Metodología empleada en el cálculo de las tablas de mortalidad de la población de España 1992-2005

Madrid, julio de 2007

Índice

Metodología empleada en el cálculo de las tablas de mortalidad de la población de España 1992-2005

1 Introducción

Las tablas de mortalidad se construyen con el fin de medir la incidencia de este fenómeno en la población que se estudia, con independencia de la estructura por edades que la misma presente.

El tipo de tabla habitualmente elaborado es el que surge del análisis transversal de la mortalidad, que estudia cómo incide dicho fenómeno en los efectivos de población clasificados por edades o grupos de edades, en un momento dado.

Dada la evolución que generalmente experimenta la mortalidad, sin cambios bruscos, estas tablas constituyen una descripción del fenómeno aceptable para períodos cortos de tiempo, próximos al momento para el que se construyen.

Para el cálculo de las funciones de una tabla completa de mortalidad corriente o de período, es preciso disponer de la información sobre fallecidos y población clasificados ambos por edades y referidos a un mismo período de tiempo.

Dado que las cifras de defunciones clasificadas por edad son de pequeña magnitud (a excepción de las edades más altas), no sólo en las provincias y las comunidades autónomas, sino también a nivel nacional, sobre ellas repercuten de forma notable tanto los errores de recuento, como las posibles perturbaciones que, en un año dado y de forma excepcional, afectan al fenómeno de la mortalidad. Por ello, se hace necesario eliminar estas anomalías que, de permanecer en los datos, darían una imagen falsa del fenómeno estudiado. Esto se lleva a cabo en una primera etapa, considerando, para el cálculo de la tabla de mortalidad referida a un momento dado, y en cada grupo de edad, la media de las defunciones correspondientes a un cierto número de años (generalmente dos o cuatro), centrada en dicho momento.

En una segunda etapa, se hace preciso eliminar las perturbaciones que, tanto en las cifras de fallecidos como en las de población, son consecuencia de errores en la declaración de la edad, y que producen un aumento de los valores observados a ciertas edades en detrimento de las contiguas, ocasionando distorsiones en la serie de probabilidades de muerte de la tabla de mortalidad. Habitualmente, se soslaya este problema mediante la aplicación de algún procedimiento de suavizado a las cifras originales.

2 Obtención de la serie de probabilidades de muerte

La probabilidad de muerte a la edad x, q_x, se define como la probabilidad que tiene un individuo perteneciente a una generación dada, con edad exacta x, de morir antes de alcanzar la edad x+1. Se han de considerar, por tanto, los casos posibles de muerte, es decir, los individuos expuestos a morir, así como los hechos reales de la misma, es decir, las muertes ocurridas en estas condiciones de edad y generación. Los casos posibles son los individuos que llegan a cumplir x años, que se calculan como suma de los habitantes que a final de año tienen esa edad y la mitad de los fallecidos con edad x durante el año considerado, ya que se supone que las muertes se distribuyen uniformemente a través del mismo. Admitida la hipótesis de que las muertes de una generación con edad x ocurren la mitad en un año y la otra mitad en el siguiente, la probabilidad de muerte vendría expresada por:

$$q_x = \frac{1/2\left(D_x^z + D_x^{z+1}\right)}{P_x^z + 1/2\left(D_x^z\right)}$$

donde:

D_x^z representa las defunciones ocurridas en el año z a la edad x.

D_x^{z+1} representa las defunciones ocurridas en el año z+1 a la edad x.

P_x^z es la población a 31 de diciembre del año z con edad x.

La anterior expresión se ha utilizado para el cálculo de las q_x correspondientes a las edades de dos a noventa años, ambas inclusive.

Dado que las defunciones de menores de un año de edad se concentran en las primeras semanas de vida, no es posible aplicar la hipótesis de distribución uniforme a través del año, por lo cual se ha calculado la probabilidad de muerte mediante la expresión:

$$q_0 = \frac{D^{z}_{0,g(z)} + D^{z+1}_{0,g(z)}}{P^{z}_0 + D^{z}_{0,g(z)}}$$

siendo:

$D^{z}_{0,g(z)}$ las defunciones ocurridas en el año z, con 0 años, de la generación de ese año.

$D^{z+1}_{0,g(z)}$ las defunciones ocurridas en el año z+1, con 0 años, de la generación nacida el año anterior.

P^{z}_0 es la población a 31 de diciembre del año z con edad 0.

Por la misma razón, a la edad uno, q_1 se ha calculado mediante:

$$q_1 = \frac{D^{z}_{1,g(z-1)} + D^{z+1}_{1,g(z-1)}}{P^{z}_1 + D^{z}_{1,g(z-1)}}$$

siendo:

$D^{z}_{1,g(z-1)}$ las defunciones ocurridas en el año z, con 1 año, de la generación z-1.

$D^{z+1}_{1,g(z-1)}$ las defunciones ocurridas en el año z+1, con 1 año, de la generación z-1.

P^{z}_1 es la población a 31 de diciembre del año z con edad 1.

El bajo número de fallecidos registrado en cada una de las edades superiores a los noventa años y una mayor repercusión de los errores en la declaración de la edad, provocan distorsiones en la serie de probabilidades de muerte a las mencionadas edades; por ello, estas últimas se han estimado mediante el ajuste de una parábola de tercer grado, por mínimos cuadrados, a partir de las q_x calculadas mediante la expresión anterior, para

x = 90, 91, 92, 93 y 94.

Para efectuar el mencionado ajuste se establecieron las siguientes condiciones: a) La parábola cúbica pasa por el punto q_{90}, lo que conlleva la continuidad de las q_x ajustadas con las calculadas a edades inferiores a los 90 años, b) el valor q_{110} = 1, truncándose la parábola a partir de ese punto, lo que supone, a priori, que no existen supervivientes mayores de ciento diez años, y c) la cúbica tiene tangente paralela al eje de las x en el punto x = 110, lo que supone un crecimiento acelerado de la mortalidad a partir del punto de inflexión de la cúbica, con objeto de tener altas mortalidades en las edades cercanas a ciento diez.

3 Obtención de las series derivadas

A partir de la serie de probabilidades de muerte pueden deducirse las funciones de las tablas de mortalidad que se describen seguidamente.

PROBABILIDAD DE VIDA O SUPERVIVENCIA A LA EDAD x, p_x

Es la probabilidad de supervivencia entre dos edades exactas. Por tanto, para cada edad x,

$$p_x = 1 - q_x$$

SUPERVIVIENTES CON x AÑOS, l_x

Es el número de individuos que alcanzan la edad exacta x de entre l_0 de partida de la tabla de mortalidad. Por tanto, a cada edad x,

$$l_x = l_{x-1} p_{x-1}$$

Habitualmente se toma l_0 = 100.000.

DEFUNCIONES TEÓRICAS CON x AÑOS, d_x

Son las defunciones ocurridas entre dos edades exactas x y x+1, deducidas de la tabla de mortalidad.

Por tanto, a cada edad x,

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

ESPERANZA DE VIDA A LA EDAD x, e_x

Es el número medio de años de vida futura a cada edad exacta x, para los supervivientes que alcanzan dicha edad, bajo el supuesto de que los años vividos por todos ellos se reparten por igual entre los mismos.

Bajo la hipótesis de que los individuos que fallecen a una cierta edad viven, por término medio, la mitad del año en que se produce la muerte, la esperanza de vida se calcula como

$$e_x = \frac{1}{2} + \frac{1}{l_x} \sum_{i=x+1}^{\infty} l_i$$

representando ∞ la edad más alta, en la cual se considera que no hay supervivientes.

4 Obtención de las series perspectivas

Además de las anteriores series o funciones biométricas clásicas de las tablas de mortalidad, se ha considerado de sumo interés la inclusión de las dos series perspectivas que se especifican a continuación.

SUPERVIVIENTES CON x AÑOS CUMPLIDOS, L_X

Representa el número de supervivientes de la tabla de mortalidad con x años cumplidos. La estimación de esta función se ha realizado mediante la siguiente fórmula (ver Introduction to the Mathematics of Population. Keyfitz. Addison-Wesley):

$$L_x = \frac{13}{24}(l_x + l_{x+1}) - \frac{1}{24}(l_{x-1} + l_{x+2})$$

para x = 1, 2, ..., 98.

Para las restantes edades

$L_0 = a_0 l_0 + a_1 l_1$, con $a_0 + a_1 = 1$

siendo

$$a_0 = \frac{D_{0,g(z)}^{z+1}}{D_{0,g(z)}^{z+1} + D_{0,g(z+1)}^{z+1}}$$

donde

$D_{0,g(z)}^{z+1}$ representa las defunciones de menores de un año ocurridas en el año z+1 entre los nacidos de la generación g(z).

Para x = 99 y x = 100

$$L_{99} = e_{99}\, l_{99} - e_{100}\, l_{100}$$

$$L_{100} = e_{100}\, l_{100}$$

donde L_{100} son los supervivientes con 100 y más años cumplidos.

PROBABILIDAD DE SUPERVIVENCIA CON x AÑOS CUMPLIDOS, T_X

Es la probabilidad de sobrevivir entre las edades x y x+1, para los individuos con x años cumplidos. Por tanto, se deduce fácilmente de la anterior mediante

$$T_x = \frac{L_{x+1}}{L_x}$$

y, para el efectivo de población con 99 y más años, la probabilidad de alcanzar 100 y más años es

$$T_{99} = \frac{L_{100}}{L_{99} + L_{100}}$$

5 Síntesis del procedimiento de suavizado empleado

Tanto los stocks de población obtenidos a partir de los censos de población y renovaciones padronales, como los datos sobre defunciones provenientes del Movimiento Natural de la Población, adolecen, en ciertas ocasiones, de defectos en la declaración de las edades, aumentando los valores en algunas de ellas en detrimento de los correspondientes a edades próximas; ello ocasiona distorsiones en las series de probabilidades de muerte calculadas. Para evitarlo, se hace preciso un suavizado de los datos originales previa a su utilización.

El procedimiento empleado para el suavizado de las cifras originales ha sido el Método de las Diferencias Variantes. Dicho método ha sido empleado por el Instituto Nacional de Estadística en la elaboración de las tablas completas de mortalidad anteriores. Una exposición completa de su aplicación, con abundante bibliografía sobre el tema, puede encontrarse en el libro de G. Tintner, The Variate Difference Method, 1940, de la colección de la Cowles Comission. En los párrafos siguientes se expone brevemente su fundamento.

La hipótesis básica de partida para la aplicación del método es que la serie observada es la superposición aditiva de otras dos series, una que expresaría el valor correcto o esperado en cada edad x, y otra que sería una perturbación aleatoria que distorsiona el valor observado. Esta última sería, en este caso, la suma de todas aquellas causas y circunstancias que provocan declaraciones erróneas de la edad.

El modelo es, por tanto, del tipo:

$$y_x = u_x + e_x$$

donde a cada edad x:

y_x es el valor observado.

u_x es el valor esperado o correcto.

e_x es el error o perturbación aleatoria.

En la presente aplicación se ha supuesto que los valores u_x siguen una tendencia suave, sin bruscos zigzags, y que los errores aleatorios son independientes entre sí, hipótesis que podría suavizarse por la de incorrelación de los errores aleatorios.

Una segunda hipótesis fundamental, que ha permitido aplicar el Método de las Diferencias Variantes, consiste en suponer que la esperanza matemática u_x no es sino una polinomial de grado n, con n desconocido. El método de las diferencias variantes determina precisamente cuál es el valor de n. Posteriormente, una vez obtenido n, se ajusta una polinomial de dicho grado a los datos observados y_x. A este respecto hay que mencionar la existencia de una estrecha relación entre el método de las medias móviles y el de las diferencias variantes. Concretamente, M.G. Kendall (A Theorem in Trend Analysis, Biometrika, vol. 48, 1.961. Advanced Theory of Statistics) ha demostrado que los cálculos de medias móviles provienen de la aplicación del método de las diferencias variantes sobre una combinación lineal de algunos términos sucesivos de los valores observados y_x. Con más precisión, toda fórmula de medias móviles se traduce en el ajuste de 2K + 1 términos sucesivos de un polinomio de grado p - 1, existiendo 2K - p + 1 números b_i (con $p - K < i < K$), tales que

$$\hat{u}_x = y_x - \Delta^p \left(\sum_{i=p-k}^{k} b_i y_{x+i} \right)$$

siendo,

Δ^p la diferencia de orden p.

$\hat{u}_x$ el valor estimado de u_x.

b_i los coeficientes de la fórmula de suavización de Sheppard.

Supuesto que la esperanza matemática u_x sigue una polinomial de grado n, se trata de determinar este último. Para ello, se aplica un proceso iterativo en el cual se calculan las sucesivas diferencias finitas. Evidentemente, llegará un momento en dicho proceso en que la esperanza matemática u_x desaparecerá, al anularse el polinomio de grado n; es decir, en cierto momento, la correspondiente diferencia será constante, anulándose, por consiguiente, las siguientes diferencias. Sin embargo, dado que los cálculos se realizan con valores observados y_x, es preciso saber en qué momento puede suponerse que la esperanza matemática se ha eliminado en este proceso de sucesivas diferencias finitas, quedando sólo un residuo proveniente de la existencia de errores aleatorios e_x. La respuesta a esta pregunta se realiza por la siguiente consideración: si se tiene una serie temporal que consiste solamente en el elemento aleatorio, entonces las varianzas de las series sucesivas de diferencias finitas son iguales, una vez corregidas por la multiplicación de un coeficiente binomial debido a que la serie, al ser aleatoria, no está ordenada en el tiempo. De aquí que la varianza de la primera y segunda diferencias sea la misma que la de la serie original.

De lo expresado anteriormente se obtiene un criterio para determinar cuándo ha desaparecido la esperanza matemática u_x. Si se calcula una cierta diferencia k tal que su varianza sea igual a la de la k + 1 diferencia, e igual a la de la k + 2, etc., entonces se puede afirmar que se ha eliminado la esperanza matemática u_x, tomando la k-ésima diferencia. Sin embargo, la igualdad entre varianzas nunca llega a alcanzarse, puesto que siempre queda un residuo de variación aleatoria pero, dado que se trabaja en un modelo de probabilidad, está demostrado que solamente se requiere que la diferencia entre las varianzas de dos series sucesivas de diferencias finitas sea más pequeña que tres veces el error estándar de la diferencia más baja.

Una vez determinado el grado de la polinomial a ajustar, sólo resta aplicar la correspondiente media ponderada con los coeficientes de la fórmula de suavización de Sheppard. El tipo de media móvil a utilizar se determina de la siguiente manera: si el elemento no aleatorio o esperanza matemática u_x queda más o menos eliminado en la primera o segunda diferencia, tomamos n = 1, o una media móvil que es equivalente a ajustar una línea recta a un cierto número (sin determinar por el método) de valores observados y_x consecutivos. Si la esperanza matemática queda eliminada tan sólo en la tercera o cuarta de las diferencias finitas, tendremos n = 2, y elegiremos una media móvil equivalente a ajustar una parábola de segundo grado a un cierto número de valores observados consecutivos. Si el elemento no aleatorio se elimina en la quinta o sexta diferencias, n = 3, realizaremos una media ponderada equivalente a ajustar una polinomial de tercer grado (cúbica) a un número seleccionado de valores observados consecutivos, etc. Si el elemento no aleatorio se elimina en la k-ésima diferencia finita, entonces n = k/2, cuando k es par, o n = (k +1)/2, cuando k es impar.

Según se ha indicado anteriormente, las medias móviles se aplican sobre un número determinado de valores observados y consecutivos, convenientemente centrados. Sin embargo, este número queda indeterminado. El criterio a seguir queda abierto a la experiencia y a la naturaleza concreta del problema a resolver. No obstante, el número de valores incluido en cada media debe tomarse respetando la longitud del ciclo principal que trata de eliminarse. En nuestro caso, se han tomado medias móviles sobre cada conjunto de cinco valores observados y_x consecutivos.

Para la aplicación a la construcción de las tablas de mortalidad, la serie de valores esperados siempre ha desaparecido en la primera o segunda diferencias, lo cual implica aplicar siempre unas simples medias móviles para la suavización de las series originales.

6 Tablas calculadas e información de base utilizada

Se han calculado, anualmente, tablas de mortalidad para la población de España y sus comunidades autónomas, para el periodo 1992-2005. En el caso de Ceuta y Melilla, se han obtenido para la población conjunta de ambas y, desde el año 2002, separadamente para las dos ciudades autónomas. En todos los ámbitos geográficos considerados se han obtenido tablas para las poblaciones de varones, de mujeres y total.

Las defunciones empleadas en el cálculo de las tablas de mortalidad(nacionales y autonómicas) de cada año, se han obtenido como promedio de las cifras por edad registradas en el Movimiento Natural de la Población de cada dos años consecutivos, el de referencia y el anterior, tanto para los varones, como para las mujeres, como para las tablas de la población total.

Las irregularidades que estas cifras originales de defunciones conllevan, como consecuencia de los posibles errores en su clasificación por edad, se han eliminado mediante el procedimiento de suavizado descrito en el punto anterior.

Las poblaciones por comunidades autónomas, sexo y edades simples, referidas a 1 de enero de cada año, que se han empleado, son las correspondientes a las *Estimaciones intercensales de población* obtenidas entre los Censos de Población de 1991 y 2001 y a las *Estimaciones de la población actual* calculadas a partir del último censo mencionado.

Las cifras empleadas se publican junto con las funciones biométricas de las tablas de mortalidad calculadas.

Glosario de Símbolos

Q(X) = Riesgo de muerte o probabilidad de fallecer entre las edades exactas X y X+1.

L(X) = Supervivientes que alcanzan la edad exacta X de entre 100.000 iniciales.

D(X) = Defunciones teóricas ocurridas entre cada dos edades exactas X y X+1.

E(X) = Esperanza de vida a la edad exacta X.

LL(X) = Supervivientes con X años cumplidos.

T(X) = Probabilidad de supervivencia entre las edades cumplidas X y X+1.