

FACTORES DE MEJORA GENERACIONAL

Guía R

Construcción factores de mejora generacional

Fecha: 31/10/2017

Tabla Mortalidad: Factores de mejora generacional

Contenido: Tratamiento y modelización

Descripción: guía detallada de los procesos R para la construcción de los factores de mejora nacionales.

[Índice](#)

01. Introducción	3
02. Modelización y proyección de la mortalidad	3
03. Construcción de los factores de mejora generacionales.....	5

01. Introducción

En el proceso R "*Factores Mejora Generacionales.R*" se realizan los procedimientos necesarios para el tratamiento de las BBDD originales obtenidas del INE y que serán la base para la construcción de los factores de mejora generacionales de la mortalidad en el ámbito nacional.

Este proceso R se desglosa en dos secciones:

1. Modelización y proyección de la mortalidad
 01. Carga de Datos.R
 02. Ajuste LeeCarter 40 a 89.R
 03. Ajuste CBD 40 a 89.R
 04. Ajuste Pspline 40 a 89.R
 05. Ajuste LeeCarter 90 a 95.R
2. Construcción de los factores de mejora generacionales
 06. Calculo Factores Mejora.R
 07. Suavizado Factores Mejora y Proyeccion.R

Adicional a los procesos de cálculo, se parametrizan el horizonte de proyección, la edad máxima que se modeliza y la proyección de los factores de mejora. Por último, se almacenan los resultados obtenidos en "*Fact_Mej_120.txt*".

Los paquetes de R necesarios para el proceso de cálculo de los factores de mejora son:

- StMoMo
- Mgcv
- MortalitySmooth
- Abind

02. Modelización y proyección de la mortalidad

Procedimiento inicial donde se realiza la lectura de la información publicada en las bases del INE, modelización de la mortalidad utilizando todo el histórico disponible y proyección de la misma.

Los apartados que componen esta primera sección son:

01. Carga de Datos.R

Definición de la ruta donde se almacenan los datos previamente descargados del INE. Estos datos son:

- Dx_ESP.txt: Defunciones teóricas para el periodo de 1991 a 2015.
- Ex_ESP: Población estacionaria para el periodo de 1991 a 2015.
- Lx_ESP: Supervivientes para el periodo de 1991 a 2015.
- Dx_ESP_1975_1990.txt: Defunciones teóricas para el periodo de 1975 a 1990.
- Ex_ESP_1975_1990: Población estacionaria para el periodo de 1975 a 1990.
- Lx_ESP_1975_1990: Supervivientes para el periodo de 1975 a 1990.

La información del INE se puede descargar de los siguientes enlaces:

- 1975 a 1990: <http://www.ine.es/jaxi/Tabla.htm?path=/t20/p319a/serie/p02/l0/&file=01001.px&L=0>
- 1991 a 2015: <http://www.ine.es/jaxi/Tabla.htm?path=/t20/p319a/serie/p02/l0/&file=02001.px&L=0>

Finalmente, se obtienen un conjunto de subtablas para la modelización de la mortalidad:

- Dth – Exp – lx: contiene los datos de defunciones, población y supervivencia, respectivamente, para el histórico de 1975 a 2015 y edades de 40 a 89 años.
- Dth_total – Exp_total – lx_total: contiene los datos de defunciones, población y supervivencia, respectivamente, para el histórico de 1991 a 2015 y edades de 40 a 95 años.

02. Ajuste LeeCarter 40 a 89.R

En este programa se ajusta un modelo Lee-Carter para las edades de 40 a 89 años utilizando el histórico de 1975 hasta 2015 para finalmente proyectar la mortalidad doce años y calcular la mortalidad estresada al 99,5% durante ese periodo.

Se introducen los datos en un objeto R tal que las librerías de trabajo puedan interpretarlo y ajustar la mortalidad. Se utiliza un objeto ya creado denominado "EWMaleData".

El modelo Lee-Carter asume una distribución Binomial, por lo que la función link es el logit. Para aplicar este modelo se utiliza la función "fit()".

Se obtienen las tasas de mortalidad ajustadas por el modelo Lee-Carter mediante la función "fitted()", cuyo argumento es el objeto que contiene el ajuste del modelo.

Para la proyección de la mortalidad se utiliza un paseo aleatorio bivalente con deriva, indicando el horizonte de proyección deseado. La función "forecast()" por defecto ajusta un paseo aleatorio con deriva.

Dado que para el modelo Lee-Carter no es posible obtener un intervalo de confianza para las predicciones, se utiliza la técnica "Bootstrap" para calcular la incertidumbre de las proyecciones. Para ello se generan 1000 muestras aleatorias de una Binomial con probabilidades dadas por las que se obtienen al ajustar el modelo a los datos.

Una vez se han generado las muestras aleatorias, se ajusta un modelo Lee-Carter a cada una de ellas, obteniéndose los parámetros del modelo para cada una. Esto se realiza con la función "bootstrap()".

Con los parámetros obtenidos se utiliza un paseo aleatorio con deriva para proyectar las tasas en el horizonte considerado. Para ello se utiliza la función "simulate()".

Finalmente, se seleccionan los percentiles 0,05% y 99,5% para obtener las tasas de mortalidad estresadas.

03. Ajuste CBD 40 a 89.R

En este programa se ajusta un modelo CBD para las edades de 40 a 89 años utilizando el histórico de 1975 hasta 2015 para finalmente proyectar la mortalidad doce años y calcular la mortalidad estresada al 99,5% durante ese periodo.

El proceso de ajuste y proyección es idéntico al de "02. Ajuste LeeCarter 40 a 89.R". El único cambio se produce en la función `"fit()"`, donde se define el ajuste `"cbd()"`.

04. Ajuste Pspline 40 a 89.R

En este programa se ajusta un modelo Pspline para las edades de 40 a 89 años utilizando el histórico de 1975 hasta 2015 para finalmente proyectar la mortalidad doce años y calcular la mortalidad estresada al 99,5% durante ese periodo.

El modelo Pspline asume una distribución Poisson, por lo que la función link es el logaritmo. Para aplicar este modelo se utiliza la función `"Mort2Dsmooth()"`, cuyos argumentos son las edades, los años y la matriz de defunciones y de población.

Para el cálculo de Pspline es necesario definir en cuantos segmentos dividir la edad y el periodo. El estándar es utilizar entre 10 y 15 divisiones. En este caso se han utilizado 10 divisiones para la edad y 11 para el periodo.

Tras el ajuste del Pspline se utiliza la función `"my.predict()"` que permite la proyección de la mortalidad y la obtención del error estándar de las predicciones para el cálculo de los intervalos de confianza.

La función `"my.predict()"`, estándar en la comunidad de programadores R, es una función para la proyección de modelos estocásticos.

05. Ajuste LeeCarter 90 a 95.R

En este programa se ajusta un modelo Lee-Carter para las edades de 40 a 95 años utilizando el histórico de 1991 hasta 2015 para finalmente proyectar la mortalidad doce años y calcular la mortalidad estresada al 99,5% durante ese periodo. No obstante se seleccionan los resultados para las edades 90 a 95 como los de interés, pues el resto de edades es modelizado en "02. Ajuste LeeCarter 40 a 89.R".

El proceso de ajuste y proyección es idéntico al de "02. Ajuste LeeCarter 40 a 89.R". El único cambio se produce en la selección de edades y periodo inicial.

03. Construcción de los factores de mejora generacionales

En este procedimiento se realiza el cálculo de los factores de mejora generacionales en el marco nacional a través de la mortalidad modelizada y proyectada en el proceso anterior.

Los apartados que componen esta segunda sección son:

06. Calculo Factores Mejora.R

En este programa se realiza el cálculo de los factores de mejora "brutos", a través de las proyecciones de mortalidad obtenidas anteriormente.

En primer lugar se define una función auxiliar para el cálculo de los factores de mejora tal que

$$\lambda_x = - \left(\frac{\hat{q}_{x,m}}{\hat{q}_{x,1}} \right)^{(-1/m)}$$

Siendo m el horizonte de proyección $\hat{q}_{x,1}$ la primera tasa de mortalidad proyectada y $\hat{q}_{x,m}$ la última.

A continuación se realiza el cálculo de los factores de mejora generacionales para las proyecciones de mortalidad y mortalidad estresada al 99,5% de Lee-Carter, que previamente se han unido las proyecciones de 40 a 89 años con las proyecciones de 90 a 95 años, CBD y Pspline.

Por último, se obtienen los factores de mejora y factores de mejora estresados medios para las edades de 40 a 89 años, que son el valor medio de los obtenidos para cada modelo.

07. Suavizado Factores Mejora y Proyeccion.R

En este programa se realiza un suavizado sobre los factores de mejora brutos obtenidos para asegurar una transición suave y que decrecen para las edades avanzadas. Por último se proyectan estos factores hasta la edad 120.

El primer paso es ajustar un GAM sobre los factores brutos, los cuales son suavizados con splines. Este suavizado se realiza con la función `"gam()"`. El suavizado se realiza tanto a los factores de mejora como a los factores de mejora estresados.

Para realizar la proyección de los factores, se observa que desde la edad 85 hasta la edad 95 estos siguen una línea recta, por lo que se ajusta esta y se proyecta. Para el ajuste se realiza la función `"lm()"` la cual ajusta un modelo lineal a los factores seleccionados tal que

$$\lambda = \alpha + \beta * edad$$

Este ajuste se realiza tanto para los factores de mejora como los factores de mejora estresados. Aplicando esta fórmula a las edades desde 96 hasta 120 se obtienen los factores correspondientes a estas edades.