

Análisis de sensibilidad de proyecciones de población

Milenka Linneth ARGOTE-CUSI

Universidad Nacional de Colombia

Resumen

¿Un ejercicio de proyección de la población es sensible a pequeños cambios en el valor de la Tasa Global de Fecundidad (TGF)? Tomando en cuenta la distribución estadística por muestreo de la TGF de Bolivia en 2003, en un ejercicio de proyección de población, se obtiene la distribución estadística por muestreo de la población proyectada a 2006. Con ambas distribuciones se realiza un análisis de sensibilidad para verificar si se presenta el efecto mariposa propio de los sistemas caóticos. Se utilizó la última Encuesta Nacional de Demografía y Salud de Bolivia disponible en línea, paralelamente se recabaron tablas de mortalidad y otras bases demográficas del Instituto Nacional de Estadística de Bolivia necesarias para la estimación de la TGF y la proyección de población. Se usaron los algoritmos y resultados de Argote de 2007 y 2009 para el remuestreo y generación de distribuciones estadísticas. El método para la proyección de población es “por componentes” y se realizó el análisis de sensibilidad por objetivo. A diferencia del efecto amplificador que encontró Argote (2011) en los nacimientos, a pequeños cambios de la TGF, el análisis de sensibilidad muestra que la distribución estadística de la población proyectada tiene un comportamiento heterogéneo; un efecto reductor, estable o de leve amplificación. Este comportamiento heterogéneo muestra la complejidad de la dinámica demográfica. El método ha probado ser robusto, coherente con los resultados y adecuado para el análisis de la incertidumbre en proyecciones de población.

Palabras clave: Sensibilidad, remuestreo, proyección de población, distribución probabilística, estimación del error, incertidumbre.

Abstract

Sensitivity analysis of population projections

Is population projection sensitive to small changes in Global Fertility Rate (GFR)? Considering the GFR sampling distribution of Bolivia in a method of population projection, sampling distribution of population projected has been obtained. With both distributions sensitivity analysis is performed to check whether the proper butterfly effect of chaotic systems is presented. The latest National Survey of Demography and Health of Bolivia, available on line, was used. For population projection, mortality tables and other demography data are collected of National Institute of Statistics. Algorithms and distributions obtained by Argote in 2007 and 2009 were used to generate new sampling distributions. This research produces projection of the resident population using a cohort-component method and assumptions about demographics component of change. Finally, the sensitivity analysis is performed by target. Unlike the amplifier effect found in sampling distribution of births to small changes in the GFR (Argote, 2011) the sensitivity analysis of the projected population is a heterogeneous behavior; in some cases, reducing, stable or amplification effect is observed. This heterogeneous behavior shows us population dynamic complexity. The method has proven to be robust, consistent with the results and suitable for the analysis of the uncertainty in population projections.

Key words: Sensibility analysis, bootstrap, population projection, uncertainty.

INTRODUCCIÓN

La presente investigación se enmarca en el análisis de la incertidumbre atada a las estimaciones. En Demografía, el estudio de la incertidumbre y la precisión de las estimaciones es reciente comparado con otras disciplinas como las ciencias puras, las ciencias biológicas o ecológicas (Barret *et al.*, 2006). Por ejemplo, en el siglo XVII se realizaron las primeras estimaciones sobre la edad de la tierra. Newton en 1687 estimó una edad de 50 mil años y fue hasta el siglo XVIII aproximadamente que De Maillet fue el primero en calcular la edad de la tierra a través de métodos científicos. De Maillet estimó para la tierra una edad de 2 400 millones de años y Georges-Louis Leclerc, por experimentación, calculó una edad de 74 832 años (Livio, 2013). Como es posible apreciar, la inquietud por medir ciertos fenómenos físicos o sociales se ha presentado a lo largo de la historia con un impacto importante en los avances de la ciencia.

El área de las ciencias sociales no queda al margen de este interés. El estimar la población total que vive en un área específica de territorio se vuelve una tarea cada vez más compleja conforme su número aumenta. Es por esta razón que cuando se realizan cálculos a mayores escalas, el error es un factor por asumir.

El método es importante y determinante para realizar una buena estimación y éste también está atado al error. Es la estadística la disciplina que más se utiliza para la estimación, ya que cuenta con los fundamentos teóricos que permiten inferir a partir de muestras, reduciendo así el error. Un supuesto que generalmente se toma en cuenta en los métodos de estimación es el comportamiento lineal, sin embargo los sistemas reales no siempre tienen este comportamiento (Bagajewicz, 2005; Caswell, 2009). Entonces es necesario no solamente abordar el estudio de estos sistemas con los métodos tradicionales, sino con nuevas metodologías provenientes de la biología, la estadística, la física, la economía o la ingeniería, que con ayuda de la computación permiten en la actualidad estudiar temas como la no linealidad, la heterogeneidad, la precisión de las estimaciones, la función de error y el análisis de sensibilidad o análisis de perturbación, como se realiza en el presente artículo (Alho y Spencer, 2005; Griffiths *et al.*, 2000; Axelrod, 1997).

LA COMPLEJIDAD DE LOS INDICADORES DEMOGRÁFICOS

De acuerdo con Escobedo (2007) los datos son elementos complejos que se construyen a partir de filosofía y metodología que provienen del investigador u observador. Entonces es ilógico considerar a los datos como simples símbolos, cuando son función de una o más variables; en sentido formal: sea Y un indicador demográfico, que depende de variables endógenas y exógenas al fenómeno demográfico y la diferencia entre el valor real y su estimación que también es una función:

$$Y = aX + bZ + \dots + cW + \varepsilon \quad (\text{Ecuación 1})$$

$$\varepsilon = Y - \hat{Y}$$

$X, Z, \dots W$ pueden ser funciones lineales o no lineales.

En este sentido, muchas controversias emergen de los indicadores demográficos debido a su difícil tarea de representar el estado de los sistemas sociales y mayor es la discusión acerca de su precisión y el manejo del error (Gottlieb, 2001; Dias *et al.*, 2001; Bagajewicz, 2005; Bagajewicz *et al.*, 2006).

La Demografía provee de un marco teórico y de métodos para la estimación de indicadores que representan la dinámica de la población. Debido a que la dinámica poblacional involucra un conjunto de variables interrelacionadas, las estimaciones están sujetas a determinado margen de error. En este ámbito los ajustes y correcciones son una tarea común y parte de la búsqueda de la mejor estimación, pero en Demografía se debate el hecho de la corrección de datos poblacionales por cientos o miles de habitantes, que podrían ser beneficiarias o no de determinada política pública.

Cuando se trabaja con indicadores resumen que cuantifican la dinámica de un evento demográfico, sus variaciones representan cambios en las variables que la componen (en el numerador y el denominador). En este sentido, por ejemplo, cuando se estima la Tasa Global de Fecundidad (TGF) u otros indicadores demográficos por diferentes métodos ¿Cuál estimación elegir? ¿Qué significan las pequeñas diferencias entre estimaciones? ¿Es posible mejorar las estimaciones y errar menos?

Existen investigaciones que han profundizado en el tema de la precisión de los indicadores y su interpretación. Griffiths *et al.* (2000) analizaron por qué razón el índice de masculinidad se ha mantenido a favor de los hombres a través del tiempo en la India, en relación a las fuentes de datos y variables culturales que pueden estar influyendo. Por otro lado, en España debido a las bajas tasas de fecundidad actual que están por debajo del nivel

de reemplazo, se han realizado análisis más detallados desde las fuentes de información, el efecto de la edad, la infertilidad, etc., para verificar si realmente la fecundidad está disminuyendo (Ortega *et al.*, 2000).

Alho y Spencer (2005) dedican varios apartados al análisis de la precisión y formalización de la función de error en el modelado de los eventos demográficos. El análisis de la precisión lleva de manera natural al análisis de sensibilidad del modelo, es decir, ya que los indicadores se pueden considerar como variables aleatorias con infinitos valores en un espacio muestral determinado, el análisis de sensibilidad busca determinar cuáles son los parámetros más sensibles a pequeñas variaciones en algunos parámetros del modelo.

ANÁLISIS DE SENSIBILIDAD

Considerando que la información de proyecciones de población se utiliza para la toma de decisiones en política pública y la incertidumbre en los escenarios futuros, el análisis de sensibilidad es un tema transcendental en la Demografía actual. Partiendo de la pregunta de investigación, es decir, ¿cuál es el impacto de pequeñas variaciones de la TGF en un ejercicio de proyección de población? el análisis de sensibilidad se constituye en la herramienta científica que permite responder a la pregunta partiendo de las distribuciones estadísticas por muestreo encontradas en Argote 2007, 2009 y 2011.

Este no es el primer estudio que realiza análisis de sensibilidad en el campo demográfico. Para Caswell (2008) la importancia de los análisis de sensibilidad radica en su aplicación en política pública y en teoría del muestreo, ya que los parámetros más sensibles son los que deberían ser estimados de forma más precisa. Este mismo autor ha venido profundizando en el modelado de los sistemas estocásticos, en los cuales considera la “individualidad estocástica” en el ámbito demográfico. Este trabajo ha desarrollado el modelado de varios tipos de población mediante matrices estocásticas, las cuales le han permitido realizar el análisis de la perturbación que incluye la sensibilidad y elasticidad (Caswell, 2009).

Un ejemplo clásico sobre sensibilidad son las investigaciones de Edward Lorenz (1963) sobre el modelado de datos atmosféricos para predecir el clima. Uno de los hallazgos más importantes, ha sido llamado el “efecto mariposa” con el cual se explica el hecho de que a pequeñas variaciones en los parámetros de un modelo de predicción se observan resultados muy diferentes. Este concepto ha dado origen a la teoría del caos que trata con

sistemas que se comportan de manera estable e inestable en determinados períodos.

Relacionado con el análisis de sensibilidad, Argote (2007) realizó un análisis de la precisión de la TGF a través su distribución estadística por remuestreo. Se observa que el comportamiento de la fecundidad de las mujeres bolivianas en 1998 tiene comportamiento normal y que el método es robusto y permite evaluar el sesgo de las estimaciones. En un segundo artículo, Argote (2012) encuentra las distribuciones por muestreo de los nacimientos por grupos quinquenales de edad de la población boliviana en 2003, con base en lo cual puede analizar el efecto que representa pequeños cambios en el valor de la TGF en los nacimientos. Entre los resultados se destaca que la distribución de los grupos de edad 20-25, 25-30, 30-35 se asemeja más a la normal y son homogéneas; de los grupos 15-20, 35-40, 40-45 las distribuciones se asemejan a la normal y son heterogéneas, sin embargo en el caso del grupo 45-49 la distribución de los nacimientos es diferente a la normal y es más heterogénea. Es para este último grupo que el porcentaje de variación es mayor a 30 con relación a los cambios en la TGF, por lo cual se presenta un efecto amplificador.

PROYECCIONES DE POBLACIÓN

La inquietud de proyectar para conocer el futuro es intrínseca a todas las disciplinas. El ejemplo más conocido es la predicción del clima. En Demografía es de mucho interés la proyección de la población y su distribución por edades con el objetivo de planificar la gestión, recursos y servicios.

El método más utilizado en la proyección de población es el método por componentes que considera la siguiente ecuación general:

$$P_{t+1} = P_t + N_t + D_t + I_t + E_t$$

Donde N_t son los nacimientos en t , D_t las defunciones, I_t la inmigración y E_t la emigración, las cuales balancean la “ecuación compensadora” para conocer la población en $t + 1$. Los cuatro componentes demográficos se representan por sus tasas de frecuencia anual (tasas específicas de fecundidad, tasa de mortalidad, tasa neta de migración) los cuales se proyectan bajo diferentes hipótesis de comportamiento (CEPAL, 2006: 29-33). Este método se utiliza de manera estándar y se adapta a diferentes contextos sociales, en particular aquellos en los que no se cuenta con información. Sin embargo cuando se relevan datos nuevos con encuestas o conteos de medio término, que retroalimentan las proyecciones realizadas, se observan

variaciones, con respecto a lo cual es frecuente ajustar las proyecciones de población (CONAPO, 2006).

Es muy importante entender el comportamiento de los indicadores demográficos, así como el método de proyección de población para reducir la incertidumbre que ataca estas estimaciones. De seguro no es posible predecir con exactitud el futuro,¹ pero sí es factible reducir el error asociado a las estimaciones y planificar por escenarios.

Alho *et al.* (2006) innovan en su trabajo de proyección de la población de 18 países europeos: “cuantifican la incertidumbre demográfica”. El grupo logra estimar la distribución probabilística de la proyección para hacer afirmaciones del tipo “El valor Y_i tiene una probabilidad P_i de ocurrir”. Esta investigación utiliza el método *socalled scaled model for error* para cuantificar la incertidumbre relacionada a la proyección de población. El método de proyección estocástica tiene la ventaja de proyectar la población futura incluyendo un intervalo de proyección probabilística. Encuentran que a diferencia de los datos oficiales, es probable que la población en general crezca y su descenso se retrase más tiempo debido a la alta esperanza de vida e incremento de la migración.

La importancia de la incertidumbre en proyecciones de población se observa en los hallazgos de Welpthon en 1947. Fue el primero en desarrollar el método por componentes para las proyecciones de población y utilizó la función logística para representar el comportamiento de la fecundidad. Sin embargo, ni sus modelos ni los de ningún otro pudieron predecir en su momento el *baby boom* que hizo que las tasas de fecundidad que iban en descenso se dispararan. De esta manera, hizo análisis exhaustivos de las tendencias de la fecundidad en varios países utilizando diversas fuentes y concluyó: “A largo plazo, la tendencia de descenso de la fecundidad es una regla universal. Aunque raramente puede ocurrir que ascienda de forma relativa y corta” (Alho y Spencer, 2005).

LA SIMULACIÓN COMO HERRAMIENTA PARA LA EXPERIMENTACIÓN

El método experimental en el estudio de sistemas sociales es importante para el avance científico. Una característica del modelado tradicional, como en Economía, ha sido la adopción de supuestos sobre las características de los sistemas como la homogeneidad, el equilibrio, tasas constantes, etc., que no siempre tienen los sistemas reales. Sin embargo, en la actualidad el desarrollo tecnológico computacional permite acceder a

¹ Para tener una justificación científica de ello se puede revisar el principio de incertidumbre del físico alemán Werner K. Heisenberg (2013).

herramientas para estudiar sistemas con los que no es fácil experimentar en la realidad (Axelrod, 1997; Barret, 2006; Botargues y Petrecolla, 1999).

La creación de escenarios en Demografía conlleva un mundo de posibilidades. En párrafos anteriores se mencionó que la función *logito*, con la cual se representa el comportamiento de la fecundidad, no se ajustó a los cambios de la fecundidad que ocurrieron con el *baby boom* (Alho y Spencer, 2005: 229-230). Considerando estos fenómenos, en proyecciones de población será mejor considerar varios escenarios para formular políticas públicas robustas que no afecten la distribución de recursos y servicios (Alho *et al.*, 2006: 6).

El actual desarrollo computacional y sus aplicaciones permiten crear escenarios. El departamento de demografía de la Universidad de Berkeley ha desarrollado varias versiones del programa SOCSIM para proyecciones de población. Cuenta con un interfaz para introducir datos como las tasas de fecundidad, mortalidad y migración, considera funciones matemáticas del comportamiento de las variables y ejecuta un ejercicio de población o varios, según requerimiento del usuario. En otro caso, Alho *et al.*, (2006) realizaron la proyección de la población de 18 países europeos utilizando un *software* programado en C++, cuyos resultados se han aplicado al análisis de la política de pensiones en Estados Unidos y Europa.

El modelaje y la simulación brindan la oportunidad de manejar la incertidumbre en el ámbito demográfico. Es una herramienta que permite un acercamiento entre la teoría y la práctica, para acceder al estudio de sistemas irreducibles como el poblacional (García, 2006: 1; Argote, 2009: 202-204; Griffiths *et al.*, 2000).

PROBLEMA DE INVESTIGACIÓN

Considerando el comportamiento no lineal de las variables demográficas (Caswell, 2008 y 2009) y el supuesto de información imperfecta ligado a estos sistemas (existencia de incertidumbre) el problema que orienta la presente investigación es ¿cuál es el impacto de pequeños cambios del estimador de la TGF en los resultados de una proyección de población? En 1963, Edward Lorenz se hizo la siguiente pregunta: ¿por qué pequeños cambios en los parámetros de un modelo de predicción, daban resultados tan diferentes? La respuesta a esta pregunta se llamó “el efecto mariposa”. El concepto de “efecto mariposa” proviene de la teoría del caos para explicar que pequeños errores provocan grandes errores (amplificación del error). En 2011, Argote encontró que a pequeños cambios de la TGF los nacimientos en los diferentes grupos quinquenales de edad variaban

ampliamente, presentándose un efecto amplificador en algunos grupos. ¿Ocurrirá lo mismo con los datos proyectados a partir de la distribución estadística por muestreo de la TGF?

La importancia de la presente investigación radica en que se trata de un estudio empírico-analítico que genera información a partir de remuestreo (*bootstrapping*) con lo cual es factible analizar la sensibilidad. La diferencia con otros estudios de sensibilidad en el ámbito demográfico (Caswell, 2008) es el método y el interés por identificar el “efecto mariposa” propio de los sistemas complejos.

HIPÓTESIS

Sea:

$T\bar{G}F_i^x$ la media de TGF_i^x una variable aleatoria que representa la distribución por muestreo de la Tasa Global de Fecundidad en la iteración i del año x , donde $i = 1, 2, \dots, 1000$, $x = 1, \dots, n$ años. Además TGF_i^x es un estimador de razón definido por:

$$TEF_{x,x+5}^{t,t+1} = \frac{Nac_{x,x+5}^{t,t+1}}{Temujeres_{x,x+5}^{t,t+1}}$$

$$TGF_i^x = 5 * \sum_{t=1}^7 TEF_{x,x+5}^{t,t+1}$$

Donde $Nac_{x,x+5}^{t,t+1}$ son los nacimientos entre el periodo $t, t+1$ y las edades $x, x+5$ y $Temujeres_{x,x+5}^{t,t+1}$ es el tiempo de exposición de las mujeres a experimentar el evento de tener un hijo en el periodo $t, t+1$ y las edades $x, x+5$.

Y sea P'_i^x la población estimada en el momento x según su distribución por muestreo, \bar{P}_i^x la media de la distribución por muestreo de la población proyectada, α_i^x la diferencia entre las dos anteriores y w_i^x la diferencia entre un valor puntual TGF'_i^x de la distribución estadística por muestreo² de la TGF_i^x y $T\bar{G}F_i^x$:

$$TGF'_i^x - T\bar{G}F_i^x = w_i^x * 100$$

$$P'_i^x - \bar{P}_i^x = \alpha_i^x * 100$$

² Se entiende por distribución estadística por muestreo a la distribución estadística de determinado indicador, en este caso la TGF, que se obtiene en mil ejercicios de estimación a partir de la misma muestra sin reemplazo. Más información de la técnica de remuestreo (*bootstrap*) en Efron y Tibshinari (1993) y en Argote (2007).

$$\alpha_i^x > Aw_i^x$$

Se espera encontrar que a pequeñas variaciones de la ΔTGF , la diferencia entre los valores de población proyectada P'^x_i , presenten el efecto mariposa en que la variación $\alpha_i^x * 100$ es mayor a $w_i^x * 100$ en A veces.³

DATOS Y MÉTODOS

Para llevar a cabo este análisis empírico-teórico se hace necesario el conocimiento del sistema poblacional y sus métodos, el conocimiento de los sistemas complejos y los temas de linealidad y no linealidad de sistemas sociales (Axelrod, 1997; Saura y Rodríguez, 2003; Ferrari, 2005; Miramontes, 1999), los métodos y técnicas provenientes de la estadística para el desarrollo científico y sus aplicaciones a la demografía (Alho *et al.*, 2006, Keyfitz y Caswell, 2005; Alho y Spencer, 2005) y las herramientas computacionales necesarias para la simulación y experimentación.

Un razonamiento inductivo y deductivo, de composición y descomposición, que permite el modelaje y simulación, se aplica para observar los cambios en los componentes de la TGF y en un ejercicio de proyección de población. La distribución estadística de la TGF generada mediante remuestreo (Argote, 2009) es utilizada como dominio para la selección de diferentes valores del estimador de la TGF en el análisis de sensibilidad por objetivo de un ejercicio de proyección de población considerando la mortalidad y migración constantes.

En cuanto al ejercicio de proyección de población, en esta primera fase de la investigación, se construye el algoritmo para la proyección de 2001 a 2006 por el método de componentes, con la información disponible de Bolivia hasta la fecha. Para la proyección de población de 2001 a 2006 se parte de la población base según el censo de 1992 y 2001 disponible en línea (INE-Bolivia). Esta población es depurada mediante el método llamado fórmula de graduación de un dieciseisavo y ajustada a mediados de año (Mina-Valdez, 2012). Se toman las tablas de mortalidad utilizadas por el INE para el periodo 2000-2005, asimismo se toma la migración registrada a través de las tasas netas de migración para el componente de migración.

³ En la hipótesis se parte de la representación formal de las tasas específicas de fecundidad, de la tasa global de fecundidad, su distribución estadística por remuestreo y la media de la distribución. Luego se representa formalmente la distribución estadística de la proyección de población y su media. Finalmente, ya que el análisis de sensibilidad considera pequeñas variaciones de la TGF, se representa formalmente la variación de la TGF que entra como parámetro en el ejercicio de proyección y se obtiene la variación del resultado de la proyección, para compararlas y determinar si el cambio es proporcional, mayor o menor al cambio realizado en el parámetro.

Para la proyección de 2001 a 2006 lo primero que se proyectó fueron los nacimientos de cero a cuatro años que componen el flujo de entrada para el año proyectado. Estos nacimientos se calcularon tomando en cuenta la estructura de fecundidad que provee la ENDSA 2003. Luego este grupo es afectado por la tasas de mortalidad y de migración (Tablas del INE 2000-2005) con lo cual se obtiene el grupo de nacimientos de cero a cuatro años en 2006.

Para los grupos entre 5 y 79 años se reconstruyen las cohortes y se ven afectados por la mortalidad y migración correspondiente a su grupo de edad. El grupo de 80 o más en 2006 queda conformado por la unión de los grupos 75-79 y 80+ en el año 2001 y es afectado por la mortalidad y la migración.

Cabe hacer notar que la estructura de la fecundidad de 15 a 49 años de las mujeres bolivianas determina el número de nacimientos proyectados sobrevivientes a la mortalidad y que no migraron a junio de 2006, por lo tanto el ejercicio de simulación para analizar el efecto de cambios en este estructura, concierne directamente a la estimación del grupo cero a cuatro años, manteniéndose la población de los demás grupos afectados solamente por la mortalidad y migración.

RESULTADOS: DISTRIBUCIÓN POR MUESTREO DE LOS NACIMIENTOS

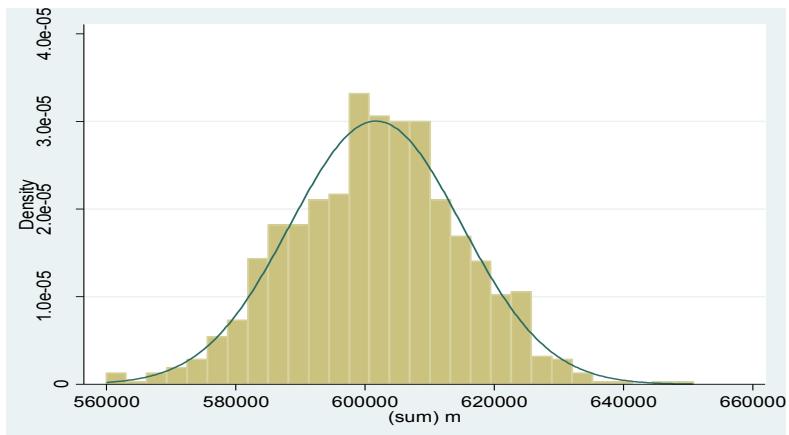
La implementación en STATA del algoritmo que permite utilizar la distribución por muestreo de la TGF como parámetro en un ejercicio de proyección de población, facilita construir el gráfico de la distribución de la población proyectada del grupo cero a cuatro años (Gráfica 1).

Se puede observar que la distribución por muestreo de la población del grupo de cero a cuatro años tiene sus propias características. El valor de -0.04318 de *skewness* indica que la distribución es asimétrica a la izquierda, es decir existen más valores que se separan de la media a la izquierda. La *kurtosis* (3.05661) indica que esta distribución es más apuntada que la normal, es decir que el peso de los valores se concentra más alrededor de la media (Gráfica 1 y Tabla 1).

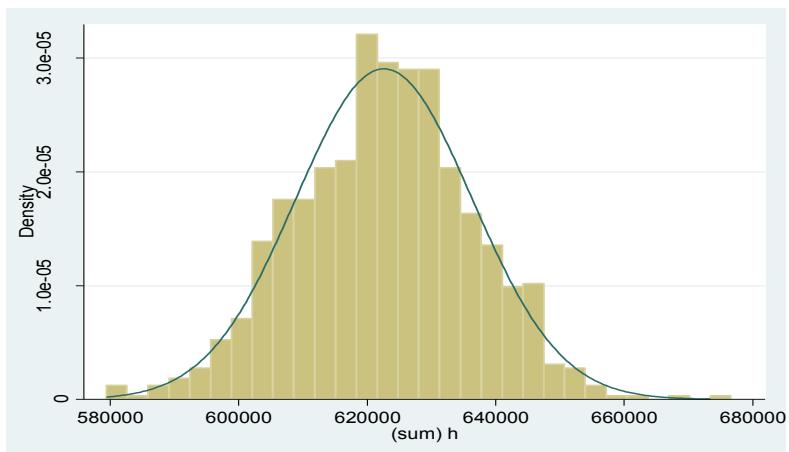
A la distribución de los nacimientos, insumo importante para la proyección, se une la información de los demás grupos que están afectados por la mortalidad y migración, para tener la distribución por muestreo de la población total proyectada a 2006, como se puede observar en la Gráfica 2.

Gráfica 1. Distribución estadística por muestreo del grupo 0-4 años según sexo, Bolivia 2006

Mujeres



Hombres



Fuente: elaboración propia, datos proyectados.

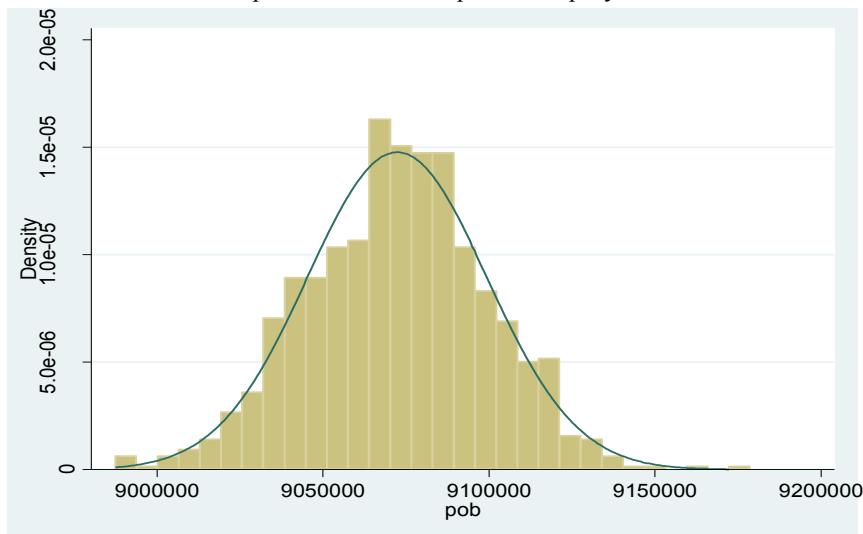
En la Gráfica 2 se observa que la distribución por muestreo de la población proyectada a 2006 tiene las mismas características que la de los nacimientos, porque la fecundidad incrementa la población del grupo 0 a 4 años de la siguiente cohorte, mientras que los otros grupos de edad se ven afectados solamente por la mortalidad y la migración.

Tabla 1. Estadísticos de la distribución de los nacimientos del grupo de 0 a 4 años

	Mujeres	Hombres	Total
Media	601 725	622 612	1 224 337
Desviación estándar	13 273.46	13 734.2	27 008
<i>Skewness</i>	-0.0431838	-0.0431866	-0.04318
<i>Kurtosis</i>	3.05670	3.05661	3.05661
Percentil 50%	601 933.3	622 827.4	1 224 761.0
Límite inferior	580 021	600 155	1 180 175
Límite superior	623 776	645 428	1 269 204

Fuente: elaboración propia, estadísticos generados en STATA.

Gráfica 2. Distribución por muestreo de la población proyectada a 2006



Fuente: elaboración propia, datos proyectados.

Tomando como año base el año 2001, la población total proyectada a 2006 en Bolivia alcanza 9 072 364, de los cuales 4 596 295 son mujeres y 4 476 069 son hombres. Esta información puede variar dentro de los límites del intervalo de confianza encontrado por el método de remuestreo a 95 por ciento de confiabilidad. Es decir que esta estimación puede variar entre [9 028 203 y 9 117 231] esto implica que la proyección puede variar en 44 161 habitantes respecto al límite inferior y en 44 867 respecto al límite superior (Tabla 2). Estos datos son diferentes de las estimaciones oficiales del INE de Bolivia que publican 9 627 269 millones de habitantes para 2006

(4 799 178 hombres y 4 828 091 mujeres) que no se han conciliado con conteos u otros censos.

Tabla 2. Estadísticos de la distribución por muestreo de la población total proyectada a 2006

	Mujeres	Hombres	Total	TGF
Media	4 596 295	4 476 069	9 072 364	3.8348
Desviación estándar	13 273.46	13 734.2	27 008	0.0896
Percentil 50%	4 596 503	4 476 285	9 072 788	3.8388
Límite inferior	4 574 591	4 453 612	9 028 203	3.6932
Límite superior	4 618 346	4 498 885	9 117 231	3.9841

Fuente: elaboración propia, datos proyectados.

Para la pregunta ¿qué tan significativo es considerar un margen de error en la estimación de aproximadamente ± 44 mil nacimientos? Aún es difícil tener una respuesta, sin embargo esta medición de la precisión de la proyección sin duda brinda mayor información al tomador de decisiones, ya que podría planificar los recursos por encima del promedio para acoger un número mayor de nacimientos, en lugar de no considerarlos.

Por otro lado, comparando la distribución por muestreo de la TGF y de la población total proyectada, llaman la atención las grandes diferencias en sus desviaciones estándar. Se tiene que en promedio la dispersión de los datos respecto de la media poblacional es de más o menos 27 mil habitantes respecto de la media y para la TGF es de 0.0896 unidades en promedio ¿Esto significa que pequeñas variaciones respecto de la media de la TGF implican grandes variaciones de la población proyectada respecto de su media? Al calcular el porcentaje que significa esta desviación respecto de la media correspondiente, se tiene que las variaciones alrededor de la media son a lo sumo de dos por ciento aproximadamente y de la población proyectada de 0.3 por ciento.

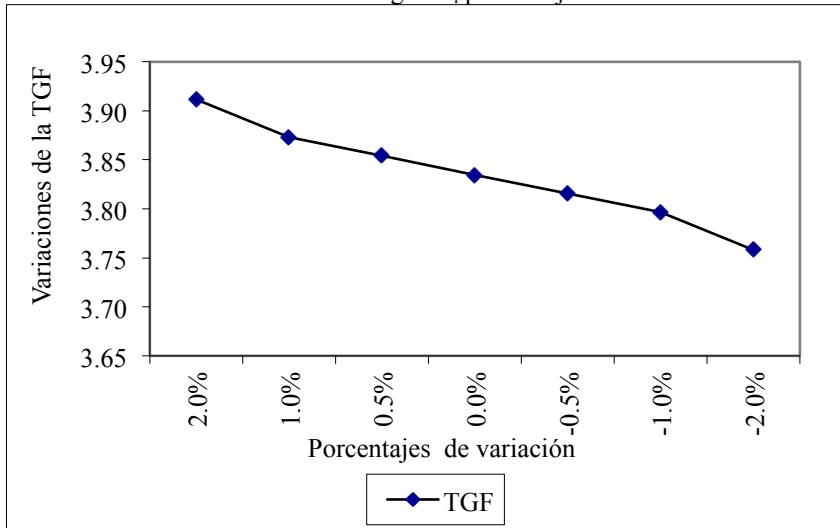
A continuación se presentan los resultados de diferentes experimentaciones variando la TGF para ver su impacto en la población total proyectada.

RESULTADOS: ANÁLISIS DE SENSIBILIDAD DE LA PROYECCIÓN DE LA POBLACIÓN RESPECTO DE LA TGF

En primera instancia se parte de variaciones Δ_i de la TGF a 0.5, uno y dos por ciento por encima y por debajo de la media (como se hizo en el análisis

de sensibilidad de los nacimientos en Argote, 2012). Esto implica que considerando la distribución por muestrado del estimador de la TGF se parte de valores que se encuentran en el área menor a dos por ciento de la curva de la distribución. Acorde a ello la probabilidad de que ocurran estos valores es alta, ya que se encuentran en el área de mayor masa de la distribución. Como para cada estimación de la TGF_i se ha calculado la correspondiente P_i de la población proyectada a 2006, se espera que en principio este tenga un comportamiento descendente como los valores de la Gráfica 3.

Gráfica 3. Variaciones de la TGF según Δ_i porcentaje de variación



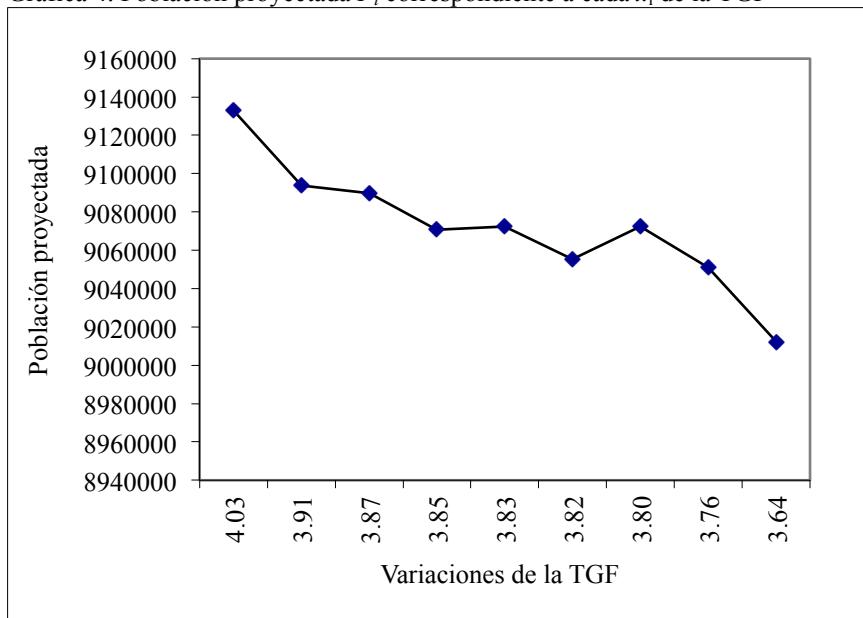
Fuente: cálculos propios.

Véase que los valores en el eje x corresponden a las Δ_i en orden descendente de izquierda a derecha.

Para valores cercanos a la media, se puede observar que el comportamiento de las P_i no es estable, como lo es el comportamiento de la TGF, a valores Δ_i definidos *a priori* (Gráfica 3). Ocurre el caso de que para un valor de uno por ciento por debajo de la media de la TGF_p , su correspondiente valor P_i en población proyectada es mayor al anterior P_{i-1} (Gráfica 4).

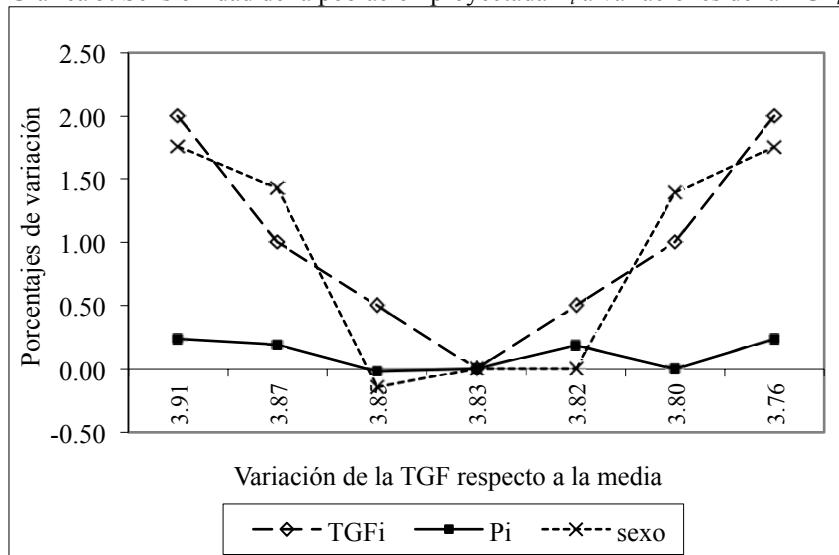
Contrario a lo que se esperaba, a variaciones de la TGF_i por encima y debajo de la media, la población proyectada P_i varía poco alrededor de la media (Gráfica 5). Por ejemplo, a una variación de dos por ciento de la TGF, la población P_i se aleja aproximadamente 0.3 por ciento de la media, es decir se observa un efecto reductor. Al parecer el método por componentes hace que los efectos se atenúen a niveles agregados.

Gráfica 4. Población proyectada P_i correspondiente a cada x_i de la TGF



Fuente: cálculos propios.

Gráfica 5. Sensibilidad de la población proyectada P_i a variaciones de la TGF_i



Fuente: cálculos propios.

Es interesante ver que el método experimental usado permite analizar el modo en que se están dando estas relaciones de forma multivariada. Para analizar la relación entre cambios en la TGF_i y la población proyectada P_r , se amplió la muestra a 30 valores por debajo de la media de la TGF (subestimación) y se tomó otra muestra de 30 valores en la cola superior de la distribución (sobreestimación); estos son valores que podría adoptar el estimador.

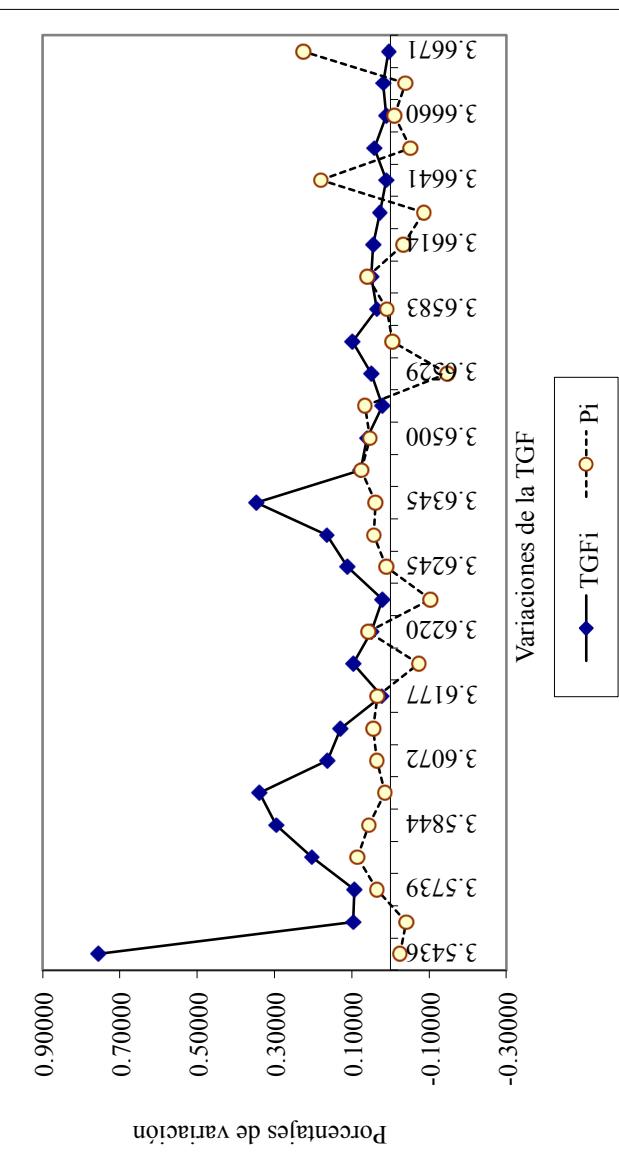
Considerando una muestra mayor de valores de la TGF, se puede observar que no se identifica una relación directa o inversa entre TGF_i y la población proyectada (Gráfica 7). Para una muestra de 30 valores de la TGF entre 3.543 y 3.667, en la Gráfica 6 se observan los porcentajes de variación entre TGF_i y TGF_{i-1} y de sus valores correspondientes de P_r .

Llama la atención que para ciertos valores de la TGF el porcentaje de variación coincide con la variación entre P_{i-1} y P_i en 2006 (puntos 3.618; 3.622; 3.654 tienen el mismo porcentaje de variación que la población total) por lo que en estos casos no se presenta el efecto amplificador. Se observan tres picos en que el porcentaje de variación de la TGF respecto de los próximos es mucho mayor que para la P_r , 0.75, 0.35, 0.36 por ciento, para estos valores la P_i varía entre cero y 0.1 por ciento por lo que se presentaría el efecto reductor. Los picos de la P_i en 3.664 y 3.667, tienen un porcentaje variación de 0.15 y 0.2 por ciento aproximadamente respecto de variaciones cercanas a cero por ciento de la TGF, que presentan un efecto de leve amplificación (Gráfica 6).

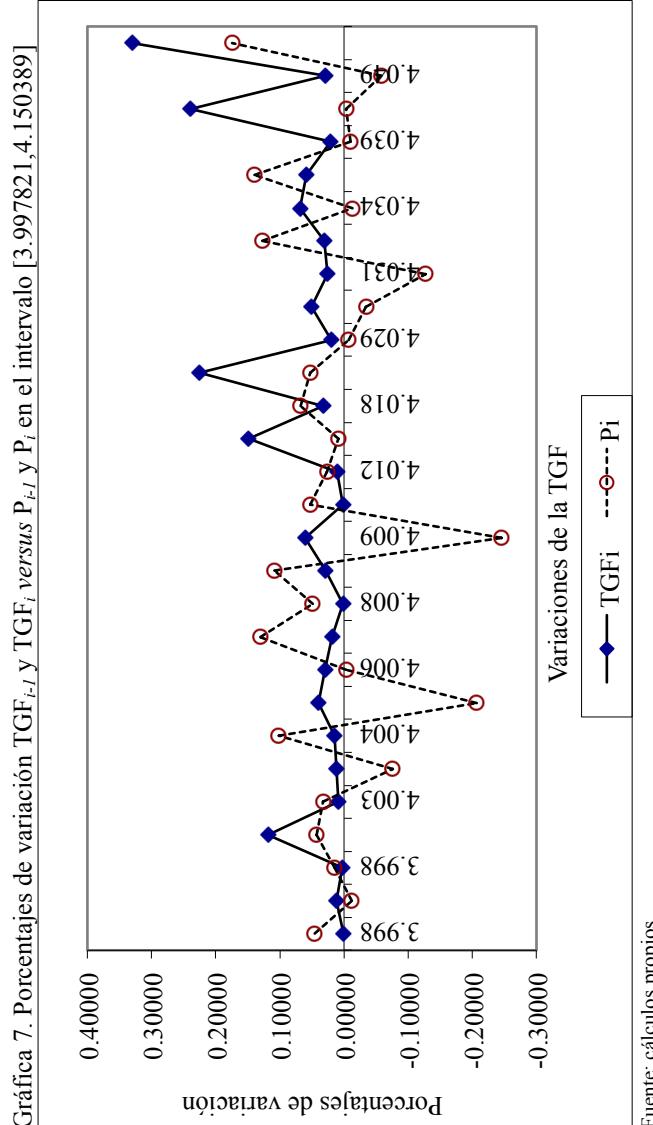
En la cola superior de la distribución por muestreo de la TGF, se observa un comportamiento diferente (Gráfica 7). Si bien se presentan mayor cantidad de picos para las P_i en que sus porcentajes de variación son mayores a los de la TGF_r , a lo sumo éste llega a ser el triple de la variación en la TGF_r , lo cual no se puede considerar amplificación del error bajo el concepto planteado en Argote (2011) que encontró que a pequeños cambios de la TGF_i los nacimientos en determinado grupo quinquenal de edad de la madre llegaron a variar incluso en 20 por ciento, lo cual se consideró efecto amplificador.

Por otro lado se observan seis picos en los que el porcentaje de variación de la TGF_i es mayor al de la P_r . En promedio un porcentaje de variación alrededor de 0.25 por ciento de la P_r que repercute en una variación menor a 0.1 por ciento en las P_r .

Gráfica 6. Porcentajes de variación entre TGF_{i-l} y TGF_i , versus P_{i-l} y P_i en el intervalo [3.543646, 3.667286]



Fuente: cálculos propios.



DISCUSIÓN

Esta investigación muestra los resultados de un ejercicio de proyección de la población boliviana de 2001 a 2006, con el objetivo de encontrar la distribución estadística por muestreo de la población proyectada a 2006 generada con la distribución por muestreo de la TGF de 2003 encontrada en Argote (2011) aplicando el método de remuestreo. Con ambas distribuciones se realizó un análisis de sensibilidad para evaluar si se presenta el efecto mariposa en la población proyectada a pequeños cambios de la TGF. La perspectiva de análisis es hacia el error y la incertidumbre atada a las estimaciones.

El método aplicado brinda información lógica, coherente e interpretable para la evaluación de la hipótesis de investigación. Según la estimación del INE de Bolivia, la población total proyectada a 2006 es de 9 627 269 millones de habitantes, de los cuales 4 799 178 son hombres y 4 828 091 son mujeres. En la presente investigación se estimó una población proyectada de 9 072 364 (4 476 069 hombres y 4 596 295 mujeres) con un intervalo de confianza de [9 028 203; 9 117 231]. Las estimaciones presentadas están seis por ciento debajo de las estimaciones oficiales proporcionadas por el INE de Bolivia. Asimismo, el método permite estimar los intervalos de confianza de la población proyectada por grupos quinquenales de edad, lo cual es importante para el análisis del error e incertidumbre de las estimaciones realizadas (Argote, 2011).

Con base en la evidencia proveniente de las distribuciones de la TGF y de la población proyectada a 2006 para los grupos quinquenales de edad entre cero y 80 o más años, no se podría afirmar que existe un efecto de amplificación del error entre estas dos distribuciones. En términos relativos, a pequeñas variaciones de la TGF no se observan grandes cambios en los resultados de proyección de la población boliviana a 2006, esto se puede atribuir a un efecto de cancelación debido a la agregación de los datos y a la estructura misma del método de proyección. Sin embargo el análisis de sensibilidad y su construcción ha permitido evidenciar la aleatoriedad y no linealidad de las relaciones entre los componentes de la proyección y el error asociado a estas estimaciones.

En términos absolutos se encontró que el intervalo de confianza en que se puede encontrar la población boliviana total proyectada a 2006 se encuentra entre más o menos 44 mil nacimientos respecto de la media, que corresponden a 0.4 por ciento del total de la población. Nuevamente surge

la pregunta ¿qué tan significativo es considerar o no estos nacimientos para una política pública orientada a los recién nacidos, por ejemplo?

Para los tomadores de decisión son muy útiles los intervalos de confianza generados porque brindan un margen de error de las estimaciones. La perspectiva de análisis cambia al considerar la distribución estadística de los indicadores. Por lo tanto, es una herramienta que permite trabajar con la incertidumbre, planificar bajo diferentes escenarios y evaluar la toma de decisiones.

BIBLIOGRAFÍA

ALHO, Juha, Maarten ALDERS, Harri CRUIJSEN, Nico KEILMAN, Timo NIKANDER, Dinh QUANG PHAM, 2006, "New forecast: Population decline postponed in Europe", en *Statistical Journal of the United Nations ECE* 23, IOS Press.

ALHO, Juha, 2005, *Statistical Demography and Forecasting*, Springer Series in Statistics.

ARGOTE, Milenka L., 2007, "Estimación de la distribución estadística de la tasa global de fecundidad", en *Papeles de Población*, 13 (54), octubre-diciembre, Toluca, Toluca.

ARGOTE, Milenka L., 2009, "Comparación y evaluación de la distribución estadística del estimador de la tasa global de fecundidad de Bolivia en 1998 y 2003", en *Papeles de Población*, 13 (62), octubre-diciembre, Toluca.

ARGOTE, Milenka L., 2012, "Análisis de sensibilidad de los nacimientos respecto a la Tasa Global de Fecundidad", en *Papeles de población*, 18 (72), Toluca.

AXELROD, Robert, 1997, "Advancing the Art of simulation in the social sciences", en Rainer HEGSELMANN and Pietro TERNA (eds.), *Simulating Social Phenomena School of Public Policy*, University of Michigan, Ann Arbor, Published in Rosario Conte, Berlin: Springer USA.

BAGAJEWICZ, Miguel J., 2005, "On a new definition of a stochastic-based accuracy concept of data reconciliation-based estimators", en *Computer Aided Chemical Engineering*, vol. 20, part 2.

BAGAJEWICZ, Miguel J. y Nguyen DUY QUANG, 2006, *Stochastic-based accuracy of data reconciliation estimators for linear systems*, University of Oklahoma.

BARRETT, Chris, Stephen EUBANK y Madhav MARATHE, 2006, *Modeling and Simulation of Large Biological, Information and Socio-Technical Systems: An Interaction Based Approach*, Virginia Polytechnic Institute and State University, To appear as a book chapter in the book titled Interactive Computation: the New Paradigm, Goldin, Smolka, Wegner, Editors, Esprtinger Verlag, Sept, Blacksburg Virginia.

- BOTARGUES, Patricia y Diego PETRECOLLA, 1999, “Estimaciones paramétricas y no paramétricas de la distribución del ingreso de los ocupados del Gran Buenos Aires, 1992-1997”, en *Económica*, vol. 45(1).
- CASWELL, Hall, 2009, *Stage, age and individual stochasticity in demography*, Oikos, Journal compilation # 2009 Oikos, Subject Editor: Per Lundberg.
- CASWELL, Hall, 2008, “Perturbation analysis of nonlinear matrix population models”, en *Demographic Research*, vol. 18.
- CANUDAS ROMO, Vladimir, 2003, *Decomposition methods in demography*.
- CEPAL, 2006, *Chile: estimaciones y proyecciones de la población. 1950-2050*, Instituto Nacional de Estadísticas/Centro Latinoamericano y Caribeño de Demografía, Santiago de Chile.
- CHAMBERS, R. L. y A.H. DORFMAN, 1994, *Robust simple survey inference via bootstrapping and bias correction: the case of ratio estimator*, Southampton Statistical Sciences, Research Institute, University of Southampton.
- COCHRAN, William, 1977, *Sampling Techniques*, Wiley Publications in statistics. Library of Congress.
- CONAPO, 2006, *Proyecciones de la Población de México, 2005-2050*, Consejo Nacional de población, México.
- DIAS, Tiago, N. ROMA y L. Sousa, 2001, *Two-Level scalable motion estimation architecture with fractional-pixel accuracy and efficient data re-usage*, Instituto Superior Técnico/INESC-ID, Lisboa.
- ESCOBEDO RIVERA, José, 2007, “El dato en la investigación demográfica: una visión epistemológica”, en *Papeles de Población*, Nueva Época año 13 núm. 54, octubre-diciembre, Toluca.
- EFRON, B. y R. TIBSHIRANI, 1993, *An introduction to the Bootstrap*, Chapman & Hall, Nueva York.
- FERRARI, Salvador Rojí, 2005, *La complejidad: un nuevo enfoque de la economía financiera*, Universidad Complutense de Madrid, núm. 216, mayo-agosto.
- GARCÍA-OCHOA, Liliana, Ana Lilia LAUREANO-CRUCES y Raymundo PERALTA HERRERA, 2006, “Economía simulada por un sistema multiagente”, en *XIX Congreso Nacional y V Congreso Internacional de Informática y Computación de la ANIEI*, 25-27 de octubre, Tuxtla Gutiérrez.
- GOTTLIEB, Alex D., 2001, *Asymptotic Accuracy of the Jackknife Variance Estimator for Certain Smooth Statistics*, en scholar.google.com.
- GRIFFITHS, Paula, Zoe MATTHEWS y Andrew HINDE, 2000, “Understanding the sex ratio in India: a simulation approach”, en *Demography*, vol. 37, núm. 4, noviembre.

GÓMEZ, R., G. NAVAS, P. CUADROS y G. GANVITANO, 2009, “Análisis computacional de modelos biológicos para su aplicación a modelos económicos”, en *Formación Universitaria*, vol. 2, núm. 5.

HEISENBERG, W., 2013, *The physical principles of the quantum theory*, Courier Corporation.

KENNETH WACHTER, John Knodel, y Mark VANLANDINGHAM, 2001, *Parental Bereavement: Heterogeneous Impacts of AIDS in Thailand*, Population Studies Center at the Institute for Social Research, University of Michigan.

KEYFITZS, Nathan y Caswell HAL, 2005, *Applied mathematical demography*, Springer.

LIVIO, Mario, 2013, *Errores geniales que cambiaron al mundo*, Editorial Ariel, España.

LORENZ, Edward, 1963, “Deterministic Nonperiodic Flow”, en *Journal of the Atmospheric Sciences*, vol. 20, march, Massachussets Institute Of Technology.

MÉNDEZ, Ignacio, 2004, *La Estadística*, Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, UNAM, México.

MIRAMONTES, Octavio, 1999. “Los sistemas complejos como instrumentos de conocimiento y transformación del mundo”, en *Perspectivas en las teorías de sistemas*.

MINA-VALDES, Alejandro, 2012, “La población mexicana censada en 2010 y su comparación con las proyecciones demográficas”, en *Papeles de población*, 18(73).

ORTEGA OSONA, José Antonio y HANS-PETER Kohler, 2000, “¿Está cayendo realmente la fecundidad española? Separación de los efectos de intensidad, calendario y varianza en el Índice Sintético de Fecundidad”, en *Revista Española de Investigaciones Sociológicas*, núm. 96, Centro de Investigaciones Sociológicas, España.

PRESSAT, Ronald, 1989, *El análisis demográfico*, Fondo de Cultura Económica, México.

RAJ, Des, 1968, *Sampling theory*, Mac Graw Hill.

SAURA BAICOA, Dulce y Ángel RODRÍGUEZ GARCÍA BRAZALES, 2003, “No linearidad y economía austriaca”, en *Rev. Econ. Ros*, 5(2), diciembre, Bogotá.

SITTER, R., 1992, “Comparing three bootstrap methods for survey data”, en *The Canadian Journal of Statistics*, núm. 20.

STOVER y KIRMEYER, 2007, *DemProj*, version 4 un programa computarizado para calcular proyecciones de población, El proyecto policy.

Milenka Linneth Argote Cusi

Estudiante de Doctorado en Estadística en la Universidad Nacional de Colombia. Maestra en Estudios de Población por la Facultad Latinoamericana de Ciencias Sociales sede México. Ingeniería de Sistemas por la Escuela Militar de Ingeniería de Bolivia. Consultora-Investigadora en la Universidad EAN de Colombia. Es miembro de los grupos de investigación de Colciencias, Colombia. Los últimos proyectos realizados son: Análisis de brechas tecnológicas en los sectores del Programa de Transformación Productiva de Colombia, Censo empresarial a estudiantes de la Universidad EAN, Colombia y Estudio de rutas de empleabilidad y generación de ingresos en un municipio de Bogotá. Entre sus publicaciones más destacadas: *Uso de Lógica difusa en proyecciones de población* (2014), *La demostración del Teorema Central del Límite* (2012), *Una mirada a las empresas de estudiantes y egresados: el caso de la Universidad EAN* (2013), *Ánalisis de sensibilidad de proyecciones de población a pequeños cambios de la Tasa Global de Fecundidad* (2012), *Comparación y evaluación de la distribución estadística del estimador de la tasa global de fecundidad de Bolivia en 1998 y 2003* (2009), *Estimación de la distribución estadística de la Tasa Global de Fecundidad* (2007). Líneas de investigación: Demografía Estadística, Demografía Empresarial, Modelaje y Simulación, Inteligencia Artificial y Análisis de la incertidumbre.

Dirección electrónica: milenqita@hotmail.com

Artículo recibido el 6 de septiembre de 2012 y aprobado el 1 de diciembre de 2014.