

# PROYECCIÓN DE POBLACIÓN TOTAL Y POR DEPARTAMENTOS DE COLOMBIA USANDO TABLAS DE VIDA MULTIESTADO

MULTI STATE LIFE TABLE PROJECTION PER  
PROVINCES AND TOTAL NATIONAL FOR COLOMBIA

*B. Piedad Urdinola<sup>1</sup>*  
*Miguel A. Lara<sup>2</sup>*

---

1 Profesora Asociada, Departamento de Estadística, Universidad Nacional de Colombia-Bogotá, <bpuardinolac@unal.edu.co>

2 Magister en Estadística, Universidad Nacional de Colombia-Bogotá, <malarag@unal.edu.co>

Este documento presenta resultados de una proyección multirregional utilizando modelos multiestado, que incorpora la desagregación regional como una transición adicional, y permite proyecciones subnacionales y nacional sin necesidad de ajuste forzoso. Para corregir las fallas de la información de mortalidad y natalidad se conformaron series de tiempo anuales por departamento desde 1921 hasta 2011 y los flujos netos de migración de los censos nacionales de población de 1973, 1993 y 2005. Esta información permite construir Matrices de Leslie por sexo y grupos de edades quinquenales para 28 departamentos y los departamentos de la región Amazonía agrupados en uno sólo, para una matriz de dimensión 986x986, correspondiente a 17 grupos de edad, dos géneros y 29 departamentos. Se encuentran diferencias sustanciales para las proyecciones de los departamentos con más y menos de un millón de habitantes, lo que sugiere convergencia entre estos dos grupos de departamentos para el caso colombiano.

**Palabras clave:** Modelos Multiregionales de Proyección. Matrices de Leslie multirregionales. Proyecciones Departamentales de Población Colombia. Series anuales demográficas departamentales para Colombia.

This document presents the results of a multiregional projection using multi-state models, which incorporates the regional desegregation as an additional transition and therefore produce both sub national and national projection, without needing any adjustment to meet both levels. To correct for mortality and birth data unusual patterns, annual time series were established by department from 1921 to 2011 and the net migration flows from national population censuses in 1973, 1993 and 2005. This allows the construction of Leslie matrices for each sex and five-year age groups for 28 departments and departments of the Amazon region are grouped into one, for a final matrix of 986x986 dimension, corresponding to the 17 age groups, the two genders and 29 departments. There are substantial differences for the projections of departments with more than one million inhabitants, suggesting convergence between these two groups of departments for the Colombian case.

**Keywords:** Multiregional Projection Models. Multiregional Leslie Matrices. Departmental Population Projections Colombia. Annual Departmental Demographic Series for Colombia.

## INTRODUCCIÓN

En plena era de globalización, cuando el mundo se encuentra más interconectado que nunca, los modelos demográficos se vuelcan sobre lo regional. Esta aparente contradicción lo único que demuestra es la necesidad por progresar en la calidad de la información de fuentes demográficas con fines prácticos de planeación y mejoramiento de la calidad de vida de las sociedades que, cada vez, queremos medir con mayor detalle. La medición de datos regionales cobra mayor importancia porque nos permite no sólo conocer las necesidades de subpoblaciones o áreas subnacionales, sino que también mejora la predicción y proyección del total de la población en modelos nacionales. El presente trabajo presenta los resultados de las proyecciones de población para el caso colombiano, a partir de las proyecciones subnacionales, lo que permite tener en completa concordancia y sin necesidad de corrección alguna ambas proyecciones: nacionales y subnacionales. Este modelo se alimenta de la información de los departamentos del país, para producir al final una gran proyección mucho más refinada de población nacional, a partir de matrices multiestado que no necesitan de ningún ajuste adicional, a diferencia de otros métodos de proyección.

Si bien el público general considera las proyecciones de población como una predicción del futuro, los demógrafos y expertos en pronósticos y proyecciones preferimos darles una lectura sobre la pregunta qué debemos hacer hoy para construir el mañana, en lugar de preguntarnos: ¿cómo será el mañana? En otras palabras, las proyecciones deben ser consideradas como un instrumento de creación más que de adivinación del futuro (INE, 2005). Con esta visión en mente, las necesidades y usos de las proyecciones de población no se ponen en duda. Son ampliamente utilizadas tanto en el sector público y privado con fines de planeación urbana, infraestructura y financiera; distribución de recursos, sobre todo en salud y educación; planes de mercadeo y pronósticos del crecimiento mismo de la población a futuro, cruciales para los balances fiscales a partir del sistema de seguridad social, entre otras (Okita et al, 2009; Tuljapurkar, 2006; Anderson et al, 2001).

Desde los primeros modelos de proyección de población propuestos en la primera mitad del siglo XX por Whelpton (1928; 1936; et al, 1947), que nos dejaron como enseñanza los escenarios bajo, medio y alto mucho se ha avanzado en el tema, hasta llegar a los métodos modernos que incorporan las proyecciones estocásticas de población y los métodos Bayesianos (por ej. Lee y Tuljapurkar, 1994; Wiśniowski et al, 2015). Sin embargo, las diferencias entre los métodos matemáticos, meramente demográficos y los más

recientes estocásticos todavía no varían diametralmente en sus resultados de proyecciones nacionales finales; la diferencia radica en que los últimos incorporan una importante medición de la incertidumbre que generan los modelos de proyección. A medida que avanzaban estas técnicas, aplicadas únicamente a totales nacionales, surgió una demanda por modelos de proyección de áreas pequeñas o subnacionales (Swanson y Tayman, 2012) que, como su nombre lo dice, se han especializado en trabajar modelos para poblaciones mucho más pequeñas que las nacionales, y para las cuales las metodologías tradicionales no producen resultados tan robustos.

No obstante, la combinación de metodologías que proyectan el total nacional y las áreas subnacionales no necesariamente se concilian. Es decir, el resultado de la proyección nacional de población utilizando modelos de orden nacional no necesariamente coincide con los resultados producidos por modelos de áreas subnacionales, una vez éstas se agregan. A pesar de ello, muchas oficinas de estadísticas nacionales siguen produciendo proyecciones de población de esta manera: con un modelo para áreas nacionales y otro para áreas subnacionales y luego se hacen coincidentes mediante algún tipo de ajuste forzoso, más allá de los ya necesarios para ajustar las fuentes de información indispensables para estas proyecciones en países, que como los Latinoamericanos, sufren de grandes deficiencias en la calidad de la información censal y de estadísticas vitales.

Ante los errores que esta práctica puede acarrear, este trabajo propone la implementación de modelos multirregionales o multiestado que garantiza la consistencia entre proyecciones a varios niveles de desagregación, por ejemplo entre las diferentes divisiones geográficas de un país — Departamentos, en el caso colombiano— y las proyecciones para el total país. Además, permite visualizar si existe algún tipo de convergencia regional entre dichas áreas y este conocimiento, sin duda ayudará a refinar las proyecciones totales, pues permite la diversidad de comportamientos de los fundamentales demográficos: fecundidad, mortalidad y migración, a las características inherentes en cada región.

En el modelo multiregional la incorporación de todas las divisiones geográficas y su proyección simultánea, garantizan que en cada una de éstas se tenga en cuenta las características propias, es decir, se asegura la consistencia interna de las cifras y los diferenciales demográficos regionales. Además, si se quieren obtener cifras a nivel nacional, éstas son obtenidas como agregación de cifras departamentales sin necesidad de ajustes. Para la aplicación de la metodología de los modelos multirregionales, se debe contar con información muy detallada de las características propias de cada departamento desagregada por edad y sexo. Esta información está

regida completamente por las tasas específicas de fecundidad, mortalidad y migración. De aquí que el modelo multiestado, originalmente por Rogers (1975) y modificado por Alho y Spencer (2005) como una generalización de las tablas de Leslie en edad y tiempo discreto, utilizando modelos multiestado, sea una generalización del modelo uniregional, en el sentido de que no solamente se plantean probabilidades de transición por edad sino también entre regiones geográficas.

Este trabajo utiliza la información de mortalidad por departamentos, tomada de *Latin American Human Mortality Database* (Urdinola y Queiroz, 2013) y los nacimientos se han recolectado de la información publicada por el Departamento Administrativo Nacional de Estadísticas (DANE); en series anuales desde 1960 hasta 2005 y los saldos migratorios se estiman a partir de la información de migrantes internos recientes de los censos de población Colombianos. Se construyen Matrices de Leslie para cada sexo y por grupos de edades quinquenales para 28 departamentos más uno que constituye la agrupación de los antiguos territorios nacionales en un solo ente regional, y al final se obtiene una matriz total con dimensión 986x986, correspondiente a los 17 grupos de edad, los dos géneros y 29 departamentos. Los resultados se contrastan con los obtenidos por la oficina oficial de estadísticas colombiana DANE, que utiliza el método demográfico de componentes en sus proyecciones oficiales subnacionales, que luego agrega y ajusta para tener coherencia entre las proyecciones subnacionales y el total nacional. Según el Censo de 2005 la población colombiana fue 41.468.384 habitantes, la proyección de este Modelo Regional Multiestado (MRM), para 2010 llega a 44.354.114 habitantes y para el año 2020 se espera observar un total de 50.826.528 habitantes. Esto sin corregir por ningún tipo de subregistro en las cifras vitales o el censo, pero sí haciendo una suavización siguiendo la tendencia temporal de las series de nacimientos que sufren de fluctuaciones muy amplias por departamentos. Esta limitación puede ser superada implementando un estudio que revise este subregistro subnacional y luego se incorpora fácilmente a este tipo de modelos, pues la corrección se haría a los insumos del modelo, mas no a la estimación de la proyección.

La preparación de las series departamentales anuales de las componentes de mortalidad y fecundidad muestran una tendencia a la convergencia entre grupos de departamentos. Es decir, en el corto plazo no se nota una convergencia total entre departamentos, sino que de momento se encuentra una convergencia entre dos grupos de departamentos, los que son más grandes, en términos poblacionales, y los que resultan con menor población. Es de destacar que los resultados a nivel nacional de este modelo multiregional, que no incorpora ningún proceso de corrección de

subregistro al censo nacional de población del año base (2005), son similares a los resultados del DANE, aunque para los años 2015 a 2050 el pronóstico de este ejercicio es mayor al de las estadísticas oficiales en cerca de 3,5%. Se encuentran diferencias importantes para las proyecciones departamentales, sobre todo para todos aquellos que tienen menos de un millón de habitantes. Este método es de fácil implementación computacional pues sólo consiste en multiplicación de matrices y a pesar de sus limitaciones puede incorporar correcciones de subregistro en el censo de población y en estadísticas vitales, lo que probablemente refine aún más las proyecciones de población sin necesidad de tener que hacer ajustes forzados entre las proyecciones nacionales y las subnacionales.

## MOTIVACIÓN

Los modelos poblacionales de proyección surgen de la necesidad de llevar una planeación, gubernamental, apropiada. De allí que los primeros modelos que se desarrollaron buscaban los totales nacionales, por las características propias que estudia la demografía: edad y sexo. Tal y como lo mencionan Alho y Spencer (2005) es mucho lo que se ha avanzado en la ciencia estadística y demográfica para refinar y alcanzar los modelos de proyección contemporáneos, al punto que tenemos una gran variedad de modelos: matemáticos, demográficos y estadísticos; más sus combinaciones. Es sólo hasta las décadas más recientes que se ha comenzado a trabajar en modelos que puedan aplicarse a datos subnacionales, o a grupos particulares como raza o etnia, niveles de educación, ingresos u otras características que subdividen la población, incluyendo ubicación geográfica. Para estos casos particulares se han desarrollado los modelos de áreas o poblaciones pequeñas, para los que un buen resumen de metodologías se presenta en Swanson y Tayman (2012).

En parte, esta deficiencia se debe a la falta de información demográfica para estas subpoblaciones y en parte porque no había una necesidad latente en la planeación gubernamental. Pero, a medida que se ha ido recolectando más información, se hacen evidentes las brechas existentes entre estratos o subpoblaciones al interior de cualquier país, de allí que la demanda por herramientas que permitan una mejor planeación subregional se ha incrementado. En particular, en los países Latinoamericanos las inequidades son realmente alarmantes, hasta el punto de ser catalogada como la región en el mundo con mayor inequidad persistentemente durante las últimas décadas, como lo demuestran las diferencias de la tasa de mortalidad infantil

por raza y residencia en la región (Jiménez et al, 2007), presentado en la Tabla 1.

Tabla 1  
América Latina, circa 1990 y 2000. Tasa de mortalidad infantil según condición étnica y zona de residencia (por mil nacidos vivos)

País	Año del censo	Total país		Área de residencia						
		Urbano			Rural					
		Indígena	Afro	Resto	Indígena	Afro	Resto	Indígena	Afro	Resto
Bolivia	1992	104,0	--	65,9	82,7	--	52,9	121,0	--	93,8
	2001	77,7	--	50,5	67,8	--	48,1	87,8	--	57,1
Brasil	1990	61,1	63,9	39,4	76,7	59,7	36,1	56,7	73,1	50,5
	2000	39,7	40,2	26,7	37,2	37,6	24,7	41,8	48,1	35,0
Chile	2002	12,8	--	11,5	12,0	--	11,4	12,7	--	12,0
Colombia	2005	39,5	31,7	23,9	30,9	30,0	23,5	39,8	33,8	25,0
Costa Rica	2000	20,9	11,2	11,5	20,4	8,2	10,7	21,5	15,1	12,3
Ecuador	1990	101,8	--	53,3	61,5	--	39,6	108,0	--	69,4
	2001	72,2	--	30,5	42,9	--	24,7	76,6	--	39,1
Guatemala	1994	61,1	--	49,7	56,9	--	41,0	62,1	--	55,2
	2002	51,1	--	41,0	47,2	--	35,3	52,6	--	46,7
Honduras	2001	43,5	31,9	34,5	27,1	30,1	25,1	45,2	33,7	41,8
México	1990	63,8	--	36,7	48,7	--	31,1	70,8	--	49,8
	2000	42,7	--	26,2	35,4	--	23,7	47,1	--	33,3
Nicaragua	2005	34,0	38,0	26,4	24,4	24,1	18,8	39,1	47,4	39,9
Panamá	1990	72,2	--	21,6	39,9	--	17,7	75,3	--	26,0
	2000	53,5	--	17,3	31,7	--	15,7	57,5	--	19,5
Paraguay	1992	96,2	--	45,8	90,8	--	45,9	96,5	--	45,8
	2002	78,5	--	37,7	72,1	--	38,8	79,1	--	36,3
Venezuela	2001	44,4	--	19,6	31,9	--	19,1	58,1	--	22,6

Fuente: Tomado de Jiménez et al, 2007

De allí, que sea fundamental contar con insumos demográficos que permitan evidenciar estas diferencias, precisamente para diseñar e implementar políticas que las corrijan. De otro lado, a medida que la transición demográfica ha avanzado en la región y los países han avanzado en niveles de desarrollo económico, ha aumentado la producción estadística

y demográfica que alimenta esta mejor planeación. Aún estamos lejos de los países con gran tradición demográfica o estadística, como los países Nórdicos, el Reino Unido o Estados Unidos. Entonces, la pregunta es cómo realizar proyecciones para un país como Colombia, con un nivel de desarrollo económico medio y que se encuentra en la fase avanzada de la transición demográfica (Tovar y Urdinola, 2014), pero que todavía cuenta con falencias importantes en la calidad de las estadísticas vitales y censales. Dependiendo de la fuente, los censos más recientes reportan niveles de subregistro de hasta el 25%, las defunciones de hasta 30% y los nacimientos de hasta 40% (Urdinola y Queiroz, 2013; Urdinola, 2011; Palloni et al, 2015; Urdinola y Ospino, 2015).

En primera instancia, esta falencia de datos de calidad no implica que debamos desistir de la idea de generar proyecciones de población subnacionales, por el contrario motiva a trabajar en modelos de proyección que corrijan estos problemas de subregistro. De allí que la propuesta de este trabajo es la incorporación de las variables regionales, no sólo para producir proyecciones de población a nivel subnacional, en el caso colombiano a nivel departamental, sino que también la definición de patrones regionales a través de los departamentos (provincias) del país permitan mayor refinación en los modelos de proyección total nacional, pues al incluir esta variable se está incorporando la información de los departamentos y la posible convergencia o divergencia entre ellos permite mejorar la proyección, aún en este escenario de calidad deficiente de la información primaria.

Bajo este escenario, la metodología propuesta es el de matrices multiestado o multiregionales, que hacen de la información subnacional un estado adicional a los tradicionales grupos de edad y sexo y que permite tener proyecciones consistentes entre los niveles nacionales y subnacionales. En 1945 Leslie popularizó los avances de Bernardelli y Lewis, como lo cita Wachter (2014), de las famosas matrices de transición. Este método ha sido ampliamente difundido y utilizado para la proyección de población y para particiones de la misma, ya que el tamaño de la población a proyectar no genera distorsiones en el modelo y sólo depende de la información recogida a partir de las componentes demográficas, pues debe incorporar las dos transiciones posibles: sobrevivir o dar vida. El modelo lineal multiestado considera la evolución que surge como consecuencia de las transiciones que ocurren entre periodos sucesivos de edad y tiempo a través de alguna categoría. Las personas que ocupan un mismo estado, pertenecientes a una misma categoría o grupo, constituyen una subpoblación. Cada grupo debe estar sujeto y gobernado por leyes de fecundidad y mortalidad propias. Por esta capacidad de generar categorías, este método ha sido ampliamente

utilizado y difundido entre demógrafos y otros científicos sociales, pues siempre y cuando se pueda hacer la analogía entre estas dos transiciones, se puede aplicar, por ejemplo, a temas tan variados como las transiciones en el sistema educativo, en el mercado laboral o el crecimiento económico por renglones.

La primera agencia oficial en introducir multidimensionalidad en sus proyecciones numéricas, basadas en tablas de vida con decrecimiento simple, y asociado con modelos de proyección de población de dimensión simple, fue el Bureau de Censos de Estados Unidos (Rogers, 1986). Una primera extensión de los modelos de proyección de poblaciones bajo estados simples enfocados en estados múltiples de residencia fueron llamados modelos de proyección multiregional (Rogers, 1975). Es claro que la proyección de poblaciones clasificada por múltiples estados de existencia podría hacerse usando una metodología común de proyecciones multiestado en la cual el modelo central de la dinámica poblacional fuera una generalización de las tablas de Leslie en edad-tiempo discreto (Rogers, 1975).

El enfoque para este trabajo es el modelo poblacional de crecimiento lineal multiestado presentado en Alho y Spencer (2005). En este modelo, se realiza una generalización de las tablas de vida con decrecimiento simple al caso en el que se tienen más de dos estados de transición. Además, en los modelos poblacionales multiestado, el tamaño y la composición de una cohorte depende de la población inicial y de las transiciones de estado realizadas por los miembros de la población sumado al comportamiento de sus tasas vitales. Detalles de este modelo se presentan a continuación.

## METODOLOGÍA

El concepto principal para la construcción de este documento es el de tablas de vida multiestado, que constituyen el insumo inicial para la construcción de las matrices multiregionales. Una tabla de vida multiestado permite incorporar la noción que existen múltiples estados, a diferencia de una tabla de vida de cohorte, donde lo único que puede representar es el decaimiento de la población desde su inicio o nacimiento de la cohorte, hasta el final, cuando muere el último integrante de la cohorte. Los estados son una transición adicional, por ejemplo, ingresos y egresos de la población, debido a migración, entendiéndose estados como transiciones. Vale la pena aclarar que individuo que inicia en un determinado estado, puede pasar a otro estado y luego volver a su estado inicial. Por ejemplo, una persona nace en el estado soltero y se casa más tarde, luego hace la transición hacia el divorcio,



de manera que puede volver al estado de casado otra vez (Alho y Spencer, 2005: 168).

Supongamos que tenemos  $J$  estados, siguiendo los desarrollos matemáticos de Keyfitz y Caswell (2005) y con el propósito de generalizar la teoría de las tablas de vida originales, en donde se analizaban las transiciones de sobrevivencia representados por el escalar  $\mu(x)$ , se pretende determinar la estructura de la matriz  $\mu(x)$ , iniciando con las tasas instantáneas de movimientos entre estados. Para su construcción se deben tener en cuenta las siguientes características: los elementos fuera de la diagonal de  $\mu(x)$  son las correspondientes tasas observadas de los movimientos en un pequeño intervalo de tiempo con signo contrario. Así,  $-\mu(x)dx$  es el cambio (negativo) de una persona que encontrándose en el estado  $j$  se transfiera al estado  $i$  durante un período corto de tiempo y edad. Cada elemento de la diagonal de  $\mu(x)$  contiene las tasas  $\mu_{\delta_i}$  de morir, con signo positivo, junto con el total de los elementos fuera de la diagonal para la columna,  $\sum_{i \neq j} \mu_{ij}$ , también con signo positivo. Por ejemplo, asumiendo que se tienen tres estados, la matriz  $\mu(x)$  tiene la forma:

$$\mu(x) = \begin{bmatrix} \mu_{\delta_1}(x) + \sum_{i=2}^n \mu_{i1}(x) & -\mu_{12}(x) & -\mu_{13}(x) \\ -\mu_{21}(x) & \mu_{\delta_2}(x) + \sum_{i=1,3}^n \mu_{i2}(x) & -\mu_{23}(x) \\ -\mu_{31}(x) & -\mu_{32}(x) & \mu_{\delta_3}(x) + \sum_{i=1,2}^n \mu_{i3}(x) \end{bmatrix} \quad (1)$$

La notación indicada en la matriz  $\mu(x)$  corresponde al caso de tres estados y permite obtener algunos detalles de la matriz. El subíndice del lado derecho es el estado de origen y el subíndice del lado izquierdo es el estado destino. Por ejemplo,  $\mu_{23}(x)$  es el movimiento del estado 3 al estado 2 de las personas en edad  $x$  durante un intervalo dado.

Podemos tener en cuenta un supuesto más general, común en todas las tablas de vida, sin el cual los procesos demográficos no podrían ser representados: la probabilidad de que un individuo realice una transición depende solamente del estado en el que se encontraba al inicio de período de transición. Por ejemplo, el cambio de estado de un hombre de 55 años quien muere antes de los 60 años en las tablas de vida comunes, depende solamente de los efectos que tuvo a partir de los 55 años y los pertenecientes a su población.

Los modelos lineales multiestado permiten la estimación de estas matrices de transición multiestado, la siguiente descripción teórica sigue de cerca a Alho y Spencer (2005). Dada una población inicial, al comienzo del período inicial  $t$ , se utilizan las mismas unidades por edad y tiempo. Por ejemplo, las tasas vitales del año  $t$ , de los tres componentes demográficos: nacimientos, defunciones y migración neta, se refieren al tiempo  $[t, t+1)$ . La población es desagregada por sexo,  $s=1$  para las mujeres y  $s=2$  para los

hombres. La edad  $x=0$  hace referencia a los individuos con edad exacta en el intervalo  $[0,1)$ , edad  $x=1$  hace referencia al intervalo  $[1,2)$ , y así sucesivamente. La edad más alta posible está denotada por  $\omega$  y se refiere al intervalo  $(\omega, \infty)$ . En este caso, por disponibilidad de la información se tomará  $\omega=80$ . Los nacimientos son atribuidos sólo a las mujeres. La menor edad para procrear se denotará como  $\alpha$  y la edad más alta como  $\beta$ . Luego se tiene que  $0 < \alpha < \beta < \omega$ . En este caso particular, se define  $\alpha=15$  y  $\beta=49$ . Los estados adicionales, en este caso departamentos colombianos, se denotan por  $j=1,2,\dots,J$ .

Los tamaños de la población en el año  $t$  son denotados por un vector de la forma:

$$V(t) = (V(0,t)^T, \dots, V(\omega,t)^T)^T \quad (2)$$

Hay tres diferentes interpretaciones que se le pueden dar al vector (2), dependiendo del contexto. Primero, supóngase que tenemos una población de mujeres en una región; en este caso  $V(x,t)$  es un escalar que representa el número de mujeres en la edad  $x$ . Segundo, suponga una población conformada por hombres y mujeres; entonces,  $V(x,t) = (V_1(x,t), V_2(x,t))^T$ , donde  $V_1(x,t)$  es el número de mujeres en edad  $x$  y  $V_2(x,t)$  es el número de hombres en edad  $x$ . Tercero, suponga un sistema cerrado conformado por hombres y mujeres en las regiones  $j=1,2,\dots,J$ . Podemos escribir  $V(x,t) = (V_1(x,t), V_2(x,t))^T$ , donde  $V_1(x,t) = (V_{11}(x,t), \dots, V_{1j}(x,t))^T$  y  $V_{ij}(x,t)$  es el número de mujeres en edad  $x$ , en la región  $j=1,2,\dots,J$ . Análogamente  $V_2(x,t) = (V_{21}(x,t), \dots, V_{2j}(x,t))^T$  corresponde a los hombres.

Aritméticamente, los componentes de la cohorte para los tres casos pueden ser escritos matricialmente como:

$$V(t+1) = R(t)V(t) \quad (3)$$

Una vez que la matriz  $R(t)$  esté propiamente definida. El supuesto requerido en la ecuación (2) es que la población sea cerrada. Una extensión en la que se incluye la migración es la siguiente:

$$V(t+1) = R(t)V(t) + N(t) \quad (4)$$

Donde,  $N(t)$  es el número de migrantes netos durante el año  $t$ , y se obtiene de manera similar a  $V(t)$ . Las ecuaciones (2) y (3) pueden ser llamadas modelos de crecimiento lineal.

Tal como se hace con las matrices de Leslie, la matriz  $R(t)$  se determina suponiendo que tenemos una población de mujeres en un estado simple. En este caso tendríamos tasas de fecundidad por edad en la primera fila de  $R(t)$  y probabilidades de sobrevivencia en un año, en la subdiagonal. En el caso general de dos sexos y  $J$ -estados se tiene la misma estructura, pero en la primera fila y la primera subdiagonal tiene una estructura de  $2J \times 2J$

bloques de tasas<sup>3</sup>, correspondiente a la información desagregada por  $J$  estados y 2 sexos. Tal como se definió anteriormente, sea  $\alpha$  el comienzo de la edad reproductiva, en este caso se definirá como 15 años, y  $\beta$  el final de la edad reproductiva, en este caso se definirá como 49 años. En el caso especial en el que la población se compone sólo de mujeres y un estado simple, la matriz de transición tiene la siguiente forma (Feeney, 1970):

$$R(t) = \begin{bmatrix} 0 & \dots & \dots & 0 & R(0,\alpha,t) & \dots & R(0,\beta,t) & 0 & \dots & 0 \\ R(1,0,t) & 0 & \dots & \dots & \dots & \dots & \dots & \dots & \dots & 0 \\ 0 & R(2,1,t) & 0 & \dots & \dots & \dots & \dots & \dots & \dots & 0 \\ 0 & 0 & R(3,2,t) & \dots & \dots & \dots & \dots & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & \dots & \dots & \dots & \dots & \dots & R(\omega,\omega-1,t) & \dots & R(\omega,\omega,t) \end{bmatrix} \quad (5)$$

En general, para el caso de la población femenina se tiene que  $R(o,x,t)$  es el número esperado de niñas, nacidas durante el período  $t$  por mujeres en la edad  $x$ , que sobreviven el siguiente año.  $R(x,x-1,t)$  es la proporción de sobrevivientes de la edad  $x-1$  en el tiempo  $t$  a la edad  $x$  en el tiempo  $t+1$ ; y  $R(\omega,\omega,t)$  es la proporción de sobrevivientes en la edad  $\omega$ . Al incluir a los hombres se tiene que:

$$R(o,x,t) = \begin{bmatrix} R_1(o,x,t) & 0 \\ R_2(o,x,t) & 0 \end{bmatrix} \quad (6)$$

Donde  $R_1(o,x,t)$  es el bloque que contiene el número esperado de niñas, nacidas durante el periodo  $t$  por mujeres en la edad  $x$ , que sobreviven al comienzo del próximo año, y  $R_2(o,x,t)$  es el bloque que contiene el número esperado de niños, nacidos durante  $t$  por mujeres en la edad  $x$ , que sobreviven al comienzo del próximo año.

Para los sobrevivientes se tiene:

$$R(x,x-1,t) = \begin{bmatrix} R_1(x,x-1,t) & 0 \\ 0 & R_2(x,x-1,t) \end{bmatrix} \quad (7)$$

Donde  $R_1(x,x-1,t)$  es la proporción de mujeres que sobreviven de la edad  $x-1$  a  $x$  en el tiempo  $t$  y  $R_2(x,x-1,t)$  es la proporción de hombres;  $R_1(\omega,\omega,t)$  es definida similarmente.

En el caso multiregional,  $R(o,x,t)$  es una matriz de dimensión  $2J \times 2J$  que contiene cuatro bloques como los de la ecuación (5). Cada bloque es una matriz  $J \times J$ . La matriz  $R_1(o,x,t)$  tiene la forma:

$$R_1(o,x,t) = \begin{bmatrix} R_{111}(o,x,t) & R_{112}(o,x,t) & \dots & R_{11J}(o,x,t) \\ R_{121}(o,x,t) & R_{122}(o,x,t) & \dots & R_{12J}(o,x,t) \\ \vdots & \vdots & \ddots & \vdots \\ R_{1J1}(o,x,t) & R_{1J2}(o,x,t) & \dots & R_{1JJ}(o,x,t) \end{bmatrix} \quad (8)$$

<sup>3</sup> También existe una matriz de dimensión  $2J \times 2J$  en la esquina inferior derecha de  $R(t)$  que corresponde a la supervivencia a la edad  $\omega$

Donde  $R_{ij}(o,x,t)$  representa el número esperado de niñas nacidas por mujer de edad  $x$ , en la región  $j$  durante el tiempo  $t$  y que están vivas en la región  $i$  al final del año.

Las matrices  $R_2(o,x,t)=R_{2ij}(o,x,t)$ , para los hombres se definen de manera similar. Los dos bloques restantes, son matrices  $J \times J$  de ceros. Para el caso de los sobrevivientes, matrices  $2J \times 2J$  son definidas como en la ecuación (6), donde  $R_1(x,x-1,t)$  es una matriz  $J \times J$  que tiene  $(i,j)$  elementos  $R_{ij}(x,x-1,t)$  definidos como la proporción de mujeres de edad  $x-1$  en la región  $j$  durante el tiempo  $t$  que sobreviven en la región  $i$  al final del año. La definición para hombres es similar.

Asumiendo que se tiene una estimación de la población inicial  $V(0)$  y se cuenta con proyecciones  $\bar{R}(t)$  para  $t=0,1,\dots,T-1$ ; entonces las proyecciones de componentes de cohortes de  $V(t)$  es:

$$V(T) = \bar{R}(T-1) \dots \bar{R}(0)V(0) \quad (9)$$

## FUENTES DE INFORMACIÓN Y TRATAMIENTO

Las fuentes de información utilizadas incluyen como acervo de población por edad y sexo a los conteos de población del Censo Nacional de Población y Vivienda más reciente, que corresponde al de 2005. Para las componentes demográficas se logró la construcción de la una serie anual desde 1921 hasta 2011 por edad y sexo por departamentos en Colombia, para las componentes demográficas del crecimiento natural: nacimientos y defunciones. Así como la información de los saldos migratorios netos y el acervo de población a partir los Censo Nacionales de Población y Vivienda de 1973, 1993 y 2005. Desafortunadamente la calidad de la información para todo este período es baja, aunque mejora a medida que pasan los años, hecho que se refleja incluso en la producción misma de la información básica demográfica, como se describe a continuación. Teniendo estas limitaciones en cuenta, se construyeron series de tiempo más largas para las componentes demográficas, con el fin de capturar las tendencias de largo plazo y poder hacer imputaciones en los casos de datos faltantes o corregir los movimientos erráticos de las series temporales. Tanto para las defunciones como para los nacimientos se construye la serie de tiempo anual desde 1921 hasta 2011 de manera que se puede establecer para cada uno de los departamentos el comportamiento de la serie temporal y se estima la tendencia de largo plazo en las dos series.

## DEFUNCIONES Y NACIMIENTOS

La información de nacimientos y defunciones hacen parte desde 1998 del sistema de registros vitales, capturadas oficialmente por la oficina oficial de estadísticas nacionales-DANE. Las defunciones anuales desde 1970 han sido tomadas de la *Latin American Human Mortality Database* (Urdinola y Queiroz, 2013) que ha construido estas series anuales por grupos de edad quinquenal, sexo y departamento, a partir de las estadísticas de mortalidad oficiales del DANE, que es la entidad que oficialmente las recopila desde 1950. Para los años antes de 1970 la información ha sido recolectada de las publicaciones impresas oficiales de estadísticas anuales del DANE. Para antes de 1950 fue la Contraloría General de la Nación la entidad encargada de las estadísticas oficiales, y de sus boletines anuales se tomó la información de mortalidad, que fue digitalizada para completar la serie que llega hasta 1921.

En el caso de la mortalidad se utilizó como insumo la serie anual construida desde 1960 hasta 2011 por edad sexo y departamento, porque en esta serie sólo había algunos años faltantes para los departamentos de Caldas, Cesar y Magdalena que se completaron extrapolando la tendencia de la serie temporal. Vale la pena aclarar que no se hace ningún tipo de corrección por sub registro de la mortalidad porque las metodologías actualmente existentes para detectar y corregir este tipo de sub registro lo hacen apropiadamente para los niveles nacionales pero no para los subnacionales (Hill, 2003), proponer una metodología de corrección de subregistro subnacional correspondería a un trabajo adicional en sí mismo. Además, el subregistro de la mortalidad es probablemente diferencial a través de las regiones, siendo las regiones más apartadas, con menor densidad poblacional y menores niveles de desarrollo las de mayores niveles de subregistro y las mujeres sufren de mayor subregistro que los hombres (Urdinola et al, 2017).

Además, al igual que sucede en el total nacional, pueden haber diferenciales de subregistro también por grupos de edad. A nivel nacional está plenamente identificado que el subregistro es mayor para las edades extremas, las menores y las mayores, pero precisamente la falta de metodologías precisas a nivel subnacional no nos permiten identificar si este patrón se replica en todos los departamentos del país. De manera que aplicar una corrección total, tomada de alguna metodología nacional, no sería apropiado para generar una corrección año a año para las defunciones departamentales. Sin un estudio juicioso o una metodología sería que en efecto capture e identifique el nivel acertado de corrección de subregistro preferimos no hacer ningún tipo de corrección a nivel departamental.

En cuanto a la información de nacimientos el DANE ha recopilado intermitentemente estos datos por departamento. El sistema de estadísticas

vitales fue sólo creado en 1998 y desde entonces es realmente un sistema y la información se encuentra en formato electrónico, obtenido desde la página web del DANE. Similarmente, desde de la década de 1950 y hasta 1987 se recolectó la información publicada en medios físicos por las fuentes oficiales, como los anuarios y boletines estadísticos del DANE. Para los años que produjeron esta información, y antes de 1950 se encuentra en los boletines oficiales de la Contraloría General de la Nación. Los datos fueron digitalizados para su uso, análogo al caso de la mortalidad. Para los años entre 1987 y 1997 fue la Registraduría General de la Nación quien recolectó esta información. La serie se logró consolidar desde 1921 y hasta 2011.

Desafortunadamente para los datos de nacimientos hay algunos faltantes, para un puñado de departamentos en ciertos años particulares, y para el año 1976 se presenta un cambio de nivel importante en todos los departamentos. Para esos años/departamentos particulares se realizaron los siguientes cambios: la información faltante para el año 1924 fue imputada como el promedio de los nacimientos en cada departamento para los años inmediatamente anterior y posterior. A pesar de que se tiene información para los años 1968-1969 y para 1974-1976, ésta presenta problemas de nivel en la mayoría de departamentos, con un evidente subregistro para al menos cinco departamentos, con niveles muy por debajo a la mitad de nacimientos reportados en años circundantes. Este problema fue solucionado mediante la inclusión de esta información a partir de los reportes hechos por el DANE en los anuarios demográficos y series estadísticas. Es así como se logró, inicialmente, incluir y completar la información en el periodo 1968 hasta 1974. Desafortunadamente, el comportamiento de estas series anuales resultó bastante errática. Para solucionar este problema se hizo una interpolación lineal por departamento, que sigue el modelo de crecimiento exponencial, que utilizó como año inicial 1967 y el año final 1977 para suavizar estas fluctuaciones exageradas.

En el periodo 1975-1986 se presentan problemas de falta de información en algunos años y para otros solo se contaba con el conteo de nacimientos para el total nacional. La corrección a este problema fue establecer la distribución porcentual de cada departamento en los nacimientos ocurridos y registrados cada año, luego, se contrastó con un estudio del DANE el número total de nacimientos registrados a total país (DANE 1993) y se desglosaron por departamentos según la representación porcentual determinada previamente. Sin embargo, la tendencia histórica se veía afectada por comportamientos erráticos para estos años. Una vez más se suavizó este comportamiento haciendo una interpolación lineal teniendo como año inicial 1975 y año final 1988.



Análogo a lo que sucede con las defunciones no existen metodologías que corrijan el subregistro a nivel subnacional o estudios detallados previos al respecto, aunque sí existen para el nivel nacional. Las solas falencias aquí descritas en la información son una evidencia más de la baja calidad de esta información a nivel subnacional, aunque no sobra recordar al lector que el tratamiento aquí descrito a las cifras de nacimientos de los departamentos no son una corrección formal del subregistro a este nivel.

## MIGRACIÓN

Para la información de migración neta, tal como lo presenta el DANE (2010), las fuentes para el análisis de este componente son escasas y el nivel de incertidumbre es muy alto con respecto a los niveles y estructuras. Hay que tener presente que esta componente demográfica se ve afectada, principalmente por aspectos socio-económicos y no tanto por los biológicos como las anteriores componentes. Influyen razones como la situación laboral, deseos de mejorar el nivel educativo, problemas de salud o problemas de orden público fuente de la migración forzada interna hacia otros departamentos. Todos estos aspectos hacen que la proyección de la migración sea aún más compleja y sus estimaciones menos acertadas, con mayor razón cuando la información necesita ser desglosada por departamentos, edad y sexo.

Esta es la única componente para la que fue imposible recolectar una serie anual histórica, en gran parte debido a que no se tiene información creíble para los departamentos de Colombia y porque la migración neta interna es mucho más difícil de capturar que la migración neta externa que cuenta con los registros de cruce de fronteras. Por esta razón se utilizó la información de migración interna recolectada en el Censo Nacional de Población y Vivienda de 2005 (DANE, 2010) y los censos de 1993 y 1973 (Martínez, 2002).

Se toma la información de las tablas de migración neta (origen-destino) que relacionan el departamento de residencia en el momento inicial, 5 años anteriores al Censo de 2005, con el departamento de residencia al momento del censo, a partir de ellas se calcularon los saldos netos migratorios por departamentos<sup>4</sup>. Esta información se complementó con la presentada por el estudio de Martínez (2002), quien utilizó la información de los censos nacionales de población y vivienda de 1973 y 1993 para generar información de migración interna a nivel departamental y municipal, el siguiente nivel de desagregación geográfica.

4 Las tablas están disponibles en el Anexo 1, alojado en <<https://sites.google.com/view/anexos-urdinola-lara/home>> para su consulta.

## RESULTADOS

Dado el alto nivel de subregistro que presentan los departamentos que conforman la Amazonía colombiana (Urdinola y Herrera, 2015; Pabón, 1993), se decidió agrupar a los departamentos de Amazonas, Guanía, Guaviare, Vaupés y Vichada en una sola entidad regional. Estos departamentos son los de menor densidad poblacional y además constituyen 0.45% del total de la población colombiana en el más reciente censo de población. Los demás departamentos también sufren de problemas de subregistro, pero que son de mucha menor envergadura por lo que se mantuvieron bajo los límites geográficos que tiene la división político administrativa de Colombia, que es estable en sus límites desde 1909 y cubre a los departamentos de: Antioquia, Arauca, Atlántico, Bogotá-Distrito Capital<sup>5</sup>, Bolívar, Boyacá, Caldas, Caquetá, Casanare, Cauca, Cesar, Chocó, Córdoba, Cundinamarca, Guajira, Huila, Magdalena, Meta, Nariño, Norte Santander, Putumayo, Quindío, Risaralda, San Andrés y Providencia, Santander, Sucre, Tolima, Valle. De allí que se construya toda la información para 29 entidades regionales, es decir los 28 departamentos originales más la región de la Amazonía. Para este trabajo no se corrigieron los problemas de subregistro del censo de población ni de los componentes de población, a pesar del conocimiento que se tiene de su existencia. Esto se explica por dos razones, la primera es que el nivel de subregistro para cada departamento es diferencial entre ellos y en el tiempo. La calidad de las estadísticas vitales viene mejorando con la implementación del sistema de estadísticas vitales. Entonces, analizar y llegar a proponer una corrección específica para cada departamento en el tiempo hacen parte de un estudio detallado y riguroso que requiere de mayor trabajo del que podía abarcar esta investigación. Segundo, se quería probar la capacidad que tienen el modelo regional multiestado aquí propuesto, de ayudar a corregir estos problemas simplemente por el hecho de incorporar la información regional que, en últimas, lo que hace es adicionar a las componentes demográficas los patrones subnacionales a través de los 29 departamentos aquí analizados.

5 Si bien Bogotá no es un departamento, constituye una entidad geográfica especial en Colombia y su importancia demográfica se refleja en que cuenta con las menores tasas de mortalidad y fecundidad en el país, aunque su crecimiento se debe prácticamente a la migración neta, al ser el principal polo de atracción de migrantes en el país.

## MATRIZ DE LESLIE MULTIESTADO

La construcción de matrices de Leslie por sexo en cada uno de los departamentos contiene información, en la primera fila, el número esperado de niños nacidos durante el periodo  $t$  por mujer en la edad  $x$  que sobreviven al siguiente año. Los elementos por debajo de la diagonal de esta matriz, representan la proporción de sobrevivientes de edad  $x-1$  a  $x$  durante el periodo  $t$ , que se obtienen a partir de tablas de vida para cada departamento, por edad y sexo, construidas a partir de la información de mortalidad presentada anteriormente. La información de los sobrevivientes en cada departamento por edades quinquenales y sexo en el año base (2005) fue obtenida a partir de la construcción de las tablas de vida, mediante el paquete *demogR* (Jones, 2007), implementado en el software de libre distribución R. Vale la pena resaltar que en este caso particular, las matrices de Leslie representan matrices de transición para las cuales los grupos de edades representan estados de transición de los individuos.

Luego de construir cada una de las 29 matrices de Leslie con la información de fecundidad y sobrevivencia, se unificaron en una sola matriz la información para cada sexo y grupos quinquenales de edad. En las matrices multiestado se tiene contemplado, teóricamente, establecer las transiciones de los individuos entre los diferentes estados, ahora departamentos, en términos del número esperado de niños nacidos y la proporción de sobrevivientes durante el periodo  $t$ . Calcular este tipo de transiciones implicaría tener conocimiento detallado sobre las características específicas de la población en términos del lugar de nacimiento en un periodo  $t$ , cantidad de personas nacidas en un departamento que migraron a otro en el mismo periodo  $t$ , según edad de la madre y el sexo. Adicionalmente, se debe tener conocimiento de la proporción de sobrevivientes de edad  $x-1$  a  $x$  durante el periodo  $t$  que residían en un departamento específico y que migraron, durante ese año, a otros departamentos.

Debido a que la información para el período de análisis no se tiene a ese nivel de detalle, se asumió que la transición de individuos entre los diferentes departamentos es baja. Por ejemplo, que el número esperado de niños nacidos en el departamento de Antioquia para las mujeres con edad 15 años a comienzos del año 2005 y que migraron hacia el departamento de Cundinamarca durante el este mismo año es cercana a cero. Caso similar se presenta para la proporción de sobrevivientes, basados en la experiencia de migración por edad y sexo descrita por los patrones de migración construidos a partir de la información censal y descrita en la sección anterior.

Las matrices de Leslie en cada departamento tienen una dimensión  $2k \times 2k$  siendo  $k=0,5,10,\dots,80$ , los 17 grupos de edades quinquenales. A partir

de éstas y teniendo en cuenta la estructura presentada en la ecuación (5), se procedió a construir la matriz de Leslie multiestado, cuya dimensión es de  $986 \times 986$ , que corresponde a los 17 grupos de edad, los dos géneros y los 28 departamentos más el grupo Amazonía.

Con esta matriz de Leslie multiestado y la población acervo, o de base, correspondiente al año 2005 se pueden estimar las proyecciones haciendo multiplicación de matrices. En esta técnica, para cada sexo, grupo de edad y departamento, los supuestos de evolución de la mortalidad, la fecundidad y las migraciones internas se expresan mediante tasas o probabilidades a partir de la forma de una matriz generalizada de Leslie (1945) y en este caso se aplicó para proyecciones quinquenales desde 2005 hasta el año 2050.

Las Tablas 2 y 3 muestran la población colombiana de base, correspondiente al año 2005, y las proyecciones para el periodo 2010-2050, junto con las tasas de crecimiento proporcional de la población. Asimismo, para el país y para la mayoría de departamentos, como se presenta a continuación, el proceso de envejecimiento comienza a manifestarse a partir de la década de 2040, cuando la población de mayores de 60 años comienza a cobrar mayor importancia, como lo reflejan las pirámides de población presentadas en el Gráfico 1.

Tabla 2  
Colombia, 2010-2050: Proyecciones de Población

Sexo	2005	2010	2015	2020	2025
Mujeres	21.132.262	22.609.295	24.253.140	25.952.281	27.648.035
Hombres	20.336.122	21.744.819	23.282.325	24.874.247	26.459.579
Total	41.468.384	44.354.114	47.535.465	50.826.528	54.107.614
Sexo	2030	2035	2040	2045	2050
Mujeres	29.283.072	30.777.146	32.153.789	33.444.736	34.707.607
Hombres	27.977.688	29.370.126	30.670.157	31.920.571	33.164.388
Total	57.260.760	60.147.272	62.823.946	65.365.307	67.871.995

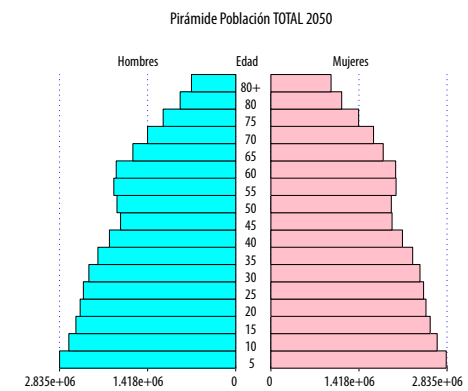
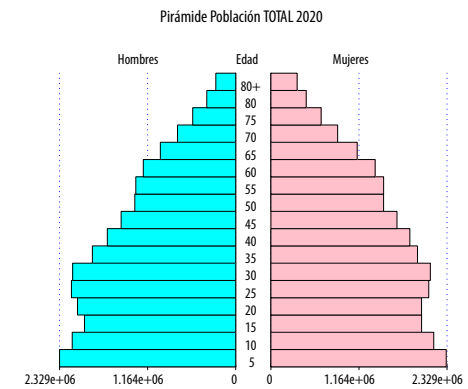
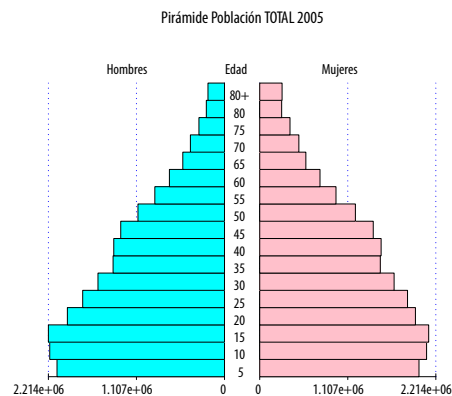
Fuente: Cálculos propios

Tabla 3  
Colombia, 2010-2050: Cambio proporcional en la población proyectada

Periodo	Cambio porcentual
2005 - 2010	7.0
2010 - 2015	7.2
2015 - 2020	6.9
2020 - 2025	6.5
2025 - 2030	5.8
2030 - 2035	5.0
2035 - 2040	4.5
2040 - 2045	4.0
2045 - 2050	3.8

Fuente: Cálculos propios

Gráfico 1  
Colombia, 2005 (observado), 2020 y 2050 (estimado): Pirámides de población



Fuente: Cálculos propios

La Tabla 4 presenta las diferencias entre esta metodología y la de componentes que sigue la oficina de estadísticas de Colombia-DANE, donde se observan diferencias de cerca de 2,5% hacia abajo entre el DANE y las proyecciones del modelo aquí propuesto (MRM) para 2010, que luego se reversan a partir de 2020. Tanto para hombres como para mujeres las proyecciones aquí presentadas producen una estimación de mayor población en ambos casos, que llega a superar a las del DANE en casi medio millón más de habitantes (cerca de 300.000 hombres y 200.000 mujeres de más). En ambos casos, proyecciones DANE y MRM, la velocidad de crecimiento de la población es similar, el crecimiento más rápido se presenta entre 2005 y 2015 y a partir de este año se presenta una desaceleración.

Tabla 4  
Colombia, 2010-2020: Contraste entre las proyecciones del modelo regional multiestado (MRM) y las proyecciones oficiales del DANE

Periodo	Total Nacional DANE		Total Nacional MRM	
	Mujeres	Hombres	Mujeres	Hombres
2010	23.042.445	22.465.760	22.609.295	21.744.819
2015	24.403.311	23.799.306	24.253.140	23.282.325
2020	25.773.706	25.138.723	25.952.281	24.874.247

Fuente: Cálculos propios

Este mismo contraste se hizo para las proyecciones departamentales, como lo sintetiza la Tabla 5, que presenta las proyecciones para este período de mayor crecimiento poblacional: 2010-2020, pero al observar las composiciones poblacionales se ven diferencias a través de los departamentos. Algunos son mucho más jóvenes que otros y por ende, el aumento de la población de adultos mayores está rezagada a la del total nacional. Por ejemplo, departamento como Nariño, Arauca y Putumayo son de mayor proporción de jóvenes, mientras que Antioquia y los departamentos del llamado eje cafetero (Risaralda, Quindío y Caldas) se envejecen a un ritmo más acelerado.

Además, las comparaciones de metodologías muestran que los departamentos con más de 1 millón de habitantes, la diferencia porcentual promedio entre la información obtenida con el método MRM y el método de componentes del DANE es cercana, a 2,92%<sup>6</sup>. Mientras que en departamentos como menos de 1 millón de habitantes, como: Arauca, Putumayo y Caquetá las diferencias se hacen muy notorias. Parte de la explicación de esta diferencia que arroja el modelo MRM es la diferencia en las tendencias de las estadísticas vitales. Los departamentos con menos de un millón de habitantes corresponden a las zonas más despobladas, con menor densidad de población y peores condiciones de vida, como los llamados Llanos Orientales, la Amazonía, Chocó, La Guajira y el sur occidente del país (Putumayo y Nariño). Estos departamentos tienen peores condiciones de mortalidad, la proporción de sobrevivientes disminuye considerablemente entre los 20 a 40 años, y la natalidad promedio es superior al resto del país. Esta tendencia se puede observar en el Gráfico 3.

6 Variaciones promedio positivas

Tabla 5  
Colombia, 2010-2020: Contraste entre las proyecciones departamentales del modelo regional multiestado (MRM) y las proyecciones oficiales del DANE

Departamentos	DANE 2010		MRM 2010		DANE 2015		MRM 2015	
	Mujeres	Hombres	Mujeres	Hombres	Mujeres	Hombres	Mujeres	Hombres
Nacional	23.042.445	22.465.760	22.609.295	21.744.819	24.403.311	23.799.306	24.253.140	23.282.325
Antioquia	3.102.390	2.963.456	3.135.598	2.936.051	3.301.418	3.154.789	3.329.299	3.115.278
Arauca	122.624	124.917	75.637	77.519	130.426	131.889	82.637	83.070
Atlántico	1.172.778	1.141.669	1.178.937	1.117.826	1.245.115	1.215.886	1.265.284	1.201.556
Bogotá	3.815.069	3.548.713	3.850.500	3.540.100	4.068.770	3.810.013	4.065.380	3.752.004
Bolívar	990.049	989.732	971.923	970.867	1.048.291	1.048.795	1.059.608	1.055.691
Boyacá	633.924	633.673	613.896	612.561	636.804	639.563	655.873	652.013
Caldas	499.335	479.027	463.262	440.637	504.633	483.370	487.866	460.966
Caquetá	221.902	225.821	171.915	174.529	238.032	239.587	191.042	190.999
Casanare	159.870	165.726	154.004	160.818	175.610	180.828	170.541	176.157
Cauca	650.723	668.260	615.314	614.456	680.534	698.536	667.007	662.738
Cesar	483.846	482.574	471.520	468.241	515.111	513.769	520.758	513.476
Chocó	239.061	237.112	200.805	199.590	249.845	250.231	227.642	225.911
Córdoba	788.356	794.362	786.025	795.383	852.621	856.982	866.146	872.785
Cundinamarca	1.240.511	1.236.525	1.220.914	1.216.715	1.343.825	1.336.216	1.312.014	1.302.160
Grupo Amazonía	154.468	164.388	98.026	102.874	168.111	176.313	111.579	115.208
Guajira	413.454	405.241	408.765	397.179	483.621	474.193	464.129	451.055
Huila	539.354	543.846	534.904	535.028	575.317	579.487	581.689	579.076
Magdalena	595.059	606.327	588.028	599.036	623.070	636.597	648.875	657.579
Meta	433.434	437.442	399.858	398.599	480.657	480.635	434.689	428.592
Norte Santander	654.035	643.807	650.874	634.568	683.777	671.946	701.873	680.343
Nariño	817.173	822.396	806.234	795.794	868.805	875.470	874.764	862.912
Putumayo	160.281	165.812	121.846	123.303	170.510	174.694	136.721	136.536
Quindío	279.896	269.728	278.525	264.424	287.762	277.504	293.744	277.971
Risaralda	474.221	450.884	466.567	439.421	488.507	463.438	491.212	459.959
San Andrés	36.847	36.473	31.894	30.783	38.401	38.041	33.978	32.759
Santander	1.017.461	992.943	1.003.823	968.391	1.042.357	1.018.738	1.063.488	1.023.296
Sucre	399.801	410.849	399.875	410.042	419.986	431.540	436.516	445.586
Tolima	690.215	697.426	669.437	659.009	703.229	705.045	714.773	699.521
Valle	2.256.308	2.126.631	2.240.389	2.061.075	2.378.166	2.235.211	2.364.013	2.167.128

Fuente: Cálculos propios

Tabla 5

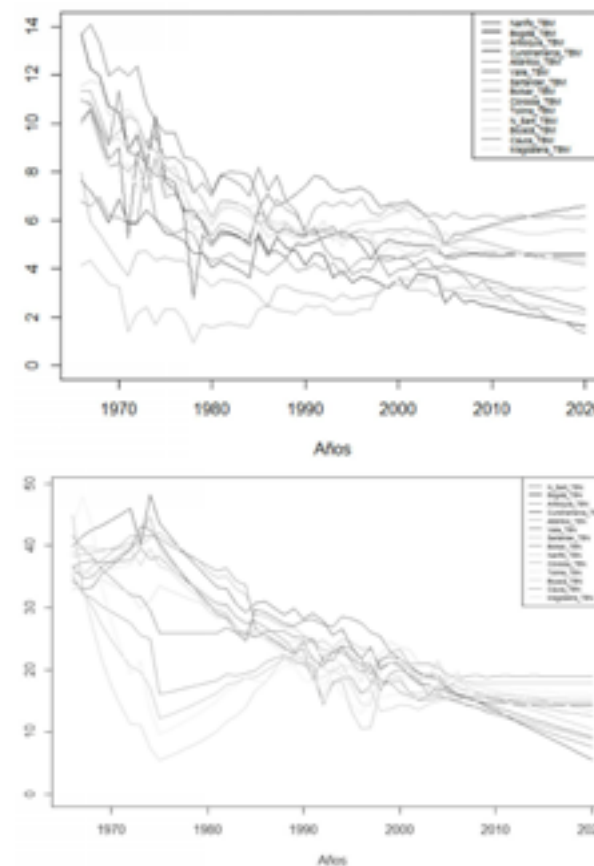
Colombia, 2010-2020: Contraste entre las proyecciones departamentales del modelo regional multiestado (MRM) y las proyecciones oficiales del DANE (conclusión)

Departamentos	DANE 2020		MRM 2020	
	Mujeres	Hombres	Mujeres	Hombres
Nacional	25.773.706	25.138.723	25.952.281	24.874.247
Antioquia	3.498.492	3.346.565	3.520.128	3.291.689
Arauca	137.620	138.194	90.417	89.400
Atlántico	1.314.636	1.286.480	1.351.069	1.284.459
Bogotá	4.316.132	4.064.669	4.263.270	3.946.674
Bolívar	1.109.529	1.109.932	1.153.109	1.146.194
Boyacá	640.777	646.219	702.387	696.054
Caldas	510.053	487.837	512.904	481.972
Caquetá	254.754	253.780	212.510	209.706
Casanare	191.691	196.131	188.631	192.782
Cauca	709.194	727.947	722.404	714.928
Cesar	545.727	544.056	575.244	563.593
Chocó	261.531	263.997	259.676	257.295
Córdoba	918.431	920.143	954.137	957.936
Cundinamarca	1.449.208	1.437.797	1.407.035	1.391.775
Grupo Amazonía	182.564	189.046	127.008	129.172
Guajira	552.240	541.493	524.614	509.970
Huila	610.735	614.525	633.672	628.198
Magdalena	655.558	670.783	717.055	723.216
Meta	529.030	524.841	471.326	459.972
Norte Santander	713.865	700.167	755.731	728.578
Nariño	922.195	929.463	946.441	933.665
Putumayo	183.221	186.111	153.125	151.216
Quindío	295.965	285.569	309.041	291.604
Risaralda	502.069	476.113	515.264	480.145
San Andrés	40.046	39.647	36.143	34.759
Santander	1.066.992	1.043.616	1.124.278	1.079.111
Sucre	441.651	453.083	477.473	485.283
Tolima	715.692	711.731	764.600	744.340
Valle	2.504.108	2.348.788	2.483.589	2.270.561

Fuente: Cálculos propios

Gráfico 2

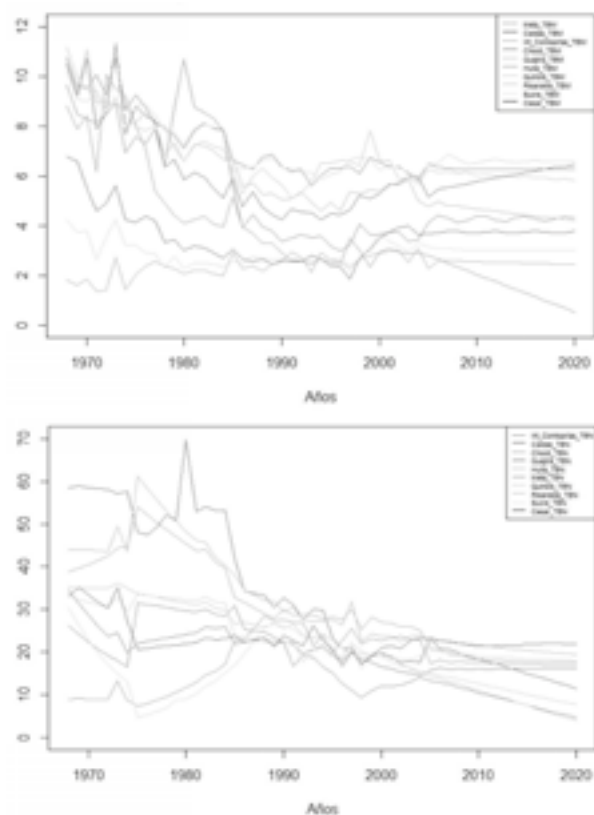
Colombia, 1968-2020: Tasas Brutas de Mortalidad y Natalidad departamentales y sus proyecciones (Departamentos de más de un millón de habitantes)



Fuente: Cálculos propios



Gráfico 3  
Colombia, 1968-2020: Tasas Brutas de Mortalidad y Natalidad departamentales y sus proyecciones (Departamentos de menos de un millón de habitantes)



Fuente: Cálculos propios

Las proyecciones para cada departamento por edad y sexo se encuentran disponibles en <<https://sites.google.com/view/anexos-urdinola-lara/home>>, para la consulta de los lectores. Contar con esta herramienta permite hacer diferentes tipos de análisis que necesiten la distribución etaria de la población, desde la planeación fiscal de los departamentos hasta la cobertura de necesidades que requiera una población objetivo como los esquemas de vacunación, la oferta de educación pública y privada, entre otras. No sobra recordar al lector que el tratamiento dado a las cifras que alimentan esta proyección no tienen una corrección por subregistro, se

eliminó la volatilidad de los nacimientos con la tendencia temporal para cada departamento y se imputaron tres valores en las defunciones que no tenían información.

Las proyecciones del modelo MRM no distan mucho de las oficiales del DANE que sí contemplan estos problemas, no quiere esto decir que se deba obviar el subregistro en las estadísticas vitales ni en los censos de población, sino por el contrario que la lectura de las proyecciones que resultan de este ejercicio deberían leerse como una cota inferior. Una vez se cuente con una corrección de subregistro a las estadísticas vitales y/o al conteo del censo de población, o se actualicen las condiciones de los patrones demográficos, como por ejemplo se refine la información de migración interna, ajustar este modelo será una tarea simple, pues solo requerirá de actualizar la información insumo con la corrección propuestas, dado que la programación para ejecutar la proyección se encuentra ya hecha.

Una de las ventajas que tienen estos modelos de proyección que se basan en las matrices de transición es que su ejecución no requiere de grandes conocimientos en programación o técnicos para elaborar. Al corresponder a una multiplicación de matrices, la mayoría de los técnicos en oficinas oficiales de estadísticas lo pueden implementar e incluso paquetes de libre distribución, como R, cuentan con librerías que corren estas multiplicaciones de matrices de Leslie. Como se explicó en secciones anteriores es mucho más engorroso, para países no desarrollados, recolectar la información que se requiere como insumo y conformar las matrices multirregionales que llevar a cabo la aplicación.

## CONCLUSIONES

La metodología propuesta del modelo regional multiestado no requiere de hacer mayores inversiones para su producción. Paquetes estadísticos de software de libre distribución pueden ejecutarlo, pues sólo se requiere de una multiplicación de matrices. Es esta simpleza del modelo lo que llama la atención de esta metodología. No obstante, requiere de insumos en las componentes demográficas que estén desagregados por edad y sexo para las entidades regionales que en el caso estudiado, el colombiano, corresponden a departamentos.

Es importante tener en cuenta que Colombia, y la mayoría de países latinoamericanos, sufren de deficiencias en la información demográfica a nivel nacional y subnacional. En el caso subnacional es más evidente: mayor cantidad de años faltantes, nacimientos sin edad de la madre, ausencia de información de migración interna, tal como lo reveló el caso

Colombiano. Una manera de sobreponer estas limitaciones de información fue la propuesta en este documento, que consiste en la elaboración de series de tiempo para las entidades subnacionales, que permita observar el comportamiento de largo plazo de las componentes a las que se les puede hacer este estudio, éstas son las series de defunciones y nacimientos, y a partir de este comportamiento completar la información faltante. Para la información de migración interna la tarea es algo más compleja, sería ideal contar con series anuales de migración por edad, sexo y zonas receptoras/expulsoras en todos los casos, preferiblemente identificando migración interna y externa. Quizás países con mayor tradición migratoria en la región como El Salvador o México cuentan con dicha información, pero para países como Colombia la mejor solución, dada la información existente, consistió en tomar los registros de los flujos migratorios netos recolectados en los censos nacionales de población. Allí se pudo establecer una baja influencia en los nacimientos de parte de migrantes en los lugares receptores. Esta misma información está disponible para todos los demás países de la región, que si bien no es la ideal, permite hacer las estimaciones de un modelo como el aquí propuesto.

Los resultados de la proyección MRM demostraron, además de las bondades para su estimación, que el modelo se asemeja bastante al oficial producido por el DANE. Para 2050 se espera que la población colombiana alcance los 68 millones de habitantes, asumiendo que las condiciones demográficas se mantienen a las observadas, siendo Bogotá el departamento más poblado con algo más de 9 millones de habitantes y el menos poblado San Andrés, con 87 mil habitantes.

Si bien en la etapa de mayor crecimiento poblacional, hasta 2020, el modelo MRM presenta estimaciones por debajo de las del DANE en un promedio departamental de -3.8%, para los siguientes años muestra un comportamiento superior en hasta 2.0%. En parte, esto se puede explicar por la ausencia de corrección de subregistro de las vitales y del censo de población. Sin embargo, no hacer esta corrección también permitió observar que el añadir esta transición regional, añade información que de alguna manera ayuda a corregir estas falencias propias de la información, por la proximidad en los resultados a los oficiales del DANE. Así mismo, se pudo observar una convergencia en patrones entre los departamentos de más o de menos de un millón de habitantes, tanto en sus componentes demográficos, como en la producción final de las proyecciones. Las diferencias con los resultados de las proyecciones oficiales así también lo reflejan.

Finalmente, y como otra propuesta de trabajo adicional sería interesante no sólo incorporar en una versión futura la corrección de subregistro,

sino que también se puede hacer el cálculo de las probabilidades de transición en los diferentes departamentos. Con esta implementación se podría garantizar más detalladamente los flujos de las variables demográficas al total Colombia y para cada departamento e incluso se podrían aplicar técnicas de proyección similares a la aquí propuesta, pero que permitan la estimación de la incertidumbre en la proyección final.

## REFERENCIAS

- ALHO, J., SPENCER, B. (2005), *Statistical Demography and Forecasting*, New York City: Springer
- ANDERSON, M. W., TULJAPURKAR, S., y LEE, R. D. (2001), “Chances are... stochastic forecasts of the social security trust fund and attempts to save it”, en *Stochastic Forecasts of the Social Security Trust Fund and Attempts to Save It (May 1, 2001)*, Michigan: Michigan Retirement Research Center Research Paper No. WP, 8.
- DEPARTAMENTO ADMINISTRATIVO NACIONAL DE ESTADÍSTICA-DANE (1993), *Las estadísticas Sociales en Colombia*, Bogotá: DANE.
- DEPARTAMENTO ADMINISTRATIVO NACIONAL DE ESTADÍSTICA-DANE (2010), *Estudios postcensales 7: Proyecciones nacionales y departamentales de población 2005-2020*, Bogotá: DANE.
- FEENEY, G. (1970), “Stable age by region distributions”, en *Demography*, Vol.7, N° 3, p. 341-348.
- HILL, K. (2003), “Métodos para estimar la mortalidad adulta en los países en desarrollo: una revisión comparativa”, en *Notas de Población*, Santiago de Chile, N° 76, p. 81-111.
- INSTITUTO NACIONAL DE ESTADÍSTICAS CHILE-INE (2005). *Chile hacia el 2050*. Santiago de Chile: INE.
- JIMÉNEZ, M., ET AL. (2007), “La reducción de la mortalidad infantil en América Latina y el Caribe: avance dispar que requiere respuestas variadas”, en *Desafíos*, Santiago de Chile, N° 6, p.1-12.
- JONES, J. H. (2007), “demogR: A Package for the Construction and Analysis of Age-structured Demographic Models in R”, en *Journal of Statistical Software*, Austria, Vol.22, N° 1, p.1-28.
- KEYFITZ, N. y CASWELL, H. (2005), *Applied mathematical demography*. New York City: Springer.
- LEE, R. D. y TULJAPURKAR, S. (1994), “Stochastic population forecasts for the United States: Beyond high, medium, and low”, en *Journal of the American Statistical Association*, United States, Vol. 89, N° 428, p. 1175-1189.
- LESLIE, P. H. (1945), “On the use of matrices in certain population mathematics”, en *Biometrika*, United Kingdom, Vol. 33, N° 3, p.183-212.

- MARTÍNEZ GÓMEZ, C. L. (2002), *Las migraciones internas en Colombia. Análisis territorial y demográfico según los censos de 1973 y 1993*. Universidad de Barcelona
- OKITA, Y., PFAU, W. D., y LONG, G. T. (2009), "A stochastic forecast model for Japan's population", en *GRIPS Policy Information Center*, Japan, Discussion Papers.
- PABÓN, A. (1993), *La Mortalidad en Colombia, 1953-1991*. Bogotá: Instituto Nacional de Salud.
- PALLONI, A., PINTO, G., y BELTRÁN-SÁNCHEZ, H. (2015), "Two Centuries of Mortality Decline in Latin America: From Hunger to Longevity", en <<https://pdfs.semanticscholar.org/6e82/6e67df57f986c424f13de42e7ec95762obba.pdf>>, acceso junio de 2013.
- ROGERS, A. (1975), *Introduction to multiregional mathematical demography*. New York City: John Wiley & Sons.
- ROGERS, A. (1986), "Parameterized multistate population dynamics and projections", en *Journal of the American Statistical Association*, United States, Vol. 81, N° 393, p. 48-61.
- SWANSON, D. A., y TAYMAN, J. (2012). *Subnational Population Estimates*. London-New York: Springer.
- TOVAR, J. A., y URDINOLA, B. P. (2014), "Inequality in National inter-generational Transfers: evidence from Colombia", en *International Advances in Economic Research*, Atlanta, Vol. 20, N° 2, p.167-187.
- TULJAPURKAR, S. (2006), "Population Forecasts, Fiscal Policy, and Risk", en *The Levy Economics Institute of Bard College*, California, Working Paper No 471.
- URDINOLA, B. P. (2011), "Determinantes socio-económicos de la mortalidad infantil en Colombia, 1993", en *Revista Colombiana de Estadística*, Colombia, Vol. 34, N° 1, p.39-72.
- URDINOLA, B. P. y HERRERA, R. (2015), "Caracterización de la Mortalidad Materna en Colombia y su Estimación Indirecta", en *Aplicaciones en Demografía*, Bogotá: Editorial Universidad Nacional de Colombia
- URDINOLA, B. P. y OSPINO, C. (2015), "Long-Term Consequences of Adolescent Fertility. The Colombian Case", en *Demographic Research*, Vol. 32, p. 1487-1518.
- URDINOLA, B. P. y QUEIROZ, B. L. (2013), "Latin American Human Mortality Database", en <[www.lamortalidad.org](http://www.lamortalidad.org)>, acceso Julio de 2013
- URDINOLA, B. P., TORRES AVILÉS, F. y VELASCO, J.A. (2017), "The Homicide Atlas in Colombia: Contagion and Under-Registration for Small Areas", en *Cuadernos de Geografía: Revista Colombiana de Geografía*, Colombia, Vol. 26, N° 1, p. 101-118.
- WACHTER, K. (2014), *Essential Demographic Methods*, Cambridge: Harvard University Press.
- WHELPTON, P. K. (1928), "Population of the United States, 1925 to 1975", en *American Journal of Sociology*, Chicago, Vol. 34, N° 2, p. 253-270.
- WHELPTON, P. K. (1936), "An empirical method of calculating future population", en *Journal of the American Statistical Association*, United States, Vol. 31, N° 195, p. 457-473.
- WHELPTON, P. K., ELRIDGE H.T. y SIEGEL J.S. (1947), *Forecasts of the population of the United States 1945-1975*. Washington D.C.: U.S.A Census Bureau
- WIŚNIEWSKI, R., et al. (2015), "Bayesian population forecasting: Extending the lee-carter method", en *Demography*, New York City, Vol. 52, N° 3, p. 1035-1059.

