UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO

FACULTAD DE CIENCIAS

ANÁLISIS Y ESTIMACIÓN DE MODELOS DE SUPERVIVENCIA TABULARES

REPORTE DE

SEMINARIO DE TITULACIÓN

QUE PARA OBTENER EL TÍTULO DE:

ACTUARIO

P R E S E N T A:

LUIS LUJÁN RANGEL

TUTOR

ACT. JAIME VÁZQUEZ ALAMILLA

2007

FACULTAD DE CIENCIAS
UNAM

Análisis y Estimación de Modelos de Supervivencia Tabulares

Luján Rangel Luis

2007

Hoja de Datos del Jurado

1.- Datos del alumno
Luján
Rangel
Luis
52 77 18 04
Universidad Nacional Autónoma de México
Facultad de Ciencias
Actuaría
09810386-1

2.- Datos del tutor
Act
Jaime
Vázquez
Alamilla

3.- Datos del sinodal 1
M en C
Salvador
Zamora
Muñoz

4.- Datos del sinodal 2
Mat
Margarita
Chávez
Cano

5.- Datos del sinodal 3
Dra
Guillermina
Eslava
Gómez

6.- Datos del sinodal 4
Act
Francisco
Sánchez
Villarreal

7.- Datos del trabajo escrito
Análisis y Estimación de Modelos de Supervivencia Tabulares
114 p
2007

Índice general

Prefacio

Los modelos de supervivencia, empezaron a tener un gran desarrollo en el siglo XX, ya que lo modelos de supervivencia solo se enfocaban a los seguros y finanzas, mientras que con los avances tecnológicos se empezó a trabajar en disminuir los riesgos de muerte.

En el primer capítulo se aborda la definición de modelo de supervivencia, así como los modelos de supervivencia actuariales (el modelos selecto y el modelo agregado), las diferentes formas de los modelos de supervivencia (forma paramétrica y forma tabular), así como su estimación y algunos ejemplos.

El capítulo dos trata del modelo de supervivencia tabular, en específico, la tabla de mortalidad o tabla de vida, las funciones derivadas de ℓ_x (el número de personas vivas a edad x) y los diferentes supuestos para edades no enteras (lineal, exponencial e hiperbólico).

En el tercer capítulo se hace un análisis, separando los modelos en estudios con datos completos e incompletos. Tambien se hace un análisis de la estimación de momentos; se estudia la diferencia entre un estudio con un medio de decremento simple y doble, utilizando distintos supuestos de distribución.

La estimación de momentos por la aproximación de Hoem en un medio de decremento doble, es vista como otro método distinto a la estimación actuarial y a la estimación máximo verosímil, así como el estimador producto-límite (estimador de Kaplan-Meier).

El último capítulo trata de aplicaciones actuariales, en específico las estimaciones con suposicones exponenciales, así como un aplicación donde se utilizan las herramientas vistas durante el trabajo, por ejemplo la exposición exacta, la exposición actuarial y la exposición programada, así como diferentes casos especiales.

Para la lectura de este trabajo se requieren conocimientos básicos de estadística, probabilidad y cálculo diferencial e integral, y puede usarse como texto de apoyo en la nueva matería de estadística III, ya que es un texto introductorio a los modelos de supervivencia tabulares.

Capítulo 1

Introducción

El análisis estadístico de supervivencia, el cual es referido como análisis de tiempo de vida, tiempo de supervivencia o tiempo de falla es un importante tema para investigadores de diversas áreas como ingeniería y ciencias biomédicas.

Por conveniencia cuando se habla de modelos de supervivencia se hace referencia a datos de tiempo de falla. En ocasiones los eventos de interés son muertes de individuos en un sentido real y el tiempo de vida es la edad actual de un individuo o quizás el tiempo de supervivencia medido desde algún punto inicial particular.

Los siguientes ejemplos ilustran algunos casos típicos en los cuales surge el planteamiento y uso de los modelos de supervivencia:

(a) Algunos tipos de artículos fabricados pueden ser reparados ante una falla. En este caso uno podría estar interesado en el periodo de tiempo entre fallas sucesivas de un artículo o pieza y referirse a éste como el tiempo de vida.

(b) En estudios médicos de enfermedades fatales el interés radica en el tiempo de supervivencia de individuos con esta enfermedad, medido desde la fecha del diagnóstico o desde algún otro punto inicial.

1.1 Definición y formas de los modelos de supervivencia

Un *Modelo de Supervivencia* es la función de distribución de probabilidad de un tipo especial de variable aleatoria, la cual se describe a continuación.

Sea T una variable aleatoria no negativa que representa el tiempo de vida (o de falla) de individuos en una población. Suponiendo que un individuo de la población se mantiene con vida en el tiempo $t = 0$, se está interesado en la probabilidad de que dicho individuo continúe con vida en el futuro (para cualquier t). Simbólicamente esta probabilidad está denotada por $S(t)$.

Así la variable aleatoria considerada, definida como el tiempo de falla de un individuo sabiendo que existe al tiempo $t = 0$, es frecuentemente llamada *variable aleatoria del tiempo de falla.* Si T denota el tiempo de falla, entonces la probabilidad de que el individuo continúe funcionando al tiempo t es la misma que la probabilidad de que el tiempo de falla exceda el valor de t. Formalmente:

$$S(t) = P(T \geq t). \tag{1.1}$$

(Ver [2], pág 11).

Por la definición de T se tienen las propiedades:

$$S(0) = 1 \tag{1.2}$$

$$\lim_{t \to \infty} S(t) = 0. \tag{1.3}$$

y $S(t)$ es una función no creciente.

Ahora, si T es el tiempo de falla de un individuo que existe en $t = 0$, entonces T es también el *tiempo futuro de vida* del individuo medido desde $t = 0$.

Las formas en que se pueden presentar los modelos de supervivencia son: paramétrica y tabular (no paramétrica). Cuando las probabilidades de supervivencia son dadas mediante una fórmula matemática se dice que el modelo $S(t)$ está en forma paramétrica, esto es porque los valores de $S(t)$ dependen de uno o más parámetros. Por ejemplo, si se supone un modelo de supervivencia $S(t) = e^{-\theta t}$, que corresponde a la función de supervivencia cuando T tiene una distribución de probabilidad exponencial, $S(t)$ depende del parámetro θ, del cual posteriormente se estima su valor. Por su parte, en los modelos de supervivencia tabulares los valores numéricos de $S(t)$ son presentados para ciertos valores seleccionados de t, más comúnmente valores enteros. El modelo de supervivencia tabular más común es el conocido como *tabla de vida* o *tabla de mortalidad*.

Dado que el modelo de supervivencia tabular $S(x)$ presenta valores sólo para números enteros $x = 0, 1, ...$ es claro que esta forma de modelo presenta una carencia de respuesta para valores de x fraccionales, por lo tanto es necesario hacer un supuesto acerca de la forma de $S(x)$ entre enteros subsecuentes. Este supuesto es llamada distribución supuesta de mortalidad, la cual nos definirá los valores de $S(x)$ para toda $x \geq 0$. Para esta nueva distribución entre enteros subsecuentes los tres supuestos más comunes son: lineal, exponencial e hiperbólico. Los modelos de supervivencia actuariales usualmente están dados en forma tabular. En la siguiente sección se describen esos modelos.

1.2 Modelos de supervivencia actuariales

En el ámbito actuarial, los modelos estudiados se dividen principalmente en dos grandes ramas: vida y no vida. Como su nombre lo menciona, en el área de vida la variable aleatoria T representa el tiempo de vida de una persona. En los modelos de supervivencia actuariales de vida, se toma en cuenta la edad cronológica de los individuos en estudio, reconociendo que la supervivencia está en función de la edad.

Por su parte, en los modelos actuariales de no vida, la variable aleatoria T para este caso representa el tiempo de ocurrencia de un evento asociado con una persona distinto de la muerte, por ejemplo: tiempo de ocurrencia de una enfermedad o el tiempo de ocurrencia de un accidente automovilístico. Para este caso también se puede plantear un modelo de supervivencia mediante la función $S(t) = P(T \geq t)$, ya que la variable T sigue representando el tiempo de ocurrencia o falla de un evento.

En la rama de vida, existen dos tipos de modelos de supervivencia actuariales los cuales son: el modelo agregado y el modelo selecto.

1.2.1 El Modelo Selecto

Considerando un modelo de supervivencia que es utilizado para el cálculo de primas de seguros con respecto a personas seleccionadas para una cierta cobertura a edad (x), la función $S(t)$ puede tomar distintos valores dependiendo de la edad (x), por lo cual necesitamos que $S(t)$ también dependa del valor (x) cuando $t = 0$. Para estos casos se hace uso del símbolo $S(t; x)$. En este nuevo contexto, la edad seleccionada (x) es llamada *variable concomitante*. En el ámbito actuarial la forma más usual de esta forma de modelo es simplemente tener por separado una función $S(t)$ para cada valor (x).

La edad (x) no es la única variable concomitante que puede tener influencia sobre el modelo de supervivencia, otra variable importante podría ser el sexo, en este caso la variable concomitante probablemente reflejaría la necesidad de tener modelos por separado de hombres y mujeres.

1.2.2 El Modelo Agregado

Este modelo de supervivencia actuarial se distingue por la particularidad de que el tiempo de inicio $t = 0$ coincide con el nacimiento del individuo $(x = 0)$. Notando que ambas variables se mueven juntas podemos ocupar cualquiera de ellas para representarse en el modelo de interés, por conveniencia se usa x. Es claro que para este caso $S(x)$ y $S(t)$ son funciones idénticas pero ambas difieren de $S(t; x)$. En este caso, la variable aleatoria X es la edad de muerte o el tiempo futuro de vida, así como en su caso T representa una variable aleatoria del tiempo de muerte o tiempo de falla.

1.3 Estimación

Una vez especificada la forma del modelo $S(t)$ (o $S(x)$), es necesario establecer una aproximación o estimación al modelo real, la cual se denota por $\widehat{S}(t)$. Se utilizarán varias aproximaciones para estimar $S(t)$ dependiendo de la naturaleza de los datos y del diseño del estudio.

Para el modelo tabular usualmente se estiman las probabilidades condicionales de supervivencia sobre pequeños intervalos unitarios (generalmente periodos de un año) y se obtiene una estimación de $S(t)$ a partir de dichas probabilidades.

Respecto al modelo paramétrico, la estimación de los parámetros desconocidos de la supuesta forma de distribución adoptada para el modelo $S(t)$ produce el modelo estimado $\widehat{S}(t)$. En el modelo paramétrico se consigue la estimación por una primera aproximación de sucesiones de intervalos de probabilidad condicional y se adecua a la forma paramétrica elegida y estas aproximaciones llevan a una prueba de hipótesis para modelos paramétricos.

1.4 Ejemplos

Los siguientes ejemplos describen situaciones en donde pueden usarse los modelos de supervivencia.

Ejemplo 1.1 *En la tabla 1.1 se muestra una tabla de vida para menores de 18 años como ejemplo de un modelo de supervivencia actuarial no paramétrico, en la cual se inicia con un radix de 100,000 personas vivas a la edad 0 y se tienen los datos de las muertes año con año. Apartir de las muertes d_x se calculan los valores para q_x y p_x. Para encontrar el valor estimado de $S(t)$ se realiza el cálculo $\widehat{S}(x) = p_0 \cdot p_1 \cdot \ldots \cdot p_{x-1}$.*

Ejemplo 1.2 *Unos científicos describieron la experiencia de supervivencia de un grupo de pacientes quienes estuvieron bajo tratamiento en conexión con un tipo de enfermedad fatal, dando como resultado la tabla 1.2.*

En este caso se calcularon las probabilidades de supervivencia en cada intervalo de unidad de tiempo, las cuales representan el modelo estimado.

Ejemplo 1.3 *En la tabla 1.3 se muestran 2 grupos de pacientes de leucemia, el tiempo de falla (tiempo hasta la muerte) en semanas y con 2 tipos de sangre y con glóbulos blancos incluidos. En el lado izquierdo se muestran los oh positivos en una muestra de N=17, mientras que del lado derecho los oh negativos con una muestra de N=16.*

En este caso se ilustra como es que se puede hacer uso de una variable concomitante al poder plantear modelos de supervivencia separados para cada tipo de Oh. (Ver [3] pág. 9).

Edad x	l_x	d_x	q_x	p_x	$\widehat{S}(x)$
0	100000	2042	0.02042	0.97958	1
1	97958	131	0.00134	0.99866	0,97958
2	97827	119	0.00122	0.99878	0,97827
3	97708	109	0.00112	0.99888	0,97708
4	97599	101	0.00103	0.99897	0,97599
5	97498	95	0.00097	0.99903	0,97498
6	97403	90	0.00092	0.99908	0,97403
7	97313	86	0.00088	0.99912	0,97313
8	97227	84	0.00086	0.99914	0,97227
9	97143	82	0.00084	0.99916	0,97143
10	97061	82	0.00084	0.99916	0,97061
11	96979	82	0.00085	0.99915	0,96979
12	96897	83	0.00086	0.99914	0,96897
13	96814	84	0.00087	0.99913	0,96814
14	96730	86	0.00089	0.99911	0,96730
15	96644	87	0.00090	0.99910	0,96644
16	96557	89	0.00092	0.99908	0,96557
17	96468	91	0.00094	0.99906	0,96468
18	96377	93	0.00096	0.999004	0,96377

Tabla 1.1

Intervalo (años)	# muertes	# retiros	# expuestos	Prob.sup.estimada
[0, 1)	90	0	374	0.759
[1, 2)	76	0	248	0.556
[2, 3)	51	0	208	0.420
[3, 4)	25	12	157	0.350
[4, 5)	20	5	120	0.291
[5, 6)	7	9	95	0.268
[6, 7)	4	9	79	0.254
[7, 8)	1	3	66	0.250
[8, 9)	3	5	62	0.237
[9, 10)	2	5	54	0.228
[10, ∞)	47	0	47	0

Tabla 1.2

GBI Oh+	Tiempo de muerte	GBI Oh-	Tiempo de muerte
2300	65	4400	56
750	156	3000	65
4300	100	4000	17
2600	134	1500	7
6000	16	9000	16
10500	108	5300	22
10000	121	10000	3
17000	4	19000	4
5400	39	27000	2
7000	143	28000	3
9400	56	31000	8
32000	26	26000	4
35000	22	21000	3
100000	1	79000	30
100000	1	100000	4
52000	5	100000	43
100000	65	-	-

Tabla 1.3

Capítulo 2

La tabla de mortalidad

Las tablas de mortalidad (también llamadas de vida) fueron desarrolladas por actuarios independientemente del desarrollo de la teoría estadística de modelos de supervivencia con distribuciones de probabilidad y su estimación. Por esta razón, la terminología y notación es diferente a la naturaleza estocástica de los demás modelos.

La tabla de vida es una tabla de valores numéricos de $S(x)$ para ciertos valores de x como la siguiente:

x	$S(x)$
0	1.00000
1	.97408
2	.97259
3	.97160
4	.97082
$\vdots$	$\vdots$
109	.00001
110	.00000

Tabla 2.1

La tabla de vida muestra valores de $S(x)$ para todos los valores enteros de x $(x = 0, 1, ...)$, por esta razón es claro poner un límite superior a x en donde a partir de este valor de x , $S(x)$ tome el valor cero. Generalmente se usa ω para este valor. Se puede ver que $S(\omega - 1) > 0$ y que $S(\omega) = 0$. En la tabla 2.1, $\omega = 110$.

2.1 La forma tradicional de la tabla de mortalidad

El modelo tabular usualmente se presenta con las siguientes características:

1.- Los valores de $S(x)$ son multiplicados generalmente por 100,000 para transformar los valores decimales en enteros.

2.- La columna de $S(x)$ es cambiada a ℓ_x donde ℓ representa "el número de vivos".

Desde $S(0) = 1$, ℓ_0 es la constante que multiplicando a las $S(x)$ las transforma en ℓ_x , esta constante es llamada **radix** (generalmente es de 100,000).

Formalmente:

$$\ell_x = \ell_0 \cdot S(x). \tag{2.1}$$

(Ver [1], pág 64)

Se transforma entonces la tabla 2.1 en la tabla 2.1a

x	ℓ_x
0	100000
1	97,408
2	97,259
3	97,160
4	97,082
$\vdots$	$\vdots$
109	1
110	0

Tabla 2.1a

La ventaja básica de esta forma tradicional de tabla es que es susceptible de interpretación. La diferencia entre $S(x)$ y ℓ_x es que las $S(x)$ son probabilidades y ℓ_x es el **número esperado** de sobrevivientes a edad x de un grupo de ℓ_0 nuevos nacimientos.

Se define :

$$d_x = \ell_x - \ell_{x+1} \tag{2.2}$$

y

$$ {}_{n}d_x = \ell_x - \ell_{x+n}. \tag{2.3} $$

(Ver [1], pág 59)

Donde ℓ_x representa el número de sobrevivientes a edad x , ℓ_{x+n} representa el número de sobrevivientes a edad $x+n$ y ${}_{n}d_x$ es el número de muertos entre las edades x y $x+n$.

Ahora:

$$ q_x = \frac{d_x}{\ell_x} \tag{2.4} $$

y

$$ {}_{n}q_x = \frac{{}_{n}d_x}{\ell_x}. \tag{2.5} $$

(Ver [1], pág 53)

Esta ${}_{n}q_x$ representa la probabilidad condicional de muerte a la edad $x+n$, dado que tiene la edad x.

Finalmente:

$$ {}_{n}p_x = 1 -_{n} q_x = \frac{\ell_{x+n}}{\ell_x}. \tag{2.6} $$

Esta ${}_{n}p_x$ representa la probabilidad de sobrevivir a la edad $x+n$, dado que ya se tiene la edad x.

2.2 Otras funciones derivadas de ℓ_x

Aunque la tabla de vida solamente toma valores ℓ_x para ciertos valores de x (usualmente enteros), se supondrá que la función ℓ_x que produce estos valores es continua y diferenciable para obtener otras funciones de mortalidad.

2.2.1 La fuerza de mortalidad

La derivada de ℓ_x se puede interpretar como la tasa de cambio instántanea de ℓ_x. Esta derivada es negativa porque ℓ_x es una función decreciente y representa el número de personas que se están muriendo a la edad x.

Se define:

$$\mu_x = \frac{-\frac{d}{dx}\ell_x}{\ell_x}, \tag{2.7}$$

como la **fuerza de mortalidad** a la edad x. Como se vio $\ell_x = \ell_0 \cdot S(x)$, y de (2.7) se obtiene:

$$\lambda(x) = \frac{-\frac{d}{dx}S(x)}{S(x)} = \frac{f(x)}{S(x)}, \tag{2.8}$$

debido a que $S(x) = P\,(X > x) = 1 - F(x)$, donde $F(x)$ representa la función de distribución de la variable aleatoria tiempo de vida X. $\lambda(x)$ es llamada la **tasa de riesgo**. Se observa que μ_x y $\lambda(x)$ son funciones idénticas, al dividir el numerador y denominador de μ_x entre ℓ_0.

Por lo tanto, de (2.8) se tiene:

$$\mu_x = \frac{-\frac{d}{dx}S(x)}{S(x)} = -\frac{d}{dx}\ln S(x), \tag{2.9}$$

(Ver [1], pág 55)

y al despejar $S(x)$ da:

$$S(x) =_x p_0 = \exp[-\int_0^x \mu_y dy], \tag{2.10}$$

que puede ser interpretado como el factor de decremento, el cual reduce el conjunto inicial de tamaño ℓ_0 al tamaño ℓ_x a la edad x.

Ahora:

$$\mu_{x+s} = \frac{-\frac{d}{ds}\ell_{x+s}}{\ell_{x+s}}. \tag{2.11}$$

Esta es una forma de la fuerza de mortalidad que será usada frecuentemente.

2.2.2 La función de densidad de probabilidad

Los datos mostrados en la tabla de vida son una representación de una distribución de la variable aleatoria del tiempo futuro de vida X (con función de densidad de probabilidad). Se tiene que $f(x) = \lambda(x) \cdot S(x)$, en la tabla de vida $\lambda(x) = \mu_x$ y $S(x) = \frac{\ell_x}{\ell_0}$, por lo tanto:

$$f(x) = \mu_x \cdot (\frac{\ell_x}{\ell_0}) = \mu_x \ {}_xp_0 \tag{2.12}$$

(Ver [1], pág 56)

Ahora que se conoce la función de densidad se puede encontrar $E[X]$ denotada por $\overset{o}{e}$ de la siguiente manera:

$$\overset{o}{e}_0 = E[X] = \int_0^\infty x \cdot f(x)\ dx = \int_0^\infty x \cdot_x p_0 \mu_x\ dx \tag{2.13}$$

(Ver [1], pág 68)

Integrando por partes con $u = x$ y $dv = \mu_x\ {}_xp_0$, se tiene que: $du = dx$ y dividiendo el numerador y denominador de (2.7) entre ℓ_0 y usando el hecho de que ${}_xp_0 = \frac{\ell_x}{\ell_0}$, se obtiene que

$$v = \int_0^\infty {}_xp_0\mu_x\ dx = \int_0^\infty \ -\frac{d}{dx}\ {}_xp_0 = -\ {}_xp_0.$$

De lo anterior, se tiene que:

$$\overset{o}{e}_0 = E[X] = \int_0^\infty {}_xp_0\ dx = \frac{1}{\ell_0}\int_0^\infty \ell_x\ dx. \tag{2.14}$$

Se define

$$T_0 = \int_0^\infty \ell_x\ dx. \tag{2.15}$$

Por lo tanto

$$\overset{o}{e}_0 = E[X] = \frac{T_0}{\ell_0}. \tag{2.16}$$

La función T_0 es un caso especial de T_x, definida por

$$T_x = \int_x^\infty \ell_y\ dy. \tag{2.17}$$

(Ver [1], pág 71)

El símbolo T_x representa el número total de años vividos después de la edad x por el grupo de sobrevivientes con ℓ_0 miembros iniciales.

El segundo momento de X está dado por

$$E[X^2] = \int_0^\infty x^2 \ _xp_0 \ \mu_x dx, \tag{2.18}$$

Integrando por partes con $u = x^2$, $du = 2x \cdot dx$, $v = -_xp_0$, $dv =_x p_0 \ \mu_x$, se obtiene

$$E[X^2] = 2 \cdot \int_0^\infty x \cdot_x p_0 dx = \frac{2}{\ell_0} \cdot \int_0^\infty x\ell_x dx,$$

volviendo a integrar por partes con $u = x$, $du = dx$, $v = -T_x$, $dv = \ell_x$, esto es porque de (2.17) se tiene que:

$$\frac{d}{dx}T_x = -\ell_x. \tag{2.19}$$

y por el teorema fundamental del cálculo, $\int \ell_x \ dx = -T_x$. Por lo tanto:

$$E[X^2] = \frac{2}{\ell_0} \cdot \int_0^\infty x \cdot \ell_x dx = \frac{2}{\ell_0} \cdot \int_0^\infty T_x dx,$$

y se define

$$Y_0 = \int_0^\infty T_x \ dx. \tag{2.20}$$

Obteniendo

$$E[X^2] = \frac{2 \cdot Y_0}{\ell_0}. \tag{2.21}$$

Finalmente

$$Var(X) = E[X^2] - \{E[X]\}^2 = \frac{2 \cdot Y_0}{\ell_0} - \left(\frac{T_0}{\ell_0}\right)^2. \tag{2.22}$$

2.2.3 Probabilidades condicionales y densidades

Las probabilidades condicionales conocidas son: ${}_np_x$ y ${}_nq_x$ pero ahora nos interesa ${}_{n|m}q_x$ la cual representará la probabilidad de que una persona viva a edad x muera entre las edades $x+n$ y $x+n+m$. En términos de probabilidad formal:

$${}_{n|m}q_x = \Pr[(x+n) < X < (x+n+m) \mid X > x].$$

En otros términos, puede expresarse como la probabilidad de que una persona de edad x sobreviva n años, pero muera dentro de los siguientes m años. Puede ser escrita por

$${}_{n|m}q_x = {}_n p_x \cdot {}_m q_{x+n}, \tag{2.23}$$

en términos de ℓ_x se tiene

$${}_{n|m}q_x = \frac{\ell_{x+n}}{\ell_x} \cdot \frac{{}_m d_{x+n}}{\ell_{x+n}} = \frac{{}_m d_{x+n}}{\ell_x}. \tag{2.24}$$

La función de densidad de probabilidad de muerte a la edad y cuando se está vivo a la edad x $(y > x)$, está dada por $f(y \mid X > x) = \frac{f(y)}{S(x)}$. Se tiene que $f(y) = \frac{1}{\ell_0}\ell_y\mu_y$ y $S(x) = \frac{\ell_x}{\ell_0}$, por lo tanto

$$f(y \mid X > x) = \frac{\ell_y\mu_y}{\ell_x} = {}_{y-x}\, p_x\mu_y, \tag{2.25}$$

cambiando $s = y - x$ como $y = x + s$ se tiene:

$$f(s \mid X > x) = {}_s\, p_x\mu_{x+s}, \tag{2.26}$$

donde la variable aleatoria S denota la cantidad de tiempo de vida futura de una persona viva a la edad x. Si tanto el numerador como el denominador de (2.11) es dividido entre ℓ_x se obtiene

$$\mu_{x+s} = \frac{-\frac{d}{ds}\ {}_sp_x}{{}_sp_x}, \tag{2.27}$$

lo cual muestra que:

$$\frac{d}{ds}\ {}_sp_x = -{}_sp_x\ \mu_{x+s}. \tag{2.28}$$

El **tiempo futuro de vida esperado** de una persona viva a la edad x está dado por

$$\overset{o}{e}_x = \int_0^\infty s \cdot_s p_x \; \mu_{x+s} \; ds = \int_0^\infty {}_sp_x \; ds, \tag{2.29}$$

esto resulta al integrar por partes con $u = s$, $du = ds$, $v = - \; {}_sp_x$ y $dv = {}_sp_x$ μ_{x+s}. Ademas ${}_sp_x = \frac{\ell_{x+s}}{\ell_x}$ por lo tanto:

$$\overset{o}{e}_x = \frac{1}{\ell_x} \cdot \int_0^\infty \ell_{x+s} \; ds = \frac{1}{\ell_x} \cdot \int_x^\infty \ell_y \; dy = \frac{T_x}{\ell_x}. \tag{2.30}$$

Analógamente a lo que se hizo en (2.18) y (2.21) se tiene

$$E[S^2] = \frac{2 \cdot Y_x}{\ell_x}, \tag{2.31}$$

donde

$$Y_x = \int_x^\infty T_y \; dy. \tag{2.32}$$

Finalmente

$$Var(S) = E[S^2] - \{E[S]\}^2 = \frac{2 \cdot Y_x}{\ell_x} - \left(\frac{T_x}{\ell_x} \right)^2. \tag{2.33}$$

2.2.4 La tasa central de mortalidad

La tasa central de mortalidad está definida por

$${}_n\mu_x = \frac{\int_0^n {}_sp_x \; \mu_{x+s} \; ds}{\int_0^n {}_sp_x \; ds}, \tag{2.34}$$

(Ver [1], pág 70), como ${}_sp_x = \frac{\ell_{x+s}}{\ell_x}$ se tiene

$$
{}_n\mu_x = \frac{\int_0^n \ell_{x+s}\ \mu_{x+s}\ ds}{\int_0^n \ell_{x+s}\ ds}, \tag{2.35}
$$

ahora

$$
\int_0^n \ell_{x+s}\mu_{x+s} ds = -\ell_{x+s}\ \Big|_0^n = \ell_x - \ell_{x+n} = \ {}_n d_x
$$

por lo tanto:

$$
{}_n\mu_x = \frac{{}_n d_x}{{}_n L_x}, \tag{2.36}
$$

donde

$$
{}_n L_x = \int_0^n \ \ell_{x+s}\ ds. \tag{2.37}
$$

Si $n = 1$, se tiene una función con intervalo de un año

$$
L_x = \int_0^1 \ \ell_{x+s}\ ds, \tag{2.38}
$$

y la tasa central en el intervalo $(x, x+1]$ está dada por

$$
\mu_x = \frac{d_x}{L_x}. \tag{2.39}
$$

(Ver [1], pág 70). La relación entre L_x y T_x está dada por

$$
\begin{aligned}
T_x &= \int_x^\infty \ell_y\ dy = \int_x^{x+1} \ell_y\ dy \ + \int_{x+1}^{x+2} \ell_y\ dy + ... \\
&= L_x + L_{x+1} + ...
\end{aligned}
$$

Entonces

$$T_x = \sum_{y=x}^{\infty} L_y. \tag{2.40}$$

2.2.5 El concepto de exposición

Se tiene que recordar que ${}_sp_x\mu_{x+s}$ es la función de densidad de probabilidad para la muerte a la edad $x+s$, dado que se está vivo a la edad x. Si se multiplica por ℓ_x, la cual se puede interpretar como el número de personas vivas en un grupo a la edad x, se obtiene $\ell_{x+s}\ \mu_{x+s}$, como la tasa de muertes ocurridas en el grupo a la edad exacta $x+s$. Ahora, $\ell_{x+s}\ \mu_{x+s}\ ds$ es la diferencial del número de muertes ocurridas a la edad exacta $x+s$. Entonces $s\cdot\ell_{x+s}\ \mu_{x+s}\ ds$ es el número total de años vividos después de llegar a edad x. Finalmente $\int_0^1 s\cdot\ell_{x+s}\ \mu_{x+s}\ ds$, da el número total de años vividos después de la edad x, por todos los que murieron entre la edad x y $x+1$.

Para el grupo ℓ_x, el número de sobrevivientes a la edad $x+1$ está dado por ℓ_{x+1}. Cada una de estas personas vive un año desde la edad x a la edad $x+1$, por lo que ℓ_{x+1} también representa el número total de años vividos, entre la edad x y $x+1$, por quienes sobrevivieron a edad $x+1$. Por lo que:

$$\ell_{x+1} + \int_0^1 s\cdot\ell_{x+s}\ \mu_{x+s}\ ds, \tag{2.41}$$

representa el número total de años vividos entre la edad x y $x+1$ por las ℓ_x personas, quienes constituyen el grupo a edad x.

Esta cantidad es medida en unidades de años vividos y es llamada **exposición**, debido a que es una medida de la extensión a la que el grupo está expuesto al riesgo de morir. Las personas que mueren contribuyen a la exposición total hasta el tiempo de muerte. Se puede simplificar (2.41) integrando por partes, obteniendo

$$\ell_{x+1} - s\cdot\ell_{x+s}\ \Big|_0^1 + \int_0^1 \ell_{x+s}\ ds = \int_0^1 \ell_{x+s}\ ds$$

La cual ha sido definida como L_x. Este resultado sugiere la posibilidad de una interpretación directa de L_x como concepto de exposición, donde la integral definida en (2.38) es por definición el límite de la suma de Riemann:

$$R = \sum_{i=1}^{n} \left(\ell_{x+s_i{}^*}\right)\left(\triangle\ s_i\right), \tag{2.42}$$

donde $\ell_{x+s_i{}^*}$ representa el número de personas vivas en el i-ésimo subintervalo de $(x, x+1]$. Multiplicando por $\triangle\ s_i$ (el ancho del subintervalo) da el número de años-vida vividos en el subintervalo. Donde ℓ_{x+s} está en constante cambio, R es solamente una aproximación del número preciso de años-vida vividos en $(x, x+1]$. Pero

$$L_x = \int_0^1 \ell_{x+s}\ ds = \lim_{\|\triangle\| \longrightarrow 0} R,$$

da el valor exacto del concepto de exposición, donde $\|\triangle\|$ es el más grande de los $\triangle\ s_i$.

Por extensión

$${}_nL_x = \int_0^n \ell_{x+s}\ ds,$$

da la exposición exacta en el intervalo $(x, x+n]$.

$$T_x = \sum_{y=x}^{\infty} L_y.$$

Claramente da la exposición exacta sobre el intervalo (x, ∞) y puede ser tomado como el número agregado total de años-vida vividos en el futuro por el grupo ℓ_x. Entonces $\frac{T_x}{\ell_x}$ da el **promedio de vida futura** por persona en el grupo ℓ_x.

$$\mu_x = \frac{d_x}{L_x},$$

μ_x definida como un promedio continuo de la función de riesgo $\lambda(y)$ sobre el intervalo $(x, x+1]$. La razón μ_x, es la **tasa de muerte** (tasa de mortalidad) en un intervalo de un año. Esta tasa difiere de la probabilidad q_x donde es posible que μ_x exceda uno, mientras que q_x no. Para el intervalo $(x, x+n]$ se tiene

$$_n\mu_x = \frac{_nd_x}{_nL_x}.$$

Una modificación del concepto de exposición, podría ser contada solamente en $(x, x+1]$ por los sobrevivientes en el intervalo. Existen ℓ_{x+1} sobrevivientes exactamente en un año, pero también estos están en la exposición. De manera similar ℓ_{x+2} es la exposición en $(x+1, x+2]$ y así sucesivamente por lo que se tiene

$$T_x^* = \sum_{y=x+1}^{\infty} \ell_y. \tag{2.43}$$

Se puede describir T_x^* como el **tiempo futuro de vida total** del grupo ℓ_x y que solamente los años enteros de tiempo de vida son contados. Entonces

$$e_x = \frac{T_x^*}{\ell_x}, \tag{2.44}$$

es el número promedio (o esperado) de años enteros de vida futura por persona en el grupo ℓ_x. Esto es llamado la **esperanza limitada de vida** a la edad x, es distinto de la esperanza completa denotada por $\overset{o}{e}_x$. Se puede derivar una fórmula alterna para e_x como

$$e_x = \frac{1}{\ell_x}\sum_{y=x+1}^{\infty} \ell_y = \frac{1}{\ell_x}\sum_{k=1}^{\infty} \ell_{x+k} = \sum_{k=1}^{\infty} {}_kp_x. \tag{2.45}$$

Finalmente se puede suponer que una persona vive aproximadamente un año y medio en el año de muerte, por lo que

$$\overset{o}{e}_x \approx e_x + \frac{1}{2},$$

Este resultado intuitivo se formalizará en el siguiente capítulo.

2.2.6 La relación entre $_nq_x$ y $_n\mu_x$

Con el concepto de exposición se desarrolló una fórmula muy usada que relaciona las funciones $_nq_x$ y $_n\mu_x$. Se explorará la relación con $n = 1$, donde $q_x = \frac{d_x}{\ell_x}$ y $\mu_x = \frac{d_x}{L_x}$ por lo tanto se relacionará ℓ_x y L_x.

$$\int_0^1 s \cdot \ell_{x+s}\mu_{x+s} ds = \int_0^1 \ell_{x+s} ds - \ell_{x+1} = L_x - \ell_{x+1}, \tag{2.46}$$

da el número agregado de años-vida vividos en $(x, x+1]$ por quienes mueren a la edad x en el intervalo. Se divide (2.46) entre d_x y este número promedio es necesariamente menor que uno y se podra llamar fracción promedio de $(x, x+1]$, se define esta fracción promedio de f_x como

$$f_x \; = \frac{L_x - \ell_{x+1}}{d_x}, \tag{2.47}$$

donde $d_x = \ell_x - \ell_{x+1}$ entonces $\ell_{x+1} = \ell_x - d_x$ por lo cual

$$f_x \; \cdot d_x = L_x - \ell_x + d_x,$$

$$L_x = \ell_x - (1 - f_x \;) \; d_x, \tag{2.48}$$

Cuando

$$\mu_x = \frac{d_x}{L_x} = \frac{d_x}{\ell_x - (1 - f_x \;) \; d_x} = \frac{q_x}{1 - (1 - f_x) \; q_x}. \tag{2.49}$$

Alternativamente

$$\ell_x = L_x + (1 - f_x \;) \; d_x, \tag{2.50}$$

entonces

$$q_x = \frac{d_x}{\ell_x} = \frac{d_x}{L_x + (1 - f_x \;) \; d_x} = \frac{\mu_x}{1 - (1 - f_x) \; \mu_x}, \tag{2.51}$$

como ℓ_x es no creciente en $(x, x+1]$, entonces $L_x \leq \ell_x$ como $\mu_x \geq q_x$. Generalizando, se considera el intervalo de n años $(x, x+n]$ cuando

$$\int_0^n s \cdot \ell_{x+s}\mu_{x+s} ds = \int_0^n \ell_{x+s} ds - n \cdot \ell_{x+1} = \; {}_nL_x - n \cdot \ell_{x+n}. \tag{2.52}$$

Análogamente

$$ {}_nf_x = \frac{{}_nL_x - n \cdot \ell_{x+1}}{n \cdot \ {}_nd_x}, \tag{2.53} $$

como $\ell_{x+n} = \ell_x - {}_nd_x$ se tiene

$$ n \cdot \ {}_nf_x \ \cdot \ {}_nd_x = \ {}_nL_x - n \cdot \ell_x + n \cdot \ {}_nd_x, $$

ó

$$ {}_nL_x = n \cdot \ell_x - n \ (1 -_n f_x \ ) \ {}_nd_x, \tag{2.54} $$

cuando

$$ {}_n\mu_x = \frac{{}_nd_x}{{}_nL_x} = \frac{{}_nd_x}{n \cdot \ell_x - n \ (1 - \ {}_nf_x \ ) \ {}_nd_x} = \frac{{}_nq_x}{n[1 - (1 - \ {}_nf_x) \ {}_nq_x]}. \tag{2.55} $$

(Ver [5], pág 55). Alternativamente

$$ n \cdot \ell_x =_n L_x + n \ (1 -_n f_x \ ) \ {}_nd_x, $$

ó

$$ \ell_x = \frac{1}{n} \ [{}_nL_x + n \ (1 -_n f_x \ ) \ {}_nd_x] \, , \tag{2.56} $$

entonces

$$ {}_nq_x = \frac{{}_nd_x}{\ell_x} = \frac{{}_nd_x}{\frac{1}{n} \ [{}_nL_x + n \ (1 -_n f_x \ ) \ {}_nd_x]} = \frac{{}_n\mu_x}{\frac{1}{n} \ [1 + n \ (1 -_n f_x \ ) \ {}_n\mu_x]}. \tag{2.57} $$

(Ver [5], pág 55). (2.57) nos da la relación general para obtener ${}_nq_x$ en términos de ${}_n\mu_x$. Esta fórmula es importante en la estimación de una tabla de vida de una polación de datos.

2.3 Métodos para edades no enteras

En el modelo de tabla de vida se supone que ℓ_{x+s} tiene una cierta forma matemática entre x y $x+1$. El supuesto de ℓ_{x+s} es que es diferenciable en el intervalo abierto $0 < s < 1$ pero no en $s = 0$ ó $s = 1$. La habilidad para determinar ℓ_{x+s} numéricamente para cualquier s, $0 \leq s \leq 1$, podrá ayudar a calcular probabilidades de la forma ${}_sp_x$ y ${}_{1-s}p_{x+s}$ y sus complementos ${}_sq_x$ y ${}_{1-s}q_{x+s}$.

Este supuesto entre edades no enteras es muy importante. Se podrán ver los elementos esenciales en la estimación del modelo y su uso para hacer cálculos del modelo de tabla de vida. Los tres supuestos de una forma matemática para ℓ_{x+s} son: lineal, exponencial e hiperbólico.

2.3.1 Forma lineal para ℓ_{x+s}

Cuando ℓ_{x+s} es una función lineal entre x y $x+1$ es de la forma $a + bs$, es continua y si $s = 0$, entonces $\ell_x = a$ y cuando $s = 1$ se tiene $\ell_{x+1} = a + b$, por lo tanto $b = \ell_{x+1} - a$ ó $b = \ell_{x+1} - \ell_x = -d_x$.

Por lo tanto, se tiene

$$\ell_{x+s} = \ell_x - s \cdot d_x, \tag{2.58}$$

una forma alternativa es usar $\ell_x - \ell_{x+1}$ en lugar de d_x obteniendo

$$\ell_{x+s} = \ell_x - s(\ell_x - \ell_{x+1}) = s \cdot \ell_{x+1} + (1-s)\ell_x, \tag{2.59}$$

todos los valores de ℓ_{x+s} vienen de ℓ_x y ℓ_{x+1} mediante una **interpolacion lineal**.

Se tiene que:

$${}_sp_x = \frac{\ell_{x+s}}{\ell_x} = 1 - s \cdot \frac{d_x}{\ell_x} = 1 - s \cdot q_x \tag{2.60}$$

y

$${}_sq_x = 1 -_s p_x = s \cdot q_x. \tag{2.61}$$

También se tiene que:

$${}_{1-s}p_{x+s} = \frac{\ell_{x+1}}{\ell_{x+s}} = \frac{\ell_{x+1}}{\ell_x - s \cdot d_x} = \frac{p_x}{1 - s \cdot q_x}, \tag{2.62}$$

y

$${}_{1-s}q_{x+s} = 1 -_{1-s} p_{x+s} = \frac{1 - s \cdot q_x - p_x}{1 - s \cdot q_x} = \frac{(1-s) \cdot q_x}{1 - s \cdot q_x}. \tag{2.63}$$

Después

$$\mu_{x+s} = \frac{-\frac{d}{ds}\ell_{x+s}}{\ell_{x+s}} = \frac{d_x}{\ell_x - s \cdot d_x} = \frac{q_x}{1 - s \cdot q_x}, \tag{2.64}$$

el cual multiplicado por (2.60) da un resultado muy conveniente.

$$f(s \mid X > x) = {}_s p_x \mu_{x+s} = q_x. \tag{2.65}$$

Se nota que μ_{x+s} y $f(s \mid X > x)$ no está definido en $s = 0$ y $s = 1$ donde ℓ_{x+s} no es diferenciable en estos puntos. Finalmente se tiene

$$L_x = \int_0^1 \ell_{x+s}\, ds = \int_0^1 (\ell_x - s \cdot d_x) ds = \ell_x - \frac{1}{2} d_x = \ell_{x+1} + \frac{1}{2} d_x, \tag{2.66}$$

y

$$T_x = \sum_{y=x}^{\infty} L_y = \sum_{y=x}^{\infty} (\ell_{y+1} + \frac{1}{2} d_y) = T_x^* + \frac{1}{2}\ell_x. \tag{2.67}$$

Cuando

$$m_x = \frac{d_x}{L_x} = \frac{d_x}{\ell_x - \frac{1}{2} d_x} = \frac{q_x}{1 - \frac{1}{2} \cdot q_x}, \tag{2.68}$$

y

$$\overset{o}{e}_x = \frac{T_x}{\ell_x} = \frac{T_x^* + \frac{1}{2}\ell_x}{\ell_x} = e_x + \frac{1}{2}, \tag{2.69}$$

el cual se había desarollado intuitivamente en el capítulo anterior.

La FDP para el tiempo de muerte durante el intervalo es una constante y la función de distribución acumulada(FDA) que es ${}_s q_x$, es una función lineal. Esto muestra que la variable aleatoria S tiene una distribución uniforme sobre $(x, x+1]$. ℓ_{x+s} decrece uniformemente (i.e. las personas están muriendo uniformemente) en el intervalo. Por esta razón el supuesto lineal para ℓ_{x+s} es tradicionalmente llamada **supuesto de distribución uniforme de muertes** (DMU). El supuesto DMU es muy utilizada para hacer cálculos a partir de la tabla de vida, la FDP es una constante, por está razón este supuesto es el más popular.

2.3.2 Forma Exponencial para ℓ_{x+s}

Cuando ℓ_{x+s} es una función exponencial entre x y $x+1$ es de la forma $\ell_{x+s} = a \cdot b^s$, es continua y si $s = 0$, entonces $\ell_x = a$ y cuando $s = 1$ se tiene $\ell_{x+1} = ab$ ó $b = \frac{\ell_{x+1}}{a} = \frac{\ell_{x+1}}{\ell_x}$.

Se tiene

$$\ell_{x+s} = \ell_x \left(\frac{\ell_{x+1}}{\ell_x} \right)^s = (\ell_{x+1})^s \cdot (\ell_x)^{1-s} \tag{2.70}$$

Todos los valores de ℓ_{x+s} vienen de ℓ_x y ℓ_{x+1} mediante una **interpolación exponencial**. Una forma alternativa de (2.70) es substituyendo $\ell_{x+1} = \ell_x \cdot p_x$ y se tiene

$$\ell_{x+s} = \ell_x (p_x)^s \tag{2.71}$$

Otras funciones son:

$$_s p_x = \frac{\ell_{x+s}}{\ell_x} = (p_x)^s \tag{2.72}$$

y

$$_s q_x = 1 - \ _s p_x = 1 - (p_x)^s = 1 - (1 - q_x)^s \tag{2.73}$$

También se tiene

$$_{1-s} p_{x+s} = \frac{\ell_{x+1}}{\ell_{x+s}} = \frac{\ell_{x+1}}{\ell_x (p_x)^s} = \frac{p_x}{(p_x)^s} = (p_x)^{1-s} \tag{2.74}$$

y

$$_{1-s} q_{x+s} = 1 - \ _{1-s} p_{x+s} = 1 - (p_x)^{1-s} = 1 - (1 - q_x)^{1-s} \tag{2.75}$$

Después

$$\mu_{x+s} = \frac{-\frac{d}{ds}\ell_{x+s}}{\ell_{x+s}} = \frac{-\ell_x (p_x)^s \cdot \ln p_x}{\ell_x (p_x)^s} = -\ln p_x \tag{2.76}$$

Una constante para $0 < s < 1$. Cuando ℓ_{x+s} es exponencial, la fuerza de mortalidad, μ_{x+s}, es constante en $(x, x+1)$, esta fuerza constante es llamada simplemente μ. Entonces (2.76) cambia a $\mu = -\ln p_x$ y se pueden reescribir las formulas anteriores como:

$$p_x = e^{-\mu} \tag{2.77}$$

de (2.72) a

$$_s p_x = e^{-\mu s} \tag{2.78}$$

y (2.74) se convierte en

$$_{1-s} p_{x+s} = e^{-\mu(1-s)} \tag{2.79}$$

La FDP condicional se encuentra multiplicando (2.79)y (2.76), obteniendo

$$f(s \mid X > x) = \ _s p_x \mu_{x+s} = \mu \cdot e^{-\mu s} \tag{2.80}$$

La FDP de S dada por (2.80) y su función de distribución de supervivencia (FDS) dada por (2.78), claramente muestra que S tiene una distribución exponencial en $(x, x+1]$. Como la fuerza de mortalidad es constante, este supuesto es llamada **supuesto de fuerza constante**.

Se puede ver que el supuesto exponencial no es tan conveniente como el supuesto lineal para hacer cálculos desde la tabla de vida, pero por otra parte es más conveniente para el propósito de la estimación del modelo.

Las funciones de exposición tienen la forma

$$L_x = \int_0^1 \ell_{x+s} \ ds = \int_0^1 \ell_x \left(p_x\right)^s \ ds = \frac{-\ell_x \cdot q_x}{\ln p_x} = \frac{-d_x}{-\mu} = \frac{d_x}{\mu} \tag{2.81}$$

La tasa central es

$$m_x = \frac{d_x}{L_x} = \frac{d_x}{d_x/\mu} = \mu. \tag{2.82}$$

μ es el valor constante de μ_{x+s}, con $0 < s < 1$ y se llamará μ_x. Por otra parte, en el intervalo $(x+1, x+2)$ se tiene una fuerza constante diferente, $\mu_{x+1} = -\ln p_{x+1}$, donde $L_{x+1} = \frac{d_{x+1}}{\mu_{x+1}}$. También se tiene que:

$$T_x = \sum_{y=x}^{\infty} L_y = \sum_{y=x}^{\infty} \frac{d_y}{\mu_y}. \tag{2.83}$$

Esta no es una forma conveniente, pero T_x puede ser precisamente calculado a partir de la tabla de vida bajo un supuesto exponencial. Finalmente $\overset{o}{e}_x$

puede ser obtenido como $\frac{T_x}{\ell_x}$, de nuevo esta fórmula simple no es conveniente para $\overset{o}{e}$.

2.3.3 Forma Hiperbólica para ℓ_{x+s}

Cuando ℓ_{x+s} es una funcion hiperbólica entre x y $x+1$ es de la forma $\ell_{x+s} = (a+bs)^{-1}$, es continua y si $s=0$, entonces $\ell_x = \frac{1}{a}$ ó $a = \frac{1}{\ell_x}$ y cuando $s=1$ se tiene $\ell_{x+1} = \frac{1}{a+b}$ ó $b = \frac{1}{\ell_{x+1}} - a = \frac{1}{\ell_{x+1}} - \frac{1}{\ell_x}$. Entonces

$$\ell_{x+s} = \left(\frac{1}{\ell_x} + s \left[\frac{1}{\ell_{x+1}} - \frac{1}{\ell_x} \right] \right)^{-1} \tag{2.84}$$

ó

$$\frac{1}{\ell_{x+s}} = \frac{1}{\ell_x} + s \left[\frac{1}{\ell_{x+1}} - \frac{1}{\ell_x} \right] = s \cdot \frac{1}{\ell_{x+1}} + (1-s) \cdot \frac{1}{\ell_x} \tag{2.85}$$

Donde los valores de $\frac{1}{\ell_{x+s}}$ pueden ser determinados por una interpolación lineal entre los recíprocos de ℓ_x y ℓ_{x+1}. La interpolación lineal en el recíproco de una función es llamada **interpolación armómica** en la función misma.

De (2.84)se pueden determinar otras funciones. Para obtener ${}_sp_x$, primero se necesita:

$$({}_sp_x)^{-1} = \frac{\ell_x}{\ell_{x+s}} = \ell_x \left(\frac{1}{\ell_x} + s \left[\frac{1}{\ell_{x+1}} - \frac{1}{\ell_x} \right] \right) = 1+s \left(\frac{1}{p_x} - 1 \right) = \frac{p_x + s\,(1-p_x)}{p_x}$$

Por lo tanto

$${}_sp_x = \frac{p_x}{p_x + s\,(1-p_x)} = \frac{1-q_x}{1-(1-s)\,q_x} \tag{2.86}$$

Se tiene

$${}_sq_x = 1 - {}_s\, p_x = \frac{s \cdot q_x}{1-(1-s)\,q_x} \tag{2.87}$$

También se tiene

$${}_{1-s}p_{x+s} = \frac{\ell_{x+1}}{\ell_{x+s}} = \ell_{x+1} \left(\frac{1}{\ell_x} + s \left[\frac{1}{\ell_{x+1}} - \frac{1}{\ell_x} \right] \right) = p_x + s\,(1-p_x) = 1-(1-s)\,q_x \tag{2.88}$$

y

$$ {}_{1-s}q_{x+s} = 1 - {}_{1-s}p_{x+s} = (1-s)\,q_x \tag{2.89} $$

Despues usando ${}_sp_x = \frac{p_x}{p_x+s(1-p_x)}$ se encontrará

$$ \mu_{x+s} = \frac{-\frac{d}{ds}\ {}_sp_x}{{}_sp_x} = \frac{p_x \cdot q_x}{(p_x + s \cdot q_x)^2} \div \frac{p_x}{p_x + s \cdot q_x} = \frac{q_x}{p_x + s \cdot q_x} = \frac{q_x}{1-(1-s)\,q_x} \tag{2.90} $$

y

$$ f(s \mid X > x) = {}_sp_x \mu_{x+s} = \frac{q_x \cdot (1-q_x)}{[1-(1-s)\,q_x]^2} \tag{2.91} $$

El supuesto hiperbólico es algo ilógica ya que μ_{x+s} decrece en $(x, x+1)$, mientras que en el supuesto lineal incrementa y en el supuesto exponencial es constante. Este supuesto no se usa para hacer cálculos a partir de la tabla de vida y tampoco es muy usada para la estimación fuera de la ciencia actuarial.

La relación ${}_{1-s}q_{x+s} = (1-s)\,q_x$ ha sido utilizada en las aproximaciones actuariales tradicionales para la estimación de modelos de supervivencia por más de un siglo. El supuesto es llamado **supuesto de Balducci** o distribución de Balducci porque el actuario italiano Gaetano Balducci hizo mayor uso de la distribución hiperbólica en sus escritos.

En las funciones de exposición primero debemos recordar que: $\ell_{x+s} = \frac{1}{a+bs}$ donde $a = \frac{1}{\ell_x}$ y $b = \frac{1}{\ell_{x+1}} - \frac{1}{\ell_x}$, entonces

$$ L_x = \int_0^1 \ell_{x+s}\ ds = \int_0^1 \frac{ds}{a+bs} = \frac{1}{b} \cdot \ln\left(\frac{a+b}{a}\right) $$

sustituyendo a y b se tiene

$$ L_x = \frac{\ell_{x+1}}{q_x} \cdot \ln\frac{1}{p_x} = -\ell_{x+1}\left(\frac{\ln p_x}{q_x}\right) \tag{2.92} $$

De L_x se pueden encontrar T_x, m_x y $\overset{o}{e}_x$de las fórmulas usuales.

En la siguiente tabla se muestran las diferencias entre las distribuciones

Función	**Lineal**	**Exponencial**	**Hiperbólica**
ℓ_{x+s}	$\ell_x - s \cdot d_x$ $= s \cdot \ell_{x+1} +$ $(1-s) \cdot \ell_x$	$\ell_x (p_x)^s$ $= (\ell_{x+1})^s \cdot$ $(\ell_x)^{1-s}$	$\left(\begin{array}{c} s \cdot \frac{1}{\ell_{x+1}} + \\ (1-s) \cdot \frac{1}{\ell_x} \end{array} \right)^{-1}$
${}_s p_x$	$1 - s \cdot q_x$	$(p_x)^s = e^{-\mu s}$	$\frac{p_x}{p_x + s(1-p_x)}$ $= \frac{1-q_x}{1-(1-s)q_x}$
${}_s q_x$	$s \cdot q_x$	$1 - (1-q_x)^s$	$\frac{s \cdot q_x}{1-(1-s)q_x}$
${}_{1-s} p_{x+s}$	$\frac{p_x}{1-s \cdot q_x}$	$(p_x)^{1-s} = e^{-\mu(1-s)}$	$p_x + s(1-p_x)$ $= 1 - (1-s) q_x$
${}_{1-s} q_{x+s}$	$\frac{(1-s)q_x}{1-s \cdot q_x}$	$1 - (1-q_x)^{1-s}$	$(1-s) q_x$
μ_{x+s}	$\frac{q_x}{1-s \cdot q_x}$	$\mu = -\ln p_x$	$\frac{q_x}{1-(1-s)q_x}$
${}_s p_x \mu_{x+s}$	q_x	$\mu \cdot e^{-\mu s}$	$\frac{q_x(1-q_x)}{[1-(1-s)q_x]^2}$
L_x	$\ell_x - \frac{1}{2} d_x$ $= \ell_{x+1} + \frac{1}{2} d_x$	$\frac{d_x}{\mu}$	$-\ell_{x+1} \left(\frac{\ln p_x}{q_x} \right)$
m_x	$\frac{q_x}{1-\frac{1}{2} \cdot q_x}$	μ	$\frac{(q_x)^2}{-p_x \cdot \ln p_x}$

Tabla 2.2

(Ver [1], pag 75].

En el siguiente capítulo se vera la estimación de modelos de supervivencia tabulares o no paramétricos, utilizando un estudio con datos completos ó con datos incompletos.

Capítulo 3

Modelos de supervivencia tabulares

3.1. Estimación para muestras con datos no censurados

Se empezará la estimación de modelos de supervivencia con un diseño muy simple, el cual se usa en estudios clínicos, donde el observador mantiene cada unidad de estudio bajo observación, durante un tiempo corto hasta que falle y no se permite que las unidades se retiren del estudio si no han fallado. En este caso se habla de **datos completos**. La estimación de modelos de supervivencia de datos completos (no censurados) y el análisis de tales estimaciones son considerablemente más sencillos que el caso con **datos incompletos** (censurados).

Como las situaciones de datos no censurados están presentes por lo general solamente en estudios clínicos, se usará $S(t)$ como la función de supervivencia a estimar, la variable de interés es, por tanto, el tiempo de supervivencia (de ahora en adelante, falla) desde algún evento inicial. El estimador de $S(t)$ será denotado por $\hat{S}(t)$ en la mayoría de los casos.

3.1.1 Diseño del estudio

Todas las unidades de estudio bajo observación, estarán definidas al tiempo $t = 0$, y serán observadas a lo largo del tiempo hasta que todas mueran; se documenta el tiempo exacto de cada muerte. El grupo inicial forma una cohorte, lo que significa que es cerrado, es decir, no hay entradas después del tiempo $t = 0$. El grupo podría tomar cualquiera de las siguientes dos formas:

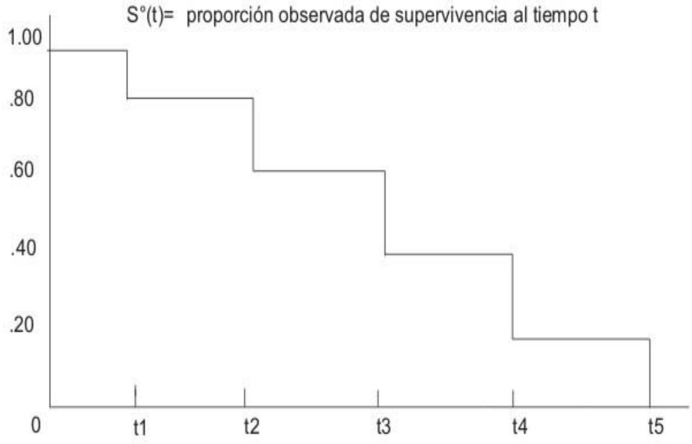
1. El tiempo $t = 0$ es obtenido el mismo día para todas las unidades de estudio y más tarde una muerte observada de duración t podría ocurrir en el $t - \acute{e}simo$ día de estudio. En este caso es fácil ver la idea de supervivencia, como el tiempo desde el evento inicial.

2. Cada persona tiene su propio valor inicial llamado $t = 0$, para después observar su propio valor t de muerte. En este caso la edad cronológica de cada persona no importa. El tiempo bajo observación (tiempo de investigación) es distinto al tiempo cronológico, ya que éste es la fecha del calendario donde empieza la investigación.

La estimación del modelo de supervivencia no dependerá si el tiempo inicial ocurre en diferentes días o en el mismo día para todas las unidades de estudio.

3.1.2 Tiempo exacto de muerte

Basado en una muestra de datos, se deberá estimar la distribución de supervivencia aplicable a las unidades de estudio. Una aproximación muy natural de $S(t)$ es la proporción observada de una muestra de supervivencia en cada punto t. Una gráfica de esta **distribución observada de superivencia,** también llamada **distribución empírica de supervivencia** se muestra a continuación.



Se usará $S^o(t)$ para denotar esta distribución. Para un conjunto inicial de tamaño n, note que:

$$S^o(t) = \begin{cases} 1 & \text{para } t < t_1 \\ \frac{n-i}{n} & \text{para } t_i < t < t_{i+1} \ i = 1, 2, \ldots, n-1 \\ 0 & \text{para } t \geq t_n \end{cases} \tag{3.1}$$

(Ver [6], pág 48)

Se espera que $S(t)$ sea una función monótona decreciente de t y el estimador $\widehat{S}(t)$ pueda mostrar esta propiedad. Se reconoce que la forma escalonada de $S^o(t)$ resulta del tamaño de muestra finita. Como resultado de incrementar el tamaño de la muestra, se pueden obtener un mayor número de etapas (o escalones) y así se tendrán valores estimados de $S(t)$ "más suaves" que aquéllos obtenidos de muestras más pequeñas. Es posible sobreponer a la gráfica de $S^o(t)$ una curva suave que ajuste el esquema escalonado. A este procedimiento se le llama **graduación**.

En el procedimiento de estimación descrito, se supone que el tiempo de muerte es documentado con suficiente exactitud, pero esto no es más que una muerte en cada punto del tiempo. Si las unidades de medición (es decir, semanas, dias, etc) no son suficientemente finas, ó la muestra es muy grande, ó ambas , se tienen **muertes múltiples** que pueden ocurrir en un mismo punto. Cuando esto ocurre los escalones de la gráfica de $S^o(t)$ se reducen en $\frac{r}{n}$, donde r es el número de muertes en el mismo punto.

Ánalisis de la distribución empírica de supervivencia

$S^o(t)$ es el estimador de $S(t)$, cada estimación depende de la aleatoriedad de la ocurrencia de las muertes previas al tiempo t. Para cualquier valor fijo de t se define $S^o(t)$ como:

$$S^o(t) = \frac{\text{número de sobrevivientes al tiempo t}}{n} = \frac{N_t}{n} \tag{3.2}$$

(Ver [4], pág 9)

Donde N_t denota el número (aleatorio) de sobrevivientes al tiempo t. (3.2) es el estimador de $S(t)$, mientras que $S(t)$ representa la probabilidad de supervivencia (real u operativa) al tiempo t para al individuo en estudio. Cuando N_t es una variable aleatoria binomial con parámetros n (tamaño de la muestra) y $S(t)$, entonces:

$$E\left[N_t\right] = n \cdot S(t) \tag{3.3}$$

y

$$Var\left[N_t\right] = n \cdot S(t) \cdot F(t) \tag{3.4}$$

donde $F(t) = 1 - S(t)$.

En este caso $S^o(t) = \frac{N_t}{n}$ es una proporción de variable aleatoria binomial con:

$$E\left[S^o(t)\right] = \frac{1}{n} \cdot E\left[N_t\right] = S(t) \tag{3.5}$$

y

$$Var\left[S^o(t)\right] = \frac{1}{n^2} \cdot Var\left[N_t\right] = \frac{S(t) \cdot F(t)}{n}. \tag{3.6}$$

(3.5) muestra que $S^o(t)$ es un **estimador insesgado** de $S(t)$.

Como $S(t)$ no es conocido, se tienen valores númericos para $Var\left[S^o(t)\right]$ obtenidas por aproximaciones o estimaciones, usando $S^o(t)$ en lugar de $S(t)$.

$$\widehat{Var}\left[S^o(t)\right] = \frac{S^o(t) \cdot F^o(t)}{n} \tag{3.7}$$

donde $F^o(t) = 1 - S^o(t)$.

3.1.3 Tiempos de muerte agrupados

Si un grupo bajo observación es muy grande, se usarán aproximaciones de grupos de muertes. Se empezará dividiendo el tiempo continuo en un número indefinido de intervalos de igual tamaño.

Notación

Como se ha dicho antes n es el tamaño del conjunto bajo estudio, todas las unidades de n son observadas hasta morir, cada una de las n muertes está en uno de los k intervalos. Sea D_t la variable aleatoria para el número de muertes que ocurren en el $(t+1)-\acute{e}simo$ intervalo (es decir, entre t y $t+1$), $0 \leq t \leq k-1$ y d_t representa la realización de D_t en el estudio actual. Se puede ver que:

$$n = \sum_{t=0}^{k-1} d_t. \tag{3.8}$$

(Ver [5], pág 86)

Sea N_t la variable aleatoria del número de sobrevivientes al tiempo t y n_t la realización de N_t en el estudio actual. Necesariamente $N_0 = n$ y $N_k = 0$ y la relación es:

$$N_{t+1} = N_t - D_t \qquad\qquad para\ t = 0, 1, \ldots, k-1. \tag{3.9}$$

En un estudio con datos no censurados y muertes agrupadas, el objetivo es estimar $S(t)$ mediante otras funciones del modelo de supervivencia y analizar los estimadores obtenidos (dependientes de las variables aleatorias).

Las distribuciones de N_t y D_t

Sea N_t una variable aleatoria binomial con parámetros n y $S(t)$. Como se vió $E\left(N_t\right) = n \cdot S\left(t\right)$ y $Var\left(N_t\right) = n \cdot S\left(t\right) \cdot F\left(t\right)$. El conjunto de variables aleatorias $D_0, D_1, \ldots, D_{k-1}$ tiene una distribución multinomial con $P\left(D_t = d_t\right) = S\left(t\right) \cdot q_t = t \mid q_0$. Así,

$$P_n\left(d_0, d_1, \ldots, d_{k-1}\right) = \frac{n!}{d_0! \cdot d_1! \cdots d_{k-1}!} \prod_{t=0}^{k-1} \left({}_{t|}\, q_o\right)^{d_t} \tag{3.10}$$

$$E\left[D_t\right] = n \cdot {}_{t|}\, q_o \tag{3.11}$$

$$Var\left(D_t\right) = n \cdot {}_{t|}\, q_o \cdot \left(1 - {}_{t|}\, q_o\right) \tag{3.12}$$

$$Cov\left(D_r, D_t\right) = -n\left({}_{r\,|}\,q_o\right)\left({}_{t\,|}\,q_o\right) \tag{3.13}$$

La distribución, la media, la varianza y covarianza son todas incondicionales, es decir, están dadas solamente por las n unidades de estudio al tiempo $t = 0$. Si se condiciona en $N_t = n_t$ sobrevivientes al tiempo t, D_t es una variable aleatoria binomial, con media $n_t \cdot q_t$ y varianza $n_t \cdot q_t \cdot p_t$.

Estimación de $S(t)$ y ${}_{t\,|}\,q_0$

$S(t)$ puede ser estimada usando la proporción observada de supervivencia al tiempo t, esto es

$$\widehat{S(t)} = \frac{N_t}{n}, \tag{3.14}$$

es una variable aleatoria de proporción binomial con media y varianza dada en (3.5) y (3.6), respectivamente, con $\widehat{S}(t)$ reemplazando a $S^o(t)$. De manera similar, ${}_{t} \mid q_0$ se estima usando las frecuencias relativas de muerte observadas entre t y $t + 1$. Esto es:

$$\widehat{{}_{t\,|}\,q_0} = \frac{d_t}{n} \qquad\qquad t = 0, 1, \ldots, k-1, \tag{3.15}$$

(Ver [5], pág 87), es una proporción multinomial, con media

$$E\left[\widehat{{}_{t\,|}\,q_0}\right] = \frac{1}{n} \cdot E\left[D_t\right] = {}_{t\,|}\,q_o, \tag{3.16}$$

y varianza

$$Var\left(\widehat{{}_{t\,|}\,q_0}\right) = \frac{1}{n^2} \cdot Var\left(D_t\right) = \frac{{}_{t\,|}\,q_0 \cdot \left(1 - {}_{t\,|}\,q_0\right)}{n}, \tag{3.17}$$

se observa que el estimador de proporción multinomial $\widehat{t \mid q_0} = \frac{d_t}{n}$ es insesgado, debido a (3.16).

Estimación de q_t y p_t

Sea q_t la probabilidad condicional de muerte entre t y $t + 1$, dado que se está vivo al tiempo t y $p_t = 1 - q_t$ la probabilidad de supervivencia. El estimador de q_t está dado por:

$$\widehat{q_t} = \frac{d_t}{n_t}, \tag{3.18}$$

(Ver [5], pág 87), si se condiciona en n_t sobrevivientes al tiempo t, D_t es binomial, por lo que (3.18) es una proporción binomial con media

$$E[\widehat{q}_t \mid n_t] = \frac{1}{n_t} \cdot E[D_t] = q_t, \tag{3.19}$$

y varianza

$$Var(\widehat{q}_t \mid n_t) = \frac{1}{n_t^2} \cdot Var(D_t) = \frac{q_t(1-q_t)}{n_t}, \tag{3.20}$$

de igual forma

$$\widehat{p}_t = 1 - \widehat{q}_t = \frac{n_t - D_t}{n_t} = \frac{N_{t+1}}{n_t}, \tag{3.21}$$

es también una proporción binomial con media

$$E[\widehat{p}_t \mid n_t] = p_t, \tag{3.22}$$

y varianza

$$Var(\widehat{p}_t \mid n_t) = \frac{p_t(1-p_t)}{n_t}, \tag{3.23}$$

donde $q_t = 1 - p_t$, entonces (3.20) y (3.23)son iguales.

Estimación de $S(t)$ por $\{\widehat{p}_i\}$

Una estimación de $S(t)$ alterna a (3.14) puede ser considerando la relación:

$$S(t) = p_0 \cdot p_1 \cdots p_{t-1}, \tag{3.24}$$

(Ver [5], pág 89), donde cada p_i es una probabilidad condicional, aunque el producto $S(t)$, es no condicional.

De acuerdo a (3.24), se obtiene el estimador:

$$\widehat{S}(t) = \widehat{p}_0 \cdot \widehat{p}_1 \cdots \widehat{p}_{t-1}, \tag{3.25}$$

por lo tanto $\widehat{S}(t)$, es un estimador de forma directa (incondicional) si es de la forma (3.14) o de forma indirecta (condicional) si es de la forma (3.25), pero con el mismo valor numérico $\widehat{S}(t)$. En estudios con datos no censurados se puede utilizar la aproximación (3.14), mientras que en estudios con datos censurados se utilizará la aproximación (3.25).

Los estimadores $\widehat{p}_t$, $\widehat{q}_t$, $t \mid \widehat{p}_0$, los cuales tienen proporciones binomiales o multinomiales, y la aproximación incondicional (3.14), son **estimadores de máxima verosimilitud**, con propiedades de insesgamiento, consistencia y eficiencia. La aproximación condicional (3.25), es también el producto de un estimador máximo verosimil $\widehat{p}_0$. $\widehat{S}(t)$ es también un estimador máximo verosimil y es insesgado.

Estimación de la función de riesgo

Se considerará la función de riesgo de T, donde $t^* = t + \frac{1}{2}$ representa el punto intermedio del intervalo $(t, t+1]$ (la tasa de riesgo está dada en un punto del tiempo, no en un intervalo). Para expresar la tasa de riesgo $\lambda(t^*)$ en términos de p_t (o q_t) se necesita un supuesto de distribución como las que se describen en la sección 2.3.

A continuación se muestra un resumen de los estimadores para las distintas funciones estudiadas y sus características.

Función	**Estimador**	**I. o C.**	**Sesgo**	**Varianza**
$t \mid q_0$	$\frac{D_t}{n}$	Incondicional	Insesgado	$\frac{t\mid q_0(1-t\mid q_0)}{n}$
$S(t)$	$\frac{N_t}{n}$	Incondicional	Insesgado	$\frac{S(t)\cdot F(t)}{n}$
p_t	$\frac{N_{t+1}}{n}$	Condicional en n_t	Insesgado	$\frac{p_t \cdot q_t}{n_t}$
q_t	$\frac{D_t}{n}$	Condicional en n_t	Insesgado	$\frac{p_t \cdot q_t}{n_t}$

3.2 Diseño del estudio para el modelo con datos censurados

Uno de los propósitos de esta sección es introducir al lector al problema de estimar q_x en el intervalo $(x, x+1]$, dado que se está con vida a la edad x, para así llegar a un estimador de $S(x)$.

También se revisará la teoría básica de los decrementos dobles con el objetivo de familiarizarse con las relaciones y la notación, antes de abordar el problema de estimación.

3.2.1 Introducción

En muchas situaciones de estimación es común encontrar personas bajo observación, en donde no todas son retenidas hasta la muerte, como en las aplicaciones actuariales de seguros y pensiones. En algunas situaciones se tienen uno o más eventos aleatorios (adicionales a la muerte), donde cada uno de los miembros de la muestra son susceptibles a ser eliminados, por ejemplo la **incapacidad**, el **retiro** o la **terminación voluntaria**. En general, se unirán todos los decrementos aleatorios (diferentes de la muerte) y se les llamará **salidas** del estudio. Es importante señalar que una retirada es un evento aleatorio en el mismo sentido que la muerte.

Cuando el periodo de observación termina, muchos miembros permanecen con vida en la muestra y son llamados **finalistas**. Es importante ver que las salidas son eventos aleatorios, mientras que los finalistas no lo son. Si se tiene la fecha en la cual el periodo de observación termina, se puede establecer la edad máxima de un miembro particular y cada miembro tiene una **edad final programada**.

La información básica que se requiere en cada miembro de la muestra para aproximar estimaciones, consiste en:

1. Fechas del periodo de observación.
2. Fecha de unión al grupo bajo observación.
3. Fecha de cumpleaños.
4. Fecha de muerte o retirada, la que ocurra.

De los primeros 3, cada miembro de la muestra tiene una edad de comienzo bajo observación y una edad programada puede ser establecida a la salida de la observación. Para el $i-ésimo$ miembro de la muestra, se refiere a estas dos edades como y_i y z_i respectivamente. Necesariamente $y_i < z_i$ y se presenta en pares ordenados (y_i, z_i) para el $i-ésimo$ miembro. Si la muerte o retirada ocurre en el periodo de observación (es decir antes de la edad z_i) , entonces la observación termina antes de la edad final programada. En ningún caso la observación puede continuar después de la edad final programada. Un problema que se presenta son las fracciones de años, por lo que se puede dividir el año en dias, para facilitar el cálculo a lo largo del estudio.

3.2.2 Contribución de un individuo al intervalo $(x, x+1]$

La aproximación para estimar $S(x)$, o su equivalente ℓ_x, se usará en la presencia de datos censurados, pero primero se deberá estimar p_x (probabilidad condicional de supervivencia a la edad $x+1$, dado que está vivo a la edad x) ya que tiene la forma:

$$\widehat{S}(x) = \widehat{p}_0 \cdot \widehat{p}_1 \cdot \cdots \cdot \widehat{p}_{x-1}, \tag{3.26}$$

donde también se puede estimar q_x de p_x como

$$\widehat{p}_x = 1 - \widehat{q}_x. \tag{3.27}$$

La notación $(x, x+1]$ es usada para indicar la muertes a la edad $x+1$ en este intervalo. Para estimar q_x, se necesita conocer qué contribución tuvo cada miembro de la muestra en el intervalo $(x, x+1]$, en otras palabras, se necesita conocer la **experiencia de supervivencia** de cada miembro entre la edad x y $x+1$. Para obtener esta experiencia se necesita ver el par ordenado (y_i, z_i) para la persona i relacionada al intervalo $(x, x+1]$. Por lo que se tienen los siguientes casos:

1. Si $z_i \leq x$, entonces la persona i no contribuye al intervalo $(x, x+1]$.
2. Si $y_i \geq x+1$, de manera similar la persona i no contribuye al intervalo.
3. Si $y_i \leq x$ y $z_i \geq x+1$, entonces la persona i está bajo observación en el intervalo entero $(x, x+1]$.
4. Finalmente, es posible que $x < y_i < x+1$ ó $x < z_i < x+1$ ó ambos.

Descartando los dos primeros casos, ya que la persona i no contribuye al intervalo, se dice que la persona i entra en $(x, x+1]$ a la edad $x+r_i$ y está programado a salir de $(x, x+1]$ a edad $x+s_i$. Si $y_i \leq x$ entonces la persona i entra en $(x, x+1]$ a la edad exacta x, así $r_i = 0$ y si $x < y_i < x+1$, entonces entra a la edad $x+r_i$, con $0 < r_i < 1$. Si $z_i \geq x+1$ entonces la persona i es programada a salir de $(x, x+1]$ a la edad exacta $x+1$, por lo que $s_i = 1$ y si $x < z_i < x+1$, entonces la salida programada es a la edad $x+s_i$ con $0 < s_i < 1$. Por lo que los intervalos serán $0 \leq r_i < 1$ y $0 < s_i \leq 1$.

Las relaciones entre la edad de entrada con la edad programada de salida del estudio (y_i, z_i) y la edad de entrada con la edad programada de salida para el intervalo estimado $(x+r_i, x+s_i)$ son muy importantes. Para estimar q_x para el intervalo $(x, x+1]$ la información necesaria está dada por $(x+r_i, x+s_i)$ o simplemente por (r_i, s_i).

Se debe notar que la persona i no puede contribuir al intervalo $(x, x+1]$ cuando $x < z_i$, ya que puede morir o retirarse antes que x y para esta persona no está definida ni r_i, ni s_i en el intervalo $(x, x+1]$.

Para los pares ordenados (r_i, s_i), que satisfacen las condiciones $0 \leq r_i < 1$ y $0 < s_i \leq 1$, en el intervalo $(x, x+1]$, se tienen varios casos especiales en los cuales las estimaciones son simplificadas.

Caso especial A

Si para un cierto intervalo $(x, x+1]$, $r_i = 0$ y $s_i = 1$ para toda i, de tal manera que todas las personas entran en el intervalo estimado al principio y son programadas a permanecer bajo observación a edad $x+1$, por lo que es sencillo estimar q_x.

Caso especial B

Si para un cierto intervalo $(x, x+1]$, $s_i = 1$ para toda i, pero en general $0 \leq r_i < 1$, las personas pueden entrar en $(x, x+1]$ a cualquier edad, pero todas son programadas a salir al final del intervalo a edad $x+1$.

Caso especial C

Si para un cierto intervalo $(x, x+1]$, $r_i = 0$ para toda i, pero en general $0 < s_i \leq 1$, todas las personas entran en $(x, x+1]$ a edad x, pero las salidas están distribuidas en el intervalo.

Los casos especiales pueden ser obtenidos como resultado del periodo de observación. En las secciones 3.3 y 3.4 se revisarán estos casos especiales en el proceso de la estimación de q_x.

3.2.3 Medios de decremento simple y doble

La persona $i-\acute{e}sima$ entra al intervalo estimado $(x, x+1]$ a la edad $x+r_i$, $0 \leq r_i < 1$, y está programada a salir del intervalo a la edad $x+s_i$, $0 < s_i \leq 1$, por lo que la probabilidad de supervivencia a edad $x+s$, para esta persona es ${}_{s-r}p_{x+r}$, y la probabilidad de muerte en este intervalo es:

$$ {}_{s-r}q_{x+r} = 1 - {}_{s-r}p_{x+r}. \tag{3.28} $$

Estas probabilidades están condicionadas a que llegue con vida a edad $x+r$. La función de densidad de muerte a edad $x+t$, dado que está vivo a edad $x+r$, está dada por:

$$ {}_{t-r}p_{x+r}\ \mu_{x+t}. \tag{3.29} $$

Esta notación es adecuada solamente si el estudio está en un medio donde la muerte es el único evento aleatorio (se hablará de un **medio de decremento simple**). Si la muerte y la retirada son eventos aleatorios que afectan el estudio se hablará de un **medio de decremento doble**, donde los cálculos para la estimación son más complejos.

Se empezará con la función básica ${}_tp_x$, la cual representa la probabilidad de no morir antes de la edad $x+t$, dado que se está vivo a edad x en un medio de decremento simple. Cuando se encuentra en un medio de decremento doble, es necesario especificar el evento aleatorio, donde ${}_tp_x$ representa la probabilidad de no ocurrencia antes de la edad $x+t$. Por lo que se usará ${}_tp_x'^{(d)}$ como la probabilidad de que la muerte no ocurra antes de la edad $x+t$ y ${}_tp_x'^{(w)}$ como la probabilidad de no retirada antes de la edad $x+t$, donde cada una de estas probabilidades están en un medio de decremento simple.

Esta ${}_tp_x'^{(d)}$ representa exactamente el mismo concepto que ya se había manejado antes como ${}_tp_x$. Similarmente se usará $q_x'^{(d)}$ como la medida de probabilidad de muerte en $(x, x+1]$ y representa el mismo concepto que ya se había manejado antes como q_x. Se asocia ${}_tp_x'^{(d)}$ y $q_x'^{(d)}$ con la **fuerza de mortalidad** $\mu_{x+t}^{(d)}$ (previamente denotado por μ_{x+t}) para tener la relación:

$$
{}_tp_x'^{(d)} = \exp\left[-\int_0^t \mu_{x+t}^{(d)}\, du\right] \tag{3.30}
$$

y

$$
q_x'^{(d)} = 1 - \exp\left[-\int_0^1 \mu_{x+t}^{(d)}\, du\right]. \tag{3.31}
$$

Ahora como también la retirada es un evento aleatorio, existe una **fuerza de retirada**, $\mu_{x+t}^{(w)}$ a la edad $x+t$ y la relación es:

$$
{}_tp_x'^{(w)} = \exp\left[-\int_0^t \mu_{x+t}^{(w)}\, du\right] \tag{3.32}
$$

(Ver [5], pág 105), y

$$
q_x'^{(w)} = 1 - \exp\left[-\int_0^1 \mu_{x+t}^{(w)}\, du\right]. \tag{3.33}
$$

Si los eventos aleatorios de muerte y retirada son independientes, entonces

$$ {}_tp_x^{(\tau)} = {}_tp_x^{\prime(d)} \cdot {}_tp_x^{\prime(w)} \tag{3.34} $$

(Ver [5], pág 105), es la probabilidad de no muerte y no retirada antes de la edad $x+t$, dado que está vivo y no se ha retirado a la edad x. Se supondrá la independencia de estos eventos para la realización de los cálculos posteriores.

La función de densidad de muerte a la edad $x+t$, dado que se está vivo a la edad x, es

$$ {}_tp_x^{(\tau)} \mu_{x+t}^{(d)}, \tag{3.35} $$

(Ver [5], pág 105), la función de densidad de retirada es

$$ {}_tp_x^{(\tau)} \mu_{x+t}^{(w)}. \tag{3.36} $$

Las relaciones más generales son

$$ {}_{s-r}q_{x+r}^{\prime(d)} = 1 - {}_{s-r}\, p_{x+r}^{\prime(d)} = 1 - \exp\left[-\int_r^s \mu_{x+r}^{(d)}\ du \right], \tag{3.37} $$

$$ {}_{s-r}q_{x+r}^{\prime(w)} = 1 - {}_{s-r}\, p_{x+r}^{\prime(w)} = 1 - \exp\left[-\int_r^s \mu_{x+r}^{(w)}\ du \right] \tag{3.38} $$

y

$$ {}_{t-r}p_{x+r}^{(\tau)} = {}_{t-r}p_{x+r}^{\prime(d)} \cdot {}_{t-r}p_{x+r}^{\prime(w)} \tag{3.39} $$

y la función de densidad de muerte es

$$ {}_{t-r}p_{x+r}^{(\tau)} \mu_{x+t}^{(d)}, \tag{3.40} $$

mientras que la función de densidad de retirada está dada por:

$$ {}_{t-r}p_{x+r}^{(\tau)} \mu_{x+t}^{(w)}. \tag{3.41} $$

Sea $q_x^{(d)}$ y $q_x^{(w)}$ las probabilidades de muerte y retirada antes de la edad $x+1$, en presencia de otro decremento (es decir, en un medio con doble decremento). Es de notar que estas probabilidades son distintas al concepto de $q_x^{\prime(d)}$ y $q_x^{\prime(w)}$. Esto es,

$$q_x^{(d)} = \int_0^1 {}_up_x^{(\tau)}\mu_{x+u}^{(d)}\, du = \int_0^1 \left[1 - {}_uq_x'^{(d)}\right]\left[1 - {}_uq_x'^{(w)}\right]\mu_{x+u}^{(d)}\, du \tag{3.42}$$

y

$$q_x^{(w)} = \int_0^1 {}_up_x^{(\tau)}\mu_{x+u}^{(w)}\, du = \int_0^1 \left[1 - {}_uq_x'^{(d)}\right]\left[1 - {}_uq_x'^{(w)}\right]\mu_{x+u}^{(w)}\, du. \tag{3.43}$$

La probabilidad de supervivencia de la edad x a la edad $x+1$ es

$$p_x^{(\tau)} = 1 - q_x^{(\tau)} = 1 - q_x^{(d)} - q_x^{(w)}. \tag{3.44}$$

De forma más general ${}_{s-r}q_{x+r}^{(d)}$ y ${}_{s-r}q_{x+r}^{(w)}$ representan la probabilidad de muerte y retirada antes de la edad $x+s$ dado que se tiene la edad $x+r$, en presencia de otro decremento. Por lo que se tiene que

$${}_{s-r}q_{x+r}^{(d)} = \int_r^s {}_{u-r}p_{x+r}^{(\tau)}\mu_{x+u}^{(d)}\, du = \int_r^s \left[1 - {}_{u-r}q_{x+r}'^{(d)}\right]\left[1 - {}_{u-r}q_{x+r}'^{(w)}\right]\mu_{x+u}^{(d)}\, du, \tag{3.45}$$

$${}_{s-r}q_{x+r}^{(w)} = \int_r^s {}_{u-r}p_{x+r}^{(\tau)}\mu_{x+u}^{(w)}\, du = \int_r^s \left[1 - {}_{u-r}q_{x+r}'^{(d)}\right]\left[1 - {}_{u-r}q_{x+r}'^{(w)}\right]\mu_{x+u}^{(w)}\, du. \tag{3.46}$$

y

$${}_{s-r}p_{x+r}^{(\tau)} = 1 - {}_{s-r}q_{x+r}^{(\tau)} = 1 - {}_{s-r}q_{x+r}^{(d)} - {}_{s-r}q_{x+r}^{(w)}. \tag{3.47}$$

Las relaciones entre los dobles decrementos de probabilidad $q_x^{(d)}$, $q_x^{(w)}$ y los correspondientes al decremento simple $q_x'^{(d)}$, $q_x'^{(w)}$ se analizan bajo el supuesto de distribuciones uniforme y exponencial.

Distribución uniforme

En cada decremento, muerte y retirada, se tiene distribuciones uniformes en el intervalo $(x, x+1]$, entonces

$$_{s}q_x^{\prime(d)} = s \cdot q_x^{\prime(d)} \tag{3.48}$$

y

$$_{s}q_x^{\prime(w)} = s \cdot q_x^{\prime(w)}. \tag{3.49}$$

De (2.64) se tiene

$$\mu_{x+s}^{(d)} = \frac{q_x^{\prime(d)}}{1 - s \cdot q_x^{\prime(d)}} \tag{3.50}$$

y

$$\mu_{x+s}^{(w)} = \frac{q_x^{\prime(w)}}{1 - s \cdot q_x^{\prime(w)}}. \tag{3.51}$$

Para encontrar $q_x^{(d)}$ en términos de $q_x^{\prime(d)}$ y $q_x^{\prime(w)}$ bajo el supuesto uniforme, se sustituye $(3.48)\,, (3.49)$ y (3.50) en (3.42), se obtiene

$$\begin{aligned} q_x^{(d)} &= \int_0^1 \left[1 - u \cdot\ q_x^{\prime(d)}\right] \left[1 - u \cdot q_x^{\prime(w)}\right] \frac{q_x^{\prime(d)}}{1 - u \cdot q_x^{\prime(d)}}\, du \\ &= q_x^{\prime(d)} \int_0^1 \left[1 - u \cdot q_x^{\prime(w)}\right] du \\ &= q_x^{\prime(d)} \left[1 - \frac{1}{2} \cdot q_x^{\prime(w)}\right]. \end{aligned} \tag{3.52}$$

(Ver [5], pág 107).

Similarmente, usando (3.43) se obtiene

$$q_x^{(w)} = q_x^{\prime(w)} \left[1 - \frac{1}{2} \cdot q_x^{\prime(d)}\right]. \tag{3.53}$$

Para encontrar $q_x^{\prime(d)}$ en términos de $q_x^{(d)}$ y $q_x^{(w)}$, se necesita resolver un par de ecuaciones cuadráticas dadas por (3.52) y (3.53). De (3.52) se tiene

$$q_x'^{(d)} = \frac{q_x^{(d)}}{1 - \frac{1}{2} \cdot q_x'^{(w)}}, \tag{3.54}$$

y de (3.53) se tiene

$$q_x'^{(w)} = \frac{q_x^{(w)}}{1 - \frac{1}{2} \cdot q_x'^{(d)}}. \tag{3.55}$$

Sustituyendo (3.55) en (3.54) se obtiene la ecuación cuadrática

$$\frac{1}{2}\left(q_x'^{(d)}\right)^2 - \left(1 - \frac{1}{2}q_x^{(w)} + \frac{1}{2}q_x^{(d)}\right) q_x'^{(d)} + q_x^{(d)} = 0,$$

la cual se resuelve por

$$q_x'^{(d)} = b - \sqrt{b^2 - 2 \cdot q_x^{(d)}}, \tag{3.56}$$

donde $b = 1 - \frac{1}{2}q_x^{(w)} + \frac{1}{2}q_x^{(d)}$. Por simetría, se tiene

$$q_x'^{(w)} = b - \sqrt{b^2 - 2 \cdot q_x^{(w)}}, \tag{3.57}$$

donde $b = 1 - \frac{1}{2}q_x^{(d)} + \frac{1}{2}q_x^{(w)}$. Se debe notar que en ambas ecuaciones se toma el radical negativo, ya que el positivo llevaría a $q_x'^{(d)} > 1$ ó $q_x'^{(w)} > 1$.

Distribución exponencial

Si tanto las muertes como las salidas se suponen que tienen distribuciones exponenciales (fuerzas constantes) en el intervalo $(x, x+1]$, entonces

$${}_sp_x'^{(d)} = e^{-s \cdot \mu^{(d)}} \tag{3.58}$$

y

$${}_sp_x'^{(w)} = e^{-s \cdot \mu^{(w)}}, \tag{3.59}$$

de acuerdo a lo establecido en (2.78). Por la propiedad de la fuerza constante se tiene

$$\mu_{x+s}^{(d)} = \mu^{(d)} \tag{3.60}$$

y

$$\mu_{x+s}^{(w)} = \mu^{(w)}. \tag{3.61}$$

Para encontrar $q_x^{(d)}$ en términos de $\mu^{(d)}$ y $\mu^{(w)}$ bajo el supuesto exponencial, se substituye $(3.58), (3.59)$ y (3.60) en (3.42), para obtener

$$\begin{aligned}
q_x^{(d)} &= \int_0^1 {}_u p_x'^{(d)} \cdot {}_u p_x'^{(w)} \cdot \mu^{(d)} du \\
&= \mu^{(d)} \int_0^1 e^{-u\left(\mu^{(d)}+\mu^{(w)}\right)} du \\
&= \frac{\mu^{(d)}}{\mu^{(d)}+\mu^{(w)}} \left[1 - e^{-\left(\mu^{(d)}+\mu^{(w)}\right)}\right], \qquad (3.62)
\end{aligned}$$

Similarmente, de (3.43) se obtiene

$$q_x^{(w)} = \frac{\mu^{(w)}}{\mu^{(d)}+\mu^{(w)}} \left[1 - e^{-\left(\mu^{(d)}+\mu^{(w)}\right)}\right]. \tag{3.63}$$

Para encontrar $q_x'^{(d)}$ en términos de $q_x^{(d)}$ y $q_x^{(w)}$, se necesita resolver un par de ecuaciones dadas por (3.62) y (3.63). De (3.62) se tiene

$$\frac{1 - e^{-u\left(\mu^{(d)}+\mu^{(w)}\right)}}{\mu^{(d)}+\mu^{(w)}} = \frac{q_x^{(d)}}{\mu^{(d)}}, \tag{3.64}$$

y de (3.63) se obtiene

$$\frac{1 - e^{-u\left(\mu^{(d)}+\mu^{(w)}\right)}}{\mu^{(d)}+\mu^{(w)}} = \frac{q_x^{(w)}}{\mu^{(w)}}, \tag{3.65}$$

de (3.64) y (3.65) se obtiene

$$\mu^{(w)} = \frac{q_x^{(w)}}{q_x^{(d)}} \cdot \mu^{(d)},$$

entonces

$$\mu^{(d)} + \mu^{(w)} = \mu^{(d)} + \frac{q_x^{(w)}}{q_x^{(d)}} \cdot \mu^{(d)} = \mu^{(d)} \left(\frac{q_x^{(d)} + q_x^{(w)}}{q_x^{(d)}} \right). \tag{3.66}$$

Entonces se sustituye (3.66) en (3.62) y se tiene

$$\frac{q_x^{(d)} + q_x^{(w)}}{q_x^{(d)}} \left[1 - \exp\left[-\mu^{(d)} \left(\frac{q_x^{(d)} + q_x^{(w)}}{q_x^{(d)}} \right) \right] \right] = q_x^{(d)}$$

ó

$$\exp\left[-\mu^{(d)} \left(\frac{q_x^{(d)} + q_x^{(w)}}{q_x^{(d)}} \right) \right] = 1 - q_x^{(d)} - q_x^{(w)},$$

por lo tanto

$$\mu^{(d)} = -\ln \left(1 - q_x^{(d)} - q_x^{(w)} \right)^{q_x^{(d)} / \left(q_x^{(d)} + q_x^{(w)} \right)}. \tag{3.67}$$

Finalmente, se tiene

$$q_x^{\prime(d)} = 1 - p_x^{\prime(d)} = 1 - e^{-\mu^{(d)}} = 1 - \left(1 - q_x^{(d)} - q_x^{(w)} \right)^{q_x^{(d)} / \left(q_x^{(d)} + q_x^{(w)} \right)}, \tag{3.68}$$

lo cual se puede escribir como

$$q_x^{\prime(d)} = 1 - \left(p_x^{(\tau)} \right)^{q_x^{(d)} / \left(q_x^{(d)} + q_x^{(w)} \right)}, \tag{3.69}$$

(Ver [5], pág 110), y por simetría

$$q_x^{\prime(w)} = 1 - \left(p_x^{(\tau)} \right)^{q_x^{(w)} / \left(q_x^{(d)} + q_x^{(w)} \right)}. \tag{3.70}$$

3.3 Procedimientos de momentos para datos censurados

El fundamento para estimar por procedimientos de momentos, es el principio estadístico de que el número de muertes observadas en $(x, x+1]$ es el mismo que el de una cierta muestra. Se supone que el número actual de muertes en el intervalo, denotado por d_x, está disponible, entonces el problema de estimación consiste en dos pasos:

1. Encontrar una expresión para el número esperado de muertes en el intervalo $(x, x+1]$;

2. Resolver la ecuación de momentos para el estimador deseado.

Generalmente se requiere un supuesto de distribución, para resolver la ecuación de momentos.

3.3.1 Estimación de momentos en un medio de decremento simple

Los datos básicos como el periodo de observación, las fechas de cumpleaños, las fechas en que el miembro se une al grupo bajo observación y las fechas de muerte, al ser analizadas en conjunto producen los pares ordenados $(x + r_i,$ $x + s_i)$, donde la persona $i - ésima$ tiene una contribución en $(x, x+1]$.

Para la persona i, la probabilidad de muerte durante la observación en un medio con decremento simple, es la probabilidad de muerte antes de la edad $x + s_i$, dado que llegó con vida a la edad $x + r_i$, lo cual es ${}_{s_i - r_i}q_{x+r_i}$, y es también el número de muertes esperado de una muestra de tamaño uno. Donde el número de muertes es uno (con probabilidad ${}_{s_i - r_i}q_{x+r_i}$), o cero (con probabilidad ${}_{s_i - r_i}p_{x+r_i}$). Entonces el número esperado es

$$1 \cdot {}_{s_i - r_i}\, q_{x+r_i} + 0 \cdot \; {}_{s_i - r_i}p_{x+r_i} = \; {}_{s_i - r_i}q_{x+r_i}. \tag{3.71}$$

La relación básica de momentos

Si n_x es el número total de personas que contribuyen en $(x, x+1]$, entonces el número total de muertes esperadas es $\sum_{i=1}^{n_x} {}_{s_i - r_i}q_{x+r_i}$. (Por conveniencia se usará n en lugar de n_x en las sumas). Cuando se iguala al número actual de muertes observadas, se obtiene la ecuación de momentos

$$E\left[D_x\right] = \sum_{i=1}^{n} {}_{s_i - r_i}q_{x+r_i} = d_x, \tag{3.72}$$

(Ver [5], pág 114), donde D_x es la variable aleatoria de muertes en $(x, x+1]$, y d_x es el número observado.

Para resolver el estimador q_x de (3.72), se usará la aproximación

$$ {}_{s_i - r_i}q_{x+r_i} \approx (s_i - r_i) \cdot q_x. \tag{3.73} $$

(Ver [5], pág 114). Entonces (3.72) se convierte en

$$ E[D_x] = q_x \cdot \sum_{i=1}^{n} (s_i - r_i) = d_x, \tag{3.74} $$

del cual es fácil obtener

$$ \widehat{q}_x = \frac{d_x}{\sum\limits_{i=1}^{n} (s_i - r_i)}, \tag{3.75} $$

que es la forma general del estimador de momentos en un medio con decremento simple.

Casos especiales

Si $r_i = 0$ y $s_i = 1$ para todas las personas (n_x) quienes contribuyen a $(x, x+1]$, entonces se tiene ${}_{s_i - r_i}q_{x+r_i} = q_x$ y (3.72) se convierte en

$$ E[D_x] = n_x \cdot q_x = d_x, \tag{3.76} $$

entonces (3.75) se convierte en

$$ \widehat{q}_x = \frac{d_x}{n_x}. \tag{3.77} $$

(Ver [5], pág 114).

Este es el **caso especial A**, donde (3.77) es un estimador de proporción binomial, donde el número de personas en la muestra, n_x, puede ser pensado como un número de ensayos binomiales. Cada ensayo es considerado independiente y la probabilidad de muerte en una prueba simple (q_x) se supone constante para todas las pruebas. En esta situación, la proporción de muertes en la muestra , dadas por (3.77), es un estimador natural para el parámetro q_x.

Si $s_i = 1$ para todas las personas, pero $r_i > 0$ para algunas de las n_x personas quienes contribuyen a $(x, x+1]$, entonces se tiene ${}_{s_i - r_i}q_{x+r_i} = {}_{1-r_i}q_{x+r_i}$, y (3.72) se convierte en

$$E\left[D_x\right] = \sum_{i=1}^{n} {}_{1-r_i}q_{x+r_i} = d_x, \tag{3.78}$$

entonces la aproximación general dada por (3.73) se convierte en

$${}_{1-r_i}q_{x+r_i} \approx (1-r_i)\cdot q_x. \tag{3.79}$$

Sustituyendo (3.79) en (3.78), se obtiene el resultado

$$\widehat{q}_x = \frac{d_x}{\sum\limits_{i=1}^{n}(1-r_i)}, \tag{3.80}$$

(Ver [5], pág 115), que es el estimador de momentos para el **caso especial B**.

Si $r_i = 0$ para todas las personas, pero $s_i < 1$ para algunas de las n_x personas quienes contribuyen a $(x, x+1]$, entonces se tiene ${}_{s_i-r_i}q_{x+r_i} = {}_{s_i}q_x$, y (3.72) se convierte en

$$E\left[D_x\right] = \sum_{i=1}^{n} {}_{s_i}q_x = d_x, \tag{3.81}$$

la aproximación general dada por (3.73) se convierte en

$${}_{s_i}q_x = s_i \cdot q_x. \tag{3.82}$$

Cuando (3.82) es sustituido en (3.81) se obtiene el resultado

$$\widehat{q}_x = \frac{d_x}{\sum\limits_{i=1}^{n}s_i}, \tag{3.83}$$

(Ver [5], pág 116), que es el estimador de momentos para el **caso especial C**.

Exposición

El par ordenado (r_i, s_i), representa el intervalo de edad dentro de $(x, x+1]$, en el cual la persona i, está bajo observación. Se dice que la persona i está **expuesta al riesgo de muerte** desde la edad $x + r_i$ hasta la edad $x + s_i$. Numéricamente, $s_i - r_i$, es la longitud de tiempo (en años) que la persona

i está expuesta, también representa la cantidad de exposición, medida en unidades de años-vividos, con la que contribuye esa persona. La exposición total contribuida por la muestra es $\sum_{i=1}^{n}(s_i - r_i)$, el denominador del estimador de (3.75). Es importante ver que el denominador del estimador de momentos representa la **exposición programada** de la muestra. Cuando un miembro de la muestra muere bajo la observación, no toda la exposición programada es realizada. La exposición realizada es la **exposición exacta** y no puede exceder a la exposición programada.

Intuitivamente, las muertes observadas divididas entre la exposición exacta proporciona un estimador de la tasa central de muerte, m_x, para el intervalo $(x, x+1]$. A partir del valor derivado de $\widehat{m}_x$, se puede estimar q_x por

$$\widehat{q}_x = \frac{\widehat{m}_x}{1 + \frac{1}{2} \cdot \widehat{m}_x}, \tag{3.84}$$

donde se hizo el supuesto de distribución lineal para ℓ_{x+s} (ver (2.68)). Alternativamente, si se supone una distribución exponencial para ℓ_{x+s} (el supuesto de fuerza de mortalidad constante) entonces esta fuerza constante, μ, es la misma que m_x, entonces $\widehat{\mu} = \widehat{m}_x$, por lo tanto

$$\widehat{q}_x = 1 - e^{-\widehat{\mu}} = 1 - e^{-\widehat{m}_x}. \tag{3.85}$$

Ver (2.77).

Agrupamiento

Se considera el caso especial B, en donde las n_x personas quienes contribuyen a $(x, x+1]$, entran a este intervalo a varias edades, pero están programadas a salir del intervalo a la edad $x+1$. Es razonable suponer que varias de estas n_x personas, entran al intervalo a la edad exacta x (es decir, se tiene $r_i = 0$); Sea b_x el subconjunto de n_x que tiene $r_i = 0$. El resto de la muestra será $k_x = n_x - b_x$, donde todas las personas tienen r_i con $0 < r_i < 1$. Se supone que el valor promedio de estas r_i's es r, con $0 < r < 1$.

Un resultado de este agrupamiento es el número esperado de muertes en el intervalo $(x, x+1]$, mediante $b_x \cdot q_x + k_x \cdot {}_{1-r}q_{x+r}$, por lo que la ecuación de momentos será

$$E\left[D_x\right] = b_x \cdot q_x + k_x \cdot {}_{1-r}\, q_{x+r} = d_x, \tag{3.86}$$

bajo un supuesto hiperbólico (ver (2.89)), ${}_{1-r}q_{x+r} = (1-r)\cdot q_x$, de (3.86) es fácil resolver para el estimador

$$\widehat{q}_x = \frac{d_x}{b_x + (1-r)\cdot k_x}. \tag{3.87}$$

(Ver [5], pág 118).

Ahora se considera el caso especial C, en donde todas las n_x personas están expuestas desde la edad x (es decir, tienen $r_i = 0$), pero están programadas a salir a distintas edades z_i con $x < z_i \leq x+1$. Sea c_x el subconjunto de n_x, que representa la salida programada a edad $z_i < x+1$ ($s_i < 1$) . El resto de la muestra, $n_x - c_x$, están programadas a permanecer bajo observación hasta la edad $x+1$. Se supone que le valor promedio de estas s_i's es s, con $0 < s < 1$.

El número esperado de muertes en $(x, x+1]$, es ahora $(n_x - c_x)\cdot q_x + c_x\cdot {}_sq_x$, por lo que la ecuación de momentos se convierte en

$$E\left[D_x\right] = (n_x - c_x)\cdot q_x + c_x\cdot\ {}_sq_x, \tag{3.88}$$

bajo un supuesto lineal, ${}_sq_x = s\cdot q_x$,de (3.88) es fácil resolver para el estimador

$$\widehat{q}_x = \frac{d_x}{n_x - (1-s)\cdot c_x}. \tag{3.89}$$

(Ver [5], pág 119).

Propiedades de los estimadores de momentos

Se explorará la media y varianza de los estimadores de momentos. Una característica de los estimadores de momentos para q_x es que son siempre insesgados bajo los supuestos lineal ó exponencial u otras aproximaciones para obtenerlos. Para investigar el insesgamineto, se necesita comparar el valor esperado del estimador con el valor real del parámetro q_x, el cual se busca estimar. Se usará $\widehat{q}_x$ para representar el valor observado de la variable aleatoria del estimador, cuando la variable aleatoria del número de muertes D_x, toma el valor d_x. Se usará $\widehat{Q}_x$ para denotar la variable aleatoria del estimador, por lo que

$$\widehat{Q}_x = \frac{D_x}{\sum\limits_{i=1}^{n}(s_i - r_i)} \tag{3.90}$$

es la relación entre las variables aleatorias $\widehat{Q}_x$ y D_x para el estimador de momentos en un medio de decremento simple. Entonces el valor esperado de $\widehat{Q}_x$ está dado por

$$E\left[\widehat{Q}_x\right] = \frac{E\left[D_x\right]}{\sum\limits_{i=1}^{n}\left(s_i - r_i\right)},$$

donde $E\left[D_x\right] = \sum\limits_{i=1}^{n} {}_{s_i-r_i}q_{x+r_i}$. Cuando ${}_{s_i-r_i}q_{x+r_i}$ es aproximada por (3.73), se obtiene

$$E\left[\widehat{Q}_x\right] = \frac{\sum\limits_{i=1}^{n} {}_{s_i-r_i}q_{x+r_i}}{\sum\limits_{i=1}^{n}\left(s_i - r_i\right)} = \frac{q_x \cdot \sum\limits_{i=1}^{n}\left(s_i - r_i\right)}{\sum\limits_{i=1}^{n}\left(s_i - r_i\right)} = q_x. \tag{3.91}$$

Esto muestra que el estimador (3.75) es insesgado bajo la aproximación usada.

Para determinar la varianza del estimador de momentos general dada por (3.75), primero se debe notar que la variable aleatoria para el número de muertes en $(x, x+1]$ D_x, puede ser escrita como:

$$D_x = \sum_{i=1}^{n} D_i, \tag{3.92}$$

donde D_i es la variable aleatoria (con valores 0 ó 1) del número de muertes para la persona i. Se supone que las D_i, $i = 1, 2, \ldots, n$, son mutuamente independientes. Donde D_i tiene una distribución binomial, con tamaño de muestra 1 y probabilidad de ocurrencia ${}_{s_i-r_i}q_{x+r_i}$, por lo que:

$$Var\left(D_x\right) = \sum_{i=1}^{n} Var\left(D_i\right) = \sum_{i=1}^{n} {}_{s_i-r_i}q_{x+r_i}\left(1 - {}_{s_i-r_i}q_{x+r_i}\right), \tag{3.93}$$

y

$$Var\left(\widehat{Q}_x \mid n_{x,} \{r_i, s_i\}\right) = \frac{\sum\limits_{i=1}^{n} {}_{s_i-r_i}q_{x+r_i}\left(1 - {}_{s_i-r_i}q_{x+r_i}\right)}{\left[\sum\limits_{i=1}^{n}\left(s_i - r_i\right)\right]^2}. \tag{3.94}$$

La notación $Var\left(\widehat{Q}_x \mid n_{x,} \{r_i, s_i\}\right)$ es usada para recordar que la varianza está condicionada al tamaño de la muestra $n_{x,}$, y a los tiempos en los que cada persona entra y está programada a salir en $(x, x+1]$. La varianza de (3.75) puede ser aproximada mediante $\widehat{Q}_x$, ya que es una proporción binomial con tamaño de muestra $\sum\limits_{i=1}^{n}(s_i - r_i)$, la aproximación de esta varianza es

$$Var\left(\widehat{Q}_x \mid n_{x,} \{r_i, s_i\}\right) \approx \frac{p_x q_x}{\sum\limits_{i=1}^{n}(s_i - r_i)}. \tag{3.95}$$

Una aproximación de momentos diferente para el caso especial C

Las muertes esperadas fuera de c_x, están programadas a salir a la edad $x+s$ y están dadas por $c_x \cdot {}_s q_x$. La fracción del intervalo de probabilidad requiere un supuesto uniforme para resolver la ecuación de momentos. Ahora se considera una diferente aproximación en donde se supone que los c_x finalistas programados tienen la misma probabilidad de muerte en $(x, x+1]$, así como los restantes $n_x - c_x$, de q_x. Se consideran todas las muertes fuera de c_x en $(x, x+1]$; se llamará d_x^c. Algunas de estas muertes ocurrirán despues del estudio y así ser observadas, mientras que otras ocurrirán antes de la salida y no son observadas. Sea d'_x el número de muertes observadas fuera de c_x y α representa la proporción de d_x^c que es observada. Entonces $d'_x = \alpha \cdot d_x^c$.

Se debe notar que d_x^c es conocido, mientras que α, no está calculado. Del grupo c_x, $c_x \cdot q_x$ son las muertes esperadas en $(x, x+1]$ y $\alpha \cdot c_x \cdot q_x$ es el número esperado de muertes observables. $(n_x - c_x) \cdot q_x$ son las muertes esperadas observables fuera de $n_x - c_x$, por lo que el total de muertes esperadas observables está dado por la ecuación de momentos

$$(n_x - c_x) \cdot q_x + \alpha \cdot c_x \cdot q_x = n_x \cdot q_x - (1-\alpha) \cdot c_x \cdot q_x = d_x. \tag{3.96}$$

Sea e_x el número actual de finalistas en la muestra. Ahora se considerará el número esperado observable de finalistas en la muestra. Se tiene $c_x(1-q_x)$ como el número esperado de no muertos en $(x, x+1]$ y $(1-\alpha) \cdot c_x \cdot q_x$ el esperado a morir, pero después de que han sido finalistas observados. Entonces el total esperado de finalistas observables, está dado por la ecuación de momentos

$$c_x(1-q_x) + (1-\alpha) \cdot c_x \cdot q_x = c_x - \alpha \cdot c_x \cdot q_x = e_x. \tag{3.97}$$

Se debe notar que estos no son los finalistas observados en $(x, x+1]$ desde el grupo $n_x - c_x$.

El estimador de q_x es obtenido de eliminar c_x de las ecuaciones (3.96) y (3.97). De la ecuación (3.97), se tiene $c_x = \frac{e_x}{1-\alpha \cdot q_x}$; sustituyendo en (3.96) se tiene que:

$$n_x \cdot q_x - \frac{(1-\alpha)\, e_x \cdot q_x}{1 - \alpha \cdot q_x} = d_x,$$

lo cual da la ecuación cuadrática

$$\alpha \cdot n_x \cdot q_x^2 - [n_x - (1-\alpha) \cdot e_x + \alpha \cdot d_x]\, q_x + d_x = 0, \tag{3.98}$$

donde la solución es

$$\widehat{q}_x = \frac{b - \sqrt{b^2 - 4\alpha \cdot n_x \cdot d_x}}{2\alpha \cdot n_x}, \tag{3.99}$$

(Ver [5], pág 123), donde $b = n_x - (1-\alpha) \cdot e_x + \alpha \cdot d_x$. El radical negativo es usado en (3.99) por que en el radical positivo se tendría que $\widehat{q}_x > 1$. Una ventaja del estimador (3.99) sobre el estimador (3.89) es que no se requiere conocer el valor de c_x, pero depende del número actual de finalistas observados, e_x; tampoco se require de un supuesto de distribución, pero se require suponer un valor para α.

3.3.2 Estimación de momentos en un medio de decremento doble

La probabilidad de muerte para la persona i bajo una observación en un medio con doble decremento doble es ${}_{s_i - r_i}q^{(d)}_{x+r_i}$, y es también el número esperado de muertes de tamaño uno para el caso de decremento simple.

La relación básica de momentos

Se consideran tanto la muerte como la retirada, eventos aleatorios. Analogamente a la ecuación de momentos (3.72) en el caso de decremento simple, se escribe un par de ecuaciones de momentos

$$E[D_x] = \sum_{i=1}^{n} {}_{s_i - r_i}q^{(d)}_{x+r_i} = d_x \tag{3.100}$$

y

$$E\left[W_x\right]=\sum_{i=1}^{n} {}_{s_i-r_i}q_{x+r_i}^{(w)}=w_x \tag{3.101}$$

donde W_x es la variable aleatoria de salidas en $(x, x+1]$, y w_x denota el número observado de salidas en la muestra. Una forma de resolver las ecuaciones (3.100) y (3.101), es usar aproximaciones análogas a (3.73), que servirán para estimar las probabilidades $q_x^{(d)}$ y $q_x^{(w)}$ en un medio de doble decremento:

$${}_{s_i-r_i}q_{x+r_i}^{(d)} \approx (s_i-r_i)\cdot q_x^{(d)} \tag{3.102}$$

y

$${}_{s_i-r_i}q_{x+r_i}^{(w)} \approx (s_i-r_i)\cdot q_x^{(w)}, \tag{3.103}$$

produciendo

$$\widehat{q}_x^{(d)}=\frac{d_x}{\sum\limits_{i=1}^{n}(s_i-r_i)} \tag{3.104}$$

y

$$\widehat{q}_x^{(w)}=\frac{w_x}{\sum\limits_{i=1}^{n}(s_i-r_i)}. \tag{3.105}$$

Sin embargo, el objetivo es estimar las probabilidades $q_x^{\prime(d)}$ y $q_x^{\prime(w)}$, en un medio con decremento simple. Se deberá estimar $q_x^{\prime(d)}$ y $q_x^{\prime(w)}$ de los estimados $\widehat{q}_x^{(d)}$ y $\widehat{q}_x^{(w)}$ bajo un supuesto (lineal o exponencial) y usando las aproximaciones (3.102) y (3.103). Estas aproximaciones normalmente resultan en ecuaciones que se deberán resolver numéricamente para $\widehat{q}_x^{\prime(d)}$ y $\widehat{q}_x^{\prime(w)}$.

Supuesto de distribución uniforme

Si se supone que los eventos aleatorios, muerte y retirada, tienen una distribución uniforme en $(x, x+1]$ en el contexto de un medio con decremento simple, de (2.63) se sabe que

$$ {}_{u-r}q'^{(d)}_{x+r} = \frac{(u-r)\cdot q'^{(d)}_x}{1-r\cdot q'^{(d)}_x}, $$

con $0 \leq r < u \leq 1$, también se sabe de (2.64) que:

$$ \mu^{(d)}_{x+u} = \frac{q'^{(d)}_x}{1-u\cdot q'^{(d)}_x}, $$

por lo que,

$$ \left(1 - {}_{u-r}q'^{(d)}_{x+r}\right)\mu^{(d)}_{x+u} = \frac{q'^{(d)}_x}{1-r\cdot q'^{(d)}_x}. $$

Entonces:

$$ \begin{aligned} {}_{s-r}q^{(d)}_{x+r} &= \int_r^s \left[1 - {}_{u-r}q'^{(w)}_{x+r}\right]\left[1 - {}_{u-r}q'^{(d)}_{x+r}\right]\mu^{(d)}_{x+u}du \\ &= \int_r^s \frac{q'^{(d)}_x}{1-r\cdot q'^{(d)}_x}\cdot\frac{1-u\cdot q'^{(w)}_x}{1-r\cdot q'^{(w)}_x}du \\ &= \frac{q'^{(d)}_x}{\left[1-r\cdot q'^{(d)}_x\right]\left[1-r\cdot q'^{(w)}_x\right]}\int_r^s \left[1-u\cdot q'^{(w)}_x\right]du \\ &= \frac{q'^{(d)}_x\left[(s-r)-\frac{1}{2}\left(s^2-r^2\right)\cdot q'^{(w)}_x\right]}{\left[1-r\cdot q'^{(d)}_x\right]\left[1-r\cdot q'^{(w)}_x\right]}. \end{aligned} \tag{3.106} $$

Similarmente, se tiene

$$ {}_{s-r}q^{(w)}_{x+r} = \frac{q'^{(w)}_x\left[(s-r)-\frac{1}{2}\left(s^2-r^2\right)\cdot q'^{(d)}_x\right]}{\left[1-r\cdot q'^{(d)}_x\right]\left[1-r\cdot q'^{(w)}_x\right]}. \tag{3.107} $$

Sustituyendo (3.106) en (3.100) resulta

$$ E\left[D_x\right] = \sum_{i=1}^{n}\frac{q'^{(d)}_x\left[(s-r)-\frac{1}{2}\left(s^2-r^2\right)\cdot q'^{(w)}_x\right]}{\left[1-r\cdot q'^{(d)}_x\right]\left[1-r\cdot q'^{(w)}_x\right]} = d_x, \tag{3.108} $$

similarmente sustituyendo (3.107) en (3.101) resulta

$$E\left[W_x\right] = \sum_{i=1}^{n} \frac{q_x'^{(w)} \left[(s-r) - \frac{1}{2}\left(s^2 - r^2\right) \cdot q_x'^{(d)}\right]}{\left[1 - r \cdot q_x'^{(d)}\right]\left[1 - r \cdot q_x'^{(w)}\right]} = w_x. \tag{3.109}$$

La complejidad de las ecuaciones anteriores requiere una solución númerica

En el caso especial A, en donde $r_i = 0$ y $s_i = 1$ para toda i, las estimaciones de $q_x'^{(d)}$ y $q_x'^{(w)}$ se simplifican considerablemente ya que las ecuaciones (3.104) y (3.105) se simplifican en

$$\widehat{q}_x^{(d)} = \frac{d_x}{n_x} \tag{3.110}$$

y

$$\widehat{q}_x^{(w)} = \frac{w_x}{n_x}, \tag{3.111}$$

por lo que se producen los estimadores:

$$\widehat{q}_x'^{(d)} = \frac{\left(n_x - \frac{1}{2}w_x + \frac{1}{2}d_x\right) - \sqrt{\left(n_x - \frac{1}{2}w_x + \frac{1}{2}d_x\right)^2 - 2n_x d_x}}{n_x} \tag{3.112}$$

(Ver [5], pág 127), y

$$\widehat{q}_x'^{(w)} = \frac{\left(n_x - \frac{1}{2}d_x + \frac{1}{2}w_x\right) - \sqrt{\left(n_x - \frac{1}{2}d_x + \frac{1}{2}w_x\right)^2 - 2n_x w_x}}{n_x}. \tag{3.113}$$

Supuesto de distribución exponencial

Si los eventos aleatorios, muerte y retirada, se supone que tienen una distribución exponencial (fuerza constante) en $(x, x+1]$, entonces $\mu_{x+u}^{(d)} = \mu^{(d)}$, $1 - {}_{u-r}q_{x+r}'^{(d)} = e^{-(u-r)\mu^{(d)}}$ y $1 - {}_{u-r}q_{x+r}'^{(w)} = e^{-(u-r)\mu^{(w)}}$ por lo que

$$
\begin{aligned}
{}_{s-r}q_{x+r}^{(d)} &= \int_r^s \left[1 - {}_{u-r}q_{x+r}^{\prime(w)}\right]\left[1 - {}_{u-r}q_{x+r}^{\prime(d)}\right]\mu_{x+u}^{(d)} du \\
&= \mu^{(d)} \int_r^s e^{-(u-r)\left(\mu^{(d)}+\mu^{(w)}\right)} du \\
&= \frac{\mu^{(d)}}{\mu^{(d)}+\mu^{(w)}}\left[1 - e^{-(u-r)\left(\mu^{(d)}+\mu^{(w)}\right)}\right] \qquad (3.114)
\end{aligned}
$$

y análogamente

$$
{}_{s-r}q_{x+r}^{(w)} = \frac{\mu^{(w)}}{\mu^{(d)}+\mu^{(w)}}\left[1 - e^{-(u-r)\left(\mu^{(d)}+\mu^{(w)}\right)}\right]. \qquad (3.115)
$$

Sustituyendo (3.114) en (3.100) resulta

$$
E\left[D_x\right] = \sum_{i=1}^{n} \frac{\mu^{(d)}}{\mu^{(d)}+\mu^{(w)}}\left[1 - e^{-(u-r)\left(\mu^{(d)}+\mu^{(w)}\right)}\right] = d_x, \qquad (3.116)
$$

análogamente sustituyendo (3.115) en (3.101) resulta

$$
E\left[W_x\right] = \sum_{i=1}^{n} \frac{\mu^{(w)}}{\mu^{(d)}+\mu^{(w)}}\left[1 - e^{-(u-r)\left(\mu^{(d)}+\mu^{(w)}\right)}\right] = w_x. \qquad (3.117)
$$

De nuevo es necesario una solución numérica para las ecuaciones $\widehat{\mu}^{(d)}$ y $\widehat{\mu}^{(w)}$.

En el caso especial A, de las ecuaciones (3.69) y (3.70) se tiene

$$
q_x^{\prime(d)} = 1 - \left[\frac{n_x - d_x - w_x}{n_x}\right]^{d_x/(d_x+w_x)} \qquad (3.118)
$$

(Ver [5], pág 128), y

$$
q_x^{\prime(w)} = 1 - \left[\frac{n_x - d_x - w_x}{n_x}\right]^{w_x/(d_x+w_x)}. \qquad (3.119)
$$

Estimación de momentos por la aproximación de Hoem en un medio con decremendo doble

La dificultad matemática de las soluciones de (3.100) y (3.101) para $\widehat{q}_x'^{(d)}$ y $\widehat{q}_x'^{(w)}$, bajo un supuesto lineal o exponencial, motiva a encontrar una aproximación diferente para la estimación de momentos en un medio con decremento doble. Una opción es la aproximación actuarial y otra es la aproximación de Hoem.

Para la persona i, quien entra a $(x, x+1]$ a la edad $x+r_i$, Hoem contempla una edad teórica programada para la retirada, llamada $x+t_i$ en la misma forma que existe una edad programada de salida al final del periodo de observación, llamado $x+s_i$. La diferencia es que $x+s_i$ es conocida, mientras que $x+t_i$ es desconocida. Para estas personas con $t_i < s_i$, los valores de t_i se vuelven conocidas posteriormente cuando la experiencia de la muestra es revelada y las salidas son observadas.

Usando la interpretación de exposición del estimador de momentos, se encuentra la exposición $(s_i - r_i)$ si la persona i sobrevive a la edad programada de salida, y la exposición $(t_i - r_i)$ si la persona i se retira a la edad $x+t_i$.

La exposición para una muerte fue la exposición programada $(s_i - r_i)$, en un medio con decremento simple. En un medio con decremento doble se usará la exposición programada para una muerte, pero ésta será la menor entre $(s_i - r_i)$ y $(t_i - r_i)$. En otras palabras, si una persona tiene un tiempo de retirada programado desconocido de $t_i < s_i$, y muere antes del tiempo t_i, este tiempo de retirada programado nunca será conocido.

Este dilema de Hoem se puede simplificar suponiendo que nadie que muera en $(x, x+1]$ tiene un valor t_i que fuese menor que s_i. En otras palabras, todas las muertes contribuyen $(s_i - r_i)$ a la exposición. Entonces el estimador de $q_x'^{(d)}$ está dado por la razón del número de muertes a la exposición.

El punto importante es que todas las personas tienen un valor conocido de s_i al entrar en $(x, x+1]$, a pesar del resultado de la perona i (sobreviva, muera o se retire). Cada valor t_i puede ser conocido **solamente** si la persona actualmente se retira.

3.3.3 La aproximación actuarial para la estimación de momentos

De las aproximaciones para la estimación de q_x sobre el intervalo $(x, x+1]$, la aproximación actuarial fue la primera en ser desarrollada, desde mediados del siglo XIX. El método fue desarrollado de manera intuitiva, en ausencia de teoría estadística y fue hecho con el interés de simplificar los cálculos y almacenar la información.

La aproximación actuarial tradicional permaneció hasta la segunda mitad del siglo XX, cuando aumentó la capacidad de procesar datos y surgió la teoría estadística moderna. El método actuarial supone un medio con decremento doble.

El concepto de exposición actuarial

La principal diferencia entre la aproximación actuarial tradicional y la aproximación de momentos, es que la aproximación actuarial por lo general no hace uso del concepto de exposición programada. Más bien, hace uso de un tipo de exposición observada, la cual se llamará **exposición actuarial**.

Se supone que la persona i entra a $(x, x+1]$ a la edad $x + r_i$, $0 \leq r_i < 1$. Bajo el método actuarial, si la persona i es un **finalista observado** a edad $x + s_i$, él contribuye una exposición de $(s_i - r_i)$. Si él es un **retirado observado** a la edad $x + t_i$, donde necesariamente, $t_i < s_i$, el contribuye una exposición de $(t_i - r_i)$. Sin embargo, la identificación de la edad programada de salida, para una persona quien muere antes, requiere un análisis extensivo de datos. Para resolver el problema de un desconocido $x + s_i$ para una muerte observada, se puede suponer que $s_i = 1$ para todas las muertes.

Propiedades del estimador actuarial

En un medio con decremento simple, el estimador actuarial para los casos especiales A y B, es el mismo que el estimador de momentos. En situaciones donde los finalistas a edad $x + s_i$ donde $s_i < 1$, varias investigaciones han mostrado en una simulación, que el estimador actuarial es negativamente sesgado. A gran escala los estudios actuariales de seguros de vida, siempre consideran un número considerable de finalistas para el periodo de observación.

El estimador de momentos siempre produce un mayor estimador $\widehat{q}_x$ que el estimador actuarial. Finalmente, todos los estimadores de la forma $\frac{\text{muertes}}{\text{exposición}}$, son aproximadamente de proporción binomial, todos son aproximadamente insesgados y con varianza dada por: $\frac{p_x q_x}{\text{exposición}}$.

3.3.4 Estimación de $S(x)$

Para estimar $S(x)$ se usará la aproximación

$$\widehat{S}(x) = \widehat{p}_0 \cdot \widehat{p}_1 \cdot \cdots \cdot \widehat{p}_{x-1}. \tag{3.120}$$

Para encontrar el valor esperado y la varianza de $\widehat{S}(x)$, se debe suponer:

1. Que cada $\widehat{p}_i$, $i = 0, 1, \ldots, x-1$, es aproximadamente una proporción binomial e insesgado, con media p_i y varianza $\frac{p_i q_i}{n'_i}$, donde n'_i es la exposición usada para calcular $\widehat{q}_i$. Se debe notar que esta exposición es al principio interpretada como una muestra binomial aproximada.

2. Que las $\widehat{p}_i$'s son mutuamente independientes.

La principal consecuencia de la primera suposición es entonces que es insesgado, $E[\ \widehat{p}_i \mid n'_i] = p_i$. La importancia de la segunda suposición es que el valor esperado de un producto de variables aleatorias independientes es el producto de sus valores esperados.

Valor esperado de $\widehat{S}(x)$

Es importante notar que este valor esperado es condicional en el conjunto de exposiciones de los estimados de $\widehat{q}_i$, para $i = 0, 1, \ldots, x-1$. Esto se refleja usando la notación $E\left[\widehat{S}(x) \mid \{n'_i\}\right]$.

Con $\widehat{S}(x)$ dado por (3.120), se tiene

$$E\left[\widehat{S}(x) \mid \{n'_i\}\right] = E\left[\widehat{p}_0 \cdot \widehat{p}_1 \cdot \cdots \cdot \widehat{p}_{x-1} \mid \{n'_i\}\right] = \prod_{i=0}^{x-1} E[\widehat{p}_i \mid \{n'_i\}], \tag{3.121}$$

como una consecuencia de la suposicion de independencia. Entonces de la suposición de que es insesgado, se tiene

$$E\left[\widehat{S}(x) \mid \{n'_i\}\right] = p_0 \cdot p_1 \cdot \cdots \cdot p_{x-1} = S(x). \tag{3.122}$$

Esto muestra que $\widehat{S}(x)$ es insesgado bajo los dos supuestos.

Varianza de $\widehat{S}(x)$

De nuevo se nota que esta varianza es condicional en el conjunto de exposiciones y se refleja en la notación $Var\left[\widehat{S}(x) \mid \{n_i'\}\right]$.

Se procede de a primera suposición,

$$\begin{aligned} Var\left[\widehat{S}(x) \mid \{n_i'\}\right] &= E\left[\widehat{S}(x)^2 \mid \{n_i'\}\right] - \left(E\left[\widehat{S}(x) \mid \{n_i'\}\right]\right)^2 \\ &= E\left[\widehat{p}_0^2 \cdot \widehat{p}_1^2 \cdot \cdots \cdot \widehat{p}_{x-1}^2 \mid \{n_i'\}\right] - \\ &\quad \left(E\left[\widehat{p}_0 \cdot \widehat{p}_1 \cdot \cdots \cdot \widehat{p}_{x-1} \mid \{n_i'\}\right]\right)^2 \end{aligned} \tag{3.123}$$

por la definición de $\widehat{S}(x)$ dada por (3.120). La suposición de independencia hace que se reescriba (3.123) como:

$$Var\left[\widehat{S}(x) \mid \{n_i'\}\right] = \prod_{i=0}^{x-1} E[\widehat{p}_i^2 \mid \{n_i'\}] - \left(\prod_{i=0}^{x-1} E[\widehat{p}_i \mid \{n_i'\}]\right)^2. \tag{3.124}$$

Ahora para cada $\widehat{p}_i$,

$$Var\left(\widehat{p}_i \mid \{n_i'\}\right) = E\left[\widehat{p}_i^2 \mid \{n_i'\}\right] - \left(E\left[\widehat{p}_i \mid \{n_i'\}\right]\right)^2,$$

por lo que

$$E\left[\widehat{p}_i^2 \mid \{n_i'\}\right] = Var\left(\widehat{p}_i \mid \{n_i'\}\right) + \left(E\left[\widehat{p}_i \mid \{n_i'\}\right]\right)^2.$$

Pero cada $\widehat{p}_i$ es tomado como una proporción binomial insesgada, por lo que el primer momento es

$$E\left[\widehat{p}_i \mid \{n_i'\}\right] = p_i, \tag{3.125}$$

y el segundo momento es

$$E\left[\widehat{p}_i^2 \mid \{n_i'\}\right] = \frac{p_i q_i}{n_i'} + p_i^2 = (p_i)^2 \cdot \left(\frac{q_i}{p_i n_i'} + 1\right). \tag{3.126}$$

Sustituyendo (3.125) y (3.126) en (3.124), se tiene

$$
\begin{aligned}
Var\left[\widehat{S}(x) \mid \{n_i'\}\right] &= \prod_{i=0}^{x-1} (p_i)^2 \cdot \left(\frac{q_i}{p_i n_i'} + 1\right) - \left(\prod_{i=0}^{x-1} p_i\right)^2 \\
&= \left(\prod_{i=0}^{x-1} p_i\right)^2 \cdot \prod_{i=0}^{x-1} \left(\frac{q_i}{p_i n_i'} + 1\right) - \left(\prod_{i=0}^{x-1} p_i\right)^2 \\
&= [S(x)]^2 \cdot \left[\prod_{i=0}^{x-1} \left(\frac{q_i}{p_i n_i'} + 1\right) - 1\right], \qquad (3.127)
\end{aligned}
$$

donde $S(x) = p_0 \cdot p_1 \cdot \cdots \cdot p_{x-1} = \prod\limits_{i=0}^{x-1} p_i$.

La expresión exacta para la varianza condicional de $\widehat{S}(x)$, dada por (3.127) , en algunas ocasiones es aproximada de la siguiente manera. Expandiendo el producto de los términos binomiales, se obtiene

$$
\begin{aligned}
\prod_{i=0}^{x-1} \left(\frac{q_i}{p_i n_i'} + 1\right) &= \left(1 + \frac{q_0}{p_0 n_0'}\right)\left(1 + \frac{q_1}{p_1 n_1'}\right) \cdots \left(1 + \frac{q_{x-1}}{p_{x-1} n_{x-1}'}\right) \\
&= 1 + \frac{q_0}{p_0 n_0'} + \frac{q_1}{p_1 n_1'} + \cdots + \frac{q_{x-1}}{p_{x-1} n_{x-1}'} + (\text{otros términos})
\end{aligned}
$$

donde los demas términos son muy pequeños y se ignoran, por lo que se puede aproximar $\prod\limits_{i=0}^{x-1} \left(\frac{q_i}{p_i n_i'} + 1\right)$ por $1 + \sum\limits_{i=0}^{x-1} \frac{q_i}{p_i n_i'}$. Entonces (3.127) , es aproximada por

$$
Var\left[\widehat{S}(x) \mid \{n_i'\}\right] \approx [S(x)]^2 \cdot \sum_{i=0}^{x-1} \frac{q_i}{p_i n_i'}. \qquad (3.128)
$$

Esta fórmula para la varianza condicional de $\widehat{S}(x)$ es conocida como la **fórmula de Greenwood**. Se debe notar que $\widehat{S}(x)$ es una función de las x variables aleatorias $\widehat{p}_0, \widehat{p}_1, \ldots, \widehat{p}_{x-1}$, específicamente del producto de ellas.

3.4 Procedimientos de máxima verosimilitud

En esta sección se estudiará la **estimación por máxima verosimilitud** (EMV) como una alternativa de los estimadores actuariales y de momentos, considerando un medio con decremento sencillo y doble.

En la estimación, la edad precisa de muerte es información que se puede incorporar para conseguir una mejor estimación. Se deben distinguir dos subdivisiones importantes del estimador máximo verosimil:

1. En el caso donde se usa la edad precisa de muerte, se dice que se tiene una **situación de datos completos**.
2. En el caso donde sólo se usa el número de muertes en el intervalo $(x, x+1]$, se dice que se tiene una **situación de datos incompletos**.

Se debe notar que en la estimación de momentos se usan los datos incompletos de las muertes, mientras que en el estimador de máxima verosimilitud se pueden usar tanto datos incompletos como datos completos. La información que se necesita para conocer la **función de verosimilitud** son los datos disponibles, incluyendo la cuestión de que se tengan datos incompletos o completos para las muertes.

3.4.1 Medio de decremento simple, caso especial A

Datos incompletos

En esta situación, de los n_x vivos a edad exacta x, d_x de ellos mueren en $(x, x+1]$, y $n_x - d_x$ sobreviven a la edad $x+1$. En este contexto, se tiene una muestra aleatoria $Y_1, ..., Y_{n_x}$, donde cada Y_i $(i = 1, ..., n_x)$ tiene una distribución Bernoulli con $P(Y_i = 1) = q_x$, la probabilidad de muerte, por lo que

$$f(y_i) = q_x^{y_i} \cdot (1-q_x)^{1-y_i} \qquad\qquad i = 1, ..., n_x$$

y la verosimilitud estaría dada por:

$$L\left(q_x \mid n_x d_x\right) = \prod_{i=i}^{n_x} f(y_i). \tag{3.129}$$

La notación $L\left(q_x \mid n_x d_x\right)$ hace recordar que la verosimilitud es una función donde q_x es desconocida, y que n_x y d_x son valores dados.

$$\begin{aligned} L\left(q_x \mid n_x d_x\right) &= q_x^{\sum\limits_{i=i}^{n_x} y_i} \cdot (1-q_x)^{\sum\limits_{i=i}^{n_x} 1 - \sum\limits_{i=i}^{n_x} y_i} \\ &= q_x^{d_x} \cdot (1-q_x)^{n_x - d_x}. \end{aligned} \tag{3.130}$$

Frecuentemente esta función se puede simplificar usando L en lugar de $L\left(q_x \mid n_x d_x\right)$. Finalmente, por conveniencia se suprime el subíndice x. Por lo que se puede reescribir la verosimilitud para el caso especial A, en una situación de datos incompletos como:

$$L = q^d \cdot (1-q)^{n-d}. \tag{3.131}$$

(Ver [5], pág 147).

El problema ahora es encontrar el valor de q, llamado $\widehat{q}$, el cual maximice (3.131). Formalmente, si $\widehat{q}$ existe, se tiene que $L\left(\widehat{q}\right) \geq L\left(q\right)$ para toda q.

Se define la **logverosimilitud** ℓ, como:

$$\ell = \ln L = d \cdot \ln q + (n-d) \cdot \ln\left(1-q\right). \tag{3.132}$$

De donde, al derivar e igualar a cero, se obtiene:

$$\widehat{q}_x = \frac{d_x}{n_x}, \tag{3.133}$$

que es el mismo estimador que (3.77).

Se tiene:

$$\left.\frac{\partial^2 \ln L}{\partial q_x^2}\right]_{\frac{d_x}{n_x}} = -\frac{1}{n_x^2 d_x} - \frac{1}{n_x^3 - n_x^2 d_x} < 0,$$

por lo si es un máximo.

Datos completos

Se supone que se tiene la edad precisa de muerte para cada una de las d_x muertes en el intervalo. Donde esta edad es diferente para cada muerte y se toma el producto de cada contribución de muerte a la función de verosimilitud. La verosimilitud para la $i-\acute{e}sima$ muerte está dada por la función de densidad de probabilidad por muerte a esa edad particular, dado que se está con vida a la edad x. Esto es, para cada muerte a la edad x_i,

$$L_i = f\left(x_i \mid X > x\right) = \frac{f\left(x_i\right)}{S\left(x\right)} = \frac{S\left(x_i\right) \cdot \lambda\left(x_i\right)}{S(x)} \tag{3.134}$$

es la contribución a L de la $i-\acute{e}sima$ muerte. Sea $s_i = x_i - x$ el tiempo de la $i-\acute{e}sima$ muerte en $(x, x+1]$, donde $0 < s_i \leq 1$, entonces se tiene

$$L_i = \frac{S(x+s_i) \cdot \lambda(x+s_i)}{S(x)} = \ {}_{s_i}p_x \left(\mu_{x+s_i}\right), \tag{3.135}$$

en notación estándar actuarial. La contribución a L por todas las muertes combinadas es:

$$\prod_{i=1}^{d} {}_{s_i}p_x \left(\mu_{x+s_i}\right), \tag{3.136}$$

donde es comunmente escrita como $\prod\limits_{D} {}_{s_i}p_x \left(\mu_{x+s_i}\right)$, y se lee como "multiplicado sobre todas las muertes".

Naturalmente, los $n_x - d_x$ sobrevivientes contribuyen $(1-q_x)^{n_x-d_x}$ a L, entonces la verosimilitud total es:

$$L = (1-q_x)^{n_x-d_x} \cdot \prod_{D} {}_{s_i}p_x \left(\mu_{x+s_i}\right), \tag{3.137}$$

para el caso especial A con una situación de datos completos.

Para resolver (3.137) para $\widehat{q}_x$, es necesario tomar un supuesto de distribución, para expresar ${}_{s_i}p_x \left(\mu_{x+s_i}\right)$ en términos de q_x. Se considerarán dos supuestos.

Bajo el supuesto lineal, o uniforme, de (2.65) se tiene que para toda s_i, $0 < s_i \leq 1$, ${}_{s_i}p_x \left(\mu_{x+s_i}\right) = q_x$. Entonces la contribución a L para todas las d_x muertes, en términos de $\prod\limits_{D} {}_{s_i}p_x \left(\mu_{x+s_i}\right)$ llega a ser $(q_x)^{d_x}$, y (3.137) llega a ser $q^d \cdot (1-q)^{n-d}$, que es (3.131). Esto es, para el caso especial A con datos completos, evaluado bajo el supuesto de distribución uniforme, produce el mismo estimador que el caso especial A con datos parciales, el cual no requiere supuesto para su evaluación.

Bajo el supuesto exponencial, se recuerda que μ_{x+s_i} es una constante llamada $\mu = -\ln p_x$ y de (2.72) se tiene que ${}_{s_i}p_x = (p_x)^{s_i} = e^{-\mu s_i}$, por lo que (3.137) se convierte en:

$$\begin{aligned} L &= (p_x)^{n_x-d_x} \cdot \prod_{D} (p_x)^{s_i} \cdot \mu \\ &= \mu^d \cdot \exp\left[-\mu(n-d) - \mu \cdot \sum_{D} s_i\right]. \end{aligned} \tag{3.138}$$

Entonces

$$\ell = \ln L = d \cdot \ln \mu - \mu \left[(n-d) + \sum_D s_i \right], \tag{3.139}$$

y

$$\frac{d\ell}{d\mu} = \frac{d}{\mu} - \left[(n-d) + \sum_D s_i \right] = 0, \tag{3.140}$$

de esta manera se tiene:

$$\widehat{\mu} = \frac{d}{(n-d) + \sum_D s_i}. \tag{3.141}$$

El denominador de (3.141) es la exposición exacta, mencionada anteriormente. Así, (3.141) estima a m_x, que es lo mismo que μ bajo el supuesto exponencial (fuerza constante). Por lo que

$$\widehat{q}_x = 1 - e^{-\widehat{\mu}} \tag{3.142}$$

es el estimador máximo verosimil de q_x. Puesto que q_x y μ tienen una correspondencia uno a uno.

Ahora se mostrará una forma alterna de escribir la función de verosimilitud para el modelo con datos completos. Se define t_i como el tiempo en el que la $i-ésima$ persona de las n_x deja la observación, ya sea que llegue a la edad $x+1$ (donde $t_i = 1$ y la persona es un sobreviviente), ó que muera (donde $t_i \leq 1$ y la persona i es un muerto). Se define δ_i como una variable indicadora donde

$$\delta_i = \begin{cases} 1 & \text{si la persona } i \text{ muere en } (x, x+1] \\ 0 & \text{en otro caso} \end{cases} \tag{3.143}$$

Entonces la contribución a L por la $i-ésima$ persona en n_x es

$$L_i = {}_{t_i}p_x \left(\mu_{x+t_i}\right)^{\delta_i}. \tag{3.144}$$

(Ver [5], pág 150). La estructura de (3.144) muestra que $L_i = {}_{t_i}p_x \left(\mu_{x+t_i}\right)$ si la $i-ésima$ persona es un muerto (donde $\delta_i = 1$), y $L_i = {}_{t_i}p_x$ si la $i-ésima$ persona es un sobreviviente (donde $\delta_i = 0$). En el caso especial A se tiene que $t_i = 1$ para toda i, donde $\delta_i = 0$.

3.4.2 Caso general en un medio de decremento simple

Se desarrollará una forma general para la función de verosimilitud en un medio de decremento simple para los casos especiales B y C.

Forma general para datos completos

Se debe recordar que n_x es el número total de personas en la muestra que contribuyen al intervalo de estimación $(x, x+1]$, y $x+r_i$ es la edad en el que la $i-ésima$ persona entra a $(x, x+1]$, $0 \leq r_i < 1$. Sea $x+t_i$ la edad en el que la $i-ésima$ persona deja a $(x, x+1]$, $0 < t_i \leq 1$, ya sea un finalista programado y observado, un sobreviviente del intervalo o un muerto. Sea δ_i la variable indicadora para la $i-ésima$ persona definida por (3.143). Entonces si $t_i = 1$ y $\delta_i = 0$, la $i - ésima$ persona es un sobreviviente; si $t_i < 1$ y $\delta_i = 0$, la $i - ésima$ persona es un finalista observado; si $t_i \leq 1$ y $\delta_i = 1$, la $i - ésima$ persona es un muerto.

Ahora se puede ver que si $r_i = 0$ para toda i, se tiene el caso especial A o el caso especial C. Similarmente, si $t_i = 1$ para toda i donde $\delta_i = 0$ (es decir, para toda i quien no muere), se tiene el caso especial A o el caso especial B. Esto se puede ver en la siguiente tabla:

	$t_i = 1$, $\forall i$, $\delta_i = 0$	$t_i < 1$, $p.a.i$, $\delta_i = 0$
$r_i = 0$, $\forall i$	Caso especial A	Caso especial C
r_i, $p.a.i$	Caso especial B	Caso general

Tabla 3.1

La forma general de la contribución a L por la $i - ésima$ persona es:

$$L_i = \ {}_{t_i - r_i}p_{x+r_i} \left(\mu_{x+t_i}\right)^{\delta_i}, \tag{3.145}$$

donde si $\delta_i = 0$, la verosimilitud es la probabilidad de sobrevivencia desde la edad $x + r_i$ a $x + t_i$, y si $\delta_i = 1$ la verosimilitud es la función de densidad de muerte a la edad $x + t_i$ dado que se está con vida a la edad $x + r_i$. La verosimilitud total es:

$$L = \prod_{i=i}^{n} {}_{t_i - r_i}p_{x+r_i} \left(\mu_{x+t_i}\right)^{\delta_i}. \tag{3.146}$$

Para evaluar (3.146) se debe hacer uso de un supuesto de distribución. De nuevo se considerarán dos casos.

Distribución exponencial para datos completos.

Bajo el supuesto exponencial (fuerza constante), se tiene

$$L = (\mu)^d \cdot \prod_{i=i}^{n} e^{-(t_i - r_i)\mu}, \tag{3.147}$$

(Ver [5], pág 152), entonces

$$\ell = \ln L = d \cdot \ln \mu - \mu \cdot \sum_{i=i}^{n} (t_i - r_i), \tag{3.148}$$

que al derivar e igualar a cero, da:

$$\widehat{\mu} = \frac{d_x}{\sum\limits_{i=i}^{n} (t_i - r_i)}. \tag{3.149}$$

(3.149) es el estimador máximo verosímil para el caso general con datos completos, y los casos especiales A, B y C están contenidos en él.

Distribución uniforme para datos completos

Bajo el supuesto de distribución uniforme

$${}_{t_i - r_i}p_{x+r_i} = 1 - {}_{t_i - r_i}q_{x+r_i} = 1 - \frac{(t_i - r_i)\, q_x}{1 - r_i \cdot q_x} = \frac{1 - t_i \cdot q_x}{1 - r_i \cdot q_x}, \tag{3.150}$$

(Ver [5], pág 152), y

$$\mu_{x+t_i} = \frac{q_x}{1 - t_i \cdot q_x}. \tag{3.151}$$

Cuando (3.150) y (3.151) son sustituidos en (3.146), se tiene que (3.146) contendrá q_x para cada muerto, $(1 - t_i \cdot q_x)$ para cada sobreviviente del intervalo $(t_i = 1)$ y para cada finalista observado $(t_i < 1)$, y $(1 - r_i \cdot q_x)^{-1}$ para todos. Entonces (3.146) se convierte en:

$$L = (q_x)^d \cdot \prod_{i=i}^{n} (1 - r_i \cdot q_x)^{-1} \cdot \prod_{S\&E} (1 - t_i \cdot q_x). \tag{3.152}$$

(S&E denota los sobrevivientes y finalistas). Entonces al derivar L con respecto a q e igualar a cero:

$$\frac{d_x}{q_x} + \sum_{i=i}^{n} \frac{r_i}{1 - r_i \cdot q_x} - \sum_{S\&E} \frac{t_i}{1 - t_i \cdot q_x} = 0. \tag{3.153}$$

En general, (3.153) se debe resolver para $\widehat{q}_x$ por iteración.

Si $r_i = 0$ para todas las n_x personas en la muestra y si $t_i = 1$ para todos los $n_x - d_x$ no muertos, entonces se tiene el caso especial A y se resuelve de manera lineal para el estimador (3.133).

Si $t_i = 1$ para todos los $n_x - d_x$ no muertos, $r_i = 0$ para las b_x de las n_x personas en la muestra, y $r_i > 0$ para los restantes $k_x = n_x - b_x$, entonces se tiene el caso especial B. Si $r_i = r$, una constante, para todas las k_x personas, entonces se tiene el caso especial B agrupado definido en la sección (3.1.4). En este caso (3.153) se convierte en:

$$\frac{d}{q} + k\left(\frac{r}{1 - r \cdot q_x}\right) - (n - d)\left(\frac{1}{1 - q}\right) = 0, \tag{3.154}$$

de donde se obtiene la ecuación cuadrática

$$r\,(n - k)\,q^2 - (n - rk + rd)\,q + d = 0, \tag{3.155}$$

que lleva al resultado:

$$\widehat{q}_x = \frac{b - \sqrt{b^2 - 4r\,(n_x - k_x)\,d_x}}{2r\,(n_x - k_x)}, \tag{3.156}$$

donde $b = n_x - rk_x + rd_x$. El radical negativo es usado en (3.156), por que el radical positivo llevaría a $\widehat{q}_x > 1$.

Similarmente, para el caso especial C agrupado con la edad común de salida $x + s$ para todos los finalistas programados. En este tiempo $r_i = 0$ para todas las n_x personas en la muestra, $t_i = s$ para todos los e_x finalistas observados y $t_i = 1$ para los $n_x - e_x - d_x$ sobrevivientes del intervalo. Entonces (3.153) se convierte en:

$$\frac{d}{q} - e\left(\frac{s}{1 - s \cdot q_x}\right) - (n - e - d)\left(\frac{1}{1 - q}\right) = 0, \tag{3.157}$$

que lleva al resultado:

$$\widehat{q}_x = \frac{b - \sqrt{b^2 - 4sn_x d_x}}{2sn_x}, \tag{3.158}$$

donde $b = n_x - (1-s)\, e_x + sd_x$. Se debe notar que (3.158) es similar a (3.156), donde r es reemplazado por s, k es reemplazado por $-e$ y $n-k$ reemplazado por n.

Caso especial C con datos incompletos

Se supone que todas las n_x personas quienes entran a $(x, x+1]$ a la edad exacta x (es decir, $r_i = 0$) y también se supone que c_x de ellos tienen una edad común (o en promedio) programada de salida $x+s$, $0 < s < 1$. Los restantes $n_x - c_x$ están programados a permanecer en la muestra a la edad $x+1$.

Se nota que c_x y $n_x - c_x$ son muestras binomiales separadas. Un miembro de c_x puede morir antes de la edad $x+s$, o puede sobrevivir hasta esta edad y salir de la observación; un miembro de $n_x - c_x$ puede morir antes de la edad $x+1$, o sobrevivir hasta esta edad y salir de la observación. Sea d'_x el número observado de muertes fuera de la muestra c_x, y sea d''_x el número observado de muertes fuera de la muestra $n_x - c_x$, por lo que el número total de muertes observadas es $d_x = d'_x + d''_x$. Se debe notar que $e_x = c_x - d'_x$ es el número observado de finalistas a la edad $x+s$.

La contribución a la verosimilitud de la muestra c_x es:

$$L_c = ({}_s q_x)^{d'} \cdot (1 - \ {}_s q_x)^{c-d'} \tag{3.159}$$

y la contribución de la muestra $n_x - c_x$ es:

$$L_{n-c} = (q_x)^{d''} \cdot (1 - \ q_x)^{n-c-d''}. \tag{3.160}$$

Donde los dos grupos son muestras binomiales independientes, por lo que la verosimilitud total es el producto de L_c y L_{n-c}. Entonces

$$L = ({}_s q_x)^{d'} \cdot (1 - \ {}_s q_x)^{c-d'} \cdot (q_x)^{d''} \cdot (1 - \ q_x)^{n-c-d''}. \tag{3.161}$$

(Ver [5], pág 155). Si la verosimilitud es evaluada bajo el supuesto uniforme, esto es ${}_s q_x = s \cdot q_x$, se tiene

$$L = (s)^{d'} \cdot (q_x)^{d'} \cdot (1 - s \cdot q_x)^{c-d'} \cdot (q_x)^{d''} \cdot (1 - \ q_x)^{n-c-d''}. \tag{3.162}$$

Eliminando la constante $(s)^{d'}$ y combinando los términos $q^{d'}$ y $q^{d''}$, se tiene que:

$$L = q^d \cdot (1 - s \cdot q)^{c-d'} \cdot (1 - \ q)^{n-c-d''}, \tag{3.163}$$

de donde se obtiene

$$\frac{d \ln L}{dq} = \frac{d}{q} - (c - d')\left(\frac{s}{1 - s \cdot q}\right) - (n - c - d'')\left(\frac{1}{1-q}\right) = 0. \tag{3.164}$$

Pero $c - d' = e$, y $n - c - d'' = n - c - (d - d') = n - d - e$, entonces se puede ver que (3.164) es lo mismo que (3.157) por lo que resulta el mismo estimador (3.158). Se tiene que para el caso especial C agrupado, el estimador máximo verosímil bajo el supuesto uniforme es el mismo para datos completos que para datos incompletos.

La verosimilitud dada por (3.161) es evaluada bajo el supuesto exponencial, con ${}_sp_x = (p_x)^s$, se tiene que:

$$L = (1 - p_x^s)^{d'} \cdot (p_x^s)^{c-d'} \cdot (q_x)^{d''} \cdot (1 - q_x)^{n-c-d''}. \tag{3.165}$$

Usando $e = c - d'$, se puede reescribir el segundo factor de (3.165) como $(p_x)^{se}$, y cuando se combina con el último factor se produce:

$$L = p^{se+n-c-d''} \cdot (1-p)^{d''} \cdot (1 - p^s)^{d'}, \tag{3.166}$$

del cual se puede obtener

$$\frac{d \ln L}{dp} = \frac{se + n - c - d''}{p} - \frac{d''}{1-p} - \frac{d' \cdot sp^{s-1}}{1 - p^s} = 0. \tag{3.167}$$

La ecuación dada por (3.167) se debe resolver por métodos numéricos, para el estimador máximo verosimil $\widehat{p}_x$ (y luego $\widehat{q}_x = 1 - \widehat{p}_x$).

Si la información disponible no incluye las muertes d'_x y d''_x por separado, o no muestra la subdivisión de n_x en c_x y $n_x - c_x$, entonces se crea una verosimilitud más complicada para datos incompletos. Se supone que los datos consisten solamente en el hecho de que fuera de las n_x personas a edad x, existen e_x finalistas observados a edad $x + s$, d_x muertes observadas, y $n_x - e_x - d_x$ sobrevivientes a edad $x + 1$.

La expresión

$$L = (q_x)^d \cdot (1 - q_x)^{n-e-d} \cdot (1 - \ {}_sq_x)^e \tag{3.168}$$

no es enteramente correcta para la verosimilitud, donde se sabe que no todos los d_x fueron muertos en el intervalo $(x, x+1]$ (no se sabe exactamente cuántos murieron y cuántos de los que han muerto en $(x, x+s]$ fueron observados).

La proporción desconocida de n_x de quienes están programados a salir a edad $x+s$ puede ser incorporada en la estructura aleatoria del modelo. Sea ϕ esta proporción desconocida; esto es, $c_x = \phi \cdot n_x$. Entonces la probabilidad de una muerte observada es $\phi \cdot \ {}_sq_x + (1-\phi) \cdot q_x$, y el primer término de la verosimilitud puede ser $[\phi \cdot_s q_x + (1-\phi) \cdot q_x]^d$.

Entonces se procede a encontrar el estimador máximo verosimil de ϕ, para despues encontrar el estimador máximo verosimil de q_x. Se debe notar que ϕ aparece solamente en el primer término de L. De donde $\ln L$ es una función decreciente de ϕ, bajo las distribuciones uniforme y exponencial, por lo que se sigue que $\ln L$, y L, es maximizada por $\phi = 0$.

Cuando se evalúa bajo el supuesto uniforme, es fácil ver que (3.168) es lo mismo que (3.163), y de nuevo resulta el estimador (3.158).

Cuando se evalúa bajo el supuesto exponencial, ocurre una pequeña simplificación. Se puede reescribir (3.168) como

$$L = (1 - p_x)^d \cdot (p_x)^{n-e-d} \cdot (p_x)^{se}, \tag{3.169}$$

y cuando se combinan los dos términos en p_x, se obtiene

$$L = p^{n-d-(1-s)e} \cdot (1-p)^d. \tag{3.170}$$

Entonces

$$\frac{d \ln L}{dp} = \frac{n - d - (1-s)\,e}{p} - \frac{d}{1-p} = 0, \tag{3.171}$$

se resuelve de manera lineal para obtener

$$\widehat{q_x} = \frac{d_x}{n_x - (1-s)\,e_x}. \tag{3.172}$$

(Ver [5], pág 157). Se puede ver que (3.172) es el estimador actuarial para el caso especial C, descrito anteriormente.

Caso especial C con censura aleatoria

Una manera alternativa de usar los valores exactos o promedios de s_i, para las c_x personas con $0 < s_i < 1$, es suponer que los valores de s_i son aleatorios y distribuidos sobre $(x, x+1]$. Entonces el tiempo programado de salida es una variable aleatoria S y $g(s)$ denota su función de densidad. En análisis de datos de supervivencia, donde la terminación de la observación causada por el término del periodo de análisis es llamada censura, esta estructura es conocida como un **mecanismo de censura aleatoria**.

Para construir la verosimilitud para el caso especial C con datos incompletos, se necesita la probabilidad de que una persona en c_x muera antes de la edad programada de salida. Si una edad programada de salida exacta no es usada, y se supone una distribución aleatoria para la edad programada de salida, entonces se debe calcular primero la probabilidad marginal de muerte antes de la edad programada de salida. Si $\overline{q}_x$ denota esta probabilidad marginal, entonces:

$$\overline{q}_x = 1 - \overline{p}_x = 1 - \int_0^1 g(s) \cdot \ {}_sp_x ds. \tag{3.173}$$

(Ver [5], pág 159). La integral en (3.173) muestra que la probabilidad marginal $\overline{p}_x$ es un gran valor promedio de ${}_sp_x$.

Para evaluar (3.173), se debe hacer un supuesto de distribución con respecto a la mortalidad y especificamente la distribución de S. Si se hace un supuesto lineal para la mortalidad y una distribución uniforme para S (donde $g(s) = 1$), entonces (3.173) se evalúa como

$$\overline{q}_x = \frac{1}{2} \cdot q_x. \tag{3.174}$$

El estimador máximo verosímil para el caso especial C es

$$L = (q_x)^{d''} \cdot (1 - q_x)^{n-c-d''} \cdot (\overline{q}_x)^{d'} \cdot (1 - \overline{q}_x)^{c-d'} \tag{3.175}$$

bajo censura aleatoria, como opuesto a (3.161) para la verosimilitud donde se supone un valor promedio (fijo) de s.

Resumen del estimador máximo verosimil en un medio con decremento simple

El estimador máximo verosimil para el caso general con datos completos y una distribución exponencial dado por (3.149), es aplicable para los casos especiales A, B y C.

Similarmente, el estimador (3.153) es para el caso general con datos completos con una distribución uniforme.

Para datos incompletos se desarrollaron resultados para el caso especial A y C bajo cada supuesto de distribución exponencial y uniforme. El estimador máximo verosimil para el caso especial B es derivado análogamente que el caso especial C. El estimador máximo verosimil en el caso general con datos parciales es considerablemente más complejo.

3.4.3 Medio de decremento doble

Si tanto la muerte como la retirada son eventos aleatorios en el intervalo $(x, x+1]$, entonces se tendrá que estimar $q_x = q_x^{\prime(d)}$ en un medio de doble decremento. Ahora se explorará el estimador de máxima verosimilitud en un medio de doble decremento.

Forma general para datos completos

Sea $x + r_i$ la edad en que la persona i entra en $(x, x+1]$, $0 \leq r_i < 1$, y n_x el número total de personas en la muestra. Sea $x+t_i$ la edad en que la persona i deja el intervalo $(x, x+1]$, $0 < t_i \leq 1$, ya sea como un sobreviviente del intervalo ($t_i = 1$), como un finalista observado ($t_i < 1$), o como resultado de uno de los eventos aleatorios (muerte o retirada).

Entonces cada persona bajo observación, de la edad $x + r_i$ a la edad $x + t_i$, tiene una probabilidad de no muerte, ni retirada de ${}_{t_i - r_i}p_{x+r_i}^{(\tau)}$. Para los sobrevivientes del intervalo y los finalistas, esta "sobrevivencia total" (es decir, no mueren ni se retiran) es la contribución a la verosimilitud. Para cada muerte y retirada, se necesitan las respectivas funciones de densidad, dadas por (3.40) y (3.41), para ser multiplicadas por la fuerza apropiada. Entonces la verosimilitud total es:

$$L = \prod_{i=1}^{n} {}_{t_i-r_i}p_{x+r_i}^{(\tau)} \cdot \prod_{D} \mu_{x+t_i}^{(d)} \cdot \prod_{W} \mu_{x+t_i}^{(w)} \tag{3.176}$$

$$= \prod_{i=1}^{n} {}_{t_i-r_i}p_{x+r_i}^{\prime(d)} \cdot \prod_{i=1}^{n} {}_{t_i-r_i}p_{x+r_i}^{\prime(w)} \cdot \prod_{D} \mu_{x+t_i}^{(d)} \cdot \prod_{W} \mu_{x+t_i}^{(w)}. \tag{3.177}$$

(Ver [5], pág 160). Se debe recordar la variable indicadora δ_i definida por (3.143). Ahora se introduce una segunda variable indicadora γ_i, donde

$$\gamma_i = \begin{cases} 1 & \text{si la persona } i \text{ se retira en } (x, x+1] \\ 0 & \text{en otro caso} \end{cases} \tag{3.178}$$

Usando estas dos variables indicadoras, se puede reescribir la verosimilitud como:

$$L = \prod_{i=1}^{n} {}_{t_i-r_i}p_{x+r_i}^{\prime(d)} \cdot \prod_{i=1}^{n} {}_{t_i-r_i}p_{x+r_i}^{\prime(w)} \cdot \left(\mu_{x+t_i}^{(d)}\right)^{\delta_i} \cdot \left(\mu_{x+t_i}^{(w)}\right)^{\gamma_i}, \tag{3.179}$$

donde $\mu_{x+t_i}^{(d)}$ es requerida en la función de verosimilitud solamente si la persona i es un muerto ($\delta_i = 1$), y $\mu_{x+t_i}^{(w)}$ es requerida solamente si la persona i es un retirado ($\gamma_i = 1$).

El objetivo es estimar $q_x^{\prime(d)}$. Suponiendo que la muerte y la retirada son independientes, entonces los términos que contengan (w) son constantes con respecto a $q_x^{\prime(d)}$, y pueden ser eliminados de la verosimilitud. Por lo anterior, se tiene:

$$L = \prod_{i=1}^{n} {}_{t_i-r_i}p_{x+r_i}^{\prime(d)} \cdot \left(\mu_{x+t_i}^{(d)}\right)^{\delta_i}. \tag{3.180}$$

Por lo que ${}_{t_i-r_i}p_{x+r_i}$ y $\mu_{x+t_i}^{(d)}$ en (3.146) son las mismas funciones de mortalidad que ${}_{t_i-r_i}p_{x+r_i}^{\prime(d)}$ y $\mu_{x+t_i}^{(d)}$ en (3.180). Así (3.180) es exactamente la misma que (3.146).

Datos completos, distribución exponencial para la mortalidad

Como (3.180) es la misma que (3.146), se sigue que la maximización de (3.180) bajo la distribución exponencial producirá el estimador de máxima verosimilitud para $\mu^{(d)}$, la fuerza constante sobre $(x, x+1)$, dado por (3.149).

Datos completos, distribución uniforme para la mortalidad

Bajo la distribución uniforme (3.180), que es la misma que (3.146), es maximizada para el valor de q_x la cual satisface que:

$$\frac{d_x}{q_x} + \sum_{i=1}^{n} \frac{r_i}{1 - r_i \cdot q_x} - \sum_{\overline{D}} \frac{t_i}{1 - t_i \cdot q_x} = 0, \tag{3.181}$$

donde la última suma es tomada sobre todas las personas que no mueren. La ecuación (3.180), al igual que en el decremento simple, se resuleve por iteración.

Datos incompletos, caso especial A

La estimación de $q_x'^{(d)}$ para datos incompletos en presencia de salidas aleatorias es más compleja que en el caso con datos completos. Ahora se considerará solamente el caso especial A, donde $r_i = 0$ y $s_i = 1$ para toda i quienes no mueren o no se retiran. Se supone que la única información disponible es la edad exacta x, el número de muertos d_x y el número de retirados w_x en $(x, x+1]$, entonces $n_x - d_x - w_x$ sobreviven a la edad $x+1$. La edad exacta de muerte y retirada no están disponibles.

La verosimilitud de esta muestra es:

$$L = \left[q_x^{(d)}\right]^{d_x} \cdot \left[q_x^{(w)}\right]^{w_x} \cdot \left[1 - q_x^{(d)} - q_x^{(w)}\right]^{n_x - d_x - w_x}, \tag{3.182}$$

(Ver [5], pág 161), donde $q_x^{(d)}$ y $q_x^{(w)}$ están definidas por (3.42) y (3.43), respectivamente. Para encontrar los estimadores de máxima verosimilitud para $q_x^{(d)}$ y $q_x^{(w)}$, primero se necesita escribir que:

$$\ln L = d_x \cdot \ln q_x^{(d)} + w_x \cdot \ln q_x^{(w)} + (n_x - d_x - w_x) \cdot \ln\left(1 - q_x^{(d)} - q_x^{(w)}\right). \tag{3.183}$$

Entonces se encuentra que

$$\frac{\partial \ln L}{\partial q_x^{(d)}} = \frac{d_x}{q_x^{(d)}} - \frac{n_x - d_x - w_x}{1 - q_x^{(d)} - q_x^{(w)}} = 0 \tag{3.184}$$

y

$$\frac{\partial \ln L}{\partial q_x^{(w)}} = \frac{w_x}{q_x^{(w)}} - \frac{n_x - d_x - w_x}{1 - q_x^{(d)} - q_x^{(w)}} = 0, \tag{3.185}$$

los cuales se resuelven simultáneamente, obteniendo:

$$\widehat{q}_x^{(d)} = \frac{d_x}{n_x} \tag{3.186}$$

resultado que coincide con (3.133) y

$$\widehat{q}_x^{(w)} = \frac{w_x}{n_x}. \tag{3.187}$$

Si el objetivo es estimar $q_x'^{(d)}$ y $q_x'^{(w)}$, en lugar de solamente $q_x^{(d)}$ y $q_x^{(w)}$, se deben adoptar nuevamente supuestos de distribuciones para los eventos aleatorios, muerte y retirada.

Datos incompletos (caso especial A), distribuciones exponenciales

De las ecuaciones (3.69) y (3.70), con $q_x^{(d)}$ y $q_x^{(w)}$ reemplazados por los estimadores dados por (3.186) y (3.187), se tiene directamente que:

$$\widehat{q}_x'^{(d)} = 1 - \left(\frac{n-d-w}{n}\right)^{d/(d+w)} \tag{3.188}$$

(Ver [5], pág 163), y

$$\widehat{q}_x'^{(w)} = 1 - \left(\frac{n-d-w}{n}\right)^{w/(d+w)}. \tag{3.189}$$

Se debe notar que los estimadores de máxima verosimilitud (3.188) y (3.189) son los mismos que los estimadores de momentos (3.118) y (3.119).

Datos incompletos (caso especial A), distribuciones uniformes

De las ecuaciones (3.56) y (3.57), con $q_x^{(d)}$ y $q_x^{(w)}$ reemplazados por los estimadores dados por (3.186) y (3.187), se tiene que:

$$\widehat{q}_x'^{(d)} = \frac{b - \sqrt{b^2 - 2n_x d_x}}{n_x}, \tag{3.190}$$

donde $b = n_x - \frac{1}{2}w_x + \frac{1}{2}d_x$, y

$$\widehat{q}_x'^{(w)} = \frac{b - \sqrt{b^2 - 2n_x w_x}}{n_x}, \tag{3.191}$$

donde $b = n_x - \frac{1}{2}d_x + \frac{1}{2}w_x$. Se debe notar que los estimadores de máxima verosimilitud (3.188) y (3.189) son los mismos que los estimadores de momentos (3.112) y (3.113).

3.4.4 El estimador producto-límite

Este estimador es frecuentemente llamado el **estimador de Kaplan-Meier**. El **estimador producto-límite** es estimador de máxima verosimilitud.

Una característica básica del estimador producto-límite es que no requiere un supuesto de distribución.

Estimación de q_x

Para el estimador producto-límite se tratarán a los finalistas y las salidas aleatorias como semejantes, por lo que a ambos se les llamarán **terminaciones**. La presencia de entrantes a cualquier edad es también considerada para la estimación.

Ahora se subdividirá el intervalo básico de estimación en subintervalos, partiendo a $(x, x+1]$ en cada punto donde ocurra una entrada o terminación. Se puede ver que cada subintervalo representa una situación del caso especial A, en donde no hay eventos en el subintervalo excepto (posiblemente) la muerte. Si q_i es la probabilidad condicional de muerte en el $i-ésimo$ subintervalo, dado que llegó con vida al principio de este subintervalo, entonces el estimador de máxima verosimilitud de q_i es la proporción binomial simple:

$$\widehat{q_i} = \frac{d_i}{n_i}, \tag{3.192}$$

donde d_i es el número observado de muertes en el $i-ésimo$ subintervalo, y n_i es el tamaño de la muestra binomial para este subintervalo. Claramente si $d_i = 0$, entonces $\widehat{q_i} = 0$.

El estimador producto-límite de q_x es definido como:

$$\widehat{q_x} = 1 - \prod_{i=1}^{m}(1 - \widehat{q_i}) = 1 - \prod_{i=1}^{m}\left(\frac{n_i - d_i}{n_i}\right), \tag{3.193}$$

(Ver [5], pág 168), donde m es el número de subintervalos en $(x, x+1]$.

La regla usual para cuando ocurra una muerte a la misma edad de una entrada o terminación, es que pertenezca al subintervalo precedente al evento. Por lo que es consistente a la notación general $(x, x+1]$.

Una alternativa a la partición de $(x, x+1]$ en cada entrada y terminación, podría ser una partición en cada punto de muerte. Sea r_j el conjunto de riesgo para el $j-ésimo$ subintervalo y está definido como el número de personas en la muestra, inmediatamente precedente a la muerte, la cual marcó el fin del $j-ésimo$ subintervalo. La probabilidad estimada de muerte en el $j-ésimo$ subintervalo es:

$$\widehat{q}_j = \frac{1}{r_j}, \tag{3.194}$$

excepto para el último subintervalo en $(x, x+1]$. Donde no hay muertos en este subintervalo, entonces $\widehat{q}_j$ para este subintervalo es cero. Finalmente

$$\widehat{q}_x = 1 - \prod_{j=1}^{m} (1 - \widehat{q}_j) = 1 - \prod_{j=1}^{m} \left(\frac{r_j - 1}{r_j} \right) \tag{3.195}$$

da el estimador para q_x.

En la partición de puntos de muerte, si una muerte y una entrada o una muerte y una terminación ocurren al mismo tiempo, se necesita interpretar el conjunto de riesgo "inmediatamente precedente" a la muerte y se considera la muerte como precedente a la entrada o terminación.

Si hay más de una muerte a la misma edad, entonces (3.194) es modificado a:

$$\widehat{q}_j = \frac{d_j}{r_j}, \tag{3.196}$$

donde d_j es el número de muertes simultáneas y que marcan el final del $j-ésimo$ subintervalo.

Propiedades del estimador

El estimador producto-límite es insesgado y consistente. Para la varianza del estimador, se considera que:

$$\widehat{p}_x = \prod_{i=1}^{m} \widehat{p}_i, \tag{3.197}$$

que viene directamente de (3.193). Se debe recordar que cada $\widehat{p}_i$ es una proporción binomial, condicional al tamaño de su muestra n_i. Entonces todas las $\widehat{p}_i$'s son condicionalmente independientes y con proporciones binomiales, por lo que son insesgadas. De pasos análogos a (3.127), se tiene que:

$$Var\left(\widehat{p}_x \mid \{n_i\}\right) = \left(p_x\right)^2 \cdot \left[\prod_{i=1}^{m}\left(\frac{q_i}{p_i n_i}+1\right)-1\right]. \tag{3.198}$$

Donde $\widehat{q}_x = 1-\widehat{p}_x$, por lo que sigue que (3.198) es también la $Var\left(\widehat{q}_x \mid \{n_i\}\right)$.

De los mismos pasos que se siguieron de (3.127) a (3.128), se obtiene la aproximación:

$$Var\left(\widehat{q}_x \mid \{n_i\}\right) = Var\left(\widehat{p}_x \mid \{n_i\}\right) \approx \left(p_x\right)^2 \cdot \sum_{i=1}^{m}\left(\frac{q_i}{p_i n_i}\right). \tag{3.199}$$

En cualquier caso, una estimación numérica para la varianza se obtiene sustituyendo los estimados $\widehat{p}_i$, $\widehat{q}_i$ y $\widehat{p}_x$ obtenidos de la muestra.

Para una muestra muy grande con muchas entradas y terminaciones, como en el caso de estudios de compañias aseguradoras, el estimador producto-límite no es generalmente usado por el gran número de subintervalos. Para estudios clínicos es muy usado, específicamente en situaciones donde no hay entradas después del tiempo $t = 0$.

Estimación de $S(t)$

Se supone que se tiene una muestra de n personas, todas observadas desde el tiempo $t = 0$. Si el estudio se detiene antes de que todos mueran, entonces se tienen terminaciones desde la muestra a varios valores de t. Si r_j es el conjunto de riesgo inmediatamente precedente del $j-\acute{e}simo$ punto de muerte, entonces se tiene que:

$$\widehat{q}_j = \frac{d_j}{r_j}.$$

Se debe notar que $\widehat{q}_j$ estima la probabilidad de muerte sobre el intervalo, al final del $j-\acute{e}simo$ punto de muerte, dado que llegó con vida al principio del intervalo. Entonces

$$\widehat{S}\left(t_m\right) = \prod_{j=1}^{m}\left(1-q_j\right) = \prod_{j=1}^{m}\left(\frac{r_j-d_j}{r_j}\right) \tag{3.200}$$

estima la probabilidad de sobrevivencia desde $t = 0$ a $t = t_m$, el tiempo del $m - \acute{e}simo$ punto de muerte. Para toda t se tiene que $t_m \leq t < t_{m+1}$, el estimador de $S(t)$ es el mismo que $\widehat{S}(t_m)$, donde no hay muertos en la muestra entre t_m y t. Entonces

$$\widehat{S}(t) = \prod_{j=1}^{m} \left(\frac{r_j - d_j}{r_j} \right) \quad \text{para } t_m \leq t < t_{m+1},\ m = 1, 2, \ldots \tag{3.201}$$

Como un caso especial, $\widehat{S}(t) = 1$ para $t < t_1$.

Análogamente a (3.199), la varianza de $\widehat{S}(t)$ es aproximada por:

$$Var\left[\widehat{S}(t) \mid \{r_j\}\right] \approx [S(t)]^2 \cdot \sum_{j=1}^{m} \left(\frac{q_j}{p_j r_j} \right), \tag{3.202}$$

para $t_m \leq t < t_{m+1}$. (3.202) es numéricamente estimado sustituyendo $\widehat{q}_j$ por q_j, $\widehat{p}_j$ por p_j y $\widehat{S}(t)$ por $S(t)$, entonces se obtiene:

$$\widehat{Var}\left[\widehat{S}(t) \mid \{r_j\}\right] \approx \left[\widehat{S}(t)\right]^2 \cdot \sum_{j=1}^{m} \left(\frac{d_j}{r_j\,(r_j - d_j)} \right), \tag{3.203}$$

donde $\widehat{q}_j = \frac{d_j}{r_j}$.

Finalmente, si no todos son terminantes (es decir, datos completos), entonces para toda j, $r_{j+1} = r_j - d_j$. Por lo que se puede ver que (3.200) llega a ser:

$$\widehat{S}(t_m) = \prod_{j=1}^{m} \left(\frac{r_{j+1}}{r_j} \right) = \frac{r_{m+1}}{r_1}, \tag{3.204}$$

(Ver [5], pág 172), donde $r_1 = n$, la muestra original, y r_{m+1} es el número de sobrevivientes antes del $(m+1) - \acute{e}simo$ punto de muerte, o justo después del $m - \acute{e}simo$ punto de muerte.

El estimador Nelson-Aalen

La función acumulativa de riesgo está definida como:

$$\Lambda(t) = \int_0^t \lambda(y)dy = -\ln S(t),$$

(Ver [5], pág 172), de donde, $S(t) = e^{-\Lambda(t)}$. Por lo que es razonable estimar $S(t)$ mediante $\Lambda(t)$ y así definir:

$$\widehat{S}(t) = e^{-\widehat{\Lambda}(t)}. \tag{3.205}$$

lo cual da la relación

$$\widehat{\Lambda}(t) = -\ln \widehat{S}(t), \tag{3.206}$$

para considerar de nuevo (3.200), el estimador producto-límite de $S(t)$, para $t_m \leq t < t_{m+1}$. Sustituyendo (3.200) en (3.206) se tiene que:

$$\begin{aligned} \widehat{\Lambda}(t) &= -\ln \left[\prod_{j=1}^{m} \left(\frac{r_j - d_j}{r_j} \right) \right] \\ &= -\sum_{j=1}^{m} \ln \left(1 - \frac{d_j}{r_j} \right), \quad t_m \leq t < t_{m+1}. \end{aligned} \tag{3.207}$$

Se puede recordar que $-\ln\left(1 - \frac{d_j}{r_j}\right) = \frac{d_j}{r_j} + \frac{1}{2}\left(\frac{d_j}{r_j}\right)^2 + \cdots$. Ignorando los términos cuadraticos y los términos pequeños de la suma, se obtiene la aproximación:

$$-\ln \left(1 - \frac{d_j}{r_j} \right) \approx \frac{d_j}{r_j}, \tag{3.208}$$

(Ver [5], pág 173), y se puede aproximar la función acumulativa por:

$$\widehat{\Lambda}(t) = \sum_{j=1}^{m} \frac{d_j}{r_j}, \quad t_m \leq t < t_{m+1}, \tag{3.209}$$

de esta forma, aproximar la función de sobrevivencia como:

$$\widehat{S}(t) \doteq \exp \left(-\sum_{j=1}^{m} \frac{d_j}{r_j} \right), \quad t_m \leq t < t_{m+1} \tag{3.210}$$

(Ver [5], pág 173), el cual es llamado el **estimador Nelson-Aalen**.

Si sólo hay una muerte en cada punto de muerte (es decir, $d_j = 1$ para toda j), y si no hay terminantes, entonces el estimador Nelson-Aalen es particularmente fácil de calcular. En este caso, $r_j = n - j + 1$, por lo que se tiene:

$$\widehat{\Lambda}(t) \doteq \sum_{j=1}^{m} \frac{1}{r_j} = \frac{1}{n} + \frac{1}{n-1} + \cdots + \frac{1}{n-m+1}, \qquad t_m \leq t < t_{m+1}.$$

Con esto termina lo relacionado con estimación de modelos de tabulares, para continuar con una pequeña aplicación de lo que se vio en este capítulo.

Capítulo 4

Aplicaciones actuariales tradicionales

En este capítulo se explorarán algunas aplicaciones de la teoría estudiada en los capítulos anteriores.

4.1. Edad actual

La edad exacta y_i es obtenida de la fecha de cumpleaños y de la fecha en que esa persona entra al periodo de observación.

4.1.1. Años decimales

Una forma conveniente de obtener las edades exactas es expresar esas fechas de eventos en términos de **años decimales**. Por lo que se puede obtener una tabla día a día, como la siguiente:

Fecha	Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
1	1	32	60	91	121	152	182	213	244	274	305	335
2	2	33	61	92	122	153	183	214	245	275	306	336
3	3	34	62	93	123	154	184	215	246	276	307	337
4	4	35	63	94	124	155	185	216	247	277	308	338
5	5	36	64	95	125	156	186	217	248	278	309	339
6	6	37	65	96	126	157	187	218	249	279	310	340
7	7	38	66	97	127	158	188	219	250	280	311	341
8	8	39	67	98	128	159	189	220	251	281	312	342
9	9	40	68	99	129	160	190	221	252	282	313	343
10	10	41	69	100	130	161	191	222	253	283	314	344
11	11	42	70	101	131	162	192	223	254	284	315	345
12	12	43	71	102	132	163	193	224	255	285	316	346
13	13	44	72	103	133	164	194	225	256	286	317	347
14	14	45	73	104	134	165	195	226	257	287	318	348
15	15	46	74	105	135	166	196	227	258	288	319	349
16	16	47	75	106	136	167	197	228	259	289	320	350
17	17	48	76	107	137	168	198	229	260	290	321	351
18	18	49	77	108	138	169	199	230	261	291	322	352
19	19	50	78	109	139	170	200	231	262	292	323	353
20	20	51	79	110	140	171	201	232	263	293	324	354
21	21	52	80	111	141	172	202	233	264	294	325	355
22	22	53	81	112	142	173	203	234	265	295	326	356
23	23	54	82	113	143	174	204	235	266	296	327	357
24	24	55	83	114	144	175	205	236	267	297	328	358
25	25	56	84	115	145	176	206	237	268	298	329	359
26	26	57	85	116	146	177	207	238	269	299	330	360
27	27	58	86	117	147	178	208	239	270	300	331	361
28	28	59	87	118	148	179	209	240	271	301	332	362
29	29		88	119	149	180	210	241	272	302	333	363
30	30		89	120	150	181	211	242	273	303	334	364
31	31		90		151		212	243		304		365

Tabla 4.1

Cualquier evento que ocurra el 29 de Febrero se supone que ocurrió el 28 de Febrero. La parte decimal de un año decimal es obtenida de dividir el dia del año entre 365.

4.1.2 Edad exacta

Se puede ver que la edad exacta se obtiene de restar el año decimal del cumpleaños al año decimal del evento.

Ejemplo 4.1 *Si la persona nació el* 11 *de Octubre de* 1953*, y entra al periodo de observación el* 1° *de Agosto de* 1992*, encontrar* y_i *la edad exacta en la cual empieza la observación.*

Solución: El 1° *de Agosto de* 1992 *es el año decimal* $1992 + 213/365 = 1992.58$*. Por lo que* $y_i = 1992.58 - 1953.78 = 38.80$*.*

(Ver [5], pág 227)

Para cada persona en la muestra, tres edades son importantes: la edad de entrada al estudio y_i, la edad programada de salida del estudio z_i, y la edad actual de salida del estudio, debido a la muerte o la retirada antes de la edad z_i, si alguna de las dos ocurre. Se denota la edad exacta de muerte como θ_i, y la edad exacta de retirada como ϕ_i, se puede ver que si $\theta_i = 0$ entonces la $i - ésima$ persona no es un muerto en estudio, y si $\phi_i = 0$ entonces la $i - ésima$ persona no es un retirado en estudio.

Cada persona en el estudio se le asigna un **vector edad** definido por $v_i' = [y_i, z_i, \theta_i, \phi_i]$ el cual contiene toda la información necesaria para el proceso de la contribución de la $i - ésima$ persona a los estimadores.

Ejemplo 4.2 *Un periodo de observación empieza a correr del* 1° *de Agosto de* 1992 *al* 31 *de Diciembre de* 1997*. La persona nació el* 11 *de octubre de* 1953*, y murió el* 12 *de abril de* 1996*. Encontrar el vector edad* v_i' *para esta persona.*

Solución: Del ejemplo 4.1 *se tiene* $y_i = 38.80$*, el* 31 *de diciembre es el año decimal* $1997+365/365 = 1998.00$*, por lo que la edad programada de salida es* $z_i = 1998-1953.78 = 44.22$*, el* 12 *de abril es el año decimal* $1996+102/365 = 1996.28$*, por lo que la edad de muerte es* $\theta_i = 1996.28 - 1953.78 = 42.50$*. La edad de salida es* $\phi_i = 0$*. Entonces* $v_i' = [38.80, 44.22, 42.50, 0]$*.*

(Ver [5], pág 228)

Para estimar q_x sobre el intervalo $(x, x+1]$, es necesario determinar la contribución de la $i - ésima$ persona. Se debe notar que si $y_i \geq x + 1$, entonces la $i - ésima$ persona entra después de la edad $x+1$ y no contribuye a $(x, x+1]$. Similarmente, si $z_i \leq x$, entonces la $i - ésima$ persona deja el estudio antes de la edad x y no contribuye a $(x, x+1]$.

Una vez que se ha eliminado el vector edad para las personas quienes no contribuyen a $(x, x+1]$, el siguiente paso es convertir cada vector edad en un **vector duración** para el intervalo $(x, x+1]$. Se denota el vector duración por $\mathbf{u}'_{i,x} = [r_i, s_i, \iota_i, \kappa_i]$, donde, necesariamente al menos ι_i ó κ_i debe de ser cero. El doble subíndice de $\mathbf{u}'$ identifica tanto la persona como el intervalo de estimación.

Las siguientes relaciones son para las personas quienes contribuyen a $(x, x+1]$.

$$r_i = \begin{cases} 0 & \text{si } y_i \leq x \\ y_i - x & \text{si } x < y_i < x+1 \end{cases} \tag{4.1}$$

$$s_i = \begin{cases} z_i - x & \text{si } x < z_i < x+1 \\ 1 & \text{si } z_i \geq x+1 \end{cases} \tag{4.2}$$

$$\iota_i = \begin{cases} 0 & \text{si } \theta_i = 0 \\ \theta_i - x & \text{si } x < \theta_i \leq x+1 \\ 0 & \text{si } \theta_i > x+1 \end{cases} \tag{4.3}$$

$$\kappa_i = \begin{cases} 0 & \text{si } \phi_i = 0 \\ \phi_i - x & \text{si } x < \phi_i \leq x+1 \\ 0 & \text{si } \phi_i > x+1 \end{cases} \tag{4.4}$$

Ejemplo 4.3 *Convertir el vector edad del ejemplo 4.3 en un vector duración para los intervalos de estimación* $(38, 39]$, $(39, 40]$ *y* $(42, 43]$.

Solución:

Para $(38, 39]$*, se tiene* $r_i = y_i - 38 = .80$*,* $s_i = 1$ *(por que* $z_i > 39$*),* $\iota_i = 0$ *(por que* $\theta_i > 39$*), y* $\kappa_i = 0$ *(por que* $\phi_i = 0$*); entonces* $u'_{i,38} = [.80, 1, 0, 0]$.

Para $(39, 40]$*,* $r_i = 0$ *(por que* $y_i < 39$*),* $s_i = 1$ *(por que* $z_i > 40$*),* $\iota_i = 0$ *(por que* $\theta_i > 40$*), y* $\kappa_i = 0$ *(por que* $\phi_i = 0$*); entonces* $u'_{i,39} = [0, 1, 0, 0]$.

Para el intervalo $(42, 43]$*,* $r_i = 0$ *(por que* $y_i < 42$*),* $s_i = 1$ *(por que* $z_i > 43$*),* $\iota_i = \theta_i - 42 = .50$ *y* $\kappa_i = 0$ *(por que* $\phi_i = 0$*); entonces* $u'_{i,42} = [0, 1, .50, 0]$*. Donde* $u'_{i,x}$ *no esta definido para para* $x < 38$ *ó para* $x > 42$*, por que la* $i-$*ésima persona no contribuye a ninguno de estos intervalos.*

(Ver [5], pág 229)

4.1.3 Cálculo de la exposición

Mucho de los estimadores tienen la forma general $\widehat{q}_x = \frac{\text{muertes}}{\text{exposición}}$. Los cuales incluyen el estimador de momentos (que utiliza la exposición programada), el estimador actuarial (que utiliza la exposición actuarial) y el estimador máximo verosimil para datos completos, bajo el supuesto exponencial (que utiliza la exposición exacta).

A los estimadores de esta forma se les pueden aplicar los datos básicos representados por $u'_{i,x}$ en el intervalo $(x, x+1]$. Se debe notar que el número de muertes observadas en $(x, x+1]$ es simplemente el número de vectores $u'_{i,x}$ con $\iota_i \neq 0$. Similarmente, el número de salidas observadas en $(x, x+1]$ es el número de vectores con $\kappa_i \neq 0$.

La exposición exacta sobre $(x, x+1]$ contribuida por la $i-ésima$ persona es:

$$(\text{exposicion exacta})_{i,x} = \begin{bmatrix} s_i \\ \iota_i \\ \kappa_i \end{bmatrix} - r_i, \tag{4.5}$$

donde $\begin{bmatrix} s_i \\ \iota_i \\ \kappa_i \end{bmatrix}$ representa el mínimo de s_i, ι_i, κ_i que exceda cero. En otras palabras, si $\iota_i = \kappa_i = 0$, entonces la $i-ésima$ persona no muere ni se retira en $(x, x+1]$, entonces se tiene $(\text{exposicion exacta})_{i,x} = s_i - r_i$. Pero si la persona muere en $(x, x+1]$, se tiene $\iota_i < s_i$ y $\kappa_i = 0$, entonces $(\text{exposicion exacta})_{i,x} = \iota_i - r_i$. Finalmente, si la persona se retira en $(x, x+1]$, se tiene $\kappa_i < s_i$ y $\iota_i = 0$, entonces $(\text{exposicion exacta})_{i,x} = \kappa_i - r_i$.

Para encontrar la exposición programada para la estimación de la probabilidad de mortalidad $q_x'^{(d)}$ bajo la aproximación de momentos de Hoem, se tiene que:

$$(\text{exposicion programada})_{i,x} = \begin{bmatrix} s_i \\ \kappa_i \end{bmatrix} - r_i, \tag{4.6}$$

donde κ_i es usado si la $i-ésima$ persona se retira en $(x, x+1]$, y s_i es usada en otro caso.

La exposición actuarial difiere de los otros dos tipos de exposición, ya que se tiene que:

$$(\text{exposicion actuarial})_{i,x} = \begin{bmatrix} s_i \\ \kappa_i \\ 1 \end{bmatrix} - r_i, \tag{4.7}$$

donde κ_i es usado si la $i-ésima$ persona se retira en $(x, x+1]$, s_i es usada si la $i-ésima$ persona no muere ni se retira en $(x, x+1]$, y 1 es usada si la $i-ésima$ persona muere en $(x, x+1]$.

Ejemplo 4.4 *Considerar el vector edad* $v_i' = [39.85, 40.75, 40.25, 0]$. *Encontrar (a) (exposicion exacta)*$_{i,40}$, *(b) (exposicion programada)*$_{i,40}$, *(c) (exposicion actuarial)*$_{i,40}$.

Solución: Primero se necesita convertir v_i' *a* $u_{i,40}' = [0, .75, .25, 0]$. *Entonces que tiene que:*

(a)(exposicion exacta)$_{i,40} = .25$, $(\iota_i < s_i)$;

(b)(exposicion programada)$_{i,40} = .75$

(c)(exposicion actuarial)$_{i,40} = 1$

(Ver [5], pág 230)

La exposición total sobre $(x, x+1]$ es sumar la contribución de todas las personas quienes contribuyen a $(x, x+1]$. Entonces:

$$(\text{exposición})_x = \sum_i (\text{exposición})_{i,x}, \tag{4.8}$$

donde (4.8) contiene los tres tipos de contribución.

4.1.4 Agrupamiento

En muchos casos una **edad promedio** puede sustituir la edad exacta de todas las personas en un cierto rango de edad. Por ejemplo,si se considera a todas las personas con edad exacta y_i entre x y $x+1$, la edad exacta puede ser sustituida por una edad comun y' para cada una de las y_i, frecuentemente se usa $y' = x + \frac{1}{2}$ como el valor promedio de las $y_i's$. Los entrantes del estudio han sido agrupados por la **edad del último cumpleaños**. Un segundo tipo de agrupamiento es el de **edad de calendario**. La edad de calendario está definida como la edad entera, obtenida en el cumpleaños en el mismo año en donde el evento toma lugar. Por ejemplo, si el cumpleaños de una persona es el 14 de Septiembre de 1960, y la fecha de retirada es el 26 de Junio de 1994, entonces la edad de calendario es 34. La edad de calendario se encuentra como:

$$CA = CYE - CYB, \tag{4.9}$$

donde CYE es el año del evento y CYB es el año del cumpleaños.

Se puede ver que una persona con edad de calendario w, tiene una edad exacta de evento que cae en $(w-1, w+1)$. La ventaja de un agrupamiento por edad de calendario es que una edad entera w puede sustituir varias edades fraccionales. Cuando las salidas son agrupadas por edad de calendario y se supone la ocurrencia de una edad entera w, entonces todas las salidas ocurren a los límites de los intervalos de estimación. La ocurrencia de salidas en los límites de los intervalos de estimación, simplifican el trabajo de estimación, creando problemas de estimación de casos especiales A.

4.2 Edades Aseguradas

Mucha gente no compra pólizas individuales de seguros, el día de su cumpleños, por lo que se tienen varias edades fraccionales. Así, una **edad asegurada** entera puede sustituir la edad actual de la fecha de la poliza. En otras palabras,

$$IA = \text{Edad actual más cercana al cumpleaños}, \tag{4.10}$$

donde IA es la edad asegurada.

4.2.1 Valuación del año del cumpleaños

Asignar una edad asegurada entera en la expedición de la poliza, implica que una fecha hipotética de cumpleños, llamada la **fecha asegurada de cumpleaños**, ha sustituido a la fecha actual de cumpleños. Claramente, el mes y día de esta fecha asegurada de cumpleaños es la misma que la fecha de expedición de la póliza. El año hipotético de cumpleaños, llamado la **valuación del año de cumpleños**(VYB), es entonces:

$$VYB = CYI - IA, \tag{4.11}$$

donde CYI es el año de calendario de la expedición de la póliza. Se puede ver que VYB es frecuentemente el mismo que CYB, pero solo puede ser un año más ó un año menos que CYB.

Ejemplo 4.5 *Una póliza de seguro es expedida el* 22 *de Agosto de* 1998. *Encontrar la* IA *y el* VYB*, si la edad actual de cumpleaños es (a)* 12 *de Enero de* 1978*, (b)* 4 *de Julio de* 1977.

Solución:

(a)Como el 22 *de Agosto de* 1998 *es el cumpleaños más cercano próximo al* 12 *de Enero de* 1999*, entonces* $IA = 21$. *Se tiene* $VYB = 1998 - 21 = 1977$*, el cual es un año menor que el* CYB *actual.*

(b)Como el 22 *de Agosto de* 1998 *es el cumpleaños más cercano pasado del* 4 *de Julio de* 1998*, entonces* $IA = 21$*. Se tiene* $VYB = 1998-21 = 1977$*, el cual es el mismo que el* CYB *actual.*

(Ver [5], pág 234)

4.2.2 Estudios Año a Año

La principal consecuencia de un estudio año a año con edades aseguradas, es que todas las personas entran al estudio a la edad entera y_i. Para los estudios de edad asegurada año a año, todos los vectores $u'_{i,x}$ son de la forma $[0, 1, \iota_i, \kappa_i]$, por lo que producen casos especiales A.

Si no existen salidas, entonces el estimador $\widehat{q}_x$ se obtiene de contar el número de vectores $u'_{i,x}$ con $r_i = 0$, que da n_x, y el número con $\iota_i \neq 0$ que da d_x, para $(x, x+1]$. Si existen salidas, el número de ellas en $(x, x+1]$, w_x es el número de vectores $u'_{i,x}$ con $\kappa_i \neq 0$. Con n_x, d_x y w_x resueltos se puede obtener $q_x'^{(d)}$ y $q_x'^{(w)}$ por (3.112), (3.113), (3.118) y (3.119). Se debe notar que para todos los $s_i = 1$, el estimador de momentos y el estimador actuarial es el mismo.

Otra aproximación que es frecuentemente usada cuando las salidas están presentes en un estudio con edad asegurada año a año, es el grupo de salidas por **edad calendárica asegurada**. Esta es similar a al edad calendárica actual, excepto que VYB es usada en vez de CYB. Esto es:

$$CIA = VYW - VYB, \tag{4.12}$$

donde CIA es la edad calendárica asegurada , y CYW es el año calendárico de retirada. Entonces todas las salidas con $CIA = w$, están asignadas a $\theta_i = w$ en lugar de sus edades actuales ϕ_i.

Ejemplo 4.6 *Para la siguiente muestra de* 10 *pólizas, estimar* q_{30} *bajo las siguientes condiciones:*

1)El periodo de observación empieza a correr año a año de 1993 *a* 1998*.*

2)Las edades aseguradas son usadas desde el principio.

3)Las salidas están agrupadas por CIA*.*

Solución: Primero se asigna una IA *y una* VYB *a cada perona. Con la* VYB*, la fecha de expedición de la poliza y las fechas del periodo de observación se obtiene una* y_i *y* z_i *para todas las personas en el estudio. La*

Persona	Fecha de cump.	Fecha de exp de pól	Fecha de muerte	Fecha de retirada
1	17-Mar-64	20-Jun-92	***	***
2	06-May-64	06-Ago-92	12-Jun-93	***
3	12-Ago-64	18-Dic-92	***	18-Jun-95
4	27-Oct-64	04-Jun-93	***	***
5	04-Jun-65	28-Abr-93	29-Ago-96	***
6	18-Abr-65	16-Jun-93	***	12-Dic-95
7	20-May-65	29-Oct-93	21-Abr-96	***
8	04-Jul-65	16-Feb-94	***	***
9	16-Sep-65	22-Ago-94	***	22-Feb-97
10	11-Dic-65	06-Mar-95	17-Feb-97	***

Figure 4.1:

suposición de agrupamiento de salidas, implica que un ϕ_i entero reemplaza el valor fraccional exacto. Las edades exactas de muerte θ_i son fraccionales. Con el vector v_i se puede obtener el vector duración $u_{i,30}$. Los datos se resumen en la siguiente tabla:

Persona	IA	VYB	y_i	z_i	θ_i	ϕ_i	r_i	s_i	ι_i	κ_i
1	*28*	*1964*	*29*	*34*	*0*	*0*	*0*	*1*	*0*	*0*
2	*28*	*1964*	—	—	—	—	—	—	—	—
3	*28*	*1964*	*29*	*34*	*0*	*31*	*0*	*1*	*0*	*1*
4	*28*	*1965*	*28*	*33*	*0*	*0*	*0*	*1*	*0*	*0*
5	*28*	*1965*	*28*	*33*	*31.34*	*0*	*0*	*1*	*0*	*0*
6	*28*	*1965*	*28*	*33*	*0*	*30*	—	—	—	—
7	*28*	*1965*	*28*	*33*	*30.48*	*0*	*0*	*1*	*.48*	*0*
8	*29*	*1965*	*29*	*33*	*0*	*0*	*0*	*1*	*0*	*0*
9	*29*	*1965*	*29*	*33*	*0*	*32*	*0*	*1*	*0*	*0*
10	*29*	*1966*	*29*	*32*	*30.95*	*0*	*0*	*1*	*.95*	*0*

Se debe notar que la persona 2 *y* 6 *no contribuyen a* $(30, 31]$*. Por que la persona* 2 *no entra al estudio, ya que muere antes del periodo de observación, y la persona* 6 *se retira a la edad* $CIA = 30$*, es decir no entra a* $(30, 31]$*. Por lo que* $n_{30} = 8$ *personas que entran a al edad de* 30 *años (es decir,* $r_i = 0$*) y permanecen hasta la edad* 31 *(es decir,* $s_i = 1$ *y* $\kappa_i \not< s_i$*), excepto para* $d_{30} = 2$ *(las personas* 7 *y* 10*). Por lo que se usa el estimador* (3.77)*, y se obtiene* $\widehat{q}_{30} = \frac{2}{8} = .25$*.*

(Ver [5], pág 236)

Ejemplo 4.7 *Repetir el ejemplo 4.8, para calcular el estimador máximo verosimil de datos completos para* q_{30}*, suponiendo una fuerza constante de mortalidad sobre* $(30, 31]$.

Solución: En este caso la exposición exacta es 1 *para las* 6 *personas quienes no mueren en* $(30, 31]$*,* .48 *para la persona* 7*, y* .95 *para la persona* 10*. Entonces* $\widehat{\mu} = \frac{2}{7.43}$*, por lo que* $\widehat{q}_{30} = 1 - e^{-2/7.43} = .23599$.

(Ver [5], pág 237)

4.2.3 Estudios selectos

Para la estimación de $S(t; x)$, se consideran únicamente las pólizas que se expidieron a la $IA = x$. Se supone un periodo de observación año a año, la fecha es similar a la descrita para el estudio de edad asegurada, con las siguientes características:

1) El vector v_i, ahora representa la **duración de la poliza** desde la entrada hasta la salida programada, muerte o retirada, en lugar de la edad asegurada. Se debe notar que para las pólizas expedidas durante el periodo de observación, la duración al entrar al estudio es 0.

2) El vector $u_{i,t}$, representa el lugar de esos eventos en el intervalo de estimación $(t, t+1]$.

3)Las salidas están agrupadas por la **duración de calendario**, en lugar de la edad asegurada calendárica. Se debe notar que la duración calendárica esta definida por:

$$CD = CYW - CYI. \tag{4.13}$$

El siguiente ejemplo mostrará la similitud de los estudios selectos a los estudios con edad asegurada.

Ejemplo 4.8 *Considerar las pólizas del ejemplo 4.8 con* $IA = 28$*. Estimar* $q_{[28]+2}$ *usando un periodo de observación que empieza en* 1994 *y termina en* 1999*, agrupando las salidas por duración calendárica.*

Solución: Del ejemplo 4.8, de las pólizas 1 *a* 7 *tienen* $IA = 28$*, y la póliza* 2 *no es parte del estudio por que muere antes de que se abra el periodo de observación. Las otras pólizas tienen un vector duración* v_i*, como sigue:*

Póliza	y_i	z_i	θ_i	ϕ_i
1	*2*	*7*	*0*	*0*
3	*2*	*7*	*0*	*3*
4	*1*	*6*	*0*	*0*
5	*1*	*6*	*3.34*	*0*
6	*1*	*6*	*0*	*2*
7	*1*	*6*	*2.48*	*0*

Para el intervalo de estimación $(2,3]$*, todas las pólizas entran a* $t = 2$ *(es decir,* $r_i = 0$*), excepto la póliza* 6 *que no contribuye a* $(2,3]$*. Todas las pólizas estan programadas a completar* $(2,3]$*, excepto la póliza* 7*. Usando el estimador máximo verosimil bajo el supuesto uniforme , se tiene que:* $\widehat{q}_{[28]+2} = \frac{1}{5} = .20$*. Usando el supuesto exponencial, se tiene:* $\widehat{q}_{[28]+2} = 1 - e^{-1/4.48} = .20005$*.*

(Ver [5], pág 238)

4.3 Edades Fiscales

En los casos de seguros grupales ó planes de pensiones grupales, un gran número de personas están bajo una sola póliza o plan. Entonces existe una fecha clave para cada plan, llamado **plan anual** ó **fecha del plan de valuación**. Todos los miembros del plan tienen una edad integral en esta fecha clave. Esta edad es llamada **edad fiscal**.

Esta situación es similar a las edades aseguradas, donde cada seguro individual tiene una edad asegurada entera en la póliza anual. Para todas las personas en un plan anual se tiene la **fecha T** que es la misma para todas las personas.

4.3.1 Año de cumpleaños fiscal

Análogamente a la definición de edad asegurada y valuación del año de cumpleaños, se asigna a cada persona en un plan grupal una edad fiscal (FA) con alguna fecha T particular. Entonces se define el **año de cumpleaños fiscal** (FYB) como:

$$FYB = z - FA. \tag{4.14}$$

FYB puede ser el mismo que el de CYB, ó un año menos ó un año más. Una vez que el FYB ha sido asignado, la fecha T en el FYB es entonces una fecha hipotética para el cumpleaños para cada persona en el plan grupal.

4.3.2 Periodo de observación para estudios de edad fiscal

La forma que corre un periodo de observación, es desde una fecha T en un cierto año a una fecha T un año más tarde. Se debe notar que un estudio T a T es tanto un estudio fecha a fecha, como un estudio año a año.

El principal beneficio de usar un periodo de observación T a T es que para todos los miembros en el plan, cuando entran en el periodo de observación entran a la edad entera y_i. Similarmente todas están programadas a salir a la edad z_i. Por lo que y_i y z_i implica que $r_i = 0$ y $s_i = 1$ para cualquier intervalo de estimación $(x, x+1]$, y por tanto tener problemas de estimación casos especiales A.

4.3.3 Nuevos miembros y salidas

Pueden entrar trabajadores nuevos en un plan grupal bajo un estudio en una fecha T, entonces estos nuevos trabajadores pueden entrar al estudio a una edad fiscal entera. Similarmente, las personas pueden retirarse del plan grupal por terminar su trabajo a cualquier fecha y a cualquier edad fiscal fracccional. Se recuerda que y_i es la edad en que la persona entra al estudio, ya sea como un nuevo miembro cuando se abre el periodo de observación (en este caso y_i es un entero), o como un nuevo miembro durante el periodo de observación (en este caso y_i puede ser fraccional).

Si se utiliza un agrupamiento **edad de calendario fiscal**, entonces una edad fiscal entera puede sustituir las edades fiscales fraccionales. Si se utiliza un agrupamiento **edad último cumpleaños fiscal**, entonces se utliza usualmente $y + \frac{1}{2}$ para todas las edades exactas fraccionales en $(y, y+1]$.

Ejemplo 4.9 *Un plan de seguros de vida grupal tiene una fecha de aniversario el* 30 *de Junio. De la siguiente muestra, estimar* q_{40} *usando un periodo de observación que corre desde el* 30 *de Junio de* 1990*, hasta el* 30 *de Junio de* 1995*. Agrupar los nuevos miembros y salidas por la edad último cumpleaños fiscal, suponer una edad promedio de* $y + \frac{1}{2}$*, y usar el estimador actuarial.*

Miem	*FYB*	*Fec. de Afi. al plan*	*Fec. de Muerte*	*Fec. de Retirada*
1	*1952*	*12/May/89*	—	*30/Sep/93*
2	*1952*	*24/Ago/89*	—	—
3	*1951*	*03/Oct/89*	*17/Mar/94*	—
4	*1951*	*30/Jun/90*	—	*30/Abr/90*
5	*1951*	*18/May/90*	—	—
6	*1950*	*03/Jun/91*	—	—
7	*1951*	*15/Jul/90*	—	*15/Feb/92*
8	*1951*	*07/Abr/92*	*22/Jun/92*	—
9	*1953*	*15/Sep/94*	—	—
10	*1953*	*01/Jul/95*	—	—

Solución: Se nota que el miembro 4 *se retira del plan antes de que el periodo de observación se abra; y el miembro* 10 *no entra hasta despues de que el periodo de observación se cierra, por lo que ninguno de estos miembros estan en el estudio. Los miembros del* 6 *al* 9 *son nuevos miembros durante el periodo de observación. La siguiente tabla muestra la edad fiscal en que entra al estudio, la edad programada fiscal para salir del estudio, la edad último cumpleaños fiscal a la muerte y el agrupamiento edad fiscal para las salidas.*

Miembro	y_i	z_i	θ_i	ϕ_i
1	*38*	*43*	*0*	*41.5*
2	*38*	*43*	*0*	*0*
3	*39*	*44*	*42*	*0*
5	*39*	*44*	*0*	*0*
6	*40.5*	*45*	*0*	*0*
7	*39.5*	*44*	*0*	*40.5*
8	*40.5*	*44*	*40*	*0*
9	*41.5*	*42*	*0*	*0*

El miembro 9 *no contribuye a* $(40,41]$*, ya que* $y_i > 41$*. Los miembros* 6 *y* 8 *entran a* $(40,41]$ *a la edad* 40.5 *y los otros* 5 *entran a la edad* 40*. El miembro* 7 *deja el estudio a la edad* 40.5 *los otros los dejan a la edad* 41*, excepto el miembro* 8*, quien muere en* $(40,41]$*. Pero bajo el estimador actuarial, el miembro* 8 *esta expuesto a la edad* 41*. Por lo que se tiene* $\widehat{q}_{40} = \frac{1}{5+2\left(\frac{1}{2}\right)-\frac{1}{2}} = \frac{1}{5.5} = .18182$*. (Ver [5], pág 241)*

4.4 Aplicación final

En la siguiente muestra, tomando en cuenta un periodo de observación del 1° de Enero de 1990 al 31 de Diciembre de 1999.

Num de Pers	Fec de Cumpl	Fec de entrada	Fec de retirad	Fec de muerte
1	12-Abr-60	16-Ene-88	***	***
2	02-Jul-56	29-Mar-88	***	04-Ene-90
3	11-Jun-58	01-Ago-88	01-Dic-90	***
4	28-Feb-54	20-Ago-88	27-Ago-89	***
5	28-Nov-61	11-Ene-89	***	25-Dic-91
6	06-Jun-51	06-Ene-89	03-May-89	***
7	15-Mar-57	09-Dic-88	***	***
8	24-Ago-58	03-May-89	04-Mar-90	***
9	25-Jun-60	31-Ene-90	31-Dic-90	***
10	23-Feb-59	22-Sep-89	***	10-May-98
11	15-Dic-57	08-Oct-89	16-May-91	***
12	06-Jul-58	18-Feb-90	10-Ago-99	***
13	10-Jun-60	05-Abr-90	***	***
14	19-Feb-59	02-Jun-90	***	14-Abr-96
15	15-Oct-60	24-Nov-90	10-May-98	***
16	14-Abr-61	10-Feb-91	***	***
17	10-Ago-58	20-Dic-90	***	10-Oct-97
18	08-Oct-61	07-Jul-91	06-Abr-96	***
19	26-Oct-59	26-May-92	***	***
20	05-Feb-60	16-Dic-91	25-Ago-94	***
21	14-Dic-61	12-Dic-91	***	06-Mar-97
22	10-Dic-60	02-May-92	***	***
23	31-Jul-62	22-Abr-92	***	***
24	24-Ene-61	09-Sep-92	09-Sep-93	***
25	18-Jul-63	28-Ene-93	16-Nov-99	***
26	02-Oct-62	14-Oct-92	***	14-Nov-94
27	05-Sep-60	24-Ago-92	04-Jun-97	***
28	21-Mar-61	12-Nov-92	***	02-Dic-97
29	10-Abr-63	04-Mar-93	***	***
30	25-May-62	30-Dic-92	21-Jun-95	***
31	06-Ago-62	26-May-93	***	***
32	20-Nov-63	13-Jun-93	***	***
33	10-Dic-61	06-Abr-93	***	15-Feb-98
34	04-Ago-63	02-Sep-93	19-Ago-94	***
35	10-Sep-63	28-Feb-94	***	***

Num de Pers	Fec de Cumpl	Fec de entrada	Fec de retirad	Fec de muerte
36	30-Jun-62	01-Nov-93	10-Mar-98	***
37	28-Feb-63	15-Jul-94	***	16-Jul-97
38	31-Ene-64	08-Ago-94	***	***
39	05-Sep-62	02-Nov-94	06-Ago-96	***
40	02-Nov-63	21-Ene-95	***	***
41	15-Jul-65	16-Mar-95	***	***
42	20-Abr-64	04-May-95	***	20-Ene-96
43	27-Jul-65	17-Sep-95	06-Jul-98	***
44	03-Dic-65	10-Abr-96	***	***
45	04-Ene-64	23-Dic-95	***	***
46	06-Nov-64	10-Mar-96	21-Jun-96	***
47	19-May-66	12-Jul-96	12-Jul-97	***
48	16-Ago-66	06-Jul-96	06-Jul-99	***
49	04-Ene-65	19-Ago-96	***	***
50	10-Mar-66	25-Jun-96	***	24-Sep-98
51	24-Sep-66	07-Feb-97	21-Dic-98	***
52	15-Nov-66	14-Abr-97	***	***
53	06-Sep-64	08-Dic-96	***	04-Jun-98
54	14-Sep-67	21-May-97	21-Sep-97	***
55	21-Nov-65	27-Ago-97	***	***
56	19-Ago-67	16-Sep-97	***	***
57	06-Ene-67	09-Ago-97	***	***
58	06-Abr-67	30-Nov-97	***	19-Ago-98
59	04-Ago-65	23-Oct-97	23-Nov-97	***
60	03-Feb-66	04-Dic-97	***	***
61	27-Abr-68	04-Ene-98	04-Jun-98	***
62	26-Feb-67	10-May-98	***	***
63	13-Ago-68	25-Jul-98	***	05-Sep-99
64	06-Jul-66	16-Sep-98	19-Nov-99	***
65	21-May-69	03-Feb-99	***	***
66	15-Sep-66	12-Dic-98	12-Dic-99	***
67	08-Nov-67	27-Abr-99	***	***
68	21-Jul-69	05-Ago-99	10-Ene-00	***
69	16-Feb-68	18-Oct-99	***	***
70	13-Nov-66	14-Ene-00	***	***

Se determino el vector edad actual v_i para cada miembro de la muestra, donde *** indica quienes no pertenecen al estudio.

i	y_i	z_i	θ_i	φ_i
1	29.72	39.721	0	0
2	33.5	43.499	33.51	0
3	31.56	41.556	0	32.47
4	***	***	***	***
5	28.09	38.09	30.07	0
6	***	***	***	***
7	32.8	42.797	0	0
8	31.35	41.353	0	31.53
9	29.6	39.518	0	30.52
10	30.85	40.852	39.21	0
11	32.04	42.044	0	33.42
12	31.62	41.488	0	41.1
13	29.82	39.559	0	0
14	31.28	40.863	37.15	0
15	30.11	39.211	0	37.57
16	29.83	38.715	0	0
17	32.36	41.392	39.17	0
18	29.75	38.23	0	34.49
19	31.58	40.181	0	0
20	31.86	39.901	0	34.55
21	29.99	38.047	35.22	0
22	31.39	39.058	0	0
23	29.73	37.419	0	0
24	31.62	38.934	0	32.62
25	29.53	36.455	0	36.33
26	30.04	37.247	32.12	0
27	31.97	39.321	0	36.75
28	31.65	38.781	36.7	0
29	29.9	36.726	0	0
30	30.6	37.603	0	33.07
31	30.8	37.403	0	0
32	29.56	36.112	0	0
33	31.32	38.058	36.18	0
34	30.08	36.408	0	31.04
35	30.47	36.307	0	0

i	y_i	z_i	θ_i	φ_i
36	31.34	37.504	0	35.69
37	31.38	36.838	34.38	0
38	30.52	35.915	0	0
39	32.16	37.321	0	33.92
40	31.22	36.162	0	0
41	29.67	34.463	0	0
42	31.04	35.699	31.75	0
43	30.14	34.43	0	32.94
44	30.35	34.077	0	0
45	31.97	35.989	0	0
46	31.34	35.151	0	31.62
47	30.15	33.619	0	31.15
48	29.89	33.375	0	32.89
49	31.62	34.989	0	0
50	30.29	33.811	32.54	0
51	30.37	33.268	0	32.24
52	30.41	33.126	0	0
53	32.25	35.318	33.74	0
54	29.68	32.296	0	30.02
55	31.76	34.11	0	0
56	30.08	32.367	0	0
57	30.59	32.984	0	0
58	30.65	32.737	31.37	0
59	32.22	34.408	0	32.3
60	31.83	33.907	0	0
61	29.69	31.679	0	30.1
62	31.2	32.844	0	0
63	29.95	31.384	31.06	0
64	32.2	33.488	0	33.37
65	29.71	30.614	0	0
66	32.24	33.293	0	33.24
67	31.47	32.145	0	0
68	30.04	30.447	0	0
69	31.67	31.871	0	0
70	***	***	***	***

Se encontro el valor de d_x, el número observado de muertes en $(x, x+1]$, para los valores de x para los cuales $d_x \neq 0$.

x	d_x
30	1
31	3
32	2
33	2
34	1
35	1
36	2
37	1
39	2

Para todos los valores de x que tienen exposición, se evaluo la: a)exposición exacta, b)exposición programada, c)exposición actuarial.

intervalo	exp. Exa	exp. Pro	exp. Act
(28,29]	0.91	0.91	0.91
(29,30]	4.98	4.98	4.98
(30,31]	24.23	25.16	25.16
(31,32]	35.32	36.52	37.14
(32,33]	44.19	45.53	45.53
(33,34]	35.81	36.56	36.56
(34,35]	28.06	28.68	28.68
(35,36]	23.82	24.6	24.6
(36,37]	16.27	17.39	17.39
(37,38]	10.54	11.39	11.39
(38,39]	8.72	8.72	8.72
(39,40]	4.72	6.34	6.34
(40,41]	2.18	2.18	2.18
(41,42]	1.1	1.1	1.1
(42,43]	0.8	0.8	0.8

Para todos los valores de x para los cuales $d_x \neq 0$, se estimó q_x con las siguientes características:

a)Estimador máximo verosimil para datos completos con un supuesto exponencial.

b)Agrupamiento edad último cumpleaños para los nuevos miembros y salidas durante el periodo de observación.

x	a):qx EM	b):qx ag
30	0.04043	0.04
31	0.08143	0.0458
32	0.04425	0.04598
33	0.05432	0.05479
34	0.03501	0.03448
35	0.04111	0.04167
36	0.11567	0.11111
37	0.09051	0.08333
39	0.3454	0.30769

De la muestra se supondrá que cada persona tiene un seguro individual de vida y la fecha de entrada es la decha en que se expidió la póliza. El periodo de observación corre de 1990 a 1999 y así, se encontró el vector v_i para cada miembro, usando la edad asegurada de calendario en lugar de la edad exacta, para ϕ_i.

i	y_i	z_i	θ_i	φ_i
1	30	39	0	0
2	***	***	***	***
3	32	41	0	32
4	***	***	***	***
5	28	37	29.95	0
6	***	***	***	***
7	34	43	0	0
8	***	***	***	***
9	30	39	0	30
10	32	41	39.63	0
11	33	42	0	34
12	32	41	0	41
13	30	39	0	0
14	31	40	36.87	0
15	30	39	0	38
16	30	38	0	0
17	32	41	38.81	0
18	30	38	0	35
19	32	40	0	0
20	32	40	0	35
21	30	38	35.23	0
22	31	38	0	0
23	30	37	0	0
24	32	39	0	33
25	30	36	0	36
26	30	37	32.08	0
27	32	39	0	37
28	32	39	37.05	0
29	30	36	0	0
30	31	38	0	34
31	31	37	0	0
32	30	36	0	0
33	31	37	35.86	0
34	30	36	0	31
35	30	35	0	0

i	y_i	z_i	θ_i	φ_i
36	31	37	0	36
37	31	36	34	0
38	31	36	0	0
39	32	37	0	34
40	31	35	0	0
41	30	34	0	0
42	31	35	31.72	0
43	30	34	0	33
44	30	33	0	0
45	32	36	0	0
46	31	34	0	31
47	30	33	0	31
48	30	33	0	33
49	32	35	0	0
50	30	33	32.25	0
51	30	32	0	31
52	30	32	0	0
53	32	35	33.49	0
54	30	32	0	30
55	32	34	0	0
56	30	32	0	0
57	31	33	0	0
58	31	33	31.72	0
59	32	34	0	32
60	32	34	0	0
61	30	31	0	30
62	31	32	0	0
63	30	31	0	0
64	32	33	0	33
65	***	***	***	***
66	32	33	0	33
67	***	***	***	***
68	***	***	***	***
69	***	***	***	***
70	***	***	***	***

Se calculó d_x y n_x para todas las x para las cuales $n_x \neq 0$.

x	d_x	n_x
28	0	1
29	1	1
30	0	23
31	2	32
32	2	43
33	1	35
34	1	29
35	2	23
36	1	15
37	1	11
38	1	7
39	1	4
40	0	2
41	0	1
42	0	1

Para cada edad asegurada 30, 31 y 32, se calculó $d_{[x]+t}$ y $n_{[x]+t}$ usando un agrupamiento de duración de calendario para las salidas.

	X=30		X=31		X=32	
t	$d_{[x]+t}$	$n_{[x]+t}$	$d_{[x]+t}$	$n_{[x]+t}$	$d_{[x]+t}$	$n_{[x]+t}$
0	0	22	2	13	0	15
1	0	18	0	11	1	13
2	2	16	0	10	0	9
3	0	11	1	9	0	7
4	0	10	1	7	0	6
5	1	8	1	4	1	5
6	0	4	0	2	1	4
7	0	3	0	1	0	3
8	0	1	1	1	0	2
9					0	1
10					0	1

De la muestra se supondrá que cada miembro pertenece a un plan grupal de seguro. Se usó una edad fiscal como la edad actual cercana al 30 de Junio siguiente a la fecha de entrada. Para un periodo de observación del 30 de Junio de 1992 al 30 de Junio del 2000. Para encontrar el conjunto de vector edad v_i para cada miembro de la muestra.

i	y_i	z_i	θ_i	φ_i
1	32	40	0	0
2	***	***	***	***
3	***	***	***	***
4	***	***	***	***
5	***	***	***	***
6	***	***	***	***
7	35	43	0	0
8	***	***	***	***
9	***	***	***	***
10	33	41	38.86	0
11	***	***	***	***
12	34	42	0	41.11
13	32	40	0	0
14	33	41	36.79	0
15	32	40	0	37.86
16	31	39	0	0
17	34	42	39.28	0
18	31	39	0	34.77
19	33	41	0	0
20	32	40	0	34.15
21	31	39	35.68	0
22	32	40	0	0
23	30	38	0	0
24	31.195	39	0	32.19
25	29.581	37	0	36.38
26	30.301	38	32.38	0
27	32.151	40	0	36.93
28	31.37	39	36.42	0
29	29.677	37	0	0
30	30.501	38	0	32.98
31	30.904	38	0	0
32	29.953	37	0	0
33	31.767	39	36.63	0
34	30.175	37	0	31.14
35	30.666	37	0	0

i	y_i	z_i	θ_i	φ_i
36	31.34	38	0	35.69
37	31.041	37	34.04	0
38	30.107	36	0	0
39	32.342	38	0	34.1
40	31.562	37	0	0
41	29.71	35	0	0
42	30.844	36	31.56	0
43	30.216	35	0	33.02
44	30.778	35	0	0
45	31.482	36	0	0
46	31.693	36	0	31.98
47	30.033	34	0	31.03
48	30.016	34	0	33.02
49	31.137	35	0	0
50	29.986	34	32.24	0
51	30.608	34	0	32.48
52	30.789	34	0	0
53	32.441	36	33.93	0
54	29.89	33	0	30.23
55	32.159	35	0	0
56	30.214	33	0	0
57	30.11	33	0	0
58	30.419	33	31.14	0
59	32.315	35	0	32.4
60	31.43	34	0	0
61	29.515	32	0	29.93
62	30.86	33	0	0
63	30.068	32	31.18	0
64	32.214	34	0	33.39
65	29.597	31	0	0
66	32.452	34	0	33.45
67	31.825	33	0	0
68	30.099	31	0	30.53
69	31.301	32	0	0
70	33.542	34	0	0

Usando el vector edad, se estimo q_x para cada toda x que cumpla que $d_x \neq 0$, bajo el supuesto exponencial para datos completos.

x	q_x
31	0.09547
32	0.05078
33	0.02876
34	0.03383
35	0.0402
36	0.145
38	0.10673
39	0.1472

Se calculó la exposición actuarial para los intervalos $(x, x+1]$ en los que la exposición fue distinto de cero. Para los nuevos miembros y salidas durante el periodo de observación, se separó bajo una suposición de agrupamiento:

a)Edad fiscal último cumpleaños

b)Edad fiscal de calendario

x	a) FALB	b) CFA
29	3.5	0
30	16.5	19
31	31.5	33
32	38.5	41
33	35.5	33
34	30.5	30
35	24.5	25
36	20	20
37	11.5	12
38	9	9
39	7	7
40	3	3
41	1.5	1
42	1	1

Comentarios Finales

Los modelos estadísticos de supervivencia son usados en aplicaciones actuariales de vida, principalmente en el ámbito de los seguros, aunque también los modelos se pueden utilizar en finanzas, fianzas, seguros de no vida, por ejemplo seguro de daños o gastos médicos mayores , por mencionar algunos que se utilizan en actuaría, pero también pueden ser utilizados en medicina, biología e ingeniería.

Una de las cosas que faltó en este trabajo, es un estudio más detallado del estimador producto-límite, del estimador Nelson-Aalen, también la aplicación de los métodos en el ámbito laboral, así como, el manejo de paquetes estadísticos y la relación entre los modelos tabulares y los modelos paramétricos.

En la aplicación final de este trabajo se utilizó el caso especial A, donde todas las personas entran a la edad x y están programas a permanecer a la edad $x + 1$, ádemas de los supuestos lineal y exponencial, para realizar la estimación de la probabilidad de muerte y así obtener la estimación de la función de supervivencia.

Los objetivos de este trabajo eran dar una introducción a los modelos de supervivencia y dar un panorama general a los modelos de supervivencia tabulares o no paramétricos, utilizando algunos de los estimadores vistos en Estadística II.

Por lo que este trabajo puede servir como apoyo a la materia de Estadistica III, en específico al tema de modelos de supervivencia.

Bibliografía

[1] Bowers Newton L. (1993). *Acturial Mathematics.* The Society of Actuaries, Schaumburg Illions.

[2] Collet (1994). *Modelling Survival Data in Medical Research.* Chapman and Hall. Texts in Statistical Science.

[3] Cox D.R. and Oakes D. (1984). *Analisys of Survival Data. Chapman and Hall*: London, New York.

[4] Lee E. T. and Wang J. W. (2003). *Statistical Methods for Survival Data Analysis.* Third Edition. John Wiley.

[5] London, D. (1997). *Survival Models and their estimation.* Third Edition. ACTEX Publications.

[6] Marubini Ettore and Grazia Valsecchi Maria. (1995). *Analysing Survival Data from Clinical Trials and Observational Studies.* John Wiley.