

UNIVERSIDAD DE COSTA RICA

CA-0404 MODELOS LINEALES

ANTEPROYECTO DE INVESTIGACIÓN

Pronóstico demográfico de Costa Rica

Profesor

Luis Barboza Chinchilla

Autores

David Zumbado Fernández

Leonardo Blanco Villalobos

Ignacio Barrantes Valerio

7 de noviembre de 2022

Índice de contenidos

	4
Anteproyecto	5
Objetivos	5
Introducción al marco teórico	5
Descripción detallada de los datos	7
Análisis descriptivo de los datos	8
Modelos estadísticos propuestos	11
Justificación modelos DLM propuestos	13
Modelo $ARIMA(p, d, q)$	13
Modelos polinomiales de primer y segundo orden	13
UVE de Gowin	14
Referencias	15

Listado de Figuras

1	Cantidad de defunciones por año para el periodo 1950-2020	9
2	Defunciones infantiles, neonatales y fetales por año	10
3	Distribución variable indicadora de la mayor tasa entre las de mortalidad fetal y la neonatal	11
4	UVE Heurística	14

Listado de Tablas

1	Resumen de cinco números para algunas variables de la tabla de datos	8
2	Estadísticos de dispersión para algunas variables de la tabla de datos	8
3	Frecuencia de defunciones, población y tasa de mortalidad	9

Anteproyecto

Objetivos

Objetivo general: Hacer un pronóstico de la cantidad de defunciones totales de Costa Rica para una ventana de tiempo de dos años (2019 y 2020) utilizando datos de 1950 al 2018.

Objetivo específico 1: Ajustar un modelo ARIMA a la serie de tiempo de defunciones totales entre 1950 y 2018 en Costa Rica.

Objetivo específico 2: Ajustar un modelo DLM polinomial a la serie de tiempo de defunciones totales entre 1950 y 2018 en Costa Rica.

Objetivo específico 3: Comparar los pronósticos obtenidos con cada modelo utilizando los datos reales como referencia.

Introducción al marco teórico

En términos generales, la demografía es “una ciencia que estudia las poblaciones humanas, su dimensión, estructura, evolución y características generales” (Bruno Ramírez, 2019) y además “estudia estadísticamente la estructura y la dinámica de las poblaciones, así como los procesos concretos que determinan su formación, conservación y desaparición [...] tales procesos son los de fecundidad, mortalidad y migración: emigración e inmigración” (Bruno Ramírez, 2019). De acuerdo al mismo autor, hay dos grandes tipos de demografía: la estática, que “estudia las poblaciones humanas en un momento de tiempo determinado desde un punto de vista de dimensión, territorio, estructura y características estructurales” (2019) y la dinámica, “que estudia las poblaciones humanas desde el punto de vista de la evolución en el transcurso del tiempo y los mecanismos por los que se modifica la dimensión, estructura y distribución geográfica de las poblaciones” (2019). Bajo estas concepciones, el presente trabajo caería en el campo de la demografía dinámica, pues se buscará realizar pronóstico sobre la serie histórica de defunciones de Costa Rica de 1950 a 2020. En ese sentido, Rees (2020) afirma que una de las mayores contribuciones de la demografía ha sido el desarrollo de modelos para la proyección de futuras poblaciones, lo cual está estrechamente relacionado con la proyección de defunciones. Según Macció et al. (1985) la mortalidad o defunción se emplea para expresar la acción de la muerte sobre la población y esta se mide en valores absolutos y sobre el año calendario. Por su parte y de acuerdo a Brownlee (2020), la realización de predicciones sobre el futuro se denomina extrapolación en el tratamiento estadístico clásico de los datos de las series temporales. Los campos más modernos que se centran en el tema, lo denominan pronóstico de series temporales. El pronóstico consiste en tomar modelos ajustados a datos históricos y utilizarlos para predecir observaciones futuras. Asimismo, una serie cronológica o crónica se le llama a “la que forman los valores sucesivos que una cierta variable ha tomado en el transcurso del tiempo” (Macció et al., 1985), como lo puede ser, por ejemplo, la serie de defunciones totales anuales para un país en específico.

Ahora bien, las teorías demográficas “pretenden explicar los patrones del crecimiento de la población en diversos países del mundo dando cuenta de la estructura y la dinámica de la población y estableciendo leyes o principios que regirían esos

fenómenos” (Bruno Ramírez, 2019). Más específicamente, una teoría demográfica “explicaría los cambios y acontecimientos de las poblaciones humanas, de su dimensión, estructura, evolución y características generales, tanto desde un punto de vista cuantitativo (estadístico) como cualitativo (biológico, sociológico, cultural y económico)” (Bruno Ramírez, 2019).

La primer teoría global de la población, y una de las más famosas, es la malthusiana, enunciada por Thomas Robert Malthus en 1798 en su Ensayo sobre el principio de población, que consistió “en una teoría donde se establece una ley general que explica el crecimiento total de la población en relación con otra variable fuera del contexto social, como lo es la disponibilidad de alimentos” (Sarrible, 2009). A su vez, explica que la teoría malthusiana propone que “el ritmo de crecimiento de la población responde a una progresión geométrica, mientras que el ritmo de aumento de los recursos para su supervivencia lo hace en progresión aritmética” (Bruno Ramírez, 2019). De esta manera, Malthus sostenía que el hambre, la guerra y otras catástrofes impedían que ocurriera la pauperización gradual de la especie humana, pues mantenían controlado el número de nacimientos (Bruno Ramírez, 2019). -Sarrible (2009) cuestiona la vigencia de esta teoría aludiendo que “hay países que crecen y otros que tienden a la población estacionaria o tienen crecimiento natural negativo [...] hay países que crecen, a pesar de que el saldo migratorio les sea desfavorable y otros que crecen gracias a las migraciones internacionales” (2009)

De acuerdo a Mariscal de Gante & Rodríguez (2018), una de las teorías demográficas más importantes es la Teoría de la Transición Demográfica (TTD). Esta consiste en una generalización empírica en función de observaciones pasadas y establece una conexión entre la evolución demográfica de la población y el crecimiento económico (Mariscal de Gante & Rodríguez, 2018). Por su parte, Alcalde (2010) define el régimen En síntesis, la TTE bajo su estudio establece que las enfermedades degenerativas modernas y las patologías creadas por el ser humano son las principales causas de mortalidad y morbilidad. Como causas secundarias, son aquellas ligadas a hambrunas, pandemias, guerras, originando acusados picos de mortalidad durante la evolución demográfica de las sociedades. Se define el régimen demográfico como “el comportamiento de una población a partir de la evolución de sus tasas de natalidad, mortalidad y el crecimiento natural o vegetativo, es decir, la diferencia entre el número de nacidos y fallecidos” (2010) y menciona que su propósito es “es explicar los cambios que se producen en la evolución de la población mundial” (2010).

El mismo autor explica que bajo el régimen demográfico, la TTD da pie a los modelos demográficos más importantes y esta “se produce cuando la natalidad y la mortalidad, o por lo menos uno de los dos fenómenos, ha dejado sus elevados niveles tradicionales para dirigirse hacia porcentajes más bajos, asociados a la fecundidad dirigida y al uso de métodos de lucha contra la natalidad, pasando de una demografía antigua y tradicional a otra moderna” (Alcalde, 2010). Elaborada durante los años 30, la TTD “ha sido profundamente revisada, debatida y criticada, pero la mayoría de los especialistas [...] la consideran útil y válida didáctica y metodológicamente [...] para analizar el distinto ritmo de crecimiento de la población a lo largo del tiempo o las diferencias existentes en la actualidad entre unos países y otros de la Tierra. Un punto muy importante para el presente trabajo es que la TTD”se cumple más fielmente en las sociedades industrializadas europeas al producirse el cambio de la dinámica demográfica a raíz de la Revolución industrial, mientras se duda de si es aplicable o no a los países del Tercer Mundo” (Alcalde, 2010). Prueba de este punto es que se han propuesto modelos distintos bajo la TTD para países desarrollados y para países subdesarrollados, donde se habla de “revolución demográfica” en los primeros, en alusión al crecimiento repentino de la población debido a los factores mortalidad y natalidad, pero de “explosión demográfica” en el caso de los segundos (Alcalde, 2010). Entre las razones para el nombre para los países subdesarrollados están que estos, entre 1950 y 1975 alcanzan su transición demográfica como del descenso de la mortalidad gracias a la difusión de los progresos sanitarios (Alcalde, 2010).

Este último punto parece aplicar al caso costarricense, según lo expone Rosero-Bixby (2004), quien afirma que “los logros en los primeros cincuenta años de salud pública en el país (de 1930 a 1980) se alcanzaron mediante el control de enfermedades transmisibles como las diarreicas, la malaria y la tuberculosis [...] estos tres grupos de causas de muerte explican por sí solos la mitad de la disminución de la mortalidad de 1930 a 1960” (2004)

La Teoría de la transición epidemiológica (TTE) surge como alternativa teórica de la ya menciona Teoría de transición demográfica (TTD) debido a la naturaleza descriptiva o evolucionista de esta. La TTE tiene un carácter más multidisciplinar y con esto multifactorial.

Esta teoría establece que los principales factores causantes de las transiciones demográficas son los factores ecobiológicos de la mortalidad, factores biopsicológicos, como el uso de contraceptivos ante el aumento de la supervivencia infantil generando un aumento de la EV, factores médicos y de salud pública, factores psicológicos o emocionales, y factores socioeconómicos, ya que los desarrollos económicos establecen los primeros sistemas sanitarios, ayudando así al descenso de la mortalidad y la reducción de la incidencia de las enfermedades infecciosas.

Es importante mencionar que tal como lo establece Mariscal de Gante & Rodríguez (2018) la TTE, tiene como objeto de estudio la mortalidad, ya que se centra en aspectos como: patrones de enfermedad, causas de las muertes y la interacción de estas con patrones demográficos, económicos y sociológicos.

Descripción detallada de los datos

La tabla de datos proviene del Instituto Nacional de Estadística y Censos (INEC) de Costa Rica y es de acceso público, descargable desde la página web del instituto, como puede consultarse en INEC (2021). Esta base presenta los principales indicadores demográficos anuales de Costa Rica durante el periodo 1950-2020. Incluye en total 18 variables, entre las cuales están el año, la población total al 30 de junio de cada año, desagregado también por sexo, así como la cantidad de defunciones. Las variables presentes en la tabla son las siguientes: **Población de estudio:** La población de estudio son aquellas personas que vivían en Costa Rica entre los años 1950-2020 y mueren en este periodo. **Muestra observada:** La muestra observada, son todas aquellas personas que vivían en Costa Rica y al morir son registrados por el Instituto Nacional de Estadística y Censo. **Unidad estadística o individuos:** La unidad estadística es el recuento anual de defunciones en Costa Rica. **Variables de estudio:** Son un total de 18 variables, las cuales según INEC (2020) se tiene: **Año:** Esta variable indica el año. **Total:** Esta variable registra la población total. **Hombres:** Esta variable registra la población total de hombres. **Mujeres:** Esta variable se encarga de registrar el total de mujeres. **Nacimientos:** Esta variable registra el total de nacimientos. **Defunciones:** Esta variable registra el total de defunciones. **Defunciones infantiles:** Esta variable registra las defunciones de infantes (niños y niñas). **Defunciones neonatales:** Esta variable registra las defunciones de recién nacidos, hace referencia a la mortalidad de los nacidos antes de alcanzar los 28 días de edad. **Defunciones fetales:** Esta variable registra las defunciones de fetos, se refiere a la mortalidad de un bebé antes o durante el parto. **Tasa de crecimiento:** Esta variable registra la tasa de crecimiento de la población costarricense. Se refiere al crecimiento de la población entre dos fechas sin contemplar la migración. **Tasa de natalidad:** Esta variable registra la tasa de nacimientos registrados en la población costarricenses. **Tasa de mortalidad:** Esta variable registra la tasa de muertes en su totalidad. **Tasa de mortalidad infantil:** Esta variable registra la tasa de muertes infantiles. **Tasa de mortalidad neonatal:** Esta variable registra la tasa de muertes neonatales. **Tasa de mortalidad fetal:** En esta variable se registra la tasa de muertes fetales. **Tasa global de fecundidad:** Esta variable registra la tasa de fecundidad global (TGF), la cual indica cantidad de hijos e hijas que en promedio tendría cada mujer al final del periodo fértil, si durante su vida tiene sus hijos e hijas de acuerdo a las tasas de fecundidad por edad observadas en el país y año de interés y, además estas mujeres no están afectadas por la mortalidad desde el nacimiento hasta el final de periodo fértil. **Tasa bruta de reproducción:** Esta variable hace referencia a el cantidad de hijas que en promedio tendría cada mujer al final del periodo fértil, si durante su vida tiene sus hijos e hijas de acuerdo a las tasas de fecundidad por edad observadas en el país y año de interés y, además estas mujeres no están afectadas por la mortalidad desde el nacimiento hasta el final de periodo fértil. **Tasa neta de reproducción:** es el número de hijas que en promedio tendría cada mujer al final del periodo fértil, si durante su vida tiene sus hijos e hijas de acuerdo a las tasas de fecundidad por edad observadas en el país y año de interés y, además estas mujeres sí están afectadas por la mortalidad por edad observada en el país y año de interés desde el nacimiento hasta el final de periodo fértil.

Tabla 1: Resumen de cinco números para algunas variables de la tabla de datos

	Población total	Población de hombres	Población de mujeres	Nacimientos	Defunciones	Defunciones infantiles	Defunciones neonatales	Defunciones fetales
Mínimo	868 934	438 185	430 749.0	37 248	8 596.0	462.0	344.0	329
Primer cuartil	1 638 283	827 590	810 692.5	59 589	9 897.5	765.0	537.0	526
Mediana	2 646 142	1 339 573	1 306 569.0	70 004	11 376.0	1 345.0	807.0	691
Tercer cuartil	4 054 418	2 053 052	2 001 366.5	75 794	15 704.5	3 543.5	1 173.5	959
Máximo	5 111 238	2 575 550	2 535 688.0	84 337	26 209.0	4 889.0	1 688.0	1 261

Tabla 2: Estadísticos de dispersión para algunas variables de la tabla de datos

	Población total	Población de hombres	Población de mujeres	Nacimientos	Defunciones	Defunciones infantiles	Defunciones neonatales	Defunciones fetales
Desviación estándar	1 340 677	677 255.2	663 434.1	11 274.6	4 428.495	1 396.199	366.604	241.5763
Rango intercuartílico	2 416 136	1 225 461.5	1 190 674.0	16 205.0	5 807.000	2 778.500	636.500	433.0000

Análisis descriptivo de los datos

En la tabla Tabla 1 se detalla el resumen de cinco números para algunas variables relevantes de la tabla de datos. Se observa un rango bastante amplio (diferencia entre el máximo y mínimo) en las defunciones totales, lo que es de esperarse dado el crecimiento poblacional que se vio a partir de los 1980's y la cantidad de años que se registran en los datos.

En la tabla Tabla 2 se detallan los estadísticos de desviación estándar y rango intercuartílico para algunas variables relevante de la tabla de datos. Se observa que la variable población de hombres presenta mayor dispersión respecto población de mujeres, esta mayor variabilidad en la población de hombres se puede deber a diversos factores, entre ellos : una diferencia en el número de defunciones, nacimientos, o número de migrantes en la población masculina en relación con la de población femenina.

Una observación importante, es que los nacimientos presentan una mayor dispersión respecto a el número de defunciones. Es decir, se observa que el número defunciones anuales presenta menor desviación respecto a la media (desviación estándar) y una menor diferencia entre tercer cuartil y primer cuartil de defunciones (IQR).

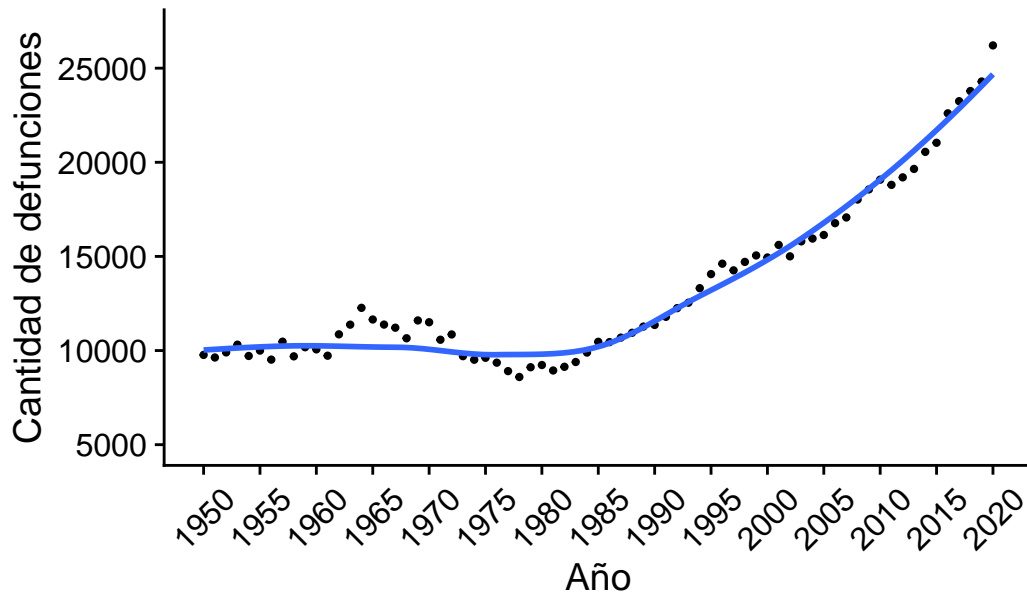
Análogamente, se observa que para los tipos de defunción: infantil, neonatales y fetales. Se evidencia una mayor dispersión para las defunciones infantiles, seguidas de las neonatales y por último, con menor dispersión las defunciones fetales.

Finalmente, se observa una diferencia significativa entre la medidas de dispersión desviación estándar y rango intercuartílico (IQR), esto se debe a la sensibilidad de la desviación estándar a valores extremos, es decir es posible que existan valores extremos (muy alto o muy bajos respecto a la media) y por esta razón ambas medidas difieran considerablemente, sin embargo para este caso en particular el orden en el grado de dispersión (observe que la dispersión disminuye al avanzar en la tabla de izquierda a derecha) se mantiene para ambos medidas.

De la Tabla 3 se observa que históricamente las defunciones totales de la mayoría de los años cae dentro del primero y segundo intervalo, con una distribución más uniforme dentro de los intervalos más altos. Por su parte la tasa de mortalidad bruta, la cual se aprecia en las columnas cinco y seis, se muestra que en la mayoría de años la tasa se ha mantenido entre 3.72 y 4.41. Esto quiere decir que la cantidad de defunciones relativo a la población ha sido más o menos constante en la mayoría de años. Por su parte, la población total se ha distribuido más uniformemente en todos los intervalos por lo que se concluye que el crecimiento población se ha movido más rápido que las defunciones.

Tabla 3: Frecuencia de defunciones, población y tasa de mortalidad

Defunciones	Frec. de defunciones	Población total	Frec. de poblacion	Tasa de mortalidad	Frec. de tasa
(8.58e+03,1.04e+04]	23	(8.65e+05,1.29e+06]	11	(3.72,4.41]	39
(1.04e+04,1.21e+04]	18	(1.29e+06,1.72e+06]	9	(4.41,5.1]	8
(1.21e+04,1.39e+04]	4	(1.72e+06,2.14e+06]	8	(5.1,5.78]	1
(1.39e+04,1.56e+04]	8	(2.14e+06,2.57e+06]	6	(5.78,6.46]	3
(1.56e+04,1.74e+04]	5	(2.57e+06,2.99e+06]	6	(6.46,7.14]	3
(1.74e+04,1.92e+04]	4	(2.99e+06,3.41e+06]	5	(7.14,7.83]	3
(1.92e+04,2.09e+04]	3	(3.41e+06,3.84e+06]	5	(7.83,8.51]	6
(2.09e+04,2.27e+04]	2	(3.84e+06,4.26e+06]	6	(8.51,9.19]	1
(2.27e+04,2.44e+04]	3	(4.26e+06,4.69e+06]	7	(9.19,9.88]	3
(2.44e+04,2.62e+04]	1	(4.69e+06,5.12e+06]	8	(10.6,11.3]	4



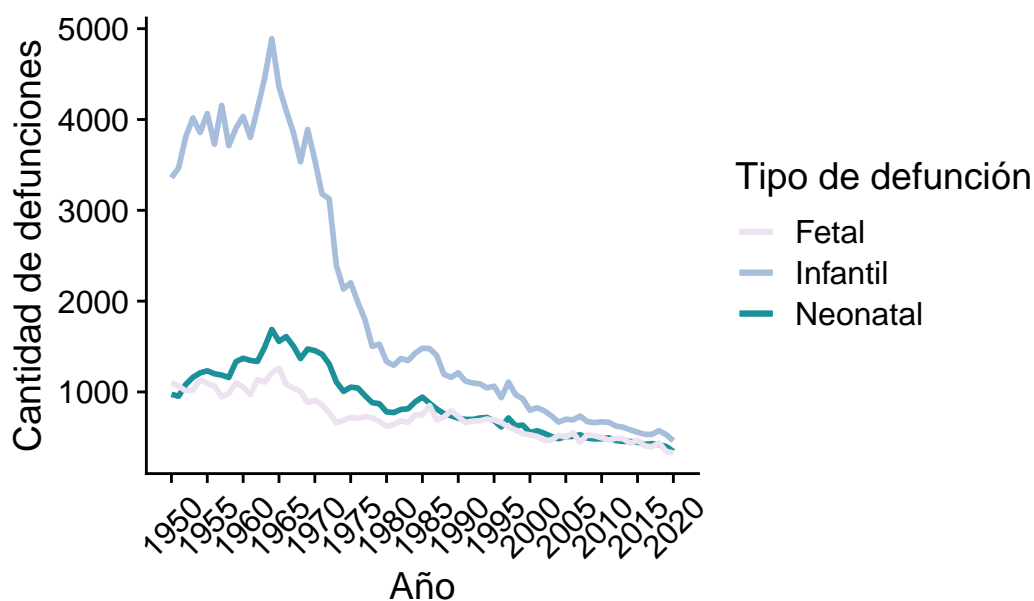
Fuente: Elaboración propia

Figura 1: Cantidad de defunciones por año para el periodo 1950-2020

En la Figura 1 se muestra la cantidad total de defunciones para el periodo 1950-2020. Destaca una tendencia creciente muy marcada a partir de cerca de 1980 y hasta el final del periodo considerado. Esto muestra que se debe hacer una transformación a los datos para lograr la estacionariedad que supone el modelo ARMA. También se resalta que para el 2020 las defunciones totales se encuentran claramente por arriba de la tendencia lo cual se debe probablemente a la pandemia del Covid-19.

En la Figura 2 se comparan las defunciones infantiles, neonatales y fetales. Cabe añadir que la distancia vertical entre las defunciones infantiles y las neonatales resulta en las llamadas defunciones posneonatales, es decir, las que ocurren a partir de

los 29 días de edad y hasta un año. Se advierte que a mediados de los años sesenta la cantidad de defunciones infantiles aparenta tener una tendencia decreciente. Al respecto, Rosero Bixby afirma que la caída más dramática en los años setenta “se logra gracias a los programas de atención primaria de la salud, ayudados por una extraordinaria reducción de la natalidad que permite un mejor desarrollo intrauterino, mejor cuidado del niño y reduce el riesgo de contagio” (2004). El mismo autor menciona que “el riesgo de morir de los menores de un año ha disminuido en forma poco menos que espectacular entre 1970-78, pues ha sido reducido a la tercera parte (de 62 a 22 muertes por cada mil nacimientos) en un lapso de apenas 8 años” (Rosero-Bixby, 2016). Por su parte, la cantidad de defunciones neonatales superó a las fetales desde mediados de los años cincuenta y hasta mediados de los años ochenta, donde se pierde un poco la noción de cuál suele ser mayor.



Fuente: Elaboración propia con datos del INEC

Figura 2: Defunciones infantiles, neonatales y fetales por año

Para estudiar la asociación entre la tasa de defunciones neonatales y la de defunciones fetales, se crea una variable categórica indicadora de si la primera es mayor a la segunda. Dado que ambas tasas se calculan respecto al mismo denominador (la cantidad de nacimientos), esta variable equivale a hacer lo mismo con la cantidad de defunciones respectiva. En la Figura 3 se muestra la distribución de esta variable y se ve que en 15 de los 71 años considerados la tasa de mortalidad fetal fue mayor, donde de la Figura 2 se sabe que esto sucedió mayoritariamente entre 1985 y 2020. En consecuencia, en 56 años fue mayor la tasa de mortalidad neonatal, lo cual sucede sobre todo en los primeros 35 años a partir de 1950. Entonces, en los 71 años de estudio, 78.87% de las ocasiones fue mayor la tasa de mortalidad neonatal, contra un 21.13% de la fetal.

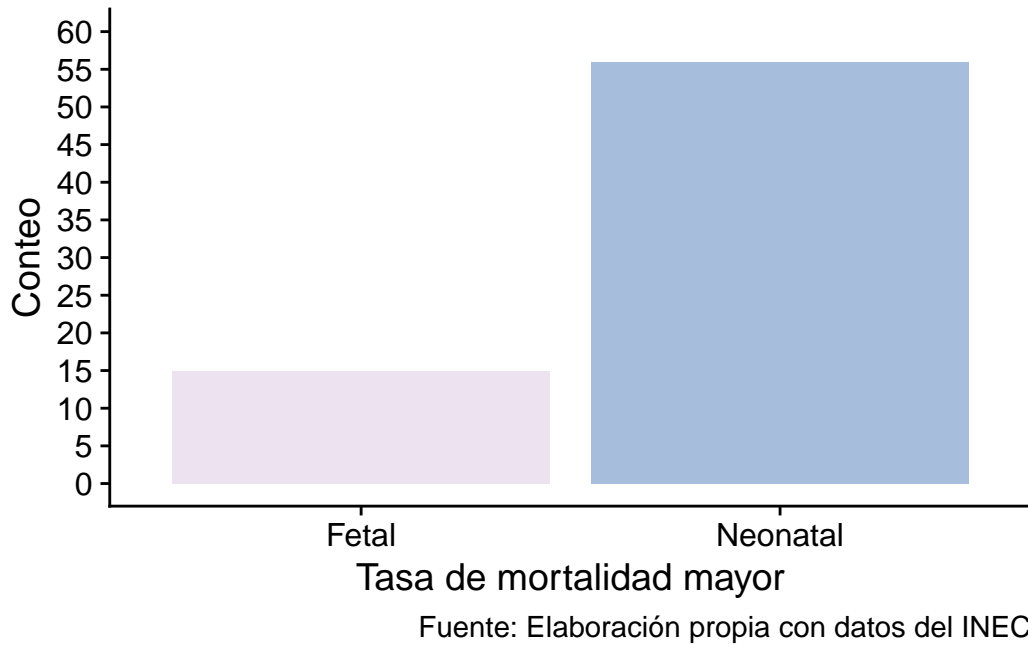


Figura 3: Distribución variable indicadora de la mayor tasa entre las de mortalidad fetal y la neonatal

Modelos estadísticos propuestos

Con la finalidad de realizar un pronóstico de la serie de defunciones totales anuales de Costa Rica, se desea implementar el modelo estadístico que mejor se ajuste a los datos.

Para nuestro estudio en cuestión, se ha optado por realizar una implementación de Modelos de Espacio-Estado. Particularmente, Modelos Dinámicos Lineales (DLM).

Tal como lo establece Petris et al. (2007), estos últimos son una clase de Modelos de Espacio-Estado también llamados Modelos de Espacio-Estado Lineales Gaussianos. Estos modelos son especificados mediante dos ecuaciones, para $t \geq 1$ se tiene:

$$Y_t = F_t \theta_t + v_t,$$

$$\theta_t = G_t \theta_{t-1} + w_t$$

$$(\theta_0 | D_0) \sim \mathcal{N}(m_0, C_0)$$

Donde la primer ecuación es llamada ecuación de observación, la segunda ecuación estado o ecuación del sistema y la última información inicial.

Es importante señalar que F_t y G_t son matrices y (v_t) , (w_t) son secuencias de ruidos blancos independientes tales que:

$$v_t \sim \mathcal{N}_m(0, V_t),$$

$$w_t \sim \mathcal{N}_p(0, W_t)$$

Los DLM poseen dos supuestos, la linealidad y el supuesto de distribuciones Gaussianas. Petris et al. (2007) señala que este último supuesto puede ser justificado mediante argumentos del teorema del límite central.

La estimación y pronóstico se pueden resolver calculando las distribuciones condicionales de las cantidades de interés, dada la información disponible. Para estimar el vector de estados es necesario computar la densidad condicional $p(\theta_t | y_1, \dots, y_t)$. En particular, nos interesa el problema de filtrado (cuando $s = t$), donde los datos se supone que llegan secuencialmente en el tiempo.

En general, el problema de pronóstico de k -pasos hacia adelante consiste en estimar la evolución del sistema θ_{t+k} para $k \geq 1$ y realizar un pronóstico de k -pasos para Y_{t+k} .

Según Petris et al. (2007) en los DLM, el filtro de Kalman proporciona las fórmulas para actualizar nuestra inferencia actual sobre el vector de estado conforme se disponga de nuevos datos.

Para un DLM, si se cumple que:

$$\theta_t | \mathcal{D}_t \sim \mathcal{N}(m_t, C_t), t \geq 1$$

Se tiene que:

La densidad de predicción de estado de k -pasos con $k \geq 1$ hacia adelante de θ_{t+k} dada la información pasada D_t es Gaussiana con media y varianza condicional dadas respectivamente por:

$$a_t(k) = E[\theta_{t+k} | D_t] = G_{t+k} a_{t,k-1}$$

$$R_t(k) = Var[\theta_{t+k} | D_t] = G_{t+k} R_{t,k-1} G_{t+k}' + W_{t+k}$$

La densidad de predicción de k -pasos con $k \geq 1$ hacia adelante de Y_{t+k} dada la información pasada D_t , es Gaussiana con media y varianza condicional dadas respectivamente por:

$$f_t(k) = E[Y_{t+k} | D_t] = F_{t+k} a_t(k)$$

$$Q_t(k) = Var[Y_{t+k} | D_t] = F_{t+k} R_t(k) F_{t+k}' + V_{t+k}$$

Justificación modelos DLM propuestos

Como se mencionó en Figura 1, la cantidad de defunciones totales siguen una cierta tendencia lineal creciente, en particular para años posteriores a 1980.

Debido a que esta es nuestra variable de interés para realizar un pronóstico, es propicio para nuestro estudio en cuestión la implementación de un modelo con supuesto de linealidad, como se mencionó justamente los DLM siguen este supuesto.

Para llevar a cabo los pronósticos se proponen por tanto tres métodos estadísticos pertenecientes a los DLM, estos son: modelo DLM polinomial de primer orden, modelo DLM polinomial de segundo orden y el modelo ARIMA(p, d, q).

Modelo ARIMA(p, d, q)

Como una primer implementación se utiliza un modelo ARIMA(p, d, q). Tal como lo menciona Petris et al. (2007) un modelo ARIMA(p, d, q) puede ser considerado un DLM, esto ya que es posible representar todo modelo ARIMA(p, d, q) (ya sea univariado o multivariado) como un DLM.

La escogencia de este modelo al ser un DLM, sigue la misma línea de justificación antes mencionado sobre la elección de modelos DLM para nuestro estudio, siendo este un caso particular de estos.

Sin embargo, es importante mencionar que la escogencia de este modelo como primera implementación también se basa en su simplicidad, y en que dada la bibliografía consultada, se observa que en múltiples investigaciones con temáticas relacionadas a nuestro estudio como el de Adekanmbi et al. (2014) y el estudio por Ordorica (2004), se implementa este tipo de modelo.

Tal como lo establece Petris et al. (2007) entre los modelos más utilizados para el análisis de series temporales se encuentra la clase de modelos de media móvil autorregresiva (ARMA). Para enteros, no negativos p y q , un modelo ARMA(p, q) es definido mediante la notación:

$$Y_t = \mu + \sum_{j=1}^p \phi_j (Y_{t-j} - \mu) + \sum_{j=1}^q \psi_j \epsilon_{t-j} + \epsilon_t$$

Donde (ϵ_t) es una ruido blanco Gaussiano con varianza σ_ϵ^2 y los parámetros $\phi_1, \phi_2, \dots, \phi_p$ satisfacen la condición de estacionariedad.

Cuando los datos no presentan estacionariedad, se suele tomar las diferencias hasta que se obtenga esta, una vez obtenida se procede a ajustar el modelo ARMA(p, q) a la data diferenciada.

Un modelo para un proceso cuya d -ésima diferencia sigue un modelo ARMA(p, q) es llamado un ARIMA(p, d, q).

La escogencia de los ordenes p y q pueden ser escogidos de una manera informal, observando la autocorrelación empírica y la autocorrelación parcial, o utilizando un criterio de selección de modelos más formal como lo es el AIC y BIC.

Modelos polinomiales de primer y segundo orden

Se propone un modelo DLM de primer orden ya que como establece Mary (2006) los DLM de primer orden son algoritmos recomendados al lidiar con datos anuales debido a que las series de tiempo es corta y no presentan patrones estacionales. Dado que nuestros datos son anuales, este modelo se presenta como un posible candidato.

Por su parte Mary (2006), señala que los DLM de segundo orden son útiles para describir tendencias. Dada la tendencia observada de la serie de defunciones totales sugiere por tanto realizar un modelo polinomial de segundo orden.

Es oportuno señalar que el desarrollo teórico de estos modelos se llevará a cabo en bitácoras posteriores, ya que se considera prudente primero realizar la implementación de estos (tal como se realizó el $ARIMA(p, d, q)$) para ver sus alcances para responder la pregunta de investigación.

UVE de Gowin

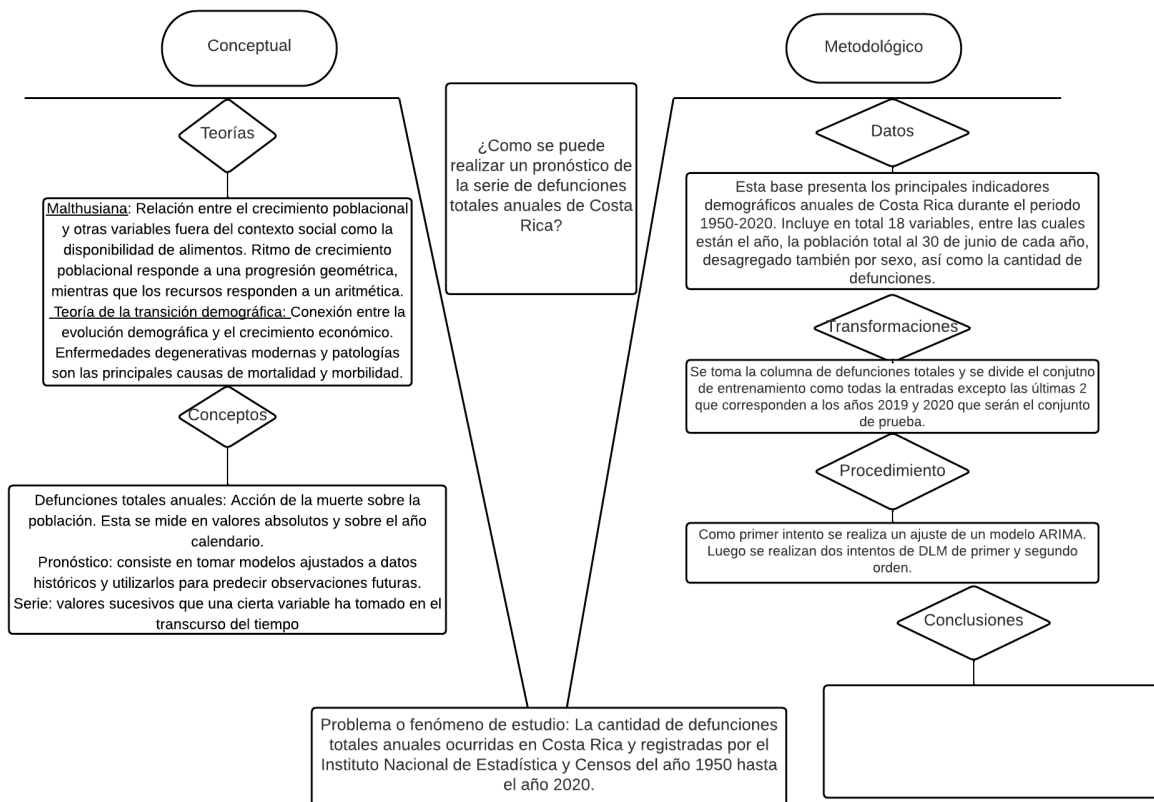


Figura 4: UVE Heurística

Referencias

- Adekanmbi, D., Ayoola, F., & Idowu, A. (2014). Demographic Time Series Modelling of Total Deaths in Nigeria. *Population Association of Southern Africa*, 15(1), 21-48.
- Alcalde, F. P. (2010). La teoría de la transición demográfica: recursos didácticos. *Enseñanza de las ciencias sociales: revista de investigación*, 129-137.
- Brownlee, J. (2020). *Introduction to time series forecasting with Python*. eBook.
- Bruno Ramírez, X. S. (2019). *Análisis del crecimiento y proyección poblacional del distrito Chulucanas, 2017-2025*.
- INEC. (2020). *Indicadores demográficos 2019*.
- INEC. (2021). *Estadísticas demográficas. 1950-2020. Principales Indicadores Demográficos*. %7B<https://www.inec.go.cr/documento/estadisticas-demograficas-1950-2020-principales-indicadores-demograficos%7D>
- Macció, G. A., Centro Latinoamericano de Demografía, S., et al. (1985). *Diccionario demográfico multilingüe*.
- Mariscal de Gante, Á., & Rodríguez, V. (2018). *Tres teorías demográficas, las evidencias disponibles y el paso de la descripción del cómo al entendimiento del porqué: Una aplicación y una crítica de tres hipótesis poblacionales en base a los casos de España y de la India (1950-2020)*.
- Mary, K. (2006). *Determining optimal architecture for dynamic linear models in time series applications*.
- Ordorica, M. (2004). Pronóstico de las defunciones por medio de los modelos autorregresivos integrados de promedios móviles. 249–264, 10(42), 21-48.
- Petris, G., Petrone, Sonia, & Patrizia, C. (2007). *Dynamic Linear Models with R*. Springer.
- Rees, P. (2020). Demography. En A. Kobayashi (Ed.), *International Encyclopedia of Human Geography (Second Edition)* (Second Edition, pp. 239-256). Elsevier. <https://doi.org/10.1016/B978-0-08-102295-5.10252-5>
- Rosero-Bixby, L. (2004). *Situación demográfica general de Costa Rica*.
- Rosero-Bixby, L. (2016). La situación demográfica en Costa Rica. *Población y Salud en Mesoamérica*, 13(2), 237-311.
- Sarrible, G. (2009). *Teoría de la población*.