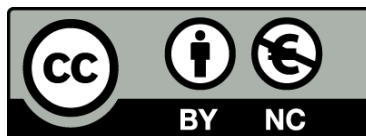




La Mortalidad y la Longevidad en la Cuantificación del Riesgo Actuarial para la Población de México

Arely Ornelas Vargas



Aquesta tesi doctoral està subjecta a la llicència **Reconeixement- NoComercial 3.0. Espanya de Creative Commons.**

Esta tesis doctoral está sujeta a la licencia **Reconocimiento - NoComercial 3.0. España de Creative Commons.**

This doctoral thesis is licensed under the **Creative Commons Attribution-NonCommercial 3.0. Spain License.**

PhD in Statistics

Título Tesis:

La Mortalidad y la Longevidad en la
Cuantificación del Riesgo Actuarial
para la Población de México

estudiante PhD:

Arely Ornelas Vargas

Dirigida por:

Montserrat Guillén i Estany
Catalina Bolancé Losilla

Fecha:

Mayo 2015



B Universitat de Barcelona

UNIVERSITAT DE BARCELONA

Departament d'Econometria, Estadística i Economia Espanyola

La Mortalidad y la Longevidad en la
Cuantificación del Riesgo Actuarial para la
Población de México

T E S I S

para obtener el grado de Doctor en Estadística

Autor:

Arely Ornelas Vargas

Dirigida por:

PhD. Montserrat Guillén i Estany

PhD. Catalina Bolancé Losilla



MAYO 2015, BARCELONA

Niuinti, nichoka, niknotlamati,
nik mati, nik itoa,
nik ilnamiki:
Ma ka aik nimiki
ma ka aik nipoliui.
In kan ajmikoa,
in kan on tepetua,
in ma onkan niau...
Ma ka aik nimiki,
ma ka aik nipoliui.

Estoy embriagado, lloro, me aflijo
pienso, digo,
en mi interior lo encuentro:
si yo nunca muriera,
si nunca desapareciera.
Allá donde no hay muerte
allá donde ella es conquistada,
que allá vaya yo...
Si yo nunca muriera,
si nunca desapareciera.

(Nezahualcóyotl 1402-1472)

Agradecimientos

Agradezco a Montserrat Guillén por guiarme en la preparación de este trabajo, por dedicarme tiempo, por tenerme paciencia y por solventar todas las dudas generadas en estos años de trabajo.

A Catalina Bolancé por su tiempo y buenos consejos. Gracias por todos los comentarios y sugerencias que me hiciste para finalizar de buena manera este proyecto.

A mis compañeros de doctorado que hacían mis días más divertidos.

A mis amigos en “ambos lados de del charco” por su apoyo y amistad.

A mi familia (incluidos los ausentes) por estar siempre conmigo . Gracias! Elvira, Nancy, Danaé y Samantha.

A Francisco por soportarme y respaldarme.

Esta Tesis doctoral no habría sido posible sin la financiación de la Agència de Gestió d’Ajuts Universitaris i de Recerca (AGAUR).

Gracias al Departament d’Econometria, Estadística i Economia Espanyola de la Universitat de Barcelona por la hospitalidad durante los años que estuve allí. En especial a todos los miembros del grupo de investigación *Riskcenter*.

Resumen

La planificación del futuro en los ámbitos demográfico, económico y actuarial es crucial. La buena planificación en programas sociales, presupuestos de gobierno, reservas actuariales, costo de seguros y pensiones, etc., depende del uso de un buen método para realizar pronósticos. Sin embargo, los constantes cambios en tecnología, estilos de vida, cambio climático, migración, por mencionar algunos, hacen que predecir ciertos fenómenos no sea una tarea fácil.

En particular, la mortalidad y la longevidad son eventos que impactan directamente en costos monetarios y por lo tanto necesitan una buena proyección a futuro. Ambos son indicadores del bienestar de la población, evalúan si los programas de salud y bienestar de la población son efectivos. Específicamente, en el mercado asegurador tener una alta mortalidad implica una subida de primas y, en general más reservas para hacer frente a los siniestros. La longevidad trae consigo situaciones similares cuando se trata el tema de anualidades y pensiones. A nivel social, alcanzar edades muy adultas implica la creación de programas de apoyo a los adultos mayores, por ejemplo, programas de dependencia, creación de albergues, lugares de ocio y campañas de salud.

El principal interés de esta Tesis es presentar opciones para una buena gestión de los costes de la mortalidad y la longevidad. La introducción presenta los antecedentes, la motivación y los objetivos. En el capítulo 1 se describen los datos que han sido utilizados para realizar los diferentes análisis: número de personas vivas y fallecimientos en México y tasas de mortalidad del sector asegurador. También se describen los cálculos necesarios para obtener la matriz con las tasas crudas de mortalidad de la población mexicana para los años 1990 a 2010 y edades de 0 a 100, siempre diferenciadas por sexos.

En el capítulo 2 se ha modelado el comportamiento de la mortalidad de la población general en México. Los modelos ajustados son el modelo Lee-Carter, el modelo Renshaw-Haberman y el modelo Age-Period-Cohort. Con las tasas de mortalidad estimadas por el modelo Lee-Carter, se ha ajustado un modelo relacional Brass-type para obtener nuevas tasas del mercado asegurador a partir de las existentes, pero corregidas por la mortalidad general.

En diciembre de 2013 entró en vigor en Europa una nueva directiva donde se establecía que los precios en bienes y servicios, lo que incluye las tarifas de seguros, no pueden ser diferentes para hombres y mujeres. En el capítulo 3 se ha abordado este tema conjuntándolo con el tema de la longevidad. Tomando

como edad inicial la edad 65, se estimó la función de supervivencia y con esta, se estimaron los parámetros ρ y λ de la distribución Weibull siguiendo la relación $S(t) = \exp(-\rho t^\lambda)$. Diferentes proporciones de hombres y mujeres fueron asumidas para este ejercicio. Con los parámetros estimados, fue posible evaluar el riesgo de longevidad, para esto se calculó el valor en riesgo (VaR, *Value-at-Risk*) a diferentes niveles de confianza α .

La Asociación Mexicana de Instituciones de Seguros (AMIS) desde el año 2000 y cada cinco años ha calculado tasas de mortalidad diferenciadas por sexo. Al disponer de esta información, en el capítulo 4 se ha decidido realizar el modelo Brass-type suponiendo que el número de muertos en cada edad y año sigue una distribución Poisson. De nueva cuenta el riesgo de longevidad se evaluó con el cálculo del VaR_α .

Después de haber utilizado distribuciones paramétricas se comprobó que es difícil describir de manera correcta la variable aleatoria número de supervivientes en cada edad, por tanto el VaR_α puede estar estimando de manera incorrecta el riesgo de longevidad. Así, en el capítulo 5 se ha calculado la función de distribución (CDF, *Cumulative Distribution Function*) utilizando métodos paramétricos y no paramétricos como el estimador núcleo clásico (CKE, *Classical Kernel Estimator*). También se propone utilizar el método semiparamétrico estimador núcleo doble transformado (DTKE, *Double Transformed Kernel Estimator*) (Alemany *et al.*, 2013). Más aún, para el ajuste de los modelos paramétricos se presenta una alternativa a la estimación máximo verosímil, que es la estimación vía minimizar el error cuadrático medio integrado (MISE, *Mean Integrated Square Error*), el cual también funciona como medida de bondad de ajuste y permite comparar tanto ajustes paramétricos como no paramétricos.

El capítulo 6 se basa en las aportaciones que la propia autora realizó en el trabajo titulado “La igualdad una exigencia para el seguro: problemas derivados de las tablas unisex”, que aparece en el libro *El seguro español ante los cambios del estado del bienestar*, publicado por la Fundación de Estudios Financieros (Guillén *et al.*, 2013).

La Tesis finaliza con conclusiones, destacando las principales aportaciones y con un capítulo dedicado a las líneas futuras de investigación.

En los anexos se incluyen los programas usados en los análisis realizados en esta Tesis. Para todos los capítulos se usó R Core Team (2014) y SAS Institute Inc. (2003).

Índice general

Agradecimientos	II
Resumen	III
Introducción	1
1. Descripción de los datos	7
1.1. Población de México	7
1.2. Conformación de la base de datos	10
1.3. Tasas de mortalidad	10
1.4. Población asegurada	11
1.4.1. Tasas de mortalidad AMIS-AMA	13
2. Mortalidad en la población general y asegurada de México. Comparación entre sexos	15
2.1. Introducción	15
2.2. Metodología	16
2.2.1. Modelos de mortalidad	16
2.2.2. Modelo relacional Brass-type	19
2.3. Resultados	20
2.3.1. Modelo Lee-Carter	24
2.3.2. Modelo Renshaw-Haberman	25
2.3.3. Modelo Age-Period-Cohort	27
2.3.4. Comparación de los modelos estimados	29
2.3.5. Modelo relacional Brass-type	31
2.4. Conclusiones	33
3. Uso de tablas de mortalidad unisex en el análisis de la longevidad	37
3.1. Introducción	37
3.2. Descripción de los datos	38
3.3. Metodología	39
3.3.1. Análisis de supervivencia	39
3.3.2. Valor en riesgo	40
3.3.3. Resultados	41
3.4. Conclusiones	43

4. Brass-type Poisson aplicado a tasas de mortalidad de asegurados	47
4.1. Introducción	47
4.2. Metodología	49
4.2.1. Modelo relacional Brass-type	49
4.3. Datos	50
4.4. Resultados	51
4.5. Conclusiones	55
5. Estimación del riesgo de longevidad, comparación de métodos paramétricos y noparamétricos	57
5.1. Introducción	57
5.2. Metodología	60
5.2.1. Distribuciones paramétricas	61
5.2.2. Estimación noparamétrica y semiparamétrica	66
5.3. Análisis de datos	73
5.4. Conclusiones	78
6. El uso de tablas de mortalidad unisex en el sector asegurador Europeo	83
6.1. El tribunal de justicia de la Unión Europea (UE) y el mercado asegurador	83
6.2. La esperanza de vida y longevidad por sexos: impacto en la tarificación unisex	86
6.3. La tarificación basada en factores de riesgo	89
6.4. Las tablas de mortalidad unisex en los sistemas de pensiones	92
6.5. Algunas consecuencias del uso de tablas unisex en el seguro de vida	94
6.6. Conclusiones	97
Conclusiones	99
Líneas futuras de investigación	103
Publicaciones	105
Bibliografía	108

Código R	119
Tasas de mortalidad	119
Ajuste de modelos con LifeMetrics	122
Elección del mejor modelo	128
Modelo relacional Brass-type lineal	132
Modelo relacional Brass-type lineal y Poisson. Datos de la AMIS . . .	136
Cálculo de la esperanza de vida y el valor en riesgo	142
Código SAS	147
Creación de la base de datos	147
Estimación de parámetros por máxima verosimilitud	148
Estimación de parámetros minimizando el error cuadrático medio integrado	149
Funciones de distribución paramétricas	152
Funciones de distribución noparamétricas	154
Funciones de distribución semiparamétricas	155
Bondad de ajuste	159
Cálculo de la función de distribución truncada	161
Cálculo del valor en riesgo	162
Cálculo del valor en riesgo empírico	163
Cálculo del valor en riesgo KQE	164
Cálculo del valor en riesgo exacto usando CKE	165
Cálculo del valor en riesgo exacto de la distribución truncada CKE .	167
Cálculo del valor en riesgo exacto usando DTKE	168
Cálculo del valor en riesgo exacto de la distribución truncada DTKE	174
Bootstrap	176
Gráficos en R	179
Histogramas	179
Parámetros estimados minimizando MISE	180
Funciones de distribución	181
Valor en riesgo	182

Índice de tablas

1.1. Datos población mexicana	9
2.1. BIC de los diferentes modelos ajustados	31
2.2. Parámetros estimados por el modelo Brass-type: $\log m_{x,t}^{aseg} = \hat{\beta}_0 + \hat{\beta}_1 \log m_{x,t}^w$. Siendo w el porcentaje de mujeres en la población. * denota p-valor ≤ 0.05	32
3.1. Parámetros estimados para la distribución Weibull según la relación lineal $\log S(t) = \log \hat{\rho} + \hat{\lambda} \log t$. Siendo w el porcentaje de mujeres en la población	41
3.2. Esperanza de vida a los 65 años para la población mexicana	42
3.3. Valor en riesgo al nivel de confianza 0.90: Número de años a sobrevivir después de la edad 65	42
3.4. Valor en riesgo al nivel de confianza 0.95: Número de años a sobrevivir después de la edad 65	42
4.1. Parámetros estimados por el modelo Brass-type de acuerdo a la regresión Poisson. DS indica desviación estándar	53
4.2. Valor en riesgo estimado: Número de años a sobrevivir después de la edad 65, con un nivel de confianza 0.90 y 0.95	55
5.1. Distribuciones paramétricas	62
5.2. Parámetros estimados por MLE y minimizando \widehat{MISE}	75
5.3. VaR_α : Número de meses estimado a sobrevivir después de la edad x	76
5.4. Media del VaR_α calculado a partir de las 500 muestras simuladas	78
5.5. Cociente de la desviación estándar del VaR_α calculado por la estimación DTKE y CKE en las 500 muestras simuladas	79
5.6. VaR_α : Número de años estimados a sobrevivir según nivel de confianza y distribución estimada	82

Índice de figuras

1.1. Diferencias de las tasas brutas de mortalidad en porcentaje entre hombres y mujeres	11
1.2. Diferencias de las tasas brutas de mortalidad en escala logarítmica entre hombres y mujeres	12
1.3. Logaritmo de las tasas brutas de mortalidad de la población asegurada mexicana en los años 2000, 2005 y 2010	14
2.1. Logaritmo de las tasas brutas de mortalidad de la población mexicana desde 1990 a 2010	22
2.2. Evolución de la mortalidad de México en hombres y mujeres desde 1990 a 2010	23
2.3. Comparativo entre el logaritmo de las tasas brutas de mortalidad para la población mexicana y de las tasas de mortalidad población asegurada (CNSF) en México	24
2.4. Parámetros estimados por el modelo Lee-Carter. Hombres y mujeres con edad 0 a 100 años en el periodo 1990 - 2010. Población mexicana	26
2.5. Parámetros estimados por el modelo Renshaw-Haberman. Hombres y mujeres con edad 0 a 100 años en el periodo 1990 - 2010. Población mexicana	28
2.6. Parámetros estimados por el modelo Age-Period-Cohort. Hombres y mujeres con edad 0 a 100 años en el periodo 1990 - 2010. Población mexicana	30
2.9. Tasas de mortalidad estimada con el modelo Lee-Carter utilizando los parámetros de la población suiza. La línea discontinua representa el logaritmo de las tasas estimadas para los asegurados por la CNSF en México	33
2.7. Residuales estandarizados $Z_{x,t}$ del ajuste de la población masculina. Los recuadros oscuros implican $Z_{x,t} < 0$ y los claros $Z_{x,t} > 0$	35
2.8. Residuales estandarizados $Z_{x,t}$ del ajuste de la población femenina. Los recuadros oscuros implican $Z_{x,t} < 0$ y los claros $Z_{x,t} > 0$	36

4.1.	Logaritmo de las tasas brutas de mortalidad de los asegurados (AMIS-AMA) en México de edad 12 a 99 en los años 2000, 2005 y 2010	52
4.2.	Comparativo entre el logaritmo de la mortalidad real de los asegurados y el logaritmo de la mortalidad estimada por el modelo Brass-type Poisson. Año 2010	54
5.1.	\widehat{MISE} para la distribución Weibull	63
5.2.	Función de distribución del tiempo de vida para los hombres. Distribución empírica (línea continua), estimación MLE (línea discontinua) y mínimo \widehat{MISE} (línea punteada)	64
5.3.	Función de distribución del tiempo de vida para las mujeres. Distribución empírica (línea continua), estimación MLE (línea discontinua) y mínimo \widehat{MISE} (línea punteada)	65
5.4.	Función de distribución del tiempo de vida. Distribución empírica (línea continua) y estimación núcleo (línea discontinua) . . .	68
5.5.	Función de distribución del tiempo de vida para hombres. Distribución empírica (línea continua), DTKE (línea discontinua) .	70
5.6.	Función de distribución del tiempo de vida para mujeres. Distribución empírica (línea continua), DTKE (línea discontinua) .	71
5.7.	Función de distribución del tiempo de vida para hombres. DTKE Lognormal (línea continua), DTKE Champernowne (línea discontinua) y DTKE Weibull (línea punteada)	72
5.8.	Función de distribución del tiempo de vida para mujeres. DTKE Lognormal (línea continua), DTKE Champernowne (línea discontinua) y DTKE Weibull (línea punteada)	73
5.9.	Histograma del número de supervivientes hombres	74
5.10.	Histograma del número de supervivientes mujeres	75
5.11.	VaR_α estimado en hombres una vez se ha alcanzado la edad 65 .	77
5.12.	VaR_α estimado en mujeres una vez se ha alcanzado la edad 65 .	78
5.13.	VaR_α estimado en hombres una vez se ha alcanzado la edad 85 .	79
5.14.	VaR_α estimado en mujeres una vez se ha alcanzado la edad 85 .	80

Introducción

En el curso del siglo XX la población a nivel mundial experimentó un cambio en la mortalidad: las tasas de mortalidad disminuyeron drásticamente. Esto se debió principalmente a la reducción de la mortalidad por enfermedades infecciosas afectando de manera notable a las edades infantiles y jóvenes. Además, se produjo una reducción de la mortalidad para edades adultas. En México, por ejemplo, la esperanza de vida al nacer se ha incrementado de 57.04 años en 1960 a 76.47 en 2009, lo que equivale a un incremento anual de 0.39 años. Si sólo se considera el cambio en la esperanza de vida de las dos últimas décadas, esta ha crecido anualmente en 0.27 y 0.33 años para mujeres y hombres, respectivamente.

La desaceleración del crecimiento de la esperanza de vida en México se puede asociar a las causas de defunción, como son las muertes por accidentes y violencia, así como el aumento de padecimientos crónicos tales como las enfermedades relacionadas con el corazón y la hipertensión, además de enfermedades de difícil y largo tratamiento como es el cáncer. Para la población masculina también se pueden atribuir los cambios al fenómeno migratorio al exterior del país, modificando de manera continua la composición y estructura de la población (Camposortega, 1997).

Es un hecho que la mortalidad ha disminuido en las edades más tempranas y se ha trasladado hacia edades más tardías, si bien ese es un logro en el aspecto demográfico, ya que se ha disminuido la mortalidad infantil y juvenil, en el área actuarial es motivo para replantear la manera en la que se ha estimado la mortalidad, esto con la finalidad de realizar una buena gestión del riesgo, hacer cálculos de reservas, rentas vitalicias, planes de pensiones y para poder calcular primas de los productos de vida.

El número de octogenarios, nonagenarios y centenarios está creciendo. En la población mexicana, el porcentaje de hombres de 80 y más años de edad ha pasado de 0.92 % en el año 2000 a 1.21 % en el año 2010, lo que representa un incremento de 53 %. En la población femenina, el porcentaje ha crecido de 1.13 % en el año 2000 a 1.49 % en el año 2010, lo que representa un incremento de 54 %. Este hecho se refleja el aumento de la esperanza de vida. De esta manera el mayor número de defunciones se ha movido cada vez más a edades más adultas.

El estudio de la edad de muerte se realiza utilizando estadísticos de orden como la mediana y el rango intercuartílico, ello se debe a la asimetría de dicha variable. La mediana de la edad de muerte en las mujeres en 2000 fue de 68 años y su rango intercuartílico (IQR, *Interquartile Range*)¹ es igual a 48 – 81, en 2010 la mediana fue de 72 años (IQR=55 – 83). Para los hombres en 2000 la mediana era de 60 años (IQR=35 – 76) y en el 2010 fue de 64 años (IQR=41 – 78). En el año 2010 la brecha en la longevidad de hombres y mujeres se mantuvo.

Naciones Unidas publica un informe anual sobre mortalidad en diferentes países, los cuales se clasifican en cuatro categorías, dependiendo de su nivel de mortalidad: baja, baja-intermedia, alta-intermedia y alta mortalidad. México pertenece a la segunda categoría; otros países de este grupo son: Belice, Corea, Barbados, Jamaica, Malasia, Mauricio, Seychelles, Sri Lanka, etc.. En estos la esperanza de vida al nacer es de alrededor de 77.4 años. En cambio, en los países con baja mortalidad, la esperanza de vida al nacer es cercana a los 85 años. También se reporta la esperanza de vida para ambos sexos combinados, a los 60 años en México fue de 20.5 años en el período 1990 – 1995 y 22.7 en el período 2010 – 2015 (Department of Economic and Social Affairs, 2013).

A lo largo de los años se han explorado diferentes modelos para describir la mortalidad. En la literatura occidental del siglo XIX ya se mencionaban conceptos relacionados con la mortalidad, las rentas vitalicias y la selección adversa de riesgos. En el Reino Unido se crearon las primeras tablas de mortalidad que incluyeron un margen para prever los cambios en la mortalidad, con la finalidad de proteger a las aseguradoras contra pérdidas. Estas tablas se construyeron con datos relativos a las aseguradoras en el período 1900-1920 (ver, por ejemplo, Pitacco *et al.*, 2009).

El estudio de la mortalidad condujo al desarrollo de los modelos de graduación, estos se han utilizado para suavizar las tasas de mortalidad crudas y para analizar el comportamiento de la mortalidad. Esta línea de estudio en la literatura actuarial está basada en gran medida en las denominadas Leyes de mortalidad de Heligman-Pollard, Gompertz-Makeham, entre otras. Con la aparición de nuevos métodos estadísticos, se revolucionó el análisis de la mortalidad. Entre estos, los Modelos Poisson, que estiman al número de muertes en cada edad, dado el número de personas expuestas al riesgo. Estos modelos

¹El rango intercuartílico se expresa a partir de los valores del cuartil uno y cuartil tres: $q_1 - q_3$

incluyen un componente de tendencia que capta la evolución de la mortalidad en el tiempo para cada edad y género. Modelos de supervivencia, que incorporan factores explicativos como es el modelo de riesgos proporcionales de Cox, que ajusta la probabilidad de que una persona sobreviva a una edad determinada; por lo tanto, es posible modelizar la duración de la vida condicionada a los valores de diversas variables explicativas. El modelo de Cox considera que la edad es una variable continua (Cox y Oakes, 1984).

Como hemos advertido anteriormente, tanto el modelo de Poisson como el de Cox pueden incluir factores de riesgo en sus análisis de la mortalidad, que no necesariamente tienen que limitarse a la variable género. Ejemplos evidentes son: características que describan el estilo de vida y el estado de salud, tendencias en el tiempo, así como interacciones entre variables.

Entre los métodos noparamétricos, se han utilizado la estimación núcleo clásica (Felipe *et al.*, 2001) y los modelos aditivos generalizados (GAM. *Generalized Additive Models*) (Debón *et al.*, 2006), la especificación ofrece el potencial para un mejor ajuste de los datos que los proporcionados por otros métodos. Esto es atribuible al hecho de que los efectos inducidos por factores de riesgo permiten formas más flexibles (Hastie y Tibshirani, 1990). Otro enfoque basado en métodos noparamétricos, que se centra en las edades de la vejez, se describe en Fledelius *et al.* (2004).

Entre los modelos estocásticos que se utilizan para ajustar la mortalidad se encuentran: el CMI Británico (1990) (CMI, *Continuous Mortality Investigation*), el de Lee y Carter (1992), el propuesto por Currie *et al.* (2004), el de Renshaw y Haberman (2006), el analizado por Cairns *et al.* (2006). Algunos de estos modelos, además de integrar el efecto del año en el que sucede la muerte, integran el efecto cohorte (es decir, grupo de personas nacidas en el mismo periodo). El efecto del año de nacimiento es importante pues se supone que distintas generaciones no están expuestas a los mismos eventos o avances tecnológicos, por ejemplo epidemias, guerras, fenómenos naturales, cura de enfermedades, vacunas, entre otros. Este enfoque facilita la predicción de la mortalidad debido a que los efectos del envejecimiento pueden variar de una cohorte a otra.

En lo que se refiere a la investigación de la mortalidad en México, en el ámbito demográfico como en el actuarial se han realizado trabajos destinados a modelizar y proyectar la mortalidad. Por ejemplo, Mina (1990) ha usado el método de Pollard para estimar las ganancias en la esperanza de vida a dife-

rentes edades en el periodo 1950 – 1980. Más recientemente en Mina (2006) se obtuvieron tablas de mortalidad abreviadas para los años 2003 – 2010 construidas a partir de la función Gompertz-Makeham.

Los splines han sido una herramienta recurrente para estimar las tasas de mortalidad en México. En Mina (2011) se utilizan polinomios de tercer orden para interpolar el número de supervivientes de una tabla de mortalidad abreviada (cada cinco años), el objetivo final consiste en hacer proyecciones de la mortalidad mexicana para los años 2010 – 2050. Dávila-Cervantes (2012) emplea el método de splines cúbicos para estimar las tasas de mortalidad en los años 2000, 2005 y 2010, en este caso se han utilizado los supervivientes para cada edad. Ambos autores destacan que el utilizar este método respeta los cambios de concavidad de la curva de mortalidad y que posibilita el cálculo de proyecciones a corto y largo plazo de manera fiable.

Aburto (2012) ha utilizado las defunciones de individuos con más de dos años de edad para ajustar el modelo Lee-Carter pero con una leve modificación, agrega una variable dependiente del tiempo a la ecuación del modelo. Además, este autor ha realizado un comparativo entre los resultados del ajuste Lee-Carter y los resultados del modelo aditivo doble multiplicativo de Wilmoth (Wilmoth, 1990), ambos ajustados a la población mexicana.

Otro artículo relacionado con la mortalidad en México es el de García y Ordorica (2012) que aplica el modelo de Lee-Carter a los datos de defunciones estimadas por el Instituto Nacional de Estadística Geografía e Informática y el Consejo Nacional de población en México en el periodo comprendido entre 1960 – 2000. Después proyecta la mortalidad a cinco años utilizando un modelo de series temporales $ARIMA(p, d, q)$ y confronta los resultados con los publicados por el Consejo Nacional de población en el año 2006.

En lo que respecta a la longevidad en los último año se han publicado el artículo Alonso *et al.* (2015), aquí se estudia la viabilidad del sistema de pensiones implantado en México en el año 1997. Los autores muestran como la economía sumergida, la poca contribución al fondo de pensiones por parte de los trabajadores y la baja tasa de reemplazo (entre personas de edad en edad de jubilación y el salario de los trabajadores en activo) son los principales inconvenientes del sistema. Su conclusión es que al incrementarse la esperanza de vida las contribuciones deberían reflejar el mismo efecto de manera automática. Este es un claro ejemplo de cuando la longevidad es un riesgo, pues de mantenerse el sistema con las facilidades actuales, como son el retiro por tiem-

po de trabajo sin importar la edad alcanzada, el dinero con el que se cuenta para pagar las pensiones sería insuficiente.

A nivel nacional el Instituto Nacional de Estadística, Geografía e Informática (INEGI) realiza proyecciones de diferentes indicadores demográficos, la más reciente publicación de resultados ha aparecido en diciembre de 2012 (Conapo, 2012), la cual presenta cifras para la planificación demográfica de los años 2010-2050. Las tasas de mortalidad han sido estimadas siguiendo el procedimiento de las Naciones Unidas para países en desarrollo, que desglosa la mortalidad utilizando componentes principales.

La mortalidad en el sector asegurador se estudia con menos frecuencia que la de la población general. Los análisis generalmente se hacen de manera interna en cada compañía utilizando su propia siniestralidad. Un problema en el sector asegurador es que las tablas de mortalidad que las compañías aseguradoras utilizan para reportar a la Comisión Nacional de Seguros y de Fianzas (CNSF) fueron hechas antes del año 2000, una mayor descripción de este hecho se encuentra en el capítulo 1.

Dada la falta de estudios de mortalidad recientes en México, en este trabajo nos planteamos alcanzar los siguientes objetivos:

1. Calcular las tasas de mortalidad para la población mexicana, para lo cual se describirá e implementará la metodología necesaria para su construcción.
2. Hacer un comparativo entre los resultados obtenidos en el punto anterior y las tablas de asegurados CNSF 2000-I.
3. Abordar la temática de la tarificación unisex.
4. Verificar si realizar tablas de mortalidad para cada sexo lleva a realizar una mejor estimación del riesgo que tener una tabla de mortalidad conjunta.
5. Presentar diferentes opciones para la estimación de la mortalidad.
6. Comparar la aproximación paramétrica y noparamétrica del riesgo de longevidad.

Además de realizar esta Tesis doctoral a la par he trabajado activamente en la investigación médica. En los últimos años he sido parte del equipo estadístico

de la *Fundació de Lluita contra la Sida* (<http://www.fl sida.org/>) donde he realizado el análisis estadístico de numerosas investigaciones relacionadas con el virus de la inmunodeficiencia humana (VIH).

La estadística es una parte fundamental en la investigación médica, su utilización va desde el diseño del estudio, por ejemplo calculando el número de muestra, pasando por la descripción de las variables, hasta métodos más sofisticados para comparar y/o relacionar variables con la finalidad de dar respuesta a la hipótesis que se desea comprobar. El artículo de Santos *et al.* (2012) es un claro ejemplo del uso de la estadística en las ciencias médicas; los autores investigan la resistencia al medicamento lopinavir/ritonavir en pacientes con el VIH, en primer lugar segmentan el número de pacientes en tres grupos según su exposición al medicamento y realizan los análisis descriptivos de las variables de interés; en segundo lugar se buscan diferencias entre grupos para una serie de variables descriptoras. Finalmente se utiliza una regresión logística multinomial para encontrar las variables que se asocian al fracaso en el uso del tratamiento con lopinavir/ritonavir.

Las técnicas estadísticas que se utilizan en estos proyectos de investigación no son tan distantes a los que se han utilizado en esta Tesis. El contexto es diferente, sin embargo la mortalidad siempre se puede traducir a otros ámbitos si se le considera como el final de un proceso, por ejemplo:

- En Tural *et al.* (2011) se desea conocer la eficacia de un tratamiento médico suministrado a un grupo de pacientes con VIH sobre el que se ha realizado el seguimiento de su evolución en doce semanas. Cada cierto periodo de tiempo se ha medido la concentración del medicamento en sangre, así como la carga viral del VIH que indica si el medicamento funciona o no. En este caso el análisis de datos longitudinales se utilizó para describir el cambio de la carga viral en los diferentes periodos e identificar las variables asociadas a dicho cambio.
- En Llibre *et al.* (2013) se estudia la mortalidad en enfermos de Neumonía (*Pneumocystis jiroveci*) en tres diferentes épocas según la disponibilidad de la Terapia antirretroviral de gran actividad (HAART, *highly active antiretroviral therapy*). Se utilizó el análisis de supervivencia para asociar diversos factores con la mortalidad mediante el modelo de riesgos proporcionales de Cox.

1 Descripción de los datos

En esta Tesis se ha decidido utilizar datos de la mortalidad de la población mexicana. En un primer análisis se han utilizado los datos demográficos de todo el país, cuya fuente son los Censos de Población y Vivienda y los Conteos de Población. En segundo lugar se utilizaron tasas de mortalidad de la población asegurada publicadas en el Diario Oficial de la Federación el 31 de marzo del 2000. Finalmente, se han conseguido tasas de mortalidad de asegurados, para los años 2000, 2005 y 2010, calculadas por dos instituciones: la Asociación Mexicana de Instituciones de Seguros (AMIS) y la Asociación Mexicana de Actuarios (AMA). A continuación se detallan tanto la obtención como el manejo de las bases de datos.

1.1. Población de México

La variable principal en este trabajo es la duración de la vida por lo que hace falta calcular la tasa de mortalidad cruda por sexo y edad. Para esto, es necesario tener el número de personas vivas y el número de defunciones registradas en cada año considerado en el estudio. Los organismos encargados de difundir esta información son el Instituto Nacional de Estadística, Geografía e Informática (INEGI) (<http://www.inegi.org.mx>) y el Consejo Nacional de Población (Conapo) (<http://www.conapo.gob.mx/>).

El Instituto Nacional de Estadística, Geografía e Informática es un organismo autónomo del gobierno mexicano, creado el 25 de enero de 1983, cuya función principal es proveer información de calidad tanto a la sociedad como al estado. El INEGI genera información estadística por medio de los censos de población cada diez años. Los censos de población y vivienda se levantan en México desde la época prehispánica, sin embargo hasta 1883, un año después de crearse la Dirección General de Estadística, no se decretó realizar censos de población cada diez años. Así para 1892 se realiza en la Ciudad de México un censo piloto de población, conocido como Censo Peñafiel, de ahí en adelante se ha realizado este ejercicio durante las pasadas 12 décadas (a excepción del año 1920 que sucedió la revolución mexicana). Los conteos de población que se han realizado en México son los de 1995 y 2005, se han planificado realizar cada cada diez años, concretamente en los años terminados en cinco; la dife-

rencia con el censo es que no se realiza a toda la población si no a una muestra aleatoria y estratificada. Los conteos representan un soporte estadístico para identificar y actualizar los cambios que el país experimenta en sus aspectos demográficos, económicos y sociales.

El Consejo Nacional de Población es una instancia gubernamental mexicana que tiene como misión la planificación demográfica del país, para ello formula programas de población y los vincula con los del desarrollo económico y social del sector público. Además, la Conapo realiza el análisis de datos que provienen de instituciones públicas como son, por un lado, nacimientos, defunciones, matrimonios y divorcios (provenientes del Registro Civil) y, por otro, datos judiciales en materia penal.

La recopilación de datos para este trabajo se hizo consultando las páginas web de las instituciones antes mencionadas. Los datos reflejan el comportamiento demográfico de 31 estados y un Distrito Federal, correspondiente al periodo 1990 – 2010. El número de fallecimientos estaba disponible para los años 1990 a 2010, por edad de 0 años a 120 años y diferenciado entre sexos.

En el caso del número de personas vivas la fuente directa ha sido el INEGI vía los Censos y Conteos de Población. Como ya se ha mencionado, los conteos de población se realizan cada 5 años, los años terminados en 5, y los censos cada 10 años, los años terminados en 0.

Los datos reportados están desagregados por edad de 0 a 100 y por sexo. Si bien en la página web del INEGI se encuentra información de todos los años donde se ha realizado censo o conteo de población, para este estudio sólo se han tomado los años 1990, 1995, 2000, 2005 y 2010, esto porque para los años anteriores a 1990 no se encuentra la información completamente separada por edades y/o sexo y principalmente porque la información de defunciones no está disponible para años previos a 1990.

El último censo de población se realizó en el año 2010, en este trabajo se utiliza la población contada casa a casa. El censo incluye 448,195 viviendas de las cuales se desconocía el número de ocupantes donde, dado la negativa debida a la inseguridad o que no se encontró a alguien que pudiera contestar el cuestionario, entre otros, estas viviendas no se incluyeron en nuestro análisis (Partida, 2008).

Una vez recolectada la información el problema siguiente, según reporta Partida (2008), es la existencia de fallos en los datos recogidos tales como la omisión de personas y la declaración de la edad. En este último caso se

puede tener el problema de que no se especifique la edad o que se declare de manera inexacta con cierta preferencia por algunos dígitos. A este respecto, la proporción de personas que no declaran su edad en los censos son 0.61 % en el año 1990, 2.11 % en el año 2000, y 0.05 % en el año 2010. Dentro de cada sexo estos porcentajes se mantienen. Para todas las viviendas censadas se conoce el género de sus habitantes.

El número de personas fallecidas de las cuales se desconocía el género en el año 1990 era de 0.25 %, mejorando sustancialmente en los años siguientes, para el año 2000 se tiene solo el 0.03 %, sin embargo para años más recientes, por ejemplo, en el año 2010 este porcentaje sube ligeramente a 0.06 %, lo cual se puede explicar con el número de muertes relacionadas con la situación actual del país, donde los fallecimientos relacionados con la delincuencia, principalmente por el narcotráfico, han ido incrementando. El mismo comportamiento se puede observar para los fallecimientos de hombres sin especificar edad, en el año 1990 este porcentaje era de 0.72 %, para el año 2000 de 0.58 % y para el año 2010 de 0.63 %. En cambio, para las mujeres ha descendido de manera constante, de 0.55 % en 1990 a 0.17 % en 2010. En la mayoría de años el desconocimiento conjunto de la edad y el sexo se mantiene por debajo del 0.01 %.

En los cálculos se excluyeron las cifras que no estaban clasificadas por sexo y/o edad. Se tomó esta decisión tratando de evitar el sesgo en las estimaciones, ya que la omisión de dicha información pudiera ser causada por dos motivos: muertes a edades muy tempranas, por lo que no se declaran el sexo, o, en edades avanzadas, porque se desconozca la edad exacta del difunto. Se sabe entonces que el error no es aleatorio y que asumir una distribución de los no especificados similar a la de la población provocaría sesgo en las estimaciones. El porcentaje de datos no clasificados es pequeño por lo que no se está perdiendo una gran cantidad de información.

La población mexicana, tanto hombres como mujeres, actualmente se compone principalmente de niños y jóvenes. Otros datos relevantes de la población mexicana se presentan en la tabla 1.1.

	Hombres	Mujeres
% Personas en edad laboral (16 – 65)	62.03 %	63.8 %
% Personas mayores a 65 años	5.39 %	6.02 %
% Muertes en edad laboral	46.21 %	32.77 %

Tabla 1.1: Datos población mexicana

1.2. Conformación de la base de datos

Una vez establecidas las características de la base de datos con la que se ha trabajado, en esta sección se describe el tratamiento que han recibido los datos antes de la estimación de modelos estocásticos de mortalidad.

En los años en donde no se realizó censo ni conteo de población se ha debido completar el número de personas vivas utilizando interpolación lineal, de acuerdo con Delwarde y Denuit (2003). Para cada sexo y edad se obtuvieron los datos para los años intermedios entre censo y contero de población, es decir, 1992 a 1994, 1996 a 1999, 2001 a 2004 y 2006 a 2010. La interpolación lineal entre dos años, t_1 y t_2 es la siguiente:

$$l_{x,t} = l_{x,t_1} + \left(\frac{t - t_1}{t_2 - t_1} \right) (l_{x,t_2} - l_{x,t_1}), \quad (1.1)$$

donde t es el año para el que se desea obtener el dato faltante. Con la interpolación de $l_{x,t}$ obtenemos las tasas de mortalidad para cualquier t .

1.3. Tasas de mortalidad

Las tasas crudas de mortalidad $m_{x,t}$ de la población general se han calculado como el cociente del número de defunciones registradas a la edad x en el tiempo t , $d_{x,t}$, y el número de personas vivas con edad cumplida x en el año t , $l_{x,t}$.

$$m_{x,t} = \frac{d_{x,t}}{l_{x,t}}. \quad (1.2)$$

Calculando el cociente entre las tasas de mortalidad de hombres y mujeres se observa que la mortalidad en los hombres es superior a la de las mujeres prácticamente para todas las edades, excepto para la etapa de la niñez y adolescencia. Las diferencias más notorias se dan en la franja de los 15 a los 50 años, principalmente la tasa de mortalidad en hombres puede ser 200% o 300% mayor a la de mujeres en edades comprendidas entre los 20 años hasta el inicio de los 30 años, notándose menos diferencias para las edades extremas: niñez y vejez. Las figuras 1.1 y 1.2 muestran las diferencias por sexos en el periodo 1990 y 2010, tanto en porcentaje como utilizando la comparación de curvas promedio de mortalidad en escala logarítmica.

En la figura 1.1 los colores más oscuros indican mayores diferencias en las

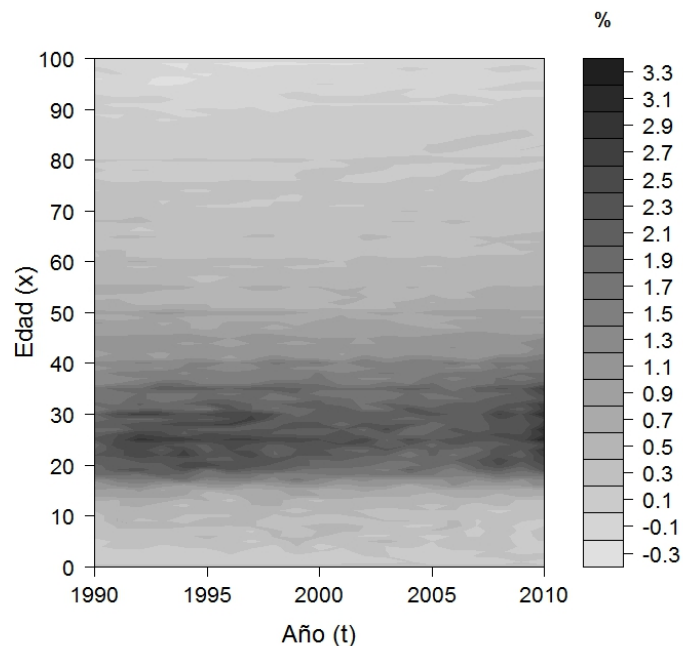


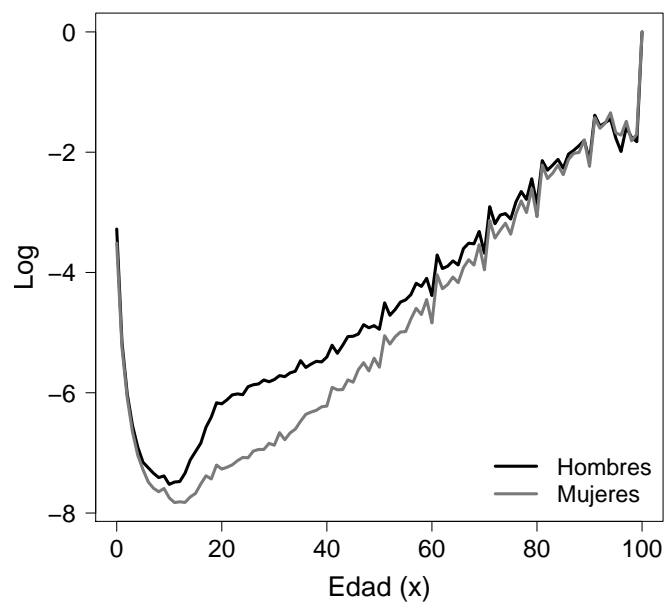
Figura 1.1: Diferencias de las tasas brutas de mortalidad en porcentaje entre hombres y mujeres

tasas brutas de mortalidad. Por tanto estas diferencias se dan entre los 18 y 40 años. En la figura 1.2 se muestra este mismo resultado, dado que las mayores diferencias entre ambas curvas se sitúan en las edades mencionadas.

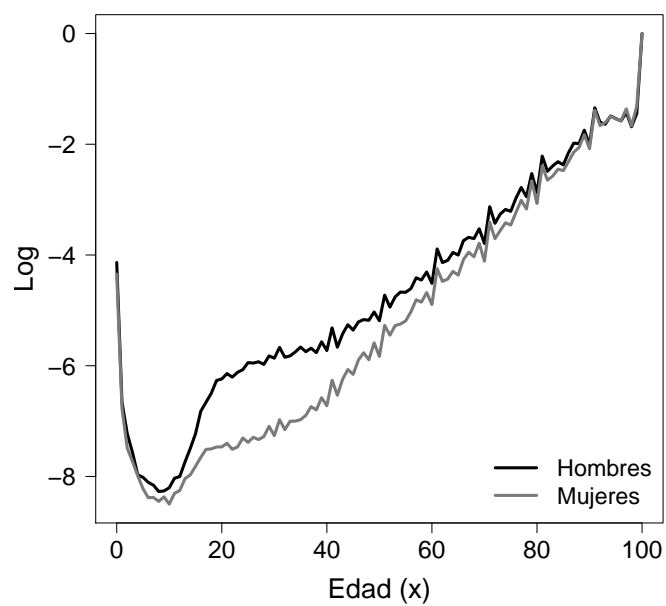
1.4. Población asegurada

Las tablas de mortalidad del sector asegurador mexicano son las publicadas en el Diario Oficial de la Federación en marzo del año 2000. La Comisión Nacional de Seguros y Fianzas (CNSF) es la entidad encargada de producir toda la información estadística de seguros, incluidas las tablas de mortalidad. Además supervisa la correcta operación de los sectores asegurador y financiero, tanto del cumplimiento de la normativa como de la solvencia necesaria para cumplir con las obligaciones contraídas.

En México, se crearon dos tablas dependiendo del tipo de seguro: Tasas de Mortalidad Individual CNSF 2000-I y Tasas de Mortalidad Grupo CNSF 2000-G. Estas tablas se utilizan en el cálculo de reservas de todas las pólizas suscritas después de abril del año 2000. Fueron construidas por Manuel Mendoza Ramírez, Ana María Madrigal Gómez y Evangelina Martínez Torres,



(a) 1990



(b) 2010

Figura 1.2: Diferencias de las tasas brutas de mortalidad en escala logarítmica entre hombres y mujeres

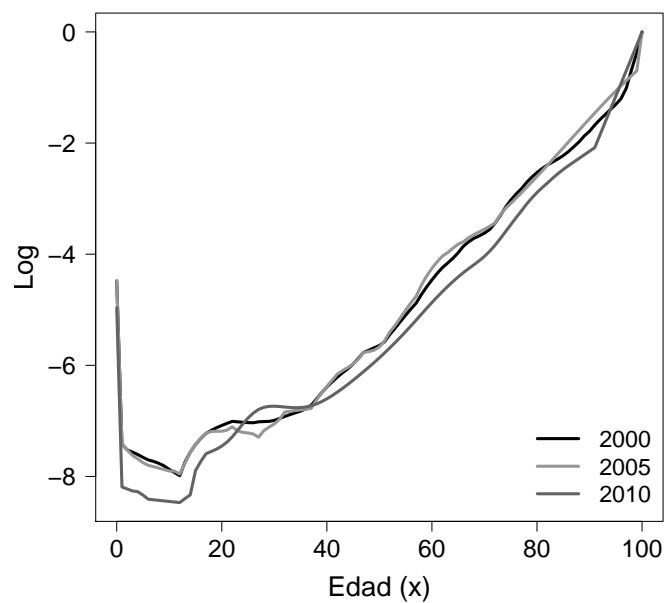
siguiendo un Análisis Bayesiano de Modelos de Regresión Logística cuya descripción se encuentra en Mendoza *et al.* (1999) y Mendoza *et al.* (2000). Para su construcción se empleó información que anualmente las aseguradoras tienen la obligación de reportar a la CNSF, mediante el Sistema Estadístico del Sector Asegurador (SESA). Los años de estudio fueron de 1991 a 1998, por lo que estas tablas ya no reflejan en su totalidad el comportamiento demográfico del país y tampoco el del mercado asegurador.

1.4.1. Tasas de mortalidad AMIS-AMA

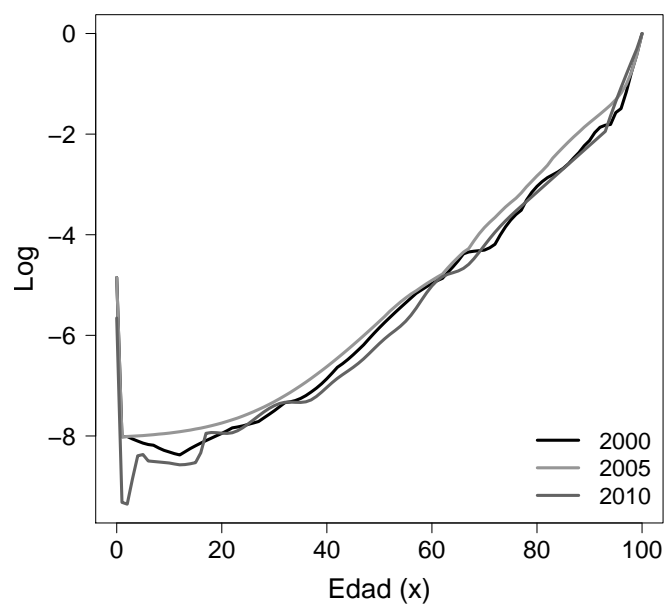
Paralelamente, dos asociaciones, la Asociación Mexicana de Instituciones de Seguros (AMIS) en convenio con la Asociación Mexicana de Actuarios (AMA) desde el año 2000 y cada cinco años se han dedicado a la tarea de crear tablas de mortalidad diferenciadas por sexos. Por un lado, el detalle de la construcción de las tablas del año 2000 se encuentra publicado en Rendón (2000) y, por otro, la metodología del cálculo de las tasas de mortalidad del año 2005 y 2010 se encuentra en Rendón (2012).

Estas tablas emplean información de las cuatro compañías de seguros más grandes en México, manteniendo pólizas con más de 4 años de antigüedad. Para su construcción, se dividió el número de muertos en un año calendario entre los vivos al principio del mismo año. Las altas y bajas dentro de las pólizas, se reflejaron con el uso de la función exponencial de decrementos múltiples hasta llevarla a una tabla con un sólo decremento. Al también modelarse la mortalidad infantil hizo falta tomar la experiencia de otros países como son Canadá y Estados Unidos. Ambas tablas contienen información desglosada por género y edad de 0 a 100. La edad máxima asumida es 100, a esta edad todos los asegurados se habrían extinguido.

Esta circunstancia es la que nos lleva a plantear un análisis más detallado de las edades más avanzadas, dado que la presencia de asegurados centenarios resulta absolutamente plausible.



(a) Hombres



(b) Mujeres

Figura 1.3: Logaritmo de las tasas brutas de mortalidad de la población asegurada mexicana en los años 2000, 2005 y 2010

2 Mortalidad en la población general y asegurada de México. Comparación entre sexos

En este capítulo se ha modelizado la mortalidad de la población general mexicana en el período 1990 a 2010, utilizando modelos de mortalidad estocásticos: Lee-Carter, Renshaw-Haberman y Age-Period-Cohort. Posteriormente, con las tasas de mortalidad ajustadas se ha realizado una comparación con las tasas de mortalidad unisex de los asegurados (CNSF-I). Además, se ha estudiado el efecto del género en la mortalidad, suponiendo que la composición de género en la cartera varía. Tratando de evidenciar el cambio en la mortalidad con diferentes proporciones de género hemos asumido diferentes porcentajes de asegurados y aseguradas y realizado un Modelo Brass-type. La conclusión es que las tablas de mortalidad utilizadas en México por las aseguradoras parecen estar basadas en una proporción de sexos sesgada a favor de los hombres.

2.1. Introducción

Las personas que pueden pagar un seguro, en su mayoría, se espera que tengan un bienestar económico mayor que el ciudadano promedio, pues tienen la posibilidad de optar por la cobertura privada en lugar o además de la pública. Del mismo modo, los miembros de la población asegurada invierten más en la medicina preventiva, lo que resulta en tasas de mortalidad más bajas que las de la población general para los mismos grupos de edad y género. Sin embargo, las carteras de seguros van cambiando constantemente integrando nuevos asegurados, por lo que tener tablas de mortalidad exclusivas de asegurados sin contemplar cómo se comporta la mortalidad de un grupo mayor puede implicar la obtención de resultados sesgados. Además, hay que considerar que los cambios en las carteras de seguros también implican diferencia en la distribución

de hombres y mujeres. Lo más común en una cartera de seguros de vida es que el asegurador no tenga el mismo porcentaje de hombres y de mujeres ni que esta composición se mantenga constante a través del tiempo.

La Directiva 2004/113/CE fue aprobada por unanimidad por el Consejo de la Unión Europea, en esta se prohíbe la discriminación por sexo en el acceso al mercado de bienes y servicios, dejando abiertas dos cuestiones respecto a la influencia del género en la mortalidad: ¿realmente el género tiene un impacto directo en las tasas de mortalidad? y ¿difiere la brecha en las tasas de mortalidad de la población general y la población asegurada entre los sexos?.

La obligación en Europa de fijar el precio del seguro de vida omitiendo el sexo del asegurado nos llevó a realizar este análisis, en el que se busca contribuir a una mejor comprensión de los modelos de mortalidad en este contexto. Para un mejor entendimiento del problema se sugiere el artículo Guillén (2012) donde se analizan los posibles efectos de los costes unisex en los seguros de vida.

A continuación se describen los modelos estadísticos que actualmente se utilizan para estimar las tendencias demográficas de diversos países. También, se presenta el modelo Brass-type que sirve para asociar la mortalidad de una población de referencia, con la mortalidad de un grupo de interés. En nuestro caso la población en general y la población de asegurados, respectivamente.

2.2. Metodología

2.2.1. Modelos de mortalidad

2.2.1.1. Modelo Lee-Carter

El modelo Lee-Carter se ha utilizado por varios autores, por ejemplo, Brouhns *et al.* (2002), Renshaw y Haberman (2003), Delwarde *et al.* (2006), Debón *et al.* (2008) y, más recientemente, Betzuen (2010) comparado las tasas de mortalidad de España con los estimados por el modelo Lee-Carter.

Las tasas de mortalidad cruda se denotan por $m_{x,t}$, donde x es la edad del individuo al tiempo t . Se estiman los parámetros $\beta_x^{(1)}$, $\beta_x^{(2)}$ y κ_t . El modelo para cada edad x se especifica de la siguiente manera:

$$\log m_{x,t} = \beta_x^{(1)} + \beta_x^{(2)} \kappa_t + \epsilon_{x,t}. \quad (2.1)$$

Este modelo explica la mortalidad para cada edad x y año de estudio t . El

parámetro $\hat{\beta}_x^{(1)}$ representa el promedio de $\log m_{x,t}$, $\beta_x^{(2)}$ representa la tasa de cambio de la mortalidad en el tiempo, ya que $\frac{\partial \log m_{x,t}}{\partial t} = \beta_x^{(2)} \frac{\partial \kappa_t}{\partial t}$. La tendencia de la mortalidad para tiempos mayores a t se modela usando una serie de tiempo a partir del parámetro κ_t . Esto sirve tanto para realizar predicciones como para el cálculo de intervalos de confianza (Lee y Carter, 1992). Los residuales $\epsilon_{x,t}$ se asumen variables aleatorias de media cero y varianza constante σ_ϵ^2 , los cuales reflejan las influencias históricas que no son recogidas por el modelo.

La solución al sistema de ecuaciones (2.1) no es única, por lo que se proponen restricciones en los parámetros:

$$\sum_x \beta_x^{(2)} = 1 \quad \text{y} \quad \sum_t \kappa_t = 0. \quad (2.2)$$

La estimación de los parámetros $\beta_x^{(2)}$ y κ_t se realiza empleando la descomposición en valores singulares (SVD, *Singular Value Decomposition*), aplicado a la matriz de logaritmos de tasas de mortalidad crudas, una vez se ha substraído su media: $\log m_{x,t} - \hat{\beta}_x^{(1)}$. Los valores estimado de κ_t pueden ser ajustados a partir de un modelo *ARIMA*(p, d, q) usando la metodología tradicional de Box-Jenkins para el análisis de series temporales.

2.2.1.2. Modelo Renshaw-Haberman

El modelo propuesto por Renshaw y Haberman (2006) es una generalización del modelo Lee-Carter, agrega un término bilineal $\beta_x^{(3)} \gamma_{t-x}$ para representar el efecto cohorte, es decir, este modelo intenta evidenciar el comportamiento de la mortalidad según los cambios generacionales. Se define una cohorte como $c = t - x$.

$$\log m_{x,t} = \beta_x^{(1)} + \beta_x^{(2)} \kappa_t + \beta_x^{(3)} \gamma_{t-x} + \epsilon_{x,t}. \quad (2.3)$$

Los parámetros se interpretan de la misma manera que el modelo Lee-Carter. El parámetro $\beta_x^{(3)}$ es la tasa de cambio de la mortalidad en el tiempo pero en referencia a la mortalidad de la cohorte γ_{t-x} . En este modelo la solución tampoco es única, por lo cual además de las restricciones definidas en las condiciones (2.2) es necesario agregar más restricciones. Por ello finalmente se

necesitan las siguientes restricciones:

$$\sum_t \kappa_t = \sum_t \gamma_t = 0 \quad \text{y} \quad \sum_x \beta_x^{(2)} = \sum_x \beta_x^{(3)} = 1. \quad (2.4)$$

Los parámetros son estimados vía proceso iterativo.

2.2.1.3. Modelo Age-Period-Cohort

El modelo Age-Period-Cohort que hemos utilizado es el propuesto por Currie (2006):

$$\log m_{x,t} = \beta_x^{(1)} + \kappa_t + \gamma_{t-x} + \epsilon_{x,t}, \quad (2.5)$$

Este modelo corresponde al modelo de Renshaw-Haberman cuando $\beta_x^{(2)}$ y $\beta_x^{(3)}$ son iguales a uno. En este modelo los efectos de la edad, el periodo y la cohorte influyen en las tasas de mortalidad de forma independiente. Se pide que los parámetros cumplan con lo siguiente:

$$\sum_t \kappa_t = 0 \quad \text{y} \quad \sum_{x,t} \gamma_{t-x} = 0. \quad (2.6)$$

2.2.1.4. Elección del modelo

Para comparar los modelos de mortalidad, se han elegido dos criterios:

- El criterio de información bayesiano (BIC, *Bayes Information Criterion*) definido en Cairns *et al.* (2009), que penaliza a los modelos que tienen más parámetros:

$$BIC_k = \hat{l}_x - \frac{1}{2} n_k \log N, \quad (2.7)$$

donde n_k es igual al número de parámetros estimados en el modelo, \hat{l}_x es la función de verosimilitud estimada y N es el número de observaciones.

- Calcular los residuales estandarizados, evaluar su comportamiento de forma gráfica y calcular su varianza. Los residuos estandarizados son:

$$Z_{x,t} = \frac{d_{x,t} - l_{x,t} \hat{m}_{x,t}}{\sqrt{l_{x,t} \hat{m}_{x,t}}}, \quad (2.8)$$

donde $d_{x,t}$ es el número de individuos que mueren a la edad x en el tiempo t y $l_{x,t}$ es el número de personas con vida a la edad x en el tiempo t y $\hat{m}_{x,t}$ es la tasa de mortalidad estimada por el modelo. Se asume que los

residuales $Z_{x,t}$ son independientes e idénticamente distribuidos según una distribución normal estándar $N(0, 1)$. Más detalles se pueden encontrar en Cairns *et al.* (2009).

2.2.2. Modelo relacional Brass-type

El modelo relacional Brass-type (Brass, 1971) supone que las tasas de mortalidad de dos poblaciones diferentes están relacionadas. Este autor trataba de establecer una relación lineal entre la transformación logística de dos curvas de supervivencia (logit), explicando de esta manera la mortalidad de un grupo a cada edad en función del otro. La idea principal es usar toda la información de mortalidad de la que se dispone, en nuestro caso, se busca explicar la mortalidad de los asegurados usando como referencia la mortalidad de la población general. Ver Brouhns *et al.* (2002).

En las ciencias actuariales, el modelo relacional Brass-type también puede ser utilizado para medir la selección adversa. Si las tasas de mortalidad de la población asegurada son más altas que las de la población general, esto significaría que los asegurados omiten información privada a la aseguradora y que adquieren seguros porque saben que tienen un mayor riesgo de mortalidad que las personas de la población general de la misma edad. Cuando no hay selección adversa, se esperaría que sucediera lo contrario, es decir, la mortalidad en la población asegurada sería menor a la mortalidad general.

Intuitivamente se espera que las personas que adquieren seguros tengan mayores ingresos que los del ciudadano promedio, gasten más dinero y recursos en asistencia sanitaria y, en consecuencia, el riesgo de mortalidad esperado sea menor al de una persona de la misma edad y sexo de la población no asegurada. Esta situación se conoce como selección positiva.

El modelo relacional Brass-type se define como sigue:

$$f(m_{x,t}) = \beta_0 + \beta_1 f(m_{x,t}^{ref}) + \epsilon_{x,t}, \quad (2.9)$$

donde $m_{x,t}^{ref}$ es la tasa de mortalidad bruta del grupo de referencia, $f(\cdot)$ es la función enlace, la cual puede ser la función logaritmo para representar una relación lineal o la función logística. Este modelo es fácilmente ampliable incorporando interacciones entre tasas de mortalidad y edad x . Otra opción es incorporar interacciones entre tasas de mortalidad y tiempo t , tal como Gatzert

y Wesker (2011) han propuesto:

$$\log m_{x,t} = \beta_0 + \beta_1 \log m_{x,t}^{ref} + \beta_2 \log m_{x,t}^{ref} \cdot t + \epsilon_{x,t}. \quad (2.10)$$

Dado que, las tasas de mortalidad de asegurados de las que disponemos son unisex, se ha decidido crear una tasa de mortalidad unisex de la población general variando el porcentaje w de mujeres: $m_{x,t}^w = w \cdot m_{x,t}^f + (1 - w) \cdot m_{x,t}^m$, donde $m_{x,t}^f$ y $m_{x,t}^m$ son las tasas brutas de mortalidad para la población femenina y masculina, respectivamente. Hemos supuesto que el porcentaje de mujeres se mantiene constante para todos los grupos de edad y para todos los años de estudio. Ciertamente, este supuesto se puede relajar variando la proporción de mujeres a cada edad, ya que sabemos que después de los 65 años la cantidad de mujeres en cada cohorte aumenta debido a que las mujeres tienen una mayor esperanza de vida.

Cuando se compara la población en general y la población asegurada, el modelo relacional Brass-type sin interacciones es:

$$\log m_{x,t}^{ins} = \beta_0 + \beta_1 \log m_{x,t}^{ref} + \epsilon_{x,t}, \quad (2.11)$$

donde $m_{x,t}^{ins}$ es la tasas de mortalidad de los asegurados de edad x en el tiempo t .

Los parámetros β en las ecuaciones (2.9), (2.10) y (2.11), deben interpretarse de la misma manera que en cualquier regresión lineal. En consecuencia, β_0 es la estimación de la intersección de la recta con el eje de ordenadas y representa el valor de $\log m_{x,t}^{ins}$ cuando $\log m_{x,t}^{ref} = 0$. Por otro lado, se tiene la pendiente de la recta β_1 que representa el incremento o decremento esperado en $\log m_{x,t}^{ins}$ por cada unidad de cambio del $\log m_{x,t}^{ref}$.

2.3. Resultados

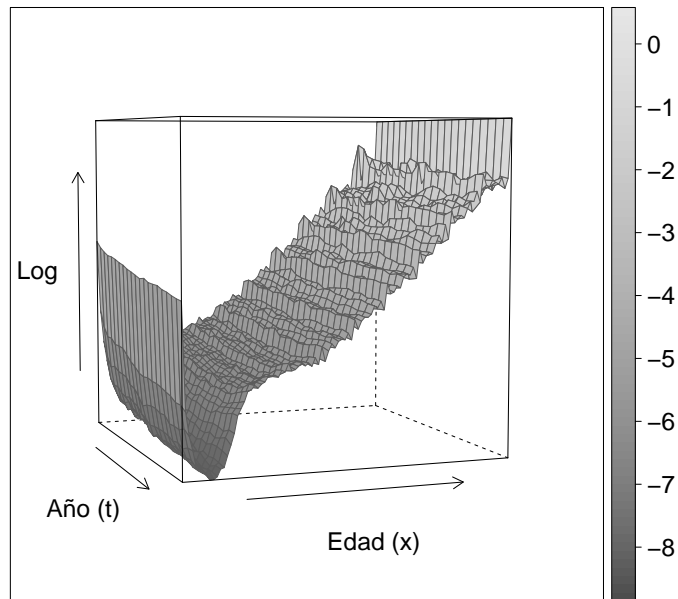
Los modelos descritos en la sección 2.2 fueron ajustados a los datos descritos en el capítulo 1. El ajuste de los modelos se realizó de forma separada para hombres y para mujeres. Las edades que se consideran van desde 0 hasta 100 años para el periodo 1990 a 2010. El modelo Brass-type considera únicamente las tasas de mortalidad del año 2000 y de los 12 a los 99 años de edad, ya que a partir de la edad 12 se pueden realizar contratos con aseguradoras en Méxi-

co. Para obtener las estimaciones de todos los modelos se utilizó el software LifeMetrics implementado en R versión 2.13 (ver Coughlan *et al.*, 2007). Los programas empleados se incluyen en el Anexo.

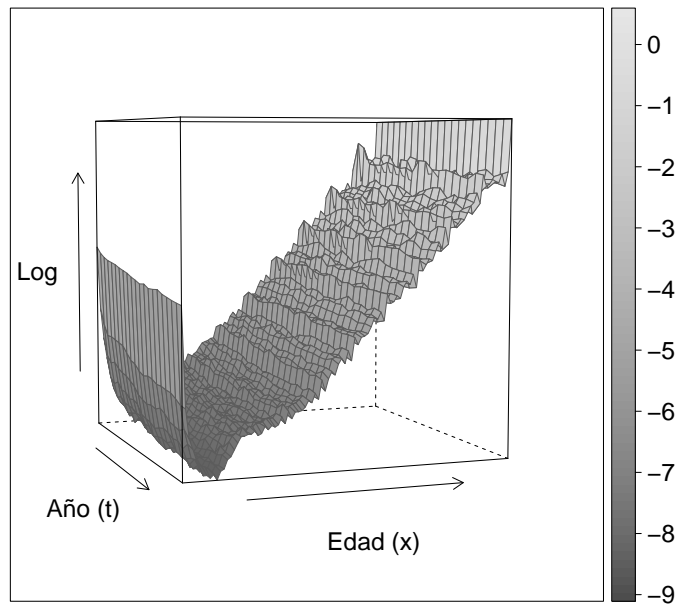
Con el logaritmo de las tasas de mortalidad se ha realizado un análisis exploratorio separando sexos, a través de los diferentes años de estudio t y para todas las edades x . En la figura 2.1 se puede ver como las curvas de mortalidad para hombres y mujeres son muy similares en la infancia y la edad adulta, las mayores diferencias se encuentran en la juventud. En esta época es cuando los hombres son más propensos a sufrir muerte por exponerse a situaciones de riesgo como, por ejemplo, morir a causa de accidentes automovilísticos. A este fenómeno se le conoce como la joroba de los accidentes. Al pasar los años esta joroba se ha mantenido presente, pero ha incrementado su amplitud, además de hacerse presente en la mortalidad femenina.

Otra manera de visualizar los cambios en la mortalidad es la que se presenta en la figura 2.2. En el eje de las x se visualizan los años y en el eje de las y las edades. El logaritmo de las tasas de mortalidad se presenta con diferentes escala de grises que indican diferentes niveles de mortalidad. Si la mortalidad evolucionara favorablemente en el tiempo las zonas de gris claro deberían difuminarse a colores más oscuros que indican menos mortalidad, según la guía adjunta al gráfico. En este caso, se pueden ver franjas horizontales del mismo tono que muestran que no ha existido una mejora sustancial en la mortalidad en los años de estudio prácticamente en todos los rangos edades, excepto para la niñez donde si se observa una reducción en las tasas de mortalidad al paso del tiempo.

Dado que contamos con las tasas de mortalidad de la población asegurada (CNSF-2000), la figura 2.3 compara el logaritmo de la tasa de mortalidad de la población mexicana en el año 2000 y el logaritmo de la tasa de mortalidad de los asegurados. Es posible ver que mortalidad de los asegurados se sitúa en medio de la mortalidad de hombres y mujeres de la población general para las edades más jóvenes pero conforme la edad crece la mortalidad de los asegurados se asemeja más a la mortalidad masculina. A partir de la edad 70 la mortalidad de una persona asegurada se mantiene por debajo de la mortalidad de la población general. Si la mortalidad de los asegurados es menor que la de la población general, entonces se espera que la longevidad de las personas aseguradas sea más grande que la esperada en la población general. Esto puede dar lugar a un sesgo en los cálculos de reservas o en el número de siniestros esperados si la

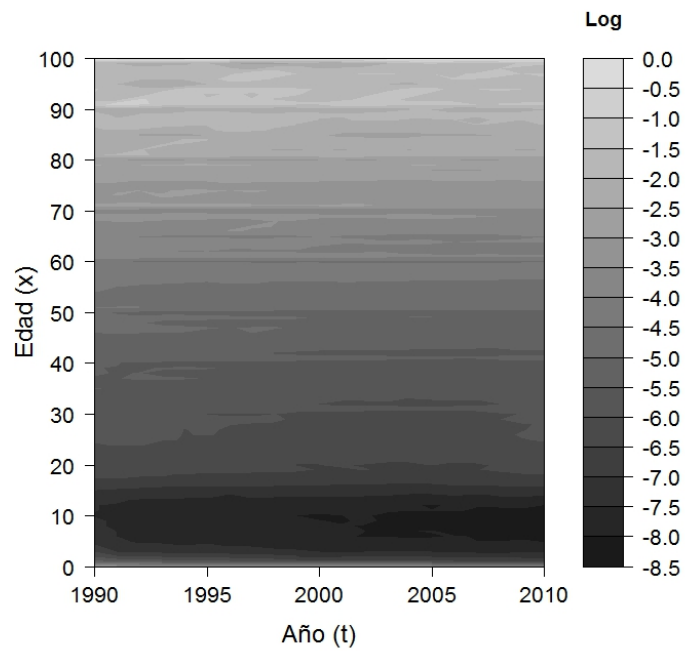


(a) Hombres

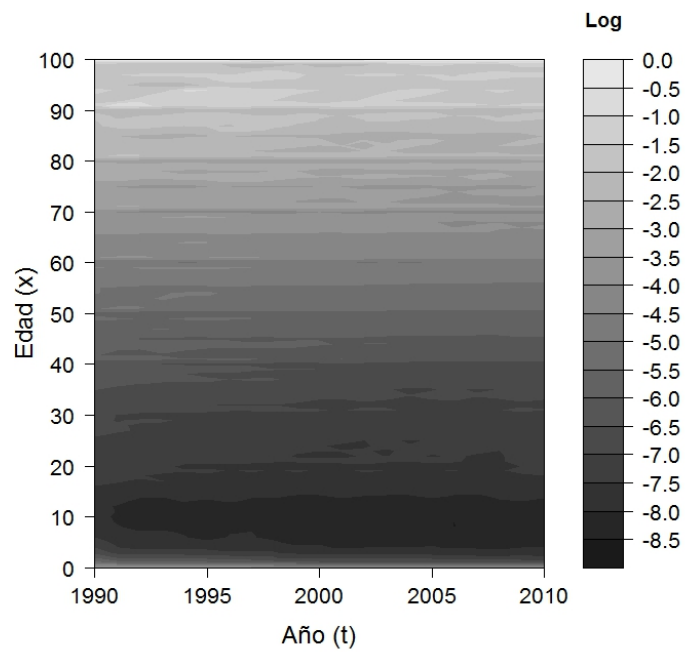


(b) Mujeres

Figura 2.1: Logaritmo de las tasas brutas de mortalidad de la población mexicana desde 1990 a 2010



(a) Hombres



(b) Mujeres

Figura 2.2: Evolución de la mortalidad de México en hombres y mujeres desde 1990 a 2010

mortalidad de la población general es considerada en los pronósticos. Más tarde abordaremos este problema mediante el estudio del modelo relacional Brass-type tal cual se describe en la ecuación (2.11). Observando los datos podemos concluir que las tasas de mortalidad de la población asegurada (CNSF-2000) no representan la forma de mortalidad actual.

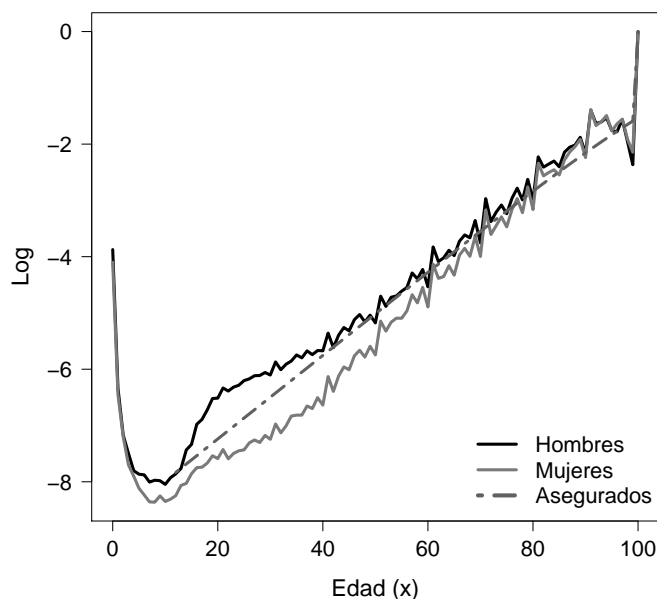


Figura 2.3: Comparativo entre el logaritmo de las tasas brutas de mortalidad para la población mexicana y de las tasas de mortalidad población asegurada (CNSF) en México

2.3.1. Modelo Lee-Carter

El primer modelo estocástico a ajustar a los datos de la población general es el modelo Lee-Carter. En la figura 2.4 se presentan de forma gráfica los parámetros estimados. El número de parámetros estimados es de 101 para cada $\beta_x^{(j)}$ $j = 1, 2$ y 21 para κ_t , resultado en un total de 223 parámetros.

Nótese que en la figura 2.4 las estimaciones de $\beta_x^{(1)}$ ($\hat{\beta}_x^{(1)}$) replican el comportamiento general de la mortalidad, es decir, la mortalidad femenina es menor a la mortalidad masculina, excepto en edades mayores a 80 años.

Desde la edad 0 años hasta la edad 95 los parámetros estimados $\hat{\beta}_x^{(2)}$ para ambos sexos tienen signo positivo, por lo tanto, dado que $\frac{\partial \kappa_t}{\partial t}$ es negativa

la mortalidad ha descendido en este tramo de edades. A partir de entonces $\hat{\beta}_x^{(2)}$ toma valores negativos indicando que la mortalidad en personas mayores a 95 años ha crecido al pasar el tiempo. En el parámetro $\hat{\beta}_x^{(2)}$ se encuentran diferencias entre hombres y mujeres en la niñez y juventud, después de los 40 años, no se observan diferencias entre sexos.

El impacto de $\hat{\beta}_x^{(2)}$ en el logaritmo estimado de la mortalidad es distinto en función del sexo. Para las mujeres el parámetro estimado $\hat{\beta}_x^{(2)}$ cambia de signo, sin embargo, $\frac{\partial \kappa_t}{\partial t}$ sigue siendo negativa, lo que implica un cambio de signo en el logaritmo de la mortalidad. Por el contrario para los hombres el cambio en $\hat{\beta}_x^{(2)}$ viene acompañado por un cambio de signo en la derivada, lo que implica que el logaritmo de la mortalidad ha seguido descendiendo en el tiempo para las edades más avanzadas.

Acerca de los valores de $\hat{\kappa}_t$ la mayor diferencia entre sexos se presenta en los últimos años de estudio, donde las $\hat{\kappa}_t$ para los hombres son mayores a las de las mujeres, reflejando un aumento de mortalidad en hombres. Es en estos años cuando inicia de la llamada “Guerra contra el narcotráfico” que ha sido causa de un número considerable de muertes. En Conapo (2012) se reporta un incremento de la mortalidad para las edades comprendidas entre 10 y 50 años, en el periodo 2005 a 2010. Los autores justifican este hecho con el aumento del número de muertes por homicidio en varias entidades federativas.

2.3.2. Modelo Renshaw-Haberman

Para explicar el efecto cohorte se tiene el modelo Renshaw-Haberman. Se estimaron un total de 445 parámetros: 101 por cada $\beta_x^{(j)}$ $j = 1, 2, 3, 21$ para $\hat{\kappa}_t$ y 121 para $\hat{\gamma}_{t-x}$.

En la figura 2.5 se presentan los parámetros estimados. A diferencia del ajuste del modelo de Lee-Carter, $\hat{\beta}_x^{(1)}$ no replica completamente el comportamiento de la mortalidad en los hombres en edades jóvenes, la joroba de los accidentes es más pronunciada y es solamente visible hasta antes de los 25 años de edad, cuando en los datos reales se aprecia hasta los 40 años de edad. La mortalidad estimada es semejante entre hombres y mujeres en casi todas las edades, excepto en las edades previas a los 22 años.

Para los hombres los parámetros estimados $\hat{\beta}_x^{(2)}$ se mantienen siempre por encima del cero, a partir de los 23 años de edad desciende paulatinamente hasta los 60 años, donde a partir de esta edad y hasta los 90 el comportamiento

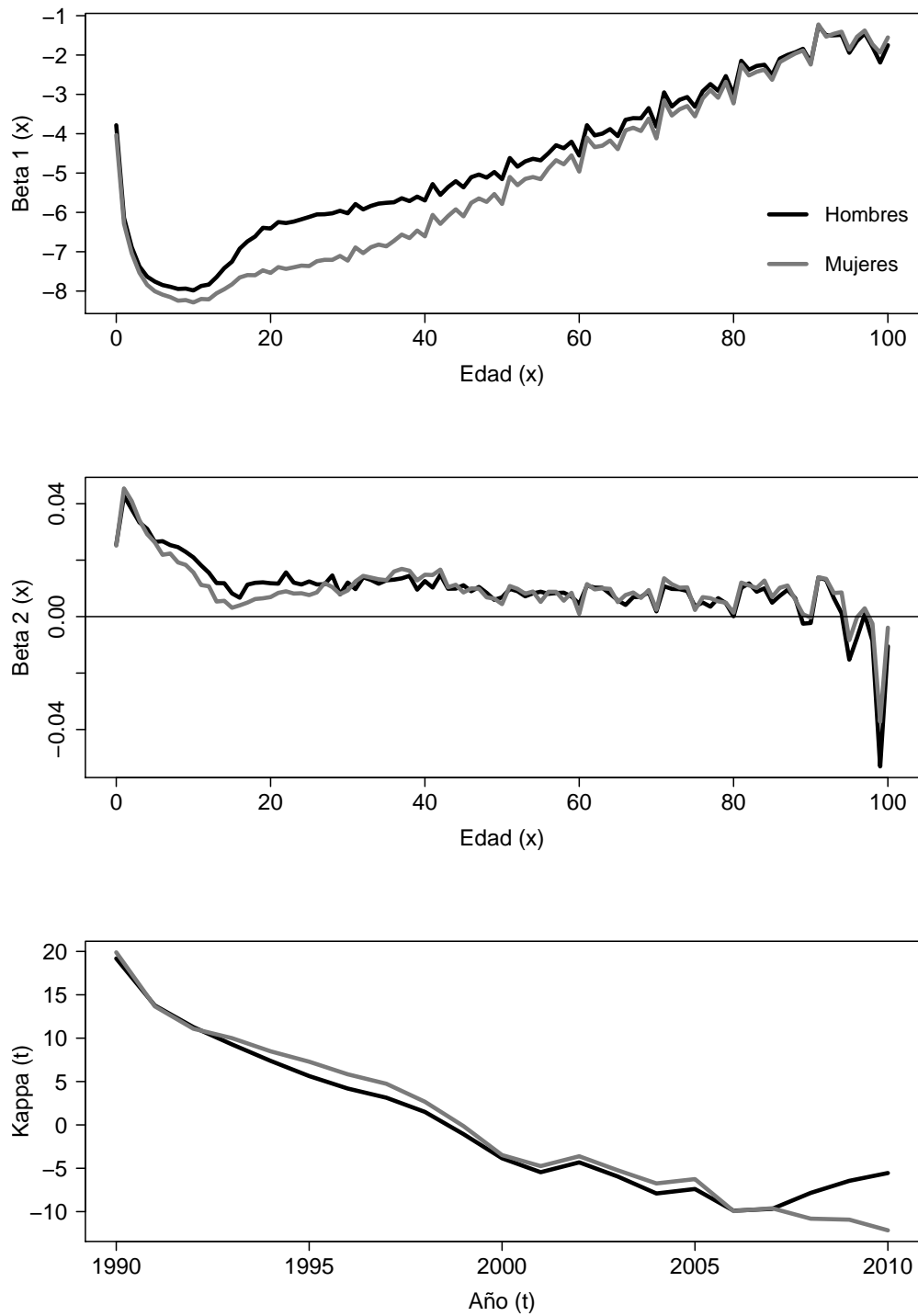


Figura 2.4: Parámetros estimados por el modelo Lee-Carter. Hombres y mujeres con edad 0 a 100 años en el periodo 1990 - 2010. Población mexicana

se mantiene estable con fluctuaciones puntuales. Después de los 90 años, las estimaciones de $\beta_x^{(2)}$ toman valores negativos.

Respecto al comportamiento de los $\hat{\beta}_x^{(2)}$ en las mujeres es similar a la de los hombres desde la edad 0 hasta la edad 60. Sin embargo, a partir de los 40 años de edad comienza un leve decaimiento hasta volverse negativas, para continuar oscilando entre valores positivos y negativos, pero muy cercanos a cero entre las edades 45 y 60. Desde los 61 años en adelante los $\hat{\beta}_x^{(2)}$ para las mujeres comienzan a ascender de forma constante para alcanzar el máximo en la edad 91. El parámetro estimado $\hat{\beta}_x^{(2)}$ del modelo Lee-Carter no reflejaba estas diferencias entre sexos.

Sabemos que el parámetro $\hat{\beta}_x^{(3)}$ indica la tasa de cambio por período de las tasas de mortalidad. El parámetro hasta antes de los 6 años de edad toma valores positivos, después toma valores negativos hasta la edad 21 para ambos sexos. De los 50 años en adelante se observa que los valores oscilan sin saltos demasiado bruscos tal y como lo hacen antes de dicha edad. Lo que indica que la mortalidad en los primeros años ha sufrido cambios al pasar del tiempo. En cuanto $\hat{\beta}_x^{(3)}$ se aleja del cero, más influencia tendrá en la tasa resultante.

La evolución de $\hat{\kappa}_t$ representa la trayectoria global de la mortalidad en el tiempo, cuando se consideran en conjunto todas las edades. El comportamiento del parámetro $\hat{\kappa}_t$ en la población masculina es prácticamente lineal desde el año 1990 hasta el año 2005, donde inicia una curvatura prolongada hasta el último año de estudio. Para las mujeres la pendiente observada es menor que la de los hombres, es decir, el decremento en la mortalidad masculina se dio a mayor velocidad que en las mujeres.

Centrándonos en el año de nacimiento, observamos que el comportamiento de $\hat{\gamma}_{t-x}$ es distinto entre sexos para los nacidos después del años 1903. La serie de los hombres es lineal con tendencia creciente en casi todo su recorrido, en tanto que para las mujeres se tiene una forma convexa antes de 1950 y cóncava para los años subsecuentes. Entre los años 1984 y 1987 se alcanzan los máximos, siendo una posible explicación de este aumento la ocurrencia de un terremoto en el año de 1985, que afectó a varios estados de la república.

2.3.3. Modelo Age-Period-Cohort

Este es un caso particular del modelo anterior con $\hat{\beta}_x^{(2)}$ y $\hat{\beta}_x^{(3)}$ igual a uno. El número de parámetros estimados es de 101 para $\hat{\beta}_x^{(1)}$, 21 para $\hat{\kappa}_t$ y 121 para

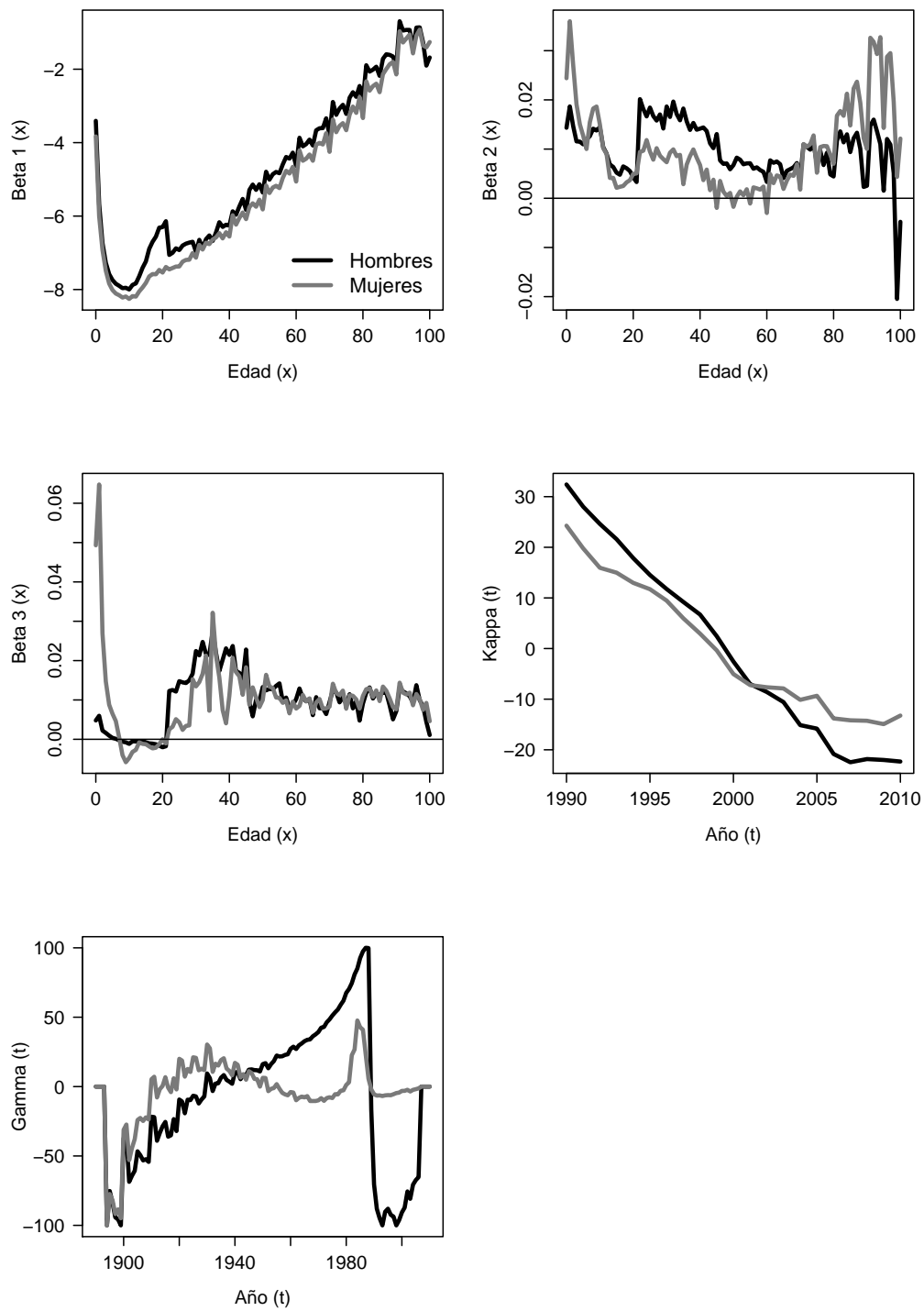


Figura 2.5: Parámetros estimados por el modelo Renshaw-Haberman. Hombres y mujeres con edad 0 a 100 años en el periodo 1990 - 2010. Población mexicana

$\hat{\gamma}_{t-x}$, resultando un total de 243 parámetros estimados.

Al igual que el modelo Lee-Carter, este modelo acentúa las diferencias entre sexos para las muertes ocurridas en la juventud y las aminora para edades más tardías, es decir, reconoce el comportamiento real del logaritmo de las tasas de mortalidad. La figura 2.6 presenta los parámetros estimados.

Los valores de los parámetros $\hat{\kappa}_t$ son similares a los obtenidos por el modelo Lee-Carter. Las curvas de hombre y mujer tienen un comportamiento parecido, se encuentran diferencias entre los años 1994 y 1999. Después del año 2008 es donde se sitúan las diferencias más grandes, lo que indica que la mortalidad en mujeres disminuyó en comparación a los hombres durante ese periodo.

Cuando estudiamos el efecto cohorte a partir de parámetro estimado $\hat{\gamma}_t$ observamos que no existen grandes diferencias entre sexos, el comportamiento es prácticamente el mismo. La mortalidad decreció a mayor velocidad para los nacidos entre los años 1990 y 2006 .

2.3.4. Comparación de los modelos estimados

Siguiendo la definición de residuo estandarizado expresado en la ecuación (2.8), se han realizado gráficos como los propuestas en Cairns *et al.* (2009). La idea es representar los residuos con la ayuda de escala de grises que varían según la magnitud de los mismos. En el eje de las abscisas se pone la edad x , en el eje de ordenadas el año de estudio t . Se desea que los residuos se distribuyan de forma aleatoria, es decir, se cumpla la hipótesis de que sean independientes e idénticamente distribuidos. Si esto se cumpliera, en la gráfica no se deberían ver patrones de diferentes tonos de grises que dibujan los residuos.

El gráfico con los residuales del modelo Lee-Carter muestra que el efecto de la cohorte no logra capturarse, ya que se presentan líneas diagonales a través de los años, principalmente después de la edad 60 en ambos sexos. Los residuos más grandes corresponden a la mortalidad estimada para las edades mas avanzadas.

En contraste, los residuales del modelo Renshaw-Haberman no muestran patrones visibles, ya que este modelo incluye un parámetro que captura los efectos de la cohorte. Errores grandes se observan entre las edades 95 y 100 años, principalmente en el inicio de la década de los 90, donde se estiman valores mayores a los observados. Pero los errores más grandes se encuentran en la estimación de la mortalidad del periodo 2007 a 2010 para los recién

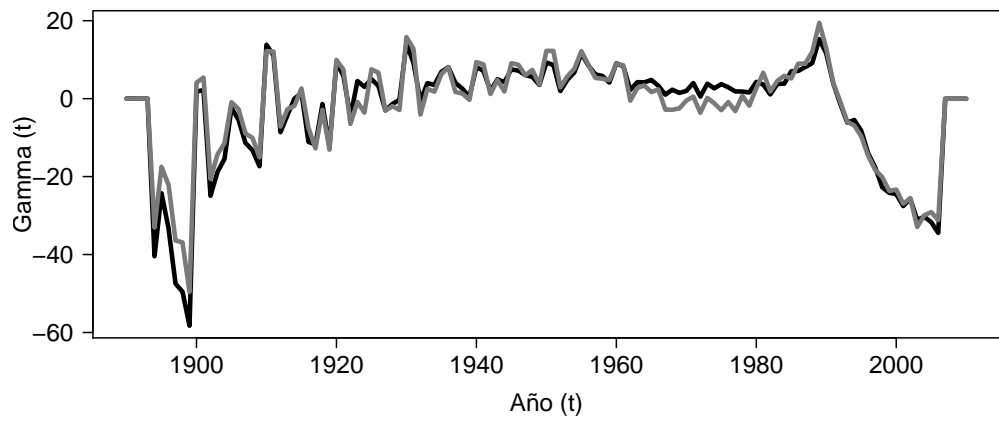
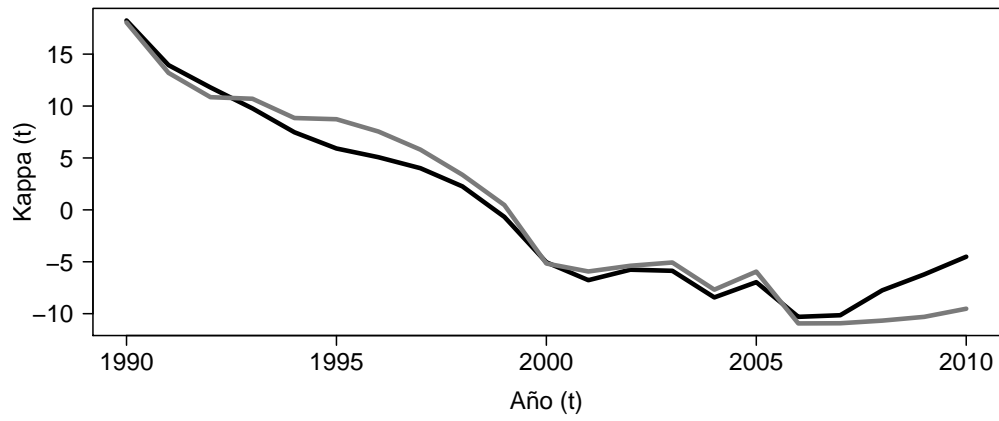
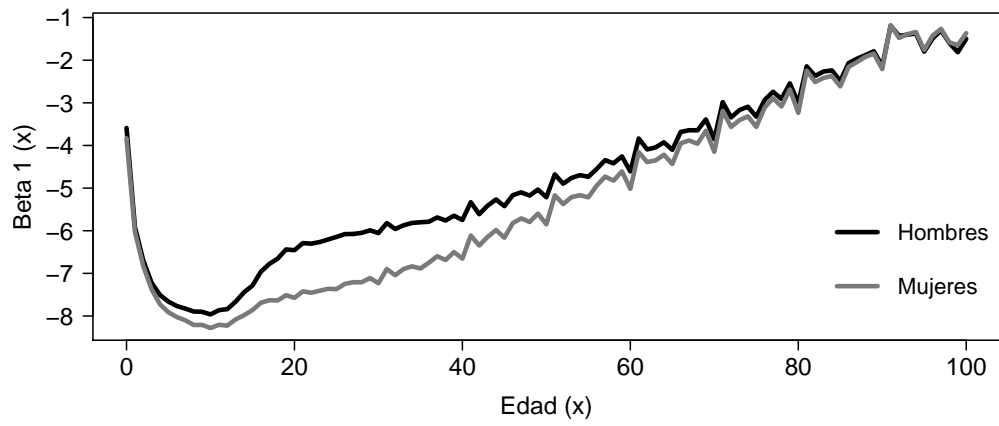


Figura 2.6: Parámetros estimados por el modelo Age-Period-Cohort. Hombres y mujeres con edad 0 a 100 años en el periodo 1990 - 2010. Población mexicana

nacidos. En general los residuos del modelo ajustado para los hombres son mayores que los del modelo ajustado para las mujeres.

Comparado los resultados del modelo Age-Period-Cohort con los dos anteriores observamos que los residuales de esta última estimación son mayores. Sin embargo, en los residuos del modelo Age-Period-Cohort no se observan las diagonales que evidencian la necesidad de incorporar el efecto cohorte.

En general los tres modelos presentan elevados errores en los últimos años analizados para los recién nacidos.

Siguiendo el criterio de información bayesiano (BIC), el modelo que mejor ajusta es el de Renshaw-Haberman, efectivamente se constata que al añadir efectos cohorte se estima mejor la mortalidad.

Modelo	Hombres	Mujeres
Lee-Carter	-25,533.31	-25,409.77
Renshaw-Haberman	-16,364.67	-15,607.63
Age-Period-Cohort	-21,379.55	-20,189.39

Tabla 2.1: BIC de los diferentes modelos ajustados

2.3.5. Modelo relacional Brass-type

Una vez se han estimado las tasas de mortalidad para hombres y mujeres en diferentes periodos de tiempo t , se procede a ajustar el modelo relacional Brass-type, descrito en la sección 2.2.2. Se han utilizado únicamente las tasas de mortalidad en las edades 12 a 99 años para el año 2000.

Hemos llevado a cabo una regresión lineal para comparar la mortalidad de los siniestros reportados en el sector asegurador con la mortalidad de la población general. Las tasas de mortalidad de referencia son las estimadas por los modelos Lee-Carter, Renshaw-Haberman y Age-Period-Cohort para la población mexicana. La idea fundamental es mostrar que el comportamiento de la mortalidad de los individuos asegurados es diferente a la mortalidad experimentada por la población en general.

Ya que disponemos de tasas de mortalidad unisex de la población asegurada, seguiremos el análisis bajo este supuesto, pero variando la proporción de hombres y mujeres en la población general. Utilizando el modelo relacional Brass-type, hemos investigado la influencia de la proporción de género en las

carteras de seguros cuando se considera una tabla unisex. Para las aseguradoras, utilizar una única tabla mortalidad puede suponer un riesgo.

Los parámetros estimados y el BIC para el modelo Brass-type se muestran en la tabla 2.2. La primera columna indica el porcentaje de mujeres en la población. Esto significa que hemos considerado diferentes proporciones w por genero en la definición de las tasas de referencia. La estimación de la pendiente es igual a uno cuando la proporción de mujeres es exactamente 25 %, lo que significa que existe una similitud entre la población asegurada y la población en general cuando se considera que una de cada cuatro personas de la población es mujer.

% Mujeres w	Lee-Carter			Renshaw-Haberman			Age-Period-Cohort		
	$\hat{\beta}_0$	$\hat{\beta}_1$	BIC	$\hat{\beta}_0$	$\hat{\beta}_1$	BIC	$\hat{\beta}_0$	$\hat{\beta}_1$	BIC
0 %	-0.06	1.04*	33.36	-0.04	1.05*	22.06	-0.01	1.05*	24.17
25 %	-0.14*	1.00*	17.88	-0.12	1.00*	3.48	-0.10	1.00*	7.79
50 %	-0.22*	0.95*	8.70	-0.20*	0.95*	-7.86	-0.19*	0.96*	-1.18
75 %	-0.30*	0.91*	6.00	-0.27*	0.91*	-11.18	-0.27*	0.91*	-2.64
100 %	-0.38*	0.87*	8.77	-0.35*	0.87*	-7.58	-0.35*	0.87*	1.96

Tabla 2.2: Parámetros estimados por el modelo Brass-type: $\log m_{x,t}^{aseg} = \hat{\beta}_0 + \hat{\beta}_1 \log m_{x,t}^w$. Siendo w el porcentaje de mujeres en la población.
* denota p-valor ≤ 0.05

Analizando el logaritmo de las tasas de mortalidad ajustadas por el modelo Brass-type, se puede ver que cuando es mayor la proporción de mujeres, más grande es la diferencia entre la tasa de mortalidad de los asegurados y la de la población en general, esto para las edades 12 a 18. Lo contrario ocurre cuando se considera el intervalo de edades ente 19 y 40, es decir, en estas edades las diferencias son pequeñas.

Según el artículo de Gatzert y Wesker (2011), los parámetros estimados para la población Suiza son $\hat{\beta}_0 = -0.3197$ y $\hat{\beta}_1 = 1.0747$. Comparando los parámetros estimados entre las poblaciones suiza y mexicana, se puede ver que $\hat{\beta}_0$ son similares cuando en el modelo Brass-type se ha considerado una población primordialmente femenina. Por otro lado, $\hat{\beta}_1$ se acerca al parámetro estimado para la población Suiza cuando se considera una mayoría de hombres.

La figura 2.9 muestra el resultado de considerar los parámetros estimados de la población suiza para obtener unas nuevas tasas de asegurados usando las tasas estimadas por el modelo Lee-Carter como población de referencia.

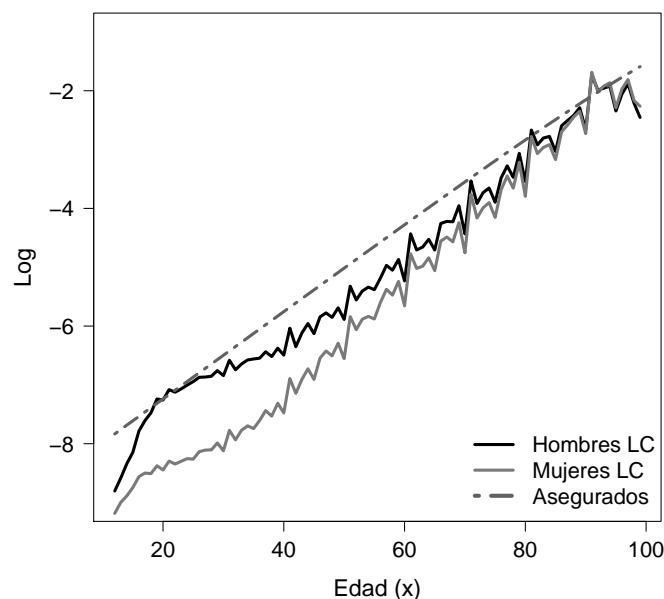


Figura 2.9: Tasas de mortalidad estimada con el modelo Lee-Carter utilizando los parámetros de la población suiza. La línea discontinua representa el logaritmo de las tasas estimadas para los asegurados por la CNSF en México

2.4. Conclusiones

Hemos descrito el comportamiento de la mortalidad de la población mexicana. Como es patente en otros países, el comportamiento de la mortalidad ha demostrado ser dependiente de la cohorte y del género, siendo este último un factor relevante en el riesgo de la longevidad y mortalidad. En particular, para la población mexicana la principal diferencia entre las tasas de mortalidad masculina y femenina ocurren en el intervalo de edad 20 a 50 años. De acuerdo con la Asociación Mexicana de Instituciones de Seguros (AMIS), en este periodo es donde se tiene la mayor proporción de asegurados. Por lo tanto, el uso de tablas de vida unisex en el ámbito de los seguros puede dar lugar a sesgo en los resultados ligados en el cálculo de primas, más aún si la proporción hombre-mujer dentro de la cartera de seguros no se tiene en cuenta.

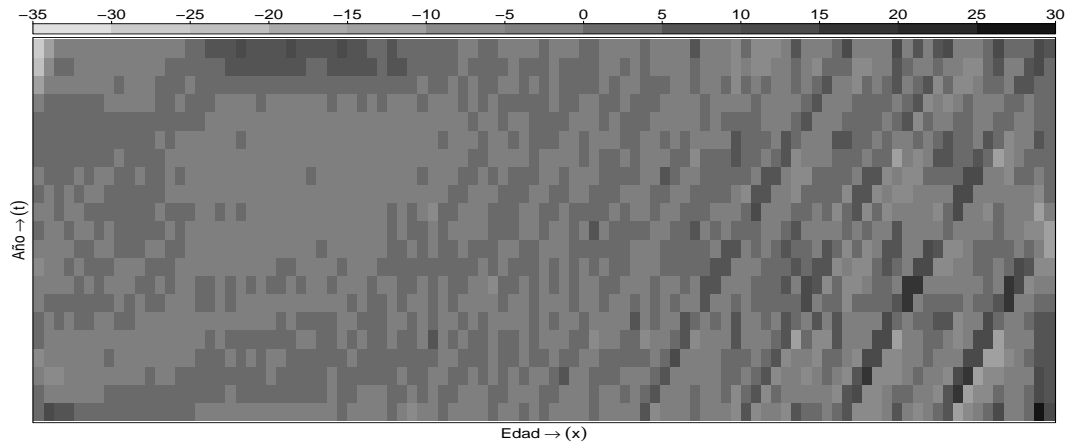
No se dispone de tasas de mortalidad históricas que reflejen exclusivamente el mercado de seguros de vida en México, una única tabla unisex ha sido utilizada por las aseguradoras desde el año 2000. Debido a esto, hemos propuesto el uso

de las experiencias en otros países, para crear nuevas tablas de asegurados para México. Por ejemplo, Gatzert y Wesker (2011) han utilizado un modelo Brass-type para relacionar las tasas de mortalidad de la población de asegurados con las tasas de mortalidad de la población general en Suiza.

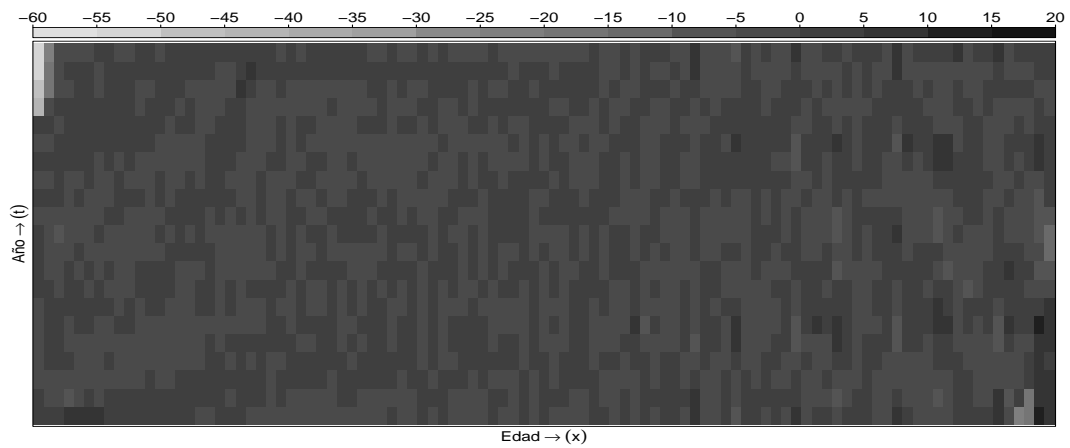
Hemos replicado el ejercicio de Gatzert y Wesker (2011) y comparado los parámetros obtenidos en el modelo relacional Brass-type para Suiza con los estimados para el caso mexicano. Con la finalidad de analizar diferentes escenarios, las tasas de mortalidad unisex se han obtenido asumiendo que la proporción hombre-mujer cambia. Se obtuvieron los coeficientes de la regresión lineal, viendo que estos son similares a los de la población Suiza. Sin embargo, teniendo en cuenta que las tasas de mortalidad para los asegurados son unisex no podemos comparar los dos conjuntos de resultados directamente.

Se recomienda que el estudio de la mortalidad en las carteras de seguros se realice de manera separada para cada sexo, ya que la proporción de hombres y mujeres en la población general y en una cartera de seguros puede diferir sustancialmente. Las tarifas unisex deben basarse en supuestos razonables, respetando la proporción de mujeres en la cartera y teniendo en cuenta la brecha entre la mortalidad de la población general y la de la población asegurada. Las tablas de vida unisex para las carteras de asegurados, así como las utilizadas actualmente por las aseguradoras en México, asumen implícitamente que hay un equilibrio entre hombres y mujeres y que este se mantiene constante en el tiempo, es decir, que el número de hombres y mujeres no varía conforme la cartera va ganando años. Por lo tanto, creemos que a pesar de que las regulaciones europeas insisten en que los precios de seguros deben ser iguales para hombres y mujeres, las estimaciones de tasas de mortalidad usadas en el mundo actuarial deben llevarse a cabo separando por género. Además, las aseguradoras tienen que tener en cuenta la composición actual de género en sus carteras al fijar el precio de su producto de seguro de vida.

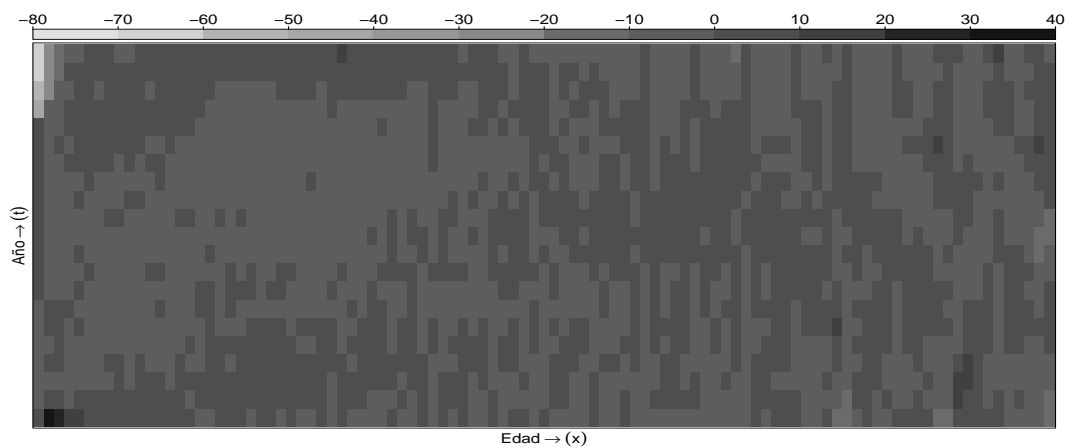
El efecto de las tablas unisex en la longevidad se analiza detalladamente en el siguiente capítulo.



(a) Lee-Carter

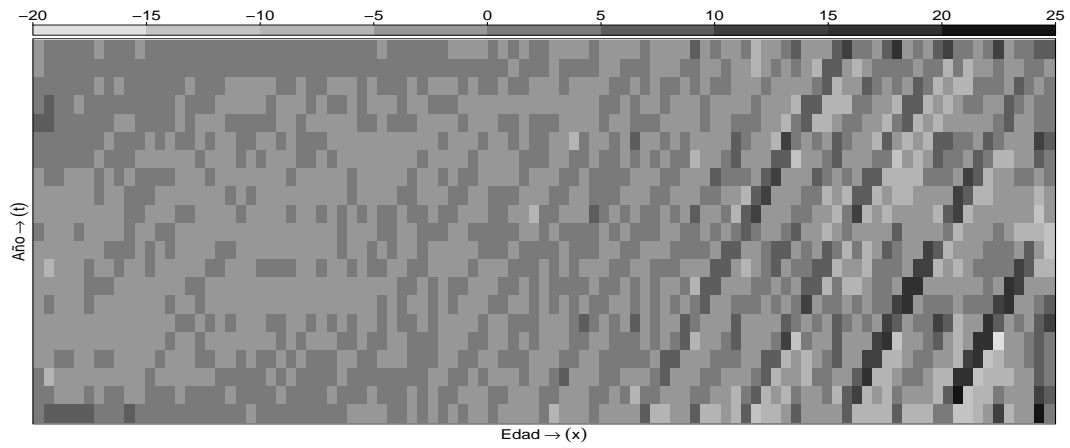


(b) Renshaw-Haberman

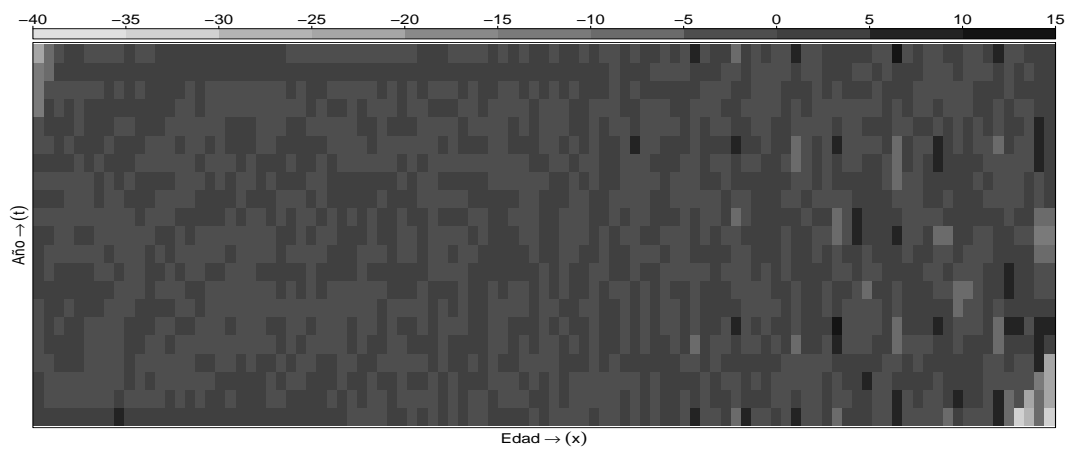


(c) Age-Period-Cohort

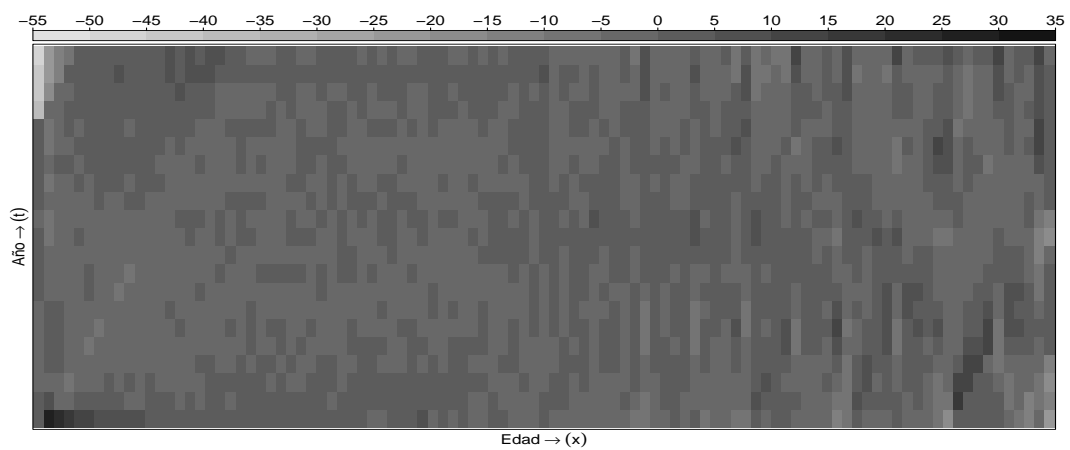
Figura 2.7: Residuales estandarizados $Z_{x,t}$ del ajuste de la población masculina. Los recuadros oscuros implican $Z_{x,t} < 0$ y los claros $Z_{x,t} > 0$



(a) Lee-Carter



(b) Renshaw-Haberman



(c) Age-Period-Cohort

Figura 2.8: Residuales estandarizados $Z_{x,t}$ del ajuste de la población femenina. Los recuadros oscuros implican $Z_{x,t} < 0$ y los claros $Z_{x,t} > 0$

3 Uso de tablas de mortalidad unisex en el análisis de la longevidad

En este capítulo se evaluará el riesgo de longevidad de la población general de México. Este ejercicio se realizará variando la proporción de hombres y mujeres en la población. Se ha decidido utilizar este recurso tratando de imitar lo que sucede en las carteras de seguros, es decir, en diferentes momentos se tienen diferentes cantidades de hombres y mujeres asegurados.

3.1. Introducción

En el capítulo 2 se ha analizado el efecto de variar la proporción entre géneros al crear una única tabla de mortalidad para ambos sexos, en este capítulo nos centraremos en evidenciar que en las últimas dos décadas las mujeres viven más que los hombres y, por lo tanto, el omitir este hecho puede causar situaciones de mayor riesgo para las aseguradoras. Por ejemplo en el seguro de vida se podría estar cobrando una prima menor a las mujeres de la que les correspondería.

La decisión de tarificar los seguros sin tener en cuenta el género del asegurado ha sido aprobado a pesar de que existen evidencias estadísticas de que las mujeres son más longevas que los hombres y que, por lo tanto, la tasa de mortalidad anual a la edad x de las mujeres es menor que la de los hombres. Esta diferencia en tasas de mortalidad se mantiene en todos los grupos de edades, aún en edades muy avanzadas. Además, este hecho ocurre en todas las poblaciones del mundo que se han estudiado.

Es en el ramo del seguro de vida donde se espera que no discriminar por sexo tendrá un gran impacto. Principalmente porque la tarificación está ligada a dos variables que son edad y género y es muy difícil reemplazar la información que esta última variable provee. La cuantificación del riesgo tendrá que estar basada en otro tipo de factores, que bien podrían ser del tipo conductual, que reflejen el estilo de vida, de tal manera que la variable género se pueda obviar. Por ejemplo: ocupación, hobbies, hábitos de salud y de consumo, etc.. Por lo

tanto, la eliminación del género en la fijación de precios supone un cambio estructural en la tarificación.

En cuanto al precio de los seguros, las mujeres se verán afectadas al pagar primas de seguros de vida más altas de lo que lo hacían en el pasado. Con la no discriminación, las tasas de mortalidad de las mujeres deben compensar a las tasas de mortalidad de los hombres, que son más altas. En cambio, los hombres se verán gravemente afectados en los productos que impliquen riesgo de longevidad, como son las anualidades de jubilación, ya que los pagos se calcularán suponiendo que pueden alcanzar las mismas edades que las mujeres.

Conocer la edad x a la que un individuo puede sobrevivir depende de su edad actual y sexo, entre otros factores como son la ocupación y los hábitos diarios. La esperanza de vida muestra el promedio de años que se espera que una persona sobreviva, se puede calcular al nacer o a cualquier edad, este último caso es el que interesa en el campo actuarial. Una medida alternativa a la esperanza de vida sería evaluar el riesgo de longevidad, entendido como el número de años que una pequeña fracción de la población sobrevive. Si es posible encontrar una variable aleatoria que represente la duración de vida, entonces es posible calcular el valor en riesgo (VaR, *Value at Risk*). En este capítulo hemos aplicado este concepto para calcular el riesgo de longevidad.

3.2. Descripción de los datos

Para realizar este análisis se han utilizado los datos referentes a la población general mexicana que se han descrito en el capítulo 1. Para enfatizar el efecto de la proporción de géneros se han calculado diferentes tasas de mortalidad variando la proporción de mujeres para crear tablas unisex. La tasa de mortalidad unisex es:

$$m_{x,t}^w = w \cdot m_{x,t}^f + (1 - w) \cdot m_{x,t}^m, \quad (3.1)$$

con $w \in \{0\%, 25\%, 50\%, 75\%, 100\%\}$, siendo $m_{x,t}^f$ y $m_{x,t}^m$ tasas de mortalidad para mujeres y hombres, respectivamente. La base de datos final considera solamente las edades entre 65 y 99, ya que como se ha mencionado anteriormente nuestro interés principal es conocer la máxima edad de supervivencia para la que se tiene información.

3.3. Metodología

3.3.1. Análisis de supervivencia

El análisis de supervivencia es una metodología que se centra en el estudio del comportamiento de la mortalidad. Dado que la definición de la mortalidad se puede aplicar en muchas otras situaciones, el análisis de supervivencia se ha utilizado ampliamente para estudiar cualquier evento relacionado con la finalización de un proceso. En bioestadística, el análisis de supervivencia se puede referir a la duración de una enfermedad. En ingeniería, al tiempo hasta que se produce un fallo en algún dispositivo o máquina. En general, el análisis de supervivencia estima el tiempo hasta la ocurrencia de un evento de interés.

Denotamos por T a la variable aleatoria “duración de la vida”. La función de supervivencia es la probabilidad de que un individuo sobreviva más allá del tiempo t , $S(t) = P[T > t] = 1 - F(t)$, donde $t > 0$. La función $S(t)$ corresponde a la probabilidad de que un individuo no muera antes de t unidades de tiempo, es decir, la probabilidad de que el evento E se produzca después de t periodos. La función de supervivencia es una función monótona decreciente y cumple que: $S(0) = 1$, $S(\infty) = 0$.

Existen dos enfoques para estimar la función de supervivencia: no paramétrico y paramétrico. En el primer caso, el método más sencillo es utilizar la distribución empírica que pasa a ser el estimador de Kaplan-Meier cuando se incorpora censura en algunas observaciones. Fledelius *et al.* (2004) realizó una comparación entre los dos enfoques en edades mayores. Los métodos paramétricos son muy útiles para estimar la mortalidad. Las distribuciones más comúnmente usadas son la Exponencial, la Weibull, la Lognormal y la Gamma. Todas estas distribuciones toman valores en el intervalo $[0, \infty)$, y tienen funciones de densidad asimétricas.

En nuestro caso, sólo hemos ajustado las tres primeras distribuciones mencionadas anteriormente. Los mejores resultados se obtuvieron usando la distribución Weibull, a continuación se describe su uso. Asumimos que la variable aleatoria T se distribuye como una Weibull (λ, ρ) , cuya función de densidad es la siguiente:

$$f(t) = \lambda \rho t^{\lambda-1} \exp(-\rho t^\lambda), \quad \rho > 0 \text{ y } \lambda > 0, \quad (3.2)$$

donde ρ es el parámetro escala y λ es el parámetro de forma. Más propiedades de la distribución de Weibull se pueden encontrar en Rinne (2010). Además,

la función de supervivencia asociada a la distribución Weibull es:

$$S(t) = \exp(-\rho t^\lambda). \quad (3.3)$$

A partir de la ecuación (3.3), es posible obtener los parámetros de la distribución Weibull, esto se consigue estimando por mínimos cuadrados ordinarios los parámetros de la siguiente relación lineal:

$$\log S(t) = \log \rho + \lambda \log t, \quad (3.4)$$

sustituyendo $S(t)$ por su aproximación empírica. Como texto de referencia del análisis de supervivencia se recomienda ver el libro de Cox y Oakes (1984).

3.3.2. Valor en riesgo

Como se ha mencionado anteriormente, la longevidad en una cartera de seguros puede representar un riesgo para una compañía, por lo que es deseable cuantificar las posibles pérdidas. En el sector de las finanzas la medida de riesgo que se utiliza con frecuencia es el VaR (Denuit *et al.*, 2005). Algunas empresas utilizan el VaR con el fin de medir y controlar el nivel de riesgo que puede soportar. El valor en riesgo se define como “La pérdida límite a partir de la cual tener una pérdida superior es muy poco probable”. Siendo α el nivel de confianza, que toma valores próximos a 1, el VaR es el valor v de tal manera que la probabilidad de que la duración de la vida T sea mayor que v es menor que $(1 - \alpha)$. El VaR se calcula directamente cuando se conoce la distribución de la variable aleatoria y los parámetros de la distribución han sido estimados. En definitiva, el VaR es el cuantil α de la función de distribución:

$$VaR_\alpha = \inf\{P(T > v) \leq (1 - \alpha)\}. \quad (3.5)$$

Para medir la longevidad, el VaR se ha utilizado como un umbral, es decir, el VaR se interpreta como la edad en la cual sólo una proporción igual a $(1 - \alpha)$ de las personas superarán dicha edad. Para evidenciar las diferencias de mortalidad y longevidad entre sexos, se ha calculado el VaR bajo diferentes escenarios, donde la proporción de hombres y mujeres no es la misma, ver sección 3.2.

3.3.3. Resultados

Con la tasa de mortalidad bruta ponderada $m_{x,t}^w$, de la población general mexicana, se ha estimado la función de supervivencia empírica para cada año t , para después obtener los parámetros de la distribución Weibull y de manera consecuente el VaR. Para calcular $S(t)$ se utilizó el paquete `demography` del programa R.

En primer lugar, ajustamos un modelo paramétrico siguiendo el enfoque que se explica en la sección 3.3. Asumimos que el comportamiento de la supervivencia sigue una distribución Weibull. Partiendo del uso de la relación lineal descrita en la ecuación (3.4) se obtuvieron los parámetros $\hat{\lambda}$ y $\hat{\rho}$. En la tabla 3.1 aparecen los valores de los parámetros estimados.

	1990		1995		2000		2005		2010	
w	$\hat{\rho}$	$\hat{\lambda}$	$\hat{\rho}$	$\hat{\lambda}$	$\hat{\rho}$	$\hat{\lambda}$	$\hat{\rho}$	$\hat{\lambda}$	$\hat{\rho}$	$\hat{\lambda}$
0%	18.09	1.62	18.24	1.66	19.59	1.61	19.14	1.67	19.33	1.66
25%	18.52	1.64	18.55	1.68	20.05	1.62	19.59	1.69	19.86	1.68
50%	18.98	1.66	18.88	1.71	20.54	1.64	20.06	1.71	20.42	1.70
75%	19.46	1.68	19.22	1.73	21.04	1.67	20.56	1.73	21.01	1.72
100%	19.97	1.71	19.57	1.76	21.58	1.69	21.09	1.75	21.64	1.75

Tabla 3.1: Parámetros estimados para la distribución Weibull según la relación lineal $\log S(t) = \log \hat{\rho} + \hat{\lambda} \log t$. Siendo w el porcentaje de mujeres en la población

El parámetro λ proporciona información acerca de la tasa de fallo. De hecho, cuando λ es menor que 1 nos indica que la tasa de mortalidad disminuye con el tiempo. Si λ es mayor que 1 nos indica que la tasa de mortalidad aumenta con el tiempo. Cuando λ es igual a 1, eso indica un comportamiento constante de la tasa de mortalidad. Como era de esperar el parámetro de forma λ estimado es mayor que uno para todos los años que se han estudiado, reflejando el incremento de la mortalidad con la edad.

Posteriormente, como medida de referencia se ha calculado la esperanza de vida a los 65 años, la cual se presenta en la tabla 3.2. La esperanza de vida a partir de 65 años ha crecido en las últimas dos décadas, la diferencia entre la esperanza de vida del año 1990 y el año 2010 es de 1.21 años para las mujeres y 1 años para los hombres. Al comparar la esperanza de vida de las mujeres y los hombres mayores de 65, la principal diferencia surge en 2010, cuando se espera que las mujeres vivan 18.30 años más después de la edad 65 y los

hombres 16.64 años. De nueva cuenta es fácil ver que las mujeres continúan viviendo más que los hombres, sin importar la época que se estudie.

	1990	1995	2000	2005	2010
Hombres	15.63	15.79	16.75	16.47	16.64
25 % Muj.	15.98	16.06	17.09	16.82	17.03
50 % Muj.	16.33	16.34	17.44	17.19	17.43
75 % Muj.	16.70	16.62	17.81	17.56	17.86
Mujeres	17.09	16.91	18.19	17.95	18.30

Tabla 3.2: Esperanza de vida a los 65 años para la población mexicana

El VaR se ha obtenido directamente del cuantil α de la función de distribución Weibull con los parámetros de la tabla 3.1. Se tomaron dos niveles de confianza: $\alpha = 0.90$ y $\alpha = 0.95$. Los resultados se muestran en la tabla 3.3 y en la tabla 3.4.

	1990	1995	2000	2005	2010
Hombres	30.23	30.14	32.93	31.50	31.94
25 % Muj.	30.78	30.46	33.51	32.07	32.63
50 % Muj.	31.36	30.79	34.11	32.67	33.36
75 % Muj.	31.95	31.12	34.72	33.29	34.11
Mujeres	32.56	31.46	35.35	33.92	34.88

Tabla 3.3: Valor en riesgo al nivel de confianza 0.90: Número de años a sobrevivir después de la edad 65

	1990	1995	2000	2005	2010
Hombres	35.55	35.32	38.80	36.86	37.42
25 % Muj.	36.14	35.62	39.41	37.47	38.17
50 % Muj.	36.74	35.92	40.03	38.11	38.94
75 % Muj.	37.36	36.23	40.67	38.75	39.74
Mujeres	37.99	36.53	41.31	39.41	40.55

Tabla 3.4: Valor en riesgo al nivel de confianza 0.95: Número de años a sobrevivir después de la edad 65

Teniendo en cuenta un nivel de confianza igual a 0.90 y observando el año 2010, el VaR calculado significa que se estima que sólo el 10 % de los hombres

sobreviva 31.94 años una vez ha alcanzado la edad 65. En cambio, para las mujeres se estima que vivan 34.88 años más allá de 65. La tabla 3.4 muestra el VaR calculado al nivel de confianza 0.95, de nueva cuenta se observa que el número de años que sobreviven las mujeres es mayor al de los hombres. El VaR crece a medida que la proporción femenina en la población se incrementa. Esta situación se observa para todos los años de estudio y es más notable en los años 2005 y 2010 con 2.56 y 3.13 años de diferencia respectivamente.

Ahora, si tomamos un nivel de confianza más estricto, 0.95, podemos decir que en 2010 cinco de cada cien mujeres vivirán 40.55 años después de haber cumplido 65 años. Para los hombres, a partir del VaR al 0.95, se tiene que cinco de cada cien hombres vivirán 37.42 años más allá de los 65 años.

El riesgo de longevidad es mayor cuando la proporción de mujeres va creciendo, comparando siempre con el caso en donde se tiene una población íntegramente masculina. Una compañía de seguros debe tener en cuenta que si su cartera está compuesta principalmente por mujeres entonces la reserva asociada a anualidades o seguros de vida deberá ser mayor a fin de prevenir pérdidas futuras.

Todos estos análisis se realizaron con información referente a la población general mexicana, sin embargo, como se ha mencionado en otros capítulos, se espera que la población asegurada tenga un nivel de vida mejor que la media poblacional, ya que pueden tener más recursos para la prevención de enfermedades, además de que su capacidad para comprar un seguro los hace ser un sector de la población con mejores condiciones de vida. En consecuencia, una persona asegurada por lo general vivirá más tiempo que un no-asegurado de edad similar y mismo sexo. Dicho esto, debemos tener en cuenta que los resultados que se han mostrado podrían ser menores a los resultados que se obtendrían si examináramos una cartera de seguros en lugar de la población en general.

Una limitación que debe destacarse aquí es que el uso de la distribución Weibull puede no ser acertada, sobre todos para el análisis de las edades más avanzadas como se verán los capítulos siguientes.

3.4. Conclusiones

A partir de la resolución de la Directiva Europea de seguros en diciembre de 2012, las compañías se enfrentaron al hecho de tener que modificar todos los

sistemas que ya tenían implementados para la gestión de riesgos y suscripción. En primera instancia se puede pensar que el camino más fácil y con menos riesgo es nivelar las tarifas de ambos sexos al precio del género con mayor costo, por ejemplo, en el seguro de vida se aumentaría la prima a las mujeres. Esta elevación en las primas se mantendría hasta tener suficiente experiencia para saber las consecuencias reales resultantes del uso de tarifas unisex. A nivel de mercado el aumentar las primas quizá podría afectar a las ventas para el sector afectado, pero también pudiera ocurrir el efecto llamada para las personas que se benefician de la bajada en costes, sin duda el marketing tendrá un papel relevante en la venta de seguros.

A menudo se argumenta que la eliminación de factores de riesgo en la suscripción puede provocar selección adversa. Esto significa que personas de bajo riesgo al verse afectadas por la subida de precios decidan terminar su contrato con la aseguradora y en cambio los asegurados que dan mayor uso a la póliza se mantendrían, lo cual crea un desequilibrio en la cartera de seguros. En el caso de los seguros de vida, ya que los hombres se benefician de precios más bajos, cancelarán menos sus pólizas y también habrá más clientes interesados en adquirir este seguro, lo contrario que pasaría con el colectivo de las mujeres. En el límite, es posible que en las compañías de seguros se tengan clientes de un sólo tipo, de hecho los que más costos generarían a la aseguradora, por lo que para poder asumir todos los gastos la aseguradora tendría que realizar un nuevo ajuste de los precios aumentándolos.

Dado que, para los años estudiados, el análisis ha reflejado que el riesgo de longevidad es diferente para hombres y para mujeres, se sugiere a las compañías aseguradoras considerar este hecho y manejar de forma separada la gestión del riesgo aunque no el precio de los seguros. De esta manera no se violaría la normativa Europea de no discriminar por sexos e internamente la compañía aseguradora seguiría con una gestión del riesgo apegada a la realidad. Por lo tanto, para calcular el riesgo de longevidad o mortalidad se debería hacer uso de tablas de mortalidad diferentes para cada sexo.

El uso de una tabla de mortalidad unisex, construida como una combinación lineal de la mortalidad femenina y de la mortalidad masculina es útil para fijar precios de las primas, pero no debe sustituir los cálculos y proyecciones realizados para cada uno de los géneros.

Este capítulo se ha centrado en encontrar diferencias en la longevidad de hombres y mujeres, se ha ajustado un modelo paramétrico a la variable dura-

ción de la vida, sin embargo es complicado encontrar una función de distribución que pueda reflejar completamente el comportamiento de dicha variable, por lo que se prevé la necesidad ajustar modelos noparamétricos.

4 Brass-type Poisson aplicado a tasas de mortalidad de asegurados

En este trabajo estudiamos la mortalidad de la población asegurada mexicana. El objetivo es cuantificar el riesgo de longevidad en la población asegurada a través de modelos que parten de la población general, estableciendo una relación entre la mortalidad de dos poblaciones.

4.1. Introducción

En el área de las ciencias actuariales es de especial importancia realizar una gestión del riesgo eficiente, que permita el correcto cálculo de reservas y tarifas para los diferentes productos del sector vida como son: las rentas vitalicias, los planes de pensiones y los seguros tradicionales con cobertura de muerte. De acuerdo a los productos del seguro de vida, los riesgos que se deben cuantificar son el riesgo de longevidad y el riesgo de mortalidad.

Desde el punto de vista de un asegurador, el riesgo de longevidad y el riesgo de mortalidad se han ido transformando al pasar los años, debido a las mejoras sanitarias y tecnológicas que se han desarrollado en las últimas décadas. La esperanza de vida al nacimiento sigue en crecimiento. En el caso de México, de acuerdo con el Banco Mundial (2013), en el año 1990 la esperanza de vida en los hombres era de 68 años y para las mujeres de 74 años. Para el año 2000 estos números se incrementaron a 72 y 77 años, respectivamente. Más recientemente, en el año 2010, la esperanza de vida para los hombres fue de 74 años y para las mujeres de 79 años. La diferencia en años entre la esperanza de vida de hombres y mujeres se ha mantenido casi inalterable en 5 años, aproximadamente, en las últimas dos décadas.

Si bien tanto en el ámbito demográfico como en el actuarial se han realizado trabajos destinados a modelar y proyectar la mortalidad en México, no hay tablas de mortalidad que reflejen su situación real. Como se mencionó en el capítulo 1, las compañías aseguradoras utilizan, para el cálculo de las reservas,

las tablas de mortalidad publicadas en Diciembre del año 2000. De manera interna algunas compañías de seguros construyen tablas de mortalidad para tarificar sus productos de seguro de vida. Recientemente la Asociación Mexicana de Instituciones de Seguros (AMIS) ha calculado tablas de mortalidad con la experiencia de algunas compañías del sector asegurador. Estas tablas de mortalidad han sido empleadas en este trabajo, un análisis más detallado se puede encontrar en la sección 1.4.1 del capítulo 1.

En la construcción de tablas de mortalidad en seguros, contemplar únicamente la mortalidad de los asegurados, es suponer que la mortalidad dentro de este subconjunto de la población se mantendrá estable, cuando en realidad es un grupo con características específicas que pueden ser muy cambiante y que no se comporta de forma análoga a la población general. Un aspecto relacionado con este hecho es la selección adversa, que se ve afectada por los hábitos de los asegurados, ya que estos poseen información que la aseguradora no podrá obtener. A este fenómeno se le llama información asimétrica, por ejemplo, la aseguradora podrá tener información obtenida a partir de los cuestionarios y en algunos casos de exámenes médicos, sin embargo no puede llegar a conocer con absoluta certeza el historial de la familia o los motivos por los cuales se ha adquirido un seguro. Si los asegurados llegan a ocultar características que indiquen un mayor riesgo de muerte, entonces la aseguradora, no siendo consciente de dicho sesgo, podría estar subestimando el verdadero riesgo. Más información al respecto se puede encontrar en Chiappori y Salanié (2000).

El caso opuesto, cuando el control de la selección adversa es efectivo, conduciría a primas demasiado elevadas, debido a la sobre estimación del riesgo, ya que sólo considera la experiencia de la población en general en lugar de la subpoblación de asegurados. Los ciudadanos que suscriben un seguro tienen un poder adquisitivo que trasciende en un estilo de vida que les permite estar más saludables.

Según Plat (2009), el problema en la información de la mortalidad en seguros no se reduce a la antiselección, sino que también hay problemas intrínsecos del manejo de información. Las bases de datos de las aseguradoras, en muchos casos, no van poco más allá de 15 años de antigüedad. Otra deficiencia es que el porcentaje de personas que contrata un seguro es pequeño. También subraya el hecho de que los asegurados con mayores sumas aseguradas son los que presentan menor mortalidad.

En el caso particular de la población asegurada en México, los dos primeros

puntos se cumplen: hay un posible efecto de selección adversa y las bases de datos con información de asegurados son insuficientes. Si bien en los últimos años la CNSF ha hecho pública las estadísticas anuales recogidas por el Sistema Estadístico del Sector Asegurador (SESA), no hay información por un largo periodo de años. En cuanto a la participación del seguro de vida en el total del mercado asegurador, según la CNSF, las primas directas del seguro de vida en México conformaban el 34.3% del total de primas recaudadas en el año 2000, para el año 2010 este porcentaje se vio incrementado al 40%. Los planes de pensiones en cambio han sufrido una caída abrupta pasando del 13% en el 2000 a 5.13% en 2010. En cuanto al número de asegurados en planes individuales de vida es de 11,678,327 según registra la SESA del año 2010. Lo que representa el 14.6% de la población censada en México en edad de adquirir un seguro en el mismo año, que asciende a 79,983,893 personas.

El método sobre cómo relacionar la experiencia de mortalidad de la población general y la experiencia de mortalidad de la población asegurada, afecta directamente a la valoración de los riesgos. De los posibles modelos relacionales hemos elegido los modelos Brass-type lineal y Poisson, que se usan frecuentemente en la literatura actuarial.

4.2. Metodología

4.2.1. Modelo relacional Brass-type

Como se ha explicado en el capítulo 2, el modelo de Brass-type fue propuesto por Brass (1971). En forma de recordatorio, en la ecuación (4.1) se reescribe el modelo, utilizando con la función logaritmo como link, entre las tasas de mortalidad de dos poblaciones. A partir de esta ecuación, se puede crear una tabla de vida nueva cambiando los parámetros que establecen la relación lineal entre ambas tasas.

$$\log m_{x,t}^{aseg} = \beta_0 + \beta_1 \log m_{x,t}^{ref} + \epsilon_{x,t}. \quad (4.1)$$

donde $m_{x,t}^{aseg}$ es la tasa de mortalidad de la población asegurada y $m_{x,t}^{ref}$ es la tasa de mortalidad de la población global. El parámetro β_1 se interpreta como el cambio en la mortalidad que hay de una población a otra. En cuanto β_0 es un incremento o decremento constante, que varía el nivel de mortalidad de acuerdo

a la de referencia. El término $\epsilon_{x,t}$ captura las fluctuaciones aleatorias entre los logaritmos de las tasas de mortalidad. La mortalidad de los asegurados $m_{x,t}^{aseg}$ de edad x en el año t se obtiene directamente despejando de la ecuación (4.1).

$$m_{x,t}^{aseg} = \exp \hat{\beta}_0 \cdot (m_{x,t}^{ref})^{\hat{\beta}_1}. \quad (4.2)$$

De acuerdo a Brouhns *et al.* (2002) el suponer un modelo lineal implica suponer que los residuos son homocedásticos y que además se distribuyen como una normal, lo cual no sucede. Por ello, los autores implementan una modificación al modelo suponiendo que las muertes $d_{x,t}$ de individuos de edad x en el periodo t , se comportan como una Poisson($l_{x,t} \cdot m_{x,t}$), donde las tasas de mortalidad $m_{x,t}$ son:

$$m_{x,t}^{aseg} = \exp\{\beta_0 + \beta_1 \log m_{x,t}^{ref}\}. \quad (4.3)$$

Con esta especificación las desviaciones estándar de los parámetros estimados son menores a los que se obtendrían con la ecuación (4.1). Los parámetros de este modelo se interpretan de la siguiente manera: la exponencial de β_0 es el valor esperado de la respuesta cuando la variable explicativa numérica vale cero, y la exponencial de β_1 representa la tasa de incremento del valor esperado de la respuesta al aumentar una unidad la variable explicativa. Esta versión del modelo Brass-type se utilizará para explorar la mortalidad de los asegurados en función de la mortalidad del total de la población mexicana.

4.3. Datos

En este capítulo se ha utilizado como población de referencia la población mexicana, los datos se encuentran descritos en el capítulo 1, sección 1.1. La población asegurada es la descrita en el apartado 1.4.1 en el mismo capítulo. Se tienen tres años de estudio: 2000, 2005 y 2010. Ambas poblaciones tienen información desglosada por género y para las edades de 0 a 100. Suponiendo cien años como la edad máxima, es decir, a esta edad todos los asegurados se han extinguido. Para el ajuste realizado en este trabajo, se ha tomado como edad mínima los doce años de edad.

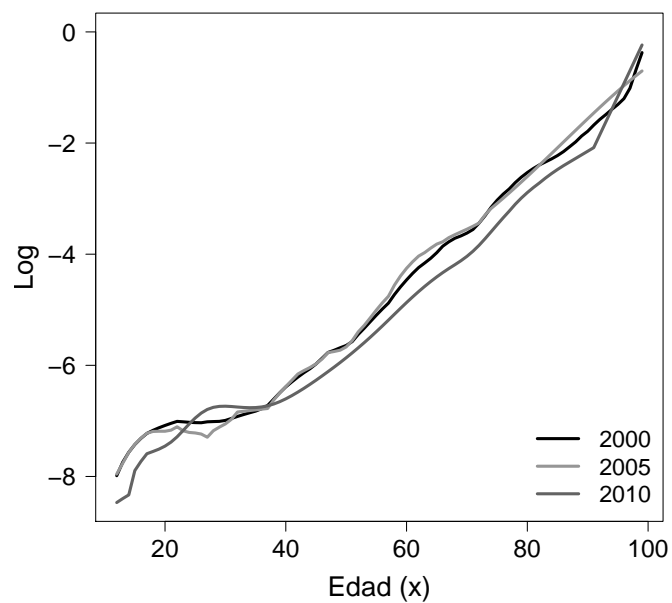
4.4. Resultados

Con los datos descritos anteriormente se han realizado los cálculos necesarios para obtener los parámetros estimados del modelo relacional Brass-type lineal y Poisson. El modelo lineal se ha estimado de la misma manera que en el capítulo 3.

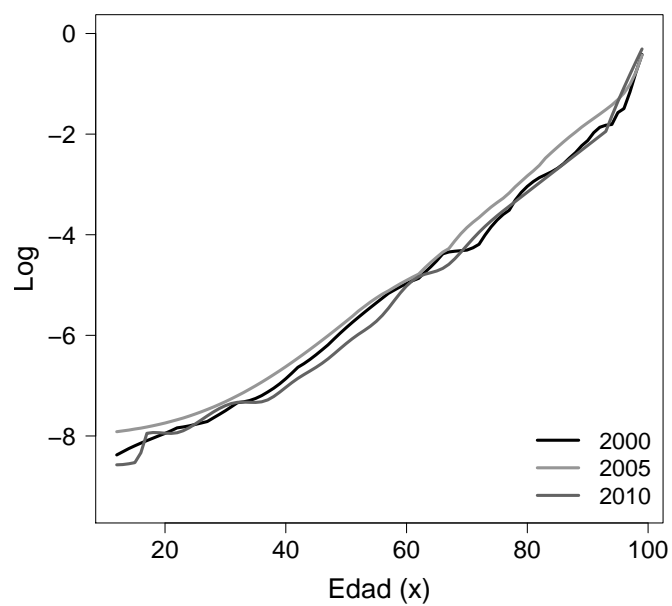
De manera descriptiva, en la figura 4.1 se muestra el logaritmo de la mortalidad de los asegurados en los tres años de estudio. Se puede ver que la joroba de los accidentes, que es característica de la mortalidad en edades jóvenes, también es visible en la población masculina que cuenta con un seguro. En el gráfico de los hombres se nota que, para el año 2010, la joroba relacionada con los accidentes, se alarga hasta la edad de 40 años, mientras que para 2000 y 2005 esta acaba en la edad 30 aproximadamente. Además, el máximo de mortalidad en este periodo se traslada de la edad 20 años a la edad 30 y la mortalidad después de los 40 años es menor en el año 2010.

Las mujeres presentan una mortalidad homogénea en los tres años analizados, aunque se nota que la mortalidad en el año 2005 es mayor a la de los otros dos años. En 2010 se observa un pico alrededor de la edad 20, que bien podría semejarse a la joroba de accidentes en los hombres. Este será un punto a revisar en la mortalidad de los siguientes años, pues con los cambios en el estilo de vida de las mujeres es posible que se incrementen el número de muertes.

En la tabla 4.1 se presentan los parámetros estimados suponiendo que el número de muertos sigue una distribución Poisson. La $\hat{\beta}_1$ que corresponde a la pendiente de la recta estimada, en todos los casos es mayor que uno, indicando que el logaritmo de la mortalidad de los asegurados crece $\hat{\beta}_1$ cuando el logaritmo de la mortalidad de la población en general aumenta. En el caso de las mujeres, los parámetros estimados son cercanos a uno en los tres años de estudio, por lo que la diferencia entre el logaritmo de la mortalidad de mujeres aseguradas y la población general es menor que la diferencia entre poblaciones masculinas.



(a) Hombres



(b) Mujeres

Figura 4.1: Logaritmo de las tasas brutas de mortalidad de los asegurados (AMIS-AMA) en México de edad 12 a 99 en los años 2000, 2005 y 2010

Mujeres				
	$\hat{\alpha}$	$\hat{\beta}$	DS $\hat{\alpha}$	DS $\hat{\beta}$
2000	- 0.030	1.065	0.007	0.002
2005	0.179	1.059	0.006	0.001
2010	- 0.025	1.081	0.006	0.002

Hombres				
	$\hat{\alpha}$	$\hat{\beta}$	DS $\hat{\alpha}$	DS $\hat{\beta}$
2000	0.684	1.229	0.008	0.002
2005	0.657	1.208	0.006	0.002
2010	0.333	1.225	0.006	0.002

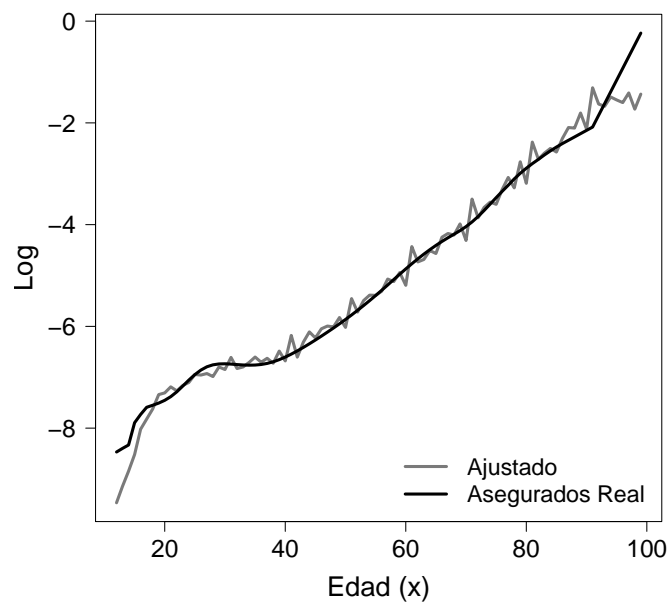
Tabla 4.1: Parámetros estimados por el modelo Brass-type de acuerdo a la regresión Poisson. DS indica desviación estándar

En la figura 4.2 se compara el logaritmo de la mortalidad en los asegurados y logaritmo de la mortalidad en los asegurados estimado por el modelo Poisson para el año 2010. Los residuos aumentan a medida que aumenta la edad; además con el incremento de la edad se observa una mayor variabilidad en los datos. Esto sucede por el menor número de datos en las edades más extremas, esta circunstancia se analizará con mayor detalle en el próximo capítulo.

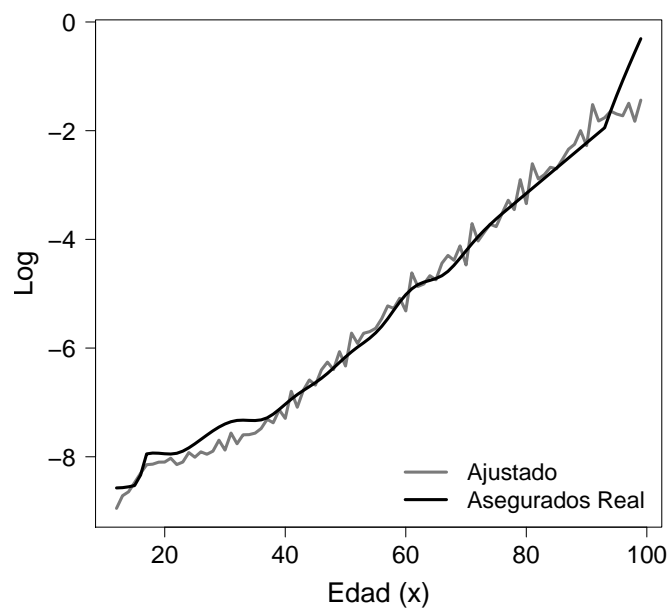
Al estar interesados en el análisis de la longevidad es útil e importante calcular el riesgo de que una persona llegue a una cierta edad. En el artículo Ornelas *et al.* (2013) se usa el valor en riesgo (VaR) como medida del riesgo de longevidad. El cuantil obtenido, según un nivel de confianza, será el número de años a partir del cual la población sobrevivirá con una probabilidad muy reducida. Para calcular el VaR se emplean las tasas de mortalidad estimadas por el modelo relacional Brass-type para edades mayores a 65 años en el periodo que transcurre entre 2000 y 2010. Se ha calculado la función de supervivencia empírica y el VaR. Se han fijado dos niveles de confianza: 0.90 y 0.95.

La tabla 4.2 muestra el VaR obtenido en tres casos distintos: la primer fila corresponde a la mortalidad real de los asegurados (AMIS-AMA), la segunda fila corresponde a la mortalidad estimada por la regresión Poisson y, finalmente, la tercera fila muestra el VaR de la mortalidad estimada por la regresión lineal. La diferencia entre hombres y mujeres es evidente. En todos los casos se espera que después de los 65 años las mujeres vivan más años que los hombres.

Considerando un nivel de confianza de 0.90, por ejemplo, en el año 2010



(a) Hombres



(b) Mujeres

Figura 4.2: Comparativo entre el logaritmo de la mortalidad real de los asegurados y el logaritmo de la mortalidad estimada por el modelo Brass-type Poisson. Año 2010

Hombres	VaR 0.90		VaR 0.95	
	2000	2010	2000	2010
Asegurados	31.11	36.76	36.30	42.58
Regresión Poisson	30.06	35.88	34.89	41.50
Regresión Lineal	31.19	35.67	36.28	41.29

Mujeres	VaR 0.90		VaR 0.95	
	2000	2010	2000	2010
Asegurados	40.14	39.14	46.50	45.05
Regresión Poisson	37.39	38.20	43.32	44.07
Regresión Lineal	37.82	38.12	43.89	44.03

Tabla 4.2: Valor en riesgo estimado: Número de años a sobrevivir después de la edad 65, con un nivel de confianza 0.90 y 0.95

es posible que las mujeres aseguradas vivan más de 39.14 años, más allá de los 65, con una probabilidad de 0.1. Sin embargo, cuando se usa el modelo Poisson este número de años baja a 38.12. El VaR para las mujeres no cambia substancialmente entre la regresión Poisson y la regresión lineal. Esto se podría explicar por el hecho de que la mortalidad de las mujeres es más estable a través de los años que la de los hombres, como se puede observar en la figura 4.1. Además, a partir de los resultados de la tabla 4.1, se deduce que la mortalidad en mujeres aseguradas se comporta de forma más similar que la de los hombres asegurados a la mortalidad de la población general.

Para los hombres asegurados el VaR con un nivel de confianza de 0.90 y para el año 2010 es 36.76 años. En cambio, si se tiene en cuenta la mortalidad de la población general mediante el modelo Poisson, este número de años baja a 35.67. En este caso las diferencias entre estimación lineal y Poisson superan el año de vida. El ajuste lineal aminora el número de años a sobrevivir en comparación con el modelo Poisson.

4.5. Conclusiones

Al existir diferencias visibles entre la mortalidad de los asegurados y la mortalidad general, una compañía aseguradora debería tener en cuenta ambas experiencias y hacer de esta actividad un ejercicio cotidiano en la construcción de tablas de mortalidad. La recomendación de estudiar la mortalidad de forma separada para hombres y mujeres sigue vigente, pues hasta la fecha no se ha

encontrado evidencia con los datos de que las diferencias en la mortalidad entre sexos haya disminuido y por lo tanto se pueda eludir el factor sexo. En el ámbito de los seguros no es suficiente con calcular la probabilidad de muerte, si no que también interesa la esperanza de vida, que define el tiempo promedio de años que se espera que una persona sobreviva. El cálculo del VaR siempre será útil para conocer la situación de la cartera, ya que indica a qué edad es probable que llegue viva un pequeño porcentaje de la población $(1 - \alpha) \%$ una vez se ha fijado un nivel de confianza α .

5 Estimación del riesgo de longevidad, comparación de métodos paramétricos y noparamétricos

En los capítulos 3 y 4 se estudió el riesgo de longevidad desde el punto de vista paramétrico, en este capítulo además se han estimado las funciones de distribución acumuladas con métodos noparamétricos y semiparamétricos. Se han comparado los tres métodos de estimación utilizados: paramétrico noparamétricos y semiparamétricos.

5.1. Introducción

Como se ha dicho en capítulos previos la mortalidad ha experimentado un decrecimiento substancial durante el siglo XX. Este hecho afecta directamente a todos los cálculos basados en la esperanza de vida, por ejemplo, en el caso de los seguros, a los planes de pensiones, los seguros de dependencia, los seguros de salud y los seguros de vida.

La relación entre mortalidad y esperanza de vida es obvia, si la mortalidad decrece, entonces la esperanza de vida aumenta. La esperanza de vida se puede definir como el promedio de años que una persona puede sobrevivir. Aunque se conozca la esperanza de vida para alguna población, en realidad no se tiene certeza de la edad máxima que un individuo pueda alcanzar. Debido a esta incertidumbre, la longevidad es una situación de riesgo cuando su aumento implica pérdidas económicas. Por ejemplo, las compañías aseguradoras deben tener especial cuidado en si su cartera de clientes envejece, las primas de seguros y los planes de pensiones se ven afectados directamente por la longevidad.

Las compañías de seguros invierten parte de sus recursos en el desarrollo de metodologías para evaluar riesgos. La precisión es algo indispensable en las estimaciones que se realizan. Por lo que, aunque la edad de muerte general-

mente se mide en años, lo ideal sería tener una aproximación que considere edades de tiempo menores y que permita una valoración más precisa de los productos aseguradores y/o financieros cuyo riesgo está ligado a la esperanza de vida, por ejemplo, utilizar los meses como unidad de medida. Las empresas necesitan tener modelos que les permitan interpolar, con precisión, todos los valores de la distribución de la variable tiempo de vida, es decir, ir más allá de la información que la muestra provee.

Nos interesa evaluar el riesgo de que un individuo sobreviva más allá del estándar previsto. En particular nos interesa la parte extrema de la distribución de la variable tiempo de vida. Sin embargo, la información en dicha región es escasa lo que dificulta la estimación del riesgo de longevidad.

El análisis se ha llevado a cabo desde la edad 65, la que se considera la edad de jubilación. El objetivo de este capítulo es estimar el riesgo de longevidad y su evolución de acuerdo con la edad que alcanza el individuo, por dicho motivo se han calculado funciones de distribución (cdf, *cumulative distribution function*) truncadas suponiendo diferentes edades alcanzadas.

Como método de estimación de los parámetros asociados a las distintas distribuciones asociadas, hemos propuesto un método basado en la minimización de la distancia entre la función de distribución teórica y su estimación. Alternativamente, también se han aplicado métodos noparamétricos para estimar la función de distribución truncada.

Para evaluar el riesgo se ha usado el valor en riesgo (VaR) definido como la inversa de la función de distribución estimada truncada, es decir, el VaR es un cuantil de la variable tiempo de vida condicionado a que el individuo ha alcanzado una determinada edad.

Algunos autores, por ejemplo Kannisto (2000) y Cheung *et al.* (2005), han observado que la esperanza de vida ha aumentado más lentamente en algunas poblaciones, este fenómeno se denomina compresión de la mortalidad (ver Fries, 1980). Esto significa que el intervalo de edades donde se concentra el mayor número de muertes ha permanecido sin cambios en los últimos años. Por el contrario, Yue (2012) ha demostrado que las muertes se presentan en edades cada vez mayores. No existe una verdad absoluta acerca de los límites de la vida, esta incertidumbre hace que evaluar el riesgo de longevidad sea necesario.

Al igual que en los capítulos 3 y 4, se ha aplicado el concepto de valor en riesgo al nivel de confianza α para los individuos con edad mayor a 65 años. El

VaR se utiliza comúnmente para evaluar las pérdidas monetarias, la longevidad no es una pérdida monetaria por si misma, pero se relaciona directamente con actividades que implican mayores gastos cuando las personas envejecen. Se desea medir el grado de envejecimiento de la población asegurada, obteniendo el número máximo de años que una pequeña proporción de la población podría llegar a sobrevivir.

La forma de la distribución asociada a la variable duración de la vida a partir de los 65 años es asimétrica a la derecha, es decir, los individuos con 65 años son frecuentes, sin embargo, los individuos con edades muy avanzadas son muy pocos. Esto supone la dificultad de encontrar una distribución que ajuste en todo el dominio de la variable tiempo de vida, especialmente en los valores extremos, sin subestimar o sobrestimar el riesgo. Ante ello, las distribuciones de valor extremo proporcionan modelos con asimetría a la derecha con formas alternativas.

Dado que nos interesa la supervivencia en la población con mayor edad, nuestra principal preocupación es obtener un buen ajuste en la cola de la distribución de la variable duración de la vida. Dado que el número de personas supervivientes disminuye considerablemente cuando la edad se aumenta, podemos decir que estamos hablando de “eventos raros”, es decir, aún cuando se simule o se disponga de una muestra de gran tamaño, el número de los datos observados en el extremo derecho será pequeño. Por tal motivo, proponemos el uso de un método de estimación paramétrica basado en la clase de estimadores de distancia mínima, que es más robusto que la estimación máxima verosímil ante la aparición de eventos raros (ver Scott, 2001) y, por tanto, proporciona mejores ajustes en la cola derecha de la distribución. Los parámetros se estiman minimizando la distancia entre la distribución empírica y la función de distribución estimada.

El tamaño de la muestra es un factor clave para determinar el uso de un método noparamétrico o paramétrico. La aproximación noparamétrica más simple es la distribución empírica (Emp), aunque no es de gran utilidad porque esta estimación solo está definida hasta la edad máxima observada en la muestra y por tanto no nos permite extrapolar el comportamiento de la distribución por encima de dicha edad máxima. Por tanto el uso de la distribución empírica podría ocasionar una subestimación del riesgo, dado que supone que la probabilidad de que un individuo sobreviviera a la edad máxima observada en la muestra es cero, lo cual no es cierto.

Además, si el VaR_α se calcula empíricamente se obtiene una edad en años, sin embargo, los costes ligados a las políticas económicas, las pensiones y las ayudas a la dependencia deben estimarse con la mayor precisión posible. A la mejora de dicha precisión podría contribuir el hecho de poder interpolar la función de distribución teniendo en cuenta que la edad es una variable continua.

Una aproximación noparamétrica alternativa a la distribución empírica es la estimación núcleo clásica (CKE, *Classical Kernel Estimator*) de la función de distribución, que suaviza la forma de la primera y permite “extrapolar” su comportamiento más allá del máximo observado. De nuevo como el número de observaciones en la cola derecha de la distribución es reducido, el CKE no suaviza la forma de la distribución empírica y, por tanto, no estima de forma eficiente la distribución empírica por encima del valor máximo observado en la muestra. Por esta razón, se propone un procedimiento de estimación que combina un ajuste paramétrico con la estimación núcleo clásica, este método se denomina estimación núcleo doble transformada (DTKE, *Double Transformed Kernel Estimator* (ver, Alemany *et al.*, 2013). Se considera al DTKE como un método semiparamétrico. Alemany *et al.* (2013) demostraron que el DTKE es un estimador consistente aunque sesgado, dicho sesgo depende de la distancia entre el modelo paramétrico utilizado en el ajuste de la distribución y la función de distribución teórica.

Existen otras aproximaciones que predicen la mortalidad ajustando su tendencia histórica, sin embargo la extrema longevidad en la población ha sido observada recientemente, por lo que no hay bases de datos históricas que reflejen dicha tendencia.

En este capítulo se presenta un procedimiento sistemático que permite evaluar el riesgo de longevidad teniendo en cuenta la edad alcanzada. En la sección 5.2 describimos las diferentes metodologías empleadas. En la sección 5.3 se describen los datos y los resultados obtenidos. Finalmente, en la sección 5.4 se presentan las conclusiones.

5.2. Metodología

Sea X la variable aleatoria que representa la duración de la vida y F es su función de distribución. Como hemos mencionado anteriormente, el mayor interés es la población que vive más de 65 años, por lo tanto, es necesario estimar la función de distribución truncada en el dominio $[a, \infty)$, donde $a \geq 65$.

La función de distribución truncada por la izquierda de la variable X se define en la ecuación (5.1) (ver Cramér, 1946).

$$F(x|X \geq a) = \frac{F(x) - F(a)}{1 - F(a)} = G(x). \quad (5.1)$$

Para estimar el riesgo de longevidad utilizamos el valor en riesgo (VaR_α) que es igual al cuantil α de la distribución:

$$\text{VaR}_\alpha(X) = G^{-1}(\alpha) = \inf\{x : G(x) \geq \alpha\}, \quad \alpha \in (0, 1), \quad (5.2)$$

donde $(1 - \alpha)$ indica la proporción esperada de personas que vivirán más allá del VaR_α calculado.

Hay diferentes maneras de estimar $G(x)$: ajustar un modelo paramétrico, usar una estimación noparamétrica que no impone ninguna forma a la distribución de la variable o usar una estimación semiparamétrica. En la siguiente sección se describen estos métodos.

5.2.1. Distribuciones paramétricas

Dada la forma asimétrica de la distribución de la variable tiempo de vida, para realizar una mejor estimación se debe utilizar una distribución de valores extremos (EVD, *Extreme Value Distribution*). Las EVDs más comunes con cola media son: Lognormal y Weibull. Una alternativa con cola pesada es la distribución de Champernowne, que fue presentada por D. G. Champernowne en 1936. La distribución Champernowne converge a una distribución de Pareto en la cola; más detalles se pueden encontrar en Buch-larsen *et al.* (2005). En la Tabla 5.1 mostramos la función de densidad y la función de distribuciones de los tres modelos utilizados en este análisis.

Para estimar los parámetros de las distribuciones se puede utilizar el método de máxima verosimilitud (MLE, *Maximum Likelihood Estimation*). En la practica el método de máximo verosímil no es robusto cuando existen valores atípicos en la muestra (ver Scott, 2001) y/o cuando el modelo no está bien especificado, esto último puede ocurrir en nuestro análisis de la longevidad, donde los modelos paramétricos pueden no ser correctos para el ajuste de la cola de la distribución del tiempo de vida. Una alternativa a la estimación máximo verosímil consiste en estimar los parámetros minimizando la distancia entre la función de distribución teórica y la estimada. Este método es más

	$f_{\theta}(x)$	$F_{\theta}(x)$	Parámetros θ
Weibull	$1 - e^{-\left(\frac{x}{\beta}\right)^{\alpha}}$	$\frac{\alpha}{\beta} \left(\frac{x}{\beta}\right)^{\alpha-1} e^{-\left(\frac{x}{\beta}\right)^{\alpha}}$	$\alpha, \beta > 0$
Lognormal	$\frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\log(x)-\mu}{\sigma}\right)^2}$	$\Phi\left(\frac{\log(x)-\mu}{\sigma}\right)$	$\mu \in R \quad \sigma > 0$
Champernowne	$\frac{\alpha(x+c)^{\alpha-1}((M+c)^{\alpha}-c^{\alpha})}{((x+c)^{\alpha}+(M+c)^{\alpha}-2c^{\alpha})^2}$	$\frac{(x+c)^{\alpha}-c^{\alpha}}{(x+c)^{\alpha}+(M+c)^{\alpha}-2c^{\alpha}}$	$\alpha, M, c > 0$

θ es el vector de parámetros

Tabla 5.1: Distribuciones paramétricas

robusto ante errores de especificación. Un criterio utilizado con frecuencia en los métodos noparamétricos consiste en minimizar el error cuadrático medio integrado (MISE, *Mean Integrated Square Error*):

$$\text{MISE}[F_{\hat{\theta}}] = E \left[\int [F_{\hat{\theta}}(t) - F_{\theta}(t)]^2 dt \right], \quad (5.3)$$

donde θ es un vector de parámetros, F_{θ} es la función de distribución teórica y $F_{\hat{\theta}}$ es la función de distribución estimada. La dificultad asociada con el cálculo de *MISE* es que depende de los parámetros teóricos θ .

Altman y Leger (1995) propusieron estimar el MISE a través de su aproximación discreta, sustituyendo la función de distribución teórica por la empírica (véase también, Bowman *et al.*, 1998). Sea X_1, \dots, X_n una muestra de la variable aleatoria X que representa el tiempo de vida. Su distribución empírica es igual a:

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I(X_i \leq x), \quad (5.4)$$

donde $I(\cdot)$ es la función indicador que toma el valor de 1 si la condición entre paréntesis es cierta y cero en caso contrario. La aproximación discreta del MISE es:

$$\widehat{\text{MISE}}[F_{\hat{\theta}}] = n^{-1} \sum_{i=1}^n [F_{\hat{\theta}}(X_i) - F_n(X_i)]^2. \quad (5.5)$$

En la figura 5.1 mostramos el $\widehat{\text{MISE}}$ asociado con la estimación de los parámetros de una distribución de Weibull¹ tal y como se define en la tabla

¹Para los ejemplos gráficos se han utilizado los datos descritos en la sección 5.3

5.2 (el parámetro de forma es α y el parámetro de escala es β). Se observa que existe un mínimo global (área más oscura). Se obtuvieron gráficos similares, usando los datos descritos en la sección 5.3, para las distribuciones Lognormal y Champernowne, donde los parámetros de forma son σ y α y los parámetros de escala son μ y M , respectivamente.

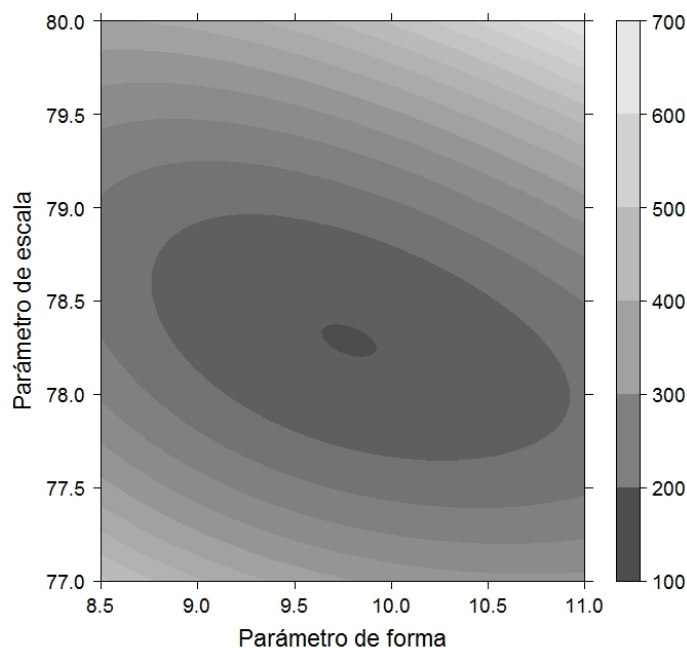
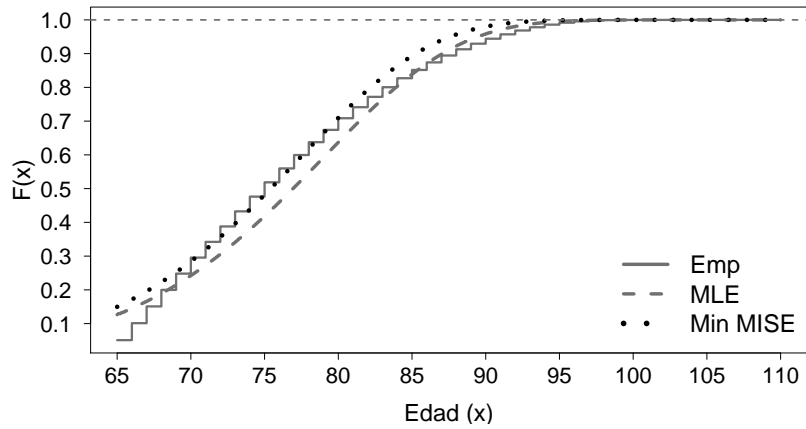


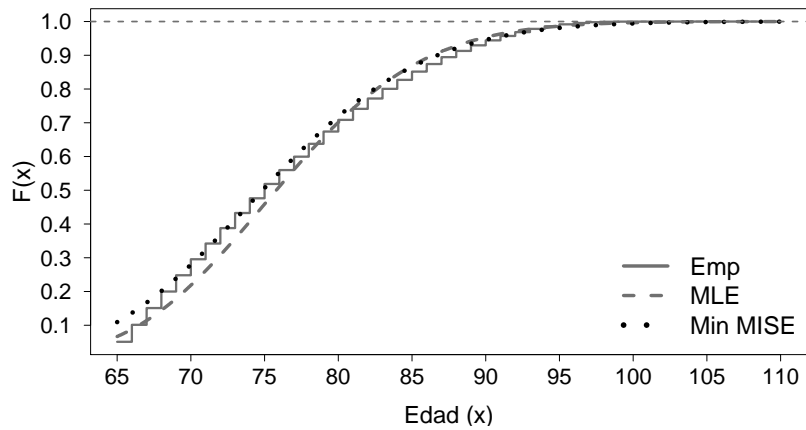
Figura 5.1: \widehat{MISE} para la distribución Weibull

Con el fin de comparar ambas estimaciones, en las figuras 5.2 y 5.3 se representan la distribución empírica junto a las distribuciones paramétricas estimadas por máxima verosimilitud y minimizando \widehat{MISE} . Centrándonos en las edades más avanzadas los resultados son diferentes dependiendo de la forma de la cola de la distribución. Para la distribución Weibull, la estimación basada en la minimización de \widehat{MISE} proporciona una cola aún más ligera que la estimación MLE, es decir, la función distribución alcanza el uno más rápidamente. Por el contrario, para la distribución Lognormal, que tiene una cola más pesada que la de Weibull, la estimación basada en la minimización de \widehat{MISE} proporciona una cola ligeramente más pesada que la estimación MLE. Finalmente, la distribución Champernowne, que es una distribución de cola pesada, tiene claramente una cola más gruesa cuando estimamos los parámetros minimizando \widehat{MISE} ².

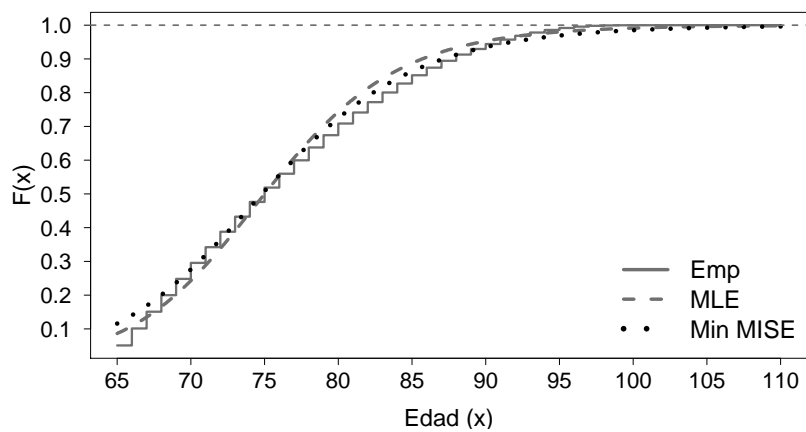
²Tener una cola más pesada implica que el crecimiento de la cdf es más lento



(a) Weibull

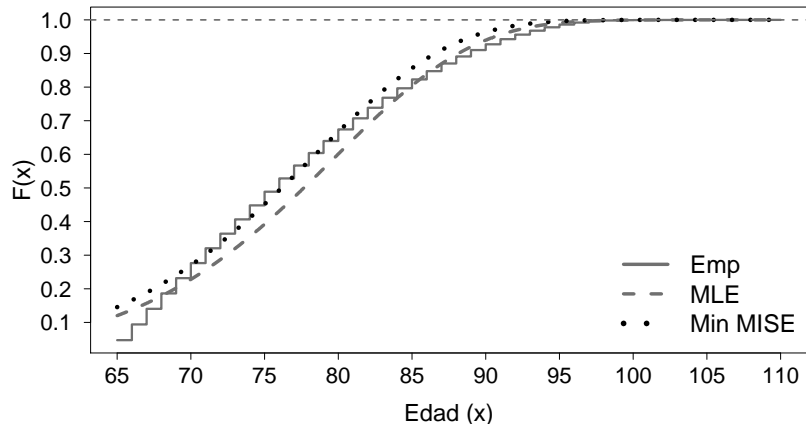


(b) Lognormal

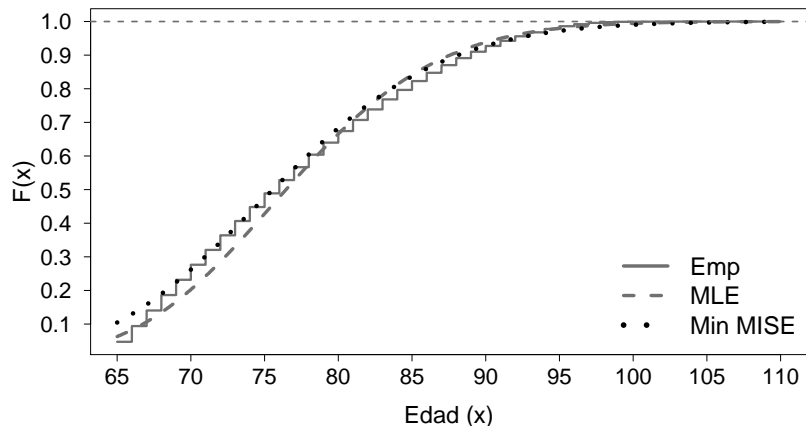


(c) Champernowne

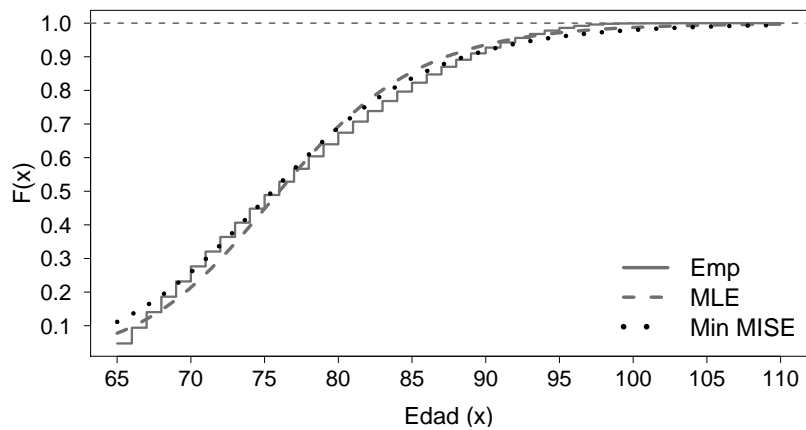
Figura 5.2: Función de distribución del tiempo de vida para los hombres. Distribución empírica (línea continua), estimación MLE (línea discontinua) y mínimo \widehat{MISE} (línea punteada)



(a) Weibull



(b) Lognormal



(c) Champernowne

Figura 5.3: Función de distribución del tiempo de vida para las mujeres. Distribución empírica (línea continua), estimación MLE (línea discontinua) y mínimo \widehat{MISE} (línea punteada)

5.2.2. Estimación no paramétrica y semiparamétrica

Los métodos no paramétricos son muy utilizados para estimar la función de densidad y la función de distribución de una variable aleatoria. Estos métodos se caracterizan porque es la información muestral la que dibuja la forma de la distribución.

La estimación no paramétrica de la función de distribución no impone una forma concreta para dicha función. Consiste en obtener una estimación del valor de la función en cada punto x del dominio de la variable aleatoria. Sin embargo, cuando los datos son escasos los métodos no paramétricos no son eficientes y los métodos semiparamétricos pueden ser una mejor alternativa.

Entre los métodos no paramétricos existentes, la distribución empírica definida en la ecuación (5.4) es el estimador de la función de distribución más sencillo. El VaR puede ser calculado directamente sustituyendo en la ecuación (5.2) la distribución empírica truncada: $G_n(x) = \frac{F_n(x) - F_n(a)}{1 - F_n(a)}$, donde $a \geq 65$.

La distribución empírica es fácil de obtener, pero no es una estimación eficiente (véase, Azzalini, 1981) ya que no puede extrapolar más allá de los datos observados. Esto es problemático cuando hay pocos datos, que es lo que ocurre cuando existen eventos raros, en este caso, se tiene información hasta una edad límite, sin embargo se puede sospechar que la probabilidad de alcanzar una edad mayor que la máxima observada no es cero. Además, si se utiliza la distribución empírica para cuantificar el riesgo, el resultado será un número entero, pues la distribución empírica sólo está definida en las edades observadas y no se puede extrapolar directamente.

Una alternativa a la distribución empírica es la estimación núcleo clásica de $F(x)$ (CKE, *Classical Kernel Estimator*):

$$\hat{F}(x) = \frac{1}{n} \sum_{i=1}^n K\left(\frac{x - X_i}{b}\right), \quad (5.6)$$

donde $K(t) = \int_{-\infty}^t k(s)ds$ es la función de distribución asociada a la función núcleo $k(\cdot)$. Algunos ejemplos de funciones núcleo son el Gaussiano y el de Epanechnikov (ver Silverman, 1986). En este trabajo se ha utilizado este último, que equivale a $K(x) = 3/4(1 - x^2)$ si $|x| \leq 1$. La elección de la función núcleo no afecta en demasía a los resultados de la estimación. El parámetro b en la ecuación (5.6) es el parámetro de alisado que es una constante positiva que controla el suavizado de la \hat{F} , valores grandes de b proporcionan una

estimación más suave.

Para estimar el valor en riesgo a partir de la estimación núcleo de la función de distribución es necesario utilizar un algoritmo numérico como es el de Newton-Raphson, que permite resolver la ecuación $\hat{G}(VaR_\alpha(X)) = \alpha$, donde $\hat{G}(x) = \hat{F}(x|X \geq a) = \frac{\hat{F}(x) - \hat{F}(a)}{1 - \hat{F}(a)}$, siendo $a \geq 65$ una edad determinada.

La elección de b sí afecta el resultado de la función de distribución estimada, además es especialmente importante cuando el objetivo de la estimación es encontrar un cuantil de la distribución estimada. Por lo tanto, el valor de b es fundamental para una estimación precisa del VaR. En la literatura, hay variedad de propuestas para el cálculo de b , en general se basan en la minimización de la distancia estimada entre $F(x)$ y $\hat{F}(x)$ (véase, por ejemplo, Altman y Leger, 1995; Bowman *et al.*, 1998). La dificultad de estos métodos es que proporcionan diferentes valores de b dependiendo del método que se utilice y, además, cuando se tiene una distribución de valor extremo, estos métodos pueden degenerar en $b = 0$. En este trabajo, para calcular parámetro de suavizado se ha utilizado la propuesta de Alemany *et al.* (2013) que se basa en el método propuesto por Silverman (1986) y que tiene en cuenta la existencia de valores extremos. En el caso clásico se utiliza:

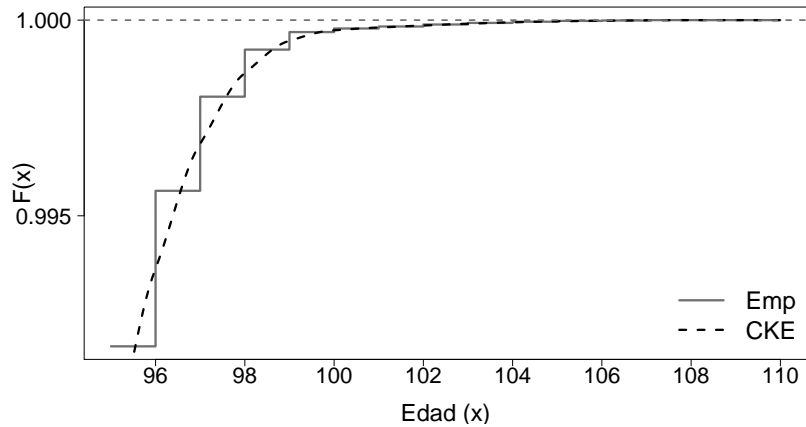
$$\hat{b} = \hat{\sigma}^{\frac{5}{3}} \left(\frac{8}{3} \right)^{\frac{1}{3}} n^{-\frac{1}{3}}, \quad (5.7)$$

donde $\hat{\sigma}$ es la desviación estándar estimada de la variable aleatoria X .

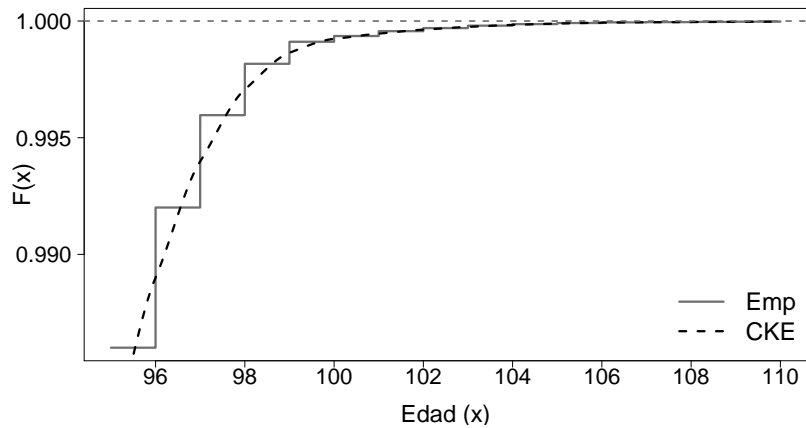
La estimación núcleo clásica de la función de distribución no es muy diferente a la distribución empírica, su principal diferencia es que la función de distribución empírica únicamente utiliza los datos por debajo de x , dando el mismo peso a todas las observaciones que influyen en la estimación de $F(x)$, mientras que el estimador núcleo utiliza los datos por encima y por debajo de x , dando más peso a las observaciones que están por debajo y que están más cerca de x .

La estimación núcleo suaviza la forma escalonada de la distribución empírica. En la figura 5.4, tanto para hombres como para mujeres, se compara la distribución empírica con la estimación núcleo de $F(x)$ de la variable tiempo de vida.

Con el objetivo de mejorar la eficiencia del estimador núcleo Alemany *et al.* (2013) propusieron el estimador núcleo doble transformado (DTKE, *Double Transformed Kernel Estimator*). El procedimiento se resume en tres pasos:



(a) Hombres



(b) Mujeres

Figura 5.4: Función de distribución del tiempo de vida. Distribución empírica (línea continua) y estimación núcleo (línea discontinua)

- **Paso 1:** Ajustar una función de distribución paramétrica minimizando $MISE$ y obtener los datos transformados $F_{\hat{\theta}}(X_1) = U_1, \dots, F_{\hat{\theta}}(X_n) = U_n$ que se distribuyen como una $Uniforme(0, 1)$.
- **Paso 2:** Los datos de la distribución uniforme se transforman usando la inversa de la distribución $Beta(3, 3)$, $M(x) = \frac{3}{16}x^5 - \frac{5}{8}x^3 + \frac{15}{16}x + \frac{1}{2}$.
- **Paso 3:** Se aplica el estimador núcleo clásico a los datos transformados en el paso 2, el resultado es a lo que se llama estimador núcleo doble

transformado:

$$\widehat{F}(x) = \frac{1}{n} \sum_{i=1}^n K \left(\frac{M^{-1}(F_{\widehat{\theta}}(x)) - M^{-1}(F_{\widehat{\theta}}(X_i))}{b} \right). \quad (5.8)$$

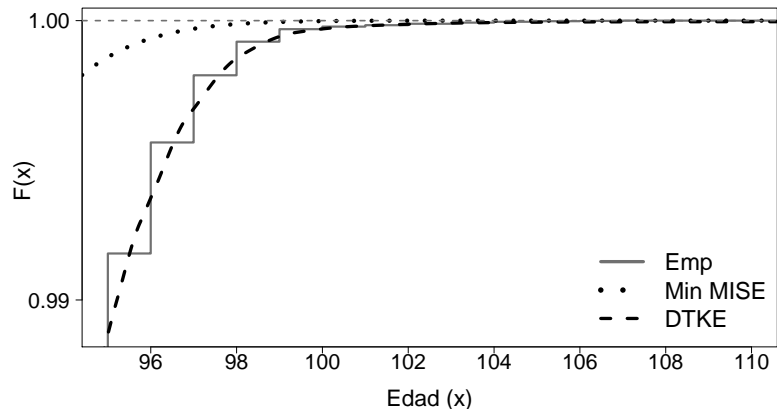
Alemaný *et al.* (2013) propusieron transformar los datos de modo que su distribución sea la de una *Beta* porque la forma de esta es la que mejor se estima utilizando el estimador núcleo clásico, es decir, el estimador núcleo doble transformado se basa en el hecho de que la función de distribución *Beta*(3, 3) se puede estimar de manera óptima usando el estimador núcleo clásico (ver Terrell, 1990). Dado que los datos transformados tienen una distribución conocida es posible calcular el valor óptimo de b para estimar el VaR, $\widehat{F}^{-1}(\alpha) = \widehat{VaR}_{\alpha}(X) = x$:

$$\widehat{b}_x = \left(\frac{m(M^{-1}(F_{\widehat{\theta}}(x))) \left(1 - \int_{-1}^1 K^2(t) dt\right)^{\frac{1}{3}}}{4 \left[\frac{1}{2} m'(M^{-1}(F_{\widehat{\theta}}(x))) \int_{-1}^1 t^2 k(t) dt\right]^2} \right)^{\frac{1}{3}} n^{-\frac{1}{3}}, \quad (5.9)$$

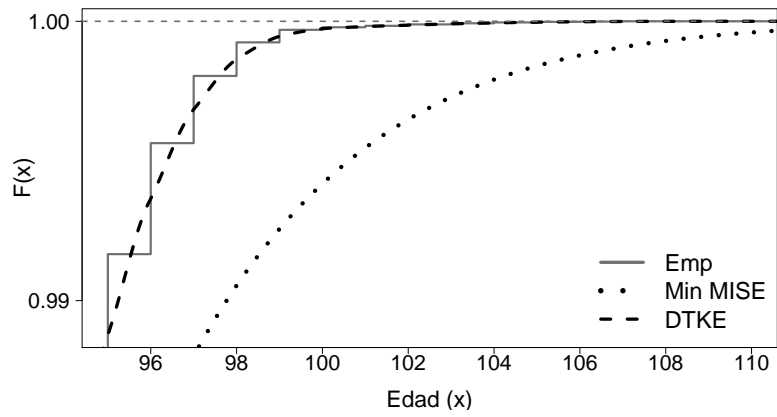
donde $m = M'$ es la función de densidad asociada a la distribución *Beta*(3, 3).

En el **Paso 1** Alemany *et al.* (2013) sugieren utilizar la estimación pseudo-máxima-verosimilitud de la distribución Champernowne. De manera alternativa, en este trabajo seleccionamos las distribuciones descritas en la tabla 5.1. Los parámetros se obtienen con el método basado en minimizar el \widehat{MISE} . Esta generalización no afecta a las propiedades asintóticas y mejora las propiedades en muestras finitas el estimador.

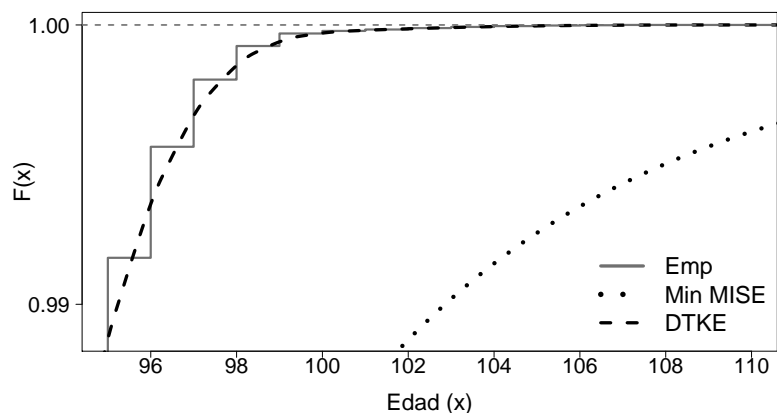
En las figuras 5.5 y 5.6 se representan de forma conjunta las distribuciones paramétricas, la distribución empírica y las estimaciones núcleo doble transformadas a partir de la Lognormal y la Champernowne. En ambas figuras se evidencia que el DTKE mejora el ajuste paramétrico mediante la reducción el sesgo. Además, este estimador es más preciso que los métodos puramente noparamétricos.



(a) Weibull

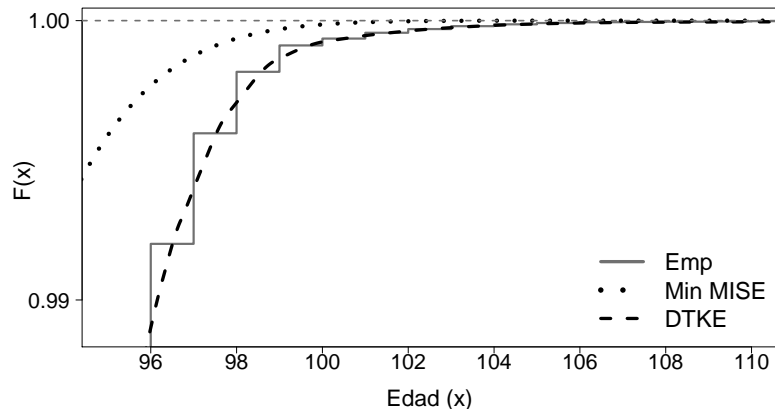


(b) Lognormal

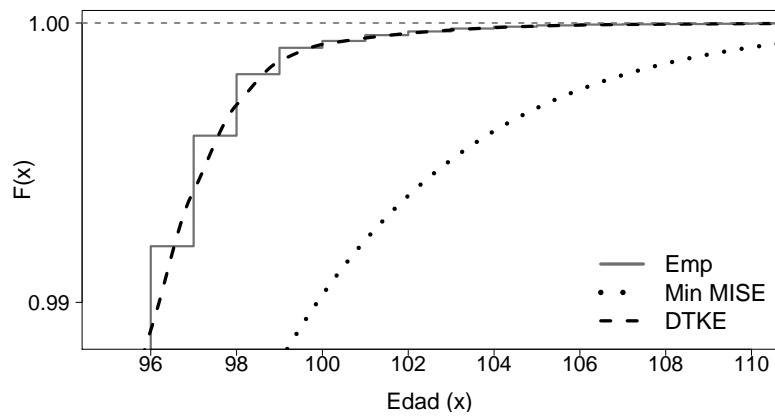


(c) Champernowne

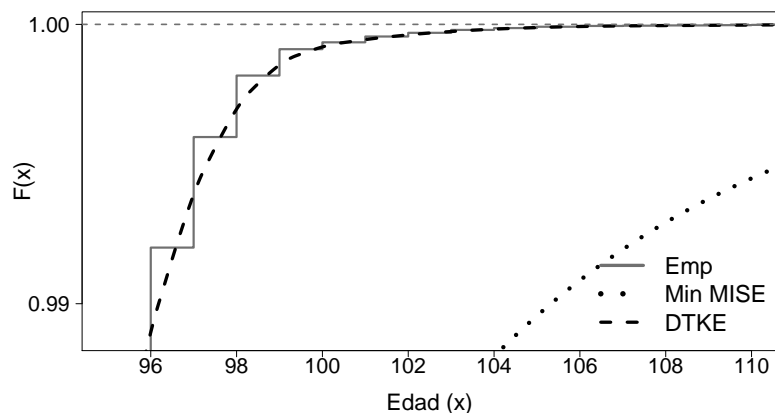
Figura 5.5: Función de distribución del tiempo de vida para hombres. Distribución empírica (línea continua), DTKE (línea discontinua)



(a) Weibull



(b) Lognormal



(c) Champernowne

Figura 5.6: Función de distribución del tiempo de vida para mujeres. Distribución empírica (línea continua), DTKE (línea discontinua)

En las figuras 5.7 y 5.8 comparamos las tres estimaciones doble transformadas obtenidas a partir de diferentes distribuciones paramétricas. Para observar las principales diferencias las funciones se han representado en dos intervalos de edad: $[95; 98)$ y $[98; 110)$. Observamos que cuando se usa una distribución paramétrica de cola pesada como primera transformación, la estimada crece más lentamente, es decir, estima un riesgo mayor. En definitiva, la distribución usada a priori tiene una mayor influencia en las edades más avanzadas.

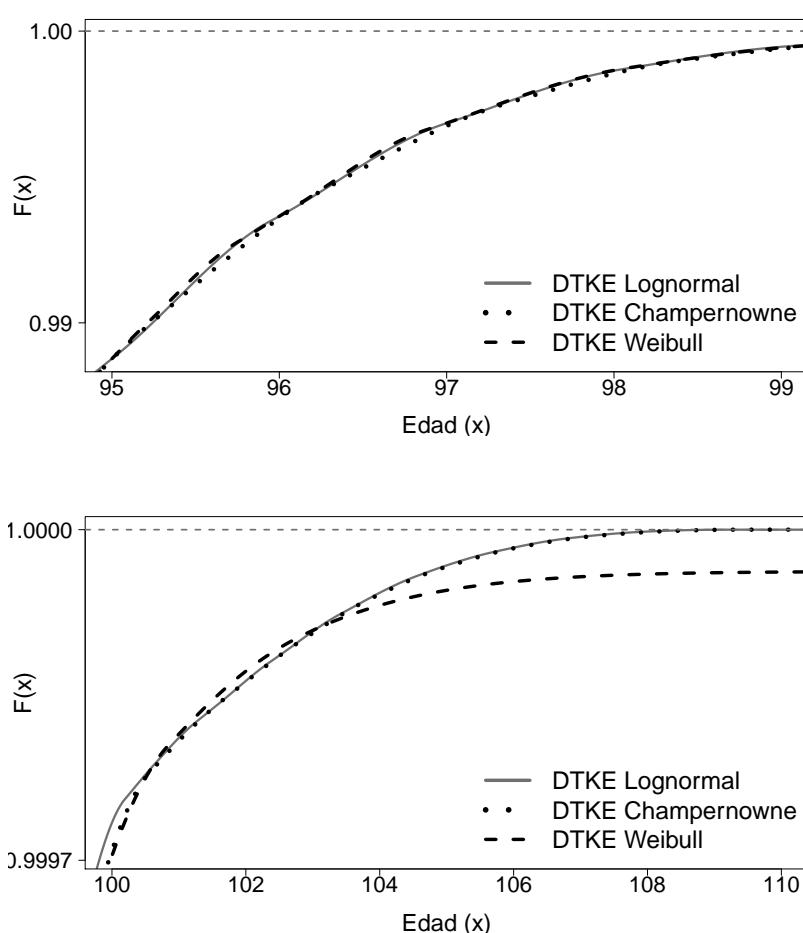


Figura 5.7: Función de distribución del tiempo de vida para hombres. DTKE Lognormal (línea continua), DTKE Champernowne (línea discontinua) y DTKE Weibull (línea punteada)

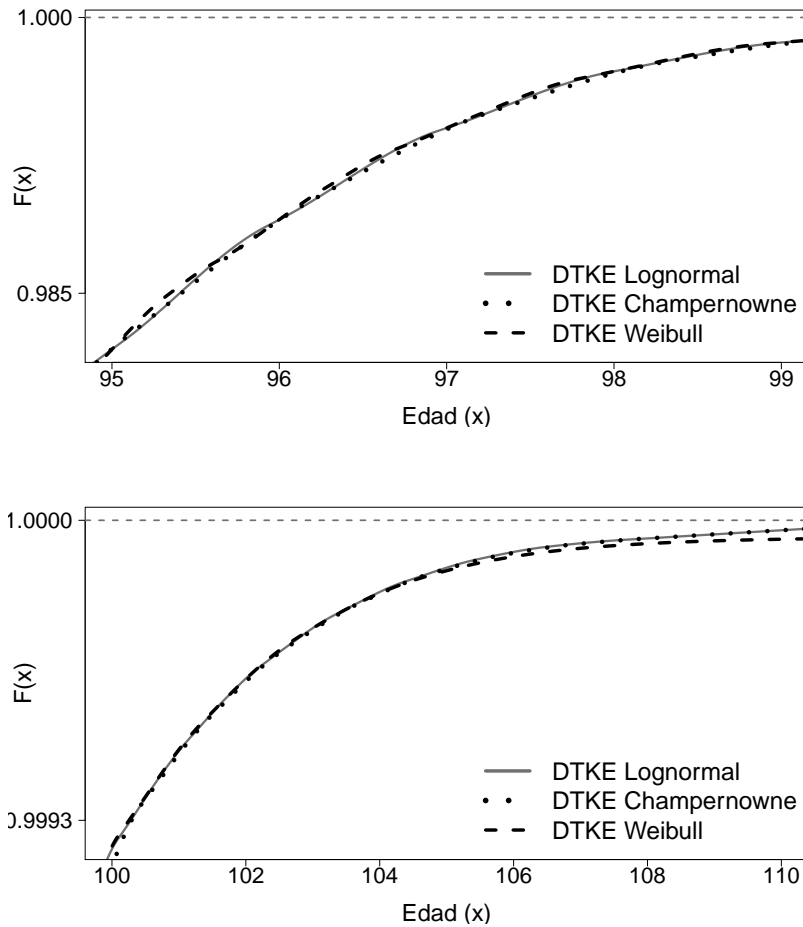


Figura 5.8: Función de distribución del tiempo de vida para mujeres. DTKE Lognormal (línea continua), DTKE Champernowne (línea discontinua) y DTKE Weibull (línea punteada)

5.3. Análisis de datos

La metodología descrita en la sección 5.2 se ha utilizado para estimar el riesgo de longevidad de la población asegurada mexicana. Como información primaria se ha utilizado la tabla de mortalidad publicada en 2010, ver la sección 1.4.1.

Dado que nuestro interés principal es evaluar el riesgo de longevidad y que las tasas de mortalidad están disponibles sólo hasta los 100 años, se ha supuesto que los supervivientes a la edad 100 morirán de manera aleatoria hasta que desaparezca toda la población. La distribución uniforme se ha utilizado para decidir si un sujeto muere o sobrevive. Con este supuesto la última edad

alcanzada para los hombres es 107 años y 112 años para las mujeres.

Con el fin de obtener el número de muertes en cada edad, las tasas de mortalidad se aplicaron a los supervivientes en cada edad x . El número de supervivientes a la edad de $x + 1$ se calculó como la diferencia entre los supervivientes a la edad x y el número de muertes en edad x . Repetimos este proceso hasta la última edad. Este procedimiento se realizó de forma independiente para cada género, suponiendo una población inicial de $N = 5000$ individuos de edad 65 años. El tamaño de la muestra es igual a la suma de los supervivientes en todas las edades.

La figuras 5.9 y 5.10 muestran el histograma de la variable aleatoria simulada X “duración de la vida” para hombres y mujeres, respectivamente. Para observar la densidad en las edades extremas el histograma se ha dividido en dos intervalos de edad, el correspondiente a $[65; 99]$ y el correspondiente a los individuos con edad > 99 . Como era de esperar la forma del histograma es asimétrica a la derecha y la probabilidad de alcanzar una edad extrema es pequeña pero no es cero.

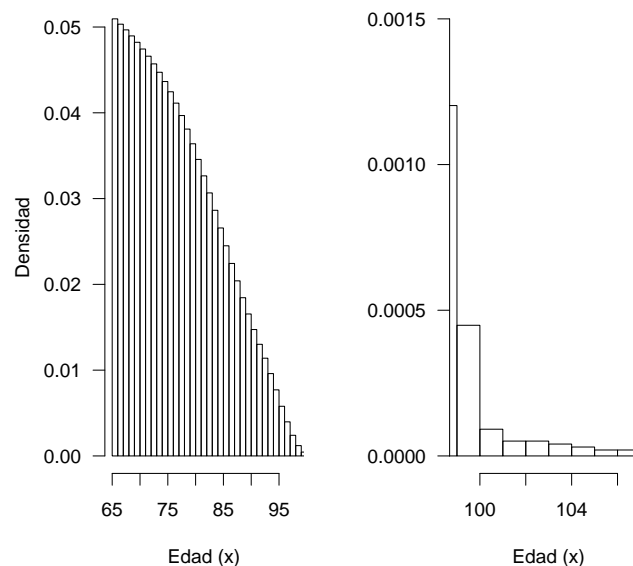


Figura 5.9: Histograma del número de supervivientes hombres

Como alternativa a la estimación máximo verosímil, en la sección 5.2 se describió el método basado en la minimización de \widehat{MISE} definido en la ecuación (5.5). La tabla 5.2 muestra los parámetros estimados utilizando ambos

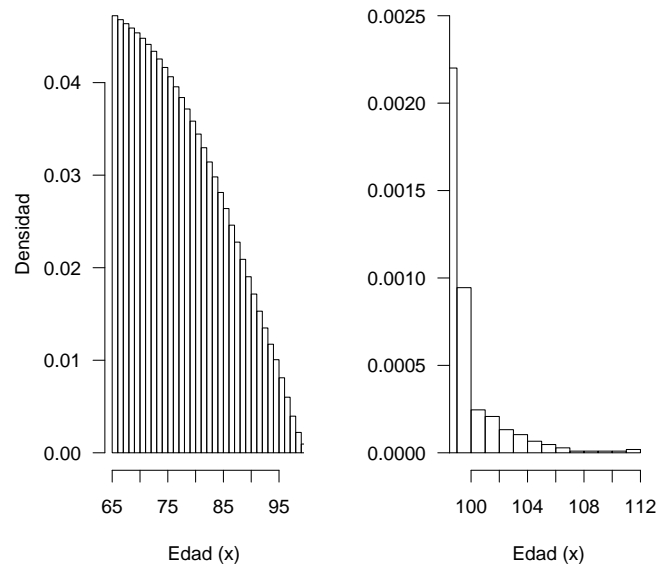


Figura 5.10: Histograma del número de supervivientes mujeres

métodos, para las distribuciones Weibull, Lognormal y Champernowne. Hemos observado que para la distribución Lognormal y la distribución Weibull los parámetros estimados por ambas metodologías son similares. Para la distribución de Champernowne, los parámetros estimados minimizando \widehat{MISE} son inferiores a los estimados por máximo verosímil; esto se produce porque la distribución de Champernowne tiene una cola derecha más pesada. En la última columna de la tabla 5.2 se presentan los valores de \widehat{MISE} para cada distribución paramétrica estimada. Se concluye que la distribución que mejor se ajusta el comportamiento de la variable tiempo de vida es la Lognormal, seguido de cerca por la Champernowne.

		MLE		Mínimo \widehat{MISE}		\widehat{MISE}
		Forma	Escala	Forma	Escala	
Hombres	Weibull	9.70	79.89	9.78	78.29	149.88
	Lognormal	0.10	4.33	0.11	4.32	37.69
	Champernowne	16.48	75.00	14.44	74.83	48.27
Mujeres	Weibull	9.49	80.72	9.36	79.19	151.28
	Lognormal	0.10	4.33	0.12	4.32	38.39
	Champernowne	15.83	76.00	13.82	75.53	51.84

Tabla 5.2: Parámetros estimados por MLE y minimizando \widehat{MISE}

La finalidad de estimar la función de distribución es calcular el VaR_α definido en la ecuación (5.2). Se ha estimado el VaR condicionado a las edades de 65 y 85 años. Los resultados obtenidos se muestran en las figuras 5.11, 5.12, 5.13 y 5.14 donde se representa el VaR con diferentes niveles de confianza, para hombres y mujeres, utilizando los ajustes paramétricos, noparamétricos y semiparamétricos. Cuando usamos los métodos noparamétricos y semiparamétricos, las curvas del VaR estimado están más cerca del VaR empírico, aunque existen diferencias entre la estimación CKE, DTKE Lognormal y DTKE Champernowne³. En general, el mayor riesgo se calcula con DTKE Champernowne. Es evidente que los modelos paramétricos proporcionan una estimación sesgada del VaR en todos los niveles de confianza, sobrestimando el riesgo de longevidad.

En la tabla 5.3 se muestra el riesgo de longevidad estimada en meses utilizando el CKE, el DTKE Lognormal y el DTKE Champernowne. Hemos estimado el VaR con tres nivel de confianza: 99 %, 99.5 % y 99.9 %, para tres distribuciones truncadas: a los 65, 75 y 85 años de edad⁴.

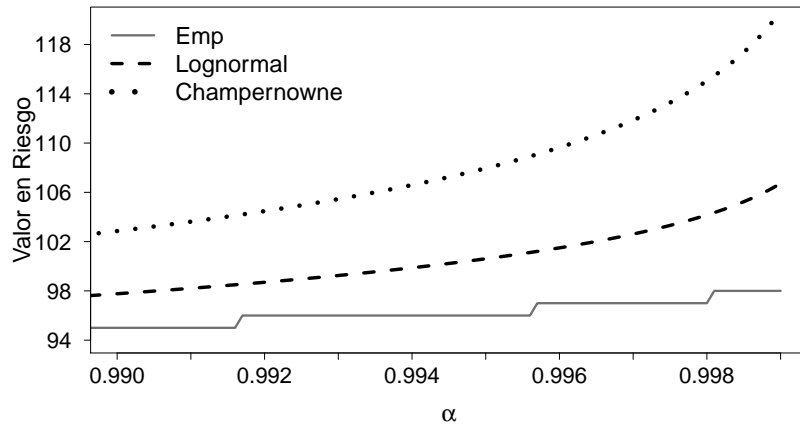
α		Hombres			Mujeres		
		65	75	85	65	75	85
99 %	CKE	362.72	256.37	153.97	373.96	266.34	162.72
	DTKE Lognormal	362.90	256.51	153.66	374.30	266.49	162.85
	DTKE Champernowne	362.88	256.70	154.49	374.29	266.50	163.56
99.5 %	CKE	376.41	267.76	163.95	387.29	277.77	175.33
	DTKE Lognormal	376.56	267.81	163.10	387.51	277.84	174.97
	DTKE Champernowne	376.76	268.00	163.93	387.65	277.86	175.98
99.9 %	CKE	400.68	289.99	196.74	414.11	312.27	224.77
	DTKE Lognormal	400.37	288.98	198.46	413.85	312.41	224.36
	DTKE Champernowne	400.78	289.55	198.48	414.69	312.56	224.71

Tabla 5.3: VaR_α : Número de meses estimado a sobrevivir después de la edad x

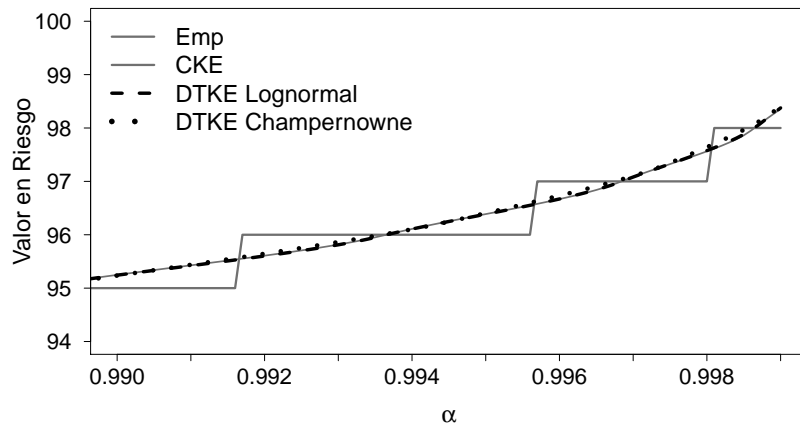
Para reforzar los resultados obtenidos, se han estimado los errores estándares utilizando el método de remuestreo “bootstrap”. Se extrajeron 500 muestras con reemplazamiento para cada una de las cuales se estimó el VaR con los diferentes niveles de confianza. En la tabla 5.4 se presenta la media de los VaR calculados en las 500 muestras bootstrap, como era de esperar estos resultados

³Denominamos DTKE Lognormal y DTKE Champernowne a las DTKE utilizando como primera transformación (Paso 1) las funciones de distribución de la Lognormal y de la Champernowne

⁴Al final del capítulo se muestra la tabla con el riesgo estimado en años para todos los métodos analizados



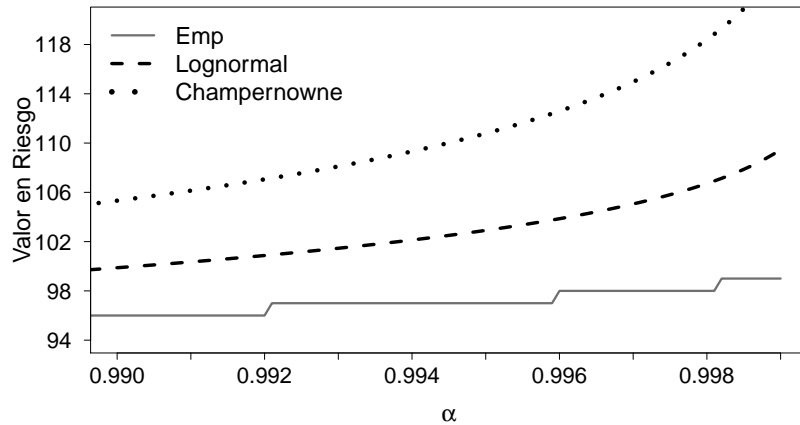
(a) Edad 65 Paramétrico



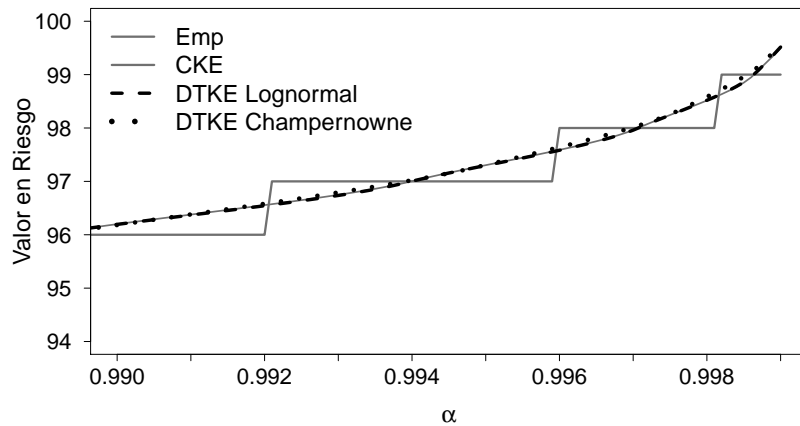
(b) Edad 65 Noparamétrico

Figura 5.11: VaR_α estimado en hombres una vez se ha alcanzado la edad 65

son similares a los presentados en la tabla 5.3. En la tabla 5.4 se presenta el cociente entre la desviación estándar del VaR estimado utilizando el DTKE Lognormal y el DTKE Champernowne versus la desviación estándar del VaR estimado por el CKE. En la mayoría de los casos los cocientes son muy próximos a uno, sin embargo, existen dos casos en los que se pone de manifiesto la ineficiencia del estimador CKE, concretamente en el cálculo del $VaR_{99.9}$ para los hombres con 85 años. Dado que el resultado del cociente es 0.66 y 0.63



(a) Edad 65 Paramétrico



(b) Edad 65 Noparamétrico

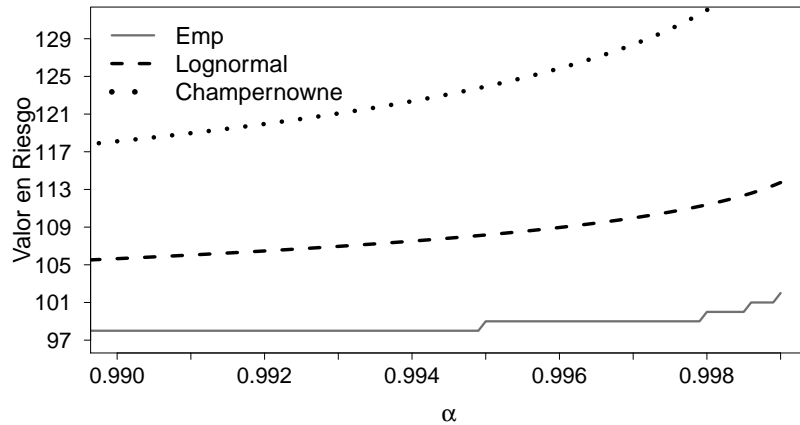
Figura 5.12: VaR_α estimado en mujeres una vez se ha alcanzado la edad 65

α		Hombres			Mujeres		
		65	75	85	65	75	85
99 %	CKE	362.93	256.51	153.37	374.33	266.50	162.78
	DTKE Lognormal	362.85	256.46	153.59	374.24	266.42	162.78
	DTKE Champernowne	362.83	256.64	154.39	374.23	266.43	163.47
99.5 %	CKE	376.60	267.73	162.58	387.58	277.78	174.68
	DTKE Lognormal	376.51	267.73	162.92	387.45	277.74	174.82
	DTKE Champernowne	376.70	267.90	163.71	387.59	277.79	175.74
99.9 %	CKE	400.12	288.48	195.08	413.76	311.62	222.34
	DTKE Lognormal	400.19	288.62	195.48	413.69	311.60	222.55
	DTKE Champernowne	400.57	289.35	195.69	414.48	311.80	222.88

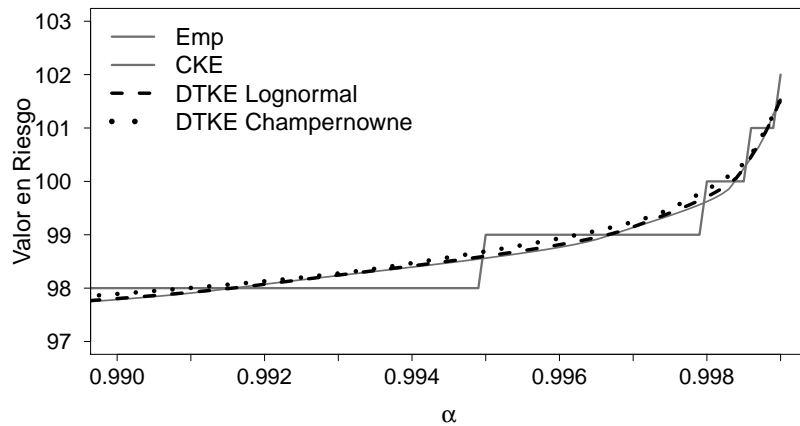
Tabla 5.4: Media del VaR_α calculado a partir de las 500 muestras simuladas

5.4. Conclusiones

En este capítulo se modelizó la variable duración de la vida siguiendo diferentes aproximaciones: métodos paramétricos y noparamétricos. Una de las



(a) Edad 85 Paramétrico



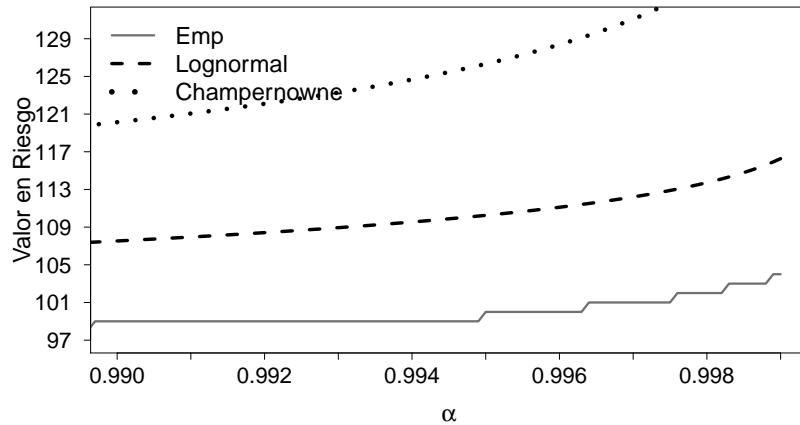
(b) Edad 85 Noparamétrico

Figura 5.13: VaR_α estimado en hombres una vez se ha alcanzado la edad 85

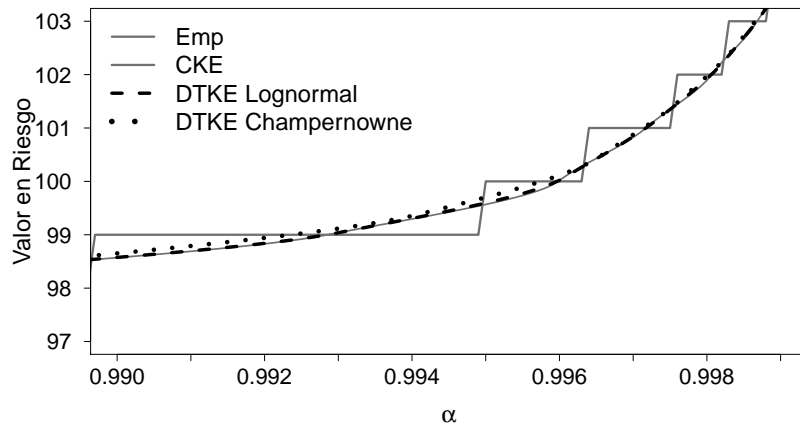
α		Hombres			Mujeres		
		65	75	85	65	75	85
99 %	DTKE Lognormal	0.98	1.00	0.96	0.98	0.98	1.01
	DTKE Champernowne	1.04	1.09	0.91	1.02	1.05	1.10
99.5 %	DTKE Lognormal	0.99	1.00	1.03	0.97	0.98	1.02
	DTKE Champernowne	1.07	1.11	1.12	1.05	0.99	1.09
99.9 %	DTKE Lognormal	1.02	1.00	0.66	1.00	1.00	0.99
	DTKE Champernowne	1.13	0.88	0.63	1.12	0.99	0.97

Tabla 5.5: Cociente de la desviación estándar del VaR_α calculado por la estimación DTKE y CKE en las 500 muestras simuladas

principales aportaciones del capítulo es la utilización de un método basado en la minimización de \widehat{MISE} que nos ha permitido comparar diferentes ajustes independientemente del método de estimación. Además, este criterio se ha



(a) Edad 85 Paramétrico



(b) Edad 85 Noparamétrico

Figura 5.14: VaR_α estimado en mujeres una vez se ha alcanzado la edad 85

utilizado en la estimación de los parámetros de las distribuciones Weibull, Lognormal y Champernowne. Los parámetros óptimos se encontraron fácilmente, además se ha observado que \widehat{MISE} converge rápidamente a un mínimo global cuando los parámetros de máxima verosimilitud se utilizan como valores iniciales.

La ventaja de utilizar \widehat{MISE} como criterio para la estimación de los parámetros radica en la mejora de la estimación en la zona de la distribución con menos información muestral. Concretamente para el caso de que la variable sea asimétrica a la derecha este método mejora la estimación en la cola de la distribución.

En general se han observado tres resultados fundamentales

- Las distribuciones paramétricas empleadas sobrestiman el riesgo de longevidad.
- Los ajustes noparamétricos como la distribución empírica y la estimación núcleo clásica son ineficientes en algunos casos
- La estimación doble núcleo transformada que proponemos es la mejor alternativa para la estimación del riesgo de longevidad. Esta proporciona unos resultados asintóticamente insesgados y mejora la eficiencia de los métodos noparamétricos.

Además los resultados de este capítulo también pueden utilizarse para calcular la función de supervivencia y en consecuencia las tasas de mortalidad.

α		Hombres			Mujeres		
		65	75	85	65	75	85
99 %	Emp	95.000	96.000	98.000	96.000	97.000	99.000
	CKE	95.248	96.380	97.785	96.199	97.214	98.571
	DTKE Weibull	95.226	96.364	97.831	96.163	97.195	98.560
	DTKE Lognormal	95.241	96.376	97.805	96.191	97.208	98.571
	DTKE Champernowne	95.240	96.391	97.874	96.190	97.208	98.630
	Weibull \widehat{MISE}	91.512	92.769	95.291	93.222	94.450	96.787
	Lognormal \widehat{MISE}	97.767	100.661	105.651	99.879	102.705	107.529
	Champernowne \widehat{MISE}	102.858	108.079	118.110	105.325	110.397	120.138
	Weibull MLE	93.509	94.588	96.783	94.818	95.850	97.900
	Lognormal MLE	96.151	98.360	103.101	97.836	99.960	104.525
Champernowne MLE	99.115	103.404	113.210	101.592	105.504	114.817	
99.5 %	Emp	96.000	97.000	99.000	97.000	98.000	100.000
	CKE	96.387	97.318	98.562	97.304	98.158	99.568
	DTKE Weibull	96.368	97.313	98.663	97.274	98.147	99.611
	DTKE Lognormal	96.380	97.317	98.592	97.292	98.153	99.580
	DTKE Champernowne	96.397	97.333	98.660	97.305	98.155	99.665
	Weibull \widehat{MISE}	92.833	93.949	96.236	94.628	95.719	97.833
	Lognormal \widehat{MISE}	100.608	103.364	108.167	102.913	105.608	110.250
	Champernowne \widehat{MISE}	107.953	113.411	123.922	110.782	116.096	126.324
	Weibull MLE	94.870	95.828	97.810	96.230	97.146	98.994
	Lognormal MLE	98.635	100.733	105.282	100.457	102.475	106.854
Champernowne MLE	103.404	107.862	118.077	106.172	110.245	119.961	
99.9 %	Emp	98.000	99.000	102.000	99.000	101.000	104.000
	CKE	98.358	99.077	101.541	99.494	101.043	103.650
	DTKE Weibull	98.390	99.166	101.395	99.509	101.023	103.731
	DTKE Lognormal	98.364	99.081	101.538	99.488	101.034	103.696
	DTKE Champernowne	98.398	99.129	101.540	99.557	101.046	103.726
	Weibull \widehat{MISE}	95.384	96.275	98.159	97.346	98.217	99.953
	Lognormal \widehat{MISE}	106.727	109.256	113.733	109.462	111.939	116.270
	Champernowne \widehat{MISE}	120.712	126.798	138.536	124.499	130.453	141.931
	Weibull MLE	97.500	98.262	99.885	98.959	99.687	101.197
	Lognormal MLE	103.963	105.874	110.087	106.087	107.928	111.984
Champernowne MLE	114.039	118.941	130.192	117.562	122.058	132.801	

Tabla 5.6: VaR_α : Número de años estimados a sobrevivir según nivel de confianza y distribución estimada

6 El uso de tablas de mortalidad unisex en el sector asegurador Europeo

Este capítulo está basado en la publicación “La igualdad una exigencia para el seguro: problemas derivados de las tablas unisex”, que aparece en el libro *El seguro español ante los cambios del estado del bienestar*, publicado por la Fundación de Estudios Financieros, (ver, Guillén *et al.*, 2013)

6.1. El tribunal de justicia de la Unión Europea (UE) y el mercado asegurador

El Tribunal de Justicia de la Unión Europea (UE) en la directiva 2004/113/EC dictaminó la no discriminación por sexo en los bienes y servicios, afectando directamente a la tarificación en seguros. A partir de diciembre del 2012 las compañías aseguradoras han debido establecer un único precio a sus productos. Inicialmente, en diciembre de 2004 el Consejo de la Unión Europea, decretó que las compañías aseguradoras podían establecer costos diferentes en sus primas para cada género, cuando se demostrara que si existían diferencias en el riesgo contraído. Esta directiva toma en cuenta el hecho de que los derechos individuales son los mismos para los hombres y para las mujeres, lo cual se traslada directamente a las prestaciones y contribuciones de los individuos.

El principio de igualdad al cual apela esta normativa, ha supuesto un reto para las compañías aseguradoras. Desde 2011, han tenido que trabajar en una estrategia para afrontar el cambio de precios. En el ramo de seguro de vida es donde más problemas se han de solventar, pues la tarificación se hace por edad y por sexo. Quitar una de las dos variables base, implica un cambio profundo en los cálculos. Además, estas dos variables son fácilmente verificables, y son capaces de proveer una buena estimación del riesgo. Las tasas de mortalidad, en cualquier población, se expresan siempre separadas por estos dos factores.

La población femenina en España, presenta en todas las edades una menor probabilidad de fallecimiento que los hombres de su misma edad, comporta-

miento que ha sido observado en muchas generaciones, por lo tanto, hasta antes de la entrada en vigor de la decisión del alto tribunal comunitario, las tarifas en el seguro de vida para mujeres era menor que la de los hombres.

Es un hecho que la mortalidad de los hombres es diferente a la de las mujeres. Tratar de explicar esta diferencia empleando otros factores explicativos del comportamiento llevaría a crear nuevas formas de recopilación y verificación de datos. Sin embargo, este tipo de factores son difíciles de corroborar. Basar el coste del seguro en las costumbres o el estilos de vida de una población puede acarrear un coste superior al que se incurre excluyéndolos.

Este problema es fácilmente ampliable a los ramos del seguro que cubren riesgos personales, como son las rentas vitalicias y los seguros de salud. Para implementar la normativa, una solución sería el aumento de las primas, ya sea haciendo un promedio entre las primas de cada género o nivelando el precio hacía el seguro más costoso. De suceder este incremento del precio en los seguros de vida, las mujeres se verían afectadas. Los hombres en cambio se verían afectados en los cálculos derivados de productos que planteen la percepción de una renta vitalicia en la jubilación. Si la compañía aseguradora opta por hacer una media de las tarifas, entonces debería considerar el número de mujeres y de hombres que conforman su cartera, es decir, optar por ponderar las primas según la proporción de cada género. Sin embargo, la aseguradora también ha de considerar que el número de personas que contratan un seguro va cambiando con el tiempo, lo que la obligaría a realizar revisiones frecuentes de la proporción diferencial entre hombres y mujeres.

En la práctica, la definición de un precio fijo, no se produjo inmediatamente al entrar en vigor la normativa, se previó un cambio generalizado de tarifas, lo que implicó un cierto riesgo a que se diera un incremento de los precios medios del seguro de automóviles en 2013, y posteriormente se hicieran ajustes a las tarifas de acuerdo a la experiencia.

A largo plazo, en los seguros de vida es donde se espera que el cambio de tarificación cause más complicaciones al sector. El mejor escenario sería desvincular completamente la gestión de riesgos y cálculo de provisiones de la tarificación, por lo tanto, la medición del riesgo debe efectuarse utilizando la tabla de mortalidad de cada sexo por separado. Según la normativa, es posible diferenciar entre gestión y tarificación del riesgo, por lo que construir una tabla de mortalidad unisex, construida como combinación lineal entre la tabla de mortalidad masculina y femenina, puede ser útil para tarificar, pero

no debe sustituir a las tablas y proyecciones que se efectúan separando ambos colectivos.

La normativa afecta a todos los ramos del seguro, en el seguro de auto, por ejemplo, usualmente se ven implicados en la tarificación diversos factores, como la antigüedad del permiso de conducir, el tipo y modelo del automóvil, el uso que se le da al auto, la edad y el sexo del conductor, etc. En este caso, el sexo es un factor más en los tantos que se usan para establece la prima, aunque según la experiencia, el sexo por sí mismo, puede definir el estilo de conducción, la velocidad y las precauciones que se toman al volante. Este sector es uno de los más afectados, pues toda persona que tenga un coche está obligada a contratar un seguro de daños a terceros. La facilidad de cambiar de compañía de seguros provoca rotaciones altas entre estas, por lo que el marketing juegue un papel importante en la adquisición de seguro de auto, es posible una intensa guerra comercial de consecuencias imprevisibles si no se impiden los desequilibrios técnicos. Más, sabiendo que incluso antes de la implementación ya existía un comportamiento del sector donde ganar cuota de mercado sin atender el riesgo constituye una prioridad.

El género ha sido una variable básica en la gestión del riesgo porque hay evidencia estadística que verifica, por ejemplo, que la longevidad de las mujeres es mayor que la de los hombres y por lo tanto sus tasas anuales de mortalidad son menores, lo cual sucede no solo en España si no, en todas las poblaciones que se han estudiado. Por lo contrario, los hombres presentan mayores tasas invalidez que las mujeres y también presentan mayor cantidad de siniestros en el seguro de automóviles en su juventud. Algunos estudios indican que en promedio los hombres recorren más kilómetros al volante que las mujeres, y a mayor velocidad, por lo que su percepción del riesgo es menor que el de las mujeres, alimentado por un temperamento más agresivo al conducir. La situación laboral de hombres y mujeres también es distinta, hasta hace algunos años, existían de manera diferenciada labores propias de los hombres y de las mujeres. Los trabajos con mayor esfuerzo físico estaban ocupados primordialmente por hombres, como son la construcción, la minería, la policía, etc. Aumentando así las tasas de invalidez en la población masculina. Los hábitos y estilos de vida distintos entre sexos han sido evidenciados en numerosos estudios.

Algunas compañías de seguros mostraron inconformidad con la nueva normativa, alegaban el hecho de que crear tarifas unisex representaba un cambio estructural en la tarificación de nuevos productos y a su coste de implementa-

ción. Al Tribunal de Justicia le fue más importante mantener el principio de igualdad entre sexos y el fomento de la competencia entre aseguradoras. Argumentando que hay factores ajenos al sexo, que explican mejor la siniestralidad en los seguros y que por lo tanto deberían ser usados. En el seguro de vida, por ejemplo, interesaría conocer las situaciones de riesgo a la que los asegurados están expuestos: fumar, beber alcohol, consumo de drogas, nivel de actividad física, enfermedades padecidas, así como el estado de salud actual e incluso el nivel de renta. Teóricamente estos datos proveen información tanto o más relevante que el género.

6.2. La esperanza de vida y longevidad por sexos: impacto en la tarificación unisex

La mayor longevidad en las mujeres se ha verificado estadísticamente, diferentes disciplinas han querido explicar el porque las mujeres viven más tiempo que los hombres, se han encontrado motivos biológicos y sociológicos. Si se arguye a la parte biológica, se dice que las mujeres al tener mayores niveles de estrógenos sufren menos enfermedades cardiovasculares que los hombres, en cambio, para los hombres los niveles altos de testosterona se asocian a mayor riesgo de enfermedades cardiovasculares. También hay algunas corrientes que aluden a una mayor resistencia femenina que se perpetúa a través de la reproducción, que por ser más fuertes, no mueren jóvenes. Se aduce también que al ser el sexo femenino el garante del éxito de la reproducción dispone de mecanismos de mantenimiento y reparación, que la llevan a ser más longeva. La sociología aborda situaciones culturales y costumbres adoptadas por cada género. Los hombres, hasta hace algunos años, eran los que participaban en profesiones de peligro, o durante décadas el consumo de tabaco y/o alcohol u otras sustancias adictivas se asociaba principalmente a los hombres. Diferentes especialistas continúan en la búsqueda del por qué de estas diferencias de longevidad entre hombres y mujeres sin embargo no existe consenso. En España la proporción de mujeres que viven más de 100 años es sistemáticamente superior a la de hombres, en todas las generaciones para las que se dispone de datos. Por todo ello la distinción por sexos en la valoración de los riesgos en seguros de vida resulta evidente.

Los factores inmediatos para calcular la probabilidad de muerte de un indi-

viduo son la edad actual y del sexo, estos han sido utilizados por los actuarios para conocer el comportamiento de la mortalidad y calcular la duración estimada del tiempo de vida a partir de la suscripción de un compromiso asegurador. La esperanza de vida evidencia el tiempo medio en años que se espera sobreviva una persona y puede ser calculada al nacimiento o a cualquier edad, que es cuando concierne al ámbito actuarial ya que se desea conocer cuántos años vivirá un asegurado a partir del momento en que contrata el seguro. En España, los datos del INE en el año 2013, dicen que la esperanza de vida al nacer es de 85.59 años y de 79.97 años, para mujeres y hombres, respectivamente.

En las aseguradoras tienen como herramienta básica a las tablas de vida o tablas de mortalidad, con ellas pueden llevar a cabo todos los cálculos necesarios para calcular el riesgo y evidenciarlo finalmente en la tarificación. Estas tasas no sólo reflejan un momento puntual de la mortalidad en una población, sino que con ellas se puede inferir el comportamiento demográfico futuro y por supuesto calcular la esperanza de vida.

En este respecto, algunos estudios han mostrado que las diferencias en la esperanza de vida entre hombres y mujeres se ha ido reduciendo en los últimos años, la normativa europea ha sustentado con esto la conveniencia de eliminar la discriminación por sexos. En España por ejemplo en 2009 la diferencia era de 6.36 años, en 2012 esta diferencia se ha reducido a 5.80, esta comparación se refiere a la esperanza de vida contada a partir del nacimiento. Pero si se examina la esperanza de vida a los 65 años, en 2005, cuando respectivamente las esperanzas de vida a partir de los 65 años eran de 21.14 años para las mujeres y de 17.19 años para los hombres. la misma diferencia que se observaba para el año 2010 se esperaba que las mujeres vivieran otros 22.29 y los hombres 18.33, una diferencia de aproximadamente cuatro años. Las variaciones en las diferencias entre hombres y mujeres no han sido notorias y por lo tanto en la población española, las evidencias para edades elevadas no son en absoluto concluyentes sobre que se produzca una convergencia en la esperanza de vida de hombres y mujeres. Calculando la probabilidad de muerte a partir de los 65 años, las mujeres se mantienen por debajo de los hombres, haciéndose más estrecha la diferencia para la edad de 90 años y a partir de ésta.

Una mejora en la esperanza de vida se haría efectiva si las condiciones de vida de la población sufriera una mejoría. Según Vaupel *et al.* (2011) hace falta disminuir las muertes a edades prematuras, aumentando la cultura de prevención y teniendo políticas sanitarias adecuadas. Observando a los países

con mayor desarrollo y calidad de vida, su población ha mejorado su esperanza de vida de forma más significativa, que en los países con condiciones precarias.

Reflejando esta situación en los seguros, se puede justificar que cuando se cubre el riesgo de fallecimiento las mujeres han venido teniendo unos precios más ventajosos y han pagado menos que los hombres. Por el contrario, dado que su esperanza de vida es mayor que la los hombres; para los productos de renta donde se contrata una prestación de supervivencia, al vivir más tiempo las mujeres deben satisfacer primas más caras. Además, si una mujer contrata una renta vitalicia, recibirán mensualmente menos dinero que los hombres por una misma prima inicial. Las mujeres, al vivir más, reciben el dinero en más pagas: obtienen la misma cantidad en valor actuarial que los hombres, pero nominalmente en cada mensualidad perciben menos.

Aplicar una tarifa unisex, lleva a un sesgo en el cálculo actuarial, lo que implica que mujeres y hombres percibirán la misma renta mensual vitalicia a partir de una idéntica prima inicial, pero las mujeres lo harán durante más tiempo, en promedio. La valoración del riesgo y el cálculo de provisiones se verán afectados por el desequilibrio entre el compromiso contraído con el asegurado y el costo que se le ha impuesto al riesgo, una tabla unisex no servirá para conocer el riesgo verdadero. A nivel monetario, las mujeres percibirán más dinero de lo que percibirían en un entorno no unisex. Algunos autores argumentan que las mujeres tienden a disponer de menores rentas por el trabajo y a contribuir menos que los hombres a planes de pensiones de carácter voluntario (Curry y O'Connell, 2004).

La forma de las curvas de mortalidad para hombres y mujeres españoles, mantiene el mismo comportamiento en las últimas décadas. La esperanza de vida en mujeres es mayor, pero hay un intervalo de edades donde, para ambos sexos, se presenta el mayor número de muertes, es decir en la vida adulta las tasas de mortalidad son similares. Hudson (2007) argumenta que decir que una mujer tiene mayor esperanza de vida que un hombre no implica que todas las mujeres vivirán más que los hombres. Quienes están a favor de las tablas unisex, dicen que bajo este principio, las mujeres están en desventaja cuando reciben sus rentas vitalicias, pues es posible que las mujeres perciban una renta vitalicia menor por más tiempo, mientras un hombre reciba una renta vitalicia mayor por menos tiempo. Los contrarios a las tablas unisex rebaten este argumento diciendo que sí hay una edad en donde la mortalidad es la más alta para ambos sexos, la compañía de seguros corre siempre el riesgo de seguir

pagando la renta mensual a las mujeres por más tiempo.

Los argumentos que se han emitido a favor y en contra de las rentas vitalicias unisex no han sido cuantificados para mostrar en qué medida benefician a las mujeres y los consumidores en general. Esta cuantificación, sin embargo, es crucial. Al poner en marcha las tablas unisex las compañías incrementan sus provisiones en tanto se revalúan los riesgos y compromisos con la nueva tarificación.

6.3. La tarificación basada en factores de riesgo

El sexo y fecha de nacimiento constituyen el punto de partida clásico en la suscripción de una póliza de vida. Como se ha mencionado anteriormente, la suscripción basada en factores de riesgo apunta a ser el futuro en el establecimiento de las primas. Los aseguradores tendrán la necesidad de recopilar información de todos sus asegurados, la manera más eficiente de hacerlo será la construcción de cuestionarios que se deberán aplicar a los nuevos y actuales asegurados de forma sistemática y no puntualmente como se venía haciendo hasta ahora para coberturas elevadas. La información obtenida debe reflejar el riesgo que contrae la compañía de seguros al cubrir a un individuo en cualquiera de sus productos. Además para llegar a una tarificación correcta, se ha de tener suficiente experiencia estadística para conocer el riesgo real al que se enfrentan con el uso de tarifas unisex.

Este nuevo procedimiento de recogida almacenamiento y tratado de datos puede conducir a gastos de operación no considerados actualmente. Este hecho reafirma que la primera reacción de las aseguradoras sea elevar los precios, hasta encontrar un equilibrio entre el costo de la prima y los gastos asociados a la gestión del seguro.

Un inconveniente de recurrir al uso de cuestionarios para conocer con detalle los hábitos del asegurado es que puede llevar a información inexacta y fácilmente manipulable. El asegurado puede falsear u omitir información relevante, La longitud y tipo de pregunta que se haga influirá directamente en la calidad de información. De todas las preguntas que se realicen, sólo un porcentaje de ellas servirán íntegramente para encontrar factores de riesgo, pues algunas cuestiones no serán fácilmente verificables por parte de la aseguradora y por

lo tanto deberían excluirse a priori de la tarificación.

A los seguros donde el precio se obtiene según diversas características del asegurado se les llama seguros preferentes (Warth, 2011), dichos productos tienen una mayor complicación técnica y pueden ser una buena alternativa en el caso de productos de vida-riesgo, aunque pueden no serlo para establecer precio en pensiones, esto debido a la dificultad de extrapolación a futuro de las tablas de mortalidad que contemplen un excesivo número de componentes de singularidad.

En general, para realizar una buena suscripción es conocer el nivel educativo, la renta familiar, la profesión, los hábitos de alimentación y toda la información individualizada del asegurado llevará mejores resultados. Además conocer al total de su entorno como puede ser su familia o la localidad donde vive sería de gran utilidad. Por ejemplo, en familias de alto nivel económico, todo su núcleo se ve beneficiado por sus ingresos, incluyendo a todos los miembros de la familia. Sin embargo si un miembro de la familia es quien tiene la mayor carga de ingresos, no se debería considerar únicamente este ingreso en el costo de su seguro, si no que el factor ingreso se debe medir como la renta total familiar ajustada por el número de miembros. Un mejor predictor del nivel de vida y, por lo tanto, de la tasa de mortalidad es el nivel educativo, este factor es más usado en la suscripción del riesgo de mortalidad, que el nivel socio-económico.

Una manera de tener información para evaluar el estado de salud del asegurado es pedir que se realicen chequeos médicos o análisis sanguíneos cuando se realiza la suscripción del contrato de seguro. Asimismo, se deben tener cuenta los antecedentes de salud de sus familiares. Estas dos fuentes de información son certeras, pues en el primer caso se tiene un reflejo real de la salud del asegurado y en el segundo caso se sabe que en estas cuestiones el asegurado no suele mentir. Todo esto es útil pues la probabilidad de fallecimiento se ve aumentada cuando se detectan enfermedades crónicas como pueden ser la hipertensión arterial y/o la diabetes. Estas enfermedades son generalmente derivadas de una predisposición genética. Otras, como los problemas cardiovasculares se ven influenciadas por el sedentarismo o una mala alimentación.

Toda la investigación que se realice para conocer a los asegurados se hace para poder discriminar entre tipos de clientes y así, clasificarlos y agruparlos en una categoría de riesgo según sus características. De esta manera la aseguradora conocerá en todo momento el riesgo que asume cada que un asegurado nuevo llegue a su cartera. Llegar a tener un método de decisión que refleje de

manera eficiente y efectiva las diferentes clases de riesgo implica un trabajo exhaustivo para el desarrollo de dichas herramientas.

La definición de los criterios de inclusión que digan si un asegurado pertenece a un grupo de riesgo alto o un grupo de riesgo bajo es una tarea ardua. Se ha de decidir los puntos de corte para cada factor de riesgo que se desee incluir, es decir, encontrar el punto donde el riesgo se convierte en sobre riesgo y deja de ser un riesgo normal. En lo que respecta a las variables médicas, el recurrir a criterios médicos siempre será la mejor solución. Para los factores sociales se ha de recurrir a la experiencia en diferentes poblaciones hasta que la propia aseguradora tenga una base de datos que refleje lo observado en sus pólizas.

Ser fumador, en algunos países desarrollados, es un factor de discriminación para calcular el coste de los seguros. En el Reino Unido, por ejemplo, se han podido construir bases de datos específicas para colectivos con características específicas como son los fumadores. De esta manera han podido distinguir cualidades de ambas poblaciones: fumadora y no fumadora. La normativa aprobada en 2011 se debe aplicar en todos los países pertenecientes a la Unión Europea, el intercambiar información entre países puede ser de gran ayuda aunque se tiene que tener en cuenta que sus características socio-económicas no son exactamente iguales. También es posible que el impacto de la nueva normativa sea diferente en cada país, dependerá de su experiencia y de las características de su mercado.

Al crear grupos de asegurados con características específicas, se espera que su mortalidad sea diferente al de la población asegurada en general y al de la población total. Por ejemplo si un factor de inclusión en el grupo selecto es el hacer deporte, entonces se puede esperar menor mortalidad en este grupo que en el grupo donde se ignora esta información.

La base de datos que se usará como fuente de información para construir la tabla de mortalidad debe estar siempre actualizada. También se han de establecer criterios sobre el periodo en que la información contenida es válida. Esto porque el riesgo de selección está basado, en parte, en características que no son estáticas, es decir, son situaciones de vida más que características intrínsecas de los asegurados. Por ejemplo, el estado de salud, fumar, hacer ejercicio, sufrir depresión, conducir muchas horas al día, etc. son comportamientos que pueden cambiar en cualquier momento. Entonces la aseguradora debe tener en cuenta un margen de tiempo durante el cual se supondrá que el asegurado continua en mismo estado que cuando entró al grupo. A este periodo se le

llama duración de la selección.

La experiencia y la buena gestión de la información, hará que cualquier factor que se considere en la suscripción, con el tiempo, dejará de ser un riesgo añadido en la suscripción. La inversión que realice la compañía de seguros en la creación de la nueva tarificación irá a la baja al pasar los años.

Dependiendo del producto que la compañía ofrezca puede pedir o no al asegurado que cumpla con ciertos requisitos, como los exámenes médicos o responder formularios extendidos con preguntas específicas sobre sus hábitos. Finalmente con la información recopilada se podrán calcular las llamadas tablas seleccionadas de mortalidad, que se definen como la experiencia estadística en mortalidad de un conjunto de individuos que cumple una serie de requisitos o ciertas características y sólo podrá ingresar al grupo el asegurado que no presente ningún agravante del riesgo. En todas los ramos de seguro a mayor riesgo, mayor prima. El costo está directamente relacionado con la probabilidad que tenga el contratante de sufrir un siniestro.

6.4. Las tablas de mortalidad unisex en los sistemas de pensiones

Las tablas unisex se han usado en algunos países cuando la renta vitalicia o pensión son fruto de una relación laboral. Es decir, si se trata de una prestación social implícita en el contrato laboral, ya que las pensiones en este caso reflejan una relación de trabajo. Ejemplo de estos sistemas son el Reino Unido, Suiza, Canadá y Estados Unidos. Además el uso de las tablas unisex es obligatorio siempre que se trate de este tipo de pensiones.

En el Reino Unido, actualmente hay productos de seguros personalizados, que han sido fruto de las exigencias del mercado. Las compañías aseguradoras han creado tarifas que dependen de alguna característica particular del asegurado, de esta manera se logra que la esperanza de vida sea diferente a la de la mayoría de la población. Es posible encontrar planes que toman en cuenta la situación económica, la ocupación o profesión, el estado de salud, los estilos de vida como son el ejercicio físico periódico e incluso la situación geográfica.

El sexo deja de ser el principal factor, y se centran en el individuo y sus características. Además las características escogidas, de manera indirecta pueden reflejar el género del contratante. Este tipo de producto se ofrece sólo a

quien cumple con determinados requisitos por lo que es factible cobrar tarifas absolutamente personalizadas al riesgo que se contraiga. Este tipo de seguros son los que se espera que acaben siendo los habituales en la EU a medio plazo.

En el Reino Unido, un problema al que se han de enfrentar es que las mujeres representan solamente el 25 % de los compradores de seguros de vida. Crear tablas unisex basadas en una igualdad de proporción de hombres y mujeres no reflejaría la situación actual del mercado. También se ha propuesto tener rentas pagaderas en un intervalo de tiempo fijo, Limited period annuities (LPA), de esta manera hombres y mujeres recibirán una misma cantidad de dinero en un mismo periodo de tiempo, dejando atrás la influencia de la esperanza de vida en el pago mensual recibido.

En Estados Unidos los sistemas de pensiones se basan en una combinación de fondos estatales y contribuciones voluntarias a través de planes de pensiones de empresa o bien incentivos fiscales al ahorro individual a la jubilación. Al finalizar la vida laboral no es obligatorio pasar a un sistema de rentas vitalicias cuando el empleador contribuyó a formar el fondo, pero en caso de optar por rentas vitalicias, el cálculo de la renta mensual que recibirá el individuo se hace con tablas unisex, evitando así la discriminación por sexo según manda la corte suprema.

En EE.UU. y Canadá, entre otros países, las pensiones previstas en los sistemas de reparto no discriminan por género. Por ejemplo, en el Reino Unido, la sentencia del Tribunal de Justicia Europeo ha establecido que las prestaciones de jubilación, como tales, forman parte de las condiciones de empleo y remuneración de una persona y por lo tanto, la remuneración debe ser idéntica al realizar un mismo trabajo, incluida la percepción de la pensión, esto se ha asentado en el artículo 119 del Tratado de Roma. Las pensiones de empleo recibidas por hombres y mujeres en los regímenes de reparto, tras el fallo en 1990, han tenido que ser iguales. A la luz de lo anterior, parece ser una anomalía que en este tipo de planes la discriminación haya seguido siendo aceptable en los planes de contribución definida (Hudson, 2007).

El tema de la desigualdad entre hombres y mujeres resurge de todos modos en el ámbito de los planes de pensiones de empleo, ya que en muchos casos, los fondos utilizados para comprar una renta vitalicia se habrán constituido por contribuciones del empleador en los regímenes de cotizaciones definidas. En estos casos, una mujer, de hecho, acaba recibiendo una pensión de jubilación más baja que un hombre habiendo trabajado los mismos años porque su salario,

es casi siempre, más bajo.

El excelente estudio de González Rabanal y Sáenz de Jáuregui Sanz (2008) se centra en averiguar las consecuencias que se derivarían de la aplicación de tablas unisex en el supuesto de los planes de pensión de prestación definida no totalmente asegurados, es decir, aquéllos que otorgan a su beneficiario una prestación determinada a priori que puede consistir en una renta vitalicia cuya cuantía puede ser un porcentaje del salario que percibía el trabajador cuando estaba en activo. Dichos autores proponen el pago de una subvención al empresario que mitigue el mayor esfuerzo que debe hacer éste al contratar una mujer y tener que realizar una mayor aportación que en el caso de los varones para sufragar la misma pensión privada.

En los planes de aportación definida las diferencias entre sexos se hacen notar cuando el capital se convierte en renta, los beneficiarios mujeres reciben menos renta pero por más tiempo que los hombres. Si el cálculo de las pensiones por jubilación anticipada o diferida se realiza con tablas diferenciadas por sexo, es natural que resulten diferencias en la cuantía de la prestación.

Las aportaciones de los planes de pensiones son deducibles de impuestos, fiscalmente esto beneficia a quien tiene mayor renta mensual, y puede dar aportaciones mayores. Las mujeres además de vivir más años y recibir pensiones vitalicias inferiores, suelen contribuir menos, pues también tienen salarios más bajos.

En los planes de pensiones de reparto o prestación definida se predetermina la cuantía de todas las prestaciones a percibir por los beneficiarios. En función de las prestaciones y de las distintas contingencias a cubrir se calcularán las cuantías de las aportaciones. La definición de la prestación podrá realizarse en términos absolutos o en función de magnitudes determinables como salarios, antigüedad de la empresa, percepciones complementarias u otras variables susceptibles de servir de referencia [art. 16.b) RPFPP].

6.5. Algunas consecuencias del uso de tablas unisex en el seguro de vida

Un posible efecto de eliminar factores de riesgo de la tarificación es la selección adversa. En el caso del seguro de vida, por ejemplo, si la tarifa nueva resultara de calcular la media entre ambos sexos, los hombres se verían benefi-

ciados en una reducción de precios lo que sería atractivo para nuevos clientes, en cambio las mujeres sufrirían un aumento en la prima que pagan actualmente. Este efecto se puede ver en cualquier tipo de seguro, los asegurados con menos riesgo podrán no re-contratar el seguro si se ven afectados por un aumento en el precio, por el contrario los asegurados con riesgos “malos” se mantendrían en la cartera al verse beneficiados de la baja de precio, esto causaría un desequilibrio en la cartera de clientes. Además, los potenciales asegurados con riesgos altos, se verán atraídos por el bajo costo de las primas.

Algunos estudios presentan la posibilidad de que no tener en cuenta ciertos factores de riesgo puede aumentar la cantidad de pérdidas que se cubren (loss coverage). El porcentaje de aumento depende directamente de la proporción de asegurados con menor riesgo, pues son estos los que se espera que, en número, disminuyan al verse afectados por el costo del seguro. Sin embargo no se puede hablar de números absolutos si no de la proporción entre asegurados de bajo y alto riesgo, en este caso el ratio entre mujeres y hombres en la cartera de la aseguradora y del riesgo relativo que existe entre dichas categorías: ratio entre el riesgo de presentar el evento en mujeres y el riesgo de presentar el evento en hombres. A continuación se replican un par de ejemplos mostrados en Guillén *et al.* (2013).

Ejemplo 1 Incremento en las pérdidas cubiertas al eliminar la discriminación.

Numéricamente se ha verificado mediante simulaciones que efectivamente una disminución leve en el porcentaje de asegurados con menor riesgo, hace que la cantidad de pérdidas que se cubre aumenten, aunque si el descenso en dicho porcentaje es dramático se puede llegar a niveles aun más bajos de cuando se tenía en cuenta el factor de riesgo. Sin embargo en la bibliografía consultada no se menciona el monto de las pérdidas, es decir la severidad de los siniestros.

Se tienen 2000 personas en situación de riesgo, de estas el 80 % son mujeres. Se tienen entonces 1600 mujeres y 400 hombres, el riesgo de presentar un siniestro para cada grupo es de 0.1 y 0.3 respectivamente. Así se tiene que del total de la población 160 mujeres y 120 hombres son proclives a presentar el evento de riesgo, en total 280. Sin embargo no toda la población adquiere un seguro, ejemplificando con tres posibles escenarios (no únicos):

1. La proporción de mujeres aseguradas es del 50 % y de hombres el 25 %, siguiendo los porcentajes de riesgo, se tiene que en total 110 personas

con seguro presentarían un siniestro (80 mujeres y 30 hombres). Así la cantidad de pérdidas cubiertas es de 39 % , calculado como el cociente entre 110 y 280.

2. La proporción de mujeres baja al 40 % y la cantidad de hombres aumenta al 50 %, ahora el total de personas en riesgo es del 124 y el loss coverage es de 44 %.
3. La proporción de mujeres baja al 25 % y la cantidad de hombres se mantiene en 50 %, ahora el total de personas en riesgo es del 100 y el loss coverage es de 36 %.

Desde el punto de vista de la aseguradora, según Thomas (2007), la disyuntiva entre cual es el mejor camino a seguir es si a la aseguradora le importa vender pólizas o tener primas que reflejen realmente el riesgo contraído. En el ejemplo 1 es fácil ver que cuando se maximiza el loss coverage (caso 2) el número de asegurados baja respecto a respecto al caso 1, donde se tarifica diferenciando por sexos. Obviamente las aseguradoras prefieren vender pólizas a personas con menor riesgo de presentar reclamaciones. Ello se consigue creando productos con restricciones de venta o productos que cubren diversa cantidad de riesgos que hacen atractiva su compra, aun si no se está expuesto a tal riesgo.

Los peligros del Ejemplo 1 quedan patentes si se calculan las coberturas por sexos, puesto que en el caso 3 sólo una de cada cuatro mujeres que tuvieran un siniestro estarían cubiertas mientras que en el caso 1 eran la mitad.

Ejemplo 2 Decremento en las pérdidas cubiertas al eliminar la discriminación.

En este ejemplo se realiza sólo un pequeño cambio respecto al Ejemplo 1, pero entonces se verifica que el cociente de cobertura disminuye siempre sea cual sea el caso de los tres considerados.

Se tienen 2000 personas en situación de riesgo, de estas el 80 % son mujeres. Se tienen entonces 1600 mujeres y 400 hombres, el riesgo de presentar un siniestro para cada grupo es de 0.3 y 0.1 respectivamente. Así se tiene que del total de la población 480 mujeres y 40 hombres son proclives a presentar el evento de riesgo, en total 520. Sin embargo no toda la población adquiere un seguro, ejemplificando con tres posibles escenarios (no únicos):

1. La proporción de mujeres aseguradas es del 50 % y de hombres el 25 %, siguiendo los porcentajes de riesgo, se tiene que en total 250 personas con seguro presentarían un siniestro (240 mujeres y 10 hombres). Así la cantidad de perdidas cubiertas es de 48 %, calculado como el cociente entre 250 y 520.
2. La proporción de mujeres baja al 40 % y la cantidad de hombres aumenta al 50 %, ahora el total de personas en riesgo es del 212 y el loss coverage es de 41 %.
3. La proporción de mujeres baja al 25 % y la cantidad de hombres se mantiene en 50 %, ahora el total de personas en riesgo es del 140 y el loss coverage es de 27 %. Incluso si la proporción de mujeres aumenta del 50 % al 51 % una drástica disminución de los hombres asegurados pasando del 25 % al 1 % (aunque se habría ganado más de dos aseguradas por cada asegurado perdido, conduciría a una cantidad de pérdidas cubiertas del 47 %, lo cual es inferior al caso 1).

6.6. Conclusiones

Una vez expuestas las diferentes circunstancias posibles en la tarificación unisex, la principal recomendación es continuar recopilando toda la información relevante del asegurado incluido el sexo, aunque no pueda usarse en la tarificación. Más cuando la experiencia y la estadística indica que el sexo es un factor referente en el conocimiento del riesgo. Todas las características de los asegurados sin duda, tendrán influencia en el riesgo que contrae la aseguradora y serán de gran utilidad en la toma de decisiones dentro de la compañía.

Las bases de datos donde se contenga los resultados de exámenes médicos y factores socio-económicos han de tener una evaluación continuada, con la finalidad de mantener su veracidad y calidad. Guillén (2012) menciona seis dimensiones a verificar en las bases de datos: validez, integridad, exhaustividad, coherencia, precisión y oportunidad. La validez implica que los valores registrados estén dentro del rango de valores posibles. La integridad es la coherencia de la información, por ejemplo, un hombre no puede recurrir a servicios ginecológicos, la fecha de ocurrencia del siniestro no puede ser anterior a la contratación del seguro. La exhaustividad apela a que los datos se recojan para todos los asegurados, además de incentivar a complementar todos los campos que se les

requiera. Estas tres magnitudes son fácilmente verificables. La coherencia, se refiere a la duplicación y realización de comprobaciones cruzadas de sistemas. La precisión garantiza que los datos han sido verificados. La oportunidad implica que la base se actualizada constantemente. Las tres últimas dimensiones son el talón de Aquiles a la práctica.

Particularmente en la tarificación, las primas unisex deberían reflejar la composición de la cartera. Cuando la información es fiable y veraz, se pueden obtener diversos factores de tarificación y de valoración del riesgo de tanto o más valor que el sexo.

El trabajo en equipo de todos los departamentos de la aseguradora, como son marketing, ventas, y el técnico actuarial, llevarán a garantizar que la prima siga siendo proporcional al riesgo cubierto. Los actuarios deberán tener claro que la valoración del riesgo y las provisiones se distancian del precio final del seguro. Los modelos estadísticos-actuariales serán el instrumento más útil para la cuantificación de riesgos, donde la tarificación se ha de ajustar a una normativa y no necesariamente a la evidencia empírica.

Conclusiones

En la Tesis se ha trabajado con la mortalidad de la población general y la población de asegurados en México. Se ha estudiado el comportamiento de la mortalidad, para después centrar nuestra atención en la longevidad, definiéndola como la máxima duración posible de la vida. Se han estimado modelos de mortalidad estocásticos que son utilizados por diversos países, entidades de seguros y agencias reguladoras para analizar y proyectar la mortalidad. Se ha enfatizado que existe evidencia para decir que la mortalidad de las mujeres es diferente a la de los hombres, teniendo como consecuencia que las mujeres a sobrevivir más años. Se ha utilizado teoría estadística paramétrica y no-paramétrica. En el primer caso, se ha utilizado el análisis de supervivencia y el ajuste de diferentes distribuciones paramétricas. En el segundo caso, se han utilizado desde los modelos más básicos como son la distribución empírica y el estimador núcleo clásico hasta técnicas más elaboradas como el estimador núcleo doble transformado. En cada capítulo se ha añadido una sección con conclusiones, así que en este apartado se presentarán las aportaciones más relevantes de la Tesis.

Al realizar este trabajo uno de los problemas que se ha presentado ha sido la falta de información pública, en particular en lo que se refiere a los datos de la mortalidad en los seguros. Si bien desde el año 2006 la Comisión Nacional de Seguros y de Fianzas hace público el número de siniestros y número de asegurados para cada edad y sexo, mas en los primeros años se presentan deficiencias en la calidad de la información. Se espera que con el tiempo estos problemas se solventen. Afortunadamente se han podido conseguir tablas de mortalidad que si bien no incluyen información sobre todas las aseguradoras, si lo hace para las más importantes. El tener esta información ha permitido realizar los capítulos 4 y 5 de esta Tesis.

En lo que concierne a los datos de la mortalidad general, ha sido fácil acceder a las bases de datos con información del número de defunciones y número de personas vivas. Además, ha sido posible también acceder a las tasas de mortalidad y proyecciones estimadas por el INEGI, tasas que no se han utilizado porque se ha querido trabajar con las tasas brutas de mortalidad sin ninguna corrección de por medio. Se puede decir que a nivel país se tiene una buena fuente de información de lo que sucede en la demografía de México. Desde la

década de los noventa se puede obtener de manera detallada dicha información.

En diferentes capítulos de la Tesis se han realizado comparativos de la mortalidad general mexicana con la de los asegurados, en todos los casos se ha observado que existen diferencias entre ambas poblaciones y que estas son más o menos visibles dependiendo del género y la edad que se estudie. En el aspecto género, la principal diferencia en tasas de mortalidad ocurren en el intervalo de los 20 – 50 años, que son las edades con mayor número de contratantes de seguros, así que el uso de tablas de vida unisex en los seguros pone en riesgo la solvencia de la compañía si realiza su tarificación asumiendo igualdad de proporción de sexos.

Una posible solución es incrementar las tarifas, en el caso del seguro de vida, a las mujeres, esta elevación en las primas sería viable sólo por unos años, hasta poder conocer las consecuencias reales resultantes del uso de tarifas unisex. Sin embargo, se corre el riesgo de una bajada de las ventas en el grupo afectado por el incremento de las primas y, por tanto, las carteras acabarían estando formadas por asegurados que se ven beneficiadas por la bajada de las primas pero, por el contrario, suponen un mayor riesgo para las compañías. Con el tiempo esto conllevaría nuevos ajustes en los precios. Usar tablas de mortalidad unisex, resultantes de promediar la mortalidad femenina y masculina es una opción para obtener precios de seguros, pero no puede sustituir a las estimaciones individuales para cada género.

A pesar de que las regulaciones insisten en que los precios de los seguros deben ser iguales para hombres y mujeres, se concluye que el estudio de la mortalidad en las carteras tiene que ser independiente del precio final del seguro. Es decir, realizar tablas de mortalidad diferenciadas por sexo y que la tarificación tenga en cuenta la proporción de hombres y mujeres dentro del total de asegurados. Además de que, para evitar la selección adversa, las tasas de mortalidad de los asegurados se han de corregir por la mortalidad de la población general. En el caso de México, donde la falta de información en seguros es un problema, se sugiere el uso de mortalidad de otros países con condiciones similares de desarrollo.

Las diferencias entre mujeres y hombres se mantienen a lo largo de las diferentes edades, por lo que en el ámbito los seguros afecta tanto a planes de vida como a planes de pensiones, donde lo que importa es la probabilidad de llegar vivo a una edad avanzada. Así, el paso siguiente ha sido modelizar la longevidad. Existen diferentes teorías sobre el final de la vida, algunos autores creen

que el aumento en la esperanza de vida será cada vez menor, hasta alcanzar un límite, sin embargo este límite aun no es visible en algunas poblaciones.

Para el análisis de la longevidad a partir de la edad 65 se han ajustado diferentes distribuciones, siendo la Weibull la que proporcionó mejor ajuste. Para evaluar el riesgo de longevidad de los hombres y de las mujeres se ha utilizado el concepto del Valor en Riesgo (cuantil α de la función de distribución). Una vez estimada la función de distribución acumulada hemos podido estimar diferentes cuantiles, variando los niveles de confianza.

Dado un nivel de confianza α , el VaR se entiende como el tiempo que una persona vivirá con una probabilidad $(1 - \alpha)$ una vez ha alcanzado la edad 65. Efectivamente se han podido encontrar diferencias en el número de años que sobrevive una mujer y un hombre. A la vez que se estudiaba la longevidad se ha seguido explorando el efecto del género, modificando la proporción de mujeres, como resultado se ha visto que el VaR crece cuando la proporción de mujeres crece y que las diferencias entre géneros se han mantenido en las últimas décadas y parece ser un comportamiento que se mantendrá por bastante más tiempo, haciendo que la gestión de la mortalidad y la longevidad sea aún más relevante.

A pesar de que la distribución Weibull es la que mejor ajusta la longevidad de la población mexicana, no recrea el comportamiento real de la variable en las edades más avanzadas, es decir, en la cola derecha de la distribución. Cuando el nivel de confianza aumenta, el VaR llega a edades que la experiencia indica son difíciles de alcanzar.

Para los niveles de confianza 0.90 y 0.95 también se han encontrado diferencias en el VaR estimado para hombres y mujeres cuando se utiliza el modelo Brass-type.

El VaR, como medida de riesgo, funciona una vez se ha encontrado un buen ajuste de la distribución de los datos, sin embargo es complicado encontrar una función de distribución paramétrica que pueda reflejar completamente el comportamiento de la variable duración de la vida, lo que nos ha llevado a utilizar métodos noparamétricos.

En el capítulo 5 se han comparado tres tipos de estimaciones: paramétricas, noparamétricas y semiparamétricas. Entre las aportaciones de este capítulo está el uso de un criterio basado en la minimización del error integrado medio al cuadrado. La ventaja de usar este criterio es que mejora el ajuste en la cola derecha de la distribución. Además, nos ha permitido comparar los ajustes ob-

tenidos con métodos paramétricos, noparamétricos y semiparamétricos. Hemos puesto de manifiesto que el método semiparamétrico basado en la doble transformación de los datos comparte algunas ventajas de los métodos paramétricos y noparamétricos.

Finalmente, aunque las primas no puedan ser diferentes entre hombres y mujeres, en el capítulo 6 se vuelve a insistir en la necesidad de utilizar tablas de mortalidad diferenciadas por sexos para la cuantificación del riesgo en los seguros de vida.

Líneas futuras de investigación

Después de modelizar la curva de mortalidad usando métodos estadísticos paramétricos y no paramétricos estamos interesados en implementar un modelo más complejo basado en una ecuación diferencial estocástica (EDE). Milevsky y Promislow (2001) han estudiado la mortalidad usando una EDE asociada con un ruido aditivo basado en un movimiento Browniano, mientras que Giacometti *et al.* (2011) hacen una generalización considerando el caso con un ruido aditivo pero con un coeficiente de difusión dependiente del tiempo y no sólo como una función constante. Los autores concluyen que el uso de una EDE para modelar la tasa de mortalidad refleja más fehacientemente el comportamiento de la mortalidad y abre un campo en el estudio de este tema.

Algunos autores en sus modelos tratan de ajustar la tasa de mortalidad asumiendo correlaciones perfectas entre generaciones, sin embargo, la intuición común sugiere que entre generaciones cercanas estas correlaciones son altas, pero no perfectas. Para explicar esta relación, Giacometti *et al.* (2012) consideran modelos de corta dependencia como son los $ARMA(p, q)$, en particular, sugieren el uso de un $AR(1)$.

El enfoque que nosotros deseamos explorar, consiste en describir la mortalidad con un modelo que tenga en cuenta la dependencia a largo alcance (Long-range dependence). Para lo cual sugerimos una generalización del Modelo Milevsky-Promislow, integrando dicha dependencia.

Se buscará ajustar una ecuación diferencial estocástica con ruido aditivo basado en un movimiento Browniano *fraccionario* (fBm), en lugar del movimiento Browniano estándar. Con lo que la dependencia a largo plazo se podrá controlar con valores del parámetro de Hurst situados en el intervalo $1/2 \leq H < 1$.

De acuerdo al modelo de Milevsky-Promislow, la mortalidad $m_x(t)$ viene dada por

$$m(t) = m_0 \exp(\alpha_0 t + \alpha_1 Y_t), \quad (6.1)$$

donde $h_0, \alpha_1, \alpha_2 > 0$. El proceso estocástico Y_t (proceso de Ornstein-Uhlenbeck fraccionario) satisface la ecuación diferencial estocástica

$$dY_t = -\lambda Y_t dt + \sigma dB_t^H, \quad Y_0 = 0, \quad (6.2)$$

siendo B_t^H un fBm con parámetro de Hurst entre $1/2 \leq H < 1$ y parámetros

$\sigma, \lambda > 0$.

Usando datos reales históricos provenientes de cualquier población, deberemos estimar los parámetros $H, \sigma, \lambda, \alpha_1, \alpha_2$. El proceso metodológico para obtener estos estimadores será el descrito en Brouste e Iacus (2013).

Otra línea futuro es generalizar el estimador núcleo doble transformado con el uso de transformaciones alternativas a las distribuciones Weibull, Lognormal y Champernowne y así poder estudiar la sensibilidad de este estimador en un marco más amplio de transformaciones paramétricas.

Publicaciones

El siguiente listado publicaciones han sido producidas conjuntamente con diferentes autores, aunque las partes empleadas en esta Tesis doctoral constituyen aportaciones de la autora de la misma a todos sus efectos.

1. Ornelas, A. (2011) “Análisis y Tarificación en Seguros de Salud Grupo”. *Investigaciones en Seguros y Gestión del Riesgo: RIESGO 2011*, 197–210.
2. Ornelas, A., Guillén, M. y Alcañiz, M. (2011) “Modelos estadísticos para los costos en carteras colectivas de seguros de salud”. *Revista Ibero-Latinoamericana de Seguros*, 20(35):233–248 .
3. Tural, C., Solà, R., Pérez-Álvarez, N., Moltó, J., Sánchez, M., Moreno-Zamora, A., Ornelas, A., Laguno, M., González, J., Von Wichmann, M.Á., Téllez, M.J., Paredes, R. y Clotet, B. (2011) “Effect of an induction period of pegylated interferon- α 2a and ribavirin on early virological response in HIV-HCV-coinfected patients: Results from the CORAL-2 study”. *Antiviral Therapy*, 16(6):833–841 .
4. Santos, J.R., Llibre, J.M., Imaz, A., Domingo, P., Iribarren, J.A., Mariño, A., Miralles, C., Galindo, M.J., Ornelas, A., Moreno, S., Schapiro, J.M., Clotet, B., Arazo, P., Crespo, M., Mateo, G., del Pozo, M.A., Puig, T. y Riera, M. (2012) “Mutations in the protease gene associated with virological failure to lopinavir/ritonavir-containing regimens”. *Journal of Antimicrobial Chemotherapy*, 67(6):1462–469.
5. Santos, J.R., Moltó, J., Llibre, J.M., Negredo, E., Bravo, I., Ornelas, A., Clotet, B. y Paredes, R. (2012) “Antiretroviral simplification with darunavir/ritonavir monotherapy in routine clinical practice: Safety, effectiveness, and impact on lipid profile”. *PLoS ONE*, 7(5).
6. Guillén, M., Ornelas, A. y Romero, M.J. (2013) “La igualdad, una exigencia para el seguro: problemas derivados de las tablas unisex”. *El Sector Asegurador ante las Transformaciones del Estado del Bienestar*, Fundación de Estudios Financieros, 48:145–157.

7. Llibre, J.M., Revollo, B., Vanegas, S., Lopez-Nuñez, J. J., Ornelas, A., Marin, J.M., Santos, J.R., Marte, P., Morera, M., Zuluaga, P., Tor, J. y Clotet, B. (2013) “Pneumocystis jirovecii pneumonia in HIV-1-infected patients in the late-HAART era in developed countries”. *Scandinavian Journal of Infectious Diseases*, 45(8):635–644.
8. Ornelas, A. y Guillén, M. (2013) “Aplicación del modelo Brass type a la mortalidad de la población asegurada mexicana”. *Investigaciones en Seguros y Gestión del Riesgo: RIESGO 2013*, 101–112.
9. Ornelas, A., Guillén, M. y Alcañiz, M. (2013) “Implications of Unisex Assumptions in the Analysis of Longevity for Insurance Portfolios”. *Lecture Notes in Business Information Processing*, 145:99–107.
10. Ornelas, A. y Guillén, M. (2013) “A comparison between general population mortality, life tables for insurance in Mexico under gender proportion inequality”. *Revista de Metodos Cuantitativos para la Economía y la Empresa*, 16(1):47–67.
11. Echeverría, P., Bonjoch, A., Puig, J., Moltó, J., Paredes, R., Sirera, G., Ornelas, A., Pérez-Álvarez, N., Clotet, B. y Negredo, E.(2014) “Randomised study to assess the efficacy, safety of once-daily etravirine-based regimen as a switching strategy in HIV-infected patients receiving a protease inhibitor-containing regimen. Etraswitch study”. *PLoS ONE*, 9(2).
12. Echeverría, P., Bonjoch, A., Moltó, J., Jou, A., Puig, J., Ornelas, A., Pérez-Álvarez, N., Clotet, B. y Negredo, E. (2014) “Pulse wave velocity as index of arterial stiffness in HIV-infected patients compared with a healthy population”. *Journal of acquired immune deficiency syndromes*, 65(1):50–56.
13. Campos, B., Camps, B., Ornelas, A., Cussó, R., Torras, A. y Arrizabalaga, P. (2015) “Rendimiento académico e inserción laboral de los titulados en medicina por la Universitat de Barcelona. Un estudio longitudinal”. *FEM Fundación Educación Médica*, 18(1):39–45.
14. Llibre, J.M., Bravo, I., Ornelas, A., Santos, J.R., Puig, J., Martín-Iguacel, R., Paredes, R. y Clotet, B. (2015) “Effectiveness of a treatment switch to nevirapine plus tenofovir and emtricitabine (or lamivudine) in adults with HIV-1 suppressed viremia”. *PLoS ONE*, aceptado.

15. Negredo, E., Bonjoch, A., Pérez-Álvarez, N., Ornelas, A., Puig, J., Herrero, C., Estany, C., Del Río, L., Di Gregorio, S., Echeverría, P. y Clotet, B. (2015) “Comparison of two different strategies of treatment with zoledronate in HIV-infected patients with low bone mineral density: single dose versus two doses in 2 years”. *HIV Medicine*, aceptado.
16. Negredo, E., Puig, J., Ornelas, A., Echeverría, P., Bonjoch, A., Estany, C., Higuera C., González-Mestre V. y Clotet, B. (2015) “Ten-year safety with polyacrylamide gel used to correct facial lipoatrophy in HIV-infected patients”. *AIDS Research and Human Retroviruses*, aceptado.

Bibliografía

- Aburto, J. (2012). Estimación de la mortalidad mexicana por el método de Lee-Carter y por el modelo aditivo doble multiplicativo propuesto por Wilmoth. En *SOMEDE, XI reunión Nacional de Investigación Demográfica en México. Aguascalientes, México 2012*.
- Alemaný, R., Bolancé, C., y Guillén, M. (2013). A nonparametric approach to calculating value-at-risk. *Insurance: Mathematics and Economics*, 52(2):255–262.
- Alonso, J., Hoyo, C., y Tuesta, D. (2015). A model for the pension system in Mexico: diagnosis and recommendations. *Journal of Pension Economics and Finance*, 14(1):76–112.
- Altman, N. y Leger, C. (1995). Bandwidth selection for kernel distribution function estimation. *Journal of Statistical Planning and Inference*, 46:195–214.
- Azzalini, A. (1981). A note on the estimation of a distribution function and quantiles by a kernel method. *Biometrika*, 68(1):326–328.
- Betzuen, A. (2010). Un Análisis sobre las posibilidades de predicción de la mortalidad futura aplicando el modelo Lee-Carter. *Anales del Instituto de Actuarios Españoles : Colegio Profesional*, (16):111–140.
- Bolancé, C., Bahraoui, Z., y Artís, M. (2014). Quantifying the risk using copulae with nonparametric marginals. *Insurance: Mathematics and Economics*, 58(0):46 – 56.
- Bolancé, C., Guillén, M., Gustafsson, J., y Nielsen, J. (2012). *Quantitative Operational Risk Models*. Chapman & Hall/CRC finance series. Taylor & Francis.
- Bolancé, C., Guillén, M., y Nielsen, J. P. (2003). Kernel density estimation of actuarial loss functions. *Insurance: Mathematics and Economics*, 32(1):19 – 36.

- Bowman, A., Hall, P., y Prvan, T. (1998). Bandwidth selection for the smoothing of distribution functions. *Biometrika*, 85(4):799–808.
- Brass, W. (1971). *On the Scale of Mortality*. Taylor & Francis, London.
- Brouhns, N., Denuit, M., y Vermunt, J. K. (2002). A Poisson log-bilinear regression approach to the construction of projected lifetables. *Insurance: Mathematics and Economics*, 31(3):373–393.
- Brouste, A. e Iacus, S. M. (2013). Parameter estimation for the discretely observed fractional ornstein-uhlenbeck process and the yuima r package. *Computational Statistics*, 28(4):1529–1547.
- Buch-larsen, T., Nielsen, J. P., Guillén, M., y Bolancé, C. (2005). Kernel density estimation for heavy-tailed distributions using the Champernowne transformation. *Statistics*, 39(6):503–516.
- Cairns, A. J., Blake, D., y Dowd, K. (2006). A Two-Factor Model for Stochastic Mortality with Parameter Uncertainty: Theory and Calibration. *Journal of Risk and Insurance*, 73(4):687–718.
- Cairns, A. J., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., Ong, A., y Balevich, I. (2009). A quantitative comparison of stochastic mortality models using data from England & Wales and the United States. *North American Actuarial Journal*, 13(1):1–35.
- Camposortega, S. (1997). Cien años de mortalidad en México. *Demos*, 10:11–13.
- Cheung, S. L. K., Robine, Jean-Marie, Jow-Ching, E., y Caselli, G. (2005). Three dimensions of the survival curve: horizontalization, verticalization, and longevity extension. *Demography*, 42(2):243–58.
- Chiappori, P. A. y Salanié, B. (2000). Testing for Asymmetric Information in Insurance Markets. *The Journal of Political Economy*, 108(1):56–78.
- Conapo (2012). Proyecciones de la población de México 2010-2050. Taller Interinstitucional para la Conciliación Demografía y las Proyecciones de Población. Consejo Nacional de Población, México. D.F.

- Coughlan, G., Epstein, D., Ong, A., Sinha, A., Hevia-Portocarrero, J., Gingrich, E., Khalaf-Allah, M., y Joseph, P. (2007). LifeMetrics: A toolkit for measuring and managing longevity and mortality risks. Technical report, JP Morgan.
- Cox, D. y Oakes, D. (1984). *Analysis of Survival Data*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis.
- Cramér, H. (1946). *Mathematical Methods of Statistics*. Princeton University Press.
- Currie, I. D. (2006). Smoothing and Forecasting Mortality Rates with P-splines. Talk given at the Institute of Actuaries.
- Currie, I. D., Durban, M., y Eilers, P. H. (2004). Smoothing and forecasting mortality rates. *Statistical Modelling*, 4(4):279–298.
- Curry, C. y O’Connell, A. (2004). An analysis of unisex annuity rates. Technical Report 16, EOC Working Paper, Pensions Policy Institute.
- Dávila-Cervantes, C. A. (2012). Ajuste matemático de la mortalidad general en México 2000, 2005 y 2010. *Papeles de Población*, 18(74):117–150.
- Debón, A., Montes, F., y Puig, F. (2008). Modelling and forecasting mortality in Spain. *European Journal of Operational Research*, 189(3):624–637.
- Debón, A., Montes, F., y Sala, R. (2006). A comparison of nonparametric methods in the graduation of mortality: application to data from the Valencia region (Spain). *International statistical review*, 74(2):215–233.
- Delwarde, A. y Denuit, M. (2003). Importance de la période d ’observation et des âges considérés dans la projection de la mortalité selon la méthode de Lee-Carter. *Belgian Actuarial Bulletin*, 3(1):1–21.
- Delwarde, A., Denuit, M., Guillén, M., y Vidiella-i Anguera, A. (2006). Application of the Poisson log-bilinear projection model to the G5 mortality experience. *Belgian Actuarial Bulletin*, 6(1):54–68.
- Denuit, M., Dhaene, J., Goovaerts, M., y Kaas, R. (2005). *Actuarial Theory for Dependent Risks: Measures, Orders and Models*. Wiley.

- Department of Economic and Social Affairs, P. D. (2013). World Mortality Report. Technical report, Nations, United.
- Felipe, A., Guillén, M., y Nielsen, J. P. (2001). Longevity studies based on kernel hazard estimation. *Insurance: Mathematics and Economics*, 28(2):191–204.
- Fledelius, P., Guillén, M., Nielsen, J. P., y Petersen, K. S. (2004). A Comparative Study of Parametric and Nonparametric Estimators of Old-Age Mortality in Sweden. *Journal of Actuarial Practice*, 11:101–106.
- Fries, J. F. (1980). Aging, natural death, and the compression of morbidity. *New England Journal of Medicine*, 303(3):130–135.
- García, V. M. y Ordorica, M. (2012). Proyección estocástica de la mortalidad mexicana por medio del método de Lee-Carter. *Estudios Demográficos y Urbanos*, 27(2):409–448.
- Gatzert, N. y Wesker, H. (2011). Mortality Risk and its Effect on Shortfall and Risk Management in Life Insurance.
- Giacometti, R., Bertocchi, M., Rachev, S. T., y Fabozzi, F. J. (2012). A comparison of the lee-carter model and ar-arch model for forecasting mortality rates. *Insurance: Mathematics and Economics*, 50(1):85 – 93.
- Giacometti, R., Ortobelli, S., y Bertocchi, M. (2011). A Stochastic Model for Mortality Rate on Italian Data. *Journal of Optimization Theory and Applications*, 149(1):216–228.
- González Rabanal, M. y Sáenz de Jáuregui Sanz, L. (2008). *La política comunitaria*. Colección: Estudios de Hacienda Pública. Instituto de Estudios Fiscales, Madrid.
- Guillén, M. (2012). Sexless and beautiful data: from quantity to quality. *Annals of Actuarial Science*, 6(2):231–234.
- Guillén, M., Ornelas, A., y Romero, M. (2013). La igualdad, una exigencia para el seguro: problemas derivados de las tablas unisex. En Blanco-Morales, P., G. M., editor, *El Sector Asegurador ante las Transformaciones del Estado del Bienestar*, volumen 48, pp. 145–157. Fundación de Estudios Financieros.

- Hastie, T. J. y Tibshirani, R. J. (1990). *Generalized Additive Models*. Chapman and Hall.
- Hudson, R. (2007). Mortality projections and unisex pricing of annuities in the UK. *Journal of Financial Regulation and Compliance*, 15(2):166–179.
- Hyndman, R. J., Booth, H., Tickle, L., y Maindonald., J. (2014). **demography**: Forecasting mortality, fertility, migration and population data. Technical report. R package version 1.18.
- Kannisto, V. (2000). Measuring the compression of mortality. *Demographic Research*, 3(6):1–24.
- Lee, R. D. y Carter, L. R. (1992). Modeling and forecasting US mortality. *Journal of the American statistical association*, 87(419):659–671.
- Llibre, J. M., Revollo, B., Vanegas, S., Lopez-Nuñez, J. J., Ornelas, A., Marin, J. M., Santos, J. R., Marte, P., Morera, M., Zuluaga, P., Tor, J., y Clotet, B. (2013). Pneumocystis jirovecii pneumonia in hiv-1-infected patients in the late-haart era in developed countries. *Scandinavian Journal of Infectious Diseases*, 45(8):635–644.
- Mendoza, M., Madrigal, A., y Martínez, E. (1999). Modelos Estadísticos de Mortalidad. Análisis de Datos. volumen 77 de *Documentos de Trabajo CNSF*.
- Mendoza, M., Madrigal, A., y Martínez, E. (2000). Tablas de Mortalidad CNSF 2000-I y CNSF 2000-G. volumen 80 de *Documentos de Trabajo CNSF*.
- Milevsky, M. A. y Promislow, S. D. (2001). Mortality derivatives and the option to annuitise. *Insurance: Mathematics and Economics*, 29(3):299 – 318. Papers presented at the 4th {IME} Conference, Universitat de Barcelona, Barcelona, 24-26 July 2001.
- Mina, A. (1990). Contribuciones en años de vida por sexo y edad en México. *Estudios Demográficos y Urbanos*, 5(1(13)):149–178.
- Mina, A. (2006). Ley de mortalidad mexicana. Funciones de supervivencia. *Estudios Demográficos y Urbanos*, 21(2(62)):431–456.
- Mina, A. (2011). La obtención y proyección de tablas de mortalidad empleando curvas spline. *Papeles de población*, 17(69):49–72.

- Ornelas, A., Guillén, M., y Alcañiz, M. (2013). Implications of Unisex Assumptions in the Analysis of Longevity for Insurance Portfolios. En Fernández-Izquierdo, M.-A., Muñoz Torres, M.-J., y León, R., editores, *Modeling and Simulation in Engineering, Economics, and Management*, volumen 145 de *Lecture Notes in Business Information Processing*, pp. 99–107. Springer Berlin Heidelberg.
- Partida, V. (2008). *Proyecciones de la población de México, de las entidades federativas, de los municipios y de las localidades 2005-2050*. Consejo Nacional de Población, México. D.F.
- Pitacco, E., Denuit, M., Haberman, S., y Olivieri, A. (2009). *Modelling Longevity Dynamics for Pensions and Annuity Business*. United States: Oxford University Press.
- Plat, R. (2009). Stochastic portfolio specific mortality and the quantification of mortality basis risk. *Insurance: Mathematics and Economics*, 45(1):123–132.
- R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rendón, J. (2000). Tablas de mortalidad México 2000, AMIS-AMA.
- Rendón, J. (2012). The construction of multiple decrement models from associated single decrement experiences. En *IAALS, Mexico city 2012*, pp. 1–6.
- Renshaw, A. E. y Haberman, S. (2003). Lee-Carter Mortality Forecasting: a Parallel Generalized Linear Modelling Approach for England and Wales Mortality Projections. *Journal of the Royal Statistical Society Series C*, 52(1):119–137.
- Renshaw, A. E. y Haberman, S. (2006). A cohort-based extension to the Lee-Carter model for mortality reduction factors. *Insurance: Mathematics and Economics*, 38(3):556–570.
- Rinne, H. (2010). *The Weibull Distribution: A Handbook*. Taylor & Francis.
- Santos, J. R., Llibre, J. M., Imaz, A., Domingo, P., Iribarren, J. A., Mariño, A., Miralles, C., Galindo, M. J., Ornelas, A., Moreno, S., Schapiro, J. M.,

- Clotet, B., Arazo, P., Crespo, M., Mateo, G., del Pozo, M. A., Puig, T., y Riera, M. (2012). Mutations in the protease gene associated with virological failure to lopinavir/ritonavir-containing regimens. *Journal of Antimicrobial Chemotherapy*, 67(6):1462–1469.
- SAS Institute Inc. (2003). *SAS/STAT Software, Version 9.2*. Cary, NC.
- Scott, D. W. (2001). Parametric Statistical Modeling by Minimum Integrated Square Error. *Technometrics*, 43(3):274–285.
- Silverman, B. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis.
- Terrell, G. R. (1990). The Maximal Smoothing Principle in Density Estimation. *Journal of the American Statistical Association*, 85(410):470–477.
- Thomas, R. G. (2007). Some Novel Perspectives on Risk Classification. *The Geneva Papers on Risk and Insurance Issues and Practice*, 32(1):105–132.
- Tural, C., Solà, R., Pérez Álvarez, N., Moltó, J., Sánchez, M., Moreno Zamora, A., Ornelas, A., Laguno, M., González, J., Von Wichmann, M. A., Téllez, M. J., Paredes, R., y Clotet, B. (2011). Effect of an induction period of pegylated interferon- α 2a and ribavirin on early virological response in hiv-hcv-coinfected patients: Results from the coral-2 study. *Antiviral Therapy*, 16(6):833–841.
- Vaupel, J. W., Zhang, Z., y van Raalte, A. A. (2011). Life expectancy and disparity: an international comparison of life table data. *BMJ open*, 1(1):e000128.
- Warth, J. (2011). Productos de Riesgos Preferentes en un Entorno Unisex. *GenRe Risk Insights*, 15(3):1–4.
- Wilmoth, J. R. (1990). Variation in vital rates by age, period, and cohort. *Sociological methodology*, 20:295–335.
- Yue, J. C. (2012). Mortality Compression and Longevity Risk. *North American Actuarial Journal*, 16(4):434–448.

Anexo

Código R

En esta sección se incluyen los scripts utilizados para obtener los resultados reportados en esta Tesis. El lenguaje de programación empleado es R Core Team (2014).

Tasas de mortalidad

- Lee el archivo *.csv que contiene número de personas vivas o muertas,
- Interpola el numero de personas vivas para los años en los cuales no se tiene información,
- Calcula las tasas crudas de mortalidad $m_{x,t}$ y el logaritmo $\log m_{x,t}$,
- Calcula la diferencia de las tasas de mortalidad entre hombres y mujeres.

Script

```
#####
##### YEARS AND AGES #####
ages<-seq(0,100,1) ; years<- seq(1990,2010,1)
ny<-length(years); na<-length(ages)

fy<-years [1]; ly<-years [ny]
fa<-ages [1]; la<-ages [na]

#### GRAPHS COLORS ####
numeros<-as.vector(round(runif(ny)*100,0))
color<-colors()[numeros]
color<-topo.colors(ny)
my.color <- colorRampPalette( c("blue", "green"))
color<-my.color(ny)

#####
##### SURVIVERS #####
hom<-read.table("hcenso.csv", header=TRUE, sep=";", na.strings="NA", dec=".")
muj<-read.table("mcenso.csv", header=TRUE, sep=";", na.strings="NA", dec=".")

rownames(muj)<-rownames(hom)<-ages
hom<-hom[,-1]
muj<-muj[,-1]

##### INTERPOLATION SURVIVERS #####
lentre<-lwoman<-lman<-mat.or.vec(nrow(hom), length(years))
```



```

colnames(lentire) <- colnames(lwoman) <- colnames(lman) <- years
rownames(lentire) <- rownames(lwoman) <- rownames(lman) <- ages

for(i in 1:101){
  lman[i,] <- round(as.vector(approx(seq(1990,2010,5), hom[i,], n=21,
    method="linear")[[2]]),0)
  lwoman[i,] <- round(as.vector(approx(seq(1990,2010,5), muj[i,], n=21,
    method="linear")[[2]]),0)
}
lentire <- lman + lwoman

#####
##### DEATHS #####
dman <- read.table("hdef.csv", header=TRUE, sep=",", na.strings="NA", dec=".")
dwoman <- read.table("mdef.csv", header=TRUE, sep=",", na.strings="NA", dec=".")

rownames(dwoman) <- rownames(dman) <- ages
dman <- dman[, -1]
dwoman <- dwoman[, -1]

dentire <- dman + dwoman

#####
##### MORTALITY RATES #####

#### CRUDE MORTALITY RATE ####
raw.ma <- dman/lman
raw.wo <- dwoman/lwoman
raw.e <- dentire/lentire

#### LAST AGE 100 ####
raw.ma[101,] <- 1
raw.wo[101,] <- 1
raw.e[101,] <- 1

#### LOG MORTALITY RATES ####
lograw.ma <- t(as.matrix(log(raw.ma)))
lograw.wo <- t(as.matrix(log(raw.wo)))
lograw.e <- t(as.matrix(log(raw.e)))

#### GRAPH ####
library(lattice)

x11(); par(las=1)
wireframe(lograw.ma, xlab="Years", ylab="Age", zlab="",
  row.values=years, column.values=ages, xlim=years, ylim=ages,
  main=paste('Log Mortality rates, Males', fy, '-', ly),
  scales=list(arrows=FALSE, x=rev(years)),
  aspect=c(2,1), drape=TRUE, colorkey=TRUE,
  screen=list(z=-50, x=-75), col='blue')

x11(); par(las=1)
wireframe(lograw.wo, xlab="Years", ylab="Age", zlab="",

```

```

row.values=years ,column.values=ages , xlim=years , ylim=ages ,
main = paste('Log Mortality rates, Females', fy, '-', ly),
scales= list(arrows=FALSE, x= rev(years)),
aspect=c(2,1), drape = TRUE, colorkey = TRUE,
screen = list(z = -50, x = -75, col='blue')

#### 2D CHANGES ####
x11(); par(las=1)
filled.contour(years , ages , lograw.ma, color = terrain.colors ,
  nlevels=20,
  plot.title = title(main=paste('Changes Mortality Rates, Males', fy, '-',
    ly),
  xlab = "Years", ylab = "Age"),
  plot.axes = {axis(1, seq(fy, ly, by = 5))
    axis(2, seq(fa, la, by = 10)) },
  key.title = title(main="Log"),
  key.axes = axis(4, seq(-8.5, 0, by = .5)))

x11(); par(las=1)
filled.contour(years , ages , lograw.wo, color = terrain.colors ,
  nlevels=20,
  plot.title = title(main = paste('Changes Mortality Rates, Females',
    fy, '-', ly),
  xlab = "Years", ylab = "Age"),
  plot.axes = {axis(1, seq(fy, ly, by = 5))
    axis(2, seq(fa, la, by = 10)) },
  key.title = title(main="Log"),
  key.axes = axis(4, seq(-8.5, 0, by = .5)))

#### COHORT CHANGES ####
raw_dif.wo<-raw_dif.ma<-raw_dif.e<-mat.or.vec(na,ny-1)

rownames(raw_dif.wo)<-rownames(raw_dif.ma)<-rownames(raw_dif.e)<-ages
colnames(raw_dif.wo)<-colnames(raw_dif.ma)<-colnames(raw_dif.e)<-years[-ny]

for (i in 1:na){
  for (j in 1:(ny-1)){
    raw_dif.ma[i,j]=(raw.ma[i,j+1]/raw.ma[i,j])-1
    raw_dif.wo[i,j]=(raw.wo[i,j+1]/raw.wo[i,j])-1
    raw_dif.e[i,j]=(raw.e[i,j+1]/raw.e[i,j])-1
  }
}

#### GRAPH ####
years8<- years[-ny]
x11(); par(las=1)
filled.contour(years8 , ages , t(raw_dif.ma), col= topo.colors(ny-1),
  plot.title = title(main = "Changes Mortality Rates. Males ",
  xlab = "Years", ylab = "Age"),
  plot.axes = {axis(1, seq(fy, ly-1, by = 1))
    axis(2, seq(fa, la, by = 10)) },
  key.title = title(main="%"),
  key.axes = axis(4, seq(-0.4, 1.3, by = 0.1)))

```

```

x11(); par(las=1)
filled.contour(years8, ages, t(raw_dif.wo), col= topo.colors(ny-1),
  plot.title = title(main = " Changes Mortality Rates. Females ",
  xlab = "Years", ylab = "Age"),
  plot.axes = {axis(1, seq(fy, ly-1, by = 1))
              axis(2, seq(fa, la, by = 10)) },
  key.title = title(main=" %"),
  key.axes = axis(4, seq(-0.4, 1.3, by = 0.1)))

##### GENDER DIFFERENCES #####
dif_raw_gender<-(raw.ma/raw.wo)-1

#### GRAPH ####
x11(); par(las=1)
filled.contour(years, ages, t(dif_raw_gender), color = topo.colors,
  nlevels=20,
  plot.title = title(main=paste('Mortality Rates', fy, '-',ly),
  xlab = "Years", ylab = "Age"),
  plot.axes = {axis(1, seq(fy, ly, by = 5))
              axis(2, seq(fa, la, by = 10)) },
  key.title = title(main=" %"),
  key.axes = axis(4, seq(-1, 3, by = .5)))

```

Ajuste de modelos con LifeMetrics

- Lee el recurso LifeMetrics para la estimación de modelos (ver Coughlan *et al.*, 2007),
- Estima los parámetros de los modelos: Lee-Carter, Renshaw-Haberman, Age-Period-Cohort, CDB 1, 2 y 3.

Script

```

#####
##### LIFEMETRICS #####

source("C:\\lifemetrics_sourcecode\\fitModels.r")
# INPUTS:
#   XV = VECTOR OF AGES, LENGTH N=NA
#   YV = VECTOR OF YEARS, LENGTH M=NY
#   ETX = M X N MATRIX OF EXPOSURES
#   DENTIREX = M X N MATRIX OF DEATHS
#   WA = M X N MATRIX OF WEIGHTS (0 OR 1)

#### ASSUME THAT WA=1 ####
wa<-matrix(1,ny,na)

```

```

# OUTPUTS:
# BETA1 = VECTOR OF LENGTH N OF ESTIMATES FOR BETA_1
# BETA2 = VECTOR OF LENGTH N OF ESTIMATES FOR BETA_2
# BETA3 = VECTOR OF LENGTH N OF ESTIMATES FOR BETA_3
# KAPPA2 = VECTOR OF LENGTH M OF ESTIMATES FOR KAPPA_1
# GAMMA3 = VECTOR OF LENGTH (M+N-1) OF ESTIMATES FOR GAMMA_3 (M2, M3, M6,
M7, M8 ONLY)
# CY = VECTOR OF COHORT YEARS OF BIRTH
# EPSILON = M BY N ARRAY OF STANDARDIZED RESIDUALS
# LL = MAXIMUM LOG-LIKELIHOOD
# BIC = BAYES INFORMATION CRITERION
# NPAR = THE EFFECTIVE NUMBER OF PARAMETERS ESTIMATED IN THE MODEL (DF)
# MTX = THE M BY N ARRAY OF CRUDE DEATH RATES
# MHATIS = THE M BY N ARRAY OF FITTED DEATH RATES
# MTXLASTYEAR = MTX[M,]

#####
##### M1 #####
##### LEE CARTER MODEL #####

# LOG M(T,X) = BETA1(X) + BETA2(X).KAPPA2(T) + POISSON ERROR

resM1.ma=fit701(ages,years,t(lman),t(dman),wa)
resM1.wo=fit701(ages,years,t(lwoman),t(dwoman),wa)
resM1.e=fit701(ages,years,t(lentire),t(dentire),wa)

names(resM1.ma)

#### GRAPH FITTED RATES ####
library(lattice)
x11(); par(las=1)
persp(resM1.ma$y, resM1.ma$x, log(resM1.ma$mhat), phi=30, theta=55, col
="chartreuse2",
main=paste('Log Mortality rates LC, Males', fy, '-', ly),
ylab=paste('Age',fa, '-', la), xlab=paste('Years',fy, '-', ly), zlab='Log
Mortality rate')

x11(); par(las=1)
persp(resM1.wo$y, resM1.wo$x, log(resM1.wo$mhat), phi=30, theta=55, col
="chartreuse2",
main=paste('Log Mortality rates LC, Females', fy, '-', ly),
ylab=paste('Age',fa, '-', la), xlab=paste('Years',fy, '-', ly), zlab='Log
Mortality rate')

#### PARAMETERS ####
x11(); op<-par(mfrow=c(3,1), las=1)
plot(resM1.ma$x, resM1.ma$beta1, xlab='Age (x)', ylab='Beta 1(x)', pch=19)
plot(resM1.ma$x, resM1.ma$beta2, xlab='Age (x)', ylab='Beta 2(x)', pch=19)
plot(resM1.ma$y, resM1.ma$kappa2, xlab='Years (t)', ylab='Kappa 2 (t)',
pch=19)
par(op)
mtext('Males', side=3, line=2, font=2)

```

```

x11(); op<-par(mfrow=c(3,1), las=1)
plot(resM1.wo$x, resM1.wo$beta1, xlab='Age (x)', ylab='Beta 1(x)', pch=19)
plot(resM1.wo$x, resM1.wo$beta2, xlab='Age (x)', ylab='Beta 2(x)', pch=19)
plot(resM1.wo$y, resM1.wo$kappa2, xlab='Years (t)', ylab='Kappa 2 (t)',
      pch=19)
par(op)
mtext('Females', side=3, line=2, font=2)

##### GENDER DIFFERENCES #####
x11(); par(mfrow=c(3,1), las=1)
plot(resM1.ma$x, resM1.ma$beta1, xlab='Age (x)', ylab='Beta 1(x)', type='l',
      col=color[1],
      ylim=c(min(resM1.ma$beta1, resM1.wo$beta1),
             max(resM1.ma$beta1, resM1.wo$beta1)), lwd=2,
      main=expression(hat(beta) * " LC Model"))
lines(resM1.wo$x, resM1.wo$beta1, xlab='Age (x)', ylab='Beta 1(x)', type='l',
       col=color[10], lwd=2)
legend('bottomright', c('Males', 'Females'), col=c(color[1], color[10]), lwd=2,
       bty='n')

plot(resM1.ma$x, resM1.ma$beta2, xlab='Age (x)', ylab='Beta 2(x)', type='l',
      col=color[1],
      ylim=c(min(resM1.ma$beta2, resM1.wo$beta2),
             max(resM1.ma$beta2, resM1.wo$beta2)), lwd=2,
      main=expression(hat(beta) * " LC Model"))
lines(resM1.wo$x, resM1.wo$beta2, xlab='Age (x)', ylab='Beta 2(x)', type='l',
       col=color[10], lwd=2)
abline(h=0)

plot(resM1.ma$y, resM1.ma$kappa2, xlab='Years (t)', ylab='Kappa 2 (t)',
      type='l', col=color[1],
      ylim=c(min(resM1.ma$kappa2, resM1.wo$kappa2),
             max(resM1.ma$kappa2, resM1.wo$kappa2)), lwd=2,
      main=expression(hat(kappa) * " LC Model"))
lines(resM1.wo$y, resM1.wo$kappa2, xlab='Years (t)', ylab='Kappa 2 (t)',
       type='l', col=color[10], lwd=2)

##### M2 #####
##### RENSHAW-HABERMAN 3-FACTOR MODEL #####

# LOG M(T,X) = BETA1(X) + BETA2(X).KAPPA2(T) +BETA3(X).GAMMA3(T-X) + POISSON
  ERROR
resM2.ma=fit702(ages, years, t(lman), t(dman), wa)
resM2.wo=fit702(ages, years, t(lwoman), t(dwoman), wa)
resM2.e=fit702(ages, years, t(lentire), t(dentire), wa)

#### GRAPH ####
x11(); par(las=1)
persp(resM2.ma$y, resM2.ma$x, log(resM2.ma$mhat), phi=30, theta=55, col
      ="chartreuse2",
      main=paste('Log Mortality rates RH, Males', fy, '-', ly),
      ylab=paste('Age', fa, '-', la), xlab=paste('Years', fy, '-', ly), zlab='Log
      Mortality rate')

```

```

x11(); par(las=1)
persp(resM2.wo$y, resM2.wo$x, log(resM2.wo$mhat), phi=30, theta=55,col
      ="chartreuse2",
      main=paste('Log Mortality rates RH, Females', fy, '-',ly),
      ylab=paste('Age',fa, '-',la), xlab=paste('Years',fy, '-',ly), zlab='Log
      Mortality rate')

#### PARAMETERS ####
x11(); op<-par(mfrow=c(3,2), las=1)
plot(resM2.ma$x, resM2.ma$beta1, xlab='Age (x)', ylab='Beta 1(x)', pch=19)
plot(resM2.ma$x, resM2.ma$beta2, xlab='Age (x)', ylab='Beta 2(x)', pch=19)
plot(resM2.ma$x, resM2.ma$beta3, xlab='Age (x)', ylab='Beta 3(x)', pch=19)
plot(resM2.ma$y, resM2.ma$kappa2, xlab='Years (t)', ylab='Kappa 2 (t)',
      pch=19)
plot(resM2.ma$cy, resM2.ma$gamma3, xlab='Years', ylab='Gamma 3 (t)', pch=19)
par(op)
mtext('Males', side=3, line=2, font=2)

x11(); op<-par(mfrow=c(3,2), las=1)
plot(resM2.wo$x, resM2.wo$beta1, xlab='Age (x)', ylab='Beta 1(x)', pch=19)
plot(resM2.wo$x, resM2.wo$beta2, xlab='Age (x)', ylab='Beta 2(x)', pch=19)
plot(resM2.wo$x, resM2.wo$beta3, xlab='Age (x)', ylab='Beta 3(x)', pch=19)
plot(resM2.wo$y, resM2.wo$kappa2, xlab='Years (t)', ylab='Kappa 2 (t)',
      pch=19)
plot(resM2.wo$cy, resM2.wo$gamma3, xlab='Years', ylab='Gamma 3 (t)', pch=19)
par(op)
mtext('Females', side=3, line=2, font=2)

##### GENDER DIFFERENCES #####
x11(); par(mfrow=c(3,2), las=1)
plot(resM2.ma$x, resM2.ma$beta1, xlab='Age (x)', ylab='Beta 1(x)', type='l',
      col=color[1],
      ylim=c(min(resM2.ma$beta1, resM2.wo$beta1),
             max(resM2.ma$beta1, resM2.wo$beta1)), lwd=2,
      main=expression(hat(beta) * " RH Model"))
lines(resM2.wo$x, resM2.wo$beta1, xlab='Age (x)', ylab='Beta 1(x)', type='l',
      col=color[10], lwd=2)
legend('bottomright',
      c('Males', 'Females'), col=c(color[1], color[10]), lwd=2, bty='n')

plot(resM2.ma$x, resM2.ma$beta2, xlab='Age (x)', ylab='Beta 2(x)', type='l',
      col=color[1],
      ylim=c(min(resM2.ma$beta2, resM2.wo$beta2),
             max(resM2.ma$beta2, resM2.wo$beta2)), lwd=2,
      main=expression(hat(beta) * " RH Model"))
lines(resM2.wo$x, resM2.wo$beta2, xlab='Age (x)', ylab='Beta 2(x)', type='l',
      col=color[10], lwd=2)
abline(h=0)

plot(resM2.ma$x, resM2.ma$beta3, xlab='Age (x)', ylab='Beta 3(x)', type='l',
      col=color[1],
      ylim=c(min(resM2.ma$beta3, resM2.wo$beta3),
             max(resM2.ma$beta3, resM2.wo$beta3)), lwd=2,

```

```

        main=expression(hat(beta) * " RH Model"))
lines(resM2.wo$x, resM2.wo$beta3, xlab='Age (x)', ylab='Beta 3(x)', type='l',
      col=color[10], lwd=2)
abline(h=0)

plot(resM2.ma$y, resM2.ma$kappa2, xlab='Years (t)', ylab='Kappa 2 (t)',
     type='l', col=color[1],
     ylim=c(min(resM2.ma$kappa2, resM2.wo$kappa2),
            max(resM2.ma$kappa2, resM2.wo$kappa2)), lwd=2,
     main=expression(hat(kappa) * " RH Model"))
lines(resM2.wo$y, resM2.wo$kappa2, xlab='Years (t)', ylab='Kappa 2 (t)',
      type='l', col=color[10], lwd=2)

plot(resM2.ma$cy, resM2.ma$gamma3, xlab='Years (t)', ylab='Gamma 3 (t)',
     type='l', col=color[1],
     ylim=c(min(resM2.ma$gamma3, resM2.wo$gamma3),
            max(resM2.ma$gamma3, resM2.wo$gamma3)), lwd=2,
     main=expression(hat(gamma) * " RH Model"))
lines(resM2.wo$cy, resM2.wo$gamma3, xlab='Years (t)', ylab='Gamma 3 (t)',
      type='l', col=color[10], lwd=2)

##### M3 #####
##### CURRIE 3-FACTOR AGE-PERIOD-COHORT MODEL #####

# LOG M(T,X) = BETA1(X) + KAPPA2(T) + GAMMA3(T-X) + POISSON ERROR
# RENSHAW-HABERMAN MODEL M2 WITH BETA2(X)=1 AND BETA3(X)=1 FOR ALL X

resM3.ma=fit703(ages, years, t(lman), t(dman), wa)
resM3.wo=fit703(ages, years, t(lwoman), t(dwoman), wa)
resM3.e=fit703(ages, years, t(lentire), t(dentire), wa)

#### GRAPH FITTED RATES ####
x11(); par(las=1)
persp(resM3.ma$y, resM3.ma$x, log(resM3.ma$mhat), phi=30, theta=55, col
      ="chartreuse2",
      main=paste('Log Mortality rates APC, Males', fy, '-', ly),
      ylab=paste('Age', fa, '-', la), xlab=paste('Years', fy, '-', ly), zlab='Log
      Mortality rate')

x11(); par(las=1)
persp(resM3.wo$y, resM3.wo$x, log(resM3.wo$mhat), phi=30, theta=55, col
      ="chartreuse2",
      main=paste('Log Mortality rates APC, Females', fy, '-', ly),
      ylab=paste('Age', fa, '-', la), xlab=paste('Years', fy, '-', ly), zlab='Log
      Mortality rate')

#### PARAMETERS ####
op <- par(mfrow=c(3,1), las=1)
plot(resM3.ma$x, resM3.ma$beta1, xlab='Age (x)', ylab='Beta 1(x)', pch=19)
plot(resM3.ma$y, resM3.ma$kappa2, xlab='Years (t)', ylab='Kappa 2 (t)',
     pch=19)
plot(resM3.ma$cy, resM3.ma$gamma3, xlab='Years', ylab='Gamma 3 (t)', pch=19)
par(op)
mtext('Males', side=3, line=2, font=2)

```

```

op <- par(mfrow=c(3,1), las=1)
plot(resM3.wo$x, resM3.wo$beta1, xlab='Age (x)', ylab='Beta 1(x)', pch=19)
plot(resM3.wo$y, resM3.wo$kappa2, xlab='Years (t)', ylab='Kappa 2 (t)',
      pch=19)
plot(resM3.wo$cy, resM3.wo$gamma3, xlab='Years', ylab='Gamma 3 (t)', pch=19)
par(op)
mtext('Females', side=3, line=2, font=2)

##### GENDER DIFFERENCES #####
x11(); par(mfrow=c(3,1), las=1)
plot(resM3.ma$x, resM3.ma$beta1, xlab='Age (x)', ylab='Beta 1(x)',
      type='l', col=color[1], ylim=c(min(resM3.ma$beta1, resM3.wo$beta1),
      max(resM3.ma$beta1, resM3.wo$beta1)), lwd=2,
      main=expression(hat(beta) * " APC Model" )
lines(resM3.wo$x, resM3.wo$beta1, xlab='Age (x)', ylab='Beta 1(x)',
      type='l', col=color[10], lwd=2)
abline(h=0)
legend(20, -2, c('Males', 'Females'), col=c(color[1], color[10]), lwd=2,
      bty='n')

plot(resM3.ma$y, resM3.ma$kappa2, xlab='Years (t)', ylab='Kappa 2 (t)',,
      type='l', col=color[1], ylim=c(min(resM3.ma$kappa2, resM3.wo$kappa2),
      max(resM3.ma$kappa2, resM3.wo$kappa2)), lwd=2,
      main=expression(hat(kappa) * " APC Model" ))
lines(resM3.wo$y, resM3.wo$kappa2, xlab='Years (t)', ylab='Kappa 2 (t)',
      type='l', col=color[10], lwd=2)

plot(resM3.ma$cy, resM3.ma$gamma3, xlab='Years (t)', ylab='Gamma 3 (t)', ,
      type='l', col=color[1], ylim=c(min(resM3.ma$gamma3, resM3.wo$gamma3),
      max(resM3.ma$gamma3, resM3.wo$gamma3)), lwd=2,
      main=expression(hat(gamma) * " APC Model" ))
lines(resM3.wo$cy, resM3.wo$gamma3, xlab='Years (t)', ylab='Gamma 3 (t)',
      type='l', col=color[10], lwd=2)

##### M5 #####
##### CBD MODEL: 2-FACTOR MODEL PUBLISHED IN JRI DEC, 2006 #####

# LOGIT Q(T,X) = A1(T)+A2(T)(X-XBAR) + POISSON ERROR

resM5.ma=fit705(ages, years, t(lman), t(dman), wa)
resM5.wo=fit705(ages, years, t(lwoman), t(dwoman), wa)
resM5.e=fit705(ages, years, t(lentire), t(dentire), wa)

#### GRAPH FITTED RATES ####
#### PARAMETERS ####
##### GENDER DIFFERENCES #####

##### M6 #####
##### CBD1 MODEL: 3-FACTOR APC MODEL #####

# LOGIT Q(T,X) = A1(T)+A2(T)(X-XBAR) +A3(T-X) + POISSON ERROR

resM6.ma=fit706(ages, years, t(lman), t(dman), wa)

```



```

resM6.wo=fit706(ages,years,t(lwoman),t(dwoman),wa)
resM6.e=fit706(ages,years,t(lentire),t(dentire),wa)

#### GRAPH FITTED RATES ####
#### PARAMETERS ####
##### GENDER DIFFERENCES #####

##### M7 #####
##### CBD2 MODEL: 3-FACTOR APC MODEL #####

# LOGIT  $q(t,x) = A1(t)+A2(t)(x-xbar)+A4(t)((x-xbar)^2-x2bar)+A3(t-x) +$ 
  POISSON ERROR

resM7.ma=fit707(ages,years,t(lman),t(dman),wa)
resM7.wo=fit707(ages,years,t(lwoman),t(dwoman),wa)
resM7.e=fit707(ages,years,t(lentire),t(dentire),wa)

#### GRAPH FITTED RATES ####
#### PARAMETERS ####
##### GENDER DIFFERENCES #####

##### M8 #####
##### CBD3 MODEL: 3-FACTOR APC MODEL #####

# LOGIT  $q(t,x) = A1(t)+A2(t)(x-xbar)+A3(t-x)(x0-x) +$  POISSON ERROR

resM8.ma=fit708(ages,years,t(lman),t(dman),wa)
resM8.wo=fit708(ages,years,t(lwoman),t(dwoman),wa)
resM8.e=fit708(ages,years,t(lentire),t(dentire),wa)

#### GRAPH FITTED RATES ####
#### PARAMETERS ####
##### GENDER DIFFERENCES #####

```

Elección del mejor modelo

Evaluando la bondad de ajuste de los modelos estimados:

- Cálculo del BIC y del Loglikelihood,
- Análisis de residuales,
- Test χ^2 .

Script

```

#####
##### GOODNESS OF FIT #####

```

```

#### BAYESIAN INFORMATION CRITERION ####
bic.ma<-as.matrix(c(resM1.ma$BIC, resM2.ma$BIC, resM3.ma$BIC, resM5.ma$BIC,
  resM6.ma$BIC, resM7.ma$BIC, resM8.ma$BIC))
bic.wo<-as.matrix(c(resM1.wo$BIC, resM2.wo$BIC, resM3.wo$BIC, resM5.wo$BIC,
  resM6.wo$BIC, resM7.wo$BIC, resM8.wo$BIC))
bic.e<-as.matrix(c(resM1.e$BIC, resM2.e$BIC, resM3.e$BIC, resM5.e$BIC,
  resM6.e$BIC, resM7.e$BIC, resM8.e$BIC))

bic<-cbind(bic.ma, bic.wo, bic.e)
rownames(bic)<-c('M1', 'M2', 'M3', 'M5', 'M6', 'M7', 'M8')
colnames(bic)<-c('Male', 'Female', 'All')
bic

#### MAXIMUM LOG-LIKELIHOOD ####
ll.ma<-as.matrix(c(resM1.ma$ll, resM2.ma$ll, resM3.ma$ll, resM5.ma$ll,
  resM6.ma$ll, resM7.ma$ll, resM8.ma$ll))
ll.wo<-as.matrix(c(resM1.wo$ll, resM2.wo$ll, resM3.wo$ll, resM5.wo$ll,
  resM6.wo$ll, resM7.wo$ll, resM8.wo$ll))
ll.e<-as.matrix(c(resM1.e$ll, resM2.e$ll, resM3.e$ll, resM5.e$ll, resM6.e$ll,
  resM7.e$ll, resM8.e$ll))

ll<-cbind(ll.ma, ll.wo, ll.e)
rownames(ll)<-c('M1', 'M2', 'M3', 'M5', 'M6', 'M7', 'M8')
colnames(ll)<-c('Male', 'Female', 'All')
ll

##### STANDARDIZED RESIDUALS #####
srM1.ma<-resM1.ma$epsilon; srM2.ma<-resM2.ma$epsilon;
  srM3.ma<-resM3.ma$epsilon; srM5.ma<-resM5.ma$epsilon;
  srM6.ma<-resM6.ma$epsilon; srM7.ma<-resM7.ma$epsilon;
  srM8.ma<-resM8.ma$epsilon

srM1.wo<-resM1.wo$epsilon; srM2.wo<-resM2.wo$epsilon;
  srM3.wo<-resM3.wo$epsilon; srM5.wo<-resM5.wo$epsilon;
  srM6.wo<-resM6.wo$epsilon; srM7.wo<-resM7.wo$epsilon;
  srM8.wo<-resM8.wo$epsilon

srM1.e<-resM1.e$epsilon; srM2.e<-resM2.e$epsilon; srM3.e<-resM3.e$epsilon;
  srM5.e<-resM5.e$epsilon; srM6.e<-resM6.e$epsilon;
  srM7.e<-resM7.e$epsilon; srM8.e<-resM8.e$epsilon

rownames(srM1.ma)<-rownames(srM2.ma)<-
  rownames(srM3.ma)<-rownames(srM1.wo)<-
  rownames(srM2.wo)<-rownames(srM3.wo)<-years

#### VARIANCE ####
vsrM1.ma<-var(as.vector(srM1.ma)); vsrM2.ma<-var(as.vector(srM2.ma))
vsrM3.ma<-var(as.vector(srM3.ma)); vsrM5.ma<-var(as.vector(srM5.ma))
vsrM6.ma<-var(as.vector(srM6.ma)); vsrM7.ma<-var(as.vector(srM7.ma))
vsrM8.ma<-var(as.vector(srM8.ma))

(vsr.ma<-cbind(vsrM1.ma, vsrM2.ma, vsrM3.ma, vsrM5.ma, vsrM6.ma, vsrM7.ma,
  vsrM8.ma))

```

```

vsrM1.wo<-var(as.vector(srM1.wo)); vsrM2.wo<-var(as.vector(srM2.wo))
vsrM3.wo<-var(as.vector(srM3.wo)); vsrM5.wo<-var(as.vector(srM5.wo))
vsrM6.wo<-var(as.vector(srM6.wo)); vsrM7.wo<-var(as.vector(srM7.wo))
vsrM8.wo<-var(as.vector(srM8.wo))

(vsr.wo<-cbind(vsrM1.wo, vsrM2.wo, vsrM3.wo, vsrM5.wo, vsrM6.wo, vsrM7.wo,
vsrM8.wo))

#### HISTOGRAM ####
x11(); par(mfrow=c(4, 2), las=1)
hist(as.vector(t(srM1.ma)), breaks=20, main='M1 Males', xlab='z(x,t)')
hist(as.vector(t(srM2.ma)), breaks=20, main='M2 Males', xlab='z(x,t)')
hist(as.vector(t(srM3.ma)), breaks=20, main='M3 Males', xlab='z(x,t)')
hist(as.vector(t(srM5.ma)), breaks=20, main='M5 Males', xlab='z(x,t)')
hist(as.vector(t(srM6.ma)), breaks=20, main='M6 Males', xlab='z(x,t)')
hist(as.vector(t(srM7.ma)), breaks=20, main='M7 Males', xlab='z(x,t)')
hist(as.vector(t(srM8.ma)), breaks=20, main='M8 Males', xlab='z(x,t)')

x11(); par(mfrow=c(4, 2), las=1)
hist(as.vector(t(srM1.wo)), breaks=20, main='M1 Females', xlab='z(x,t)')
hist(as.vector(t(srM2.wo)), breaks=20, main='M2 Females', xlab='z(x,t)')
hist(as.vector(t(srM3.wo)), breaks=20, main='M3 Females', xlab='z(x,t)')
hist(as.vector(t(srM5.wo)), breaks=20, main='M5 Females', xlab='z(x,t)')
hist(as.vector(t(srM6.wo)), breaks=20, main='M6 Females', xlab='z(x,t)')
hist(as.vector(t(srM7.wo)), breaks=20, main='M7 Females', xlab='z(x,t)')
hist(as.vector(t(srM8.wo)), breaks=20, main='M8 Females', xlab='z(x,t)')

#### CONTOUR PLOT ####
x11(); par(las=1)
levelplot(t(srM1.ma), colorkey=list(space='top', col.regions=terrain.colors),
xlab='Age', ylab='Years', col.regions=terrain.colors, main='z(x,t) LC
Males', pretty=T, aspect='fill')
x11(); par(las=1)
levelplot(t(srM2.ma), colorkey=list(space='top', col.regions=terrain.colors),
xlab='Age', ylab='Years', col.regions=terrain.colors, main='z(x,t) RH
Males', pretty=T, aspect='fill')
x11(); par(las=1)
levelplot(t(srM3.ma), colorkey=list(space='top', col.regions=terrain.colors),
xlab='Age', ylab='Years', col.regions=terrain.colors, main='z(x,t) APC
Males', pretty=T, aspect='fill')

x11(); par(las=1)
levelplot(t(srM1.wo), colorkey=list(space='top', col.regions=terrain.colors),
xlab='Age', ylab='Years', col.regions=terrain.colors, main='z(x,t) LC
Females', pretty=T, aspect='fill')
x11(); par(las=1)
levelplot(t(srM2.wo), colorkey=list(space='top', col.regions=terrain.colors),
xlab='Age', ylab='Years', col.regions=terrain.colors, main='z(x,t) RH
Females', pretty=T, aspect='fill')
x11(); par(las=1)
levelplot(t(srM3.wo), colorkey=list(space='top', col.regions=terrain.colors),
xlab='Age', ylab='Years', col.regions=terrain.colors, main='z(x,t) APC
Females', pretty=T, aspect='fill')

```

```

#####
##### NESTED MODELS #####

#### DEGREES OF FREEDOM ####
glresM1<-resM1.wo$npar; glresM2<-resM2.wo$npar; glresM3<-resM3.wo$npar;
  glresM5<-resM5.wo$npar; glresM6<-resM6.wo$npar; glresM7<-resM7.wo$npar;
  glresM8<-resM8.wo$npar

#### CHI-SQUARE ####
alpha<-0.95
qchisM1_M2<-qchisq(alpha, (glresM2-glresM1)); qchisM3_M2<-qchisq(alpha,
  (glresM2-glresM3)); qchisM5_M6<-qchisq(alpha, (glresM6-glresM5));
qchisM5_M7<-qchisq(alpha, (glresM7-glresM5)); qchisM6_M7<-qchisq(alpha,
  (glresM7-glresM6)); qchisM5_M8<-qchisq(alpha, (glresM8-glresM5));
qchisM6_M8<-qchisq(alpha, (glresM8-glresM6))

#### LIKELIHOOD - RATIO TEST ####
(llrtM1_M2.ma<- 2*abs((resM2.ma$ll-resM1.ma$ll)))
(llrtM3_M2.ma<- 2*abs((resM2.ma$ll-resM3.ma$ll)))
(llrtM5_M6.ma<- 2*abs((resM6.ma$ll-resM5.ma$ll)))
(llrtM5_M7.ma<- 2*abs((resM7.ma$ll-resM5.ma$ll)))
(llrtM6_M7.ma<- 2*abs((resM7.ma$ll-resM6.ma$ll)))
(llrtM5_M8.ma<- 2*abs((resM8.ma$ll-resM5.ma$ll)))
(llrtM6_M8.ma<- 2*abs((resM8.ma$ll-resM6.ma$ll)))

(pvalM1_M2<-pchisq(llrtM1_M2.ma, (glresM2-glresM1), lower.tail = FALSE))
(pvalM3_M2<-pchisq(llrtM3_M2.ma, (glresM2-glresM3), lower.tail = FALSE))
(pvalM5_M6<-pchisq(llrtM5_M6.ma, (glresM6-glresM5), lower.tail = FALSE))
(pvalM5_M7<-pchisq(llrtM5_M7.ma, (glresM7-glresM5), lower.tail = FALSE))
(pvalM6_M7<-pchisq(llrtM6_M7.ma, (glresM7-glresM6), lower.tail = FALSE))
(pvalM5_M8<-pchisq(llrtM5_M8.ma, (glresM8-glresM5), lower.tail = FALSE))
(pvalM6_M8<-pchisq(llrtM6_M8.ma, (glresM8-glresM6), lower.tail = FALSE))

nested.ma<-matrix(c(llrtM1_M2.ma, llrtM3_M2.ma, llrtM5_M6.ma , llrtM5_M7.ma ,
  llrtM6_M7.ma , llrtM5_M8.ma, llrtM6_M8.ma,
qchisM1_M2, qchisM3_M2, qchisM5_M6, qchisM5_M7, qchisM6_M7, qchisM5_M8,
  qchisM6_M8,
pvalM1_M2, pvalM3_M2, pvalM5_M6, pvalM5_M7, pvalM6_M7, pvalM5_M8,
  pvalM6_M8), nrow= 7, ncol= 3)

(llrtM1_M2.wo<- 2*abs((resM2.wo$ll-resM1.wo$ll)))
(llrtM3_M2.wo<- 2*abs((resM2.wo$ll-resM3.wo$ll)))
(llrtM5_M6.wo<- 2*abs((resM6.wo$ll-resM5.wo$ll)))
(llrtM5_M7.wo<- 2*abs((resM7.wo$ll-resM5.wo$ll)))
(llrtM6_M7.wo<- 2*abs((resM7.wo$ll-resM6.wo$ll)))
(llrtM5_M8.wo<- 2*abs((resM8.wo$ll-resM5.wo$ll)))
(llrtM6_M8.wo<- 2*abs((resM8.wo$ll-resM6.wo$ll)))

(pvalM1_M2<-pchisq(llrtM1_M2.wo, (glresM2-glresM1), lower.tail = FALSE))
(pvalM3_M2<-pchisq(llrtM3_M2.wo, (glresM2-glresM3), lower.tail = FALSE))
(pvalM5_M6<-pchisq(llrtM5_M6.wo, (glresM6-glresM5), lower.tail = FALSE))
(pvalM5_M7<-pchisq(llrtM5_M7.wo, (glresM7-glresM5), lower.tail = FALSE))
(pvalM6_M7<-pchisq(llrtM6_M7.wo, (glresM7-glresM6), lower.tail = FALSE))

```

```

(pvalM5_M8<-pchisq(llrtM5_M8.wo, (glresM8-glresM5), lower.tail = FALSE))
(pvalM6_M8<-pchisq(llrtM6_M8.wo, (glresM8-glresM6), lower.tail = FALSE))

nested.wo<-matrix(c(llrtM1_M2.wo, llrtM3_M2.wo, llrtM5_M6.wo , llrtM5_M7.wo ,
  llrtM6_M7.wo , llrtM5_M8.wo, llrtM6_M8.wo,
qchisM1_M2, qchisM3_M2, qchisM5_M6, qchisM5_M7, qchisM6_M7, qchisM5_M8,
qchisM6_M8,
pvalM1_M2, pvalM3_M2, pvalM5_M6, pvalM5_M7, pvalM6_M7, pvalM5_M8,
pvalM6_M8), nrow= 7, ncol= 3)

colnames(nested.ma)<-colnames(nested.wo)<-c('LR test', 'chi', 'p-value')
rownames(nested.ma)<-rownames(nested.wo)<-c('M1_M2', 'M3_M2', 'M5_M6', 'M5_M7',
  'M6_M7', 'M5_M8', 'M6_M8')
nested.ma; nested.wo

library(xtable); xtable(.)

```

Modelo relacional Brass-type lineal

- Modelo relacional Brass-type: regresión lineal entre las tasas de mortalidad población mexicana y tasas de mortalidad de los asegurados CNSF,
- Tasas de mortalidad calculadas usando los parámetros estimados en la población suiza.

Script

```

#####
##### ASSURED LIVES #####

insure<-read.table("cnsf_i.csv", header=TRUE, sep="," , na.strings="NA",
  dec=".")
rownames(insure)<-insure[,1]
insure<-insure[-89,]

#### GRAPH ####
x11();par(las=1)
plot(ages[(fa+1):(la+1)], log(resM1.ma$mhat[py,(fa+1):(la+1)]), type='l',
  lwd=2, col=color[1], ylim=c(-9,-1), ylab='Log Mortality Rate',
  xlab=paste('Age', fa, '-', la))
lines(ages[(fa+1):(la+1)], log(resM1.wo$mhat[py,(fa+1):(la+1)]),
  col=color[10], lwd=2)
lines(ages[(fa+1):(la+1)], log(insure[,2]), col=color[17], lwd=2)
legend(8,-2,c('Males LC', 'Females LC', 'Assured lives'),
  col=c(color[1], color[10], color[17]), lwd=2, bty='n')

#####
##### LINEAR REGRESSION ASSURED ~ MEXICAN POPULATION #####

```

```

#### FROM AGE 12 TO 99; YEAR 2000 ####
fa<-12; la<-99; py<-11 #2000 YEAR POSITION

#### WEIGHTS ####
w<-seq(0,1,.05)
rw<-mat.or.vec(na,1) #RECALCULATED RATES

##### WITH LC RATES #####
lm.M1<-mat.or.vec(length(w),py)
rownames(lm.M1)<-w
colnames(lm.M1)<-c('beta0','beta1','pvb0','pvb1','rs','ars','rstd',
                 'stdb0','stdb1','BIC','loglik')
res.lm.M1<-fit.lm.M1<-mat.or.vec(length(ages[(fa+1):(la+1)]),length(w))
rownames(res.lm.M1)<-rownames(fit.lm.M1)<-ages[(fa+1):(la+1)]
colnames(res.lm.M1)<-colnames(fit.lm.M1)<-w

#### WEIGHTED LINEAR REGRESSION ####
for(i in 1:length(w)){
  rw<-log(resM1.ma$mhat[py,(fa+1):(la+1)]*(1-w[i]) +
          log(resM1.wo$mhat[py,(fa+1):(la+1)]*w[i])

  lm.M1[i,1]<-round(summary(lm(log(insure[,2])~rw))$coefficients[1],4)
  lm.M1[i,2]<-round(summary(lm(log(insure[,2])~rw))$coefficients[2],4)

  lm.M1[i,3]<-round(summary(lm(log(insure[,2])~rw))$coefficients[7],4)
  lm.M1[i,4]<-round(summary(lm(log(insure[,2])~rw))$coefficients[8],4)

  lm.M1[i,5]<-round(summary(lm(log(insure[,2])~rw))$r.squared,4)
  lm.M1[i,6]<-round(summary(lm(log(insure[,2])~rw))$adj.r.squared,4)

  lm.M1[i,7]<-round(summary(lm(log(insure[,2])~rw))$sigma,4)

  lm.M1[i,8]<-round(summary(lm(log(insure[,2])~rw))$coefficients[3],4)
  lm.M1[i,9]<-round(summary(lm(log(insure[,2])~rw))$coefficients[4],4)

  lm.M1[i,10]<-round(BIC(lm(log(insure[,2])~rw))[1],4)
  lm.M1[i,11]<-round(logLik(lm(log(insure[,2])~rw))[1],4)

  fit.lm.M1[,i]<-round(fitted(lm(log(insure[,2])~rw)),4)
  res.lm.M1[,i]<-round(residuals(lm(log(insure[,2])~rw)),4)
  print(summary(lm(log(insure[,2])~rw)))
}
lm.M1

#### FITTED VS. OBSERVED ####
x11(); par(las=1)
plot(fit.lm.M1[,1]~log(insure[,2]), type='l', lwd=2, col=color[1],
     ylim=c(min(as.vector(log(insure[,2])), as.vector(fit.lm.M1[,1])),
            max(as.vector(log(insure[,2])), as.vector(fit.lm.M1[,1]))),
     xlim=c(min(as.vector(log(insure[,2])), as.vector(fit.lm.M1[,1])),
            max(as.vector(log(insure[,2])), as.vector(fit.lm.M1[,1]))),
     xlab='Log mortality rate insurance (CNSF)', ylab='Log mortality rate
         adjusted by Brass Type model')
for(i in c(6, 11, 16, 21)){

```

```

        lines(fit.lm.M1[,i]~log(insure[,2]), col=color[i-2], lwd=2)
    }
abline(0, 1, lwd=2, lty=2)
legend(-8, -2, c('0% F', '25% F', '50% F', '75% F', '100% F'),
       col=c(color[1],color[4],color[9],color[14],color[19]), lwd=2, bty='n')

##### WITH RH RATES #####
lm.M2<-mat.or.vec(length(w),py)
rownames(lm.M2)<-w
colnames(lm.M2)<-c('beta0','beta1','pvb0','pvb1','rs','ars','rstd',
                  'stdb0','stdb1','BIC','loglik')
res.lm.M2<-fit.lm.M2<-mat.or.vec(length(ages[(fa+1):(la+1)]),length(w))
rownames(res.lm.M2)<-rownames(fit.lm.M2)<-ages[(fa+1):(la+1)]
colnames(res.lm.M2)<-colnames(fit.lm.M2)<-w

#### WEIGHTED LINEAR REGRESSION ####
for(i in 1:length(w)){
  rw<-log(resM2.ma$mhat[py,(fa+1):(la+1)])*(1-w[i]) +
    log(resM2.wo$mhat[py,(fa+1):(la+1)])*w[i]

  lm.M2[i,1]<-round(summary(lm(log(insure[,2])~rw))$coefficients[1],4)
  lm.M2[i,2]<-round(summary(lm(log(insure[,2])~rw))$coefficients[2],4)

  lm.M2[i,3]<-round(summary(lm(log(insure[,2])~rw))$coefficients[7],4)
  lm.M2[i,4]<-round(summary(lm(log(insure[,2])~rw))$coefficients[8],4)

  lm.M2[i,5]<-round(summary(lm(log(insure[,2])~rw))$r.squared,4)
  lm.M2[i,6]<-round(summary(lm(log(insure[,2])~rw))$adj.r.squared,4)

  lm.M2[i,7]<-round(summary(lm(log(insure[,2])~rw))$sigma,4)

  lm.M2[i,8]<-round(summary(lm(log(insure[,2])~rw))$coefficients[3],4)
  lm.M2[i,9]<-round(summary(lm(log(insure[,2])~rw))$coefficients[4],4)

  lm.M2[i,10]<-round(BIC(lm(log(insure[,2])~rw))[1],4)
  lm.M2[i,11]<-round(logLik(lm(log(insure[,2])~rw))[1],4)

  fit.lm.M2[,i]<-round(fitted(lm(log(insure[,2])~rw)),4)
  res.lm.M2[,i]<-round(residuals(lm(log(insure[,2])~rw)),4)
  print(summary(lm(log(insure[,2])~rw)))
}
lm.M2

##### WITH APC RATES #####
lm.M3<-mat.or.vec(length(w),py)
rownames(lm.M3)<-w
colnames(lm.M3)<-c('beta0','beta1','pvb0','pvb1','rs','ars','rstd',
                  'stdb0','stdb1','BIC','loglik')
res.lm.M3<-fit.lm.M3<-mat.or.vec(length(ages[(fa+1):(la+1)]),length(w))
rownames(res.lm.M3)<-rownames(fit.lm.M3)<-ages[(fa+1):(la+1)]
colnames(res.lm.M3)<-colnames(fit.lm.M3)<-w

#### WEIGHTED LINEAR REGRESSION ####
for(i in 1:length(w)){

```

```

rw<-log(resM3.ma$mhat[py,(fa+1):(la+1)])*(1-w[i]) +
  log(resM3.wo$mhat[py,(fa+1):(la+1)])*w[i]

lm.M3[i,1]<-round(summary(lm(log(insure[,2])~rw))$coefficients[1],4)
lm.M3[i,2]<-round(summary(lm(log(insure[,2])~rw))$coefficients[2],4)

lm.M3[i,3]<-round(summary(lm(log(insure[,2])~rw))$coefficients[7],4)
lm.M3[i,4]<-round(summary(lm(log(insure[,2])~rw))$coefficients[8],4)

lm.M3[i,5]<-round(summary(lm(log(insure[,2])~rw))$r.squared,4)
lm.M3[i,6]<-round(summary(lm(log(insure[,2])~rw))$adj.r.squared,4)

lm.M3[i,7]<-round(summary(lm(log(insure[,2])~rw))$sigma,4)

lm.M3[i,8]<-round(summary(lm(log(insure[,2])~rw))$coefficients[3],4)
lm.M3[i,9]<-round(summary(lm(log(insure[,2])~rw))$coefficients[4],4)

lm.M3[i,10]<-round(BIC(lm(log(insure[,2])~rw))[1],4)
lm.M3[i,11]<-round(logLik(lm(log(insure[,2])~rw))[1],4)

fit.lm.M3[,i]<-round(fitted(lm(log(insure[,2])~rw)),4)
res.lm.M3[,i]<-round(residuals(lm(log(insure[,2])~rw)),4)
print(summary(lm(log(insure[,2])~rw)))
}
lm.M3

#####
##### RATES FITTED, USING SWISS COHEFICIENTS #####

fit.lm.M3_sw<-fit.lm.M2_sw<-fit.lm.M1_sw<-
  mat.or.vec(length(ages[(fa+1):(la+1)]), length(w))
colnames(fit.lm.M3_sw)<-colnames(fit.lm.M2_sw)<-colnames(fit.lm.M1_sw)<- w
rownames(fit.lm.M3_sw)<-rownames(fit.lm.M2_sw)<-rownames(fit.lm.M1_sw)<-
  ages[(fa+1):(la+1)]

for(i in 1:length(w)){
  rw1<-log(resM1.ma$mhat[py,(fa+1):(la+1)])*(1-w[i]) +
    log(resM1.wo$mhat[py,(fa+1):(la+1)])*w[i]
  fit.lm.M1_sw[,i] <- -0.3197 + 1.0747*rw1

  rw2<-log(resM2.ma$mhat[py,(fa+1):(la+1)])*(1-w[i]) +
    log(resM2.wo$mhat[py,(fa+1):(la+1)])*w[i]
  fit.lm.M2_sw[,i] <- -0.3197 + 1.0747*rw2

  rw3<-log(resM3.ma$mhat[py,(fa+1):(la+1)])*(1-w[i]) +
    log(resM3.wo$mhat[py,(fa+1):(la+1)])*w[i]
  fit.lm.M3_sw[,i] <- -0.3197 + 1.0747*rw3
}

#### GRAPH ####
x11();par(las=1)
plot(ages[(fa+1):(la+1)], fit.lm.M1_sw[,1], type='l', lwd=2, col=color[1],
  ylim=c(-9,-1), ylab='Log Mortality Rate: -0.3197 + 1.0747 log(m(x))',
  xlab=paste('Edad', fa, '-', la))

```



```

lines(ages[(fa+1):(la+1)], fit.lm.MI_sw[,21], col=color[10], lwd=2)
lines(ages[(fa+1):(la+1)], log(insure[,2]), col=color[17], lwd=2)
legend(8,-2,c('Males LC','Females LC','Assured lives'),
      col=c(color[1],color[10],color[17]), lwd=2, bty='n')

#### FITTED BETA SW VS. FITTED BRASS-TYPE ####
x11();par(las=1)
plot(fit.lm.MI_sw[,1]~fit.lm.MI[,1], type='l', lwd=2, col=color[1],
      ylim=c(-9,-1), xlim=c(-9,-1), ylab='Log Mortality Rate: -0.3197 +
      1.0747 log m(x)',
      xlab='Log Mortality Rate - Brass-Type Weighted', lty=2)
for(i in 1:length(w)){
  lines(fit.lm.MI_sw[,i]~fit.lm.MI[,i], col=color[i], lwd=2, lty=4)
}
abline(0,1, lwd=2, lty=3)
legend(-8,-2, c('Males', 'Females'), lwd=2, bty='n', lty=c(4,5),
      col=c(color[1],color[21]))

```

Modelo relacional Brass-type lineal y Poisson.

Datos de la AMIS

- Modelo Brass-type: regresión lineal y regresión Poisson entre las tasas de mortalidad población mexicana y las tablas de la Asociación Mexicana de Instituciones de Seguros (AMIS).

Script Lineal

```

#####
##### ASSURED LIVES AMIS-AMA #####

insure2<-read.table(file='amis.csv', header=T, sep=',')

#### GRAPH ####
x11(); par(las=1)
plot(ages, log(insure2[, 'ma00']), type='l', lwd=2, col=2, ylim=c(-10, 0),
      ylab='Log Mortality Rate', xlab='Age')
lines(ages, log(insure2[, 'ma05']), lwd=2, col=3)
lines(ages, log(insure2[, 'ma10']), lwd=2, col=4)
lines(ages, log(insure2[, 'wo00']), lwd=2, col=5, lty=2)
lines(ages, log(insure2[, 'wo05']), lwd=2, col=6, lty=2)
lines(ages, log(insure2[, 'wo10']), lwd=2, col=8, lty=2)
legend('topleft', c('Male 2000', 'Male 2005', 'Male 2010', 'Female 2000', 'Female
2005', 'Female 2010'), col=c(2,3,4,5,6,8), lwd=2, bty='n')

##### MATRIX OF RATES #####
ma00<-cbind(ages, c(rep(0,12),insure[,2]), insure2[,2], raw.ma[,11],
  resM1.ma$mhat[11,], resM2.ma$mhat[11,], resM3.ma$mhat[11,],
  resM5.ma$mhat[11,],
  resM6.ma$mhat[11,],resM7.ma$mhat[11,],resM8.ma$mhat[11,])

```

```

ma05<-cbind(ages ,c(rep(0,12),insure[,2]), insure2[,4], raw.ma[,16],
  resM1.ma$mhat[16,], resM2.ma$mhat[16,], resM3.ma$mhat[16,],
  resM5.ma$mhat[16,],
  resM6.ma$mhat[16,],resM7.ma$mhat[16,],resM8.ma$mhat[16,])
ma10<-cbind(ages ,c(rep(0,12),insure[,2]), insure2[,6], raw.ma[,21],
  resM1.ma$mhat[21,], resM2.ma$mhat[21,], resM3.ma$mhat[21,],
  resM5.ma$mhat[21,],
  resM6.ma$mhat[21,],resM7.ma$mhat[21,],resM8.ma$mhat[21,])

wo00<-cbind(ages ,c(rep(0,12),insure[,2]), insure2[,3], raw.wo[,11],
  resM1.wo$mhat[11,], resM2.wo$mhat[11,], resM3.wo$mhat[11,],
  resM5.wo$mhat[11,], resM6.wo$mhat[11,], resM7.wo$mhat[11,],
  resM8.wo$mhat[11,])
wo05<-cbind(ages ,c(rep(0,12),insure[,2]), insure2[,5], raw.wo[,16],
  resM1.wo$mhat[16,], resM2.wo$mhat[16,], resM3.wo$mhat[16,],
  resM5.wo$mhat[16,], resM6.wo$mhat[16,], resM7.wo$mhat[16,],
  resM8.wo$mhat[16,])
wo10<-cbind(ages ,c(rep(0,12),insure[,2]), insure2[,7], raw.wo[,21],
  resM1.wo$mhat[21,], resM2.wo$mhat[21,], resM3.wo$mhat[21,],
  resM5.wo$mhat[21,], resM6.wo$mhat[21,], resM7.wo$mhat[21,],
  resM8.wo$mhat[21,])

colnames(wo10)<- colnames(wo05)<- colnames(wo00)<- colnames(ma10)<-
  colnames(ma05)<- colnames(ma00)<-
c('age', 'cnsf', 'amis', 'raw', 'm1', 'm2', 'm3', 'm5', 'm6', 'm7', 'm8')

#### COMPARING INSURANCE RATES AND REAL RATES ####
x11(); par(las=1)
plot(ages, log(ma10[, 'cnsf']), type='l', lwd=2, col=color[1], ylim=c(-9,0),
  xlab='Ages', ylab='Log mortality')
lines(ages, log(ma10[, 'amis']), col=color[20], lwd=2)
lines(ages, t(lograw.ma)[, 'y2010'], col=color[12], lwd=2)
legend('topleft', c('CNSF', 'AMIS', 'Male Population'),
  col=c(color[1], color[20], color[12]), lwd=2, bty='n')
title('Year 2010')

x11(); par(las=1)
plot(ages, log(wo10[, 'cnsf']), type='l', lwd=2, col=color[1], ylim=c(-9,0),
  xlab='Ages', ylab='Log mortality')
lines(ages, log(wo10[, 'amis']), col=color[20], lwd=2)
lines(ages, t(lograw.wo)[, 'y2010'], col=color[12], lwd=2)
legend('topleft', c('CNSF', 'AMIS', 'Female Population'),
  col=c(color[1], color[20], color[12]), lwd=2, bty='n')
title('Year 2010')

#### COMPARING INSURANCE RATES AND FITTED RATES LC-RH-APC ####
x11(); par(las=1)
plot(ages, log(ma10[, 'amis']), type='l', lwd=2, col=color[1], ylim=c(-9,0),
  xlab='Ages', ylab='Log mortality')
lines(ages, log(ma10[, 'm1']), col=color[7], lwd=2)
lines(ages, log(ma10[, 'm2']), col=color[13], lwd=2)
lines(ages, log(ma10[, 'm3']), col=color[18], lwd=2)
lines(ages, t(lograw.ma)[, 'y2010'], col=2, lwd=2, lty=2)

```

```

legend('topleft', c('AMIS', 'LC', 'RH', 'APC', 'Male population'),
      col=c(color[1], color[7], color[13], color[18], 2), lwd=2, bty='n')
title('Year 2010')

#### COMPARING INSURANCE RATES AND FITTED RATES CBD MODELS ####
x11()
plot(ages, log(ma10[, 'amis']), type='l', lwd=2, col=color[1], ylim=c(-11,0),
     xlab='Ages', ylab='Log mortality')
lines(ages, log(ma10[, 'm5']), col=color[7], lwd=2)
lines(ages, log(ma10[, 'm6']), col=color[13], lwd=2)
lines(ages, log(ma10[, 'm7']), col=color[18], lwd=2)
lines(ages, log(ma10[, 'm8']), col=color[12], lwd=2)
lines(ages, t(lograw.ma)[, 'y2010'], col=2, lwd=2, lty=2)
legend('topleft', c('AMIS', 'CBD', 'CBD2', 'CBD2.1', 'CBD3', 'Male population'),
      col=c(color[1], color[7], color[13], color[18], color[12], 2), lwd=2, bty='n')
title('Year 2010')

#####
##### LINEAR REGRESSION ASSURED AMIS 2010 ~ MEXICAN POPULATION #####
#### FROM AGE 12 TO 99; YEAR 2000 ####
fa<-12; la<-99

#### MATRIX WITH RESULTS ####
lm.Ma.a10<-lm.Wo.a10<-mat.or.vec(8,11)
rownames(lm.Ma.a10)<-rownames(lm.Wo.a10)<-
  c('Real', 'LC', 'RH', 'APC', 'CBD', 'CBD2', 'CBD2.1', 'CBD3')
colnames(lm.Ma.a10)<-colnames(lm.Wo.a10)<-
  c('beta0', 'beta1', 'pvb0', 'pvb1', 'rs', 'ars', 'rstd', 'stdb0', 'stdb1',
    'BIC', 'loglik')
fit.lm.Ma.a10<-fit.lm.Wo.a10<-mat.or.vec(na,8)
rownames(fit.lm.Ma.a10)<-rownames(fit.lm.Wo.a10)<-seq(fa, la, 1)
colnames(fit.lm.Ma.a10)<-colnames(fit.lm.Wo.a10)<-
  c('Real', 'LC', 'RH', 'APC', 'CBD', 'CBD2', 'CBD2.1', 'CBD3')
res.lm.Ma.a10<-res.lm.Wo.a10<-mat.or.vec(na,8)
rownames(res.lm.Ma.a10)<-rownames(res.lm.Wo.a10)<-seq(fa, la, 1)
colnames(res.lm.Ma.a10)<-colnames(res.lm.Wo.a10)<-
  c('Real', 'LC', 'RH', 'APC', 'CBD', 'CBD2', 'CBD2.1', 'CBD3')

#### LINEAR REGRESSION ALL MODELS####
for(j in 4:11){
  i=j-3
  ## MALES ##
  lm.Ma.a10[i,1]<-lm(log(ma10[fa:la+1, 'amis'])~log(ma10[fa:la+1, j]))$coef[1]
  lm.Ma.a10[i,2]<-lm(log(ma10[fa:la+1, 'amis'])~log(ma10[fa:la+1, j]))$coef[2]

  lm.Ma.a10[i,3]<-summary(lm(log(ma10[fa:la+1, 'amis'])~log(ma10[fa:la+1, j])))
    )$coef[7]
  lm.Ma.a10[i,4]<-summary(lm(log(ma10[fa:la+1, 'amis'])~log(ma10[fa:la+1, j])))
    )$coef[8]

  lm.Ma.a10[i,5]<-summary(lm(log(ma10[fa:la+1, 'amis'])~log(ma10[fa:la+1, j])))
    )$r.squared
  lm.Ma.a10[i,6]<-summary(lm(log(ma10[fa:la+1, 'amis'])~log(ma10[fa:la+1, j])))
    )$adj.r.squared
}

```

```

lm.Ma.a10[i,7]<-summary(lm(log(ma10[fa:la+1,'amis'])~log(ma10[fa:la+1,j]))
)$sigma

lm.Ma.a10[i,8]<-summary(lm(log(ma10[fa:la+1,'amis'])~log(ma10[fa:la+1,j]))
)$coef[3]
lm.Ma.a10[i,9]<-summary(lm(log(ma10[fa:la+1,'amis'])~log(ma10[fa:la+1,j]))
)$coef[4]

lm.Ma.a10[i,10]<-BIC(lm(log(ma10[fa:la+1,'amis'])~log(ma10[fa:la+1,j]))) [1]
lm.Ma.a10[i,11]<-logLik(lm(log(ma10[fa:la+1,'amis'])~log(ma10[fa:la+1,j]))) [1]

fit.lm.Ma.a10[,i]<-fitted(lm(log(ma10[fa:la+1,'amis'])~log(ma10[fa:la+1,j])))
res.lm.Ma.a10[,i]<-residuals(lm(log(ma10[fa:la+1,'amis'])~log(ma10[fa:la+1,j])))

## FEMALES ##
lm.Wo.a10[i,1]<-lm(log(wo10[fa:la+1,'amis'])~log(wo10[fa:la+1,j]))$coef[1]
lm.Wo.a10[i,2]<-lm(log(wo10[fa:la+1,'amis'])~log(wo10[fa:la+1,j]))$coef[2]

lm.Wo.a10[i,3]<-summary(lm(log(wo10[fa:la+1,'amis'])~log(wo10[fa:la+1,j]))
)$coef[7]
lm.Wo.a10[i,4]<-summary(lm(log(wo10[fa:la+1,'amis'])~log(wo10[fa:la+1,j]))
)$coef[8]

lm.Wo.a10[i,5]<-summary(lm(log(wo10[fa:la+1,'amis'])~log(wo10[fa:la+1,j]))
)$r.squared
lm.Wo.a10[i,6]<-summary(lm(log(wo10[fa:la+1,'amis'])~log(wo10[fa:la+1,j]))
)$adj.r.squared

lm.Wo.a10[i,7]<-summary(lm(log(wo10[fa:la+1,'amis'])~log(wo10[fa:la+1,j]))
)$sigma

lm.Wo.a10[i,8]<-summary(lm(log(wo10[fa:la+1,'amis'])~log(wo10[fa:la+1,j]))
)$coef[3]
lm.Wo.a10[i,9]<-summary(lm(log(wo10[fa:la+1,'amis'])~log(wo10[fa:la+1,j]))
)$coef[4]

lm.Wo.a10[i,10]<-BIC(lm(log(wo10[fa:la+1,'amis'])~log(wo10[fa:la+1,j]))) [1]
lm.Wo.a10[i,11]<-logLik(lm(log(wo10[fa:la+1,'amis'])~log(wo10[fa:la+1,j]))) [1]

fit.lm.Wo.a10[,i]<-fitted(lm(log(wo10[fa:la+1,'amis'])~log(wo10[fa:la+1,j])))
res.lm.Wo.a10[,i]<-residuals(lm(log(wo10[fa:la+1,'amis'])~log(wo10[fa:la+1,j])))
}
lm.Ma.a10
lm.Wo.a10

```

Script Poisson

```

#####
##### ASSURED LIVES AMIS-AMA #####
#####

#USING THE OBJECTS CREATED IN THE SCRIPT BRASS_TYPE_AMIS_RL:

```

```

# MATRIX MA00, WO00, ETC (RATES MATRIX)

#####
##### POISSON REGRESSION ASSURED AMIS 2010 ~ MEXICAN POPULATION #####
#### FROM AGE 12 TO 99; YEAR 2000 ####
fa<-12; la<-99; na<-length(seq(fa,la,1))

#### MATRIX OF DEATHS CALCULATED WITH AMIS RATES ####
dman.amis<-dwo.amis<-mat.or.vec(na,3)
rownames(dman.amis)<-rownames(dwo.amis)<-seq(fa,la,1)
colnames(dman.amis)<-colnames(dwo.amis)<-c('2000','2005','2010')

dman.amis[,1]<-round(ma00[fa:la+1,3]*lman[fa:la+1,11],0)
dman.amis[,2]<-round(ma05[fa:la+1,3]*lman[fa:la+1,16],0)
dman.amis[,3]<-round(ma10[fa:la+1,3]*lman[fa:la+1,21],0)

dwo.amis[,1]<-round(wo00[fa:la+1,3]*lwoman[fa:la+1,11],0)
dwo.amis[,2]<-round(wo05[fa:la+1,3]*lwoman[fa:la+1,16],0)
dwo.amis[,3]<-round(wo10[fa:la+1,3]*lwoman[fa:la+1,21],0)

#### MATRIX WITH RESULTS ####
glmp.Ma.a10<-glmp.Wo.a10<-mat.or.vec(8,10)
rownames(glmp.Ma.a10)<-rownames(glmp.Wo.a10)<-
  c('Real','LC','RH','APC','CBD','CBD2','CBD2.1','CBD3')
colnames(glmp.Ma.a10)<-colnames(glmp.Wo.a10)<-
  c('beta0','beta1','pvb0','pvb1','deviance','aic',
    'stdb0','stdb1','BIC','loglik')
fit.glmp.Ma.a10<-fit.glmp.Wo.a10<-mat.or.vec(na,8)
rownames(fit.glmp.Ma.a10)<-rownames(fit.glmp.Wo.a10)<-seq(fa,la,1)
colnames(fit.glmp.Ma.a10)<-colnames(fit.glmp.Wo.a10)<-
  c('Real','LC','RH','APC','CBD','CBD2','CBD2.1','CBD3')
res.glmp.Ma.a10<-res.glmp.Wo.a10<-mat.or.vec(na,8)
rownames(res.glmp.Ma.a10)<-rownames(res.glmp.Wo.a10)<-seq(fa,la,1)
colnames(res.glmp.Ma.a10)<-colnames(res.glmp.Wo.a10)<-
  c('Real','LC','RH','APC','CBD','CBD2','CBD2.1','CBD3')

#### POISSON REGRESSION ALL MODELS####
for(j in 4:11){
  i=j-3
  ## MALES ##
  glmp.Ma.a10[i,1]<-glm(dman.amis[, '2010']~offset(log(lman[fa:la+1,21]))
    +log(ma10[fa:la+1,j]), family=poisson(link=log))$coef[1]
  glmp.Ma.a10[i,2]<-glm(dman.amis[, '2010']~offset(log(lman[fa:la+1,21]))
    +log(ma10[fa:la+1,j]), family=poisson(link=log))$coef[2]

  glmp.Ma.a10[i,3]<-summary(glm(dman.amis[, '2010']~offset(log(lman[fa:la+1,21]))
    +log(ma10[fa:la+1,j]), family=poisson(link=log)))$coef[7]
  glmp.Ma.a10[i,4]<-summary(glm(dman.amis[, '2010']~offset(log(lman[fa:la+1,21]))
    +log(ma10[fa:la+1,j]), family=poisson(link=log)))$coef[8]

  glmp.Ma.a10[i,5]<-summary(glm(dman.amis[, '2010']~offset(log(lman[fa:la+1,21]))
    +log(ma10[fa:la+1,j]), family=poisson(link=log)))$deviance
  glmp.Ma.a10[i,6]<-summary(glm(dman.amis[, '2010']~offset(log(lman[fa:la+1,21]))
    +log(ma10[fa:la+1,j]), family=poisson(link=log)))$aic
}

```

```

glmp.Ma.a10[i,7]<-summary(glm(dman.amis[, '2010']~offset(log(lman[fa:la+1,21]))
+log(ma10[fa:la+1,j]), family=poisson(link=log)))$coef[3]
glmp.Ma.a10[i,8]<-summary(glm(dman.amis[, '2010']~offset(log(lman[fa:la+1,21]))
+log(ma10[fa:la+1,j]), family=poisson(link=log)))$coef[4]

glmp.Ma.a10[i,9]<-BIC(glm(dman.amis[, '2010']~offset(log(lman[fa:la+1,21]))
+log(ma10[fa:la+1,j]), family=poisson(link=log)))[1]
glmp.Ma.a10[i,10]<-logLik(glm(dman.amis[, '2010']~offset(log(lman[fa:la+1,21]))
+log(ma10[fa:la+1,j]), family=poisson(link=log)))[1]

fit.glmp.Ma.a10[i]<-fitted(glm(dman.amis[, '2010']~
offset(log(lman[fa:la+1,21])) +log(ma10[fa:la+1,j]),
family=poisson(link=log)))
res.glmp.Ma.a10[i]<-residuals(glm(dman.amis[, '2010']~
offset(log(lman[fa:la+1,21])) +log(ma10[fa:la+1,j]),
family=poisson(link=log)))

## FEMALES ##
glmp.Wo.a10[i,1]<-glm(dwo.amis[, '2010']~offset(log(lwoman[fa:la+1,21]))
+log(wo10[fa:la+1,j]), family=poisson(link=log))$coef[1]
glmp.Wo.a10[i,2]<-glm(dwo.amis[, '2010']~offset(log(lwoman[fa:la+1,21]))
+log(wo10[fa:la+1,j]), family=poisson(link=log))$coef[2]

glmp.Wo.a10[i,3]<-summary(glm(dwo.amis[, '2010']~offset(log(lwoman[fa:la+1,21]))
+log(wo10[fa:la+1,j]), family=poisson(link=log)))$coef[7]
glmp.Wo.a10[i,4]<-summary(glm(dwo.amis[, '2010']~offset(log(lwoman[fa:la+1,21]))
+log(wo10[fa:la+1,j]), family=poisson(link=log)))$coef[8]

glmp.Wo.a10[i,5]<-summary(glm(dwo.amis[, '2010']~offset(log(lwoman[fa:la+1,21]))
+log(wo10[fa:la+1,j]), family=poisson(link=log)))$deviance
glmp.Wo.a10[i,6]<-summary(glm(dwo.amis[, '2010']~offset(log(lwoman[fa:la+1,21]))
+log(wo10[fa:la+1,j]), family=poisson(link=log)))$aic

glmp.Wo.a10[i,7]<-summary(glm(dwo.amis[, '2010']~offset(log(lwoman[fa:la+1,21]))
+log(wo10[fa:la+1,j]), family=poisson(link=log)))$coef[3]
glmp.Wo.a10[i,8]<-summary(glm(dwo.amis[, '2010']~offset(log(lwoman[fa:la+1,21]))
+log(wo10[fa:la+1,j]), family=poisson(link=log)))$coef[4]

glmp.Wo.a10[i,9]<-BIC(glm(dwo.amis[, '2010']~offset(log(lwoman[fa:la+1,21]))
+log(wo10[fa:la+1,j]), family=poisson(link=log)))[1]
glmp.Wo.a10[i,10]<-logLik(glm(dwo.amis[, '2010']~offset(log(lwoman[fa:la+1,21]))
+log(wo10[fa:la+1,j]), family=poisson(link=log)))[1]

fit.glmp.Wo.a10[i]<-fitted(glm(dwo.amis[, '2010']~
offset(log(lwoman[fa:la+1,21])) +log(wo10[fa:la+1,j]),
family=poisson(link=log)))
res.glmp.Wo.a10[i]<-residuals(glm(dwo.amis[, '2010']~
offset(log(lwoman[fa:la+1,21])) +log(wo10[fa:la+1,j]),
family=poisson(link=log)))
}

glmp.Ma.a10
glmp.Wo.a10

```

Cálculo de la esperanza de vida y el valor en riesgo

Con las tasas de mortalidad de la población asegurada (AMIS) y las tasas de mortalidad ajustadas según la regresión Poisson y lineal se ha creado una tabla de vida a partir de la edad 65:

- Cálculo de la esperanza de vida usando el programa Demography (Hyndman *et al.*, 2014),
- Cálculo del VaR a partir de las función de supervivencia Weibull y Log-normal,
- Gráficos.

Script

```
#####
##### LIFETABLE : AFTER AGE 65 #####
library(demography)
ratesRFit<-read.demogdata("rates_amis_fit.txt",type="mortality", label="Mex",
  skip=1)
rates.65<-extract.ages(ratesRFit, 65:99, combine.upper=F)

#### GRAPH FITTED RATES ####
x11(); par(las=1, mfrow=c(2,3))
plot(rates.65, series='maleamis', main='Male: AMIS')
plot(rates.65, series='male_poi', main='Male: Poisson')
plot(rates.65, series='male_lin', main='Male: Linear')
plot(rates.65, series='femaleamis', main='Female: AMIS')
plot(rates.65, series='female_poi', main='Female: Poisson')
plot(rates.65, series='female_lin', main='Female: Linear')

#### LIFE TABLE - LIFE EXPENCTANCY####
(lt.Ma.a.65<-lifetable(rates.65, series="maleamis"))
(lt.Wo.a.65<-lifetable(rates.65, series="femaleamis"))
(lt.Ma.poi.65<-lifetable(rates.65, series="male_poi"))
(lt.Wo.poi.65<-lifetable(rates.65, series="female_poi"))
(lt.Ma.lin.65<-lifetable(rates.65, series="male_lin"))
(lt.Wo.lin.65<-lifetable(rates.65, series="female_lin"))

#### GRAPH - LIFE EXPECTANCY ####
x11(); par(las=1)
plot(life.expectancy(rates.65, series="femaleamis",), ylab="Remaining Time",
  col=2, ylim=c(15,23), lwd=2, xaxt='n', yaxt='n', main='Life Expectancy -
  65')
lines(life.expectancy(rates.65, series="female_poi"), col=3, lwd=2)
```

```

lines(life.expectancy(rates.65, series="female_lin"),col=4, lwd=2)
lines(life.expectancy(rates.65, series="maleamis"),col=5, lwd=2)
lines(life.expectancy(rates.65, series="male_poi"),col=6, lwd=2)
lines(life.expectancy(rates.65, series="male_lin"),col=7, lwd=2)
axis(2,seq(15,23,2),lwd=1)
mtext('2000      -      2005      -      2010', 1)
legend('topright',c('Female AMIS', 'Female Fit-R.Poisson', 'Female
Fit-R.lineal', 'Male AMIS', 'Male Fit-R.Poisson', 'Male Fit-R.lineal'),
col=2:7,lwd=2,bty='n')

#### LX - SX LOOKING FOR A PARAMETRIC FIT ####
cum.haz.Ma.a.65<- -log(lt.Ma.a.65$Lx); cum.haz.Wo.a.65<-
-log(lt.Wo.a.65$Lx)
cum.haz.Ma.poi.65<- -log(lt.Ma.poi.65$Lx); cum.haz.Wo.poi.65<-
-log(lt.Wo.poi.65$Lx)
cum.haz.Ma.lin.65<- -log(lt.Ma.lin.65$Lx); cum.haz.Wo.lin.65<-
-log(lt.Wo.lin.65$Lx)

#### GRAPH ####
age65<-1:35
for( i in 1:3){ x11(); par(mfrow=c(1,3))
plot(cum.haz.Ma.poi.65[,i]~ age65, xlab = "Age", ylab = "cum hazard", pch =
19, main='Exponential')
abline(lm(cum.haz.Ma.poi.65[,i] ~ age65 - 1), col = 2, lwd = 3)
plot(log(cum.haz.Ma.poi.65[,i]) ~ log(age65), xlab = "Log(Age)", ylab =
"Log(cum hazard)", pch = 25,col='slategray3', main='Weibull')
abline(lm(log(cum.haz.Ma.poi.65[,i]) ~ log(age65)), lwd = 3, col =
'olivedrab3')
plot(qnorm(1 - exp(-cum.haz.Ma.poi.65[,i])) ~ log(age65), xlab = "Log(Age)",
ylab = "Normal quantiles 1-exp(-cum hazard)", pch = 19, ,
main='Lognormal')
abline(lm(qnorm(1- exp(-cum.haz.Ma.poi.65[,i])) ~ log(age65)), col = 2, lwd
= 3)
}

#####
##### SURVIVAL FIT #####
survfit.Ma.a.65<-survfit.Ma.poi.65<-survfit.Ma.lin.65<-mat.or.vec(3,11)
rownames(survfit.Ma.a.65)<-rownames(survfit.Ma.poi.65)
<-rownames(survfit.Ma.lin.65) <-seq(2000, 2010, 5)
colnames(survfit.Ma.a.65)<-colnames(survfit.Ma.poi.65)
<-colnames(survfit.Ma.lin.65)
<-c('ro.E', 'ro.W', 'alpha.W', 'mu.LN', 'sigma.LN', 'W 90', 'W 95', 'W 99', 'LN
90', 'LN 95', 'LN 99')

survfit.Wo.a.65<-survfit.Wo.poi.65<-survfit.Wo.lin.65<-mat.or.vec(3,11)
rownames(survfit.Wo.a.65)<-rownames(survfit.Wo.poi.65)
<-rownames(survfit.Wo.lin.65) <-seq(2000, 2010, 5)
colnames(survfit.Wo.a.65)<-colnames(survfit.Wo.poi.65)
<-colnames(survfit.Wo.lin.65)
<-c('ro.E', 'ro.W', 'alpha.W', 'mu.LN', 'sigma.LN', 'W 90', 'W 95', 'W 99', 'LN
90', 'LN 95', 'LN 99')

#### PARAMETER ESTIMATION ####

```



```

for( i in 1:3){
  survfit.Ma.a.65[i,1] <- -summary(lm(cum.haz.Ma.a.65[,i] ~ age65 - 1
  ))$coef[1]
  survfit.Ma.a.65[i,2] <- exp(summary(lm(log(cum.haz.Ma.a.65[,i]) ~
  log(age65)) )$coef[1])^(-1/summary(lm(log(cum.haz.Ma.a.65[,i]) ~
  log(age65)) )$coef[2])
  survfit.Ma.a.65[i,3] <- summary(lm(log(cum.haz.Ma.a.65[,i]) ~ log(age65))
  )$coef[2]
  survfit.Ma.a.65[i,4] <- -summary(lm(qnorm(1- exp(-cum.haz.Ma.a.65[,i]) ~
  log(age65))) $coef[1] / summary(lm(qnorm(1- exp(-cum.haz.Ma.a.65[,i])
  ~ log(age65))) $coef[2]
  )
  survfit.Ma.a.65[i,5] <- 1/summary(lm(qnorm(1- exp(-cum.haz.Ma.a.65[,i]) ~
  log(age65))) $coef[2]

  survfit.Ma.poi.65[i,1] <- -summary(lm(cum.haz.Ma.poi.65[,i] ~ age65 - 1
  ))$coef[1]
  survfit.Ma.poi.65[i,2] <- exp(summary(lm(log(cum.haz.Ma.poi.65[,i]) ~
  log(age65)) )$coef[1])^(-1/summary(lm(log(cum.haz.Ma.poi.65[,i]) ~
  log(age65)) )$coef[2])
  survfit.Ma.poi.65[i,3] <- summary(lm(log(cum.haz.Ma.poi.65[,i]) ~
  log(age65)) )$coef[2]
  survfit.Ma.poi.65[i,4] <- -summary(lm(qnorm(1-
  exp(-cum.haz.Ma.poi.65[,i]) ~ log(age65))) $coef[1] /
  summary(lm(qnorm(1- exp(-cum.haz.Ma.poi.65[,i]) ~ log(age65)))
  )$coef[2]
  survfit.Ma.poi.65[i,5] <- 1/summary(lm(qnorm(1-
  exp(-cum.haz.Ma.poi.65[,i]) ~ log(age65))) $coef[2]

  survfit.Ma.lin.65[i,1] <- -summary(lm(cum.haz.Ma.lin.65[,i] ~ age65 - 1
  ))$coef[1]
  survfit.Ma.lin.65[i,2] <- exp(summary(lm(log(cum.haz.Ma.lin.65[,i]) ~
  log(age65)) )$coef[1])^(-1/summary(lm(log(cum.haz.Ma.lin.65[,i]) ~
  log(age65)) )$coef[2])
  survfit.Ma.lin.65[i,3] <- summary(lm(log(cum.haz.Ma.lin.65[,i]) ~
  log(age65)) )$coef[2]
  survfit.Ma.lin.65[i,4] <- -summary(lm(qnorm(1-
  exp(-cum.haz.Ma.lin.65[,i]) ~ log(age65))) $coef[1] /
  summary(lm(qnorm(1- exp(-cum.haz.Ma.lin.65[,i]) ~ log(age65)))
  )$coef[2]
  survfit.Ma.lin.65[i,5] <- 1/summary(lm(qnorm(1-
  exp(-cum.haz.Ma.lin.65[,i]) ~ log(age65))) $coef[2]
}

#####
##### VAR #####

#### WEIBULL ####
survfit.Ma.a.65[,6] <- qweibull(.90, survfit.Ma.a.65[,3] , scale
= survfit.Ma.a.65[,2], lower.tail = TRUE, log.p = FALSE)
survfit.Ma.a.65[,7] <- qweibull(.95, survfit.Ma.a.65[,3] , scale
= survfit.Ma.a.65[,2], lower.tail = TRUE, log.p = FALSE)
survfit.Ma.a.65[,8] <- qweibull(.99, survfit.Ma.a.65[,3] , scale
= survfit.Ma.a.65[,2], lower.tail = TRUE, log.p = FALSE)

```

```

survfit.Ma.poi.65[,6]<-qweibull(.90, survfit.Ma.poi.65[,3] , scale
=survfit.Ma.poi.65[,2], lower.tail = TRUE, log.p = FALSE)
survfit.Ma.poi.65[,7]<-qweibull(.95, survfit.Ma.poi.65[,3] , scale
=survfit.Ma.poi.65[,2], lower.tail = TRUE, log.p = FALSE)
survfit.Ma.poi.65[,8]<-qweibull(.99, survfit.Ma.poi.65[,3] , scale
=survfit.Ma.poi.65[,2], lower.tail = TRUE, log.p = FALSE)

survfit.Ma.lin.65[,6]<-qweibull(.90, survfit.Ma.lin.65[,3] , scale
=survfit.Ma.lin.65[,2], lower.tail = TRUE, log.p = FALSE)
survfit.Ma.lin.65[,7]<-qweibull(.95, survfit.Ma.lin.65[,3] , scale
=survfit.Ma.lin.65[,2], lower.tail = TRUE, log.p = FALSE)
survfit.Ma.lin.65[,8]<-qweibull(.99, survfit.Ma.lin.65[,3] , scale
=survfit.Ma.lin.65[,2], lower.tail = TRUE, log.p = FALSE)

#### LOGNORMAL ####
survfit.Ma.a.65[,9]<-qlnorm(.90, survfit.Ma.a.65[,4] , survfit.Ma.a.65[,5],
lower.tail = TRUE, log.p = FALSE)
survfit.Ma.a.65[,10]<-qlnorm(.95, survfit.Ma.a.65[,4] , survfit.Ma.a.65[,5],
lower.tail = TRUE, log.p = FALSE)
survfit.Ma.a.65[,11]<-qlnorm(.99, survfit.Ma.a.65[,4] , survfit.Ma.a.65[,5],
lower.tail = TRUE, log.p = FALSE)

survfit.Ma.poi.65[,9]<-qlnorm(.90, survfit.Ma.poi.65[,4] ,
survfit.Ma.poi.65[,5], lower.tail = TRUE, log.p = FALSE)
survfit.Ma.poi.65[,10]<-qlnorm(.95, survfit.Ma.poi.65[,4] ,
survfit.Ma.poi.65[,5], lower.tail = TRUE, log.p = FALSE)
survfit.Ma.poi.65[,11]<-qlnorm(.99, survfit.Ma.poi.65[,4] ,
survfit.Ma.poi.65[,5], lower.tail = TRUE, log.p = FALSE)

survfit.Ma.lin.65[,9]<-qlnorm(.90, survfit.Ma.lin.65[,4] ,
survfit.Ma.lin.65[,5], lower.tail = TRUE, log.p = FALSE)
survfit.Ma.lin.65[,10]<-qlnorm(.95, survfit.Ma.lin.65[,4] ,
survfit.Ma.lin.65[,5], lower.tail = TRUE, log.p = FALSE)
survfit.Ma.lin.65[,11]<-qlnorm(.99, survfit.Ma.lin.65[,4] ,
survfit.Ma.lin.65[,5], lower.tail = TRUE, log.p = FALSE)

survfit.Ma.a.65; survfit.Ma.poi.65; survfit.Ma.lin.65

```


Código SAS

En esta sección se incluyen los scripts utilizados para obtener los resultados reportados en el capítulo 5 de esta Tesis. El lenguaje de programación empleado es SAS Institute Inc. (2003).

Los programas que se presentan a continuación están elaborados en el contexto del libro Bolancé *et al.* (2012) y de los artículos Alemany *et al.* (2013), Bolancé *et al.* (2014) y Bolancé *et al.* (2003).

Creación de la base de datos

- Lee la base de datos que contiene el número de supervivientes a cada edad x ,
- Genera una nueva base de datos repitiendo la edad x tantas veces como número de personas vivas existan a cada edad.

Script

```
libname simul '.';

/*Reading data */
proc iml;
use simul.Data;
read all var{x,hom10,muj10} into x;

/*Subset number of survivors at each age*/
create simul.alive from x [colname={'x' 'hom10'
'muj10' }];
append from x;
close simul.alive;

/*Final Data base: repeating each age many times
as the number of survivors at that age */
use simul.alive;
read all into datos;

n=nrow(datos);
age=datos[,1];
h=datos[,2];
m=datos[,3];
nh=sum(h);
nm=sum(m);

print nh nm;
matrixh=j(nh,1,0);
do j=1 to h[1];
matrixh[j]=age[1];
end;
n1=h[1];
do i=2 to n;
ni=n1;
if h[i]>0 then do;
do j=1 to h[i];
matrixh[ni+j]=age[i];
end;
n1=ni+h[i];
end;
end;

matrixm=j(nm,1,0);
do j=1 to m[1];
matrixm[j]=age[1];
end;
n1=m[1];
do i=2 to n;
ni=n1;
if m[i]>0 then do;
```

```

do j=1 to m[i];
matrixm[ni+j]=age[i];
end;
n1=ni+m[i];
end;
end;
create simul.man65 from matrixh [colname={'x'}];
append from matrixh;

close simul.man65;
create simul.woman65 from matrixm
[colname={'x'}];
append from matrixm;
close simul.woman65;
run;
QUIT;

```

Estimación de parámetros por máxima verosimilitud

- Histograma del número de personas vivas a cada edad x ,
- Estimación de los parámetros de las distribuciones Weibull, Lognormal y Champernowne por máxima verosimilitud.

Script

```

libname simul '.';

/* Parameter Estimation: Weibull & Lognormal */
proc univariate data=simul.man65;
var x;
histogram x/ midpoints =65 to 120 by 1
lognormal
weibull ;
run;
proc univariate data=simul.woman65;
var x;
histogram x/ midpoints =65 to 120 by 1
lognormal
weibull ;
run;

/* Parameter Estimation: Champernowne */
proc iml symsize=50000;

/*Reading data*/
use simul.man65;
read all into xx;

* use simul.woman65;
* read all into xx;

/* The number of variables in the sample */
nr=nrow(xx);

/* The number of sample */
nc=ncol(xx);

/* Function: Median */
start median(x);
n=nrow(x);
r=rank(x);
xsort=j(n,1,0);
do i=1 to n;
xsort[r[i]]=x[i];
end;
m=(xsort[n/2]+xsort[(n/2)+1])/2;
return(m);
finish;

/* Transformation function: Generalised Mobius
Distr. Func */
start Tcap(v,a,c,M);
n=nrow(v);
rr=1/2*(((v+j(n,1,c))##a-j(n,1,(M+c)**a))
/((v+j(n,1,c))##a+j(n,1,(M+c)**a)-
2*j(n,1,c**a))+1);
return(rr);
finish;

/* Likelihood */
start like(v,a,c,M);

```

```

N=nrow(v);
rr=N*log(a)+N*log((M+c)**a-c**a)+
  (a-1)*sum(log(v+j(n,1,c)))
  -2*sum(log((v+j(n,1,c))##a
  +j(n,1,(M+c)**a)-2*j(n,1,c**a) ));
return(rr);
finish;

/* Variables */
par=j(3,nc,0);
L=j(21,1,0);
aopt=j(nc,1,0);
x=j(nr,1,0);
jmax=j(10,1,0);
cvalg=t(do(0,0.5,0.0001));
ncvalg=nrow(cvalg);
avalg=j(ncvalg,1,0);
Lvalg=j(ncvalg,1,0);

/* Optimal Parameters */
do i=1 to nc;
  x=xx[,i];
  M=median(x);
  gridc=cvalg*M;  *print gridc;
  print M;      *print cvalg L;

/*Possible value of c */
/*Finding the optimal a for every value of c*/
do k=1 to ncvalg;
  c=cvalg[k]*M;
  *print c;

  astart=0.1;
  aslut=21;
  do decimal=1 to 3;
    do j=1 to 21;
      a=astart+(j-1)*(aslut-astart)/20;
      L[j]=like(x,a,c,M);
    end;
    *print a;
    jmax=loc(L=max(L));

    amax=astart+(jmax[1]-1)*(aslut-astart)/20;
    astart=max(0.0001,astart+
      (jmax[1]-2)*(aslut-astart)/20);
    aslut=astart+jmax[1]*(aslut-astart)/20;
  end;
  avalg[k]=amax;
  Lvalg[k]=L[jmax[1]];
end;

/*Finding the optimal a and c for each sample*/
optk=loc(Lvalg=max(Lvalg));
optk=optk[1];
par[1,i]=avalg[optk];
par[2,i]=cvalg[optk]*M;
par[3,i]=M;
end;
logLik=Lvalg[optk];
print par logLik;
print c;

QUIT;

```

Estimación de parámetros minimizando el error cuadrático medio integrado

- Estimación de los parámetros de las distribuciones Weibull, Lognormal y Champernowne minimizando \widehat{MISE} ,
- Una opción más eficiente para minimizar el \widehat{MISE} es usar la función `optimize` de R project.

Script Weibull

```

libname simul '.';

proc iml SYMSIZE=50000;

  use simul.Ec;
  read all var{emp} into x1;

```

```

use simul.Ec;
read all var{x} into x3;

use simul.alive;
read all var{hom10} where(x>=65) into x2;
*read all var{muj10} where(x>=65) into x2;

/* Grid Points */
min=min(x3);
last=max(loc(x2>0));
gv=t(do(min,min+last-1,1));
ngrid=nrow(gv);

/* Parameters Male population*/
/* Weibull */
*s= 79.88856862 ; *MLE scale;
*a= 9.70099632 ; *MLE shape;

/* Parameters Female population*/
/* Weibull */
*s= 80.71985568 ; *MLE scale;
*a= 9.4868903 ; *MLE shape;

/* Moving MISE parameters */
aa=T(do(9, 11 , .01));
na=nrow(aa);

ss=T(do(78, 80, .01));
*ss=s;
ns=nrow(ss);

/* Variables */
n=sum(x2);
suma=j(na*ns,3,0);
r=1;

xy=j(ngrid,1,0);
fx=j(ngrid,1,0);

/* min MISE, finding optimal parameters */
do k= 1 to ns;
do j=1 to na ;

```

```

fx=cdf('WEIBULL',gv,aa[j],ss[k]);
do i=1 to ngrid;
xy[i,1]=((fx[i,1]-x1[i,])**2)*x2[i,];

*xy[i,1]=(((fx[i,1]-x1[i,])**2)*x2[i,])*gv[i];

*xy[i,1]=(((fx[i,1]-x1[i,])**2)*x2[i,])*gv[i]**2;

*xy[i,1]=(((fx[i,1]-x1[i,])**2)*x2[i,])*sqrt(gv[i]);
end;
suma[r,1]=ss[k];
suma[r,2]=aa[j];
suma[r,3]=xy[+,1];
r=r+1;
end;
end;
pos=loc(suma[,3]=min(suma[,3]));
min=min(suma[,3]);

/* Optimal parameters */
a=suma[pos,2];
s=suma[pos,1];
print a s min;

/* Graph a parameter vs. min suma */
res= suma[,2] || suma[,3];
MATTRIB res
colname = {'shape', 'MISE' };
call pgraf (res, '+', 'shape', 'MISE');

/* Graph s parameter vs. min suma */
res= suma[,1] || suma[,3];
MATTRIB res
colname = {'shape', 'MISE' };
call pgraf (res, '+', 'shape', 'MISE');

/* Matrix of parameters - MISE */
create simul.sumawei from suma [colname =
{'ss','aa','suma' }];
append from suma;
close simul.sumawei;

QUIT;

```

Script Lognormal

```

libname simul '?';

proc iml SYMSIZE=50000;

use simul.Ec;

```

```

read all var{emp} into x1;

use simul.Ec;
read all var{x} into x3;

```

```

use simul.alive;
read all var{hom10} where(x >=65) into x2;
*read all var{muj10} where(x>=65) into x2;

/* Grid Points */
min=min(x3);
last=max(loc(x2>0));
gv=t(do(min,min+last-1,1));
ngrid=nrow(gv);

/* Parameters Male population*/
/* Lognormal */
*mu= 4.328013635; *MLE scale;
*sigma= 0.102264784; *MLE shape;

/* Parameters Female population*/
/* Lognormal */
*mu= 4.33669801; *MLE scale;
*sigma= 0.105999045; *MLE shape;

/* Moving MISE parameters */
aa=T(do(.1, .13, .001));
na=nrow(aa);

ss=T(do(4.3,4.33,.001));
*ss=mu;
ns=nrow(ss);

/* Variables */
n=sum(x2);
suma=j(na*ns,3,0);
r=1;

xy=j(ngrid,1,0);
fx=j(ngrid,1,0);

/* min MISE, finding optimal parameters */
do k=1 to ns;
do j=1 to na ;
fx=cdf('LOGNORMAL',gv,ss[k],aa[j]);
do i=1 to ngrid;

```

```

xy[i,1]=((fx[i,1]-x1[i,])**2)*x2[i,];
*xy[i,1]=(((fx[i,1]-x1[i,])**2)*x2[i,])*gv[i];
*xy[i,1]=(((fx[i,1]-x1[i,])**2)*x2[i,])*gv[i]**2;
*xy[i,1]=(((fx[i,1]-x1[i,])**2)*x2[i,])*sqrt(gv[i]);
end;
suma[r,1]=ss[k];
suma[r,2]=aa[j];
suma[r,3]=xy[+,1];
r=r+1;
end;
end;
pos=loc(suma[,3]=min(suma[,3]));
min=min(suma[,3]);

/* Optimal parameters */
a=suma[pos,2];
s=suma[pos,1];
print a s min;

/* Graph a parameter vs. min suma */
res= suma[,2] || suma[,3];
MATTRIB res
colname = {'shape', 'MISE' };
call pgraf (res, '+', 'shape', 'MISE');

/* Graph s parameter vs. min suma */
res= suma[,1] || suma[,3];
MATTRIB res
colname = {'shape', 'MISE' };
call pgraf (res, '+', 'shape', 'MISE');

/* Graph s parameter vs. min suma */
create simul.sumaln from suma [colname =
{'ss','aa','suma' }];
append from suma;
close simul.sumaln;

QUIT;

```

Script Champernowne

```

libname simul ' ';

proc iml SYMSIZE=50000;

use simul.Ec;
read all var{emp} into x1;

```

```

use simul.Ec;
read all var{x} into x3;

use simul.alive;
read all var{hom10} where(x >=65) into x2;
*read all var{muj10} where(x>=65) into x2;

```



```

/* Grid Points */
min=min(x3);
last=max(loc(x2>0));
gv=t(do(min,min+last-1,1));
ngrid=nrow(gv);

/* Parameters Male population*/
/* Champernowne */
    *M=75      ; *MLE scale;
    *a=16.48194 ; *MLE shape;

/* Parameters Female population*/
/* Champernowne */
    *M=76      ; *MLE scale;
    *a=15.83281 ; *MLE shape;

/* Moving MISE parameters */
aa=T(do(13, 17, .01));
na=nrow(aa);

ss=T(do(74,76,.001));
*ss=M;
ns=nrow(ss);

/* Variables */
n=sum(x2);
suma=j(na*ns,3,0);
r=1;

xy=j(ngrid,1,0);
fx=j(ngrid,1,0);

/*Transformation1:Champernowne*/
lim=1;
c=0;

start Tcap1(v,a,c,M,lim);
/*v is a data vector*/
n=nrow(v);
rr=1/2*(((v+j(n,1,c))##a-j(n,1,(M+c)**a))
/((v+j(n,1,c))##a
+j(n,1,(M+c)**a)-2*j(n,1,c**a))+1);
rr=(rr#(lim-(1-lim)))+(1-lim);
return(rr);
finish;

/* min MISE, finding optimal parameters */
do k= 1 to ns;
do j=1 to na ;
fx=Tcap1(gv,aa[j],c,ss[k],lim);
do i=1 to ngrid;
xy[i,1]=((fx[i,1]-x1[i,])**2)*x2[i,];

*xy[i,1]=(((fx[i,1]-x1[i,])**2)*x2[i,])*gv[i];

*xy[i,1]=(((fx[i,1]-x1[i,])**2)*x2[i,])*gv[i]**2;

*xy[i,1]=(((fx[i,1]-x1[i,])**2)*x2[i,])*sqrt(gv[i]);
end;
suma[r,1]=ss[k];
suma[r,2]=aa[j];
suma[r,3]=xy[+,1];
r=r+1;
end;
end;
pos=loc(suma[,3]=min(suma[,3]));
min=min(suma[,3]);

/* Optimal parameters */
a=suma[pos,2];
M=suma[pos,1];
print a M min;

/* Graph a parameter vs. min suma */
res= suma[,2] || suma[,3];
MATTRIB res
colname = {'shape', 'MISE' };
call pgraf (res, '+', 'shape', 'MISE');

/* Graph s parameter vs. min suma */
res= suma[,1] || suma[,3];
MATTRIB res
colname = {'shape', 'MISE' };
call pgraf (res, '+', 'shape', 'MISE');

/* Matrix of parameters - MISE */
create simul.sumach from suma [colname =
{'ss','aa','suma' }];
append from suma;
close simul.sumach;

QUIT;

```

Funciones de distribución paramétricas

- Cálculo de las funciones de distribución Weibull, Lognormal y Champernowne,

- Para el cálculo se han empleado los parámetros estimados por MLE y los parámetros estimados minimizando \widehat{MISE} .

Script

```

libname simul '.';

proc iml symsize=50000;

use simul.man65;
read all into xx;

* use simul.woman65;
* read all into xx;

/* Grid Points */
min=min(xx);
gv=t(do(min,200,1));
ngrid=nrow(gv);

/* Variables */
fhat_wei=j(ngrid,1,0);
fhat_ln=j(ngrid,1,0);
fhat_champ=j(ngrid,1,0);

/* Parameters Male population*/
/* Weibull */
s= 79.88856862; *MLE scale;
a= 9.70099632; *MLE shape;
*a= 8.577084; *MISE s fijo;
*s= 78.28556; *MISE;
*a= 9.782975; *MISE;

/* Lognormal */
mu= 4.328013635; *MLE scale;
sigma= 0.102264784; *MLE shape;
*sigma= 0.118113; *MISE mu fijo;
*mu= 4.315576; *MISE;
*sigma= 0.114779; *MISE;

/* Champernowne */
c=0;
M= 75; *MLE scale;
aa= 16.48194; *MLE shape;
*aa= 14.393958; *MISE M fijo;
*M= 74.827807; *MISE;
*aa= 14.442683; *MISE;

/* Parameters Female population*/
/* Weibull */
*s= 80.71985568; *MLE scale;
*a= 9.4868903; *MLE shape;
*a= 8.3552098; *MISE s fijo;
*s= 79.194232; *MISE;
*a= 9.364897; *MISE;

/* Lognormal */
*mu= 4.33669801; *MLE scale;
*sigma= 0.105999045; *MLE shape;
*sigma= 0.122632; *MISE mu fijo;
*mu= 4.324958; *MISE;
*sigma= 0.119931; *MISE;

/* Champernowne */
*c=0;
*M= 75; *MLE scale;
*a= 15.83281; *MLE shape;
*a= 13.696014; *MISE M fijo;
*M= 75.534849; *MISE;
*a= 13.821771; *MISE;

/* CDF Calculation */
fhat_wei=cdf('WEIBULL',gv,a,s);
print fhat_wei;

fhat_ln=cdf('LOGNORMAL',gv,mu,sigma);
print fhat_ln;

start Tcap1(v,a,c,M,lim);
/* v is a data vector */
n=nrow(v);
rr=1/2*(((v+j(n,1,c))##a-j(n,1,(M+c)**a))
/((v+j(n,1,c))##a
+j(n,1,(M+c)**a)-2*j(n,1,c**a))+1);
rr=(rr#(lim-(1-lim)))+(1-lim);
return(rr);
finish;

lim=1;

fhat_champ=Tcap1(gv,aa,c,M,lim);
print fhat_champ;

/* Matrix of CDF */
fhat_para=gv||fhat_wei||fhat_ln||fhat_champ;
print fhat_para;

```

```

create simul.fhatp2 from fhat_para [colname={'x',
      'Wei','Ln','Champ'}];
append from fhat_para;
close simul.fhatp2;
QUIT;

```

Funciones de distribución noparamétricas

- Cálculo de la función de distribución empírica,
- Cálculo de la función de distribución con el estimador núcleo clásico.

Script

```

libname simul '?';

proc iml SYMSIZE=50000;

use simul.man65;
read all into xx;

* use simul.woman65;
* read all into xx;

nr=nrow(x);

/* Function: Ordena, quartile 1-3 */
start ordena(v);
  v0=v;
  v[rank(v)]=v0;
  rr=v;
  return(rr);
finish;
start q1(v,n);
  v0=v;
  v[rank(v)]=v0;
  rr=v[int(n/4)];
  return(rr);
finish;
start q3(v,n);
  v0=v;
  v[rank(v)]=v0;
  rr=v[int((n/4)*3)];
  return(rr);
finish;

/* Epanechnikoc Kernel */
start kintd(u);
  rr=-0.25#((1+u)##2)#(u-2);
  return (rr);

finish;

/* z is a scalar, v is a vector */
start fhatz(z,v,nr,b);
  rr=Kintd((z-v)/b)/nr;
  return(rr);
finish;

/* Grid Points */
reset fw=30;
min=min(x);
gv=t(do(min,200,1));
ngrid=nrow(gv);

/*Empirical*/
emp=j(ngrid,1,0);
do j=1 to ngrid;
  ind=(x<=gv[j]);
  emp[j]=sum(ind)/nr;
end;
res_emp=gv||emp;
print res_emp;

/*Clasic Kernel*/
fhat_clas=j(ngrid,1,0);
sigma_x=sqrt((sum((x-(sum(x)/nr)#
  j(nr,1,1)##2))/nr);
iqr=q3(x,nr)-q1(x,nr);
s=min(sigma_x,iqr/1.349);
b_clas=(s**(5/3))*((8/3)**(1/3))*(nr**(-1/3));
suma=j(nr,1,0);

do j=1 to ngrid;
  do i=1 to nr;
    if abs(gv[j]-x[i])/b_clas<=1 then do;
      suma[i]=fhatz(gv[j],x[i],nr,b_clas);
    end;
  end;
end;

```

```

        end;
    if (gv[j]-x[i])/b_clas<=-1 then do;
        suma[i]=0;
    end;
    if (gv[j]-x[i])/b_clas>=1 then do;
        suma[i]=1/nr;
    end;
end;
fhat_clas[j]=sum(suma);
end;
print b_clas;
res_clas=gv||fhat_clas;

print res_clas;

/* Matrix of CDF*/
res=gv||emp||fhat_clas;
print res;

create simul.ec from res [colname={'x''emp''cke'}];
append from res;
close simul.ec;

QUIT;

```

Funciones de distribución semiparamétricas

- Cálculo de la función de distribución con el estimador núcleo doble transformado,
- Para el cálculo se han empleado los parámetros estimados por MLE y los parámetros estimados minimizando \widehat{MISE} ,
- Como primera transformación se han usado las distribuciones: Weibull, Lognormal y Champernowne.

Script Weibull

```

libname simul '.';

proc iml SYMSIZE=50000;

use simul.man65;
read all into x1;

* use simul.woman65;
* read all into x1;

nr=nrow(x1);
print nr;

/* Transformation: Weibull */
/*v = data vector*/
start Tcap1(v,a,s,c,lim);
n=nrow(v);
rr=cdf('WEIBULL',v-c,a,s);
rr=(rr#(lim-(1-lim)))+(1-lim);
return(rr);
finish;

/* Transformation: Inverse Beta distribution */
start Tcap2(v,w);
n=nrow(v);
do i=1 to n;
    v[i]=(betainv(w[i],3,3)*2)-1;
end;
rr=v;
return(rr);
finish;

start dTcap2(v);
ll=v-0.000000000001;
dv=(((betainv(v,3,3)*2)-1)-
    ((betainv(ll,3,3)*2)-1))/(v-ll);
rr=dv;
return(rr);
finish;

/* Epanechnikoc Kernel */
start kintd(u);

```

```

    rr=-0.25#((1+u)##2)#(u-2);
    return (rr);
finish;

/*z is a scalar, v is a vector*/
start fhatz(z,v,nr,b);
    rr=Kintd((z-v)/b)/nr;
    return(rr);
finish;

/* Grid Points */
reset fw=30;
min=min(x1);
gv=t(do(min,200,1));
ngrid=nrow(gv);

/* Parameters Male population*/
/* Weibull */
    s= 79.88856862; *MLE scale;
    a= 9.70099632; *MLE shape;
    *a= 8.577084; *MISE s fijo;
    *s= 78.28556; *MISE;
    *a= 9.782975; *MISE;

/* Parameters Female population*/
/* Weibull */
    *s= 80.71985568; *MLE scale;
    *a= 9.4868903; *MLE shape;
    *a= 8.3552098; *MISE s fijo;
    *s= 79.194232; *MISE;
    *a= 9.364897; *MISE;

/* Beta Kernel */
/* Transformation: Weibull */
c=0;
lim=1;
z=Tcap1(x1,a,s,c,lim);
gvz=Tcap1(gv,a,s,c,lim);
fhat_wei=gvz;

/* Transformation: inverse Beta */
aa=0;
aa=Tcap2(aa,lim);
gvy=j(ngrid,1,0);
gvy=Tcap2(gvy,gvz);
y=j(nr,1,0);

```

```

y=Tcap2(y,z);

/*Values for the kernel density in the variable
shifted grid points*/
suma=j(nr,1,0);
fhat_Betaw=j(ngrid,1,0);
der2=j(ngrid,1,0);
*b_Beta=1.0874*(nr**(-1/3));

do j=1 to ngrid;
    my=(15/16)*(1-gvy[j]**2)**2;
    dmy=((15/4)*gvy[j]*(gvy[j]**2-1)**2;

    b_beta=((my*(9/35))/((1/5)**2*dmy))**(1/3)
    *(nr**(-1/3));
    *print b_beta j;
do i=1 to nr;
    if abs((gvy[j]-y[i])/b_beta)<=1 then do;
        suma[i]=fhatz(gvy[j],y[i],nr,b_beta);
    end;
    if (gvy[j]-y[i])/b_beta<=-1 then do;
        suma[i]=0;
    end;
    if (gvy[j]-y[i])/b_beta>=1 then do;
        suma[i]=1/nr;
    end;
end;
fhat_Betaw[j]=sum(suma);
if fhat_Betaw[j]=0 then do;
    fhat_Betaw[j]=1;
end;
end;

/* Matrix of CDF*/
print b_beta;
res_betaw=gv||fhat_Betaw;
print res_betaw;

create simul.Bwei from res_betaw
    [colname={'x''DTKE Wei'}];
append from res_betaw;
close simul.Bwei;

QUIT;
*Bwei2 = CDF using mise shape parameter, Bwei3
        = CDF using mise parameters;

```

Script Lognormal

```
libname simul '?';
```

```
proc iml SYMSIZE=50000;
```

```

use simul.man65;
read all into x1;

* use simul.woman65;
* read all into x1;

nr=nrow(x1);

/*Transformation: Lognormal*/
/*v = data vector*/
start Tcap1(v,mu,sigma,lim);
  n=nrow(v);
  rr=cdf('LOGNORMAL',v,mu,sigma);
  rr=(rr#(lim-(1-lim)))+(1-lim);
return(rr);
finish;

/*Transformation: Inverse Beta distribution*/
start Tcap2(v,w);
  /*v is a vector*/
  n=nrow(v);
  do i=1 to n;
    v[i]=(betainv(w[i],3,3)*2)-1;
  end;
  rr=v;
  return(rr);
finish;

start dTcap2(v);
  ll=v-0.000000000001;
  dv=((betainv(v,3,3)*2)-1)-
    ((betainv(ll,3,3)*2)-1)/(v-ll);
  rr=dv;
  return(rr);
finish;

/*Epanechnikoc Kernel*/
start kintd(u);
  rr=-0.25#((1+u)##2)#(u-2);
  return (rr);
finish;

/*z is a scalar, v is a vector*/
start fhatz(z,v,nr,b);
  rr=Kintd((z-v)/b)/nr;
  return(rr);
finish;

/* Grid Points */
reset fw=30;
min=min(x1);
gv=t(do(min,200,1));
ngrid=nrow(gv);

/* Parameters Male population*/
/* Lognormal */
  mu= 4.328013635; *MLE scale;
  sigma= 0.102264784; *MLE shape;
  *sigma= 0.118113; *MISE mu fiyo;
  *mu= 4.315576; *MISE;
  *sigma= 0.114779; *MISE;

/* Parameters Female population*/
/* Lognormal */
  *mu= 4.33669801; *MLE scale;
  *sigma= 0.105999045; *MLE shape;
  *sigma= 0.122632; *MISE mu fiyo;
  *mu= 4.324958; *MISE;
  *sigma= 0.119931; *MISE;

/*Beta Kernel*/
/* Transformation: Lognormal */
lim=1;
z=Tcap1(x1,mu,sigma,lim);
gvz=Tcap1(gv,mu,sigma,lim);
fhat_ln=gvz;

/* Transformation: inverse Beta */
aa=0;
aa=Tcap2(aa,lim);
gvy=j(ngrid,1,0);
gvy=Tcap2(gvy,gvz);
y=j(nr,1,0);
y=Tcap2(y,z);

/*Values for the kernel density in the variable
  shifted grid points*/
suma=j(nr,1,0);
fhat_Betaln=j(ngrid,1,0);
der2=j(ngrid,1,0);
*b_Beta=1.0874*(nr**(-1/3));

do j=1 to ngrid;
  my=(15/16)*(1-gvy[j]**2)**2;
  dmy=((15/4)*gvy[j]*(gvy[j]**2-1))**2;

  b_beta=((my*(9/35))/((1/5)**2*dmy))**(1/3)
  *(nr**(-1/3));

do i=1 to nr;
  if abs((gvy[j]-y[i])/b_beta)<=1 then do;
    suma[i]=fhatz(gvy[j],y[i],nr,b_beta);
  end;
  if (gvy[j]-y[i])/b_beta<=-1 then do;
    suma[i]=0;
  end;
  if (gvy[j]-y[i])/b_beta>=1 then do;
    suma[i]=1/nr;
  end;
end;

```

```

        end;
    end;
    fhat_Betaln[j]=sum(suma);
    if fhat_Betaln[j]=0 then do;
        fhat_Betaln[j]=1;
    end;
end;

/* Matrix of CDF*/
print b_beta;
res_betaln=gv||fhat_Betaln;

```

```

print res_betaln;

create simul.Bln from res_betaln
    [colname={'x'DTKE Ln'}];
append from res_betaln;
close simul.Bln;

QUIT;
*Bl2 = CDF using mise shape parameter, Bln3 =
    CDF using mise parameters

```

Script Champernowne

```

libname simul '?';

proc iml SYMSIZE=50000;

use simul.man65;
read all into x1;

* use simul.woman65;
* read all into x1;

nr=nrow(x1);

/* Transformation: Champernowne */
/*v = data vector*/
start Tcap1(v,a,c,M,lim);
    n=nrow(v);
    rr=1/2*(((v+j(n,1,c))##a-j(n,1,(M+c)**a))
        /((v+j(n,1,c))##a+
            j(n,1,(M+c)**a)-2*j(n,1,c**a))+1);
    rr=(rr#(lim-(1-lim)))+(1-lim);
return(rr);
finish;

/* Transformation: Inverse Beta distribution */
start Tcap2(v,w);
    /*v is a vector*/
    n=nrow(v);
    do i=1 to n;
        v[i]=(betainv(w[i],3,3)*2)-1;
    end;
    rr=v;
return(rr);
finish;

start dTcap2(v);
    ll=v-0.000000000001;
    dv=(((betainv(v,3,3)*2)-1)-
        ((betainv(ll,3,3)*2)-1))/(v-ll);

```

```

    rr=dv;
    return(rr);
finish;

/* Epanechnikoc Kernel */
start kintd(u);
    rr=-0.25#((1+u)##2)#(u-2);
    return (rr);
finish;

/*z is a scalar, v is a vector*/
start fhatz(z,v,nr,b);
    rr=Kintd((z-v)/b)/nr;
    return(rr);
finish;

/* Grid Points */
reset fw=30;
min=min(x1);
gv=t(do(min,200,1));
ngrid=nrow(gv);

/* Parameters Male population*/
/* Champernowne */
c=0;
M= 75; *MLE scale;
aa= 16.48194; *MLE shape;
*aa= 14.393958; *MISE M fijs;
*M= 74.827807; *MISE;
*aa= 14.442683; *MISE;

/* Parameters Female population*/
/* Champernowne */
*c=0;
*M= 76; *MLE scale;
*a= 15.83281; *MLE shape;
*a= 13.696014; *MISE M fijs;
*M= 75.534849; *MISE;

```

```

    *a= 13.821771;    *MISE;

/*Beta Kernel*/
/* Transformation: Champernowne */
lim=1;
z=Tcap1(x1,a,c,M,lim);
gvz=Tcap1(gv,a,c,M,lim);
fhat_champ=gvz;

/* Transformation: inverse Beta */
aa=0;
aa=Tcap2(aa,lim);
gvy=j(ngrid,1,0);
gvy=Tcap2(gvy,gvz);
y=j(nr,1,0);
y=Tcap2(y,z);

/*Values for the kernel density in the variable
  shifted grid points*/
suma=j(nr,1,0);
fhat_Beta=j(ngrid,1,0);
der2=j(ngrid,1,0);
*b_Beta=1.0874*(nr**(-1/3));

do j=1 to ngrid;
    my=(15/16)*(1-gvy[j]**2)**2;
    dmy=((15/4)*gvy[j]*(gvy[j]**2-1))**2;

    b_beta=((my*(9/35))/((1/5)**2*dmy))**(1/3)
    * (nr**(-1/3));
do i=1 to nr;
    if abs((gvy[j]-y[i])/b_beta)<=1 then do;
        suma[i]=fhatz(gvy[j],y[i],nr,b_beta);
        end;
    if (gvy[j]-y[i])/b_beta<=-1 then do;
        suma[i]=0;
        end;
    if (gvy[j]-y[i])/b_beta>=1 then do;
        suma[i]=1/nr;
        end;
    end;
    fhat_Beta[j]=sum(suma);
    if fhat_Beta[j]=0 then do;
        fhat_Beta[j]=1;
    end;
end;

/* Matrix of CDF*/
print b_beta;
res_beta=gv||fhat_beta;
print res_beta;

create simul.Bchamp from res_beta
    [colname={'x' 'DTKE Champ'}];
append from res_beta;
close simul.Bchamp;

QUIT;

```

Bondad de ajuste

- Para facilitar la manipulación de los datos se ha creado una matriz que contiene todas las funciones de distribución estimadas,
- Se calcula el \widehat{MISE} con el fin de comparar todas funciones de distribución estimadas vs. la distribución empírica.

Script CDF

```

libname simul '?';

proc iml SYMSIZE=50000;

/* Reading all CDF's */

use simul.ec;    *emp, clas;

read all into x1;

use simul.Fhatp;    *we, ln, ch;
read all into x2;

use simul.Bwei;    *bw;
read all into x3;

```



```

use simul.Bln; *bln;
read all into x4;

use simul.Bchamp; *bch;
read all into x5;

use simul.Fhatp2; *we, ln, ch mise;
read all into x6;

use simul.Bwei2; *bw mise;
read all into x7;

use simul.Bln2; *bln mise;
read all into x8;

use simul.Bchamp2; *bch mise;
read all into x9;

use simul.Fhatp3; *we, ln, ch mise2;
read all into x10;

use simul.Bwei3; *bw mise2;
read all into x11;

use simul.Bln3; *bln mise2;
read all into x12;

```

```

use simul.Bchamp3; *bch mise2;
read all into x13;

allcdf= x1|| x2[,2:4] || x3[,2] || x4[,2] || x5[,2]
|| x6[,2:4] || x7[,2] || x8[,2] || x9[,2]
|| x10[,2:4]|| x11[,2]|| x12[,2]|| x13[,2];

MATTRIB allcdf
colname = {'x','emp','Clas', 'Wei', 'Ln', 'Champ',
'Wei', 'BLn', 'BChamp', 'WeiISE', 'LnISE',
'ChampISE', 'BWeiISE', 'BLnISE',
'BChampISE', 'WeiISE2', 'LnISE2',
'ChampISE2', 'BWeiISE2', 'BLnISE2',
'BChampISE2'};
print allcdf;

create simul.FX from allcdf [colname =
{'x','emp','Clas', 'Wei', 'Ln', 'Champ',
'Wei', 'BLn', 'BChamp', 'WeiISE', 'LnISE',
'ChampISE', 'BWeiISE', 'BLnISE',
'BChampISE', 'WeiISE2', 'LnISE2',
'ChampISE2', 'BWeiISE2', 'BLnISE2',
'BChampISE2'}];
append from allcdf;
close simul.FX;

QUIT;

```

Script MISE

```

libname simul '.';

proc iml SYMSIZE=50000;

use simul.Fx;
read all into x1;

use simul.alive;
read all var{hom10} where(x>=65) into x2;
*read all var{muj10} where(x>=65) into x2;

/* Grid Points */
min=min(x1[,1]);
last=max(loc(x2>0));
gv=t(do(min,min+last-1,1));
ngrid=nrow(gv);

/* Variables */
nc=ncol(x1)-2;
n=sum(x2);

```

```

freq=j(ngrid,nc,0);
suma=j(nc,1,0);

/* MISE */
do j=1 to nc ;
do i=1 to ngrid;
freq[i,j]=((x1[i,j+2]-x1[i,2])**2)*x2[i,];
end;
suma[j,1]=freq[+,j];
end;

MATTRIB suma
rowname = {'Clas', 'Wei', 'Ln', 'Champ', 'BWei',
'BLn', 'BChamp', 'WeiISE', 'LnISE',
'ChampISE', 'BWeiISE', 'BLnISE',
'BChampISE', 'WeiISE2', 'LnISE2',
'ChampISE2', 'BWeiISE2', 'BLnISE2',
'BChampISE2'};
print suma;

create simul.mise from suma ;

```

```
append from suma;
close simul.mise;
```

```
QUIT;
```

Cálculo de la función de distribución truncada

- Crea una matriz por cada edad de truncamiento (de 70 a 100, cada 5 años). Cada matriz contiene las funciones de distribución truncadas para todos los métodos de estimación.

Script

```
libname simul '.';

proc iml SYMSIZE=50000;

use simul.FX;
read all into x1;

/* Grid Points */
grid=1;
gv=t(do(65,200,grid));
ngrid=nrow(gv);

/* Variables */
nc=ncol(x1);
fx=j(ngrid,nc,0);
fx[,1]=gv;
fx[1,]=x1[1,];

y=do(70,100,5); *num trunca;
ny=ncol(y);

/* Densidad */
do j=2 to nc;
  do i=2 to ngrid;
    fx[i,j]=x1[i,j] -x1[i-1,j];
  end;
end;
*print fx;

/* CDF Truncada */

do j=1 to ny;
  yy=y[j];
  p=loc(x1[,1]=yy);
  nrg=ngrid-p+1;
  gx=j(nrg,nc,0); *PDF;

  gxx=j(nrg,nc,0); *CDF;
  do k=2 to nc;
    do i=1 to nrg ;
      gx[i,k]=fx[i+p-1,k]/(1-x1[p,k]);
      gx1=gx[2:nrg, k];
      gxx[i,k]=gx1[+,k];
    end;
  end;
  gx[,1]=t(do(yy,200,grid));
  gxx[,1]=t(do(yy,200,grid));
  namex = "pgx"+strip(char(j));
  namexx = "cgx"+strip(char(j));
call valset(namex, gx);
call valset(namexx, gxx);
*print gx gxx;
end;

print cgx1 cgx2 cgx3 cgx4 cgx5 cgx6 cgx7;

create simul.gx70 from cgx1 [colname =
  {'x','emp', 'Clas', 'Wei', 'Ln', 'Champ',
  'BWei', 'BLn', 'BChamp', 'WeiISE', 'LnISE',
  'ChampISE', 'BWeiISE', 'BLnISE',
  'BChampISE',
  'WeiISE2', 'LnISE2', 'ChampISE2', 'BWeiISE2',
  'BLnISE2', 'BChampISE2'}];
append from cgx1;
close simul.gx70;

create simul.gx75 from cgx2 [colname =
  {'x','emp', 'Clas', 'Wei', 'Ln', 'Champ',
  'BWei', 'BLn', 'BChamp', 'WeiISE', 'LnISE',
  'ChampISE', 'BWeiISE', 'BLnISE',
  'BChampISE',
  'WeiISE2', 'LnISE2', 'ChampISE2', 'BWeiISE2',
```

```

        'BLnISE2', 'BChampISE2'}];
append from cgx2;
close simul.gx75;

create simul.gx80 from cgx3 [colname =
    {'x','emp', 'Clas', 'Wei', 'Ln', 'Champ',
'BWei', 'BLn', 'BChamp', 'WeiISE', 'LnISE',
'ChampISE', 'BWeiISE', 'BLnISE',
'BChampISE',
'WeiISE2', 'LnISE2', 'ChampISE2', 'BWeiISE2',
'BLnISE2', 'BChampISE2'}];
append from cgx3;
close simul.gx80;

create simul.gx85 from cgx4 [colname =
    {'x','emp', 'Clas', 'Wei', 'Ln', 'Champ',
'BWei', 'BLn', 'BChamp', 'WeiISE', 'LnISE',
'ChampISE', 'BWeiISE', 'BLnISE',
'BChampISE',
'WeiISE2', 'LnISE2', 'ChampISE2', 'BWeiISE2',
'BLnISE2', 'BChampISE2'}];
append from cgx4;
close simul.gx85;

create simul.gx90 from cgx5 [colname =
    {'x','emp', 'Clas', 'Wei', 'Ln', 'Champ',
'BWei', 'BLn', 'BChamp', 'WeiISE', 'LnISE',
'ChampISE', 'BWeiISE', 'BLnISE',
'BChampISE',
'WeiISE2', 'LnISE2', 'ChampISE2', 'BWeiISE2',
'BLnISE2', 'BChampISE2'}];
append from cgx5;
close simul.gx90;

create simul.gx95 from cgx6 [colname =
    {'x','emp', 'Clas', 'Wei', 'Ln', 'Champ',
'BWei', 'BLn', 'BChamp', 'WeiISE', 'LnISE',
'ChampISE', 'BWeiISE', 'BLnISE',
'BChampISE',
'WeiISE2', 'LnISE2', 'ChampISE2', 'BWeiISE2',
'BLnISE2', 'BChampISE2'}];
append from cgx6;
close simul.gx95;

create simul.gx100 from cgx7 [colname =
    {'x','emp', 'Clas', 'Wei', 'Ln', 'Champ',
'BWei', 'BLn', 'BChamp', 'WeiISE', 'LnISE',
'ChampISE', 'BWeiISE', 'BLnISE',
'BChampISE',
'WeiISE2', 'LnISE2', 'ChampISE2', 'BWeiISE2',
'BLnISE2', 'BChampISE2'}];
append from cgx7;
close simul.gx100;

QUIT;

```

Cálculo del valor en riesgo

- Se calcula el valor en riesgo directamente de las funciones de distribución truncadas estimadas,
- Se calcula el valor en riesgo a diferentes niveles de confianza: de 0.9 a 0.999 cada 0.001,
- Se emplea la interpolación lineal para aproximar el cuantil exacto.

Script

```

libname simul '.';

proc iml SYMSIZE=50000;

*use simul.Fx;    *65;
*read all into x5;

use simul.gx85;
read all into x5;

/* Variables */
mvar=do(0.9,0.999,0.001);
nalpha=ncol(mvar);

```

```

nr=nrow(x5);
nc=ncol(x5);

pos=j(nc-1,nalpha,0);
y1=j(nc-1,nalpha,0);
xint = j(nc-1,nalpha,0);

/* Function: Interpolation */
start LinInterpPt(x1, x2, y1, y2, x);
  m = (y2-y1)/(x2-x1); /* slope */
  return( m#(x-x1) + y1 ); /* point-slope
    formula */
finish;

/* Age where FX is <=alpha */
do j=2 to nc;
  do i = 1 to nalpha;
    pos[j-1,i] =max(loc(x5[,j] <= mvar[,i]
    ));
    y1[j-1,i]= x5[pos[j-1,i],1];
  end;
end;

/* Finding VaR */
do j=2 to nc;
  do i = 1 to nalpha;
    if (pos[j-1,i]+1 >= nr) then
      do ;
        xint[j-1,i] = y1[j-1,i];
      end;
    else
      do;
        xx1=x5[pos[j-1,i],j];
        xx2=x5[pos[j-1,i]+1,j];
        yy1=x5[pos[j-1,i],1];
        yy2=x5[pos[j-1,i]+1,1];
        *print i xx1 xx2 yy1 yy2;
        xint[j-1,i] =
          LinInterpPt(xx1,xx2,yy1,yy2, mvar[,i]);
      end;
    end;
  end;
end;

print y1;

xint=t(xint); print xint;

create simul.var65 from xint [colname =
  {'emp','Clas', 'Wei', 'Ln', 'Champ', 'BWei',
  'BLn', 'BChamp', 'WeiISE', 'LnISE',
  'ChampISE', 'BWeiISE', 'BLnISE',
  'BChampISE', 'WeiISE2', 'LnISE2',
  'ChampISE2', 'BWeiISE2', 'BLnISE2',
  'BChampISE2'} ];
append from xint;
close simul.var65;

create simul.var85 from xint [colname =
  {'emp','Clas', 'Wei', 'Ln', 'Champ', 'BWei',
  'BLn', 'BChamp', 'WeiISE', 'LnISE',
  'ChampISE', 'BWeiISE', 'BLnISE',
  'BChampISE', 'WeiISE2', 'LnISE2',
  'ChampISE2', 'BWeiISE2', 'BLnISE2',
  'BChampISE2'} ];
append from xint;
close simul.var85;

*/
QUIT;

```

Cálculo del valor en riesgo empírico

- Se calcula el valor en riesgo empírico directamente de las funciones de distribución estimadas,
- Se calcula el valor en riesgo a diferentes niveles de confianza: de 0.9 a 0.999 cada 0.001.

Script

```
libname simul '?';
```

```
proc iml SYMSIZE=50000;
```

```

use simul.Fx;
read all var{emp} into x2;

use simul.gx70;
read all var{emp} into x3;

use simul.gx75;
read all var{emp} into x4;

use simul.gx80;
read all var{emp} into x5;

use simul.gx85;
read all var{emp} into x6;

use simul.gx90;
read all var{emp} into x7;

use simul.gx95;
read all var{emp} into x8;

use simul.gx100;
read all var{emp} into x9;

/* Grid Points */
gv=t(do(65,200,1));
ngrid=nrow(gv);

/* alpha */
mvar=do(0.9,0.999,0.001);
nalpha=ncol(mvar);

/* Matrix of Empirical CDF */

```

```

x33=t(j(1,5,0))//x3;
x44=t(j(1,10,0))//x4;
x55=t(j(1,15,0))//x5;
x66=t(j(1,20,0))//x6;
x77=t(j(1,25,0))//x7;
x88=t(j(1,30,0))//x8;
x99=t(j(1,35,0))//x9;

xx=gv||x2||x33||x44||x55||x66||x77||x88||x99;
nc=ncol(xx);

/* Variables */
pos=j(nc-1,nalpha,0);
y1=j(nc-1,nalpha,0);

/* Finding Empirical VaR */

do j=2 to nc;
  do i = 1 to nalpha;
    pos[j-1,i]=max(loc(xx[,j]<= mvar[,i]));
    y1[j-1,i]=xx[pos[j-1,i]+1,1];
  end;
end;

y1=t(y1); print y1;

create simul.vemp from y1 [colname ={
  '65''70''75''80''85''90''95''100'}];
append from y1 ;
close simul.vemp;

QUIT;

```

Cálculo del valor en riesgo KQE

- Calculo del valor en riesgo usando el KQE, *Kernel Quantile Estimator*

Script

```

libname simul '.';

proc iml SYMSIZE=5000;

use simul.man65;
read all var{x} where(x>=85) into xx;

* use simul.woman65;
* read all into x1;

/* Function: Order */
start ordena(v);
  v0=v;
  v[rank(v)]=v0;
  rr=v;
  return(rr);
finish;

/* Kernel */
start kernel(u);

```

```

/*z is a scalar, v is a vector*/
rr=(1/sqrt(2#3.141592654))#exp(-(u##2)/2);
return(rr);
finish;

/* Variables */
nr=nrow(xx);
nc=ncol(xx);
maxiter2=100000;

/* alpha */
alpha=do(0.9,0.999,0.001);
nalpna=ncol(alpha);
Var_kqe=j(nalpna,1,0);

/* KQE */
do l=1 to nalpna;
  b_kqe=((alpha[l]*(1-alpha[l]))/(nr+1))**(1/2);
  x=xx[,l];
  x=ordena(x);

  num=j(nr,1,0);
  den=j(nr,1,0);
  do i=1 to nr;
    arg=(((i-(1/2))/nr)-alpha[l])/b_kqe;
    num[i]=kernel(arg)*x[i];
    den[i]=kernel(arg);
  end;
  Var_kqe[l]=sum(num)/sum(den);
end;
print Var_kqe;

a=t(alpha);
result=a||Var_kqe;

create simul.kqe85 from result [colname ={'85'}];
append from result ;
close simul.kqe85;

QUIT;
*KQE for all ages 70:100 each 5;

```

Cálculo del valor en riesgo exacto usando CKE

- Calculo del VaR exacto usando el estimador núcleo.

Script

```

libname simul '.';
proc iml SYMSIZE=5000;

use simul.man65;
read all into xx;

* use simul.woman65;
* read all into xx;

/* Function: Ordena, quartile 1-3 */
maxiter2=100000;
start ordena(v);
  v0=v;
  v[rank(v)]=v0;
  rr=v;
  return(rr);
finish;
start q1(v,n);
  v0=v;
  v[rank(v)]=v0;
  rr=v[int(n/4)];
  return(rr);

finish;
start q3(v,n);
  v0=v;
  v[rank(v)]=v0;
  rr=v[int((n/4)*3)];
  return(rr);
finish;

/*Epanechnikoc Kernel*/
start K(u);
  rr=0.75*(1-(u#u))#(abs(u)<=1);
  return (rr);
finish;

start kintd(u);
  rr=-0.25#((1+u)##2)#(u-2);
  return (rr);
finish;

/* Functions */
start DERIV(z,v,nr,b);
  rr=sum(K((z*j(nr,1,1)-v)/b))/(nr*b);

```

```

    return(rr);
finish;

start cdfhatz(z,v,nr,b);
/*z is a scalar, v is a vector*/
rr=Kintd((z-v)/b)/nr;
return(rr);
finish;

start FUN(z,v,nr,b,alpha);
/*z is a scalar, v is a vector*/
suma=j(nr,1,0);
do i=1 to nr;
if abs(z-v[i])/b<=1 then do;
suma[i]=cdfhatz(z,v[i],nr,b);
end;
if (z-v[i])/b<=-1 then do;
suma[i]=0;
end;
if (z-v[i])/b>=1 then do;
suma[i]=1/nr;
end;
end;
rr=sum(suma)-alpha;
return(rr);
finish;

/* Newton */
converge=1E-10;
start newton;
/*z is a scalar, v is a vector*/
ff=FUN(z,v,nr,b,alpha);
do iter2=1 to maxiter2 while
(abs(ff)>converge);
dd=DERIV(z,v,nr,b);
delta=-solve(dd,ff);
z=z+delta;
ff=FUN(z,v,nr,b,alpha);
end;
finish;

/* Variables */
nr=nrow(xx);
nc=ncol(xx);

reset fw=30;
emp=j(nr,1,0);
alpha=0.90;

rati=j(nc,1,0);
Var_clas=j(nc,1,0);
fun=j(nc,1,0);
it=j(nc,1,0);
*nc=2;

/* VaR Calculation*/
do l=1 to nc;
/*Original data*/
x=xx[,l];
x=ordena(x);
do j=1 to nr;
ind=(x<=x[j]);
emp[j]=sum(ind)/nr;
end;
*Estimation;
ind=(emp>alpha);
sind=sum(ind);
if sind>0 then do;
ni=nr-sind+1;
Var_hat=x[ni];
end;
if sind=0 then do;
ni=nr-sind;
Var_hat=x[ni];
end;
ini=var_hat;
/*Clasic kernel*/
sigma_x=sqrt((sum((x-(sum(x)/nr)
#j(nr,1,1))#2))/nr);
iqr=q3(x,nr)-q1(x,nr);
s=min(sigma_x,iqr/1.349);
b_clas=(s**(5/3))*((8/3)**(1/3)) *(nr**(-1/3));
z=ini;
v=x;
b=b_clas;
run newton;
Var_clas[l]=z;
fun[l]=ff;
it[l]=iter2;
end;

maf=max(ff); mif=min(ff);
mait=max(it); miit=min(it);

result=Var_clas;
print 'alpha=0.95';
print var_hat result ;

QUIT;

```

Cálculo del valor en riesgo exacto de la distribución truncada CKE

- Cálculo del valor en riesgo exacto usando el estimador núcleo cuando se tiene la función de distribución truncada.

Script

```

libname simul ' ';
proc iml SYMSIZE=5000;

use simul.man65;
read all into xx;

*use simul.woman65;
*read all into xx;

/* Function: Ordena, quartile 1-3 */
maxiter2=100000;
start ordena(v);
  v0=v;
  v[rank(v)]=v0;
  rr=v;
  return(rr);
finish;
start q1(v,n);
  v0=v;
  v[rank(v)]=v0;
  rr=v[int(n/4)];
  return(rr);
finish;
start q3(v,n);
  v0=v;
  v[rank(v)]=v0;
  rr=v[int((n/4)*3)];
  return(rr);
finish;

/*Epanechnikoc Kernel*/
start K(u);
  rr=0.75*(1-(u#u))#(abs(u)<=1);
  return (rr);
finish;

start kintd(u);
  rr=-0.25#((1+u)##2)#(u-2);
  return (rr);
finish;

/* Functions */

start cdfhatz(z,v,nr,b);
  /*z is a scalar, v is a vector*/
  rr=Kintd((z-v)/b)/nr;
  return(rr);
finish;

start FUN(z,tru,v,nr,b,alpha);
  /*z is a scalar, v is a vector*/
  suma=j(nr,1,0);
  do i=1 to nr;
    if abs(z-v[i])/b<=1 then do;
      suma[i]=cdfhatz(z,v[i],nr,b);
    end;
    if (z-v[i])/b<=-1 then do;
      suma[i]=0;
    end;
    if (z-v[i])/b>=1 then do;
      suma[i]=1/nr;
    end;
  end;
  sumat=j(nr,1,0);
  do i=1 to nr;
    if abs(tru-v[i])/b<=1 then do;
      sumat[i]=cdfhatz(tru,v[i],nr,b);
    end;
    if (tru-v[i])/b<=-1 then do;
      sumat[i]=0;
    end;
    if (tru-v[i])/b>=1 then do;
      sumat[i]=1/nr;
    end;
  end;
  rr=((sum(suma)-sum(sumat))/(1-sum(sumat)))-alpha;
  return(rr);
finish;

start DERIV(z,tru,v,nr,b);
  /*z is a scalar, v is a vector*/
  rr=sum(K((z*j(nr,1,1)-v)/b))/(nr*b);
  sumat=j(nr,1,0);
  do i=1 to nr;

```



```

        if abs(tru-v[i])/b<=1 then do;
            sumat[i]=cdfhatz(tru,v[i],nr,b);
        end;
    if (tru-v[i])/b<=-1 then do;
        sumat[i]=0;
    end;
    if (tru-v[i])/b>=1 then do;
        sumat[i]=1/nr;
    end;
end;
rr=rr/(1-sum(sumat));
return(rr);
finish;

/* Newton */
converge=1E-10;
start newton;
    /*z is a scalar, v is a vector*/
    ff=FUN(z,75,v,nr,b,alpha);
    do iter2=1 to maxiter2 while(abs(ff)>converge);
        dd=DERIV(z,75,v,nr,b);
        delta=-solve(dd,ff);
        z=z+delta;
        ff=FUN(z,75,v,nr,b,alpha);
    end;
finish;

/* Variables */
nr=nrow(xx);
nc=ncol(xx);
emp=j(nr,1,0);

reset fw=30;
alpha=0.99;

rati=j(nc,1,0);
Var_clas=j(nc,1,0);
fun=j(nc,1,0);
it=j(nc,1,0);

/* VaR Calculation*/

do l=1 to nc;
    /*Original data*/
    x=xx[,l];
    x=ordena(x);
do j=1 to nr;
    ind=(x<=x[j]);
    emp[j]=sum(ind)/nr;
end;
/*Estimation;
ind=(emp>alpha);
sind=sum(ind);
if sind>0 then do;
ni=nr-sind+1;
Var_hat=x[ni];
end;
if sind=0 then do;
ni=nr-sind;
Var_hat=x[ni];
end;
ini=var_hat;
/*Clasic kernel*/
sigma_x=sqrt((sum((x-(sum(x)/nr)#
j(nr,1,1))##2))/nr);
iqr=q3(x,nr)-q1(x,nr);
s=min(sigma_x,iqr/1.349);
b_clas=(s**(5/3))*((8/3)**(1/3))*(nr**(-1/3));
z=ini;
v=x;
b=b_clas;
run newton;
Var_clas[l]=z;
fun[l]=ff;
it[l]=iter2;
end;
maf=max(ff); mif=min(ff);
mait=max(it); miit=min(it);

result=Var_clas;
print 'alpha=0.99';
print result ;

QUIT;

```

Cálculo del valor en riesgo exacto usando DTKE

- Cálculo del valor en riesgo exacto usando el estimador núcleo doble transformado, asumiendo como primera transformación a las distribuciones Weibull, Lognormal y Champernowne.

```

libname simul '.';
proc iml SYMSIZE=5000;

use simul.man65;
read all into xx;

* use simul.woman65;
* read all into x1;

/* Function: Order */
maxiter2=100000;
start ordena(v);
v0=v;
v[rank(v)]=v0;
rr=v;
return(rr);
finish;

/*Transformation: Weibull*/
start Tcap1(v,a,s,c,lim);
/*v is a data vector*/
n=nrow(v);
rr=cdf('WEIBULL',v-c,a,s);
rr=(rr#(lim-(1-lim)))+(1-lim);
return(rr);
finish;

/*Transformation: Inverse Beta distrtribution*/
start Tcap2(v,w);
/*v is a vector*/
n=nrow(v);
do i=1 to n;
v[i]=(betainv(w[i],3,3)*2)-1;
end;
rr=v;
return(rr);
finish;

/*Epanechnikoc Kernel*/
start K(u);
rr=0.75*(1-(u#u))#(abs(u)<=1);
return (rr);
finish;

start kintd(u);
rr=-0.25#((1+u)##2)#(u-2);
return (rr);
finish;

/* Functions */
start DERIV(z,v,nr,b);
/*z is a scalar, v is a vector*/
rr=sum(K((z*j(nr,1,1)-v)/b))/(nr*b);
return(rr);
finish;

start cdfhatz(z,v,nr,b);
/*z is a scalar, v is a vector*/
rr=Kintd((z-v)/b)/nr;
return(rr);
finish;

start FUN(z,v,nr,b,alpha);
/*z is a scalar, v is a vector*/
suma=j(nr,1,0);
do i=1 to nr;
if abs(z-v[i])/b<=1 then do;
suma[i]=cdfhatz(z,v[i],nr,b);
end;
if (z-v[i])/b<=-1 then do;
suma[i]=0;
end;
if (z-v[i])/b>=1 then do;
suma[i]=1/nr;
end;
end;
rr=sum(suma)-alpha;
return(rr);
finish;

/* Newton */
converge=1E-10;
start newton;
/*z is a scalar, v is a vector*/
ff=FUN(z,v,nr,b,alpha);
do iter2=1 to maxiter2 while
(abs(ff)>converge);
dd=DERIV(z,v,nr,b);
delta=-solve(dd,ff);
z=z+delta;
ff=FUN(z,v,nr,b,alpha);
end;
finish;

/* Variables */
nr=nrow(xx);
nc=ncol(xx);

reset fw=30;
alpha=0.95;
emp=j(nr,1,0);

rati=j(nc,1,0);
Var_Beta=j(nc,1,0);

```

```

fun=j(nc,1,0);
it=j(nc,1,0);

/* Inverse Beta distribution */
lim=1;
aa=0;
aa=Tcap2(aa,lim);

/* Beta*/
*b_Beta=1.0874*nr**(-1/3);
*b_Beta=1.4422*nr**(-1/3);
gy=(betainv(alpha,3,3)*2)-1;
my=(15/16)*(1-gy**2)**2;
dmy=((15/4)*gy*(gy**2-1)**2;
b_beta=((my*(9/35))/((1/5)**2*dmy))**(1/3)
      *(nr**(-1/3));

/* VaR Calculation*/
do l=1 to nc;
  /*Original data*/
  x=xx[,l];
  x=ordena(x);
  /* Parameters Male population*/
  /* Weibull */
  s= 79.88856862; *MLE scale;
  a= 9.70099632; *MLE shape;
  *a= 8.577084; *MISE s fijo;
  *s= 78.28556; *MISE;
  *a= 9.782975; *MISE;

  /* Parameters Female population*/
  /* Weibull */
  *s= 80.71985568; *MLE scale;
  *a= 9.4868903; *MLE shape;
  *a= 8.3552098; *MISE s fijo;
  *s= 79.194232; *MISE;
  *a= 9.364897; *MISE;

  lim=1;
  do j=1 to nr;
    ind=(x<=x[j]);
    emp[j]=sum(ind)/nr;
  end;
  /*Estimation */
  ind=(emp>alpha);
  sind=sum(ind);
  if sind>0 then do;
    ni=nr-sind+1;
    Var_hat=x[ni];
  end;
  if sind=0 then do;
    ni=nr-sind;
    Var_hat=x[ni];
  end;
  Vin=Tcap1(Var_hat,a,s,c,lim);
  ini=0;
  ini=Tcap2(ini,vin);
  /*Weibull transformation kernel*/
  w=Tcap1(x,a,s,c,lim);
  y=j(nr,1,0);
  y=Tcap2(y,w);
  z=ini;
  v=y;
  b=b_Beta;
  run newton;
  zy=(z+1)/2;
  zy=CDF('BETA',zy,3,3);
  zx=quantile('WEIBULL',zy,a,s)+c;
  Var_Beta[l]=zx;
  fun[l]=ff;
  it[l]=iter2;
end;

maf=max(ff); mif=min(ff);
mait=max(it); miit=min(it);

result=Var_Beta;
print 'alpha=0.95';
print result;

QUIT;

```

Script Lognormal

```

libname simul '.';
proc iml SYMSIZE=5000;

use simul.man65;
read all into xx;

* use simul.woman65;
* read all into x1;

/* Function: Order */
maxiter2=100000;
start ordena(v);
  v0=v;
  v[rank(v)]=v0;
  rr=v;
  return(rr);

```

```

finish;

/*Transformation: Lognormal*/
start Tcap1(v,mu,sigma,lim);
/*v is a data vector*/
n=nrow(v);
rr=cdf('LOGNORMAL',v,mu,sigma);
rr=(rr#(lim-(1-lim)))+(1-lim);
return(rr);
finish;

/*Transformation: Inverse Beta distribution*/
start Tcap2(v,w);
/*v is a vector*/
n=nrow(v);
do i=1 to n;
v[i]=(betainv(w[i],3,3)*2)-1;
end;
rr=v;
return(rr);
finish;

/*Epanechnikoc Kernel*/
start K(u);
rr=0.75*(1-(u#u))#(abs(u)<=1);
return (rr);
finish;

start kintd(u);
rr=-0.25#((1+u)##2)#(u-2);
return (rr);
finish;

/* Functions */
start DERIV(z,v,nr,b);
/*z is a scalar, v is a vector*/
rr=sum(K((z*j(nr,1,1)-v)/b))/(nr*b);
return(rr);
finish;

start cdfhatz(z,v,nr,b);
/*z is a scalar, v is a vector*/
rr=Kintd((z-v)/b)/nr;
return(rr);
finish;

start FUN(z,v,nr,b,alpha);
/*z is a scalar, v is a vector*/
suma=j(nr,1,0);
do i=1 to nr;
if abs(z-v[i])/b<=1 then do;
suma[i]=cdfhatz(z,v[i],nr,b);
end;
if (z-v[i])/b<=-1 then do;

```

```

suma[i]=0;
end;
if (z-v[i])/b>=1 then do;
suma[i]=1/nr;
end;
end;
rr=sum(suma)-alpha;
return(rr);
finish;

/* Newton */
converge=1E-10;
start newton;
/*z is a scalar, v is a vector*/
ff=FUN(z,v,nr,b,alpha);
do iter2=1 to maxiter2 while
(abs(ff)>converge);
dd=DERIV(z,v,nr,b);
delta=-solve(dd,ff);
z=z+delta;
ff=FUN(z,v,nr,b,alpha);
end;
finish;

/* Variables */
nr=nrow(xx);
nc=ncol(xx);

reset fw=30;
alpha=0.95;
emp=j(nr,1,0);

rati=j(nc,1,0);
Var_Beta=j(nc,1,0);
fun=j(nc,1,0);
it=j(nc,1,0);

/* Inverse Beta distribution */
lim=1;
aa=0;
aa=Tcap2(aa,lim);

/* Beta*/
*b_Beta=1.0874*nr**(-1/3);
*b_Beta=1.4422*nr**(-1/3);
gy=(betainv(alpha,3,3)*2)-1;
my=(15/16)*(1-gy**2)**2;
dmy=((15/4)*gy*(gy**2-1))**2;
b_beta=((my*(9/35))/((1/5)**2*dmy))**(1/3)*(nr**(-1/3));

/* VaR Calculation*/
do l=1 to nc;
/*Original data*/
x=xx[,l];

```

```

        x=ordena(x);
/* Parameters Male population*/
/* Lognormal */
        mu= 4.328013635; *MLE scale;
        sigma= 0.102264784; *MLE shape;
        *sigma= 0.118113; *MISE mu fijo;
        *mu= 4.315576; *MISE;
        *sigma= 0.114779; *MISE;

/* Parameters Female population*/
/* Lognormal */
        *mu= 4.33669801; *MLE scale;
        *sigma= 0.105999045; *MLE shape;
        *sigma= 0.122632; *MISE mu fijo;
        *mu= 4.324958; *MISE;
        *sigma= 0.119931; *MISE;

lim=1;
do j=1 to nr;
        ind=(x<=x[j]);
        emp[j]=sum(ind)/nr;
end;
/*Estimation */
ind=(emp>alpha);
sind=sum(ind);
if sind>0 then do;
        ni=nr-sind+1;
        Var_hat=x[ni];
end;
if sind=0 then do;
        ni=nr-sind;
        Var_hat=x[ni];
end;
Vin=Tcap1(Var_hat,mu,sigma,lim);
ini=0;
ini=Tcap2(ini,vin);
/*Lognormal transformation kernel*/
w=Tcap1(x,mu,sigma,lim);
y=j(nr,1,0);
y=Tcap2(y,w);
z=ini;
v=y;
b=b_Beta;
run newton;
zy=(z+1)/2;
zy=CDF('BETA',zy,3,3);
zx=quantile('LOGNORMAL',zy,mu,sigma);
Var_Beta[l]=zx;
fun[l]=ff;
it[l]=iter2;
end;
maf=max(ff); mif=min(ff);
mait=max(it); miit=min(it);

result=Var_Beta;
print 'alpha=0.95';
print result;

QUIT;

```

Script Champernowne

```

libname simul '.';
proc iml SYMSIZE=5000;

use simul.man65;
read all into xx;

* use simul.woman65;
* read all into x1;

/* Function: Order */
maxiter2=100000;
start ordena(v);
v0=v;
v[rank(v)]=v0;
rr=v;
return(rr);
finish;

/*Transformation: Champernowne*/
start Tcap1(v,a,c,M,lim);
/*v is a data vector*/
n=nrow(v);
rr=1/2*(((v+j(n,1,c))##a-j(n,1,(M+c)**a))
/((v+j(n,1,c))##a+j(n,1,(M+c)**a)
-2*j(n,1,c**a)+1);
rr=(rr#(lim-(1-lim)))+(1-lim);
return(rr);
finish;

/*Transformation: Inverse Beta distrtribution*/
start Tcap2(v,w);
/*v is a vector*/
n=nrow(v);
do i=1 to n;
        v[i]=(betainv(w[i],3,3)*2)-1;
end;

```

```

rr=v;
return(rr);
finish;

/*Epanechnikoc Kernel*/
start K(u);
rr=0.75*(1-(u#u))#(abs(u)<=1);
return (rr);
finish;

start kintd(u);
rr=-0.25#((1+u)##2)#(u-2);
return (rr);
finish;

/* Functions */
start DERIV(z,v,nr,b);
/*z is a scalar, v is a vector*/
rr=sum(K((z*j(nr,1,1)-v)/b))/(nr*b);
return(rr);
finish;

start cdfhatz(z,v,nr,b);
/*z is a scalar, v is a vector*/
rr=Kintd((z-v)/b)/nr;
return(rr);
finish;

start FUN(z,v,nr,b,alpha);
/*z is a scalar, v is a vector*/
suma=j(nr,1,0);
do i=1 to nr;
if abs(z-v[i])/b<=1 then do;
suma[i]=cdfhatz(z,v[i],nr,b);
end;
if (z-v[i])/b<=-1 then do;
suma[i]=0;
end;
if (z-v[i])/b>=1 then do;
suma[i]=1/nr;
end;
end;
rr=sum(suma)-alpha;
return(rr);
finish;

/* Newton */
converge=1E-10;
start newton;
/*z is a scalar, v is a vector*/
ff=FUN(z,v,nr,b,alpha);
do iter2=1 to maxiter2 while
(abs(ff)>converge);
dd=DERIV(z,v,nr,b);
delta=-solve(dd,ff);
z=z+delta;
ff=FUN(z,v,nr,b,alpha);
end;
finish;

/* Variables */
nr=nrow(xx);
nc=ncol(xx);

reset fw=30;
alpha=0.95;
emp=j(nr,1,0);

rati=j(nc,1,0);
Var_Beta=j(nc,1,0);
fun=j(nc,1,0);
it=j(nc,1,0);

/* Inverse Beta distrtribution */
lim=1;
aa=0;
aa=Tcap2(aa,lim);

/* Beta*/
*b_Beta=1.0874*nr**(-1/3);
*b_Beta=1.4422*nr**(-1/3);
gy=(betainv(alpha,3,3)*2)-1;
my=(15/16)*(1-gy**2)**2;
dmy=((15/4)*gy*(gy**2-1)**2);
b_beta=((my*(9/35))/((1/5)**2*dmy))**(1/3)
*(nr**(-1/3));

/* VaR Calculation*/
do l=1 to nc;
/*Original data*/
x=xx[,l];
x=ordena(x);
/* Parameters Male population*/
/* Champernowne */
c=0;
M= 75; *MLE scale;
aa= 16.48194; *MLE shape;
*aa= 14.393958; *MISE M fiyo;
*M= 74.827807; *MISE;
*aa= 14.442683; *MISE;

/* Parameters Female population*/
/* Champernowne */
*c=0;
*M= 76; *MLE scale;
*a= 15.83281; *MLE shape;
*a= 13.696014; *MISE M fiyo;
*M= 75.534849; *MISE;

```

```

*a= 13.821771; *MISE;

lim=1;
do j=1 to nr;
    ind=(x<=x[j]);
    emp[j]=sum(ind)/nr;
end;
/*Estimation */
ind=(emp>alpha);
sind=sum(ind);
if sind>0 then do;
    ni=nr-sind+1;
    Var_hat=x[ni];
end;
if sind=0 then do;
    ni=nr-sind;
    Var_hat=x[ni];
end;
Vin=Tcap1(Var_hat,a,c,M,lim);
ini=0;
ini=Tcap2(ini,vin);
/*Champernowne transformation kernel*/
w=Tcap1(x,a,c,M,lim);

y=j(nr,1,0);
y=Tcap2(y,w);
z=ini;
v=y;
b=b_Beta;
run newton;
zy=(z+1)/2;
zy=CDF('BETA',zy,3,3);
zx=((zy*(M+c)**a-(2*zy-1)*c**a)/
(1-zy))**(1/a)-c;
Var_Beta[l]=zx;
fun[l]=ff;
it[l]=iter2;
end;

maf=max(ff); mif=min(ff);
mait=max(it); miit=min(it);

result=Var_Beta;
print 'alpha=0.95';
print result;

QUIT;

```

Cálculo del valor en riesgo exacto de la distribución truncada DTKE

- Calculo del VaR exacto usando el estimador núcleo doble transformado, asumiendo como primera transformación a la distribución Lognormal.

Script Lognormal Truncada

```

libname simul '?';
proc iml SYMSIZE=5000;

use simul.man65;
read all into xx;

* use simul.woman65;
* read all into xx;

/* Function: Order */
maxiter2=100000;
start ordena(v);
v0=v;
v[rank(v)]=v0;
rr=v;

return(rr);
finish;

/*Transformation: Lognormal*/
start Tcap1(v,mu,sigma,lim);
/*v is a data vector*/
n=nrow(v);
rr=cdf('LOGNORMAL',v,mu,sigma);
rr=(rr#(lim-(1-lim)))+(1-lim);
return(rr);
finish;

/*Transformation: Inverse Beta distrtribution*/
start Tcap2(v,w);
/*v is a vector*/

```

```

n=nrow(v);
do i=1 to n;
  v[i]=(betainv(w[i],3,3)*2)-1;
end;
rr=v;
return(rr);
finish;

/*Epanechnikoc Kernel*/
start K(u);
  rr=0.75*(1-(u#u))#(abs(u)<=1);
  return (rr);
finish;

start kintd(u);
  rr=-0.25#((1+u)##2)#(u-2);
  return (rr);
finish;

/* Functions */
start DERIV(z,v,nr,b);
  /*z is a scalar, v is a vector*/
  rr=sum(K((z*j(nr,1,1)-v)/b))/(nr*b);
  return(rr);
finish;

start cdfhatz(z,v,nr,b);
  /*z is a scalar, v is a vector*/
  rr=Kintd((z-v)/b)/nr;
  return(rr);
finish;

start FUN(z,tru,v,nr,b,alpha);
  /*z is a scalar, v is a vector*/
  suma=j(nr,1,0);
  do i=1 to nr;
    if abs(z-v[i])/b<=1 then do;
      suma[i]=cdfhatz(z,v[i],nr,b);
      end;
    if (z-v[i])/b<=-1 then do;
      suma[i]=0;
      end;
    if (z-v[i])/b>=1 then do;
      suma[i]=1/nr;
      end;
  end;
  sumat=j(nr,1,0);
  do i=1 to nr;
    if abs(tru-v[i])/b<=1 then do;
      sumat[i]=cdfhatz(tru,v[i],nr,b);
      end;
    if (tru-v[i])/b<=-1 then do;
      sumat[i]=0;
      end;
  end;

```

```

  if (tru-v[i])/b>=1 then do;
    sumat[i]=1/nr;
    end;
  end;
  rr=((sum(suma)-sum(sumat)))/
  (1-sum(sumat))-alpha;
  return(rr);
finish;

start DERIV(z,tru,v,nr,b);
  /*z is a scalar, v is a vector*/
  rr=sum(K((z*j(nr,1,1)-v)/b))/(nr*b);
  sumat=j(nr,1,0);
  do i=1 to nr;
    if abs(tru-v[i])/b<=1 then do;
      sumat[i]=cdfhatz(tru,v[i],nr,b);
      end;
    if (tru-v[i])/b<=-1 then do;
      sumat[i]=0;
      end;
    if (tru-v[i])/b>=1 then do;
      sumat[i]=1/nr;
      end;
  end;
  rr=rr/(1-sum(sumat));
  return(rr);
finish;

/* Newton */
converge=1E-10;
start newton;
  /*z is a scalar, v is a vector*/
  ff=FUN(z,tru,v,nr,b,alpha);
  do iter2=1 to maxiter2 while(abs(ff)>converge);
    dd=DERIV(z,tru,v,nr,b);
    delta=-solve(dd,ff);
    z=z+delta;
    ff=FUN(z,tru,v,nr,b,alpha);
  end;
finish;

/* Variables */
nr=nrow(xx);
nc=ncol(xx);

alpha=0.99;
emp=j(nr,1,0);
reset fw=30;

rati=j(nc,1,0);
Var_Beta=j(nc,1,0);
fun=j(nc,1,0);
it=j(nc,1,0);

```



```

/* Inverse Beta distribution */
lim=1;
aa=0;
aa=Tcap2(aa,lim);

/* Beta*/
gy=(betainv(alpha,3,3)*2)-1;
my=(15/16)*(1-gy**2)**2;
dmy=((15/4)*gy*(gy**2-1))**2;
b_beta=((my*(9/35))/((1/5)**2*dmy))**
(1/3)*(nr**(-1/3));

/* VaR Calculation */
do l=1 to nc;
  /*Original data*/
  x=xx[,l];
  x=ordena(x);
  /* Parameters Male population*/
  /* Lognormal */
  mu= 4.328013635; *MLE scale;
  sigma= 0.102264784; *MLE shape;
  *sigma= 0.118113; *MISE mu fijo;
  *mu= 4.315576; *MISE;
  *sigma= 0.114779; *MISE;

  /* Parameters Female population*/
  /* Lognormal */
  *mu= 4.33669801; *MLE scale;
  *sigma= 0.105999045; *MLE shape;
  *sigma= 0.122632; *MISE mu fijo;
  *mu= 4.324958; *MISE;
  *sigma= 0.119931; *MISE;

lim=1;
do j=1 to nr;
  ind=(x<=x[j]);
  emp[j]=sum(ind)/nr;
end;
/*Estimation */

```

```

ind=(emp>alpha);
sind=sum(ind);
if sind>0 then do;
  ni=nr-sind+1;
  Var_hat=x[ni];
end;
if sind=0 then do;
  ni=nr-sind;
  Var_hat=x[ni];
end;
Vin=Tcap1(Var_hat,mu,sigma,lim);
ini=0;
ini=Tcap2(ini,vin);
/*Lognormal transformation kernel*/
w=Tcap1(x,mu,sigma,lim);
y=j(nr,1,0);
y=Tcap2(y,w);
tru=Tcap1(75,mu,sigma,lim);
truy=1;
tru=Tcap2(truy,tru);
z=ini;
v=y;
b=b_Beta;
run newton;
zy=(z+1)/2;
zy=CDF('BETA',zy,3,3);
zx=quantile('LOGNORMAL',zy,mu,sigma);
Var_Beta[]=zx;
fun[l]=ff;
it[l]=iter2;
end;
maf=max(ff); mif=min(ff);
mait=max(it); miit=min(it);

result=Var_Beta;
print 'alpha=0.99';
print result;

QUIT;

```

Bootstrap

Script

```

libname simul '?';
proc iml;
  use simul.men;
  read all into x;

```

```

n=nrow(x);
b=1000;
xb=j(n,b,0);

do i=1 to b;

```

```
rat=int((uniform(j(n,1,i+100))#(n-1)+1));
xb[,i]=x[rat,];
end;
```

```
create cost.men_boots from xb;
append from xb;
close cost.men_boots;
quit;
```


Gráficos en R

Los gráficos del capítulo 5 se realizaron con el software R Core Team (2014).

Histogramas

- Crea histogramas superponiendo la función de densidad Weibull, Lognormal y Champernowne, con los parámetros estimados por MLE y minimizando el \widehat{MISE} .

Script

```
library(MASS)
library(gdata)
library(sas7bdat)
library(FAdist)

#####
#### CHAMPERNOWNE DENSITY ####
dchamp = function(x,par){
  a=par[1]
  M=par[2]
  c=par[3]
  (a*(x+c)^(a-1)*((M+c)^a-c^a))/((x+c)^a+(M+c)^a-2*c^a)^2
}
#####
getwd(); setwd("/Path/")

#### DATA ####
alive<-as.matrix(read.sas7bdat("man65.sas7bdat"))
#ALIVE<-AS.MATRIX(READ.SAS7BDAT("WOMAN65.SAS7BDAT"))

#### HISTOGRAM WITH DENSITY FUNCTION ####
x11()
par(mfrow=c(1,3),las=1)
hist(alive,breaks = seq(65, 107, by=1),right=F,main=' ', xlab='Male',freq=F,xlim=c(65,110))
curve(dweibull(x, scale=79.88856862, shape=9.70099632), col = 1, lty = 2, lwd = 2, add = TRUE)
curve(dweibull(x, scale=78.28556, shape=9.782975), col = 1, lty = 3, lwd = 2, add = TRUE)
legend('topright', c('Weibull MLE', 'Weibull MISE'), lty=2:3, col=1, bty='n',lwd=2)

hist(alive,breaks = seq(65, 107, by=1),right=F,main=' ', xlab='Male',freq=F,xlim=c(65,110))
curve(dlnorm(x, meanlog = 4.328013635, sdlog = 0.102264784), col = 1, lty = 2, lwd = 2, add = TRUE)
curve(dlnorm(x, meanlog = 4.315576, sdlog = 0.114779), col = 1, lty = 3, lwd = 2, add = TRUE)
legend('topright', c('Lognormal MLE', 'Lognormal MISE'), lty=2:3, col=1, bty='n',lwd=2)

hist(alive,breaks = seq(65, 107, by=1),right=F, main=' ', xlab='Male',freq=F,xlim=c(65,110))
```

```

curve(dchamp(x, c(a=16.48194, M=75,c=0)), col = 1, lty = 2, lwd = 2, add = TRUE)
curve(dchamp(x, c(a=14.442683, M=74.827807,c=0)), col = 1, lty = 3, lwd = 2, add = TRUE)
legend('topright', c('Champernowne MLE', 'Champernowne MISE'), lty=2:3, col=1, bty='n',lwd=2)

#### TWO HISTOGRAMS ###
x11()
par(mfrow=c(1,2), las=1, cex=1.2, oma=c(0,0,0,0))
hist(alive, breaks = seq(65, 107, by=1),right=F, freq=F, main="", xlim=c(65,98), xlab='Age (x)',
      ylab='Density')
hist(alive, breaks = seq(65, 107, by=1),right=F, freq=F, main="", xlim=c(99,107), ylim= c(0, 0.0015),
      xlab='Age (x)', ylab="")

```

Parámetros estimados minimizando MISE

- Crea el gráfico “contour” del valor del \widehat{MISE} cuando los parámetros de las diferentes distribuciones varían hasta encontrar el mínimo error.

Script

```

library(sas7bdat)
library(lattice)

#####
#### READING SUMA=VECTOR CREATED BY 8.MISE.SAS ####
x<-read.sas7bdat("sumawei.sas7bdat"); z<-read.sas7bdat("sumaln.sas7bdat");
y<-read.sas7bdat("sumach.sas7bdat")

#### PLOT PARAMETERS VS. MISE ####
# WEIBULL # #GR\AFICAS DE NIVEL
x11()
par(lab=c(3,20,1),cex.axis=1.1,las=1,lwd=1,cex.lab=1.1,oma=c(0,0,0,0),bty='o')
levelplot(x$suma ~ x$aa * x$ss,xlab='Shape Parameter',ylab='Scale Parameter',col.regions =
  gray.colors(50), colorkey=list(at=seq(100,700,100)),
  scales=list(x=list(at=seq(8.5,11,.5)), y=list(at=seq(77,80,.5))))

# LOGNORMAL #
x11()
par(lab=c(3,20,1),cex.axis=1.1,las=1,lwd=1,cex.lab=1.1,oma=c(0,0,0,0),bty='o')
levelplot(z$suma ~ z$aa * z$ss,xlab='Shape Parameter',ylab='Scale Parameter',col.regions =
  gray.colors(50), colorkey=list(at=seq(37,38,.1)),
  scales=list(x=list(at=seq(.114,0.115,.0001)), y=list(at=seq(4.315,4.316,.0001))))

# CHAMPERNOWNE #
x11()
par(lab=c(3,20,1),cex.axis=1.1,las=1,lwd=1,cex.lab=1.1,oma=c(0,0,0,0),bty='o')
levelplot(y$suma ~ y$aa * y$ss,xlab='Shape Parameter',ylab='Scale Parameter',col.regions =
  gray.colors(50), colorkey=list(at=seq(40,180,20)),
  scales=list(x=list(at=seq(13,16,.2)), y=list(at=seq(74,76,.2))))

```

Funciones de distribución

- Compara las funciones de distribución obtenidas por los dos métodos de estimación de parámetros: MLE y minimizando \widehat{MISE} ,
- Compara las funciones de distribución obtenidas por el estimador núcleo y el estimador núcleo doble transformado: para las distribuciones Weibull, Lognormal y Champernowne calculadas con los parámetros estimados minimizando \widehat{MISE} .

Script Parametric CDF

```
library(sas7bdat)

#####
#### READING CDF ####
cdf<-read.sas7bdat("fx.sas7bdat")

#### DEFINING RANGE OF VALUES TO GRAPH ####
cdf<-subset(cdf, cdf$x>64.9 & cdf$x <110.1)
x<-cdf[,1]

#####          GRAPHS          #####
#### WEIBULL ####
x11()
par(mfrow=c(1,1), lab=c(10,10,2), cex.axis=1.2, las=1, lwd=1.5, cex.lab=1.3, oma=c(0,0,0,0), bty='o',
    pch=8)
plot(x, cdf[, 'emp'], type="l", xlab="Age", ylab="CDF", col=1, lwd=2, lty=1, bg=1, main='Weibull')
  lines(x, cdf[, 'Wei'], col=1, lwd=2, lty=2)
  lines(x, cdf[, 'WeiISE2'], col=1, lwd=2, lty=6)
legend('bottomright', c('Emp', 'MLE', 'Min MISE'), lty=c(1,2,6), col=1, bty='n', lwd=2, cex=1.4)
grid()

#### LOGNORMAL ####
x11()
par(mfrow=c(1,1), lab=c(10,10,2), cex.axis=1.2, las=1, lwd=1.5, cex.lab=1.3, oma=c(0,0,0,0), bty='o',
    pch=8)
plot(x, cdf[, 'emp'], type="l", xlab="Age", ylab="CDF", col=1, lwd=2, lty=1, bg=1, main='Lognormal')
  lines(x, cdf[, 'Ln'], col=1, lwd=2, lty=2)
  lines(x, cdf[, 'LnISE2'], col=1, lwd=2, lty=6)
legend('bottomright', c('Emp', 'MLE', 'Min MISE'), lty=c(1,2,6), col=1, bty='n', lwd=2, cex=1.4)
grid()

#### CHAMPERNOWNE ####
x11()
par(mfrow=c(1,1), lab=c(10,10,2), cex.axis=1.2, las=1, lwd=1.5, cex.lab=1.3, oma=c(0,0,0,0), bty='o',
    pch=8)
plot(x, cdf[, 'emp'], type="l", xlab="Age", ylab="CDF", col=1, lwd=2, lty=1, bg=1,
    main='Champernowne')
```

```

lines(x, cdf[, 'Champ'], col=1, lwd=2, lty=2)
lines(x, cdf[, 'ChampISE2'], col=1, lwd=2, lty=6)
legend('bottomright', c('Emp', 'MLE', 'Min MISE'), lty=c(1,2,6), col=1, bty='n', lwd=2, cex=1.4)
grid()

```

Script Nonparametric CDF

```

library(sas7bdat)

#####
#### READING CDF ####
cdf<-read.sas7bdat("fx.sas7bdat")

#### DEFINING RANGE OF VALUES TO GRAPH ####
cdf<-subset(cdf, cdf$x>94.9 & cdf$x <105.1)
x<-cdf[,1]

#####          GRAPHS          #####
#### CKE - DTKE WEIBULL ####
x11()
par(mfrow=c(1,1), lab=c(6,3,1), cex.axis=1.2, las=0, lwd=1.5, cex.lab=1.3, oma=c(0,0,0,0), bty='o',
    pch=8)
plot(x,cdf[, 'emp'], type="l", xlab="Age", ylab="CDF", col=1, lwd=2, lty=1, bg=1)
  lines(x, cdf[, 'Clas'], col=1, lwd=2, lty=2)
  lines(x, cdf[, 'BWeIISE2'], col=1, lwd=2, lty=6)
legend('bottomright', c('Emp', 'CKE', 'DTKE Weibull'), lty=c(1,2,6), col=1, bty='n', lwd=2, cex=1.4)

#### CKE - DTKE LOGNORMAL ####
x11()
par(mfrow=c(1,1), lab=c(6,5,1), cex.axis=1.2, las=0, lwd=1.5, cex.lab=1.3, oma=c(0,0,0,0), bty='o',
    pch=8)
plot(x,cdf[, 'emp'], type="l", xlab="Age", ylab="CDF", col=1, lwd=2, lty=1, bg=1)
  lines(x, cdf[, 'Clas'], col=1, lwd=2, lty=2)
  lines(x, cdf[, 'BLnISE2'], col=1, lwd=2, lty=6)
legend('bottomright', c('Emp', 'CKE', 'DTKE Lognormal'), lty=c(1,2,6), col=1, bty='n', lwd=2,
    cex=1.4)

#### CKE - DTKE CHAMPERNOWNE ####
x11()
par(mfrow=c(1,1), lab=c(6,5,1), cex.axis=1.2, las=0, lwd=1.5, cex.lab=1.3, oma=c(0,0,0,0), bty='o',
    pch=8)
plot(x,cdf[, 'emp'], type="l", xlab="Age", ylab="CDF", col=1, lwd=2, lty=1, bg=1)
  lines(x, cdf[, 'Clas'], col=1, lwd=2, lty=2)
  lines(x, cdf[, 'BChampISE2'], col=1, lwd=2, lty=6)
legend('bottomright', c('Emp', 'CKE', 'DTKE Champernowne'), lty=c(1,2,6), col=1, bty='n', lwd=2,
    cex=1.4)

```

Valor en riesgo

Script

```

library(sas7bdat)

#####
#### READING EMPIRICAL VAR ####
vemp<-read.sas7bdat("vemp.sas7bdat")

#### READING TRUNCATED VAR - AGE 65 AND 85 ####
var65<-read.sas7bdat("var65.sas7bdat")
var65[,1]<-vemp[,1]

var85<-read.sas7bdat("var85.sas7bdat")
var85[,1]<-vemp[,5]

#### KQE VAR - AGE 65 AND 85 ####
kqe65<-read.sas7bdat("kqe65.sas7bdat")
kqe85<-read.sas7bdat("kqe85.sas7bdat")

#### ALPHA ####
(alpha<-seq(.9,.999,.001) )
la<-length(alpha)

#### VAR FROM PARAMETRIC DISTRIBUTIONS####
var65[, 'Wei'] <-as.vector(VaRWmvh)
var65[, 'WeiISE2'] <-as.vector(VaRWISE2h)

var65[, 'Ln'] <-as.vector(VaRLnmvh)
var65[, 'LnISE2'] <-as.vector(VaRLnISE2h)

var65[, 'Champ'] <-as.vector(VaRChmvh)
var65[, 'ChampISE2'] <-as.vector(VaRChISE2h)

#### LABELS AND COLOR ####
nopara_name<- c("EDF", "CKE", "KQE", "DTKE Weibull", "DTKE LN", "DTKE Champernowne")
lcol<-c('chartreuse3', 'aquamarine4', 'dodgerblue3', 'firebrick3', 'darkslateblue', 'darkgoldenrod2')
posnamep<- c("EDF", "Weibull", "LN", "Champernowne", "Weibull LNO", "LN LNO",
"Champernowne LNO" )
lcol<-c('chartreuse3', 'aquamarine4', 'dodgerblue3', 'firebrick3',
'darkslateblue', 'darkgoldenrod2', 'deppink1')

##### AGE 65 #####
#### NONPARAMETRIC ####
a<-95; b<-100 #### DEFINING RANGE OF AGES TO GRAPH ####
x11()
par(mfrow=c(1,1), lab=c(3,20,1), cex.axis=1.2, las=1, lwd=1.5, cex.lab=1.3, oma=c(0,0,0,0), bty='o',
pch=8)
plot(alpha[a:b], var65[a:b,1], xlab="Confidence Level", ylab="Value at Risk", xaxt='n', yaxt='n',
lwd=2, bg=1, lty=1, type='l', main=65, col=lcol[1], ylim=c(a,b))
lines(alpha[a:b], var65[a:b,2], col=lcol[2], type='l', lty=2, lwd=2)
lines(alpha[a:b], kqe65[a:b,1], col=lcol[3], type='l', lty=3, lwd=2)
lines(alpha[a:b], var65[a:b,18], col=lcol[4], type='l', lty=4, lwd=2)
lines(alpha[a:b], var65[a:b,19], col=lcol[5], type='l', lty=5, lwd=2)
lines(alpha[a:b], var65[a:b,20], col=lcol[6], type='l', lty=2, lwd=2)

```



```

axis(1, c(.9,.93,.95,.97,.99,.995,.999))
axis(2,seq(a,b,2))
legend('topleft', nopara_name,lwd=2, col=lcol, bty='n',cex=.9,lty=c(1:5,2))

#### PARAMETRIC ####
c<-85; d<-110      #### DEFINING RANGE OF AGES TO GRAPH ####
x11()
par(mfrow=c(1,1), lab=c(3,20,1), cex.axis=1.2, las=1, lwd=1.5, cex.lab=1.3, oma=c(0,0,0,0), bty='o',
    pch=8)
plot(alpha[c:d], var65[c:d,1], xlab="Confidence Level", ylab="Value at Risk", xaxt='n', yaxt='n',
     lwd=2, bg=1, lty=1, type='l', main=65, col=lcol[1], ylim=c(c,d))
lines(alpha[c:d],var65[c:d,3],col=lcol[2],type='l',lty=2, lwd=1)
lines(alpha[c:d],var65[c:d,4],col=lcol[3],type='l',lty=3, lwd=1)
lines(alpha[c:d],var65[c:d,5],col=lcol[4],type='l',lty=4, lwd=1)
lines(alpha[c:d],var65[c:d,15],col=lcol[5],type='l',lty=2, lwd=2)
lines(alpha[c:d],var65[c:d,16],col=lcol[6],type='l',lty=3, lwd=2)
lines(alpha[c:d],var65[c:d,17],col=lcol[7],type='l',lty=4, lwd=2)
axis(1, c(.9,.93,.95,.97,.99,.995,.999))
axis(2,seq(c,d,2))
legend('topleft', posnamep,lwd=c(rep(1,4),rep(2,3)), col=lcol, bty='n',cex=.9,lty=c(1,2:4,2:4))

##### AGE 85 #####
#### NONPARAMETRIC ####
a<-98; b<-103     #### DEFINING RANGE OF AGES TO GRAPH ####
x11()
par(mfrow=c(1,1), lab=c(3,20,1), cex.axis=1.2, las=1, lwd=1.5, cex.lab=1.3, oma=c(0,0,0,0), bty='o',
    pch=8)
plot(alpha[a:b], var85[a:b,1], xlab="Confidence Level", ylab="Value at Risk", xaxt='n',yaxt='n', lwd=2,
     bg=1, lty=1, type='l', main=85, col=lcol[1], ylim=c(a,b))
lines(alpha[a:b], var85[a:b,2], col=lcol[2], type='l', lty=2, lwd=2)
lines(alpha[a:b], kqe85[a:b,1], col=lcol[3], type='l', lty=3, lwd=2)
lines(alpha[a:b], var85[a:b,18], col=lcol[4], type='l', lty=4, lwd=2)
lines(alpha[a:b], var85[a:b,19], col=lcol[5], type='l', lty=5, lwd=2)
lines(alpha[a:b], var85[a:b,20], col=lcol[6], type='l', lty=2, lwd=2)
axis(1, c(.9,.93,.95,.97,.99,.995,.999))
axis(2,seq(a,b,2))
legend('topleft', posname, lwd=2, col=lcol, bty='n', cex=.9, lty=c(1:5,2))

#### PARAMETRIC ####
c<-94; d<-120     #### DEFINING RANGE OF AGES TO GRAPH ####
x11()
par(mfrow=c(1,1), lab=c(3,20,1), cex.axis=1.2, las=1, lwd=1.5, cex.lab=1.3, oma=c(0,0,0,0), bty='o',
    pch=8)
plot(alpha[c:d], var85[c:d,1], xlab="Confidence Level", ylab="Value at Risk", xaxt='n', yaxt='n',
     lwd=2, bg=1, lty=1, type='l', main=85, col=lcol[1], ylim=c(c,d))
lines(alpha[c:d], var85[c:d,3], col=lcol[2], type='l', lty=2, lwd=1)
lines(alpha[c:d], var85[c:d,4], col=lcol[3], type='l', lty=3, lwd=1)
lines(alpha[c:d], var85[c:d,5], col=lcol[4], type='l', lty=4, lwd=1)
lines(alpha[c:d], var85[c:d,15], col=lcol[5], type='l', lty=2, lwd=2)
lines(alpha[c:d], var85[c:d,16], col=lcol[6], type='l', lty=3, lwd=2)
lines(alpha[c:d], var85[c:d,17], col=lcol[7], type='l', lty=4, lwd=2)
axis(1, c(.9,.93,.95,.97,.99,.995,.999))
axis(2,seq(c,d,2))
legend('topleft', posnamep, lwd=c(rep(1,4), rep(2,3)), col=lcol, bty='n', cex=.9, lty=c(1,2:4,2:4))

```