



Escuela de Graduados
de la Facultad de
Ciencias Económicas

Universidad
Nacional de
Córdoba



DOCTORADO EN DEMOGRAFÍA

Métodos probabilísticos de pronóstico de la mortalidad y su aplicación a tres Departamentos de la Provincia de Córdoba

Por Lucía ANDREOZZI

Tesis doctoral presentada para optar al grado de Doctora en Demografía

Director: Eduardo Torres

Co-Directora: María Teresa Blaconá

Córdoba, septiembre de 2018



Métodos probabilísticos de pronóstico de la mortalidad y su aplicación a tres Departamentos de la Provincia de Córdoba por Lucía ANDREOZZI se distribuye bajo una [Licencia Creative Commons Atribución – No Comercial – Sin Obra Derivada 4.0 Internacional](https://creativecommons.org/licenses/by-nc-nd/4.0/).



Dedicado a la memoria de mi padre,

Raúl H. Andreozzi,

y de mis abuelos,

Susana D. Añon y Hugo L. Paviotti.



Agradecimientos

Gracias a mi director Eduardo por acompañarme en este arduo camino,
a Dora por animarme a seguir,
a Leandro por enseñarme a ver más allá de las cifras,
a MT (Profe) mi codirectora, por acompañarme y guiarme, por tercera vez,
a los cuatro por su calidez humana y su calidad científica,
a mi familia, Mamá y Dui ¡Se viene la tercera foto de los tres!,
a mis hermanas de la vida, Ivana, Ana, no estaría aquí sin ustedes,
a Ceci, por compartir tantas alegrías,
a Claudia, Felipe y Nati, mis amigos de la demografía,
a todo el equipo humano del Doctorado en Demografía de la UNC, por acompañarme en una etapa maravillosa de mi vida,
a Rosana, Lía y Luciana, gracias, gracias, gracias,
a mi equipo de investigación: Guille, Patri, Miriam y Javier, porque trabajar día a día con ustedes es un gusto,
a mis compañeros de la facultad, por entender mis cursos, mis tiempos y más,



Índice

1. Introducción	5
2. Antecedentes	12
2.1. Los modelos probabilísticos	13
2.1.1 El Modelo de Lee-Carter	13
2.1.2 Modelo para Datos Funcionales	16
2.2. Ejemplos de aplicaciones	18
2.3. La crítica pendiente	20
2.4. Un análisis de las nuevas metodologías más allá de la técnica	23
3. Marco Teórico	27
3.1. La relación entre la demografía y la estadística	28
3.2. Pronósticos probabilísticos y ciencias sociales	35
3.3. Hipótesis y Objetivos	36
4. Métodos	38
4.1. Modelos estadísticos para el pronóstico de las tasas de mortalidad	39
4.1.1. Modelo de Lee-Carter	39
4.1.2. Modelo para Datos Funcionales	40
4.2. Otros aspectos metodológicos	42
4.2.1. Formas de intervención del demógrafo	42
4.3. Fuentes de Datos, Depuración y Procesamiento a través del software R	43
5. Resultados	47
5.1. Estimación de los modelos	48
5.1.1. Departamento Capital	49
5.1.2. Departamento Río Cuarto	59
5.1.3. Departamento Sobremonte	69
5.2. Pronósticos	72



6. Conclusiones	78
7. Glosario	86
8. Referencias bibliográficas	91
9. Anexo	99



Índice de figuras

Tabla 1. Tasas de mortalidad por grandes grupos de edad para departamentos seleccionados, años 1980, 1990, 2000 y 2010	48
Gráfico 1. Logaritmo de las tasas de mortalidad observadas. Departamento Capital (1980-2013)	50
Gráfico 2. Suavizado del logaritmo de las tasas de mortalidad observadas. Departamento Capital (1980-2013)	51
Gráfico 3. Modelo para Datos Funcionales. Departamento Capital	52
Gráfico 3.1 Modelo para Datos Funcionales. Departamento Capital. Primer par base-coeficiente	53
Gráfico 3.2 Modelo para Datos Funcionales. Departamento Capital. Segundo par base-coeficiente	54
Gráfico 3.3 Modelo para Datos Funcionales. Departamento Capital. Tercer par base-coeficiente	55
Figura 1. Esquematización del proceso de obtención de una escala de colores a partir de una superficie de nivel.	56
Gráfico 4. Residuos del modelo MDF. Departamento Capital	57
Gráfico 5. Modelo de Lee-Carter. Departamento Capital	58
Gráfico 6. Residuos del modelo LC. Departamento Capital	59
Gráfico 7. Logaritmo de las tasas de mortalidad observadas. Departamento Río Cuarto (1980-2013)	60
Gráfico 7.b. Logaritmo de las tasas de mortalidad observadas. Departamento Río Cuarto (1980-	



2013)(recorte)	61
Gráfico 8. Suavizado del logaritmo de las tasas de mortalidad observadas. Departamento Río Cuarto (1980-2013)	62
Gráfico 9. Modelo para Datos Funcionales. Río Cuarto.	63
Gráfico 9.1. Modelo para Datos Funcionales. Río Cuarto. Primer par base-coeficiente	64
Gráfico 9.2. Modelo para Datos Funcionales. Río Cuarto. Segundo par base-coeficiente	65
Gráfico 9.3. Modelo para Datos Funcionales. Río Cuarto. Tercer par base-coeficiente	66
Gráfico 10. Residuos del modelo MDF. Río Cuarto	67
Gráfico 11. Modelo de Lee-Carter. Río Cuarto	68
Gráfico 12. Residuos del modelo LC. Río Cuarto	69
Gráfico 13. Logaritmo de las tasas observadas Departamento Sobremonte (1980-2013)	70
Tabla 2. Comparación del porcentaje de variabilidad explicada entre edades simples y agrupadas. Departamentos Capital y Río Cuarto, mediante Modelo para Datos Funcionales y Modelo de Lee-Carter.	71
Tabla 3. 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.	72
Tabla 4. Tasas de mortalidad pronosticadas para el Departamento Capital a través de MDF (selección de edades y años)	73
Tabla 5. Tasas de mortalidad pronosticadas para el Departamento Capital a través de LC (selección de	



edades y años)	74
Tabla 6. Tasas de mortalidad pronosticadas para el departamento Río Cuarto a través de MDF (selección de edades y años)	75
Tabla 7. Tasas de mortalidad pronosticadas para el departamento Río Cuarto a través de LC (selección de edades y años)	76
Tabla 8. Esperanza de Vida al Nacer pronosticada. Departamento Capital y Río Cuarto. Modelos LC y MDF para años seleccionados.	77
Tabla 9. Esperanza de vida al nacer según sexo para la provincia de Córdoba 2015-2040	77

1. Introducción

Cuando en el año 1996 las Naciones Unidas proyectaron la población mundial para el año 2050 en base a la variante media, la cifra obtenida resultó menor en 466 millones de personas respecto a la proyección realizada 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. Lo más importante es que ambas proyecciones tenían intervalos de incertidumbre que se superponían, lo que significa que en algún caso ambas proyecciones son iguales o muy próximas, y si bien los productores de las proyecciones comprendían este punto, la mayoría de los consumidores de los datos, no.

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 otorgan 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 esta característica y remarcaba la necesidad de desarrollar métodos que la cuantifiquen.

La fecundidad, la mortalidad y la migración constituyen las componentes de crecimiento de la población y 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 cálculos se obtiene la proyección de la población con dicha especificación. Esta

metodología es conocida como el método de las componentes y representa el modo más tradicional de realizar una proyección demográfica.

Es inevitable que las proyecciones contengan incertidumbre y si bien la situación demográfica actual es conocida, las tendencias en los nacimientos, muertes y migraciones están sujetas a cambios impredecibles. Así, los errores cometidos en la elaboración de proyecciones varían de acuerdo al horizonte de pronóstico, el nivel de desarrollo del país y su tamaño. Mientras que las tendencias generales 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 a la tendencia particular de un país o región.

Es importante señalar que el método de las componentes presenta una inconsistencia probabilística cuando se pretende considerar la incertidumbre asociada a las proyecciones mediante la construcción de escenarios con distintas hipótesis baja-medio-alta, dado que resulta imposible construir escenarios que reflejen simultáneamente la incertidumbre en todas las variables de interés: si el rango alto-bajo se diseña para contener las fluctuaciones a corto plazo en las tasas demográficas correspondientes a las tres componentes, luego será demasiado amplio para indicar la incertidumbre asociada al tamaño de la población pronosticado, es decir, es decir a largo plazo, dado que algunas fluctuaciones a corto plazo suelen cancelarse en el largo plazo.

Aún cuando estas fluctuaciones se mantienen dentro de los límites de los escenarios bajo y alto, otros parámetros demográficos derivados (tales como la proporción de la población de una determinada edad o ciertas relaciones de dependencia) pueden verse afectados mucho más y pueden exceder los niveles definidos por estos escenarios. En síntesis, los escenarios alto y bajo no representan extremos consistentes a lo largo de todos los parámetros demográficos.

Esta inconsistencia probabilística y la dependencia de los métodos tradicionales a fuertes supuestos elaborados por demógrafos y expertos, entre otros aspectos, propiciaron el desarrollo de métodos probabilísticos de pronóstico

que, cada vez con mayor aceptación, son implementados por las agencias de estadísticas oficiales para producir sus pronósticos nacionales, como las de Holanda y Estados Unidos, por mencionar algunos ejemplos. Los mismos poseen su principal ventaja en la consistencia probabilística a través de todas las variables pronosticadas y sus índices derivados (Lee y Tuljapurkar, 1994).

Si bien los modelos probabilísticos constituyen un aporte contundente a la Demografía, tanto desde la diversidad como la calidad de los insumos que estas metodologías generan, se destaca que dicha metodología proviene de la Ciencia Estadística y se constituye en una herramienta puesta 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 mortalidad, tienen como principal función la generación de pronósticos y, en este punto, es necesario evaluar de qué forma proveen información válida, que permita desarrollar teorías demográficas o que ayude a comprender la dinámica de la mortalidad de modo que le sea útil tanto a los organismos oficiales como 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.

Finalmente, si bien las proyecciones o pronósticos probabilísticos en demografía surgen con la finalidad de cuantificar la incertidumbre asociada a los resultados, en contraste con las metodologías clásicas que carecen de dicha medida asociada, terminan por fomentar la generación de una gran variedad de extensiones y posibilidades de estimación, (a partir de Lee-Carter y el Modelo para Datos Funcionales se generan numerosas propuestas, algunas con leves y otras con notorias variaciones en la metodología tanto de estimación como de pronóstico) de modo tal que otro tipo de incertidumbre aparece: la de la de seleccionar una metodología oficial.

La mayoría de los estudios se reducen a un análisis estadístico para elegir una metodología de acuerdo a criterios técnicos cómo la reducción del error (ya sea de estimación o de pronóstico), por mencionar uno, quedan múltiples facetas por conocer, por ejemplo, si estos modelos son aplicables a las diversas regiones con sus diversas complejidades demográficas y de disponibilidad de datos base, si hay una relación con la calidad de los datos o con la magnitud de la región a pronosticar y, principalmente, si los resultados que arroja son coherentes con la historia que la componente expone desde un análisis sociológico y si es realmente un aporte frente a metodologías más simples y clásicas, teniendo en cuenta además que a la hora de su aplicación práctica las nuevas propuestas requieren conocimientos estadísticos y computacionales profundos.

El foco principal de estos métodos persigue cuantificar la incertidumbre: esto sería de gran utilidad para los usuarios de las proyecciones y dado que los escenarios alternativos que se elaboran como parte de una proyección demográfica tienen diferentes implicancias en el futuro, conocer una medida de la incertidumbre asociada a cada resultado permite decidir en base a mayor información. Actualmente la posible variación en los resultados se expresa proveyendo escenarios alternativos con variantes para la fecundidad (y con menos frecuencia en la mortalidad y la migración). De este modo los escenarios alto y bajo se utilizan para indicar un rango de posibles futuros, no obstante, ninguna probabilidad específica se asocia a estos resultados.

De esta manera, mejorar las proyecciones es de gran importancia ya que las tendencias del tamaño de la población, la estructura de edad, nacimientos, muertes y otras variables demográficas resultan de interés para una amplia gama de analistas, políticos, científicos y planificadores de la industria y del gobierno. Contar con información de calidad en estos aspectos constituye un insumo básico para la toma de decisiones en las más diversas áreas. Por ejemplo, las tendencias demográficas a nivel nacional y mundial constituyen información necesaria para proyectar las demandas futuras de agua, alimentos y energía, además de que permiten estimar el impacto ambiental debido al consumo de recursos naturales. Asimismo, las proyecciones subnacionales contribuyen, por

mencionar alguna de sus múltiples utilidades, a la planificación de la inversión, dado que en base a ellas se establecen los sitios donde deben construirse escuelas o caminos. Puntualmente, las proyecciones se utilizan, en conjunto con otros indicadores, para determinar la asignación de fondos destinados a infraestructura en general.

En la literatura se han presentado diferentes modelos, aunque no muchos trabajos hablan sobre sus virtudes y desventajas más allá de una magnitud del error que se comete al pronosticar y, como ya se ha mencionado, es aquí donde se centra el interés de este trabajo. Es por ello que en esta tesis se presentarán los modelos probabilísticos, sus principales características y se revisarán sus virtudes y limitaciones. Cabe resaltar que hasta el momento esta metodología no es de uso frecuente en Argentina y para contribuir a su estudio se presentará en este trabajo una aplicación dirigida a proyectar las tasas específicas de mortalidad de tres departamentos de la Provincia de Córdoba; La Capital, Río Cuarto y Sobremonte, con información disponible para un período de estudio que abarca desde el año 1980 hasta 2013. La aplicación en los tres casos se realiza mediante dos modelos: el tradicional Modelo de Lee-Carter y una de las últimas propuestas, el Modelo para Datos Funcionales. La información base proviene, por un lado de los Censos Nacionales de 1980, 1991, 2001 y 2010 para las cifras de población, y de la dirección de Estadística e Información de Salud para las cifras de muertes; de este modo se construyen las tasas específicas de mortalidad por edad y género que permiten llevar adelante las mencionadas aplicaciones.

La tesis está estructurada en secciones, a continuación (Sección 2) se presentan los antecedentes, un análisis exhaustivo de las aplicaciones existentes y las ventajas y desventajas que refieren los investigadores, luego se detalla el marco teórico (Sección 3), el cual hace un recorrido a través de las dos ciencias que confluyen para la aplicación de los modelos probabilísticos: la Demografía y la Estadística. Además analiza cómo es posible enmarcar estas técnicas estadísticas en el contexto de las ciencias sociales, entendiendo su génesis, sus limitaciones y sus ventajas prácticas. Dentro del mismo apartado se plantean las hipótesis y objetivos de trabajo y en la Sección 4 se introducen las metodologías;

Modelo de Lee-Carter y Modelo para Datos Funcionales. En el mismo apartado describe el proceso de depuración de los datos base, y luego en la Sección 5 se presentan los resultados obtenidos para los tres departamentos seleccionados, mediante los dos modelos probabilísticos antes mencionados. Finalmente en la Sección 6 se detallan las principales conclusiones del trabajo.

En las Secciones 7 a 9 se pueden hallar el glosario, las Referencias Bibliográficas y el Anexo, respectivamente.

2. Antecedentes

2.1. Los modelos probabilísticos

El cálculo de más de una variante de pronóstico es la práctica estándar entre las agencias oficiales de estadística (Keilman y Crujisen, 1992). Esta práctica pretende dar cuenta del hecho de que el futuro es intrínsecamente incierto, y que las diferentes suposiciones de pronóstico conducirán a diferentes resultados de pronóstico. Sin embargo, un problema importante es que el enfoque convencional es completamente determinista, es decir, no se incluyen distribuciones estadísticas en el modelo de pronóstico. Por lo tanto, no se asocian probabilidades a cada variante dejando en manos del usuario la elección de una de ellas para emplear como insumo para su análisis. En cambio los pronósticos probabilísticos presentan a las pirámides de tamaño de población y edad futuras, a las tasas específicas de mortalidad, a las de fecundidad no como un número (o tal vez unos pocos, dependiendo del número de variantes), sino como un rango completo de distribuciones de probabilidad. El futuro es inherentemente incierto, sin embargo, algunos resultados demográficos son más probables que otros y son las distribuciones de probabilidad las que dicen cuánto más probable son. Sin embargo, casi todos los pronósticos oficiales son deterministas, es decir, no probabilísticos.

En resumen, en los modelos deterministas, el resultado está totalmente determinado por los valores de los parámetros y las condiciones iniciales. Mientras que, los modelos estocásticos poseen cierta aleatoriedad inherente. Un mismo conjunto de valores de parámetros y de condiciones iniciales conducirán a un conjunto de resultados. El mundo, la sociedad, los seres humanos presentan aleatoriedad, pero, los modelos estocásticos son considerablemente más complicados. Es por ello que resulta apropiado evaluar cuando los modelos deterministas proporcionan una aproximación útil de procesos intrínsecamente estocásticos.

2.1.1. El modelo de Lee-Carter

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 (Lee y Carter, 1992), el cual posee actualmente numerosas variantes y extensiones. Originalmente los autores proponen una metodología que permite modelar y extrapolar las tendencias observadas en las tasas específicas de mortalidad a largo plazo e implementan dicha metodología para pronosticar la mortalidad de los Estados Unidos hasta el año 2065. Lee y Miller (2001) y Booth *et al.* (2002) han propuesto modificaciones al método de Lee-Carter, como la elección del período de ajuste, el método para el ajuste del parámetro de nivel y la elección de las tasas base para el pronóstico. La propuesta de Lee y Miller (2001) es ampliamente utilizada y la variante propuesta por Booth *et al.* (2002) ha demostrado ser al menos tan precisa como la de Lee y Miller en el corto plazo, ver Booth *et al.* (2005) y Booth *et al.* (2006a).

Varios desarrollos han incorporado una estructura de error heterocedástica Poisson (como Brouhns *et al.* (2002) y Wilmoth (1993)). Otros autores han extendido la aplicabilidad del modelo de Lee-Carter a los factores de reducción de la mortalidad, ver Renshaw y Haberman (2003a) o examinado el uso de más de un término en el modelo, como en el caso de Booth *et al.* (2002) y Renshaw y Haberman (2003b). También existen dos extensiones recientes que incluyen suavizados semi-paramétricos en el modelo; Jong y Tickle (2006) combinan un suavizado por *splines* y una estimación por medio del filtro de Kalman para ajustar una versión generalizada del modelo de Lee-Carter. Hyndman y Ullah (2007) proponen utilizar el paradigma de los datos funcionales para modelar las componentes demográficas; el mismo considera que los logaritmos de las tasas de cada año son una función continua de las edades. Hyndman y Ullah (2007) suavizan la mortalidad a través de regresiones *spline* penalizadas para luego ajustar un modelo mediante una descomposición en componentes principales. Estos métodos son comparados por Booth *et al.* (2006b).

Es así que los mayores avances en relación al desarrollo de metodologías específicas se presentaron en relación a la mortalidad, en cambio, la fecundidad y las migraciones han recibido menor atención por parte de los investigadores. En cuanto al pronóstico de la fecundidad se destaca, en primer lugar, que los métodos elaborados para dicha componente se encuentran menos

desarrollados que los destinados al análisis de la mortalidad. Una explicación está basada en que la fecundidad presenta dificultades a la hora de ser pronosticada debido a los cambios estructurales en su comportamiento están altamente asociados a cambios en las pautas culturales.

En ese sentido, Lee (1993) encontró necesario pre especificar el valor medio a largo plazo de la fecundidad total e imponer valores límites para reducir la amplitud del intervalo de predicción e implementó un método paralelo al del Lee y Carter. Por otro lado, un enfoque de componentes principales fue empleado por Bozik y Bell (1987), quienes utilizan en su análisis los primeros cuatro componentes principales y modelos ARIMA¹ multivariados. Un método similar emplean Hyndman y Ullah (2007) como parte de un análisis funcional, el cual es el que se utiliza en el presente trabajo, donde los autores emplean un mismo modelo base para analizar la mortalidad y la fecundidad.

Por su parte, los métodos para pronosticar la migración están aún menos desarrollados y son habitualmente mucho más simples (George y Perreault, 1992). La razón principal de esta falta de desarrollo radica en la carencia de series temporales confiables, detalladas y que representen adecuadamente los flujos migratorios internacionales. Además, las tendencias observadas en esta componente son más volátiles, debido a los acontecimientos políticos y económicos y los cambios en la legislación. Un método que puede solucionar la falta de datos es estimar la migración neta como la diferencia entre el incremento del tamaño de la población y el crecimiento natural utilizando la ecuación compensadora, también llamada ecuación de crecimiento. Más aún, para pronósticos de población subnacionales que involucran efectos de la migración interna, este método es el único disponible ya que es frecuente que no se registren datos sobre migración a ese nivel de desagregación (ver Miller (2003); Miller y Lee (2004)). Además, el uso de la migración neta no contempla un aspecto importante: la bondad de pronóstico depende de la desagregación de

¹ Box y Jenkins desarrollaron modelos estadísticos para series temporales que tiene en cuenta la dependencia existente entre los datos, esto es, cada observación en un momento dado es modelada en función de los valores anteriores. Los modelos se conocen con el nombre genérico ARIMA (Autoregressive Integrated Moving Average) que deriva de sus tres componentes AR: autorregresivo, I: integrado y MA: promedio móvil.

acuerdo a diferentes grupos o tipos de inmigrantes y emigrantes (Hilderink *et. al.*, 2002 y Rogers, 1990). Por otro lado, Beer (1997) encontró consistencia entre pronósticos de series de tiempo de inmigración total, emigración y migración neta. También es posible obtener los pronósticos de migración neta desagregados por edad (Keilman y Pham, 2004), utilizando una versión reducida del modelo multi-exponencial desarrollado en Rogers y Castro (1981) y Rogers y Little (1994).

2.1.2. Modelo para Datos Funcionales.

La metodología propuesta por Hyndman y Ullah (2007) difiere en varios aspectos importantes con respecto al modelo de Lee-Carter; en primer lugar está enmarcado en el paradigma de datos funcionales (Ramsay y Silverman, 2005), por otro lado emplea suavizados semi-paramétricos con el fin de reducir la aleatoriedad inherente a los datos observados y en la etapa de descomposición de las tasas suavizadas permite utilizar componentes principales clásicos o robustos. Estos últimos tienen en cuenta dentro del análisis, los posibles *outliers*. En relación al modelado de *outliers*, Li y Chan (2005) proponen una versión robusta del modelo de Lee-Carter, pero ésta no se ubica dentro del paradigma de los datos funcionales.

Durante los últimos años se han desarrollado múltiples enfoques para pronósticos probabilísticos (Booth *et al.*, 2006a) y dentro de esta categoría han cobrado una particular relevancia los métodos para datos funcionales² (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 para todas las componentes demográficas (Hyndman y Ullah, 2007). Un aporte interesante a este enfoque lo hacen Hyndman *et al.* (2013) al introducir la idea de pronósticos coherentes en el paradigma de datos funcionales. La idea principal de esta propuesta radica en que la diferencia entre los pronósticos de grupos de interés debe permanecer constante a través

² La filosofía básica del análisis funcional radica en pensar a los datos observados como funciones en lugar de una secuencia de observaciones individuales. El término funcional se refiere a la estructura intrínseca de los datos, más que a su forma explícita.

del tiempo, reproduciendo la relación presente en los datos observados. Los grupos de interés pueden ser subregiones geográficas o géneros, por mencionar algunas posibilidades.

El método de Hyndman y Ullah (2007) sigue el paradigma de datos funcionales y propone suavizar las tasas específicas de mortalidad -mediante regresiones *spline* penalizadas- para luego ajustar un modelo utilizando una descomposición en componentes principales. El enfoque de datos funcionales permite modelar la mortalidad, la fecundidad y las migraciones, para obtener como producto final la proyección estocástica de población, acerca de la mencionada metodología García Guerrero y Ordorica (2012) afirman:

“El último y más reciente exponente es el trabajo de Hyndman y Booth (2008). Estos autores proponen un método muy innovador para pronosticar las poblaciones basándose en modelos de datos funcionales. A grandes rasgos, su método trata de modelar las tasas vitales por edad y sexo de acuerdo con modelos funcionales con coeficientes estimados con series de tiempo. La migración es estimada como la diferencia entre los datos poblacionales históricos y los pronosticados para cada año. Para eliminar el error de pronóstico se utiliza la transformación de Box-Cox. Después, con el fin de hacer más robusto el pronóstico, se asume que los nacimientos y las defunciones siguen una distribución de Poisson. Finalmente, se utiliza el método de simulación de Montecarlo para pronosticar la población total. Esta nueva técnica parece ser muy robusta y sería muy interesante que se desarrollara en trabajos futuros con el fin de comparar los pronósticos aquí propuestos”. (Página 413)

Finalmente, si bien resulta interesante el análisis y pronóstico de cada componente por separado, el objetivo último suele ser integrar los resultados para obtener una proyección probabilística de la población. De este modo, los componentes pronosticados (mortalidad, fecundidad y migraciones) son utilizados para generar pronósticos probabilísticos de población a través de la

renovación estocástica de la población usando el método de las componentes Preston *et al.* (2001).

El pronóstico de población se obtiene analíticamente a partir de la matriz estocástica de Leslie (Alho y Spencer, 1985, Lee y Tuljapurkar, 1994 y Sykes, 1969) o de un modo más simple por simulación de Monte Carlo, generando una distribución de los posibles resultados. Para ambos enfoques es necesario especificar la media (o mediana), una estructura de variancia-covariancia y la forma de la distribución de cada componente demográfica.

2.2. Ejemplos de aplicaciones

Lee y Rofman (1994) realizan una aplicación a datos de Chile, y en sus conclusiones sostienen que la síntesis de técnicas estadísticas de series de tiempo y modelos demográficos tiene grandes ventajas. Así, los autores afirman que un modelo simple es útil para completar vacíos en los datos de edad y períodos de tiempo y además en algunos casos se emplea para ajustar el subregistro y proyectar la mortalidad; con obvias ventajas dentro de la estructura “estadística-demográfica” integrada.

Además, Guerrero y González (2007) desarrollan un trabajo en base a la aplicación del modelo a datos de Méjico evaluando el impacto de los resultados sobre el sistema previsional.

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

“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." (Página 445)

En Argentina se difundió el primer trabajo sobre el modelo de Lee y Carter con datos a nivel país en el año 2011 (Blaconá y Andreozzi, 2011), posteriormente se realizó una comparación entre distintos métodos de estimación (Andreozzi y Blaconá, 2012). En el trabajo de 2011 se aplica el modelo de Lee-Carter a las tasas de mortalidad específicas por edad y género de la República Argentina, y el período analizado abarca desde el año 1979 hasta el año 2006. En dicha investigación se trabaja con grupos etarios de 0 a 4 años, 5 a 14, 15 a 24, 25 a 34, 35-44, 45 a 54, 55-64, 65-74 y 75 y más, dadas las limitaciones de la información³. A partir de los resultados se obtienen tasas específicas de mortalidad por edad y género y esperanzas de vida al nacer pronosticadas con intervalos de pronóstico de reducida amplitud indicando una buena calidad en los resultados obtenidos.

En relación a la aplicación de datos funcionales la primera aplicación se realiza para Argentina en la tesis de 2016, "Proyecciones probabilísticas en Demografía" presentada por la propia autora para obtener el grado de Magister en Estadística Aplicada.

La aplicación a datos de Argentina demuestra que los modelos para datos funcionales se pueden emplear con éxito para pronosticar tasas de mortalidad específicas por edad y género y tasas globales de fecundidad por edad simple cuando se cuenta con información para períodos no muy extensos de tiempo y una calidad aceptable de las fuentes de datos (en regiones populosas). Es importante remarcar que este enfoque de modelado es altamente adaptable; el mismo modelo básico se aplica a las tres componentes permitiendo transformaciones y suavizados específicos para cada una de ellas aunque la aplicación a datos de migración resulta defectuosa.

Los pronósticos obtenidos para mortalidad y fecundidad mediante datos funcionales presentan estrechos intervalos de confianza indicando una buena

³ Se basa en la tesis de grado de la tesista el cual tenía como principal objetivo la comparación con la ciudad de Rosario y dicha ciudad presentaba la información agregada, imponiendo luego este formato a los datos de Argentina a fines comparativos.

calidad de los pronósticos obtenidos. Luego a partir de las proyecciones de mortalidad y fecundidad (con dos hipótesis: estable y descendente), más un supuesto de migración nula (debido a que la modelización de dicha componente no arroja buenos resultados) se obtienen proyecciones de población. Los métodos empleados son útiles para pronosticar a corto y a mediano plazo, pero la amplitud de los intervalos muestra que los métodos se vuelven menos informativos a lo largo del tiempo y recomiendan no utilizarlos con un horizonte mayor a los 20 años.

Por otro lado se realizó una aplicación de Modelos de Datos Funcionales, una variante Modelo para Datos Funcionales Coherente y Lee-Carter a datos de la Ciudad Autónoma de Buenos Aires (Andreozzi, 2016), el artículo concluye

“Los modelos estadísticos permiten medir la incertidumbre de las proyecciones de la mortalidad (o eventualmente otra componente demográfica) y calcular la esperanza de vida. Proporcionar un intervalo de pronóstico de estos indicadores demográficos resulta más informativo que la mera proyección puntual, principalmente porque en contraposición con la lógica de los escenarios, provienen de una distribución de probabilidad cuantificando así la incertidumbre asociada al resultado mediante un coeficiente de confianza. Los tres métodos proporcionan intervalos de confianza con una precisión similar.” (Página 42)

En este punto es importante remarcar que el hecho de medir la incertidumbre se traduce en obtener para el pronóstico un intervalo de pronóstico asociado. Junto al coeficiente de confianza elegido (tradicionalmente 80%, 90% o 95%) la amplitud del intervalo de pronóstico obtenido es la que da cuenta de la incertidumbre.

2.3. La crítica pendiente

Si bien se han presentado los modelos, no muchos trabajos hablan sobre sus virtudes y desventajas más allá de una magnitud del error que se comete al pronosticar y, como ya se ha mencionado, es aquí donde se centra el interés de

este trabajo. En este sentido, puede apreciarse cómo en los últimos años la Demografía ha avanzado fuertemente en relación a los datos y las técnicas de análisis sobre la temática de la mortalidad. Sin dudas, esto es influencia del positivismo lógico y de la tradición estadística que se remonta a Karl Pearson (Burch, 2003).

En este punto John Hobcraft (2003) introduce una expresión interesante: "retorno teórico", al referirse a la información útil que generan los análisis comparativos de las encuestas de fecundidad. Los modelos probabilísticos de pronóstico de las componentes demográficas, ¿generan un retorno teórico? Si bien Ullah y Finch, realizan un análisis exhaustivo del uso de modelos para datos funcionales en el período 1995 a 2010, tan sólo 3 de 84 estudios se realizan en el área demográfica, destacándose que la mayoría de ellos se aplican en biomedicina, biomecánica y medicina, áreas del conocimiento donde los datos se obtienen de un modo muy distinto al registro de hechos vitales humanos, como sucede en la demografía.

Si bien se destacan las bondades técnicas de los modelos para datos funcionales, su flexibilidad en relación a que no requiere datos equiespaciadas y otras ventajas técnicas, al mismo tiempo no se hace referencia al aporte que ellos pueden hacer para explicar un fenómeno demográfico, para generar teoría o la utilidad comprobada de los pronósticos que generan. En una revisión del estado del arte de la incertidumbre presente en las proyecciones de población, Guimaraes (2014) habla acerca de los tres principales enfoques, el método clásico de componentes, los modelos probabilísticos frecuentistas y el enfoque bayesiano.

Sobre los modelos probabilísticos, que son los que a este trabajo le competen, rescata un trabajo de Girosi y King del año 2008, el cual sostiene que el modelo desarrollado por Lee y sus colegas en año 1992 no conseguirá adaptarse a los postulados de los demógrafos como así también tampoco a los patrones futuros de mortalidad, una crítica no menor, teniendo en cuenta que los defensores de dichas metodologías sostienen que el pronóstico (más que el

modelo descriptivo de la componente) es el fuerte de las nuevas propuestas. Los autores sostienen que a pesar de su maleabilidad y aplicabilidad en múltiples temas demográficos, el modelo de Lee-Carter tiene sus limitaciones, ya que si bien pronostica correctamente lo hace a un corto plazo y no al largo plazo, que sería el insumo que se necesita para evaluar la planificación políticas públicas.

Por su parte, Burch (2003) sostiene que el modelo tiende a “olvidar” los patrones de edad presentes en las tasas de mortalidad cuando se acerca a niveles bajos de dicha componente, además que nada evita que la mortalidad se vuelva nula, técnicamente, cuando este hecho resulta imposible en la realidad. Incluso el propio Lee reconoce que los intervalos de confianza tienden a ser estrechos (lo que se traduciría en pronósticos confiables) como resultado de una baja entropía en las curvas de supervivencia en contextos de altos niveles de esperanza de vida (Lee y Miller, 2001).

Una crítica interesante aporta el informe del año 2013 sobre la incertidumbre en las proyecciones de población realizado por la Oficina Parlamentaria de Ciencia y Tecnología del Parlamento Inglés: en el mismo se sostiene que quienes diseñan políticas públicas tienden a recurrir a los demógrafos para que les provean “la mejor” estimación de la población futura, de su estructura y tamaño (y por ende de la mortalidad que compone dicha estimación). Un rango amplio de futuros, que se informan a través de los intervalos de confianza, no resultan definitivos y muchas veces no resultan útiles. Asimismo, sostiene que los modelos probabilísticos de pronóstico siguen siendo altamente dependientes de la calidad y confiabilidad de los datos. Incluso menciona que los enfoques probabilísticos se basan en el conocimiento de las experiencias pasadas, y que no tienen en cuenta eventos imprevistos tales como la pandemia del HIV o la aparición de nuevas enfermedades.

Sin embargo en este punto es interesante destacar cómo el informe equipara experiencia pasada a datos del pasado, cuando en realidad esa experiencia estaría resumida en las teorías demográficas que se han derivado a partir del análisis de datos provenientes de registros vitales. En este aspecto Sautu

(2003) señala que la tentación de juntar datos es grande (ya sea que se trate de cualquier tipo), entre ellos los secundarios o estadísticos, siendo éstos últimos la única base de los modelos probabilísticos. La autora destaca que esto no constituye un problema siempre y cuando el propósito de dicha recolección (en este caso estadísticas vitales) y producción de datos (en este caso, estimaciones y pronósticos) esté especificado en el propósito del diseño de investigación y apoyado en un marco teórico. Y, es en este aspecto que las metodologías probabilísticas de pronóstico son débiles.

En síntesis las aplicaciones detalladamente enumeradas en esta sección van acompañadas de un análisis de sus resultados, de comparaciones en relación al error cometido pero no se enmarcan en un contexto socio-político, en el que las proyecciones tienen un fin y un uso.

2.4. Un análisis de las nuevas metodologías más allá de la técnica

Por todo lo antes mencionado el ánimo de esta tesis es realizar un aporte en ese vacío que limita los modelos a una herramienta técnica, desconociendo que responden a uno de los interrogantes más necesarios de cubrir en la actualidad (más aún en un contexto de marcado envejecimiento poblacional) y aún cuando su aparición pretende mejorar las cifras y establecer una incertidumbre asociada a los resultados. En el afán de ser cada investigador protagonista y quien contribuya con una nueva metodología se pierde el foco de la verdadera necesidad, arrojar luz sobre el comportamiento de las dinámicas demográficas y aportar una metodología que genere pronósticos útiles y claros (la complejidad de algunos modelos probabilísticos van en detrimento de este aspecto) acordes a los datos y recursos con los que se cuenta en diversas regiones del mundo. Sautu (2003) sostiene que en la práctica, cita como ejemplo la economía, no constituye una preocupación principal el explicitar el marco teórico y suele confundirse la producción de datos con investigación, cayendo en una "datomanía". La estimación y proyección de tasas de mortalidad o cifras de población son una colección de datos ordenados, un conjunto de datos, una serie o una medición proveniente de una encuesta en sí misma no es una

investigación, son datos. En cambio la decisión de utilizar una cierta metodología de investigación no es una cuestión menor, todo lo contrario implica una decisión filosófica, teórica, política y operativa (Tonón, 2015).

En este punto surge uno de los puntos clave, y es válido recordar para darle un contexto a Oscar Varsavsky (1969):

"... La clásica respuesta es que esos no son problemas científicos: la ciencia da instrumentos neutros, y son las fuerzas políticas quienes deben usarlos justicieramente. Si no lo hacen, no es culpa de la ciencia. Esta respuesta es falsa: la ciencia actual no crea toda clase de instrumentos, sino sólo aquellos que el sistema le estimula a crear."
(pp. 8-9)

Estos modelos surgen en un contexto histórico particular, en la década de los 90, cuando el envejecimiento poblacional se cataloga de "problema", en aparente contraposición a la metodología clásica que se basa en supuestos de los demógrafos, 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, una mención a este aspecto puede verse en el trabajo de Minoldo y Pérez Díaz (2018):

"Sin embargo, cabe puntualizar que la idea de una tendencia a la inversión de la estructura por edades, con predominio de las edades mayores, se asienta más sobre proyecciones que sobre la descripción de realidades demográficas contemporáneas (a excepción del caso de Japón). Y al respecto es importante apuntar que las proyecciones de población a futuro no pueden ser interpretadas como estimaciones de alto nivel de certeza, puesto que dependen del cumplimiento, a lo largo de muchos años, de muchos supuestos, que

tienden a asumir que las tendencias actuales de las dinámicas reproductivas tendrán continuidad a lo largo del tiempo." (p. 1)

Y si bien los métodos probabilísticos no se basan en supuestos teóricos-demográficos se basan en supuestos estadísticos 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). Entonces 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.

Haciendo un breve repaso de los resultados, a través de los distintos países se puede observar un patrón en los resultados al comparar los pronósticos obtenidos con la metodología propuesta por cada investigador y los resultados oficiales. Si se comienza con el trabajo de Lee y Carter del año 1992, los autores concluyen que la esperanza de vida obtenida para los Estados Unidos en el año 2065 por ellos es (para ambos sexos) de 86.05 años, mientras que la misma cifra obtenida por actuarios y considerada la oficial es de 80.45 años. Luego, en el trabajo ganador del Premio de Pensiones 2007 otorgado por la Comisión Nacional del Sistema de Ahorro para el Retiro en México, realizado por Guerrero Guzmán y González Pérez, el pronóstico estocástico para el año 2050, es de 79.56 años, mientras que el pronóstico de Naciones Unidas de 78.90 años. Li, Lee y Tuljaprkar presentan los resultados de Corea del Sur y China, para el primer caso concluyen que los resultados son significativamente más elevados a las cifras que provee Naciones Unidas (2002-2050), luego, para el caso de China, el modelo probabilístico arroja inicialmente valores más elevados que UN, y luego a largo plazo las estimaciones se ubican por debajo. Para el caso de Argentina, Buenos Aires y Rosario los métodos probabilísticos arrojan en todos los casos cifras superiores a las oficiales. Si bien se listan varios ejemplos de distintas regiones, la mayoría de las aplicaciones generan esperanzas de vida más altas cuando se emplean métodos probabilísticos y estos son comparados con los pronósticos de

la oficina de estadística de la región o los de UN, por ellos y dado que el surgimiento se da en un contexto histórico en que el envejecimiento poblacional se presenta como “un problema”; es importante tener en cuenta que estos “aumentos” que se obtienen con la metodología más moderna pueden resultar de sustento para las teorías que abogan por la elevación de la edad previsual en pos de la sustentabilidad del sistema previsual.

3. Marco Teórico

3.1. La relación entre la demografía y la estadística

La heterogeneidad o variabilidad es fundamental en cualquier ciencia social. Mayr (1982) sostiene que una de las más importantes y perdurables influencias de la teoría de la evolución de Darwin es la distinción entre "pensamiento poblacional" y "pensamiento tipológico", este último prevalente en la física, es el que intenta descubrir la "esencia" o "verdad" en la sociedad humana. Una vez comprendido, un concepto científico o ley es siempre generalizable a todos los tiempos. Si el fenómeno observado aparece en desorden se toman promedios y no se tienen en cuenta sus desviaciones. En la historia de la estadística, el pensamiento tipológico está asociado al "hombre promedio" de la física social de Quetelet. En contraste, el "pensamiento poblacional" hace de las variaciones el objeto mismo a ser estudiado. Fue Francis Galton, primo de Darwin quien introdujo el pensamiento poblacional en el campo de la estadística y al hacerlo descubrió la regresión y la correlación. (Xie, 2000)

La demografía se ve muy enriquecida al aproximarse a los fenómenos sociales a través del pensamiento poblacional, enfocándose en las variaciones por grupo y las características individuales. Según Hauser y Duncan (1959) la demografía es el estudio del tamaño, la distribución territorial y la composición de la población, sus cambios y los componentes de dichos cambios. Por fortuna; Hauser y Duncan incluyeron explícitamente "composición de la población" y "los cambios en esta" en su definición, ampliando la demografía de modo que incluya dos tipos de demografía: la demografía formal y los estudios de población. La demografía formal cuyo origen puede rastrearse al año 1662 con John Graunt, quien se interesa por la fecundidad, mortalidad, estructura de edad y distribución espacial de la población humana.

Los estudios de población se ocupan de la composición de la misma y sus cambios desde puntos de vista sustantivos, anclados en otras disciplinas, sociológico, económico, biológico o antropológico y su origen puede situarse con Thomas Malthus en 1798. Por definición los estudios de población son interdisciplinarios, en el límite entre la demografía formal y una disciplina sustantiva pero no necesariamente una ciencia social. Definida de este modo, la

demografía provee una base empírica sobre la que se construyen otras ciencias sociales. Es difícil imaginar que una ciencia social pueda avanzar firmemente sin saber primero la información básica sobre la población humana que esta estudia.

Como un campo de la investigación, la demografía gozó de un rápido crecimiento durante el siglo XX. Por ejemplo los miembros de la *Population Association of América* (PAA), la principal asociación de demógrafos de los Estados Unidos fundada en 1931 creció de unos 500 socios en 1956 a más de 3000 en 1999. Este crecimiento es notable dada la virtual ausencia de departamentos de demografía en las universidades norteamericanas (a excepción de Berkley). Para reconocer las contribuciones hechas por demógrafos, uno solamente necesita recordar información fáctica sobre las sociedades contemporáneas, mucho de lo que se conoce como "hechos estadísticos", "estadísticas" sobre la sociedad norteamericana, por ejemplo, ha sido estudiado y proporcionado por demógrafos; desigualdades socioeconómicas por raza (Farley, 1984), género (Bianchi y Spain, 1986), segregación residencial por raza (Duncan, 1957; Massey y Denton, 1993), morbilidad intergeneracional (Blau y Duncan 1967, Fatherman and Hauser 1978-9) por mencionar algunos ejemplos.

Pero más allá de proveer estadísticas, la demografía, también ha sido fundamental para pronosticar las situaciones futuras de las sociedades humanas. Aunque el pronóstico demográfico está sujeto a incertidumbre, como cualquier tipo de pronóstico, los demógrafos pueden predecir el tamaño futuro de la población por edad con un alto grado de confianza, utilizando información relativa a la fecundidad del pasado, regularidades en patrones etarios y probables niveles futuros de mortalidad. Un ejemplo destacado de pronóstico demográfico es el trabajo de Lee y Tuljapurkar (1994,1997) quienes demostraron como las "fuerzas demográficas" impactan dramáticamente en las demandas futuras de la seguridad social.

La demografía formal y los estudios de población no solamente tienen diferentes temas de interés, sino que se basan en diferentes enfoques metodológicos. La demografía formal está basada en la matemática y por ello

estrechamente vinculada a la demografía matemática. La misma brinda un rico arsenal de poderosas herramientas de investigación, como las tablas de vida y la teoría de poblaciones estables, usualmente atribuidas a Alfred J. Lotka en 1922. Es importante señalar que los modelos matemáticos de la demografía formal, incorporan algunas veces procesos estocásticos.

El refinamiento y la formalización de la demografía matemática y su aplicación exitosa a las poblaciones humanas puede hallarse en trabajos de Coale (1972), Keyfitz (1985), Preston y Campbell (1993), Rogers (1975) y Sheps Y Menken (1973). En estas aplicaciones la demografía matemática presupone el acceso a datos de población y se ocupa de la heterogeneidad a través de la desagregación (dividiendo la población en subpoblaciones).

Los métodos usados en los estudios de población son eclécticos, pidiendo mucho prestado de las disciplinas substantivas de las ciencias sociales. Dado el amplio uso de datos de encuestas y el rol predominante de la inferencia estadística en las ciencias sociales desde 1960, no resulta sorprendente que el método más característico de los estudios de población sea estadístico. Debe notarse que también se utilizan métodos cualitativos en demografía. Las metodologías usadas por demógrafos varían mucho y cambian rápidamente, abarcando ecuaciones estructurales, modelos log-lineales y modelos econométricos.

Una substantiva cantidad de trabajos en estudios de población comprenden análisis estadísticos de muestras (como opuesto a datos poblacionales) en un esquema multivariado, a veces se desarrollan modelos estadísticos para probar hipótesis derivadas de un modelo de comportamiento a nivel individual.

Por esta razón, los estudios de población están estrechamente relacionados a la demografía estadística. Aunque es útil la distinción entre la demografía formal y los estudios de población los límites entre ambas no son fijos ni impermeables. Es más, este límite presenta múltiples temas excitantes para la investigación. Por ejemplo, es posible y también deseable estimar tasas

demográficas de datos muestrales con modelos estadísticos antes de proporcionarlas como *input* para un análisis matemático en la economía formal.

Hay buenas razones para usar herramientas estadísticas en combinación con modelos matemáticos. Primero el avance de la demografía ha traído consigo, datos más ricos y mejores formas de muestreo, el uso de datos muestrales requiere herramientas estadísticas, ya que tratar a los datos muestrales como cantidades conocidas corre el peligro de contaminarse con errores muestrales. En segundo lugar, los modelos estadísticos se adaptan mejor para examinar diferencias entre grupos a través del uso de covariables, el método de desagregación en cambio supone un modelo de interacción completa y puede llevar a una estimación imprecisa debida al tamaño pequeño de los grupos. Tercero, debido a que los datos observados pueden ser irregulares o simplemente estar incompletos, los modelos estadísticos pueden servir para suavizar o imputar datos (Little and Rubin, 1987).

A la inversa, las técnicas en la demografía formal (estimación indirecta, comparaciones de cohortes intercensales, tablas de vida modelo) pueden utilizarse para mejorar el análisis estadístico. Por ejemplo, las relaciones matemáticas entre las componentes demográficas son utilizadas habitualmente para corregir datos incorrectos y/o proveer datos faltantes antes de analizarlos con técnicas estadísticas.

Tabutin (2007) sostiene que hace 30 o 40 años la demografía era una disciplina claramente delimitada, definida en forma precisa en términos de sus temas y metodología. Sus objetos esenciales eran la estructura por edad, sexo y estado civil, fecundidad, mortalidad y migración interna. Sus herramientas básicas eran el análisis demográfico estándar y la estadística elemental; la mayoría de los datos proveían de sistemas de información administrativa (registros civiles, censos, registros ocasionales). Medía fenómenos y tasas de crecimiento, describía niveles y tendencias, proyectaba (lo mejor que podía) y trabajaba principalmente a nivel agregado. Esto puede verse en alguna de las definiciones que se daban por esos

días, como ejemplo la de Bogue (1969): “La demografía es el estudio empírico, estadístico y matemático de las poblaciones humanas”.

Llamada “*wild frontier science*” (que podría traducirse como “ciencia salvaje”) por Alfred Sauvy en 1945 y más recientemente “ciencia rehén” por Michel Loriaux (1996) la demografía se ha vuelto una ciencia en el verdadero sentido de la palabra, con su cuerpo de objetos de investigación, métodos y paradigmas.

El término “Ciencia Demográfica” se usa algunas veces o más ampliamente “Estudios de población”. La demografía no es, o no es más, simplemente un técnica para calcular tasas. Sus objetos de estudio (el comportamiento de las poblaciones humanas del nivel individual al social), ponen a la demografía entre las ciencias sociales, aún cuando la estadística, la biología y la salud pública son componentes integrales de la disciplina, si bien los demógrafos o quienes practican la demografía provienen de variados orígenes académicos.

Si se aborda la habitual y simplista dicotomía entre ciencias “blandas” y “duras”, la demografía es innegablemente una de las más duras de las ciencias sociales, interesada esencialmente en la cuantificación y en la aplicación de estrictos estándares metodológicos para recolección y análisis de datos. Es una ciencia inductiva, algunas veces descrita como formal o empírica.

Según Preston (1993) en parte por su cercanía a la producción de datos, los demógrafos son los más inductivos de las ciencias sociales, enfocada más que otras ciencias sociales en la medición e interpretación cuidadosa de los fenómenos poblacionales.

De un modo similar Caldwell (1996) sostiene que la demografía seguirá siendo una disciplina diferente por su enfoque: ésta demanda que las conclusiones sean compatibles con datos observables y probadas en el mundo real, que los datos sean usados del modo más inteligente posible para obtener resultados concretos y que los estudios sean representativos de poblaciones considerables, significativas y definibles.

El análisis demográfico tradicional sigue siendo una herramienta clave para el demógrafo, funciona bien con sistemas de información estándar y pueden emplearse para medir la dinámica de la población a nivel agregado. Su búsqueda de rigor es legítima, el diagrama de Lexis es aún útil y muchas veces esencial y por 30 años ha sellado la reputación de la escuela francesa de demografía. Pero la demografía ha dejado de ser solamente una técnica de investigación en ausencia de riesgos competitivos, ahora utiliza múltiples herramientas estadísticas más allá de los análisis demográficos simples, usualmente prestados de afuera (modelos log-lineales, modelos multinivel) y adaptados a sus necesidades sin mucha creatividad (en su opinión)

Tabutin (2007) se pregunta si la demografía se está volviendo (una vez más) una ciencia estadística puramente cuantitativa, y sostiene que a pesar de ser considerada largamente una ciencia de la medida, una disciplina puramente descriptiva, que no puede ni debe proveer explicaciones, la demografía actualmente busca en mayor medida comprender el comportamiento individual y colectivo.

Este cambio reciente en la perspectiva es el cambio mayor. El individuo, por ejemplo, no más visto como una trayectoria simple de varios eventos demográficos, cada uno analizado separadamente, sino como un curso de vida complejo comprendiendo varios tipos de eventos interconectados (demográficos, ocupacionales, residenciales) una vida de etapas y períodos con causalidades. De explicar la demografía por la demografía, se ha pasado a un enfoque multifactorial dando lugar a la familia, círculo social y redes sociales o factores sociales, económicos y culturales. Se está aún lejos de una ciencia explicativa pero se han hecho grandes avances.

Según Alho y Spencer (1985), más allá de los asuntos globales, la demografía tiene un rol importante en la toma de decisiones diarias de los gobiernos nacionales y locales. Desde tiempos bíblicos los datos demográficos han servido de base para fijar impuestos, la conscripción militar, las proporciones en la representación política, y ubicación de fondos. Los sesgos sistemáticos en los

datos pueden causar inequidades a través de dominios étnicos y regiones geográficas. Cuando se consideran áreas pequeñas, las variaciones aleatorias pueden causar desigualdades en el tratamiento.

Métodos matemáticos relativamente simples han sido utilizados tradicionalmente para abordar las tendencias demográficas y su rol en la sociedad. Estos métodos han estado basados típicamente en la medida de tasas demográficas por edad y sexo. También se calculan medidas resumen tales como la tasa global de fecundidad o la esperanza de vida.

Una substantiva línea de investigación trata de explicar la variación en las tasas a través de grupos sociales, regiones o tiempo, en términos de conceptos sociológicos o económicos. Otra línea de investigación menos ambiciosa trata de dilucidar las implicancias a largo plazo de las tasas actuales.

Métodos simples han servido, sin dudas, y continuaran sirviendo a la demografía. De todos modos, hay varias razones para expandir la caja de herramientas del demógrafo en la dirección de la estadística. Inicialmente, como se dice más arriba, hay un interés considerable en explorar las variaciones en las tasas demográficas en poblaciones pequeñas. Por ejemplo, si encontramos que jóvenes viudas tienen un riesgo elevado de muerte, pero las cifras son pequeñas, resulta necesario demostrar que no se debe al azar. También se puede pretender evaluar si la duración del desempleo está asociada con la mortalidad, y para ellos pueden emplearse diversos modelos estadísticos o bien calcular un coeficiente de correlación y evaluar luego su significación. Las tablas cruzadas son un modo clásico, pero simple y claro, útil a la hora de estudiar diversas variables y sus relaciones, si bien en epidemiología las tablas cruzadas se han reemplazado ampliamente por técnicas estadísticas de regresión del riesgo relativo.

La estadística en las últimas décadas ha tenido avances muy significativos en nuevas metodologías, tanto en las teorías como la parte computacional, por ello es muy importante que la demografía use dichas herramientas y el avance de las dos disciplinas marche conjuntamente.

3.2. Pronósticos probabilísticos y ciencias sociales

Esta tesis se basa en estudio de los métodos de pronósticos probabilísticos, un concepto que proviene de la estadística, no de la demografía, pero que define a una metodología desarrollada (en sus múltiples versiones) para el pronóstico de tasas específicas de mortalidad por edad y género. En este punto el mismo concepto es definitorio, dado que la fuente principal de conceptos de las ciencias sociales es la redefinición de palabras preexistentes. Según Dogan y Pahre (1993) no siempre se puede trasponer directamente estos conceptos; pero quien se esfuerce en adaptarlos tiene buenas probabilidades de inventar herramientas. Claramente este no es el caso, dado que la herramienta existe, los modelos probabilísticos de pronóstico constituyen un tema en si mismo dentro de la estadística, pero, “embeber” y “ubicar” este concepto dentro de un contexto de una ciencia social resulta innovador y si bien en demografía existe el concepto de proyección, este se haya asociado a cifras de población y no a cifras de mortalidad o fecundidad como sucede en el caso de los pronósticos. Por ello esta tesis introduce un concepto nuevo en el área y pretende evaluarlo desde diversos puntos de vista a fin de conocer sus ventajas y desventajas, sus fortalezas y virtudes, y su utilidad desde la perspectiva de las ciencias sociales.

Sin embargo, introducida la discusión que enmarca la metodología, es posible atacar el problema de investigación cuestionando los hechos supuestamente comprobados y proponiendo ajustes a los modelos para reinterpretar la realidad. Según Mills (1994) los problemas de la ciencia social, adecuadamente formulados, deben comprender inquietudes personales y cuestiones públicas, biografía e historia, y el ámbito de sus intrincadas relaciones, ya que dentro de ellas ocurren las vidas de los individuos y la actividad de las sociedades; y es dentro de este ámbito donde tiene la imaginación sociológica su oportunidad para diferenciar la calidad de la vida humana en estos tiempos. Dicho esto, es de vital interés enmarcar las nuevas metodologías, cuestionarlas en su aspecto de generadoras de cifras al de servicio a la sociedad en su conjunto y no reducirlas a una mera técnica que se expande en miles de variantes sin aportar respuestas a un mundo que envejece y las reclama de modo urgente.

Finalmente, como enuncian Strauss y Corbin (2002) cualquier tecnología, ya sea de carácter cualitativo o cuantitativo es un medio para construir teorías útiles. No hay una supremacía de uno modo de hacer investigación sobre otro y un instrumento es un instrumento, un modelo estadístico es un modelo estadístico en sí mismo y el asunto principal es cuando es útil para teorizar. De modo que las metodologías se tornan “útiles” cuando son capaces de enmarcarse en un paradigma y ofrecen la posibilidad de un desarrollo teórico.

Luego de exponer los conceptos teóricos y las complejas relaciones que los vinculan se describen a continuación las hipótesis y objetivos de trabajo.

3.3. Hipótesis y Objetivos

El principal aspecto que aborda este trabajo es el empleo de las proyecciones probabilísticas para estimar las tasas específicas de mortalidad en algunos Departamentos de la Provincia de Córdoba: Capital, Río Cuarto y Sobremonte. Esto constituye un aporte importante a la Demografía debido a que se trata de una novedosa metodología que no se ha aplicado ampliamente en nuestro país y cuyos resultados permitirían obtener insumos de calidad para realizar otros análisis. Sin embargo, resulta relevante estudiar las limitaciones y el cumplimiento de supuestos en los que estas técnicas se basan.

De este modo, el objetivo general es evaluar el desempeño de las proyecciones probabilísticas de la mortalidad empleando dos grandes exponentes de las metodologías probabilísticas: el modelo de Lee y Carter (el cual es pionero en el área) y el Modelo para Datos Funcionales, que constituye uno de los últimos desarrollos realizados en este ámbito. Debido a que no es una metodología habitual, es importante considerar tanto sus ventajas como sus desventajas desde el punto de vista demográfico. Así, para ello, se llevará a cabo la aplicación de ambos modelos en base a datos correspondientes a un subconjunto de departamentos de la provincia de Córdoba, utilizando información correspondiente al período 1980 – 2013.

Asimismo, también se plantean dos objetivos secundarios

- 1) Evaluar el desempeño de metodologías de pronóstico probabilístico en mortalidad en regiones geográficas pequeñas; como son los departamentos de la provincia de Córdoba. Dado que la metodología presenta excelentes resultados en diversos países, se pretende estresar los modelos bajo una condición en la que no han sido empleados.
- 2) Analizar aspectos puntuales tales como: la intervención del demógrafo en la metodología, la calidad de la fuente de datos y su relación con los modelos bajo estudio, y aspectos intrínsecos tales como suavizados y la cuantificación de la incertidumbre asociada a los resultados.

4. Métodos

4.1. Modelos Estadísticos para el pronóstico de las tasas de mortalidad

4.1.1. Modelo de Lee-Carter

Como se ha mencionado, en este trabajo se aplica el modelo de Lee y Carter (1992) que es el precursor de los modelos probabilísticos de pronóstico en mortalidad y –en segundo término– uno de los últimos modelos propuestos en el área: el modelo para datos funcionales desarrollado por Hyndman y Ullah (2008).

En relación al modelo de Lee y Carter, la premisa básica que plantea 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 .

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.

La estimación se realiza utilizando Descomposición en Valores Singulares (DVS)⁴, 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. Para garantizar la unicidad de la solución a la suma de los valores de b_x se le impone el valor 1, y a la suma de los valores k_t se les impone sumar cero.

4.1.2. Modelo para Datos Funcionales

En relación al Modelo de Datos Funcionales (o MDF) se supone el siguiente modelo para las observaciones transformadas (dado que el modelo postulado permite emplear múltiples transformaciones, se habla, en general, de observaciones transformadas, pero en este caso se trata de logaritmos) $y_t(x)$:

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

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

⁴ Descomposición en valores singulares: Tanto los valores singulares como la descomposición en valores singulares de una matriz son conceptos de gran utilidad en las aplicaciones del álgebra lineal a diversos problemas prácticos y teóricos. Es la descomposición de una matriz en tres factores. Dos de ellos son matrices ortogonales, lo cual le confiere un interés especial. La razón es que las transformaciones ortogonales preservan normas y ángulos; en particular, preservan las longitudes de los vectores de error que son inevitables en los cálculos numéricos.

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 observemos en la realidad 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) .

El primer paso para la aplicación de los modelos estadísticos (Lee-Carter y Modelo para Datos Funcionales), consiste en la aplicación de un suavizado a las tasas observadas, a través de las edades. Si bien en principio, Lee-Carter no requiere este suavizado (no es parte de la formulación teórica del modelo), esto

permite obtener una función suave a través de las edades, eliminando las variaciones que en principio se considerarían estocásticas, es decir, que no responden al comportamiento de la mortalidad sino a desviaciones producto del azar.

Asimismo es importante destacar que el proceso de suavizado se puede aplicar a tasas informadas por grupos etarios, llevándolas de edades agrupadas a edades simples. Las edades presentadas en grupos presentan sus ventajas y desventajas: por un lado, el hecho de que estén agregadas disminuye la variabilidad y genera menos tasas nulas, favoreciendo la estimación en regiones pequeñas. Sin embargo si se trabaja en regiones de tamaño poblacional considerable, trabajar con edades agregadas implica una pérdida de información. En el glosario puede encontrarse una definición extensa del concepto de suavizado.

4.2. Otros aspectos metodológicos

También se consideran otros aspectos importantes, como ser las posibilidades de intervención del demógrafo en la metodología y las fuentes de datos y sus dificultades (cantidad de años necesarios, edades simples o grupos, omisión etc). Por otro lado a partir de la aplicación a tres departamentos de Córdoba se analizarán aspectos tales como los residuos, los insumos que el modelo brinda como resultado (tasas, intervalos, tablas de vida), la cuantificación de la incertidumbre como el mejor rasgo de la metodología y el tema de la magnitud de las regiones.

4.2.1. Formas de intervención del demógrafo

En esta sección se pretende evaluar los diferentes modos en los que el demógrafo puede intervenir y modificar los diversos aspectos del modelo.

En primer lugar, en la etapa de suavizado se especifica un número a partir del cual las tasas suavizadas van a crecer monótonamente, es decir, se puede fijar una edad a partir de la cual se considera que la mortalidad ya no vuelve a decrecer. Este procedimiento permite en cierto modo adaptar la metodología a

una población particular, de acuerdo a su edad jubilatoria o sus patrones específicos.

Si bien la metodología permite elegir entre distintas técnicas de suavizado, estas no constituyen más que diferentes propuestas matemáticas para reducir la variabilidad en los datos, por otro lado, de acuerdo a la componente que se modele (en este caso la mortalidad) los autores ya han establecido una técnica de suavizado que mejor se adapta a cada componente.

Ambas metodologías permiten constituir el modelado en combinación con dos técnicas de pronósticos de series de tiempo; modelos ARIMA y modelos de innovación de espacio de estados, sin embargo dado el corto período de tiempo (en comparación con años de información disponibles para otros países) son recomendables los primeros modelos, y que los últimos insumen muchos datos para la estimación de valores iniciales. Además la implementación en el paquete *Demography* de R, permite que al modelo de series de tiempo que se selecciona para los coeficientes que dependen del tiempo (k , para Lee-Carter y β para el Modelo para Datos Funcionales) se le imponga la estimación de una pendiente o no, pero si la selección es automática el modelo se selecciona mediante el criterio de Akaike⁵ entre todos los modelos posibles, con y sin pendiente.

De todo lo aquí mencionado se desprende que el demógrafo que emplea estas técnicas está limitado, en su accionar, a un par de opciones técnicas. No es posible introducir hipótesis de tendencias, ni valores posibles para años determinados, así como tampoco es posible seleccionar la información de entrada de ambos modelos.

4.3. Fuentes de Datos, Depuración y Procesamiento a través del software R

El modelo para datos funcionales y el modelo de Lee y Carter se aplican a tasas específicas de mortalidad, para ello los datos requeridos son las defunciones y población por edad simple y género. Es importante destacar que durante la estimación es necesario contar con las cantidades netas tanto de eventos como

⁵ El criterio de información de Akaike (AIC de "Akaike's Information Criterion") sirve para comparar la plausibilidad relativa de un conjunto de modelos.

de población y por ello no es la tasa en sí el dato base, sino que el mismo se construye a partir de los registros de defunciones y las cifras de población por edad.

En relación a los datos provenientes del registro de hechos vitales, para el caso que nos ocupa, los mismos son suministrados por la Dirección de Estadísticas e Información de Salud (DEIS), a través de bases de datos que contienen el registro de los hechos individuales que permiten generar la matriz por edades para un período de tiempo de aproximadamente 30 años. En cambio en relación a las cifras de población se genera una dificultad, no existe una única población oficial por edades simples para los últimos treinta años. Por otro lado se interpolan las cifras de población para los períodos intercensales, dado que no existe una estimación oficial de la población en base a los cuatro últimos Censos, menos aún por edades simples y por departamento.

En cuanto a las defunciones se construye una matriz, en la que las filas se corresponden a las edades simples mientras que las columnas a los códigos de los departamentos, construyendo una matriz por cada año de 1980 a 2013. Sin embargo, es posible que la información suministrada oficialmente requiera ser depurada para servir de entrada en la programación de los modelos en lenguaje R, como por ejemplo cuando se presentan defunciones con lugar y edad desconocida y se procede, como en este caso, a la distribución proporcional de las mismas.

Finalmente, a partir de las 34 matrices se diseñan las correspondientes a cada uno de los tres departamentos (ya que el modelo se calcula a dicho nivel de desagregación) quedando definidas inicialmente matrices de 99 (edades simples) por 34 años. Se dice "inicialmente" dado que la presencia de ceros puede determinar que haya que colapsar las edades simples en algún tipo de agrupación etaria.

Por otro lado se cuenta con las cifras de población provenientes de los censos, y como el modelo requiere tanto muertes como población a nivel edad simple y anualmente se procede a realizar una programación por bucle en R a fin

de interpolar sin la intervención manual del usuario sobre las poblaciones que se requieren. Para la interpolación se hace un supuesto de crecimiento exponencial entre censos (para mayor detalle ver Glosario)

Años faltantes y depuración de códigos se tienen en cuenta a la hora de interpolar defunciones mediante una técnica de suavizado de *Splines*, que permite subsanar la falta de datos. Yasmeen y Mahmood en un artículo del 2012 emplean un *spline* cúbico que permite estimar valores faltantes en una aplicación de datos funcionales. La misma metodología se emplea en R para completar las series (defunciones) faltantes de los departamentos de Córdoba.

Se trabaja en primer lugar con los departamentos más poblados, Capital y Río Cuarto, finalmente se obtienen resultados para el Departamento Sobremonte, uno de los menos poblados de acuerdo al Censo 2010. Esta decisión se basa en conocer el grado de precisión que tiene la técnica en un departamento con gran población (Capital), uno intermedio (Río Cuarto) y uno pequeño (Sobremonte). Esto último representa un aporte importante en la materia ya que la técnica no es habitualmente empleada en este tipo de situaciones, siendo éste un primer indicio del desempeño de la metodología en áreas pequeñas.

A continuación se detalla paso a paso la etapa de modelización y de pronóstico:

- 1) En base a las matrices de defunciones y de población por edad y sexo se calculan los logaritmos de las tasas específicas de mortalidad para cada región (en este caso departamento) y se representan gráficamente.
- 2) Luego se procede a suavizar las mismas y también se presentan en forma grafica.
- 3) A continuación se realiza la estimación de los modelos:
 - a. A través de la metodología de Componentes principales Funcionales se descompone la variabilidad inherente a las tasas suavizadas en bases y coeficientes.
 - b. A través del modelo de Lee-Carter, mediante la descomposición en valores singulares, se obtiene una base que resume el

comportamiento a través de las edades y un índice general de mortalidad que refleja la trayectoria temporal de la mortalidad en el período en estudio.

- 4) Se grafican los residuos de los modelos en gráficas de contorno para evaluar si ha quedado algún patrón sin explicar.
- 5) A partir de cada modelo se pueden obtener pronósticos de las tasas específicas de mortalidad, manteniendo estable el patrón por edad (bases) y empleando modelos de series temporales sobre los coeficientes (en el caso del Modelo para Datos Funcionales) y el índice general de mortalidad (en el caso del modelo de Lee-Carter).
- 6) Asimismo es posible derivar la esperanza de vida, con sus intervalos de pronóstico para cada conjunto de tasas específicas de mortalidad pronosticadas.

5. Resultados

5.1. Estimación de los modelos

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. Inicialmente se tabulan las tasas de mortalidad por grandes grupos de edad de los tres departamentos seleccionados para cuatro años seleccionados (Tabla 1.) estos presentan diferentes perfiles de mortalidad y principalmente de variabilidad, que 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.

Tabla 1. Tasas específicas de mortalidad por mil habitantes según grandes grupos de edad para departamentos seleccionados, 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 (Gráficos 1, 8 y 15) 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 la Organización Mundial de la Salud (OMS) la adolescencia es el período de la vida en el cual el individuo adquiere su madurez

reproductiva, transita los patrones psicológicos de la niñez a la adultez y establece su independencia socio-económica comprendiendo varones y mujeres cuyas edades están entre los 10 y 24 años. Serfaty et al. (2007) sostiene que en Argentina este grupo asciende a 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.

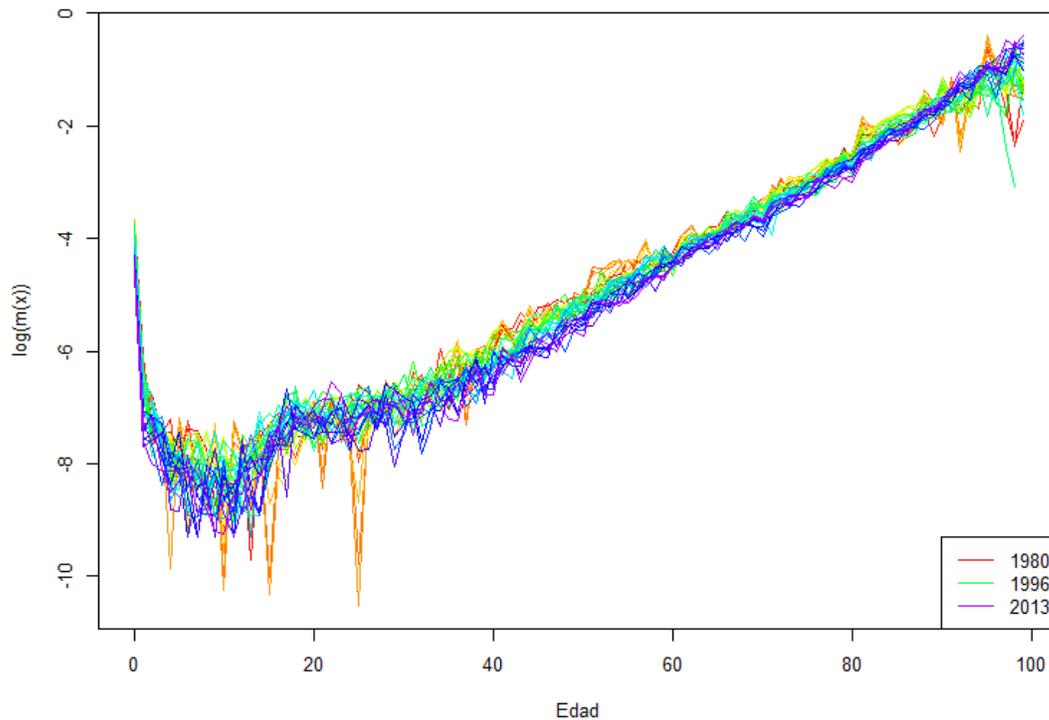
Un aspecto que se destaca visualmente en los gráficos mencionados, es la caída marcada en los niveles de mortalidad a través del tiempo, que se da en todas las edades, excepto para el pico de los 20 años, franja en la que parece existir una caída más leve o fluctuante. La caída en general, en los niveles 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 el presente período de tiempo 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.

5.1.1. Departamento Capital

Inicialmente se presentan los logaritmos de los datos observados. Para la representación de las curvas se emplea la escala "arco iris", donde cada línea constituye un año calendario del período en estudio, la escala comienza con colores cálidos como el rojo, luego anaranjado, amarillo para luego pasar por el verde y terminar con azul y magenta. La idea de emplear esta escala proviene del hecho que al graficar múltiples años no es posible individualizar cada uno, y en su lugar se prefiere emplear tonalidades que se asocien al paso del tiempo. Los picos rojos, anaranjados y amarillos revelan que la mayor variabilidad la presentan los primeros años incluidos en el análisis y a medida que transcurren los

años estos se ven más concentrados, como sucede al llegar a la tonalidad azul de los últimos años.

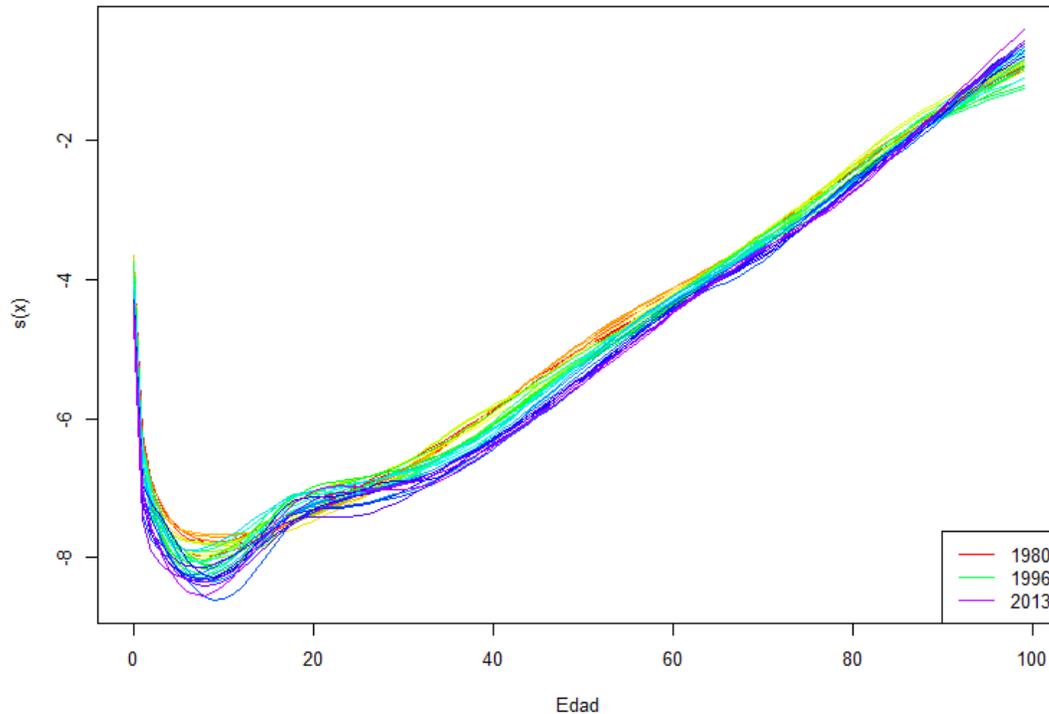
**Gráfico 1. Logaritmo de las tasas específicas de mortalidad observadas.
Departamento Capital (1980-2013)**



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

Luego, en un segundo paso se realiza un suavizado de los mismos, eliminando variaciones espurias a fin de obtener un patrón más nítido. El conjunto de curvas suavizadas permiten percibir mejor el patrón general en las tasas específicas y constituyen el insumo base del modelo.

Gráfico 2. Suavizado del logaritmo de las tasas específicas de mortalidad observadas. Departamento Capital (1980-2013)



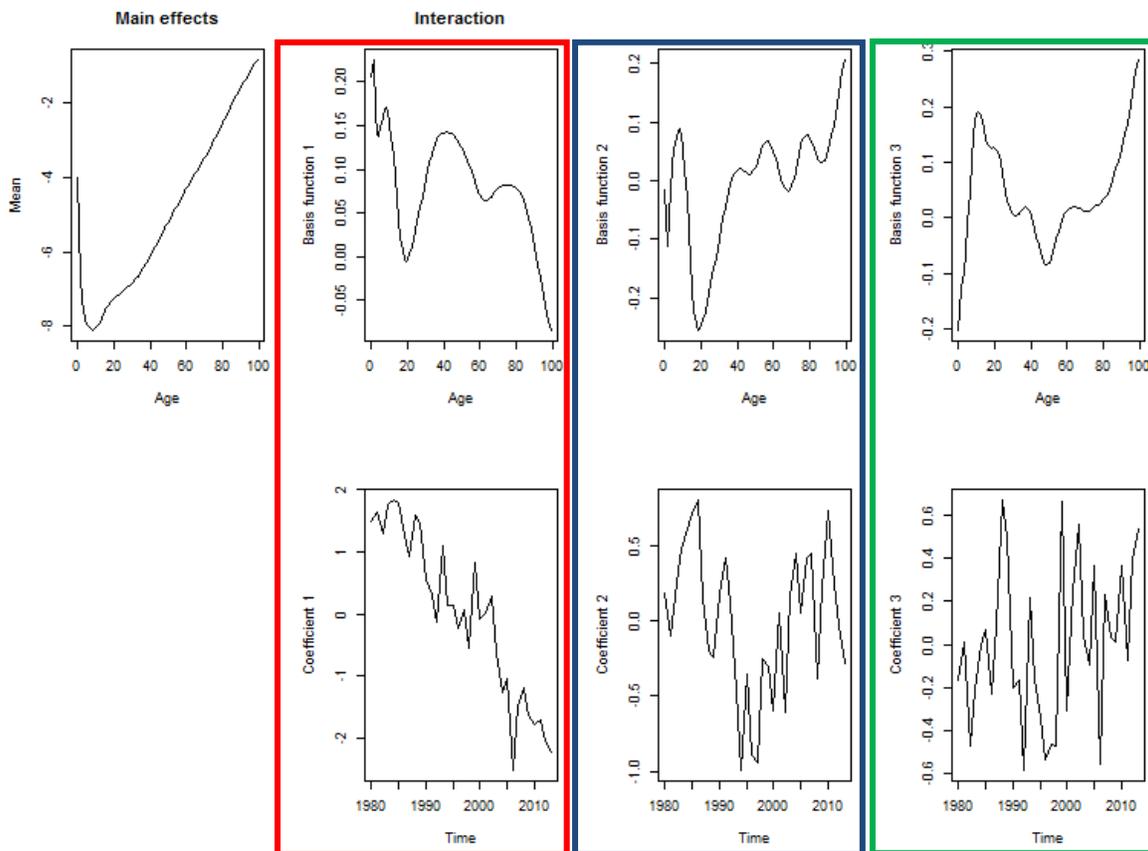
Fuente: Elaboración propia en base a datos del INDEC y DEIS

Como ya se mencionó en base a las tasas suavizadas se estima el Modelo para Datos Funcionales. El modelo estima una función media y un conjunto (pares) de bases y coeficientes, las mismas son representadas en forma gráfica en el gráfico 3. Estas figuras 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. El gráfico 3⁶ contiene la estimación de la media (mean) y tres pares de bases y coeficientes (se estimaron

⁶ Los gráficos 3.1, 3.2 y 3.3 son sólo réplicas del gráfico 3 pero presentados de manera individual para facilitar su lectura e interpretación.

seis, pero luego de la tercer base resulta difícil evaluar algún tipo de correlato con un comportamiento demográfico de las tasas).

Gráfico 3. Modelo para Datos Funcionales. Departamento Capital

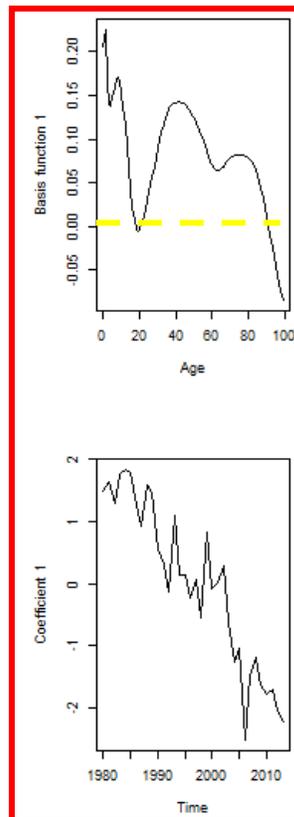


Fuente: Elaboración propia en base a datos del INDEC y DEIS

El primer par base-coeficiente explica el 73% de la variabilidad (gráfico 3.1). Este par puede interpretarse, en primer lugar a través del coeficiente que expresa una tendencia general decreciente, con fluctuaciones moderadas (recuadro superior), y luego su base asociada nos indica que grupos etarios contribuyen en mayor o menor medida a dicho comportamiento (recuadro inferior). Así, la mortalidad en las edades iniciales es la que más contribuye al descenso de la mortalidad general, luego el grupo etario en torno a los 40 años y en menor medida el grupo de 70 a 80 años. La misma base indica que las edades que menos contribuyen son las edades avanzadas y el grupo en torno a los 20 años. Si bien se trabaja con el total de la población, sería interesante discriminar este

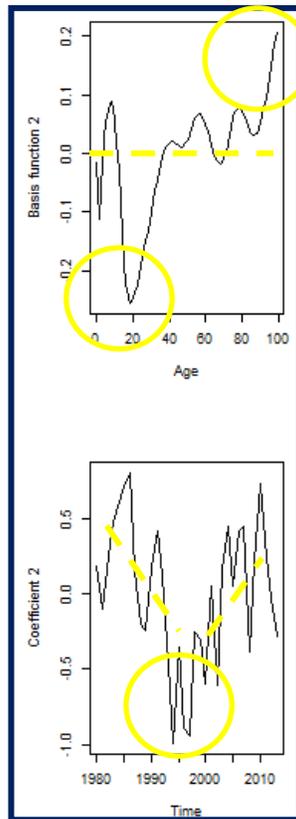
comportamiento por género, ya que este “valle” alrededor de los 20 años se corresponde a varones más que a mujeres. Pero, si se tiene en cuenta que el área geográfica de análisis es pequeña, la subdivisión por género arrojaría tasas más pequeñas y/o nulas y el método puede comenzar a presentar problemas en la estimación.

**Gráfico 3.1 Modelo para Datos Funcionales
Departamento Capital. Primer par base-coeficiente**



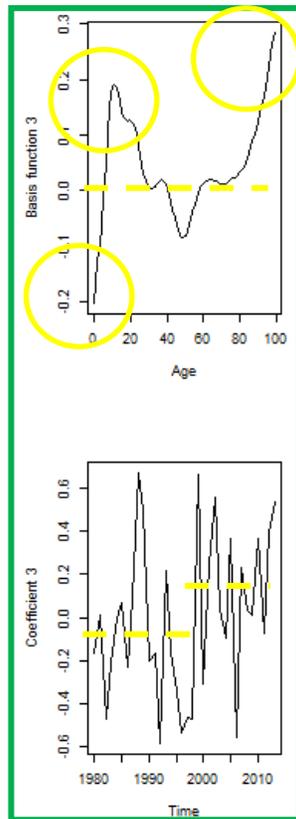
El segundo par base-coeficiente explica el 9.8% de la variabilidad presente en los datos, es importante destacar que esto significa que de la variabilidad inherente a las tasas específicas de mortalidad para este período, casi un 10% se podría explicar mediante el comportamiento etario-temporal que describe este par.

**Gráfico 3.2 Modelo para Datos Funcionales.
Departamento Capital. Segundo par base-coeficiente**



En este caso el coeficiente asociado comienza con un descenso general y finaliza con una tendencia creciente, mientras que la base asociada presenta dos edades alejadas del cero: un “valle” alrededor de los 20 años y un “pico” en las edades avanzadas, por lo que puede pensarse este par en relación a un diferencial entre ambas edades y su comportamiento a través del tiempo. Y si se analiza en relación al coeficiente puede pensarse que la mortalidad de jóvenes y adultos mayores se “diferencia” más entre los años 1990 y 2000 (período para el cual los valores del coeficiente se hallan más alejados del cero).

**Gráfico 3.3 Modelo para Datos Funcionales
Departamento Capital. Tercer par base-coeficiente**

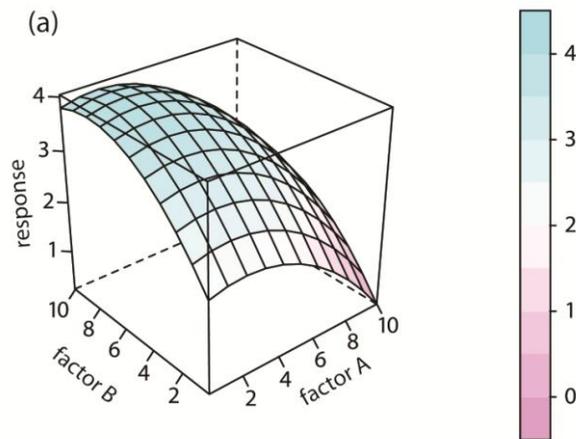


Luego, a partir del tercer par resulta difícil describir el comportamiento ya que el coeficiente presenta una alta fluctuación en torno a dos niveles distintos, siendo más bajo en la primera mitad del período y más alto en la segunda parte del mismo. Los grupos de edad que pueden asociarse a este comportamiento son el grupo en torno a los cinco años y las edades avanzadas, no así las edades iniciales. Sin embargo no resulta un patrón lógico desde el punto de vista demográfico, pero dado que dicho par explica sólo alrededor del 5% de la variabilidad puede pensarse que dicha variación, o comportamiento es fruto del azar y no de una tendencia demográfica definida. En total el modelo explica en las seis bases el 96% de la variabilidad presente en los datos: un 4.0% en la cuarta base, un 1.8% en la quinta y finalmente un 1.8% en la sexta base.

Por otra parte, con la finalidad de observar el comportamiento de los residuos se construye un gráfico de contorno. En un gráfico para residuos

independientes se espera observar zonas pequeñas y que alternan colores opuestos (celeste y rosa). La idea que subyace a la realización de estos gráficos es que si el modelo logra captar de manera adecuada los patrones relacionados, tanto a las edades como al tiempo, es esperable observar un gráfico completamente aleatorio, es decir alternancia de manchas blancas, celestes y rosas que no determinen ningún tipo de patrón. La presencia de un patrón, franja o punto puede deberse a una relación que no se ha captado o a la presencia de un posible *outlier*, entre otras causas.

Figura 1. Esquemmatización del proceso de obtención de una escala de colores a partir de una superficie de nivel.



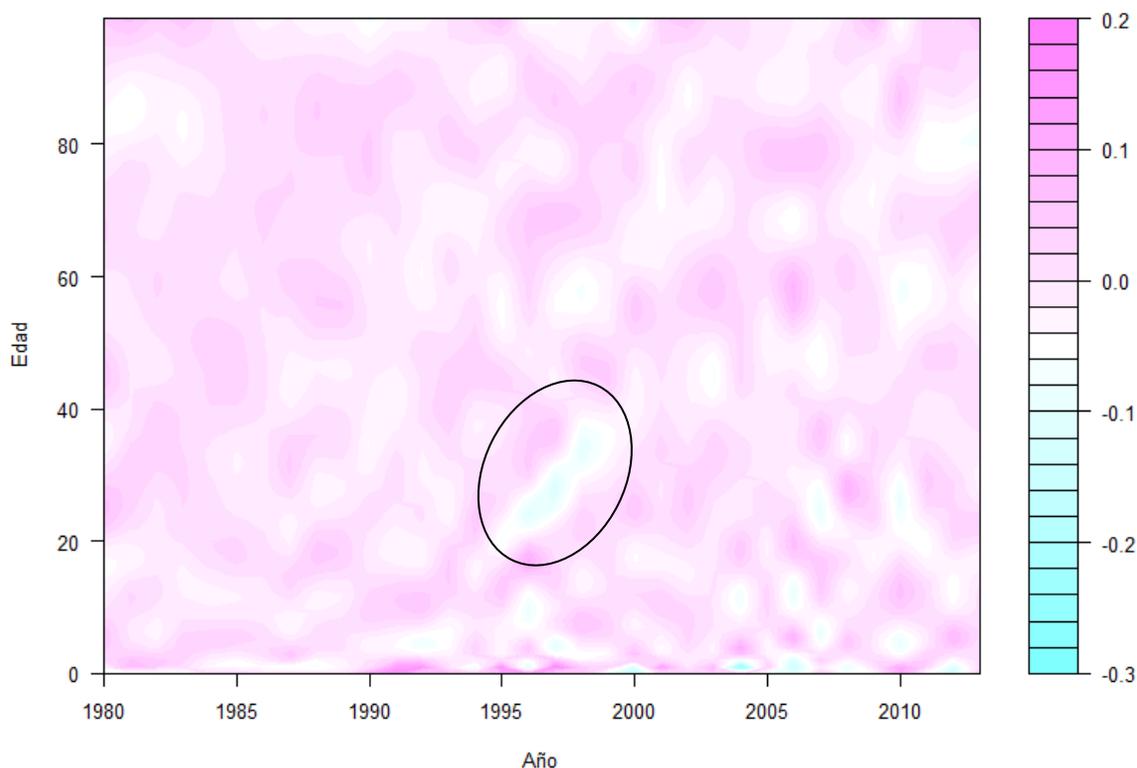
El gráfico de contornos es la simplificación de una superficie, si hay relación entre las variables de los ejes (ver figura 1). De este modo, si el factor A y B no están relacionados se espera ver residuos cercanos a cero, en cambio si la superficie tiene algún patrón particular -como sucede en la figura- se esperará ver áreas marcadas y diferenciadas de distintos colores pero no ubicadas de forma aleatoria. A valores bajos de A y valores altos de B se corresponden valores altos de la respuesta, en el caso de este trabajo los residuos.

En el gráfico realizado para los residuos observados del modelo para datos funcionales se presentan grupos o bandas que podrían indicar falta de independencia, una banda diagonal entre 1995 y 2000 y las edades entre 20 y 40 años. Además el gráfico presenta en su mayoría un tono rosado indicando que

los residuos son positivos, siendo un indicador de subestimación, es decir las tasas estimadas resultan menores a las observadas. Sin embargo se percibe una franja alrededor del año 1997 entre los 20 y los 40 años, que a diferencia del resto de los residuos presenta tonalidad celeste, por lo que sería interesante evaluar si algún hecho histórico modificó en esos años y para esas edades las tasas produciendo sobreestimaciones. (Gráfico 4.)

Por otro lado, como medida resumen comparativa se emplea el Error Medio Absoluto Porcentual (conocido por sus siglas en inglés MAPE) promediado a través de las edades y este resulta 0,007%, siendo un valor bajo, ya que suele pensarse en valores menores a 5% como límite superior aceptable. Este resultado implica que el modelo presenta una bondad de ajuste alta, es decir que capta el comportamiento de las tasas observadas con un bajo porcentaje de error. (Ver salidas en Anexo)

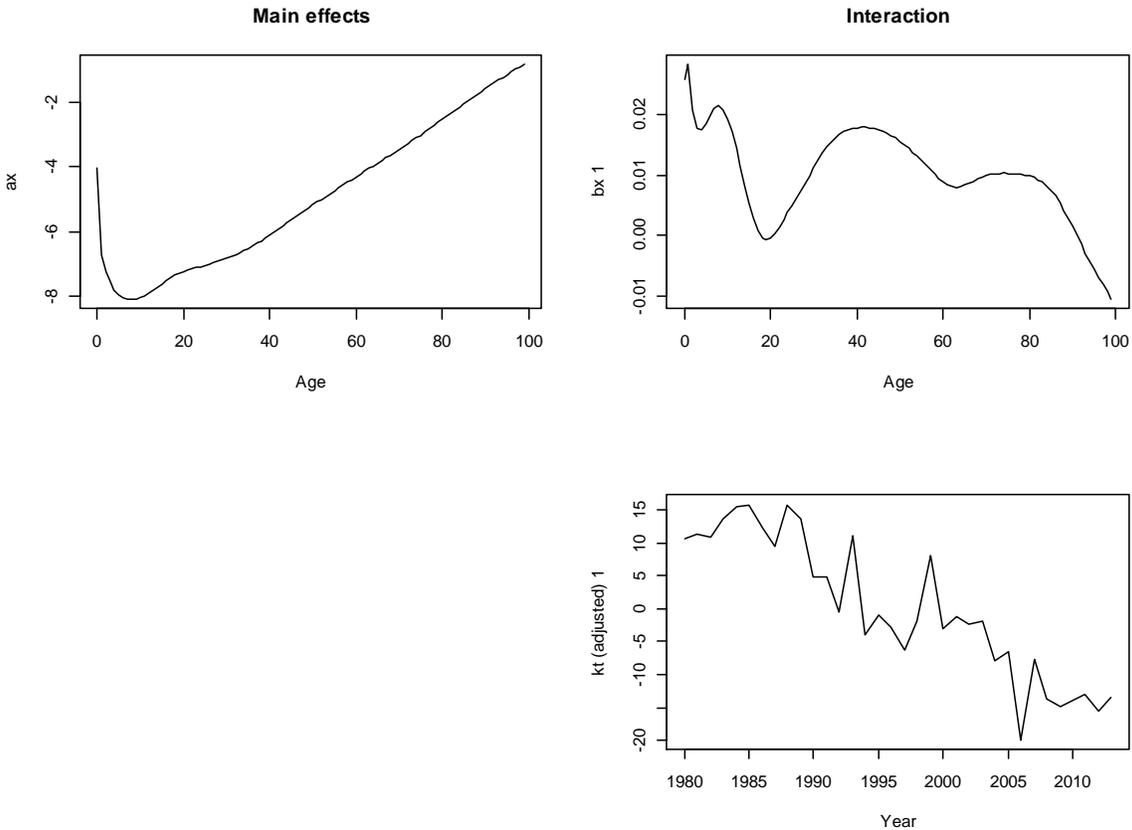
Gráfico 4. Residuos del modelo MDF. Departamento Capital.



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

En segundo lugar y para el mismo departamento se estima el Modelo de Lee-Carter. El mismo comprende un parámetro a que cumple el mismo rol de la media en el modelo funcional y un parámetro b , llamado de sensibilidad cuyo comportamiento es prácticamente el mismo que presenta la primera base del modelo antes mencionado, ver gráfico 5. Además siendo este un modelo sencillo explica el 76.5% de la variabilidad presente en la mortalidad. Por otro lado el MAPE asciende al 0.018%, siendo, del mismo modo que en el caso del modelo funcional, un valor bajo.

Gráfico 5. Modelo de Lee-Carter. Departamento Capital

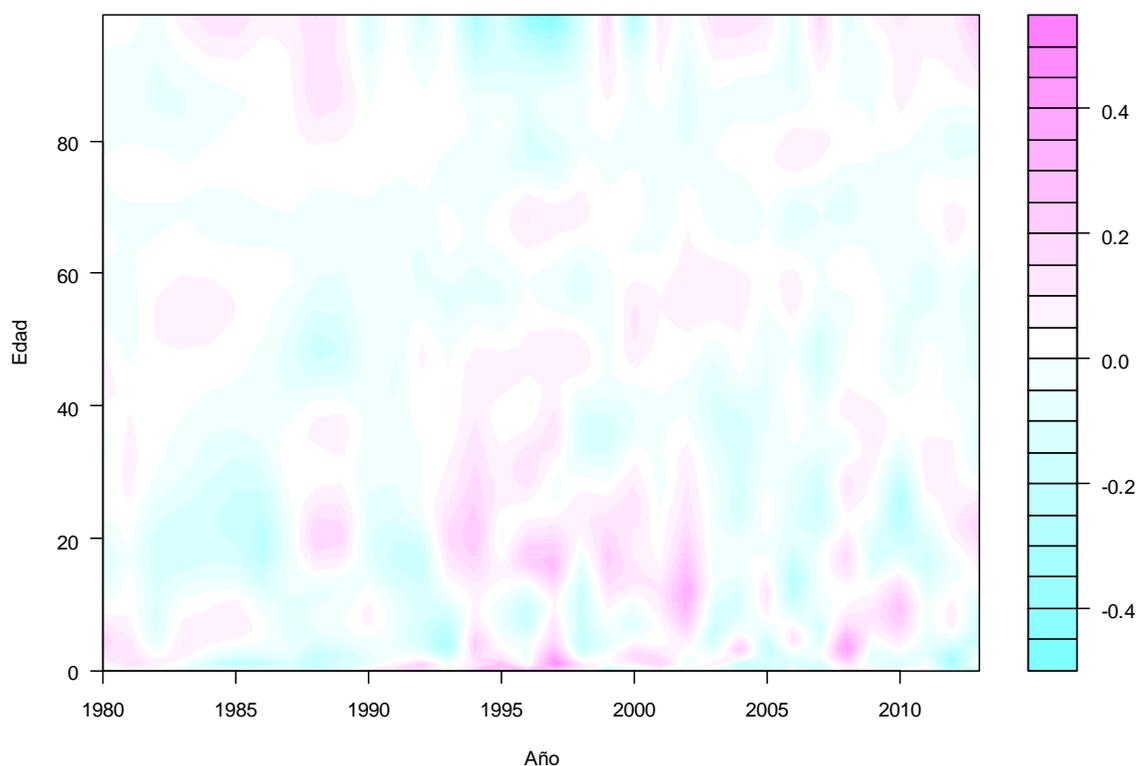


Fuente: Elaboración propia en base a datos del INDEC y DEIS

En el Gráfico 6 se presentan los residuos del Modelo de Lee-Carter, en el mismo, a diferencia del Modelo para Datos Funcionales, se observa una alternancia de manchas rosadas y celestes, que indica un patrón aleatorio; esto es esperab

le bajo una estimación que se ajusta a los supuestos. Esto significa que la edad y el tiempo, (en años calendario son factores que explican el comportamiento de las tasas específicas de mortalidad y que no quedan “patrones” que puedan indicar un comportamiento que no haya sido captado por el modelo.

Gráfico 6. Residuos del modelo LC. Departamento Capital

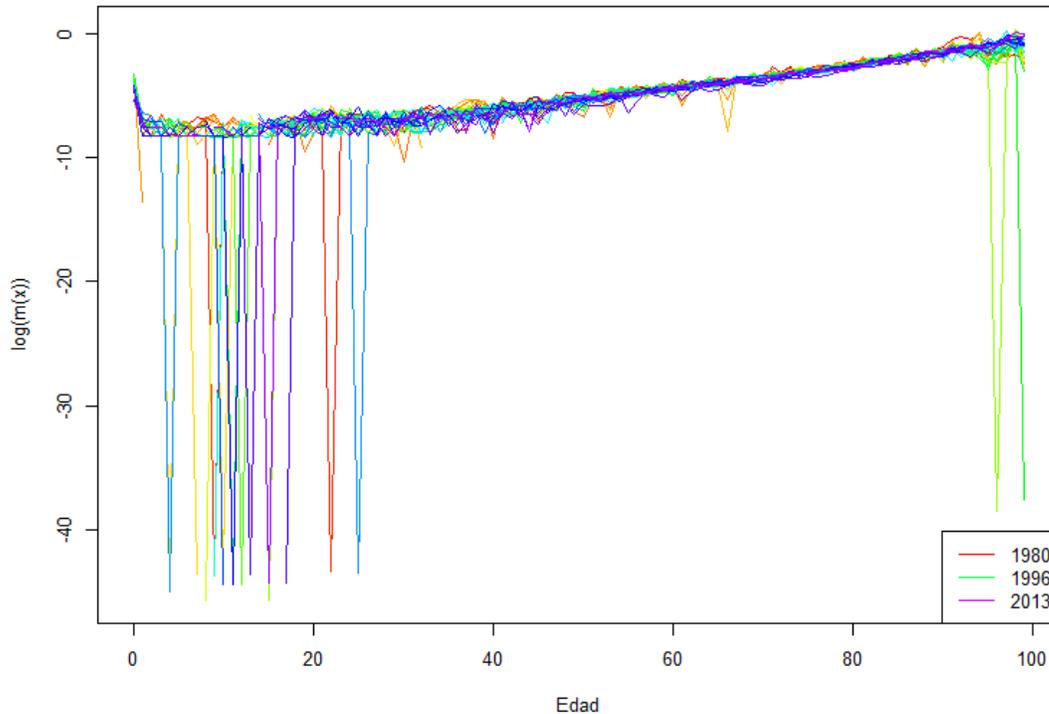


Fuente: Elaboración propia en base a datos del INDEC y DEIS

5.1.2. Departamento Río Cuarto

Se realizan las mismas estimaciones para el segundo departamento más poblado de la provincia, Río Cuarto. A diferencia del departamento Capital, Río Cuarto presenta una moderada presencia de tasas bajas y nulas, sin embargo el proceso de suavizado las corrige posibilitando la estimación del modelo.

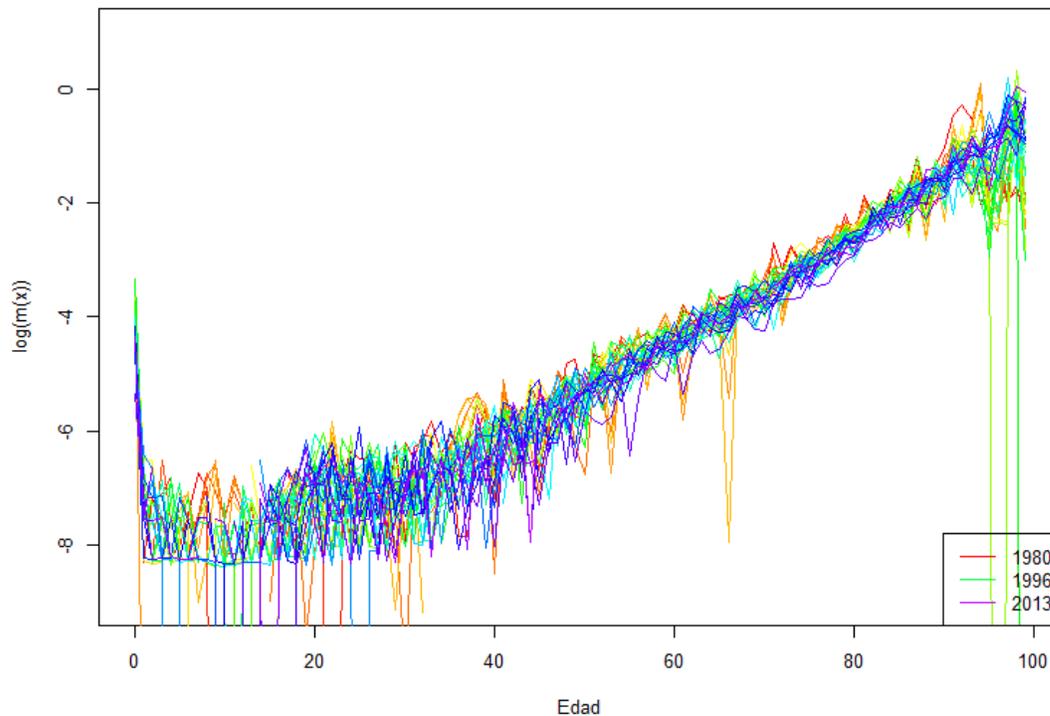
**Gráfico 7. Logaritmo de las tasas específicas de mortalidad observadas.
Departamento Río Cuarto (1980-2013)**



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

El gráfico 7 presenta el logaritmo de las tasas observadas, allí no puede evaluarse el patrón de las curvas dado que existen valores extremadamente bajos que hacen que el eje de las ordenadas se extienda demasiado, por ello se presenta un gráfico 7.b. que es un recorte que excluye los valores extremos a fin de permitir la visualización clara de las curvas bajo estudio. Se observa una variabilidad más marcada y sostenida a través de las edades, además de un pico menor en torno a los 18, 20 años.

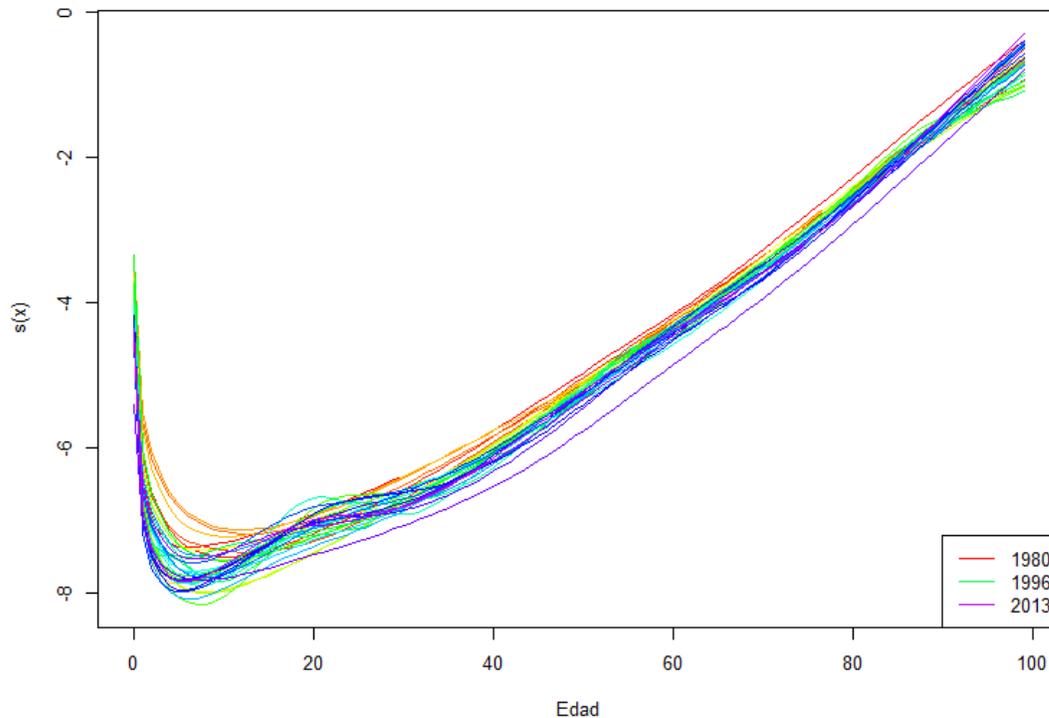
Gráfico 7.b. Logaritmo de las tasas de mortalidad observadas. Departamento Río Cuarto (1980-2013) (recorte)



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

A partir de los datos suavizados, el modelo para datos funcionales estimado para el departamento Río Cuarto explica en total un 96,6% de la variabilidad presente en los datos; de la primera a la sexta base explican: 54.5%, 18.3%, 12.9%, 4.9%, 3.5% y 2.5% respectivamente. En la media puede verse que la elevación alrededor de los 20 años es menor que para el departamento Capital, esto podría deberse a que se supone que cuanto menos "cosmopolita" es la región bajo estudio esta presenta menores índices de suicidio y criminalidad.

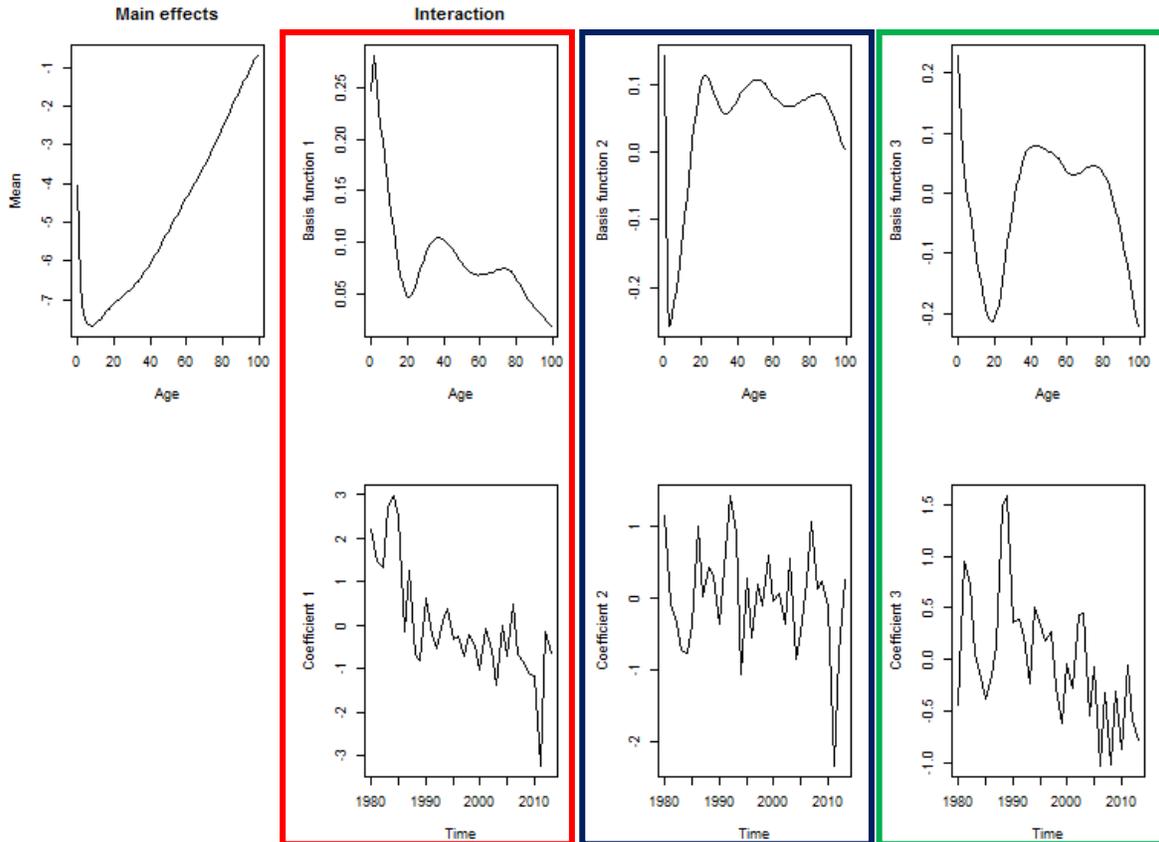
**Gráfico 8. Suavizado del logaritmo de las tasas de mortalidad observadas.
Departamento Río Cuarto (1980-2013)**



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

El primer par base coeficiente explica un porcentaje sustancialmente menor que lo explicado para el mismo par por el modelo funcional para el departamento Capital. El coeficiente presenta un amesetamiento o un leve descenso con la presencia de dos valores, marcadamente alto y bajo hacia el inicio y el final del periodo, siendo probablemente valores atípicos y que merecen un análisis más profundo.

Gráfico 9. Modelo para Datos Funcionales. Río Cuarto.

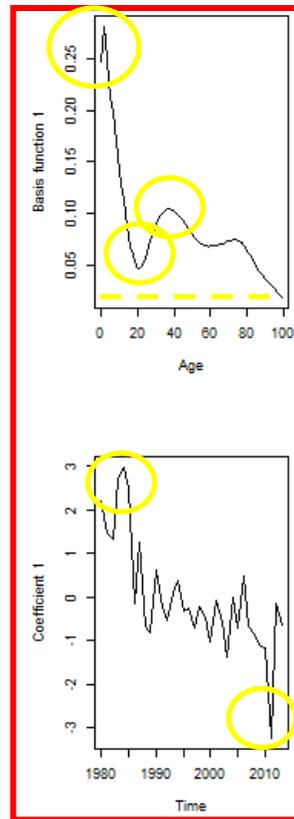


Fuente: Elaboración propia en base a datos del INDEC y DEIS

La base asociada presenta valores altos para edades cercanas a los cinco años y levemente altos alrededor de los 30, mientras que el valle alrededor de los 20 resulta cercano a cero, pero no es nulo como sucede para el departamento Capital. Esto estaría indicando que dicho grupo etario, aunque en mucha menor proporción que las edades cercanas a los cinco y a los treinta, aporta al descenso general de la mortalidad, o en otras palabras; el grupo de edades cercanas a los 18, a pesar de presentar una mortalidad alta, describe un leve descenso a través del tiempo.

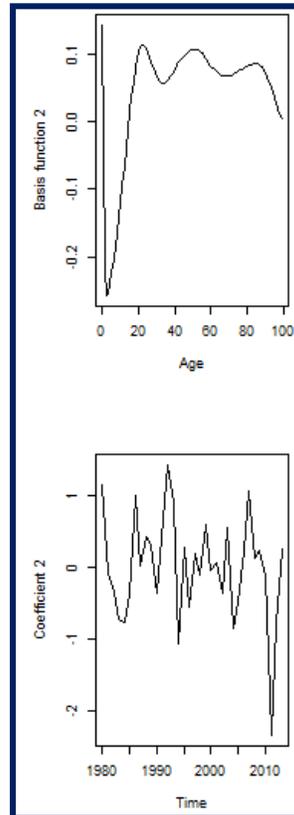
Gráfico 9.1. Modelo para Datos Funcionales. Río Cuarto.

Primer par base-coeficiente



El segundo par presenta un coeficiente que oscila en torno a un nivel medio y que presenta un posible valor atípico hacia el final del período en estudio. La base asociada parece indicar que se trata de un diferencial de las edades cercanas a los 5 años versus las demás.

**Gráfico 9.2. Modelo para Datos Funcionales. Río Cuarto.
Segundo par base-coeficiente**



Finalmente el tercer par presenta una base fluctuante con una tendencia histórica descendente (representada por el coeficiente). Nuevamente resulta difícil identificar algún tipo de patrón demográfico. El modelo presenta un MAPE del 0,007%, es decir un valor bajo. Asimismo en el gráfico 10 se presentan los residuos que indican sobreestimación; dado que los mismos toman valores negativos (representados en celeste como indica la escala).

**Gráfico 9.3. Modelo para Datos Funcionales. Río Cuarto.
Tercer par base-coeficiente**

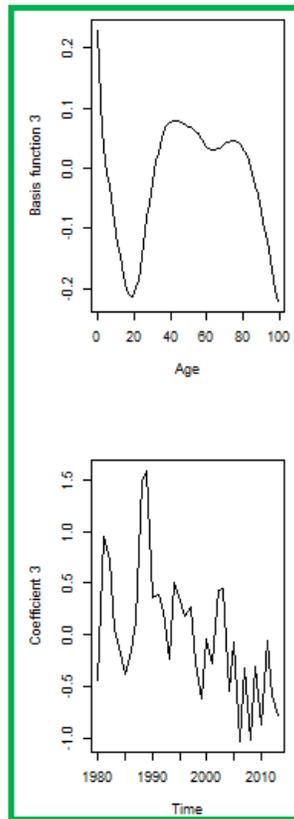
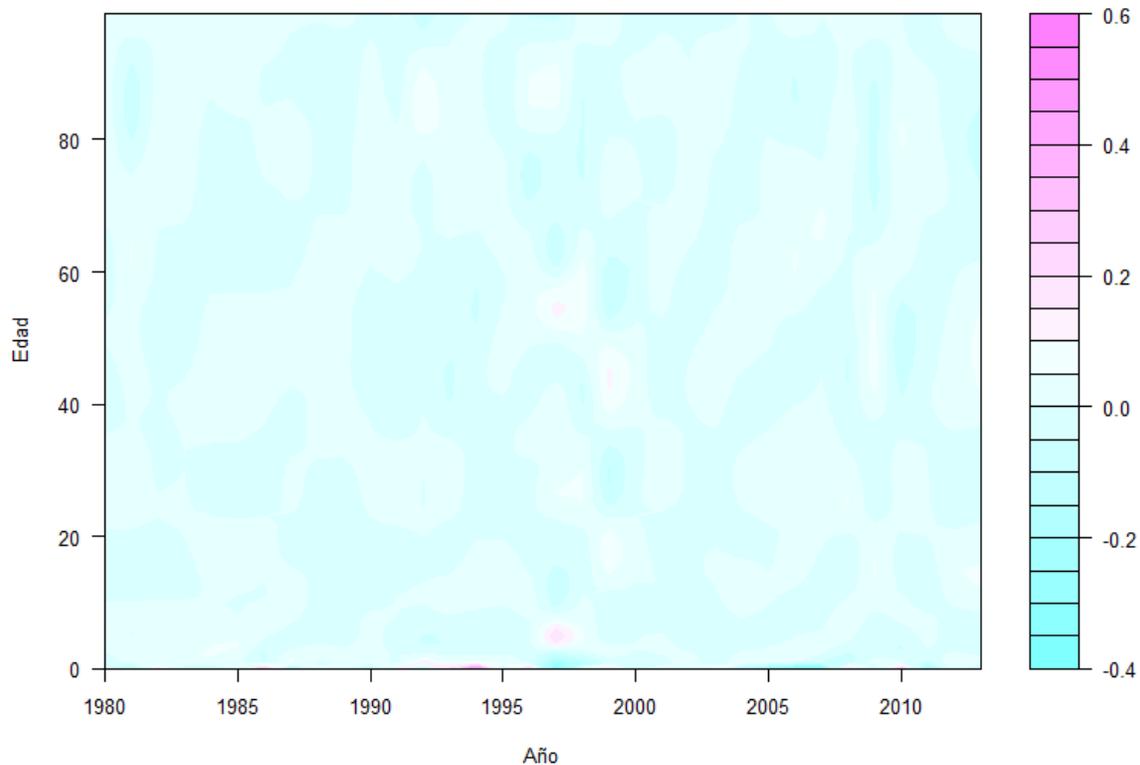


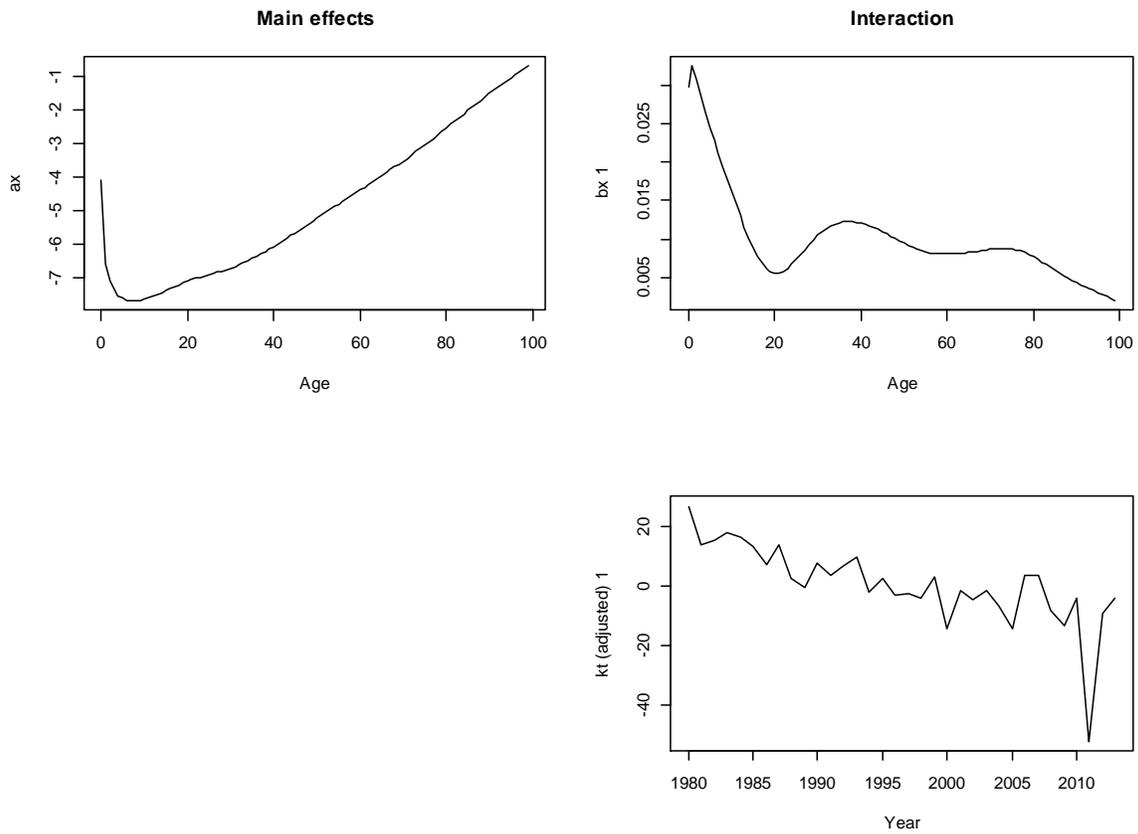
Gráfico 10. Residuos del modelo MDF. Río Cuarto.



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

En segundo lugar se estima el Modelo de Lee-Carter para el departamento Río Cuarto, el cual explica el 57.9% de la variabilidad inherente a las tasas. Al igual que para el departamento Capital el comportamiento es similar al modelo para Datos Funcionales (Gráfico 11.); el parámetro a coincide con la media y el parámetro b y el índice general de mortalidad k coincide con el primer par base-coeficiente respectivamente. En relación a la bondad de ajuste relativa a este modelo, el MAPE asciende al 0,027%.

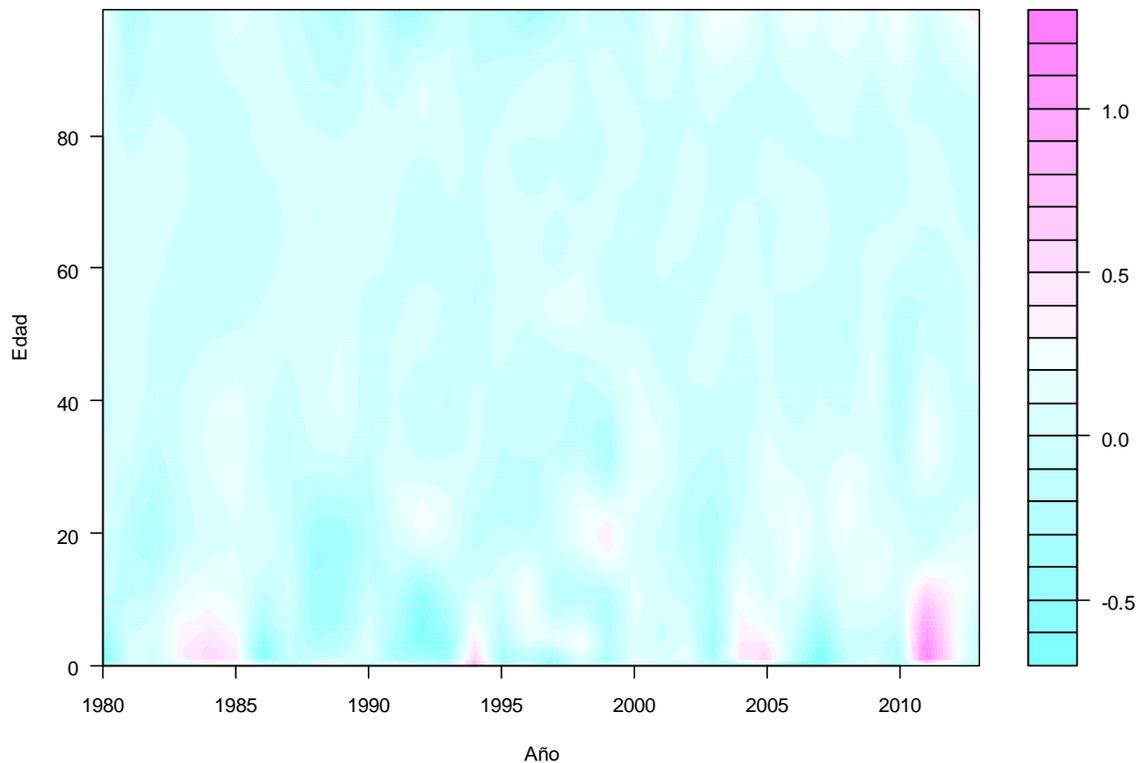
Gráfico 11. Modelo de Lee-Carter. Río Cuarto



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

Finalmente el gráfico de los residuos obtenidos para el modelo Lee-Carter del departamento Río Cuarto (Gráfico 12) presenta hacia el final de período y para edades iniciales una franja rosada y en el resto del mismo el color celeste, que indica que el modelo sobreestima las tasas observadas.

Gráfico 12. Residuos del modelo LC. Río Cuarto.

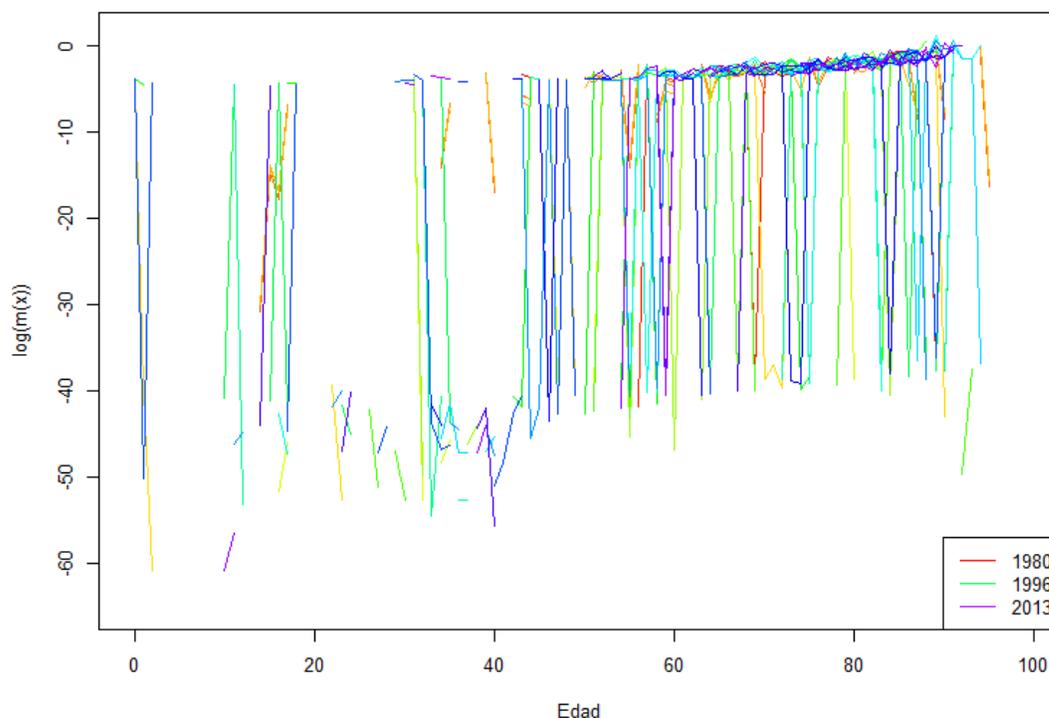


Fuente: Elaboración propia en base a datos del INDEC y DEIS

5.1.3. Departamento Sobremonte

Finalmente se selecciona el departamento Sobremonte, ya que es uno de los menos poblados según el censo 2010. El departamento presenta un alto porcentaje de tasas nulas, que son desestimadas a la hora de modelizar ya que se emplea la transformación logaritmo, función para la cual su valor en el punto cero no está definido. A continuación, en el gráfico 13 se presentan los logaritmos de las tasas observadas.

Gráfico 13. Logaritmo de las tasas observadas Departamento Sobremonte (1980-2013)



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

El proceso de suavizado no se puede ejecutar debido a la alta presencia de tasas nulas, es por ellos que no se puede aplicar ninguno de los modelos propuestos a las tasas del mencionado departamento.

Esto revela una limitación tanto del modelo de Lee-Carter como del Modelo para Datos Funcionales cuando se pretende trabajar con regiones geográficas que contienen poca población dado que presentan una importante cantidad de tasas específicas de mortalidad nulas.

Luego, dado los problemas en la estimación que presenta este departamento se procede a evaluar el desempeño de ambos modelos para los tres departamentos para las tasas específicas de mortalidad, pero en grupos quinquenales en lugar de edades simples (0, 1, 2, 3, 4, 5-9, ... ,95 y más). Si bien el objetivo central de esta tesis es trabajar en las condiciones más "estresantes" posibles para los modelos; edades simples, regiones pequeñas, etc, es interesante

evaluar si los modelos son estimables al agrupar las edades. En el caso de el 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 2.) 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 2. Comparación del porcentaje de variabilidad explicada entre edades simples y agrupadas. Departamentos Capital y Río Cuarto, mediante Modelo para Datos Funcionales y Modelo de Lee-Carter.

Dpto.	Modelo	Edades		Dpto.	Modelo	Edades	
		Simplees	Agrupadas			Simplees	Agrupadas
Capital (6 bases)	MDF	73,0	59,2	Río Cuarto	MDF (6 bases)	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.

Por otro lado para ambos departamentos y mediante ambos modelos estimados (MDF y LC) 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%, ver tabla 3.

Tabla 3. 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 (%).

	Capital		Río Cuarto	
Modelo/Edades	Simple	Agrupadas	Simple	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.

A continuación en la siguiente sección se presentan los pronósticos de mortalidad y de esperanza de vida al nacer con sus respectivos intervalos de confianza, para ambos dominios de estimación.

5.2. Pronósticos

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 95% de confianza.

Los modelos aquí empleados generan valores pronosticados de las tasas para edades simples (si bien por cuestiones de espacio se seleccionan algunas edades) y para un horizonte de 40 años, sin embargo no es recomendable utilizar un horizonte de pronósticos tan extendido y este hecho se ve reflejado en la amplitud del intervalo, ya que dicha amplitud resulta tan grande que el intervalo ya no es informativo.

En la tabla 4 se presentan los intervalos de pronóstico obtenidos a través del Modelo para Datos Funcionales para el departamento Capital, para edades y años seleccionados, si bien se generan para todos los años por edad simple. 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 evalúa el intervalo obtenido para los 95 y 99 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.

Tabla 4. Tasas de mortalidad específicas pronosticadas para el Departamento Capital a través de MDF (selección de edades y años)

Edad/Año	2020	2025	2030	2040
0	9,73(6,95;13,61)	8,65(5,75;13,00)	7,69(4,81;12,29)	6,08(3,48;10,77)
5	0,23(0,14;0,40)	0,21(0,12;0,37)	0,20(0,11;0,35)	0,16(0,09;0,31)
10	0,20(0,10;0,41)	0,19(0,09;0,38)	0,17(0,08;0,35)	0,14(0,07;0,31)
15	0,44(0,26;0,74)	0,43(0,26;0,72)	0,42(0,25;0,71)	0,40(0,24;0,65)
20	0,73(0,52;1,04)	0,73(0,52;1,04)	0,73(0,52;1,04)	0,74(0,52;1,04)
25	0,77(0,50;1,18)	0,75(0,49;1,16)	0,74(0,48;1,14)	0,71(0,46;1,09)
30	0,82(0,58;1,16)	0,78(0,55;1,12)	0,74(0,51;1,08)	0,67(0,45;1,00)
35	1,00(0,71;1,40)	0,93(0,64;1,34)	0,86(0,58;1,28)	0,74(0,47;1,16)
40	1,45(1,06;1,97)	1,34(0,94;1,89)	1,23(0,84;1,80)	1,05(0,67;1,63)
45	2,31(1,75;3,05)	2,13(1,55;2,94)	1,97(1,38;2,81)	1,68(1,10;2,56)
50	3,86(3,03;4,92)	3,60(2,72;4,76)	3,35(2,45;4,58)	2,91(2,01;4,21)
55	6,52(5,30;8,02)	6,16(4,87;7,80)	5,82(4,49;7,55)	5,20(3,84;7,05)
60	10,93(9,28;12,86)	10,50(8,75;12,58)	10,08(8,27;12,29)	9,30(7,40;11,68)
65	16,85(14,61;19,44)	16,28(13,82;19,06)	15,63(13,09;18,65)	14,49(11,80;17,80)
70	24,67(21,10;28,85)	23,58(19,70;28,23)	22,54(18,45;27,55)	20,60(16,26;26,10)
75	38,19(32,61;44,72)	36,45(30,35;43,76)	34,78(28,33;42,69)	31,67(24,84;40,37)
80	63,08(53,81;73,95)	60,33(50,30;72,36)	57,68(47,13;70,60)	52,73(41,58;66,87)
85	109,92(95,22;126,88)	106,27(90,72;124,49)	102,73(86,53;121,98)	96,01(78,93;116,78)
90	200,53(172,85;232,63)	199,20(171,60;231,23)	197,84(170,32;229,79)	195,14(167,80;226,94)
95	362,37(279,43;469,94)	371,86(285,39;484,53)	381,40(291,36;499,28)	401,21(303,72;529,98)
99	559,63(381,08;821,83)	587,58(395,58;872,76)	616,43(410,44;925,79)	678,39(442,19;1040,77)

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

La tabla 5 presenta los resultados para el mismo departamento pero mediante el modelo de Lee-Carter. En primer lugar resultan en general para todas las edades y años más altas que las obtenidas por el método anterior. Por otro lado, presentan intervalos más amplios y se destaca que para las edades avanzadas los intervalos presentan un error de estimación (los límites superior e inferior son incoherentes) esto podría deberse a algún problema de convergencia. Asimismo a través del tiempo la amplitud de los intervalos se hace más grande y mayor también en comparación con el Modelo para Datos Funcionales.

Tabla 5. Tasas de mortalidad específicas pronosticadas para el Departamento Capital a través de LC (selección de edades y años)

Edad/Año	2020	2025	2030	2040
0	11,11(6,11;20,21)	10,10(4,40;23,19)	9,19(3,24;26,05)	7,59(1,80;32,01)
5	0,25(0,17;0,39)	0,24(0,13;0,43)	0,22(0,11;0,47)	0,19(0,07;0,54)
10	0,23(0,14;0,35)	0,21(0,11;0,39)	0,20(0,09;0,42)	0,17(0,06;0,45)
15	0,45(0,40;0,52)	0,45(0,37;0,53)	0,44(0,35;0,54)	0,42(0,31;0,57)
20	0,73(0,74;0,72)	0,73(0,74;0,72)	0,73(0,75;0,72)	0,74(0,76;0,72)
25	0,79(0,71;0,89)	0,78(0,67;0,91)	0,76(0,63;0,93)	0,74(0,56;0,97)
30	0,87(0,67;1,13)	0,84(0,59;1,20)	0,80(0,51;1,26)	0,74(0,40;1,38)
35	1,09(0,75;1,58)	1,02(0,61;1,72)	0,96(0,50;1,85)	0,86(0,35;2,11)
40	1,59(1,05;2,40)	1,49(0,84;2,64)	1,39(0,68;2,86)	1,22(0,45;3,29)
45	2,53(1,69;3,80)	2,37(1,35;4,17)	2,23(1,10;4,51)	1,96(0,74;5,18)
50	4,19(2,92;5,99)	3,95(2,40;6,51)	3,73(2,00;6,98)	3,33(1,41;7,90)
55	6,96(5,23;9,26)	6,65(4,47;9,88)	6,35(3,86;10,45)	5,80(2,92;11,53)
60	11,45(9,34;14,04)	11,09(8,35;14,72)	10,73(7,52;15,31)	10,06(6,16;16,43)
65	17,6(14,55;21,30)	17,07(13,10;22,25)	16,56(11,88;23,09)	15,59(9,86;24,66)
70	25,99(20,69;32,66)	25,06(18,25;34,41)	24,17(16,24;35,98)	22,48(12,98;38,92)
75	40,32(31,81;51,12)	38,83(27,94;53,97)	37,39(24,74;56,51)	34,68(19,61;61,33)
80	66,46(52,96;83,39)	64,10(46,77;87,85)	61,83(41,64;91,81)	57,52(33,32;99,29)
85	114,33(96,35;135,66)	111,26(87,73;141,10)	108,27(80,37;145,88)	102,54(67,95;154,74)
90	202,32(195,60;209,26)	201,23(192,01;210,89)	200,15(188,72;212,28)	198,01(182,57;214,77)
95	352,74(401,73;309,71)	360,11(431,41;300,60)	367,64(461,16;293,09)	383,18(523,94;280,23)
99	531,36(679,10;415,77)	552,51(776,78;393,00)	574,50(880,88;374,69)	621,14(1120,62;344,29)

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

A continuación se presentan los resultados para el departamento Río Cuarto, tabla 6 para los obtenidos por Modelo para Datos Funcionales y tabla 7 para los obtenidos por Lee-Carter. El nivel estimado de los pronósticos resulta

inferior en el caso del Modelo para Datos Funcionales, pero se destaca para estos resultados intervalos extremadamente amplios para edades de 5 a 15 y de 90 años en adelante. La menor población del departamento y la mayor variabilidad en las tasas parecen implicar un mal funcionamiento en ambos algoritmos.

Tabla 6. Tasas de mortalidad específicas pronosticadas para el departamento Río Cuarto a través de MDF (selección de edades y años)

Edad/Año	2020	2025	2030	2040
0	8,81(4,41;17,61)	7,80(3,44;17,70)	6,91(2,73;17,49)	5,43(1,78;16,55)
5	0,33(0,00;198,88)	0,30(0;18,42)	0,27(0;16,86)	0,22(0,00;13,73)
10	0,40(0;69,32)	0,38(0;65,05)	0,35(0;60,63)	0,31(0;58,58)
15	0,63(0,00;227,94)	0,61(0,00;221,77)	0,59(0,00;215,75)	0,54(0,00;204,19)
20	0,89(0,37;2,13)	0,87(0,34;2,25)	0,85(0,31;2,36)	0,81(0,26;2,57)
25	0,99(0,25;3,89)	0,96(0,23;3,84)	0,93(0,22;3,80)	0,87(0,21;3,71)
30	1,05(0,27;4,10)	1,01(0,26;3,94)	0,96(0,25;3,77)	0,88(0,23;3,47)
35	1,28(0,24;6,97)	1,22(0,22;6,65)	1,16(0,21;6,34)	1,05(0,19;5,77)
40	1,80(0,55;5,92)	1,72(0,52;5,69)	1,63(0,49;5,46)	1,48(0,43;5,04)
45	2,76(1,59;4,77)	2,63(1,49;4,65)	2,52(1,40;4,53)	2,30(1,23;4,29)
50	4,39(1,06;18,14)	4,22(1,01;17,56)	4,06(0,97;16,99)	3,75(0,87;15,92)
55	6,94(4,84;9,97)	6,71(4,60;9,80)	6,49(4,37;9,62)	6,06(3,96;9,26)
60	10,71(7,92;14,48)	10,36(7,58;14,14)	10,02(7,27;13,80)	9,37(6,68;13,15)
65	16,29(12,44;21,32)	15,74(11,92;20,78)	15,21(11,43;20,25)	14,21(10,50;19,22)
70	24,87(19,16;32,28)	24,00(18,24;31,57)	23,15(17,37;30,86)	21,55(15,79;29,42)
75	39,76(31,15;50,75)	38,36(29,55;49,79)	37,00(28,06;48,80)	34,43(25,36;46,76)
80	68,28(54,25;85,95)	66,16(52,02;84,14)	64,10(49,91;82,33)	60,18(45,99;78,74)
85	122,66(98,60;152,59)	119,73(96,01;149,30)	116,86(93,49;146,08)	111,34(88,66;139,83)
90	217,57(156,20;303,04)	213,69(149,59;305,26)	209,88(143,49;306,99)	202,47(132,56;309,24)
95	375,56(199,61;706,60)	370,63(186,14;738,00)	365,76(174,33;767,39)	356,21(154,41;821,74)
99	575,20(106,11;3118,14)	569,89(100,44;3233,67)	564,63(95,18;3349,56)	554,25(85,75;3582,53)

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

Tabla 7. Tasas de mortalidad específicas pronosticadas para el departamento Río Cuarto a través de LC (selección de edades y años)

Edad/Año	2020	2025	2030	2040
0	12,49(2,73;57,24)	10,87(1,31;90,07)	9,47(0,67;134,37)	7,18(0,18;279,54)
5	0,39(0,11;1,35)	0,35(0,06;1,97)	0,31(0,03;2,73)	0,25(0,01;5,00)
10	0,41(0,18;0,95)	0,38(0,12;1,21)	0,35(0,08;1,51)	0,30(0,04;2,26)
15	0,58(0,37;0,91)	0,56(0,30;1,04)	0,53(0,24;1,17)	0,49(0,17;1,46)
20	0,79(0,59;1,04)	0,77(0,52;1,13)	0,75(0,46;1,22)	0,71(0,36;1,39)
25	0,93(0,64;1,36)	0,90(0,54;1,51)	0,87(0,46;1,67)	0,82(0,33;2,00)
30	1,10(0,65;1,88)	1,05(0,50;2,21)	1(0,39;2,54)	0,91(0,25;3,29)
35	1,44(0,77;2,68)	1,36(0,57;3,23)	1,28(0,43;3,80)	1,14(0,26;5,13)
40	2,06(1,11;3,82)	1,95(0,83;4,58)	1,84(0,63;5,39)	1,65(0,37;7,25)
45	3,12(1,78;5,48)	2,97(1,36;6,48)	2,82(1,06;7,51)	2,54(0,66;9,84)
50	4,89(3,01;7,95)	4,68(2,39;9,19)	4,48(1,92;10,44)	4,10(1,28;13,18)
55	7,63(4,98;11,67)	7,34(4,06;13,25)	7,06(3,36;14,81)	6,53(2,35;18,18)
60	11,56(7,65;17,48)	11,14(6,27;19,77)	10,73(5,22;22,04)	9,95(3,68;26,89)
65	17,53(11,49;26,74)	16,87(9,38;30,32)	16,23(7,78;33,88)	15,03(5,44;41,52)
70	27,05(17,36;42,15)	25,98(14,03;48,10)	24,95(11,52;54,05)	23,02(7,92;66,91)
75	43,39(27,75;67,84)	41,67(22,40;77,50)	40,01(18,36;87,16)	36,88(12,59;108,07)
80	73,28(49,40;108,70)	70,70(40,89;122,24)	68,21(34,31;135,58)	63,49(24,59;163,90)
85	126,26(93,29;170,88)	122,83(80,69;187,00)	119,50(70,53;202,47)	113,10(54,62;234,20)
90	211,31(169,36;263,65)	207,10(152,30;281,62)	202,97(138,03;298,48)	194,97(114,49;332,02)
95	341,15(292,17;398,34)	336,37(271,24;417,16)	331,67(253,18;434,49)	322,45(222,11;468,14)
99	493,54(446,67;545,32)	489,08(425,80;561,77)	484,66(407,32;576,69)	475,95(374,39;605,06)

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

Finalmente y dado que Sobremonte dado su poca población no permitió la implementación de los modelos, 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 (para más detalle ver Glosario). La tabla 8 presenta resultados plausibles, para ambos departamentos los intervalos resultan más estrechos para los intervalos obtenidos por Modelo para Datos Funcionales y como es de esperar presenta una esperanza mayor coherente con los niveles inferiores obtenidos anteriormente en las tasas de mortalidad pronosticadas.

Tabla 8. Esperanza de Vida al Nacer pronosticada. Departamento Capital y Río Cuarto. Modelos LC y MDF para años seleccionados

Año	Capital		Río Cuarto	
	MDF	LC	MDF	LC
2020	76,74(75,07;77,97)	76,08(72,83;78,51)	75,93(74,07;77,4)	74,92(67,52;79,68)
2025	77,29(75,41;78,69)	76,54(72,56;79,59)	76,39(74,28;78,05)	75,5(66,67;81,52)
2030	77,81(75,57;79,51)	76,99(72,39;80,68)	76,84(74,82;78,43)	76,05(61,94;82,71)
2040	78,81(76,34;80,83)	77,85(72,46;81,75)	77,7(75,01;79,41)	77,08(58,59;84,91)

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

Se realiza 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.

Tabla 9. Esperanza de vida al nacer según sexo para la provincia de Córdoba 2015-2040

	2015	2020	2025	2030	2035	2040
Varones	74,07	75,20	76,20	77,10	77,88	78,57
Mujeres	80,67	81,72	82,66	83,49	84,22	84,86
Promedio	77,37	78,46	79,43	80,30	81,05	81,72

Fuente: elaboración propia en base a datos INDEC.

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 para 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.

6. Conclusiones

Retomando el objetivo general de este trabajo; la evaluación del desempeño de las proyecciones probabilísticas de la mortalidad empleando dos grandes exponentes de las metodologías probabilísticas, el modelo para Datos Funcionales y el de Lee y Carter, a datos correspondientes a tres departamentos de la provincia de Córdoba, se observa que los resultados a este nivel geográfico no son los esperados, ya que a medida que la magnitud de población descende 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”.

Los objetivos secundarios comprendían la evaluación de las metodologías de pronóstico probabilístico en mortalidad en regiones geográficas pequeñas analizando aspectos puntuales tales como: la intervención del demógrafo en la metodología, la calidad de la fuente de datos y su relación con los modelos bajo estudio, y aspectos intrínsecos tales como suavizados y la cuantificación de la incertidumbre asociada a los resultados.

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, como sucede al llegar a la tonalidad azul de los últimos años (recordar la escala “arco iris”). 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 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.

Un punto que vale la pena destacar es que las fuentes de datos en general van mejorando su calidad a través de los años, por ello al trabajar con datos históricos hay que contemplar la existencia de una variabilidad al inicio del período en estudio vinculada en mayor medida al proceso de generación del dato que al comportamiento inherente a la mortalidad misma. Estos métodos se tornarían más adecuados para su uso con la consolidación de las fuentes de datos, la mejora en la captación y la disminución de la omisión y de los errores de registro.

Si bien los métodos aquí aplicados no fueron diseñados para áreas pequeñas resulta útil evaluarlos en dicho contexto, ya que, una metodología de pronóstico requiere ser empleada a distintos niveles geográficos y es deseable conocer sus limitaciones. Las metodologías intentan hacer una descripción, una simplificación o reducción del comportamiento complejo de la mortalidad, tienen como principal función la generación de pronósticos y, en este punto, la información que proveen se puede considerar útil, que permita desarrollar teorías demográficas o que ayude a comprender la dinámica de la mortalidad de modo que le sea útil tanto a los organismos oficiales como provechoso para la sociedad en general. Los modelos por lo general no arrojan buenos resultados, ya que el Modelo para Datos Funcionales que presenta intervalos estrechos para las tasas de mortalidad pronosticadas para Capital, no arroja los mismos resultados para Río Cuarto y más aún ninguno de los modelos arroja resultados para Sobremonte. Ahora bien es importante destacar que la población del departamento Capital equivale a la población total de algunos países europeos, y las metodologías aquí propuestas se presentan como una propuesta para obtener estimaciones y

pronósticos a nivel país y ciudad. 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, además de permitir elegir entre algunas estrategias de suavizado.

Sin embargo las diversas aplicaciones a otros niveles geográficos presentan buenos resultados. Lee y Rofman (1994) en su aplicación del modelo de Lee-Carter a datos de Chile, y los autores afirman que se trata de un modelo simple es útil para completar vacíos en los datos de edad y períodos de tiempo y además en algunos casos se emplea para ajustar el subregistro y proyectar la mortalidad; con obvias ventajas dentro de la estructura "estadística-demográfica" integrada. Guerrero y González (2007) desarrollan un trabajo en base a la aplicación del modelo a datos de Méjico evaluando el impacto de los resultados sobre el sistema previsional. Luego, García Guerrero y Ordorica (2012) concluyen sobre su aplicación del modelo de Lee y Carter a Méjico que 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.

En Argentina en Blaconá y Andreozzi (2011) y Andreozzi y Blaconá, (2012) se obtienen esperanzas de vida al nacer pronosticadas con intervalos de pronóstico de reducida amplitud indicando una buena calidad en los resultados obtenidos. También se obtienen resultados adecuados en relación a la aplicación de los

modelos para datos funcionales. En la primera aplicación que se realiza para Argentina (tesis de 2016: "Proyecciones probabilísticas en Demografía") se demuestra que los modelos para datos funcionales se pueden emplear con éxito para pronosticar tasas de mortalidad específicas por edad y género y tasas globales de fecundidad por edad simple cuando se cuenta con información para períodos no muy extensos de tiempo y una calidad aceptable de las fuentes de datos (en regiones populosas). Remarcando además que dicho enfoque de modelado es altamente adaptable; el mismo modelo básico se aplica a las tres componentes demográficas permitiendo transformaciones y suavizados específicos para cada una de ellas aunque la aplicación a datos de migración resulta defectuosa.

En relación a dimensiones geográficas más pequeñas se realizó una aplicación de Modelos de Datos Funcionales, una variante Modelo para Datos Funcionales Coherente y Lee-Carter a datos de la Ciudad Autónoma de Buenos Aires (Andreozzi, 2016). El artículo concluye que los modelos estadísticos proporcionan un intervalo de pronóstico para tasas de mortalidad y esperanza de vida que resulta más informativo que la mera proyección puntual.

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 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. Sin embargo luego del proceso de estimación se destacan resultados atípicos en la obtención de tasas específicas de mortalidad, en las edades 95 y 99 para el Modelo de Lee-Carter estimado para Capital y edades 5, 10 y 15, 95 y 99 en las estimaciones que se obtienen a través del Modelo para Datos

Funcionales aplicado a Río Cuarto. Esto refuerza la necesidad de agrupar edades en intervalos.

Más allá de los resultados técnicos, como ya se mencionó, estos modelos surgen en un contexto histórico particular, en la década de los 90, cuando el envejecimiento poblacional se cataloga de "problema", en aparente contraposición a la metodología clásica que se basa en supuestos de los demógrafos, 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 únicamente en datos cuantitativos no proporciona resultados mejores. 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 este punto retomamos el concepto introducido por John Hobcraft (2003): "retorno teórico", al referirse a la información útil que generan determinados análisis. 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 para 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 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 relativamente fácil implementación. 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 clásicas a las cuáles se les critica en este punto, existe un límite en la medición. Como sostuvo el gran sociólogo francés Alain Desrosières, cuantificar es convenir, acordar, consensuar, luego medir. Convenir es ponerse de acuerdo sobre lo que merece ser medido, es definir, clasificar; la medida viene siempre después (Beaud, 2016). Por ejemplo, ningún método, matemático, estadístico o demográfico pudo prever el descenso en la esperanza de vida debido a la aparición del HIV, como así tampoco en la actualidad ningún método pronosticó el descenso de la esperanza de vida durante dos años consecutivos que se observa en Estados Unidos a raíz de la crisis de los opiáceos. Asimismo, otra limitación importante de destacar es que los métodos resultan complejos, en su lógica y en su implementación, requieren gran volumen de datos y si bien generan resultados útiles es necesario poner en perspectiva su implementación.

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 para horizontes de tiempo muy extensos (aproximadamente, 30 años), en el caso de la demografía y la matemática porque directamente no la calculan y en el caso de la estadística porque la amplitud de los intervalos de pronóstico es tan amplia que la incertidumbre es muy grande y los datos no son informativos. Sólo los métodos estadísticos ponen de manifiesto el grado de incertidumbre que conllevan los pronósticos.

Como sostenía Heisenberg “Cuando uno desea conocer el “futuro” a partir del “presente” sólo puede obtener resultados estadísticos, puesto que nunca puede descubrir todos los detalles del presente”⁷.

⁷ 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.

7. Glosario

Tasas

El término tasa se usa en muchos campos y su significado no es consistente en todos ellos. En demografía, las tasas se definen comúnmente como tasas de ocurrencia/exposición (Preston *et al.*, 2001). El numerador de este tipo de tasas contabiliza el número de ocurrencias de un evento de interés, mientras que el denominador combina dos factores: el número de personas en la población y la longitud del tiempo que enmarca el estudio. El denominador de una tasa de "ocurrencia/exposición", se conoce con el nombre de personas-tiempo en riesgo y es la suma del tiempo que cada persona permanece bajo observación y en riesgo de convertirse en un caso. El tiempo debe expresarse en unidades de medida apropiadas, por ejemplo, personas-año, en cuyo caso se dice que la tasa está anualizada. Cuando los casos se obtienen a partir del registro de estadísticas vitales no se dispone de medidas directas de personas-tiempo en riesgo. Una estimación de dicho valor, para una población en un período dado, se puede calcular como el producto entre el número total de personas que componen la población en el punto medio del período de interés y la duración total del período. Este método proporciona estimaciones adecuadas siempre que la población permanezca estable durante ese período.

Las tasas se pueden calcular para una población entera o para subgrupos específicos de la misma. En el primer caso se las denomina tasas crudas o brutas y en el segundo tasas específicas por subgrupo de interés. Por ejemplo, en el estudio de la mortalidad resulta de interés el cálculo de tasas específicas por sexo y edad.

En las tasas específicas, el denominador de la tasa se define por separado para cada grupo. Por ejemplo, la tasa específica de mortalidad según edad se define como el cociente entre el número de defunciones acaecidas en un grupo de edad específica de la población de un área geográfica dada durante un período determinado y el correspondiente valor de personas-tiempo en riesgo en ese grupo específico de edad del área geográfica y período bajo estudio. Si las defunciones se obtienen a partir de Registros de Estadísticas Vitales, el denominador se estima a partir de los datos censales.

Suavizados

Los modelos estadísticos son siempre simplificaciones, e incluso los modelos más complicados serán una pálida imitación de la realidad. Dicho esto, puede parecer un esfuerzo inútil estimar modelos estadísticos, pero cómo George Box lo describió sucintamente: "Todos los modelos están equivocados, algunos son útiles." A pesar de que los modelos siempre están equivocados, las estadísticas nos proporcionan una visión considerable en el mundo político, económico y sociológico que habitamos. Ellos se convierten en simplificaciones de la realidad porque debemos hacer suposiciones sobre diversos aspectos de la realidad. Si no existe una idea acerca de la forma funcional de los datos en estudio, en lugar de suponer una, la mejor alternativa es estimar la forma funcional apropiada a partir de los datos. En la ausencia de teoría fuerte para la forma funcional, ésta es a menudo la mejor manera para proceder. Para estimar la forma funcional de los datos, debemos reemplazar la estimación global por la estimación local.

Con las estimaciones globales, el analista asume una forma funcional para el modelo; con estimaciones locales, la forma funcional se estima a partir de los datos. Los estimadores locales utilizados en esta tesis se denominan modelos de regresión no paramétricos o suavizados. La regresión no paramétrica permite estimar los ajustes no lineales entre variables continuas con pocas hipótesis sobre la forma funcional de la no linealidad. Las técnicas lowess y splines son los modelos de regresión no paramétricos más conocidos y difundidos y se basan en estimaciones locales para estimar la forma funcional de los datos. (Keele, 2008)

En la estimación local, la dependencia estadística entre dos variables se describe no con un solo parámetro como una media o un coeficiente β , sino con una serie de estimaciones locales. Es decir, la estimación de una media o una regresión se estima entre Y y X para algunos rangos restringidos de X e Y . Esta estimación local de la dependencia entre X e Y se repite en el rango de X e Y . Esta estimación no paramétrica resultante, que puede ser lineal o cuadrática, no impone una forma funcional particular a la relación entre X e Y . Debido a la naturaleza local del proceso de estimación, la regresión no paramétrica proporciona ajustes muy flexibles entre X e Y . Los modelos locales, como la

regresión no paramétrica, estiman la forma funcional entre dos variables, mientras que los modelos globales imponen una forma funcional datos.).

Crecimiento poblacional exponencial

En el modelo aritmético el supuesto básico consiste en que la población crece en un mismo monto cada unidad de tiempo. En el modelo exponencial se mantiene constante es el porcentaje de crecimiento por unidad de tiempo y no el monto. Supongamos que r es la tasa de crecimiento por unidad de tiempo. Generalizando para el momento $t + k$, la población final (N_f) sería

$$N_f = N_i(1 + r)k,$$

siendo N_i la población inicial.

En cambio, el modelo exponencial supone que el crecimiento se produce en forma continua y no cada unidad de tiempo. Este supuesto obliga a sustituir la expresión $(1 + r)k$ por $\exp(rt)$. El tamaño de la población en el momento $t + k$ viene dado por:

$$N_f = N_i \cdot \exp(r \cdot k)$$

entonces, la tasa de crecimiento poblacional bajo este supuesto viene dada por:

$$r = \frac{1}{k} \cdot \ln\left(\frac{N_f}{N_i}\right)$$

Metodología de Tablas de vida de Chiang

Si bien para el cálculo de la esperanza de vida se emplea la función disponible en el paquete Demography del software libre R. Es importante remarcar que dicho cálculo se basa en la fórmula de 1984 propuesta por Chiang,

$$q_x = \frac{m_x}{(1 + [1 - a_x]m_x)}$$

Donde q_x es la probabilidad de que un individuo de la edad exacta x muera antes de la edad exacta $x + 1$., m_x es la tasa de mortalidad a la edad x y a_x es el

número de años del intervalo que, en promedio, viven las personas que mueren en el intervalo.

8. Referencias bibliográficas

- Alho, J. y Spencer, B. (1985). Uncertain population forecasting. *Journal of the American Statistical Association*, 80:306–314.
- Andreozzi, L. (2016). "Proyecciones probabilísticas en Demografía" presentada por la propia autora para obtener el grado de Magister en Estadística Aplicada.
- 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.
- Blau, P. M., and Duncan, O. D. (1967), *The American Occupational Structure*, New York: Wiley
- Beer, J. D. (1997). The effect of uncertainty of migration on national population forecasts: the case of the Netherlands. *Journal of Official Statistics*, 13:227–243.
- Bianchi, S. M., and Spain, D. (1986), *American Women in Transition*, New York: Russell Sage Foundation.
- Bogue, D. J., *Principles of Demography*. New York: John Wiley. 1969.
- 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.

Bozik, J. y Bell, W. (1987). Forecasting age-specific fertility using principal components.

Brouhns, N., Denuit, M., y Vermunt, J. (2002). A Poisson log-bilinear regression approach to the construction of projected lifetables. *Insurance: Mathematics and Economics*, 31(3):373–393.

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.

Caldwell J.C., 1996, "Demography and social science", *Population Studies*, vol. 50, pp. 305-333

Coale, A. J. (1972), *The Growth and Structure of Human Populations: A Mathematical Investigation*. Princeton, NJ: Princeton University Press.

Farley, R. (1984), *Blacks and Whites: Narrowing the Gap?*, Cambridge, MA: Harvard University Press.

Featherman, D. L., and Hauser, R. M. (1978), *Opportunity and Change*, New York: Academic Press.

Duncan, O. D. (1957), *The Negro Population of Chicago: A Study of Residential Succession*, Chicago: University of Chicago Press.

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>>.

Giroi, 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>

George, M. y Perreault, J. (1992). Methods of external migration projections and forecasts. *National population forecasting in industrialized countries*, p. 87103.

Guimaraes, R.R.M. (2014) Uncertainty in population projections: the state of the art. *Revista Brasileira de Estudos de População* 31 (2), 277-290.

Hauser, P. and Duncan, O. (eds), 1959. *The Study of Population*, University of Chicago, Chicago.

Hilderink, H., der Gaag, N. V., Wissen, L. V., Jennissen, R., Roman, A., Salt, J., Clarke, J., y Pinkerton, C. (2002). Analysis and forecasting of international migration by major groups, part iii. *Working papers and studies*, 3.

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.

Jong, P. D. y Tickle, L. (2006). Extending Lee-Carter mortality forecasting. *Mathematical Population Studies*, 13(1):1–18.

INDEC. Proyecciones elaboradas en base a resultados del Censo Nacional de Población, Hogares y Viviendas 2010. Cuadro 4. en https://www.indec.gob.ar/nivel4_default.asp?id_tema_1=2&id_tema_2=24&id_tema_3=85. Consultada el 20/06/18.

Kalben, B. (2000). Why men die younger. *North American Actuarial Journal*, 4(4):83–111.

Keilman, N. (2008). Using deterministic and probabilistic population forecasts. *Interdisciplinary Communications*, University of Oslo, Department of Economics.

- Keilman, N. and H. Crujisen (1992). *National population forecasting in industrialized countries*. Amsterdam: NIDI/CBGS (Publication No. 24).
- Keilman, N. y Pham, D. (2004). Empirical errors and predicted errors in fertility, mortality and migration forecasts in the European economic area. *Discussion Papers, Research Department of Statistics Norway*, 386.
- Keyfitz, N. (1985), *Applied Mathematical Demography* (2nd ed.), New York: Springer-Verlag.
- 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.
- Little, R. J. A., and Rubin, D. B. (1987), *Statistical Analysis With Missing Data*, New York: Wiley
- Loriaux m., 1996, "L'apport de l'école louvaniste de démographie: réflexions critiques sur le passé et nouveaux défis pour l'avenir des sciences de la population", *Population*, 51(2), pp. 405-416.
- Massey, D. S., and Denton, N. A. (1993), *American Apartheid: Segregation and the Making of the Underclass*, Cambridge, MA: Harvard University Press.

- Mayr, E. (1982) *The Growth of Biological Thought: Diversity, Evolution, and Inheritance* (Belknap Press of Harvard Univ. Press, Cambridge, MA)
- 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.
- Mills, C.W. (1994) *La imaginación sociológica*. México, Fondo de Cultura Económica
- 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
- 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.
- Renshaw, A. y Haberman, S. (2003a). Lee-Carter mortality forecasting: A parallel generalized linear modelling approach for England and Wales mortality projections. *Applied Statistics*, 52(1):119–137.
- Renshaw, A. y Haberman, S. (2003b). Lee-Carter mortality forecasting with age-specific enhancement. *Insurance: Mathematics and Economics*, 33(2):255–272.
- Renshaw, A. y Haberman, S. (2003c). On the forecasting of mortality reduction factors. *Insurance: Mathematics and Economics*, 32(3):379–401.
- Rogers, A. (1975), *Introduction to Multiregional Mathematical Demography*, New York: Wiley

- Rogers, A. (1990). Requiem for the net migrant. *Geographical Analysis*, 22:283–300.
- Rogers, A. y Castro, L. (1981). Model migration schedules. *Research Report, International Institute for Applied Systems Analysis*, pp. 81–30.
- Rogers, A. y Little, J. S. (1994). Parameterizing age patterns of demographic rates with the multiexponential model schedule. *Mathematical Population Studies*, 4:175–195.
- Sautu R., (2005). *Todo es Teoría. Objetivos y métodos de investigación*. Buenos Aires: Lumiere.
- Sheps, M. C., and Menken, J. A. (1973), *Mathematical Models of Conception and Birth*, Chicago: University of Chicago Press.
- 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.
- Sykes, Z. (1969). Some stochastic versions of the matrix model for population dynamics. *Journal of the American Statistical Association*, 44:111–130.
- Tabutin, D. y Depledge, R. Whither. (2007) Demography? Strengths and Weaknesses of the Discipline over Fifty Years of Change. *Population (English Edition, 2002-)* Vol. 62, No. 1, pp. 15-31
- Tonon, G. ed. (2015). *Qualitative Studies in Quality of Life. Methodology and Practice*. New York, London: Springer-Heilderberg-Dordrecht, Social Indicators Research Series 55. ISBN 978-3-319-13778-0.
- Ullah, F. y Finch, C.F. (2013) Applications of functional data analysis: a sistematic review. *BMC Medical Research Methodology*. 13(43) 1-12.
- Varsavsky, O. (1969) *Ciencia, política y cientificismo*. Buenos Aires, CEAL.
- Wilmoth, J. (1993). Computational methods for fitting and extrapolating the Lee–Carter model of mortality change technical report department of demography. University of California, Berkeley.

Xie, Y. (2000) "Demography: Past, Present, and Future". *Journal of the American Statistical Association* 3r150093.

9. Anexo (Programación y Salidas)

Programación

Edades Simples

```
years <- as.vector(1980:2013)
ages <- c(0:99)
library(demography)

#Capital
pop<-Mat_Int_2[,1:34]
def<-d_Cap2
rate<- def/pop
rateCap<-rate
Mortalidad_Cap <- demogdata(rate, pop, ages, years, "mortality",
"Capital", "total")
Mortalidad_Cap$lambda <- 0 # Use logs for transformation
length(which(Mortalidad_Cap$rate$total==0))
length(which(Mortalidad_Cap$rate$total<0))

win.graph(width = 8, height = 6, pointsize = 10)
plot(Mortalidad_Cap)
str(Mortalidad_Cap)

mort_Cap.sm <- smooth.demogdata(Mortalidad_Cap,age.grid=0:99,
obs.var="theoretical")
plot(mort_Cap.sm,transform=TRUE, main="", xlab="Edad",
ylab="log(m(x))", plotlegend=T,legendpos = "bottomright")
str(mort_Cap.sm)
length(which(mort_Cap.sm$rate$total==0))
length(which(mort_Cap.sm$rate$total<0))

# Estimación y pronóstico usando ARIMA
Cap_Mort.fit <- fdm(mort_Cap.sm,order=6,method="classical")
summary(Cap_Mort.fit)
```

```

plot(Cap_Mort.fit,components=3)
Cap_Mort.fcast1<-
forecast(Cap_Mort.fit,h=30,level=80,method="arima",adjust=FALSE)
plot(residuals.fdm(Cap_Mort.fit),xlab="Año", ylab="Edad", type=
"filled.contour" )
# Pronóstico
Cap_fdm_fcast
forecast(Cap_Mort.fit,h=37,method="arima",level=80,adjust=FALSE)
rate_Cap_fdm<-
(cbind(Cap_fdm_fcast$rate$total*1000,Cap_fdm_fcast$rate$lower*1000
,Cap_fdm_fcast$rate$upper*1000))
format(rate_Cap_fdm,digits=2)
write.table(rate_Cap_fdm, file =
"rate_Cap_fdm.csv",row.names=TRUE)
e0_t_mdf_Cap<-e0(Cap_fdm_fcast, PI=T, h=37)
write.table(e0_t_mdf_Cap, file =
"e0_t_mdf_Cap.csv",row.names=TRUE)

# Estimación LC
Cap_Mort_lca <- lca(mort_Cap.sm,max.age=99,interpolate=TRUE)
summary(Cap_Mort_lca)
plot(Cap_Mort_lca)
plot(residuals(Cap_Mort_lca),xlab="Año", ylab="Edad", type=
"filled.contour" )

# Pronóstico
Cap_lca_fcast <-
forecast(Cap_Mort_lca,h=37,method="arima",level=80,adjust=FALSE)
rate_Cap_lca<-
(cbind(Cap_lca_fcast$rate$total*1000,Cap_lca_fcast$rate$lower*1000
,Cap_lca_fcast$rate$upper*1000))
format(rate_Cap_lca,digits=2)
write.table(rate_Cap_lca, file =
"rate_Cap_lca.csv",row.names=TRUE)
e0_t_lca_Cap<-e0(Cap_lca_fcast, PI=T, h=37)

```

```

write.table(e0_t_lca_Cap, file =
"e0_t_lca_Cap.csv", row.names=TRUE)
dim(rate_Cap_lca)

#Río Cuarto
Mortalidad_RC <- demogdata(rate, pop, ages, years, "mortality",
"Río Cuarto", "total")
Mortalidad_RC$lambda <- 0 # Use logs for transformation

win.graph(width = 8, height = 6, pointsize = 10)
plot(Mortalidad_RC, ylim=c(-9,1))
limit<- (-10)
table(lapply(as.matrix(Mortalidad_RC$rate$total), log)=="-Inf")
mort_RC.sm <- smooth.demogdata(Mortalidad_RC, age.grid=0:99,
obs.var="theoretical")
plot(mort_RC.sm, transform=TRUE, main="", xlab="Edad",
ylab="log(m(x))", plotlegend=T, legendpos = "bottomright")
length(which(mort_RC.sm$rate$total==0))
length(which(mort_RC.sm$rate$total<0))

# Estimación y pronóstico ARIMA
RC_Mort.fit <- fdm(mort_RC.sm, order=6, method="classical")
summary(RC_Mort.fit)
plot(RC_Mort.fit, components=3)
RC_Mort.fcast1 <-
forecast(RC_Mort.fit, h=30, level=80, method="arima", adjust=FALSE, all
owdrift=FALSE)
plot(residuals.fdm(RC_Mort.fit), xlab="Año", ylab="Edad", type=
"filled.contour" )

# Pronóstico
RC_fdm_fcast <-
forecast(RC_Mort.fit, h=37, method="arima", level=80, adjust=FALSE)

```

```

rate_RC_fdm<-
(cbind(RC_fdm_fcast$rate$total*1000,RC_fdm_fcast$rate$lower*1000,R
C_fdm_fcast$rate$upper*1000))
format(rate_RC_fdm,digits=2)
write.table(rate_RC_fdm, file = "rate_RC_fdm.csv",row.names=TRUE)
e0_t_mdf_RC<-e0(RC_fdm_fcast, PI=T, h=37)
write.table(e0_t_mdf_RC, file = "e0_t_mdf_RC.csv",row.names=TRUE)

# Estimación usando LC
RC_Mort_lca <- lca(mort_RC.sm,max.age=99,interpolate=TRUE)
summary(RC_Mort_lca)
win.graph(width = 8, height = 6, pointsize = 10)
plot(RC_Mort_lca)
plot(residuals.fdm(RC_Mort_lca),xlab="Año", ylab="Edad", type=
"filled.contour" )

# Pronóstico
RC_lca_fcast <-
forecast(RC_Mort_lca,h=37,method="arima",level=80,adjust=FALSE)
rate_RC_lca<-
(cbind(RC_lca_fcast$rate$total*1000,RC_lca_fcast$rate$lower*1000,R
C_lca_fcast$rate$upper*1000))
format(rate_RC_lca,digits=2)
write.table(rate_RC_lca, file = "rate_RC_lca.csv",row.names=TRUE)
e0_t_lca_RC<-e0(RC_lca_fcast, PI=T, h=37)
write.table(e0_t_lca_RC, file = "e0_t_lca_RC.csv",row.names=TRUE)

#Sobremonte
Mortalidad_S <- demogdata(rate, pop, ages, years, "mortality",
"Sobremonte", "total")
Mortalidad_S$lambda <- 0 # Use logs for transformation

win.graph(width = 8, height = 6, pointsize = 10)
plot(Mortalidad_S)

```

```
mort_S.sm <- smooth.demogdata(Mortalidad_S,age.grid=0:99,  
obs.var="theoretical")
```

```
write.table(  
rbind(  
(colSums(defCap[1:15,])/colSums(popCap[1:15,]))*1000,  
(colSums(defCap[16:65,])/colSums(popCap[16:65,]))*1000,  
(colSums(defCap[66:99,])/colSums(popCap[66:99,]))*1000,  
  
(colSums(defRC[1:15,])/colSums(popRC[1:15,]))*1000,  
(colSums(defRC[16:65,])/colSums(popRC[16:65,]))*1000,  
(colSums(defRC[66:99,])/colSums(popRC[66:99,]))*1000,  
  
(colSums(defS[1:15,])/colSums(popS[1:15,]))*1000,  
(colSums(defS[16:65,])/colSums(popS[16:65,]))*1000,  
(colSums(defS[66:99,])/colSums(popS[66:99,]))*1000),  
file="tasa_gg.csv",row.names=T)
```

Salidas

```
> summary(Cap_Mort.fit)  
Functional demographic model  
  
Call: fdm(data = mort_Cap.sm, order = 6, method = "classical")
```

```
Region: Capital  
Data type: mortality  
Years in fit: 1980 - 2013  
Ages in fit: 0 - 99
```

```
Order: 6  
Percentage variation due to basis functions:  
73.0% 9.6% 5.8% 4.0% 1.8% 1.8%
```

```
Average across ages:  
MAPE  
0.00657
```

```
> summary(Cap_Mort_lca)  
Lee-Carter analysis
```

```
Call: lca(data = mort_Cap.sm, max.age = 99, interpolate = TRUE)
```

```
Region: Capital  
Years in fit: 1980 - 2013  
Ages in fit: 0 - 99
```

```
Percentage variation explained: 72.9%
```

```
ERROR MEASURES BASED ON MORTALITY RATES
```

```
Average across ages:  
MAPE  
0.06225
```

```
ERROR MEASURES BASED ON LOG MORTALITY RATES
```

```
Average across ages:  
MAPE  
0.01833
```

```
> summary(RC_Mort.fit)  
Functional demographic model
```

```
Call: fdm(data = mort_RC.sm, order = 6, method = "classical")
```

```
Region: Río Cuarto  
Data type: mortality  
Years in fit: 1980 - 2013  
Ages in fit: 0 - 99
```

```
Order: 6  
Percentage variation due to basis functions: 54.5% 18.3% 12.9%  
4.9% 3.5% 2.5%
```

```
Average across ages:  
MAPE  
0.00670
```

```
> summary(RC_Mort_lca)  
Lee-Carter analysis
```

```
Call: lca(data = mort_RC.sm, max.age = 99, interpolate = TRUE)
```

```
Region: Río Cuarto
```

Years in fit: 1980 - 2013

Ages in fit: 0 - 99

Percentage variation explained: 53.9%

ERROR MEASURES BASED ON MORTALITY RATES

Average across ages:

MAPE

0.09104

ERROR MEASURES BASED ON LOG MORTALITY RATES

Average across ages:

MAPE

0.02653

Edades Agrupadas

```
> summary(Cap_Mort.fit)
Functional demographic model

Call: fdm(data = mort_Cap.sm, order = 6, method = "classical")

Region: Capital
Data type: mortality
Years in fit: 1980 - 2013
Ages in fit: 0 - 99

Order: 6
Percentage variation due to basis functions: 59.2% 16.9% 10.1%
4.5% 2.4% 1.8%

Average across ages:
MAPE
0.00727

> summary(Cap_Mort_lca)
Lee-Carter analysis

Call: lca(data = mort_Cap.sm, max.age = 99, interpolate = TRUE)

Region: Capital
Years in fit: 1980 - 2013
Ages in fit: 0 - 99

Percentage variation explained: 59%

ERROR MEASURES BASED ON MORTALITY RATES

Average across ages:
MAPE
0.07771

ERROR MEASURES BASED ON LOG MORTALITY RATES

Average across ages:
MAPE
0.02569

> summary(RC_Mort.fit)
Functional demographic model
```

```
Call: fdm(data = mort_RC.sm, order = 6, method = "classical")
```

```
Region: Río Cuarto  
Data type: mortality  
Years in fit: 1980 - 2013  
Ages in fit: 0 - 99
```

```
Order: 6  
Percentage variation due to basis functions: 50.1% 16.7% 12.7%  
8.7% 3.8% 3.2%
```

```
Average across ages:  
MAPE  
0.00979
```

```
> summary(RC_Mort_lca)  
Lee-Carter analysis
```

```
Call: lca(data = mort_RC.sm, max.age = 99, interpolate = TRUE)
```

```
Region: Río Cuarto  
Years in fit: 1980 - 2013  
Ages in fit: 0 - 99
```

```
Percentage variation explained: 49.5%
```

```
ERROR MEASURES BASED ON MORTALITY RATES
```

```
Average across ages:  
MAPE  
0.11395
```

```
ERROR MEASURES BASED ON LOG MORTALITY RATES
```

```
Average across ages:  
MAPE  
0.03879
```