

## PRONÓSTICOS PROBABILÍSTICOS DE LA ESPERANZA DE VIDA PARA LA CIUDAD DE ROSARIO, ARGENTINA

Lucía Andreozzi\*

Nora Ventroni\*\*

**Resumen.** La esperanza de vida al nacer es uno de los índices más útiles para medir el nivel general de la mortalidad y estima el nivel de la mortalidad con más precisión que la tasa bruta de mortalidad porque es independiente de la estructura por edad de la población. Además permite comparar los niveles de distintas poblaciones en distintos momentos históricos. Los modelos de pronósticos probabilísticos generan tasas de mortalidad por edad para períodos futuros, y a partir de estos resultados es posible, derivar los pronósticos de la esperanza de vida al nacer, y sus correspondientes intervalos de confianza. En este trabajo se aplica uno de los primeros modelos probabilísticos de pronóstico en mortalidad como es el modelo de Lee y Carter (1992) y una de las últimas metodologías propuestas en el área, el modelo para datos funcionales desarrollado por Hyndman y Ullah (2008). Ambos modelos pronostican la mortalidad y permiten, a través de tablas de vida basadas en los pronósticos de las tasas de mortalidad, derivar la esperanza de vida al nacer. La aplicación se realiza a datos de mortalidad de la Ciudad de Rosario, en el período 1980-2015, para obtener pronósticos puntuales y por intervalos de la esperanza de vida al nacer.

*Palabras clave:* Datos funcionales; Modelo de Lee-Carter; Intervalo de pronóstico.

\* Docente-Investigadora Universidad Nacional de Rosario (UNR), Argentina. Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina.

**Contacto:** andreozzi.lu@gmail.com

\*\* Docente-Investigadora Universidad Nacional de Rosario (UNR), Argentina. Municipalidad de Rosario, Argentina.

**Contacto:** nventroni@gmail.com

## CLIFE EXPECTANCY'S PROBABILISTIC FORECASTS FOR ROSARIO CITY, ARGENTINA

**Abstract.** Life expectancy at birth is one of the most useful indices for measuring the overall level of mortality and estimates the level of mortality more accurately than the crude mortality rate because it is independent of the age structure of the population. It also allows to compare the levels of mortality for different populations in different historical moments. Probabilistic forecasting models generate age-specific mortality rates for future periods, and from these results it is possible to derive forecasts of life expectancy at birth and their corresponding confidence intervals. In this paper two models are applied; the forerunner of probabilistic models of mortality prediction: the Lee and Carter model (1992) and another one, the last proposed in the area: the functional data model developed by Hyndman and Ullah (2008). Both models predict mortality, and allow, through living tables based on mortality forecasts, to derive life expectancy at birth. The application is made to mortality data of Rosario City, from 1980 to 2015, to obtain specific forecasts and intervals for life expectancy at birth with their respective intervals.

*Keywords:* Functional data; Lee-Carter model; Forecast interval.

**Original recibido el 29-11-2016**

**Aceptado para su publicación el 03-04-2016**

## 1. Introducción

Entre la Dirección General de Estadística de la Municipalidad de Rosario y la Escuela de Estadística de la Facultad de Ciencias Económicas y Estadística de la UNR se lleva adelante un proyecto<sup>1</sup> que, entre otros objetivos, se plantea obtener pronósticos probabilísticos para la Esperanza de Vida al Nacer (EVN) para la Ciudad de Rosario. Resulta de una gran importancia el análisis de la mortalidad y la esperanza de vida de una ciudad, región o país porque no sólo brinda indicios sobre la calidad de vida y el estado sanitario de una población, sino que también, el estudio de la evolución y la proyección de dichos indicadores constituyen insumos indispensables para delinear políticas públicas.

La información que se emplea para construir medidas relativas a la mortalidad proviene por lo general de dos fuentes: estadísticas de las defunciones (registro de estadísticas vitales) y estadísticas de la población a la que le ocurren dichas defunciones (generalmente censos).

Las tasas de mortalidad específicas por edad miden, como lo hace la tasa Bruta de Mortalidad, la frecuencia de muertes cada mil habitantes de una población, pero en lugar de basarse en el total de la población, lo hace en la población de una edad determinada o un grupo pequeño de edades. Las tasas específicas miden adecuadamente el nivel de la mortalidad en cada grupo de edad o edad simple, pero, por sí solas, no proporcionan el nivel general de la mortalidad de la población de interés. Un índice que posee gran aceptación para representar el nivel general de la mortalidad es la esperanza de vida al nacer, medida que se obtiene, en la mayoría de los casos a partir de la construcción de una tabla de vida. La EVN permite además comparar los niveles de distintas poblaciones en distintos momentos históricos (Arriaga, 2014). Sin embargo, como Arriaga (2014) también explica, la esperanza de vida mide el nivel de la mortalidad con años de vida y este hecho tiene implicaciones cuando se quiere medir el cambio en la mortalidad.

Por otro lado, la proyección del comportamiento de la mortalidad resulta de utilidad para la planificación en diversas áreas del Estado. Sin embargo, resulta inevitable que una proyección o pronóstico contenga incertidumbre. Si bien en demografía resulta más sencillo describir la situación actual a través de determinados indicadores, las tendencias en los nacimientos, muertes y migraciones están sujetas a cambios impredecibles. Por ésta y otras razones, combinar los métodos demográficos con los estadísticos enriquece los resultados que se pueden obtener.

En este sentido, vale aclarar que los métodos probabilísticos para el pronóstico de la población rápidamente están ganando reconocimiento. Los mismos poseen su principal ventaja en la llamada consistencia probabilística a través de todas las variables pronosticadas y sus índices derivados (Lee y Tuljapurkar, 1994). Esto propició el desarrollo de métodos probabilísticos de pronóstico que, cada vez con mayor aceptación (Alho, 2000), son implementados por las agencias de estadísticas oficiales para producir sus pronósticos nacionales (Holanda y Estados Unidos son ejemplos en este sentido). Otra propuesta ampliamente difundida e implementada por las Naciones Unidas es la de Raftery, Li, Sevč'ikov'a, Gerland y Heilig (2012) que consiste

<sup>1</sup> En el marco de un convenio de colaboración mutua, sin ningún tipo de financiamiento.

en calcular una proyección probabilística de la esperanza de vida mediante modelos jerárquicos bayesianos.

Los mayores avances en modelos probabilísticos de pronóstico se presentan para la mortalidad, en cambio, para la fecundidad y las migraciones existe un número menor de metodologías propuestas. Un pronóstico que incluye distribuciones de probabilidad permite incorporar la noción de incertidumbre de manera explícita a través de una probabilidad, proporcionando de ese modo resultados más informativos. Se pueden pensar los resultados demográficos futuros como valores desconocidos con una distribución de probabilidad. La irrupción más clara y concisa de los métodos probabilísticos de pronóstico en el área demográfica la concretó el método propuesto por Lee y Carter (1992). El mismo posee actualmente numerosas variantes y extensiones. Originalmente los autores proponían una metodología que permite modelar y extrapolar las tendencias observadas en las tasas de mortalidad a largo plazo e implementar dicha metodología para pronosticar hasta el año 2065.

Como síntesis García Guerrero y Ordorica (2012) concluyen de la aplicación del modelo de Lee y Carter a datos de México:

Finalmente es importante resaltar la virtud del método presentado, al permitir tener intervalos de confianza de las estimaciones de la mortalidad. El enfoque estocástico reconoce que el comportamiento demográfico no sigue leyes determinadas, sino que la demografía, que al final de todo estudia conjuntos humanos, está sujeta a las leyes del azar. Además, el método permite realizar ajustes continuos a los pronósticos, pensando también en el número de muertes que se podrían evitar para alcanzar una meta en la esperanza de vida (p. 445).

Además, Guerrero Guzmán y González Pérez (2007) desarrollan un trabajo en base a la aplicación del modelo a datos de México evaluando el impacto de los resultados sobre el sistema previsional. En Chile, Lee y Rofman (1994) realizan una aplicación del modelo, y en sus conclusiones sostienen que la síntesis de técnicas estadísticas de series de tiempo y modelos demográficos tiene múltiples ventajas, y para el caso del modelo de Lee-Carter, este constituye un modelo sencillo y útil para completar datos faltantes y además en algunos casos se emplea para ajustar el subregistro y proyectar la mortalidad.

En un artículo de Andreozzi y Blaconá (2011) se presenta la primera aplicación del modelo de Lee y Carter (1992) a datos de Argentina. Luego, en un avance de la misma línea de trabajo presentan una comparación entre distintos métodos de estimación que surgieron luego como variantes al modelo original (Andreozzi y Blaconá, 2012).

Anteriormente, Hyndman y Ullah (2007) proponen suavizar la mortalidad mediante regresiones spline penalizadas para luego ajustar un modelo utilizando una descomposición en componentes principales, siguiendo un enfoque llamado de datos funcionales, dado que el insumo básico de los modelos es el conjunto de tasas de mortalidad específicas suavizadas.

Si bien los métodos antes presentados generan pronósticos de las tasas de mortalidad, como ya se mencionó, uno de los índices más útiles para medir el nivel general de la mortalidad es la esperanza de vida al nacer. Ésta estima el nivel de la mortalidad con más precisión que la tasa bruta de mortalidad porque es independiente de la estructura por edad de la población.

En este trabajo se aplican, por un lado, el precursor de los modelos probabilísticos de pronóstico en mortalidad: el modelo de Lee y Carter (1992); y por otro lado, uno de los últimos modelos propuestos en el área: el modelo para datos funcionales propuesto por Hyndman y Ullah (2007), sobre datos de mortalidad de la Ciudad de Rosario en el periodo que va desde 1980 a 2015. A través de estos modelos se pretende obtener pronósticos puntuales y por intervalos de las tasas de mortalidad por edad en el periodo 2015-2025. Asimismo a partir de las tasas de mortalidad calculadas se obtienen los pronósticos para la esperanza de vida al nacer con sus respectivos intervalos.

## **2. Metodología**

### **2.1 Modelo de Lee y Carter**

El modelo de Lee y Carter (1992), denotado como LC de aquí en adelante, es un modelo demográfico-estadístico que permite realizar proyecciones de las tasas de mortalidad. En términos generales, se puede definir como un modelo de carácter extrapolativo para una serie de tiempo univariada, ya que no incorpora información acerca de efectos tecnológicos o sociales en la mortalidad, esto es, no busca incorporar información externa (variables exógenas) ni opiniones sobre posibles acontecimientos futuros. El modelo básicamente proyecta la tendencia histórica observada durante el periodo de estudio en el que se dispone de datos. El modelo LC combina un modelo demográfico de pocos supuestos con métodos estadísticos de series de tiempo, por lo cual es capaz de proveer una base probabilística a las proyecciones así como intervalos de predicción. El cambio en el nivel de mortalidad general se representa en el modelo por el cambio de un sólo índice. El modelo LC permite a cada tasa específica de mortalidad decrecer exponencialmente sin restricciones.

Para utilizar el modelo LC se requiere inicialmente revisar la calidad de la información disponible y sus limitaciones, ya que el único insumo es el registro histórico de tasas de mortalidad por grupos de edad. Es importante señalar que el modelo, por construcción, tiene la ventaja de que las tasas no necesitan tener la misma periodicidad. No obstante, se requiere que los datos abarquen un periodo considerable de tiempo, una regla propuesta es que cubran al menos un periodo de 30 años de historia reciente.

Una vez que se han obtenido los datos y se ha validado su calidad, la etapa siguiente del método consiste en ajustar el modelo a los datos originales y evaluar su desempeño histórico.

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

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

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,

$\omega$ : es el inicio del último intervalo de edad.

Siendo  $a_x$  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 le 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 el modelo LC el índice  $k_t$  estimado es lineal lo cual hace que la mortalidad en cada grupo de edad cambie a su propia tasa exponencial. Por lo tanto, en este modelo no pueden presentarse valores negativos de las tasas de mortalidad. Al hacer todas las tasas de mortalidad dependientes del parámetro  $k_t$ , el método difiere mucho de pronosticar cada tasa específica de mortalidad de manera independiente. Esto evita tener que calcular las distintas covariancias entre los errores. Además, todas las tasas individuales están restringidas a pertenecer al sistema general de mortalidad que se ajusta a los datos históricos observados.

Posteriormente, a través de métodos de series de tiempo se pronostica el índice general de mortalidad y, finalmente, se pronostican las tasas de mortalidad específicas por edad, tanto en forma puntual como por intervalos de confianza.

## 2.2 Modelos para datos funcionales

Durante los últimos años se han desarrollado múltiples enfoques para pronósticos

probabilísticos (Booth, Hundman, Tickle y de Jong, 2006) y dentro de esta categoría han cobrado una particular relevancia los métodos para datos funcionales (Ramsay y Silverman, 2005). Estos métodos, de reciente aparición, constituyen un nuevo marco para el análisis de series de tiempo, que ha sido adoptado, entre otras finalidades para realizar pronósticos de todas las componentes demográficas (Hyndman y Ullah, 2007). El modelo para datos funcionales es una extensión del modelo LC, en el sentido de que descompone la mortalidad de acuerdo a su comportamiento en el tiempo y los patrones etarios. La diferencia radica en dos puntos. En primer lugar, el modelo para datos funcionales se aplica siempre sobre los datos suavizados (funciones de los datos observados); y, en segundo lugar, se emplean múltiples bases y no un único índice como se plantea en el caso de LC, denominado  $kt$ .

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 (primera ecuación). La segunda ecuación describe la dinámica de  $s_t(x)$  a través del tiempo donde  $\mu(x)$  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  $[\hat{s}(x) - \hat{\mu}(x)]$  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, Booth y Yasmeen (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.

### 3. Aplicación a la ciudad de Rosario

Los insumos básicos de ambos modelos los constituyen las cifras de población y el número de defunciones<sup>2</sup> por edad y sexo en el período 1980-2015, proporcionados por la Dirección General de Estadística de la Ciudad de Rosario en el caso de las

<sup>2</sup> Resulta interesante destacar que la ciudad de Rosario posee desde hace décadas un nivel de cobertura de los hechos vitales de gran calidad, debido a la coordinación y articulación existente entre los distintos sectores responsables de la elaboración de la información.

defunciones y por el INDEC en el caso de las cifras de población.

La información se haya desagregada en 19 grupos<sup>3</sup>. Para la realizar los cálculos necesarios se emplea el software R (R Development Core Team, 2008).

Inicialmente para la aplicación de los modelos estadísticos (LC y MDF) es necesario realizar un suavizado de las tasas observadas a través de los grupos etarios. En la Figura 1 se presentan las tasas de mortalidad para el total de la población, los varones y las mujeres, insumo base para ambos modelos de pronóstico. Las curvas presentan una mortalidad alta al inicio de la vida, con una reducción marcada luego del nacimiento llegando a los niveles más bajos entre los 5 y 15 años y a partir de ese punto aumenta hasta las edades más avanzadas. En Argentina y también en Rosario se presenta un máximo relativo en torno a las edades de 18 a 25 años en varones. Este valor alto está vinculado, en ésta y en la mayoría de las poblaciones, a accidentes de tránsito, muertes relacionadas al consumo de drogas y muertes violentas en general. En Serfaty, Foglia, Masautis y Negri (2007) se sostiene que en Argentina el grupo de 10 a 25 años abarca un 27% de la población y que en los últimos tiempos se ha profundizado el conocimiento que se tiene acerca de ellos. En ese sentido, concluyen que si bien no se enferman clínicamente con frecuencia, son más vulnerables a las causas de mortalidad vinculadas a la violencia: los accidentes, el suicidio y el homicidio. Luego de este pico las tasas presentan un leve descenso para volver a subir, de forma sostenida, hasta las edades avanzadas.

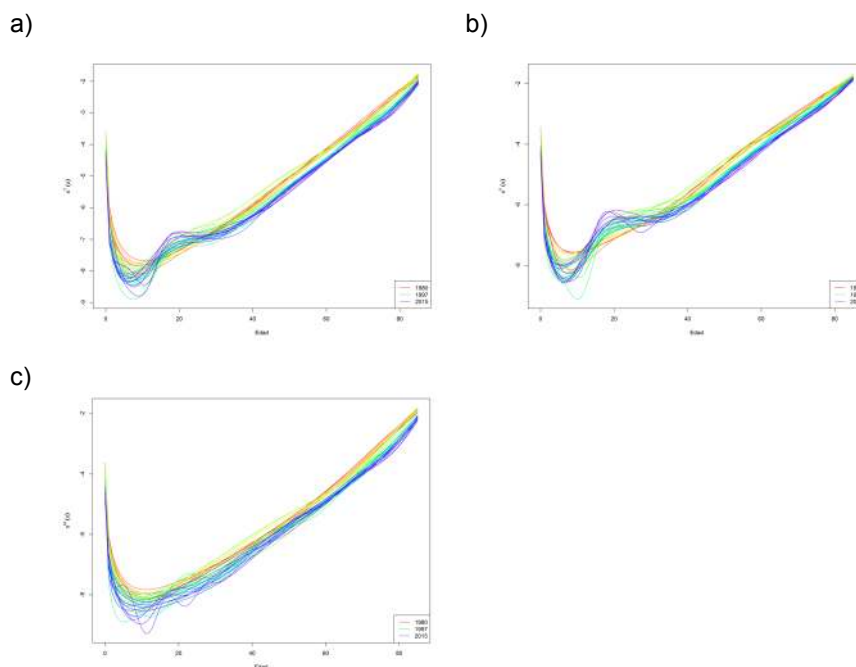
Visualmente las curvas suavizadas permiten ver la caída marcada en los niveles de mortalidad a través del tiempo, que se da en todas las edades, exceptuando al grupo de varones anteriormente mencionado, para el cual dicho comportamiento parece más leve o fluctuante. La caída en general en los niveles de la mortalidad se atribuye principalmente a las mejoras en la medicina cuando se evalúan dinámicas de largo plazo (períodos de estudio de 100 años o más), por lo cual el descenso que se observa en el presente período de tiempo puede obedecer en parte a estas causas y a otras más específicas del proceso histórico particular de la Argentina.

---

3 Un primer grupo de 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"



**Figura 1. Suavizado de las tasas de mortalidad de Rosario, 1980-2015. (a) Total de la población, (b) Varones y (c) Mujeres.**



**Nota:**

Se emplea una escala arco iris implementada por Hyndman y se le agrega una leyenda que incluye referencias al valor mínimo, medio y más alto en la escala de colores, para orientar la lectura.

**Fuente:** elaboración propia en base a datos de la Dirección General de Estadística de la Municipalidad Ciudad de Rosario.

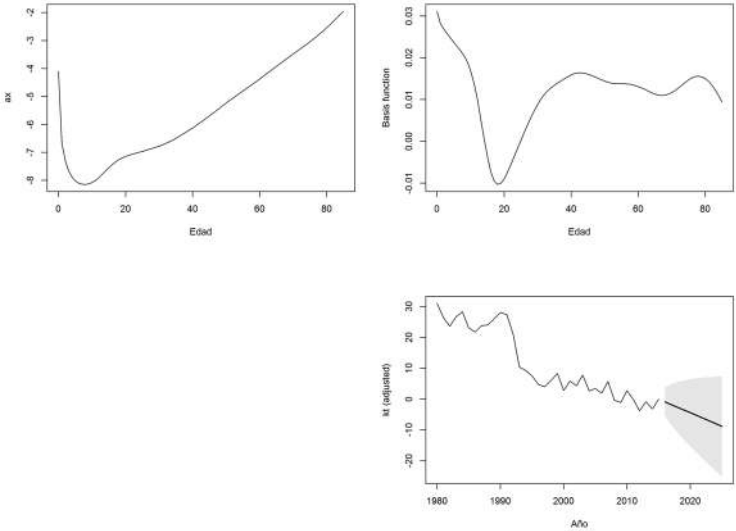
**3. Modelo de Lee y Carter para la mortalidad de la ciudad de Rosario.**

Al aplicar el modelo de LC se obtiene un índice general de mortalidad  $kt$ , que representa el comportamiento global de la mortalidad a través del tiempo (Figura 2). En gris sombreado se representa el intervalo de pronóstico y se puede observar que, más allá de las fluctuaciones, la mortalidad descende en el período en estudio con una estructura similar para varones y mujeres. Se destaca para los tres casos evaluados un leve aumento de la mortalidad alrededor de 1990, hecho que resultaría interesante evaluar demográficamente dado que, para otros casos como el de la ciudad de Buenos Aires, se encontró que en los años 90 se había producido una sobremortalidad por HIV (hecho que se daba principalmente en varones). Por ello resultaría adecuado analizar la mortalidad por causas e incluso evaluar la existencia de un hecho puntual en la historia que haya sido el causal de este máximo relativo, o si es que éste se debe meramente al azar.

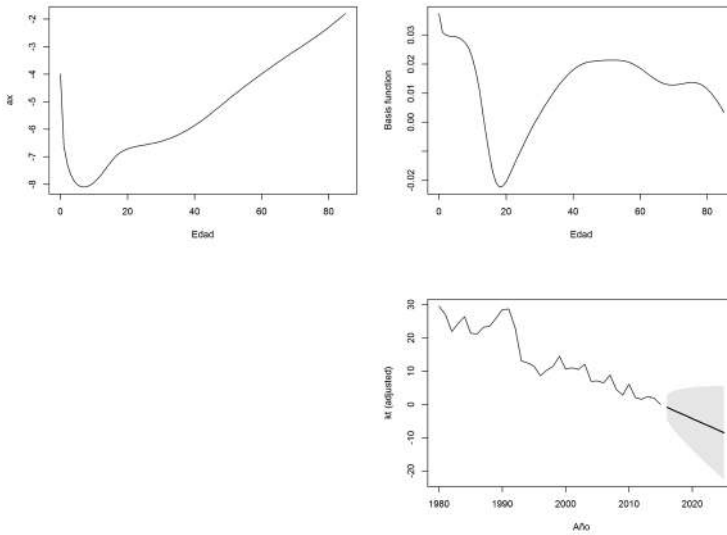
Por otra parte, el parámetro  $b$  indica qué edades son las que contribuyen en mayor o menor medida al descenso que general que se manifiesta a través de  $k$ . En los varones puede observarse un mínimo en torno a los 18 años lo que significa que dicha edad y sus valores colindantes son los que menos contribuyen en el descenso, sustentando el aumento en la mortalidad presente en edades jóvenes antes mencionado. El parámetro  $b$  en las mujeres presenta un comportamiento distinto, las edades iniciales, los treinta y los ochenta años presentan valores altos, indicando que son dichos grupos etarios los que contribuyen en mayor medida al descenso de la mortalidad en las mujeres. Con el fin de calcular los pronósticos, se predice el índice general mediante modelos para series de tiempo y suponiendo que el comportamiento del parámetro  $b$  permanece estable, se construyen los valores estimados de las tasas de mortalidad para años futuros.

**Figura 2. Estimaciones de la media, función base e índice de mortalidad general (con su intervalo de pronóstico del 80%) a través del Modelo de Lee Carter. Rosario, 1980-2015. (a) Total de la población, (b) Varones y (c) Mujeres**

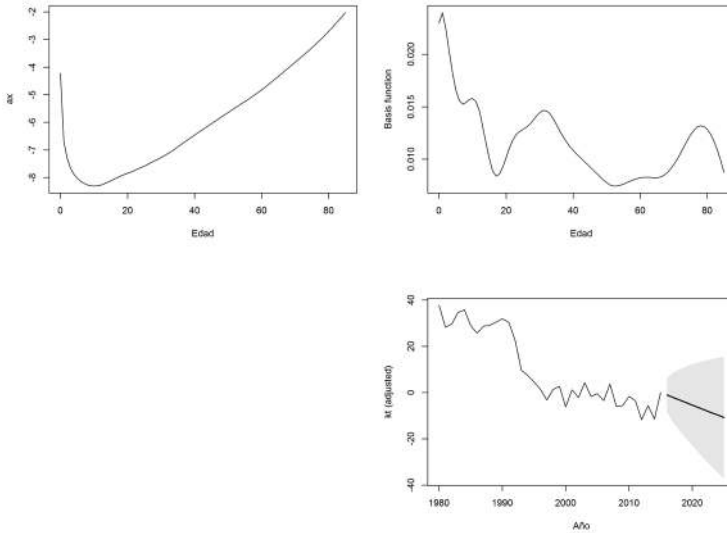
a)



b)



c)



**Fuente:** elaboración propia en base a datos de la Dirección General de Estadística de la Municipalidad de Rosario.

A partir de los pronósticos de las tasas de mortalidad es posible obtener, a través del uso de tablas de vida, los pronósticos de las esperanzas de vida al nacer. En este caso se presentan los valores hasta el año 2024 (Tabla 1). Si bien los valores obtenidos son el verdadero producto de estos modelos y en ellos radica el interés al momento de diseñar y planificar políticas públicas, el gráfico de los resultados puede ayudar a evaluar la concordancia de los resultados con los datos observados. Por

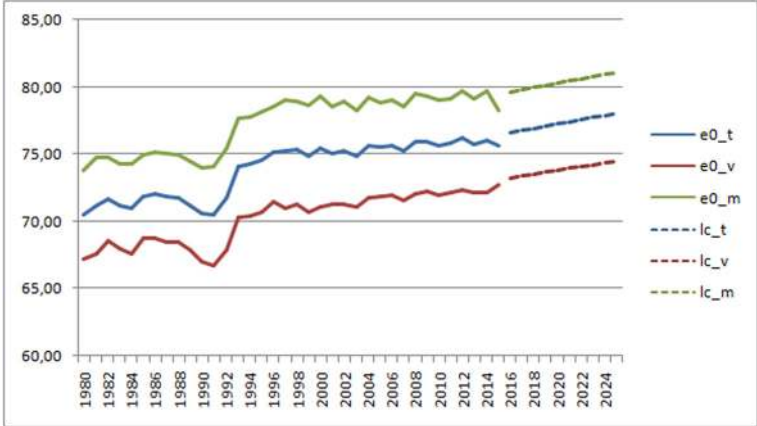
ello en la Figura 3 pueden observarse las series que representan la esperanza de vida al nacimiento para el período disponible (1980-2015), llamadas “esperanzas observadas” y las líneas que representan los pronósticos de las mismas, “esperanzas pronosticadas”, para los tres grupos: total de la población, varones y mujeres. En los tres grupos puede observarse la leve tendencia creciente en las líneas que representan los pronósticos, alcanzando los 78,06 años de esperanza de vida para el total de la población, 74,49 años los varones y 81,08 las mujeres. Un modo de evaluar la calidad de los pronósticos es a través de la amplitud de sus intervalos (cuanto más amplio sea menos informativo es el pronóstico). Si se calcula la amplitud promedio de los intervalos obtenidos por LC, ésta resulta de 4,49 años, luego es de interés evaluar este dato en forma comparativa con los obtenidos para los siguientes modelos.

**Tabla 1. Pronósticos LC de la Esperanza de vida al nacer e intervalos de pronóstico del 80%. Rosario, 1980-2015. Total de la población, Varones y Mujeres.**

| Año  | Total              | Varones            | Mujeres            |
|------|--------------------|--------------------|--------------------|
| 2016 | 76,60(75,66-77,50) | 73,24(72,41-73,92) | 79,63(78,51-80,64) |
| 2017 | 76,66(75,53-77,99) | 73,39(72,38-74,28) | 79,79(78,15-81,39) |
| 2018 | 76,93(75,55-78,32) | 73,53(72,24-74,60) | 79,95(78,02-81,88) |
| 2019 | 77,09(75,45-78,55) | 73,67(72,27-75,01) | 80,11(77,83-82,35) |
| 2020 | 77,25(75,51-79,02) | 73,81(72,27-75,26) | 80,27(77,63-82,72) |
| 2021 | 77,41(75,37-79,57) | 73,95(72,22-75,53) | 80,43(77,53-83,48) |
| 2022 | 77,57(75,29-79,65) | 74,09(72,27-75,63) | 80,60(77,51-83,73) |
| 2023 | 77,74(75,25-79,88) | 74,22(72,26-75,92) | 80,76(77,20-84,24) |
| 2024 | 77,90(75,11-80,03) | 74,35(72,46-76,06) | 80,92(77,09-84,34) |
| 2025 | 78,06(75,34-80,67) | 74,49(72,33-76,32) | 81,08(77,37-84,82) |

Fuente: elaboración propia en base a datos de la Dirección General de Estadística de la Municipalidad de Rosario.

**Figura 3. Esperanza de vida al nacer observada (1980-2015) y pronosticada a través del modelo LC (2016-2025). Total de la población, Varones y Mujeres**



Fuente: elaboración propia en base a datos de la Dirección General de Estadística de la Municipalidad de Rosario.

### 3.1. Modelo de datos funcionales de la mortalidad de la ciudad de Rosario

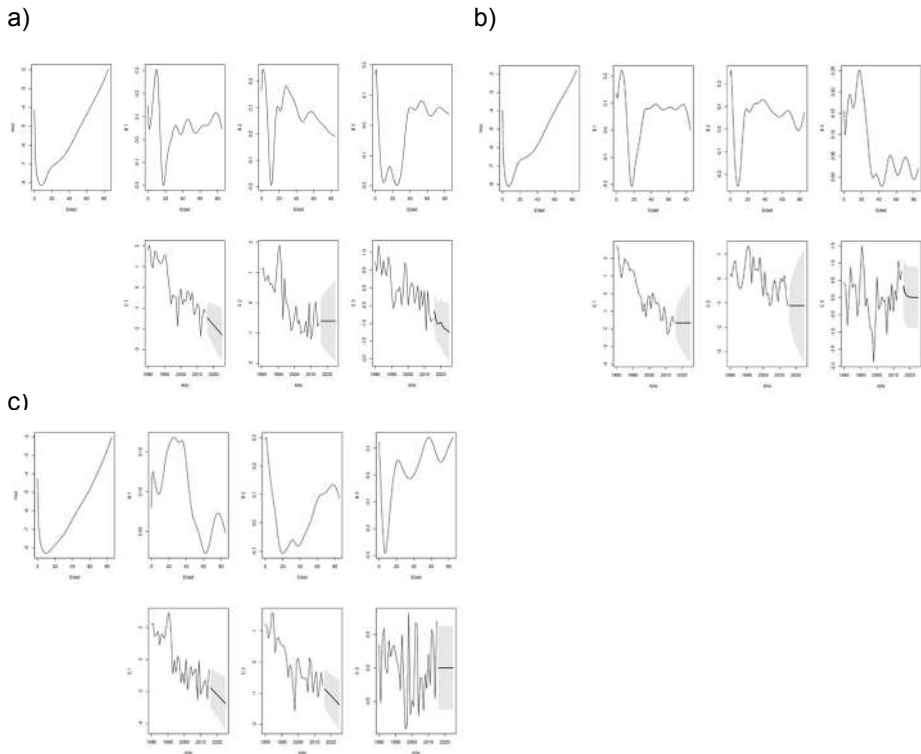
La aplicación del modelo de datos funcionales (MDF) a las tasas de mortalidad permite descomponer las tasas mediante coeficientes, bases y una media. Esta descomposición se realiza para el total de la población, los varones y las mujeres.

La Figura 4 presenta la media, las bases (sus intervalos de pronóstico del 80%), los coeficientes funcionales estimados para el total de la población, varones y mujeres. Estas sub-figuras (una para el total y otra para cada sexo) se deben interpretar por columna. En primer lugar se representa el comportamiento promedio de la mortalidad a través de las edades, y luego, a partir de la segunda columna, la fila superior contiene las bases y la inferior los coeficientes. Cada columna contiene entonces un par coeficiente-base que debe interpretarse en forma conjunta. Por ejemplo, en el caso de la mortalidad del total de la población, el primer coeficiente muestra un decrecimiento de la mortalidad a través del tiempo (segunda columna, fila inferior), pero para interpretarlo correctamente este comportamiento debe ponerse en correspondencia con su base asociada (recuadro de la fila superior) ya que la misma indica en qué edades se manifiesta este descenso y con qué intensidad, del mismo modo que se relacionaban  $b$  y  $k$  para el modelo LC.

En este caso la media estimada representa el perfil promedio de la mortalidad a lo largo de la vida (obtenido como el promedio de las funciones a través de los años). Si se analizan los pares base-coeficiente, para los varones se observa que al descenso con fluctuaciones presente en el primer coeficiente contribuyen (según la base correspondiente) las edades de 10 a 14, mientras que las edades cercanas a los 18 son las que menos lo hacen. Observando el segundo par base-coeficiente, puede representar (según la base) el diferencial en la mortalidad para edades iniciales versus 10 años ya que se presenta un valor marcadamente alto para el primer grupo y uno marcadamente bajo para el segundo, y el coeficiente correspondiente indicaría que éste presenta un descenso en el tiempo aunque con fluctuaciones. Más allá del tercer par base-coeficiente resulta complejo obtener un correlato demográfico de los resultados.

En el caso de las mujeres el primer par puede interpretarse que son las mujeres de entre 20 y 40 años son las que en mayor medida contribuyen al descenso en la mortalidad expresado por el coeficiente. Mientras que el segundo par base-coeficiente puede relacionarse al descenso que manifiestan las edades iniciales (menos de 5) y las avanzadas (80 y más), dado que para ambas edades la base presenta valores altos.

**Figura 4. Estimación de la media, funciones base (con sus intervalos de pronóstico del 80%) y coeficientes a través del Modelo para Datos Funcionales. Rosario, 1980-2015. (a) Total de la población, (b) Varones y (c) Mujeres**



**Fuente:** elaboración propia en base a datos de la Dirección General de Estadística de la Municipalidad de Rosario.

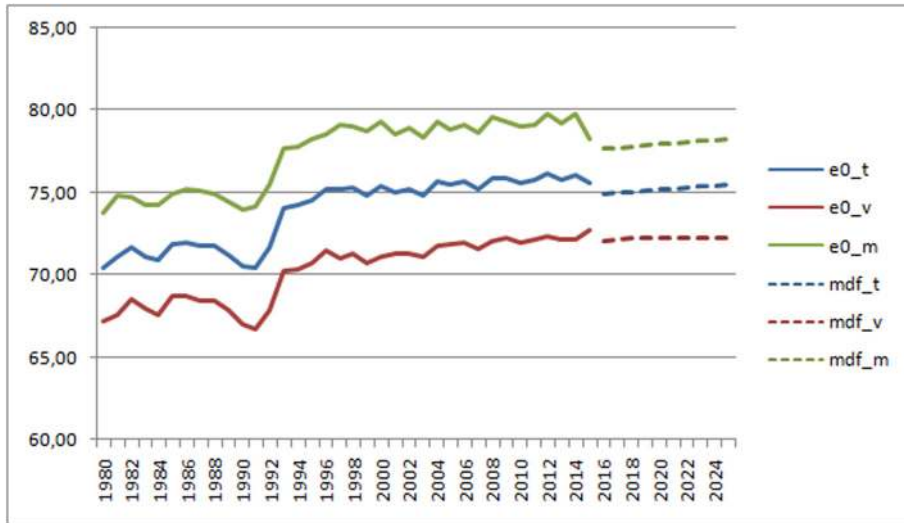
Al igual que en el modelo LC, a partir de las tasas de mortalidad se calculan las esperanzas de vida para años futuros por sexo con sus respectivos intervalos de pronóstico mediante MDF. En la Tabla 2 pueden observarse las esperanzas de vida pronosticadas para el período 2016-2025, donde la amplitud promedio calculada para estos intervalos es de 4,46 años. Además la Figura 5 muestra una clara divergencia en los comportamientos de los pronósticos, una esperanza de vida que aumenta con el tiempo en las mujeres y una esperanza de vida estancada para varones, resultados que parecen no condecir el comportamiento de dicha medida obtenidos para diferentes regiones y ciudades del país, (Andreozzi y Blaconá, 2012).

**Tabla 2. Pronósticos MDF de la Esperanza de vida al nacer e intervalos de pronóstico del 80%. Rosario, 1980-2015. Total de la población, Varones y Mujeres**

| Año  | Total              | Varones            | Mujeres            |
|------|--------------------|--------------------|--------------------|
| 2016 | 76,34(75,27-77,46) | 73,05(71,91-74,24) | 80,83(79,26-82,65) |
| 2017 | 76,55(75,36-77,84) | 73,09(71,60-74,46) | 80,97(79,29-82,86) |
| 2018 | 76,69(75,37-77,98) | 73,10(71,37-74,70) | 81,14(79,31-82,95) |
| 2019 | 76,76(75,29-78,15) | 73,10(70,94-74,80) | 81,31(79,50-83,11) |
| 2020 | 76,82(75,46-78,38) | 73,10(70,85-75,01) | 81,47(79,64-83,52) |
| 2021 | 76,96(75,41-78,41) | 73,10(70,69-75,18) | 81,64(79,73-83,61) |
| 2022 | 77,07(75,33-78,61) | 73,10(70,53-75,30) | 81,81(79,78-83,92) |
| 2023 | 77,16(75,38-78,80) | 73,10(70,12-75,55) | 81,98(80,02-84,16) |
| 2024 | 77,25(75,36-79,03) | 73,10(70,25-75,64) | 82,15(80,20-84,24) |
| 2025 | 77,36(75,41-79,04) | 73,10(70,01-75,75) | 82,32(80,15-84,48) |

Fuente: elaboración propia en base a datos de la Dirección General de Estadística de la Municipalidad de Rosario.

**Figura 5. Esperanza de vida al nacer observada (1980-2015) y pronosticada a través de MDF (2016-2025). Total de la población, Varones y Mujeres**



Fuente: elaboración propia en base a datos de la Dirección General de Estadística de la Municipalidad de Rosario.

Por último se aplica un MDF con restricciones que permiten obtener pronósticos estimados coherentes por sexo. 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) para los que se puede hallar nuevos valores y sus respectivos intervalos de pronóstico. En otras palabras, el método de pronósticos coherentes pretende asegurar que los pronósticos para poblaciones relacionadas mantengan cierta

relación estructural basada en información histórica y consideraciones teóricas. Por ejemplo, se ha observado que la mortalidad masculina es más alta que la mortalidad femenina para todas las edades y mientras la evidencia disponible sustenta tanto la hipótesis biológica-genética como la social-cultural-ambiental-conductual, Kalben (2000) concluye que el factor determinante es el biológico. Esta diferencia en la mortalidad entre los sexos es de esperar que persista en el futuro y que además permanezca dentro de los límites observados.

De este modo se obtienen entonces los pronósticos de las esperanzas de vida que respetan la diferencia entre sexos antes mencionada. Los mismos se presentan junto a sus intervalos de pronóstico del 80% en la Tabla 3. La amplitud promedio de estos intervalos es de 2,18 años resultando la menor de las tres medidas obtenidas. Por otro lado, en la Figura 6 puede verse un comportamiento creciente en las líneas que representan las esperanzas de vida para los tres grupos. Con esta metodología se pronostica una esperanza de vida al nacer de 77,17 años para el total de la población, 73,82 años para varones y 81,36 años para mujeres para el año 2025.

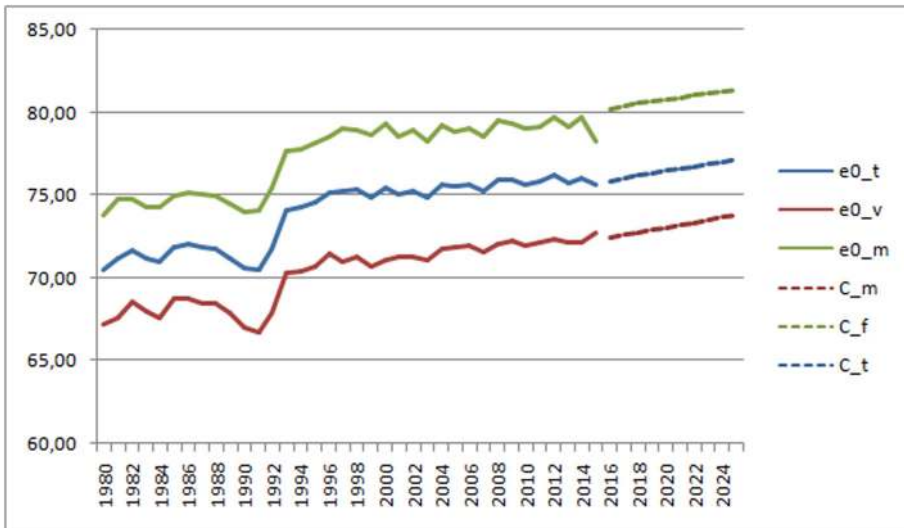
**Tabla 3. Pronósticos MDF coherentes de la Esperanza de vida al nacer e intervalos de pronóstico del 80%. Rosario, 1980-2015. Total de la población, Varones y Mujeres**

| Año  | Total              | Varones            | Mujeres            |
|------|--------------------|--------------------|--------------------|
| 2016 | 75,82(74,88-76,53) | 72,41(71,51-73,25) | 80,14(79,20-81,06) |
| 2017 | 76,03(75,12-76,93) | 72,61(71,70-73,44) | 80,37(79,45-81,35) |
| 2018 | 76,17(75,20-77,03) | 72,73(71,68-73,65) | 80,53(79,66-81,51) |
| 2019 | 76,31(75,37-77,19) | 72,88(71,83-73,87) | 80,66(79,63-81,79) |
| 2020 | 76,46(75,39-77,37) | 73,04(71,88-74,20) | 80,78(79,67-81,98) |
| 2021 | 76,60(75,51-77,57) | 73,20(72,14-74,25) | 80,89(79,78-81,95) |
| 2022 | 76,74(75,72-77,95) | 73,35(72,30-74,48) | 81,01(79,82-82,33) |
| 2023 | 76,89(75,80-78,10) | 73,51(72,35-74,82) | 81,13(79,96-82,33) |
| 2024 | 77,03(75,99-78,10) | 73,66(72,47-74,81) | 81,24(79,98-82,47) |
| 2025 | 77,17(75,96-78,43) | 73,82(72,61-75,16) | 81,36(80,11-82,68) |

Fuente: elaboración propia en base a datos de la Dirección General de Estadística la Municipalidad de Rosario.



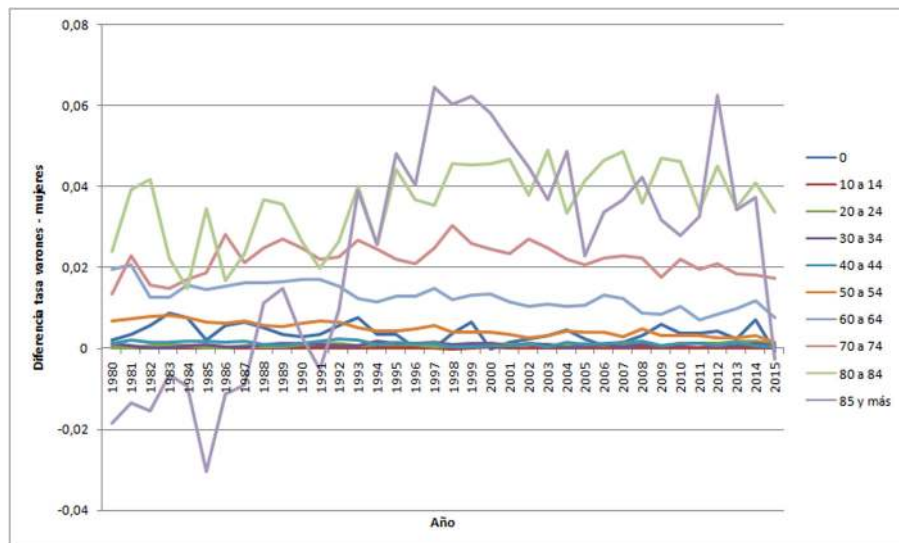
**Figura 6. Esperanza de vida al nacer observada (1980-2015) y pronosticada a través de MDF coherentes (2016-2025). Total de la población, Varones y Mujeres.**



**Fuente:** elaboración propia en base a datos de la Dirección General de Estadística de la Municipalidad de Rosario.

Finalmente, la Figura 7 presenta las diferencias entre las tasas de varones y mujeres para el período 1980-2015. Del gráfico de dichas series se desprende que para la mayoría de las edades la diferencia entre las tasas observadas para ambos sexos se mantiene constante en el tiempo con excepción del último intervalo de edad analizado. Sin embargo este intervalo en la actualidad resulta amplio ya que abarca edades que por el aumento en la esperanza de vida presentan comportamientos disímiles y sería adecuado pensar a futuro en contar con datos con mayor nivel de desagregación. Obviando este análisis del intervalo abierto final, es claro que las restantes edades presentan diferencias constantes entre sexos reforzando de este modo la utilización del modelo que impone coherencia.

**Figura 7. Diferencia en las tasas de mortalidad observadas para varones y mujeres. (Edades seleccionadas) Rosario, 1980-2015**



**Fuente:** elaboración propia en base a datos de la Dirección General de Estadística de la Municipalidad de Rosario.

Asimismo, pueden calcularse medidas de bondad de pronóstico, las cuales evalúan el desempeño de los modelos a la hora de estimar los valores futuros. Para ello se deja fuera de la modelización un conjunto de 5 valores, de 2011 a 2015 y luego se les calcula el error medio absoluto porcentual (Tabla 4). Los valores resultan menores para el modelo de pronósticos coherentes indicando que el mismo es el que presenta mejor calidad a la hora de efectuar pronósticos.

**Tabla 4. Error medio absoluto porcentual (2011-2015)**

|         | LC   | MDF  | MDF<br>Coherente |
|---------|------|------|------------------|
| Total   | 6.43 | 0.35 | 0.39             |
| Mujeres | 6.17 | 2.00 | 1.63             |
| Varones | 0.22 | 0.46 | 0.45             |

**Fuente:** elaboración propia en base a datos de la Dirección General de Estadística de la Municipalidad de Rosario.

#### 4. Consideraciones finales

El objetivo final de este trabajo es obtener estimaciones de la esperanza de vida al nacer para la ciudad de Rosario durante el período 2016-2025. Es por ello que la selección del mejor resultado comprende criterios tanto estadísticos como demográficos. El modelo que presenta mejores resultados es el MDF coherente dado que posee los intervalos de pronóstico más estrechos, trayectorias que respetan la diferencia entre sexos y, por otro lado, resulta adecuado desde el punto de vista demográfico en contraste con MDF que presenta un estancamiento para el caso de los varones. Además presenta la mejor calidad de pro-

nósticos medida por el MAPE.

Es importante destacar que el modelo MDF coherente es una metodología moderna empleada por diversos organismos de estadísticas oficiales del mundo para generar pronósticos de las diversas componentes y medidas demográficas. Sin embargo, vale aclarar que se encuentra en pleno desarrollo y su empleo en las diversas regiones depende de múltiples factores, entre ellos la calidad y exactitud de los insumos base: estadísticas vitales y datos censales. En este sentido resulta importante continuar con el estudio de los resultados analizando las interpretaciones de los pares base-coeficiente en relación a mortalidad por causa y hechos históricos a fin de potenciar los resultados aquí obtenidos.

Un punto a destacar es que los modelos funcionales pueden emplearse también para analizar y pronosticar la fecundidad y la componente migratoria, y, estos resultados, junto con los obtenidos para la mortalidad, permiten desarrollar proyecciones probabilísticas de la población. Además permiten obtener pronósticos de medidas derivadas de las tasas, como la tasa global de fecundidad que se obtiene a partir de las tasas de fecundidad, o pronósticos de la relación de dependencia, determinada a partir de la población por edades, entre otros. Asimismo cada medida derivada se estima y pronostica en forma puntual y por intervalos de confianza.

Estos resultados son especialmente importantes al momento de delinear políticas públicas porque los intervalos proveen información acerca de los límites inferiores y superiores asociados a una estimación puntual, y por ende, un grado de incertidumbre. La implementación de estos métodos por parte de los organismos oficiales de estadística es accesible ya que se utiliza como único insumo los datos que la ciudad recaba y los cálculos se realizan utilizando el software libre R. Por tal motivo este trabajo constituye una propuesta para emplear una metodología moderna que ha sido incorporada por las agencias de estadística de diversos países del mundo con gran éxito.

## Referencias bibliográficas

- Alho, J. M. (2000). A statistical look at Modeen's forecast of the population of Finland in 1934. *Yearbook of Population Research in Finland*, Vol. 36, Helsinki. Väestöliitto, 2000: 107-120.
- Andreozzi, L. y Blaconá, M. T. (2011). The Lee Carter method, for estimating and forecasting mortality: an application for Argentina. Documento presentado al ISF 2011 – Prague - Proceedings. Disponible en: <<http://www.forecasters.org/submissions/>> ANDREOZZILUCIAISF2011.pdf
- Arriaga, E. (2014). Análisis demográfico de la mortalidad. Centro de investigaciones y estudios sobre Cultura y Sociedad. CONICET y UNC.
- Booth, H., Hyndman, R., Tickle, L., y de Jong, P. (2006). Lee-Carter mortality forecasting: a multi-country comparison of variants and extensions.

- Blaconá, M. T y Andreozzi, L. (2012). Comparación de métodos de estimación del modelo de Lee y Carter. *Estadística*, 64(182 y 183), 57-84.
- García Guerrero, V. M. y Ordorica Mellado, M. (2012). Proyección estocástica de la mortalidad mexicana por medio del método de Lee-Carter. *Estudios Demográficos y Urbanos*, 27(2), 409-448.
- Guerrero Guzmán, V. M. y González Pérez, C. (2007). Pronósticos estadísticos de mortalidad y su impacto sobre el sistema de pensiones de México. Trabajo ganador del Premio de Pensiones 2007 otorgado por la Comisión Nacional del Sistema de Ahorro para el Retiro, 2007. Disponible en: [http://www.consar.gob.mx/premio\\_pension/es/premio\\_2007.shtml](http://www.consar.gob.mx/premio_pension/es/premio_2007.shtml)
- Hyndman, R. J. con contribuciones de Booth, H., Tickle, L. y Maindonald, J. (2017). demography: Forecasting Mortality, Fertility, Migration and Population Data. R package version 1.19. Disponible en <https://CRAN.R-project.org/package=demography>
- Hyndman, R. y Ullah, H. (2007). Robust forecasting of mortality and fertility rates: A functional data approach. *Computational Statistics and Data Analysis*, 51, 4942-4956.
- Hyndman, R. y Booth, H. (2008). Stochastic population forecast using functional data models for mortality, fertility and migration. *International Journal of Forecasting*, 24, 323-342.
- Hyndman, R., Booth, H. y Yasmeen, F. (2013). Coherent Mortality Forecasting: The Product-Ratio Method With Functional Time Series Models. *Demography*, 50(1), 261-283.
- Kalben, B. (2000). "Why men die younger". *North American Actuarial Journal*, 4(4), 83-111.
- Lee, R. D. y Carter, L. (1992). Modeling and Forecasting the Time Series of U.S. Mortality. *Journal of the American Statistical Association*, 87, 659-671.
- Lee, R. D. y R. Rofman (1994). Modelación y Proyección de la Mortalidad en Chile. *Notas de Población*, XXII, 59, 182-213.
- Lee, R. y Tuljapurkar, S. (1994). Stochastic population forecasts for the United States: Beyond high, medium, and low. *Journal of the American Statistical Association*, 89, 1175-1189.
- Raftery, A. E., Li, N., Sevč'ikov'a, H., Gerland, P., y Heilig, G. K. (2012). Bayesian probabilistic population projections for all countries. *Proceedings of the National Academy of Sciences*, 109, 13915-13921
- Ramsay, J. O. y Silverman, B. W. (2005). Functional data analysis. 2nd ed. Springer, New York.
- R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Viena. Disponible en: < <http://www.R-project.org/>>

Serfaty, E., Foglia, L., Masautis, A. y Negri, G. (2007), Mortalidad por causas violentas en adolescentes y jóvenes de 10- 24 años, en Revista Vertex., 40, 25-30.