



Universidad Nacional Mayor de San Marcos

Universidad del Perú. Decana de América

Dirección General de Estudios de Posgrado

Facultad de Ciencias Matemáticas

Unidad de Posgrado

**Estimación de una tabla de vida para el Sistema
Nacional de Pensiones Peruano**

TESIS

Para optar el Grado Académico de Magíster en Estadística
Matemática

AUTOR

José Artemio VALDERRAMA TORRES

ASESOR

Antonio BRAVO QUIROZ

Lima, Perú

2019



Reconocimiento - No Comercial - Compartir Igual - Sin restricciones adicionales

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

Usted puede distribuir, remezclar, retocar, y crear a partir del documento original de modo no comercial, siempre y cuando se dé crédito al autor del documento y se licencien las nuevas creaciones bajo las mismas condiciones. No se permite aplicar términos legales o medidas tecnológicas que restrinjan legalmente a otros a hacer cualquier cosa que permita esta licencia.

Referencia bibliográfica

Valderrama, J. (2019). *Estimación de una tabla de vida para el Sistema Nacional de Pensiones Peruano*. [Tesis de maestría, Universidad Nacional Mayor de San Marcos, Facultad de Ciencias Matemáticas, Unidad de Posgrado]. Repositorio institucional Cybertesis UNMSM.

Metadatos complementarios

Datos de autor	
Nombres y apellidos	JOSE ARTEMIO VALDERRAMA TORRES
DNI	10815474
URL de ORCID	https://orcid.org/0000-0002-2736-6526
Datos de asesor	
Nombres y apellidos	ANTONIO BRAVO QUIROZ
DNI	10130035
URL de ORCID	https://orcid.org/0000-0001-9051-2808
Datos de investigación	
Línea de investigación	A.3.2.5. Inferencia y métodos estadísticos (Gráficos, nuevas medidas)
Grupo de investigación	No aplica.
Agencia de financiamiento	Sin financiamiento.
Ubicación geográfica de la investigación	Distrito de Pueblo Libre Latitud: -12.074099 Longitud: -77.069884
Año o rango de años en que se realizó la investigación	Abril 2018 - abril 2019
URL de disciplinas OCDE	1.01.03 - Estadísticas, probabilidad https://purl.org/pe-repo/ocde/ford#1.01.03

ACTA DE SUSTENTACIÓN DE TESIS DE GRADO ACADÉMICO DE MAGÍSTER

Siendo las, 17:45 horas del día miércoles trece de noviembre del dos mil diecinueve, en el Auditorio de la Facultad de Ciencias Matemáticas, el Jurado Evaluador de Tesis, Presidido por el Mg. Wilfredo Eugenio Domínguez Cirilo e integrado por los siguientes miembros, Mg. Olga Lidia Solano Dávila (Jurado Informante), Mg. Rosa Ysabel Adriazola Cruz (Jurado Evaluador), y el Mg. Antonio Bravo Quiroz como Miembro Asesor, se reunieron para la sustentación de la tesis titulada: «ESTIMACIÓN DE UNA TABLA DE VIDA PARA EL SISTEMA NACIONAL DE PENSIONES PERUANO» presentada por el Bachiller **José Artemio Valderrama Torres** para optar el Grado Académico de Magíster en Estadística Matemática.

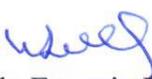
Luego de la exposición del graduando, los Miembros del Jurado hicieron las preguntas correspondientes, así como las observaciones e inquietudes acerca del trabajo de tesis, a las cuales el Bachiller **José Artemio Valderrama Torres** respondió con acierto y solvencia, demostrando pleno conocimiento del tema.

A continuación se realizó la calificación correspondiente, según tabla adjunta, resultando el Bachiller **José Artemio Valderrama Torres** aprobado con el calificativo de 16..... (BUENO)...

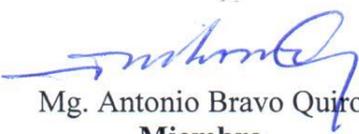
Habiendo sido aprobada la sustentación de la Tesis, el Jurado Evaluador recomienda para que el Consejo de Facultad apruebe el otorgamiento del Grado Académico de **MAGÍSTER EN ESTADÍSTICA MATEMÁTICA** al **Bachiller José Artemio Valderrama Torres**.

Siendo las 18:45 horas, se levantó la sesión, firmando para constancia la presente Acta.


Mg. Olga Lidia Solano Dávila
Miembro


Mg. Wilfredo Eugenio Domínguez Cirilo
Presidente


Mg. Rosa Ysabel Adriazola Cruz
Miembro


Mg. Antonio Bravo Quiroz
Miembro

Índice

1	Introducción	1
1.1	Situación problemática	1
1.2	Formulación del problema	3
1.2.1	Problema principal	3
1.2.2	Problemas secundarios	3
1.3	Justificación de la investigación	4
1.4	Objetivos de la investigación	4
1.4.1	Objetivo general	4
1.4.2	Objetivos específicos	5
2	Marco Teórico	6
2.1	Antecedentes del problema	6
2.2	Bases teóricas	7
2.2.1	Modelos de sobrevivencia	7
2.2.2	Tabla de vida	10
2.2.3	Modelo aleatorio de censura y fallecimiento	19
2.3	Marco conceptual	22
2.3.1	Conceptualización de una tabla de vida	22
2.3.2	Reserva matemática	24
2.4	Hipótesis y variables	26
2.4.1	Hipótesis general	26
2.4.2	Hipótesis específicas	26
2.4.3	Identificación de variables	26
3	Metodología	27
3.1	Tipo y diseño de investigación	27
3.2	Unidad de análisis	28
3.3	Población de Estudio	28
3.4	Tamaño de Muestra	28
3.5	Selección de Muestra	28
3.6	Técnicas de recolección de datos	30

3.7	Análisis	30
3.7.1	Tasa bruta de mortalidad	32
3.7.2	Técnicas de graduación	36
3.7.3	Estimación en edades tempranas y avanzadas	41
3.7.4	Validación	42
4	Resultados y discusión	43
4.1	Estimación según etapas	43
4.1.1	Tasa cruda o bruta de mortalidad	43
4.1.2	Graduación	44
4.1.3	Estimación en edades tempranas y avanzadas	47
4.1.4	Validación	48
4.2	Análisis de robustez	49
5	Impactos	57
6	Conclusiones	59
7	Recomendaciones	59
A	Anexos	61
A.1	Muestra de datos utilizados	61
A.2	Criterios de información para selección de modelo para extrapolación	62
A.3	Función Kannisto: Estimación	63
A.4	Metodología SP 2005	64
A.5	Tabla de mortalidad	65
	Bibliografía	68

Índice de tablas

1	Tabla de vida hipotética	24
2	Descripción de los datos utilizados	30
3	Tests estadísticos seleccionadas. P-Values	46
4	Expectativa de vida ^{1,2}	48
5	Esperanza de vida según escenario	58
6	Impacto en las reservas matemáticas 2017 (en millones de soles) .	58
7	Muestra de datos utilizados	61
8	AIC según modelos y rango de edades	62
9	BIC según modelos y rango de edades	62
10	Estimación de modelos ^a	63
11	Tabla de mortalidad-Hombres. (SNP 2017)	66
12	Tabla de mortalidad-Mujeres.(SNP 2017)	67

Índice de figuras

1	Ratio entre probabilidades y esperanza de vida a los 65 años en países seleccionados.	2
2	Tasa bruta según tipo de estimador. MV:Máxima verosimilitud; MM:Método de momentos; MA: Método actuarial; y KM: Kaplan Meier.	44
3	Graduación de acuerdo a la regresión polinómica local.	45
4	Probabilidades graduadas e intervalos de confianza. Intervalos con una, dos y tres desviaciones estándar. El valor central corresponde a la variable graduada y los intervalos de confianza corresponden a las tasas brutas con errores estándar correspondientes al estimador Kaplan-Meier.	51
5	Proyección en colas inferior y superior según sexo.	52
6	Logaritmo de probabilidades de fallecimiento según tablas.	53
7	Esperanza de vida según tablas.	54
8	Sensibilidad de las probabilidades de fallecimiento. Panel superior son probabilidades en logaritmos y panel inferior son ratios respecto a metodología propuesta.	55
9	Sensibilidad de las esperanzas de vida. Panel superior son esperanzas de vida y panel inferior son ratios respecto a metodología propuesta.	56

Resumen

Una tabla de mortalidad es estimada empleando la estrategia no paramétrica desarrollada por Kaplan-Meier para la estimación de las tasas brutas mientras que el suavizamiento o graduación de estos estimadores se calculan empleando una regresión polinómica local. Los datos provienen de registros administrativos y acumulan alrededor de 5 millones de personas, información que es colectada entre los años 2013 al 2017. De acuerdo a los resultados el diferencial en esperanza de vida entre mujeres y hombres es menor en relación a los resultados oficiales. De esta manera, la diferencia en esperanza de vida a favor de la mujer, se reduce de 7 a 2 años, brecha que se explica por un descenso de 2.4 años en las mujeres y un aumento de 2.1 años en el caso de los hombres (medido a los 65 años de edad). Como consecuencia de lo anterior se encuentra que el pasivo actuarial está sobreestimado en $S/2,03$ mil millones en relación al cálculo oficial, ello como consecuencia de dos efectos opuestos: una disminución en el valor del pasivo en el caso de las mujeres, y un aumento para los hombres, siendo el primero el de mayor magnitud.

Palabras clave: Kaplan y Meier, regresión polinómica local, Tabla de vida.

Clasificación JEL: H55, J11.

Abstract

A mortality table is estimated using the non-parametric strategy developed by Kaplan-Meier for the estimation of gross rates while the smoothing or graduation of these estimators are calculated using a local polynomial regression. The data come from administrative records and accumulate around 5 million people, information that is collected between 2013 and 2017. According to the results, the differential in life expectancy between women and men is smaller in relation to the official results. In this way, the difference in life expectancy in favor of women is reduced from 7 to 2 years, a gap that is explained by a decrease of 2.4 years in women and an increase of 2.1 years in the case of men (measured at 65 years of age). As a consequence of the foregoing, it is found that the actuarial liability is overestimated at $S/2.03$ billion in relation to the official calculation, as a result of two opposite effects: a decrease in the value of the liability in the case of women, and an increase for men, the first being the largest.

Keywords: Kaplan and Meier, local polynomial regression, Life table.

Clasificación JEL: H55, J11.

1 Introducción

1.1 Situación problemática

El Sistema Nacional de Pensiones (SNP) es un sistema de reparto regido por el Decreto Legislativo N° 19990, por la lógica de su funcionamiento los afiliados, mediante sus aportes, contribuyen con el financiamiento de la planilla de pensionistas. No todos los afiliados tienen derecho a una pensión, además del requisito mínimo de 65 años de edad, el afiliado debe haber aportado al menos veinte años durante su vida laboral en un monto de al menos el 13 % del salario mínimo legal vigente al momento de realizar el aporte.

A la fecha el sistema es deficitario. Los ingresos de los aportes no son suficientes para financiar el pago de las pensiones. En el año 2018 el 13 % del pago de las pensiones fue financiado con transferencias del tesoro. Aunque esta información es anual, el pasivo o deuda que el estado adquiere con los pensionistas se manifiesta con el otorgamiento de una pensión vitalicia. Como la fecha de fallecimiento es incierta, calcular el dinero necesario para financiar todos los flujos futuros de pensiones es un reto con el que debe lidiar cualquier institución encargada de gestionar pensiones. La tabla de vida o tabla de mortalidad es el instrumento en el que se modelan las probabilidades de sobrevivencia o fallecimiento las mismas que son usadas para estimar los flujos futuros.

En ese sentido una tabla de mortalidad es el producto de un modelo actuarial o estadístico que informa cómo el tamaño de un universo hipotético (100,000 personas por ejemplo) se va desgastando desde el momento del nacimiento hasta que fallezca la última persona. La tabla muestra la probabilidad condicional de muerte en los siguientes doce meses cuando el individuo tiene x años. En la tradición actuarial este indicador es denotado por q_x para cada año de edad desde los 0 hasta una edad máxima conocido como infinito actuarial (110 años en el caso del SNP). La actual tabla de mortalidad, denominada SP2005, se encuentra vigente desde el año 2006, la misma que fue calculada usando información de afiliados al sistema nacional como el privado con datos de los años 1993 al 2005. A partir de esta instrumento se tiene que la expectativa de vida es de 90 y 83

años para mujeres y hombres, respectivamente (esperanza medida a los 65 años).

El diferencial en probabilidades de fallecimiento entre hombres y mujeres exhibida por la tabla vigente es amplia. Así, en promedio, si se considera el rango de edades de 65 a más, la probabilidad de fallecimiento de los hombres duplica a la de las mujeres; incluso, en edades tempranas como los 20 años, esta diferencia llega a ser tres veces más. Si bien la estadística internacional observada tiende a favorecer a las mujeres en este indicador, esta diferencia no es tan amplia como la presentada en la SP2005 (ver figura N° 1).

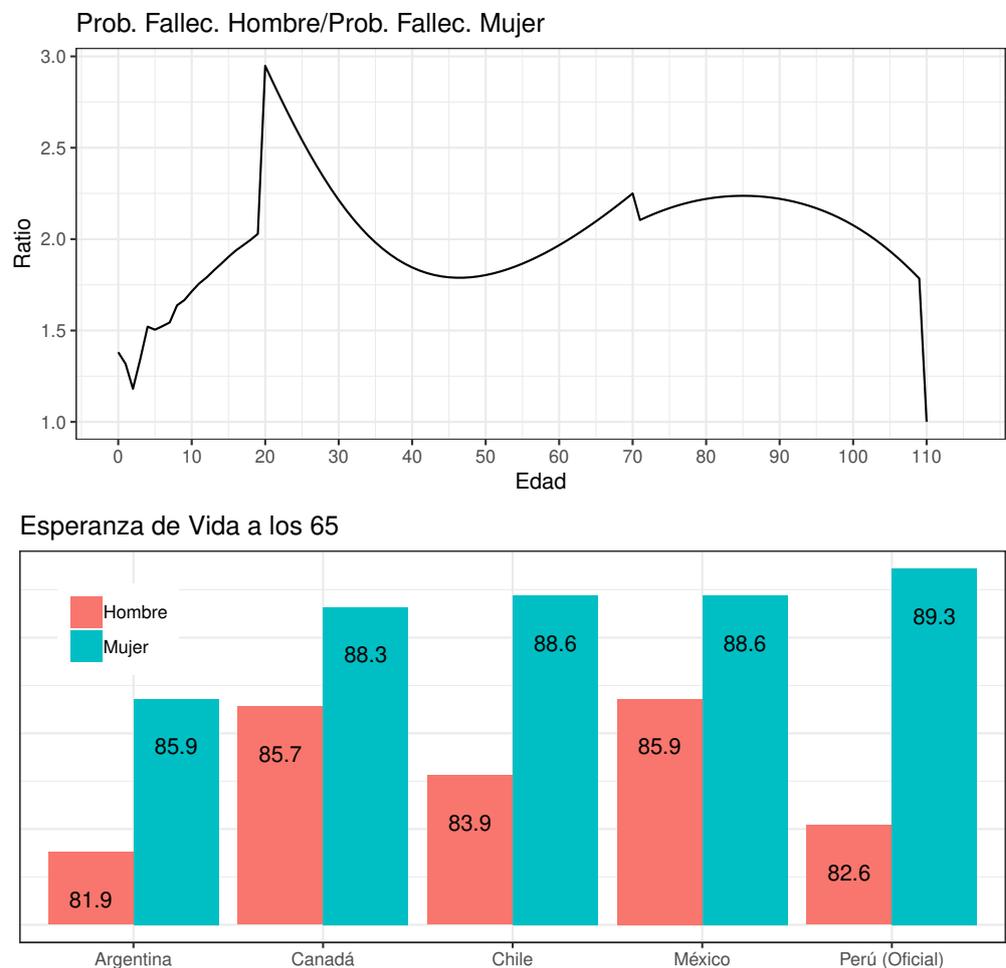


Figura 1. Ratio entre probabilidades y esperanza de vida a los 65 años en países seleccionados.

Verificar si esta brecha se condice con los datos recientemente observados de

los afiliados del SNP es la hipótesis a contrastar. Así, si las probabilidades de sobrevivencia de las mujeres están sobreestimadas y la de los hombres subestimadas, como es la hipótesis que motiva este estudio, entonces la reserva matemática de una pensión vitalicia estaría sobre y subestimada en el caso de las mujeres y hombres, respectivamente. El efecto neto de estos dos efectos contrapuestos depende del tamaño de los sesgos, la composición demográfica de los afiliados y el tamaño de las pensiones.

En resumen, una actualización permitirá contar con estimaciones más cercanas a la realidad de lo que el Estado debería estar provisionando para enfrentar al pago de las futuras pensiones; además, permitirá conocer la estructura de una población en un momento dado, hacer proyecciones y estimar la esperanza de vida de los pensionistas y afiliados de la ONP a cualquier edad.

Es en este contexto que se aborda como problemática el modelamiento de una nueva tabla de vida para el SNP, pues las probabilidades ahí reportadas inciden directamente sobre el valor que el Estado debe reservar para hacer frente a las obligaciones monetarias que tiene con sus pensionistas.

1.2 Problema

1.2.1 Problema principal

¿Será que un nuevo modelo para la construcción de la tabla de vida permitirá mejorar las estimaciones actuariales como la reserva matemática o la esperanza de vida?

1.2.2 Problemas secundarios

1. ¿Cómo estimar la nueva tabla de vida para el SNP?
2. ¿Cómo verificar si la estimación de las probabilidades de sobrevivencia se encuentran sesgadas o no?
3. ¿Cómo estimar el efecto de la nueva tabla sobre la esperanza de vida y las necesidades de financiamiento de la planilla de pensiones?

1.3 Justificación del estudio

Tres son los argumentos que justifican esta investigación.

1. Pasados ya 12 años desde la vigencia de la última tabla, es mandatorio la actualización de la misma, esto teniendo en cuenta además que las recomendaciones internacionales para la revisión sugieren que esta se debe hacer al menos cada 5 años.
2. La actualización permitiría obtener el valor real del pasivo que tiene el Estado para con los afiliados al SNP.
3. La tabla también permitiría hacer un diagnóstico real de la población asegurada, conocer la estructura de la misma, su esperanza de vida, y ser insumo para el diseño de políticas públicas.

Además, el desarrollo de la presente tesis permitirá alimentar la discusión académica en un contexto en el cual se suelen imponer metodologías mas no discutir las; más aún, en un escenario en el que tanto el SNP como el Sistema Privado de Pensiones (SPP), adolecen de tablas recientes.

El presente estudio es de interés para instituciones como el Ministerio de Economía y Finanzas (MEF), quien es la entidad que subsidia al SNP pagando la planilla de pensiones que no logra ser financiada por los aportes recibidos; también es de interés para la ONP, institución que administra el SNP y realiza estudios actuariales basados en la tabla; para la SBS, institución que reglamenta el uso de las tablas en el SPP, y que puede ver en este estudio un referente para la actualización de sus tablas; y la comunidad académica en general, que podrá encontrar en este estudio las metodologías disponibles con sus ventajas y desventajas.

1.4 Objetivos del estudio

1.4.1 Objetivo general

Desarrollar un modelo estadístico para las tablas de mortalidad que son de uso del Sistema Nacional de Pensiones tal que sea comparable con otros de uso previsional

en el país.

1.4.2 Objetivos específicos

1. Construcción de la tabla de vida a partir de la población asegurada en el SNP.
2. Discriminar la tabla de mortalidad considerando sexo y comprobar, de acuerdo a la tabla vigente, si las probabilidades de sobrevivencia son subestimadas en el caso de los hombres y sobreestimadas en el caso de las mujeres.
3. Estimar el efecto sobre las reservas matemáticas necesarias para honrar las pensiones de los afiliados como consecuencia de la actualización de la tabla.

2 Marco Teórico

2.1 Antecedentes

La tabla de mortalidad vigente para los afiliados al SNP es conocida como SP2005. Esta rige desde el año 2006 y para su construcción se empleó información que va del año 1993 al 2005. En Melinsky (2006) se encuentra detallada la metodología vigente. El objetivo de este estudio fue proponer una tabla de vida relevante solo para toda la población de afiliados al régimen público, empleando para ello registros administrativos tanto del Sistema Nacional como el Sistema Privado de Pensiones, y donde además la información ha sido validada con datos provenientes del Registro Nacional de Identificación y Estado Civil (RENIEC).

En Melinsky (2006) se estiman tres tablas de mortalidad, una siguiendo la Base de Datos de Afiliados a la ONP (Asegurados y Pensionados), otra considerando la población total (afiliados de ambos sistemas), siendo esta última la que reporta mayores expectativas de vida¹. Una tercera tabla es calculada ajustando los resultados para que esta tenga niveles similares de expectativa de vida a los 20 años reportada por la tabla chilena conocida como RV-2004. Esta última tabla es la que finalmente se recomienda para la valuación de las reservas del SNP.

La estrategia seguida para la obtención de la tabla se resume a continuación:

1. Estimación de tasas muestrales de mortalidad en el rango de edades que va de 20 a 95 años.
2. Suavizamiento o graduación empleando la metodología de ajuste conocido como Whittaker-Henderson, donde el parámetro de suavizamiento es elegido de tal forma que se superen ciertas pruebas estadísticas.
3. Las probabilidades para menores de 20 se obtiene usando la tabla de mortalidad B-85 (tabla anterior usada en el sistema previsional peruano) ajustando

¹El autor también señala que los valores obtenidos difieren sustancialmente de los correspondientes a los Estudios de CELADE sobre el total de la Población de Perú. Se argumenta que ello se debe a que la población analizada responde a características de selección, en cuanto a aspectos distintivos de nivel socioeconómico y formalidad respecto del total de la Población.

los valores en función de la relación entre las probabilidades de muerte de ambas tablas.

4. Las probabilidades para mayores de 95 se obtiene estimando los parámetros de una función biométrica (primera ley de Makeham) estimados con información del rango de edades de 65 a 95 (tasas ajustadas previamente por Whittaker-Henderson). Con los parámetros estimados se extrapolan las probabilidades hasta los 110 años.
5. La tabla resultante se ajusta para que la esperanza de vida medida a los 20 años sea igual a la reportada por la tabla chilena RV2004. Para esto se consideran los tramos de 20 a 70 años y 71 a 110 años. El ajuste es hecho asumiendo la función biométrica de Makeham, primera ley. Los parámetros A y B de la fuerza de mortalidad $\mu(x) = A + BC^x$ son modificados proporcionalmente para alcanzar tal fin.

No existe una actualización oficial de la tabla ni esfuerzos en el ambiente académico por proponer una nueva tabla para el SNP. En el caso del SPP, las tablas empleadas a la fecha son adaptaciones a las tablas chilenas, no estando disponible la metodología.

2.2 Bases teóricas

2.2.1 Modelos de sobrevivencia

Como se señala en London (1988, p. 3) “Un modelo de sobrevivencia es una distribución de probabilidad para una clase especial de variable aleatoria”. Así, sea la variable aleatoria T definido como la edad del fallecimiento y x la edad en la que se evalúa la probabilidad de la ocurrencia de la muerte más allá de x . La función de sobrevivencia queda definida entonces como:

$$S(x) = Pr(T > x)$$

Es decir, la función de sobrevivencia mide la probabilidad que el evento aleatorio ‘muerte’ o ‘falla’ ocurra después del momento x , esto es la función $S(x)$ mide

la probabilidad que el individuo sobreviva al menos hasta x . Por la naturaleza de T es claro que $T \geq 0$, que $S(0) = 1$ y que $\lim_{t \rightarrow \infty} S(t) = 0$.

Siempre que las probabilidades de supervivencia, $S(x)$, sean dados por una función matemática de x , decimos que $S(x)$ es una forma paramétrica. Siendo este nombre explicado por el hecho que los valores de $S(x)$ dependen de uno o más parámetros además de x . Por ejemplo, si T es una variable aleatoria distribuida exponencialmente, con $S(x) = e^{-\lambda x}$, entonces λ es un parámetro en el modelo de supervivencia (London, 1988, p. 6).

Como también es señalado por London (1988, p. 6) los modelos actuariales de supervivencia típicamente no han sido planteados de forma paramétrica pues la forma de $S(x)$, sugerida por los datos empíricos, es compleja para ser representada adecuadamente por una distribución de un solo parámetro como la exponencial, por ejemplo. El ajuste no es satisfactorio incluso si se emplean otras funciones como la Gompertz o Weibull (dos parámetros), o Makeham (tres parámetros)² . En su lugar, los actuarios han tendido a usar modelos tabulares mas que los paramétricos. Un modelo tabular de supervivencia es uno en el que los valores numéricos de $S(x)$ son presentados para ciertos valores seleccionados de x , más comúnmente valores enteros.

Desde el punto de vista conceptual, como es indicado por Kalbfleisch y Prentice (2002, pp. 6-7) una función de supervivencia (FDS) $S(x)$ puede ser representada con otras tres funciones (i) la función de distribución acumulada (FDA), (ii) la función de densidad (FDP) y (iii) el ratio de riesgo (FRR), las mismas que se listan a continuación.

1. Función de distribución acumulada (FDA). Esta expresa la probabilidad que el fallecimiento ocurra antes de cumplir los x años. Esto es:

$$F(x) = Pr(T \leq x) = 1 - S(x)$$

y que por tanto $F(0) = 0$ y $\lim_{x \rightarrow \infty} F(x) = 1$.

²Donde la función de ratio de riesgo viene definida por $\lambda(x) = A + BC^x$

2. Función de Densidad de Probabilidad (FDP), que es definido como la derivada de $F(x)$, esto es:

$$f(x) = \frac{\partial F(x)}{\partial x} = -\frac{\partial S(x)}{\partial x}$$

Consecuentemente, se tiene que:

$$F(x) = \int_0^x f(y)dy$$

y que:

$$S(x) = \int_x^\infty f(y)dy$$

3. Función de ratio de riesgo (FRR). Para su construcción se define el intervalo en el que se encuentra T por $[x, x + \Delta]$. Teniendo en cuenta que $T \geq x$, se tiene que la probabilidad de ocurrencia del fallecimiento vendría dado por:

$$Pr(x \leq T \leq x + \Delta | T \geq x) = \frac{Pr([x \leq T \leq x + \Delta] \cap [T \geq x])}{Pr(T \geq x)}$$

Es decir

$$\frac{Pr(x \leq T \leq x + \Delta)}{Pr(T \geq x)} = \frac{S(x) - S(x + \Delta)}{S(x)}$$

Se define la probabilidad instantánea de fallecimiento, o la probabilidad de que ocurra tal evento en un intervalo infinitesimal de Δ ($\Delta \rightarrow 0$). Matemáticamente puede ser representado por:

$$\lim_{\Delta \rightarrow 0} \frac{1}{\Delta} Pr(x \leq T \leq x + \Delta | T \geq x) = \lim_{\Delta \rightarrow 0} \frac{F(x + \Delta) - F(x)}{\Delta \cdot S(x)} = \frac{f(x)}{S(x)}$$

A esta probabilidad instantánea se le conoce como función de ratio de riesgo (FRR) en la literatura de sobrevivencia ($\lambda(x)$) y fuerza de mortalidad (FOM

por sus siglas en inglés) en la literatura actuarial ($\mu(x)$). Así se tiene que

$$\lambda(x) = \mu(x) = \frac{f(x)}{S(x)}$$

Desde que $f(x) = -\partial S(x)/\partial(x)$ se tiene que:

$$\lambda(x) = -\frac{\frac{\partial S(x)}{\partial x}}{S(x)} = -\frac{\partial \ln S(x)}{\partial x}$$

Integrando y despejando $S(x)$

$$S(x) = \exp - \left(\int_0^x \lambda(y) dy \right)$$

2.2.2 Tabla de vida

Una tabla de vida representa como un colectivo va decreciendo debido a la defunción de sus integrantes. De esta manera una tabla modela como un conjunto de personas son seguidas desde su nacimiento hasta que se extinga la totalidad de dicho colectivo. Los datos para la estimación pueden ser recolectados de dos maneras: transversales o longitudinales.

En el caso del primer método de recolección los datos son todos aquellos con los que se cuenta en un momento en el tiempo. De otro lado, en el caso de los datos longitudinales, un grupo es seguido hasta el fallecimiento; diseños de este tipo son típicos de ensayos clínicos o, en general, en diseños que no requieran seguir por mucho tiempo a la población estudiada (Bueno et al. 2008, p. 110).

Las tablas de vida fueron desarrollados por actuarios independientemente del desarrollo de la teoría estadística de los modelos de sobrevivencia y su estimación (un siglo antes). Por esta razón, la notación tradicional de las tablas de vida y terminología no tienden a revelar la naturaleza estocástica del modelo como claramente si lo hace el modelo conceptual de sobrevivencia.

Desde el punto de vista actuarial una tabla de vida se circunscribe a lo que se conoce como un modelo tabular, un instrumento que consigna las probabilidades de sobrevivencia para cada edad. En lo que sigue se desarrolla el modelo tabular, se presenta también las funciones derivadas a partir de asumir continuidad y

diferenciabilidad en la sobrevivencia y los principales supuestos que se suelen asumir para garantizar continuidad en las tablas, vale decir los distintos tipos de interpolación entre los valores de sobrevivencia discretos conocidos. Esta sección está basado en London (1988).

A. Modelo tabular

Como se señala en London (1988, p. 40) el modelo tabular difiere de aquel en el que se muestran las probabilidades de sobrevivencia. En lugar de presentar los valores reportados por $S(x)$, estos valores son multiplicados por números tales como 100,000, y de este modo presentar los valores de $S(x)$ como números enteros. Desde que estos enteros no son probabilidades, la columna $S(x)$ es renombrada por l_x , donde l_x representa al número de sobrevivientes. Desde que $S(0) = 1$, entonces l_0 es la misma constante que multiplicada a $S(x)$ y que permite obtener l_x . Esta constante es llamado la base o *radix* de la tabla. Formalmente:

$$l_x = l_0 \cdot S(x)$$

La tabla modular o también conocida como tabla de vida también asume un límite superior práctico para x , tal que los valores de $l(x)$ mas allá de dicho límite deberían ser iguales a cero. Este límite es conocido por los actuarios como infinito actuarial y es representado por la letra griega ω . Esto es, $l(\omega) > 0$, pero $l(\omega + 1) = 0$.

La principal ventaja de la forma tradicional de la tabla de vida es la simplicidad en su interpretación. Por ejemplo, de un hipotético grupo de recién nacidos conformados por $l_0 = 100,000$ personas, cada valor de l_x representa los sobrevivientes del grupo a la edad x , con lo cual la tabla o modelo tabular tiene una conveniente interpretación determinística. Por supuesto, desde que $l_x = l_0 S(x)$, y $S(x)$ es una probabilidad, entonces $l(x)$ es la cantidad esperada de sobrevivientes a la edad x a partir de un colectivo original de l_0 nuevos nacimientos (London 1988, p. 41).

Como es estándar en las ciencias actuariales, a partir del conocimiento de l_x se pueden definir los principales elementos que contiene una tabla. Así por ejemplo

como se señala en Dickson, Hardy y Waters (2009, pp. 26-28, 42-43), la cantidad de personas que fallecen a la edad de x representada por d_x se define como:

$$d_x = l_x - l_{x+1}$$

y por tanto, las probabilidades de fallecer (q_x) o sobrevivir en los siguientes 12 meses después de cumplir x son respectivamente:

$$q_x = \frac{d_x}{l_x}$$

$$p_x = \frac{l_{x+1}}{l_x}$$

expresiones que pueden ser generalizadas para considerar que el evento aleatorio de fallecimiento sea medido en intervalos distintos a los siguientes 12 meses. Así por ejemplo, si se define ${}_n d_x$ como la cantidad de fallecidos que ocurre entre los x y los $x + n$ años, entonces la probabilidad de fallecimiento en los siguientes n años después de cumplir los x se denota como:

$${}_n q_x = \frac{{}_n d_x}{l_x}$$

B. Funciones derivadas

Aunque una tabla de vida presenta valores de l_x para ciertos valores enteros de x , es conveniente adoptar el supuesto de que esta es una función continua y diferenciable, donde en la tabla de vida solo se muestran algunos valores de esta. La conveniencia se justifica en que hay otras funciones de mortalidad que pueden ser derivadas de l_x si esta es continua y diferenciable.

Una aplicación de este supuesto es la obtención de la Fuerza de Mortalidad (Un desarrollo matemático amplio puede encontrarse en Dickson et al. (2009, pp. 21-26)). La derivada de l_x puede ser interpretada como la tasa de cambio instantáneo absoluto anual de l_x . Desde que l_x representa la cantidad de sobrevivientes a la edad x , entonces la derivada representa la cantidad de fallecidos instantáneos en el punto en el que se defina.

Esta derivada es negativa pues l_x es una función decreciente, por lo que para obtener la magnitud absoluta de esta tasa se usa la negativa de esta derivada. Finalmente, desde que la magnitud de la derivada depende del tamaño de l_x , para obtener la tasa instantánea relativa se divide el negativo de la derivada de l_x por l_x . Así se tiene:

$$\mu_x = \frac{-\frac{dl_x}{dx}}{l_x}$$

Desde que $l_x = l_0 \cdot S(x)$, esta puede ser presentada como el negativo de la derivada del logaritmo de la función de sobrevivencia.

$$\lambda_x = \frac{-\frac{dS(x)}{dx}}{S(x)} = \frac{f(x)}{S(x)} = -\frac{dLnS(x)}{dx}$$

De esta manera la función de sobrevivencia puede ser expresada en términos de λ :

$$S(t) = \exp\left[-\int_0^t \lambda(y)dy\right]$$

Con lo que finalmente se tiene que l_x también puede ser presentado en términos de la fuerza de mortalidad:

$$l_x = l_0 \cdot S(x) = l_0 \cdot \exp\left[-\int_0^x \mu(y)dy\right]$$

En ese sentido, como hace notar London (1988, p. 44), $S(x)$ puede ser leído como un factor decremental que reduce la cohorte inicial de tamaño l_0 al tamaño l_x . De otro lado la fuerza de mortalidad puede ser generalizado para un valor $x+s$ de la tabla como:

$$\mu_{x+s} = \frac{-dl_{x+s}/ds}{l_{x+s}}$$

Un detalle a tener en cuenta es que se puede demostrar que la fuerza de mortalidad es el valor límite de la probabilidad de muerte sobre un intervalo infinitesimal dividido por la amplitud de dicho intervalo. Esto es:

$$\mu_x = \lim_{\Delta \rightarrow 0} \frac{\Delta q_x}{\Delta}$$

La siguiente función a ser desarrollada a partir de l_x es la FDP de la variable aleatoria T , la edad de fallecimiento. Teniendo en cuenta que $f(x) = \lambda(x) \cdot S(x)$. En el contexto de una tabla de vida, se tiene que $\lambda(x) = \mu_x$ y $S(x) = l_x/l_0$ (London, 1988, p. 44). Entonces se tiene que:

$$f(x) = \mu_x \cdot (l_x/l_0) = \mu_x \cdot {}_x p_0$$

Como $dl_x/dx = -l_x\mu_x$, luego dividiendo ambos lados por l_0 da:

$$\frac{d_x p_0}{dx} = -\mu_x \cdot {}_x p_0$$

Con la FDP, se puede calcular $E(x)$, que es denotado por \dot{e}_0 :

$$\dot{e}_0 = E[x] = \int_0^{\infty} x \cdot f(x) dx = \int_0^{\infty} x \cdot {}_x p_0 \mu_x$$

Integrando por partes se tiene la siguiente expresión alternativa

$$\dot{e}_0 = E[x] = \int_0^{\infty} {}_x p_0 dx = \frac{1}{l_0} \cdot \int_0^{\infty} l_x dx$$

Un desarrollo matemático amplio de la expectativa de vida que considera además el caso discreto puede ser encontrado en Dickson et al. (2009, pp. 29-34)).

Hasta ahora tenemos solo consideradas medidas de probabilidad de la edad $x = 0$, denotando tales probabilidades por $S(x)$ o $F(x)$. Específicamente, tales probabilidades fueron incondicionales, desde que solo se sabe que la persona estuvo viva en $x = 0$. Ahora se considerará el caso de una persona que estuvo viva en $x > 0$, y buscamos probabilidades (y densidades) de sobrevivencia (o fallecimiento) medido a partir de la edad x . Como se establece en Gil et al. (1999, p. 4), siguiendo la notación actuarial internacional y asumiendo conocida la distribución de la variable aleatoria vida residual, se definen a continuación las probabilidades condicionales de sobrevivencia y fallecimiento.

La probabilidad de un sobreviviente a la edad x y que se mantiene viva n años después (i.e. en la edad $x + n$), representado por $\Pr(\text{Sobrevivir a } x + n, \text{ dado que sobrevivió a la edad } x)$, puede ser deducida si multiplicamos esta probabilidad

condicional por $S(x)$, para obtener la probabilidad no condicional de sobrevivir a la edad $x + n$, que es $S(x + n)$, esto es $S(x) \cdot {}_n p_x = S(x + n)$. Por lo tanto la probabilidad deseada, denotada por los actuarios como ${}_n p_x$, es dado por:

$${}_n p_x = \frac{S(x + n)}{S(x)}$$

La probabilidad condicional de fallecimiento antes de que se cumpla la edad $x + n$, dado que estuvo vivo en la edad x , es dado entonces por:

$${}_n q_x = 1 - {}_n p_x = \frac{S(x) - S(x + n)}{S(x)}$$

Cabe resaltar que cuando hablamos de probabilidades o densidades condicionales de sobrevivencia a la edad x , estamos lidiando con la distribución de un subconjunto del espacio muestral de la variable aleatoria T , aquella cuyos valores de T caen por encima de x . Esta distribución es llamada la distribución de T truncada por debajo de x . Como se señala en Gil et al. (1999, p. 4) la probabilidad condicional de sobrevivencia ${}_n p_x$ puede ahora ser presentado formalmente como:

$${}_n p_x = Pr(T > x + n | T > x) = S(x + n | T > x) \quad (1)$$

En palabras, esta pregunta por la probabilidad que la edad de fallecimiento exceda $x + n$, dado que esta excede a x . Es fácil ver que este es el mismo concepto de "probabilidad de sobrevivir a $x + n$, dado que sobrevivió hasta x ". Así, se tiene que:

$$S(x + n | T > x) = \frac{S(x + n)}{S(x)}$$

Similarmente,

$${}_n q_x = Pr(T \leq x + n | T > x) = Pr(x < T \leq x + n | T > x) = F(x + n | T > x)$$

Así, se tiene que:

$$F(x+n|T > x) = \frac{S(x) - S(x+n)}{S(x)} = \frac{F(x+n) - F(x)}{1 - F(x)}$$

desde que $S(x) = 1 - F(x)$ y además $P(A|B) \cdot P(B) = P(A \cap B)$.

C. Interpolación

De otro lado, la distribución de T puede obtenerse a partir de una tabla de vida mediante interpolación, para lo cual es necesario asumir alguna hipótesis sobre el comportamiento de los sobrevivientes $l(x)$ para edades fraccionarias, esto es $l(x+s)$ para $0 < s < 1$. Las interpolaciones usuales son del tipo lineal, exponencial e hiperbólico (London, 1988, pp. 58-65), o equivalentemente, como se indica en Debon (2003, pp. 7-9): linealidad de ${}_s q_x$, μ_{x+s} constante o linealidad de ${}_{1-s} q_{x+s}$. El desarrollo siguiente sigue al primer autor.

1. Interpolación lineal para l_{x+s} (linealidad de ${}_s q_x$)

Si l_{x+s} es una función lineal entre x y $x+1$, esto es, de la forma $a + bS$. Para dar continuidad a l_{x+s} , $l_x = a$ y $l_{x+1} = a + b$, como se señala en London (1988) se tiene que:

$$l_{x+s} = l_x - s \cdot d_x$$

o lo que es lo mismo:

$$l_{x+s} = s \cdot l_{x+1} + (1-s) \cdot l_x$$

De esta manera los valores de l_{x+s} pueden ser obtenidos a partir de la interpolación de los valores l_x y l_{x+1} . Los probabilidades de sobrevivencia y fallecimiento entonces quedan definidas como:

$${}_s p_x = \frac{l_{x+s}}{l_x} = 1 - s \cdot q_x$$

y

$${}_s q_x = s \cdot q_x$$

también se tiene que:

$${}_{1-s} p_{x+s} = \frac{l_{x+1}}{l_{x+s}} = \frac{p_x}{1 - s \cdot q_x}$$

y

$${}_{1-s} q_{x+s} = \frac{(1-s) \cdot q_x}{1 - s \cdot q_x}$$

De otro lado la fuerza de mortalidad queda definido como:

$$\mu_{x+s} = \frac{-\frac{\partial l_{x+s}}{\partial s}}{l_{x+s}} = \frac{q_x}{1 - s \cdot q_x}$$

la cual es creciente en s . La naturaleza como l_{x+s} decrece uniformemente (esto es, las personas fallecen uniformemente) sobre el intervalo hacen que este supuesto de distribución de fallecidos sea conocido como UDD por sus siglas en inglés (*Uniform Distribution of Deaths Assumption*).

2. Interpolación exponencial (μ_{x+s} constante)

Si l_{x+s} es una función exponencial entre x y $x + 1$, luego este es de la forma $l_{x+s} = a \cdot b^s$. Para asegurar continuidad se tiene que $l_x = a$ y $l_{x+1} = ab$, por tanto se tiene que:

$$l_{x+s} = (l_{x+1})^s \cdot (l_x)^{1-s}$$

es decir, la interpolación se consigue como un promedio geométrico de los valores l_x y l_{x+1} . Las probabilidades se definen entonces como:

$${}_s p_x = \frac{l_{x+s}}{l_x} = (p_x)^s$$

y

$${}_s q_x = 1 - (1 - q_x)^s$$

también se tiene que:

$${}_{1-s} p_{x+s} = \frac{l_{x+1}}{l_{x+s}} = (p_x)^{1-s}$$

y

$${}_{1-s} q_{x+s} = 1 - (1 - q_x)^{1-s}$$

De otro lado la fuerza de mortalidad queda definido como:

$$\mu_{x+s} = \frac{-\frac{\partial l_{x+s}}{\partial s}}{l_{x+s}} = \frac{-l_x (p_x)^s \cdot \ln p_x}{l_x (p_x)^s} = -\ln p_x$$

esto es, una constante para $0 < s < 1$, así, si l_{x+s} es exponencial entonces la fuerza de mortalidad μ_{x+s} , es constante sobre el intervalo $[x, x + 1 > y$, como se indica en London (1988), puede ser representado por μ_x con lo cual se tiene que:

$$\ln p_x = -\mu_x$$

$$p_x = e^{-\mu_x}$$

y

$$q_x = 1 - e^{-\mu_x}$$

de esta forma se tiene que:

$${}_s p_x = (p_x)^s = e^{-\mu_x s}$$

3. Interpolación hiperbólica (linealidad de ${}_{1-s} q_{x+s}$)

Si l_{x+s} es una función hiperbólica entre x y $x + 1$, entonces este es de la forma $l_{x+s} = (a + bs)^{-1}$. Para asegurar continuidad, en $s = 0$, se tiene que $l_x = 1/a$ y en $s = 1$, se tiene que $l_{x+1} = 1/(a + b)$. Reemplazando, como se indica en London (1988), se tiene que:

$$\frac{1}{l_{x+s}} = s \cdot \frac{1}{l_{x+1}} + (1 - s) \cdot \frac{1}{l_x}$$

Expresión con la cual se pueden obtener los demás indicadores, ${}_s p_x$ por ejemplo:

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

Luego,

$${}_s q_x = 1 - {}_s p_x = \frac{s \cdot q_x}{1 - (1 - s)q_x}$$

También, se tiene el siguiente conveniente resultado:

$${}_{1-s} q_{x+s} = (1 - s)q_x$$

Luego, usando la expresión para ${}_s p_x$ se tiene que:

$$\mu_{x+s} = \frac{-d_s p_x / ds}{{}_s p_x} = \frac{q_x}{1 - (1 - s)q_x}$$

Expresión que es conocida como supuesto de Balducci y que pese a su popularidad tiene el problema de presentar μ_{x+s} como una función decreciente de s cuando lo lógico es que sea creciente como en el supuesto de interpolación lineal.

2.2.3 Modelo aleatorio de censura y fallecimiento

El modelo presentado a continuación se encuentra desarrollado en Broffitt (1984, pp. 79-82). El modelo es útil para comprender formalmente el problema de la estimación de la probabilidad de fallecimiento. En el modelo se tienen dos variables aleatorias independientes Z e Y , de la censura y fallecimiento respectivamente, donde:

$$q_x = P(x + Y \leq 1 + x|x) = P(Y \leq 1|x)$$

es la probabilidad de fallecer antes del cumplimiento de los $t + x$ años; del mismo modo la probabilidad de censura se define como:

$$r_x = P(x + Z \leq t + x|x) = P(Z \leq 1|x)$$

siendo el escenario ideal para la estimación aquel en el que se siguen a un conjunto N_x de personas hasta el cumplimiento de los $x + 1$ años, es decir, se tendría un experimento binomial en el que el estimador de la probabilidad de fallecimiento sería

$$\frac{d_x}{N_x}$$

,

con d_x como la cantidad de fallecidos en dicho rango. La limitante en lo anterior es que típicamente no se tiene un experimento, mas bien se cuenta con registros administrativos que se caracterizan por tener información censurada, esto es, no se tiene historias completas de las personas, el fin de la observación es determinado por el período de observación o por la salidad de la persona de la muestra por una razón distinta al fallecimiento.

Se tiene entonces, como se reconoce en Broffitt (1984), que el estimador d_x/N estaría subestimando el verdadero valor, pues hay fallecimientos no observados que no se encuentran contabilizados en d_x . Un estimador popular entre los actuarios ofrece una corrección para la censura, este consiste en descontar del total de observados la parte que por motivo de la censura no es observada. Así, en lugar de asumir como el estimador binomial que los censurados se encuentran presentes durante todo el intervalo a analizar, se asume que estos se encuentran expuestos al riesgo de fallecimiento en 0.5 años, por tanto la corrección al estimador propuesto por este estimador sería la siguiente corrección al estimador actuarial donde C_x la cantidad de observaciones censuradas en el intervalo $[x, x + 1 >$.

$$\frac{d_x}{N_x - C_x/2}$$

,

El modelo parte del hecho que tanto el fallecimiento como la censura son variables aleatorias independientes. Se observa lo primero que ocurra. En ese sentido, si T es el momento en el que se deja de observar a una persona ya sea por fallecimiento o censura, se tiene entonces que:

$$T = \text{mín}(Y, Z, 1)$$

Si Y y Z se distribuyen con funciones de distribución acumulada F y H , respectivamente y se define Q_x como la probabilidad observada de fallecimiento antes de cumplir $x + t$ años, entonces esta se tiene que:

$${}_tQ_x = P(Y < Z, Y \leq t|x)$$

y

$${}_tq_x = P(Y < t|x) = F(t)$$

También, como se detalla en Broffitt (1984), teniendo en cuenta que la probabilidad de fallecimiento antes del cumplimiento de los $x + t$ años se puede descomponer como la suma de la probabilidad observada y no observada:

$$P(Y < t|x) = P(Y < Z, Y < t|x) + P(Y > Z, Y < t|x)$$

donde el segundo sumando, dada las funciones de distribución puede ser expresado como:

$$P(Y > Z, Y < t|x) = \int_0^t H(\mu) \cdot f(\mu) d\mu = \int_0^t H(\mu) \cdot dF(\mu)$$

entonces se tiene que:

$${}_tq_x = {}_tQ_x + \int_0^t H(\mu) \cdot dF(\mu)$$

teniendo en cuenta que

$$\int_0^t H(\mu) \cdot dF(\mu) = H(t) \cdot F(t) - \int_0^t F(\mu) \cdot dH(\mu) = {}_t r_x \cdot {}_t q_x - \int_0^t F(\mu) \cdot dH(\mu)$$

reemplazando se tiene que:

$${}_t Q_x = {}_t q_x [1 - {}_t r_x (1 - c)] \quad (2)$$

donce $c = \int_0^t F(\mu) \cdot dH(\mu) / (q \cdot r) = P(Y < Z) / (P(Y < t) \cdot P(Z < t)) = P(Y < Z | P(Y < t, Z < t))$ es la probabilidad observada de fallecimiento dado que el fallecimiento y censura ocurren antes de $x + t$ años.

$${}_t R_x = {}_t r_x [1 - {}_t q_x \cdot c] \quad (3)$$

${}_t Q_x$ y ${}_t R_x$ pueden ser estimados consistentemente conociendo al total de personas, la cantidad de fallecidos y el tamaño de la censura (N_x , D_x y C_x , respectivamente), no obstante el sistema de ecuaciones dado por (2) y (3) se encuentra subestimado, pues los parámetros a estimar son 3 (q , r y c). En ese sentido, se dice que bajo las condiciones antes descritas no es posible construir un estimador consistente de q_x , a menos que se imponga supuestos adicionales sobre el modelo. En particular, se puede demostrar que el estimador actuarial antes señalado es consistente siempre y cuando $c = \frac{1}{2-q}$.

2.3 Marco conceptual

La estimación de la tabla de mortalidad para el SNP y sus implicancias contables o monetarias requiere la conceptualización previa de los siguientes dos elementos.

1. Los elementos que componen una tabla de vida, y
2. La definición de las reservas matemáticas

2.3.1 Conceptualización de una tabla de vida

Siguiendo a Debon (2003, p. 7) se define T como una variable aleatoria que representa la vida residual. Dicho de otra manera es la cantidad de años que

vivirán las personas. Su distribución de probabilidad viene dado por:

$$P(K = k) = P(k \leq T \leq k + 1) = {}_k p_x \cdot q_{x+k}, \quad k = 0, 1..$$

Donde la esperanza de T es el valor esperado de vida abreviada representada por:

$$e_x = \sum_{k=1}^{\omega-x} k P(T = k) = \sum_{k=1}^{\omega-x} k ({}_k p_x \cdot q_{x+k})$$

o lo que es lo mismo

$$e_x = \sum_{k=1}^{\omega-x} P(T \geq k) = \sum_{k=1}^{\omega-x} ({}_k p_x)$$

Sea S la fracción del año de su muerte durante la cual la persona vive, por tanto $T' = T + S$. Si se asume que S es una variable aleatoria con una distribución continua que va de cero a uno, y se asume además que esta es simétrica, entonces su valor esperado es $1/2$, con lo cual se define la esperanza de vida exacta como:

$$\dot{e}_x = E[T + S] = e_x + \frac{1}{2}$$

La distribución de probabilidad de la vida residual se estima mediante un modelo matemático cuyos valores se reportan en lo que se denomina una tabla de vida. Específicamente, en una tabla de vida se muestra la cantidad de sobrevivientes en cada año de vida, con esta información base se pueden construir las probabilidades de fallecimiento q_x o las de sobrevivencia p_x definiendo completamente la distribución de T , donde l_x es la cantidad de sobrevivientes a la edad x y d_x son los fallecidos a dicha edad.

Una definición formal puede ser encontrada en Dickson et al. (2009, pp. 41-43), de acuerdo a esta, dado un modelo de sobrevivencia, con probabilidades ${}_t p_x$, se puede construir la tabla de vida a partir de una edad inicial x_0 hasta una edad máxima de ω . Se define una función l_x para $x_0 \leq x \leq \omega$ como sigue. Sea un l_{x_0} un numero positivo arbitrario (llamado también radix) y, para $0 \leq t \leq \omega - x_0$, $l_{x_0+t} = l_{x_0} \cdot {}_t p_{x_0}$, que se puede generalizar como:

$$l_{x+t} = l_x \cdot {}_t p_x$$

y con ello construir una tabla como la que se muestra a continuación.

Tabla 1

Tabla de vida hipotética

x	l_x	d_x	q_x	p_x
0	100,000	10,000	10 %	90 %
1	90,000			
·	·			
·	·			
110	2			
111	0			

2.3.2 Reserva matemática

El indicador por excelencia para medir la situación financiera de un sistema de pensiones como el SNP es el déficit actuarial o reserva neta. Este se constituye como el valor presente de las pensiones corrientes (reserva pensionaria), más el valor presente de las pensiones de los futuros jubilados (reserva bruta), menos el valor presente de los aportes de los asegurados.

En ONP (2018) se describen las fórmulas empleadas en cada uno de los componentes arriba citados. En todos los casos la lógica financiera estándar de valor presente es modificado para tomar en cuenta la probabilidad de sobrevivencia, esto es, se considera la probabilidad de que el flujo se realiza siempre y cuando el individuo esté vivo. Así, la reserva pensionaria del individuo j de edad x y cuando la pensión es vitalicia sería:

$$Reserva = 12 \cdot R \left[\sum_{t=1}^{\omega-x} \frac{l_{x+t}}{l_x} \cdot \left(\frac{1}{1+i} \right)^t + \frac{11}{24} \right]$$

Donde R es la pensión vitalicia mensual, ω es la edad máxima de vida, x es la edad de la persona en cuestión, i la tasa de interés o descuento y l_{x+t}/l_x es la probabilidad de que un individuo de edad x sobreviva t años más, y el factor $11/24$ es un factor de ajuste que considera el hecho que las flujos son mensuales y las demás factores anuales (ONP, 2018, p. 8).

La expresión antes señalada considera la reserva matemática de un titular, a este valor se le puede añadir la reserva necesaria para financiar la pensión de un derechohabiente como la esposa. En ese caso, la probabilidad relevante es la que se obtiene de multiplicar la probabilidad de fallecimiento del titular con la probabilidad de sobrevivencia de la mujer. El principio es el mismo en el caso del valor presente de los aportes (ONP, 2018, pp. 8-16).

En suma se definen las siguientes expresiones que permiten explicar el efecto contable de los cambios en la sobrevivencia o mortalidad:

1. **Reserva pensionaria (RP):** Es el valor presente del flujo de pensiones de los pensionistas vigentes.
2. **Reserva bruta (RB):** Es el valor presente de los afiliados actuales que en su momento serán pensionistas.
3. **Valor presente de aportes VPA:** Es el valor presente de los aportes futuros estimados de los afiliados actuales.
4. **Reserva actuarial neta (RAN):** Es el pasivo u obligación que tiene el estado en valor presente descontado de los aportes futuros también expresados en valor presente; vale decir: $RAN = RP + RB - VPA$.

En todos los casos, las expresiones anteriores toman en cuenta el valor del dinero en el tiempo al descontar los flujos futuros por una tasa de interés o descuento; tomando en cuenta además, el hecho que los pagos futuros se realizan siempre y cuando el individuo esté vivo, de ahí que se le asigne a cada edad una probabilidad de sobrevivencia hasta un máximo teórico de ω , edad a partir de la cual se asume que la probabilidad de sobrevivencia es cero.

Desde el punto de vista contable, la importancia de las reservas actuariales radica en que al no conocer la fecha de fallecimiento, tampoco se tiene certeza de cuánto dinero se tiene que reservar para pagar la pensión vitalicia de una persona. La implicancia de que las personas vivan más por ejemplo, puede tener consecuencias de política económica, pues financiar las pensiones podría por ejemplo

requerir ajustes en los parámetros que caracterizan al sistema, como la tasa de aporte (13 % en la actualidad), o el aumento de la cantidad de años mínimos que se exige para solicitar la pensión de jubilación (20 años en la actualidad).

2.4 Hipótesis y variables

2.4.1 Hipótesis general

Las probabilidades vigentes en la tabla de mortalidad oficial son estimadores insesgados de los verdaderos valores.

2.4.2 Hipótesis específicas

1. Las probabilidades de fallecimiento de las mujeres se encuentran subestimadas.
2. Las probabilidades de fallecimiento de los hombres se encuentran sobrestimadas.
3. El déficit actuarial neto se encuentra sobrestimado debido a que el efecto de la subestimación de las probabilidades de fallecimiento de la mujer es superior al efecto de la sobrestimación de las probabilidades de los hombres.

2.4.3 Identificación de variables

Existen tres tipos de variables en este estudio. Uno del tipo dependiente, que en este caso es la ocurrencia del evento aleatorio ‘fallecimiento’; otra del tipo dependiente como lo es la edad; y finalmente una del tipo cualitativo como el sexo.

En el caso de la edad, claramente es un determinante del fallecimiento, pues, salvo en edades tempranas, donde la probabilidad disminuye después del primer año de vida, la probabilidad de defunción aumenta con la edad. El sexo por otra parte, no es una variable que explícitamente sea modelado en la estimación de las tablas, lo que se suele hacer es estimar un modelo para los hombres y otro para las mujeres bajo la hipótesis que las expectativas de vida son distintas.

3 Metodología

3.1 Tipo y diseño de investigación

De acuerdo a London (1988, p. 8) los diseños pueden ser de corte transversal o longitudinal. En el caso del primer tipo se empieza con la definición del grupo de estudio, es decir un grupo identificable de personas cuya comportamiento de sobrevivencia sea estudiado (población de una ciudad o los clientes de un plan de seguros, por ejemplo). Sobre las unidades de estudio, London, señala que una vez que el período de observación es escogido, al inicio de este período hay muchas personas que ya son miembros del grupo de estudio y son observadas desde ese momento, mientras que otras personas podrían unirse al grupo durante el período de observación, y algunas otras podrían salir por motivos distintos al fallecimiento como migración (1998, p. 9).

De otro lado, en el caso de los estudios longitudinales, en lugar de seleccionar un intervalo de tiempo, y observar la ocurrencia del evento dentro del mismo, estos estudios seleccionan a un grupo de estudio y lo siguen en el tiempo. Estos diseños son propios de estudios clínicos donde en general el tiempo hasta la ocurrencia del evento es pequeño. Un caso particular de estos estudios tienen lo que se conoce como un ‘diseño completo de cohortes’, en este diseño, una cohorte, o un grupo inicial, de unidades de estudio son seleccionadas en el momento $t = 0$ y se mantienen en observación hasta que todos fallecen. Es claro que para desarrollar este tipo de estudio, el investigador debe tener la habilidad para controlar al grupo de estudio, en el sentido que ninguna de las unidades estudiadas salga del estudio sin un motivo distinto al del fallecimiento. Estos estudios son también llamados ‘estudios de datos controlados’ y por su naturaleza, no es posible que las muestras de estudio sean grandes (London 1988, p. 9).

Como también es señalado por London, en algunos casos el estudio longitudinal puede ser terminado antes que todas las unidades estudiadas mueran con el fin de evitar un período de estudio excesivamente largo. En tales casos se tendría un estudio con ‘datos incompletos’. Existen dos enfoques para terminar el estudio. Si el estudio es terminado en un fijo y predeterminado momento, entonces decimos

que los datos han sido truncados. De otro lado, si un estudio es continuado hasta un predeterminado número de muertes es observada, entonces decimos que los datos han sido censurados (1998, pp. 9-10).

Por sus características el estudio aquí desarrollado es de tipo no experimental, específicamente del tipo de corte transversal donde las variables de estudio son la tasa de mortalidad, la edad y el sexo, en la que se asume que la primera viene determinada por las restantes. La fecha de inicio y de fin de estudio son determinados exógenamente. Por su naturaleza, las unidades observadas son estudiadas en el rango en el que las variables de interés son observadas, se admiten ingresos después de la fecha de inicio y la fecha de fin de estudio representa un truncamiento en la información, es decir, no se siguen a las personas hasta el acontecimiento del fallecimiento, la fecha de fin de cada observación está determinada por el fallecimiento si este ocurre antes de la fecha de corte o la sobrevivencia hasta la fecha de fin.

3.2 Unidad de análisis

Un afiliado al Sistema Nacional de Pensiones durante los años 2013 hasta noviembre de 2017.

3.3 Población de Estudio

Afiliados al Sistema Nacional de Pensiones durante los años 2013 hasta noviembre de 2017.

3.4 Tamaño de Muestra

Se emplean los datos observados desde 2013 hasta noviembre 2017. Estos corresponden a afiliados y pensionistas que pertenecieron o pertenecen al SNP. Alrededor de 5 millones de registros como se muestra en la tabla 2.

3.5 Selección de Muestra

La muestra considera alrededor de 5 años, desde enero de 2013 a noviembre de 2017. Los datos proceden de los registros administrativos de la ONP correspon-

diente a los afiliados activos y pasivos (pensionistas) del Sistema Nacional de Pensiones (tanto sobrevivientes como fallecidos). De la ONP se conoce la fecha en la que se empieza a observar a las personas y si los afiliados pasaron a ser pensionistas. El sexo y fechas de nacimiento y fallecimiento provienen del RENIEC.

Respecto a los afiliados activos estos pueden haber realizado aportes antes de la fecha de inicio de observación, en cuyo caso la fecha de inicio consignada en la muestra se fija en enero de 2013. Si las personas realizaron su primer aporte después de enero de 2013, la fecha del primer aporte coincide con la fecha de inicio de observación. La fecha de salida es la censura determinada por el final del período de estudio o la fecha de fallecimiento si aplica. En caso el afiliado se jubile, la información se concilia para evitar duplicidades.

De otro lado, en el caso de los pensionistas no identificados como afiliados de acuerdo a lo señalado en el párrafo previo (estos corresponden a derechohabientes y pensionistas que siendo titulares son pensionistas desde antes de julio de 1999). La fecha de inicio para estos pensionistas se iguala al inicio del período de estudio si la primera pensión se paga antes que dicha fecha y se iguala a la fecha de la primera pensión en otro caso. La fecha de salida, sigue las mismas reglas de censura y fallecimiento señaladas en el caso de los afiliados.

El estudio considera solo registros válidos, esto es, no se incluyen a aquellas personas cuya fecha final de observación es previo al inicio del período de estudio. Se excluyen también a aquellas personas que tienen una fecha de inicio posterior al final del horizonte de estudio. Una persona aporta información en la medida que esta es observada en el período de estudio, así, se excluyen las observaciones cuya censura ocurrió antes de enero de 2013. Se excluye a los asegurados que se encuentran afiliados al Sistema Privado de Pensiones; también se excluyeron los registros de las personas con alguna invalidez.

La descripción de los datos finalmente empleados en el análisis se muestra en la tabla 2, estos corresponden a alrededor de cinco millones de personas. La muestra se caracteriza por que la mayoría son asegurados activos (87%), predominan más

los hombres (59%), y la cantidad de fallecidos es de 170 mil. En el anexo A.1 se muestran una parte de los datos utilizados.

Tabla 2

Descripción de los datos utilizados

	En %	Personas
SEXO		
Hombres	58.9 %	2,987,095
Mujeres	41.1 %	2,086,466
TIPO DE DATO		
Afiliado a pensionista	4.3 %	218,189
Solo afiliado	87.2 %	4,423,209
Solo pensionista	8.5 %	432,163
CONDICIÓN DE SOBREVIVENCIA		
Sobreviviente	96.6 %	4,903,231
Fallecido	3.4 %	170,330
TOTAL	100.0 %	5,073,561

3.6 Técnicas de recolección de datos

La información procede de la Oficina de Normalización Previsional. La investigación está hecha sobre base de la observación directa de los hechos.

3.7 Análisis

Estimar una tabla de mortalidad siguiendo a un grupo de personas en el tiempo no es factible, hacerlo significaría tener que esperar más de cien años para poder usar los datos. Convencionalmente una tabla de mortalidad se modela empleando información de la edad y el sexo, pues es un hecho estilizado que la edad es uno de los determinantes de la sobrevivencia y que las mujeres tienden a vivir más que los hombres.

Conductas habituales como ejercicios físicos, fumar, dietas alimenticias y cuidados de salud son también factores que influyen en el riesgo de fallecimiento. Aunque no es usual contar con esta información, esto no constituye un problema si es que la información corresponde a una población homogénea, tal como lo de los afiliados al sistema nacional, ya que estos que se caracterizan por tener menor remuneración respecto a sus pares privados.

Existe un problema de sesgo de selección si los resultados obtenidos a partir de una población se extrapolan a otra, como por ejemplo asumir que los patrones de sobrevivencia del SNP son los mismos que en el sistema privado. La evidencia indica que distintas poblaciones podrían tener también distintas probabilidades de fallecimiento. Por ejemplo, en Sánchez (2002), se da cuenta de diferencias entre las tablas de vida empleadas para toda la población respecto a otra tabla referida solo a la subpoblación de asegurados (p. 111).

Para estimar una una tabla de mortalidad Charpentier (2015, pp. 383-384) recomienda los siguientes pasos. En primer lugar estimar la tasa bruta o cruda de mortalidad; luego, graduar o suavizar dicha tasa; y finalmente validar el modelo. A este conjunto de pasos se le debe agregar el tratamiento a las estimaciones en las colas, pues es de esperar que no existan la suficiente cantidad de fallecimientos que permita tener un estimador razonable para las colas.

De acuerdo al orden en la estimación, calcular la tasa cruda es el primer paso. Esta puede seguir un enfoque actuarial o estadístico. En lo que respecta al enfoque estadístico se tiene el estimador de Kaplan y Meier (1958). De otro lado, de acuerdo al enfoque actuarial, también es posible tener estimadores que tienen como base el tiempo a la exposición al riesgo del fallecimiento.

En el caso de la graduación, como señalan Debon, Montes y Sala (2005, p. 270), existen dos enfoques. El paramétrico y el no paramétrico, en el primero se asume un comportamiento de la probabilidad de fallecimiento definido por una función que depende de parámetros que deben ser estimados con alguna técnica de optimización, de otro lado, los métodos no paramétricos no exigen el conocimiento de alguna función, su lógica más bien se relaciona con las diversas técnicas de suavizamiento que existen en la literatura (Debón, Montes y Sala, 2005, p. 270).

Siguiendo a Charpentier (2015, pp. 383-384) se propone como estrategia los siguientes pasos para modelar la tabla de mortalidad: (i) estimar la tasa bruta; (ii) suavizar o graduar estos resultados para tener así una representación sin irregularidades de la tasa; (iii) estimar las probabilidades en la cola inferior y

superior; y (iv) validación de los resultados, entendido esto último como la evidencia que de cuenta de la razonabilidad de los resultados. En lo que sigue se detalla la estrategia de análisis en cuanto a la determinación de la tasa bruta de mortalidad y la graduación.

3.7.1 Tasa bruta de mortalidad

Cualquiera sea el método, un concepto fundamental para la estimación de las probabilidades es el de exposición al riesgo (Sánchez, 2002, p. 112). Intuitivamente la probabilidad observada de fallecimiento toma en cuenta la cantidad de fallecidos a la edad x y cuántos de ellos estuvieron expuestos a la ocurrencia de dicho evento (muerte).

Como se define en London (1988, pp. 50-51), la exposición al riesgo se define como el tiempo en el que una persona estuvo afecto de sufrir el evento aleatorio. En ese sentido, para una determinada edad x una persona podría contribuir completamente a la exposición al riesgo o contribuir parcialmente, si es que en este último caso por ejemplo sale del estudio antes de cumplir los $x + 1$ por la afiliación al sistema privado, el fallecimiento o el período de corte del estudio.

La definición de la tasa bruta depende de los datos y la estrategia de estimación. En London (1988, p. 85) por ejemplo, el autor hace notar que si los datos no se encuentran censurados, escenario improbable en el caso de tablas de mortalidad, la tasa se obtiene directamente de los datos y es definida por:

$$\hat{q}_x = \frac{d_x}{N_x}$$

Donde d_x es la cantidad de decesos que ocurren en el intervalo $\langle x, x + 1 \rangle$ y N_x la cantidad de personas vivas a la edad x . Esto es, la exposición al riesgo de fallecer es todo el intervalo $\langle x, x + 1 \rangle$, incluso si se sabe el momento en el que alguna persona fallece al interior del intervalo.

Bajo esta lógica el fallecimiento se podría asumir como una variable aleatoria gobernada por una distribución binomial. Esto es, sigue una distribución $Bi(N_x, q_x)$, donde q_x es la probabilidad poblacional de fallecer y N_x los expuestos

al riesgo (London, 1988, p. 85). En este sentido, el estimador antes propuesto es insesgado y de mínima varianza (Debón y Sala, 2001, p. 9).

Como los datos típicamente se encuentran censurados debido a que no es posible hacer un seguimiento de las personas hasta su fallecimiento, o debido a que algunas personas salen de la muestra por razones distintas al fallecimiento, la definición de la tasa tiene que tomar en cuenta que la muestra a partir de la cual se va estimar los datos se encuentra censurada. Varias son las estrategias para estimar la mortalidad (ver London (1988), Agoston (2001) y Debon (2003), por ejemplo). Esta investigación emplea los estimadores que de acuerdo a Agoston (2001, p. 151) son los más comunmente empleados: Estimador de Kaplan-Meier, estimador por máxima verosimilitud y estimador Actuarial. También se incluye el estimador de momentos, los mismos que a continuación se detallan:

A. Estimador de Kaplan y Meier

También conocido como estimador producto límite, propuesto originalmente por Kaplan-Meier (1958), esta estrategia se basa en la siguiente observación, si una persona ha sobrevivido de x a $x + 1$ años, esto es equivalente a señalar que ha sobrevivido a los subintervalos que se definan al interior del intervalo $< x, x + 1]$. De esta manera el estimador de \hat{p}_x se define como la productoria de las probabilidades de sobrevivir en cada subintervalo (London, 1988, p. 166). Formalmente:

$$\hat{q}_x = 1 - \prod_i^m \hat{p}_i$$

Donde m es la cantidad de subintervalos en el rango $< x, x + 1]$, los mismos que se definen por la ocurrencia por lo menos un deceso y \hat{p}_i es la probabilidad de sobrevivencia en el subintervalo i . Con $\hat{p}_i = 1 - d_i/n_i$; donde d_i es la cantidad de fallecimientos observados y n_i es la cantidad de observaciones en el subintervalo i .

B. Estimador Máxima Verosimilitud

Como se señala en London (1988, p. 147-148), en este caso genéricamente se tiene que cada persona aparece a la edad de $x + r_i$ y sale a la edad de $x + s_i$,

donde $s_i > r_i$ y ambos sucesos ocurren en el intervalo $[x, x + 1]$. Si se define la variable binaria δ_i que toma el valor de 1 si la persona fallece en el intervalo de interés y cero en otro caso, entonces se tiene la siguiente tipología:

- Si $s_i = 1$ y $\delta_i = 0$, entonces la persona i es un sobreviviente.
- Si $s_i < 1$ y $\delta_i = 0$, entonces la persona i es una observación censurada.
- Si $s_i \leq 1$ y $\delta_i = 1$, entonces la persona i falleció en el intervalo.

De este modo, teniendo en cuenta que la función de sobrevivencia se denota por $S()$; la distribución acumulada es $F() = 1 - S()$; y la densidad se denota por $f()$; se tiene que el aporte a la verosimilitud de la persona i puede ser expresada como:

$$L_i = \frac{S(x + s_i)}{S(x + r_i)} \left[\frac{f(x + s_i)}{S(x + s_i)} \right]^{\delta_i}$$

Si se asume que la interpolación de la función de sobrevivencia es del tipo exponencial³, entonces se tiene que $S(x) = e^{-\mu x}$. Maximizar la función de verosimilitud con dicho supuesto permite estimar:

$$\hat{\mu}_x = \frac{d_x}{\sum_{i=1}^{n_x} (s_i - r_i)}$$

y obtener finalmente el estimador deseado:

$$\hat{q}_x = 1 - e^{-\hat{\mu}_x}$$

C. Estimador Método de Momentos

Este estimador se obtiene al asumir que el valor esperado de muertes es igual a la cantidad observada de muertes. Si por ejemplo la edad de entrada y salida es la misma para todos los individuos ($r_i = r$ y $s_i = s$) entonces el número esperado de fallecidos igualado al número observado se puede representar como:

³Es decir, $l_{x+s} = (l_{x+1})^s (l_x)^{1-s}$

$$n_x(s-rq_{x+r}) = d_x$$

Que, de acuerdo a London (1988, p. 108) puede ser generalizado por:

$$\sum_{i=1}^{n_x} (s-r_i q_{x+r_i}) = d_x$$

Asumiendo que $s_i-r_i q_{x+r_i} \approx (s_i - r_i)q_x$ el estimador queda definido como:

$$\hat{q}_x = \frac{d_x}{\sum_{i=1}^{n_x} (s_i - r_i)}$$

D. Estimador actuarial

Similar a lo que se realiza con el método de momentos bajo esta estrategia se iguala las muertes esperadas a las observadas. Así, asumiendo $r_i = r$, entonces todas las muertes esperadas se definen como:

$$A = n_x(1-rq_{x+r})$$

y que los fallecidos no observados, asumiendo $s_i = s$ y e_x como el total de sobrevivientes hasta $x + s$, se puede estimar como:

$$B = e_x(1-sq_{x+s})$$

Entonces se tiene que:

$$A = B + d_x$$

Generalizando, siguiendo al desarrollo planteado por Debón (2003, p. 15-16), donde I es el conjunto de todos los fallecimientos observados:

$$\sum_{i=1}^{n_x} (1-r_i q_{x+r_i}) = \sum_{i \notin I} (1-s_i q_{x+s_i}) + d_x$$

Si se asume conveniente el supuesto de interpolación hiperbólica (supuesto de Balducci), el mismo que trae como consecuencia que:

$${}_{1-\mu}q_{x+\mu} = (1 - \mu)q_x$$

Si se define $E_x = \sum_{i=1}^{n_x} (s_i - r_i)$ como la exposición al riesgo, se tiene que:

$$\hat{q}_x = \frac{d_x}{\sum_{i \notin I} (s_i - r_i) + \sum_{i \in I} (1 - r_i)} = \frac{d_x}{E_x + \sum_{i \in I} (1 - s_i)}$$

Donde dicho estimador es un caso particular del método de momentos, en el que se asume que la contribución en la exposición al riesgo de los fallecidos equivale a asumir que estos llegan a los $x + 1$ años de edad. Vale la pena señalar que estimador actuarial $d_x / (N_x - 0,5C_x)$ señalado en la sección previa es un caso particular del estimador aquí presentado, ello ocurre cuando $r_i = r = 0$ y $s_i = s = 0,5$.

3.7.2 Técnicas de graduación

Graduación es como se conoce en la literatura actuarial al suavizamiento de una estimación previa. Así, después de estimar la tasa bruta de mortalidad, las irregularidades que podrían presentar son suavizadas bajo la lógica que estas responden a fluctuaciones ajenas al comportamiento real de la variable en cuestión. Matemáticamente se puede recurrir a Tomas (2012) para una justificación formal. De acuerdo a este autor si las probabilidades \hat{q}_{x_1} y \hat{q}_{x_2} son independientes, es decir que una no influye en la otra, de esta manera las tasas brutas o crudas son los estimadores definitivos. No obstante esta premisa podría no estar cumpliéndose, pues es razonable que las tasas deberían estar relacionadas a sus vecinos. Así, la tasa bruta de mortalidad se representa a través de la siguiente relación funcional:

$$\hat{q}_{x_i} = g(x_i) + \mu_i \tag{4}$$

donde μ_i son errores aleatorios que representan variabilidad de fuentes no observadas o errores aleatorios en las observaciones; entonces un estimador \hat{q}_{x_1} cercano a \hat{q}_{x_2} contiene información de g en x . Sobre este asunto Benjamin y Pollard (1980) apelan a la máxima “*Natura non agit per saltum*” para señalar que “las fuerzas naturales operan gradualmente y sus efectos se ponen de manifiesto de

manera continua y no en saltos repentinos” (Citado en Tomas (2012), p. 11). Aplicado a las probabilidades de fallecimiento se espera entonces que las edades que cercanas a la edad x_i tienen probabilidades similares a sus vecinos y por tanto contienen información que puede ser usada para mejorar la estimación original de \hat{q}_{x_i} .

De este modo, con el fin reducir alguna fluctuación aleatoria se puede emplear la información de la vecindad alrededor del estimador con las técnicas de graduación de las tasas. Dependiendo del tipo de estrategia los métodos empleados para el suavizamiento o graduación pueden ser catalogados como paramétricos (se asume que g es predeterminada que sigue a alguna ley biométrica) o no paramétricos (se asume que g no adopta ninguna forma específica).

Típicamente las tasas de mortalidad son altas y decrecientes en el primer año de vida, a partir de dicho valor crecen y suelen tener una “joroba” alrededor de los 20 años. Debido a este comportamiento es difícil que alguna función biométrica se ajuste razonablemente a todas las edades. Entre los esfuerzos por tratar de encontrar alguna ley de mortalidad se pueden citar a Heligman y Pollard (1980).

De otro lado, el enfoque no paramétrico no asume un comportamiento predeterminado de la curva. Dentro de la tradición actuarial el enfoque más popular es la técnica conocida como Whittaker-Henderson. Otras estrategias son la de “Spline Cúbico” y “la Regresión Polinómica Local”. Como se mencionó, el método de Whittaker-Henderson es el de mayor popularidad entre los actuarios, no obstante, como es señalado por Li y Chan (2004, p. 135) su aplicación requiere la asignación subjetiva del valor de dos parámetros. A continuación se resumen las técnicas citadas.

A. Spline cúbico

Como señala Debon (2003, p. 40), mediante un spline cúbico se aproxima una función como la agregación de polinomios cúbicos definidos en subintervalos de la muestra. La continuidad se garantiza imponiendo la restricción de que las interpolaciones en los límites de cada subintervalo sean iguales. De manera general,

para un spline de orden k , en una partición del intervalo $[a, b]$ en subintervalos definidos por n puntos, tales que:

$$a = t_1 < t_2 < \dots < t_n = b$$

Se dice entonces que la función spline de orden k con nodos en t_1, t_2, \dots, t_n es una función S que satisface las siguientes condiciones:

1. En cada subintervalo $[t_{i-1}, t_i)$, S es un polinomio de grado menor o igual a k .
2. Los polinomios S_{i-1} y S_i interpolan el mismo valor en el punto t_i por lo que se garantiza que S sea continua en $[t_1, t_n]$.
3. S es $(k - 1)$ veces diferenciable.

Como se señala en Debon (2003, pp. 40-41) los splines cúbicos tienen una propiedad conocida e interesante. Dado una función $g(x)$ con dos derivadas continuas, se puede expresar la función objetivo como aquella que minimiza la siguiente sumatoria restringida:

$$\sum (y_i - g(x))^2 + \alpha \int_a^b (g''(t))^2 dt \quad (5)$$

Donde el primer término mide el ajuste entre lo observado y lo estimado (error al cuadrado), y el segundo penaliza los cambios de curvatura $g(x)$, en otras palabras, mediante esta representación se entiende que los splines cúbicos maximizan la bondad del ajuste (primer sumando) procurando que el ajuste sea suave (segundo sumando). El parámetro α es el parámetro que determina la importancia de cada sumando. Como se muestra a continuación, bajo ciertas condiciones, esta representación de la técnica de los splines coincide con la estrategia desarrollada por Whittaker- Henderson, método que se muestra a continuación.

B. Whittaker - Henderson

Como señala Debon (2003, pp. 49-50) el método propuesto por Whittaker-Henderson busca minimizar una función que toma en cuenta simultáneamente los criterios de ajuste y suavidad. Sea $\tilde{\theta}_x$ la estimación bruta de la probabilidad de fallecimiento y θ_x la cantidad a estimar, entonces mediante este método se busca el estimador de θ_x , $\hat{\theta}_x$, que minimiza:

$$WH(\theta) = \sum_{x=1}^n w_x (\tilde{\theta}_x - \theta_x)^2 + \alpha \sum_{x=z+1}^n (\Delta^z \theta_x)^2 \quad (6)$$

donde w_x es el ponderador, α es el parámetro de suavizamiento, Δ es el operador diferencial definido por:

$$\Delta \theta_x = \theta_x - \theta_{x-1} \quad (7)$$

y Δ^z es el diferencial aplicado z veces. El primer término de la función a minimizar mide el ajuste y el segundo la suavidad. En el caso particular que $z = 2$ la función a optimizar se vuelve:

$$\sum (y - g(x))^2 + \alpha \sum (\Delta^2 g(x))^2 \quad (8)$$

Donde $\Delta^2 g(x)$ representa el segundo diferencial en las estimaciones. El parámetro α , a medida que este es más grande, la importancia relativa de la varianza respecto al sesgo es mayor.

C. Regresión polinómica local

Si en (4) la perturbación aleatoria se asume idéntica e independientemente distribuida de media nula y varianza finita; $g(x_i)$ se aproxima con una función lineal en la vecindad del punto x_0 ; y se estima la regresión de tal forma que se da más importancia a las observaciones más cercanas a x_0 , entonces el tipo de regresión mediante el cual se estiman los parámetros se circunscribe al problema siguiente de mínimos cuadrados ponderados:

$$\min_{a,b} \sum_i [y_i - \beta_0(x_0) - \beta_1(x_0)x_i]^2 K\left(\frac{x_i - x_0}{h}\right) \quad (9)$$

Donde K en la literatura especializada es conocida como un Kernel, la cual cumple el rol de una función ponderadora (por ejemplo con distribución normal en la vecindad de x_0), h es la mayor distancia entre x_0 y algún otro punto de la vecindad $V(x_0)$. De esta manera, el valor estimado en esta ventana es de:

$$\hat{f}(x_0) = \hat{\beta}_0(x_0) + \hat{\beta}_1(x_0) \cdot x_0$$

el cual se constituye como el primer valor graduado. El siguiente valor graduado se obtiene al mover la ventana alrededor del siguiente x , y así sucesivamente con los otros valores de x . El método se generaliza si en lugar de una función lineal en (11) se emplea un polinomio de grado p , el procedimiento así definido recibe el nombre de regresión polinómica local.

Mediante esta estrategia de suavizamiento se requieren tomar 3 decisiones: (i) el tipo de función ponderadora, (ii) el grado del polinomio a usar, y (iii) el tamaño o ancho de banda en donde se realizarán las regresiones locales. Respecto al grado del polinomio, como se indica en Hastie et al. (2008, p. 197) el ajuste local lineal tiende a ser sesgado en regiones donde la verdadera función presenta curvaturas, mientras que el ajuste cuadrático normalmente es capaz de corregir este sesgo. Una inspección gráfica del ajuste medido como el tamaño del error es una herramienta útil para la elección del grado del polinomio.

Respecto a la forma funcional, conviene citar a Yatchew (2003, p. 32). De acuerdo a este autor la elección del ancho de banda es más importante que la selección de la función ponderadora. Para efectos de esta investigación se emplea un Kernel Tricúbico, el cual se caracteriza por dar mayor ponderación a la proximidad en relación a otras opciones usuales como el Kernel Gaussiano o el de tipo Epanechnikov⁴.

⁴La forma funcional del Kernel tricúbico es: $K(\mu) = (1 - |\mu|^3)^3$ cuando $\mu < 1$ y 0 en otro caso.

Finalmente, en el caso de la ventana o ancho de banda, su elección se puede automatizar tal que el error promedio de predicción sea el menor posible. En Tomas (2012, p. 33) se señala que la técnica de validación cruzada puede ser empleada para ello; por medio de esta técnica cada una de las observaciones es omitida usando las $n - 1$ observaciones restantes para predecirla. El ancho de banda óptimo se elige como aquel que genera el menor error cuadrático medio. La estrategia conocida como *Generalized Cross Validation* (GCV) es empleada como una aproximación de la validación cruzada (Hastie et al. 2008, p. 244).

3.7.3 Estimación en edades tempranas y avanzadas

En edades tempranas se emplea como insumo la tabla de vida calculada por el Instituto Nacional de Estadística e Informática (INEI) correspondiente al período 2010 al 2015, INEI (2009). La estrategia de estimación en esta parte de la distribución tiene tres etapas:

1. Las probabilidades presentadas de manera quinquenal en la tabla abreviada del INEI⁵ se anualizan empleando la siguiente relación: $q_x = 1 - (1 - {}_5q_x)^{0,2}$.
2. Una vez anualizado este valor se ajusta con un factor de proporcionalidad tal que la probabilidad obtenida a los 20 años de edad sea la misma que la obtenida también a los años con la graduación en la etapa de estimación previa.
3. Conocida las probabilidades en los puntos 0, 5, 10 y 15, las probabilidades entre estos valores se interpolan con una regresión polinómica local cuya variable dependiente es el logaritmo de la probabilidad que depende de la edad. El ejercicio fue realizado diferenciado por sexo.

Por su parte, en el caso de edades avanzadas, las probabilidades son extrapoladas usando una función biométrica; se plantean 5 funciones alternativas: Kannisto, Exponencial, Makeham, Gompertz, y Heligman-Pollard. El criterio de elección de los modelos es el de la parsimonia, esto es los que generan los menores errores

⁵A diferencia de una tabla completa en una tabla abreviada la información se presenta por rangos de edades.

con la menor cantidad de estimadores. Las funciones en el mismo orden antes listados se muestran a continuación:

$$q_x = 1 - \exp\left(-\left(\frac{\gamma_1 \cdot \exp(\gamma_2 \cdot x)}{1 + \gamma_1 \cdot \exp(\gamma_2 \cdot x)} + \gamma_3\right)\right)$$

$$q_x = 1 - \exp^{-\gamma x}$$

$$q_x = 1 - \gamma_2 \cdot (\gamma_0)^{-\gamma_1^x (\gamma_1 - 1)}$$

$$q_x = 1 - (\gamma_0)^{-\gamma_1^x (\gamma_1 - 1)}$$

$$q_x = \gamma_0^{(x+\gamma_1)^{\gamma_2}} + \gamma_3 \cdot \exp(-\gamma_4(\ln x - \ln \gamma_5)^2) + \frac{\gamma_6 \cdot \gamma_7^x}{1 + \gamma_6 \cdot \gamma_7^x}$$

3.7.4 Validación

En esta etapa se analiza la razonabilidad de los resultados contrastando los hallazgos con las tablas de vida vigentes para el SNP, también se comparan los resultados con la tabla de mortalidad estimada en el 2018 para uso del sistema privado de pensiones, SBS(2018).

Los resultados esperados en esta sección es una reducción y un aumento en la esperanza de vida de las mujeres y hombres, respectivamente; en el caso de los afiliados al sistema privado, se espera una menor esperanza de vida tanto en hombres como mujeres.

4 Resultados y discusión

En esta sección se plasma lo que metodológicamente se ha expuesto en la sección previa. Así, se estiman las tablas de acuerdo a los cuatro pasos antes descritos, a saber: (i) estimación la tasa bruta, (ii) graduación estos resultados para tener así una representación sin irregularidades de la tasa, (iii) estimar las probabilidades en la cola inferior y superior, y (iv) validación de los resultados.

También se incluye una sección con un análisis de robustez o sensibilidad ante cambios en la forma de estimar la tasa bruta de mortalidad, el método de graduación y la misma estrategia usada en la tabla vigente (SP2005).

4.1 Estimación según etapas

4.1.1 Tasa cruda o bruta de mortalidad

En la figura 2 se tiene la comparación de las cuatro estrategias señaladas en la sección previa. A continuación se resumen los principales hallazgos:

1. Los estimadores son similares, las diferencias son perceptibles solo cuando se consideran las diferencias en términos relativos, específicamente cuando se comparan respecto al estimador Máxima Verosimilitud (MV), encuyo caso las diferencias más grandes se estiman en alrededor del 7,5% en edades avanzadas.
2. Alta volatilidad en edades inferiores a los 20 años y un cambio de tendencia de la tasa a partir de los 95 años. La hipótesis para este comportamiento es la reducida cantidad de datos que se tiene en las colas de la distribución.
3. Aunque con mayor volatilidad respecto a sus pares, el estimador de Kaplan Meier (KM) tiende a estar ligeramente por encima de sus pares, con lo cual la expectativa de vida calculada con este estimador tendería a ser también ligeramente menor que sus pares. Específicamente este estimador es en promedio 1% mayor respecto al estimador actuarial y el de máxima verosimilitud.

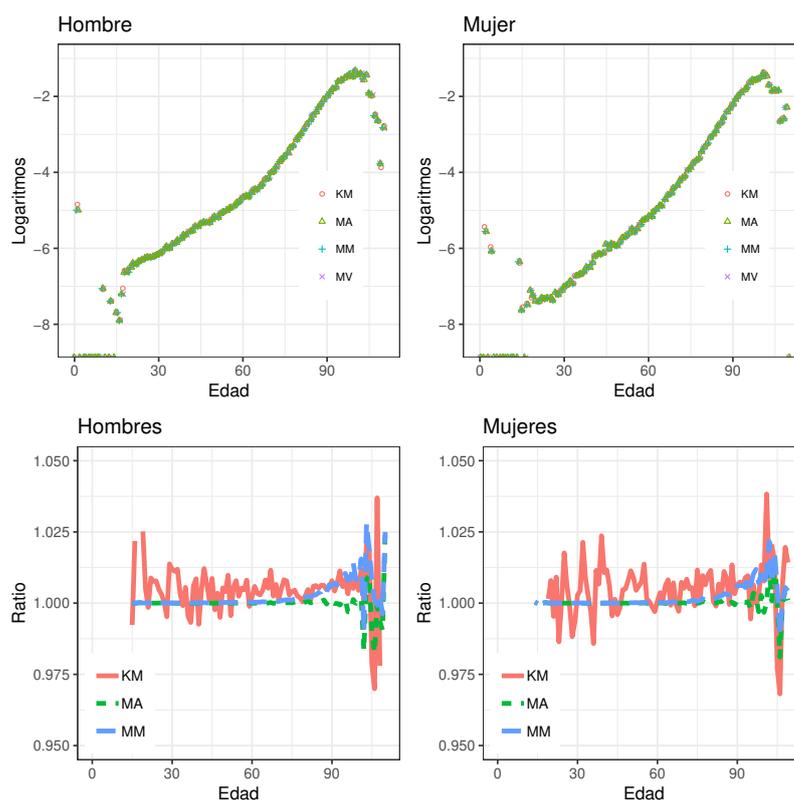


Figura 2. Tasa bruta según tipo de estimador.

MV:Máxima verosimilitud; MM:Método de momentos;

MA: Método actuarial; y KM: Kaplan Meier.

En esta investigación se emplea como rango relevante para el estudio los años entre los 20 y los 95; además, se ha seleccionado la estrategia KM debido a que es el único con la propiedad estadística de consistencia de los estimadores (Kaplan Meier,1958 y Charpentier, 2015).

4.1.2 Graduación

Como se señaló en la metodología existen distintas formas de graduar una tabla, siendo la estrategia conocida como Whittaker-Henderson la que goza de mayor popularidad entre los actuarios. Como ya se ha comentado, la estrategia tiene en la elección de al menos uno de los parámetros que la define su principal limitación. En esta investigación se emplea la regresión polinómica local, técnica propuesta por Cleveland (1979), que se caracteriza porque en ella se pueden establecer criterios objetivos que permiten determinar los parámetros que exige el método.

En la figura 3 se presenta la primera diferencia de los valores graduados. Para la determinación de los anchos de banda se emplea el criterio de validación cruzada. Se presentan los resultados en logaritmos para polinomios de orden cero, uno y dos; también se muestra la primera diferencia de los probabilidades en niveles. En logaritmos las diferencias no son visibles (tampoco lo son en niveles), estas se manifiestan en variaciones que son reveladas cuando se examinan las probabilidades en diferencias. De otro lado, respecto al grado del polinomio, no existiendo diferencias significativas en relación a los polinomios de grado cero y uno, se usa el polinomio de grado dos siguiendo a Hastie et al. (2008, p. 197).

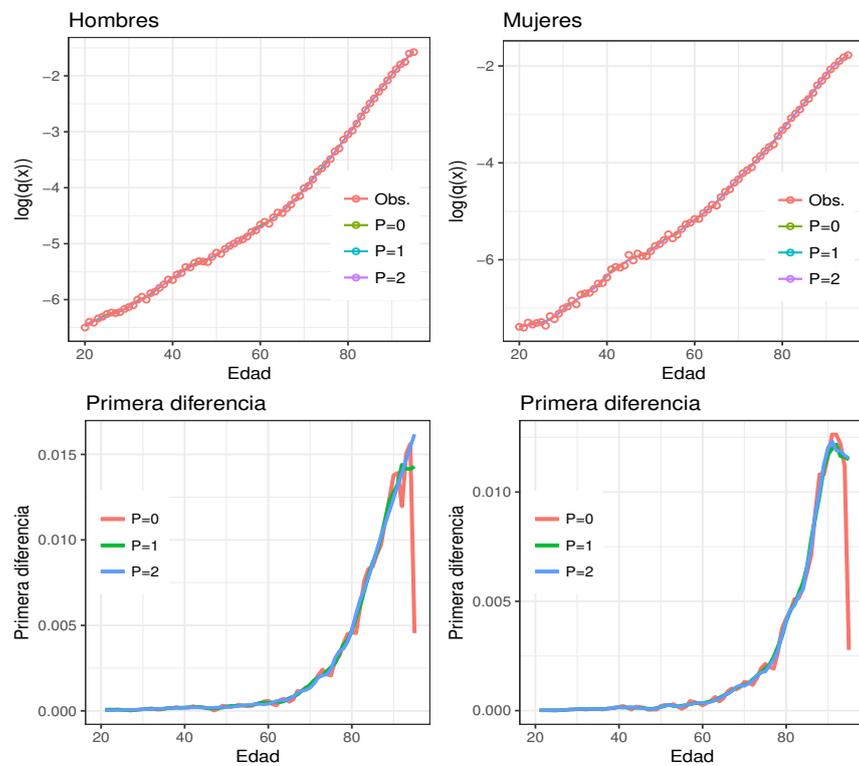


Figura 3. Graduación de acuerdo a la regresión polinómica local.

En la siguiente etapa se evalúa si las probabilidades graduadas gozan de buenas propiedades estadísticas. Tradicionalmente los actuarios aplican un conjunto de pruebas, en Forfar et al. (1988, pp. 40-48) se puede encontrar los detalles de las mismas, aunque en Konstantopoulos (2006, p. 59) se cuestiona la razonabilidad de alguna de las pruebas, independientemente de ello, lo que al menos se espera de las tasas ya graduadas o suavizadas es que preserven el comportamiento tendencial de las probabilidades originales, que los errores definidos como la diferencia entre valores originales y graduados se encuentren alrededor de 0, y que las probabilidades graduadas no presenten sesgos sistemáticos como superar a los valores brutos en edades tempranas y ser inferior a esta en el resto de la distribución, por ejemplo.

De acuerdo a la figura 4 y al cuadro 3 se tiene que la graduación realizada es la pertinente. En la figura 4 se muestran los intervalos de confianza de los valores brutos con sus respectivas probabilidades graduadas; en la mayoría de los casos las probabilidades graduadas están dentro de los intervalos; también, desde un punto de vista gráfico, no se aprecia evidencia de algún tipo sistemático de sesgo. Complementando esto, según las pruebas estadísticas mostradas en el cuadro 3, se concluye que tanto las probabilidades originales (valores brutos) y las probabilidades graduadas se encuentran estadísticamente correlacionadas, además de que en promedio las brechas entre valores originales y graduados se encuentran alrededor de cero.

Tabla 3

Tests estadísticos seleccionadas. P-Values

Prueba	Hipótesis nula	Mujer	Hombre
Correlación	Correlación igual a 0.	0.000	0.000
Prueba de medias	Promedios iguales.	0.997	0.998
Prueba de rangos ¹	Medianas iguales.	0.271	0.454
Prueba de signos ²	Medianas iguales.	0.731	0.909

¹ Prueba de signos de Wilcoxon. El test también considera el tamaño de la diferencia (rango).

² Si los signos de las diferencias no nulas son similares la mitad de estas son aproximadamente positivas y la otra mitad negativas.

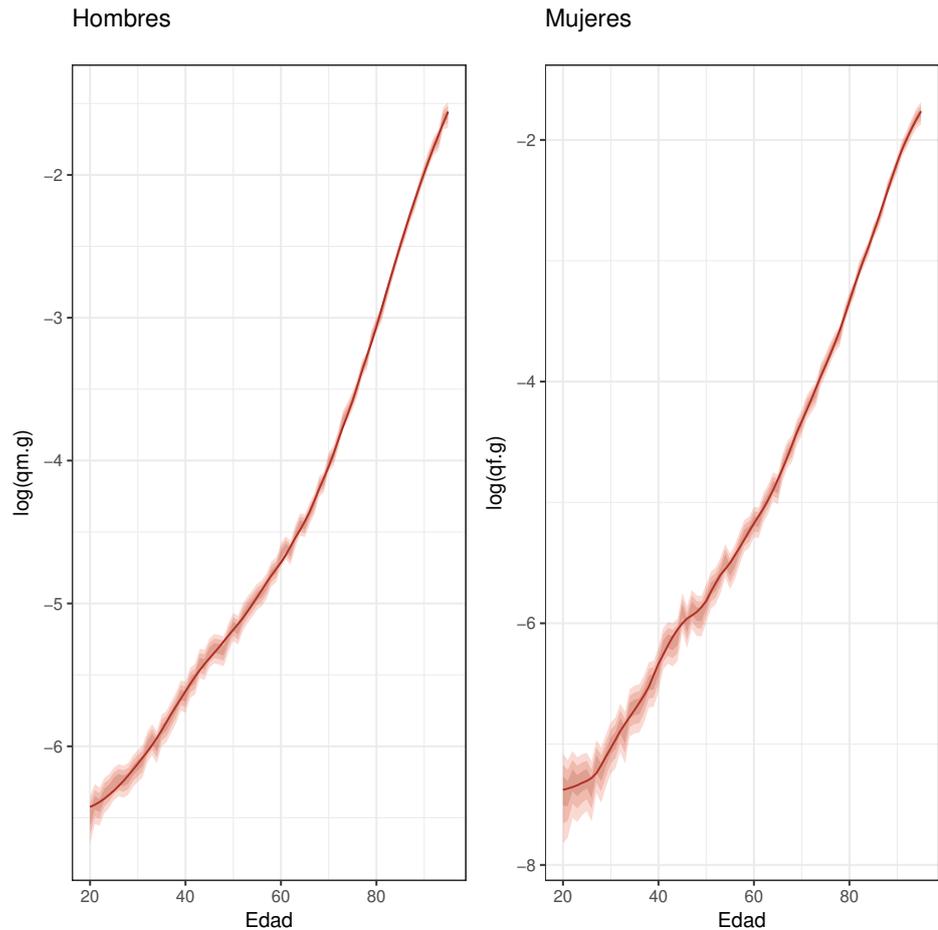


Figura 4. Probabilidades graduadas e intervalos de confianza. Intervalos con 1, 2 y 3 desviaciones estándar. La línea central es la variable graduada, los intervalos de confianza corresponden a las tasas brutas con errores estándar del estimador Kaplan-Meier.

4.1.3 Estimación en edades tempranas y avanzadas

En el caso de edades tempranas, siguiendo lo señalado en el desarrollo metodológico, se escalan las probabilidades de fallecimiento de todo el país para que estas sean consistentes con la población SNP menor a 20 años de edad. Vale la pena señalar que este grupo poblacional representa el 1 % en el caso de los pensionistas y el 0,8 % en el caso de los afiliados⁶.

De otro lado, en el caso de edades avanzadas, tal como lo indicado en la sec-

⁶Valores estimados para el año 2015 para el SNP.

ción metodológica, se elige la función biométrica que mejor ajuste a los datos observados. Específicamente se estiman los parámetros de las funciones empleando mínimos cuadrados no lineales⁷. La variable a explicar son las probabilidades graduadas.

Para la elección de los modelos se usó los criterios de parsimonia. Esto es, se eligen los modelos con menor error de ajuste y menor cantidad de parámetros empleados para ello. Los indicadores empleados para la elección fueron los criterios de información de Akaike y Schwarz (Ver Hastie et al. 2008, pp. 230-235 para una explicación extendida). En el caso del rango de edades a elegir para la estimación del modelo, de acuerdo a los resultados presentados en el anexo A.2, independiente de ello y el criterio de información a usar, siempre se eligen la función Kannisto. En el anexo A.3 se muestran los resultados de la estimación de dicha función tanto para hombres como para mujeres, se usó como muestra para la estimación los rango de edades 60 a 95 años.

Con el modelo elegido, el siguiente paso es la extrapolación de las probabilidades hasta una edad considerada como límite o edad máxima que se asume podrían vivir las personas (infinito actuarial en la jerga de los actuarios). Como se muestra en lo que sigue, la elección de dicho valor respecto a los valores usuales que se suelen asumir no tienen efectos significativos sobre la expectativa de vida, y por ende tampoco en estimaciones actuariales tales como el valor de la reserva matemática.

A fin de ilustrar el efecto marginal que se da al asumir distintos valores para el último valor de la tabla se presenta el cuadro 4. En este se muestran las expectativas de vida medida a los 20 y 65 años de edad considerando tres opciones: 110 (valor actual oficial), 115 y 120 años. Como se puede observar en el mencionado cuadro, no se dan diferencias significativas entre las alternativas consideradas. Así, mientras que en los hombres, el indicador es de 77.9 a los 20 años de edad y 84.7 si se calcula a los 65 años; en el caso de las mujeres la estimación es de 83 y de 86.9, a los 20 y 65 años, respectivamente. En ambos casos en la mayoría de veces las diferencias son visibles a partir del cuarto decimal. Siendo así, se fija la

⁷Se empleó para ello el algoritmo Levenberg-Marquardt.

edad máxima de la tabla en 110 años.

En la figura 5 se agregan a los valores graduados entre los 20 y 95 años las estimaciones para la cola inferior (menos de 20 años) y cola superior (más de 95 hasta los 110 años).

Tabla 4

Expectativa de vida^{1,2}

Medido a los 20 años		
Edad máxima asumida	Mujer	Hombre
$\omega=110$	82.9633197233986	77.8893140935905
$\omega=115$	82.9639907563457	77.8894596902662
$\omega=120$	82.9640094519793	77.8894629075012
Medido a los 65 años		
Edad máxima asumida	Hombre	Mujer
$\omega=110$	86.8560238265777	84.6733577409578
$\omega=115$	86.8567771715841	84.6735376224666
$\omega=120$	86.8567981605122	84.6735415972901

¹ Expresión matemática puede ser consultada en London (1988, p. 53).

² En todos los casos la extrapolación usa la función Kannisto pertinente.

4.1.4 Validación

Resultado de las estimaciones previas se tiene una nueva tabla que de aquí en adelante se denominará SNP 2017, en la figura 6 se compara este resultado con la tabla de mortalidad vigente (SNP 2005) y con la tabla empleada por las instituciones que supervisa la Superintendencia de Banca y Seguros (SBS), conocida como SPP 2017. Los resultados muestran que las probabilidades de fallecimiento q_x tienen el comportamiento esperado, creciente a partir de los 20 años y creciendo a tasas decrecientes en la cola superior; siendo esto último consistente con los resultados encontrados por investigaciones como las desarrolladas por Thatcher (1999) y Horiuchi-Wilmoth (1998).

En cuanto a la comparación entre las tablas SNP2017 y SPP2017, se tiene que las probabilidades de deceso son mayores en la SNP 2017, lo cual es consistente con la evidencia internacional que indica que las personas con mayor nivel

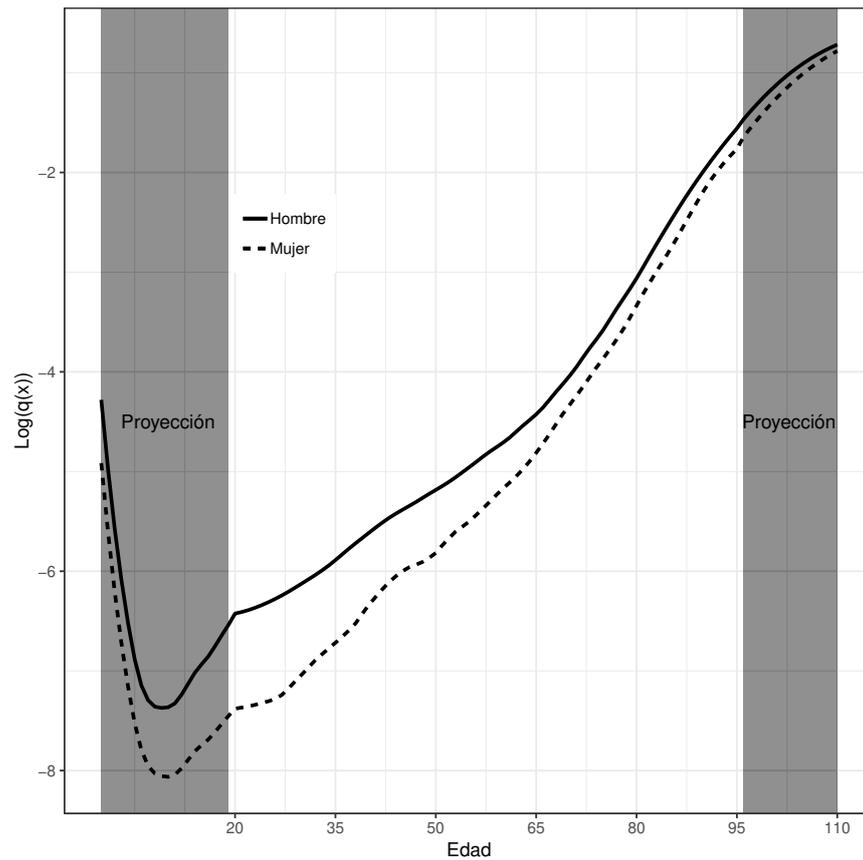


Figura 5. Proyección en colas inferior y superior según sexo.

socioeconómico tienen mayor esperanza de vida respecto los de menor nivel socioeconómico (Duleep (1986) y Deaton y Paxson (1998), por ejemplo) siendo que los afiliados al SNP tienen ingresos laborales que en promedio son la mitad que la de sus pares del régimen privado.

En la figura 7 se puede apreciar también que la expectativa de vida para los hombres es superior en casi todos los casos que lo informado por la tabla vigente, pero inferior que lo estimado para el SPP. En el caso de las mujeres, la brecha entre esperanzas de vida entre la estimación actual y la oficial es amplia, en todos los casos la estimación oficial sobrestima a su verdadero valor; visto de otra manera, la actual tabla asume esperanzas de vida para el caso de las mujeres como si estas formaran parte de la población femenina del SPP.

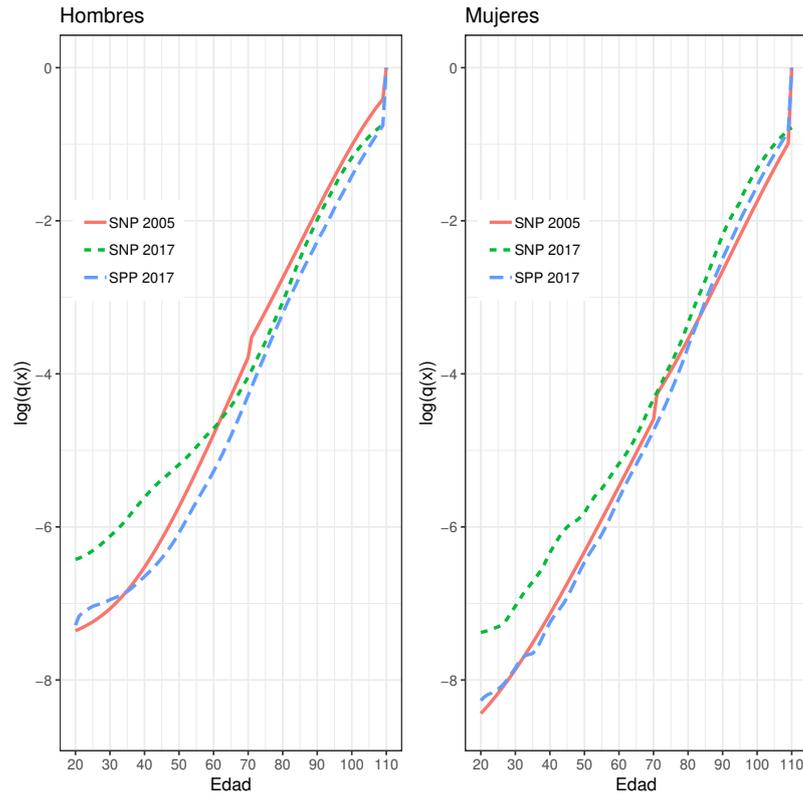


Figura 6. Logaritmo de probabilidades de fallecimiento según tablas.

4.2 Análisis de robustez

Se presentan los resultados sobre las probabilidades de fallecimiento y expectativa de vida de modificar en tres aspectos la estimación de las probabilidades propuestas. Las dos primeras tienen que ver con la forma tradicional con la que los actuarios estiman tablas (Ver por ejemplo SBS (2018)), bajo esta concepción se suele usar como estimador de tasa bruta el método actuarial y como método de graduación la técnica Whittaker-Henderson. El tercer aspecto a evaluar las implicancias de estimar la tabla usando la misma estrategia empleada para estimar la tabla SP2005 estimada por Melinsky (2006).

En el caso de los dos primeros aspectos se trata de un efecto parcial atribuible exclusivamente al cambio en la estrategia de estimación del método de estimación de tasa bruta o método de graduación según corresponda, todos los demás pasos se mantienen según lo planteado en el presente estudio. De otro lado, en el caso del tercer aspecto, se trata de un efecto combinado, pues en el enfoque planteado

en la metodología SP2005 se contempla el método actuarial como estrategia para estimar la tasa bruta y el método de Wittaker-Henderson para la graduación; además, para la extrapolación en edades avanzadas se emplea la función biométrica de Makeham, entre las características más importantes. El anexo A.4 es informativo respecto a dicha metodología.

En las figuras 8 y 9 se muestran las probabilidades y esperanzas de vida bajo cada escenario, además se presenta el ratio en relación a la nueva tabla de mortalidad estimada (SNP 2017). De acuerdo a los resultados las diferencias son marginales y solo se notan cuando la comparación se realiza respecto a ratios, particularmente en el caso de la expectativa de vida. De esta manera, de los tres aspectos evaluados, en términos relativos la metodología oficial (SP2005) es la que más dista de lo aquí propuesto (SNP2017), siendo las diferencias más significativas las que se dan en edades más avanzadas.

Por su parte, respecto a la tasa bruta estimada con el criterio actuarial, se tiene también menores esperanzas de vida, en este caso las brechas se reducen hasta ser similares en las edades más altas, consistente con lo anterior la diferencia más importante se da a los 20 años. Así, la esperanza de vida a los 20 años se calcula en 83 y 78 para mujeres y hombres, respectivamente, mientras que las mismas métricas pero estimados a los 65 años dan como resultado 87 y 85. Por último, en el caso de las opciones de graduación no se aprecia diferencias significativas entre el método de regresión polinómica local y Whittaker-Henderson, siendo la ventaja de la regresión que su implementación es más simple al no requerir elección arbitraria de parámetros, lo cual si se necesita hacer bajo la estrategia Whittaker-Henderson.

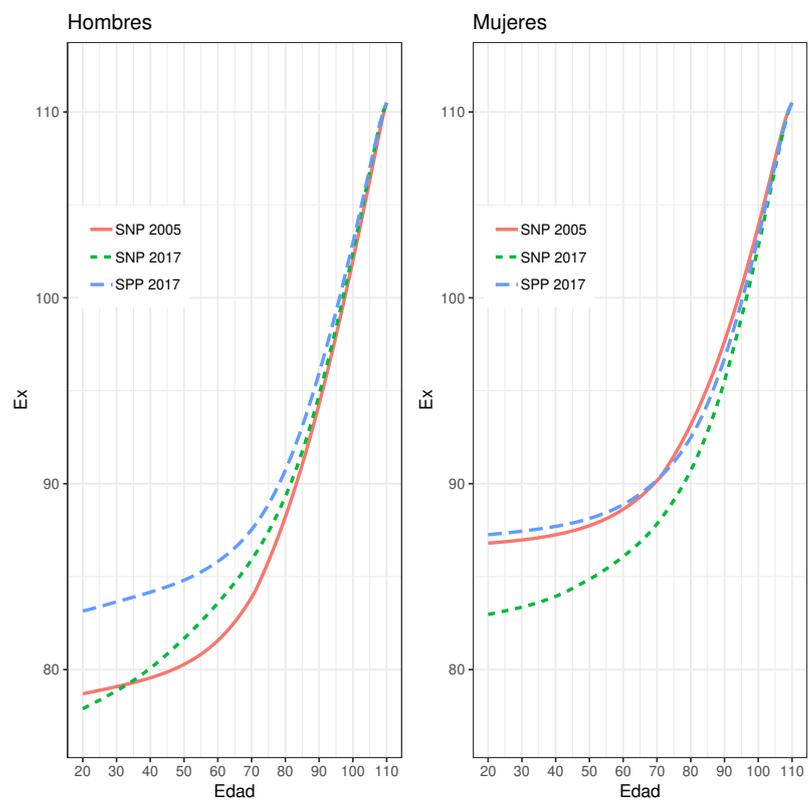


Figura 7. Esperanza de vida según tablas.

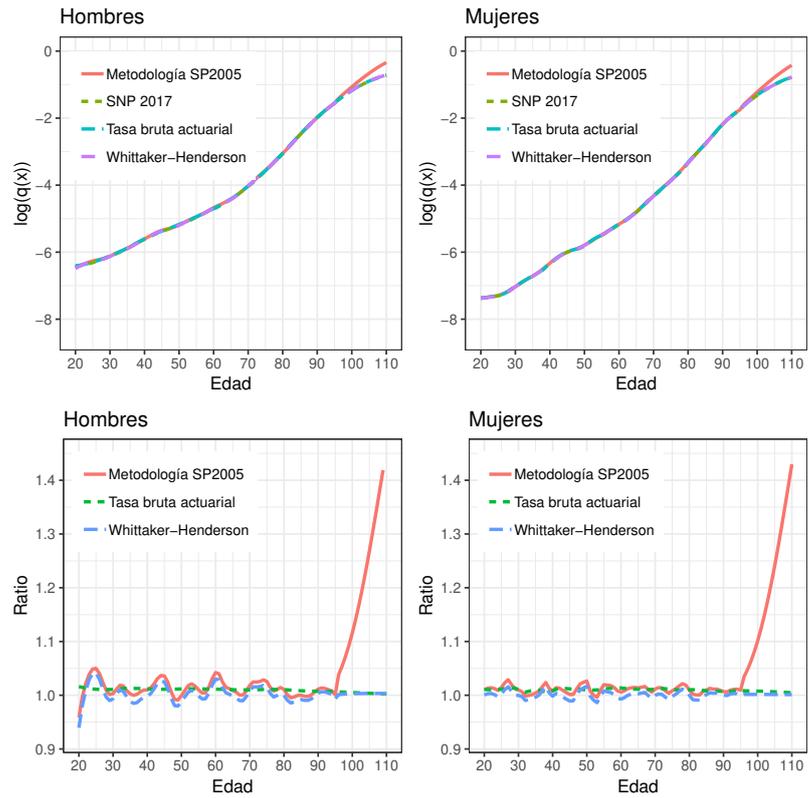


Figura 8. Sensibilidad de las probabilidades de fallecimiento. Panel superior son probabilidades en logaritmos y panel inferior son ratios respecto a metodología propuesta.

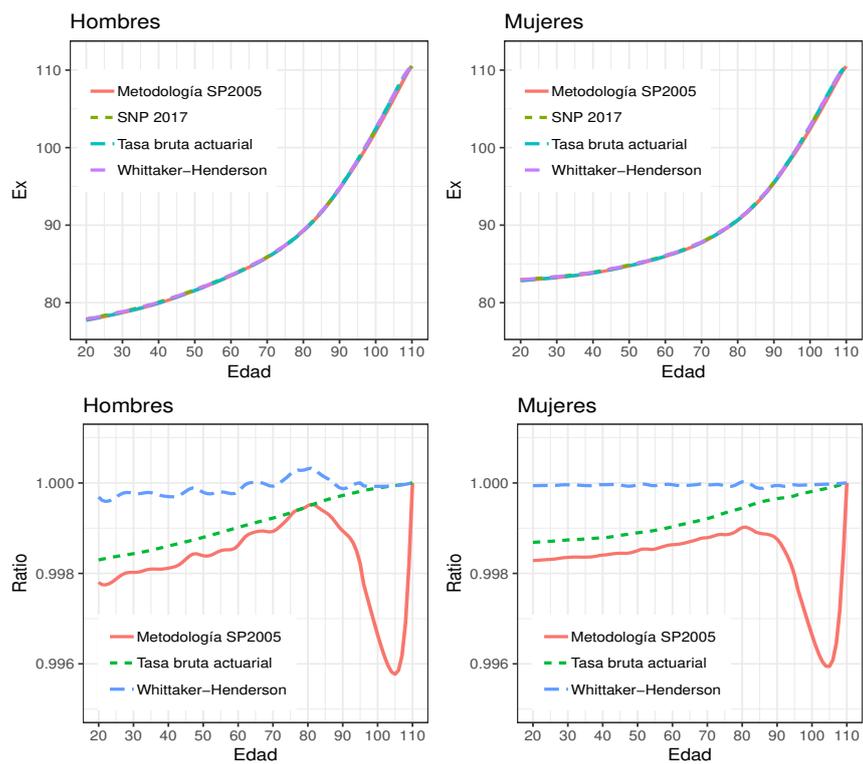


Figura 9. Sensibilidad de las esperanzas de vida. Panel superior son esperanzas de vida y panel inferior son ratios respecto a metodología propuesta.

5 Impactos

De acuerdo al cuadro 5 la esperanza de vida de las mujeres y hombres a los 65 años en la nueva tabla son de 87 y 85, respectivamente; significando una reducción y un aumento de 3 y 2 años respecto a la tabla oficial, brecha similar a lo observado en la tabla del sistema privado de pensiones (SPP 2017). La mayor esperanza de vida en los afiliados al SPP es consistente con la literatura que sostiene que las personas de mayor nivel socioeconómico tienden a ser más longevas (Ver por ejemplo Von Gaudecker y Scholz (2007), Dowd et al. (2011) y Belloni et al. (2013)), así se tiene que la diferencia medida a los 65 años es de 2 y 3 años para hombres y mujeres, respectivamente.

Establecido ya que las esperanzas de vida de las mujeres estarían sobrestimadas y la de los hombres subestimadas de continuar usando las tablas SP2005, lo que sigue es determinar sus implicancias financieras. En el caso de las mujeres una menor esperanza de vida tiene un impacto doble, un impacto directo corresponde a una menor reserva si es que estas son titulares, y uno indirecto en el caso que el derecho sea por viudez (reducción también). Por su parte, un aumento en la expectativa masculina se corresponde solo con un efecto directo de incremento en la reserva solo de los titulares, pues en el régimen pensionario del SNP solo las viudas son beneficiarias de una pensión vitalicia.

Como consecuencia de corroborar la hipótesis de una menor esperanza de vida para las mujeres y una mayor esperanza para los hombres, se tiene que el efecto neto sobre las reservas actuariales netas es de alrededor $S/2$ mil millones según se detalla en el cuadro 6. Con información del año 2017 se tiene que la reserva actuarial neta descendería en más de $S/2$ mil millones, es decir, una disminución de 2%; descenso explicado por el resultado obtenido en las mujeres las mismas que experimentan una reducción de $S/4$, 217 millones mientras que en los hombres este indicador se incrementa en $S/2$, 185 millones.

Tabla 5

Esperanza de vida según escenario

Edad	Hombre			Mujer		
	SNP17 ¹	SP2005 ²	SPP17 ³	SNP17 ¹	SP2005 ²	SPP17 ³
0	75	77	82	81	86	86
20	78	79	83	83	87	87
65	85	83	87	87	89	90

¹ Nueva tabla.

² Tabla vigente en el sistema nacional de pensiones.

³ Tabla vigente en el sistema privado de pensiones.

Tabla 6

Impacto en las reservas matemáticas 2017 (en millones de soles)

	Hombres y mujeres			
	SP 2005	SNP2017	Δ	Δ %
Aportes	83,406	82,070	-1,336	-1.6 %
Reserva bruta	154,707	152,728	-1,979	-1.3 %
Reserva pensionaria	42,125	40,736	-1,389	-3.3 %
Reserva neta	113,426	111,394	-2,032	-1.8 %
Hombres				
	SP 2005	SNP2017	Δ	Δ %
Aportes	53,409	52,388	-1,021	-1.9 %
Reserva bruta	102,388	103,254	866	0.8 %
Reserva pensionaria	73,096	75,281	298	1.2 %
Reserva neta	73,096	75,281	2,185	3.0 %
Mujeres				
	SP 2005	SNP2017	Δ	Δ %
Aportes	29,997	29,682	-315	-1.1 %
Reserva bruta	52,319	49,474	-2,845	-5.4 %
Reserva pensionaria	18,008	16,321	-1,687	-9.4 %
Reserva neta	40,330	36,113	-4,217	-10.5 %

6 Conclusiones

En esta investigación se utilizó información de aproximadamente 5 millones de personas con la que se estimó una tabla de mortalidad para la población afiliada al sistema público de pensiones peruano, Sistema Nacional de Pensiones. Los datos corresponden a registros administrativos recolectados por la ONP durante cinco años, información que es enriquecida con datos provenientes de RENIEC.

Como metodología se contempló una estrategia de tres etapas. En primer lugar se estima la tasa bruta (o tasa cruda) de mortalidad usando el estimador producto límite propuesto por Kaplan y Meier (1958). En una segunda etapa las tasas brutas se gradúan o suavizan empleando el enfoque de regresión polinómica local. En el último paso, para tratar con la cantidad reducida de información en las colas, las probabilidades fueron calculadas con otra estrategia: en edades inferiores a los 20 años se escalaron los cálculos para toda la población realizados por el INEI; para edades más allá de los 95 años, las probabilidades fueron extrapoladas hasta los 110 empleando una función del tipo Kannisto. Con el fin de evaluar la relevancia del último valor de la tabla se evaluó el efecto sobre la esperanza de vida al cambiar el valor máximo a 115 o 120, se encontró que este indicador no cambia significativamente.

Los resultados revelan que las probabilidades de fallecimiento se encuentran subestimadas y sobreestimadas en el caso de los hombres mujeres, respectivamente. Producto de esto la diferencia entre mujeres y hombres en expectativa de vida medido a los 65 años disminuye de 7 a 2 años. Como consecuencia de lo anterior el valor presente de los recursos monetarios necesarios para financiar el pago de las pensiones de los hombres aumentaría y el de las mujeres se reduciría. Se estima que para el año 2017 el pasivo actuarial neto se reduciría en poco más de S/2 mil millones, es decir, un 2% menos en relación a la estimación oficial.

7 Recomendaciones

Pasos siguientes a partir de este estudio pueden ampliar el mismo al considerar estimaciones que tomen en cuenta que cada vez las personas viven más, la litera-

tura que se ha enfocado en este tipo de estimaciones suelen denominar a este tipo de instrumentos como tablas de mortalidad dinámicas. Sin importar la estrategia para este tipo de estimación, los procedimientos clásicos como los planteados por Lee y Carter (1992) o similares son intensivos en el uso de información histórica, con lo cual ello se constituye en un reto en países como el nuestro en el que la disponibilidad de información histórica es reciente. La tabla aquí presentada es la primera tabla estimada con población única y exclusivamente perteneciente al SNP, por lo que cualquier estrategia de ajuste en la tabla que busque convertir esta en dinámica, o construir lo que los actuarios denominan “factores de mejoramiento”, debe lidiar con las limitaciones de información existente, mostrando las opciones factibles con sus ventajas y desventajas.

A Anexos

A.1 Muestra de datos utilizados

Tabla 7

Muestra de datos utilizados

id	Nacimiento	Inicio Obs.	Fin Obs.	Sexo	Fallec.=1	Tipo
21	1/02/1932	1/01/2013	13/10/2017	M	0	Aseg. a Pens.
22	1/11/1940	1/01/2013	13/10/2017	M	0	Aseg. a Pens.
23	13/09/1945	1/01/2013	13/10/2017	M	0	Aseg. a Pens.
24	26/02/1952	1/01/2013	13/10/2017	F	0	Aseg. a Pens.
25	14/10/1967	1/01/2013	17/07/2013	M	1	Asegurados
26	14/07/1947	1/01/2013	13/10/2017	M	0	Asegurados
27	24/09/1994	9/09/2013	4/01/2017	M	1	Asegurados
28	9/03/1982	1/01/2013	1/04/2014	M	1	Asegurados
29	11/09/1948	1/01/2013	13/10/2017	M	0	Aseg. a Pens.
30	26/09/1950	1/01/2013	13/10/2017	M	0	Aseg. a Pens.
31	16/05/1963	1/01/2013	21/09/2014	F	1	Asegurados
32	20/09/1952	1/01/2013	13/10/2017	M	0	Aseg. a Pens.
33	29/03/1949	1/01/2013	13/10/2017	F	0	Asegurados
34	2/01/1962	1/01/2013	26/12/2014	F	1	Asegurados
35	19/02/1960	1/01/2013	13/10/2017	F	0	Asegurados
36	2/09/1955	1/01/2013	13/10/2017	M	0	Aseg. a Pens.
37	28/02/1953	1/01/2013	13/10/2017	M	0	Aseg. a Pens.
38	17/07/1950	1/01/2013	13/10/2017	M	0	Aseg. a Pens.
39	15/09/1947	1/01/2013	13/10/2017	F	0	Aseg. a Pens.
40	5/11/1951	1/01/2013	13/10/2017	M	0	Aseg. a Pens.
41	27/10/1941	1/01/2013	20/01/2016	M	1	Asegurados
42	11/01/1942	1/01/2013	13/10/2017	M	0	Aseg. a Pens.
43	24/03/1934	1/01/2013	13/10/2017	M	0	Asegurados
44	20/07/1930	1/01/2013	20/01/2017	M	1	Asegurados

A.2 Criterios de información para selección de modelo para extrapolación

Tabla 8

AIC según modelos y rango de edades

Hombres					
Rango de edad	Gompertz	Makeham	H-P	Kannisto	Exponencial
30 a 95	-600.16	-650.06	-596.71	-720.02	-619.26
40 a 95	-506.96	-542.11	-504.37	-625.60	-515.16
50 a 95	-421.79	-435.99	-421.49	-508.71	-415.14
60 a 95	-339.30	-337.86	-347.18	-389.78	-323.12
Mujeres					
Rango de edad	Gompertz	Makeham	H-P	Kannisto	Exponencial
30 a 95	-635.96	-647.54	-641.18	-685.63	-628.45
40 a 95	-531.61	-539.34	-536.51	-576.62	-522.86
50 a 95	-431.58	-432.99	-437.13	-464.15	-420.21
60 a 95	-332.43	-330.44	-339.63	-352.83	-321.71

Tabla 9

BIC según modelos y rango de edades

Hombres					
Rango de edad	Gompertz	Makeham	H-P	Kannisto	Exponencial
30 a 95	-593.59	-641.31	-590.14	-711.26	-610.50
40 a 95	-500.88	-534.01	-498.29	-617.50	-507.06
50 a 95	-416.30	-428.68	-416.01	-501.39	-407.82
60 a 95	-334.55	-331.52	-342.43	-383.44	-316.79
Mujeres					
Rango de edad	Gompertz	Makeham	H-P	Kannisto	Exponencial
30 a 95	-629.39	-638.79	-634.61	-676.87	-619.69
40 a 95	-525.53	-531.24	-530.43	-568.52	-514.76
50 a 95	-426.10	-425.68	-431.64	-456.84	-412.89
60 a 95	-327.68	-324.10	-334.88	-346.49	-315.38

A.3 Función Kannisto: Estimación

Tabla 10

Estimación de modelos^a

Parámetro	Hombre	Mujer
γ_1	2.213e-06*** (.4875e-07)	1.569e-06*** (.6348e-07)
γ_2	.125*** (.214e-03)	.126e*** (.389e-03)
γ_3	.435e-02*** (.2396e-03)	.2377e-02*** (.3831e-03)
$\hat{\sigma}_\epsilon$	0.000828	0.00139
Grad. de Lib.	48	48

**** significancia al 0.0001.

^a Función biométrica de Kannisto:

$$\hat{q}_x = 1 - \exp\left(-\left(\frac{\gamma_1 \cdot \exp(\gamma_2 \cdot x)}{1 + \gamma_1 \cdot \exp(\gamma_2 \cdot x)} + \gamma_3\right)\right).$$

A.4 Metodología SP 2005

La metodología empleada se encuentra documentada en el informe final presentado por Eduardo Melinsky el 14 de noviembre de 2016, documento elaborado en el marco de la Cooperación Técnica no reembolsable destinada al proyecto de la “Elaboración de Tablas de Mortalidad en Perú”, convenio aprobado por Resolución Suprema Nro. 002-2006-EF. Los datos corresponden a los registros administrativos tanto del SNP como del SPP, el período de observación es de 1993 a 2005. La estrategia de estimación se resume a continuación.

En el rango de edades de 20 a 95 años se estima la tasa bruta de mortalidad con el método actuarial, se gradúan las tasas observadas mediante la técnica de Whittaker-Henderson tipo B, estrategia que busca minimizar la siguiente fórmula:

$$\sum_{x=1}^n W_x \cdot (q_x - \hat{q}_x)^2 + k \cdot \sum_{x=1}^{n-2} (\Delta^z q_x)^2$$

donde:

- $W(x)$, son coeficientes de ponderación.
- q_x , son los valores observados.
- \hat{q}_x , son los valores suavizados.
- Δ^z , es la diferencia z de las $q(x)$ suavizadas.

En el caso del orden de las diferencias Z se asume el valor de 3 y respecto al valor de k (constante que regula la importancia que se le asigna a la suavidad contra la proximidad con los datos observados) el estudio señala que se selecciona un valor que le permita realizar el mejor suavizamiento y que cumpla con los siguientes test: Chi Cuadrado, Barnett, Desvios Estandarizado, Signos, Stevens, Desvios Acumulados y Correlacion Serial. De este modo se determina que los valores de k son de 0.84 para los hombres y 5.6 para el caso de las mujeres.

De otro lado, para las edades menores a 20 años se tomó como base al comportamiento observado en la Tabla de Mortalidad B-85 (Tabla de Mortalidad usada en ese entonces en el SPP para los beneficiarios) ajustando los valores en función de la relación entre las probabilidades de muerte de ambas tablas.

En el caso de las edades mayores a 95 años se extrapola las tasas obtenidas mediante el ajuste de la Ley de Makeham para las edades de 65 a 95 años, las cuales fueron ajustadas previamente por el método Whittaker-Henderson. En ese sentido, el estudio señala que dado que el ajuste previamente mencionado explicaba razonablemente las probabilidades de muerte de 65 a 95 años, es posible extrapolar las mismas hasta la edad de 110 años.

Estimadas las probabilidades para los tres tramos de edades se ajustaron las mismas para las edades que se encuentran en el rango que se encuentra entre los 20 y 110 años con el fin que la expectativa de vida sea similar a la observada en la tabla de mortalidad chilena conocida como RV 2004. Dicho ajuste se hizo en dos tramos, entre los 20 y 70 años, y entre los 71 y los 110 años, para dicho objetivo se estimaron funciones biométricas de acuerdo a la ley de mortalidad de Makeham diferenciando por sexo y ajustando los parámetros hasta conseguir el objetivo de conseguir las mismas esperanzas de vida.

A.5 Tabla de mortalidad

Tabla 11

Tabla de mortalidad-Hombres. (SNP 2017)

Edad	q_x	Edad	q_x	Edad	q_x
0	0.013838	37	0.003106	74	0.025375
1	0.006954	38	0.003279	75	0.027827
2	0.003817	39	0.003456	76	0.030923
3	0.002282	40	0.003644	77	0.034420
4	0.001476	41	0.003839	78	0.038050
5	0.001025	42	0.004035	79	0.042085
6	0.000788	43	0.004228	80	0.046795
7	0.000679	44	0.004412	81	0.052451
8	0.000637	45	0.004588	82	0.058959
9	0.000629	46	0.004766	83	0.066091
10	0.000633	47	0.004955	84	0.073915
11	0.000658	48	0.005166	85	0.082502
12	0.000718	49	0.005385	86	0.091920
13	0.000802	50	0.005602	87	0.102139
14	0.000896	51	0.005831	88	0.113089
15	0.000975	52	0.006084	89	0.124776
16	0.001055	53	0.006376	90	0.137209
17	0.001168	54	0.006698	91	0.150394
18	0.001303	55	0.007039	92	0.164333
19	0.001445	56	0.007406	93	0.179022
20	0.001621	57	0.007806	94	0.194456
21	0.001646	58	0.008201	95	0.210635
22	0.001679	59	0.008578	96	0.230880
23	0.001718	60	0.008981	97	0.249537
24	0.001765	61	0.009453	98	0.268740
25	0.001818	62	0.010040	99	0.288350
26	0.001878	63	0.010666	100	0.308219
27	0.001945	64	0.011280	101	0.328188
28	0.002020	65	0.011962	102	0.348092
29	0.002105	66	0.012792	103	0.367772
30	0.002194	67	0.013849	104	0.387074
31	0.002287	68	0.015027	105	0.405854
32	0.002388	69	0.016237	106	0.423985
33	0.002500	70	0.017590	107	0.441361
34	0.002625	71	0.019195	108	0.457894
35	0.002771	72	0.021162	109	0.473519
36	0.002934	73	0.023257	≥ 110	0.488191

Tabla 12

Tabla de mortalidad-Mujeres.(SNP 2017)

Edad	q_x	Edad	q_x	Edad	q_x
0	0.007332	37	0.001367	74	0.019119
1	0.003695	38	0.001467	75	0.020900
2	0.002030	39	0.001615	76	0.022949
3	0.001212	40	0.001771	77	0.025287
4	0.000781	41	0.001914	78	0.028045
5	0.000540	42	0.002066	79	0.031473
6	0.000412	43	0.002216	80	0.035582
7	0.000352	44	0.002354	81	0.040190
8	0.000326	45	0.002477	82	0.044998
9	0.000317	46	0.002577	83	0.050285
10	0.000315	47	0.002644	84	0.055815
11	0.000322	48	0.002719	85	0.062390
12	0.000343	49	0.002832	86	0.069908
13	0.000374	50	0.002978	87	0.078896
14	0.000407	51	0.003222	88	0.088918
15	0.000433	52	0.003459	89	0.100163
16	0.000458	53	0.003700	90	0.112185
17	0.000494	54	0.003883	91	0.124507
18	0.000536	55	0.004083	92	0.136406
19	0.000579	56	0.004347	93	0.148284
20	0.000624	57	0.004636	94	0.159957
21	0.000631	58	0.004958	95	0.171517
22	0.000639	59	0.005311	96	0.193539
23	0.000649	60	0.005665	97	0.210902
24	0.000661	61	0.006022	98	0.229047
25	0.000672	62	0.006417	99	0.247871
26	0.000689	63	0.006891	100	0.267250
27	0.000717	64	0.007453	101	0.287046
28	0.000765	65	0.008132	102	0.307103
29	0.000822	66	0.008911	103	0.327259
30	0.000881	67	0.009802	104	0.347348
31	0.000945	68	0.010870	105	0.367204
32	0.001016	69	0.012031	106	0.386669
33	0.001078	70	0.013154	107	0.405599
34	0.001143	71	0.014393	108	0.423864
35	0.001210	72	0.015774	109	0.441355
36	0.001284	73	0.017371	≥ 110	0.457984

Referencias

- ÁGOSTON, K. (2001), *Comparison of estimators for probability of death used in actuarial science*. Journal of the Hungarian Central Statistical Office, 150-160.
- BELLONI, M., ALESSIE, R., KALWIJ, A. Y MARINACCI, C. (2013) *Lifetime income and old age mortality risk in Italy over two decades*. Demograph Research 29 (45), 1261-1298.
- BROFFITT, J. D. (1984), *Maximum likelihood alternatives to actuarial estimators of mortality rates*, Transactions of Society of Actuaries, 36(1), 77-142.
- BUENO, J.; ESCUDER, R.; PAVIA, J. y GUILLÉN, M. (2008), *Determinación de los tantos brutos de mortalidad*, Anales del Instituto de Actuarios Españoles, Nº 14 3ra época, 109-133.
- CANADIAN INSTITUTE OF ACTUARIES, (2014) *Final Report: Canadian Pensioners' Mortality*, Document 214013, Canada-Ottawa,.
- CHARPENTIER, A. (ED.) (2014), *Computational Actuarial Science With R*, CRC Press.
- CLEVELAND, W.S. (1979), *Robust Locally Weighted Regression and Smoothing Scatterplots*, Journal of the American Statistical Association 74(368), 829-836.
- DEATON, A. y PAXSON, C. (1998), *Aging and Inequality in Income and Health*, The American Economic Review 88(2), 248-253.
- DEBON, A. y SALA, R. (2001), *Tablas dinámicas de mortalidad y sobrevivencia [Mimeo]*. Universidad de Valencia.
- DEBON, A. (2003), *Graduación de tablas de mortalidad. Aplicaciones actuariales. (tesis doctoral)*. Universidad de Valencia, Valencia, España.
- DEBON, A.; MONTES, F. y SALA, R. (2005), *A comparison of Parametric Models for Mortality Graduation. Application to mortality data for the Valencia Region*, Statistics and Operations Research Transactions 29(2), 269-288.

- DICKSON, D., HARDY, M. y WATERS, H. (2009), *Actuarial Mathematics for Life Contingent Risks*. First Edition, New York: Cambridge University Press.
- DOWD, J., ALBRIGHT, J., KALWIJ, A., RAGHUNATHAN, T., SCHOENI, R., LECLERE, F. y KAPLAN, G. (2011) *Deeper and wider: income and mortality in the USA over three decades*. International Journal of Epidemiology 40, 183-188.
- DULEEP, H. O. (1986), *Measuring the Effect of Income on Adult Mortality Using Logitudinal Administrative Record Data*, Journal of Human Resources 21(2), 238-251.
- FORFAR, D.O.; MCCUTCHEON, J.J. y WILKIE, A.D. (1988), *On Graduation by Mathematical Formula*, Journal of the Institute of Actuaries, 115(1), 1-149.
- GIL, J.; HERAS, A. y VILAR, J. L. (1999), *Matemática de los Seguros de Vida*, Madrid: Editorial MAPFRE.
- HASTIE, T., TIBSHIRANI, R. y FRIEDMAN, J. (2008), *The Elements of Statistical Learning*. Second Edition, New York, USA, Springer.
- HELIGMAN, L. y POLLARD, J. (1980) *The Age Pattern of Mortality*. Journal of the Institute of Actuaries 107(1), 49-80.
- HORIUCHI, S. y WILMOTH, J.(1998) , *Deceleration in the Age Pattern of Mortality at Older Ages*. Demography, 35(4), 391-412.
- I MASSONS, J. M. D. (1992), *Una aplicación del análisis de la supervivencia en ciencias de la salud*, Anuario de psicología/The UB Journal of psychology, (55), 109-142.
- INEI (2009), *Perú: Estimaciones y Proyecciones de Población 1950-2050*, Boletín de Análisis Demográfico 36,
- KALBFLEISCH, J. y PRENTICE, R. (2002), *The Statistical Analysis of Failure Time Data*. Second Edition, New Yersey: John Wiley and Sons, Inc.
- KAPLAN, E. y MEIER, P. (1958). *Nonparametric estimation from incomplete observations*. Journal of the American statistical association, 53(282), 457-481.

- KONSTANTOPOULOS, T. (2006), *Notes on Survival Models*, School of Mathematical and Computer Sciences Heriot-Watt University,
- LEE, R. y CARTER, L. (1992) *Modeling and Forecasting U.S. Mortality*. Journal of the American Statistical Association 87(419), 659-671.
- LI, SIU-HANG y CHAN, WAI-SUM (2004) *Estimation of Complete Period Life Tables for Singaporeans*. Journal of Actuarial Practice 1(11) 129-146.
- LONDON, D. (1988) *Survival Models and Their Estimation*. Second Edition, ACTEX Publications.
- MELINSKY, E. (2006) *Elaboración de Tablas de Mortalidad en Perú*. Lima, Perú. Manuscrito no publicado.
- OFICINA DE NORMALIZACIÓN PREVISIONAL (ONP) (2018), *Guía Técnica Para el Cálculo de Reservas Actuariales*. Manuscrito no publicado. Disponible en: http://www.onp.gob.pe/seccion/centro_de_documentos/Documentos/2477.pdf
- SÁNCHEZ, J. (2002). *Estimación Actuarial Versus Estimación por el Método de los Momentos para la Probabilidad de Muerte*. Anales del Instituto de Actuarios Españoles, 8, 111-135.
- SUPERINTENCIA DE BANCA Y SEGUROS (SBS) (2018), *Desarrollo de Tablas de Mortalidad Aplicables al Sistema Privado de Pensiones del Perú*. Manuscrito disponible en: <http://www.sbs.gob.pe/tmortalidad/paginas/home.html>
- THATCHER, A. (1999), *The Long-Term Pattern of Adult Mortality and the Highest Attained Age*. Journal of the Royal Statistical Society, 162(1), 5-43.
- TOMAS, J. (2012), *Univariate Graduation of Mortality by Local Polynomial Regression*. Bulletin Francais d'Actuariat 12, 23, 5-58.
- VON GAUDECKER, H. Y SCHOLZ, R. (2007) *Differential mortality by lifetime earnings in Germany*. Demograph Research 17 (4), 83-108.
- YATCHEW, A. (2003), *Semiparametric Regression for the Applied Econometrician*, Cambridge University Press.

ZARRUK, A., VILLEGAS, A. y ORTIZ, F. (2011). *La Industria Aseguradora en Colombia*. Colombia: FASECOLDA,.