

# Tasa de mortalidad infantil interanual en Costa Rica: Un análisis de la serie temporal 1989 - 2017

César Gamboa Sanabria<sup>\*,a</sup>

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

## RESUMEN

**Palabras clave:** TMII, mortalidad infantil, R, series temporales, regresión, ARIMA

La mortalidad infantil involucra algunos factores implícitos tanto a nivel socioeconómico como a nivel educativo y adquisitivo. La tasa de mortalidad infantil es un indicador demográfico utilizado para interpretar la calidad del sistema de salud de un país. El presente artículo analiza la tasa de mortalidad infantil interanual desde el periodo Febrero 1999 – Enero 2000 al periodo Enero 2017 – Diciembre 2017 mediante métodos de series temporales. La serie presenta una tendencia decreciente a lo largo del tiempo y carece de efectos estacionales. Se utilizó una descomposición aditiva de la serie debido a los cambios en la variancia que presenta la misma a lo largo del tiempo. El modelo final para realizar el pronóstico es el  $ARIMA(3, 1, 4)(1, 0, 1)_{12}$  con intervención en cinco periodos; dicho pronóstico para los siguientes 12 periodos con un 95 % de confianza abarca valores mínimos de 6.91, mientras que para el límite superior hasta 9.16.

## 1. INTRODUCCIÓN

La Tasa de Mortalidad Infantil (TMI) es uno de los indicadores demográficos más importantes, pues es utilizado como un parámetro de referencia sobre la calidad del sistema de salud, tanto a nivel nacional como regional. Si bien este indicador se construye relacionando las defunciones de menores de un año con el total de nacimientos, también involucra de manera implícita otras condiciones tales como las económicas, sociales y culturales, así como la efectividad en los métodos preventivos y curativos de esta categoría poblacional (León 1998). Debido a esto, el fallecimiento de un niño menor de un año se traduce en una falla del sistema de salud, por lo que estos casos son sujetos de estudio con el fin de conocer las causas que desencadenaron el evento.

En algunos países en vías de desarrollo de Asia, África y América Latina, la mortalidad infantil alcanza valores elevados pues la desnutrición, ausencia de asistencia médica y mala calidad de las condiciones sanitarias son, a diferencia de los países más desarrollados, algo muy común (Donoso S 2004). En el caso de Costa Rica, la unidad de estadísticas demográficas del Instituto Nacional de Estadística y Censos<sup>1</sup> (INEC) es el ente encargado de reportar este indicador con el fin de dar seguimiento y control al comportamiento del mismo a lo largo del tiempo con el objetivo de llegar a los niveles más bajos posibles.

En el INEC, cada mes se publica el boletín de la TMI interanual (TMII), que analiza la TMI de un mes y los 11 meses previos para comparar los periodos correspondientes (INEC 2004). El presente artículo busca hacer un análisis de la TMII para los 12 periodos desde el año 1989 y hasta 2017, y no de manera mensual simple, pues dada la volatilidad del fenómeno de estudio, hacer un estudio interanual permite analizar de una mejor manera los cambios entre periodos. Es decir, se analizará la TMII desde el periodo Febrero 1989 – Enero 1990 hasta el periodo Enero 2017 – Diciembre 2017.

La importancia de este proceso, aparte de servir de parámetro para evaluar el sistema de salud, está en su estrecha relación con las proyecciones de población, pues como se mencionó previamente, la TMII analiza la mortalidad en

---

\*email: cesar.gamboasanabria@ucr.ac.cr  
<https://github.com/cgamboasanabria>

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

LinkedIn: <https://linkedin.com/in/cgamboasanabria> | Github:

el grupo de edad de menores de un año, que es el primer grupo al generar tablas de mortalidad, ya sea de la forma clásica o mediante la mortalidad óptima (Villalón 2006). Uno de los métodos más conocidos para realizar estas estimaciones es el método de los componentes de cambio demográfico, que son la fecundidad, la mortalidad y la migración. En el caso de la mortalidad, uno de los puntos de partida es la estimación de las tasas de mortalidad por grupos de edad, siendo de particular interés la de menores de cinco años, pues esta a su vez se subdivide en los grupos de menores de un año y el de uno a cuatro años. Conocer el comportamiento de la mortalidad infantil es importante porque es en este grupo de edad en el que pueden existir cambios muy bruscos en la mortalidad y la fecundidad (Rincon 2000).

El objetivo general de este estudio consiste en analizar el comportamiento de la TMII a lo largo del tiempo para, haciendo uso de distintas herramientas de series temporales para así generar un pronóstico de este indicador para el año 2018. Para lograr esto, se plantean como objetivos específicos la identificación el mejor modelo de regresión, de suavizamiento exponencial, ARIMA identificado manualmente y de forma automática, y un modelo ARIMA con una intervención debido a comportamientos particulares en algunos tramos de la serie.

## 2. METODOLOGÍA

Los datos utilizados para esta investigación fueron obtenidos en la Unidad de Estadísticas Demográficas del INEC, y corresponden a información mensual de la cantidad de nacimientos y defunciones de menores de un año en Costa Rica desde Febrero 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 utilizó el paquete `lubridate` (Grolemund and Wickham 2011) y `forecast` (Hyndman R 2018) para realizar los pronósticos. En cuanto a la visualización, los paquetes `ggplot2` (Wickham 2016) y `ggseas` (Ellis 2018) serán empleados para generar distintos tipos de gráficos.

El método de análisis consiste en estimar, a partir las cantidades de nacimientos y defunciones, las tasas de mortalidad infantil interanual de la forma que indica la ecuación (1),

$$\sum_{k=0}^{11} \frac{D_{t-k}}{N_{t-k}} \quad (1)$$

donde  $D$  representa la cantidad de defunciones de menores de un año,  $N$  la cantidad de nacimientos y  $t$  el mes de arranque, así, si se toma  $t = 1$  se tendría la TMII del período Febrero – Enero. Una vez obtenidos estos valores se realizará un análisis descriptivo de la serie, para posteriormente realizar una descomposición de esta. 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 a los periodos Febrero 1989 – Enero 1990 hasta Enero 2016 – Diciembre 2016 el conjunto de entrenamiento y de Febrero 2016 – Enero 2017 hasta Enero 2017 – Diciembre 2017 para la validación. Se estimarán distintas variantes de los modelos de regresión, de suavizamiento exponencial y ARIMA.

En el caso de los modelos de regresión, se estimará uno solo 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 en (2).

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

---

<sup>2</sup><http://www.inec.go.cr/>

<sup>3</sup><https://cran.r-project.org/>

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()` y 3) aplicando un ARIMA con intervención.

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 y el automático, y el ARIMA con intervención. Dicha selección se hará analizando los criterios de información indicados en las ecuaciones (3), (4) y (5) donde los valores más bajos indican cuál es el mejor modelo, mientras que las medidas de rendimientos utilizadas se especifican en las ecuaciones (6), (7), (8) y (9) donde las más cercanas a cero sirven para indicar cuál modelo es el más adecuado,

$$AIC = -2\log L(\hat{\theta}) + 2k \quad (3)$$

$$AICc = -2\log L(\hat{\theta}) + 2k + \frac{2k + 1}{n - k - 1} \quad (4)$$

$$BIC = -2\log L(\hat{\theta}) + k \cdot \log(n) \quad (5)$$

$$MAE = n^{-1} \sum_{t=1}^n |y_t - f_t| \quad (6)$$

$$RMSE = \sqrt{n^{-1} \sum_{t=1}^n (y_t - f_t)^2} \quad (7)$$

$$MAPE = 100n^{-1} \sum_{t=1}^n \frac{|y_t - f_t|}{|y_t|} \quad (8)$$

$$MASE = n^{-1} \sum_{t=1}^n \frac{|y_t - f_t|}{q}; q = (n - m)^{-1} \sum_{t=2}^n |y_t - y_{t-1}| \quad (9)$$

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

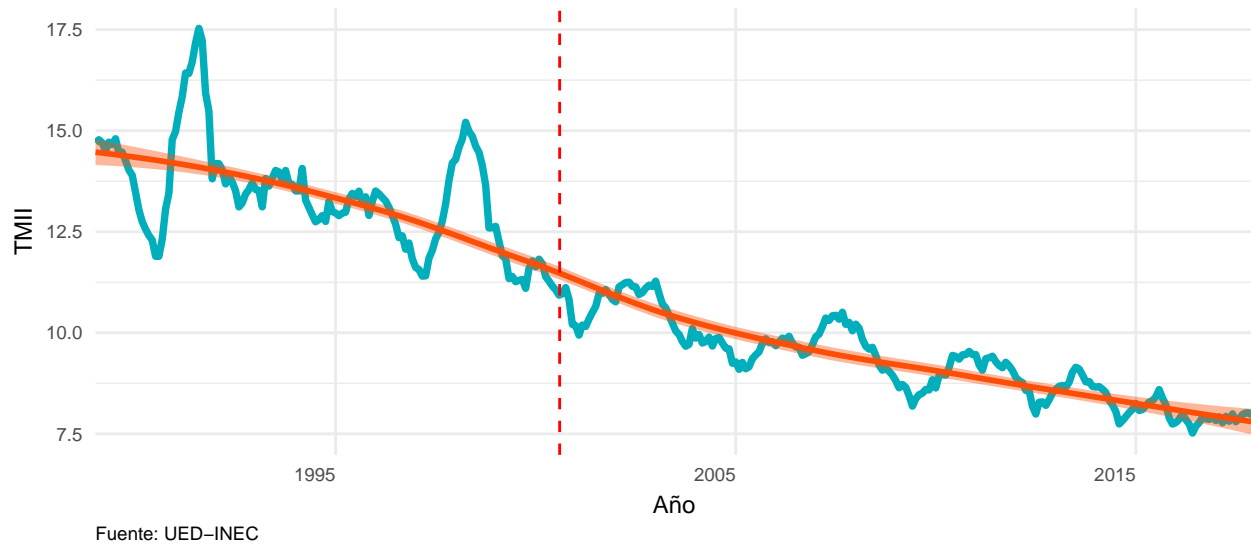
Finalmente se escogerá el modelo más adecuado en relación a tasa de mortalidad por causa externa (TMII) para posteriormente realizar un pronóstico de los posibles valores que alcanzará este indicador en el futuro, así como un contextualización del fenómeno con respecto a la actualidad.

### 3. RESULTADOS

Como fue mencionado en la introducción, un factor determinante en la mortalidad infantil son las condiciones de higiene y salubridad general que presenta un país. Con el paso de los siglos las condiciones previamente mencionadas han mejorado, de modo que en nuestros días la promoción de salud es un tema de interés público en casi todos los países (???). Debido a este mejoramiento es que la mortalidad infantil debido a la falta de higiene ha venido disminuyendo, y cada vez más, los casos se deben a causas endógenas (???). En base a estas condiciones es que es esperable observar una tendencia decreciente en la TMII.

Dado que la medición de la TMII se hace partiendo de un determinado mes y a partir de éste se consideran los 11 meses anteriores, el primer valor de la base de datos fue medido a partir de Enero de 2000, que corresponde al período interanual Febrero 1999 – Enero 2000. Todos los periodos siguientes se muestran en el gráfico 1.

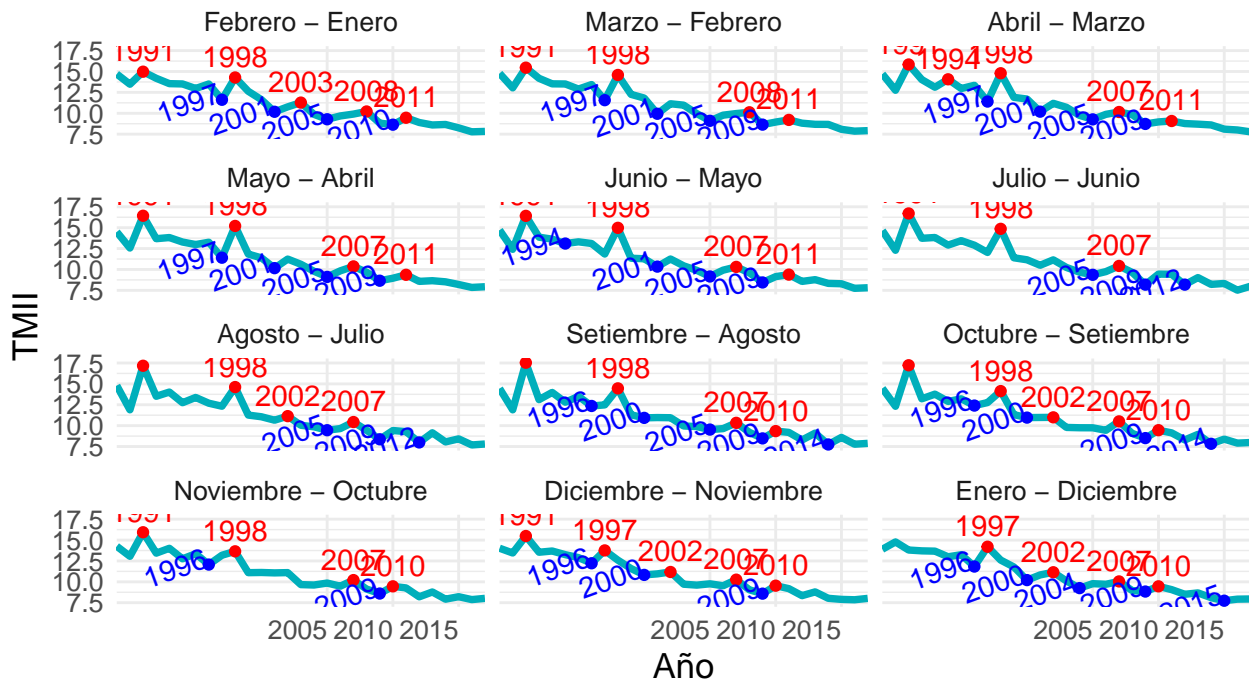
Gráfico 1: Tasa de Mortalidad Infantil Interanual 1989 – 2017



La serie muestra picos y valles pronunciados a lo largo de todo el periodo. A modo de visualización, se ajustó un suavizamiento de Loess para buscar señales de tendencia y concavidad en los datos temporales. La línea roja punteada se ubica aproximadamente en el mes de Julio del año 2000, pues a partir de ese punto el suavizamiento de Loess muestra un ligero cambio en la concavidad, lo cual sugiere que a partir ese punto será más difícil que la TMII vuelva a alcanzar valores similares a los mostrados al inicio de la serie. Además, al presentarse dos caídas y subidas abruptas en la TMII, esta tiende a estabilizarse.

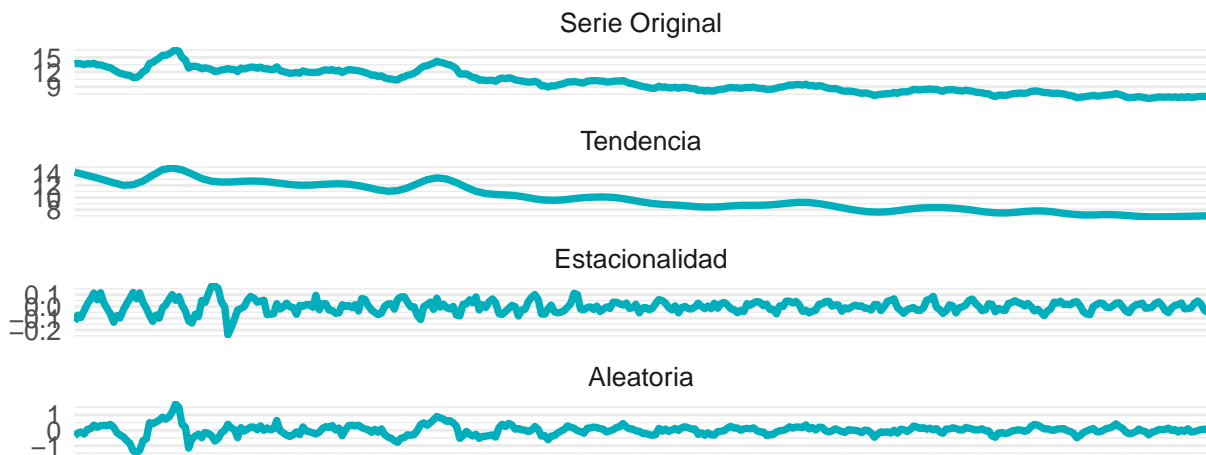
Mediante un análisis visual, el gráfico 2 parece respaldar el supuesto de que la mortalidad no posee efectos estacionales determinantes, pues para cada uno de los 12 períodos, en ninguno parecen existir mayores diferencias. El efecto que se mantiene en cada uno de los períodos es el de la tendencia, pues en cada uno ésta sigue descendiendo con el pasar de los años. Este hecho coincide con lo observado en el gráfico 1.

Gráfico 2: Tasa de Mortalidad Infantil Interanual 1989 – 2017 según período



Para hacer la descomposición de la serie se hizo una transformación de Box-Cox con  $\lambda = 1$  para aplicarla de manera multiplicativa. Esto se debe a que en el gráfico 1 pueden observarse cambios considerables en la variabilidad de la serie a lo largo del tiempo. El gráfico 3 muestra, como se mencionó previamente, una tendencia decreciente y una estacionalidad que no es reiterada a lo largo del tiempo. Además, el componente aleatorio muestra como los errores no son constantes durante todo el período.

Gráfico 3: Descomposición de la TMII en el periodo 2000 – 2017



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 las 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, los resultados pueden verse en cuadro 3 de la sección de anexos. El mejor modelo en este caso es el de variante exponencial, expresado mediante la ecuación (10), pues es el que cuenta con mejores medidas en el conjunto de entrenamiento y es muy competente en el conjunto de validación.

$$\begin{aligned} \log(\hat{y}_t) = & 2,7099 - 0,0020t + 0,0002s_2 + 0,0009s_3 + 0,0002s_4 + 0,0011s_5 + 0,0003s_6 \\ & + 0,0007s_7 + 0,0006s_8 + 0,0018s_9 + 0,0015s_{10} + 0,0014s_{11} + 0,0010s_{12} \end{aligned} \quad (10)$$

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 4 presente en la sección de anexos. Tras esta comparación, el mejor modelo de suavizamiento exponencial es que tiene errores aditivos, sin tendencia y con estacionalidad aditiva descrito en la ecuación (11), pues presenta mejores medidas de rendimiento en el set de validación y medidas cercanas al mínimo en el set de entrenamiento.

$$\hat{y}_{t+h|t} = l_t + s_{t+h-m(k+1)} \quad (11)$$

Donde  $l_t = 0,956(y_t - s_{t-m}) + (1 - 0,956)l_{t-1}$ ;  $s_t = 0,0113(y_t - l_{t-1}) + (1 - 0,0113)s_{t-m}$ .

Antes de estimar los modelos ARIMA es importante realizar una inspección en base a los autocorrelogramas, y para esto es necesario verificar la estacionaridad de la serie, es decir, que el promedio y la variancia de la serie se mantengan a lo largo del tiempo. Con este fin es que se analiza el gráfico 5, en el cual se puede ver que para volver la serie estacionaria es necesario aplicar el logaritmo y hacer una doble diferenciación cada 12 periodos. El gráfico 4 muestra los autocorrelogramas parcial y total para la serie estacionaria, el cual parece sugerir que se tiene un MA(1) en la parte estacional, sin embargo ante la poca claridad de este criterio conviene realizar una sobreparametrización para encontrar el modelo más adecuado.

Para el proceso de sobreparametrización, se hace uso de la función `sarima()`, en la cual ese 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 llegaron a estimar 556 modelos ARIMA, de los cuales el más adecuado es el  $ARIMA(3, 1, 4)(2, 0, 3)_{12}$ , el cual posee mejores medidas de rendimiento que el mejor modelo seleccionado mediante la función `auto.arima()`, la cual sugiere utilizar un  $ARIMA(1, 1, 2)(0, 0, 1)_{12}$ ; el mejor modelo se muestra en la ecuación mostrado en la ecuación (12).

$$\begin{aligned} y_t = & -0,1924y_{t-1} - 0,0866y_{t-2} + 0,7667y_{t-3} - 1,3291y_{t-12} - 0,9808y_{t-24} \\ & + 0,3980y_{t-1} + 0,4177y_{t-2} - 0,5650y_{t-3} + 0,0607y_{t-4} + 0,5458y_{t-12} - 0,0119y_{t-24} - 0,7683y_{t-36} + u_t \end{aligned} \quad (12)$$

El gráfico 4 indica que en la serie existen 5 niveles, estos cambios se dan en Marzo de 1994, Febrero de 1999, Junio de 2006, Julio de 2008 y Agosto de 2013. Al incorporar estas intervenciones en el modelo, algunos términos previamente estimados dejan de ser significativos, quedando así un  $ARIMA(3, 1, 4)(1, 0, 1)_{12}$ , descrito en la siguiente ecuación:

$$\begin{aligned} y_t = & 0,1185y_{t-1} + 0,9935y_{t-2} - 0,1127y_{t-3} - 0,0001y_{t-12} \\ & + 0,0547y_{t-1} - 0,7314y_{t-2} - 0,0353y_{t-3} - 0,2484y_{t-4} - 0,8864y_{t-12} \\ & - 0,1829I_{1994} - 0,2950I_{1999} + u_t \end{aligned} \quad (13)$$

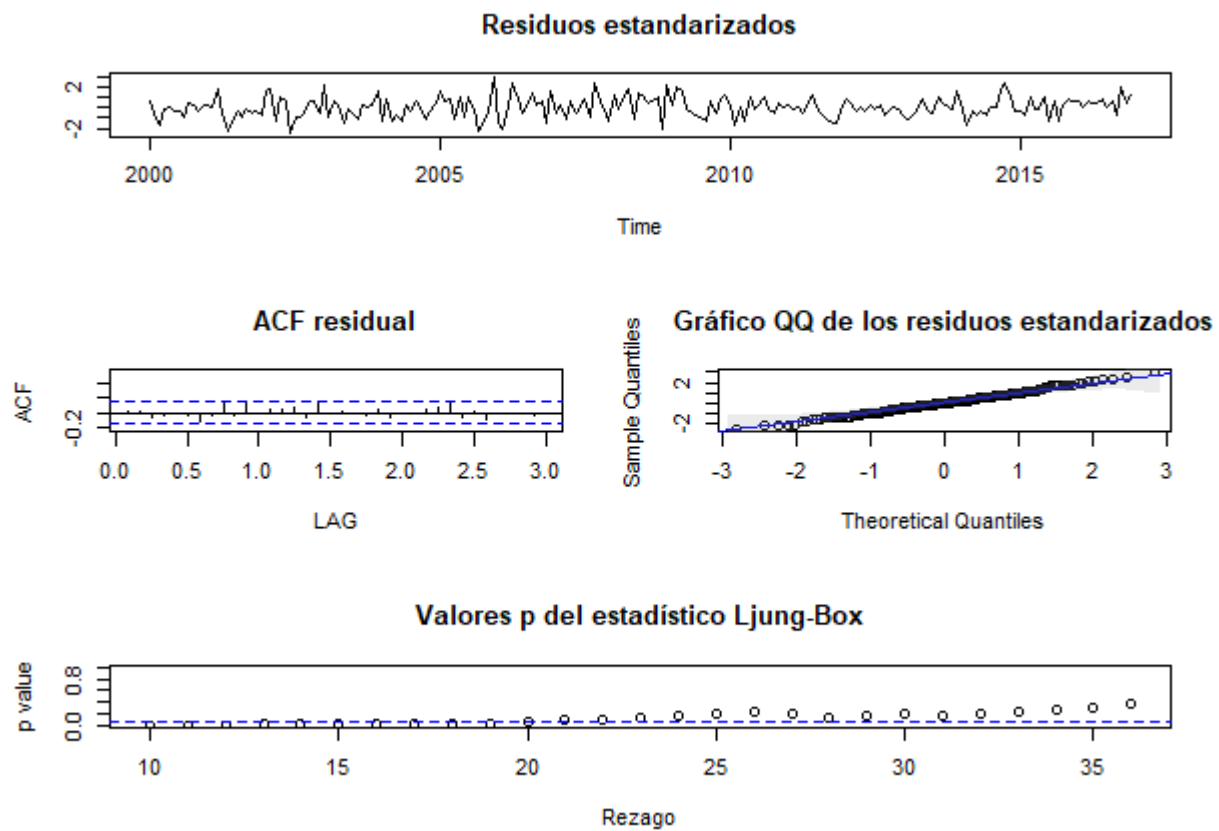
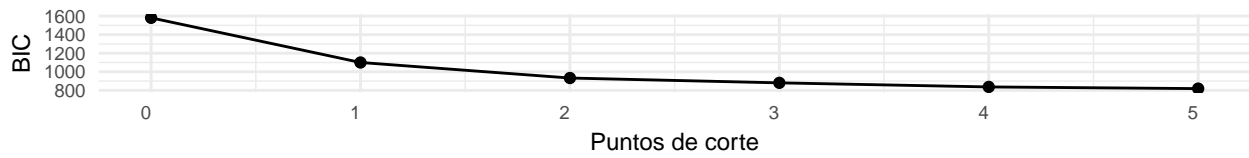


Gráfico 4: Número de intervenciones en la Tasa de Mortalidad Infantil Interanual 1989 – 2017



Fuente: UED-INEC



Fuente: UED-INEC

Con los 4 mejores modelos seleccionados en cada categoría, se procede a comparar sus medidas de rendimiento y, posteriormente, el pronóstico para cada uno de ellos y verificar, mediante sus pronósticos, cual de todos es el más adecuado. El cuadro 1 muestra los criterios de información y las medidas de rendimiento para los datos de entrenamiento (rojo) y de validación (azul), siendo el modelo ARIMA con intervención el mejor en el conjunto de entrenamiento y bastante competente en el de validación, pues queda ligeramente superado por el modelo de suavizamiento exponencial. Para el ARIMA con intervención se muestran los resultados del pronóstico.

Modelo	AIC	AICc	BIC	MAE	MAPE	RMSE	MASE
Reg. tendencia exponencial	-1785.19	-1783.89	-1731.75	0.57	4.92	0.84	0.58
Reg. tendencia exponencial Validacion				0.25	3.14	0.28	0.26
(A,N,A)	1113.99	1115.49	1171.25	0.19	1.71	0.27	0.2
(A,N,A) Validacion				0.07	0.89	0.09	0.07
ARIMA(3,1,4)(2,0,3)[12]	-128.18	-127.05	-78.6	0.13	1.21	0.18	0.14
ARIMA(3,1,4)(2,0,3)[12] Validacion				0.08	1.03	0.1	0.08
Arima con intervención	-143.94	-142.97	-98.17	0.13	1.19	0.18	0.14
Arima con intervención Validacion				0.27	3.38	0.3	0.28



Gráfico 7: Pronóstico de la TMCE para el 2018



Cuadro 1: Pronóstico de la TMCE para 2018 e intervalos de confianza de 95 por ciento

	Pronóstico	Límite Inferior	Límite Superior
Ene 2018	7.927547	7.552867	8.302227
Feb 2018	7.855485	7.265884	8.445085
Mar 2018	7.901780	7.075568	8.727993
Abr 2018	7.835802	6.798904	8.872700
May 2018	7.976087	6.743009	9.209165
Jun 2018	7.953698	6.528464	9.378932
Jul 2018	8.034495	6.431593	9.637396
Ago 2018	7.911296	6.133925	9.688667
Sep 2018	7.850503	5.910405	9.790601
Oct 2018	7.902747	5.803567	10.001926
Nov 2018	8.001486	5.752624	10.250349
Dic 2018	8.066496	5.671733	10.461260

Fuente: UED-INEC

#### 4. CONCLUSIONES Y RECOMENDACIONES

Con todo lo descrito en las secciones previas, se han cubierto algunos de los aspectos más fundamentales en un análisis de series temporales, como lo son los análisis descriptivos, la estimación de modelos junto con sus medidas de rendimiento y, lo más importante de todo, el pronóstico. Los resultados obtenidos ofrecen una guía para futuros estudios relacionados con la tasa de mortalidad infantil interanual.

La mortalidad, particularmente a edades tempranas, es una serie muy volátil, y aunque se busque llegar a los valores más bajos posibles en este indicador, existen causas naturales que impedirían llegar a un cierto punto, es decir, eventualmente la mortalidad infantil debe converger a un cierto valor. En décadas anteriores mucha de la mortalidad se debía a condiciones inadecuadas en la higiene pública, sin embargo, en la actualidad la gran mayoría de los casos se debe a enfermedades congénitas que difícilmente puedan ser controladas.

Los análisis descriptivos ofrecen una visualización de la tendencia decreciente de la TMII, lo cual es congruente con el párrafo anterior pues las condiciones de salud pública han mejorado con los años. De todos los modelos que

se estimaron los mejores son los de suavizamiento exponencial que consideran un amortiguamiento y excluyen el efecto de la estacionalidad, esto se debe a que al final de la serie los datos muestran una cierta convergencia.

El pronóstico realizado ofrece un intervalo de confianza que es congruente con el comportamiento de la serie en general, pues alcanza como valores máximos los que se han venido mostrando desde 2015, mientras que para los valores mínimos es posible que se llegue a niveles más bajos que los presentados en toda la serie, esto es esperable debido a la tendencia decreciente de la serie.

A pesar de estos resultados conviene tener presente que, en Costa Rica, desde Mayo del año 2018 los registros de defunciones también se realizan de manera digital, y al estar en una fase inicial de la implementación, pueden existir subregistros de las muertes infantiles o bien, un sobre registro de las mismas. Este hecho puede afectar las estimaciones de la TMII, por lo que es recomendable reproducir este análisis de manera mensual para llevar un registro adecuado de todos los períodos interanuales.

## 5. ANEXOS

Cuadro 2: Medidas de rendimiento y criterios de informacion para modelos de regresion

Modelo	AIC	AICc	BIC	MAE	MAPE	RMSE	MASE
Regresion T	-95.71	-95.63	-84.26	0.61	5.4	0.86	0.62
Regresion T Validacion				0.67	8.44	0.68	0.69
Regresion T+E	-73.75	-72.44	-20.31	0.61	5.41	0.86	0.63
Regresion T+E Validacion				0.67	8.45	0.68	0.69
Regresión exponencial	-1785.19	-1783.89	-1731.75	0.57	4.92	0.84	0.58
Regresión exponencial Validacion				0.25	3.14	0.28	0.26

Fuente: UED-INEC

## 6. REFERENCIAS

- Donoso S, Enrique. 2004. “Desigualdad en mortalidad infantil entre las comunas de la provincia de Santiago.” *Revista médica de Chile* 132 (April). scieloc: 461–66. [https://scielo.conicyt.cl/scielo.php?script=sci\\_arttext&pid=S0034-98872004000400008&nrm=iso](https://scielo.conicyt.cl/scielo.php?script=sci_arttext&pid=S0034-98872004000400008&nrm=iso).
- Ellis, Peter. 2018. *Ggseas: 'Stats' for Seasonal Adjustment on the Fly with 'Ggplot2'*. <https://CRAN.R-project.org/package=ggseas>.
- 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>.
- INEC. 2004. “Documento Metodológico de Defunciones Infantiles.” INEC.
- León, B. ; Etevez, R.; Gallegos. 1998. “Mortalidad infantil: Análisis de un decenio.” *Revista Cubana de Medicina General Integral* 14 (December). scielocu: 606–10. [http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S0864-21251998000600017&nrm=iso](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0864-21251998000600017&nrm=iso).
- Rincon, M. 2000. *Métodos Para Proyecciones Demográficas*.
- Villalón, S.; Ortega, G.; Vera. 2006. “Tabla de Vida Por Método de Mortalidad óptima.” INE.
- 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>.