XIV Jornadas Argentinas de Estudios de Población. Asociación de Estudios de Población de la Argentina, Santa Fe, 2017.

Análisis comparativo de la mortalidad por edad y género (1980-2014) y pronósticos a 20 años para CABA y Rosario, mediante el uso de modelos probabilísticos. .

Lucía Andreozzi.

Análisis comparativo de la mortalidad por edad y género (1980-2014) y pronósticos a 20 años para CABA y Rosario, mediante el uso de modelos probabilísticos.

Mg. Lucía Andreozzi (UNR – CONICET)

Objetivos

Este trabajo constituye una propuesta, plausible de ser empleada por organismos oficiales, la cual combina la teoría demográfica con métodos estadísticos de series de tiempo para describir el comportamiento histórico de la tasa de mortalidad por edad simple y sexo en la Ciudad Autónoma de buenos Aires (CABA) y Rosario, y, además pronosticar, con métodos probabilísticos la esperanza de vida.

Metodología

En este trabajo se aplican; el precursor de los modelos probabilísticos de pronóstico en mortalidad: el modelo de Lee y Carter (1992) y uno de los últimos modelos propuestos en el área: el modelo para datos funcionales propuesto por Hyndman y Ullah (2008).

Al aplicar LC a tasas de mortalidad, las mismas se descomponen en dos parámetros a y b (también llamado base) y un índice k_t que representa el nivel general de la mortalidad. La premisa básica del modelo es que existe una relación lineal entre el logaritmo de las tasas específicas de mortalidad $m_{x,t}$ y dos factores explicativos: el intervalo de edad, x , y el tiempo, t. La ecuación que describe esto es

$$m_{x,t} = \exp(a_x + b_x k_t + e_{x,t}) \quad t = 1, \dots, n \quad x = 1, \dots, \omega$$

aplicando logaritmo

$$f_{x,t} = \ln(m_{x,t}) = a_x + b_x k_t + e_{x,t} \quad t = 1, \dots, n \quad x = 1, \dots, \omega$$

donde,

$m_{x,t}$: tasa específica de mortalidad para el intervalo de edad x y año t,

a_x: parámetro de forma, el conjunto de dichos parámetros describe el patrón de mortalidad según la edad,

b_x: parámetro de sensibilidad. Representa el cambio en la mortalidad en el intervalo que se inicia a la edad x, frente a cambios en el índice k_t,

$e_{x,t}$: error aleatorio,

ω: es el inicio del último intervalo de edad.

Con a_x se denomina el parámetro de "forma" y el conjunto de las estimaciones de dicho parámetro para cada uno de los grupos etarios describe la forma general o el patrón de las tasas de mortalidad específicas por edad. Dicho parámetro se estima como el promedio aritmético simple sobre el tiempo, de los logaritmos de las tasas específicas de cada edad.

El parámetro b_x (o base, o función base), llamado de "sensibilidad" describe el cambio en la mortalidad en el intervalo de edad x, frente a cambios en el índice k_t. El parámetro de sensibilidad representa la intensidad en el crecimiento o decrecimiento de la tasa de mortalidad, para un grupo de edad a través del tiempo. Para garantizar la unicidad de la solución a la suma de los valores de b_x se le impone el valor 1, y a la suma de los valores k_t se les impone sumar cero. La estimación se realiza utilizando Descomposición en Valores Singulares (DVS), y si bien existen otras propuestas para la estimación, tales como mínimos cuadrados o máxima verosimilitud, no se obtienen grandes ganancias con respecto al uso de DVS con un costo computacional superior.

En relación al MDF Se supone el siguiente modelo para las observaciones transformadas $y_t(x)$:

$$y_t(x) = s_t(x) + \sigma_t(x)\varepsilon_{t,x}$$
$$s_t(x) = \mu(x) + \sum_{k=1}^{K} \beta_{t,k}\phi_k(x) + e_t(x),$$

donde $s_t(x)$ es una función suave subyacente de $x, \varepsilon_{t,x}$ son variables aleatorias, independientes e idénticamente distribuidas y la definición de $\sigma_t(x)$ permite a la variancia cambiar con la edad y con el tiempo. Esto significa que las observaciones transformadas son la suma de la cantidad a modelar, $s_t(x)$, una función suave de la edad y un error (Primer ecuación). La segunda ecuación describe la dinámica de $s_t(x)$ a través del tiempo, en esta ecuación $\mu(x)$, es la media de $s_t(x)$ a través de los años, $\{\phi_k(x)\}$ es un conjunto de K funciones base ortogonales calculadas utilizando una descomposición en componentes principales funcionales de la matriz $[\hat{s}_t(x) - \hat{\mu}(x)]$ y $e_t(x)$ es el error del modelo (el cual se supone no correlacionado serialmente). La dinámica del proceso está controlada por los coeficientes $\{\beta_{t,k}\}$, los cuales tienen un comportamiento independiente uno de otro (garantizado por la utilización del método de componentes principales).

Un aporte interesante a este enfoque lo hacen Hyndman et al. (2013) al introducir la idea de pronósticos coherentes en el paradigma de datos funcionales. La idea principal de esta propuesta radica en que la diferencia entre los pronósticos de grupos de interés debe permanecer constante a través del tiempo, reproduciendo la relación presente en los datos observados. Los grupos de interés pueden ser subregiones geográficas o sexos, por mencionar algunas posibilidades.

Fuentes

Los insumos básicos de ambos modelos (LC y MDF) los constituyen las cifras de población y el número de defunciones por edad y género en el período 1980-2014, ambos proporcionados por la Dirección General de Estadística y Censos de la Ciudad de Buenos Aires, en el caso de CABA. En relación a Rosario, las defunciones las proporciona la Dirección General de Estadística de la Ciudad de Rosario y las cifras de población son suministradas por el INDEC La información se haya desagregada en 19 grupos 0 años, 1 a 4 años y grupos quinquenales, 5 a 9, 10 a 14, etc. hasta el grupo abierto final "85 años y más". Para la realizar los cálculos necesarios se emplea el software R (Development Core Team, 2008).

Resultados esperados de la investigación.

En primer lugar se realiza una breve reseña acerca de la mortalidad en ambas ciudades. La misma sirve para situar el modelo en un contexto demográfico adecuadamente descripto. Luego se aplicarán a ambas matrices de datos los modelos LC y MDF. De ellos se obtiene una descomposición en bases y coeficientes que permite distinguir que edades dominan el comportamiento de la mortalidad a través del tiempo y en qué forma lo hacen. Por otro lado se pretende pronosticar la componente temporal, mediante modelos para series de tiempo, para, de este modo, obtener las tasas de mortalidad por edad y género para un horizonte de 20 años y la esperanza de vida al nacer para el mismo período, ambas medidas acompañadas de sus intervalos de confianza. Por otro lado el trabajo evaluará que modelo se adapta mejor a cada región evaluando medidas de bondad de ajuste y de pronóstico.

Para ser considerado para sesión regular – acepto Póster