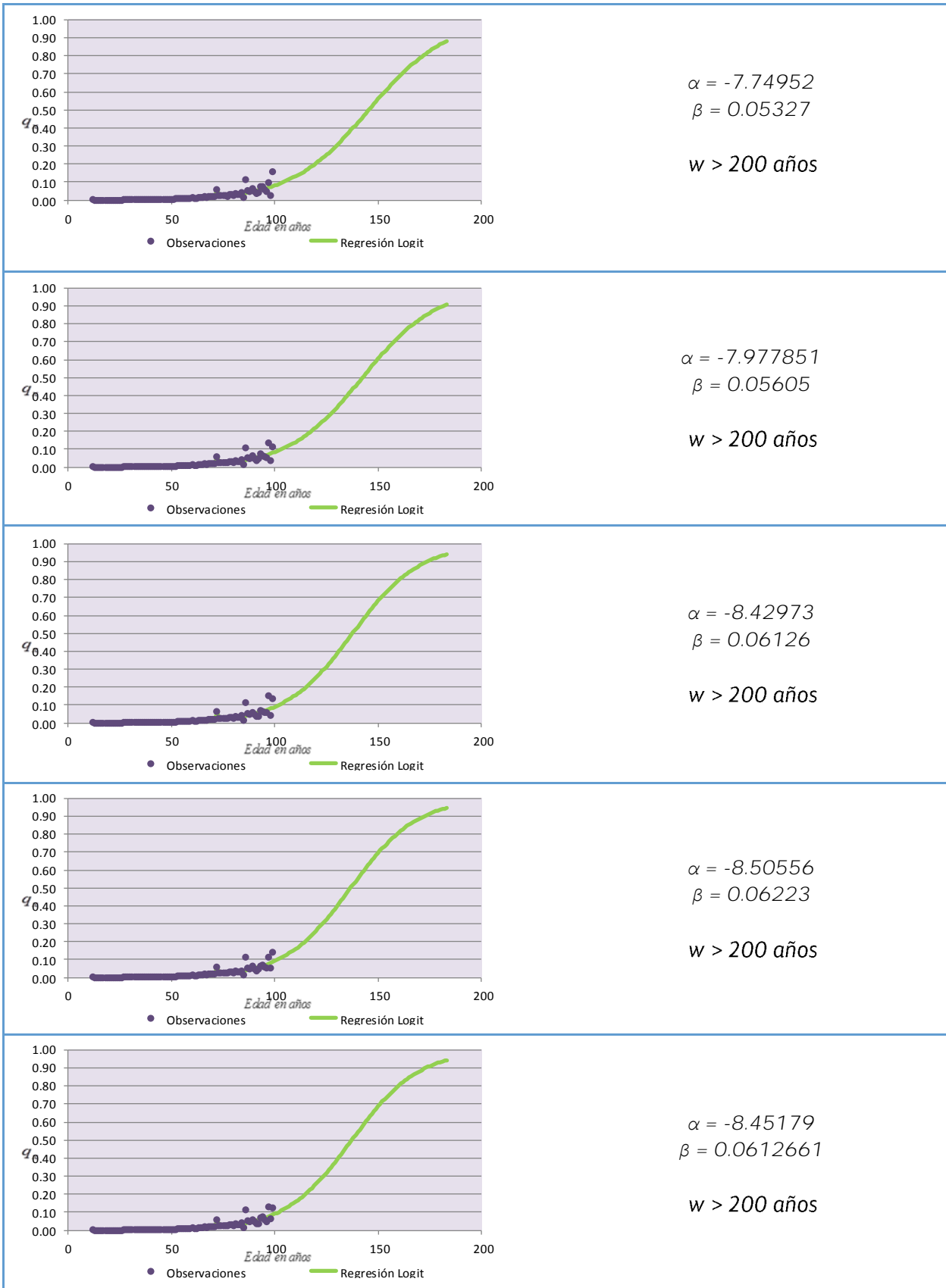


³³ Tabla de mortalidad mexicana 2000-I ajustada mediante el Modelo logit por la CNSF se encuentra en el anexo D.8

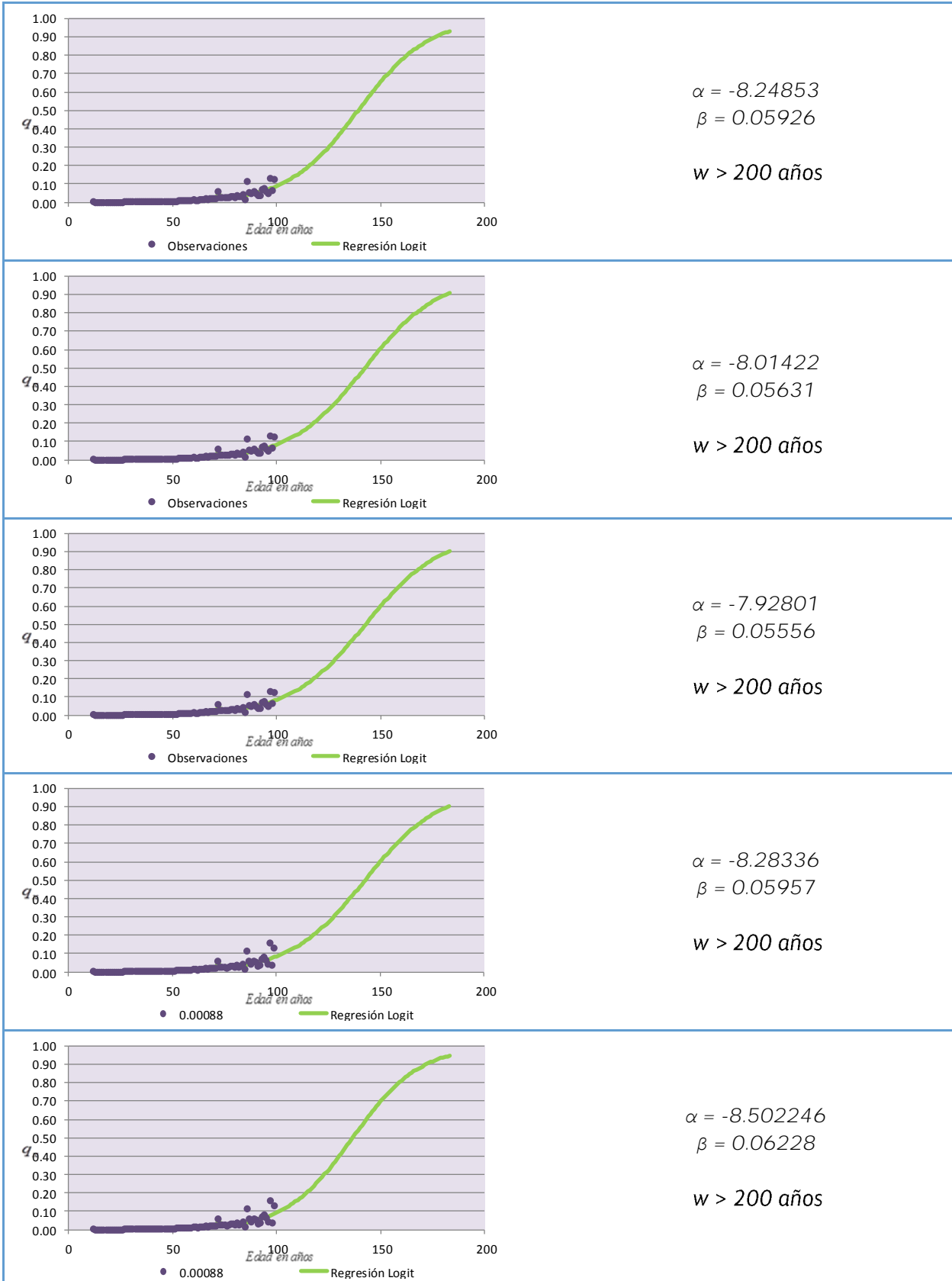
³⁴ Las simulaciones se explican más detalladamente en el Anexo A.5

³⁵ El código que se utilizó para realizar las simulaciones de la población se encuentra en el Anexo A.5

Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.



Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.



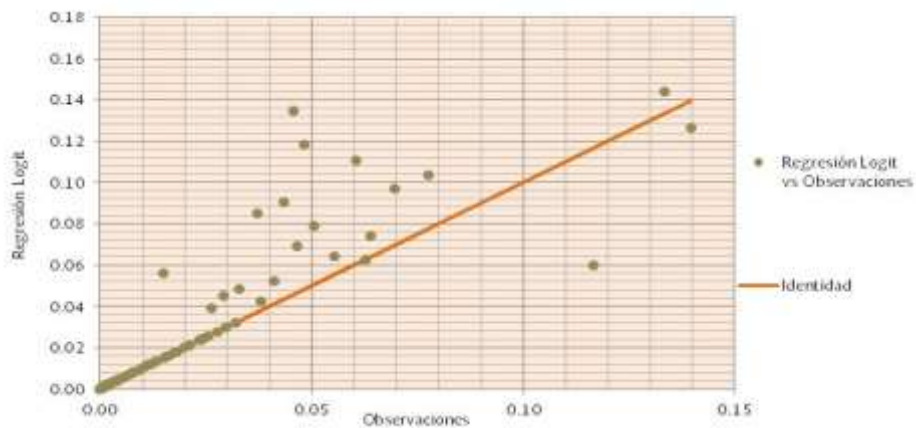
Fuente: Elaboración propia

Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.

Las simulaciones realizadas permitieron encontrar funciones que suavizan la curva que describe la mortalidad mexicana; sin embargo, estas curvas son asintóticas y que superan los 200 años para la edad máxima estimada, por esta razón aun cuando no se realizaron pruebas de bondad de ajuste se eliminaron como una estimación óptima de la mortalidad mexicana 2000-I.

Finalmente se realizó un análisis cuantil a cuantil, el cual permite observar que a mayores edades se incrementará la dispersión de los datos observados, con lo que se puede concluir que no es la metodología más óptima para edades avanzadas, justo el objetivo de dicha investigación.

Figura 2.7.2 Cuantil-Cuantil regresión logit (CNSF) vs observaciones aplicado a la mortalidad mexicana 2000-I.



Fuente: Elaboración propia

La información en el extremo de la edad (de 90 y más), tiene una dispersión muy importante debida a la falta de información (asegurados en esas edades), lo cual hace que las extrapolaciones a la derecha de esas edades de baja calidad, incluso el hacer uso de las estadísticas de INEGI o de CONAPO en las que se maneja toda la población y consideran muchos más casos en estas edades extremas, existen también, edades para las que la variación entre una edad y otra será muy alta por los pocos elementos de la muestra en dichas edades.

Capítulo 3

Aplicación del Teorema del Valor Extremo al cálculo de la edad máxima esperada de la Tabla de Mortalidad Mexicana CNSF 2000-I

Siempre que se realiza un análisis estadístico de datos, se tiene la inquietud de realizar un ajuste de los mismos que ayuden a entender el comportamiento en general y así poder trabajar con ellos, es decir, se busca crear un modelo que se asemeje a dicha información. Desde el punto de vista estadístico, el problema de ajuste de la cola de una distribución es un problema de extrapolación, debido a que se pretende encontrar un modelo paramétrico para la cola del proceso que genera los datos y luego ajustar este modelo a las observaciones extremas. La importancia del modelo dependerá de qué también describe el comportamiento de la cola para la distribución de los datos observados.

Generalmente el análisis estadístico se concentra en masas y no en la ocurrencia de ciertos sucesos atípicos (extremos) debido a la poca importancia aparente que representan en el conjunto de datos. La idea general de esto es eliminar la información atípica y facilitar el ajuste de alguna distribución de probabilidad, al convertirse en una curva más suave y uniforme que finalmente podrá informar mejor de la tendencia del conjunto, más allá de la ocurrencia de sucesos extremos.

Desafortunadamente esta idea presenta una carencia importante, pues está rechazando la idea de la existencia de valores con baja probabilidad de ocurrencia pero con alto impacto o incluso la existencia de eventos que exceden cierto umbral, por esta razón no se debe asegurar que un modelo de estadística clásica describirá precisamente los eventos que son atípicos y extremos porque son generalmente ignorados. Considerando esto, se han desarrollado metodologías que intentan dar solución al problema descrito, entre ellas la Teoría del Valor Extremo.

Hay quienes definen la teoría de valores extremos como una disciplina única por desarrollar técnicas y modelos para describir sucesos poco comunes, otros la definen como la teoría que describe el máximo y el mínimo y esta definición engloba atributos tales como “excepcional” y “catastrófico”.

Como se puede observar resulta complicado encontrar una definición adecuada para esta teoría pues engloba la subjetividad de quien la defina, se ha decidido caracterizar la teoría de valores extremos por sus propiedades estadísticas, predictibilidad, mecanismos, observaciones, etc.

3.1 Teoría de Extremos³⁶

Las últimas décadas se ha desarrollado la teoría y las aplicaciones de los modelos extremos, en varias áreas del conocimiento como: la ingeniería, las ciencias ambientales, las finanzas y

³⁶ Cebrián Guajardo, A. (2011). Análisis, modelización y predicción de episodios de sequía. España: Universidad de Zaragoza, Tesis Doctoral.

Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.

por supuesto la actuaría, estos nuevos modelos se han vuelto muy populares y son más sofisticados en cuanto a la matemática ocupada. Actualmente existe una rama de la estadística llamada Teoría de Extremos (también conocida como Teoría los Valores Extremos) que se debe al desarrollo de los trabajos de Von Mises, Gnedenko o Fisher y Tippet (1928), los cuales permitieron la creación de una distribución asintótica para modelar los mínimos o en su debido caso los máximos de una variable aleatoria (por ejemplo la edad máxima alcanzada). El objetivo central de la teoría de extremos es desarrollar procedimientos, estadísticamente justificables, para estimar la cola de una distribución desconocida, F , a partir de una muestra de datos.

Los principales problemas que se pretende realizar la estimación de valores extremos son:

- La escasez de datos en la cola de la distribución; los datos extremos son por definición deficientes, por lo que el número de observaciones disponibles es, generalmente, muy pequeño, para la mortalidad mexicana 2000-I por ejemplo, se carece de suficiente información para edades avanzadas.
- La necesidad de extrapolar la inferencia a situaciones no observadas, ya que las estimaciones de interés corresponden a valores que sobrepasan los máximos de la muestra, por ejemplo en la mortalidad mexicana 2000-I, se pretende encontrar la edad máxima que no se encuentra en la muestra, seguramente.
- La imposibilidad de usar la mayor parte de las técnicas de estimación estándar, que suelen producir estimadores de las colas muy sesgadas, como anteriormente se ha observado, el uso de herramientas estándar produce resultados no satisfactorios para el ajuste de mortalidad mexicana 2000-I.

Los principales modelos paramétricos utilizados en la Teoría de Extremo son:

- Distribución Generalizada del Valor Extremo.
- Distribución Generalizada de Pareto.

3.1.1 Distribución Generalizada de Extremos (GEVD)³⁷

En particular, el análisis de los valores extremos se concentra en eventos que tienen poca frecuencia al ser observados. Los métodos estadísticos para evaluar eventos extremos necesitan poder modelar adecuadamente la cola de la distribución que se analiza. La Teoría de Extremos (TVE) es la que se encarga de este análisis. La TVE no solo genera modelos para la muestra que se está trabajando, también puede ser utilizada para extrapolar la probabilidad de un evento aún más extremo (ver e.g. Coles, 2001).

Así en este apartado se entenderá por Distribución Generalizada del Valor extremo (GEVD) a:

“Es la distribución que ajusta bloques de observaciones adecuadamente normalizados que surgió a partir del desarrollo de la Teoría del valor Extremo (Fisher y Tippet 1928). Según esté sea un conjunto de variables aleatorias

³⁷ García Perez, A. (2004). La Teoría Del Valor Extremo: Una Aplicación Al Sector Asegurador. Castilla: VII congreso Hispano-Italiano de Matemáticas Financieras y Actuariales.

Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.

*independientes e idénticamente distribuidas con una función de distribución común $F(x)$, donde M_x representa el máximo de todas ellas. Así la cuestión a la que se da respuesta con esta distribución es **obtener la distribución del máximo.**"*

El modelo asintótico para el máximo de la función está sustentado en el siguiente teorema:

Teorema 3.1.2.I Sea $X_1, X_2, X_3, \dots, X_n$ una secuencia de variables aleatorias independientes e idénticamente distribuidas, con una función de distribución F común y sea:

$$M_n = \max(X_1, X_2, X_3, \dots, X_n)$$

Si se denota un término arbitrario en la secuencia X_i , por X , y si se supone que F satisface la que si existen secuencias de constantes $\{a_n > 0\}$ tales que:

$$P \left[\frac{M_n - b_n}{a_n} \right] \rightarrow G(z) \text{ cuando } n \rightarrow \infty$$

Del desarrollo asintótico de la expresión anterior aparecen tres tipos de distribución límite:

- TIPO 1: Distribución de Gumbel- colas medias
- TIPO 11: Distribución de Frechet - cola gruesas
- TIPO 111: Distribución de Weibull - colas suaves

Esta aproximación es fundamental, ya que la distribución límite $G(x)$ siempre pertenece a una de estas tres distribuciones sea cual sea la distribución original de los datos $F(x)$, es decir, la distribución asintótica de los máximos se puede estimar sin realizar suposiciones acerca de la naturaleza de la distribución original de las observaciones, generalmente desconocida.

3.1.2 Distribución Generalizada de Pareto. (GPD)

La segunda distribución importante en la teoría de extremos es la distribución generalizada de Pareto ésta es importante en los modelos de umbrales y en la teoría de extremos debido a que es una distribución límite de las excedencias de umbrales.

Este modelo tiene una justificación asintótica siempre que se aplique a valores que superan un umbral lo suficientemente grande, así como para defectos por debajo de un umbral lo suficientemente pequeño. En 1994 con el Teorema de Balkema-de Haan- Pickands³⁸ (también conocido como el segundo Teorema del Valor Extremo), se demostró que bajo ciertas condiciones, la función que describe el exceso sobre un umbral lo suficientemente grande, se puede aproximar con un Distribución Generalizada de Pareto.

El teorema considera una función de distribución desconocida F , de una variable aleatoria x , que excede un umbral N lo suficientemente grande, esta función es llamada también como función de distribución condicional de exceso y es definida como:

³⁸ Balkema, A. Haan, L. Pickands J.

Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.

$$P[X > N + y | X > N] = \frac{1-F(N+y)}{1-F(N)} \quad \text{para } 0 < y$$

Si se conoce la distribución F , entonces la función de distribución condicional de exceso también es conocida. Sin embargo, desafortunadamente la mayor parte del tiempo esta distribución es desconocida y por esta razón generalmente se recurre a realizar aproximaciones para los valores que exceden el umbral. Complementando el teorema 3.2.1:

Para una función de distribución no degenerada G , entonces G es un miembro de la familia de distribuciones generalizada de extremos (GEV). Entonces para un n grande:

$$P[M_n \leq x] \approx H(x)$$

Donde:

$$H(x) = e^{\left\{-\left[1+\gamma\left[\frac{x-N}{\theta}\right]\right]^{-1/\gamma}\right\}}$$

Para algún $\theta > 0$ y $\gamma > 0$. Entonces para algún N , lo suficientemente grande, la función de distribución de $X-N$, condicionado a que $X > N$, es aproximadamente:

$$G(x) = 1 - \left[1 + \gamma \left[\frac{x - N}{\theta}\right]\right]^{-1/\gamma} \quad 3.1$$

Definida para los $N \leq x$ y $1 + \gamma \left[\frac{x-N}{\theta}\right] > 0$. Donde N es el parámetro de localización, θ es el parámetro de escala y γ es el parámetro de forma³⁹.

La familia de curvas que fueron definidas en la ecuación 1.5 recibe el nombre de familia Generalizada de Pareto. Su función de densidad está dada por:

$$g(x) = \frac{1}{\theta \left[1 + \gamma \left[\frac{x - N}{\theta}\right]\right]^{1+\frac{1}{\gamma}}}$$

3.2 Especificaciones del modelo

La elección de un umbral N lo “suficientemente” grande o pequeño es una cuestión subjetiva que depende del criterio y los propósitos de la persona que esté realizando el análisis de los datos, además la elección de un umbral diferente producirá estimaciones diferentes de los

³⁹ Para conocer más acerca de la distribución generalizada de Pareto ver el Anexo B.1

parámetros de la distribución (Hernández G., A. 2001) que se ajuste sin que por eso podamos asegurar que una es mejor que otra.

De acuerdo a las observaciones que se tienen para el caso de la mortalidad mexicana CNSF 2000-I, se puede elegir entre alguno de los siguientes escenarios:

- Un umbral lo suficientemente pequeño que permita trabajar con un mayor número de observaciones tal que la inferencia realizada proporcione información más ajustada a los datos observados.
- Un umbral lo suficientemente grande, que reduzca el sesgo del modelo, aunque esto implique que el ajuste incrementará la varianza de los estimadores.

Sin embargo, lo anterior tiene la desventaja de que N será un número fijo y arbitrario, lo que no garantiza que se haya elegido el umbral adecuado para el análisis que se desee hacer.

Recientemente se han propuesto modelos que mezclan diferentes funciones de distribución a los datos observados con el propósito de mezclar el centro con los valores extremos de la distribución y así evitar la elección de un umbral fijo. Estos modelos consideran al umbral de la distribución como un parámetro más, que finalmente deberá ser estimado al igual que los otros parámetros de la distribución que se quiere ajustar, y de esta manera solucionar la elección arbitraria del umbral.

Asumiendo que la función de supervivencia antes de la edad umbral sigue una distribución de Gompertz y la distribución que excede la edad umbral sigue una distribución de Pareto Generalizada, de acuerdo con el teorema de Balkema-de Haan- Pickands, ésta quedará de la siguiente manera:

$x \leq N$

$$F(x) = 1 - e^{-\frac{BC^x}{\ln(C)}(C-1)}$$

$x > N$

$$F(x) = 1 - S(N) \left[1 + \gamma \left[\frac{x - N}{\theta} \right] \right]^{-\frac{1}{\gamma}}$$

Donde N es la edad umbral de la función. Para asegurar que se esté realizando el ajuste de una función de distribución de probabilidad (i.e. que cumpla con los axiomas de Kolmogorov) se requiere que los parámetros cumplan: $B > 0$, $C > 1$ y $\theta > 0$.

3.2.1 Estimación de parámetros

Para la estimación de los parámetros de la función que describe la edad máxima esperada, se puede utilizar alguno de los siguientes métodos:

- Máxima verosimilitud
- Mínimos cuadrados

La edad umbral será elegida tal que la cola (la parte que es descrita por la función Generalizada de Pareto) y el cuerpo (la parte que es descrita por la función de Gompertz) sean consistentes entre sí.

Se inicia con la hipótesis de que la edad máxima de la tabla de mortalidad CNSF 2000-I debiera superar los 100 años de edad considerando que el umbral para nuestros cálculos se encuentra después de los 65 años de edad, debido a que el crecimiento de la función comienza a ser mayor a partir de este momento.

3.2.1.1 Máxima verosimilitud

El método de máxima verosimilitud, es uno de los más utilizados en estadística para la estimación de parámetros de funciones de probabilidad.

Aplicando este concepto al propósito inicial, se afirma que la función de máxima verosimilitud para N puede ser escrita como:

$$L(B, C, \gamma, \theta; N) = \left[\prod_{x=65}^{99} \left(\frac{S(x) - S(x+1)}{S(65)} \right)^{d_x} \right] \left(\frac{S(100)}{S(65)} \right)^{l_{100}}$$

Las razones que justifican la expresión anterior son las siguientes:

1. La razón por la cual se presenta la expresión $\left(\frac{S(x) - S(x+1)}{S(65)} \right)^{d_x}$, es debido a que expresa la contribución de probabilidad por intervalo de un año, considerando sólo a los que llegaron a la edad de 65 años, siendo elevado a la d_x por representar el número de fallecimientos durante este periodo.

De lo anterior:

$$\sum_{x=65}^{99} d_x \ln \left[\frac{S(x) - S(x+1)}{S(65)} \right] + l_{100} \left[\ln \left[\frac{S(100)}{S(65)} \right] \right] \text{ para } 65 < N < 99 \text{ y además } N \in Z$$

$$\sum_{x=65}^{N-1} d_x \ln \left[\frac{S(x) - S(x+1)}{S(65)} \right] + \sum_{x=N}^{99} d_x \ln \left[\frac{S(x) - S(x+1)}{S(65)} \right] + l_{100} \left[\ln \left[\frac{S(100)}{S(65)} \right] \right] =$$

Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.

$$\sum_{x=65}^{N-1} d_x \ln \left[\frac{S(x)-S(x+1)}{S(65)} \right] + l_N \left[\ln \left[\frac{S(N)}{S(65)} \right] \right] + \sum_{x=N}^{99} d_x \ln \left[\frac{S(x)-S(x+1)}{S(65)} \right] - l_N \left[\ln \left[\frac{S(N)}{S(65)} \right] \right] + l_{100} \left[\ln \left[\frac{S(100)}{S(65)} \right] \right] =$$

La idea de la separación en dos sumandos de la expresión original surge a partir del Teorema de Balkema–de Haan–Pickands, el cual combina dos funciones de distribución que describirán de mejor manera el comportamiento de la mortalidad. La expresión anterior está compuesta por dos sumandos:

El primer sumando que es igual a:

$$l_1 = \sum_{x=65}^{N-1} d_x \ln \left[\frac{S(x) - S(x + 1)}{S(65)} \right] + l_N \left[\ln \left[\frac{S(N)}{S(65)} \right] \right]$$

$$= \sum_{x=65}^{N-1} d_x \ln(S(x) - S(x + 1)) + l_N(S(N)) - l_{65}(S(65))$$

Y el segundo:

$$l_2 = \sum_{x=N}^{99} d_x \ln \left[\frac{S(x) - S(x + 1)}{S(65)} \right] - l_N \left[\ln \left[\frac{S(N)}{S(65)} \right] \right] + l_{100} \left[\ln \left[\frac{S(100)}{S(65)} \right] \right]$$

$$= \sum_{x=N}^{99} d_x \ln \left[\frac{S(x) - S(x + 1)}{S(N)} \right] + l_{100} \left[\ln \left[\frac{S(100)}{S(N)} \right] \right]$$

Tal que el proceso para la estimación de parámetros pueda ser resumido en el siguiente algoritmo⁴⁰:

- Para $N=98$,
 - Encontrar los valores de B y C que maximizan l_1 .

$$l_1 = \sum_{x=65}^{N-1} d_x \ln(S(x) - S(x + 1)) + l_N(S(N)) - l_{65}(S(65))$$

- Encontrar los valores de γ y θ que maximizan l_2 .

$$l_2 = \sum_{x=N}^{99} d_x \ln \left[\frac{S(x) - S(x + 1)}{S(N)} \right] + l_{100} \left[\ln \left[\frac{S(100)}{S(N)} \right] \right]$$

⁴⁰ El desarrollo de la igualdad se encuentra en el Anexo B.2

Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.

- Calcular el valor del perfil de log-verosimilitud l_p . Que involucra a toda la función.
- Repetir el paso anterior para $N= 97, 96, 95, \dots, 85$
- Se encuentra entonces el valor de N a partir del cual se encuentra la máxima verosimilitud.

Una vez determinado el valor de máxima verosimilitud de N , este determinará los estimadores óptimos de las funciones de distribución ajustadas (Gompertz y Pareto Generalizada).

Para el desarrollo de este algoritmo se requieren los valores de d_x y l_x , estos valores pueden ser calculados a partir de la selección de un radix y los valores de q_x de la tabla de mortalidad. En el ejercicio podría elegirse una cohorte de 1,000,000 para el radix, es decir, $l_0=10,000,000$, y a partir de este valor se pueden encontrar $d_x = l_x q_x$ y $l_{x+1} = l_x - d_x$, no hay que perder de vista que independientemente del radix, los parámetros que describen la mortalidad (B, C, γ, θ) serán los mismos.

3.2.1.2 Mínimos cuadrados

Consiste en minimizar la suma de los cuadrados de las diferencias entre los valores de la muestra y los valores estimados por el modelo sugerido, esto con el propósito de asegurar que las distancias sean las mínimas.

Para el caso de la mortalidad mexicana CNSF-2000-I tenemos la existencia de índices que facilitan la estimación de parámetros mediante el método de mínimos cuadrados. Dado que se tiene que la fuerza de mortalidad está dada por:

$$x \leq N$$

$$\mu(x) = BC^x$$

$$x > N$$

$$\mu(x) = \frac{F'(x)}{1 - F'(x)} = \frac{S(N) \left[\frac{1}{\gamma} \right] \left[\frac{\gamma}{\theta} \right] \left(1 + \gamma \left(\frac{x - N}{\theta} \right) \right)^{\frac{1}{\gamma} - 1}}{S(N) \left(1 + \gamma \left(\frac{x - N}{\theta} \right) \right)^{\frac{1}{\gamma}}} = \frac{1}{\theta \left(1 + \gamma \left(\frac{x - N}{\theta} \right) \right)}$$

Se puede abordar la estimación de manera más sencilla teniendo presente las siguientes igualdades:

$$x \leq N$$

$$\mu(x) = BC^x$$

$$\ln(\mu(x)) = \ln(B) + x \ln(C)$$

$$x > N$$

$$\ln(\mu(x)) = -\ln(\theta) - \ln \left[1 + \gamma \left(\frac{x - N}{\theta} \right) \right]$$

Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.

Lo que se pretende ahora es minimizar la diferencia entre los valores reales y los estimados así que se proponen las siguientes sumas para ser minimizadas:

$$\begin{aligned} \text{Suma de Gompertz} &= \sum_{x=12}^N E_x [\ln(\mu(x)) - \ln(B) + x \ln(C)] \\ \text{Suma de Pareto} &= \sum_{x=N+1}^{100} E_x \left[\ln(\mu(x)) + \ln(\theta) + \ln \left(1 + \gamma \frac{(x-N)}{\theta} \right) \right]^2 \end{aligned}$$

$$\text{Suma Óptima} = \text{Suma de Pareto} + \text{Suma de Gompertz.}$$

El proceso para realizar la estimación puede ser resumido en un algoritmo más simple:

- Para $N=98$,
 - Encontrar los valores de B y C que minimizan la *Suma de Gompertz*.
 - Encontrar los valores de γ y θ que minimizan la *Suma de Pareto*.
 - Calcular el valor que minimiza la *Suma Óptima*.
- Repetir el paso anterior para $N= 97, 96, \dots, 85$.
- Se encuentra el valor de N a partir del cual se minimiza la función.

El valor de N para el cual la suma de la diferencia de cuadrados es mínima, define los estimadores óptimos de la función.

En la presente investigación se realizó la estimación de parámetros sólo con la metodología de mínimos cuadrados sin embargo mediante ambos métodos los resultados son muy similares. Aquí se presenta la implementación del algoritmo para el caso en el que $N=90$ (edad umbral), debido a que se observó que la mayor variabilidad de datos se situaba entre las edades 85 y 100 años⁴¹. La descripción del proceso se encuentra en una hoja de cálculo (Excel)⁴².

Los resultados (para $N=90$) obtenidos fueron los siguientes:

Parámetro	Valor	Representación en el modelo general (DGP)
B	0.0010000	Parámetro de Gompertz
C	1.0435086	Parámetro de Gompertz
Γ	- 1.6501000	Parámetro Pareto Generalizada
Θ	18.178376	Parámetro Pareto Generalizada

De acuerdo a los parámetros θ , γ , B , C y N (antes encontrados) la probabilidad de fallecimiento de la población mexicana CNSF-2000-I puede ser expresada con la siguiente función:

Para $x \leq N$ (aplicado para $N=90$)

$$q_x = 1 - e^{-\int_0^1 BC^{(x+s)} ds}$$

⁴¹ El resultado de la simulación $N=85$ se encuentran en el Anexo B.3.

⁴² Si se desea ver el desarrollo dirigirse al Anexo C.5

Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.

$$q_x = 1 - e^{-\frac{BC^x}{\ln(C)}(C-1)}$$

$$q_x = 1 - e^{-(1.0435086)^x(0.0010216)}$$

$x > N$ (aplicado para $N=90$)

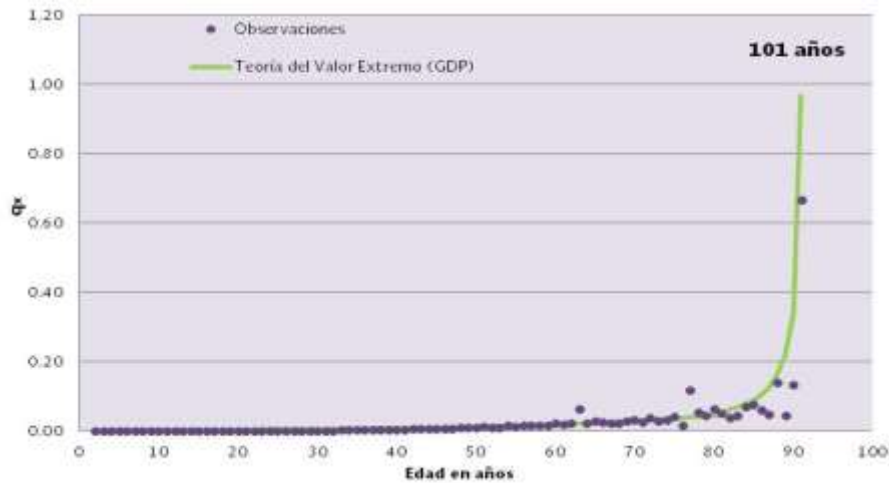
$$q_x = 1 - e^{-\int_{x-N}^{x-N+1} \frac{1}{\theta + \gamma s} ds}$$

$$q_x = 1 - e^{-\frac{1}{\gamma} [\ln(\theta + \gamma(x-N+1)) - \ln(\theta + \gamma(x-N))]}$$

$$q_x = 1 - e^{\frac{1}{1.652} [\ln(18.1783 - 1.625(x-90+1)) - \ln(1.783 - 1.625(x-90))]}$$

Con lo anterior se puede realizar el llenado de la columna G que representa la probabilidad de fallecimiento estimada. Con los resultados obtenidos se generó la siguiente gráfica que muestra el ajuste por la teoría del valor extremo usando en particular la distribución Pareto Generalizada, para la mortalidad mexicana CNSF-2000-I.

Figura 3.1.1 Teoría de extremos (GPD) aplicada a los datos base de mortalidad mexicana CNSF-2000-I⁴³.



Fuente: Elaboración propia

Esta teoría permitió realizar la estimación de la edad máxima de la población mexicana 2000-I suavizando el umbral encontrado y dando consistencia a la curva que se generó, además logró representar la mortalidad mexicana 2000-I de manera precisa.

Sin embargo, antes de concluir que la estimación logra describir de manera adecuada la mortalidad de la población mexicana es necesario realizar pruebas de bondad de ajuste que respalden estas observaciones o en su debido caso se descarte con sustentos matemáticos el resultado.

⁴³ Tabla de mortalidad mexicana CNSF-2000-I ajustada mediante la teoría de extremos se encuentra en el Anexo D9,

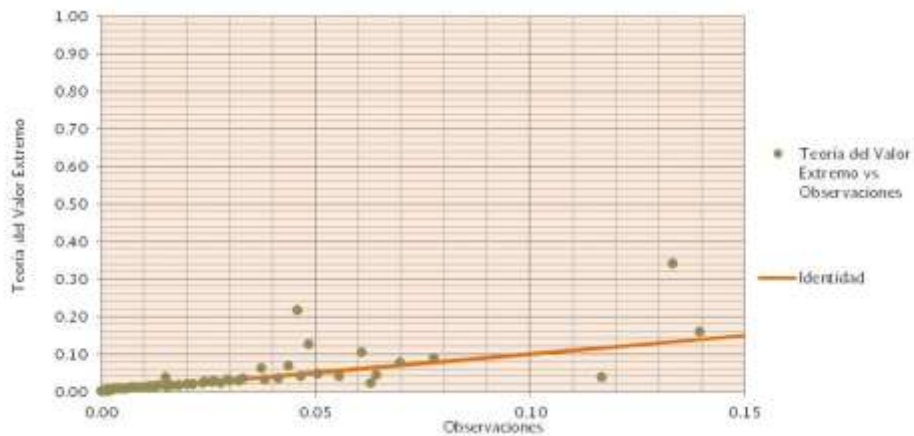
Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.

Para comenzar este análisis se realizó una prueba cuantil a cuantil en la Figura 3.1.2 se puede observar el resultado. Esta gráfica muestra que la variabilidad de las observaciones con las estimaciones mediante la metodología de la teoría del valor extremo es menor que cualquiera de los casos anteriores, por lo que se procede a realizar el test de Kolmogorov-Smirnov, con el propósito de comparar la estimación obtenida con los modelos anteriores de manera analítica.

Es importante resaltar que la curva debe analizarse en dos partes debido a que se construyó con dos distribuciones de probabilidad:

- La distribución de Gompertz
- La distribución Pareto Generalizada.

Figura 3.1.2 Teoría de extremos (GPD) vs observaciones aplicada a los datos base de mortalidad mexicana CNSF-2000-I



Fuente: Elaboración propia

Para la curva estimada con la distribución de Gompertz:

Observados & EVT
 $D = 0.2064$
 $p\text{-value} = 0.9997$

Para el caso de la distribución de EVT D debe ser menor a 0.2067 lo cual se cumple, además p -value es muy alto, por lo tanto no se rechaza la idea de que q_x queda bien descrita por q'_{x-EVT} . Incluso se consideró que es la mejor estimación realizada por ser un resultado que contrasta de manera adecuada con la realidad.

Capítulo 4

Importancia de la edad máxima en la Tabla de Mortalidad Mexicana CNSF-2000-I

Las tablas de mortalidad se caracterizan por finalizar en el momento en que todos los individuos han fallecido (es decir, la edad x a partir de la cual $q_x = 1$), así que la diferencia entre los diferentes métodos que se utilizan es básicamente la rapidez con la que estas alcanzan el final. Las tablas de mortalidad no restringen a la población pues éstas se pueden realizar:

- Por sexo hombres y/o mujeres.
- Por enfermedades importantes.
- Por lugar de residencia, etc.

La manera más simple de generar estas tablas es a partir de la mortalidad que se presenta por edad debido a que los resultados son atractivos por ser valores que ayudan a medir la sobrevivencia y la esperanza de vida.

Las tablas de mortalidad son una base fundamental en los temas de:

1. Salud pública:

Se utilizan para conocer la importancia de alguna causa de fallecimiento. Se realiza primero la construcción de la tabla de mortalidad que registre la mortalidad sin restricción y después se realiza una tabla que no considera la causa que se está analizando, de esta manera se logra conocer el impacto que tiene la causa de fallecimiento. También se utiliza para conocer la efectividad de algún tratamiento ante alguna patología importante.

2. Seguridad Pública:

En México particularmente en el IMSS (Instituto Mexicano del Seguro Social) y el ISSSTE (Instituto de Seguridad y Servicios Sociales de los Trabajadores del Estado) se utilizan para el cálculo de pensiones y jubilaciones. Así a mayor mortalidad, los recursos que deben almacenarse para hacer frente a las obligaciones que se generarán por concepto de pensiones y jubilaciones serán menores, en caso contrario, los niveles de ahorro por las instituciones son mayores.

3. Demografía:

Se utilizan para conocer el crecimiento de la población, la fecundidad, la longevidad, etc. De esta manera se pueden realizar proyecciones de la población.

4. Economía:

Las tasas de mortalidad que se presentan, son combinadas con datos demográficos esto permite construir un modelo más complejo que permite medir el efecto de la mortalidad y los cambios que ésta sufre bajo transformaciones socioeconómicas, para que la idea sea más clara basta mencionar que a mayor nivel socioeconómico del individuo, la esperanza de vida que tiene será mayor.

Las tablas de mortalidad también tienen importancia actuarial, se presenta una explicación más profunda de este caso y la aplicación de la investigación que se desarrolló.

4.1 Importancia en aplicaciones actuariales

Debido a que muchos eventos son inevitables, como la muerte, como actuario se deben tomar las medidas necesarias para minimizar el impacto financiero que esto produce. Por esta razón se realizó el análisis para la estimación de la edad máxima alcanzada por la población mexicana 2000-I lo que permite un cálculo adecuado de rentas vitalicias, seguros de vida, rentas fraccionadas, planes de pensiones, etc, estos se irán explicando en los siguientes apartados, con el propósito de observar las ventajas y desventajas que presente un ajuste de mortalidad.

4.1.1 Anualidades Contingentes

Una anualidad es una serie de pagos realizados de manera continua o en intervalos iguales (como quincenas, meses, bimestres, etc.) mientras se sobrevive. Clasificadas respecto a la duración en tres tipos:

- Anualidades Temporales⁴⁴.
- Anualidades Vitalicias.
- Anualidades Diferidas.

Como consecuencia de lo anterior se pueden definir los factores que intervienen en el cálculo de las anualidades, estos factores son los siguientes:

- Factor financiero. Aquí se valuarán los capitales en el tiempo, en donde intervendrá una tasa de interés.
- Factor de aleatoriedad. Este factor está determinado por las probabilidades de supervivencia del individuo.

De lo anterior la importancia de un buen ajuste de la curva que describe la mortalidad mexicana 2000-I (cálculo de las probabilidades de fallecimiento) y en particular para la edad máxima estimada en las Anualidades Vitalicias (y como caso particular las pensiones).

4.1.1.1 Plan de pensiones. Un caso particular de las anualidades

Un plan de pensiones según Bowers (1997) puede ser considerado como un sistema de compra de anualidades vitalicias diferidas (pagaderas durante el retiro), mediante alguna forma de anualidades temporales (Durante el servicio activo del individuo realiza las aportaciones).

⁴⁴ Para conocer el significado de cada una de las anualidades dirigirse al apartado 1.2 del capítulo 1.

De acuerdo con Lecina (1989), un plan de pensiones es una operación actuarial en la que una serie de individuos pertenecientes a un grupo aportan una prima durante una parte de su vida laboral, con el fin de recibir una cantidad a partir de su jubilación. Los siguientes elementos son básicos, para el cálculo de pensiones:

- La función de supervivencia, debido a que no todos los individuos asegurados alcanzarán a recibir pagos y algunos otros no los recibirán todo el período que comprenda la anualidad.
- Interés técnico, que debe ser mayor a la inflación.
- El porcentaje anual de crecimiento del salario.
- Edad de jubilación. En México generalmente son 65 años.

La información anterior es requerida con el objetivo de conocer el impacto que tendrá la mortalidad en la cuestión económica (los costos y beneficios con los que la pensión será otorgada), dado que a mayor sobrevivencia se requerirán mayor cantidad de recursos para poder garantizar la pensión vitalicia.

Para que sea más sencillo observar la importancia que tiene una buena estimación de la probabilidad de fallecimiento de la población mexicana 2000-I en el tema de anualidades y las pensiones como un caso particular, se realizó un ejemplo de anualidad vitalicia vencida tomado los siguientes supuestos:

- Tasa de interés anual del 15%.
- Se pagaran \$100,000 pesos al final del año de sobrevivencia, de manera vitalicia.
- Edad del asegurado 85 años.
- Se realizaron tres cálculos de esta anualidad para la curva de mortalidad estimada por la CNSF, el ajuste del teorema del valor extremo y el ajuste de Coale Kisker.

Para la metodología de la CNSF se obtuvo el siguiente resultado:

$$\ddot{a}_{85} = \$100,000 * \sum_{r=0}^{100-85-1} V^r {}_r p_{85}^{(CNSF)} = \$100,000 * \sum_{r=0}^{14} V^r {}_r p_{85}^{(CNSF)}$$

$$a_{85} = \$448,494.05$$

- Se deberán ahorrar \$448,494.05 pesos para poder obtener una anualidad de \$100,000.00 pesos de manera vitalicia a partir del fin del año en que se tiene edad 85, Donde la edad máxima esperada es de 100 años.

Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.

$$\ddot{a}_{85} = \$100,000 * \sum_{r=0}^{101-85-1} V^r {}_r p_{85}^{(TVE)} = \$100,000 * \sum_{r=0}^{15} V^r {}_r p_{85}^{(TVE)}$$

$$a_{85} = \$547,626.77$$

- Se deberán ahorrar \$547,626.77 pesos para poder obtener una anualidad de \$100,000.00 pesos de manera vitalicia a partir del fin del año en que se tiene edad 85. Donde la edad máxima esperada es de 101 años.

Finalmente para el modelo de Coale-Kisker se obtuvo:

$$\ddot{a}_{85} = \$100,000 * \sum_{r=0}^{109-85-1} V^r {}_r p_{85}^{(CK)} = \$100,000 * \sum_{r=0}^{23} V^r {}_r p_{85}^{(CK)}$$

$$a_{85} = \$492,497.16$$

- Se deberán ahorrar \$492,497.16 pesos para poder obtener una anualidad de \$100,000.00 pesos de manera vitalicia a partir del fin del año en que se tiene edad 85. Donde la edad máxima esperada es de 109 años.

Intuitivamente no es un resultado esperado, debido a que entre más tiempo viva una persona se espera que mayor será el número de pagos que habrá que realizar a su favor, sin embargo es importante observar la manera en que se comporta la curva de mortalidad de la población que se observe.

Ahora, si una compañía decide adoptar la estimación de la CNSF como la mejor estimación de la curva de mortalidad mexicana 2000-I y solicita una pago de \$192,982.27 pesos para dar a cada beneficiario \$100,000.00 peso cada año hasta su muerte y la curva que mejor describe la mortalidad mexicana 2000-I es la curva obtenida por medio la metodología del valor extremo, la compañía habrá perdido \$26,087 pesos por cada beneficiario, de ahí la preocupación de realizar una estimación lo más adecuada posible.

4.1.2 Seguros

Los sistemas de seguros se establecieron con el objetivo de reducir el impacto financiero negativo que tienen algunos eventos aleatorios como la invalidez, el fallecimiento, etc. Así que un seguro de vida está diseñado para reducir el efecto financiero del evento aleatorio la edad de fallecimiento.

Al igual que las anualidades, se definen los factores que intervienen en el cálculo de seguros, estos son:

- Factor financiero. Aquí se valuarán los capitales en el tiempo, en donde intervendrá una tasa de interés.

Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.

- Factor de aleatoriedad. Este factor está determinado por las probabilidades de supervivencia del individuo, debido a que la aseguradora debe calcular la prima que cobrará al asegurado, para que ésta le permita reservar lo necesario para hacer frente a sus obligaciones dentro del periodo establecido.

Así como las anualidades, existen varios tipos de seguro, por mencionar algunos de estos se tienen⁴⁵:

- Seguros temporales.
- Seguros vitalicios

Que pueden ser pagados en varias modalidades, dos de estas son:

- Pagados con Prima Única.
- Pagados con Prima Neta Nivelada por n-periodos.

Para ver el impacto que tiene una estimación no adecuada de la mortalidad mexicana 2000-I, se realizó un ejercicio con los siguientes supuestos, para un seguro temporal 10 años:

- Tasa de interés anual del 20%
- Monto deseado \$1,000,000.
- Edad del asegurado 85 años.
- Pago de primas netas niveladas por 3 años.

Para la metodología de la CNSF se obtuvo el siguiente resultado:

$$PNU = A_{85} = \$ 415,007.76$$

$$PNN = \frac{A_{85}}{\ddot{a}_{85:3}} = \frac{\$415,007.76}{2.43} = \$ 170,813.73$$

- Se deberán pagar tres primas anuales por \$170,813.73 pesos para obtener un seguro vitalicio en caso de fallecimiento por \$1'000,000.00 pesos de dado que el individuo tiene edad 85, Donde la edad máxima esperada es de 100 años.

Para el caso del Teorema del Valor Extremo (TVE), se obtuvo:

$$PNU = A_{85} = \$ 285,704.21$$

$$PNN = \frac{A_{85}}{\ddot{a}_{85:3}} = \frac{\$253,554.87}{2.54} = \$ 112,638.94$$

⁴⁵ Para conocer la notación y significado de los seguros mencionados dirigirse al apartado 1.2, del capítulo 1

Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.

- Se deberán pagar tres primas anuales por \$112,638.94 pesos para obtener un seguro vitalicio en caso de fallecimiento por \$1'000,000.00 pesos de dado que el individuo tiene edad 85, Donde la edad máxima esperada es de 101 años.

Finalmente para el modelo de Coale-Kisker se obtuvo:

$$PNU = A_{85} = \$ 357,612.41$$

$$PNN = \frac{A_{85}}{\ddot{a}_{85:3}} = \frac{\$357,612.41}{2.50} = \$ 143,174.94$$

- Se deberán pagar tres primas anuales por \$135,585.59 pesos para obtener un seguro vitalicio en caso de fallecimiento por \$1'000,000.00 pesos de dado que el individuo tiene edad 85, Donde la edad máxima esperada es de 109 años.

En este caso resulta importante observar que para un seguro temporal el impacto sería negativo en caso de que la compañía estuviera realizando el cálculo de primas con la curva obtenida mediante el teorema del valor extremo y la curva de mortalidad estimada que mejor describe a las observaciones de mortalidad mexicana sea la curva calculada por la CNSF, debido a que se cobrarían \$58,174.79 pesos menos por cada pago de prima única.

Conclusiones

Una tabla de mortalidad o supervivencia es una guía para realizar análisis temporales y comparaciones especiales sobre la incidencia de algún fenómeno con independencia de la estructura por edad de las poblaciones estudiadas.

Las tablas de mortalidad tienen múltiples aplicaciones en diferentes áreas, por ejemplo en:

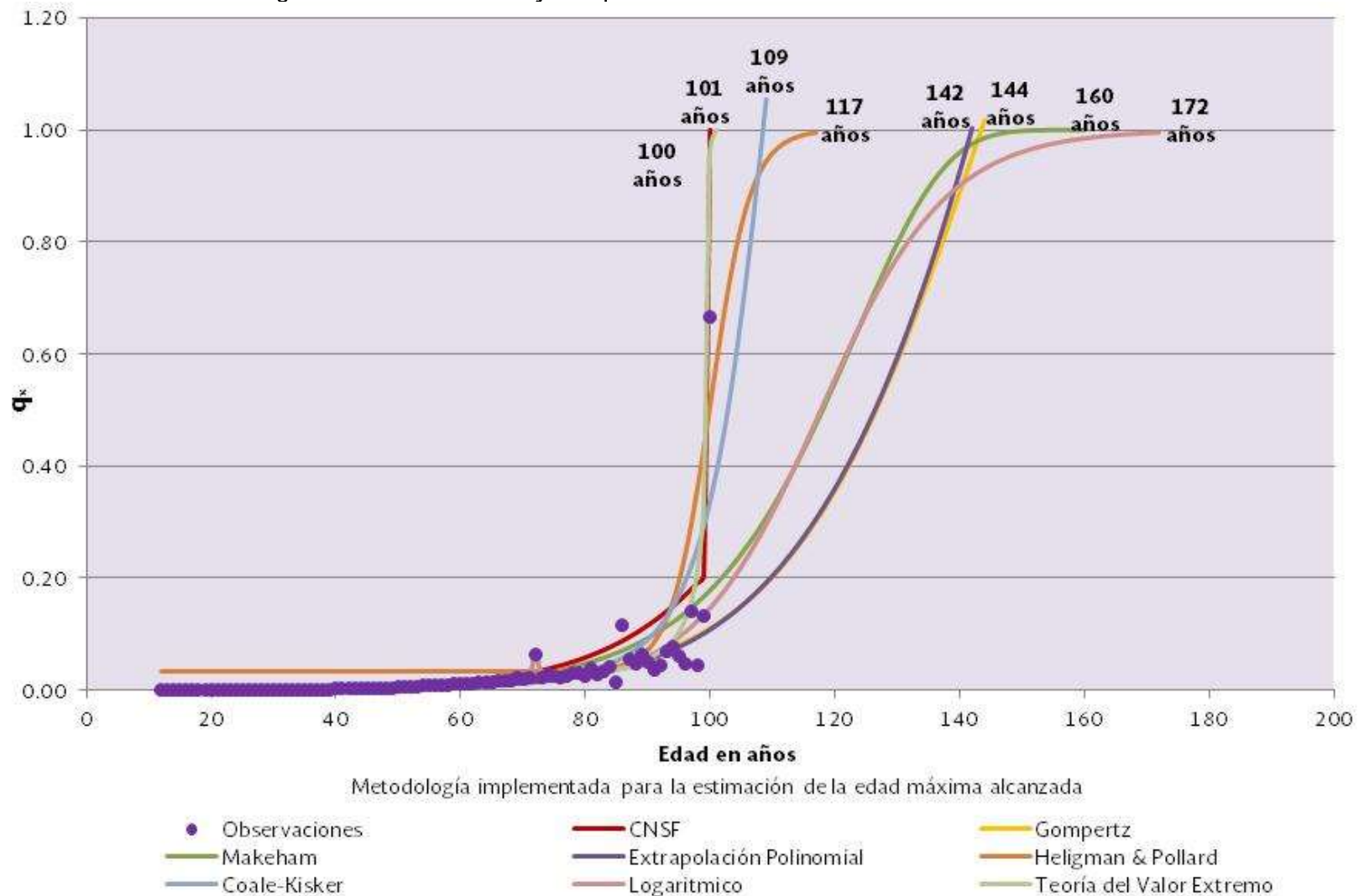
1. Salud pública
2. Seguridad Pública
3. Demografía
4. Economía
5. Actuaría.

En México las tablas de mortalidad al ser construidas, son utilizadas hasta por más de una década, lo que ocasiona que a partir de cierto periodo la q'_x (probabilidad de fallecimiento estimada) no describa de manera adecuada a la q_x , por esta razón como actuarios especialistas en calcular el impacto de riesgos e incertidumbre y cuantificar los resultados para minimizar las pérdidas, el análisis de un buen ajuste de la tabla de mortalidad y de la edad máxima alcanzada es uno de los temas más importantes que debe ser estudiado, implementando algún modelo que logre describir la mortalidad mexicana de manera más adecuada.

Durante la investigación se utilizaron seis metodologías diferentes para el ajuste de la tabla de mortalidad y la edad máxima alcanzada, con el propósito de que las tasas de mortalidad, fueran consistentes con la finalización de la curva que representará la mortalidad mexicana CNSF-2000-I, además se buscó, que la finalización de la curva que describiera la mortalidad mexicana 2000-I, fuera consistente con la experiencia que se tiene, es decir que la edad máxima alcanzada no superara los 125 años que es justo la edad máxima que se ha registrado en México.

En la Figura 4.1, se puede observar de manera gráfica un contraste que se hace de las curvas que intentan describir la mortalidad mexicana 2000-I, aun cuando se sabe que no es una prueba estadística formal, esta gráfica podría dar al menos una general de los resultados.

Figura 4.1 Modelos de ajuste para la mortalidad mexicana CNSF-2000-I.



Fuente: Elaboración propia

Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.

Además para hacer un comparativo consistente entre las curvas obtenidas más allá de la Figura 4.1, en el Cuadro 1.3 se encuentran las estimaciones realizadas con los valores de D (distancia máxima entre los valores estimados y las observaciones de mortalidad mexicana 2000-I) y los de p-value, para hacer un contraste de los modelos realizados y tomar la decisión que cubra de manera más adecuada el propósito de la presente investigación, es decir, describir de manera más adecuada la mortalidad mexicana 2000-I, y así encontrar la edad máxima que alcanza esta población.

Cuadro 1.3 Comparativo de los modelos ajustados a la mortalidad mexicana 2000-I, para la estimación de la edad máxima alcanzada.

Modelo implementado	Edad máxima estimada	Prueba Kolmogorov-Smirnov ⁴⁶	Decisión
Extrapolación Polinomial	142 años	D = 0.070554 p-value = 1	Rechazado por la edad
Modelo de Gompertz	144 años	D = 0.07058 p-value = 0.0001	Rechazado por p-value
Modelo de Makeham	160 años	D=06908.0 p-value =0.3763	Rechazado por p-value
Heligman & Pollard	117 años	D=0.31638 p-value =1	Rechazado por D
Coale-Kisker	109 años	D=0.04391 p-value = 0.991	NO rechazado
Modelo Logit	172 años	D=0.07817 p-value = 0	Rechazado por p-value
Modelo Logit CNSF	100 años	D=0.1464 p-value = 0.7064	NO rechazado
Teoría del Valor Extremo (DPG)	101 años	D=0.2064 p-value =0.9997	NO rechazado

Fuente: Elaboración propia

Por el número de observaciones con que cuenta la muestra, se debe tener un parámetro que permita decidir si la distribución estimada describe de manera adecuada a la mortalidad mexicana 2000-I, prueba de Kolmogorov-Smirnov menciona como parámetro D, el cual debe ser menor a 0.1728⁴⁷, se tomó en cuenta también el p-value el cual para las observaciones rechazadas está por debajo de 0.838 por esta razón en la columna llamada decisión se colocó el resultado de la prueba de Kolmogorov-Smirnov.

⁴⁶ Si se desea saber la justificación de los valores para D y p-value ver el Anexo A A.3 pág. 81.

⁴⁷ Ver el cálculo del valor de D en el Anexo A.3

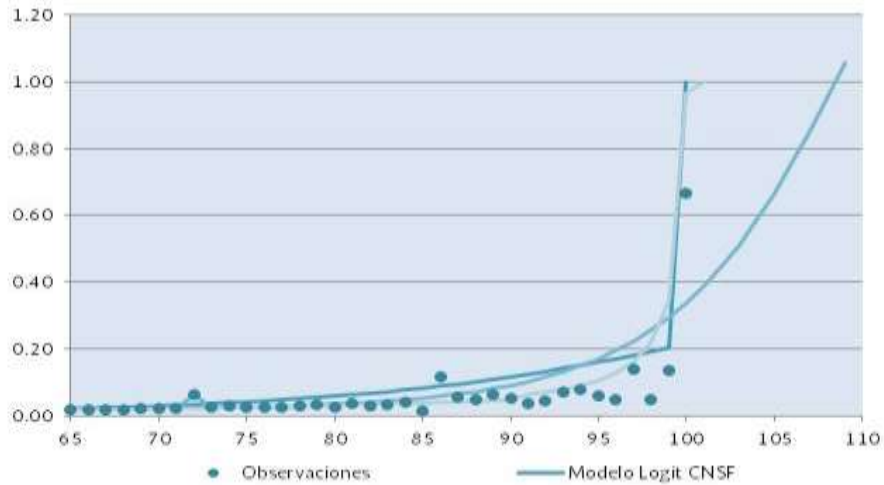
Se puede observar en el Cuadro 1.3 que a pesar de ser que algunos modelos suavizan la curva de mortalidad mexicana 2000-I, presentaron deficiencias pues la prueba de Kolmogorov-Smirnov las ha rechazado, además la estimación de la edad máxima alcanzada es inaceptable si se toma como referencia la persona más longeva de México (125 años) aun cuando este es un caso excepcional.

Existen además tres estimaciones de la curva de mortalidad mexicana 2000-I que aun cuando cumplen el objetivo deseado, tienen carencias importantes:

- El ajuste propuesto por Coale-Kisker, este modelo que logró estimar la edad máxima alcanzada, suavizar la curva que describe de manera consistente la mortalidad, por tal motivo no es rechazada por la prueba de Komogorov-Smirnov debido a que también minimiza las variaciones en edades seniles para la mortalidad mexicana observada 2000-I, el p-value resulta ser más elevado que 0.95; sin embargo, la carencia de la metodología radica en que los supuestos de los que parte (la edad en la que y la edad en la cual comienzan a presentar mayor varianza las observaciones), pues al ser modificados, alteran de manera importante la curva estimada. Además no existe alguna prueba que ayude a la elección de dichos supuestos. Para contrarrestar los efectos de dichas elecciones se realizaron ajustes con la misma metodología modificando los dos supuestos, como resultado de estas modificaciones se llegó a la conclusión de que la edad máxima estimada mediante la metodología de Coale-Kisker es de 109 años para las observaciones de mortalidad mexicana 2000-I.
- El modelo logit CNSF, es una estimación que parte del supuesto que no existen individuos en la población mexicana que superen los 100 años de edad esto provoca que la curva no sea suavizada entre edad 99 años y 100 años, porque no se realizó algún modelo que describa la mortalidad mexicana en este intervalo, lo cual ocasiona que la estimación no sea consistente con la información que se tiene de los últimos años. Gracias a los avances en la ciencia y la tecnología, ahora las personas tienen una esperanza de vida mayor, lo que se entiende también como una mejora en la edad máxima alcanzada por la población mexicana 2000-I, por esta razón el modelo propuesto por la CNSF, no se consideró el más adecuado para describir la mortalidad mexicana 2000-I.
- Utilizando la teoría de extremos y en particular lo que se refiere a la distribución de Pareto Generalizada se tomó como edad umbral $N= 90$ años, así se llegó a la conclusión de que la edad máxima alcanzada por la población mexicana 2000-I es de 101 años, esta curva de ajuste cumple con ser suavizada y no es rechazada por la prueba de Kolmogorov-Smirnov, incluso es la curva con el p-value más alto, cubriendo así los requerimientos de la investigación. Por tal motivo se propone como la mejor metodología para ser aplicada en la mortalidad mexicana CNSF-2000-I, debido a que permite realizar el proceso de manera iterativa y rápida con la modificación de la elección del umbral, es decir su elección no tendría que ser a consideración de quien realice el ajuste, lo que facilita la modificación de las hipótesis en caso de que alguna de ellas sufra una variación y la simulación de varios escenarios ante dichas variaciones resulta simple.

Un punto que se considero es importante analizar también, es el contraste de las estimaciones para los resultados obtenidos por el modelo de Coale-Kisker, el estimado por la CNSF y el modelo resultante de la aplicación del modelo de la teoría de extremos (DGP) para las probabilidades de fallecimiento a edades seniles en comparación con los valores observados.

Figura 4.2 Análisis de probabilidades de fallecimiento observadas vs modelo de Coale-Kisker, logit CNSF y teoría de extremos (GPD).



Fuente: Elaboración propia

Como bien la prueba de Kolmogorov-Smirnov nos dice, la curva mejor ajustada a la mortalidad mexicana observada 2000-I es la curva estimada mediante la distribución Pareto Generalizada. Sin embargo, se puede observar que:

El comportamiento de las curvas de mortalidad es el siguiente:

- Para $x < 95$ años se tiene que $q_{x-TVE} < q_{x-Coale-Kisker} < q_{x-CNSF}$.
- Para $95 \text{ años} \leq x < 98$ años la mortalidad estimada se comporta de la siguiente manera $q_{x-TVE} < q_{x-CNSF} < q_{x-Coale-Kisker}$.
- Para $98 \text{ años} \leq x < 99$ años se observa que $q_{x-CNSF} < q_{x-TVE} < q_{x-Coale-Kisker}$ y
- Para $x > 99$ años $q_{x-Coale-Kisker} < q_{x-CNSF} < q_{x-TVE}$.

En el último capítulo mostré que la curva bien ajustada y por ende la edad máxima estimada de la curva de mortalidad mexicana adecuada, son fundamentales para el cálculo de seguros, anualidades (pensiones), etc. Y en caso de no estar modeladas adecuadamente podían originar pérdidas importantes para una compañía, así que de acuerdo a todas las desigualdades mostradas en el párrafo anterior se debe tener cuidado en la elección adecuada de la curva, se puede decir además, que el uso de tablas de mortalidad ajustadas de manera correcta y con edad máxima estimada adecuada nos brinda la oportunidad de hacer mejoras en las aplicaciones actuariales que tiene, no solo para las compañías aseguradoras, financieras, etc. sino también para los beneficiarios de estos productos.

Por estas razones es recomendable la actualización de las tablas de mortalidad de manera continua para que de esta forma logren reflejar mejor la evolución y los niveles reales de mortalidad de una población, cabe destacar que los datos observados de mortalidad mexicana 2000-I que permitieron la elaboración de las tablas de mortalidad en México, es información sesgada e incompleta, pues presenta la limitación de ser truncada por la izquierda, lo que podría afectar de manera importante los resultados de la construcción de la tabla de mortalidad que represente de forma general a toda la población mexicana.

Anexos

Anexo A (Códigos de análisis)

A.1 Código de extrapolación polinomial en R

##Abrir la tabla

```
l<-read.csv('C:/Users/Karina/Dropbox/Tesis/Avances/Splines.csv')
probabilidad<-numeric(140);edad<-numeric(140)
```

##Cálculo los coeficientes del ajuste cúbico para la información de la tabla de Mendoza

```
x<-array(c(l[,1]));y<-array(c(l[,4]))
f <- splinefun(x, y,method="hyman")
```

```
ls(envir=environment(f))
splinecoef <- get("z", envir = environment(f))
curve(f(x),ylab="qx",,type="p", xlim=c(12,135),ylim=c(0,1),
      col = "chocolate2", lwd = 1,main="Extrapolación Cúbica, CNSF 2000-I")
legend(12,1, c("CSF200","Spline R"),col=c("chocolate2","chocolate4"), pch=1)
points(splinecoef, col = "chocolate2", cex = 1)
points(100,1,col="chocolate2",lwd=1)
for(i in 1:100){if(max(f(90+i))>1){Mn=90+i-1;break}
points(90+i,f(90+i),col="chocolate4",lwd=2}}
```

##Edad Máxima Esperada con extrapolación Cúbica para la información de la CNSF 2000-I.

Mn

##Convertir a Excel

```
for (i in 1:140){probabilidad[i]<-f(i); edad[i]<-11+i}
Splines <- data.frame(x = edad, q_x = probabilidad)
write.table(Splines, file="C:/Users/Karina/Dropbox/Tesis/Avances/splines.txt")
```

A.2 Comparación de muestras mediante la prueba Cuantil-Cuantil

En estadística, un gráfico Cuantil Cuantil es un método gráfico para el diagnóstico de diferencias entre la distribución de probabilidad de una población de la que se ha extraído una muestra aleatoria (para la presente investigación la muestra aleatoria son las observaciones de fallecimiento que se obtuvieron) y una distribución usada para la comparación (en esta investigación serán las estimaciones realizadas para mortalidad mexicana 2000-I). Para decir que ambas funciones tienen relación deben parecerse a la función identidad, es por esa razón, que dentro de todas las Figuras cuantil a cuantil de esta investigación se encuentra trazada la recta identidad.

A.3 Prueba Kolmogorov-Smirnov⁴⁸

Los test de bondad de ajuste tratan de decidir si los datos observados en la muestra se ajustan a una distribución de probabilidad determinada. Generalmente esta prueba es utilizada para funciones de probabilidad discretas; sin embargo, existen también para funciones continuas. Este tipo de pruebas se realizan entre otros motivos, para predecir sucesos de interés, en el caso de la presente investigación en la mortalidad mexicana 2000-I, el test se realiza para edades donde se pudo recolectar una muestra con el objetivo de predecir el momento en que la población mexicana alcanza $q_x = 1$.

- Entre las pruebas de bondad de ajuste encontramos:
- Contraste Chi-cuadrado.
- Test de Kolmogorov-Smirnov.
- Criterio de Cramér-von Mises
- Test de Anderson–Darling
- Test de Shapiro–Wilk
- Criterio de información de Akaike

Dependiendo de las características que tenga la muestra a la cual se requiere realizar el análisis será la prueba de bondad que se aplica. En el caso de la mortalidad mexicana 2000-I se utilizó el test de Kolmogorov-Smirnov por ser una prueba para $F(x)$ continua que no necesita una muestra grande y además no requiere de realizar agrupaciones de datos, al contrario del contraste de Chi-cuadrado.

La idea fundamental para la prueba de Kolmogorov-Smirnov consiste en comparar las frecuencias acumuladas teóricas $F_n(x)$ contra las frecuencias acumuladas observadas $F_o(x)$, sin agrupar los datos es decir, para cada uno de los datos de la muestra se calculan ambos valores, y se restan uno del otro. La más grande (en valor absoluto) de estas diferencias es el valor de la prueba. El test queda descrito de la siguiente manera:

Sea X una variable aleatoria continua con función de distribución $F(x)$. Se plantea el contraste:

$$H_0 : F(x) = F_o(x) \text{ para todo } x.$$
$$H_1 : F(x) \neq F_o(x) \text{ para algún } x.$$

Donde $F_o(x)$ es la distribución estimada mediante el modelo de ajuste es decir, se propone un modelo de distribución para los datos, y como alternativa que los datos no se distribuyan según este modelo.

Como expresión matemática, la estadística $D_n = \max\{F_n^*(x) - F_o(x)\}$, calculada para todo x se distribuye de acuerdo con la función definida por Kolmogorov-Smirnov. El estadístico calcula la discrepancia máxima entre la función empírica y la propuesta en H_0 , al realizar este contraste obtenemos como salida el valor D y el p -value. Se rechaza H_0 si el p -value es menor al nivel de significancia elegido y si el valor de D es mayor al valor representativo elegido de la siguiente tabla:

⁴⁸http://books.google.com.mx/books?id=FPdEyTzh6c4C&pg=PA58&lpg=PA58&dq=prueba+de+bondad+y+ajuste+en+excel+para+muestras+grandes&source=bl&ots=X4WhZ2ycD1&sig=9P_hbSheRzZOd3czgslU_ifwd9I&hl=es-419&sa=X&ei=TaddUv6vJ8GhkQeGpYHYCg&redir_esc=y#v=onepage&q=prueba%20de%20bondad%20y%20ajuste%20en%20excel%20para%20muestras%20grandes&f=false

Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.

Distribución del estadístico de Kolmogorov-Smirnov D_n .
Se tabula d tal que $P(D_n > d) = \alpha$

n	α					n	α				
	0'2	0'1	0'05	0'02	0'01		0'2	0'1	0'05	0'02	0'01
1	0'900	0'950	0'975	0'990	0'995	21	0'226	0'259	0'287	0'321	0'344
2	0'684	0'776	0'842	0'900	0'929	22	0'221	0'253	0'281	0'314	0'337
3	0'565	0'636	0'780	0'785	0'829	23	0'216	0'247	0'275	0'307	0'330
4	0'493	0'565	0'624	0'689	0'734	24	0'212	0'242	0'269	0'301	0'323
5	0'447	0'509	0'563	0'627	0'669	25	0'208	0'238	0'264	0'295	0'317
6	0'410	0'468	0'519	0'577	0'617	26	0'204	0'233	0'259	0'290	0'311
7	0'381	0'436	0'483	0'538	0'576	27	0'200	0'229	0'254	0'284	0'305
8	0'358	0'410	0'454	0'507	0'542	28	0'197	0'225	0'250	0'279	0'300
9	0'339	0'387	0'430	0'480	0'513	29	0'193	0'221	0'246	0'275	0'295
10	0'323	0'369	0'409	0'457	0'489	30	0'190	0'218	0'242	0'270	0'290
11	0'308	0'352	0'391	0'437	0'468	31	0'187	0'214	0'238	0'266	0'285
12	0'296	0'338	0'375	0'419	0'449	32	0'184	0'211	0'234	0'262	0'281
13	0'285	0'325	0'361	0'404	0'432	33	0'182	0'208	0'231	0'258	0'277
14	0'275	0'314	0'349	0'390	0'418	34	0'179	0'205	0'227	0'254	0'273
15	0'266	0'304	0'338	0'377	0'404	35	0'177	0'202	0'224	0'251	0'269
16	0'258	0'295	0'327	0'366	0'392	36	0'174	0'199	0'221	0'247	0'265
17	0'250	0'286	0'318	0'355	0'381	37	0'172	0'196	0'218	0'244	0'262
18	0'244	0'279	0'309	0'346	0'371	38	0'170	0'194	0'215	0'241	0'258
19	0'237	0'271	0'301	0'337	0'361	39	0'168	0'191	0'213	0'238	0'255
20	0'232	0'265	0'294	0'329	0'352	40	0'165	0'189	0'21	0'235	0'252
						> 40	$\frac{1'07}{\sqrt{n}}$	$\frac{1'22}{\sqrt{n}}$	$\frac{1'36}{\sqrt{n}}$	$\frac{1'52}{\sqrt{n}}$	$\frac{1'63}{\sqrt{n}}$

Fuente: <http://dm.udc.es/asignaturas/estadistica2/documentos-pdf/dmtablas.pdf>. (Consulta: 24-Oct-2013).

El parámetro para esta prueba es n , el tamaño de la muestra. La prueba es unilateral derecha, por lo que si el valor de prueba cae a la derecha del valor crítico, la hipótesis nula se rechazará.

Para el caso de la mortalidad mexicana 2000-I el tamaño de la muestra es de 89 observaciones.

	80%	90%	95%	98%	99%
>40	0.1134	0.1293	0.1442	0.1611	0.2067

El nivel de significancia que se adoptó para la presente investigación es del $1 - \alpha = 99\%$ es decir, se busca que:

$$1 - p\text{-value sea lo menor posible y } D < 0.2067$$

El código que se utilizó para obtener en el programa R los valores del p-value y los valores de D es el siguiente:

```
simulaciones<-read.csv('C:/Users/Karina/Dropbox/Tesis/Versión Final/Tabla.csv')
```

Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.

observados<-simulaciones[,1]
Estimados<-simulaciones[,2]

ks.test(observados,Estimados)

A.4 Desarrollo del modelo de Coale-Ksiker

Para analizar el impacto que tiene el supuesto de la edad a la cual $m_x = 1$, se generalizó la expresión que calcula el valor de R con el propósito de facilitar el proceso iterativo y realizar la comparación de varias simulaciones. Partiendo de la ecuación original:

$$k(x) = k(x-1) - R \text{ con } x \in \{n, n+1, n+2, \dots, n+(w-n-1), n+(w-n) | n \in N\}$$

$$\text{con } n \leq w$$

Donde n es la edad a partir de la cual hay una variabilidad importante y w es la edad para el cual $m_w = 1$. De lo anterior:

$$k(n+1) = k(n) - R = k(n-1) - 2R$$

$$k(n+2) = k(n+1) - R = k(n-1) - 3R$$

$$k(n+3) = k(n+2) - R = k(n-1) - 4R$$

$$\dots$$

$$k(w-1) = k(n+(w-n-2)) - R = k(n-1) - (w-n)R$$

$$k(w) = k(n+(w-n-1)) - R = k(n-1) - (w-n+1)R$$

De lo anterior se observa que:

$$k(n) + k(n+1) + k(n+2) + \dots + k(w+1) + k(w) = (w-n+1) * k(n-1) - \sum_{i=1}^{w-n+1} iR$$

$$k(n) + k(n+1) + \dots + k(w+1) + k(w) = (w-n+1) * k(n-1) - \left[\frac{(w-n+1)(w-n+2)}{2} \right] R$$

Por otra parte se tiene que:

$$k(n) + k(n+1) + \dots + k(w+1) + k(w) = \ln \left[\frac{m_n}{m_{n-1}} \right] + \ln \left[\frac{m_{n+1}}{m_n} \right] + \dots + \ln \left[\frac{m_{n+(w-n-1)}}{m_{n+(w-n-2)}} \right] + \ln \left[\frac{m_{n+(w-n)}}{m_{n+(w-n-1)}} \right]$$

$$k(n) + k(n+1) + \dots + k(w+1) + k(w) = \ln(m_w) - \ln(m_{n-1})$$

Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.

Por lo tanto:

$$\ln(m_w) - \ln(m_{n-1}) = (w - n + 1) * k(n - 1) - \left[\frac{(w - n + 1)(w - n + 2)}{2} \right] R$$

Despejando R se obtiene que que:

$$R = \frac{2 \left[(w - n + 1)k(n - 1) - \frac{\ln(m_w)}{\ln(m_{n-1})} \right]}{(w - n + 1)(w - n + 2)}$$

Con $x \in \{n, n + 1, n + 2, \dots, n + (w - n - 1), n + (w - n) | n \in N\}$ y $n \leq w$.

A.5 Simulaciones binomiales del modelo logit, para la estimación de la curva que describe la mortalidad mexicana 2000-I

```
##Programa para la simulación de escenarios para la tabla de mortalidad Mexicana 2000-I
simulaciones<-read.csv('C:/Users/Karina/Dropbox/Tesis/Avances/Tabla de Mendoza
simulación de valores.csv')
```

```
x<-simulaciones[,1]
Ex<-simulaciones[,2]
dx<-simulaciones[,3]
qx<-simulaciones[,4]
buscando<- matrix(0,length(qx),10)
ajuste_fin<-numeric(length(qx))
for (i in 1:length(qx)){
    for (j in 1:10){buscando[i,j]<-(rbinom(1,Ex[i],qx[i])/Ex[i])}
    ajuste_fin[i]<-mean(buscando[i,])}
```

```
plot(ajuste_fin, type="p",main="Simulación Tabla de Mortalidad Mexicana 2000-I", col="blue",
cex=0.6)
```

```
##Convertir a Excel
```

```
Simulación <- data.frame(x = x, q_x = ajuste_fin)
write.table(Simulación, file="C:/Users/Karina/Dropbox/Tesis/Avances/Simulación Tabla de
Mortalidad.txt")
```

Anexos B (Observaciones de la distribución Pareto)

B.1 Distribución Pareto generalizada

La distribución Pareto tiene tres parámetros, los cuales son:

- parámetro de localización.

Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.

- parámetro de forma.
- parámetro de escala.

B.1.1 Características para la distribución del máximo

La función de densidad para el máximo está dada por:

$$f(x) = \frac{1}{\theta \left[1 + \gamma \left(\frac{x-N}{\theta}\right)\right]^{1+\frac{1}{\gamma}}} \begin{cases} x \geq N \text{ cuando } \gamma \geq 0 \\ x \leq N - \frac{\theta}{\gamma} \text{ cuando } \gamma < 0 \end{cases}$$

La función de distribución para el máximo

$$F(x) = \begin{cases} 1 - e^{-\frac{x-N}{\theta}} & \text{si } \gamma = 0 \\ 1 - \left[1 + \gamma \left(\frac{x-N}{\theta}\right)\right]^{-\frac{1}{\gamma}} & \text{si } \gamma \neq 0 \end{cases} \quad \text{para } \begin{cases} x \geq N \text{ cuando } \gamma \geq 0 \\ x \leq N - \frac{\theta}{\gamma} \text{ cuando } \gamma < 0 \end{cases}$$

La media para la distribución del máximo es: $N + \frac{\theta}{1-\gamma}$ siempre que $\gamma < 1$.

La varianza para la distribución del máximo es: $\frac{\theta^2}{(1-\gamma)^2(1-2\gamma)}$ siempre que $\gamma < 1/2$.

La mediana para la distribución del máximo es: $N + \frac{\theta(2^\gamma - 1)}{\gamma}$.

También existe información que describe al mínimo de una muestra.

B.1.2 Características para la distribución del mínimo

La función de densidad para el mínimo está dada por:

$$f(x) = \begin{cases} \frac{1}{\theta} e^{\frac{x}{\theta}} & \text{si } \gamma = 0, x \leq 0, \theta > 0 \\ \theta \left[1 + \gamma \left(\frac{x}{\theta}\right)\right]^{-\frac{1}{\gamma}+1} & \text{si } \gamma \neq 0, \gamma \left(\frac{x}{\theta}\right) \geq 0, \theta > 0. \end{cases}$$

La función de distribución para el mínimo es:

$$F(x) = \begin{cases} e^{\frac{x}{\theta}} & \text{si } \gamma = 0, x \leq 0, \theta > 0 \\ \left[1 + \gamma \left(\frac{x}{\theta}\right)\right]^{-\frac{1}{\gamma}} & \text{si } \gamma \neq 0, \gamma \left(\frac{x}{\theta}\right) \geq 0, \theta > 0. \end{cases}$$

B.1.3 Casos particulares

B.1.3.1 Para el máximo

- Cuando $\gamma=0$, la Distribución Pareto Generalizada con dos parámetros para el máximo es la distribución exponencial con media θ .
- Cuando $\gamma=1$, la Distribución Pareto Generalizada con dos parámetros para el máximo es la distribución uniforme con parámetros 0 y θ .

B.1.3.2 Para el mínimo.

- Cuando $\gamma=0$, la Distribución Pareto Generalizada con dos parámetros para el mínimo es la distribución exponencial inversa con media θ .
- Cuando $\gamma=1$, la Distribución Pareto Generalizada con dos parámetros para el máximo es la distribución uniforme con parámetros 0 y $-\theta$.

B.2 Desarrollo de la igualdad para la estimación de parámetros del modelo construido a partir de la teoría de extremo (GPD)

El primero que es igual a:

$$\begin{aligned} & \sum_{x=64}^{N-1} \left[\frac{S(x) - S(x+1)}{S(65)} \right] + l_N \left[\ln \frac{S(N)}{S(65)} \right] = \\ & \sum_{x=65}^{N-1} d_x \ln(S(x) - S(x+1)) - \sum_{x=65}^{N-1} d_x \ln(S(65)) + l_N(S(N)) - l_N(S(65)) = \\ & \sum_{x=65}^{N-1} d_x \ln(S(x) - S(x+1)) + l_N(S(N)) - \sum_{x=65}^{N-1} d_x \ln(S(65)) - l_N(S(65)) = \\ & \sum_{x=65}^{N-1} d_x \ln(S(x) - S(x+1)) + l_N(S(N)) - \left[l_N + \sum_{x=65}^{N-1} d_x \right] \ln(S(65)) = \end{aligned}$$

Recordando que $l_{65} = l_N + d_{N-1} + d_{N-2} + \dots + d_{N-65}$

$$\sum_{x=65}^{N-1} d_x \ln(S(x) - S(x+1)) + l_N(S(N)) - l_{65} \ln(S(65))$$

Al tratarse de las edades $65 \leq x < N$ con $x \in \mathbb{Z}$, se deben calcular los parámetros considerando a la función de supervivencia como: $S(x) = e^{-\frac{BC^x}{\ln(C)}(C-1)}$ por considerar que sigue la ley de Gompertz. Y el segundo sumando es igual a:

$$\sum_{x=N}^{99} d_x \ln \left[\frac{S(x) - S(x+1)}{S(65)} \right] - l_N \left[\ln \left[\frac{S(N)}{S(65)} \right] \right] + l_{100} \left[\ln \left[\frac{S(100)}{S(65)} \right] \right] =$$

Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.

$$= \sum_{x=N}^{99} d_x [\ln(S(x) - S(x+1)) - \ln(S(65))] + l_{100}(S(100) - \ln(S(65))) - l_N(\ln(S(N)) - \ln(S(65)))$$

Recordando que $l_{100} = l_{100} + d_N + d_{N+1} + \dots + d_{N+99}$, entonces $\sum_{x=N}^{99} d_x = l_N - l_{100}$

$$\begin{aligned} & \sum_{x=N}^{99} d_x [\ln(S(x) - S(x+1))] + l_{100}(S(100)) - l_N(\ln(S(N))) = \\ & \sum_{x=N}^{99} d_x [\ln(S(x) - S(x+1))] + \left[l_N - \sum_{x=N}^{99} d_x \right] (S(100)) - l_N(\ln(S(N))) = \\ & \sum_{x=N}^{99} d_x [\ln(S(x) - S(x+1)) - S(N)] + l_{100}(S(100) - \ln(S(N))) = \\ & = \sum_{x=N}^{99} d_x \ln \left[\frac{S(x) - S(x+1)}{S(N)} \right] + l_{100} \left[\ln \left[\frac{S(100)}{S(N)} \right] \right] \end{aligned}$$

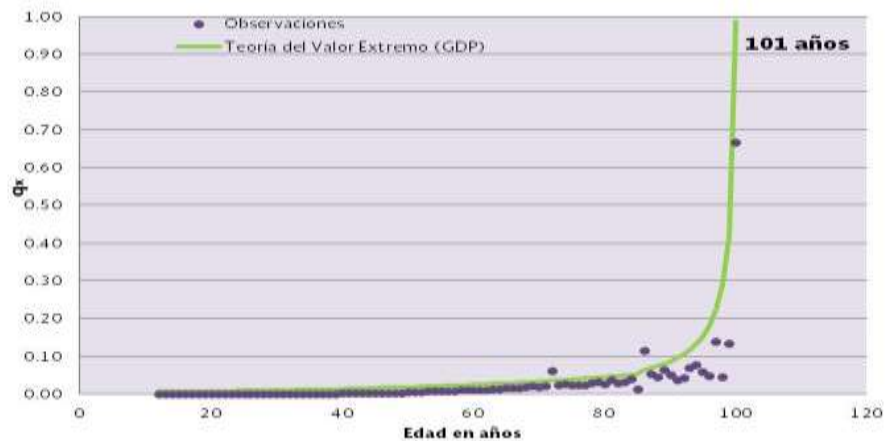
Donde $S(x) = S(N) \left[1 + \gamma \left(\frac{x-N}{\theta} \right) \right]^{-\frac{1}{\gamma}}$ para edades $N \leq x \leq 99$. De lo anterior que:

$$\frac{S(x)}{S(N)} = \frac{S(N) \left[1 + \gamma \left(\frac{x-N}{\theta} \right) \right]^{-\frac{1}{\gamma}}}{S(N) \left[1 + \gamma \left(\frac{N-N}{\theta} \right) \right]^{-\frac{1}{\gamma}}} = \left[1 + \gamma \left(\frac{x-N}{\theta} \right) \right]^{-\frac{1}{\gamma}}$$

B.3 Simulación de la curva de mortalidad cuando el umbral es de 85 años

Los parámetros estimados (mediante el algoritmo desarrollado en el capítulo 3 apartado 3.3.1.2) que describen la mortalidad mexicana 2000-I para la edad umbral $N=85$, son

Figura 5.1 Teoría de extremos aplicada a los datos base de mortalidad mexicana CNSF-2000-I



Fuente: Elaboración propia

Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.

El presente ajuste tiene una variación mayor entre la curva ajustada y las observaciones, por esta razón la curva no pudo ser suavizada, para cumplir el objetivo principal de esta investigación, lo que significa que debemos realizar un mayor número de repeticiones variando la edad umbral, antes de considerar alguno de los ajustes como el indicado para modelar la mortalidad mexicana 2000-I.

Anexos C (Metodologías de cálculo)

C.1 Ajuste de parámetros del modelo de Gompertz

1.- Se crea un archivo en Excel con una tabla que contenga las columnas:

- Edad, que debe comenzar en 12 años y aumenta en una unidad hasta llegar a los 150 años, los registros que proporciona la Comisión Nacional de Seguros y Fianzas llegan a 100, pero como se busca explorar la función para establecer una edad máxima alcanzada se extiende esta columna.
- q_x representará la mortalidad sin ajustar⁴⁹.
- m_x representa el índice de mortalidad central, se elige este índice debido a la facilidad que representa la estimación de sus parámetros. Es sustituida por la ecuación 1.1.
- $\ln(m_x)$, $\ln(m'_x)$ y $\ln(m_x - m'_x)^2$, columnas de apoyo para el ajuste.
- m'_x representa el índice de mortalidad central estimado por el modelo del ajuste.
- q'_x representa la mortalidad del modelo ajustado.

Se deberán colocar en las columnas de la B2 a la I2, respectivamente.

	A	B	C	D	E	F	G	H	I
1			Valores Originales						
2		Edad	q_x	m_x	$\ln(m_x)$	$\ln(m'_x)$	$\ln(m_x - m'_x)^2$	m'_x	q'_x
3		12	0.00173705	0.0017					
4		13	0.00028543	0.0003					
5		14	0.00017890	0.0002					

2.- Se debe recordar que: $\mu(x) = BC^x$. Por este motivo, se pueden estimar los parámetros B y C apoyados de la fuerza de mortalidad, y el índice de mortalidad central, debido a que $m_x \cong \mu(x)$, de lo anterior se puede escribir que $\ln(m_x) \cong \ln(\mu(x))$, de esta forma los parámetros a estimar pertenecen a la regresión lineal: $\ln(m_x) \cong \ln(\mu(x)) = \ln(B) + x \ln(C)$, entonces $\ln(B) = B^*$ y $\ln(C) = C^*$ en la ecuación $\ln(m_x) \cong \ln(B) + x \ln(C)$. Es importante observar que las condiciones para el modelo de Gompertz se han modificado de la siguiente manera:

- Originalmente $C > 1$, por ser el logaritmo una función creciente se respeta la desigualdad, lo que significa que $\ln(C) > \ln(1)$, es decir $C^* = \ln(C) > 0$. Esta condición se colocará en la casilla O11, para iniciar el proceso se debe dar un valor inicial, este se elige arbitrariamente.
- Para $B > 0$, se tendría una condición similar pero el logaritmo de cero diverge, lo que produce que la condición se haya eliminado. El valor de este parámetro se colocará en la

⁴⁹ La muestra de mortalidad mexicana para 2000-I se encuentran en el Anexo D1

Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.

casilla O10, para iniciar el proceso se debe dar un valor inicial, este se elige arbitrariamente.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1			Valores Originales													
2			edad	q_x	m_x	$\ln(m_x)$	$\ln(e_i - F_i)$	$\ln(m'_x)$								
3			12	0.00173793	0.0017											
4			13	0.00129563	0.0013											
5			14	0.00097996	0.0010											
6			15	0.00073285	0.0008											
7			16	0.00056000	0.0006											
8			17	0.00042887	0.0005											
9			18	0.00032790	0.0004											
10			19	0.00025732	0.0003											
11			20	0.00021603	0.0003											
12			21	0.00019972	0.0003											
13			22	0.00019585	0.0003											

- Se procede al llenado de la columna F. Para la casilla $F_i = \ln(q_x + B_i)$, con $i \in \{3, 4, \dots, 14\}$ es decir, se obtendrá el valor de $\ln(m'_x)$.
- La columna G, para la casilla $G_i = (E_i - F_i)^2$, con $i \in \{3, 4, \dots, 14\}$. La celda objetivo es la O12 = Suma(G:G), ya que lo buscado es que exista la mínima diferencia entre los datos reales y los datos estimados.

	I	J	K	L	M	N	O	P	Q
1									
2			q'_x						
3			0.000379						
4			0.000404						
5			0.000431						
6			0.000461						
7			0.000492						
8			0.000525						
9			0.000560						
10			0.000598						
11			0.000638						
12			0.000681						
13			0.000727						

- Se sustituirán las casillas L10 = exp(O10) que representará el valor de los parámetros para estimar el índice central de mortalidad m'_x .

3.- Los parámetros de la función se aproximaron mediante la función Solver con el código de optimización Simplex LP que se encuentra como una herramienta en el programa Excel de Microsoft Office 2007 (en adelante referido como Simplex LP).

Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.

- Aquí se colocará como objetivo minimizar la casilla O12, con la restricción de que la casilla O11, (que representará al parámetro C*). Además es importante mencionar que sólo deben moverse los parámetros a estimar, en este caso son los ubicados en las casillas O10, O11.
- Así que para calcular la columna H que corresponde a m'_x es decir, $\exp(\ln(m'_x))$ basta con aplicar la función exponencial a la columna F.

	B	C	D	E	F	G	H	I
1		Valores Originales						
2	Edad	q_x	m_x	$\ln(m_x)$	$\ln(m'_x)$	$\ln(m_x) - \ln(m'_x) \times 2$	m'_x	q'_x
3	12	0.00173705	0.0017	-6.356433	-7.8785	2.316687501	=EXP(F3)	0.000379
4	13	0.00028563	0.0003	-8.160947	-7.8133	0.120858247	0.000404322	0.000404
5	14	0.00017890	0.0002	-8.628794	-7.7481	0.775622265	0.000431562	0.000431
6	15	0.00038285	0.0004	-7.868063	-7.6829	0.034285298	0.000460637	0.000461

Se realizó entonces el cálculo de q_x que se expresa como $q_x = \frac{m_x}{1 + \frac{1}{2}m_x}$, hasta la casilla I141.

	C	D	E	F	G	H	I	J	K	L	M	N
1		Valores Originales										
2	q_x	m_x	$\ln(m_x)$	$\ln(m'_x)$	$\ln(m_x) - \ln(m'_x) \times 2$	m'_x	q_x					
3	0.00173705	0.0017	-6.356433	-7.8785	2.316687501	0.00037901	=H3/(1+(H3/2))					
4	0.00028563	0.0003	-8.160947	-7.8133	0.120858247	0.000404322	0.000404					
5	0.00017890	0.0002	-8.628794	-7.7481	0.775622265	0.000431562	0.000431					
6	0.00038285	0.0004	-7.868063	-7.6829	0.034285298	0.000460637	0.000461					

C.2 Ajuste de parámetros del modelo de Makeham

1.- Se crea un archivo en Excel con una tabla que contenga las columnas:

- Edad, que debe comenzar en 12 años y aumenta en una unidad hasta llegar a los 160 años, debido a que se busca explorar la función para establecer una edad máxima alcanzada se extiende esta columna y al no conocer el resultado se parte de la hipótesis que es una edad muy avanzada (Celda B2).
- q_x , que representara la mortalidad sin ajustar (Celdas Ci con $i \in \{3,91\}$).
- q'_x , que será el ajuste propuesto (Celda D2).

En las celdas G1 a G3 se colocan las letras A, B, C que serán las referencias para los parámetros que representarán la función. Así las celdas H1 a H3 contendrán los valores numéricos de dichos parámetros.

Apoyados en los parámetros del ajuste anterior y al ser un modelo muy parecido, la columna D contendrá el cálculo: $D_i = 1 - \text{EXP}(-\$H\$1 - (\$H\$2 * (\$H\$3^{\wedge} B_i) * (\$H\$3 - 1) / \text{LN}(\$H\$3)))$, con $i=3,4,5,\dots,151$.

Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.

	A	B	C	D	E	F	G	H	I
1					Máxima		A	0.000175228	$A \geq -B$
2		Edad	q_x	q'_x	0.54528		B	0.000115228	$B > 0$
3		12	0.00173705	0.00047	0.000185		C	1.072372478	$C > 1$
4		13	0.00028563	0.00047	0.000185		Celda objetivo		0.557490
									$x \geq D$

La columna Diferencias será: $E_j = \text{ABS}(C_j - D_j)$. Con $j=3,4,5,\dots, 91$. Pues es hasta edad 100 donde se tienen valores de muestra. La Celda objetivo será: $H4 = \text{SUMA}(E3:E90)$. Lo anterior debido a que se busca que exista la mínima diferencia entre la función ajustada y los valores reales.

	A	B	C	D	E
1					Máxima
2		Edad	q_x	q'_x	0.54528
3		12	0.00173705	0.00047	=ABS(C3-D3)
4		13	0.00028563	0.00047	0.000185
5		14	0.00017890	0.00047	0.000514
6		15	0.00018285	0.00047	0.000133

2.- Se procede a realizar la estimación de los parámetros del modelo. Los parámetros de la función se aproximaron mediante la función Solver con el código de optimización GRG Nonlinear que se encuentra como una herramienta en el programa Excel de Microsoft Office 2007.

The image shows the Solver Parameters dialog box in Excel. The target cell is H4 with the value 0.557490. The constraints are: A >= -B, B > 0, C > 1, and x >= D. The Solver engine is set to GRG Nonlinear. The graph shows the fitted curve for the mortality rate.

3.- Se arrastó la fórmula de la columna D (contemplando hasta 5 decimales), hasta la celda D151, que representa la edad 160 años, aquí la probabilidad ajustada de fallecer q_x es 1, es decir el propósito ha sido encontrado.

C.3 Ajuste de parámetros del modelo de Heligman & Pollard

1. Se crea una hoja de Excel que contenga los siguientes datos:

- Encabezados de la columna B a la G: Edad, q_x , $\frac{q_x}{1-q_x}$, $\frac{q_{1x}}{1-q'_{1x}}$, q'_x y $\left(\frac{q_x - q_{1x}}{q_x}\right)^2$ respectivamente, de las cuales sólo se podrá hacer el llenado de Edad, q_x , $\frac{q_x}{1-q_x}$ y $\left(\frac{q_x - q_{1x}}{q_x}\right)^2$ con los datos base de la

Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.

tabla de mortalidad mexicana 2000-I, a excepción del valor a edad 100, al cual se le asignará una $q_x = 0.5$.

- También de la columna K a la S se colocarán las variables A, B, C, D, E, F, G, H y x_0 , respectivamente. Dado que no se cuenta con los valores para estas variables, se puede asignar el valor que se desee, las únicas condiciones que debe cumplir son:
 - A, B y C son cero, pues no se tienen registros para estas edades.
 - D.- Toma valores en el intervalo (0,1)
 - E.- Toma valores en el intervalo (0, ∞)
 - F.- Toma valores en el intervalo (15,30]
 - G.- Toma valores en el intervalo (0,1)
 - H.- Toma valores en el intervalo (0,10)
 - X_0 .- Se le asigna el valor de 100.

Para concluir el paso 1, se debe tener una tabla como la siguiente:

	F	G	H	J	K	L	M	N	O	P	Q	R
1			Máximo		A	B	C	D	E	F	G	H
2	q'_x	q_x	0.31438	Celda objetivo	0.00000	0.00000	0.50000	0.03479	0.00010	21.00296	0.99999	1.37
3	0.03362	0.89932717	0.03188	-0.20								
4	0.03362	0.98307919	0.03333									
5	0.03362	0.98938543	0.03344									
6	0.03362	0.97735345	0.03324									
7	0.03362	1.00000000	0.03362									
8	0.03362	0.97431037	0.03318									

2.- Se completa la tabla:

- Para la columna E, se utiliza la ecuación:

$$\frac{q_x}{1 - q_x} = A^{(x+B)^C} + De^{-E[\ln(x)-\ln(F)]^2} + GH^{(x-x_0)}$$

- Debido a que son las probabilidades estimadas con los parámetros se desean encontrar.

	B	C	D	E	F	G	H
1							Máximo
2	Edad	q_x	$\frac{q_x}{1-q_x}$	$\frac{q'_x}{1-q'_x}$	q'_x	$\frac{q_x - q'_x}{q_x}$	0.31438
3	12	0.00173705	0.00174008	0.034786761	0.03362	0.89932717	0.03188
4	13	0.00028563	0.00028571	0.034787147	0.03362	0.98307919	0.03333

- Para la columna q'_x (columna F) se completa con ayuda de la columna que contiene los valores de $\frac{q'_x}{1-q'_x}$, mediante un despeje de la misma.

	B	C	D	E	F	G	H
1							Máximo
2	Edad	q_x	$\frac{q_x}{1-q_x}$	$\frac{q'_x}{1-q'_x}$	q'_x	$\frac{q_x - q'_x}{q_x}$	0.31438
3	12	0.00173705	0.00174008	0.034786761	=E3/(1+E3)	0.89932717	0.03188
4	13	0.00028563	0.00028571	0.034787147		0.98307919	0.03333
5	14	0.00017890	0.00017893	0.034787466		0.98938543	0.03344

Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.

- El objetivo es minimizar la diferencia entre q_x y q'_x , por tal motivo, se utiliza la siguiente expresión: $\left(\frac{q_x - q'_x}{q_x}\right)^2$, la cual reflejará la diferencia entre la probabilidad de fallecimiento estimada y la real.

	B	C	D	E	F	G	H
1							Máximo
2	Edad	q_x	$\frac{q_x}{1-q_x}$	$\frac{q'_x}{1-q'_x}$	q'_x	$\left(\frac{q_x - q'_x}{q_x}\right)^2$	0.31638
3	12	0.00173705	0.00174008	0.034786761	0.03362	=(C3-F3)/F3^2	0.03188
4	13	0.00028563	0.00028571	0.034787147	0.03362		0.03333
5	14	0.00017890	0.00017893	0.034787466	0.03362		0.03344
6	15	0.00038285	0.00038300	0.034787727	0.03362		0.03324
7	16	0.00000000	0.00000000	0.034787942	0.03362		0.03362

3.- Finalmente como ya se había mencionado la minimización de la suma $\left(\frac{q_x - q'_x}{q_x}\right)^2$, será la celda objetivo.

	C	D	E	F	G	H	I	J	K
1						Máximo			
2	q_x	$\frac{q_x}{1-q_x}$	$\frac{q'_x}{1-q'_x}$	q'_x	$\left(\frac{q_x - q'_x}{q_x}\right)^2$	Máximo	0.31638		A
3	0.00173705	0.00174008	0.034786761	0.03362	0.89932717	0.03188		Celda objetivo	0.00000
4	0.00028563	0.00028571	0.034787147	0.03362	0.98307919	0.03333		52.09	0.00000

4.- En la pestaña Datos, donde se encuentra la herramienta Solver, se le colocarán todas las restricciones antes mencionadas, el método de resolución que se utilizó fue Evolutionary.

De esta manera se encontraron los parámetros del modelo de Heligman & Pollard.

C.4 Ajuste de parámetros del modelo de logit

1.- Como se carece de varias muestras, para realizar el ajuste se consideró que el primer paso, es la creación de una tabla en Excel, la cual contendrá los siguientes campos: x (es la edad del individuo), q_x , $\ln\left(\frac{q_x}{1-q_x}\right)$, $\ln\left(\frac{q'_x}{1-q'_x}\right)$, (la cual representa el valor estimado, de la columna anterior),

Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.

$\left(\ln\left(\frac{q_x}{1-q_x}\right) - \ln\left(\frac{q'_x}{1-q'_x}\right) \right)^2$, (representará la diferencia entre la estimación y los datos reales, justo la que deseamos minimizar), y finalmente q'_x el valor estimado para las probabilidades de muerte a edad x . Se colocan entre las columnas B y G, en el orden anterior.

	A	B	C	D	E	F	G	H
1								Máximo
2		x	q_x	$\ln\left(\frac{q_x}{1-q_x}\right)$	$\ln\left(\frac{q'_x}{1-q'_x}\right)$	$\ln\left(\frac{q_x}{1-q_x}\right) - \ln\left(\frac{q'_x}{1-q'_x}\right)$	q'_x	0.51950
3		12	0.0017371				0.00174	0.00000

Es importante mencionar que se deben asignar 3 casillas en la columna K que contendrán los parámetros estimados: α, β , así como la celda objetivo.

	C	D	E	F	G	H	I	J	K
1									Máximo
2		q_x	$\ln\left(\frac{q_x}{1-q_x}\right)$	$\ln\left(\frac{q'_x}{1-q'_x}\right)$	$\ln\left(\frac{q_x}{1-q_x}\right) - \ln\left(\frac{q'_x}{1-q'_x}\right)$	q'_x	0.51950		
3		0.0017371				0.00174	0.00000		
4		0.0002834				0.00029	0.00000		
5		0.0001789				0.00018	0.00000		
6		0.0007424				0.00074	0.00000		
7		0.0000000				0.00000	0.00000		
8		0.0004346				0.00043	0.00000		
9		0.0004479				0.00044	0.00000		

Estimador

β 0.099134

α -11.670418

Celda Objetivo 11.226579

2.- Se llenan las celdas de la columna D, es importante recordar que esto sólo se hace a partir de edad 80, ya que es en donde se observa una varianza mayor, así como asigna valores cualesquiera a las celdas que corresponden a los parámetros a estimar, la tabla debe verse así:

	A	B	C	D	E	F	G	H	I	J
1										
2		x	q_x	$\ln\left(\frac{q_x}{1-q_x}\right)$	$\ln\left(\frac{q'_x}{1-q'_x}\right)$	$\ln\left(\frac{q_x}{1-q_x}\right) - \ln\left(\frac{q'_x}{1-q'_x}\right)$	q'_x	0.51950		
71		80	0.0263404	=LN(C71/(1-C71))	3.7397184	0.0168378	0.02321	0.00333		
72		81	0.0379950	-3.23156504	3.6405846	0.1672970	0.02557	0.01241		
73		82	0.0293088	-3.50012058	3.5414509	0.0017082	0.02818	0.00115		

3.- Inmediatamente se llena la columna E, recurriendo a la fórmula encontrada bajo la celda objetivo.

	A	B	C	D	E	F	G	H	I	J
1										
2		x	q_x	$\ln\left(\frac{q_x}{1-q_x}\right)$	$\ln\left(\frac{q'_x}{1-q'_x}\right)$	$\ln\left(\frac{q_x}{1-q_x}\right) - \ln\left(\frac{q'_x}{1-q'_x}\right)$	q'_x	0.51950		
71		80	0.0263404	-3.6099577	=LN(C71/(1-C71))	0.0168378	0.02321	0.00333		
72		81	0.0379950	-3.23156504	3.6405846	0.1672970	0.02557	0.01241		

4.- Se llena la columna F como (columna E - columna D)², La cual finalmente se utilizará para la obtención de la celda objetivo, pues esta última será la suma desde $x = 80$, hasta $x = 100$.

	A	B	C	D	E	F	G	H	I	J	K
1											
2		x	q_x	$\ln\left(\frac{q_x}{1-q_x}\right)$	$\ln\left(\frac{q'_x}{1-q'_x}\right)$	$\ln\left(\frac{q_x}{1-q_x}\right) - \ln\left(\frac{q'_x}{1-q'_x}\right)$	q'_x	0.51950			
3		12	0.0017371				0.00174	0.00000			
4		13	0.0003956				0.00029	0.00000			
5		14	0.0001789				0.00018	0.00000			
6		15	0.0007424				0.00074	0.00000			
71		80	0.0263404	-3.60995787	3.7397184	=D71-E71^2	0.02321	0.00333			
72		81	0.0379950	-3.23156504	3.6405846	0.1672970	0.02557	0.01241			
73		82	0.0293088	-3.50012058	3.5414509	0.0017082	0.02818	0.00115			
74		83	0.0329670	-3.37872556	3.4423171	0.0040429	0.03100	0.00197			

Estimador

β 0.099134

α -11.670418

Celda Objetivo 11.226579

Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.

5.- De la pestaña Datos del menú se selecciona Solver, para estimar los parámetros α y β , esto se realiza mediante la minimización de la celda objetivo, cambiando sólo las celdas que corresponden a los parámetros y el método Simplex LP. Se procede a llenar la columna G, con la ecuación $q_x = \frac{e^{\alpha+\beta x}}{(e^{\alpha+\beta x}+1)}$ es importante recordar que las probabilidades de fallecer entre edad 12 y 79 no serán modificadas.

C.5 Ajuste de parámetros de la teoría de extremos

1.- En una hoja de Excel se colocan en las celdas de la B4 a la G4 los siguientes encabezados: Edad, q_x , m_x , E_x , *Sumando*, q'_x . En la columna B, entre las celdas B5: B93, se colocarán las edades de 12 años hasta los 100 años, en las celdas C5:C93 se colocaran las probabilidades de fallecimiento, que corresponderán a la edad de la celda en B. En la columna D, entre las celdas D5:D93, se aplicará para la celda $D_i = C_i/(1-(C_i/2))$ con $i \in \{5,6,\dots, 92,93\}$ para calcular m_x . E_x representa el número de expuestos al riesgo⁵⁰.

En las celdas F4 y F5 se colocará un valor inicial, es indiferente la elección del valor, cabe mencionar que será el valor de los parámetros B y C (respectivamente) del modelo de Gompertz.

2.-En las celdas L1 y L2 se colocará un valor inicial, al igual que los dos anteriores es indiferente su elección, cabe mencionar que será el valor de los parámetros γ y β (respectivamente) de la función Generalizada de Pareto. La pantalla que se debe observar debe ser la siguiente:



En la Celda G2 se coloca el valor de la edad umbral para la que se realizó el ajuste (N = 90). Se procede al llenado de la columna F, de la siguiente manera:

Intervalo de edad	Celdas en Excel	Función que le corresponde	Fórmula de Excel
12 a 90 años	F5:F83	Gompertz	$F_i = E_i * (\ln(D_i) - \ln(\$F\$1) - B_i * \ln(\$F\$2))^2$ para $i \in \{5, 6, \dots, 83\}$
91 a 100 años	F84:F93	Generalizada de Pareto	$= E_i * (\ln(D_i) + \ln(\$L\$2) + \ln(1 + (\$L\$1 * (B_i - \$G\$2) / \$L\$2)))^2$ para $i \in \{84, 85, \dots, 93\}$

La celda C2= SUMA(F5:F83), que representará:

$$\text{Suma de Gompertz} = \sum_{x=12}^N E_x [\ln(\mu(x)) - \ln(B) + x \ln(C)]^2$$

⁵⁰ Expuestos al riesgo Anexo D.10

Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.

La celda I2=SUMA(F84:F92), que representará:

$$\text{Suma de Gompertz} = \sum_{x=N+1}^N E_x \left[\ln(\mu(x)) + \ln(\theta) + \ln\left(1 + \gamma \frac{(x-N)}{\theta}\right) \right]^2$$

Y finalmente N2=C2+I2, que será nuestro objetivo final. Al terminar lo anterior debe verse la siguiente pantalla:

The screenshot shows an Excel spreadsheet with the following data:

Edad	q _x	m _x	l _x	Suma de	q _x estimada
75	0.029309	0.0297447	379	37.70	0.037013
76	0.02967	0.0315196	322	3.67	0.034424
77	0.041249	0.0421175	258	68.77	0.031895
78	0.04953	0.050639	184	4.8113	0.027427
79	0.116679	0.1239079	125	1.8510	0.029023
80	0.053379	0.0569551	91	10.7	0.046666
81	0.046629	0.0477419	77	10.7	0.043418
82	0.064024	0.0661417	63	102.32	0.044223
83	0.050473	0.0527799	48	8.0	0.046102
84	0.0372399	0.0379464	44	97.5	0.061762
85	0.0425414	0.0443104	37	57.31	0.068785
86	0.0696217	0.0721849	29	0.4	0.077611
87	0.0775194	0.0806432	18	0.8	0.089040
88	0.060061	0.0625000	10	36.59	0.104422
89	0.0483972	0.0495050	10	80.6	0.126247
90	0.1395349	0.1300000	6	0.0	0.159459
91	0.0433015	0.0448750	5	136.12	0.171532

The Solver window shows the following parameters:

- Objetivo: \$C\$2
- Por: Min
- Restricciones: \$B\$1 > \$B\$1, \$C\$2 > 1.00001
- Método de resolución: Simplex LP

3.- La idea del algoritmo es estimar los parámetros como modelos independientes y después, realizar una estimación conjunta, por tal motivo ajustamos la función de Gompertz primero, es decir los parámetros de la función se aproximaron mediante la función Solver con el código de optimización Simplex LP. Es importante mencionar que las restricciones serán las mismas que anteriormente para el modelo de Gompertz es decir: B > 0, C > 1.

The screenshot shows an Excel spreadsheet with the following data:

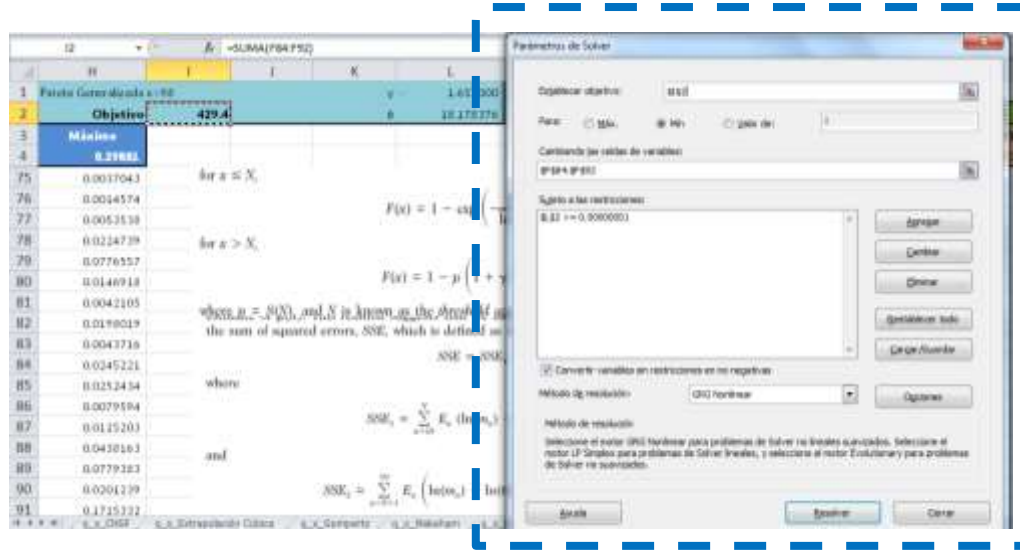
Edad	q _x	m _x	l _x	Suma de	q _x estimada
75	0.029309	0.0297447	379	37.70	0.037013
76	0.02967	0.0315196	322	3.67	0.034424
77	0.041249	0.0421175	258	68.77	0.031895
78	0.04953	0.050639	184	4.8113	0.027427
79	0.116679	0.1239079	125	1.8510	0.029023
80	0.053379	0.0569551	91	10.7	0.046666
81	0.046629	0.0477419	77	10.7	0.043418
82	0.064024	0.0661417	63	102.32	0.044223
83	0.050473	0.0527799	48	8.0	0.046102
84	0.0372399	0.0379464	44	97.5	0.061762
85	0.0425414	0.0443104	37	57.31	0.068785
86	0.0696217	0.0721849	29	0.4	0.077611
87	0.0775194	0.0806432	18	0.8	0.089040
88	0.060061	0.0625000	10	36.59	0.104422
89	0.0483972	0.0495050	10	80.6	0.126247
90	0.1395349	0.1300000	6	0.0	0.159459
91	0.0433015	0.0448750	5	136.12	0.171532
92	0.1323323	0.1428571	3	19.83	0.141911
93	0.066667	1.0000000	1	0.961480	0.961480
94	0.0	0.0	1	1.000000	1.000000

The Solver window shows the following parameters:

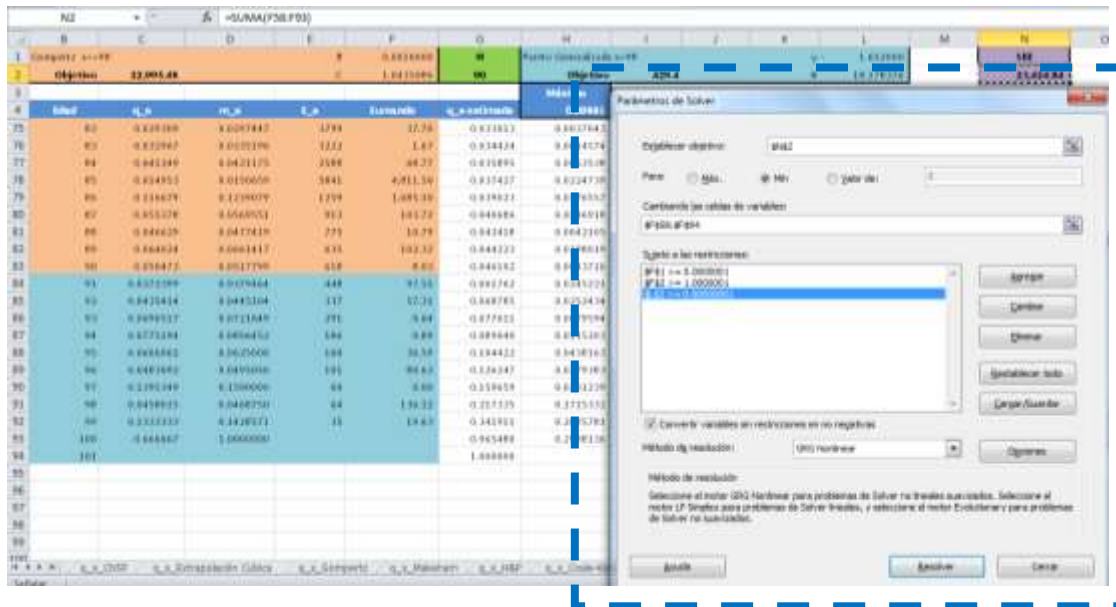
- Objetivo: \$C\$2
- Por: Min
- Restricciones: \$B\$1 > \$B\$1
- Método de resolución: Simplex LP

Se continúa con la Distribución Pareto Generalizada que se aproximaron mediante la función Solver con el código de optimización Simplex LP. Es importante mencionar que las restricciones serán las mismas que anteriormente para la Distribución Pareto Generalizada es decir: θ > 0.

Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.



Finalmente se realiza la estimación conjunta que contempla las condiciones de las dos funciones mediante la función Solver con el código de optimización Simplex LP.



Los parámetros ajustados proporcionaron el comportamiento de la curva de mortalidad mexicana 2000-I.

Anexos D (Tabla de mortalidad mexicana 2000-I, ajustada por diferentes metodologías)

Se presentarán las diferentes tablas de mortalidad ajustadas por cada una de las metodologías descritas en la presente investigación. Cada una de ellas surgió a partir de la estimación de parámetros obtenidos al resolver cada una de las metodologías anteriormente descritas.

Anexos D.1 Muestra de mortalidad mexicana CNSF-2000-I. Datos base

x	q _x	x	q _x	x	q _x
12	0.00173555	44	0.00333434	76	0.0236478
13	0.00028559	45	0.00338074	77	0.02393617
14	0.00017888	46	0.00378611	78	0.02945642
15	0.00038278	47	0.00409647	79	0.03151148
16	0	48	0.00441878	80	0.02599795
17	0.00043454	49	0.00449774	81	0.03728661
18	0.00048778	50	0.00476001	82	0.02888548
19	0.00073447	51	0.00585803	83	0.03243243
20	0.00055584	52	0.00642784	84	0.04041528
21	0.00068899	53	0.00731363	85	0.0148423
22	0.0005357	54	0.00769883	86	0.11024735
23	0.00049599	55	0.00771448	87	0.05388601
24	0.00080354	56	0.00780264	88	0.0455665
25	0.00086567	57	0.00909015	89	0.0620384
26	0.00072942	58	0.01000472	90	0.04923077
27	0.00110351	59	0.01100962	91	0.03655914
28	0.00098588	60	0.01240635	92	0.04261364
29	0.00115522	61	0.01156762	93	0.06730769
30	0.00121365	62	0.01156022	94	0.07462687
31	0.00119163	63	0.01508277	95	0.05882353
32	0.00114252	64	0.01329229	96	0.04716981
33	0.00124875	65	0.0153311	97	0.13043478
34	0.00125002	66	0.01761262	98	0.04477612
35	0.00132664	67	0.01626784	99	0.125
36	0.00139602	68	0.01785309	100	1
37	0.00150096	69	0.02108164		
38	0.00176652	70	0.01965096		
39	0.00183761	71	0.02110456		
40	0.00222425	72	0.06092079		
41	0.00202826	73	0.02332257		
42	0.00240188	74	0.02742846		
43	0.00283573	75	0.02528121		

Fuente: Construcción a partir de los datos obtenidos de Mendoza, M., Madrigal, A. & Gutiérrez-Peña, E., 2000. Predictive Mortality Graduation and the Value At Risk: A Bayesian Approach, México pp. 17.

Anexos D.2 Tabla de mortalidad mexicana ajustada mediante el modelo de extrapolación polinomial

x	q _x	x	q _x	x	q _x	x	q _x
12	0.00018958	46	0.002186	80	0.02672	114	0.25854978
13	0.00019815	47	0.002354	81	0.028724	115	0.27372836
14	0.00020876	48	0.002535	82	0.030874	116	0.28959699
15	0.00022141	49	0.00273	83	0.03318	117	0.30617365
16	0.00023609	50	0.00294	84	0.035651	118	0.32347636
17	0.00025282	51	0.003166	85	0.0383	119	0.3415231
18	0.00027159	52	0.00341	86	0.041136	120	0.36033189
19	0.00029239	53	0.003672	87	0.044174	121	0.37992072
20	0.00031523	54	0.003954	88	0.047424	122	0.4003076
21	0.00034012	55	0.004258	89	0.050902	123	0.42151051
22	0.00036704	56	0.004585	90	0.054619	124	0.44354747
23	0.000396	57	0.004938	91	0.058592	125	0.46643646
24	0.000427	58	0.005317	92	0.062834	126	0.4901955
25	0.00046	59	0.005725	93	0.067362	127	0.51484258
26	0.000495	60	0.006164	94	0.07219	128	0.5403957
27	0.000533	61	0.006637	95	0.077337	129	0.56687287
28	0.000575	62	0.007145	96	0.082817	130	0.59429207
29	0.000619	63	0.007693	97	0.088649	131	0.62267132
30	0.000667	64	0.008282	98	0.09485	132	0.65202861
31	0.000718	65	0.008915	99	0.101436	133	0.68238194
32	0.000773	66	0.009597	100	0.108424	134	0.71374931
33	0.000833	67	0.01033	101	0.115832	135	0.74614872
34	0.000897	68	0.011119	102	0.12367804	136	0.77959817
35	0.000966	69	0.011967	103	0.13198012	137	0.81411567
36	0.001041	70	0.012879	104	0.14075625	138	0.84971921
37	0.001121	71	0.01386	105	0.15002442	139	0.88642679
38	0.001207	72	0.014914	106	0.15980262	140	0.92425641
39	0.0013	73	0.016048	107	0.17010887	141	0.96322607
40	0.0014	74	0.017265	108	0.18096116	142	1.00335377
41	0.001508	75	0.018574	109	0.1923775		
42	0.001624	76	0.01998	110	0.20437587		
43	0.001749	77	0.02149	111	0.21697429		
44	0.001884	78	0.023111	112	0.23019074		
45	0.002029	79	0.024851	113	0.24404324		

Fuente: Elaboración propia

Anexos D.3 Tabla de mortalidad mexicana ajustada mediante el modelo de Gompertz

x	q _x	x	q _x	x	q _x	x	q _x
12	0.000379	46	0.003471	80	0.031407	114	0.255445
13	0.000404	47	0.003704	81	0.033487	115	0.270329
14	0.000431	48	0.003953	82	0.035703	116	0.285938
15	0.000461	49	0.004219	83	0.038063	117	0.302290
16	0.000492	50	0.004502	84	0.040575	118	0.319404
17	0.000525	51	0.004805	85	0.043250	119	0.337294
18	0.000560	52	0.005128	86	0.046097	120	0.355974
19	0.000598	53	0.005472	87	0.049126	121	0.375454
20	0.000638	54	0.005840	88	0.052349	122	0.395744
21	0.000681	55	0.006232	89	0.055777	123	0.416849
22	0.000727	56	0.006651	90	0.059424	124	0.438772
23	0.000776	57	0.007097	91	0.063301	125	0.461512
24	0.000828	58	0.007574	92	0.067421	126	0.485064
25	0.000884	59	0.008082	93	0.071801	127	0.509420
26	0.000943	60	0.008624	94	0.076453	128	0.534568
27	0.001007	61	0.009202	95	0.081394	129	0.560490
28	0.001075	62	0.009819	96	0.086641	130	0.587165
29	0.001147	63	0.010477	97	0.092209	131	0.614568
30	0.001224	64	0.011179	98	0.098116	132	0.642668
31	0.001307	65	0.011928	99	0.104382	133	0.671431
32	0.001395	66	0.012726	100	0.111024	134	0.700816
33	0.001488	67	0.013578	101	0.118062	135	0.730779
34	0.001589	68	0.014486	102	0.125517	136	0.761273
35	0.001696	69	0.015454	103	0.133409	137	0.792245
36	0.001810	70	0.016487	104	0.141760	138	0.823640
37	0.001931	71	0.017588	105	0.150592	139	0.855397
38	0.002061	72	0.018762	106	0.159926	140	0.887455
39	0.002200	73	0.020013	107	0.169786	141	0.919749
40	0.002348	74	0.021347	108	0.180195	142	0.952213
41	0.002506	75	0.022769	109	0.191174	143	0.984778
42	0.002675	76	0.024285	110	0.202749	144	1.017375
43	0.002855	77	0.025899	111	0.214940		
44	0.003047	78	0.027620	112	0.227772		
45	0.003252	79	0.029454	113	0.241266		

Fuente: Elaboración propia

Anexos D.4 Tabla de mortalidad mexicana ajustada mediante el modelo de Makeham

x	q _x	x	q _x	x	q _x	x	q _x
12	0.00059	50	0.00613	88	0.08162	126	0.70155
13	0.00063	51	0.00656	89	0.08725	127	0.72655
14	0.00066	52	0.00702	90	0.09325	128	0.75104
15	0.00069	53	0.00751	91	0.09964	129	0.77487
16	0.00073	54	0.00804	92	0.10645	130	0.79790
17	0.00077	55	0.00861	93	0.11368	131	0.81998
18	0.00081	56	0.00922	94	0.12138	132	0.84099
19	0.00086	57	0.00987	95	0.12956	133	0.86080
20	0.00091	58	0.01056	96	0.13825	134	0.87931
21	0.00096	59	0.01131	97	0.14746	135	0.89644
22	0.00102	60	0.01211	98	0.15724	136	0.91211
23	0.00108	61	0.01297	99	0.16760	137	0.92629
24	0.00115	62	0.01389	100	0.17857	138	0.93897
25	0.00122	63	0.01488	101	0.19017	139	0.95015
26	0.00129	64	0.01593	102	0.20243	140	0.95987
27	0.00137	65	0.01706	103	0.21537	141	0.96821
28	0.00146	66	0.01827	104	0.22901	142	0.97523
29	0.00155	67	0.01957	105	0.24338	143	0.98104
30	0.00165	68	0.02096	106	0.25849	144	0.98577
31	0.00176	69	0.02245	107	0.27435	145	0.98954
32	0.00187	70	0.02404	108	0.29099	146	0.99248
33	0.00199	71	0.02574	109	0.30841	147	0.99472
34	0.00213	72	0.02757	110	0.32662	148	0.99639
35	0.00227	73	0.02952	111	0.34561	149	0.99760
36	0.00242	74	0.03161	112	0.36538	150	0.99845
37	0.00258	75	0.03385	113	0.38591	151	0.99903
38	0.00275	76	0.03624	114	0.40720	152	0.99941
39	0.00294	77	0.03880	115	0.42921	153	0.99966
40	0.00314	78	0.04154	116	0.45190	154	0.99981
41	0.00335	79	0.04446	117	0.47523	155	0.99990
42	0.00358	80	0.04759	118	0.49915	156	0.99995
43	0.00383	81	0.05094	119	0.52359	157	0.99997
44	0.00409	82	0.05451	120	0.54848	158	0.99999
45	0.00438	83	0.05832	121	0.57372	159	0.99999
46	0.00468	84	0.06240	122	0.59923	160	1.00000
47	0.00501	85	0.06675	123	0.62489		
48	0.00535	86	0.07139	124	0.65058		
49	0.00573	87	0.07634	125	0.67618		

Fuente: Elaboración propia.

Anexos D.5 Tabla de mortalidad mexicana ajustada mediante el modelo de Heligman & Pollard

x	q _x	x	q _x	x	q _x
12	0.0336173	49	0.0336177	86	0.0449051
13	0.0336177	50	0.0336177	87	0.0490120
14	0.0336180	51	0.0336176	88	0.0545801
15	0.0336182	52	0.0336176	89	0.0621013
16	0.0336184	53	0.0336176	90	0.0722106
17	0.0336186	54	0.0336176	91	0.0857082
18	0.0336187	55	0.0336177	92	0.1035700
19	0.0336188	56	0.0336179	93	0.1269313
20	0.0336189	57	0.0336181	94	0.1570198
21	0.0336190	58	0.0336185	95	0.1950164
22	0.0336190	59	0.0336190	96	0.2418225
23	0.0336191	60	0.0336198	97	0.2977484
24	0.0336191	61	0.0336209	98	0.3621861
25	0.0336191	62	0.0336224	99	0.4333947
26	0.0336191	63	0.0336245	100	0.5085444
27	0.0336191	64	0.0336275	101	0.5840961
28	0.0336190	65	0.0336316	102	0.6564352
29	0.0336190	66	0.0336372	103	0.7225340
30	0.0336190	67	0.0336449	104	0.7803999
31	0.0336189	68	0.0336555	105	0.8291899
32	0.0336189	69	0.0336701	106	0.8690408
33	0.0336188	70	0.0336901	107	0.9007541
34	0.0336188	71	0.0337176	108	0.9254728
35	0.0336187	72	0.0337552	109	0.9444294
36	0.0336186	73	0.0338067	110	0.9587871
37	0.0336186	74	0.0338773	111	0.9695592
38	0.0336185	75	0.0339741	112	0.9775839
39	0.0336184	76	0.0341066	113	0.9835304
40	0.0336184	77	0.0342881	114	0.9879196
41	0.0336183	78	0.0345366	115	0.9911500
42	0.0336182	79	0.0348767	116	0.9935224
43	0.0336181	80	0.0353422	117	1.0587870
44	0.0336181	81	0.0359791		
45	0.0336180	82	0.0368501		
46	0.0336179	83	0.0380406		
47	0.0336178	84	0.0396663		
48	0.0336178	85	0.0418841		

Fuente: Elaboración propia

Anexos D.6 Tabla de mortalidad mexicana ajustada mediante el modelo de Coale Kisker.

x	q _x	x	q _x	x	q _x
12	0.001737	48	0.004429	84	0.04817698
13	0.000286	49	0.004508	85	0.05296457
14	0.000179	50	0.004771	86	0.05847466
15	0.000383	51	0.005875	87	0.0648273
16	0.000409	52	0.006449	88	0.07216427
17	0.000435	53	0.007340	89	0.08065296
18	0.000488	54	0.007729	90	0.09049064
19	0.000735	55	0.007744	91	0.10190944
20	0.000556	56	0.007833	92	0.1151816
21	0.000689	57	0.009132	93	0.13062509
22	0.000536	58	0.010055	94	0.14860912
23	0.000496	59	0.011071	95	0.16955896
24	0.000804	60	0.012484	96	0.19395935
25	0.000866	61	0.011635	97	0.22235501
26	0.000730	62	0.011627	98	0.25534669
27	0.001104	63	0.015197	99	0.29358016
28	0.000986	64	0.013381	100	0.33772546
29	0.001156	65	0.015450	101	0.38844313
30	0.001214	66	0.017769	102	0.44633462
31	0.001192	67	0.016401	103	0.51187532
32	0.001143	68	0.018014	104	0.58533147
33	0.001250	69	0.021306	105	0.66666667
34	0.001251	70	0.019846	106	0.75544996
35	0.001328	71	0.021330	107	0.85078349
36	0.001397	72	0.062835	108	0.95127167
37	0.001502	73	0.023598	109	1.05505257
38	0.001768	74	0.027810		
39	0.001839	75	0.025605		
40	0.002227	76	0.023931		
41	0.002030	77	0.024226		
42	0.002405	78	0.029897		
43	0.002840	79	0.032016		
44	0.003340	80	0.034438		
45	0.003386	81	0.037208		
46	0.003793	82	0.040378		
47	0.004105	83	0.044010		

Fuente: Elaboración propia

Anexos D.7 Tabla de mortalidad mexicana ajustada mediante el modelo de Logit.

x	q _x	x	q _x	x	q _x	x	q _x	x	q _x
12	0.0017	46	0.0038	80	0.0232	114	0.4087	148	0.9526
13	0.0003	47	0.0041	81	0.0256	115	0.4329	149	0.9569
14	0.0002	48	0.0044	82	0.0282	116	0.4574	150	0.9608
15	0.0004	49	0.0045	83	0.0310	117	0.4821	151	0.9644
16	0.0004	50	0.0048	84	0.0341	118	0.5068	152	0.9676
17	0.0004	51	0.0059	85	0.0375	119	0.5316	153	0.9706
18	0.0005	52	0.0064	86	0.0413	120	0.5562	154	0.9733
19	0.0007	53	0.0073	87	0.0454	121	0.5805	155	0.9758
20	0.0006	54	0.0077	88	0.0499	122	0.6044	156	0.9780
21	0.0007	55	0.0077	89	0.0548	123	0.6279	157	0.9800
22	0.0005	56	0.0078	90	0.0602	124	0.6507	158	0.9819
23	0.0005	57	0.0091	91	0.0660	125	0.6729	159	0.9836
24	0.0008	58	0.0101	92	0.0724	126	0.6943	160	0.9851
25	0.0009	59	0.0111	93	0.0794	127	0.7150	161	0.9865
26	0.0007	60	0.0125	94	0.0869	128	0.7347	162	0.9877
27	0.0011	61	0.0116	95	0.0951	129	0.7536	163	0.9889
28	0.0010	62	0.0116	96	0.1040	130	0.7715	164	0.9899
29	0.0012	63	0.0152	97	0.1136	131	0.7885	165	0.9909
30	0.0012	64	0.0134	98	0.1240	132	0.8046	166	0.9917
31	0.0012	65	0.0154	99	0.1351	133	0.8197	167	0.9925
32	0.0011	66	0.0178	100	0.1472	134	0.8339	168	0.9932
33	0.0012	67	0.0164	101	0.1600	135	0.8472	169	0.9938
34	0.0013	68	0.0180	102	0.1738	136	0.8596	170	0.9944
35	0.0013	69	0.0213	103	0.1885	137	0.8711	171	0.9949
36	0.0014	70	0.0198	104	0.2042	138	0.8818	172	1.0000
37	0.0015	71	0.0213	105	0.2207	139	0.8918		
38	0.0018	72	0.0628	106	0.2383	140	0.9010		
39	0.0018	73	0.0236	107	0.2567	141	0.9095		
40	0.0022	74	0.0278	108	0.2761	142	0.9173		
41	0.0020	75	0.0256	109	0.2963	143	0.9245		
42	0.0024	76	0.0239	110	0.3174	144	0.9312		
43	0.0028	77	0.0242	111	0.3393	145	0.9373		
44	0.0033	78	0.0299	112	0.3618	146	0.9428		
45	0.0034	79	0.0320	113	0.3850	147	0.9480		

Fuente: Elaboración Propia

Anexos D.8 Tabla de mortalidad mexicana ajustada mediante el modelo de CNSF.

x	q _x	x	q _x	x	q _x
12	0.000396	51	0.007145	90	0.115832
13	0.000427	52	0.007693	91	0.123677
14	0.000460	53	0.008282	92	0.131973
15	0.000495	54	0.008915	93	0.140737
16	0.000533	55	0.009597	94	0.149983
17	0.000575	56	0.010330	95	0.159723
18	0.000619	57	0.011119	96	0.169970
19	0.000667	58	0.011967	97	0.180733
20	0.000718	59	0.012879	98	0.192020
21	0.000773	60	0.013860	99	0.203837
22	0.000833	61	0.014914	100	1.000000
23	0.000897	62	0.016048		
24	0.000966	63	0.017265		
25	0.001041	64	0.018574		
26	0.001121	65	0.019980		
27	0.001207	66	0.021490		
28	0.001300	67	0.023111		
29	0.001400	68	0.024851		
30	0.001508	69	0.026720		
31	0.001624	70	0.028724		
32	0.001749	71	0.030874		
33	0.001884	72	0.033180		
34	0.002029	73	0.035651		
35	0.002186	74	0.038300		
36	0.002354	75	0.041136		
37	0.002535	76	0.044174		
38	0.002730	77	0.047424		
39	0.002940	78	0.050902		
40	0.003166	79	0.054619		
41	0.003410	80	0.058592		
42	0.003672	81	0.062834		
43	0.003954	82	0.067362		
44	0.004258	83	0.072190		
45	0.004585	84	0.077337		
46	0.004938	85	0.082817		
47	0.005317	86	0.088649		
48	0.005725	87	0.094850		
49	0.006164	88	0.101436		
50	0.006637	89	0.108424		

Fuente: M Mendoza Ramírez, M., Madrigal Gómez, A. M. & Martínez Torres, E., 2000 "Tablas de Mortalidad CNSF2000-I y CNSF 2000-G", CNSF, 2000.

Anexos D.9 Tabla de mortalidad mexicana ajustada mediante el modelo de Valores extremos.

x	q _x	x	q _x	x	q _x
12	0.001702	46	0.00721981	80	0.03035892
13	0.001776	47	0.00753275	81	0.03165867
14	0.001853	48	0.0078592	82	0.03301311
15	0.001933	49	0.00819974	83	0.03442446
16	0.002017	50	0.00855497	84	0.03589502
17	0.002105	51	0.00892551	85	0.03742717
18	0.002197	52	0.00931204	86	0.03902339
19	0.002292	53	0.00971522	87	0.04068623
20	0.002392	54	0.01013576	88	0.04241836
21	0.002495	55	0.01057442	89	0.04422251
22	0.002604	56	0.01103195	90	0.04610154
23	0.002717	57	0.01150916	91	0.06176197
24	0.002835	58	0.01200689	92	0.06878473
25	0.002958	59	0.01252601	93	0.07761119
26	0.003087	60	0.01306742	94	0.08903969
27	0.003221	61	0.01363207	95	0.10442232
28	0.003361	62	0.01422095	96	0.12624745
29	0.003507	63	0.01483507	97	0.15965874
30	0.003659	64	0.0154755	98	0.21733471
31	0.003818	65	0.01614336	99	0.34191145
32	0.003984	66	0.01683979	100	0.96548023
33	0.004157	67	0.01756599	101	1.002634
34	0.004337	68	0.01832322		
35	0.004525	69	0.01911277		
36	0.004722	70	0.01993599		
37	0.004927	71	0.0207943		
38	0.005141	72	0.02168915		
39	0.005364	73	0.02262206		
40	0.005596	74	0.02359461		
41	0.005839	75	0.02460845		
42	0.006092	76	0.02566527		
43	0.006357	77	0.02676685		
44	0.006632	78	0.02791504		
45	0.006920	79	0.02911173		

Fuente: Elaboración Propia

Anexos D.10 Tabla de mortalidad mexicana ajustada mediante el modelo de Valores extremos.

x	E_x	x	E_x	x	E_x
12	9203	46	162347	80	4758
13	7001	47	151459	81	4286
14	16768	48	141042	82	3799
15	10446	49	128595	83	3222
16	9654	50	117505	84	2588
17	16102	51	106745	85	5841
18	18442	52	98154	86	1259
19	20408	53	87818	87	913
20	25173	54	79525	88	775
21	33359	55	72288	89	635
22	50374	56	65234	90	618
23	74561	57	59083	91	448
24	99480	58	52445	92	337
25	122342	59	48508	93	291
26	147954	60	44817	94	186
27	165651	61	40417	95	160
28	184425	62	36168	96	101
29	201460	63	32128	97	60
30	218085	64	28208	98	64
31	231340	65	24342	99	35
32	242170	66	20805		
33	249538	67	18746		
34	254078	68	17329		
35	261217	69	15277		
36	260378	70	13819		
37	258114	71	14286		
38	250332	72	12116		
39	242261	73	10888		
40	231921	74	9822		
41	224367	75	8752		
42	211408	76	7762		
43	201843	77	6973		
44	188313	78	6392		
45	174223	79	5440		

Fuente: Mendoza, M., Madrigal, A. & Gutiérrez-Peña, E., 2000. Predictive Mortality Graduation and the Value At Risk: A Bayesian Approach, México pp. 17

Glosario

Cohorte. Es un conjunto de personas que comparten un mismo suceso dentro de un cierto periodo temporal.

Esto quiere decir que una cohorte, por ejemplo, puede estar formada por todas las personas nacidas en una ciudad X entre 1970 y 1972.

Contingente. (Del lat. contingens, -entis, part. act. de contingere, tocar, suceder).

1. adj. Que puede suceder o no suceder.
2. m. Parte que cada uno paga o pone cuando son muchos quienes contribuyen para un mismo fin.

Daño. ~ emergente. 1. m. Der. Valor de la pérdida sufrida o de los bienes destruidos o perjudicados.

Dominio de una función. Está formado por aquellos valores de x (para el caso de la mortalidad mexicana son las edades) para los que se puede calcular la imagen $f(x)$.

Estocástico. (Del gr. στοχαστικός, hábil en conjeturar).

1. adj. Perteneciente o relativo al azar.
2. f. Mat. Teoría estadística de los procesos cuya evolución en el tiempo es aleatoria, tal como la secuencia de las tiradas de un dado.

Muestra. (De mostrar).

1. f. Parte o porción extraída de un conjunto por métodos que permiten considerarla como representativa de él.

Muestra sesgada. Aquella en la que sectores importantes de la población no están debidamente representados. Si la muestra es sesgada no se pueden generalizar los resultados a la población.

Por ejemplo, en una encuesta telefónica decidir quién va a ganar las elecciones.

Recurrente. (Del ant. part. act. de recurrir; recurrens, -entis).

1. adj. Mat. Dicho de un proceso: Que se repite.

Recursivo, va. (Del lat. recursus, part. pas. de recurrere 'recurrir', e -ivo).

1. adj. Sujeto a reglas o pautas recurrentes. La capacidad recursiva del lenguaje.

Riesgo. (Del it. risico o rischio, y este del ár. clás. rizq, lo que depara la providencia).

1. m. Contingencia o proximidad de un daño.
2. m. Cada una de las contingencias que pueden ser objeto de un contrato de seguro.

Simulación. f. Acción de simular. Reproducción de un fenómeno real mediante otro más sencillo y más adecuado para ser estudiado.

Varianza. 1. f. Estad. Media de las desviaciones cuadráticas de una variable aleatoria, referidas al valor medio de esta.

Bibliografía

- Alley., W. M. (1984). "The Palmer Drought Severity Index: Limitations and Assumptions". U.S Geological Survey, Reston.
- Borges Peña, R. E. (2005). *Análisis de Sobrevivencia utilizando el lenguaje R*. Paipa, Boyacá, Colombia.: Consejo de Desarrollo Científico, Humanístico y Tecnológico de la Universidad de los Andes.
- Botella-Rocamora, P., Martínez-Beneito, M., & Alacreú-García, M. (Febrero de 2013). Regresión Logística. Valencia: Universidad Cardenal Herrera.
- Carriere, J. (1992). Parametric Models for life tables. *Transactions of society of Actuaries*, 44, 77 a 99.
- Carrillo Armenta, J. (Junio 2005). México, ¿Un país de viejos? *Gaceta Universitaria*, *jufracarrillo@hotmail.com*, 9.
- Castillo, E., & Moreno, E. (1983). Nuevos modelos de distribuciones de extremos basados en aproximaciones en las ramas. *Trabajos de Estadística e Investigación Operativa*, 34(3), 6 a 24.
- Castorena, G. M. (1980). *Análisis historioco de las sequías en México*. México.
- Cebrián Guajardo, A. (2011). *Análisis, modelización y predicción de episodios de sequía*. España: Universidad dde Zaragoza, Tesis Doctoral.
- Coelho, E., Graca Magalhanes, M., & Miguel Bravo, J. (2005). *Mortality and Longevity projections for the oldest-old in Portugal*. Portugal: Statistic Portugal, University or Évora.
- Correa, J., & González, N. (2002). *Gráficos Estadísticos con R*. Medellín: Universidad Nacional-Sede Medellín.
- de Vicente Merino, A., Hernández Mach, J., Albarrán Lozano, I., & Ramírez Pérez, C. (Febrero de 2000). "PROYECCIÓN Y ESTUDIO DE UNA POBLACIÓN. EL PAPEL DE LA MORTALIDAD". España: Swiss Re, Munchener, Bayerische y Nacional de reaseguros.
- Debón , A., Montes, F., & Sala, R. (8-11 Abril 2003). Graduación de datos de Mortalidad. 27 *Congreso Nacional de Estadística e investigación Operativa* (págs. 4-7). Valencia, España: Leida.
- Debón , A., Montes, F., & Sala, R. (Abril 2003). Graduación de datos de mortalidad. España: Congreso Nacional de Estadística e Investigación Operativa.
- Deubon Acejo, A., Montes Suay, F., & Sala Garrido, R. (Julio de 2008). Tablas de mortalidad dinámicas para España. Una aplicación a la hipoteca inversa. Valencia: ICO.
- Domínguez, R. e. (2007). *Sequías* (2 ed.). México: CENAPRED.
- Erdely, A. (2011). Simulación y graduación de las tablas de mortalidad por medio de cópulas. México: XXV Congreso Nacional de Actuarios 2011.
- Espin García , O. (2009). ANÁLISIS DE SUPERVIVENCIA PARA DIABETES EN LA POBLACIÓN MEXICANA, UN ENFOQUE BAYESIANO. México: Universidad Nacional Autónoma de México.
- Estadística, I. N. (2009). Tablas de mortalidad Metodología. España: INE.
- Felipe Checa, A., Guillén Estany, M., & Betzuen Zalbideigitia, A. (1997). Modelos de tablas de mortalidad en España Situación Actual. Barcelona: Universidad de Barcelona.

Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.

- García Perez , A. (2004). LA TEORÍA DEL VALOR EXTREMO: UNA APLICACIÓN AL SECTOR ASEGURADOR. Castilla: VII congreso Hispano-Italiano de Matemáticas Financieras y Actuariales.
- Gil Fana, J. A., Heras Martínez, A., & Vilar Zanón , J. L. (1999). *Matemáticas de los Seguros de Vida*. Madrid, España: Mapfre.
- Gobierno de los Estado Unidos Mexicanos, P. d. (2007). *Plan Nacional de Desarrollo 2007-2012*. México.
- Godoy Aguilar , A. (2009). Introducción a análisis de supervivencia con R. México: Universidad Nacional Autónoma de Mexico.
- Gompertz, B. (1825). *On the nature of the Function Expressive of the law of Human Mortality and on a New Mod of Determining the Value of Life Contingencies*. Recuperado el 5 de Noviembre de 2013, de Royal Society Publishing: <http://www.jstor.org/stable/107756>
- Gonzalo, J., & Olmo, J. (2003). Which Extreme Values are Really Extremes ? Montreal: DGCYT Conference on Extremal Events in Finance.
- Graca Magallanes, M., & Coelho, E. (2006). MORTALITY AND LONGEVITY PROJECTIONS FOR THE OLDEST-OLD IN PORTUGAL. Portugal, Portugal, Portugal.
- Graunt, J. (1683). OBSERVATIONS NATURE THE Dublin-Bills OF MORTALITY. London: Polit. Set. Quart.
- Guerreschi, F. (2009/2010). Longevity risk e tavole di mortalità: una proposta per la popolazione Italiana. Milano: Università Cattolica del sacro cuore di Milano.
- Gumbel, E. (1958). *Statics of extremes*. Columbia Univ. Press.
- Gutierrez Robledo, L. (2002). México y la Revolución de la Longevidad. México: Instituto de Geriatria.
- Hernández G., A. (2001). *Análisis Bayesiano de Eventos Extremos vía el Model Generalizado de Pareto (GPD) con Estimación del Umbral*. Venezuela: Uniersidad de Carabobo Facultad de Ciencias y Tegnología.
- Himes, C., Preston, S., & Condran , G. (1994). A Relational Model of Mortality at Older Age in Low Mortality Countries. *Population Studies*, 48(Downloaded by UNAM Ciudad Universitaria at 8:32, 21 January 2013), 269 a 291.
- Horiuchi, S., & Coale, A. (1990). AGE PATTERNS OF MORTALITY FOR OLDER WOMEN: AN ANALYSIS USING THE AGE-SPECIFIC RATE OF MORTALITY CHANGE WITH AGE . *Mathematical Population Studies* , 2(4), 245 a 267.
- Ibañes Rosales , A. (Octubre 2011.). Análisis estadístico de valores extremos y aplicaciones. Granada, España: Universidad de Granada.
- León Vásquez R., Conztanzo J. (2006). Splines Cúbicos. (Valparaíso). Forsythe, G. E., Malcolm, M. A. and Moler, C. B. (1977) Computer Methods for Mathematical Computations.
- Linus Fang-Shu Chan, Cary Chi-Liang Tsai, & Chenghsien Tsai. (2010). Empirical Tests on a Relational Model of Mortality Rates with Applications to Internal Hedging. UK: NAIC.
- MC. Neil, A. (March 1997). Estimating the tails of loss severity Distributions using extreme Value Theory . Zurich: Deparment Mathematik.
- Medina Moral , E. (Diciembre de 2003). *Modelos de Elección Discreta*. Recuperado el 18 de Marzo de 2013, de http://www.uam.es/personal_pdi/economicas/eva/pdf/logit.pdf

Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.

- Mendoza García, M., & Tapia Colocía, G. (2010). Situación Demográfica de México 1910-2010. México: http://www.unfpa.org.mx/publicaciones/cuadro_4.pdf.
- Mendoza Ramírez, M., Madrigal Gómez, A., & Martínez Torrez, E. (2000). Tablas de Mortalidad CNSF 2000-I y CNSF 2000-G. México: Comisión Nacional de Seguros y Fianzas.
- Mendoza, M., Madrigal, A., & Gutierrez-Peña, E. (Enero 2000). Predictive Mortality Graduation and the Value At Risk A Bayesian Approach. México: Sistema Nacional de Investigadores.
- Muñoz Pradas, F. (2001). Modelos e historia de la mortalidad: una evaluación crítica. *Demografía Histórica XIX, II*(Segunda época), 183 a 224.
- Norström, F. (1997). *The Gompertz-Makeham distribution*. London: Inglaterra.
- Pavilla Milláres, J., & Escuder Vallés, R. (2003). El proceso estocástico de muerte. Diferentes estrategias para la elaboración de tablas recargadas. Análisis de sensibilidad. *Estadística Española, 45*(153), 253 a 274.
- Romero Ramírez, L. (2012). Predicción de las tablas de vida dinámicas hasta el año 2025 para México. México: Colegio de Postgraduados.
- SECRETARIA DE GOBERNACION, C. (2001). *DIAGNÓSTICO DE PELIGROS E IDENTIFICACION DE RIESGOS DE DESASTRES EN MÉXICO*. México.
- Sui-Hang Li, J., Hardy, M., & Seng Tan, K. (2008 Octubre). THRESHOLD LIFE TABLES AND THEIR APPLICATIONS. *NORTH AMERICAN ACTUARIAL JOURNAL, 12*(2), 99 a 115.
- Veres Ferrer, F. (2000). Obtención de Tablas de Mortalidad por comparación con las de otros ámbitos pasados. *Estudios de Economía Aplicada*(15), 187 a 196.
- Witten, M., & Satzer, W. (1992). Gompertz survival model parameters: Estimation and sensitivity. *Applicaton Mathematica Lett., 5*(1), 7 a 12.
- Zarruk, A., Villegas, A. M., & Ortiz, F. (2010). Aspectos Teóricos Tablas de Mortalidad Evolución en el sector Asegurador Colombiano. Colombia.

Referencias

- Carrill Armenta, J., 2005. *México, ¿un país de viejos?*. Gaceta Universitaria, 20 Junio, p. 9.
- http://academic.uprm.edu/eddiem/psic3001/HTMLobj-329/Microsoft_PowerPoint_-_poblaciones-muestras.pdf. Principios de Psicología. Prof. Eddie Marrero, Ph. D. UPR-RUM, Departamento de Ciencia Sociales, 06 de septiembre de 2013.

Páginas web

- CNN Expansión, 2011. CNN Expansion. [Online] Available at: <http://www.cnnexpansion.com/economia/2009/05/18/aseguradoras-suman-el-17-de-pib> [Vienes 07 Junio 2013].
- Instituto Nacional de Estadística. Madrid España, 2011. Instituto Nacional de Estadística.. [Online] <http://www.ine.es/jaxi/menu.do?type=pcaxis&path=%2Ft20%2Fp319a&file=inebase> [07 Junio 2013].

Cálculo de la edad máxima estimada de la tabla de mortalidad mexicana CNSF 2000-I, su importancia y sus aplicaciones.

- L. Himes, C., H. Preston, S. & A. Condran, G., 2010. Taylor & Francis Online. [Online] <http://dx.doi.org/10.1080/003247203100014779>. [Enero 2013].
- Naciones Unidas, F. d. P. d. I. N., n.d. UNFPA México. [Online] <http://www.unfpa.org.mx/situacion%20en%20mexico.php> [Marzo 2013].
- <http://cuentame.inegi.org.mx/poblacion/habitantes.aspx?tema=P>. Censo de Población y Vivienda 2010, realizado por el INEGI. [Agosto 07 Junio 2013]
- <http://cran.r-project.org>
- <http://definicion.de/cohorte/#ixzz2dNAYcBj8>.
- <http://www.rae.es/rae.html> [Septiembre 2013].
- <http://www.juntadeandalucia.es/averroes/centros-tic/41009822/helvia/aula/archivos/repositorio/500/563/Dominiodefunciones.pdf> [Septiembre 2013].