

# Estimación de los nacimientos en Costa Rica

César Gamboa Sanabria<sup>\*\*,a</sup>, Stefany Matarrita Muñoz<sup>\*\*,a</sup>

<sup>a</sup>*Escuela de Estadística, Universidad de Costa Rica*

---

\*email: [cesar.gamboasanabria@ucr.ac.cr](mailto:cesar.gamboasanabria@ucr.ac.cr) | LinkedIn: <https://linkedin.com/in/cgamboasanabria> | Github: <https://github.com/cgamboasanabria>  
\*\*email: [stefmatarrita@hotmail.com](mailto:stefmatarrita@hotmail.com)

## RESUMEN

El monitoreo de los procesos demográficos es vital debido a su estrecha relación con el desarrollo económico y social de un país. Desde hace varios años, numerosos estudios han abordado el pronóstico de los nacimientos, pues su disminución junto con el aumento en la esperanza de vida causan el envejecimiento de la población. Este fenómeno, obliga a los gobiernos a generar políticas públicas dirigidas a los sistemas de seguridad social y servicios de salud, pero para ello se debe contar con estimaciones precisas del futuro demográfico. Dado lo anterior se pretende analizar los nacimientos mensuales en Costa Rica desde enero de 1988 hasta diciembre 2017, con datos provenientes del Instituto Nacional de Estadística y Censos, con el fin de pronosticar hasta el 2022. Los nacimientos se estimaron con 4 modelos: la regresión lineal, el suavizamiento exponencial, el ARIMA y las redes neuronales. La selección del mejor dentro de ellos se escogió mediante las medidas de rendimiento (MAE, RMSE, MAPE, MASE) y criterios de información (AIC, BIC y AICc). Se encontró que el  $ARIMA(3, 1, 4)(1, 1, 3)$  fue el modelo que obtuvo mejor desempeño. El pronóstico revela que el nivel medio del número de nacimientos seguirá decreciendo. Por ende si los nacimientos siguen disminuyendo tal y como indica el modelo, y la esperanza de vida sigue en aumento como se ha observado en la última década, eventualmente la población de Costa Rica envejecerá y los jóvenes serán menos. Por tanto se invita a las autoridades a tomar diferentes medidas como: la inclusión del adulto mayor en el mercado laboral, mejoría de servicios de salud y política migratoria, promoción de ahorros en la población, así como inversión en investigaciones sobre envejecimiento que dirijan la toma decisiones.

**Palabras clave:** nacimientos, envejecimiento población, series de tiempo, regresión lineal, suavizamiento exponencial, ARIMA, redes neuronales.

## 1. INTRODUCCIÓN

Durante las Primeras Jornadas de Investigación del Centro Centroamericano de Población de la Universidad de Costa Rica (CCP-UCR), se llegó a concluir que dado el panorama demográfico actual es muy improbable que Costa Rica llegue a tener 6 millones de habitantes en un futuro cercano, pues para ello se necesitaría una tasa de fecundidad de 1,8 (Mayorga 2018). La tendencia de los indicadores demográficos indica que en los próximos años la población en Costa Rica envejecerá y los jóvenes serán menos, es decir, que se produzca el envejecimiento acelerado de la población (Domínguez 2018).

Para el 2017 la tasa de fecundidad registrada fue de 1,67 hijos por mujer, tasa ubicada por debajo de la tasa de fecundidad de remplazo (2,1 hijos por mujer) (Ávalo 2018; Domínguez 2018; Mayorga 2018), esto dificulta que cada generación sea remplazada por otra del mismo tamaño; según datos del Instituto Nacional de Estadística y Censo (INEC) la disminución de la tasa global de fecundidad ha venido ocurriendo desde los años setenta.

La mortalidad es otro factor al que se le atribuye el envejecimiento de la población, desde hace varios años, la tasa de mortalidad ha venido disminuyendo, para el año 2017 la tasa de mortalidad general fue de 4,97 defunciones por cada 1000 habitantes, lo cual se explica en parte, por el descenso en las defunciones de jóvenes y menores de un año, siendo la tasa de mortalidad infantil en el 2017 de 7,9. Como consecuencia, la esperanza de vida al nacer ha venido en ascenso, siendo de 77 años para los hombres y 82 para las mujeres, en el 2017 (Ávalo 2018).

Este panorama demográfico conlleva implicaciones sociales y económicas importantes. Cuando un país llega a tener una pirámide poblacional inversa, se cuenta con menos personas para activar el desarrollo económico del país, lo cual provoca que los sistemas de pensiones y los servicios de salud presenten serias dificultades para sostenerse, pues los ingresos del gobierno se reducen en momentos en que se requiere una mayor inversión en servicios, principalmente los que demandan los adultos mayores (Población 2008)(Redondo 2013).

Si a una sociedad envejecida se le adiciona una económico débil, que es el caso actual de Costa Rica, se dan las condiciones perfectas para una crisis económica, pues los gastos en que se deben incurrir para la atención de adultos mayores pueden agravar la deuda del gobierno, como fue el caso de Portugal en el 2010-2014(Albuquerque 2018).

Es por ello que en este estudio se analiza uno de las principales causas del envejecimiento poblacional, como lo son los nacimientos en Costa Rica, con el fin de pronosticar para los próximos cinco años cuál es el comportamiento esperado, y brindar recomendaciones en cuanto al tipo de política pública que se debe promover.

## **2. MARCO DE LA CUESTIÓN**

El envejecimiento de la población es el crecimiento del porcentaje de adultos mayores debido a la disminución del porcentaje de menores (Financiero 2018). Este fenómeno ha sido ampliamente analizado,

debido a su importancia en la creación de política pública. Según las convenciones demográficas, un país con 10 y 19 % de adultos mayores se cataloga como envejecido, si está entre 20 y 29 %, como muy envejecido; y mayor del 30 %, híper envejecido (Fernández 2016).

La disminución del número de hijos que tienen las mujeres y el aumento en la esperanza de vida modifican de manera importante la estructura por edad de la población (Población 2008). Por lo general, en la naturaleza ningún extremo es deseado, se sabe que una alta fertilidad sostenida en el tiempo puede colapsar los sistemas educativos y el mercado laboral, mientras que caso contrario, se colapsarían los servicios de salud y pensiones (Tiit 2018).

Es bien conocido que todos los procesos demográficos son interdependientes, y como ya se mencionó, existe una fuerte conexión con el desarrollo general y el nivel económico de un país, allí es donde radica la importancia de estudiar estos procesos (Tiit 2018).

Al disminuir la mortalidad las personas sobreviven más tiempo, esto incide en un aumento de la población de 65 y más años (Ver Anexo 1). En el caso de Costa Rica, el decremento de la tasa de mortalidad general observada en los últimos años (Ver Anexo 2), ha sido producto de la disminución de la mortalidad de las edades jóvenes y en particular de la mortalidad infantil, esto gracias a la mejora en los servicios de salud, esta es una de las razones por las que se vislumbra una mayor esperanza de vida al nacer (Ver Anexo 3) (Población 2008).

Por otro lado, la natalidad depende directamente del número promedio de hijos por mujer. En la década de 1950, el panorama nacional distaba mucho del actual, fue en este período cuando ocurrió el primer *Baby Boom*. En ese entonces, el número anual de nacimientos pasó de ser 34 mil en 1949 a 63 mil en 1963 (Bixby.L. 1997), esto sucedió por el aumento en el número de hijos promedio por mujer, que llegó a ser de 7,1 a inicios de la década de los setenta (Población 2008). En años posteriores, cuando las personas nacidas en esa época estaban en edades reproductivas, hubo otra expansión de nacimientos, pero a diferencia del primer *Baby Boom*, el número de hijos promedio por mujer era menor, pasando de 7,1 en 1961 a 3,6 en 1976, es decir la expansión de nacimientos se atribuyó a una mayor cantidad de mujeres con hijos (Población 2008).

Después de los cambios abruptos en los nacimientos de ese período, el patrón que se ha observado es una disminución sin pausa, datos oficiales confirman que entre el 2008 y el 2017, la tasa de natalidad disminuyó un 9 % (Financiero 2018).

Desde 1997, ya los expertos en la materia habían advertido que las oleadas demográficas históricas iban a colapsar los sistemas pensiones y los servicios de salud entre los años 2010 a 2023 (Bixby.L. 1997). Contrario a lo que se pronosticaban en 1997, actualmente se prevé que en los años venideros habrá una menor presión sobre el mercado laboral y vivienda, esto a consecuencia de la baja tasa global de fecundidad registrada en los últimos años (Ver Anexo 4) (Bixby.L. 1997), siendo la tasa global de fecundidad nacional para el 2017 una de las más bajas del continente (1,67 hijos por mujer) (Mayorga 2018).

Entre las causas directas de la baja de la fecundidad en Costa Rica, el Dr. Luis Rosero Bixby, señala que la caída en la maternidad adolescente y la Ley de Paternidad Responsable, aplicada desde el 2000 fueron factores claves (Mayorga 2018). Otros especialistas apuntan a que las parejas posponen el nacimiento del primer hijo para alcanzar una mayor preparación académica (Domínguez 2018), además el acceso, conocimiento y uso de métodos de planificación, así como el deseo expreso de tener menos hijos son factores que pesan mucho en el panorama demográfico actual (Financiero 2018) (Shitan 2015).

Como se mencionó anteriormente, la estructura por edad de la población presenta una estrecha vinculación con el desarrollo económico. Dentro de una población se pueden distinguir: la población de menores, población en edad de trabajar y los adultos mayores, dependiendo de la etapa en la que se encuentren las personas, su capacidad de producción y de consumo varía, a este proceso se lo conoce como la **economía del ciclo vital**. En la primera fase de este ciclo las personas crecen, estudian, viven y dependen de los miembros mayores del hogar, en una segunda etapa se da la incorporación al sistema productivo del país, y en la última fase, las personas subsisten de ahorros, pensiones, apoyo de la familia o transferencias del gobierno (Población 2008).

Como consecuencia de los pocos nacimientos en la última década, dentro de unos años, la población en edad de trabajar podría ser insuficiente, es decir no se percibirá el beneficio del *bono demográfico*, pues el volumen de los dependientes (niños y ancianos) será mayor que la población en edad de trabajar. Lo anterior impacta directamente la contribución fiscal, ya que la base de contribuyentes efectivos será menor que la de receptores efectivos de las transferencias del gobierno (Población 2008). Sin embargo este panorama puede ser distinto si en la actualidad las personas ahorran y cotizan al sistema de pensiones para subsistir en su vejez, pues pese al cese de los ingresos laborales, estas personas podrán hacerle frente a sus gastos sin depender directamente de las transferencias del gobierno (Población

2008).

Las previsiones económicas que tomen las personas no son el único factor que podría variar el ciclo de la economía vital, la planificación y oferta de buenos programas de salud, educación, empleo y pensiones son claves para hacer frente a este fenómeno, pues el éxito de las personas durante sus años de preparación y de participación en el sistema productivo pueden impactar positivamente la economía nacional futura (Población 2008).

El envejecimiento de la población no es único de Costa Rica, a nivel internacional, este fenómeno se están produciendo sin distinguir entre países desarrollados y en desarrollo (M 2016). En el caso de América Latina, para el año 2016, no se registraron países hiper envejecidos ni muy envejecidos, pero sí envejecidos, entre esos países se encontraban: Uruguay, Cuba, Chile, Argentina, Costa Rica, Brasil, El Salvador, Panamá, Colombia y Perú (M 2016).

En los otros continentes, el envejecimiento es mayor, particularmente en los países asiáticos, donde el porcentaje de la población que tiene 65 años o más, pasó de 15 % a 20 % en aproximadamente 15 años, mientras que en Europa y Estados Unidos este aumento requirió en promedio 26 y 50 años respectivamente, es decir los países asiáticos han tenido menos tiempo para adaptar su economía a su sociedad envejecida (Financiero 2017).

Para el 2050, Japón (42,5 %) seguirá encabezando la lista de países híper envejecidos, siguiéndole Corea del Sur (41,5 %), España (41,4 %), Portugal (41,2 %), Hong Kong (40,9 %), Grecia (40,8 %), Italia (40,7 %), Bosnia y Herzegovina (40,5 %), Singapur (40,4 %) y Cuba (39,7 %) (M 2016). Para ese año se pronostica que en Costa Rica este porcentaje será de 21 %, pasando de ser un país envejecido a muy envejecido (Mayorga 2018).

Bajo el contexto anterior los analistas incitan a una adecuada planificación en los diferentes ámbitos. De hecho uno de los más reconocidos demógrafos, Rosero Bixby, señala que en el 2080 solo habrá dos personas trabajadoras por cada pensionado, es decir, para entonces, las personas trabajadoras deberán aportar cerca del 50 % de su salario para mantener el sistema de pensiones (Mayorga 2018).

Ante este panorama muchos de los países mencionados han creado leyes y políticas preventivas a nivel económico y de seguridad social, pero pocos han hecho hincapié en que los sistemas tradicionales de apoyo familiar están en decadencia, principalmente porque las familias son cada vez más pequeñas,

acentuándose el problema con la movilización de los posibles cuidadores a las ciudades, en busca de oportunidades educativas y laborales. Esta es una de las razones por la que se ha encontrado que el número de personas adultas mayores viven solas (M 2016).

A nivel nacional el enfoque de la política pública para atacar la baja de nacimientos, ha sido la generación de política pública dirigida a mejorar la calidad de vida de las personas adultas mayor, como lo es la *Política Pública sobre Envejecimiento y Vejez en Costa Rica 2011-2021*, la cual cuenta con cinco líneas estratégicas para su aplicación: I. Línea estratégica de protección social, ingresos y prevención de la pobreza, II. Línea estratégica de abandono, abuso y maltrato en contra de las personas adultas mayores, III. Línea estratégica de participación social e integración intergeneracional, IV Línea estratégica de consolidación de derechos y V Línea estratégica de salud integral (M 2016).

Sin embargo, en otros países como Rusia la política está enfocada en promover la fertilidad, brindando incentivos por tener un segundo o más hijos; pese a esto un estudio ha encontrado mediante modelos probabilísticos que las políticas pronatalistas no tendrán un impacto significativo en la dinámica futura de la tasa de natalidad como lo tuvo en el año 2007, esto se debe a que la crisis actual y los cambios de coyunturales de este país no se conciben como condiciones propicias para tener hijos. Además los investigadores acotan que se debe hacer frente a la crisis demográfica mediante la promoción de la inmigración (Chichkanov 2018) (Aragóñez 2017).

Dado el contexto anterior es evidente que el pronóstico de los nacimientos es crucial para una toma decisiones sobre el futuro del país, por ello en diferentes estudios de varias regiones del mundo se ha modelado tanto los nacimientos como la tasa de fertilidad global mediante diferentes técnicas de series de tiempo (Essuman 2017)(Baker 2017)(Carter 1986)(Shitan 2015).

En un estudio de los nacimientos de la franja de Gaza, entre los modelos comparados, las redes neuronales fueron las que brindaba estimaciones más precisas, siendo más eficientes que los ARIMA estacionales(Baker 2017). Para el pronóstico de la tasa global de fecundidad, en un estudio en Malasia, el modelo más apropiado para el pronóstico fue el ARAR (Shitan 2015), en Estados Unidos se utilizó la técnica de Box-Jenkins para pronosticar la fertilidad marital, nupcialidad, nacimientos y matrimonios, y mencionan que este método presenta importantes ventajas en las estimaciones de eventos como de tasas (Carter 1986). Lo mismo sucedió en un análisis de series de tiempo de los nacimientos , en Ghana, donde un ARIMA estacional fue el presentó mejores medidas de ajuste(Essuman 2017).

Bajo el contexto anterior se pretende analizar los nacimientos mensuales en Costa Rica desde enero de 1988 hasta diciembre 2017, mediante diferentes modelos de series de tiempo, utilizando datos provenientes del Instituto Nacional de Estadística y Censos, con el fin de pronosticar con un horizonte de cinco años.

### 3. METODOLOGÍA

Los datos utilizados para esta investigación fueron obtenidos en la Unidad de Estadísticas Demográficas del INEC<sup>1</sup>, y corresponden a información mensual de la cantidad de nacimientos en Costa Rica desde Enero del año 1988 hasta Diciembre de 2017. Estos datos serán procesados mediante el lenguaje R<sup>2</sup> y su interfaz RStudio<sup>3</sup>. Para la manipulación de los datos se utilizarán los paquetes de *R* *readxl*(Wickham and Bryan 2018), *tidyr*(Wickham and Henry 2018), y *dplyr*(Wickham et al. 2018), mientras que para el manejo de datos temporales se utilizaron los paquetes *lubridate* (Grolemund and Wickham 2011), *astsa* (Stoffer 2017) y *forecast*(Hyndman R 2018) para realizar los pronósticos. En cuanto a la visualización, los paquetes *ggplot2* (Wickham 2016), *kableExtra* (Zhu 2018) y *ggseas*(Ellis 2018) serán empleados para generar distintos tipos de gráficos.

Con los datos de los registros administrativos procesados se realizará un análisis descriptivo de la serie para posteriormente realizar una descomposición de la misma. La serie completa será particionada en dos conjuntos, uno servirá de entrenamiento para la construcción de los modelos para ajustar la serie y otro servirá de validación una vez que el modelo sea generado, esto con el fin de conocer qué tan bien se ajustan las estimaciones a los datos reales; siendo los datos correspondientes al periodo Enero 1988 – Diciembre 2016 el conjunto de entrenamiento y de Enero 2017 – Diciembre 2017 para la validación. Se estimarán distintas variantes de los modelos de regresión, de suavizamiento exponencial, ARIMA y redes neuronales.

En el caso de los modelos de regresión, se estimará uno solamente con la tendencia, otro con la tendencia y la estacionalidad, y un tercero que consiste en una variante exponencial, todos ellos estimados mediante la máxima verosimilitud maximizando los coeficientes respecto a los parámetros de la función expresada

---

<sup>1</sup><http://www.inec.go.cr/>

<sup>2</sup><https://cran.r-project.org/>

<sup>3</sup><https://www.rstudio.com/>



en (1).

$$\ln(L) = -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}(Y - X\beta)'(Y - X\beta) \quad (1)$$

Para los modelos de suavizamiento exponencial, se estimarán distintas variantes considerando errores aditivos y multiplicativos, sin tendencia y con tendencia aditiva y multiplicativa, y sin estacionalidad y con estacionalidad aditiva y multiplicativa, todo esto además de considerar el amortiguamiento.

La estimación de los modelos ARIMA se abordará de tres maneras: 1) identificando los modelos de forma manual tanto en su parte estacional como en la no estacional, 2) de forma automática mediante la función `auto.arima()` del paquete `forecast` y 3) aplicando un ARIMA con intervención.

La estimación de la red neuronal se hará mediante la función `nnetar()` del paquete `forecast`.

Una vez estimados estos modelos, se seleccionarán los mejores de cada clase, es decir, el mejor modelo de regresión, el mejor de suavizamiento exponencial, el mejor ARIMA entre los identificados, el automático y el ARIMA con intervención; y la red neuronal más apropiada. Dicha selección se hará analizando los criterios de información indicados en las ecuaciones (2), (3) y (4) donde los valores más bajos indican cuál es el mejor modelo, mientras que las medidas de rendimientos utilizadas se especifican en las ecuaciones (5), (6), (7) y (8) donde las más cercanas a cero sirven para indicar cuál modelo es el más adecuado,

$$AIC = -2\log L(\hat{\theta}) + 2k \quad (2)$$

$$AICc = -2\log L(\hat{\theta}) + 2k + \frac{2k + 1}{n - k - 1} \quad (3)$$

$$BIC = -2\log L(\hat{\theta}) + k \cdot \log(n) \quad (4)$$

$$MAE = n^{-1} \sum_{t=1}^n |y_t - f_t| \quad (5)$$

$$RMSE = \sqrt{n^{-1} \sum_{t=1}^n (y_t - f_t)^2} \quad (6)$$

$$MAPE = 100n^{-1} \sum_{t=1}^n \frac{|y_t - f_t|}{|y_t|} \quad (7)$$

$$MASE = n^{-1} \sum_{t=1}^n \frac{|y_t - f_t|}{q}; q = (n - m)^{-1} \sum_{t=2}^n |y_t - y_{t-1}| \quad (8)$$

donde  $k$  es el número de parámetros y  $n$  es el número de datos.

Finalmente se escogerá el modelo más adecuado en relación a la cantidad de nacimientos para posteriormente realizar un pronóstico de los posibles valores que alcanzará esta variable en los próximos dos años, así como una contextualización del fenómeno con respecto a la actualidad.

#### 4. RESULTADOS

##### OJO: PONER PÁRRAFO INTRODUCTORIO

Dado que las certificaciones de nacimientos se realizan diariamente, conviene analizar su comportamiento de manera mensual de una manera más general. El gráfico 1 muestra picos y valles pronunciados a lo largo de todo el periodo. A modo de visualización, se ajustó un suavizamiento de Loess, o regresión local, para buscar señales de tendencia y concavidad en los datos temporales. La línea roja punteada se ubica aproximadamente en el mes de Abril del año 2001, pues a partir de ese punto la regresión local muestra un ligero cambio en la concavidad. Una primera impresión muestra indicios de estacionalidad.

Gráfico 1: Nacimientos en Costa Rica para el periodo 1989 – 2017

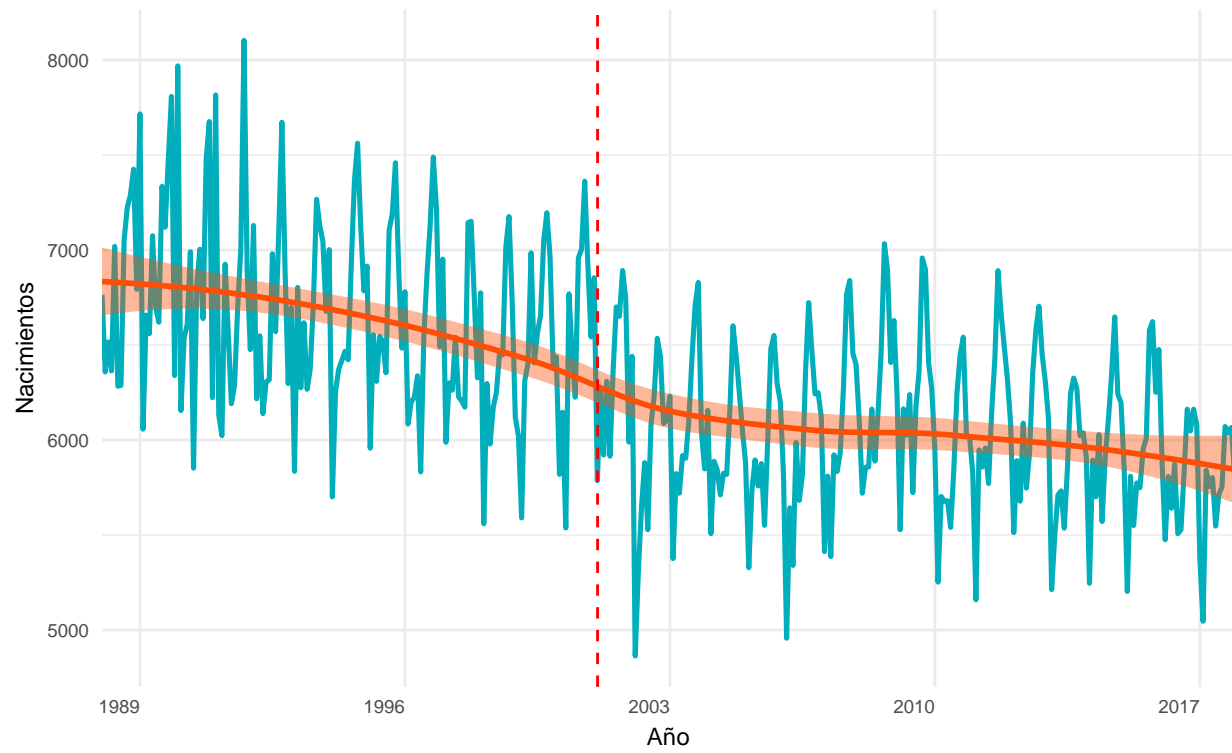
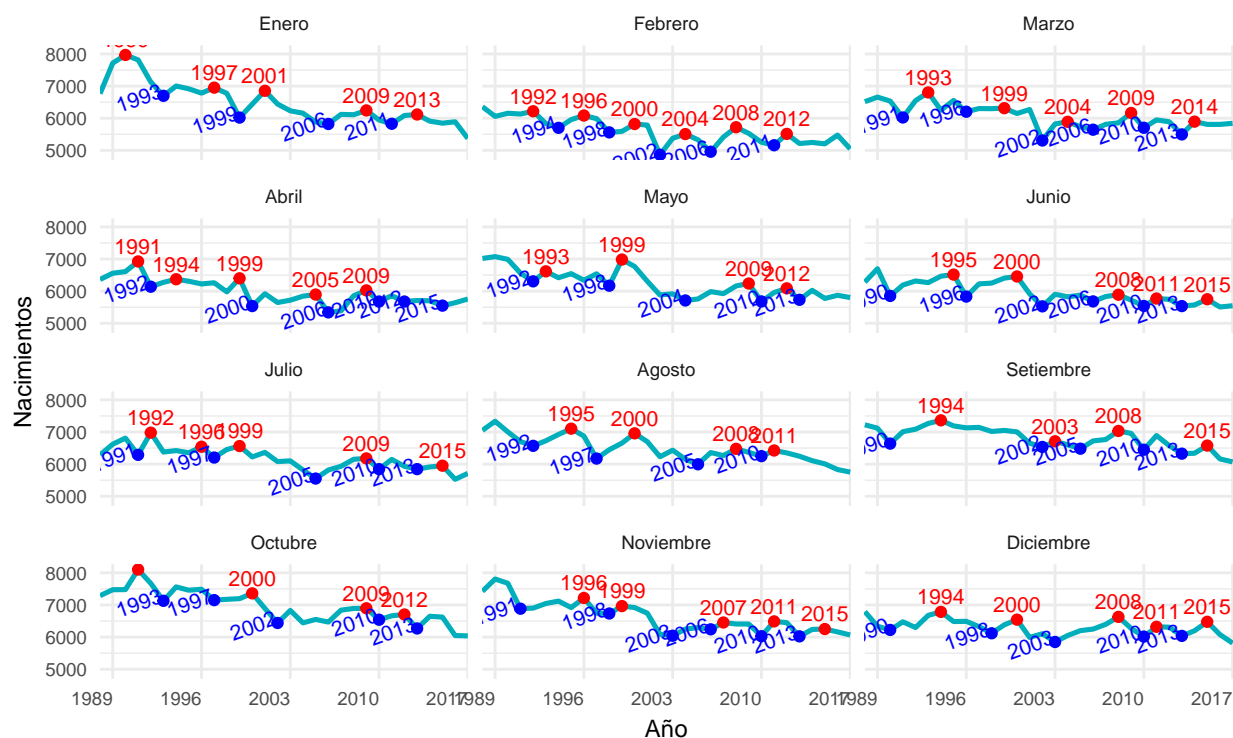


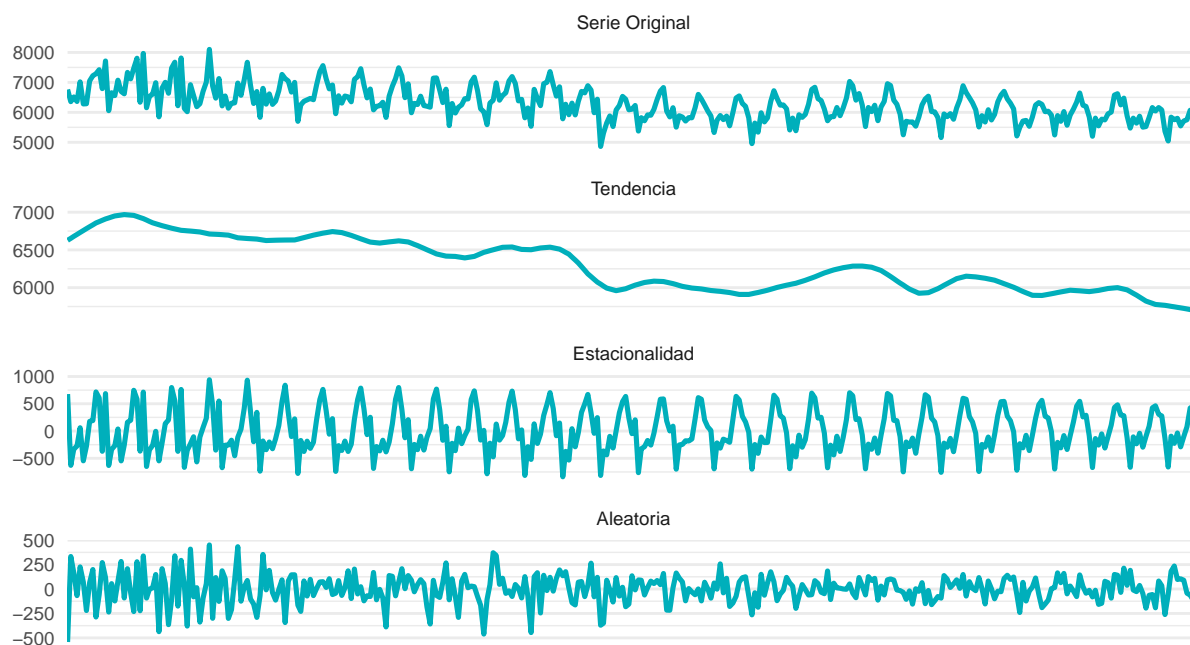
Gráfico 2: Nacimientos en Costa Rica en el periodo 1989 – 2017 según mes



Fuente: UED-INEC

Mediante un análisis visual, el gráfico 2 parece respaldar el supuesto de que los nacimientos poseen un efecto estacional, pues de todos los periodos, los meses de Setiembre y Octubre poseen un nivel medio más alto que todos los demás meses; además, el nivel parece ser más bajo en cada mes a partir del 2001. Este hecho coincide con lo observado en el gráfico 1.

Gráfico 3: Descomposición de la serie de Nacimientos en Costa Rica en el periodo 1989 – 2017



Fuente: UED-INEC

La descomposición de la serie se hará de forma aditiva debido a que en el gráfico 1 no se observan grandes cambios en la variabilidad a lo largo del tiempo. El gráfico 3 muestra que la tendencia descende a lo largo del tiempo, especialmente a mitad del periodo, mientras que parece haber estacionalidad en los últimos meses del año. Además, el componente aleatorio muestra como los errores no son constantes a lo largo de todo el período.

Tras tener un panorama más claro del comportamiento de la serie mediante el análisis descriptivo anterior, es importante realizar las estimaciones de los modelos. Para la estimación de los mismos y todos los aspectos relacionados, se dividió la base de datos en dos subconjuntos: 1) la base de entrenamiento son los datos desde el año 1989 y hasta el 2016 inclusive, y 2) la base de validación, que corresponde a los datos del año 2017.

La primera estimación hecha es la que corresponde a los modelos de regresión. Las tres variantes descritas en el apartado metodológico de este artículo son calculadas para luego contrastarlos mediante las medidas de rendimiento y criterios de información, los resultados pueden verse en cuadro 1 de la sección de anexos. El mejor modelo en este caso es el de variante exponencial, expresado mediante la

ecuación (9), pues es el que cuenta con mejores medidas en el conjunto de entrenamiento y en el de validación.

$$\begin{aligned} \log(\hat{y}_t) = & 8,8558 - 0,0005t - 0,1428s_2 - 0,0639s_3 - 0,0796s_4 - 0,0345s_5 - 0,0784s_6 \\ & - 0,0443s_7 + 0,0074s_8 + 0,0565s_9 + 0,0771s_{10} + 0,0299s_{11} - 0,0199s_{12} \end{aligned} \quad (9)$$

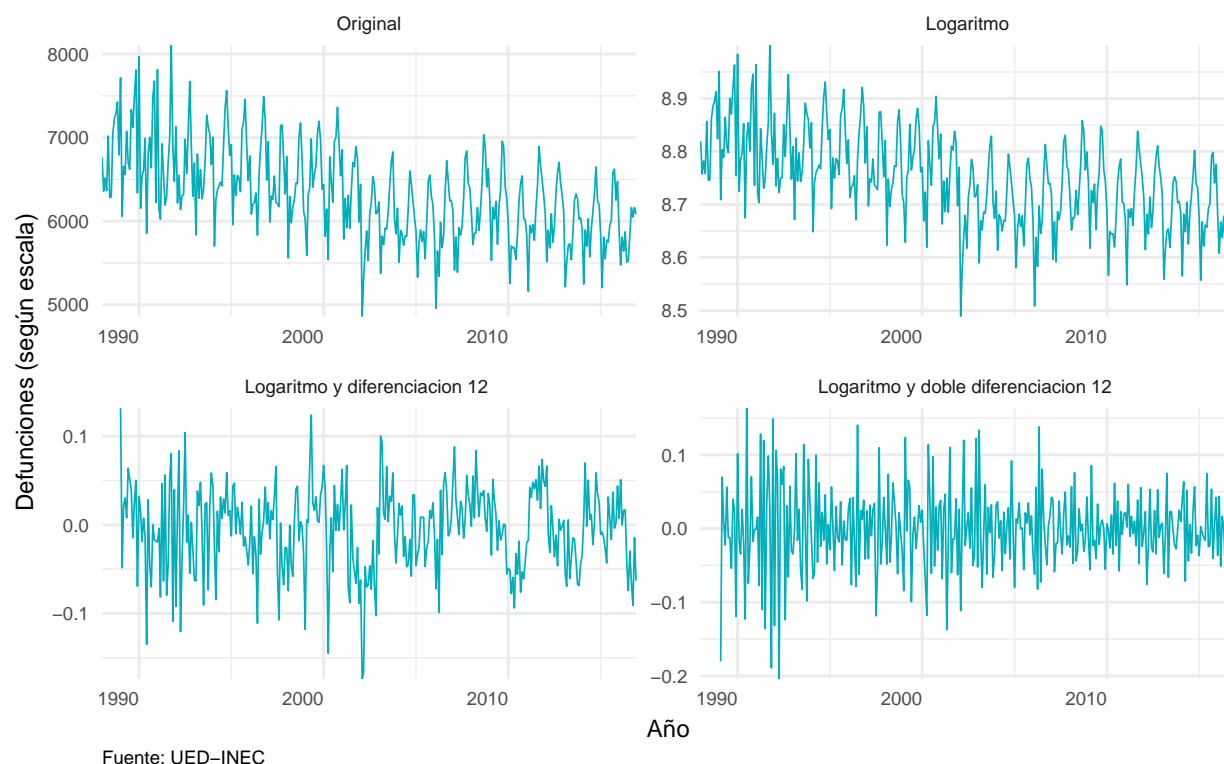
El segundo gran grupo de estimaciones corresponde a los modelos de suavizamiento exponencial. Para todas las variantes mencionadas en el apartado metodológico, se calcularon las medidas de rendimiento y los criterios de información para cada uno de los modelos, dichos resultados pueden verse en el cuadro 2 presente en la sección de anexos. Tras esta comparación, el mejor modelo de suavizamiento exponencial es que tiene errores, tendencia y estacionalidad multiplicativa descrito en la ecuación (10), pues presenta mejores medidas de rendimiento en el set de validación y medidas cercanas a las óptimas en el set de entrenamiento.

$$\hat{y}_{t+h|t} = l_t b_t^h s_{t-m+h_m^+} \quad (10)$$

Donde  $l_t = 0,1697 \left( \frac{y_t}{s_{t-m}} \right) + (1 - 0,1697)l_{t-1}b_{t-1}$ ,  $b_t = 0,0001 \left( \frac{l_t}{l_{t-1}} \right) + (1 - 0,0001)b_{t-1}$  y  $s_t = 0,1661 \left( \frac{y_t}{l_{t-1}b_{t-1}} \right) + (1 - \gamma)s_{t-m}$ .

Antes de estimar los modelos ARIMA es necesario realizar una inspección en base a los autocorrelogramas, y para esto se verifica la estacionaridad de la serie, es decir, que el promedio y la varianza de la serie se mantengan a lo largo del tiempo. Con este fin es que se analiza el gráfico 4, en el cual se puede ver que para volver la serie estacionaria es necesario aplicar el logaritmo y hacer una doble diferenciación cada 12 meses. El gráfico 5 muestra los autocorrelogramas parcial y total para la serie estacionaria, el cual parece sugerir que se tiene un MA(3) tanto en la no parte estacional y un MA(1) en la estacional, sin embargo ante la poca claridad de este criterio conviene realizar una sobreparametrización para encontrar el modelo más adecuado.

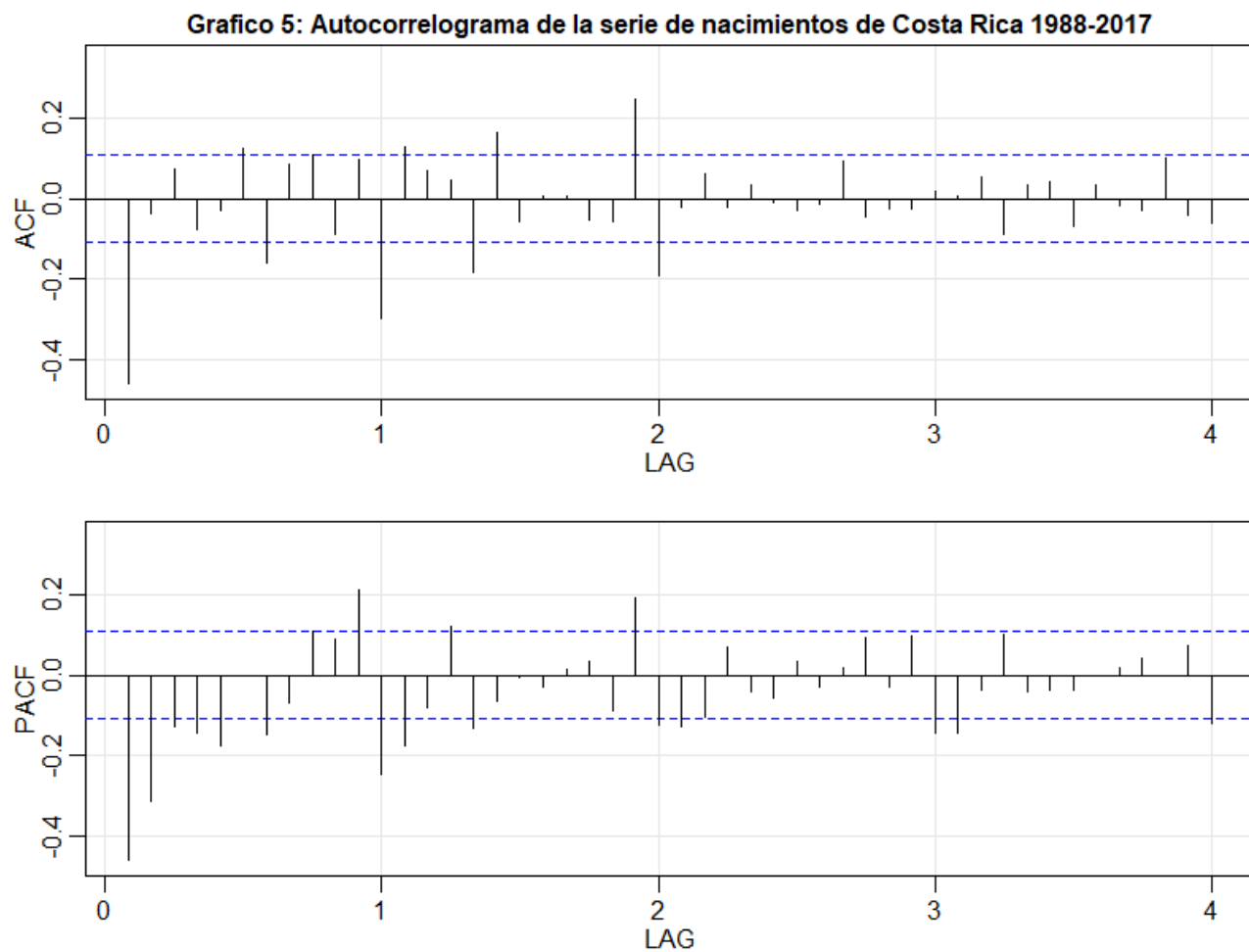
Gráfico 4: Nacimientos en Costa Rica en el periodo 1988 – 2017



Para el proceso de sobreparametrización, se especifican los valores tanto en la parte estacional como en la no estacional, y a partir de esto se obtienen las probabilidades asociadas a los coeficientes y se descartan aquellos modelos cuyos coeficientes no sean significativamente distintos de cero. Al ser este un proceso iterativo, se comparan 184 modelos ARIMA, de los cuales el más adecuado es el  $ARIMA(3, 1, 4)(1, 1, 3)_{12}$  mostrado en la ecuación (11). Los resultados de los mejores cinco modelos ARIMA pueden encontrarse en el cuadro 3 en la sección de anexos.

$$y_t = -0,2015y_{t-1} + 0,0872y_{t-2} + 0,9398y_{t-3} - 0,5642u_{t-1} - 0,2931u_{t-2} - 0,8318u_{t-3} + 0,6892u_{t-4} \\ + 0,8091y_{t-12} - 1,5892u_{t-12} + 0,5072u_{t-24} + 0,1520u_{t-36} + u_t \quad (11)$$

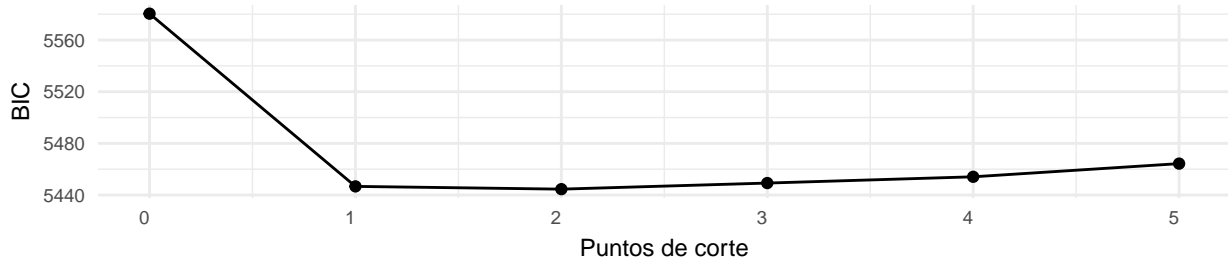
Además de los modelos ARIMA estimados previamente, es importante considerar intervenir la serie en cierto tramos. El gráfico 4 indica que, según el criterio del BIC se deben utilizar una intervención en el



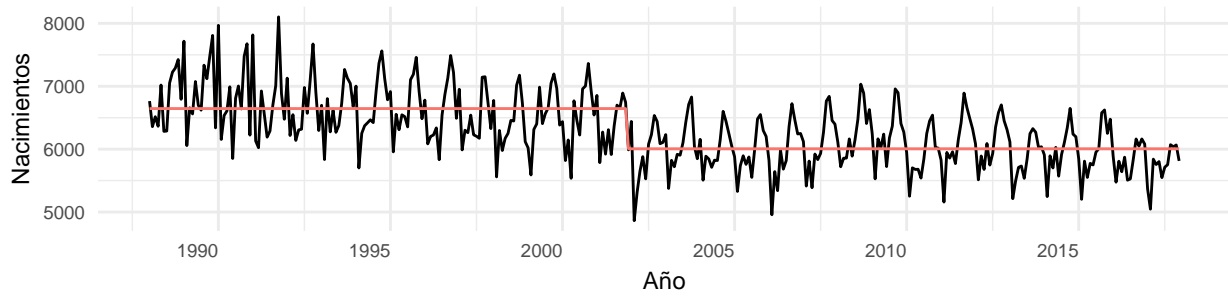
año 2001. Al estimar los modelos con intervención, el mejor es un  $ARIMA(5, 1, 3)(1, 1, 3)_{12}$ , descrito en la ecuación (12), el cual es inferior al modelo  $ARIMA(3, 1, 4)(1, 1, 3)$  sin intervención estimado previamente. Los cinco mejores modelos ARIMA con intervención pueden verse en el cuadro 4 en la sección de anexos.



Gráfico 5: Número de intervenciones para la serie de nacimientos en el periodo 1988 – 2017



Fuente: UED-INEC



Fuente: UED-INEC

$$\begin{aligned}
 y_t = & -0,9419y_{t-1} - 0,5431y_{t-2} + 0,5775y_{t-3} + 0,3603y_{t-4} + 0,1253y_{t-5} + 0,1565u_{t-1} \\
 & - 0,1884u_{t-2} - 0,9682u_{t-3} + 0,8172y_{t-12} - 1,6209u_{t-12} + 0,5100u_{t-24} \\
 & + 0,1775u_{t-36} + u_t
 \end{aligned} \tag{12}$$

El último modelo a estimar consiste en una red neuronal con dos capas ocultas estimada mediante la función `nnetar()` del paquete `forecast`. Los resultados de las comparaciones de los cuatro modelos y la visualización de la red se muestran a continuación:

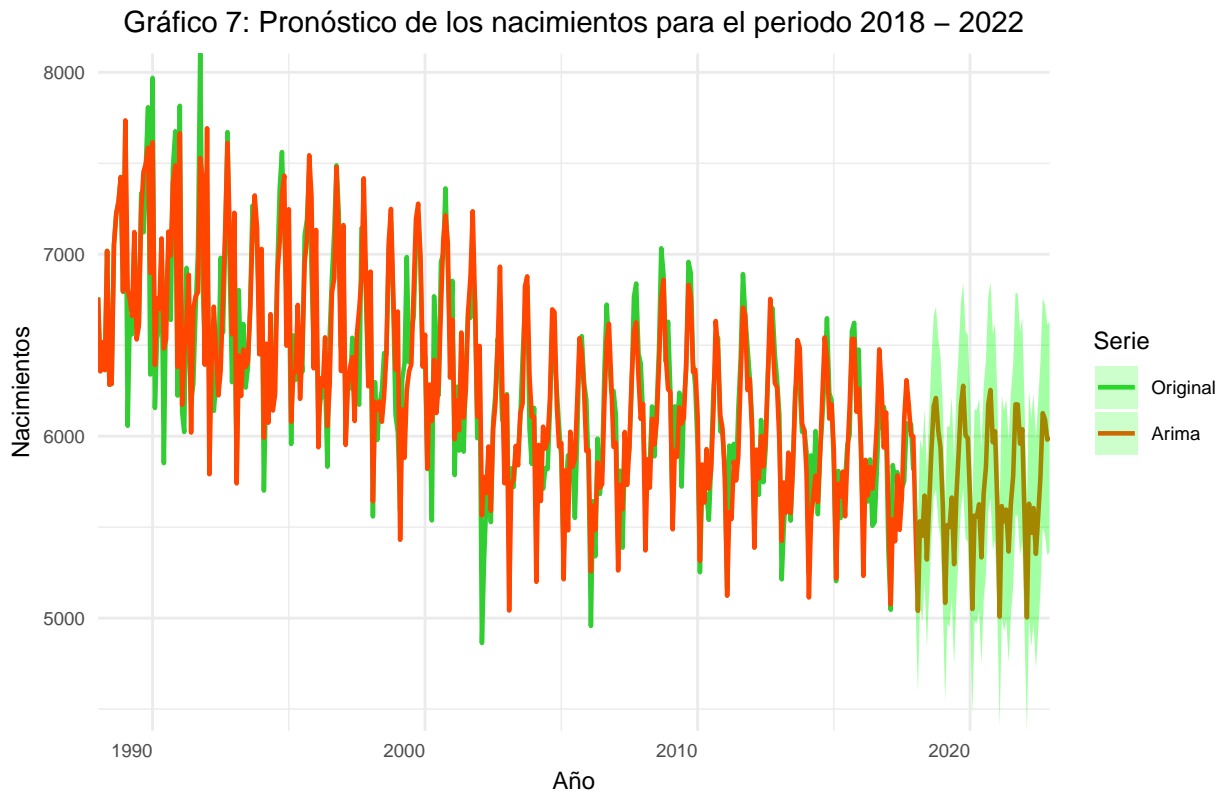
Para los datos de entrenamiento, la red neuronal es el modelo que ofrece mejores resultados, sin embargo esta no es capaz de obtener un buen ajuste en el conjunto de validación pues sus medidas de rendimiento son las peores de los cuatro modelos. El modelo  $ARIMA(3, 1, 4)(1, 1, 3)$  ofrece los mejores resultados en el conjunto de entrenamiento después de la red neuronal, y además cuenta con medidas muy competentes en el conjunto de validación, debido a este equilibrio es que se selecciona dicho modelo para realizar el

Cuadro 1: Criterios de información y medidas de rendimiento de los mejores modelos

Modelo	AIC	AICc	BIC	MAE	MAPE	RMSE	MASE
Regresion	-2201.01	-2199.75	-2147.08	200.35	3.17	259.43	0.85
Regresion Validacion				184.07	3.24	241.68	0.78
(M,M,M)	5852.27	5854.12	5917.76	178.67	2.83	232.62	0.76
(M,M,M) Validacion				178.72	3.12	205.27	0.76
ARIMA(3,1,4)(1,1,3)[12]	4607.83	4608.8	4653.6	161.07	2.56	214.29	0.68
ARIMA(3,1,4)(1,1,3)[12] Validacion				185.5	3.24	204.55	0.79
NNAR(24,1,12)[12]	0	0	0	10.49	0.17	14.26	0.04
NNAR(24,1,12)[12] Validacion				218.37	3.88	254.5	0.93

Fuente: UED-INEC

pronóstico para los siguientes 60 periodos como se muestra a continuación:



Fuente: UED-INEC

## 5. CONCLUSIONES Y RECOMENDACIONES

Entre los hallazgos más importantes, se encuentra que la entrada de la Ley de Paternidad responsable, aplicada desde el 2000 fue uno de los factores más influyentes para el descenso de la natalidad. Además,

Cuadro 2: Pronóstico de los nacimientos para el periodo 2018-2022 e intervalos de confianza de 95 por ciento

	Periodo	Pronostico	LI	LS	Periodo	Pronostico	LI	LS
Ene 2018	Ene 2018	5490	5054	5926	Jul 2020	5683	5095	6272
Feb 2018	Feb 2018	5039	4591	5488	Ago 2020	5831	5242	6420
Mar 2018	Mar 2018	5532	5077	5987	Sep 2020	6177	5586	6768
Abr 2018	Abr 2018	5452	4985	5919	Oct 2020	6254	5662	6846
May 2018	May 2018	5673	5199	6148	Nov 2020	5967	5375	6559
Jun 2018	Jun 2018	5322	4843	5801	Dic 2020	6027	5433	6621
Jul 2018	Jul 2018	5621	5133	6108	Ene 2021	5618	5017	6219
Ago 2018	Ago 2018	5920	5428	6412	Feb 2021	5008	4406	5610
Sep 2018	Sep 2018	6165	5669	6660	Mar 2021	5616	5011	6221
Oct 2018	Oct 2018	6210	5709	6712	Abr 2021	5532	4927	6138
Nov 2018	Nov 2018	6030	5526	6535	May 2021	5596	4990	6203
Dic 2018	Dic 2018	5937	5430	6443	Jun 2021	5366	4758	5974
Ene 2019	Ene 2019	5617	5081	6152	Jul 2021	5629	5020	6237
Feb 2019	Feb 2019	5083	4544	5623	Ago 2021	5782	5173	6392
Mar 2019	Mar 2019	5511	4968	6055	Sep 2021	6175	5564	6786
Abr 2019	Abr 2019	5501	4951	6051	Oct 2021	6174	5563	6785
May 2019	May 2019	5663	5111	6215	Nov 2021	5959	5347	6571
Jun 2019	Jun 2019	5296	4741	5851	Dic 2021	6038	5425	6651
Jul 2019	Jul 2019	5668	5108	6227	Ene 2022	5534	4913	6155
Ago 2019	Ago 2019	5907	5346	6468	Feb 2022	5004	4381	5627
Sep 2019	Sep 2019	6171	5607	6735	Mar 2022	5629	5004	6254
Oct 2019	Oct 2019	6277	5710	6844	Abr 2022	5469	4843	6095
Nov 2019	Nov 2019	6005	5438	6573	May 2022	5605	4978	6233
Dic 2019	Dic 2019	5990	5420	6560	Jun 2022	5353	4725	5982
Ene 2020	Ene 2020	5656	5078	6235	Jul 2022	5554	4925	6183
Feb 2020	Feb 2020	5050	4471	5629	Ago 2022	5776	5145	6406
Mar 2020	Mar 2020	5564	4982	6146	Sep 2022	6127	5496	6758
Abr 2020	Abr 2020	5558	4973	6142	Oct 2022	6092	5461	6724
May 2020	May 2020	5626	5041	6211	Nov 2022	5981	5348	6614
Jun 2020	Jun 2020	5334	4746	5921	Dic 2022	5995	5362	6629

Fuente: UED-INEC

se encontró que los nacimientos poseen patrones estacionales marcados, siendo los meses de setiembre y octubre los que presentan mayores nacimientos registrados, producto de las fiestas de fin y principio de año del país (Rodríguez 2015).

Se pronosticaron los nacimientos de Costa Rica hasta el año 2022, con un  $ARIMA(3, 1, 4)(1, 1, 3)$ , este modelo permite un alto nivel de precisión en la predicción los nacimientos y las mejores medidas de ajuste. El pronóstico revela que el número de nacimientos seguirán decreciendo. Por ende si los

nacimientos siguen disminuyendo tal y como indica el modelo, y la esperanza de vida siga en aumento como se ha observado en la última década, eventualmente la población de Costa Rica envejecerá y los jóvenes serán menos, por tanto se invita a las autoridades a tomar diferentes medidas en diferentes ámbitos como: la inclusión del adulto mayor en el mercado laboral, mejoría de servicios de salud y política migratoria, promoción de ahorros en la población, así como inversión en investigaciones sobre envejecimiento que dirijan la toma decisiones.

A continuación se resumen algunas recomendaciones concretas recopiladas de la bibliografía consultada (Financiero 2017)(Población 2008)(Albuquerque 2018)(M 2016)(Redondo 2013):

- Incluir políticas para proteger a los adultos mayores y mejorar su calidad de vida,
- mejorar las redes de seguridad social,
- ofrecer beneficios mínimos garantizados a los inmigrantes,
- facilitar las condiciones a los trabajadores de más edad para que participen en la fuerza laboral,
- atraer más inversión extranjera directa, impulsar la inversión interna y expandir su capacidad para adoptar nuevas tecnologías,
- promover políticas que permitan el acceso a empleos formales y bien remunerados.
- aumentar fuerza trabajo mujeres, promoviendo redes formales de trabajo y el cuidado de otros miembros del hogar, ya sea de menores o personas adultas mayores,
- no imponer una edad de pensionarse tan temprana, para sí mejorar la sostenibilidad del sistema de pensiones,
- propiciar el ahorro,
- favorece una adecuada planificación gerontológica,
- incentivar la promoción, prevención, atención, curación y la rehabilitación en relación con estilos de vida saludable a auto cuidado,
- impulsar investigaciones sobre la salud, causas de muerte, pobreza en la vejez,
- estudiar las características de las futuras personas jubiladas, pues dada la alta escolaridad del país, se esperaría mayor cantidad de profesionales, lo que implica pensiones más caras,
- adaptar estructuras y servicios para que sean accesibles a todas las personas de edad,

Cabe resaltar que la creación de políticas tomando en cuenta el envejecimiento poblacional, puede mitigar las consecuencias negativas para el futuro desarrollo del país, además se pueden obtener avances

en la ciencia, la tecnología desarrollada para la adecuación de los servicios (Redondo 2013) . Es importante recordar que tener pocas personas en edad de trabajar, también tiene sus ventajas, entre las que se destacan la disminución de la saturación en el mercado laboral y de vivienda (Bixby.L. 1997).

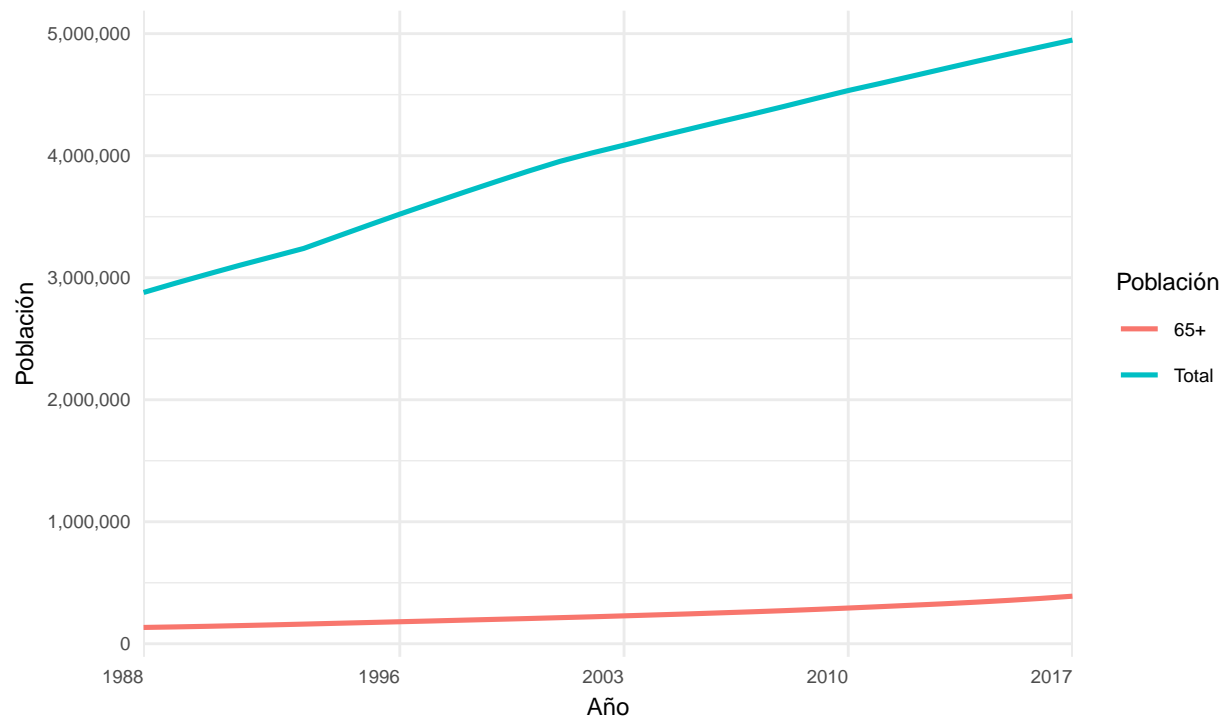
No obstante, si se ignoran las proyecciones sobre el envejecimiento de la población, la crisis fiscal actual, en vez de mejorar puede acentuarse como está sucediendo en otros países, como Rusia (Aragóñez 2017).

Asimismo, se considera importante visibilizar este fenómeno para que la población tome previsiones, ya que en la actualidad no hay suficiente conciencia, un ejemplo claro de ello es que hoy en día hay más pediatras que geriatras, a pesar de que la pirámide poblacional puede invertirse en un futuro cercano (M 2016).

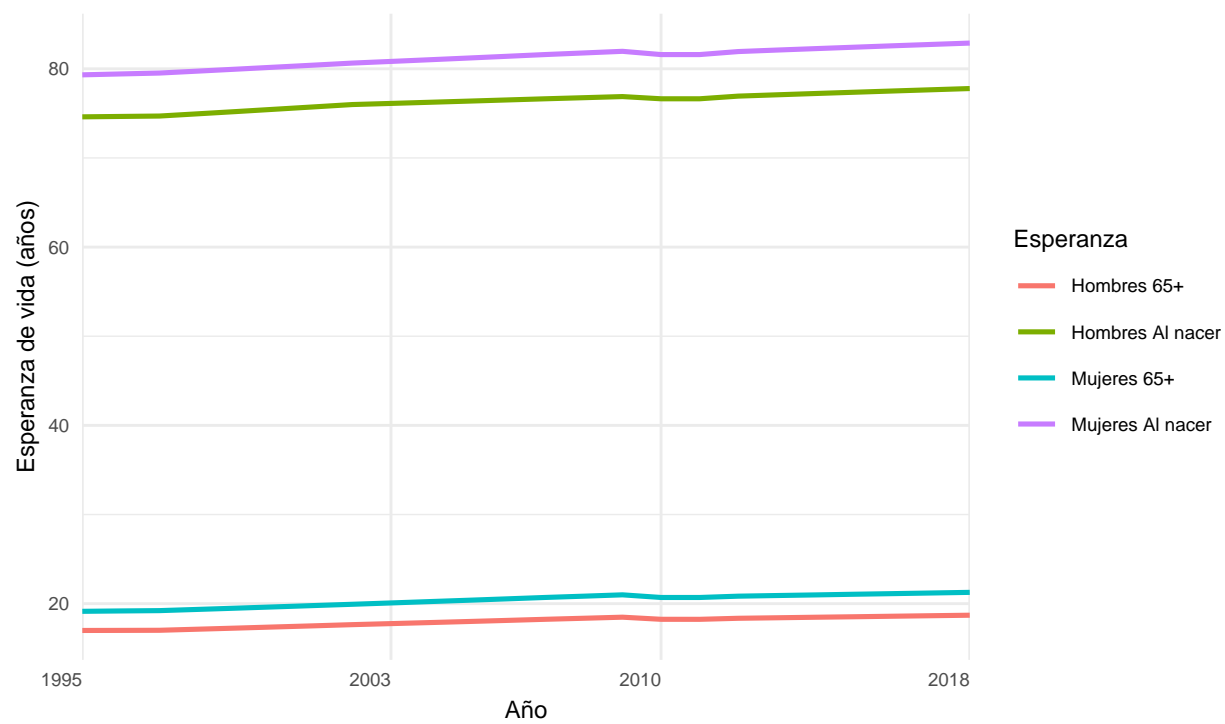
Para futuras investigaciones se recomienda analizar de forma multivariadas las diferentes series demográficas, para así obtener un panorama integral, pues como se me mencionó, estos procesos son interdependientes. Algunos indicadores que se pueden incluir son: la tasa de natalidad, global de fecundidad, mortalidad general, mortalidad infantil, la esperanza de vida al nacer y a los 65 años, tasa de dependencia, y la población de 65 años o más.

## 6. ANEXOS

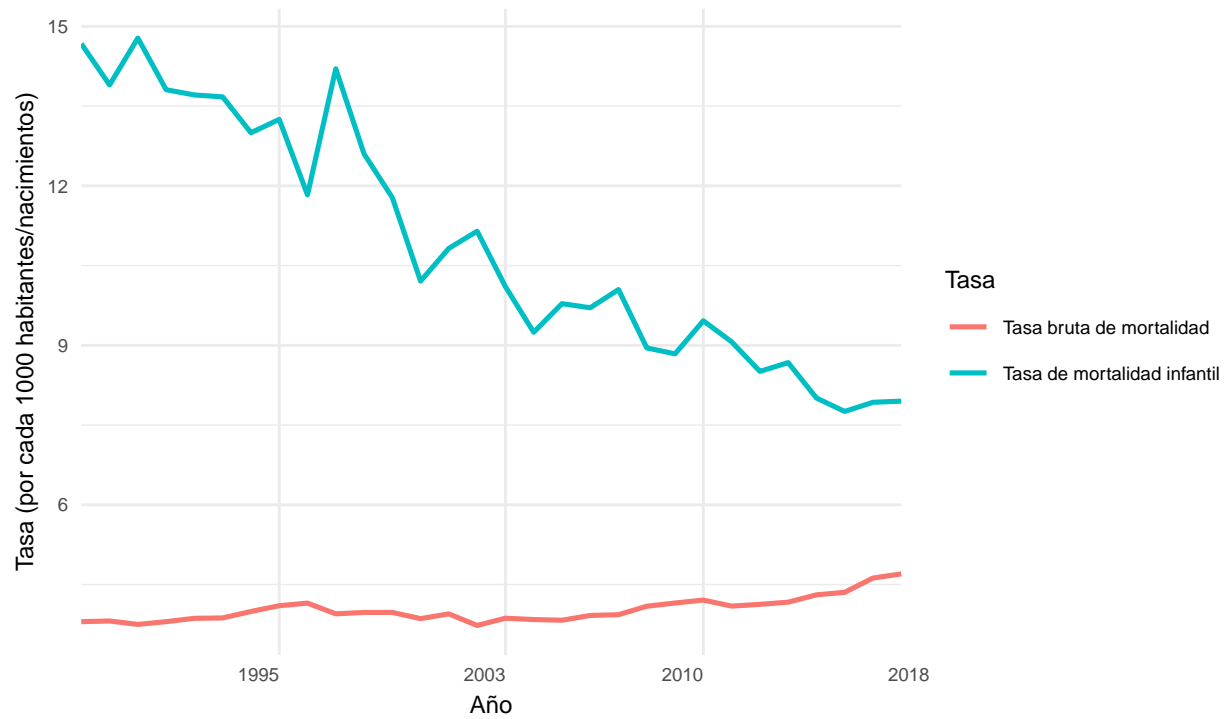
Anexo 1: Población total y mayores de 65 años  
en Costa Rica para el periodo 1988 – 2017



Anexo 2: Esperanza de vida al nacer y para mayores de 65 años en Costa Rica para el periodo 1988 – 2017 según sexo

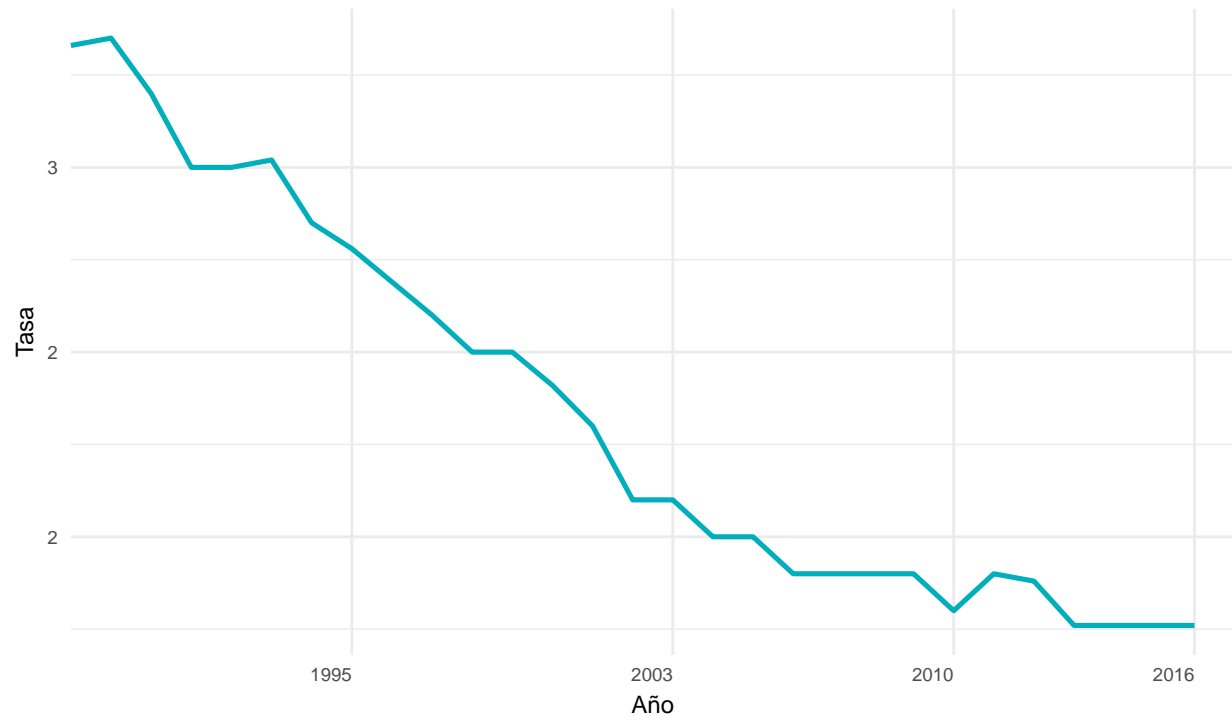


Anexo 3: Tasa bruta de mortalidad y de mortalidad infantil en Costa Rica para el periodo 1988 – 2017 según sexo

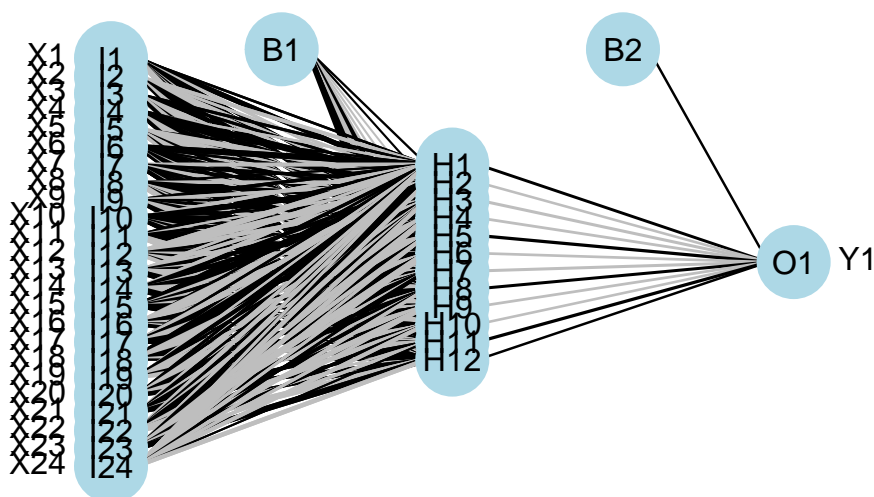




Anexo 4: Tasa global de fecundidad  
en Costa Rica para el periodo 1988 – 2017



## Anexo 5: Estructura de la red



Cuadro 3: Criterios de información y medidas de rendimiento de modelos de regresión

Modelo	AIC	AICc	BIC	MAE	MAPE	RMSE	MASE
Regresion T	4275.01	4275.08	4286.57	371.4	5.9	461.11	1.58
Regresion T Validacion				209.29	3.79	297.9	0.89
Regresion T+E	3903.04	3904.3	3956.97	202.26	3.21	261.8	0.86
Regresion T+E Validacion				202.23	3.53	259.71	0.86
Regresión exponencial	-2201.01	-2199.75	-2147.08	200.35	3.17	259.43	0.85
Regresión exponencial Validacion				184.07	3.24	241.68	0.78

Fuente: UED-INEC

Cuadro 4: Criterios de información y medidas de rendimiento de modelos de suavizamiento exponencial

Modelo	AIC	AICc	BIC	MAE	MAPE	MASE	RMSE
(A,N,N)	6299.32	6299.39	6310.88	350.77	5.58	1.49	453.06
(A,N,N) Validacion				354.73	6.47	1.51	455.58
(A,A,N)	6303.37	6303.55	6322.63	351.08	5.58	1.49	453.09
(A,A,N) Validacion				344.31	6.29	1.46	449.82
(A,Ad,N)	6305.34	6305.58	6328.45	351.09	5.59	1.49	453.07
(A,Ad,N) Validacion				354.72	6.46	1.51	455.55
(A,N,A)	5877.44	5878.88	5935.22	180.5	2.86	0.77	238.74
(A,N,A) Validacion				190.54	3.32	0.81	222.89
(A,A,A)	5883.11	5884.97	5948.6	179.67	2.84	0.76	239.32
(A,A,A) Validacion				185.74	3.24	0.79	215.64
(A,Ad,A)	5885.61	5887.69	5954.95	180.47	2.86	0.77	239.49
(A,Ad,A) Validacion				190.4	3.32	0.81	222.72
(M,N,N)	6276.83	6276.9	6288.38	351.23	5.59	1.49	453.13
(M,N,N) Validacion				356.24	6.49	1.51	456.75
(M,M,N)	6280.83	6281.01	6300.09	350.96	5.59	1.49	453.46
(M,M,N) Validacion				385.58	7	1.64	474.12
(M,Md,N)	6282.39	6282.64	6305.51	351.33	5.6	1.49	453.76
(M,Md,N) Validacion				356.7	6.5	1.51	457.05
(M,N,M)	5849.68	5851.13	5907.47	179.01	2.84	0.76	233.26
(M,N,M) Validacion				184.12	3.22	0.78	214.78
(M,M,M)	5852.27	5854.12	5917.76	178.67	2.83	0.76	232.62
(M,M,M) Validacion				178.72	3.12	0.76	205.27
(M,Md,M)	5866.79	5868.87	5936.13	177.42	2.81	0.75	237.63
(M,Md,M) Validacion				183.23	3.2	0.78	217.41
(M,N,A)	5864.33	5865.77	5922.11	179.49	2.84	0.76	238.57
(M,N,A) Validacion				190.26	3.31	0.81	221.44

Fuente: UED-INEC

Cuadro 5: Criterios de información y medidas de rendimiento de modelos ARIMA

Modelo	AIC	AICc	BIC	MAE	MAPE	RMSE	MASE
ARIMA(3,1,4)(1,1,3)[12]	4607.83	4608.8	4653.6	161.07	2.56	214.29	0.68
ARIMA(3,1,4)(1,1,3)[12] Validacion				185.5	3.24	204.55	0.79
ARIMA(1,1,0)(2,1,0)[12]	4751.59	4751.71	4766.85	206.66	3.29	280.3	0.88
ARIMA(1,1,0)(2,1,0)[12] Validacion				144.65	2.58	178.66	0.61
ARIMA(5,1,3)(5,1,3)[12]	4617.04	4618.97	4681.88	162.14	2.59	217.88	0.69
ARIMA(5,1,3)(5,1,3)[12] Validacion				172.1	3.02	196.07	0.73
ARIMA(0,1,1)(2,1,0)[12]	4680.15	4680.27	4695.41	187.03	2.99	251.5	0.79
ARIMA(0,1,1)(2,1,0)[12] Validacion				145.15	2.58	174.15	0.62
ARIMA(3,1,1)(0,1,0)[12]	4742.15	4742.33	4761.22	206.04	3.3	275.34	0.88
ARIMA(3,1,1)(0,1,0)[12] Validacion				145.79	2.62	181.57	0.62

Fuente: UED-INEC

Cuadro 6: Criterios de información y medidas de rendimiento de modelos ARIMA con intervención

Modelo	AIC	AICc	BIC	MAE	MAPE	RMSE	MASE
ARIMA(5,1,3)(1,1,3)[12]	4609.85	4611.36	4667.07	157.82	2.51	210.83	0.67
ARIMA(5,1,3)(1,1,3)[12] Validacion				214.34	3.73	260.65	0.91
ARIMA(1,1,0)(2,1,0)[12] errors	4755.5	4755.76	4778.39	206.67	3.29	280.25	0.88
ARIMA(1,1,0)(2,1,0)[12] Validacion				144.74	2.58	178.69	0.61
ARIMA(5,1,3)(2,1,2)[12] errors	4610.16	4611.67	4667.38	157.24	2.5	210.5	0.67
ARIMA(5,1,3)(2,1,2)[12] Validacion				217.41	3.78	265.1	0.92
ARIMA(6,1,6)(1,1,1)[12] errors	4611.29	4613.23	4676.13	158.19	2.52	211.14	0.67
ARIMA(6,1,6)(1,1,1)[12] Validacion				193.66	3.39	218.78	0.82
ARIMA(5,1,3)(2,1,3)[12] errors	4612.83	4614.54	4673.85	160.23	2.55	213.06	0.68
ARIMA(5,1,3)(2,1,3)[12] Validacion				200.97	3.51	235.92	0.85

Fuente: UED-INEC

## 7. REFERENCIAS

- Albuquerque, P. C. A. M. de. 2018. "Population Ageing and the Labour Market During the Recent Crisis in Portugal." *Portuguese Journal of Social Science*. [https://doi-org.ezproxy.sibdi.ucr.ac.cr/10.1386/pjss.17.1.105pass:\[\\_\]1](https://doi-org.ezproxy.sibdi.ucr.ac.cr/10.1386/pjss.17.1.105pass:[_]1).
- Aragón, G. 2017. "Rusia Se Asoma a Un Precipicio Demográfico." *La Vanguardia*. <https://www.lavanguardia.com/internacional/20170926/431566417533/rusia-precipicio-demografico.html>.
- Ávalo, A. 2018. "Un Tico Que Nazca Hoy Vivirá Casi 80 Años Y Un Poco Más Si Es Mujer." [https://www.teletica.com/210623\\_baja-en-tasa-de-natalidad-podria-provocar-que-nunca-lleguemos-a-6-millones-de-habitantes-en-el-paisf](https://www.teletica.com/210623_baja-en-tasa-de-natalidad-podria-provocar-que-nunca-lleguemos-a-6-millones-de-habitantes-en-el-paisf).
- Baker, & Iqelan, S. A. 2017. "Comparative Approach of Box-Jenkins Models and Artificial Neural Network Models on Births Per Month in Gaza Strip Using R." *IUG Journal of Natural Studies*. <http://journals.iugaza.edu.ps/index.php/IUGNS/article/viewFile/2616/1818>.
- Bixby, L. 1997. "Boom Demográfico En Costa Rica." Edited by La Nación. <https://ccp.ucr.ac.cr/investi/boom/boom.htm>.
- Carter, & Lee, L. R. 1986. "Joint Forecasts of US Marital Fertility, Nuptiality, Births, and Marriages Using Time Series Models." *Journal of the American Statistical Association*. <https://www.tandfonline.com/doi/abs/10.1080/01621459.1986.10478348>.
- Chichkanov, Vasilyeva, V. P. 2018. "Russia's Birth Rate Dynamics Forecasting." *R-Economy*. [http://elar.urfu.ru/bitstream/10995/47780/1/r-economy\\_2015\\_v1\\_2\\_17.pdf](http://elar.urfu.ru/bitstream/10995/47780/1/r-economy_2015_v1_2_17.pdf).
- Domínguez, J.D. 2018. "Baja En Tasa de Natalidad Podría Provocar Que Nunca Lleguemos a 6 Millones de Habitantes En El País." [https://www.teletica.com/210623\\_baja-en-tasa-de-natalidad-podria-provocar-que-nunca-lleguemos-a-6-millones-de-habitantes-en-el-paisf](https://www.teletica.com/210623_baja-en-tasa-de-natalidad-podria-provocar-que-nunca-lleguemos-a-6-millones-de-habitantes-en-el-paisf).
- Ellis, Peter. 2018. *Ggseas: 'Stats' for Seasonal Adjustment on the Fly with 'Ggplot2'*. <https://CRAN.R-project.org/package=ggseas>.
- Essuman, Nortey, R. 2017. "Modeling Births at a Tertiary Health-Care Facility in Ghana: Box-Jenkins Time Series Approach." *Journal of Biostatistics and Epidemiology*. <http://jbe.tums.ac.ir/index.php/jbe/>

[article/view/135](#).

Financiero, El. 2017. “Cómo Mantener El Crecimiento En Una Asia Que Envejece.” <https://www.elfinancierocr.com/opinion/como-mantener-el-crecimiento-en-una-asia-que-envejece/QTCBIKRW2JB2VLOKC2JD46VCLU/story/>.

———. 2018. “EF Lo Explica: Envejecimiento de La Población.” <https://www.elfinancierocr.com/economia-y-politica/video-ef-lo-explica-envejecimiento-de-la-ce015c78-0272-41a0-9c1f-2e9200701f1b/video/>.

Grolemund, Garrett, and Hadley Wickham. 2011. “Dates and Times Made Easy with lubridate.” *Journal of Statistical Software* 40 (3): 1–25. <http://www.jstatsoft.org/v40/i03/>.

Hyndman R, Bergmeir C, Athanasopoulos G. 2018. *Forecast: Forecasting Functions for Time Series and Linear Models*. <http://www.jstatsoft.org/article/view/v027i03>.

M, Fernández. 2016. “Datos Y Desafíos Del Envejecimiento Poblacional: Una Reflexión Sobre Políticas Públicas.” *Revista Nuevo Humanismo*. Universidad Nacional. doi:<http://dx.doi.org/10.15359/rnh.4-1.1>.

Mayorga, G. 2018. “La Población de Costa Rica No Llegará a Los Seis Millones"Y Otros Datos Que Nos Retan Como País.” Oficina de Divulgación e información de la Universidad de Costa Rica. <https://www.ucr.ac.cr/noticias/2018/11/16/la-poblacion-de-costa-rica-no-llegara-a-los-seis-millones-y-otros-datos-que-nos-retan-como-pais.html?fbclid=IwAR0GwoFAQmVoWQOHfZCQmbOdXAuZarwhUiYtNhiDnfH1g0K>

Población, Centro Centroamericano de. 2008. “I Informe Estado de Situación de La Persona Adulta Mayor.” <http://envejecimiento.csic.es/documentos/documentos/costarica-iinforme-01.pdf>.

Redondo, & Garay, N. 2013. “El Envejecimiento En América Latina: Evidencia Empírica Y Cuestiones Metodológicas.” *Reis*. [http://seminarioenvejecimiento.unam.mx/Publicaciones/libros/envejecimiento\\_AL.pdf](http://seminarioenvejecimiento.unam.mx/Publicaciones/libros/envejecimiento_AL.pdf).

Shitan, & Ng, M. 2015. “Forecasting the Total Fertility Rate in Malaysia.” *Pak. J. Statist.* [http://www.pakjs.com/journals/31\(5\)/31\(5\)8.pdf](http://www.pakjs.com/journals/31(5)/31(5)8.pdf).

Stoffer, David. 2017. *Astsas: Applied Statistical Time Series Analysis*. <https://CRAN.R-project.org/package=astsa>.

Tiit, E. M. 2018. “Measuring and Forecasting Fertility.” *Papers on Anthropology*. <http://ojs.utlib.ee/>

<index.php/PoA/article/view/poa.2018.27.1.07>.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <http://ggplot2.org>.

Wickham, Hadley, and Jennifer Bryan. 2018. *Readxl: Read Excel Files*. <https://CRAN.R-project.org/package=readxl>.

Wickham, Hadley, and Lionel Henry. 2018. *Tidyr: Easily Tidy Data with 'Spread()' and 'Gather()' Functions*. <https://CRAN.R-project.org/package=tidyr>.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2018. *Dplyr: A Grammar of Data Manipulation*. <https://CRAN.R-project.org/package=dplyr>.

Zhu, Hao. 2018. *KableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. <https://CRAN.R-project.org/package=kableExtra>.