

, vol. 26, 2020, pp. 6-19.

Mortality projection using two probabilistic models. Applications and limitations.

Andreozzi, Lucía y Torres, Eduardo.

Cita:

Andreozzi, Lucía y Torres, Eduardo (2020). *Mortality projection using two probabilistic models. Applications and limitations.* , 26, 6-19.

Dirección estable: <https://www.aacademica.org/lucia.andreozzi/15>

ARK: <https://n2t.net/ark:/13683/preH/myU>



Esta obra está bajo una licencia de Creative Commons.
Para ver una copia de esta licencia, visite
<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.es>.

Acta Académica es un proyecto académico sin fines de lucro enmarcado en la iniciativa de acceso abierto. Acta Académica fue creado para facilitar a investigadores de todo el mundo el compartir su producción académica. Para crear un perfil gratuitamente o acceder a otros trabajos visite: <https://www.aacademica.org>.

Proyección de la mortalidad utilizando dos modelos probabilísticos. Aplicaciones y limitaciones

Mortality projection using two probabilistic models. Applications and limitations

*Lucía Andreozzi¹, Eduardo Torres¹

¹Universidad Nacional de Rosario. Rosario, Argentina.

Recibido: 30/09/2019

Aceptado: 25/02/2020

RESUMEN

En este artículo se presenta el modelo de Lee-Carter y el Modelo de Datos Funcionales (dos modelos probabilísticos, ambos provenientes de la Estadística) y con cada uno se elabora la proyección de la mortalidad y las esperanzas de vida al nacimiento en tres Departamentos de la Provincia de Córdoba para el período 2020 – 2040. Además, se analizan las fortalezas y debilidades de ambos métodos. Para ello se utilizan datos provenientes de Estadísticas Vitales y Censos Nacionales de Población de Argentina. Asimismo, se aborda la relación entre las metodologías estadísticas y la Demografía.

PALABRAS CLAVE: Proyecciones, demografía, modelos-probabilísticos, mortalidad

ABSTRACT

This paper is focused on Lee-Carter Model and Functional Data Model, both of which were developed in the field of Statistics. These models are applied in order to carry out the projection of mortality and life expectancies at birth on three departments of the province of Córdoba (Argentina) from 2020 to 2040. The strengths and weakness of said models are also analyzed. In order to do that, data from Vital Statistics and National Census in Argentina were used. Moreover, there is reference to the relationship between Demography and Statistics as well.

KEY WORDS: Demographic, projections, probabilistic-models, mortality

AUTOR CORRESPONDIENTE: *Dra. Lucía Andreozzi. Universidad Nacional de Rosario (UNR) – Consejo Nacional de Investigaciones Científicas y Técnicas de Argentina (CONICET). Rosario, Argenti-na. Email: andreozzi.lu@gmail.com

INTRODUCCIÓN

En este trabajo se ha elaborado una proyección de la mortalidad utilizando dos modelos probabilísticos (Lee-Carter y Modelo de Datos Funcionales) para tres Departamentos de la Provincia de Córdoba (Capital, Río Cuarto y Sobremonte). La elección de estos obedece a que cada uno de ellos presenta diferencias tanto en sus características demográficas como socio-económicas. Para ello se han empleado datos provenientes de los Censos Nacionales de Población, Hogares y Viviendas de Argentina correspondientes a los años 1980, 1991, 2001 y 2010 y las cifras provenientes de Estadísticas Vitales del período comprendido entre 1980 y 2013.

Si bien es cierto que las proyecciones demográficas se vienen realizando desde hace tiempo, puede señalarse que cuando Naciones Unidas en el año 1996 proyectó la población mundial para el año 2050 -según la hipótesis de variante media- la cifra obtenida resultó menor en 466 millones de personas respecto a la proyección que se realizó para el mismo año en 1994. Este cambio en las cifras fue tomado como una evidencia de que el crecimiento en la población no era un problema tan grave como se pensaba. Pero, más allá de esto, lo más importante es que ambas proyecciones tenían intervalos de incertidumbre que se superponían, lo que significa que en algún caso ambas proyecciones eran iguales o muy próximas; y si bien los productores de las proyecciones comprendían este punto, la mayoría de los consumidores de los datos no tenían esto en cuenta.

Luego, en 1998, el Consejo Nacional de Investigaciones de la Academia de Ciencias de los Estados Unidos conformó un panel de expertos para analizar en profundidad el tema de las proyecciones (Bongaarts y Bulatao, 2000). En dicho encuentro se concluyó que las proyecciones elaboradas para el año 2050 se basaban en supuestos razonables y proveían pronósticos plausibles para las tendencias demográficas en las próximas décadas. Sin embargo, esta conclusión no abarcaba las proyecciones realizadas para países específicos, dado que el panel solamente examinó la metodología general de las proyecciones a nivel mundial. Además, entre las conclusiones se destacaba que las proyecciones elaboradas no otorgaban la importancia necesaria a un aspecto relevante: la incertidumbre asociada a los pronósticos, de modo que el panel indicó que sería adecuado que las cifras de proyecciones contemplaran de forma explícita este aspecto y se remarcó la necesidad de desarrollar métodos que la cuantifiquen.

Como herramienta principal y más tradicional para realizar proyecciones de población, en Demografía se utiliza desde hace tiempo el Método de las Componentes. Para emplear esta metodología es necesario trabajar con datos relacionados a fecundidad, mortalidad y migración, los que constituyen las componentes de crecimiento de la población. En la actualidad organismos tales como las Naciones Unidas y otras instituciones referentes en el área son los encargados de

determinar los niveles y probables patrones para cada una de ellas. Luego de establecer ciertos supuestos, los niveles y patrones se combinan con información relativa a la estructura existente por edad y sexo y mediante una serie de operaciones matemáticas se obtiene la proyección de la población con las hipótesis incorporadas.

Pero es inevitable que las proyecciones contengan incertidumbre, ya que si bien la situación demográfica actual es conocida, las tendencias futuras en los nacimientos, muertes y migraciones siempre están sujetas a cambios impredecibles. Así, los errores cometidos en la elaboración de proyecciones varían de acuerdo al horizonte de pronóstico, la calidad de las estadísticas, el nivel de desarrollo del país y su tamaño y -sobre todo- de las hipótesis que se tengan en cuenta para elaborar la proyección. Mientras que la tendencia general en la mortalidad, la fecundidad y la migración pueden discernirse y proyectarse para el futuro con una confianza (en el uso no estadístico de la palabra) razonable, una sustancial incertidumbre se haya asociada cuando se trata de elaborar una proyección para un país o una región de tamaño pequeño.

A lo largo de las siguientes páginas se realiza una referencia a la relación entre las metodologías estadísticas y las Ciencias Sociales, haciendo hincapié en una crítica pendiente sobre la articulación entre ambas donde los modelos son limitados a ser sólo una herramienta técnica. Luego se presentarán los principales antecedentes de los dos modelos utilizados y sus características teóricas y conceptuales más importantes. Posteriormente, mediante el uso de los datos y a través del lenguaje de programación R, se desarrolla el objetivo principal del trabajo: realizar la proyección de las tasas específicas de mortalidad y obtener evidencias acerca de las fortalezas y limitaciones de ambas metodologías en regiones geográficas sub-nacionales que presentan disimilitudes entre sí, lo cual constituye un sustancial aporte ya que no se dispone de antecedentes de aplicación con características similares. Asimismo, se elaboran los pronósticos de la esperanza de vida al nacimiento, que son el insumo de mayor relevancia que entregan estos modelos, como propuesta inicial se elaboran para el período 2020-2040, sin embargo se evalúa la bondad (y su capacidad predictiva) de los resultados en base a amplitud de los intervalos asociados. Finalmente, se plasma un conjunto de reflexiones finales que procuran sintetizar lo expuesto en el artículo.

Si bien los modelos probabilísticos constituyen un importante aporte a la Demografía, tanto por la diversidad como la calidad de los insumos que estas metodologías generan, se destaca que dichas propuestas provienen de la Ciencia Estadística y constituyen en una herramienta orientada al servicio de una Ciencia Social. Es por ello que resulta relevante estudiar sus limitaciones, ya que si bien las metodologías intentan hacer una descripción, una simplificación o reducción del comportamiento complejo de la cualquiera de las componentes demográficas (fecundidad, mortalidad o migración), tienen como principal función generar pronósticos y, en este punto, es necesario evaluar de qué manera proveen información válida que permita desarrollar teorías demográficas o que ayude a comprender la dinámica de las tres componentes, de modo tal que le sea útil tanto a los organismos oficiales como también resulte provechoso para la sociedad en general.

Por otro lado, en relación a la función de pronóstico, resulta fundamental evaluar si las cifras que se obtienen tienen un horizonte límite; una cantidad específica de años a pronosticar más allá de la cual la amplitud del intervalo asociado ya no resulta informativa. Además, la información que brindan los modelos debe ser útil y oportuna en relación a que esté disponible en el momento en que la misma se necesite con una calidad aceptable. Para más información haga click aquí

MATERIALES Y MÉTODOS

Se describen a continuación las principales características de los métodos empleados en el presente artículo: el de Lee-Carter y el Modelo de Datos Funcionales.

El Modelo de Lee-Carter

El modelo de Lee y Carter (1992) es el precursor de los modelos probabilísticos de pronóstico en mortalidad, en donde la premisa básica que se plantea es la existencia de 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 .

Así, al aplicar el modelo a tasas específicas de mortalidad, las mismas se descomponen, a través de un proceso de estimación, en dos parámetros a y b (llamado base) y un índice k_t llamado índice general de mortalidad. Sin embargo, es importante no confundir este índice con el nivel que determina la esperanza de vida, sino que se trata de un índice que refleja la tendencia general que presenta la mortalidad. Matemáticamente, la ecuación que describe el modelo es

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

aplicando logaritmo

$$f_{x,t} = \ln(m_{x,t}) = a_x + b_x k_t + e_{x,t} \quad t = 1, \dots, n \quad x = 1, \dots, \omega \quad (2)$$

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

específicas de mortalidad. Dicho parámetro se estima como el promedio de los logaritmos de las tasas específicas de cada edad en el tiempo.

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 . Este parámetro representa la intensidad en el crecimiento o decrecimiento de la tasa de mortalidad, para un grupo de edad a través del tiempo.

Modelo de Datos Funcionales

Este modelo –desarrollado por Hyndman y Ullah (2008) constituye uno de los últimos desarrollos propuestos en el área, donde se supone el siguiente postulado para las observaciones transformadas (dado que el modelo presentado permite emplear múltiples transformaciones, se habla, en general, de observaciones transformadas, pero en este caso se trata de logaritmos) $y_t(x)$:

$$y_t(x) = s_t(x) + \sigma_t(x)\varepsilon_{t,x} \tag{3}$$

$$s_t(x) = \mu(x) + \sum_{k=1}^K \beta_{t,k}\phi_k(x) + e_t(x), \tag{4}$$

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, 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).

En otras palabras, se piensa que las tasas son funciones suaves y que la variabilidad (que puede ser distinta para cada edad) es lo que hace que en la realidad se observen curvas con picos y valles; no obstante, la verdadera forma es la suavizada, esto puede verse en la ecuación (3). Luego, el modelo postula a través de la ecuación (4) que el conjunto de funciones suavizadas puede descomponerse en una media -o comportamiento común a todas las edades- y un conjunto de pares base-coeficientes, donde las bases atrapan el comportamiento a través de las edades y los coeficientes, captan el comportamiento a través del tiempo (estos son los que luego se pronostican, ya que son las únicas componentes del modelo que dependen del tiempo, y en base a ellos se reconstruyen las tasas en futuro, pensando que tanto la media como las bases que recogen el comportamiento a través de las edades permanecen estables).

RESULTADOS Y DISCUSIÓN

Para elaborar los modelos se utilizó el lenguaje de programación R, el cual es un software libre para el cálculo estadístico y la elaboración de gráficos, disponible para múltiples plataformas. Inicialmente se tabulan las tasas de mortalidad por grandes grupos de edad de los tres departamentos escogidos para cuatro años seleccionados (Tabla 1), donde se observa que éstos presentan diferentes perfiles de mortalidad y principalmente de variabilidad: claramente resulta más alta en el departamento Sobremonte, ya que al ser un área de baja población cualquier fluctuación repercute en forma más visible en los resultados.

Los departamentos elegidos tienen perfiles socio demográficos contrastantes (CENSO 2010), Capital tiene una superficie de 576 km² con una densidad de 2.367 hab/km² y del total de hogares que lo componen un 5.8% con NBI (Necesidades básicas insatisfechas). Río Cuarto es el más extenso de la provincia con 18.399 km² y una densidad media de 13.4 hab/km², presenta un bajo porcentaje de hogares con NBI; 4,5% y, finalmente, Sobremonte con una superficie de 3.307 km² y una densidad poblacional de 1,39 hab/km² presenta un 14,1% de hogares NBI. Sobremonte integra una región de baja densidad, junto a los departamentos de Río Seco y Tulumba, en el norte de la provincia. Dicha región presenta valores altos de porcentaje de NBI, sin embargo, no son los más altos de la provincia (Minas y Pocho presentan más del 20% de hogares con NBI). En cuanto a niveles de productividad Río Cuarto es la provincia que más se destaca con presencia de ganado, faena y variedades de cultivos. El departamento Capital se caracteriza por el turismo, el sector automotriz y producción de trigo. Según informes oficiales Sobremonte no se destaca marcadamente en ningún sector específico de la producción. Del mismo modo, en la Tabla 1, se presenta un conjunto de indicadores que permiten conocer de manera resumida la estructura demográfica que cada Departamento tiene como así también el total de la Provincia.

Tabla 1. Indicadores demográficos seleccionados. Departamento Capital, Río Cuarto, Sobremonte y Total Provincia – Argentina. Año 2010.

Variables	Capital	Río Cuarto	Sobremonte	Provincia
Población	1.329.604	246.393	4.591	3.308.876
Porcentaje de población respecto al total provincial	40,2	7,4	0,1	100
Edad promedio (en años)	32,81	34,63	32,25	33,39
Edad mediana (en años)	29	31	28	30
Relación de sexo	91,2	94,5	107,5	94,3
Relación de dependencia	51,06	55,07	63,79	55,07
Porcentaje de población materno infantil (menores de 6 años y mujeres de 10 a 49 años)	40,3	37,9	37,2	39,0
Porcentaje de población femenina	52,3	51,4	48,2	51,5
Porcentaje de población sin cobertura médica	31,6	25,1	49,2	32,8
Porcentaje de población de 0 a 14 años	23,6	23,0	28,3	24,3
Porcentaje de población de 15 a 64 años	66,2	64,5	61,1	64,5
Porcentaje de población de 65 años y más	10,2	12,5	10,7	11,2
Porcentaje de población de 80 años y más	2,4	3,0	2,4	2,7
Porcentaje de población mayor de 16 años con Secundario o Polimodal completo o más	68,1	53,2	35,9	59,3
Porcentaje de población nacida en el extranjero	2,4	0,80	0,13	1,5
Tasa de analfabetismo (personas mayores de 5 años de edad)	1,5	2,2	5,1	1,9

Fuente: Censo Nacional de Población, Hogares y Viviendas (INDEC, 2012).

A continuación, se presentan los resultados para tres departamentos seleccionados, Capital y Río Cuarto; los dos más poblados. Seguido luego por los resultados obtenidos para el departamento Sobremonte, uno de los menos poblados de la provincia (tabla 2).

Tabla 2. Tasas específicas de mortalidad por mil habitantes según grandes grupos de edad para departamentos seleccionados en Argentina. Años 1980, 1990, 2000 y 2010.

Departamento	Grupo de Edad	1980	1990	2000	2010
Capital	0- 14	2,559	2,115	1,424	1,166
	15- 64	3,716	3,478	3,365	2,724
	65 y más	52,614	52,324	52,677	54,527
Río Cuarto	0- 14	2,578	1,431	1,211	1,138
	15- 64	4,305	3,729	3,171	3,071
	65 y más	59,118	52,965	48,315	59,716
Sobremonte	0- 14	3,791	1,368	2,071	1,509
	15- 64	6,864	4,684	5,359	2,515
	65 y más	76,301	55,388	78,833	112,497

Fuente:Elaboración propia en base a datos del INDEC y DEIS.

Para cada departamento se presenta un gráfico que muestra los logaritmos de las tasas de mortalidad observadas, que constituyen el insumo básico de ambos modelos probabilísticos. En cada año las curvas muestran la forma típica del patrón de la mortalidad: alta al inicio de la vida, luego un descenso hasta antes de los 10 años seguido de un aumento hasta su pico alrededor de los 20 años, fenómeno presente principalmente 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. Según Serfaty et al. (2007) en Argentina aquellos entre 10 y 24 años de edad representan un 27% de la población y que en los últimos tiempos se ha profundizado el conocimiento que se tiene acerca de ellos; 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.

Un aspecto que se destaca visualmente en los gráficos mencionados, es la caída marcada en los niveles de mortalidad a través del tiempo, que se da en todas las edades, excepto para el pico de los 20 años, franja en la que parece existir una caída 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 puede obedecer en parte a estas causas y a otras más específicas del proceso histórico particular de la Argentina. En el caso del departamento Río Cuarto es necesario presentar los logaritmos en dos escalas distintas dado que aumenta para esta región la frecuencia de tasas muy pequeñas que distorsionan la figura (para más información ver aquí).

En el caso del departamento Capital y Río Cuarto el porcentaje de variancia explicada resulta inferior para ambos modelos, en el caso del Modelo para Datos Funcionales, 96% para edades simples versus 94,9% con edades agrupadas para Capital y 96,6% para edades simples versus 95,2% en edades agrupadas para Río Cuarto. En el caso del modelo de Lee-Carter la diferencia es

más marcada 72,9% para edades simples versus 59% con edades agrupadas para Capital y 53,9% para edades simples versus 49,5% en edades agrupadas para Río Cuarto (Tabla 3.) Estas cifras indicarían que al emplear grupos quinquenales de edad el modelo permite explicar el comportamiento de las tasas específicas de mortalidad en base a las edades y el tiempo (años calendario) en menor medida.

Tabla 3. Comparación del porcentaje de variabilidad explicada entre edades simples y agrupadas. Departamentos Capital y Río Cuarto- Argentina, mediante Modelo para Datos Funcionales y Modelo de Lee-Carter.

Dpto.	Modelo	Edades		Dpto.	Modelo	Edades	
		Simples	Agrupadas			Simples	Agrupadas
Capital (6 bases)	MDF	73,0	59,2	Río Cuarto (6 bases)	MDF	54,5	50,1
		9,6	16,9			18,3	16,7
		5,8	10,1			12,9	12,7
		4,0	4,5			4,9	8,7
		1,8	2,4			3,5	3,8
		1,8	1,8			2,5	3,2
Total		96,0	94,9	Total		96,6	95,2
LC		72,9	59,0	LC		53,9	49,5

Fuente: Elaboración propia en base a datos del INDEC y DEIS. (INDEC 1980, 1991, 2001 y 2010 – DEIS 1980-2013)

Por otro lado para ambos departamentos y mediante ambos modelos estimados el Error Medio Absoluto Porcentual (MAPE) resulta levemente más alto al usar edades agrupadas, sin embargo ninguna de estas medidas supera el límite recomendable del 5% (tabla 4).

Tabla 4. Comparación del Error Medio Absoluto Porcentual (Mape) entre edades simples y agrupadas. Departamentos Capital y Río Cuarto- mediante Modelo para Datos Funcionales y Modelo de Lee-Carter (%).

Modelo/Edades	Capital		Río Cuarto	
	Simples	Agrupadas	Simples	Agrupadas
MDF	0,007	0,007	0,007	0,010
LC	0,018	0,026	0,027	0,039

Fuente: Elaboración propia en base a datos del INDEC y DEIS. (INDEC 1980, 1991, 2001 y 2010 – DEIS 1980-2013).

Por otro lado, el insumo de mayor relevancia que producen estos modelos son los pronósticos acompañados de sus intervalos de confianza. Para estos últimos se plantea utilizar un intervalo de pronóstico del 80%, el mismo coeficiente de confianza que emplea Hyndman en sus diversos trabajos. Esta elección se basa en que los datos provenientes de registros y relativos a fenómenos sociales poseen habitualmente una variación mayor a los datos de laboratorio, para los cuales se emplea el clásico intervalo de confianza del 95%. En relación a las cifras obtenidas, no resulta sencillo dictaminar si la amplitud de un intervalo es informativa, ¿informativa en términos de utilidad?, ¿informativa en términos comparativos? Por ejemplo, para el año 2020 la tasa mortalidad para personas de 50 años será de 3,859 por mil, entre 3,025 y 4,923. Puede pensarse que es un intervalo relativamente estrecho, sin embargo, cuando se evalúan los intervalos obtenidos para edades más allá de los 90 años, se detecta una amplitud más marcada, resultado esperable dado que dichas edades presentan tasas más variables a lo largo del tiempo y menor número de casos.

Asimismo, los intervalos se amplían a medida que aumenta el horizonte de pronóstico, y en este punto, resulta interesante evaluar hasta qué punto del tiempo los intervalos ofrecen una amplitud aceptable, es decir, hasta que año resulta útil pronosticar con la presente metodología (Andreozzi, 2018).

Finalmente, y dado que para Sobremonte no fue posible la implementación de los modelos debido a su población pequeña, se presenta una tabla resumen de los pronósticos de las esperanzas de vida al nacer para Capital y Río Cuarto en años seleccionados. Los cálculos de la esperanza de vida se basan en la metodología propuesta por Chiang (1984) para la construcción de tablas de vida. Se obtienen resultados para ambos departamentos, los intervalos obtenidos por el Modelo de Datos Funcionales resultan más estrechos y como es de esperar presenta una esperanza mayor coherente con los niveles inferiores obtenidos anteriormente en las tasas de mortalidad pronosticadas. Para el año 2025, el departamento Capital arroja una E.V.N. de 77,29 (75,41;78,69) años por MDF y de 76,54 (72,56;79,59) por LC. Para el mismo año y departamento Río Cuarto, la E.V.N. es de 76,39 (74,28;78,05) y 75,5 (66,67;81,52) por MDF y LC respectivamente. Si se extiende el horizonte, para año 2040, por ejemplo, el departamento Capital arroja una E.V.N. de 78,81(76,34;80,83) años por MDF y de 77,85(72,46;81,75) por LC. Para el mismo año y departamento Río Cuarto, la E.V.N. es de 77,7(75,01;79,41) y 77,08(58,59;84,91) por MDF y LC respectivamente.

Se realizó luego una búsqueda de proyecciones oficiales a fin de poder comparar resultados, sin embargo el INDEC solamente proporciona resultados a nivel provincia, mediante el cuadro disponible en su web “Esperanza de vida al nacer por sexo, según provincia. Período 2015-2040”. El mismo arroja proyecciones elaboradas en base a resultados del Censo Nacional de Población, Hogares y Viviendas 2010. Para los años antes mencionados la E.V.N es de 77,10 en varones, 83,49 en mujeres en 2030 y 78,57 en varones y 84,86 en mujeres para el año 2040.

Si bien los resultados obtenidos por el INDEC se presentan a nivel provincia y para ambos sexos, puede observarse que el promedio de ambas cifras para todos los años proyectados se encuentra por encima de las estimaciones de ambos modelos, pero más similares a las obtenidas por el Modelo de Datos Funcionales (incluso cercanas al extremo superior de los intervalos de pronóstico). Es importante tener en cuenta que no existen cifras estimadas ni pronosticadas de la esperanza de vida al nacer a nivel departamental. Sin embargo, el departamento Capital presenta cifras pronosticadas cercanas a las obtenidas por el INDEC para la provincia. Resulta luego dicha medida uno de los insumos más útiles y precisos (en términos de la estrecha amplitud del intervalo de pronóstico asociado) que arrojan los modelos bajo estudio.

CONCLUSIONES

Se observa que los resultados a este nivel geográfico presentan dificultades, ya que a medida que la magnitud de la población desciende se presentan tasas nulas que impiden la modelización. En línea con este aspecto, se evaluó el uso de edades agrupadas, sin embargo, no se logra subsanar la presencia de tasas nulas y nuevamente resulta imposible la estimación para el departamento Sobremonte. Por otro lado, los métodos empleados producen una amplia variedad de resultados,

sin embargo, bases y coeficientes matemáticos (o base e índice en el caso del modelo de Lee-Carter) detallan comportamientos demográficos que resulta difícil poner en concordancia con la historia, en este aspecto resulta clave no caer en el vicio de forzar la realidad a los modelos.

En relación a los datos básicos para la aplicación de los modelos, para el departamento Capital la mayor variabilidad la presentan los primeros años calendario incluidos en el análisis y a medida que estos transcurren la variabilidad disminuye. En segundo lugar, el departamento Río Cuarto presenta una variabilidad mayor durante todo el período y a diferencia del departamento Capital, presenta una moderada presencia de tasas bajas y nulas, sin embargo, el proceso de suavizado las corrige posibilitando la estimación del modelo. La presencia de tasas iguales a cero es un punto que merece ser destacado, ya que, frente a una cantidad moderada, los algoritmos funcionan, dado que mediante el proceso de suavizado los valores nulos se convierten en valores pequeños. Sin embargo, cuando la presencia de valores nulos es alta, la estimación resulta imposible, como sucede con los datos pertenecientes al departamento Sobremonte. Una posible línea de investigación futura consistiría en simular escenarios de distintos porcentajes de tasas nulas a fin de evaluar las cantidades que resultan críticas para la estimación. Para el departamento Sobremonte ninguna de las metodologías propuestas, aplicadas tanto a edades simples como agrupadas, arroja resultados. La estimación resulta imposible en todos los casos.

Finalmente, y en relación a la esperanza de vida, se observa una clara superioridad de los pronósticos obtenidos mediante el Modelo para Datos Funcionales, reflejado en la estrechez de sus intervalos asociados. La esperanza de vida al nacer resulta ser el mejor producto de los modelos y los resultados obtenidos presentan, además, un comportamiento lógico y coherente con la estimación a nivel provincia realizada por el INDEC.

La intervención del demógrafo en este tipo de modelos es limitada, no es posible seleccionar conjuntos de tasas más confiables o adecuadas de acuerdo a algún criterio, ya que requiere de todas las series históricas. La metodología permite simplemente fijar una edad a partir de la cual se considera que la mortalidad ya no vuelve a decrecer (por convención se elige la edad de jubilación) y emplear variedades de modelos de series de tiempo pudiendo definir la inclusión de determinados parámetros.

En cuanto a los resultados obtenidos en este trabajo, el modelo de Lee-Carter explica en ambos casos (Capital y Río Cuarto) el mismo porcentaje de variabilidad que la primera base del Modelo para Datos Funcionales, siendo que para estos últimos no se encuentran comportamientos identificables más allá de la segunda base. Los cuatro modelos, Lee-Carter y Funcionales para Capital y Río Cuarto, presentan una medida de bondad de ajuste (MAPE) menor al 5%. Sin embargo, para el departamento Capital los intervalos obtenidos para tasas específicas de mortalidad son más estrechos para el Modelo de Datos Funcionales, mientras que, para Río Cuarto, los más acotados resultan los que se obtienen por el modelo de Lee-Carter. Luego del pronóstico de las tasas específicas de mortalidad se destacan intervalos de amplitudes tan extensas que no resultan en absoluto informativos, reforzando la necesidad de agrupar las edades simples en intervalos. Sin embargo, un ejercicio realizado sobre los mismos departamentos, trabajando con

grupos quinquenales de edad, no arrojó ninguna mejoría con respecto a los resultados obtenidos previamente.

Más allá de los resultados técnicos, como ya se mencionó, estos modelos surgieron en un contexto histórico particular, en la década de 1990, cuando el envejecimiento poblacional se catalogó de “problema”, y en aparente contraposición a la metodología clásica, presenta como “inocuo” basarse sólo en la información y la tendencia que brindan las cifras, como si la intervención de una opinión de experto sesgara los resultados. Una de las críticas que se le hace a la metodología clásica es que esta se basa en supuestos, que son tendencias postuladas por el demógrafo de acuerdo a su experiencia. Sin embargo, el basarse sólo en datos cuantitativos no proporciona resultados mejores, necesariamente. Y si bien los métodos probabilísticos no se basan en supuestos teóricos-demográficos se basan en supuestos matemáticos sobre una cierta regularidad en los datos que permite aplicar modelos de series temporales (que constituyen una de las etapas de todo modelo probabilístico de proyección).

Puede pensarse que el punto clave es que las proyecciones se basan en una historia de aumento de la esperanza de vida, y si se proyecta en base a esta lógica, las cifras obtenidas por los métodos probabilísticos son siempre superiores a las obtenidas por métodos clásicos. En el caso de los modelos aplicados podría esperarse que las curvas que representan bases y coeficientes revelaran patrones subyacentes en la historia de la mortalidad de las regiones en estudio, sin embargo, no es así. Las interpretaciones se hacen difíciles y resultan forzadas más allá del segundo par (Modelo de Datos Funcionales).

Si bien entre sus ventajas el Modelo para Datos Funcionales cuenta con que es aplicable a las tres componentes de la población, y que los pronósticos de cada una de ellas pueden combinarse para producir pronósticos probabilísticos de la población por edades, mientras que Lee-Carter se emplea únicamente para la mortalidad, es necesario tener en cuenta, en primer lugar, que es deseable emplear una metodología de pronóstico que permita múltiples niveles de desagregación y que sea de una implementación relativamente fácil. Sin embargo, las limitaciones en ciertos casos van más allá de la metodología puntual, es decir, a pesar de cuantificar la incertidumbre asociada a los pronósticos, en contraposición con las herramientas.

Hasta el momento, desde cualquiera de las tres perspectivas (matemática, estadística o demográfica) e incluso combinando estos métodos, es posible conseguir pronósticos con un grado de incertidumbre estimado y controlado a mediano plazo (aproximadamente, 30 años). Sólo los métodos estadísticos ponen de manifiesto el grado de incertidumbre que conllevan los pronósticos.

REFERENCIAS BIBLIOGRÁFICAS

- Andreozzi, L. (2016). “Proyecciones probabilísticas en Demografía” (Tesis de Maestría). Facultad de Ciencias Económicas y Estadística. Universidad Nacional de Rosario.
- Andreozzi, L. (2018). “Métodos probabilísticos de pronóstico de la mortalidad y su aplicación a tres departamentos de la provincia de Córdoba” (Tesis Doctoral). Escuela de Graduados de Ciencias Económicas. Doctorado en Demografía. Universidad Nacional de Córdoba.

- Andreozzi, L. and 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
- Blaconá, M. T y L. Andreozzi (2012), "Comparación de métodos de estimación del modelo de Lee y Carter", en Estadística, 64 (182 y 183), Santiago de Chile, Instituto Interamericano de Estadística, pp.57-84.
- Bongaarts, J. y Bulatao, R. A. (2000). Beyond Six Billion: Forecasting the World's Population. Panel on Population Projections, Committee on Population, National Research Council. NATIONAL ACADEMY PRESS.
- Booth, H., Hyndman, R., Tickle, L., y de Jong, P. (2006a). Lee-Carter mortality forecasting: a multi-country comparison of variants and extensions. Demographic Research, 15(9):289–310.
- Booth, H., Hyndman, R. J., Tickle, L., y de Jong, P. (2006b). Lee-Carter mortality forecasting: a multi-country comparison of variants and extensions.
- Booth, H., Maindonald, J., y Smith, L. (2002). Applying Lee-Carter under conditions of variable mortality decline. Population Studies, 56(3):325–336.
- Booth, H., Tickle, L., y Smith, L. (2005). Evaluation of the variants of the Lee-Carter method of forecasting mortality: A multi-country comparison. New Zealand Population Review, 31(1):13–34.
- Burch, T. K. (2003) Data, models, theory and reality: the structure of demographic knowledge, en: F. C. Billari y A. Prskawetz (Eds) Agent-based computational demography: using simulation to improve our understanding of demographic behaviour (Heidelberg, Physica-Verlag), 19–40.
- Dogan, M. y Pahre, R. (1993), Las nuevas ciencias sociales. La marginalidad creadora, México DF, Grijalbo.
- García Guerrero, V. M. y M. Ordorica Mellado (2012), "Proyección estocástica de la mortalidad mexicana por medio del método de Lee-Carter", en Estudios Demográficos y Urbanos, vol. 27, núm. 2, mayo-agosto, México, El colegio de México, pp. 409-448. Disponible en: <<http://www.redalyc.org/articulo.oa?id=31226408004>>.
- Girosi, F. y King, G. (2008), Demographic Forecasting Princeton University Press, Princeton.
- 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_pensiones/premio_2007.shtml>
- Guimaraes, R.R.M, (2014) Uncertainty in population projections: the state of the art. Revista Brasileira de Estudos de População 31 (2), 277-290.
- Hobcraft, J. (2002) Moving beyond elaborate description: towards understanding choices about parenthood, in: M. Macura & G. Beets (Eds) The dynamics of fertility and partnership in

- Europe: insights and lessons from comparative research, Vol. I (New York and Geneva, The United Nations), 131–143.
- Hyndman, R. y Booth, H. (2007). Stochastic population forecasts using functional data models for mortality, fertility and migration.
- 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.
- Hyndman, R. y Ullah, M. (2007). Robust forecasting of mortality and fertility rates: A functional data approach. *Computational Statistics and Data Analysis*, 51:4942–4956.
- Lee, R. (1993). Modeling and forecasting the time series of us fertility: age distribution, range, and ultimate level. *International Journal of forecasting*, 9:187–202.
- Lee, R. y Carter, L. (1992). Modeling and Forecasting U. S. Mortality. *Journal of the American Statistical Association*, 87:659–671.
- Lee, R. y Miller, T. (2001). Evaluating the performance of the Lee-Carter Method for Forecasting Mortality. *Demography*, 38:537–549.
- Lee, R. D. y R. Rofman (1994), “Modelación y Proyección de la Mortalidad en Chile”, en *Notas de Población*, XXII, núm. 59, Santiago de Chile: CEPAL, pp. 182-213, junio.
- 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.
- Lee, R. D., and Tuljapurkar, S. (1997), “Death and Taxes: Longer Life, Consumption, and Social Security,” *Demography*, 34, 67–81.
- Li, S. y Chan, W. (2005). Outlier analysis and mortality forecasting: The United Kingdom and Scandinavian countries. *Scandinavian Actuarial Journal*, 2005(3):187–211.
- Miller, T. (2003). California's uncertain population future. technical appendix. Special Report: the growth and aging of California's population: demographic and fiscal projections, characteristics and service needs. center of Economics and Demography of aging. CEDA.
- Miller, T. y Lee, R. (2004). A probabilistic forecast of net migration to the United States. Report I in *Stochastic infinite horizon forecasts for social security and related studies*, 10917.
- Minoldo, S. y Pérez Díaz (2018). Separando aguas en el debate de las pensiones: Del abordaje metodológico-científico a la discusión política”, ponencia presentada en el XIII Congreso Internacional de Ciencias Sociales Interdisciplinarias, Granada.
- Preston, S. H., and Campbell, C. (1993), “Differential Fertility and the Distribution of Traits: The Case of IQ,” *American Journal of Sociology*, 98, 997–1019
- Dugald Murdoch (1987), *Niels Bohr's Philosophy of Physics*, Cambridge University Press, Cambridge (Reino Unido), p 47, citando los archivos de Niels Bohr, correspondencia científica. Extraído del libro *Einstein su vida y su Universo* de Walter Isaacson, 2014, p 368
- Preston, S., Heuveline, P., y Guillot, M. (2001). *Demography, Measuring and Modelling Population Processes*. Blackwell, Oxford.
- Ramsay, J. O. y Silverman, B. W. (2005). *Functional data analysis* 2nd ed. Springer, New York.
- Sautu R., (2003). *Todo es Teoría. Objetivos y métodos de investigación*. Buenos Aires: Lumiere.

-
- Schumacher, E. F. en: Box, G. E. P., Hunter, W. G. and Hunter, J. S. (2002). Estadística para investigadores. Introducción al diseño de experimentos, análisis de datos y construcción de modelos. Barcelona: Editorial Reverté.
- Strauss, A & Corbin, J. (2002). Bases de la investigación cualitativa. Técnicas y procedimientos para desarrollar la teoría fundamentada. Bogotá: Editorial Universidad de Antioquia.
- Tonon, G. ed. (2015). Qualitative Studies in Quality of Life. Methodology and Practice. New York, London: Springer-Heilderberg-Dordretch, Social Indicators Research Series 55. ISBN 978-3-319-13778-0.
- Varsavsky, O. (1969) Ciencia, política y cientificismo. Buenos Aires, CEPAL.