ESTIMACIÓN ACTUARIAL VERSUS ESTIMACIÓN POR EL MÉTODO DE LOS MOMENTOS PARA LA PROBABILIDAD DE MUERTE

José María Sánchez López
Profesor del Área de Métodos Cuantitativos (Estadística y Matemáticas) de la Facultad de Ciencias Jurídicas y Sociales de la Universidad Rey Juan Carlos

Sumario: I.- Planteamiento del problema: la exposición al riesgo. II.- Método de los momentos. III.- Método de estimación actuarial. IV.- Ejemplo numérico abreviado. V.- Comentarios y nuevos enfoques. V.- Bibliografía.

Palabras clave: exposición al riesgo, estimador actuarial, probabilidad de muerte, criterio de implicación.

I.- PLANTEAMIENTO DEL PROBLEMA: LA EXPOSICIÓN AL RIESGO.

Desde un **enfoque actuarial**, la estimación de la probabilidad de muerte para un perfil de individuos se basa en los datos brutos observados. Una característica deseable en estos datos está en que combinen dos importantes características deseables: ser muy numerosos (como datos tipo censo) y recoger información detallada y precisa (como datos tipo experimental). Además, la medida de la mortalidad se debe tomar a partir de colectivos análogos a los que se aseguran: Betzuen Zalbidegoitia, Amancio (1999) presenta interesantes resultados que muestran la diferencia entre las tablas de mortalidad que se utilizan en el sector asegurador (obtenidas con información de la población general) y la medida de la mortalidad de un colectivo de activos ocupados (perfil habitual de los individuos asegurados).

Los elementos de interés para una correcta recopilación de los **datos brutos** son: la precisión, el intervalo de registro, el periodo elegido para la recogida de datos, la independencia entre los individuos, y la recogida y análisis de varios tipos de variables. Estos elementos están relacionados. La precisión en los datos se puede alterar por hipótesis adoptadas para facilitar los cálculos con las fechas, éstas fechas dependen a su vez del intervalo de registro o seguimiento, conjuntamente y adecuadamente tratados evitan el sesgo en las investigaciones de mortalidad. Los errores también pueden surgir si se eligen periodos de recogidas de datos con oscilaciones inusuales de mortalidad o si se toman individuos dependientes en cuanto a su exposición al riesgo de muerte. Todos los elementos mencionados intervienen en la correcta determinación de las variables fecha necesarias para calcular la exposición al riesgo.

De todos los elementos mencionados, en este artículo interesa la precisión en las **fechas** relevantes en el cálculo de la exposición individual al riesgo de muerte, ya que serán decisivas en la técnica de estimación: influirán en la dificultad de registro-tratamiento de los datos y en la correcta estimación. Las dificultades asociadas al registro exacto y a los cálculos para la estimación condujeron, inicialmente, al uso de la estimación actuarial como alternativa al método de los momentos. Posteriormente, se estudiaron las ventajas, los inconvenientes, las relaciones con otros métodos de estimación y las posibles interpretaciones.

Para plantear el problema es necesario analizar la **exposición individual al riesgo de muerte** como paso previo a la estimación. Partiendo de la contribución individual a la exposición al riesgo de muerte se estiman probabilidades por periodos recogiendo datos de todos los individuos. Según el tipo de exposición considerada se pueden utilizar distintas técnicas de estimación. La exposición se calculará mediante la transformación de la información relativa a las fechas relevantes respecto del intervalo de estimación. Cuando se obtiene la información se debe contemplar la necesidad de recoger tres edades importantes para los análisis posteriores: la edad de entrada al estudio y_i, la edad de salida programada del estudio z_i y la edad real de salida del periodo de estudio. Esta última situación puede darse por muerte a una edad θ_i o puede darse por retirada a una edad ϕ_i, en ambos casos la situación debe producirse antes de fin del periodo de estudio u observación. Las variables θ_i y ϕ_i toman valores cero si no se da la situación de muerte o retirada que las define. Así, a cada persona objeto de estudio se le puede asignar un vector que la caracterice de la forma $v_i' = [y_i, z_i, \theta_i, \phi_i]$ que recoge la información necesaria sobre la edad (ya sea real o virtual) para los posteriores procesos de estimación.

Un primer tratamiento de los datos llevaría a determinar la contribución de cada persona al intervalo de estimación que, en principio, se considera unitario y de la forma $(x,x+1]$. Se eliminan los vectores de los individuos que no contribuyen o permanecen en el intervalo, esto es, si $y_i \geq x+1$ o si $z_i \leq x$ o si $\theta_i \leq x$ o si $\phi_i \leq x$. Por último, se transforma cada vector v_i' en un vector de duración en $(x,x+1]$, de la forma $u_{i,x}' = [r_i, s_i, l_i, k_i]$:

$$r_i \begin{cases} 0 & si \quad y_i \leq x \\ y_i - x & si \quad x < y_i < x+1 \end{cases}$$

$$s_i \begin{cases} z_i - x & si \quad x < z_i < x+1 \\ 1 & si \quad z_i \geq x+1 \end{cases}$$

$$l_i \begin{cases} 0 & si \quad \theta_i = 0 \\ \theta_i - x & si \quad x < \theta_i \leq x+1 \\ 0 & si \quad \theta_i > x+1 \end{cases}$$

$$k_i \begin{cases} 0 & si \quad \phi_i = 0 \\ \phi_i - x & si \quad x < \phi_i \leq x+1 \\ 0 & si \quad \phi_i > x+1 \end{cases}$$

Se puede ya, a partir de los datos así presentados, obtener la exposición al riesgo de muerte para un individuo en el intervalo de estudio u observación, esto es, la amplitud del periodo en el cual un individuo está bajo observación, condicionando la posibilidad de muerte.

Se considera la **exposición exacta** como $\begin{bmatrix} s_i \\ l_i \\ k_i \end{bmatrix} - r_i$, eligiendo el menor valor distinto de cero en el vector columna.

Se considera la **exposición programada** como $\begin{bmatrix} s_i \\ k_i \end{bmatrix} - r_i$, eligiendo el menor valor distínto de cero en el vector columna. Existe la posibilidad, en los casos de muerte, de utilizar una exposición distinta a la anterior: en vez de considerar la exposición hasta la salida programada del estudio se toma la exposición hasta una retirada programada fijada (la fecha de finalización de la póliza por ejemplo) o hasta una retirada aleatoria (edad máxima aleatoria en función del comportamiento de toda la cartera).

Se considera la **exposición actuarial** como $\begin{bmatrix} s_i \\ k_i \\ 1 \end{bmatrix} - r_i$, eligiendo el menor valor distinto de cero en el vector columna, salvo en casos de muerte que se toma el valor uno.

Para hallar la exposición total se han de considerar todos los individuos sumando sus exposiciones al riesgo de muerte. Los distintos tipos de exposiciones al riesgo de muerte se utilizarán en los procesos de estimación en las situaciones de datos incompletos.

Los datos normalmente disponibles para los cálculos actuariales en las empresas de seguros vida son incompletos. En primer lugar, existen otros fenómenos (distintos de la muerte) de naturaleza aleatoria, que pueden concretarse en sucesos que conlleven la "retirada" de los individuos de la muestra observada antes del momento de la muerte. Aunque no se considere significativo el número de retiradas, en la mayoría de los estudios actuariales se encuentra el problema de observaciones de las que no se realiza un seguimiento hasta la total extinción del grupo seleccionado. Los estudios longitudinales, a edades no avanzadas, sólo se consideran viables si se analizan muestras pequeñas con corto tiempo de vida (análisis de supervivencia en ensayos clínicos o investigaciones médicas). Los estudios actuariales son mayoritariamente de corte transversal. Al finalizar el periodo de observación son muchos los individuos de la muestra que permanecen vivos, son los llamados "finalistas", que tendrán (en el momento de cierre de observación) una "edad de finalización predeterminada".

Con el fin de diferenciar ciertos casos, se estudia al individuo i-ésimo mediante el par (y_i,z_i). La primera coordenada indica la edad de entrada al proceso de observación y la segunda la edad que se tendrá al final del periodo de observación (edad de finalización predeterminada). Si hay muerte o retirada durante la observación, se tiene que dicha observación para este individuo concreto de la muestra termina antes del momento prefijado. Para un intervalo de estimación genérico $(x,x+1]$ no hay contribución a la exposición si $z_i \leq x$ o si $y_i \geq x+1$. Hay posible contribución en todo el intervalo si $y_i \leq x$ y $z_i \geq x+1$. Hay posible contribución en una parte si $x<y_i<x+1$ o/y $x<z_i<x+1$.

También se puede caracterizar al individuo i-ésimo, que entra en $x+r_i$ y se prevé salga en $x+s_i$, mediante el par (r_i,s_i) siendo: $r_i=0$ si $y_i \leq x$; $0<r_i<1$ si $x<y_i<x+1$; $s_i=1$ si $z_i \geq x+1$; $0<s_i<1$ si $x<z_i<x+1$. Se deben dejar fuera los casos que no presenten exposición en el intervalo establecido.

Según los valores del par (r_i,s_i), podemos distinguir cuatro casos (o tipos de individuos), tres especiales y uno genérico.

<u>Caso A</u>, con $r_i=0$ y $s_i=1$ para todo i: contribución en todo el intervalo.

<u>Caso B</u>, con $0 \leq r_i<1$ y $s_i=1$ para todo i: hay incorporaciones tras el inicio del intervalo y todos permanecen hasta el final del intervalo.

<u>Caso C</u>, con $0<s_i \leq 1$ y $r_i=0$ para todo i: todos se incorporan al inicio del intervalo y pueden darse salidas programadas antes de finalizar el intervalo.

Caso D, con $0 \leq r_i < 1$ y $0 < s_i \leq 1$ para todo i: hay incorporaciones tras el inicio del intervalo y pueden darse salidas programadas antes de finalizar el intervalo, es el "caso más genérico".

En la **exposición programada,** utilizada en la estimación por el **método de los momentos**, se sigue que cuando una persona es finalista observado en $x+s_i$ contribuye s_i-r_i, cuando una persona se observa se retira en $x+k_i$ contribuye k_i-r_i, y, cuando una persona muere en $(x,x+1]$ contribuye hasta la edad de salida programada del periodo de estudio (o bien se puede considerar una retirada fijada según el momento de expiración de la póliza).

En ciertos planteamientos se intenta distinguir entre retiradas planeadas y no planeadas (tal como anulaciones de pólizas), buscándose un tratamiento aleatorio especial para las retiradas no planeadas: se pasa a un entorno conocido como de "doble decremento". Ante esta situación parece adecuado el comentario que aparece en Elandt-Johnson, Regina and Johnson Norman, L. (1980):

"En nuestra opinión, si la mencionada mortalidad no depende del tipo de retirada, una hipótesis muy común, distinguir entre retiradas planeadas y no planeadas es irrelevante, una vez que los datos han sido recogidos". En todo caso, cabe añadir que únicamente para las situaciones o datos de muerte podría plantearse este problema y que no interesa en este estudio el proceso de retirada en sí mismo. Todo lo anterior lleva a trabajar en un entorno de "simple decremento".

El método actuarial, también conocido como **aproximación actuarial** al método de los momentos, simplifica los cálculos mediante el uso de una **exposición actuarial**, se distingue de la exposición programada en que para el caso de muerte se toma una exposición de $1-r_i$.

En la exposición observada se utilizan los datos derivados únicamente de lo observado, aplicándose en la estimación por máxima verosimilitud.

En este trabajo no se tratará este último método, exclusivamente se presentan el método de los momentos (utiliza la denominada "exposición programada") y el método de estimación actuarial (utiliza la denominada "exposición actuarial").

II.- MÉTODO DE LOS MOMENTOS.

Se identifica el número de muertes esperado (calculado a través de la variable aleatoria número de muertes) con el número de muertes observado (obtenido con la información muestral) en un determinado intervalo de tiempo fijado. Se debe proponer alguna hipótesis que lleve a una expresión que refleje el número de muertes esperado y se resuelve la ecuación para hallar el estimador deseado. Este procedimiento de estimación se justifica, siguiendo el teorema de Kintchine, por la convergencia en probabilidad de los momentos muestrales a los poblacionales, supuesta la independencia e igual distribución de las variantes que se tienen en los elementos de la muestra, cuya esperanza finita se supone igual a la de la población.

Interesa, para cada persona, genéricamente la i-ésima, la probabilidad condicionada de morir antes de $x+s_i$ (edad programada de salida del estudio o de retirada si anterior) estando vivo en $x+r_i$, esto proporciona el número de muertes esperado para la muestra de tamaño uno que se representa como

$$ {}_{s_i-r_i}q_{x+r_i} \,. $$

Si se toma toda la muestra, la esperanza de la variable aleatoria queda:

$$ E(D_x) = \sum_{i=1}^{n_x} {}_{s_i-r_i}q_{x+r_i} \,. $$

Para obtener probabilidades estimadas para años enteros hay que utilizar alguna **hipótesis de distribución** de los valores de la probabilidad en el interior del intervalo. Se presentan tres situaciones: fuerza de mortalidad decreciente, fuerza de mortalidad constante y fuerza de mortalidad creciente.

Si se sigue la llamada hipótesis de Balducci, que presupone un comportamiento hiperbólico de la función de supervivencia y un decremento de la fuerza de mortalidad dentro del intervalo de estimación (situación poco realista pero conveniente por el resultado analítico), queda:

$$_{s_i-r_i}q_{x+r_i} \approx (s_i-r_i)q_x \quad \Rightarrow \quad E(D_x)=q_x\sum_{i=1}^{n_x}(s_i-r_i).$$

Si se sigue, dentro del intervalo de estimación, la primera ley de Dormoy, que presupone un comportamiento exponencial de la función de supervivencia y una fuerza de mortalidad constante (situación algo más realista que la anterior), queda:

$$_{s_i-r_i}q_{x+r_i} \approx 1-(1-q_x)^{s_i-r_i} \quad \Rightarrow \quad E(D_x)=n_x-\sum_{i=1}^{n_x}(1-q_x)^{s_i-r_i}.$$

Si se presupone un comportamiento lineal de la función de supervivencia, tal como expresa la ley de Moivre, que conduce a una fuerza de mortalidad creciente dentro del intervalo de estimación (situación más realista y equivalente a una interpolación lineal entre valores consecutivos de la función de supervivencia), queda:

$$_{s_i-r_i}q_{x+r_i} \approx \frac{(s_i-r_i)q_x}{1-(r_i\bullet q_x)} \quad \Rightarrow \quad E(D_x)=q_x\sum_{i=1}^{n_x}\frac{s_i-r_i}{1-(r_i\bullet q_x)}.$$

Por otra parte, el número de muertes observado en la muestra será representado por d_x.

Al igualar esperanza de la variante y número de muertes observado, la estimación de la probabilidad de muerte durante el año x se obtendrá bajo los tres supuestos considerados:

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

$$E(D_x)=d_x \quad \Rightarrow \quad n_x-\sum_{i=1}^{n_x}(1-\hat{q}_x)^{s_i-r_i}=d_x,$$

$$E(D_x)=d_x \quad \Rightarrow \quad \hat{q}_x\sum_{i=1}^{n_x}\frac{s_i-r_i}{1-(r_i\bullet \hat{q}_x)}=d_x.$$

Sólo la primera expresión tiene una solución analítica, para el resto se utilizarán algoritmos matemáticos de aproximación o métodos de tanteo recursivos mediante ordenador.

Si se presta atención a la primera expresión, es claro que el sumatorio que aparece en el denominador representa la exposición prevista o programada de la muestra y que si un individuo de la muestra muere en periodo de observación no se da toda la exposición programada.

En cuanto a las propiedades del estadístico estimador es claro que al estimarse un momento respecto del origen de la distribución poblacional se llega a que el estimador obtenido es insesgado. Por ejemplo, para el estimador

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

si se admite la hipótesis de hiperbólica o de Balducci empleada al estimar, se llega a:

$$E\left(\hat{Q}_x\right) = \frac{E(D_x)}{\sum_{i=1}^{n_x}(s_i - r_i)} = \frac{\sum_{i=1}^{n_x} {}_{s_i - r_i}q_{x + r_i}}{\sum_{i=1}^{n_x}(s_i - r_i)} \approx \frac{q_x \sum_{i=1}^{n_x}(s_i - r_i)}{\sum_{i=1}^{n_x}(s_i - r_i)} = q_x .$$

El sesgo sólo aparece si la hipótesis no se cumple. Si para hallar el estimador se emplea una hipótesis de mortalidad decreciente y la población sigue una fuerza de mortalidad constante o creciente se llega a una esperanza del estimador mayor que el valor del parámetro estimado.

Por el mencionado teorema de Kintchine se tienen estimaciones consistentes y con el teorema de Lindeberg-Levi se comprueba la normalidad asintótica. En cuanto a la varianza del estimador, si se admite el comportamiento binomial de la variante número de muertes, y con la aproximación hiperbólica quedará:

$$V\left(\hat{Q}_x\right)=\frac{V(D_x)}{\left[\sum_{i=1}^{nx}(s_i-r_i)\right]^2}=\frac{\sum_{i=1}^{nx}{}_{s_i-r_i}q_{x+r_i}\bullet\left(1-{}_{s_i-r_i}q_{x+r_i}\right)}{\left[\sum_{i=1}^{nx}(s_i-r_i)\right]^2}=$$

$$=\frac{q_x\sum_{i=1}^{nx}(s_i-r_i)-(q_x)^2\sum_{i=1}^{nx}(s_i-r_i)^2}{\left[\sum_{i=1}^{nx}(s_i-r_i)\right]^2}$$

Una aproximación se consigue si se considera el estimador de la probabilidad de muerte a la edad x como proporción binomial de una muestra con tamaño equivalente a la exposición programada:

$$V\left(\hat{Q}_x\right)=\frac{V(D_x)}{\left[\sum_{i=1}^{nx}(s_i-r_i)\right]^2}\approx\frac{q_x\bullet(1-q_x)}{\sum_{i=1}^{nx}(s_i-r_i)}.$$

Ambas expresiones dependen del tamaño de la muestra y, de la entrada y salida programada al intervalo de estimación.

III.- MÉTODO DE ESTIMACIÓN ACTUARIAL.

El Método Actuarial es del siglo XIX, previo al desarrollo formal de la Teoría de Estimación estadística. Tuvo un desarrollo intuitivo y busca simplificar los cálculos (por carecer de medios informáticos o de información precisa sobre la salida programada). Utiliza el concepto de exposición actuarial (en lugar de exposición programada).

Cuando se trata la denominada "aproximación actuarial al método de los momentos" o simplemente el "estimador actuarial" surge el problema de fijar, para los casos de muerte dentro del intervalo de estudio, la edad de salida considerada como programada, $x+s_i$. Se necesitaría un análisis de datos extenso para resolver el problema. La solución actuarial, por simplicidad de cálculo y de análisis, establece un s_i=1 para todas las muertes (se supone a todos programados para finalizar el intervalo unitario de tiempo objeto de estudio). Con esto, se identifica $x+s_i$ a posteriori de la finalización del intervalo (según el individuo muera o no). Si, realmente, s_i fuera igual a uno en todos los individuos considerados, no sería una aproximación. Esto último ocurre en las situaciones denominadas Caso especial A y Caso especial B.

Aparecieron intentos, ya clásicos, de **justificación estadística** (por Cantelli y Balducci). Se intenta identificar o asociar el método de los momentos al método actuarial. Así, se considera:

$$E(D_x) = n_x {}_{1-r}q_{x+r} - e_x {}_{1-s}q_{x+s} = d_x\,;$$

para un grupo de tamaño n_x, donde todas las muertes que se espera ocurran antes de x+1 sumarán la cantidad de $n_x {}_{1-r}q_{x+r}$.

Esta cifra se separa en dos, las muertes observables, d_x, y, las muertes que se espera ocurran antes de x+1, pero no observables al ocurrir entre los e_x individuos que dejan la muestra en $x+s$, esto es, $e_x {}_{1-s}q_{x+s}$. Como se consideraba que las edades de finalización programadas no estaban disponibles no se recurre a ellas, pero sí al número real observado de finalistas. Si se dieran las condiciones del Caso especial A o del Caso especial B la aproximación actuarial al método de los momentos y el método de los momentos conducirían a expresiones idénticas. En general, se tratará de una aproximación basada en la filosofía del método de los momentos que utiliza los escasos datos disponibles.

En cambio, esta aproximación actuarial se **critica** en los casos en que se esperaran finalistas en el periodo de estimación (Caso especial C) tal como dice Hoem, Jan M. (1984). Se condiciona al grupo n_x a estar vivo en $x+r$ y, al grupo e_x, a estarlo en $x+s$ ($x+s>x+r$). No se especifica, sin embargo, la supervivencia entre $x+r$ y $x+s$. De hecho, e_x o finalistas observados serían la realización de una variante binomial del fenómeno número de supervivientes a $x+s$ de una muestra con edad de salida programada en situación de supervivencia a $x+s$.

En los casos de finalistas a $x+s_i$ con $s_i<1$ se muestra mediante simulaciones y atendiendo al análisis del problema, tal como se tiene en Broffit, James D. (1984), que el estimador actuarial es sesgado negativamente e inconsistente. En los estudios actuariales basados en grandes cantidades de datos aportados por compañías de seguros aparecerán un número significativo de finalistas. Para evitarlo, puede diseñarse un estudio cuyos límites de intervalo coincidan con periodos de finalistas programados (coincidir límite de intervalo con periodo de renovación de la póliza).

Críticas parecidas se tratan en Hoem, Jan M. (1984), cuya base se encuentra en no asignar una función de probabilidad para los individuos que se retiran.

IV.- EJEMPLO NUMÉRICO ABREVIADO.

A pesar de las críticas anteriores, se comprueba que en algunas situaciones las estimaciones obtenidas por el método de los momentos y el método actuarial no difieren mucho.

Una forma de reflejar las posibles situaciones y las consecuencias en cuanto a exposición total y valores de mortalidad estimados consiste en plantear un **ejemplo que incluya gran cantidad de casos diferentes**. Por este motivo se expone una tabla con 70 casos que serán analizados con el método expuesto. Cada caso (o posible póliza referida a una persona asegurada) se identifica con un número y contiene los siguientes cuatro registros: fecha de nacimiento, fecha de entrada o inicio de aseguramiento con la compañía, fecha de retirada o fin de aseguramiento si canceló o no renovó la póliza, y fecha de muerte si ocurrió en el periodo de aseguramiento. El **periodo de observación** o estudio que se considera en el análisis será el comprendido **entre el 1-1-1994 y el 31-12-2003** (estudio transversal de periodo reciente).

El fichero de partida es el siguiente:

TABLA 1. Datos Brutos sobre Fechas.

	Nacim.	Inicio	Retirada	Muerte		Nacim.	Inicio	Retirada	Muerte
1	11-4-64	15-6-92			36	29-6-66	31-10-97	9-3-02	
2	1-7-60	28-3-92		3-1-94	37	27-2-67	14-7-98		15-7-01
3	10-6-62	31-7-92	30-11-94		38	30-1-68	7-8-98		
4	27-2-58	19-8-92	26-8-93		39	4-9-66	1-11-98	5-8-00	
5	27-11-65	10-1-93		24-12-95	40	1-11-67	20-1-99		
6	5-6-55	5-1-93	2-5-93		41	14-7-69	15-3-99		
7	14-3-61	8-12-92			42	19-4-68	3-5-99		19-1-00
8	23-8-62	2-5-93	3-3-94		43	26-7-69	16-9-99	5-7-02	
9	24-6-64	30-1-94	30-12-94		44	2-12-69	9-4-00		
10	22-2-63	21-9-93		9-5-02	45	3-1-68	22-12-99		
11	14-12-61	7-10-93	15-5-95		46	5-11-68	9-3-00	20-6-00	
12	5-7-62	17-2-94	9-8-03		47	18-5-70	11-7-00	11-7-01	
13	9-6-64	4-4-94			48	15-8-70	5-7-00	5-7-03	
14	18-2-63	1-6-94		13-4-00	49	3-1-69	18-8-00		
15	14-10-64	23-11-94	9-5-02		50	9-3-70	24-6-00		23-9-02
16	13-4-65	9-2-95			51	23-9-70	6-2-01	20-12-02	
17	9-8-62	19-12-94		9-10-01	52	14-11-70	13-4-01		
18	7-10-65	6-7-95	5-4-00		53	5-9-68	7-12-00		3-6-02
19	25-10-63	25-5-95			54	13-9-71	20-5-01	20-9-01	
20	4-2-64	15-12-95	24-8-98		55	20-11-69	26-8-01		
21	13-12-65	11-12-95		5-3-01	56	18-8-71	15-9-01		
22	9-12-64	1-5-96			57	5-1-71	8-8-01		
23	30-7-66	21-4-96			58	5-4-71	29-11-01		18-8-02
24	23-1-65	8-9-96	8-9-97		59	3-8-69	22-10-01	22-11-01	
25	17-7-67	27-1-97	15-11-03		60	2-2-70	3-12-01		
26	1-10-66	13-10-96		13-11-98	61	26-4-72	3-1-02	3-6-02	
27	4-9-64	23-8-96	3-6-01		62	25-2-71	9-5-02		
28	20-3-65	11-11-96		1-12-01	63	12-8-72	24-7-02		4-9-03
29	9-4-67	3-3-97			64	5-7-70	15-9-02	18-11-03	
30	24-5-66	29-12-96	20-6-99		65	20-5-73	2-2-03		
31	5-8-66	25-5-97			66	14-9-70	11-12-02	11-12-03	
32	19-11-67	12-6-97			67	7-11-71	26-4-03		
33	9-12-65	5-4-97		14-2-02	68	20-7-73	4-8-03	9-1-04	
34	3-8-67	1-9-97	18-8-98		69	15-2-72	17-10-03		
35	9-9-67	27-2-98			70	12-11-70	13-1-04		

Se observa diversas situaciones: entradas y salidas antes y después del periodo de estudio, retiradas y muertes antes y después del periodo de estudio, y casos sin retirada ni muerte.

Estos datos de partida han de **transformarse en edades** referenciadas consideradas según el periodo de estudio considerado, en este caso entre el 1-1-1994 y el 31-12-2003. Cada caso, identificado con su número, se acompañará de cuatro registros que forman el vector $v_i' = [y_i, z_i, \theta_i, \phi_i]$ que recoge la información: edad de entrada al estudio y_i, la edad de salida programada del estudio z_i , edad real de salida del periodo de estudio por muerte θ_i y edad real de salida del periodo de estudio por retirada ϕ_i (las variables θ_i y ϕ_i toman valores cero si no se da la situación que las define).

El fichero de generado a partir del anterior (mostrado en la TABLA1) y considerando las fechas de inicio y fin del periodo de estudio queda como se expone:

TABLA 2. Edades con relación al Periodo de Estudio.

	Entrada	Salida	Muerte	Retirada
1	29,72485	39,72074	0	0
2	33,50308	43,49897	33,50856	0
3	31,56194	41,55784	0	32,47365
4				
5	28,09582	38,09172	30,07255	0
6			0	0
7	32,80219	42,79808	0	0
8	31,35934	41,35524	0	31,52635
9	29,60164	39,51814	0	30,51608
10	30,85832	40,85421	39,20876	0
11	32,04928	42,04517	0	33,41547
12	31,62218	41,48939	0	41,09514
13	29,81793	39,55921	0	0
14	31,28268	40,86516	37,14990	0
15	30,10815	39,21150	0	37,56605
16	29,82615	38,71595	0	0
17	32,36140	41,39357	39,16769	0
18	29,74401	38,23135	0	34,49418

	Entrada	Salida	Muerte	Retirada
19	31,58111	40,18344	0	0
20	31,86037	39,90418	0	34,55168
21	29,99316	38,04791	35,22519	0
22	31,39220	39,05818	0	0
23	29,72758	37,42094	0	0
24	31,62491	38,93498	0	32,62423
25	29,53320	36,45722	0	36,33128
26	30,03422	37,24846	32,11773	0
27	31,96715	39,32101	0	36,74470
28	31,64682	38,78166	36,70089	0
29	29,90007	36,72827	0	0
30	30,60096	37,60438	0	33,07324
31	30,80356	37,40452	0	0
32	29,56331	36,11499	0	0
33	31,32101	38,05886	36,18344	0
34	30,08077	36,41068	0	31,04175
35	30,46954	36,30938	0	0
36	31,34018	37,50582	0	35,69336
37	31,37577	36,84052	34,37919	0
38	30,51882	35,91786	0	0
39	32,15880	37,32238	0	33,91923
40	31,21971	36,16427	0	0
41	29,66735	34,46407	0	0
42	31,03628	35,69884	31,75086	0
43	30,14100	34,43121	0	32,94182
44	30,35181	34,07803	0	0
45	31,96715	35,99179	0	0
46	31,34018	35,15127	0	31,62218
47	30,14921	33,62081	0	31,14853
48	29,88912	33,37714	0	32,88706
49	31,62218	34,98973	0	0
50	30,29432	33,81246	32,54209	0
51	30,37372	33,27036	0	32,24093
52	30,41205	33,12799	0	0
53	32,25462	35,31828	33,74127	0
54	29,68378	32,29843	0	30,02053
55	31,76454	34,11088	0	0
56	30,07803	32,36961	0	0
57	30,59001	32,98563	0	0

	Entrada	Salida	Muerte	Retirada
58	30,65298	32,73922	31,37029	0
59	32,21903	34,40931	0	32,30390
60	31,83299	33,90828	0	0
61	29,68925	31,67967	0	30,10267
62	31,20055	32,84600	0	0
63	29,94661	31,38398	31,06092	0
64	32,19713	33,48939	0	33,37166
65	29,70568	30,61465	0	0
66	32,24093	33,29500	0	33,24025
67	31,46612	32,14784	0	0
68	30,03970	30,44764	0	0
69	31,66872	31,87406	0	0
70				

Los valores aparecen para su mejor visualización truncados a cinco decimales. Si el caso no aparece es que no coincide temporalmente con el periodo de estudio.

A continuación, se debe estudiar, considerando **por separado cada intervalo anual** de estimación, la aportación o duración o contribución de cada caso o individuo.

Se necesitará obtener para cada año, de forma genérica (x,x+1], los elementos que permitan obtener la exposición exacta, programada y actuarial que, posteriormente guíen la estimación por los diferentes métodos. Para esto, como se dijo, se explora si existe contribución al intervalo anual tratado y se transforma cada vector v_i' en un vector de duración en $(x,x+1]$, de la forma $u_{i,x}' = [r_i,s_i,l_i,k_i]$.

Se presenta solamente el intervalo (31,32] para evitar una extensión excesiva y por ser el de mayor interés al generar distintos valores para los distintos tipos de exposición tratada. Esto es así porque en este mismo intervalo de edades se da una situación que provocará diferencias entre exposición programada y actuarial: existe muerte y salida del estudio en el mismo individuo entre los 31 y 32 años (caso 63).

De igual forma se han efectuado las operaciones para el resto de intervalos.

TABLA 3. Contribución al Intervalo (31,32].

	T(31)	r(31)	s(31)	l(31)	k(31)		T(31)	r(31)	S(31)	l(31)	k(31)
1	1	0	1	0	0	36	1	0,34018	1	0	0
2	0					37	1	0,37577	1	0	0
3	1	0,56194	1	0	0	38	1	0	1	0	0
4	0					39	0				
5	0					40	1	0,21971	1	0	0
6	0					41	1	0	1	0	0
7	0					42	1	0,03628	1	0,75086	0
8	1	0,35934	1	0	0,52635	43	1	0	1	0	0
9	0					44	1	0	1	0	0
10	1	0	1	0	0	45	1	0,96715	1	0	0
11	0					46	1	0,34018	1	0	0,62218
12	1	0,62218	1	0	0	47	1	0	1	0	0,14853
13	1	0	1	0	0	48	1	0	1	0	0
14	1	0,28268	1	0	0	49	1	0,62218	1	0	0
15	1	0	1	0	0	50	1	0	1	0	0
16	1	0	1	0	0	51	1	0	1	0	0
17	0					52	1	0	1	0	0
18	1	0	1	0	0	53	0				
19	1	0,58111	1	0	0	54	0				
20	1	0,86037	1	0	0	55	1	0,76454	1	0	0
21	1	0	1	0	0	56	1	0	1	0	0
22	1	0,39220	1	0	0	57	1	0	1	0	0
23	1	0	1	0	0	58	1	0	1	0,37029	0
24	1	0,62491	1	0	0	59	0				
25	1	0	1	0	0	60	1	0,83299	1	0	0
26	1	0	1	0	0	61	0				
27	1	0,96715	1	0	0	62	1	0,20055	1	0	0
28	1	0,64682	1	0	0	63	1	0	0,38398	0,06092	0
29	1	0	1	0	0	64	0				
30	1	0	1	0	0	65	0				
31	1	0	1	0	0	66	0				
32	1	0	1	0	0	67	1	0,46612	1	0	0
33	1	0,32101	1	0	0	68	0				
34	1	0	1	0	0,04175	69	1	0,66872	0,87406	0	0
35	1	0	1	0	0	70	0				

Ya con los datos que recoge la TABLA 3 se realizan los cálculos individuales de exposición al riesgo (exacta, programada y actuarial) y seguidamente se suman todos los casos.

Con las exposiciones y siguiendo la **hipótesis de Balducci** sobre el comportamiento de la mortalidad intraintervalo se obtienen estimaciones de mortalidad por los métodos de momentos y actuarial. La siguiente tabla recoge los resultados para las edades entre treinta y cuarenta años.

TABLA 4. Exposiciones y Estimación Mortalidad Anual.

x	d_x	Expos. Exacta	Expos. Programada	Expos. Actuarial	$q_{(x)}$ M.Momentos	$q_{(x)}$ M. Actuarial
30	1	24,21697467	25,14442163	25,14442163	0,039770253	0,039770253
31	3	35,34086242	36,54277892	37,15879535	0,082095563	0,080734587
32	2	44,19712526	45,53730322	45,53730322	0,043920036	0,043920036
33	2	35,80287474	36,55304586	36,55304586	0,054715003	0,054715003
34	1	28,06776181	28,68856947	28,68856947	0,034857088	0,034857088
35	1	23,82819986	24,60301164	24,60301164	0,040645431	0,040645431
36	2	16,27720739	17,39288159	17,39288159	0,114989571	0,114989571
37	1	10,54140999	11,39151266	11,39151266	0,087784654	0,087784654
38	0	8,715947981	8,715947981	8,715947981	0	0
39	2	4,714579055	6,338124572	6,338124572	0,315550756	0,315550756

Se observa que el método de momentos y actuarial coinciden en su estimación en todos los intervalos excepto en el ya señalado (edad de 31 años). También se destaca que si no hay fallecimientos la estimación es cero (en general se necesitan muestras muy grandes para estimaciones aceptables). Los resultados obtenidos en las estimaciones no corresponden a la realidad dada la imposibilidad de presentar datos reales (demasiado numerosas para presentar las tablas). Se desea resaltar e ilustrar los métodos. Estos métodos permiten, con una cartera de pólizas, obtener adecuadas estimaciones de mortalidad que permiten **analizar desviaciones respecto de las tablas estándar** utilizadas. Además, si se emplean simultáneamente técnicas multivariantes (regresión logística por ejemplo) se pueden hacer estudios según factores de riesgos (condicionado al tamaño de la cartera y a los subgrupos que se puedan formar).

Una vez entendido el sentido del estimador actuarial y, según datos brutos, la posible **"semejanza numérica"** a posteriori con el estimador obtenido por el método de los momentos, se pretende profundizar en las justificaciones teóricas. En el siguiente punto se analizan estas otras posibles justificaciones del estimador actuarial.

V- COMENTARIOS Y NUEVOS ENFOQUES.

Debido a la importancia y a la gran aplicación del llamado "estimador actuarial convencional (clásico)" se han realizado muchos estudios. A pesar de esto, no hay consenso unánime en la interpretación y en la posible validez teórica del método actuarial clásico.

Una sencilla e interesante defensa de la aproximación actuarial convencional a la estimación del riesgo por exposición se tiene en **Dorrington, Robert Edwin and Slawski, Janina Krystyna (1993).** Basa sus argumentos en la consistencia del método actuarial de cálculo de exposición con el número esperado agregado de muertes.

Supone que la variable aleatoria que representa la muerte o vida para el individuo i-ésimo, cuya observación comienza a la edad $x+r_i$, tiene una esperanza de la forma

$$E[D_i] = {}_{\alpha_i - r_i}q_{x+r_i}$$

con lo que la esperanza para el conjunto de los individuos considerados será

$$E[\theta_x] = \sum_i {}_{\alpha_i - r_i}q_{x+r_i} = \sum_i {}_{1-r_i}q_{x+r_i} - \sum_i {}_{\alpha_i - r_i}p_{x+r_i}\,{}_{1-\alpha_i}q_{x+\alpha_i} =$$
$$= \sum_i {}_{1-r_i}q_{x+r_i} - \sum_i (1 - E[D_i])_{1-\alpha_i}q_{x+\alpha_i} =$$
$$= \sum_i {}_{1-r_i}q_{x+r_i} - E\left[\sum_i (1 - D_i)_{1-\alpha_i}q_{x+\alpha_i}\right] =$$
$$= \sum_i {}_{1-r_i}q_{x+r_i} - E\left[\sum_{i:D_i=0} {}_{1-\alpha_i}q_{x+\alpha_i}\right]$$

Basta ahora sustituir la edad máxima $x+\alpha$ por la edad real de salida $x+t$ ya que aparece sólo en los casos de individuos no fallecidos (valor cero para la variante dicotómica "vida muerte"). Se tiene entonces,

$$E[\theta_x] = \sum_i {}_{1-r_i}q_{x+r_i} - E\left[\sum_{i:D_i=0} {}_{1-t_i}q_{x+t_i}\right],$$

que se presenta como la correcta interpretación para las muertes esperadas según la aproximación convencional o actuarial. Tiene su adecuado sentido cuando se involucran todas las N vidas, sin separar o descomponer la expresión para individuos determinados.

Si en el desarrollo anterior, en concreto en

$$E[\theta_x] = \sum_i {}_{1-r_i}q_{x+r_i} - \sum_i (1 - E[D_i])_{1-\alpha_i}q_{x+\alpha_i}$$

se sustituye la variante dicotómica por su esperanza se tiene la aproximación convencional o actuarial:

$$\sum_i D_i = \sum_i {}_{1-r_i}q_{x+r_i} - \sum_i (1 - D_i)_{1-\alpha_i}q_{x+\alpha_i} = \sum_i {}_{1-r_i}q_{x+r_i} - \sum_{i:D_i=0} {}_{1-t_i}q_{x+t_i}.$$

Por tanto, aparece una relación racional y coherente entre el estimador que se tiene por el método de los momentos y el estimador actuarial. Son estas razones las que llevan a afirmar a Robert Edwin Dorrington y a Janina Krystyna Slawski que "la aproximación convencional es una aceptable aproximación práctica que permite cálculo directo de la estimación requerida sin necesitar hipótesis explícitas sobre la edad máxima de la muerte".

Recientes trabajos apoyan el método de estimación actuarial, suministrando un nuevo enfoque. Se propone una nueva racionalidad que se relaciona con el estimador límite producto. Destaca el trabajo desarrollado por **Puzey, Anthony S. (1995)**. El autor pretende reorientar la discusión, evitando la referencia continua al método de los momentos. Tal como dice, hasta ahora "la crítica y la justificación a la idiosincrasia del método de cálculo de muertes esperadas parece surgir desde las comparaciones del método de cálculo del estimador convencional con el método de cálculo del estimador del método de momentos". Se conseguirá expresar el criterio subyacente en el estimador actuarial sin recurrir a una posible relación con el método de los momentos, a la que se llegaría modificando hipótesis o estableciendo restricciones.

Se parte, al igual que en el estimador actuarial, de la relación que iguala el número observado de muertes a la diferencia entre número de muertes esperadas en el total de vivos que están presentes (en algún momento) en el periodo y edad de estudio, y número de muertes esperadas en los vivos que se retiran de observación (según retirados realmente observados):

$$D = \sum_{N} {}_{1-r_i} q_{x+r_i} - \sum_{W} {}_{1-t_i} q_{x+t_i}$$

Este punto de partida puede criticarse por no emplear, a priori, una función de probabilidad para los individuos que se retirarán.

A continuación, se observa que la expresión

$$
\begin{aligned}
{}_{1-r} p_{x+r} &= P\left(X \geq x+1 \middle/ X \geq x+r\right) = \frac{P(X \geq x+1)}{P(X \geq x+r)} = \\
&= \frac{P(X \geq x+1)/P(X \geq x)}{P(X \geq x+r)/(X \geq x)} = \frac{P\left(X \geq x+1 \middle/ X \geq x\right)}{P\left(X \geq x+r \middle/ X \geq x\right)} = \frac{p_x}{{}_r p_x}
\end{aligned}
$$

permite sustituir en la ecuación anterior:

$$D=\sum_N {}_{1-r_i}q_{x+r_i}-\sum_W {}_{1-t_i}q_{x+t_i}=\sum_N\left(1-\frac{p_x}{{}_{r_i}p_x}\right)-\sum_W\left(1-\frac{p_x}{{}_{t_i}p_x}\right)=$$

$$=N-\sum_N\left(\frac{p_x}{{}_{r_i}p_x}\right)-W+\sum_W\left(\frac{p_x}{{}_{t_i}p_x}\right)$$

Finalmente, considerando los supervivientes como *S=N-W-D*, simplificando y despejando se llega a:

$$\sum_N\frac{1}{{}_{r_i}p_x}=\sum_{WS}\frac{1}{{}_{t_i}p_x}.$$

No se emplea una posible distribución de probabilidad para retirados. El método de los momentos, en sentida estricto, si requiere estimar tal distribución (o utilizar retiradas planeadas).

El lado izquierdo de la ecuación refleja el número de vidas que intervienen para establecer el número de supervivientes a la mortalidad. El lado derecho de la ecuación refleja el número de vidas que salen no por muerte, de las consideradas anteriormente. Esta asociación, entre probabilidades relacionadas con número de individuos implicados al inicio y probabilidades relacionadas con número de individuos saliendo no fallecidos implicados al final, es lo que se llamará **"criterio de implicación"**. Este criterio se deriva de la aproximación actuarial clásica o convencional de estimación del riesgo de exposición.

También se debe destacar la relación con el estimador límite producto. Se pueden fijar las probabilidades de la anterior expresión

$$\sum_N\frac{1}{{}_{r_i}p_x}=\sum_{WS}\frac{1}{{}_{t_i}p_x}$$

sin recurrir a leyes de mortalidad paramétricas. Se cumple el criterio de implicación y la ecuación anterior si se recurre al estimador límite producto. Para ello, se parte el periodo en intervalos limitados por los puntos donde aparecen entradas o salidas no debidas a muertes, y las probabilidades se estimarán con la proporción de vidas que sobrevivan en el subperiodo estudiado. La consideración de todos los subperiodos permite hallar la probabilidad de supervivencia en todo el intervalo.

Aunque se puedan unificar los criterios iniciales entre estimación actuarial y límite producto, la forma final de fijar la estimación establecerá diferencias. Destaca que el estimador actuarial es ponderado según el número de vidas mientras que el estimador límite producto es ponderado según tiempo de subperiodos. La desviación típica en este último caso es generalmente mayor debida a la importancia de los datos en cada punto y según su volumen.

A modo de **resumen** se puede decir que el método de estimación actuarial empezó siendo un método informal, previo a la teoría de estimación estadística clásica, desarrollado para estimar probabilidades de muerte en la cartera de pólizas de una empresa aseguradora. Cuando surgió el método de los momentos, se observó (en la práctica actuarial) la similitud entre ambos métodos. En el ejemplo numérico presentado se intenta reflejar esta característica empírica (se confirma fácilmente con una simulación).

Las críticas al estimador actuarial se propusieron según ventajas e inconvenientes respecto del método de los momentos. Las ventajas se refieren a la posibilidad de aplicar la filosofía del método de los momentos a pesar de no tener todos los datos necesarios disponibles, buscando la consistencia del método en el cálculo de la exposición con el número esperado agregado de muertes. Los inconvenientes se refieren a los casos con finalistas en el periodo de estimación ya que no se asigna una función de probabilidad a los individuos que se retiran.

Posteriormente, se reorienta el estudio evitando la comparación con el método de los momentos, esto lleva a buscar una racionalidad propia en el estimador actuarial en base al criterio de implicación y al estimador límite producto.

V.- BIBLIOGRAFÍA.

Betzuen Zalbidegoitia, Amancio (1999). La medida de la mortalidad en un colectivo de activos ocupados. Actuarios. Núm.17, Dossier, Págs.:1-8.

Broffit, James D. (1984). Maximum likelihood alternatives to actuarial estimators of mortality rates. Transactions of the Society of Actuaries. Vol.36, Págs.:77-122.

Dorrington, Robert Edwin and Slawski, Janina Krystyna (1993). A defence of the conventional actuarial approach to the estimation of the exposed-to-risk. Scandinavian Actuarial Journal. Núm.2, Págs.:107-113.

Elandt-Johnson, Regina C. and Johnson Norman, L. (1980). Survival models an data analysis. Wiley. New York.

Felipe Gómez Rojas, Felipe (2003). Algunas ideas sobre un sistema de gestión de riesgos técnicos y de mercado en una compañía de seguros de vida. Actuarios. Núm.21, Págs.:26-33.

Frostig, Esther (2003). Properties of the power family of fractional age approximations. Insurance: Mathematics and Economics. Vol.33, Págs.:163-171. North-Holland.

Hoem, Jan M. (1984). A flaw in actuarial exposed to risk theory. Scandinavian Actuarial Journal. Núm.3, Págs.:187-194.

Hosmer, D.W. (2003). Applied Survival Analysis. Wiley, New York.

Jones, Bruce L. and Mereu, John A. (2002). A critique of fractional age assuptions. Insurance: Mathematics and Economics. Vol.30, Págs.:363-370. North-Holland.

London, D. (1997-3ªedición). Survival models and their estimation. Actex Publications. Connecticut.

Miller, John H. and Courant, Simon T. (1988). Anderson's attributed exposure method. Congress International of Actuaries. Págs.373-391.

Puzey, Anthony S. (1995). A reappraisal of the principle underlying the conventional actuarial estimator of q_x. Insurance, Mathematics and Economics. Núm.17, Págs.:125-132. North-Holland.

Puzey, Anthony S. (1997). A general theory of mortality rate estimators. Actuarial Research Paper. Núm.98. Department of Actuarial Science and Statistics, City University. London.

Puzey, Anthony S. (1997). On the bias of the conventional actuarial estimator of q_x. Actuarial Research Paper. Núm.99. Department of Actuarial Science and Statistics, City University. London.
Renshaw, A.E. and Haberman, S. (2003). Lee-Carter mortality forecasting with age-specific enhancement. Insurance: Mathematics and Economics. Vol.33, Págs.:255-272. North-Holland.
Renshaw, A.E. and Haberman, S. (2003). On the forecasting of mortality reduction factors. Insurance: Mathematics and Economics. Vol.32, Págs.:379-401. North-Holland.
Sánchez López, José María (2001). Cuantificación de riesgos y análisis global de la empresa aseguradora de vida, Dykinson, S.L., Madrid.
Sánchez López, José María (2001). Construcción de tablas de mortalidad. Revista Española de Seguros. Núm.106 Págs.:297-306. Madrid.
Smith, P. (2002). Analysis of Failure & Survival Data, Ed. Chapman and Hall, London.
Wunchs, G. (2002). Life Table Modelling Survival & Death. Kluwer Academic Publishers, Boston.