UNIVERSIDAD DE COSTA RICA

MODELOS LINEALES

Pronóstico demográfico de Costa Rica

Autores
David Zumbado
Leonardo Blanco
Ignacio Barrantes

5 de noviembre de 2022

Indice de contenidos

Listado de Figuras

Listado de Tablas

Capítulo 1

Teorías demográficas

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 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 alución 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)

Capítulo 2

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:

$$\begin{aligned} Y_t &= F_t\theta_t + v_t, \\ \theta_t &= G_t\theta_{t-1} + w_t \end{aligned}$$

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

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:

$$\begin{aligned} v_t &\sim \mathcal{N}_m(0, V_t), \\ w_t &\sim \mathcal{N}_p(0, W_t) \end{aligned}$$

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 $p(\theta_t|y_1, .., y_t)$. 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:

$$f_t(k) = E[Y_{t+k}|D_t] = F_{t+k}a_t(k)$$
$$Q_t(k) = Var[Y_{t+k}|D_t] = F_{t+k}R_t(k)F^{'}_{t+k} + V_{t+k}$$

2.1 Justificación modelos DLM propuestos

Como se mencionó en **?@fig-defunciones_ano** , 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).

2.2 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:

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

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

2.3 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.

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.

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 serires 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.

Rosero-Bixby, L. (2004). *Situación demográfica general de Costa Rica*.