

Análisis de sensibilidad de los nacimientos respecto a la Tasa Global de Fecundidad

Milenka L. ARGOTE

Corpotalentos ONG

Resumen

En el presente trabajo se realiza un análisis de sensibilidad de los nacimientos a pequeños cambios de la Tasa Global de Fecundidad (TGF) para determinar si se presenta el efecto mariposa. En cuanto al método, se toma como fuente de datos la muestra de la Encuesta Nacional de Demografía y Salud (ENDSA) de Bolivia en 2003, que es una muestra probabilística a nivel nacional, estratificada y bifásica. Por otro lado, se crean los algoritmos de estimación de la TGF, de remuestreo y de generación de las distribuciones estadísticas de los nacimientos por grupos quinquenales de edad a partir de la muestra. Se construyó un índice para evaluar la normalidad y homogeneidad de las distribuciones y se realizó un análisis de sensibilidad por objetivo de los nacimientos con base en las distribuciones. Los resultados, de acuerdo al índice la distribución de los grupos de edad 20-25, 25-30, 30-35 se asemejan más a la normal y son homogéneos; de los grupos 15-20, 35-40, 40-45 las distribuciones se asemejan a la normal y son heterogéneas y del grupo 45-49 es diferente a la normal y es el más heterogéneo. Según el análisis de sensibilidad de los nacimientos a variaciones de medio, uno y dos por ciento por encima y por debajo de la media de la TGF se observó un comportamiento no lineal y heterogéneo en los diferentes grupos quinquenales, y para el último grupo el porcentaje de variación es mayor a 30 por ciento, por lo cual se presenta un efecto que amplifica el error en los nacimientos; entonces la TGF es una variable sensible, que según estos resultados es factor a tomar en cuenta en proyecciones de población.

Palabras clave: análisis de sensibilidad, remuestreo, distribución estadística de nacimientos, efecto mariposa, amplificación del error.

Abstract

Sensitivity analysis of births in respect to Global Fertility Rate

In the present article we perform a sensitivity analysis of births as for small changes in the change of global fertility rate (GFR) to determine whether there is butterfly effect. In relation to the method, we take as a data source the sample of the national Survey on Demography and Health (ENDSA) of Bolivia in 2003, which is a probabilistic sample at national level, stratified and two-phased; separately, the algorithms of estimation of GFR, of resampling and generation of statistical distributions of the births are created by quinquennial age groups from the sample. An index was built to assess the normality and homogeneity of distributions and a sensibility analysis was carried out by objective of the births based on the distributions. The results, according to the index the distribution of the 20-25, 25-30, 30-35 age groups are more similar to the normal and homogeneous; for the 15-20, 35-40, 40-45 age groups the distributions are similar to the normal and heterogeneous; and the 45-49 group is different the normal and the most heterogeneous. According to the sensitivity analysis of births at variations of half, one and two percent over and under the mean GFR there was a nonlinear heterogeneous behavior in the different quinquennial age groups, and for the last group the variation percentage is higher than 30%, so there is an effect that amplifies the error in births; therefore, GFR is a sensitive variable, which according to these results is a factor to be accounted for in demographic projections.

Key words: sensitivity analysis, resampling, statistical distribution of births, butterfly effect, error amplification.

INTRODUCCIÓN

La medición de los sistemas siempre ha sido una preocupación de las diferentes ciencias. Por ejemplo, la teoría de la medida y las probabilidades nace con el fin de asignar de forma apropiada una medida probabilística a determinado evento o conjunto de eventos. En realidad cualquier cosa es susceptible de ser medible, sin embargo, algunos sistemas son más difíciles de medir que otros, entonces podemos estimar.

Los primeros trabajos de medición se han realizado en las ciencias naturales a través de experimentos, sin embargo, la inquietud de cuantificar ciertas variables o fenómenos de los sistemas sociales no es nueva (Barrett *et al.*, 2006; Bagajewicz, 2005; Ortega y Kohler, 2000). Los científicos sociales cada vez más se inclinan a representar mediante indicadores o estadísticas sus hallazgos, sean estos cuantitativos, cualitativos o mixtos. Especialmente en demografía se calculan tasas, cocientes, indicadores que permiten realizar una evaluación cuantitativa de los fenómenos poblacionales. En este sentido, es importante interpretar los indicadores de forma multidimensional atendiendo al origen de su construcción, por ello, muchas controversias emergen en la interpretación de los indicadores demográficos debido a la dificultosa tarea de representar el estatus de la población, y mayor es la discusión acerca de su precisión (Argote, 2009).

En este marco de referencia se realiza la presente investigación. La Tasa Global de Fecundidad (TGF) es un indicador resumen de la fecundidad de las mujeres en edad fértil de determinada población, cuyo numerador son los nacimientos de la mujeres en determinado grupo de edad quinquenal y el denominador el tiempo de exposición de las mujeres al riesgo de quedar embarazada en años persona, si contamos con la distribución estadística por muestreo de la TGF generada a partir de la muestra de mujeres de la Encuesta Nacional de Demografía y Salud (ENDSA) 2003 de Bolivia, ¿es posible evaluar como cambian los nacimientos por grupos quinquenales a pequeños cambios de la TGF? ¿Se presenta un efecto amplificador en esta relación? Estas son las preguntas de investigación que se responderán en los siguientes apartados.

Considerando las investigaciones anteriores de Argote (2007 y 2009) se utilizan los algoritmos de estimación de la TGF y de remuestreo para generar paralelamente la distribución por muestreo del estimador de los nacimientos por grupos quinquenales de edad en 2003. Esto último con el objetivo de realizar un análisis de sensibilidad de los nacimientos a pe-

queños cambios de la TGF, con lo cual se pueda evidenciar si se presenta el efecto mariposa que amplifica el error (propio de los sistemas caóticos).

Dentro los hallazgos no cabe duda que a mayor información menor incertidumbre podemos esperar (Caswell, 2008; Alho, 2006). Acorde a nuestras hipótesis de investigación, la fecundidad en Bolivia se comporta de manera heterogénea para los diferentes grupos quinquenales de edad. Si bien Argote (2009) encontró que la distribución del estimador de la TGF a nivel nacional tenía un comportamiento semejante a la distribución normal en 2003, en esta investigación se verifica que cada grupo quinquenal tiene un comportamiento diferente. Producto del análisis de sensibilidad de la TGF se observa que se presenta el efecto mariposa, ya que a pequeños cambios de la TGF los cambios en los nacimientos son diferentes y superiores, y se amplifican en los dos últimos grupos. Por lo tanto, la fecundidad es un fenómeno no lineal y heterogéneo, por lo cual es muy probable que forme parte de un sistema complejo, el poblacional.

¿QUÉ ES UNA ESTIMACIÓN?

Los conceptos básicos bien definidos son la base para el desarrollo de teorías sólidas. Un concepto al parecer simple encierra grandes ideas acerca del modelado de sistemas reales, por ello el concepto de estimación es clave.

Cuando deseamos contar cosas las contamos, pero, ¿qué pasa si el número de objetos crece en gran medida? Ocurre que se nos dificulta contar. En este caso lo que hacemos es “adivinar” el número que deseamos conocer, es decir, realizamos una estimación. Pero, ¿qué tanto me acerco al número real? Se deben aplicar ciertas metodologías o técnicas para realizar una buena estimación. La disciplina que nos proporciona métodos y técnicas de medición y estimación son la estadística y la teoría de probabilidades. De esta manera tenemos un aparato sólido de métodos de estimación para aplicar a sistemas inalcanzables como el poblacional.

Por ejemplo, para la estimación de la TGF es usual adoptar el supuesto de la normalidad basado en el Teorema Central del Límite.¹ En general es usual utilizar este teorema para sustentar que determinado fenómeno tiene comportamiento normal en promedio. Según la definición teórica del teorema: “Sea X, X_1, X_2, \dots una sucesión de variables aleatorias independientes e idénticamente distribuidas con $E[X_n] = \mu$ y $\text{Var}[X_n] = \sigma^2 < \infty$ ”. Supóngase

¹ Se denomina Teorema Central del Límite por su importancia central en la teoría de la medida y probabilidades, en lugar de la denominación Teorema del Límite Central (Blaiotta y Delieutraz, 2004).

además que N es una variable aleatoria con distribución $N(0,1)$. Si $S_n = X_1 + X_2 + \dots + X_n$, $n \geq 1$, entonces:

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} N(0,1)$$

Es decir, la suma de la distribución de variables aleatorias idénticamente distribuidas y tiene comportamiento normal cuando n tiene a infinito.

Sin embargo, en la actualidad, con ayuda de los procesadores, podemos disponer de otros métodos que nos ayuden a no sujetarnos por defecto a los supuestos o verificarlos. En este ámbito juegan un papel de gran importancia los métodos de repetición, simulación y las probabilidades. En los sistemas reales debemos jugar con las probabilidades para acercarnos a ese mundo real, que concebimos desde nuestro punto de vista. En este sentido, los métodos de repetición y simulación por computadora en la ciencia nos permiten realizar experimentaciones con datos empíricos de sistemas inalcanzables, elaborar escenarios y manejar la incertidumbre. El presente trabajo es un ejemplo del manejo de la incertidumbre en el estimador de la TGF para la población de mujeres bolivianas en edad fértil durante 2003.

Siempre me he preguntado cuando estimábamos la TGF u otros indicadores demográficos por diferentes métodos en mis clases de posgrado, ¿cuál estimación elegimos? Al parecer si redondeamos obtenemos el mismo número, pero, esos decimales redondeados nos dicen algo? ¿Se puede redondear así de fácil los indicadores de sistemas complejos en donde pequeños errores pueden provocar grandes errores en la proyección de la información? ¿Será que estos sistemas tienen comportamiento complejo? ¿Será posible mejorar las estimaciones, errar menos?

En demografía se debate bastante el hecho de corregir las estimaciones por cientos o miles de personas estimadas para determinado momento. Y para efectos de planificación seguramente será mejor estimar con menor error y que no falte un número significativo de personas no estimadas que podrían ser beneficiarias o no de determinada política de salud o de mitigación de la pobreza, por ejemplo. Sin embargo, esto no es tan fácil, ya que trabajamos con sistemas inaccesibles y de gran escala, por esta razón es interesante explorar otras teorías para abordar esta complejidad y, a partir del conocimiento, aplicar nuevas herramientas y técnicas.

LA POBLACIÓN COMO SISTEMA COMPLEJO

El ser humano observa su entorno y la ciencia ha emergido de los cuestionamientos que se ha hecho este individuo y de lo observado y percibido

con sus sentidos. Asimismo, la dinámica del ser humano, producto de sus diversas actividades, ha conformado una naturaleza social con dinámica propia cada vez más compleja, acorde al desarrollo de las sociedades y de la tecnología. En un principio los sistemas son accesibles a una escala pequeña, factibles de contabilizar, ordenar y organizar; en tal escenario los métodos tradicionales se han aplicado muy bien. Sin embargo, con el crecimiento poblacional y la cada vez más compleja red de actividades que conlleva, se hace necesaria la construcción de nuevos marcos teóricos que consideren la complejidad de estos sistemas mayores.

Un camino para analizar estos sistemas complejos es a través de la teoría general de sistemas y, dentro de ella, la teoría de los sistemas complejos cuyo concepto base es la definición de sistema (conjunto de elementos interconectados que se rigen bajo una dinámica propia). En este concepto están implícitos otros dos, el de estructura y función. La estructura hace referencia a una cantidad de elementos interrelacionados (como la estructura de un edificio), en cambio el término de función hace alusión a las leyes que determinan el movimiento o dinámica de dicha estructura; del todo en su conjunto emerge un comportamiento mayor a la simple suma de sus partes (Miramontes, 1999).

La ingeniería industrial, la ingeniería de sistemas y la economía son disciplinas que aplican los conceptos de la teoría de sistemas complejos con el objetivo de modelar y estudiar la realidad (Ramírez, 1999). En demografía es posible aplicar esta perspectiva de análisis, ya que los componentes de la dinámica poblacional pueden ser vistos como parte de un sistema complejo, compuesto por elementos (individuos) que se relacionan y generan una dinámica propia. En un escenario los componentes del sistema tienen diferentes características demográficas como edad, sexo, estado civil, etc., pero estos individuos no se encuentran estáticos, sino que se reproducen, se movilizan en el espacio pequeñas o grandes distancias y cumplen un ciclo de vida. Si imaginamos esta complejidad, también podemos pensar que cuantificar o medir este sistema tiene su grado de dificultad. La demografía ha desarrollado un aparato teórico e instrumental que nos permite realizar análisis demográfico a partir de la construcción de indicadores, tasas, cocientes, razones, etc. (Pressat, 2000), sin embargo es pertinente enriquecer las herramientas ya establecidas con nuevos paradigmas, que mejoren nuestras estimaciones y su interpretación.

Recordemos que los diferentes programas sociales utilizan los datos de proyecciones de población en general y por edades como denominadores en la construcción de indicadores, con base en esto se diseñan políticas

sociales para mejorar los indicadores o, en su defecto, disminuirlos para beneficio de la condiciones de vida de la población. *A posteriori* se construyen también indicadores para evaluar los resultados de la aplicación de dichas políticas que retroalimentan al ejecutor. De ahí la importancia de la construcción e interpretación adecuada de estos indicadores demográficos y de procurar disminuir el error en nuestras estimaciones y manejar la incertidumbre.

Por esta razón resulta interesante analizar la complejidad de la población considerando la TGF como indicador demográfico. La construcción de este indicador relaciona los nacimientos y los tiempos de exposición, y a partir de estos cálculos podemos explorar la incertidumbre que esta detrás de las estimaciones, con el fin de que esta investigación tenga utilidad en planificación.

LA TASA GLOBAL DE FECUNDIDAD

Dentro de todos los indicadores demográficos se estudia la fecundidad por su importancia en proyecciones de población. Acorde al documento de la CEPAL (2006:15), la fecundidad es el componente que más incide en el tamaño, la composición y la distribución de la población, por lo cual en una proyección de población generalmente se toman cuatro hipótesis de la evolución probable de su nivel: constante, baja, alta y media. En este sentido, según estos documentos, estimar el número de nacimientos futuros es el aspecto más importante y complejo de las proyecciones de población.

Su importancia se debe, principalmente, a la influencia de este número en el volumen y composición final de la población por edades y su complejidad, a la incertidumbre que generalmente se tiene de los factores que inciden en los niveles de la fecundidad de la sociedad. Se ha observado que ambos aspectos, muchas veces provocan cambios muy difíciles de predecir en el transcurso del tiempo (CEPAL, 2006: 28).

Las medidas de fecundidad son numerosas, considerando el interés en ciertos grupos poblacionales, las escalas de tiempo usadas para el calendario del fenómeno y la distinción según el orden del nacimiento o la diversidad de las definiciones del orden (Pressat, 2000).

Dentro de las medidas por periodo son de principal interés en esta investigación las Tasas Específicas de Fecundidad (TEF) por edad y, su indicador resumen, la Tasa Global de Fecundidad (TGF). Las TEF por grupos quinquenales de edad se definen como el cociente de los nacimientos ocurridos a

mujeres de un grupo de edad $Nac_{x,x+5}$ entre los años-persona vividos en exposición al riesgo de las mujeres $TeMujeres_{x,x+5}$ del mismo grupo de edad.

$$TEF_{x,x+5}^{t,t+1} = \frac{Nac_{x,x+5}^{t,t+1}}{TeMujeres_{x,x+5}^{t,t+1}} \quad (1.1)$$

$$TGF^{t,t+1} = 5 * \sum_{i=1}^7 TEF_i^{t,t+1} \quad (1.2)$$

La TGF representa el número de hijos por mujer al final de su vida reproductiva, bajo el supuesto que a lo largo de su vida tendrá la fecundidad presente.² Se calcula como la sumatoria de las tasas específicas de fecundidad en los siete grupos quinquenales (15 a 49), multiplicado por cinco. También podemos decir que la TGF es una combinación lineal de las TEF; desde el punto de vista estadístico se trata de una combinación lineal de razones.³

En virtud de la complejidad de medición de una tasa que pretende cuantificar la dinámica reproductiva de una mujer, su interpretación no es sencilla. También hay que considerar que el supuesto de fecundidad constante a lo largo de la vida reproductiva de una mujer no siempre es cierto en los sistemas dinámicos como el poblacional. A ello se añade que para estimar la TGF es más común disponer de muestras complejas como fuente de información, que poseen un margen de error. Es decir, el indicador resumen de la TGF esta determinado por varios factores: métodos, supuestos, fuentes de información, lo cual añade incertidumbre al estimador de una tasa y, por ello, es interesante analizar si pequeñas variaciones en su valor son significativas y si representan grandes cambios en la estructura por edades de las TEF, lo cual estaría evidenciando la presencia del “efecto mariposa” que es nuestra hipótesis de investigación (CEPAL, 2006: 28-15).

LA TEORÍA DEL CAOS Y EL EFECTO MARIPOSA

Bajo la perspectiva de los sistemas complejos se han desarrollado nuevos paradigmas para el estudio de sistemas dinámicos, los cuales pueden ser estables, inestables y caóticos.

² Se hace referencia al supuesto de cohorte ficticia.

³ Tenemos en el numerador el conjunto de los números enteros (nacimientos) y en el denominador el conjunto de los reales positivos (tiempo de exposición), por lo tanto, el dominio de la tasa es el de los números reales positivos, que es nuestra variable aleatoria.

Un sistema estable tiende a lo largo del tiempo a un punto, u órbita, según su dimensión (atractor). Un sistema inestable se escapa de los atractores. Y un sistema caótico manifiesta los dos comportamientos. Por un lado, existe un atractor por el que el sistema se ve atraído, pero a la vez, hay “fuerzas” que lo alejan de este. De esa manera, el sistema permanece confinado en una zona de su espacio de estados, pero sin tender a un atractor fijo (Tejada, 2003).

Bajo estos conceptos se ha desarrollado, y se sigue desarrollando, la llamada Teoría del Caos, para describir en particular los sistemas caóticos.

Una de las principales características de los sistemas caóticos es su sensibilidad a las condiciones iniciales. Uno de los ejemplos más importantes para ilustrar esta característica es el modelo de predicción del tiempo atmosférico desarrollado por Edward Lorenz. A partir de un sistema de ecuaciones diferenciales y determinadas condiciones iniciales, el modelo podía predecir el tiempo atmosférico, sin embargo, a pequeñas variaciones de las condiciones iniciales se obtenían resultados muy diferentes; a esta característica de los sistemas dinámicos caóticos se le llamó “efecto mariposa”. Esta característica se presenta en varios sistemas como el atmosférico, en la física, en economía e inclusive en el crecimiento poblacional (Ramírez, 1999).

En demografía para las proyecciones de población se parte de tres variables importantes: la fecundidad, la mortalidad y la migración, que determinan el nivel de población en un tiempo t . Podemos decir que definir estas variables al inicio de una proyección de población, es definir las condiciones iniciales del modelo de predicción. Lo que se asume cuando se adoptan diferentes hipótesis de la fecundidad es que pequeñas variaciones en el parámetro no afectaran en gran medida el resultado de la población proyectada, si bien esto puede ser cierto por el teorema central del límite para variables aleatorias idénticamente distribuidas, a medida que el tamaño muestral se hace grande, ¿qué ocurre al desagregar la información por nacimientos en grupos quinquenales de edad?

Considerando este apartado, uno de los objetivos de la presente investigación es evaluar la sensibilidad de los nacimientos a pequeñas variaciones de la TGF. Ya que el estimador de la TGF no tiene un comportamiento lineal, resulta interesante analizar si sus componentes tienen un comportamiento estable o inestable propio de los sistemas caóticos.

DATOS Y MÉTODOS

El centro de la presente investigación es realizar un análisis de sensibilidad de los nacimientos respecto a pequeños cambios de la TGF, en apoyo a la

toma de decisiones en escenarios de incertidumbre. Bajo esta premisa, para abordar, analizar y verificar las hipótesis de investigación se recurrió de forma transversal a la teoría de sistemas complejos, a la teoría de la medida y probabilidades, a herramientas computacionales para la simulación, a la estadística y a la demografía.

La presente investigación es una consecución de las investigaciones anteriores de Argote (2007 y 2009) alrededor de la distribución estadística por muestreo de la TGF y los métodos de remuestreo. La principal fuente de información de estas investigaciones son las muestras de la Encuesta Nacional de Demografía y Salud de Bolivia de 1998 y 2003 que forman parte del programa de Encuestas de Demografía y Salud (DHS) que Macro International Inc. ejecuta en varios países en desarrollo, se trata de muestras probabilísticas a nivel nacional, estratificadas y bietápicas. En este caso se utiliza la muestra de 2003, que es la información más reciente disponible en línea.⁴

Los métodos de repetición en estadística, con ayuda de la computación, permiten un razonamiento inductivo y deductivo a la vez. Los algoritmos para la estimación de la tasa y para el método de remuestreo permiten, repetir y detallar los resultados, almacenar información generada, con lo cual se puede observar los cambios a nivel general y particular del proceso de remuestreo, de las TEF, los nacimientos y los tiempos de exposición por grupos quinquenales de edad.

Se ejecutan los algoritmos para la estimación clásica de la TGF y el algoritmo de remuestreo, considerando la complejidad de la muestra de Argote (2007: 93-94, 99), y se añade un módulo de código en *stata*⁵ para desagregar la información por nacimientos y tiempos de exposición por grupos quinquenales de edad, y para el cálculo de los sesgos, desviación estándar, coeficiente de variación, intervalos de confianza a 95 por ciento, tanto de la distribución por muestreo de la TGF, como de los nacimientos.

El grupo de distribuciones estadísticas de los nacimientos por grupos quinquenales de edad tiene sus propias características y variaciones, por lo cual se construyó un índice de normalidad y dispersión, que considera los estadísticos, *skewnes* (asimetría), *kurtosis* (apuntamiento) y el coeficiente de variación para determinar el grado de semejanza de las distribuciones a la distribución la normal y grado de homogeneidad de los datos.

La metodología aplicada para el análisis de sensibilidad es por objetivo, es decir, con base en las distribuciones estadísticas de la TGF y de

⁴ La información detallada de las muestras esta disponible en los documentos técnicos de la encuesta en la página de la Instituto Nacional de Estadística de Bolivia: www.ine.gob.bo.

⁵ *Software* para el análisis estadístico. Véase www.stata.com.

los nacimientos por grupos quinquenales de edad, se parte del objetivo de encontrar los nacimientos a determinado valor de la TGF. Se consideran variaciones de 0.5 por ciento uno y dos por ciento de la TGF por encima y por debajo. Mediante la búsqueda del valor objetivo de la TGF en su distribución por muestreo, se encuentran sus correspondientes siete valores de nacimientos por grupos quinquenales de edad en otra matriz desglosada. Luego se calcula el porcentaje de variación de los nacimientos, que bajo el supuesto de linealidad tendría que tener el mismo valor que la variación que se hizo de la TGF *a priori*, sin embargo, esto no sucede así (véase Cuadro 3).

RESULTADOS

A través del remuestreo de la ENDSA 2003, el cálculo de los nacimientos, tiempos de exposición, tasas específicas de fecundidad y la tasa global de fecundidad para cada iteración, se obtienen las distribuciones del numerador, denominador y los estadísticos de error de las TGF (error estándar, intervalos de confianza, sesgos). Este conjunto de datos conforman la fuente básica de nuestro análisis de sensibilidad de la TGF.

Debido a que la estimación de los nacimientos es de vital importancia en proyecciones de población (CEPAL, 2006: 15), centraremos nuestro análisis en el denominador de la tasa, sin tomar en cuenta en este caso la distribución de los tiempos de exposición o de las tasas específicas de fecundidad.

En las gráficas 1 y 2 se muestran las distribuciones del estimador de los nacimientos y los sesgos encontrados respecto del valor promedio para los siete grupos quinquenales de edad, de los cuales se promedia la TGF. Al ser un estimador de razón, se trata de un estimador cuya función depende de los nacimientos y tiempos de exposición acorde a la definición teórica de una tasa que mide la dinámica del comportamiento demográfico. Asimismo, la TGF es un indicador sintético que promedia las tasas específicas de fecundidad de los siete grupos quinquenales 15-20 (G1), 20-25 (G2), 25-30 (G3), 30-35 (G4), 35-40 (G5), 40-45 (G6) y 45-49 (G7), que de aquí en adelante se enumerarán de 1 a 7.⁶

Se puede observar que las distribuciones por muestreo de los siete grupos quinquenales de edad son diferentes, reflejando pequeños cambios que aportan al análisis de la precisión de esta investigación. Si bien, según el teorema central del límite, a un mayor tamaño muestra (en este caso se

⁶ Se considera este rango de edad ya que el periodo reproductivo de la mujer abarca de 15 a 49 años.

utilizaron mil iteraciones⁷) se espera que las distribuciones se aproximen a la distribución normal, podemos observar que este método brinda mayor información de lo que ocurre con los datos de los nacimientos por grupos quinquenales de edad que solamente adoptar el supuesto *a priori*. Es más, los nacimientos del grupo 45-49 es el que menos se parece a la normal.

En el Cuadro 1 se puede observar una comparación de los intervalos de confianza estimados bajo el supuesto de normalidad y bajo el método de remuestreo (*bootstrap*). Para tener idea de la variabilidad de los datos tenemos que la desviación típica (sd) bajo el supuesto de normalidad (4ta columna del Cuadro 1) es mayor en el grupo 20-25 años y menor el grupo 45-49, lo cual coincide con la estimación por remuestreo y refleja la coherencia del método experimentado. Por otro lado se observa que la desviación típica es mayor para todos los grupos mediante remuestreo que por la estimación clásica, lo cual se ve reflejado en la estimación de los intervalos de confianza. Los intervalos de confianza por el método de remuestreo se ven muy pegados en el límite inferior con las estimaciones clásicas y más separados por el límite superior.

Atendiendo a la forma de las distribuciones de los nacimientos con respecto a la normal, la Gráfica 3 nos muestra la asimetría, *kurtosis* y cv. Podemos observar que el grupo 30-35 es el que tiene mayor simetría a la normal al acercarse más a cero. Los grupos 20-25 y 25-30 también están muy cercados al anterior. Los grupos en inicio a la edad reproductiva y 35-40 se alejan más, mientras que el último grupo se aleja mucho de la simetría.

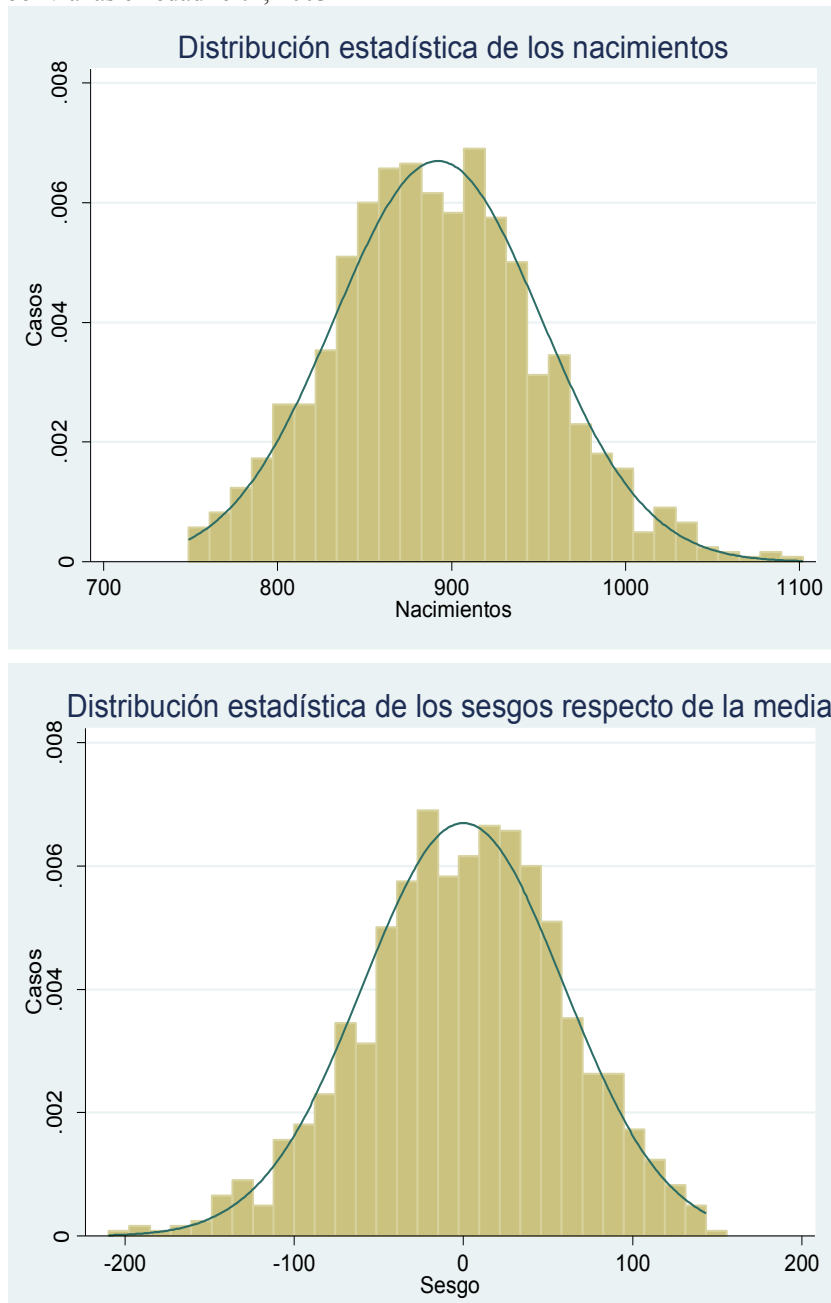
Los grupos de 15 a 35 años tienen una distribución tan apuntada como la normal, ya que la *kurtosis* está en el rango de 2.9 a 3.1, mientras que los grupos 35-40, 40-45 y 45-49 se alejan en forma ascendente de estos parámetros.

El Coeficiente de Variación (cv),⁸ al representar el porcentaje de variación con relación a la media, muestra que en los grupos de 15 a 45 años de edad varía entre seis y once por ciento, aproximadamente, mientras que en el último grupo se registra una variación de más de 30 por ciento, en las estimaciones por ambos métodos (véase Gráfica 3-C).

⁷ Número óptimo bajo el cual se obtienen buenas estimaciones (Efron, 1993).

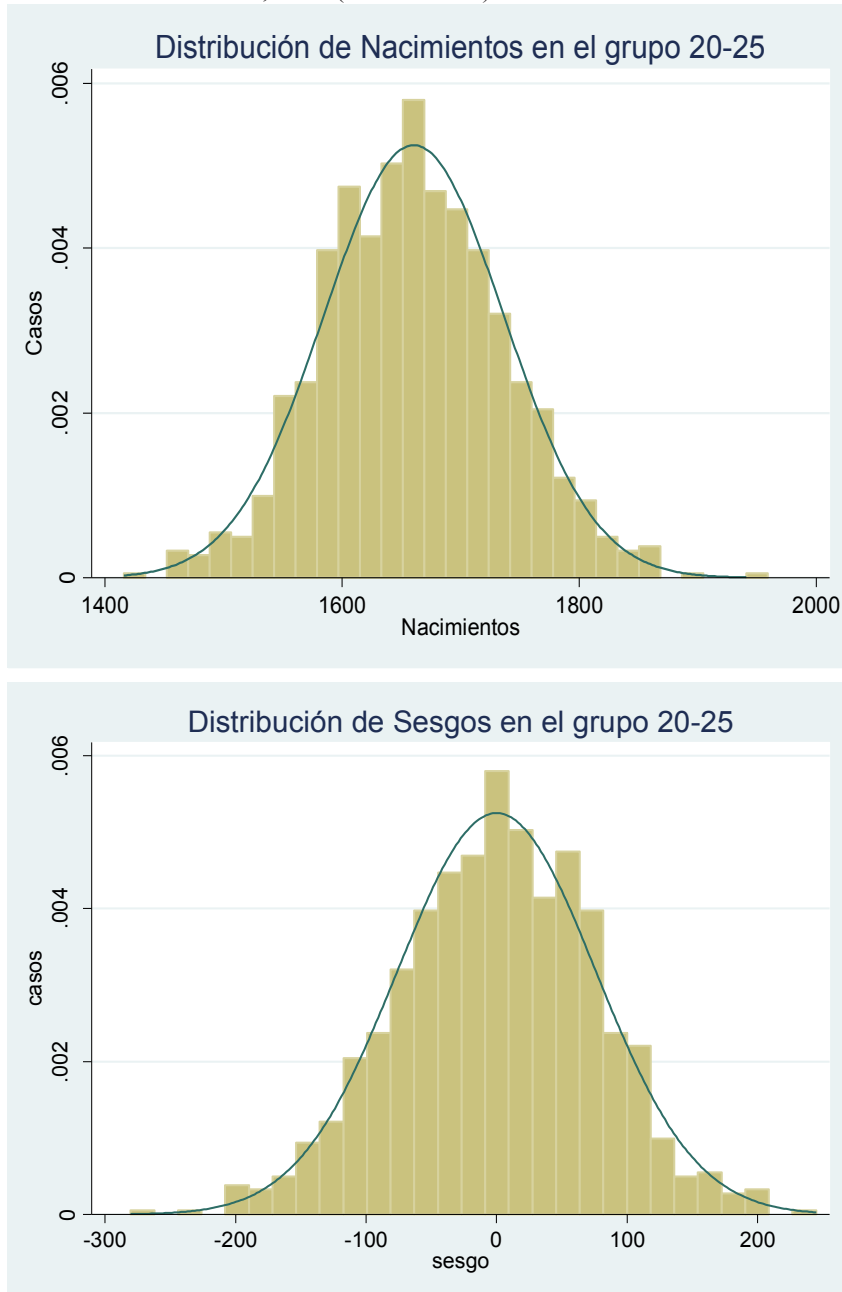
⁸ En estadística es el coeficiente de variación a distintas escalas, pero que están correlacionadas estadísticamente y sustantivamente con un factor en común. Es decir, ambas variables tienen una relación causal con ese factor. Su fórmula expresa la desviación estándar como porcentaje de la media aritmética, mostrando una mejor interpretación porcentual del grado de variabilidad que la desviación típica o estándar. Por otro lado, presenta problemas, ya que, a diferencia de la desviación típica, este coeficiente es variable ante cambios de origen. Por ello es importante que todos los valores sean positivos y su media dé, por tanto, un valor positivo. A mayor valor de cv mayor heterogeneidad de los valores de la variable, y a menor cv mayor homogeneidad en los valores de la variable.

Gráfica 1. Distribución por muestreo de los nacimientos y sesgos de las mujeres bolivianas en edad fértil, 2003



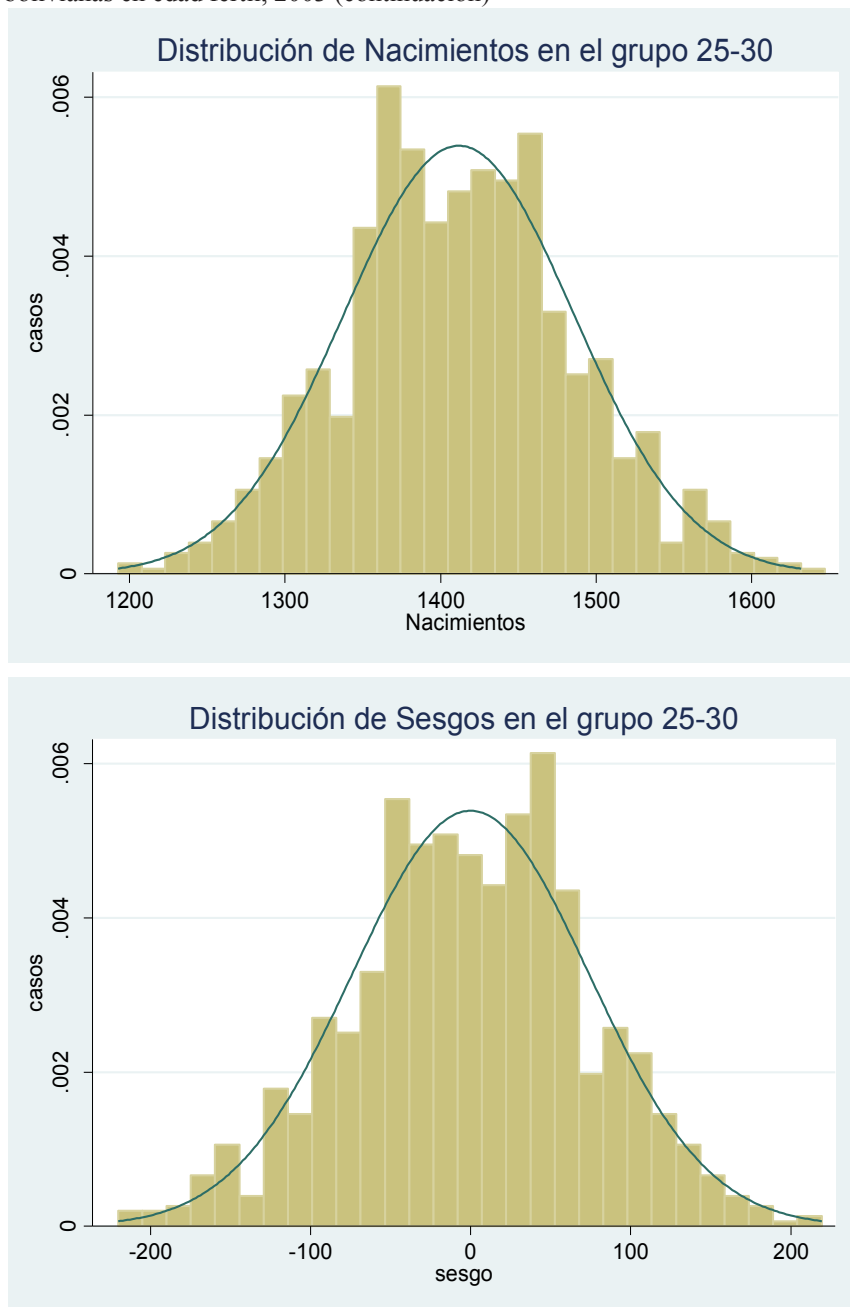
Fuente: estimaciones propias mediante remuestreo con base en la ENDSA 2003.

Gráfica 1. Distribución por muestreo de los nacimientos y sesgos de las mujeres bolivianas en edad fértil, 2003 (continuación)



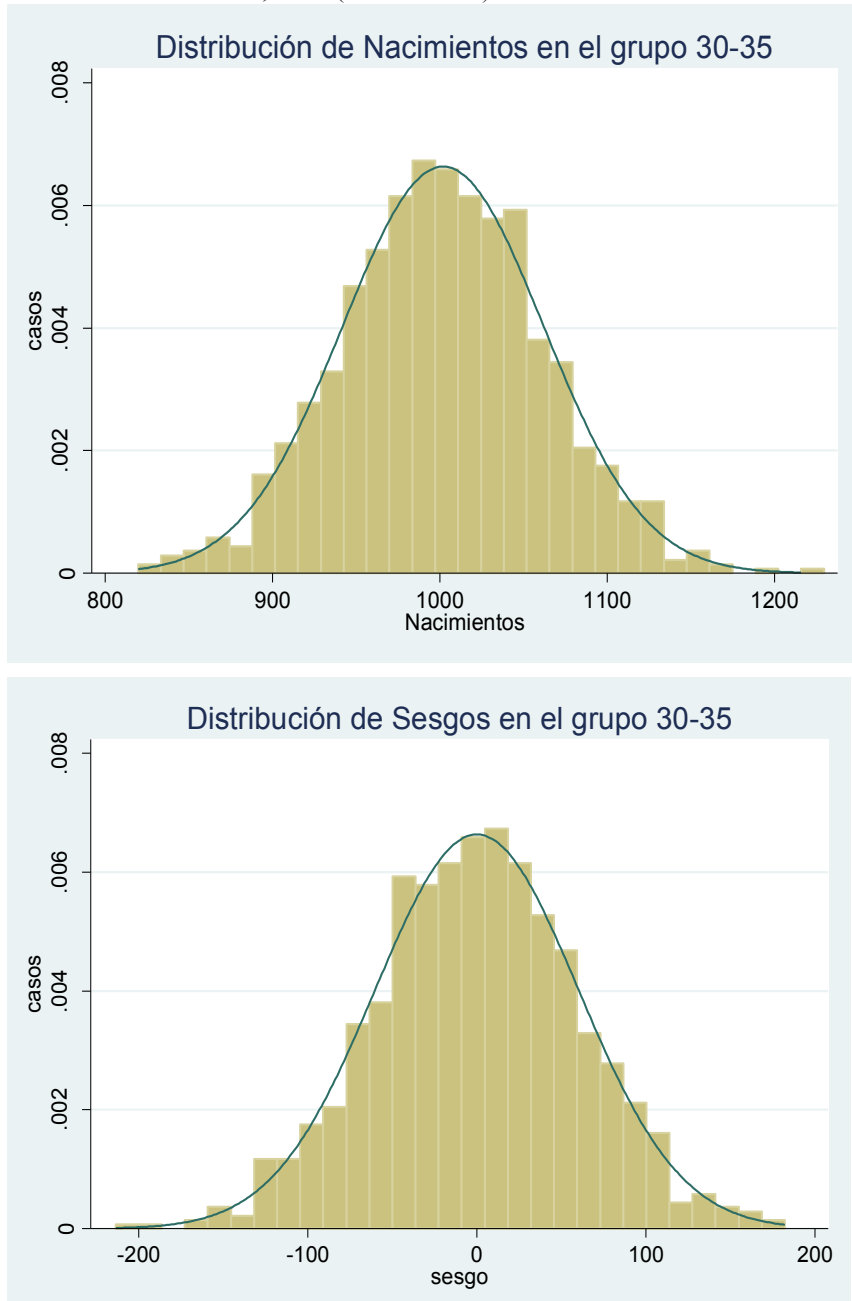
Fuente: estimaciones propias mediante remuestreo con base en la ENDA 2003.

Gráfica 1. Distribución por muestreo de los nacimientos y sesgos de las mujeres bolivianas en edad fértil, 2003 (continuación)



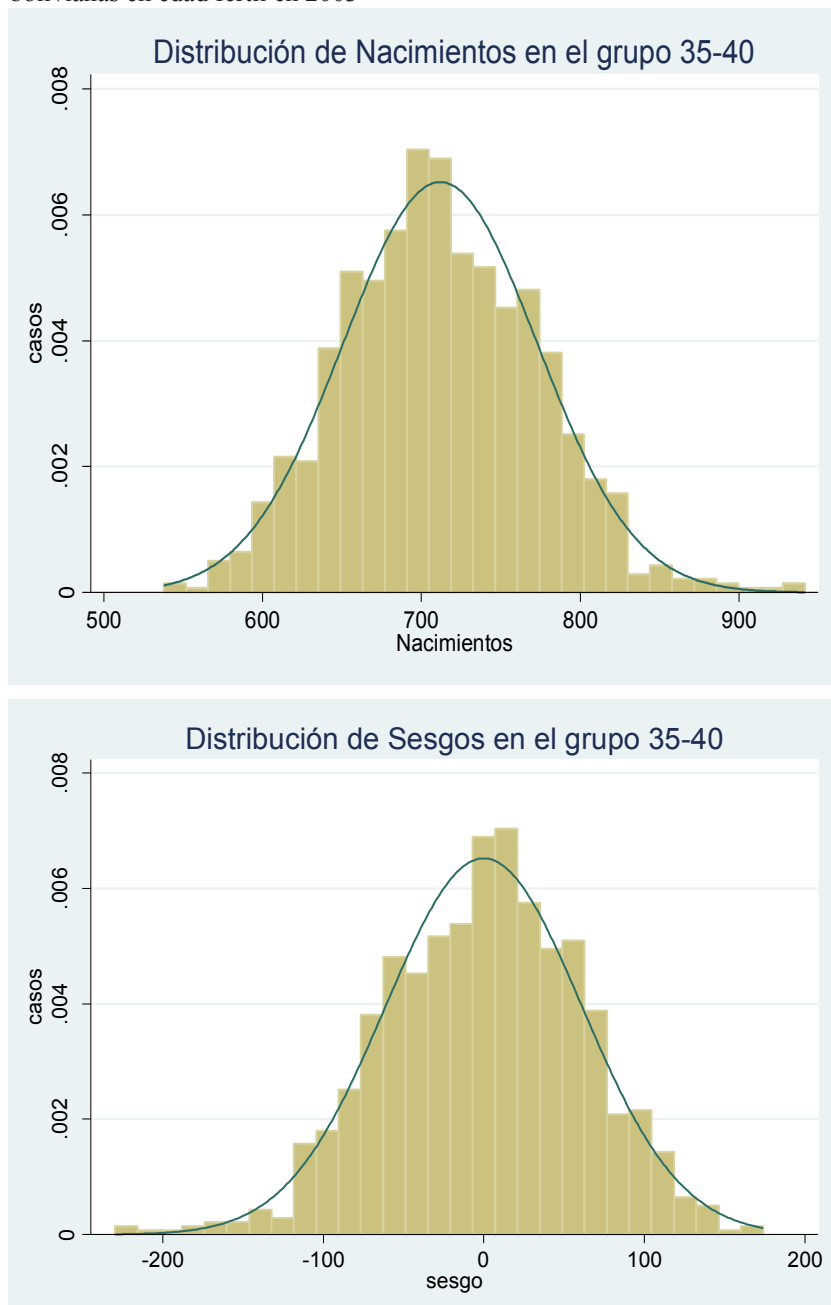
Fuente: estimaciones propias mediante remuestreo con base en la ENDA 2003.

Gráfica 1. Distribución por muestreo de los nacimientos y sesgos de las mujeres bolivianas en edad fértil, 2003 (continuación)



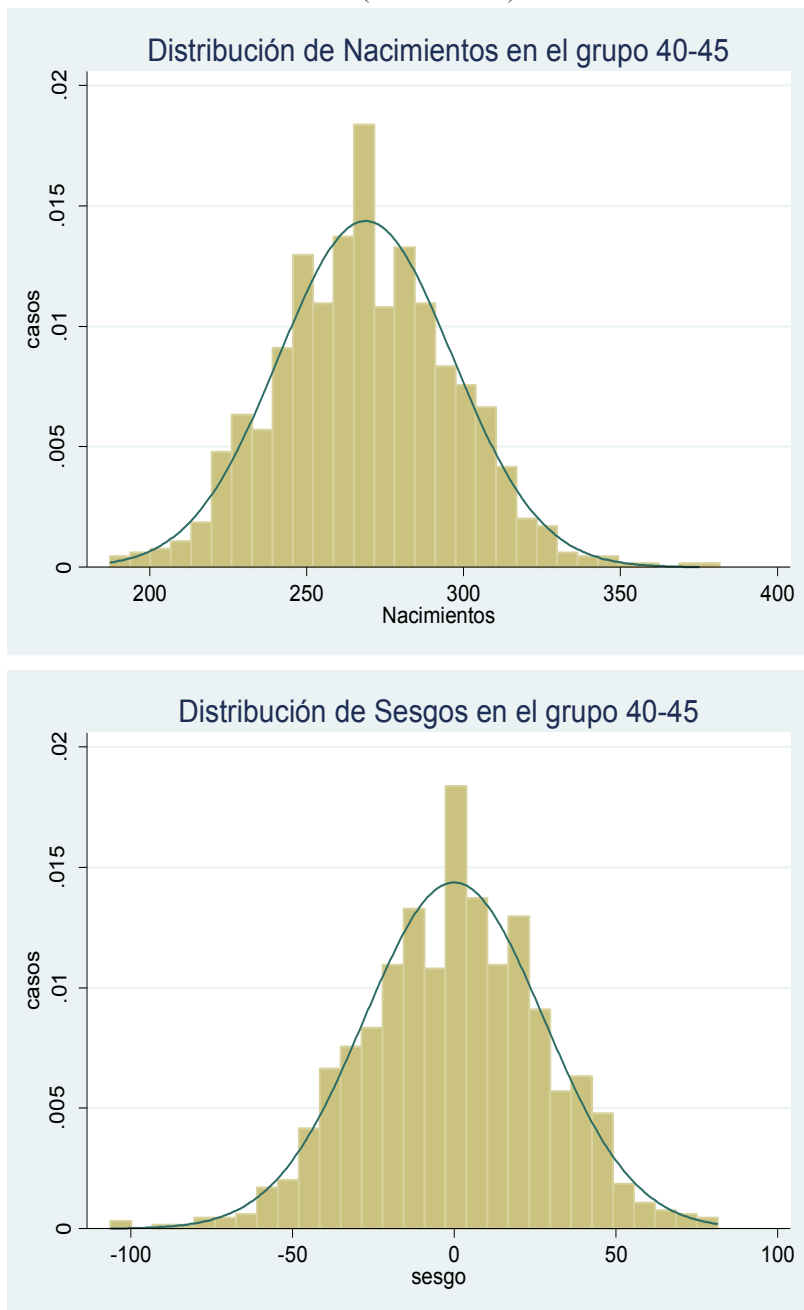
Fuente: estimaciones propias mediante remuestreo con base en la ENDSA 2003.

Gráfica 2. Distribución por muestreo de los nacimientos y sesgos de las mujeres bolivianas en edad fértil en 2003



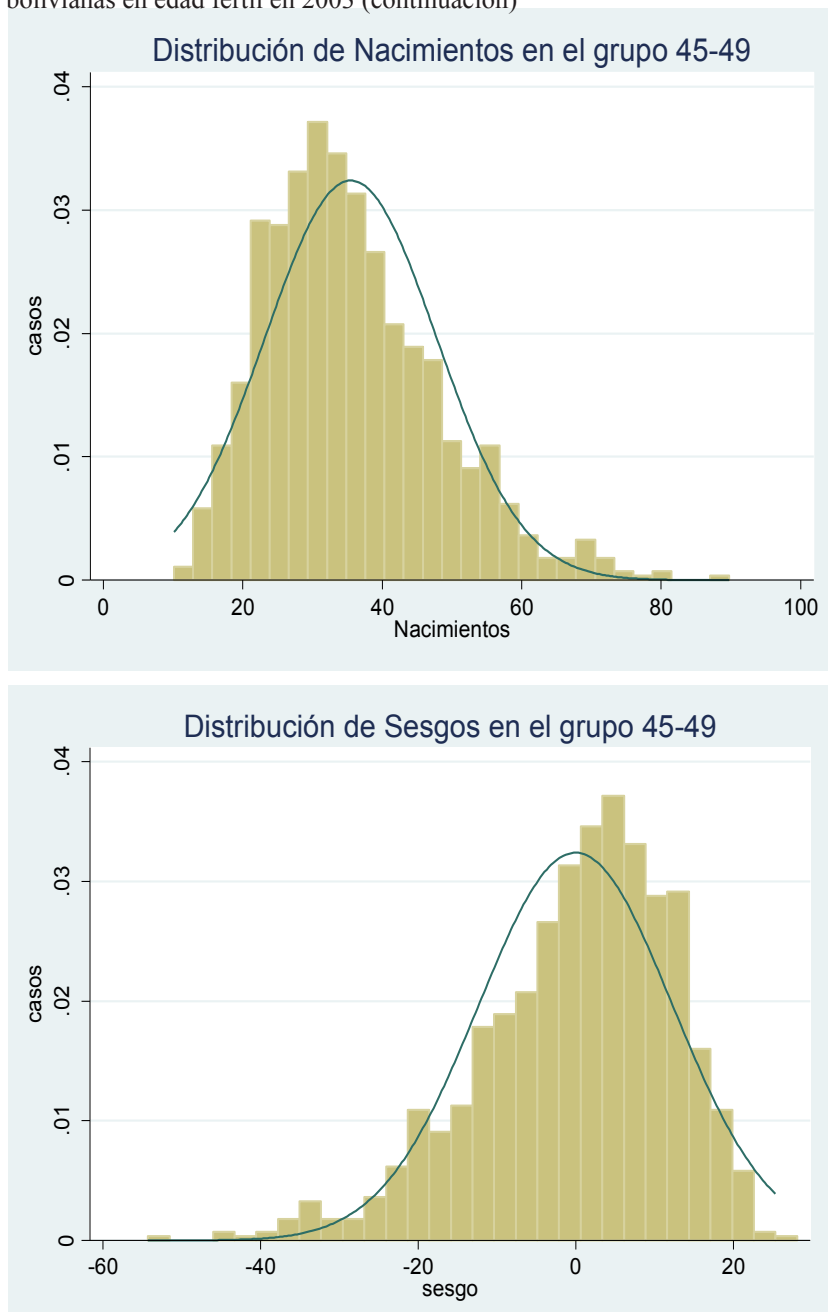
Fuente: estimaciones propias mediante remuestreo con base en la ENDSA 2003.

Gráfica 2. Distribución por muestreo de los nacimientos y sesgos de las mujeres bolivianas en edad fértil en 2003 (continuación)



Fuente: estimaciones propias mediante remuestreo con base en la ENDSA 2003.

Gráfica 2. Distribución por muestreo de los nacimientos y sesgos de las mujeres bolivianas en edad fértil en 2003 (continuación)



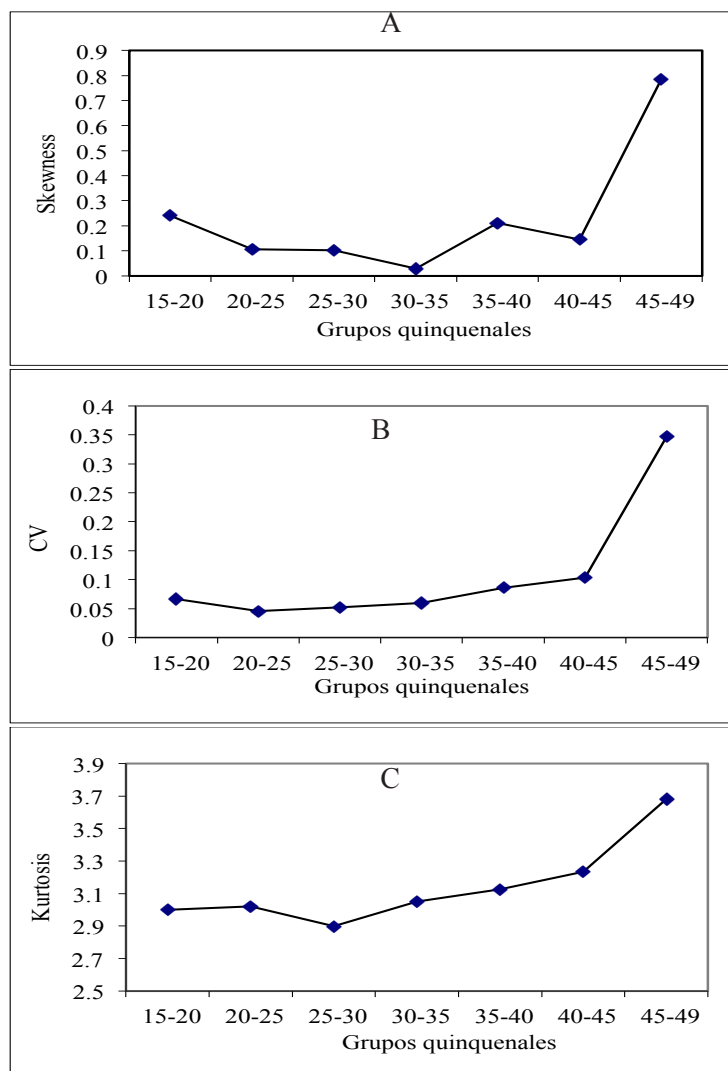
Fuente: estimaciones propias mediante remuestreo con base en la ENDSA 2003.

Cuadro 1. Datos estadísticos de las distribuciones por muestreo del estimador de los nacimientos por grupos quinquenales de edad (Mujeres bolivianas en 2003)

Grupo	Media	V(x)	Sd	Skewness	Kurtosis	CI-Normal		
						CV	Inf	Sup
15-20	892.125	3 550.119	59.583	0.241	2.999	0.067	888.427	895.822
20-25	1 660.581	5 773.570	75.984	0.106	3.020	0.046	1 655.866	1 665.296
25-30	1 411.770	5 480.324	74.029	0.102	2.899	0.052	1 407.176	1 416.363
30-35	1 001.848	3 616.432	60.137	0.029	3.051	0.060	998.116	1 005.580
35-40	711.881	3 743.953	61.188	0.212	3.127	0.086	708.084	715.678
40-45	268.792	770.590	27.760	0.146	3.235	0.103	267.070	270.515
45-49	35.481	151.637	12.314	0.783	3.681	0.347	34.717	36.245
Grupo	Media	V(x)	Sd	CV	CI-Bootstrap			
					CV	Inf	Sup	
15-20	892.125	3 550.119	66.511	0.075	0.075	889.752	996.440	
20-25	1 660.581	5 773.570	84.819	0.051	0.051	1 658.754	1 790.962	
25-30	1 411.770	5 480.324	83.703	0.059	0.059	1 410.283	1 536.476	
30-35	1 001.848	3 616.432	67.129	0.067	0.067	1 000.740	1 103.043	
35-40	711.881	3 743.953	68.302	0.096	0.096	709.146	814.114	
40-45	268.792	770.590	30.987	0.115	0.115	267.758	312.969	
45-49	35.481	151.637	13.746	0.387	0.387	33.603	57.862	

Fuente: cálculos propios.

Gráfica 3. Estadísticos de asimetría y normalidad de la distribución de los nacimientos por grupos quinquenales de edad



Fuente: cálculos propios con base a los estadísticos de los distribuciones.

Nota:

A, skewness = 0, tan alta como la distribución normal.

B, Kurtosis = 3, tan apuntada como la distribución normal.

C, Coeficiente de Variación = mayor valor, mayor heterogeneidad; menor valor, mayor homogeneidad.

Considerando los dos estadísticos de normalidad (*skewnes* y *kurtosis*) y el coeficiente de variación que dan cuenta de la dispersión de los datos junto a la desviación estándar, se construyó un índice de normalidad y dispersión de los datos; si es igual a tres o cercano significa que tiene una distribución muy semejante a la normal. En el Cuadro 2 se puede observar que la distribución del grupo 20-25 (índice = seis) es la que más se asemeja a la normal y es más homogénea; este grupo tiene el menor valor del índice construido. Le sigue el grupo 25-30 y el 30-35 que se encuentran en la misma tónica. En ese orden los grupos 15-20, 35-40 y 40-45 tienen el siguiente nivel promedio de semejanza a la normal y mediana heterogeneidad. Mientras que el último grupo 45-49 por sí solo es la distribución que no se parece a la normal y sus datos presentan la mayor heterogeneidad de todos.

ANÁLISIS DE SENSIBILIDAD

El análisis de sensibilidad nos permite encontrar variables críticas y el efecto que causan estas en las variables relacionadas y el modelo. Un análisis de sensibilidad es pertinente porque los datos de entrada poseen error, tenemos infinitas posibilidades de valores reales de la variable, un valor posible de un conjunto de posibilidades. El objetivo no es tanto encontrar el verdadero valor, sino evaluar los puntos críticos y encontrar la estimación más probable para la toma de decisiones.

A variaciones de medio, uno y dos por ciento menores que cero respecto a la media de la TGF (3.835), los valores de los nacimientos varían en diferente porcentaje, como se puede observar en la Gráfica 4.

Por ejemplo, a una variación de -0.5 por ciento de la TGF los nacimientos en el grupo uno tienen un cambio de 11.7 por ciento, en los grupos 2, 3, 4 y 5 menor a dos por ciento y cercano a cero, pero en los grupos seis y siete este porcentaje se dispara tomando una diferencia de 14 y 22 por ciento, respectivamente. Para las tres variaciones de las TGF se observa que las estimaciones se disparan en los dos últimos grupos quinquenales.

Analizando las variaciones por encima de la media de la TGF a 0.5, uno y dos por ciento, también se encuentra que las variaciones en porcentaje son superiores y diferentes para los grupos quinquenales con la coincidencia de que en los grupos 6 y 7 este porcentaje de variación se dispara mayor a 20 por ciento, al igual que en el anterior grupo de análisis.

Las Gráficas 4 y 5 nos permiten observar claramente la variabilidad de comportamiento de los nacimientos respecto a pequeños cambios en el valor del estimador de la TGF.

Cuadro 2. Índice de normalidad y dispersión de los datos respecto a la media

Grupo	Skewness	Rank(s)	Kurtosis	Dif de 3	Rank(k)	cv	Rank(cv)	Índice
15-20	0.2412965	6	2.999395	0.000605	1	0.066787621	4	11
20-25	0.1057632	3	3.020162	-0.020162	2	0.045757485	1	6
25-30	0.1021855	2	2.899343	0.100657	4	0.05243716	2	8
30-35	0.0293866	1	3.050596	-0.050596	3	0.060025852	3	7
35-40	0.2119847	5	3.12729	-0.12729	5	0.085952333	5	15
40-45	0.1459215	4	3.234972	-0.234972	6	0.103274946	6	16
45-49	0.7834932	7	3.680716	-0.680716	7	0.347063776	7	21

Nota: la metodología de construcción del índice es propia.
Considerando las características de los estadísticos de normalidad respecto de sus valores de referencia, se procedió a *rankear* los datos tomando en cuenta que el primer lugar se asigna al valor que más se acerca a la distribución normal. Así, los valores van de 1 a 7, donde 1 es más parecido a la normal y 7 más diferente. El índice que toma los tres estadísticos considera la sumatoria por lo cual un total de 3 sería el que más se parece a la normal y un total de 21, el que menos. Se aclara que *skewness* y *kurtosis* se relacionan a la forma de la normal y el Coeficiente de Variación (CV) a la heterogeneidad u homogeneidad de los datos, que es un estadístico adecuado al tratarse de un estimador de razón en el cual sus variables están relacionadas.

Cuadro 3. Análisis de sensibilidad de la TGF con respecto a los nacimientos por grupos quinquenales de edad

Variación de la TGF al:	TGF	15-20	%	20-25	%	25-30	%	30-35	%	35-40	%	40-45	%	45-49	%
-0.02	3.835	891	0.1	1 703	2.5	1 345	4.7	986	1.6	671	5.7	242	10.1	42	19.2
-0.01	3.835	876	1.8	1 775	6.9	1 271	10.0	938	6.4	704	1.1	283	5.3	42	20.5
-0.005	3.950	787	11.7	1 679	1.1	1 423	0.8	989	1.3	718	0.8	307	14.1	43	22.0
0	3.835	892	0.0	1 661	0.0	1 412	0.0	1 002	0.0	712	0.0	269	0.0	35	0.0
0.005	3.950	898	0.7	1 628	2.0	1 420	0.5	956	4.6	713	0.1	278	3.4	26	24.9
0.01	3.950	942	5.6	1 565	5.8	1 409	0.2	1 008	0.6	713	0.2	252	6.2	68	30.0
0.02	3.950	901	1.1	1 794	8.0	1 361	3.6	1 016	1.4	733	3.0	306	13.7	30	15.4

Fuente: cálculos propios.
Nota: la fila de color gris es el valor promedio de referencia respecto a la cual se analizan las variaciones.

No se identifica una secuencia o patrón en los valores de los nacimientos por grupos quinquenales de edad tanto para los valores por debajo y por encima de la media de la TGF. Sí, se identifica que el porcentaje de variación en el caso del grupo de 45 y 49, en la terminación de la edad reproductiva de las mujeres bolivianas, se amplifica respecto a variaciones de 0.5, uno y dos por ciento de la media de la TGF.

DISCUSIÓN

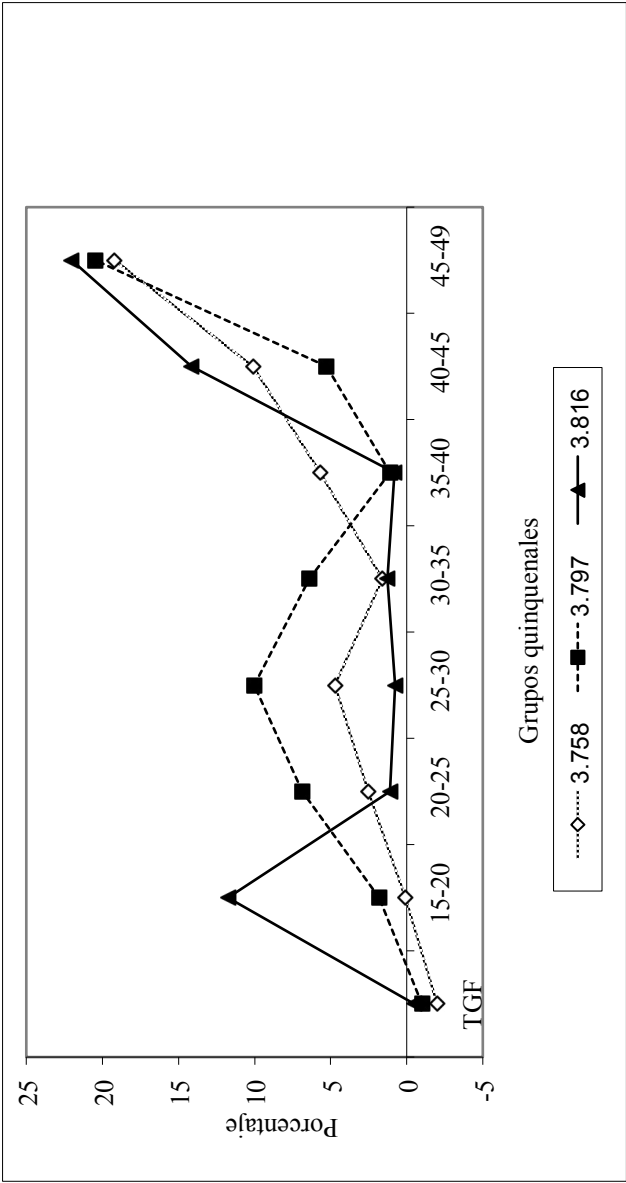
Si bien los métodos de estimación tradicionales acerca de la población proyectada tienen una importante base teórica, la estadística no paramétrica a través de modelos de simulación y remuestreo permite experimentar con la incertidumbre detrás de nuestras estimaciones, es decir, evaluar y conocer el error de nuestras estimaciones; de esta manera, al reducir la incertidumbre, podemos mejorar nuestras estimaciones de la población en determinado momento. A través de estos métodos también podemos evaluar el impacto de diferentes grados de error de manera experimental y conocer si se reduce o se amplifica.

El análisis de sensibilidad nos ha dado interesantes resultados que no hubiésemos podido observar con los métodos tradicionales. En primer lugar se cuenta con información base de las distribuciones por muestreo de la TGF, así como de los nacimientos de los siete grupos quinquenales de edad, es decir, se ha generado mayor información acerca de las distribuciones de los nacimientos de las mujeres bolivianas en 2003. Por otro lado, se ha podido construir un índice para evaluar la normalidad, así como la heterogeneidad y dispersión de la distribución del estimador de los nacimientos por grupos quinquenales de edad.

Finalmente, con lo anterior se ha dado pie a que sea factible realizar un análisis de sensibilidad de los nacimientos a pequeñas variaciones de la TGF (0.5, uno y dos por ciento).

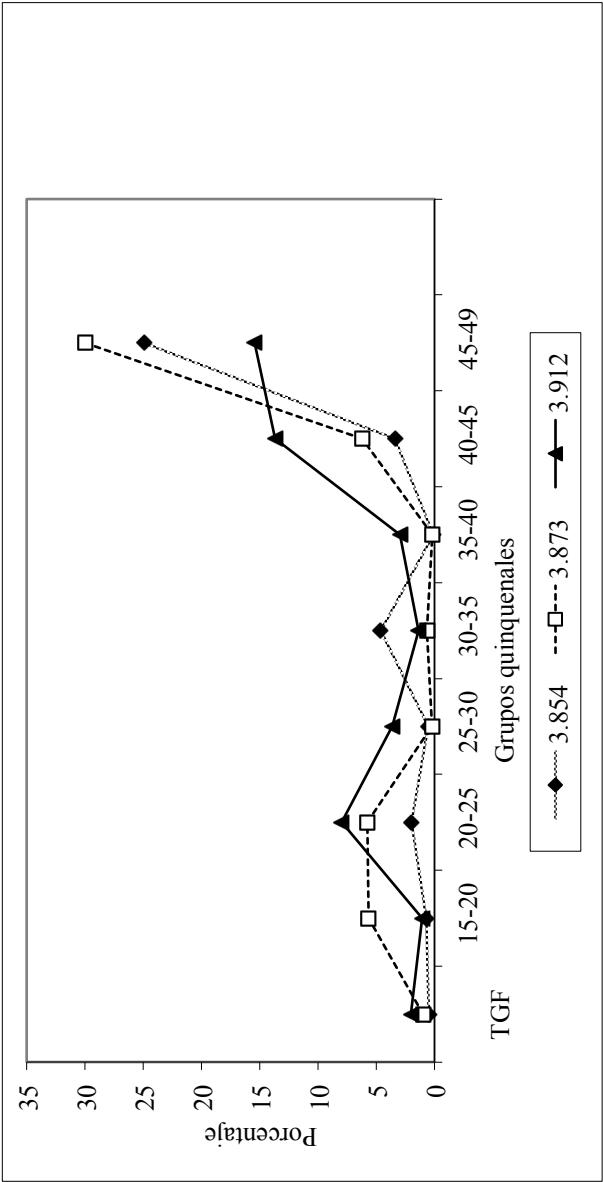
Con todo ello, no cabe duda que a mayor información menor incertidumbre podemos esperar. Acorde a nuestras hipótesis de investigación, por un lado, la fecundidad en Bolivia se comporta de manera heterogénea para los diferentes grupos quinquenales de edad. Si bien en estudios anteriores Argote (2007 y 2009) analizó la normalidad de la distribución de la TGF, en esta investigación se verifica que cada grupo quinquenal tiene un comportamiento diferente. En cuanto a normalidad y heterogeneidad, los grupos 2, 3 y 4 se parecen más a la normal y son más homogéneos, le siguen los grupos 1, 5 y 6, pero el último grupo no se asemeja a la normal.

Gráfica 4. Porcentaje de variación respecto a la media de la TGF y los nacimientos por grupos quinquenales de edad (valores inferiores)



Fuente: elaboración propia.

Gráfica 5. Porcentaje de variación respecto a la media de la TGF y los nacimientos por grupos quinquenales de edad (valores superiores)



Fuente: elaboración propia.

Además, este análisis nos permite encontrar los puntos críticos del modelo que pueden afectar futuras estimaciones. En este caso se ha encontrado que los grupos de 40-45 y 45-49 son los más sensibles, es decir, a pequeños cambios de la TGF los nacimientos de la mujeres al terminar su periodo reproductivo pueden dar sorpresas y presentarse una mayor fecundidad, esto probablemente debido al retraso en la edad a tener hijos como ocurre en otros países (Ortega y Kohler, 2000) y que la probabilidad de fecundar a estas edades no sea muy baja, como se esperaría, en la población de mujeres bolivianas en 2003.

Por otro lado, según las Gráficas 4 y 5 producto del análisis de sensibilidad de la TGF, se observa que se presenta el efecto mariposa, ya que a pequeños cambios de la TGF, los cambios en los nacimientos son diferentes y superiores, y se amplifican en los dos últimos grupos. Por lo tanto la fecundidad es un fenómeno no lineal, heterogéneo y se comporta de forma compleja, por lo cual es muy probable que forme parte de un sistema complejo que es el poblacional.

Por último, resulta interesante aplicar estos resultados al análisis de sensibilidad de una proyección de población. A pequeños cambios de la TGF, ¿cómo cambian los resultados de proyección de población? Respondiéndonos a esta pregunta, pretendemos dar un paso más en el conocimiento del modelado bajo incertidumbre del sistema poblacional.

BIBLIOGRAFÍA

ALHO, Juha, Maarten ALDERS, Harri CRUIJSEN, Nico KEILMAN, Timo NIKANDER y Pham DINH QUANG, 2006, "New forecast: population decline postponed in Europe", en *Statistical Journal of the United Nations*, vol. 23, núm. 1.

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

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*, año 13, núm. 62, octubre-diciembre, Toluca, México.

AXELROD, Robert, 1997, "Advancing the art of simulation in the social sciences", en Rosario CONTE, Rainer HEGSELMANN y Pietro TERNA (eds.), *Simulating social phenomena*, School of Public Policy/Universidad de Michigan, Ann Arbor.

BARRETT, Chris, Stephen EUBANK y Madhav MARATHE, 2006, "Modeling and simulation of large biological, information and socio-technical systems: an

interaction based approach”, en GOLDIN, SMOLKA y WEGNER (eds.), *Interactive computation: the new paradigm*, Virginia Polytechnic Institute/State University, Blacksburg, Virginia.

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

CASWELL, Hall, 2008, “Perturbation analysis of nonlinear matrix population models”, en *Demographic Research*, vol. 18.

CEPAL, 2006, *Chile: estimaciones y proyecciones de la población. 1950-2050*, serie OI, núm. 208, Instituto Nacional de Estadísticas/Centro Latinoamericano y Caribeño de Demografía, Santiago de Chile.

CHAMBERS, Raymond y Ajan DORFMAN, 1994, *Robust simple survey inference via bootstrapping and bias correction: the case of ratio estimator*, Southampton Statistical Sciences Research Institute, Southampton.

DIAS, Tiago, Nuno ROMA y Leonel SOUSA, s/f, *Two-level scalable motion estimation architecture with fractional-pixel accuracy and efficient data re-usage*, Instituto Superior Técnico, Lisboa, Portugal.

EFRON, Bradley y Robert TIBSHIRANI, 1993, *An introduction to the bootstrap*, Chapman y Hall, Nueva York.

ESCOBEDO RIVERA, José, 2007, “El dato en la investigación demográfica: una visión epistemológica”, en *Papeles de Población*, año 13, núm. 54, octubre-diciembre, Toluca, México.

GARCÍA OCHOA, Liliana, Ana Lilia LAUREANO CRUCES y Raymundo PERALTA HERRERA, 2006, *Economía simulada por un sistema multiagente*, XIX Congreso Nacional y V Congreso Internacional de Informática y Computación de la ANIEI, 25-27 de octubre, Tuxtla Gutiérrez.

GOTTLIEB, Alex, 2001, *Asymptotic accuracy of the jackknife variance estimator for certain smooth statistics*, Internet, recuperado de: www.alexgottlieb.com/papers/lastjack.pdf.

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

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

ROJÍ FERRARI, Salvador, 2005, *La complejidad: un nuevo enfoque de la Economía Financiera*, Universidad Complutense de Madrid, Madrid.

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

MIRAMONTES, Octavio, 1999, "Los sistemas complejos como instrumentos de conocimiento y transformación del mundo", en Santiago RAMÍREZ *et al.*, *Perspectivas en las teorías de sistemas*.

MARJORAM, Paul y Simón TAVARÉ, 2006, "Modern computational approaches for analysing molecular genetic variation data. Focus on statistical analysis", en *Nature Reviews Genetics*, Nature publishing Group, vol. 7, october.

SAURA BAICOA, Dulce y Ángel RODRÍGUEZ GARCÍA-BRAZALES, 2003, "No linealidad y economía austriaca", en *Revista Economía del Rosario*, vol. 5, núm. 2, Bogotá, Colombia.

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. XLV, núm. 1.

ORTEGA OSONA, José Antonio y Hans-Peter KOHLER, 2001, "¿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.

GRIFFITHS, Paula, Zoe MATTHEWS y Andrew HINDE, 2000, "Understanding the sex ratio in India: a simulation approach", en *Demography*, vol. 37, núm. 4, november.

GUT, Allan, 2005, *Probability: a graduate course*, Springer, Nueva York.

TEJADA DÍAZ, Juan Carlos Rafael, 2003, *Alcances y limitaciones de la teoría del caos aplicada al análisis del comportamiento organizacional, cultura y la necesidad del cambio con la finalidad de afrontar la turbulencia del entorno de las organizaciones*, Universidad Peruana de Ciencias Aplicadas, Lima, Perú.

Milenka Linneth Argote Cusi

Maestra en Estudios de Población por Flacso-México. Es investigadora en la institución Corpotalentos ONG. Líneas de investigación: demografía estadística, monitoreo y evaluación de indicadores, minería de datos, modelaje y simulación. Publicaciones recientes: 2009, *Análisis macroeconómico del Sida en tres países latinoamericanos: México, Brasil y Argentina 1995-2005*, en The Latin American Studies Association (LASA), del 11 al 14 de junio en Río de Janeiro, Brasil; 2008, *Análisis comparativo de la distribución estadística de la tasa global de fecundidad de Bolivia en 1998 y 2003*, en IX Reunión Nacional de Demografía, Mérida-Yucatán, 8 al 11 de octubre; 2007, "Estimación de la distribución estadística de la tasa global de fecundidad", en *Papeles de Población*, año 13, núm. 54, octubre-diciembre, Toluca.

Dirección electrónica: milenkalinneth@yahoo.com.mx

Este artículo fue recibido el 17 de agosto de 2011 y aprobado el 16 de mayo de 2012.